25-26

GRADO EN PSICOLOGÍA
PRIMER CURSO

GUÍA DE ESTUDIO PÚBLICA

INTRODUCCIÓN AL ANÁLISIS DE DATOS

CÓDIGO 62011037

UNED

25-26

INTRODUCCIÓN AL ANÁLISIS DE DATOS

CÓDIGO 62011037

ÍNDICE

PRESENTACIÓN Y CONTEXTUALIZACIÓN
REQUISITOS Y/O RECOMENDACIONES PARA CURSAR LA ASIGNATURA
EQUIPO DOCENTE
HORARIO DE ATENCIÓN AL ESTUDIANTE
TUTORIZACIÓN EN CENTROS ASOCIADOS
COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE
RESULTADOS DE APRENDIZAJE
CONTENIDOS
METODOLOGÍA
SISTEMA DE EVALUACIÓN
BIBLIOGRAFÍA BÁSICA
BIBLIOGRAFÍA COMPLEMENTARIA
RECURSOS DE APOYO Y WEBGRAFÍA
IGUALDAD DE GÉNERO

NOMBRE DE LA ASIGNATURA	INTRODUCCIÓN AL ANÁLISIS DE DATOS
CÓDIGO	62011037
CURSO ACADÉMICO	2025/2026
DEPARTAMENTO	METODOLOGÍA DE LAS CIENCIAS DEL COMPORTAMIENTO
TÍTULO EN QUE SE IMPARTE CURSO - PERIODO - TIPO	GRADO EN PSICOLOGÍA - PRIMER - SEMESTRE 1 - FORMACIÓN BÁSICA
TÍTULO EN QUE SE IMPARTE	PRUEBA DE APTITUD PARA HOMOLOGACIÓN DE GRADO EN PSICOLOGÍA (COMPLEMENTO)
Nº ETCS	6
HORAS	150.0
IDIOMAS EN QUE SE IMPARTE	CASTELLANO

PRESENTACIÓN Y CONTEXTUALIZACIÓN

La aplicación de la Estadística a las distintas disciplinas recibe con frecuencia la denominación de "Análisis de datos". La asignatura "Introducción al análisis de datos" constituye, según su nombre indica, una primera aproximación a la descripción e inferencia de datos en el ámbito psicológico.

El Análisis de datos ha ido adquiriendo cada vez más importancia dentro de las distintas áreas científicas y, muy especialmente, dentro de las denominadas Ciencias Sociales y de la Salud. El objeto de esta asignatura es introducir y presentar las ideas y los conceptos fundamentales del análisis de datos, con ejemplo concretos, tanto en la investigación como en la psicología aplicada. Tiene, por tanto, un carácter instrumental, y proporciona las herramientas estadísticas básicas para ayudar a los futuros psicólogos a analizar y entender los resultados obtenidos tanto en la investigación como en la práctica psicológica, así como generar sus propios análisis de datos en aquellos campos que resulten de su interés. Se pretende que el alumnado alcance un entendimiento correcto de la información expresada de manera cuantitativa, un conocimiento adecuado de los términos, conceptos y análisis estadísticos y, como consecuencia de ello, la capacidad de evaluar los resultados obtenidos tanto en la investigación como en la práctica diaria del profesional.

Esta asignatura pertenece a la materia "Métodos, diseños y técnicas de investigación en Psicología". Se trata de una "herramienta" de carácter metodológico que enlaza directamente con las asignaturas de "Fundamentos de Investigación" (también de primer curso) y "Diseños de investigación y Análisis de datos" (de segundo curso) y encuentra su aplicación en prácticamente todas las asignaturas de la titulación del Grado en Psicología.

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR LA ASIGNATURA

Al alumno no se le presupone conocimientos especiales de matemáticas salvo conocer los fundamentos de análisis de datos recogidos en el Curso de Acceso a la Universidad y/o en la Enseñanza Secundaria en sus distintos planes de estudio.

EQUIPO DOCENTE

Nombre y Apellidos	MARIA DEL PILAR POZO CABANILLAS (Coordinador de asignatura)
Correo Electrónico	ppozo@psi.uned.es
Teléfono	91398-6684
Facultad	FACULTAD DE PSICOLOGÍA
Departamento	METODOLOGÍA DE LAS CIENCIAS DEL COMPORTAMIENTO
Nombre y Apellidos	BELEN FERNANDEZ CASTILLA
Correo Electrónico	bfcastilla@psi.uned.es
Teléfono	91398-7932
Facultad	FACULTAD DE PSICOLOGÍA
Departamento	METODOLOGÍA DE LAS CIENCIAS DEL COMPORTAMIENTO
Nombre y Apellidos	JOSE ANGEL MARTINEZ HUERTAS
Correo Electrónico	jamartinez@psi.uned.es
Teléfono	91398-7669
Facultad	FACULTAD DE PSICOLOGÍA
Departamento	METODOLOGÍA DE LAS CIENCIAS DEL COMPORTAMIENTO
Nombre y Apellidos	CELIA SERRANO MONTILLA
Correo Electrónico	cserrano@psi.uned.es
Teléfono	91398-9547
Facultad	FACULTAD DE PSICOLOGÍA
Departamento	METODOLOGÍA DE LAS CIENCIAS DEL COMPORTAMIENTO
Nombre y Apellidos	DANIEL VICENTE MORILLO CUADRADO
Correo Electrónico	dvmorillo@psi.uned.es
Teléfono	913986584
Facultad	FACULTAD DE PSICOLOGÍA
Departamento	METODOLOGÍA DE LAS CIENCIAS DEL COMPORTAMIENTO

HORARIO DE ATENCIÓN AL ESTUDIANTE

Los alumnos estarán atendidos presencialmente por el Tutor de su Centro Asociado quien les orientará y resolverá sus dudas sobre la asignatura. También deben utilizarse los grupos de Tutoría disponibles en el Curso Virtual de la asignatura.
Por otra parte, los alumnos dispondrán en el Curso Virtual de Foros dedicados a cada uno de los bloques temáticos. En estos Foros, los estudiantes podrán formular sus dudas, que serán atendidas por el equipo docente. Los Foros constituyen la principal herramienta de comunicación entre los profesores y los estudiantes.

Otras vías de comunicación son el correo postal (C/ Juan del Rosal, 10, 28040 Madrid) y la atención telefónica en el siguiente horario:

Dra. Pilar Pozo Cabanillas
Martes y Miércoles de 10:30 a 14:30 horas.
Teléfono: 91398 6684

Dra. Belén Fernández Castilla
Martes y Miércoles de 10:00 a 14:00 horas.
Teléfono: 91 398 7932

Dr. José Ángel Martínez Huertas
Martes y Miércoles de 10:00 a 14:00 horas.
Teléfono: 91398 7999

Dra. Celia Serrano Montilla
Martes y Miércoles de 10:00 a 14:00 horas.
Teléfono: 91398 9547

Dr. Daniel Morillo Cuadrado
Martes y Jueves de 10:00 a 14:00 horas.
Teléfono: 91398 6584

TUTORIZACIÓN EN CENTROS ASOCIADOS

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

Competencias Generales

•CG1.1.1 - Iniciativa y motivación
•CG1.1.2 - Planificación y organización
•CG1.1.3 - Manejo adecuado del tiempo
•CG1.2.4 - Pensamiento creativo
•CG2.1.4 - Comunicación y expresión matemática, científica y tecnológica
•CG2.2.1 - Competencia en el uso de las TIC

Competencias Específicas

•CE 1.4. - Ser capaz de identificar, discriminar y utilizar de forma pertinente al ámbito de la ciencia psicológica, los distintos diseños de investigación, procedimientos de formulación, contrastación de hipótesis e interpretación de resultados
•CE 2.4. - Desarrollar el interés tanto por la investigación teórica como por la aplicada.

- CE 2.5. - Ser capaz de aplicar procedimientos para organizar y describir los datos recopilados en una investigación en el ámbito de las ciencias sociales y de la salud.
- CE 2.6. - Saber razonar científicamente y poder poner en relación la teoría con la evidencia
- CE 3.1.2. - Ser capaz de describir y medir variables (personalidad, inteligencia y otras aptitudes, actitudes, etc.) y procesos cognitivos, emocionales, psicobiológicos y conductuales.
- CE 3.1.7. - Ser capaz de organizar la información obtenida: integrar resultados dentro de una formulación teórica y formular conclusiones
- CE 3.1.9. - Ser capaz de interpretar los resultados obtenidos en investigaciones previas y poner en relación unos resultados con otros.

RESULTADOS DE APRENDIZAJE

Al finalizar la asignatura, el estudiante debe:

- Ser capaz de recopilar, organizar, presentar e interpretar datos numéricos.
- Distinguir el nivel de medida con el que se han obtenido unos datos, como requisito imprescindible, para seleccionar adecuadamente los correspondientes análisis gráficos y los estadísticos o índices a calcular.
- Manejar con soltura los índices estadísticos correspondientes con el fin de resumir los datos e interpretar correctamente los resultados obtenidos.
- Poder identificar patrones de covariación y relación lineal entre variables, interpretar su relación y efectuar predicciones.
- Saber desenvolverse en situaciones de incertidumbre estadística, aplicando los conceptos básicos y los modelos de probabilidad más habituales al campo de la Psicología.
- Conocer las distintas formas de obtener una muestra y conocer el procedimiento básico para realizar estimaciones por intervalos de los parámetros de la población a partir de los estadísticos obtenidos en una muestra.

CONTENIDOS

Tema 1. Conceptos básicos y organización de datos

Tema 2. Índices de tendencia central y de posición

Tema 3. Medidas de variabilidad y de forma

Tema 4. Relación entre variables I

Tema 5. Relación entre variables II

Tema 6. Nociones básicas de probabilidad

Tema 7. Variables aleatorias y modelos discretos de probabilidad

Tema 8. Modelos continuos de probabilidad

Tema 9. Muestreo y distribución muestral de un estadístico

Tema 10. Estimación de parámetros y cálculo del tamaño muestral

METODOLOGÍA

Esta asignatura se imparte en la modalidad de enseñanza a distancia. Además de las unidades didácticas de la asignatura, los alumnos podrán consultar los materiales y realizar las actividades de aprendizaje recogidos en el curso virtual (material multimedia, ejercicios resueltos, ejercicios de autoevaluación, participación en los foros,....).
En los foros, los alumnos podrán formular sus dudas sobre el contenido de la asignatura que serán atendidas por el equipo docente. Por lo tanto, los foros constituyen la principal herramienta de comunicación entre el equipo docente y los alumnos.

SISTEMA DE EVALUACIÓN

TIPO DE PRUEBA PRESENCIAL

Tipo de examen	Examen tipo test
Preguntas test	25
Duración del examen	120 (minutos)
Material permitido en el examen	

Libro de Formulario y Tablas estadísticas de Introducción al Análisis de Datos en Psicología y Ciencias de la Salud (ver bibliografía básica). Este formulario, editado en papel (no sirven fotocopias), debe utilizarse para la realización del examen pero SIN ANOTACIONES por parte del alumno. Por este motivo es conveniente adquirirlo con suficiente antelación.

Calculadora no programable.

Criterios de evaluación

La calificación en el examen, en una escala de 0 a 10, se obtendrá aplicando la fórmula:

Calificación Examen = (0,4 x Aciertos) - (0,2 x Errores)

Para superar la asignatura es necesario obtener una puntuación igual o superior a 5 en el examen.

Recordamos que es responsabilidad del estudiante acceder con regularidad al curso virtual de la asignatura. Entre otros contenidos, el estudiante encontrará

Una *actividad voluntaria* relacionada con la herramienta gratuita de análisis de datos Jamovi (valoración máxima de 0,5 puntos).

Investigaciones que puede publicar el equipo docente, en las que el estudiante puede participar voluntariamente, recibiendo algún incentivo en la calificación.

Estas puntuaciones adicionales se sumarán a la nota final, siempre y cuando se obtenga una calificación de 5 o más en el examen. La calificación final en la asignatura no podrá ser superior al 10.

Una vez celebradas las pruebas presenciales, los exámenes resueltos podrán consultarse en el curso virtual de la asignatura.

% del examen sobre la nota final	100
Nota del examen para aprobar sin PEC	0
Nota máxima que aporta el examen a la calificación final sin PEC	0
Nota mínima en el examen para sumar la PEC	0

Comentarios y observaciones

La prueba presencial (examen) se celebrará en los Centros Asociados, según el calendario previsto. Versará sobre cuestiones tanto teóricas como prácticas y todos los temas del programa estarán representados en el examen con al menos una pregunta. El examen constará de 25 preguntas o ejercicios, con tres alternativas de respuesta, de las que sólo una es correcta.

Además, al final del examen se incluirá una "pregunta de reserva", identificada como tal. Esta pregunta (nº 26) será tenida en cuenta sólo en el caso de que el equipo docente considere necesaria la anulación de alguna de las 25 preguntas principales. Es decir, esta pregunta podrá sustituir, en caso de anulación, a cualquier pregunta del examen. El alumno no podrá solicitar que esa pregunta se compute para su nota en ninguna otra situación.

Este formato se utilizará en todos los modelos de examen (sean originales o de reserva) y para todos los Centros. El alumno deberá contestar en la hoja de respuestas facilitada al inicio del examen. Para la corrección del examen, únicamente se considerarán las respuestas dadas en la hoja de respuestas.

En caso de que la situación sanitaria no permita que los exámenes se realicen de forma presencial, y se tengan que hacer de forma online a través de la herramienta AvEx, el formato de examen cambiará a 20 preguntas tipo test (de tres alternativas) y el tiempo máximo para su realización será de 90 minutos. Este formato será aplicable a todos los estudiantes, incluidos los de Centros Penitenciarios, Centros en el Extranjero y Unidis.

Información sobre las Calificaciones y Revisión de Exámenes

El alumno puede conocer su calificación a través de la secretaría virtual a la que tiene acceso desde la web de la UNED. En el caso que no aparezca su calificación deberá contactar con el equipo docente a través del correo electrónico.

Si desea solicitar una Revisión de su calificación deberá hacerlo mediante la Aplicación informática de revisión de exámenes de grado, que se encuentra en el espacio "Mis calificaciones" de la Secretaría Virtual. El plazo máximo para solicitarla será de *7 días naturales contados a partir del día siguiente a la publicación de la calificación* en la Secretaría Virtual.

La solicitud deberá ir dirigida al profesor o profesora responsable de la calificación y ha de estar debidamente motivada.

PRUEBAS DE EVALUACIÓN CONTINUA (PEC)

¿Hay PEC?	No
Descripción	
No hay PEC.	
Criterios de evaluación	
Ponderación de la PEC en la nota final	0
Fecha aproximada de entrega	No procede
Comentarios y observaciones	

OTRAS ACTIVIDADES EVALUABLES

¿Hay otra/s actividad/es evaluable/s? Si

Descripción

Actividad voluntaria sobre la herramienta de análisis de datos Jamovi, publicada en el curso virtual de la asignatura.

Criterios de evaluación

La actividad se evaluará con una puntuación de entre 0 y 0,5 puntos, según el rendimiento del estudiante en la misma. Esta puntuación solo se sumará a la nota del examen si este está aprobado (es decir, si se obtiene una calificación de 5 o más).

En el campus virtual de la asignatura se ofrecerá información detallada sobre las características de esta actividad y los criterios de evaluación.

Ponderación en la nota final 0

Fecha aproximada de entrega Se indicará en el curso virtual

Comentarios y observaciones

¿CÓMO SE OBTIENE LA NOTA FINAL?

La calificación final se obtiene a partir de la nota del examen. Si el examen está aprobado, se sumará a esa nota la puntuación obtenida en la actividad voluntaria de Jamovi, en caso de que el estudiante la haya realizado.

BIBLIOGRAFÍA BÁSICA

ISBN(13):9788417765422
Título:INTRODUCCIÓN AL ANÁLISIS DE DATOS: APLICACIONES EN PSICOLOGÍA Y CIENCIAS DE LA SALUD 2ª
Autor/es:Suárez Falcón, J. C. ; Pozo Cabanillas, P. ; San Luis Costas, C. ; Recio Saboya, P. ;
Editorial:SANZ Y TORRES, S.L.

ISBN(13):9788417765439
Título:FORMULARIO Y TABLAS ESTADÍSTICAS DE INTRODUCCIÓN AL ANÁLISIS DE DATOS2ª
Autor/es:Suárez Falcón, J. C. ; Pozo Cabanillas, P. ; San Luis Costas, C. ; Recio Saboya, P. ;
Editorial:SANZ Y TORRES, S.L.

La bibliografía básica de la asignatura es:

•Suárez Falcón, J.C., Recio Saboya, P., San Luis Costas, M.C., y Pozo Cabanillas, P. (2019). *Introducción al análisis de datos: Aplicaciones en psicología y ciencias de la salud (2ª Ed.)*. Madrid: Sanz y Torres. ISBN: 978-84-17765-42-2.

•Suárez Falcón, J.C., Recio Saboya, P., San Luis Costas, M.C., y Pozo Cabanillas, P. (2019). *Formulario y tablas estadísticas de introducción al análisis de datos en psicología y ciencias de la salud (2ª Ed.).* Madrid: Sanz y Torres. ISBN: 978-84-17765-43-9.

IMPORTANTE

La 1ª Edición del libro básico y del Formulario y tablas, publicados en 2017, también son válidos para el estudio de la asignatura.
El Formulario y Tablas estadísticas de Introducción al Análisis de Datos en Psicología y Ciencias de la Salud, **editado en papel**, debe utilizarse para la realización del examen pero SIN ANOTACIONES por parte del alumno. Por este motivo es conveniente adquirirlo con suficiente antelación.

BIBLIOGRAFÍA COMPLEMENTARIA

ISBN(13):9788436800821
Título:ESTADÍSTICA PARA PSICÓLOGOS I15
Autor/es:Amón Hortelano, Jesús ;
Editorial:EDICIONES PIRÁMIDE, S.A.

ISBN(13):9788436801538
Título:ESTADÍSTICA PARA PSICÓLOGOS II9ª ed
Autor/es:Amón Hortelano, Jesús ;
Editorial:PIRÁMIDE

El texto básico y el material disponible en el curso virtual son suficientes para la preparación de la asignatura. Aquellos estudiantes que lo deseen pueden consultar los dos libros de la bibliografía complementaria que explican, clara y extensamente, los temas tratados en las unidades didácticas de la asignatura profundizando más en los desarrollos matemáticos y aportando un gran número de ejercicios.

RECURSOS DE APOYO Y WEBGRAFÍA

Para el estudio de la asignatura, los alumnos encontrarán en el Curso Virtual herramientas y materiales didácticos complementarios (material multimedia, preguntas más frecuentes, ejercicios resueltos, exámenes resueltos, foros,....).

IGUALDAD DE GÉNERO

En coherencia con el valor asumido de la igualdad de género, todas las denominaciones que en esta Guía hacen referencia a órganos de gobierno unipersonales, de representación, o miembros de la comunidad universitaria y se efectúan en género masculino, cuando no se hayan sustituido por términos genéricos, se entenderán hechas indistintamente en género femenino o masculino, según el sexo del titular que los desempeñe.

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2025 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Material: **Adenda (Formulario y Tablas) y calculadora (no programable)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados. La pregunta 26 es de reserva y solamente se utilizará si se anula alguna pregunta del examen.

Tabla 1. Resultados obtenidos por 240 estudiantes de 3º de la ESO en un test de motivación (X) divididos en dos grupos.

Motivación (X)	Repetidores (n_i)	No repetidores (n_i)
80-99	20	32
60-79	22	40
40-59	33	23
20-39	30	15
1-19	15	10

Figura 1. Diagrama de caja de la variable "Horas de estudio al día" para repetidores y no repetidores

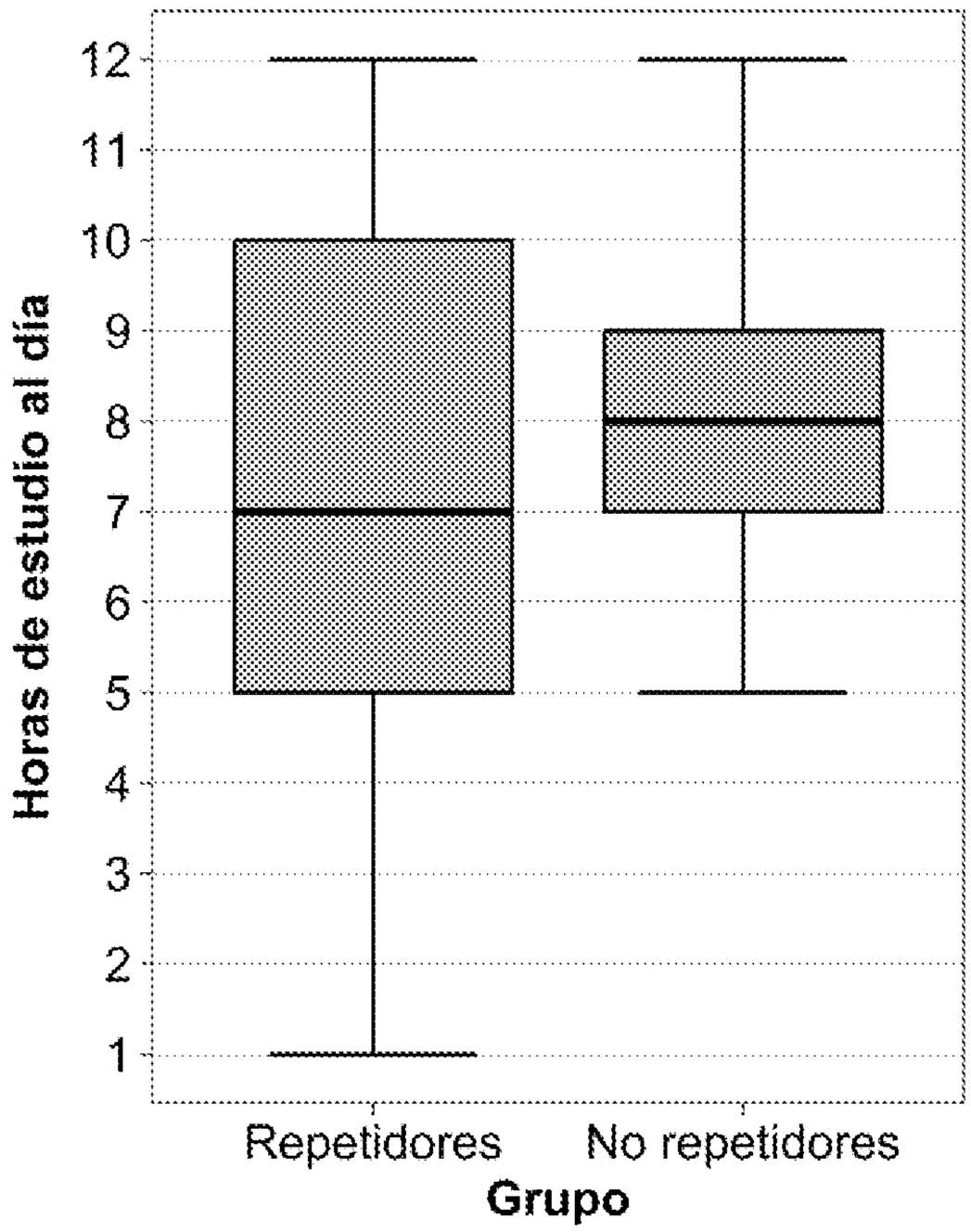


Tabla 2. Resultados de un estudio sobre la relación entre el contacto intergrupal y el prejuicio hacia los inmigrantes

	Prejuicio Bajo	Prejuicio Medio	Prejuicio Alto
Con contacto intergrupal	150	60	30
Sin contacto intergrupal	40	70	130

Tabla 3. Descriptivos de las variables bienestar (X) y rendimiento académico (Y) de 132 alumnos/as.

	Media	**Varianza**	**Covarianza**
X	6,88	1,95	1,771
Y	6,90	3,10	

1. ¿Qué nivel de medida presenta la variable "Motivación" de la Tabla 1?: A) Nominal; B) Intervalo; C) Ordinal.
2. ¿Qué porcentaje del total de personas repetidoras puntuaron igual o menor a 59,5 en el test de motivación de la Tabla 1? A) 37,51%; B) 83,33%; C) 65%.

3. ¿Cuál es la moda en la distribución de frecuencias de la Tabla 1 para los no repetidores? A) 60-79; B) 10; C) 69,5.
4. La frecuencia relativa para la moda del grupo de los repetidores es: A) 0,275; B) 0,137; C) 0,125
5. Atendiendo a la Tabla 1, y sabiendo que las varianzas del grupo de repetidores y no repetidores son respectivamente 640,22 y 611,08, calcula el coeficiente de variación del grupo de repetidores y no repetidores. Con los resultados se puede concluir que: A) los datos de ambos grupos tienen una dispersión similar; B) los datos de repetidores presentan una mayor dispersión que los datos de no repetidores; C) los datos de no repetidores presentan una mayor dispersión que los datos de repetidores.
6. En la Figura 1, el valor de la amplitud intercuartil de la variable "Horas de estudio al día" para repetidores y no repetidores es: A) 5 para repetidores y 2 para no repetidores; B) 7 para repetidores y 8 para no repetidores; C) 11 para repetidores y 7 para no repetidores.
7. En la Figura 1, el diagrama de caja y bigotes para la variable "horas de estudio al día" del grupo de no repetidores muestra una: A) distribución simétrica; B) distribución asimétrica negativa; C) distribución asimétrica positiva.
8. En la Tabla 2, ¿cuál es la probabilidad de tener un prejuicio bajo condicionada a que se ha tenido contacto intergrupal? A) 0,625; B) 0,789; C) 0,313.
9. Con los datos de la Tabla 2, la V de Cramer es: A) 0,51; B) 0,32; C) 0,61.
10. En la Tabla 2, el número de personas que no tienen contacto intergrupal y presentan un prejuicio medio es una frecuencia: A) condicional; B) marginal; C) conjunta.
11. Con los datos de la Tabla 3, el coeficiente de correlación de Pearson es igual a: A) 0,72; B) -0,72; C) 0,90.
12. Para conocer la proporción de varianza de la variable pronosticada o criterio (Y) que es explicada por la variable predictora (X) en un modelo lineal, hay que calcular: A) el coeficiente C de Contingencia; B) el coeficiente de determinación; C) el coeficiente de correlación de Spearman.
13. Atendiendo a la Tabla 3, ¿cuál sería el rendimiento académico que se predeciría para un alumno que ha puntuado 5 en bienestar? A) 4,75; B) 4,11; C) 5,20.
14. Indica qué afirmación es correcta sobre el Teorema de la Suma: A) cuando los sucesos A y B son dependientes, la probabilidad de ocurrencia de A y B es igual a la probabilidad de ocurrencia de A por la probabilidad de ocurrencia de B; B) se emplea para calcular la probabilidad de unión entre dos sucesos; C) se emplea para calcular la probabilidad de intersección entre dos sucesos.
15. Un determinado trastorno del sueño puede estar provocado solo por el estrés laboral (L) en el 40% de los casos, solo por malos hábitos de sueño (H) en el 35% de los casos, y por la combinación (C) de ambos factores en el 25% de los casos. El tratamiento de este trastorno requiere terapia conductual (T) en el 15% de los casos si la única causa es el estrés laboral, en el 40% de los casos si la única causa son los malos hábitos de sueño, y en el 30% de los casos si es producido por la combinación de ambos factores. ¿Cuál es la probabilidad de que un paciente cualquiera no necesite terapia conductual? A) 0,357; B) 0,275; C) 0,725.
16. Selecciona la función de distribución correcta para una variable aleatoria discreta:
A) F(0)=0,15, F(1)=0,20, F(2)=0,30, F(3)=0,80, F(4)=1; B) F(0)=0,15, F(1)=0,50, F(2)=0,10, F(3)=0,50, F(4)=0,20; C) F(0)=0,15, F(1,5)=0,50, F(2,5)=0,10, F(3,5)=0,50, F(4,5)=0,20.

17. En una ciudad, se sabe que un 15% de las personas son alérgicas a cierto tipo de polen. Si elegimos aleatoriamente una muestra de 10 personas, ¿cuál es la probabilidad de que exactamente 2 de ellas sean alérgicas? A) 0,2759; B) 0,8202; C) 0,1969.
18. Una variable aleatoria se distribuye según una *t* de Student con 12 grados de libertad. ¿Cuál es el valor de *t* que deja por encima de sí el 10% de la distribución?
A) 1,782; B) 1,356; C) 1,771.
19. Las puntuaciones de 1200 estudiantes en una prueba de aptitud verbal se distribuyen normalmente con una media de 110 y una desviación estándar de 12. ¿Cuál es la probabilidad de obtener una puntuación menor o igual que 95? A) 0,1056; B) 0,3944; C) 0,2119.
20. En una distribución F con 8 grados de libertad en el numerador y 9 grados de libertad en el denominador, ¿cuál es el valor de F que corresponde al percentil 95? A) 6,693; B) 2,602; C) 3,230.
21. Un investigador desea estudiar el nivel de actividad física en adultos mayores. Decide ir a un parque cercano y medir la frecuencia cardíaca y actividad física de las primeras 50 personas mayores de 60 años a las que tiene fácil acceso y encuentra durante su visita. El método de muestreo es: A) incidental; B) intencional; C) bola de nieve.
22. En la población de adultos de Sevilla, se sabe que el porcentaje de personas que realizan actividad física al menos 3 veces por semana es del 30%. Se selecciona una muestra aleatoria de tamaño 200. ¿Cuál es el error típico de estimación de la proporción? A) 0,032; B) 0,045; C) 0,065.
23. Si una variable Y tiene una media $\mu=50$ y una desviación típica $\sigma=12$ en la población, ¿cuáles son los valores de μ y $\sigma_{\bar{y}}$ (error estándar de la media) para una muestra de tamaño 49? A) 50 y 2,00; B) 50 y 3,00; C) 50 y 1,71.
24. Señala cuál de las siguientes afirmaciones es verdadera en relación a los intervalos de confianza: A) muestras heterogéneas implican un error típico de medida pequeño, intervalos de menor amplitud y más precisos; B) un nivel de confianza del 95% conlleva intervalos de confianza más precisos que un nivel de confianza del 99%; C) Un mayor tamaño muestral conlleva un mayor error típico de medida.
25. Una variable X se distribuye normalmente en la población con una varianza poblacional igual a 49. Si se admite un error máximo de estimación de 0,2 y se fija un nivel de confianza del 95%, ¿qué tamaño muestral aproximado es necesario para estimar la media poblacional asumiendo población infinita?: A) 2750; B) 3210; C) 4706.

Reserva

26. Un investigador desea realizar un estudio sobre el porcentaje de estudiantes que consumen bebidas energéticas durante los exámenes. Se asume que la población es infinita. Con un nivel de confianza del 95% y un margen de error del 5%, ¿cuál debe ser el tamaño de la muestra aproximado, si estudios previos indican que la proporción de consumidores es del 15%? A) 248; B) 196; C) 323.

FORMULARIO Y TABLAS

Introducción al Análisis de Datos:

Aplicaciones en la Psicología y las Ciencias de la Salud

FE DE ERRATAS 1ª Edición (2017)

Página	**Localización**	**Dice**	**Debería decir**
11	En la segunda fórmula de Función de distribución:	$P(x_1 \leq x \leq x_2) = F(x_2) - F(x_1)$	$P(x_1 < x \leq x_2) = F(x_2) - F(x_1)$
16	En la fórmula de la varianza de F de Fisher-Snedecor	$\sigma^2 = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-4)^2(n_2-2)^2}$	$\sigma^2 = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-4)(n_2-2)^2}$
18	En la primera línea de texto	La variable X no se distribuye normalmente y ($n \geq 30$) σ conocida:	La variable X no se distribuye normalmente y ($n \geq 30$) (Teorema del límite central) σ conocida:
19	En las fórmulas de la aproximación a la normal de la distribución muestral de la varianza para muestras grandes	$Z = \frac{\left(\frac{nS_X{}^2}{\sigma^2}\right) - (n-1)}{\sqrt{(2n-1)}}$	$Z = \frac{\left(\frac{nS_X{}^2}{\sigma^2}\right) - (n-1)}{\sqrt{2(n-1)}}$
		$Z = \frac{\left(\frac{(n-1)S_{n-1}^2}{\sigma^2}\right) - (n-1)}{\sqrt{(2n-1)}}$	$Z = \frac{\left(\frac{(n-1)S_{n-1}^2}{\sigma^2}\right) - (n-1)}{\sqrt{2(n-1)}}$
23	En tamaño muestral para el parámetro proporción Poblaciones infinitas:	$E_{max} = Z_{\alpha/2}\sqrt{\frac{P(1-P)}{E^2{}_{max}}}$	$E_{max} = Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}$
41	En la figura de la distribución t de Student, en el eje horizontal	1,383	1,372

INTRODUCCIÓN AL ANÁLISIS DE DATOS
Código asignatura: 62011037
EXAMEN TIPO TEST MODELO B DURACION: 2 HORAS

Material: **Adenda (Formulario y Tablas) y calculadora (no programable)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados. La pregunta 26 es de reserva y solamente se utilizará si se anula alguna pregunta del examen.

Tabla 1. Resultados obtenidos en un test de autoconcepto (X) aplicado a 40 adolescentes.

X	n_i
26-28	7
23-25	11
20-22	9
17-19	7
14-16	6

Figura 1. Resultados de la mejora o no del colesterol tras la administración de tres fármacos distintos en un grupo de 150 personas hipertensas. Los participantes fueron asignados aleatoriamente a uno de los tres tratamientos.

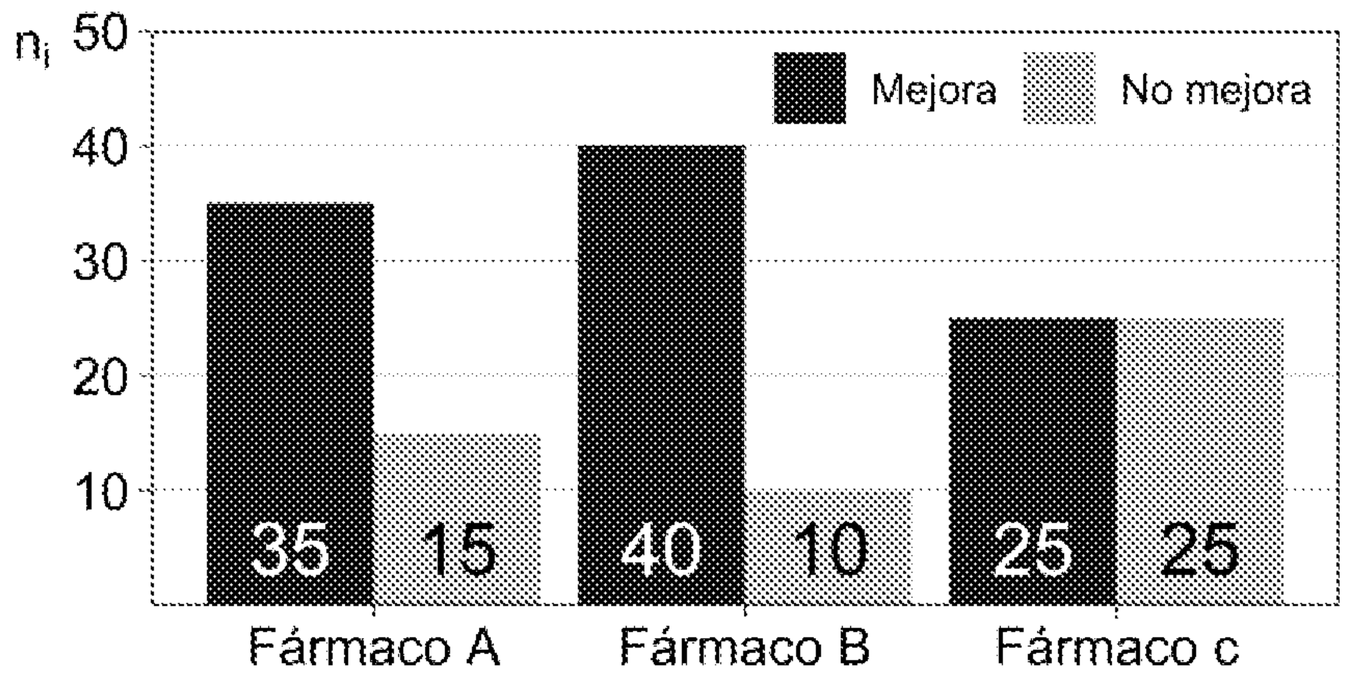


Tabla 2. Para dos variables cuantitativas, X e Y, disponemos de los siguientes datos:

X	Y	XY
$\overline{X}=8$	$\overline{Y}=100$	$Cov(X,Y)=S_{XY}=18$
$S_X=2$	$S_Y=10$	

Tabla 3. Variables X e Y, y su tipo de distribución.

Variable	**Distribución**	
X	$N(20,5)$	Normal con media 20 y desviación típica 5
Y	X^2_{40}	Ji-cuadrado con 40 grados de libertad

1. La variable "Tipo de Fármaco" de la Figura 1 es: A) cuasicuantitativa; B) cuantitativa; C) politómica.
2. La amplitud del intervalo de los datos de la Tabla 1 es: A) 3; B) 2,5; C) 2.
3. ¿Cuál es el gráfico más adecuado para representar los datos de la Tabla 1? A) Diagrama de barras; B) Diagrama de dispersión; C) Histograma.
4. La media aritmética de la distribución de X en la Tabla 1 es: A) 21,45; B) 20,83; C) 27,48.
5. Atendiendo a los datos de la Tabla 1, el valor del Percentil 75 es: A) 30,25; B) 20,04; C) 24,68.
6. La moda es un índice de: A) posición; B) tendencia central; C) variabilidad.
7. Con los datos de la Tabla 1, la amplitud intercuartil es: A) 8,21; B) 6,47; C) 4,68.
8. Según los datos de la Figura 1, asumiendo que ambas variables son independientes, ¿cuál es la frecuencia teórica de mejorar y recibir el fármaco B? A) 16,76; B) 33,33; C) 40.
9. Con los datos de la Figura 1, hemos obtenido un valor de X^2 igual a 10,5. El coeficiente *C* de Contingencia tiene un valor que está comprendido entre: A) 0,1 y 0,4; B) 0,5 y 0,8; C) 0,9 y 1,1.
10. Teniendo en cuenta la Tabla 2, la recta de regresión que permite pronosticar la variable *Y* en función de la variable *X* es: A) Y' = 64 + 4,5X; B) Y' = 70 - 0,5X; C) Y' = 35 + 6X.

11. Se quiere observar la relación entre la posición en un ranking educativo (X) y la posición en un ranking socioeconómico (Y). Para ello, se obtiene la información sobre ambas variables de 10 personas. Los rangos de X son {4, 3, 8, 5, 6, 2, 9, 7, 1, 10}, y los rangos de Y son {6, 4, 2, 7, 8, 9, 3, 5, 1, 10}. ¿Qué procedimiento podría usarse para calcular la relación entre ambas variables? A) Solamente el coeficiente de correlación de Pearson; B) Solamente el coeficiente de correlación de Spearman; C) El coeficiente de correlación tanto de Pearson como de Spearman.
12. Con los datos de la pregunta 11, ¿qué valor toma el coeficiente de correlación entre X e Y? A) -0,16; B) 0,16; C) 0,21.
13. En un grupo de 100 estudiantes, se está realizando un análisis sobre el interés en dos actividades extracurriculares: asistir a clases de música (M) y participar en un club de debate (D). Se sabe que el 60% de los estudiantes asisten a clases de música, el 40% de los estudiantes participan en el club de debate, y el 15% de los estudiantes asisten tanto a clases de música como participan en el club de debate. ¿Cuál es la probabilidad de que un estudiante asista a clases de música, o participe en el club de debate, o realice ambas actividades? A) 0,85; B) 0,70; C) 0,15.
14. Teniendo en cuenta la representación de la Figura 1, elegida una persona al azar ha resultado que se le administró el fármaco A. ¿Cuál es la probabilidad aproximada de que mejore? A) 0,23; B) 0,50; C) 0,70.
15. Con los datos de la Figura 1, si elegimos una persona al azar ¿cuál es la probabilidad aproximada de que presente mejoría en su colesterol?: A) 0,33; B) 0,67; C) 0,70.
16. Sea X una variable aleatoria discreta con valores 1, 2, 3 y 4. Si los cuatro valores de X son equiprobables, su media es: A) 2,50; B) 1,50; C) 0,25.
17. En el lanzamiento de un dado una sola vez se ha definido la variable X "obtener número impar". La variable X se distribuye según: A) Bernoulli con parámetro p = 1/2; B) Binomial con parámetro p = 1/6; C) Binomial con parámetro p = 1/2.
18. La función de distribución de una variable aleatoria discreta: A) asocia a cada valor de la variable la probabilidad de que ésta adopte ese valor; B) asocia a cada valor de la variable la probabilidad de que ésta adopte ese valor o cualquier otro inferior; C) asocia a cada valor de la variable la probabilidad de que ésta adopte cualquier valor inferior a éste.
19. Teniendo en cuenta los datos de la Tabla 3, P (X > 25) es aproximadamente: A) 0,2500; B) 0,8413; C) 0,1587.
20. Con los datos de la Tabla 3, para la variable Y, el valor 51,8051 es el percentil: A) 10; B) 50; C) 90.
21. La distribución F: A) es asimétrica negativa; B) puede adoptar cualquier valor entre -∞ y +∞; C) posee la propiedad recíproca.
22. Calcule el error típico de la proporción sabiendo que las muestras (extraídas por muestreo aleatorio simple) tienen un tamaño de 80 y el valor esperado es 0,8. A) Se puede calcular y su valor es aproximadamente 0,045; B) No se puede calcular dado que falta el nivel de significación; C) Se puede calcular y su valor es aproximadamente 0,022.
23. Sea X una variable con distribución normal en la población con N(150; 8). Se extrae una muestra aleatoria simple de tamaño 15, ¿cuál es la probabilidad aproximada de que su media sea menor o igual a 145? A) 0,0078; B) -2,4206; C) 0,9922.
24. Se quiere estimar la media de la variable bienestar físico en un grupo de personas que practican pilates. Se extrae una muestra de 74 personas, de una población de 2500. Sabiendo que el bienestar físico se distribuye N (μ; 17), ¿cuál es el valor aproximado del error máximo que se comete con este tamaño muestral con un nivel de significación α = 0,05? A) 1,273; B) 3,814; C) 5,016.
25. Sabemos que el tiempo promedio que los estudiantes universitarios dedican a estudiar por día se distribuye normalmente con una desviación estándar de σ=3 horas. En una muestra aleatoria de 64 estudiantes, se observa una media muestral de 5 horas. ¿Entre qué valores estimaremos que se encuentra la media del tiempo de estudio en la población con un nivel de confianza del 99%? A) 4,03 y 5,97; B) 4,03 y -5,97; C) -4,03 y 5,97.

Pregunta de Reserva:

26. Se hace un muestreo aleatorio simple de 60 estudiantes de la UNED. Un tercio ha realizado una PEC. ¿Cuál es el intervalo de confianza (α=0,05) para estimar la proporción de alumnos que han realizado esta PEC? A) (0,211; 0,449); B) (0,211; 0,614); C) (0,032; 0,614).

FORMULARIO Y TABLAS

Introducción al Análisis de Datos:

Aplicaciones en la Psicología y las Ciencias de la Salud

FE DE ERRATAS 1ª Edición (2017)

Página	Localización	Dice	Debería decir
11	En la segunda fórmula de Función de distribución:	$P(x_1 \le x \le x_2) = F(x_2) - F(x_1)$	$P(x_1 < x \le x_2) = F(x_2) - F(x_1)$
16	En la fórmula de la varianza de F de Fisher-Snedecor	$\sigma^2 = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-4)^2(n_2-2)^2}$	$\sigma^2 = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-4)(n_2-2)^2}$
18	En la primera línea de texto	La variable X no se distribuye normalmente y $(n \ge 30)$ σ conocida:	La variable X no se distribuye normalmente y $(n \ge 30)$ (Teorema del límite central) σ conocida:
19	En las fórmulas de la aproximación a la normal de la distribución muestral de la varianza para muestras grandes	$Z = \frac{\left(\frac{nS_X{}^2}{\sigma^2}\right) - (n-1)}{\sqrt{(2n-1)}}$	$Z = \frac{\left(\frac{nS_X{}^2}{\sigma^2}\right) - (n-1)}{\sqrt{2(n-1)}}$
		$Z = \frac{\left(\frac{(n-1)S_{n-1}^2}{\sigma^2}\right) - (n-1)}{\sqrt{(2n-1)}}$	$Z = \frac{\left(\frac{(n-1)S_{n-1}^2}{\sigma^2}\right) - (n-1)}{\sqrt{2(n-1)}}$
23	En tamaño muestral para el parámetro proporción Poblaciones infinitas:	$E_{max} = Z_{\alpha/2}\sqrt{\frac{P(1-P)}{E^2{}_{max}}}$	$E_{max} = Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}$
41	En la figura de la distribución t de Student, en el eje horizontal	1,383	1,372

Código asignatura	Nombre asignatura
62011037	**INTRODUCCIÓN AL ANÁLISIS DE DATOS**
Fecha alta y origen	Convocatoria
17/09/2024	**SEPTIEMBRE 2024 – TIPO A**
Curso virtual	

bINTRODUCCIÓN AL ANÁLISIS DE DATOS
Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Material: **Adenda (Formulario y Tablas) y calculadora (no programable)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

c

Tabla 1. Puntuaciones de 70 adultos en una escala de *depresión* (X) agrupadas en intervalos y sus frecuencias absolutas acumuladas (n_a).

X	n_a
81-100	70
61-80	62
41-60	40
21-40	25
1-20	10

Figura 1. Diagrama de caja y bigotes de las puntuaciones en una prueba de atención de 500 participantes.

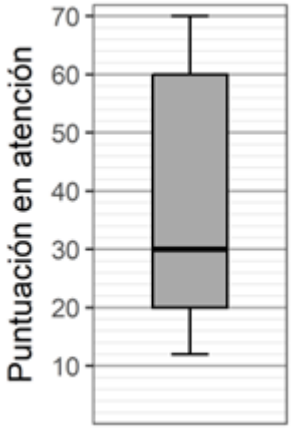


Tabla 2. Resultados obtenidos en un estudio (*n*=80) sobre la relación entre la presencia de estrés y la frecuencia con la que se realiza actividad física.

		Frecuencia deporte		
		Alta	Media	Baja
Presencia de estrés	No	20	10	10
	Sí	15	5	20

Tabla 3. Datos sobre la variable "responsabilidad" (X) y "rendimiento académico" (Y) de 100 estudiantes.

Σ(XY)	47,93
Σ(X)	-3,33
Σ(Y)	288,55
$\Sigma(X^2)$	98,06

$\bar{X}$	-0,03
$\bar{Y}$	2,89

1. La variable de la Tabla 1 "puntuación en depresión" es: A) Nominal; **B) Cuantitativa**; C) Ordinal.
2. A las propiedades descriptivas de una población se les denomina: A) Estadísticos; **B) Parámetros**; C) Muestra.
3. ¿Cuál es la amplitud de los intervalos de la Tabla 1? A) 19; B) 21; **C) 20.**
4. En la Figura 1, la mayor concentración de puntuaciones se produce en la parte baja de la escala. Por lo tanto, la distribución de estas puntuaciones es: **A) Asimétrica positiva**; B) Asimétrica negativa; C) Simétrica.
5. El 30% de los participantes de la Tabla 1 con mayores puntuaciones en depresión serán derivados a una clínica especializada. ¿A partir de qué puntuación se derivará al 30% de los participantes a una clínica especializada? A) 35,17; B) 56,98; C) 69,18. **Pregunta anulada.**
6. ¿Cuál es la desviación típica de la variable *depresión* en la Tabla 1? A) 622,86; **B) 24,96**; C) 58,74.
7. ¿Cuál es la amplitud intercuartil de las puntuaciones de la Figura 1? **A) 40**; B) 30; C) 10.
8. ¿Cuál de los siguientes índices sería el más correcto para estudiar la asociación entre las variables de la Tabla 2? **A) El Coeficiente C de Contingencia**; B) Coeficiente φ (phi); C) Correlación de Spearman.

9. ¿Cuál es la distribución marginal de la variable "presencia de estrés" en la Tabla 2? A) 13,33 y 13,33; B) 35,15 y 30; **C) 40 y 40.**
10. La covarianza entre las variables X e Y es 2450, y la covarianza entre las variables Z y W es 245. Dado que tienen diferentes unidades de medida, podemos concluir de estos datos: A) Las variables X e Y están más relacionadas que las variables Z y W; B) Las variables X e Y están 10 veces más relacionadas que las variables Z y W; **C) Ninguna de las opciones anteriores es correcta.**
11. Utilizando los datos de la Tabla 3 ¿Cuál es la recta de regresión de la variable responsabilidad" sobre la variable "rendimiento académico"? A) Y = 0,387 + 1,890·X; **B) Y = 2,908 + 0,587·X;** C) Y = 0,397 + 2,890·X
12. Teniendo en cuenta que S_X = 0,99 y S_Y = 2,06, ¿qué porcentaje de la variabilidad en el "rendimiento académico" de la Tabla 3 se puede explicar por la variable "responsabilidad"? A) 28; B) 0,28; **C) 8.**
13. Atendiendo a los datos de la Tabla 2, dado que una persona realiza mucho deporte, ¿cuál es la probabilidad de que sufra estrés? **A) 0,429**; B) 0,150; C) 0,352.
14. Atendiendo a los datos de la Tabla 2, ¿cuál es la probabilidad de que una persona sufra estrés o realice una actividad física medianamente frecuente? A) 0,688; B) 0,357; **C) 0,625.**
15. Si dos sucesos A y B son independientes: **A) P(A ∩ B) = P(A) · P(B)**; B) P(A ∩ B) = P(A) + P(B); C) P(A ∩ B) = P(A) + P(B) - P(A ∪ B).
16. Señala cuál es una función de probabilidad correcta: **A) *f*(0) = 0,16; *f*(2) = 0,52; *f*(4) = 0,32;** B) *F*(0) = 0,16; *F*(2) = 0,68; *F*(4) = 1; C) *f*(-1) = 0,06; *f*(0) = 0,22; *f*(1) = 0,32.
17. Según la Organización Mundial de la Salud, el 55% de la población ha sufrido problemas de insomnio en algún momento de su vida. Si elegimos una muestra aleatoria de 12 personas, ¿Cuál es la probabilidad de que más de 8 personas nunca hayan sufrido problemas de insomnio? A) 0,9644; B) 0,0730; **C) 0,0356**.
18. Una variable sigue una distribución X^2 con 30 grados de libertad. ¿Qué valor de X^2 deja por encima una probabilidad de 0,10? **A) 40,256**; B) 50,892; C) 20,499.
19. La variable Y sigue una distribución *t* de Student con 40 grados de libertad. ¿Qué valor de *t* corresponde al percentil 30 de la distribución?: **A) -0,529**; B) 0,851; C) -0,432.
20. ¿Cuál de estos tipos de muestreo es no probabilístico? A) Estratificado; **B) Por cuotas**; C) Sistemático.
21. El trastorno obsesivo compulsivo afecta al 7% de la población. Se recoge muestra de 120 personas. ¿Cuál es el valor de la desviación típica de la distribución muestral de esta proporción? A) 0,001; B) 0,045; **C) 0,023.**
22. La variable extraversión se distribuye normalmente con media 35 y varianza 76. Se seleccionan muestras aleatorias de 96 participantes. ¿Entre qué valores se encuentra el 50% central de las medias en la distribución? A) -15,92 y 85,92; B) 29,16 y 40,84; C) 16,28 y 56,34. **Pregunta anulada.**
23. Dados dos estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$ del mismo parámetro θ, el menos eficiente será: A) El que tenga una distribución muestral menos dispersa; B) El que tenga una media en la distribución muestral más cercana al valor del parámetro; **C) El que tenga una distribución muestral con mayor varianza.**
24. Una muestra de 78 estudiantes ha completado un test de cultura general, obteniendo una media de 7,8 puntos. Esta variable se distribuye normalmente, y se conoce que su varianza poblacional es 4,5. ¿Cuál es el valor del intervalo de confianza de la media de esta variable con un nivel de riesgo de 0,05? **A) (7,33 y 8,27);** B) (6,80 y 8,80); C) (7,12 y 8,34).
25. En una muestra aleatoria de 230 profesores, 155 están a favor de incorporar tecnología en las aulas. Con un nivel de confianza del 99%, el error de estimación máximo de la proporción de los que están a favor del cambio de reglamento en la población es de: **A) 0,080**; B) 0,101; C) 0,157.

SOLUCIONES:

1. B

2. B. Página 10 del manual

3. C. La amplitud de un intervalo se obtiene restando los límites exactos. Cogiendo los límites exactos del primer intervalo: 100,5 – 80,5 = 20

4. A. Página 38 del manual: *Una distribución tiene **asimetría positiva** cuando la mayor concentración de puntuaciones se produce en la parte baja de la escala y **asimetría negativa** cuando la mayor parte de las puntuaciones se sitúan en la parte alta de la escala*.

5. Pregunta anulada.

6. B

Intervalo de depresión	n_a	n_i	X_i	$X_i\ n_i$	X_i^2	$X_i^2\ n_i$
81-100	70	8	90,5	724	8190,25	65522
61-80	62	22	70,5	1551	4970,25	109345,5
41-60	40	15	50,5	757,5	2550,25	38253,75
21-40	25	15	30,5	457,5	930,25	13953,75
1-20	10	10	10,5	105	110,25	1102,5
				3595	16751,25	228177,5

$$\bar{X} = \frac{\sum \text{Xi ni}}{n} = \frac{3595}{70} = 51{,}35$$

$$S_X^2 = \frac{\sum n_i X_i^2}{n} - \bar{X}^2 = \frac{228177{,}5}{70} - 51{,}35^2 = 3259{,}68 - 2636{,}82 = 622{,}86$$

$$S_X = \sqrt{S_X^2} = \sqrt{622{,}86} = 24{,}96$$

7. A. La amplitud intercuartílica se obtiene restando el cuartil 3 (60) del cuartil 1 (20)

8. A.

9. C. La distribución marginal es el número de observaciones para cada valor de una de las variables, prescindiendo de la información sobre los valores de las demás variables. Por lo tanto, la distribución marginal de la variable “presencia de estrés” es la suma de las filas.

10. C. El valor de la covarianza entre dos variables depende de las unidades de medida de las variables involucradas, por lo tanto, no se pueden comparar dos covarianzas de diferentes conjuntos de variables.

11. B. Y = 2,908 + 0,587·X

$$b = \frac{100 \cdot 47{,}93 - (-3{,}33 \cdot 288{,}55)}{100 \cdot 98{,}06 - (-3{,}33)^2} = \frac{5753{,}87}{9794{,}91} = 0{,}587$$

$a = 2{,}89 - 0{,}587 \cdot (-0.03) = 2{,}908$

12. C

$b = r_{XY}\frac{S_Y}{S_X} \rightarrow r_{XY} = \frac{b \cdot S_X}{S_Y} = \frac{0{,}587 \cdot 0{,}99}{2{,}06} = 0{,}282$

$r_{XY}^2 = 0{,}282^2 = 7{,}997 \approx 8$

13. A

P(estrés/alta actividad) = $\frac{\text{P(estrés} \cap \text{alta actividad)}}{\text{P(alta actividad)}} = \frac{\frac{15}{80}}{\frac{35}{80}} = 0{,}429$

14. C

$P(\text{estrés} \cup \text{actividad}_{\text{media}}) = P(\text{estrés}) + \text{P}(\text{actividad}_{\text{media}}) - P(\text{estrés} \cap \text{actividad}_{\text{media}}) =$
$= \frac{40}{80} + \frac{15}{80} - \frac{5}{80} = 0{,}625$

15. A. Ver ejercicio resuelto 6.3 del manual.

16. A

17. C. La probabilidad de nunca haber sufrido problemas de insomnio es 1-0.55 = 0.45
Tabla II: P(X > 8) = 1 – P(X ≤ 8) = 1-0,9644 = 0,0356.

18. A. Tabla V.

19. A. Tabla VI

20. B. Página 379.

21. C. La desviación típica de la distribución muestral de una proporción para muestras grandes es:

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0{,}07(0{,}93)}{120}} 0{,}023$$

22. Pregunta anulada.

23. C

24. A

$\alpha =$ 0,05 → $\left|\alpha/2\right| = |1{,}96|$ (Tabla III y IV);

$L_{inf} = \bar{X} - \left|Z_{\frac{\alpha}{2}}\right| \sigma/\sqrt{n} = 7{,}8 - 1{,}96 \cdot \frac{2{,}12}{\sqrt{78}} = 7{,}8 - 0{,}470 = 7{,}33$

$L_{sup} = \bar{X} + \left|Z_{\frac{\alpha}{2}}\right| \sigma/\sqrt{n} = 7{,}8 + 1{,}96 \cdot \frac{2{,}12}{\sqrt{78}} = 7{,}8 + 0{,}470 = 8{,}27$

Por tanto, los límites del intervalo son: (7,33; 8,27)

25. A

n.c. = 0,99 → $z_{1-\alpha/2} = z_{0,995} = 2,58$ (Tabla IV). La proporción de profesores a favor de la tecnología es 0,67.

$$E_{máx} = z_{1-\frac{\alpha}{2}}\sqrt{\frac{P(1-P)}{n}} = 2,58 \times \sqrt{\frac{0,67\cdot(1-0,67)}{230}} = 0,080$$

INTRODUCCIÓN AL ANÁLISIS DE DATOS
Código asignatura: 62011037
EXAMEN TIPO TEST MODELO B DURACION: 2 HORAS

Material: **Adenda (Formulario y Tablas) y calculadora (no programable)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Tabla 1. "Resiliencia familiar" (X) en una muestra de madres de niños con autismo.

X	n_i
110-114	2
105-109	6
100-104	17
95-99	30
90-94	45

Figura 1. Distribución de un grupo de 800 profesores por sexo y etapa educativa (200 en cada una de las etapas)

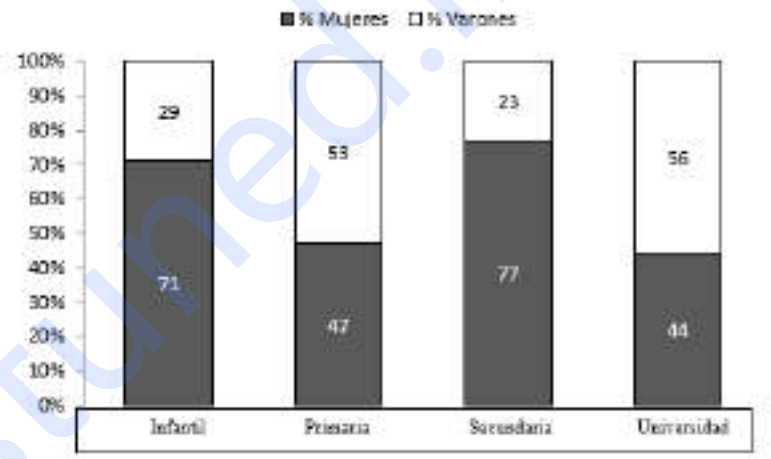
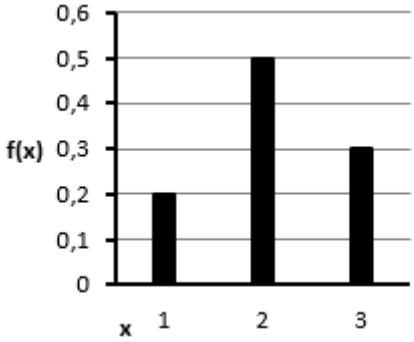


Tabla 2. Para dos variables cuantitativas, X e Y, disponemos de los siguientes datos:

X	Y	
$\overline{X} = 100$	$\overline{Y} = 8$	$Y' = 0{,}1X - 2$
$S_X = 16$	$S_Y = 2$	

Figura 2: Función de probabilidad, *f*(x), de una variable aleatoria discreta X.



1. La variable *resiliencia familiar* de la Tabla 1 es de tipo: A) nominal politómica; **B) cuantitativa**; C) cuasicuantitativa.
2. De acuerdo con los datos de la Tabla 1, la frecuencia relativa (p_i) asociada al intervalo 105-109 es: **A) 0,06**; B) 6; C) 0,6.
3. La amplitud del intervalo de la distribución de frecuencias de la Tabla 1 es: A) 6; B) 5,5; **C) 5**.
4. Según las propiedades de la media aritmética, si tenemos 30 puntuaciones cuya media es 15, si sumamos un 5 a cada una de las puntuaciones, la media de las nuevas puntuaciones es: A) 15; **B) 20**; C) 30.
5. Atendiendo a los datos de la Tabla 1, la mediana de la distribución es igual a: **A) 95,333**; B) 91,525; C) 108,452.
6. Con los datos de la Tabla 1, ¿cuál es el valor de la moda? A) 97; B) 102; **C) 92**.
7. Atendiendo a los datos de la Tabla 1, ¿es exactamente simétrica la distribución? A) sí; B) no, tiene una asimetría negativa; **C) no, tiene una asimetría positiva.**

Psicotestuned.info

8. Con los datos de la Tabla 1, la varianza de la variable resiliencia familiar en la muestra de madres es: **A) 25,75**; B) 5,07 C) 66,30.
9. La representación de la Figura 1 se denomina: A) histograma acumulado; B) diagrama de barras adosadas; **C) diagrama de barras apiladas.**
10. Según la información de la Figura 1, el número de profesores varones de universidad es: A) 56; **B) 112**; C) 44.
11. Para calcular el grado de asociación entre las variables "etapa educativa" y "sexo", recogidas en el Figura 1, utilizaremos: **A) el coeficiente *C* de contingencia**; B) el coeficiente de correlación lineal de Pearson; C) el coeficiente phi (φ).
12. Teniendo en cuenta los datos de la Tabla 2 y que las variables X e Y tienen unidades de medida diferentes, ¿cuál de las dos variables presenta mayor variabilidad? A) X; **B) Y**; C) Ambas tienen la misma variabilidad.
13. Atendiendo a los datos de la Tabla 2, ¿cuánto vale el coeficiente de correlación de Pearson entre X e Y? A) 0,1; **B) 0,8**; C) 0,9.
14. Con los datos de la Tabla 2, ¿qué puntuación directa pronosticaremos a una persona con una puntuación en X igual a 95? A) 0,1; B) 9,5; **C) 7,5.**
15. Según los datos de la Figura 1, elegida una persona al azar ha resultado ser profesor de infantil, ¿cuál es la probabilidad de que sea mujer? **A) 0,71**; B) 0,50; C) 0,29.
16. En un centro especializado infantil, un 33% de los niños recibe una intervención grupal e individual para mejorar la comprensión lectora (grupo 1), mientras que el 67% restante recibe sólo intervención grupal (grupo 2). Tras la aplicación de las respectivas intervenciones, siguen mostrando problemas de comprensión lectora el 20% de los niños del grupo 1 y el 60% de los niños del grupo 2. Se escoge un niño al azar, y se observa que todavía tiene problemas de comprensión lectora. ¿Cuál es la probabilidad de que este niño se encuentre en el grupo 2? A) 0,4667; **B) 0,8571**; C) 0,6667.
17. En una prueba diagnóstica denominamos "sensibilidad" a: **A) la probabilidad de discriminar a los verdaderos positivos;** B) la probabilidad de detectar a los verdaderos negativos; C) la probabilidad de que todos los que den positivo tengan un trastorno.
18. Según los datos de la Figura 2, ¿cuál es la media o esperanza matemática de X? A) 1,2; B) 2,9; **C) 2,1.**
19. Siguiendo los datos de la Figura 2, la probabilidad de sacar una puntuación igual o menor que 2 es: **A) 0,7**; B) 0,5; C) 0,9
20. Si una variable se distribuye normalmente, ¿cuál es la probabilidad de obtener valores mayores que z = 0,58? A) 0,7190; B) 0,4320; **C) 0,2810**.
21. En una distribución X^2 con 9 grados de libertad, el valor 14,6837 se corresponde con: **A) el percentil 90;** B) el percentil 9; C) el decil 90.
22. Indica cuál de las siguientes afirmaciones <u>NO</u> es correcta. En el muestreo aleatorio estratificado: A) existen estratos o grupos heterogéneos entre sí; **B) hay gran heterogeneidad dentro del estrato**; C) cada estrato funciona de forma independiente.
23. Se sabe que en una determinada población el 60% de las personas utiliza el coche para los desplazamientos. Si se extrae una muestra aleatoria de 15 personas de dicha población, ¿cuál sería el valor aproximado del error típico de la proporción de personas que utilizan el coche? **A) 0,127**; B) 0,022; C) 0,407.
24. Se selecciona una muestra aleatoria de 90 sujetos de una población de jóvenes con TDAH de entre 20 y 25 años. Un tercio de ellos presenta baja autoestima. Con α = 0,01 el intervalo de confianza para estimar la proporción de sujetos con baja autoestima es: A) (0,23 y 0,43); **B) (0,20 y 0,46)**; C) (0,23 y 0,46).
25. En una muestra de psicólogos del ámbito de la salud seleccionada al azar, los resultados señalan que el 45% no está de acuerdo con el nuevo plan de intervención que propone el Ministerio de Sanidad. Asumiendo población infinita, ¿cuál es el tamaño muestral necesario para que el error máximo asociado a un intervalo de confianza del 95% sea 0,03?: A) 1333; B) 1481; **C) 1057.**

Psicotestmed.info

SOLUCIONES

1. **B) cuantitativa continua**
2. **A) 0,06**

$$p_i = \frac{n_i}{n} = \frac{6}{100} = 0{,}06$$

3. **C) 5**

La amplitud del intervalo es la diferencia entre el límite superior exacto y límite inferior exacto. Haciendo el cálculo con el intervalo 110-114 tenemos: 109,5 – 114,5 = 5

4. **B) 20**

$$\overline{X} = 15 \quad a = 5 \quad \rightarrow \quad \overline{Y} = 15 + 5 = 20$$

5. **A) 95,333**

Calculamos las frecuencias acumuladas n_a a partir de las frecuencias absolutas n_i:

X	n_i	n_a
110-114	2	100
105-109	6	98
100-104	17	92
95-99	30	75
90-94	45	45

El intervalo crítico para $Md = Q_2 = P_{50}$ es 95-99, ya que $\frac{n \cdot k}{100} = \frac{100 \times 50}{100} = 50$

$$Md = Q_2 = P_{50} = L_i + \left(\frac{\frac{n \cdot k}{100} - n_d}{n_c}\right) \cdot I = 94{,}5 + \left(\frac{\frac{100 \cdot 50}{100} - 45}{30}\right) \cdot 5 = 95{,}333$$

6. **C) 92**

La moda se corresponde con el valor de la variable con mayor frecuencia absoluta. En este caso, el intervalo con la frecuencia más alta es (90-94) con *ni* = 45, por lo que la moda es el punto medio del intervalo que es igual a 92.

7. **C) no, tiene una asimetría positiva.**

$$A_p = \frac{\bar{X} - Mo}{S_x} = \frac{96.5 - 92}{5{,}074} = 0{,}8868 \cong 0{,}89$$

8. **A) 25,75**

X	n_i	X_i	n_i	n_iX_i	X_i^2	$n_iX_i^2$
110-114	2	112	100	224	12544	25088
105-109	6	107	98	642	11449	68694
100-104	17	102	92	1734	10404	176868
95-99	30	97	75	2910	9409	282270
90-94	45	92	45	4140	8464	380880
Σ	100			9650		933800

Psicotestuned.info

$$\bar{X} = \frac{\sum n_i X_i}{n} = \frac{9650}{100} = 96{,}5$$

$$S_x^2 = \frac{\sum n_i X_i^2}{n} - \bar{X}^2 = \frac{933800}{100} - 96{,}5^2 = 25{,}75$$

9. **C) diagrama de barras apiladas.**

10. **B) 112**

11. **A) el coeficiente *C* de contingencia**

12. **B) Y**

$$CV_X = \frac{16}{100} \cdot 100 = 16$$

$$CV_Y = \frac{2}{8} \cdot 100 = 25$$

13. **B) 0,8**

$$Y' = bX + a$$

$$b = r_{XY}\frac{S_Y}{S_X} \Rightarrow 0{,}1 = r_{XY}\frac{2}{16} \Rightarrow 0{,}1 = r_{XY} \cdot 0{,}125 \Rightarrow r_{XY} = \frac{0{,}1}{0{,}125} = 0{,}8$$

14. **C) 7,5**

$$Y' = 0{,}1\,X - 2 = 0{,}1 \cdot 95 - 2 = 9{,}5 - 2 = 7{,}5$$

15. **A) 0,71**

	Varón (V)	Mujer (M)	
Infantil (I)	58	142	200
Primaria (P)	106	94	200
Secundaria (S)	46	154	200
Universidad (U)	112	88	200
	322	478	800

$$P(M/I) = \frac{\frac{142}{800}}{\frac{200}{800}} = 0{,}71$$

16. **B)** Esta pregunta es similar al Ejemplo 6.11 C) (página 280 del manual)

La probabilidad de que niño se encuentre en el grupo 2 (G2) una vez que se ha observado que continúa teniendo problemas de comprensión lectora (PCL) se obtiene mediante el teorema de Bayes:

$$P(PCL) = P(G1) \cdot P(PCL|G1) + P(G2) \cdot P(PCL|G2) = 0{,}33 \cdot 0{,}20 + 0{,}67 \cdot 0{,}60 = 0{,}0667 + 0{,}40 = 0{,}4667$$

$$P(G2|PCL) = \frac{P(G2) \cdot P(PCL|G2)}{P(PCL)} = \frac{0{,}67 \cdot 0{,}60}{0{,}4667} = 0{,}8571$$

17. **A) la probabilidad de discriminar a los verdaderos positivos**

18. **C) 2,1**

x	***f(x)***	***x f(x)***
1	0,2	0,2
2	0,5	1,0
3	0,3	0,9
	1	2,1

$$\mu = \sum x \cdot f(x) = 2{,}1$$

Psicotestuned.info

19. **A) 0,7**

F(2) = P (X ≤ 2) = P (X = 1) + P (X = 2) = f(1) + f(2) = 0,2 + 0,5 = 0,7

20. **C) 0,2810**.

Se mira en la tabla IV y se ve que la puntuación típica z = 0,58 deja por debajo de sí una probabilidad de 0,7190. Como lo que nos preguntan es la probabilidad que queda por encima, y habría que calcular: 1 – 0,7190 = 0,2810

21. **A) el percentil 90**

Ver Tabla V de X^2 con 9 g.l.

22. **B) hay gran heterogeneidad dentro del estrato**

23. **A) 0,127**

$$\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0{,}60(0{,}40)}{15}} = 0{,}127$$

24. **B)** (**0,20 y 0,46)**

La variable aleatoria es una variable del tipo Bernoulli. Se conoce la proporción π en la población (o el valor *P* en la muestra). Respecto al tamaño de la muestra es suficientemente grande como para usar la aproximación a la normal, por tanto:

$$P \rightarrow N\left(\pi, \sqrt{\frac{\pi(1-\pi)}{n}}\right); \text{ P= 30/90= 0,33}$$

$$L_{inf} = P - \left|Z\alpha_{/2}\right| \pi, \sqrt{\frac{P(1-P)}{n}} = 0{,}33 - 2{,}58 \cdot \sqrt{\frac{0{,}33(1-0{,}33)}{90}} = 0{,}33 - 0{,}13 = 0{,}20$$

$$L_{sup} = P + \left|Z\alpha_{/2}\right| \pi, \sqrt{\frac{P(1-P)}{n}} = 0{,}33 + 2{,}58 \cdot \sqrt{\frac{0{,}33(1-0{,}33)}{90}} = 0{,}33 + 0{,}13 = 0{,}46$$

25. **C) 1057**

$p = 0{,}45$; $E_{max} = 0{,}03$

$$n = \frac{Z^2_{\alpha/2}\, P(1-P)}{E^2_{max}} = \frac{1{,}96^2 \cdot 0{,}45 \cdot 0{,}55}{0{,}03^2} = 1056{,}44 \gg 1057$$

Psicotestuned.info

INTRODUCCIÓN AL ANÁLISIS DE DATOS
Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Material: **Adenda (Formulario y Tablas) y calculadora (no programable)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Tabla 1. Nivel de ansiedad (X) de 260 adultos y sus frecuencias absolutas (n_i).

Nivel de ansiedad	X	n_i
Muy Bajo	0	30
Bajo	1	50
Medio	2	90
Alto	3	80
Muy alto	4	10

Figura 1. Diagrama de caja para la variable inteligencia de dos grupos de participantes.

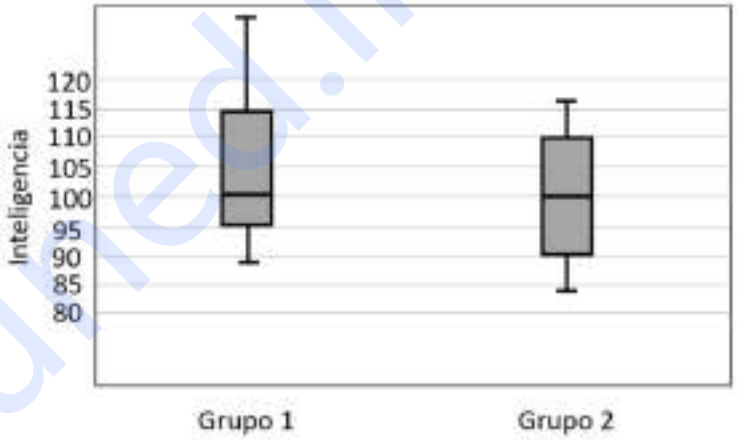


Tabla 2. Datos sobre la obtención de un título universitario (sí, no) y la satisfacción con el trabajo (alta, baja) en una muestra.

	Satisfacción alta	Satisfacción baja
Con título universitario	100	30
Sin título universitario	100	300

1. Un grupo de investigadores de la UNED estudia la comprensión lectora de los estudiantes de la Facultad de Psicología. En su investigación, se evaluó a 2000 estudiantes. ¿Cuál es la población objeto de estudio? A) Los 2000 estudiantes; **B) Los estudiantes de Psicología de la UNED**; C) Los estudiantes de la UNED.
2. Las variables con escala de medida nominal: A) sólo admiten dos valores posibles; B) admiten relaciones de orden; **C) admiten dos o más valores**.
3. En un mismo test, el grupo A obtiene una media de 15 mientras que el grupo B obtiene una media de 50. ¿Qué grupo muestra una mayor heterogeneidad en sus puntuaciones? A) El grupo A; B) El grupo B; **C) No se puede saber.**
4. ¿Cuál es la media de la variable X de la Tabla 1? **A) 1,96;** B) 2,22; C) 2,94.
5. ¿Cuál es el primer cuartil de la distribución de la variable X de la Tabla 1? A) 0,75; **B) 1,2;** C) 2,5.
6. El quinto decil de la distribución de una variable es igual a: A) P_{05}; B) Q_5; **C) La mediana**.
7. ¿Cuál es la desviación típica de la variable X de la Tabla 1? **A) 1,06**; B) 2,24; C) 4,96.

Psicotestuned.info

8. ¿Qué afirmación sobre la variable inteligencia de la Figura 1 es FALSA? A) La mediana del grupo 1 es igual a la mediana del grupo 2; B) No hay valores atípicos; **C) La amplitud intercuartil es diferente en ambos grupos.**
9. ¿Cuál es la forma de la distribución de la variable inteligencia para el grupo 1 de la Figura 1? **A) Asimétrica positiva**; B) Asimétrica negativa; C) Simétrica positiva.
10. ¿Qué valor aproximado toma el estadístico X^2 que permite relacionar las variables de la Tabla 2? A) 37,57; B) 87,57; **C) 112,57.**
11. Aproximadamente, ¿qué valor máximo podría tomar el coeficiente de contingencia con los datos de la Tabla 2? **A) 0,707;** B) 0,821; C) 0,978.
12. ¿Cuál es la diferencia que existe entre el coeficiente phi (φ) y el coeficiente χ^2 de independencia? A) El rango de valores que pueden adoptar ambos coeficientes; B) El número de niveles de las variables que se pueden analizar; **C) Ambas opciones son correctas.**
13. Cov(A, B) = 9,27; Cov(C, D) = 12,57. ¿Qué grupo de variables muestra una correlación más fuerte? A) Variables A y B; B) Variables C y D; **C) No se puede saber.**
14. En una investigación, donde $\bar{A} = 100$ y $\bar{B} = 20$, se estandarizan ambas variables predictoras (A y B) y se obtiene esta ecuación de regresión lineal múltiple: $Y' = 0{,}37 + 0{,}89 * A + 1{,}21 * B$. ¿Cuál será el pronóstico (Y') de un participante que tiene una puntuación media en ambas variables? **A) 0,37;** B) 113,57; C) Falta información para estimar el pronóstico.
15. El coeficiente de determinación de la relación existente entre la variable X y la variable Y vale 0,40. ¿Qué valor toma la correlación entre ambas variables? A) 0,16; B) 0,40; **C) -0,63.**
16. Un nuevo test de diagnóstico de TDAH tiene una alta probabilidad de detectar verdaderos negativos. ¿Qué caracteriza a este test? A) Alto valor predictivo positivo; **B) Alta especificidad;** C) Alta sensibilidad.
17. De acuerdo a la Tabla 2, ¿cuál es, aproximadamente, la probabilidad de tener un título universitario? **A) 0,245;** B) 0,333; C) 0,491.
18. Z es una variable aleatoria discreta con valores equiprobables: 0, 1, 2, y 3. ¿Cuál es su media? **A) 1,5;** B) 2; C) 2,5.
19. ~~Z es una variable aleatoria discreta con valores equiprobables: 0, 1, 2, y 3. ¿Cuál es su desviación típica? A) 1; **B) 1,25;** C) 1,50.~~
20. La variable inteligencia se distribuye normalmente con media 100 y desviación típica de 15. ¿Cuál es, aproximadamente, la probabilidad de que una persona obtenga una puntuación igual o superior a 120? A) 0,0092; **B) 0,0918;** C) 0,1918.
21. ¿Qué valor del estadístico χ^2_{20} deja por debajo de sí mismo una probabilidad de 0,95? A) 7,43; B) 10,85; **C) 31,41.**
22. Una variable se distribuye normalmente en una población con media μ y desviación típica 25. Se seleccionan distintas muestras de 380 participantes cada una. ¿Cuál es, aproximadamente, el error típico de la distribución muestral de las medias? A) 0,98; **B) 1,28;** C) 1,58.
23. El 70% de la población de profesionales sanitarios de España afirma sufrir burnout. ¿Cuál es la probabilidad de extraer una muestra de 500 sanitarios en la que solo la mitad o menos tengan burnout? **A) Aproximadamente 0;** B) Aproximadamente 0,0248; C) Aproximadamente 0,0498.
24. Un grupo de 400 estudiantes de una asignatura de la UNED son seleccionados a través de muestreo aleatorio simple y se ve que han obtenido una nota media de 6. La variable se distribuye normalmente *N*(μ, 2). ¿Cuál es el intervalo de confianza para μ con una confianza del 99%? **A) (5,742; 6,258);** B) (5,042; 6,958); C) (5,942; 6,058).
25. En cuanto a las propiedades de los estimadores, ¿qué nos indica la expresión $E(\hat{\theta}) = \theta$? A) Eficiencia; B) Consistencia; **C) Carencia de sesgo.**

Protestmedi.info

RESPUESTAS

1. B.

2. C.

3.C. La media solo informa de la tendencia central, no de la variabilidad (homogeneidad o heterogeneidad) de la variable.

4. A. $\bar{X} = \frac{\sum n_i X_i}{n} = \frac{0*30+1*50+2*90+3*80+4*10}{260} \simeq 1{,}96.$

5. B. El primer cuartil es el percentil 25. Sabemos que $\frac{n*25}{100} = \frac{260*25}{100} = 65$. Por lo tanto, el intervalo crítico es [0,5-1,5] con n_a=80, n_c=50 y n_d=30. Solamente tenemos que aplicar la fórmula:

$$P_{25} = L_i + \left((\frac{n*k}{100} - n_d)/n_c \right) * I = 0{,}5 + \left(\frac{\frac{260*25}{100} - 30}{50} \right) * 1 = 0{,}5 + 0{,}7 = 1{,}2.$$

6. C.

7. A. Por el ejercicio 4, sabemos que la media de la variable es 1,96. Entonces procedemos a completar la Tabla 1:

Nivel de ansiedad	**X_i**	**n_i**	**X_i^2**	**$n_i X_i^2$**
Muy Bajo	0	30	0	0
Bajo	1	50	1	50
Medio	2	90	4	360
Alto	3	80	9	720
Muy alto	4	10	16	160
				1290

Ya podemos aplicar la fórmula: $S_X^2 = \frac{\sum n_i X_i^2}{n} - (\bar{X}^2)$; $S_X^2 = \frac{1290}{260} - (1{,}96^2) \simeq 1{,}120$. Entonces, podemos calcular la desviación típica aplicando la raíz cuadrada de la varianza: $S_X = \sqrt{S_X^2} = \sqrt{1{,}120} \simeq 1{,}058$.

8. C.

9. A.

10. C. Completamos la Tabla 2:

	Satisfacción alta	Satisfacción baja	
Con título universitario	100 (49,06)	30 (80,94)	130
Sin título universitario	100 (150,94)	300 (249,06)	400
	200	330	530

Aplicamos la fórmula del cálculo de $X^2 = \sum\sum \frac{(n_e - n_t)^2}{n_t} = \frac{(100-49{,}06)^2}{49{,}06} + \frac{(30-80{,}94)^2}{80{,}94} + \frac{(100-150{,}94)^2}{150{,}94} + \frac{(300-249{,}06)^2}{249{,}06} \simeq 112{,}56.$

11. A. Solamente necesitamos saber el número (k) de filas y columnas. Aplicamos la fórmula: $C_{max} = \sqrt{\frac{k-1}{k}} = \sqrt{\frac{2-1}{2}} \simeq 0{,}707.$

Psicotesunedi.info

12. C.

13. C. No se puede saber porque no se tienen datos de las desviaciones típicas de las variables.

14. A. Dado que ambas variables predictoras, A y B, fueron estandarizadas antes del análisis, sus medias son cero. Entonces el pronóstico de un participante con puntuaciones medias será igual a: $Y' = 0{,}37 + 0{,}89 * A + 1{,}21 * B = 0{,}37 + 0{,}89 * 0 + 1{,}21 * 0 = 0{,}37$.

15. Dado que el coeficiente de determinación es igual a r_{XY}^2, el único valor que da como resultado 0,40 es -0,63.

16. B. Ver ejercicio 6.17 del manual.

17. A. P=130/530=0,245.

18. A. Dado que son valores equiprobables:

y	0	1	2	3
f(y)	0,25	0,25	0,25	0,25

Por tanto: $\mu_y = 0 * 0{,}25 + 1 * 0{,}25 + 2 * 0{,}25 + 3 * 0{,}25 = 1{,}50$.

19. Ninguna alternativa de respuesta es correcta. Por el ejercicio anterior, conocemos la media de la variable. Podemos aplicar la fórmula: $\sigma^2 = V(X) = \sum(x-\mu)^2 f(x) = (0-1{,}5)^2\, 0{,}25 + (1-1{,}5)^2\, 0{,}25 + (2-1{,}5)^2\, 0{,}25 + (3-1{,}5)^2\, 0{,}25 = 1{,}25$; $\sigma = 1{,}12$

20. B. $z = \frac{120-100}{15} = 1{,}333$; Tabla IV: P(Z > 1,33) = 1 – P(Z ≤ 1,33) = 1 - 0,9082 = 0,0918.

21. C.

22. B. $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{25}{\sqrt{380}} \simeq 1{,}28$.

23. A. Reportar o no burnout es una variable de Bernoulli con dos posibilidades: éxito (presentar burnout) o fracaso (no presentar burnout). Siendo el tamaño muestral n=500 y π=0,70. Dadas estas características, podemos aplicar la aproximación a la normal de la distribución de la probabilidad que tiene parámetros: π = 0,70 y $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0{,}70(1-0{,}70)}{500}} \simeq 0{,}0205$. Así, podemos aproximar a una normal con media 0,70 y desviación típica 0,0205. Entonces, $Z = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0{,}50-0{,}70}{0{,}0205} \simeq -9{,}76$.

Tabla IV: P(Z < -9,76) ≃ 0.

24. A. Dadas las características descritas, podemos asumir que la distribución muestral de la media es $N(\mu, \frac{\sigma}{\sqrt{n}})$ y el intervalo de confianza viene definido por estos límites inferior y superior:

$$L_{inferior} = 6 - 2{,}58 \cdot \left(\frac{2}{20}\right) = 6 - 0{,}258 = 5{,}742$$

$$L_{superior} = 6 + 2{,}58 \cdot \left(\frac{2}{20}\right) = 6 + 0{,}258 = 6{,}258$$

25. C. Página 434.

Psicotestuned.info

FORMULARIO Y TABLAS Introducción al Análisis de Datos: Aplicaciones en la Psicología y las Ciencias de la Salud FE DE ERRATAS 1ª Edición (2017)			
Página	**Localización**	**Dice**	**Debería decir**
11	En la segunda fórmula de Función de distribución:	$P(x_1 \leq x \leq x_2) = F(x_2) - F(x_1)$	$P(x_1 < x \leq x_2) = F(x_2) - F(x_1)$
16	En la fórmula de la varianza de F de Fisher-Snedecor	$\sigma^2 = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-4)^2(n_2-2)^2}$	$\sigma^2 = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-4)(n_2-2)^2}$
18	En la primera línea de texto	La variable X no se distribuye normalmente y $(n \geq 30)$ σ conocida:	La variable X no se distribuye normalmente y $(n \geq 30)$ (Teorema del límite central) σ conocida:
19	En las fórmulas de la aproximación a la normal de la distribución muestral de la varianza para muestras grandes	$Z = \frac{\left(\frac{nS_X{}^2}{\sigma^2}\right) - (n-1)}{\sqrt{(2n-1)}}$	$Z = \frac{\left(\frac{nS_X{}^2}{\sigma^2}\right) - (n-1)}{\sqrt{2(n-1)}}$
		$Z = \frac{\left(\frac{(n-1)S_{n-1}^2}{\sigma^2}\right) - (n-1)}{\sqrt{(2n-1)}}$	$Z = \frac{\left(\frac{(n-1)S_{n-1}^2}{\sigma^2}\right) - (n-1)}{\sqrt{2(n-1)}}$
23	En tamaño muestral para el parámetro proporción Poblaciones infinitas:	$E_{max} = Z\alpha/_2\sqrt{\frac{P(1-P)}{E^2{}_{max}}}$	$E_{max} = Z\alpha/_2\sqrt{\frac{P(1-P)}{n}}$
41	En la figura de la distribución t de Student, en el eje horizontal	1,383	1,372

psicotestuned.info

Gracias a Javi por compartir las plantillas de corección.

INTRODUCCIÓN AL ANÁLISIS DE DATOS
Código asignatura: 62011037
EXAMEN TIPO TEST MODELO B DURACION: 2 HORAS

Material: **Adenda (Formulario y Tablas) y calculadora (no programable)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados. La pregunta 26 es de reserva y solamente se utilizará si se anula alguna pregunta del examen.

Tabla 1. Resultados obtenidos en un test de autoconcepto (X) aplicado a 40 adolescentes.

X	n_i
26-28	7
23-25	11
20-22	9
17-19	7
14-16	6

Figura 1. Resultados de la mejora o no del colesterol tras la administración de tres fármacos distintos en un grupo de 150 personas hipertensas. Los participantes fueron asignados aleatoriamente a uno de los tres tratamientos.

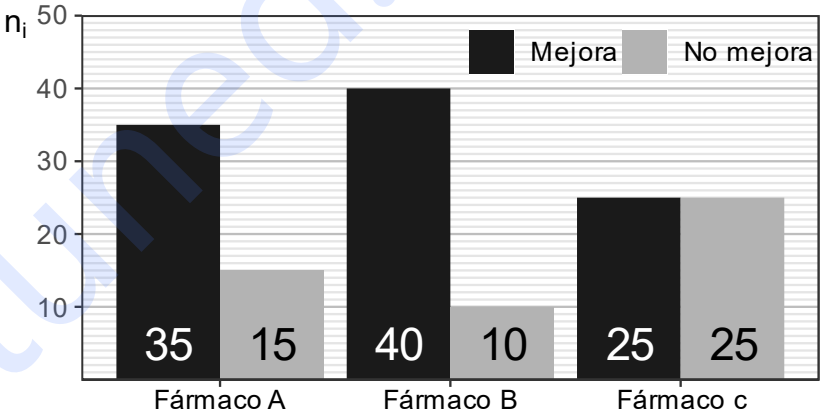


Tabla 2. Para dos variables cuantitativas, X e Y, disponemos de los siguientes datos:

X	Y	XY
$\bar{X}=8$	$\bar{Y}=100$	$Cov(X,Y)=S_{XY}=18$
$S_X=2$	$S_Y=10$	

Tabla 3. Variables X e Y, y su tipo de distribución.

Variable	Distribución	
X	$N(20,5)$	Normal con media 20 y desviación típica 5
Y	X^2_{40}	Ji-cuadrado con 40 grados de libertad

1. La variable "Tipo de Fármaco" de la Figura 1 es: A) cuasicuantitativa; B) cuantitativa; C) politómica.
2. La amplitud del intervalo de los datos de la Tabla 1 es: A) 3; B) 2,5; C) 2.
3. ¿Cuál es el gráfico más adecuado para representar los datos de la Tabla 1? A) Diagrama de barras; B) Diagrama de líneas; C) Histograma.
4. La media aritmética de la distribución de X en la Tabla 1 es: A) 21,45; B) 20,83; C) 27,48.
5. Atendiendo a los datos de la Tabla 1, el valor del Percentil 75 es: A) 30,25; B) 20,04; C) 24,68.
6. La moda es un índice de: A) posición; B) tendencia central; C) variabilidad.
7. Con los datos de la Tabla 1, la amplitud intercuartil es: A) 8,21; B) 6,47; C) 4,68.
8. Según los datos de la Figura 1, asumiendo que ambas variables son independientes, ¿cuál es la frecuencia teórica de mejorar y recibir el fármaco B? A) 16,76; B) 33,33; C) 40.
9. Con los datos de la Figura 1, hemos obtenido un valor de X^2 igual a 10,5. El coeficiente *C* de Contingencia tiene un valor que está comprendido entre: A) 0,1 y 0,4; B) 0,5 y 0,8; C) 0,9 y 1,1.
10. Teniendo en cuenta la Tabla 2, la recta de regresión que permite pronosticar la variable *Y* en función de la *variable X* es: A) Y' = 64 + 4,5X; B) Y' = 70 - 0,5X; C) Y' = 35 + 6X.

psicotestuned.info

11. Se quiere observar la relación entre la posición en un ranking educativo (X) y la posición en un ranking socioeconómico (Y). Para ello, se obtiene la información sobre ambas variables de 10 personas. Los rangos de X son {4, 3, 8, 5, 6, 2, 9, 7, 1, 10}, y los rangos de Y son {6, 4, 2, 7, 8, 9, 3, 5, 1, 10}. ¿Qué procedimiento podría usarse para calcular la relación entre ambas variables? A) Solamente el coeficiente de correlación de Pearson; B) Solamente el coeficiente de correlación de Spearman; C) El coeficiente de correlación tanto de Pearson como de Spearman.
12. Con los datos de la pregunta 11, ¿qué valor toma el coeficiente de correlación entre X e Y? A) -0,16; B) 0,16; C) 0,21.
13. En un grupo de 100 estudiantes, se está realizando un análisis sobre el interés en dos actividades extracurriculares: asistir a clases de música (M) y participar en un club de debate (D). Se sabe que el 60% de los estudiantes asisten a clases de música, el 40% de los estudiantes participan en el club de debate, y el 15% de los estudiantes asisten tanto a clases de música como participan en el club de debate. ¿Cuál es la probabilidad de que un estudiante asista a clases de música, o participe en el club de debate, o realice ambas actividades? A) 0,85; B) 0,70; C) 0,15.
14. Teniendo en cuenta la representación de la Figura 1, elegida una persona al azar ha resultado que se le administró el fármaco A. ¿Cuál es la probabilidad aproximada de que mejore? A) 0,23; B) 0,50; C) 0,70.
15. Con los datos de la Figura 1, si elegimos una persona al azar ¿cuál es la probabilidad aproximada de que presente mejoría en su colesterol?: A) 0,33; B) 0,67; C) 0,70.
16. Sea X una variable aleatoria discreta con valores 1, 2, 3 y 4. Si los cuatro valores de X son equiprobables, su media es: A) 2,50; B) 1,50; C) 0,25.
17. En el lanzamiento de un dado una sola vez se ha definido la variable X “obtener número impar”. La variable X se distribuye según: A) Bernoulli con parámetro p = 1/2; B) Binomial con parámetro p = 1/6; C) Binomial con parámetro p = 1/2.
18. La función de distribución de una variable aleatoria discreta: A) asocia a cada valor de la variable la probabilidad de que ésta adopte ese valor; B) asocia a cada valor de la variable la probabilidad de que ésta adopte ese valor o cualquier otro inferior; C) asocia a cada valor de la variable la probabilidad de que ésta adopte cualquier valor inferior a éste.
19. Teniendo en cuenta los datos de la Tabla 3, P (X > 25) es aproximadamente: A) 0,2500; B) 0,8413; C) 0,1587.
20. Con los datos de la Tabla 3, para la variable Y, el valor 51,8051 es el percentil: A) 10; B) 50; C) 90.
21. La distribución F: A) es asimétrica negativa; B) puede adoptar cualquier valor entre -∞ y +∞; C) posee la propiedad recíproca.
22. Calcule el error típico de la proporción sabiendo que las muestras (extraídas por muestreo aleatorio simple) tienen un tamaño de 80 y el valor esperado es 0,8. A) Se puede calcular y su valor es aproximadamente 0,045; B) no se puede calcular dado que falta el nivel de significación; C) Se puede calcular y su valor es aproximadamente 0,022.
23. Sea X una variable con distribución normal en la población con N(150; 8). Se extrae una muestra aleatoria simple de tamaño 15, ¿cuál es la probabilidad aproximada de que su media sea menor o igual a 145? A) 0,0078; B) -2,4206; C) 0,9922.
24. Se quiere estimar la media de la variable bienestar físico en un grupo de personas que practican pilates. Se extrae una muestra de 74 personas, de una población de 2500. Sabiendo que el bienestar físico se distribuye N (μ; 17), ¿cuál es el valor aproximado del error máximo que se comete con este tamaño muestral con un nivel de significación α = 0,05? A) 1,273; B) 3,814; C) 5,016.
25. Sabemos que el tiempo promedio que los estudiantes universitarios dedican a estudiar por día se distribuye normalmente con una desviación estándar de σ=3 horas. En una muestra aleatoria de 64 estudiantes, se observa una media muestral de 5 horas. ¿Entre qué valores estimaremos que se encuentra la media del tiempo de estudio en la población con un nivel de confianza del 99%? A) 4,03 y 5,97; B) 4,03 y -5,97; C) -4,03 y 5,97.

Picotestuned.info

Pregunta de Reserva:

26. Se hace un muestreo aleatorio simple de 60 estudiantes de la UNED. Un tercio ha realizado una PEC. ¿Cuál es el intervalo de confianza (α=0,05) para estimar la proporción de alumnos que han realizado esta PEC? A) (0,211; 0,449); B) (0,211; 0,614); C) (0,032; 0,614).

SOLUCIONES

1. C
2. A

La diferencia de los límites exactos del primer intervalo es 25,5 – 28,5 = 3

3. C (página 32)
4. A

X	n_i	X_i	n_iX_i	X_i^2	$n_iX_i^2$
26-28	7	27	189	729	5103
23-25	11	24	264	576	6336
20-22	9	21	189	441	3969
17-19	7	18	126	324	2268
14-16	6	15	90	225	1350
Σ	40		858		19026

$$\overline{X} = \frac{\sum n_i X_i}{\sum n_i} = \frac{\sum n_i X_i}{n} = 858/40 = 21{,}45$$

5. C

El valor del Percentil 75 es 24,68.

nk/100= 40 x 75/100= 30; I= 25,5 - 22,5 = 3

X	n_i	n_a
26-28	7	40
23-25	**11**	**33**
20-22	9	**22**
17-19	7	13
14-16	6	6
Σ	40	

$$P_{75} = L_i + \left(\frac{\frac{n \cdot k}{100} - n_d}{n_c} \right) \cdot I = 22{,}5 + \left(\frac{30-22}{11}\right) 3 = 24{,}68$$

6. B
7. B

La amplitud intercuartil es 6,47.

$A_{IQ} = Q_3 - Q_1 = P_{75} - P_{25} = 24{,}68 - P_{25}$

nk/100= 40 x 25/100= 10; I= 3

Psicotestuned.info

X	n_i	n_a
26-28	7	40
23-25	11	33
20-22	9	22
17-19	**7**	**13**
14-16	6	**6**
Σ	40	

$P_{25} = 16{,}5 + \left(\frac{10-6}{7}\right) 3 = 18{,}21$

$A_{IQ} = 24{,}68 - 18{,}21 = 6{,}47$

8. B

	Fármaco A	Fármaco B	Fármaco C	
Mejora	35	40	25	100
No mejora	15	10	25	50

$$\text{Frecuencia teórica} = n_t = \frac{\text{total fila x total columna}}{\text{n}} = \frac{100 \text{ x } 50}{150} = 33.33$$

9. A

0,1 y 0,4

$$C = \sqrt{\frac{X^2}{X^2 + n}} = \sqrt{\frac{10{,}5}{10{,}5 + 150}} = \sqrt{0{,}06542} = 0{,}25$$

10. A

$$Y' = bX + a$$

$$b = r_{XY}\frac{S_Y}{S_X} = 0{,}9\frac{10}{2} = 4{,}5 \qquad a = \overline{Y} - b\overline{X} = 100 - (4{,}5 \cdot 8) = 64$$

$$Y' = 4{,}5X + 64$$

11. C)

Página 174 y 207. Como no hay valores empate, los resultados tienen que ser similares aplicando Correlación tanto Correlación de Pearson como Correlación de Spearman

Psicotestuned.info

12. B)

Página 171

Rango X	Rango Y	d	d^2
4	6	-2	4
3	4	1	1
8	2	6	36
5	7	-2	4
6	8	-2	4
2	9	-7	49
9	3	6	36
7	5	2	4
1	1	0	0
10	10	0	0
			Σ 138

$$r_s = 1 - \frac{6\sum d_i^2}{n\left(n^2 - 1\right)}$$

$$r_s = 1 - \frac{6\ x\ 138}{10\ (100-1)} = 0{,}16$$

13. A

Página 265

La probabilidad de que un estudiante participe en al menos una de las dos actividades (unión de los sucesos M y D) se calcula usando la fórmula:

P (M) = 0,6; P(D) = 0,4; P(M∩D) = 0,15

P(M∪D) = P(M)+P(D)−P(M∩D)

Sustituyendo los valores: P(M∪D) = 0,6+0,4−0,15 =0,85

Esto significa que hay un 85% de probabilidad de que un estudiante participe en al menos una de las dos actividades.

Psicotestuned.info

14. C

	Fármaco A	Fármaco B	Fármaco C	
Mejora	35	40	25	100
No mejora	15	10	25	50
	50	50	50	150

$$P(M|A) = \frac{35}{50} = 0{,}7$$

15. B

	Fármaco A	Fármaco B	Fármaco C	
Mejora	35	40	25	100
No mejora	15	10	25	50
	50	50	50	150

$$P(M) = \frac{100}{150} = 0{,}67$$

16. A

La función de probabilidad es:

x	1	2	3	4
f(x)	0,25	0,25	0,25	0,25

Por tanto,

$\mu_x = \sum x \cdot f(x) = 1 \times 0{,}25 + 2 \times 0{,}25 + 3 \times 0{,}25 + 4 \times 0{,}25 = 2{,}50$

17. A

El espacio muestral es *E* = {1, 2, 3, 4, 5 y 6}

El éxito (1) sería sacar los números 1, 3 ó 5. Su probabilidad es *p* = 3/6 = ½ = 0,5

El fracaso (0) sería sacar 2, 4 ó 6. Cuya probabilidad q = 1- p = 0,5 = 1/2

Dado que se trata de un solo lanzamiento la distribución que se sigue es Bernoulli con parámetro *p* = 1/2

18. B

19. C

$$P\,(X \leq 25) = P\left(z \leq \frac{25-20}{5}\right) = P(z \leq 1) = 0{,}8413$$

P (x > 25) = 1 – 0,8413 = 0,1587

20. C (ver Tabla V). Percentil 90.

Psicotestuned.info

21. C (Ver página 361)

22. A

Ejercicio parecido al 9.13 del libro (página 423)

$\pi = E\,(p) = 0{,}8$;

$\sigma_{\,p} = \sqrt{\frac{\pi\,(1-\pi)}{n}} = \sqrt{\frac{0{,}8\,(1-0{,}8)}{80}} = 0{,}0447 \approx 0{,}045$

23. A

$$P(\ \bar{X}\ \leq 145) = P\,(z \leq \frac{145-150}{8/\sqrt{15}}) = P\,(z\ \leq \frac{-5}{2{,}0656}) = P(z \leq -2{,}42)$$

Ver Tabla III es 0,0078

24. B

$Emáx = \frac{z_{1-\alpha/2}\sigma}{\sqrt{n}} = \frac{1{,}96 \cdot 17}{\sqrt{74}} = \frac{33{,}32}{8{,}602} = 3{,}873$

Como la población es finita hay que corregir multiplicando por el factor de corrección $\sqrt{\frac{N-n}{N-1}}$

entonces $Emáx = 3{,}873\sqrt{\frac{2500-74}{2499}} = 3{,}814$

25. A Página 446.

Media muestral = 5

σ=3 (desviación estándar de la población),
n = 64

Z es el valor crítico de la distribución normal estándar para un nivel de confianza del 99%, que es aproximadamente Z = 2,58

L_{inf} = 5 – 2,58 *3/$\sqrt{64}$ = 4,03; L_{sup} = 5 + 2,58 *3/$\sqrt{64}$ = 5,97

Psicotestuned.info

Pregunta de reserva

26. La variable “realizar la PEC” es una variable de Bernoulli (hacer vs. no hacer la PEC). Se conoce π (que vale, aproximadamente, 0,33). El tamaño de la muestra es grande al ser mayor de 30. Por tanto, podemos aproximar el cálculo utilizando la normal:

$$P \rightarrow N\left(\pi, \sqrt{\frac{\pi(1-\pi)}{n}}\right)$$

P = 1/3 = 0,33.

$$L_{inf} = P - \left|z_{\frac{\alpha}{2}}\right| \sqrt{\frac{P(1-P)}{n}} = \frac{20}{60} - 1{,}96\sqrt{\frac{0{,}33(1-0{,}33)}{60}} = 0{,}211$$

$$L_{sup} = P + \left|z_{\frac{\alpha}{2}}\right| \sqrt{\frac{P(1-P)}{n}} = \frac{20}{60} + 1{,}96\sqrt{\frac{0{,}33(1-0{,}33)}{60}} = 0{,}449$$

Psicotestuned.info

Gracias a Ana por compartir las plantillas de corección.

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2025 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Material: **Adenda (Formulario y Tablas) y calculadora (no programable)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados. La pregunta 26 es de reserva y solamente se utilizará si se anula alguna pregunta del examen.

Tabla 1. Resultados obtenidos por 240 estudiantes de 3º de la ESO en un test de motivación (X) divididos en dos grupos.

Motivación (X)	Repetidores (n_i)	No repetidores (n_i)
80-99	20	32
60-79	22	40
40-59	33	23
20-39	30	15
1-19	15	10

Figura 1. Diagrama de caja de la variable "Horas de estudio al día" para repetidores y no repetidores

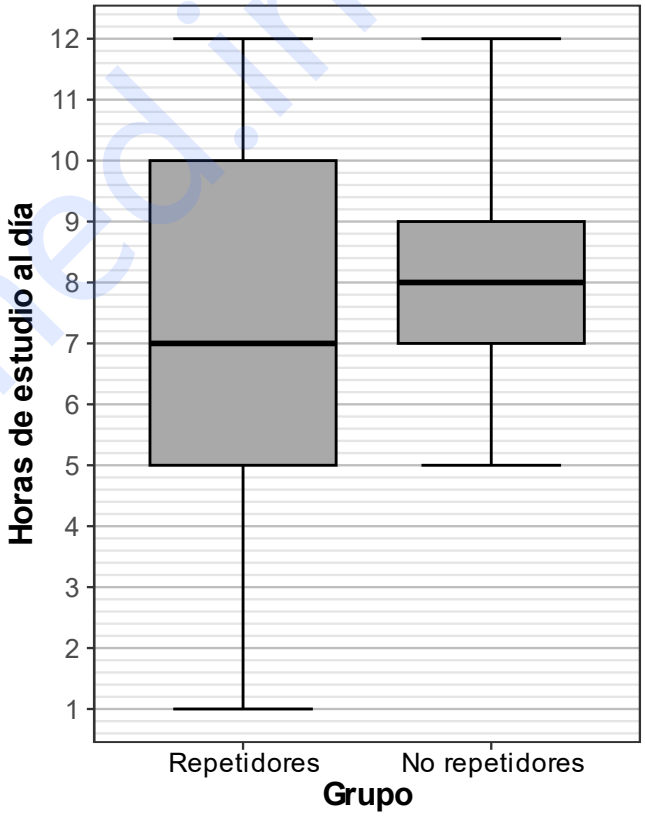


Tabla 2. Resultados de un estudio sobre la relación entre el contacto intergrupal y el prejuicio hacia los inmigrantes

	Prejuicio Bajo	Prejuicio Medio	Prejuicio Alto
Con contacto intergrupal	150	60	30
Sin contacto intergrupal	40	70	130

Tabla 3. Descriptivos de las variables bienestar (X) y rendimiento académico (Y) de 132 alumnos/as.

	Media	**Varianza**	**Covarianza**
X	6,88	1,95	1,771
Y	6,90	3,10	

1. ¿Qué nivel de medida presenta la variable "Motivación" de la Tabla 1?: A) Nominal;
B) Intervalo; C) Ordinal.

Psicotestuned.info

2. ¿Qué porcentaje del total de personas repetidoras puntuaron igual o menor a 59,5 en el test de motivación de la Tabla 1? A) 37,5%; B) 83,33%; C) 65%.
3. ¿Cuál es la moda en la distribución de frecuencias de la Tabla 1 para los no repetidores? A) 60-79; B) 10; C) 69,5.
4. La frecuencia relativa para la moda del grupo de los repetidores es: A) 0,275; B) 0,137; C) 0,125
5. Atendiendo a la Tabla 1, y sabiendo que las varianzas del grupo de repetidores y no repetidores son, respectivamente, 640,22 y 611,08, calcula el coeficiente de variación de repetidores y no repetidores. Con los resultados se puede concluir que: A) los datos de ambos grupos tienen una dispersión similar; B) los datos de repetidores presentan una mayor dispersión que los datos de no repetidores; C) los datos de no repetidores presentan una mayor dispersión que los datos de repetidores
6. En la Figura 1, el valor de la amplitud intercuartil de la variable "Horas de estudio al día" para repetidores y no repetidores es: A) 5 para repetidores y 2 para no repetidores; B) 7 para repetidores y 8 para no repetidores; C) 11 para repetidores y 7 no repetidores.
7. En la Figura 1, el diagrama de cajas y bigotes para la variable "horas de estudio al día" del grupo de no repetidores muestra una: A) distribución simétrica; B) distribución asimétrica negativa; C) distribución asimétrica positiva.
8. En la Tabla 2, ¿cuál es la probabilidad de tener un prejuicio bajo condicionada a que se ha tenido contacto intergrupal? A) 0,625; B) 0,789; C) 0,313.
9. Con los datos de la Tabla 2, la V de Cramer es: A) 0,51; B) 0,32; C) 0,61.
10. En la Tabla 2, el número de personas que no tienen contacto intergrupal y presentan un prejuicio medio es una frecuencia: A) condicional; B) marginal; C) conjunta.
11. Con los datos de la Tabla 3, el coeficiente de correlación de Pearson es igual a: A) 0,72; B) -0,72; C) 0,9.
12. Para conocer la proporción de varianza de la variable pronosticada o criterio (Y) que es explicada por la variable predictora (X) en un modelo lineal, hay que calcular: A) el coeficiente *C* de Contingencia; B) el coeficiente de determinación; C) el coeficiente de correlación de Spearman.
13. Atendiendo a la Tabla 3, ¿cuál sería el rendimiento académico que se predeciría para un alumno que ha puntuado 5 en bienestar? A) 4,75; B) 4,11; C) 5,19.
14. Indica qué afirmación es correcta sobre el Teorema de la Suma: A) cuando los sucesos A y B son dependientes, la probabilidad de ocurrencia de A y B es igual a la probabilidad de ocurrencia de A por la probabilidad de ocurrencia de B; B) se emplea para calcular la probabilidad de unión entre dos sucesos; C) se emplea para calcular la probabilidad de intersección entre dos sucesos.
15. Un determinado trastorno del sueño puede estar provocado solo por el estrés laboral (L) en el 40% de los casos, solo por malos hábitos de sueño (H) en el 35% de los casos, y por la combinación (C) de ambos factores en el 25% de los casos. El tratamiento de este trastorno requiere terapia conductual (T) en

Psicotesunedi.info

el 15% de los casos si la única causa es el estrés laboral, en el 40% de los casos si la única causa son los malos hábitos de sueño, y en el 30% de los casos si es producido por la combinación de ambos factores. ¿Cuál es la probabilidad de que un paciente cualquiera no necesite terapia conductual? A) 0,357; B) 0,275; C) 0,725.

16. Selecciona la función de distribución correcta para una variable aleatoria discreta: A) $F(0)=0{,}15$, $F(1)=0{,}20$, $F(2)=0{,}30$, $F(3)=0{,}80$, $F(4)=1$; B) $F(0)=0{,}15$, $F(1)=0{,}50$, $F(2)=0{,}10$, $F(3)=0{,}50$, $F(4)=0{,}20$; C) $F(0)=0{,}15$, $F(1{,}5)=0{,}50$, $F(2{,}5)=0{,}10$, $F(3{,}5)=0{,}50$, $F(4{,}5)=0{,}20$.

17. En una ciudad, se sabe que un 15% de las personas son alérgicas a cierto tipo de polen. Si elegimos aleatoriamente una muestra de 10 personas, ¿cuál es la probabilidad de que exactamente 2 de ellas sean alérgicas? A) 0,2759; B) 0,8202; C) 0,1969.

18. Una variable aleatoria se distribuye según una *t* de Student con 12 grados de libertad. ¿Cuál es el valor de *t* que deja por encima de sí el 10% de la distribución? A) 1,782; B) 1,356; C) 1,771.

19. Las puntuaciones de 1200 estudiantes en una prueba de aptitud verbal se distribuyen normalmente con una media de 110 y una desviación estándar de 12. ¿Cuál es la probabilidad de obtener una puntuación menor o igual que 95? A) 0,1056; B) 0,3944; C) 0,2119.

20. En una distribución F con 8 grados de libertad en el numerador y 9 grados de libertad en el denominador, ¿cuál es el valor de F que corresponde al percentil 95? A) 6,69; B) 2,60; C) 3,23.

21. Un investigador desea estudiar el nivel de actividad física en adultos mayores. Decide ir a un parque cercano y medir la frecuencia cardíaca y actividad física de las primeras 50 personas mayores de 60 años que encuentra durante su visita. El método de muestreo es: A) incidental; B) intencional; C) bola de nieve.

22. En la población de adultos de Sevilla, se sabe que el porcentaje de personas que realizan actividad física al menos 3 veces por semana es del 30%. Se selecciona una muestra aleatoria de tamaño 200. ¿Cuál es el error típico de estimación de la proporción? A) 0,032; B) 0,045; C) 0,065.

23. Si una variable Y tiene una media $\mu=50$ y una desviación típica $\sigma=12$ en la población, ¿cuáles son los valores de μ y $\sigma_{\bar{y}}$ (error estándar de la media) para una muestra de tamaño 49? A) 50 y 2,00; B) 50 y 3,00; C) 50 y 1,71.

24. Señala cuál de las siguientes afirmaciones es verdadera en relación a los intervalos de confianza: A) muestras heterogéneas implican un error típico de medida pequeño, intervalos de menor amplitud y más precisos; B) un nivel de confianza del 95% conlleva intervalos de confianza más precisos que un nivel de confianza del 99%; C) Un mayor tamaño muestral conlleva un mayor error típico de medida.

25. Una variable X se distribuye normalmente en la población con una varianza poblacional igual a 49. Si se admite un error máximo de estimación de 0,2 y se fija un nivel de confianza del 95%, ¿qué tamaño muestral es necesario para estimar la media poblacional asumiendo población infinita?: A) 2750; B) 3210; C) 4706.

Psicotestuned.info

Reserva

26. Un investigador desea realizar un estudio sobre el porcentaje de estudiantes que consumen bebidas energéticas durante los exámenes. Se asume que la población es infinita. Con un nivel de confianza del 95% y un margen de error del 5%, ¿cuál debe ser el tamaño de la muestra, si estudios previos indican que la proporción de consumidores es del 15%? A) 248; B) 196; C) 323.

Soluciones

1. B

Páginas 15 y 18 (Tabla 1.1).

2. C

X_i =59,5

Motivación (X)	Repetidores (n_i)	n_a	p_a	P_a
80-99	20	120	1	100
60-79	22	100	0,83	83
40-59	33	78	0,65	65
20-39	30	45	0,375	37,5
1-19	15	15	0,125	12,5

65% de personas.

3.C

Página 73. La Moda en una distribución de frecuencias es el punto medio del intervalo al que corresponde la mayor frecuencia. En el caso de No Repetidores es el intervalo (60-79) ya que tiene la mayor frecuencia n_i = 40

60+79/2 = 69,5

4. A

Página 26

5. B

Página 52. Media aritmética a partir de una distribución de frecuencias absolutas

Psicotestuned.info

$$\overline{X} = \frac{\sum n_i X_i}{\sum n_i} = \frac{\sum n_i X_i}{n}$$

Página 112. Coeficiente de variación.

$$CV = \frac{S_X}{\overline{X}} \cdot 100$$

$S_{xrepetidores}$ = 25,303

$S_{xnorepetidores}$ =24,720

X	Punto medio del intervalo (X_i)	Repetidores (n_i)	Repetidores $n_i X_i$	No repetidores (n_i)	No Repetidores $n_i X_i$
80-99	89,5	20	1790	32	2864
60-79	69,5	22	1529	40	2780
40-59	49,5	33	1633,5	23	1138,5
20-39	29,5	30	885	15	442,5
1-19	10	15	150	10	100
			Σ 5987,5		Σ 7325
			120		120
			49,896		61,042

Repetidores: CV = (25,303/49,896) x 100 = 50,711

No repetidores: CV = (24,720/61,042) x 100 = 40,497

6.A

$$A_{IQ} = Q_3 - Q_1 = P_{75} - P_{25}$$

La amplitud intercuartil se calcula restando el cuartil 3 del cuartil 1 ($Q_3 - Q_1$). Los datos se obtienen de la gráfica.

Repetidores: 10 – 5 = 5

No repetidores: 9 – 7 = 2

7. C

Psicotestuned.info

Dado que los bigotes tienen diferente longitud es una distribución asimétrica (página 125). En el grupo de los repetidores, la asimetría es positiva (concentrándose el 75% de valores en 9 o menos) (página 118, Figura 3.2).

8. A

Página 150.

Habría que calcular el porcentaje de prejuicio bajo sobre el total de la fila (contacto intergrupal).

P_i = (150 x 100)/ 240 = 62,5

P = 0,625

9. A

$$\chi^2 = \sum\sum \frac{(n_e - n_t)^2}{n_t}$$

n_e: frecuencia empírica u observada

$$n_t = \frac{Total\ fila \times Total\ columna}{n}$$

	Prejuicio Bajo	Prejuicio Medio	Prejuicio Alto	Total
Contacto intergrupal	150 (95)	60 (65)	30 (80)	240
Ausencia de contacto intergrupal	40 (95)	70 (65)	130 (80)	240
Total	190	130	160	480

$n_{t11} = \frac{240+190}{480} = 95$

$n_{t12} = \frac{240+130}{480} = 65$

$n_{t13} = \frac{240+160}{480} = 80$

$n_{t21} = \frac{240+190}{480} = 95$

$n_{t22} = \frac{240+130}{480} = 65$

Psicotestuned.info

$n_{t23} = \frac{240+160}{480} = 80$

$$\chi^2 = \sum\sum \frac{(n_e - n_t)^2}{n_t} = \frac{(150-95)2}{95} + \frac{(40-95)2}{95} + \frac{(60-65)2}{65} + \frac{(70-65)2}{65} + \frac{(30-80)2}{80} + \frac{(130-80)2}{80} =$$

= 31,842 + 31,842 + 0,385 + 0,385 + 31,25 + 31,25 = 126,95

$$V = \sqrt{\frac{\chi^2}{n(m-1)}}$$

$$V = \sqrt{\frac{\chi^2}{n(m-1)}} = \sqrt{\frac{126,95}{480\ (2-1)}} = 0,51$$

10. C

Página 149-151. Explicación de frecuencias conjuntas, marginales y condicionales.

11. A

$$r_{XY} = \frac{Cov(X,Y)}{S_X S_Y} = \frac{S_{XY}}{S_X S_Y}$$

S_{xy} = 1,771

R_{xy} = 1,771/ 1,396 x 1,761 = 0,719 (aprox. 0,72).

12. B

Página 229.

13. C

Páginas 194, 197, 220, 221

$$S_{XY} = \frac{\sum X_i Y_i}{n} - \bar{X}\bar{Y} = \frac{6500,09}{132} - (6,88 \cdot 6,90) = 1,771$$

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{1,771}{1,396 \cdot 1,761} = 0,72$$

Psicotestuned.info

$b = r_{XY} \cdot \frac{S_Y}{S_X}$ = 0,721 x $\frac{1,768}{1,396*1,76}$ = 0,91

$a = \bar{Y} - b\bar{X} = 6,90 - 0,91 \cdot 6,88$ = 0,64

Y = 0,64+ 0,91·5 = 5,19 (aprox. 5,20)

14. B
Página 264

15. C
Teorema de Bayes (página 277)

0,4 x 0,85 + 0,35 x 0,6 + 0,25 x 0,7 = 0,725

16. A
Página 308

17. A
Página 322. Tabla I del formulario

N =10
X = 2
p = 0,15

18. B
Tabla VI

Al ser simétrica con media 0, se busca directamente en la tabla t, la probabilidad 0,90, con 12 grados de libertad.

19.A
Página 349

Z=1295−110=12−15=−1.25

P (X < 95) = 0,1056

20. C
Página 360
Tabla VII

n_1 = 8
n_2 = 9
Percentil 95

Psicotestuned.info

21. A

Página 385

22. A

Ejercicio 9.10. Error típico de estimación de la proporción.

$\sqrt{0{,}3(1-0{,}3)}$ = 0,458

$\frac{0{,}458}{\sqrt{200}}$ = 0,032

23. C

Ejercicio 9.14

σ_y = 12/ $\sqrt{49}$ =12/7= 1,71

μ=50

24. B

Página 458. Factores que afectan al intervalo de confianza

25 C

Conocida la varianza poblacional asumiendo población infinita.

$n.c. = 0{,}95 \rightarrow \alpha = 0{,}05 \rightarrow Z_{\alpha/2} = Z_{0{,}025} = 1{,}96$ (Tabla III)

$\sigma^2 = 49$, $E_{máx} = 0{,}2$

$$n = \frac{\sigma^2 Z_{\alpha/2}^2}{E_{máx}^2} = \frac{49 \times 1{,}96^2}{0{,}2^2} = \frac{188{,}2384}{0{,}04} = 4705{,}96 \approx 4706$$

Pregunta Reserva

26. B

n= $(1{,}96)^2$ x 0,15 x (1-0,15) / $(0{,}05)^2$

n = 195,92

Aprox. 196.

Psicotestuned.info

INTRODUCCIÓN AL ANÁLISIS DE DATOS
SEPTIEMBRE 2023. Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A. DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Tabla 1. Tabla de frecuencias absolutas de la variable Z.

Z	1	2	3	4	5
n_i	20	30	50	40	30

Tabla 2. Resultados obtenidos en un estudio (n=500) sobre la relación entre el grado de sintomatología ansiosa y estar desempleado.

		Sintomatología	
		Leve	Severa
Desempleo	Sí	50	100
	No	250	100

Tabla 3. Datos obtenidos para relacionar las puntuaciones de un test de inteligencia (X) y el rendimiento laboral (Y).

	Media	Desviación típica	Valor de la correlación
X	50	15	r_{XY}=0,64
Y	10	5	

1. En la Tabla 2, ¿qué escala de medida tiene la variable "sintomatología"? A) Politómica; **B) Ordinal**; C) Cuantitativa discreta.
2. En un examen puntuado de 0 a 10, una muestra de 1000 participantes tiene una media igual a 5, pero el 75% de los participantes han obtenido una nota entre 1 y 3. ¿Cómo es la distribución de esta variable? A) Leptocúrtica; B) Asimétrica negativa; **C) Asimétrica positiva.**
3. ¿Cuál de los siguientes índices de posición indica la puntuación más alta en una variable cuantitativa? A) Primer cuartil; **B) Tercer decil**; C) Percentil 20.
4. ¿Cuál es la media de la variable Z de la Tabla 1? A) 2,88; **B) 3,18**; C) 3,48.
5. Según los datos de la Tabla 1, la varianza de la variable Z es: **A) 1,56**; B) 1,96; C) 2,36.
6. Si una persona obtiene una puntuación de 20 en el test de inteligencia de la Tabla 3, ¿qué puntuación tipificada le corresponde? **A) z_X = -2**; B) z_X = 2; C) z_X = 6.
7. En la Tabla 3, ¿qué variable tiene un mayor coeficiente de variación? A) La variable X; **B) La variable Y**; C) Ambas variables muestran el mismo coeficiente de variación.
8. En la Tabla 2, ¿qué medida de asociación hay que calcular para comprobar si existe relación o no entre las variables "sintomatología" y "desempleo"? **A) Estadístico χ^2 de independencia**; B) Coeficiente de correlación por rangos de Spearman; C) Coeficiente de correlación biserial-puntual.
9. ¿Qué valor toma el estadístico que permite decidir sobre la relación entre las variables "sintomatología" y "desempleo" según los datos de la Tabla 2? **A) 63,49**; B) 53,49; C) 43,49.

PSICOTEST UNED

10. En la Tabla 3, ¿cuál es la ecuación de la recta de regresión que predice el rendimiento laboral (Y) a partir del test de inteligencia (X)? Redondea los resultados de cada una de las operaciones que realices a dos decimales. A) Y' = 0,21 + 0,64X; B) Y' = 10 - 0,50X; **C) Y' = -0,50 + 0,21X**.
11. Asumiendo que se cumplen todos los supuestos estadísticos, ¿cuál es la media esperada de los errores de predicción o residuos de la recta de regresión Y' = 0,50 - 0,50X? **A) 0**; B) 0,50; C) -0,50.
12. ¿Qué valor toma la covarianza entre las variables X e Y de la Tabla 3? A) 0,64; B) 24; **C) 48**.
13. Un investigador quiere analizar la relación entre la variable cuantitativa "cantidad de presencia de proteínas del virus COVID-19" y la variable "diagnóstico de COVID-19" (positivo, negativo). ¿Qué coeficiente de correlación deberá utilizar este investigador? A) Coeficiente de correlación de Pearson; **B) Correlación biserial-puntual**; C) Coeficiente de correlación de Spearman.
14. Si seleccionamos una persona al azar de la Tabla 2, ¿cuál es la probabilidad de que esté desempleado? A) 0,20; **B) 0,30**; C) 0,70.
15. Un investigador puso a prueba un test para el diagnóstico de problemas de alimentación en 200 participantes (100 con diagnóstico de trastornos de la conducta alimentaria, y 100 sin diagnóstico). De los 100 participantes que tenían diagnóstico de trastorno de la conducta alimentaria, 80 fueron correctamente detectados por el test. Ninguno de los 100 participantes sin trastorno de la conducta alimentaria fue clasificado como tal por el test. ¿Cuál es la sensibilidad del test diagnóstico de este investigador? A) 100%; **B) 80%**; C) 40%.
16. Si sabemos que una variable aleatoria toma dos valores (cero y uno) con *E*(X)=0,60, ¿cuál es la probabilidad de que la variable tome el valor "cero"? A) 0,60; **B) 0,40**; C) 0,80.
17. La función de probabilidad de una variable aleatoria discreta es la siguiente: *f*(1)=0,10; *f*(2)=0,25; *f*(3)=0,25; *f*(4)=0,40. ¿Cuánto vale la media de esta variable? A) 1,95; B) 2,5; **C) 2,95**.
18. Si un estudiante responde al azar a un ítem de tres alternativas de respuesta (habiendo sólo una respuesta correcta), y definimos la variable "acertar el ítem". ¿Cuál será la desviación típica de esta variable aleatoria? A) 0,1089; B) 0,3333; **C) 0,4667**.
19. ¿Cuántos puntos de inflexión tiene la distribución normal tipificada o estandarizada? A) Uno; **B) Dos**; C) Cuatro.
20. Una variable se distribuye según *t* con 10 grados de libertad. ¿Cuál es la probabilidad de obtener una puntuación *t* igual o menor a -0,70? **A) 0,25**; B) 0,70; C) 0,75.
21. La variable "responsabilidad" se distribuye *N*(30; 5) en la población de estudiantes de Psicología de la UNED. Si se seleccionan distintas muestras aleatorias de 100 estudiantes, ¿entre qué valores se encuentra el 50% central de las medias? **A) 29,67; 30,33**; B) 25,67; 34,33; C) 20,67; 29,33.
22. La distribución de la variable "comprensión lectora" es normal con varianza (S_X^2) 1. Se extraen muestras de tamaño 11. ¿Cuál es, aproximadamente, la probabilidad de que las varianzas muestrales de la variable "comprensión lectora" se encuentren entre 0,196 y 2,290? A) 0,675; B) 0,800; **C) 0,990.**
23. Si sabemos que la varianza poblacional del test de inteligencia (X) de la Tabla 3 es 225 y se distribuye normalmente, ¿cuál es el intervalo de confianza (α = 0,05) para la media estimada de esta variable en una muestra de 100 personas? **A) [47,06; 52,94]**; B) [41,06; 58,94]; C) [49,06; 50,94].
24. Utilizar la estimación puntual de un parámetro tiene riesgos. ¿Qué podemos hacer para mejorar la interpretación de nuestros resultados? **A) Calcular los intervalos de confianza de la estimación en la muestra**; B) Estimar la varianza de la muestra; C) Aumentar el nivel de riesgo o significación α.
25. ¿Qué característica es más deseable para un estimador insesgado? **A) Que la varianza de su distribución muestral sea lo más pequeña posible**; B) Que la varianza de su distribución muestral sea lo más grande posible; C) Que la varianza poblacional de la variable sea lo más grande posible.

PSICOTEST UNED

SOLUCIONES:

1. B.
2. C.
3. B. El tercer decil deja por debajo al 30% de las observaciones, mientras que el primer cuartil y el percentil 20 dejan un 25% y un 20%, respectivamente.
4. B. n = 170. Entonces, conociendo las frecuencias absolutas (n_i) de cada categoría, podemos aplicar la siguiente fórmula:
$$\bar{X} = \frac{\sum n_i X_i}{n} = \frac{1 \text{ x } 20 + 2 \text{ x } 30 + 3 \text{ x } 50 + 4 \text{ x } 40 + 5 \text{ x } 30}{170} = 3{,}18$$
5. A. Para calcularla, aplicamos la siguiente fórmula:
$$S_X^2 = \frac{\sum n_i (X_i - \bar{X})^2}{n} = \frac{20\,(1-3{,}18)^2 + 30\,(2-3{,}18)^2 + 50\,(3-3{,}18)^2 + 40\,(4-3{,}18)^2 + 30\,(5-3{,}18)^2}{170} = 1{,}56$$
6. A. $z_x = \frac{(X_j - \bar{X})}{S_x} = \frac{(20-50)}{15} = -2$
7. B. El coeficiente de variación de X es $CV = \frac{15}{50} = 0{,}30$. El coeficiente de variación de Y es $CV = \frac{5}{10} = 0{,}50$.
8. A.
9. A. Hay que calcular el estadístico χ^2. La frecuencia esperada en cada casilla de la Tabla 2 es la siguiente:
$$n_{t11} = \frac{(300 \text{ x } 150)}{500} = 90$$
$$n_{t12} = \frac{(200 \text{ x } 150)}{500} = 60$$
$$n_{t21} = \frac{(300 \text{ x } 350)}{500} = 210$$
$$n_{t22} = \frac{(200 \text{ x } 350)}{500} = 140$$
Luego aplicamos el cálculo del estadístico χ^2:
$$\chi^2 = \sum\sum \frac{(n_e - n_t)^2}{n_t} = \frac{(50-90)^2}{90} + \frac{(100-60)^2}{60} + \frac{(250-210)^2}{210} + \frac{(100-140)^2}{140} = 63.49$$
10. C. Para responder, hay que calcular los coeficientes a y b de la ecuación de regresión: Y' = *a* + *b*X.
$$b = r_{XY}\frac{S_Y}{S_X} = 0{,}64\frac{5}{15} = 0{,}21$$
$$a = \bar{Y} - b * \bar{X} = 10 - (0{,}21 \text{ x } 50) = -0{,}50$$
Por tanto: Y' = -0,50 + 0,21X.
11. A. Cuando se cumplen todos los supuestos, la media de los errores de predicción o residuos de los modelos de regresión lineal siempre es 0. Página 230 del manual.
12. C. Dado que r_{XY} es 0.64 y tenemos los valores de las desviaciones típicas de X e Y, entonces podemos conocer el valor de la covarianza a partir de la fórmula de la correlación:
$r_{XY} = \frac{Cov(X,Y)}{S_X S_Y}$ → $0{,}64 = \frac{Cov(X,Y)}{5 \text{ x } 15}$ → Cov(X,Y) = 48.
13. B. La correlación biserial-puntual es la técnica de correlación adecuada para calcular la relación entre una variable continua y una variable dicotómica.
14. B: $P(Desempleo) = \frac{150}{500} = 0{,}30$.
15. B. $P(+|T) = \frac{80}{100} = 0{,}80$
16. B. Ejercicio similar resuelto en la página 330 del libro. El enunciado nos informa de que la variable toma valores de cero y uno, con una esperanza o media igual a 0,60. Entonces, $f(1) = 0{,}60$. Además, sabemos que $\sum f(x) = 1$ porque es una función de probabilidad. Por tanto, $f(0) + f(1)$ = 1,00 → $f(0) + 0{,}60 = 1$ → $f(0) = 1 - 0{,}60 = 0{,}40$.
17. C. $E(X) = \sum x\, f(x) = 1 \text{ x } 0{,}10 + 2 \text{ x } 0{,}25 + 3 \text{ x } 0{,}25 + 4 \text{ x } 0{,}40 = 2{,}95$.

PSICOTEST UNED

18. C. Esta variable solo tiene 2 resultados: acertar por azar (1) con p=0,33, o fallar (0) con q=1-0,33=0,66, Dado que sabemos que esta variable se distribuye según Benoulli con parámetro p=0,33: $\sigma = \sqrt{\sigma^2} = \sqrt{pq} = \sqrt{0{,}33 \text{ x } 0{,}66} = \sqrt{0{,}2178} = 0{,}4667$.
19. B. Página 344 del manual.
20. A. Según las tablas estadísticas de la distribución t de Student (Tabla VI), $P(X \leq -0{,}70) = 0{,}25$.
21. A. Ejercicio similar al 9.15 del manual. La variable responsabilidad se distribuye *N*(30; 5) en la población. La variable se distribuye normalmente y conocemos la media y la varianza poblacional, así que se puede tipificar la distribución normal de la variable. Como queremos averiguar qué puntuación Z deja por debajo al 25% y qué Z deja por encima al 25%:

    $P(z < Z) = 0{,}25$, cuyo punto de corte es: -0,67

    $P(z < Z) = 0{,}75$, cuyo punto de corte es: 0,67

    Volvemos a transformar la variable a su métrica original:

    $Z_{\bar{X}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \rightarrow -0{,}67 = \frac{\bar{X}-30}{5/\sqrt{100}} \rightarrow \bar{X} = (-0{,}67) * 0{,}5 + 30 = 29{,}67$

    $Z_{\bar{X}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \rightarrow 0{,}67 = \frac{\bar{X}-30}{5/\sqrt{100}} \rightarrow \bar{X} = 0{,}67 * 0{,}5 + 30 = 30{,}33$

22. C. Ejercicio igual que el 9.16 del manual. Sabemos que $\chi^2 = \frac{nS_X^2}{\sigma^2}$ se distribuye χ^2_{n-1}. Por tanto, calculamos valores de χ^2_{n-1} para $S_X^2 = 0{,}196$ y para $S_X^2 = 2{,}290$.

    $$\chi^2_{n-1} = \frac{nS_X^2}{\sigma^2} = \frac{11 * 0{,}196}{1} \approx 2{,}1559$$
    $$\chi^2_{n-1} = \frac{nS_X^2}{\sigma^2} = \frac{11 * 2{,}290}{1} \approx 25{,}1882$$

    Para $\chi^2_{10} = 2{,}1559$, la probabilidad correspondiente es de 0,005, aproximadamente.

    Para $\chi^2_{10} = 25{,}1882$, la probabilidad correspondiente es de 0,995, aproximadamente.

    Por tanto, la probabilidad es: $P(0{,}196 \leq S_X^2 \leq 2{,}290) = 0{,}995 - 0{,}005 = 0{,}990$.

23. A. Dado S_x^2=σ_x^2=225, y α=0,05… Como conocemos la varianza poblacional de la variable y sabemos que se distribuye normalmente:
    Nivel del intervalo de α: [-1.96; 1,96]

    $$L_{inf} = \bar{X} - \frac{1{,}96\,\sigma}{\sqrt{n}} = 50 - \frac{1{,}96 \text{ x } 15}{\sqrt{100}} = 50 - 2{,}94 = 47{,}06$$
    $$L_{sup} = \bar{X} - \frac{1{,}96\,\sigma}{\sqrt{n}} = 50 + \frac{1{,}96 \text{ x } 15}{\sqrt{100}} = 50 + 2{,}94 = 52{,}94$$

24. A.
25. A.

PSICOTEST UNED

Compartida por Bea E.
Antes de imprimir, piensa en tu responsabilidad con la naturaleza.
Quizás no pueda salvar el planeta, pero sí puedes dejar de destruirlo.

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2023 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO B DURACION: 2 HORAS

Material: **Adenda (Formulario y Tablas) y calculadora (no programable)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Tabla 1. Frecuencia relativa del tiempo de reacción de 120 sujetos en una prueba de atención.

Tiempo de reacción (segundos)	Frecuencia relativa (p_i)
0-9	0,042
10-19	0,291
20-29	0,358
30-39	0,242
40-49	0,067

Figura 1. Diagrama de caja de la variable "Minutos invertidos en Instagram al día"

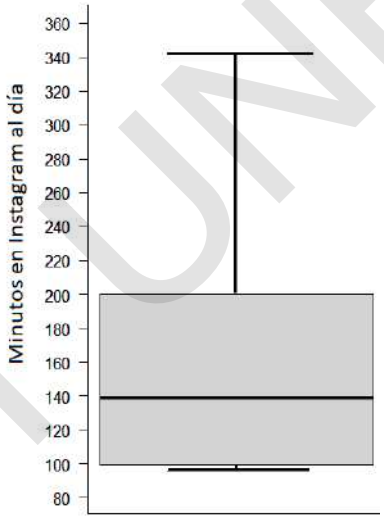


Tabla 2. Recuperación de un Trastorno de Ansiedad Generalizado, en función de si ha recibido tratamiento o no.

		Grupo experimental	
		Tratamiento	Control
Resulta-do	Se recupera	29	17
	No se recupera	8	20

Tabla 3. En la siguiente tabla se informa de algunos índices descriptivos de las variables horas de estudio (X) y rendimiento académico (Y) de 87 alumnos.

	Media	**Varianza**	
X	9,96	2,38	$\sum XY =$ 4259,73
Y	4,80	5,63	

1. El índice que describe alguna característica de la población se denomina: A) estadístico; **B) parámetro;** C) frecuencia.
2. La variable "Tiempo de reacción" de la Tabla 1 es una variable: A) cuantitativa de intervalos; **B) cuantitativa de razón;** C) cuantitativa discreta.
3. Atendiendo a la Tabla 1, ¿por debajo de qué segundo ha reaccionado el 75% de la muestra? A) 25,57; B) 17,64; **C) 31,91.**
4. Calcula la media y la mediana de los tiempos de reacción de la Tabla 1. Atendiendo a esos valores, se puede concluir que: **A) no parece que haya datos atípicos, ya que la media y la mediana son muy similares (se diferencian en menos de un segundo);** B) hay datos atípicos, ya que la media y la mediana difieren mucho entre sí (se diferencian en más de un segundo); C) no se puede saber si hay o no datos atípicos mirando solo los valores de la media y la mediana.
5. En base a la Figura 1, el valor de la amplitud intercuartil de la variable "minutos pasados al día en *Instagram*" es: **A) 100;** B) 242; C) 40.
6. En base a la Figura 1, el índice de asimetría de Fisher de la variable "minutos pasados al día en *Instagram*" será: A) igual a 0; B) menor que cero; **C) mayor que 0.**
7. En una muestra de 60 recién nacidos, la desviación típica de la variable peso es 1,34 y la desviación típica de la variable altura es 5,23. De estos datos, se puede concluir que: A) existe más variabilidad entre las alturas de los recién nacidos que entre su peso; **B) no se pueden comparar entre sí las**

PSICOTESTUNED

desviaciones típicas de estas variables; C) para comparar la variabilidad de estas dos variables, es necesario saber sus varianzas.

8. En la Tabla 2 ¿cuál es el valor del coeficiente V de Cramer y qué indica este valor? A) 0,671, por lo que pertenecer al grupo de tratamiento o al grupo control no está relacionado con recuperarse o no; **B) 0,334, por lo que pertenecer al grupo de tratamiento o al grupo control está moderadamente relacionado con recuperarse o no**; C) 0,193, por lo que pertenecer al grupo de tratamiento o al grupo control no está relacionado con recuperarse o no.
9. En la Tabla 2, el número total de personas que han recibido el tratamiento es una frecuencia: **A) marginal;** B) conjunta; C) condicional.
10. Con los datos de la Tabla 3, ¿cuál es el valor de la correlación de Pearson? A) 0,465; **B) 0,316;** C) 0,086.
11. Con los datos de la Tabla 3 ¿cuál sería el rendimiento académico que se predeciría a un alumno que ha estudiado 14,15 horas? **A) 6,88;** B) 8,53; C) 6,12.
12. Si la variable X de la Tabla 3 fuera "pertenecer a la clase A o a la clase B" en lugar de "horas de estudios", ¿cuál habría sido el mejor coeficiente para estudiar la relación entre las dos variables? A) La correlación lineal de Spearman; B) El coeficiente phi – φ; **C) La correlación biserial puntual.**
13. Atendiendo a la Tabla 2, la probabilidad de recuperarse habiendo recibido previamente el tratamiento es: A) 0,682; B) 0,392; **C) 0,784.**
14. La probabilidad de estudiar Psicología es 0,14, y la probabilidad de tener una mascota en casa es 0,45. Si la probabilidad de la intersección de estos dos sucesos es 0,063, los sucesos son: A) mutuamente excluyentes; **B) independientes;** C) dependientes.
15. Siendo X una variable aleatoria discreta, selecciona la función de probabilidad correcta: **A) *f*(-1)=0,50, *f*(0)=0,35, *f*(1)=0,15;** B) *f*(1)=0,34, *f*(2)=0,75, *f*(3)=0,05; C) *f*(0)=0,10, *f*(1)=0,40, f(2)=0,75.
16. Un estudio reciente ha encontrado que 44 de cada 110 mujeres ha sufrido algún episodio de abuso sexual en la vía pública. Si elegimos aleatoriamente una muestra de 12 mujeres, ¿cuál es la probabilidad de que más de 5 mujeres hayan sufrido algún episodio de abuso sexual en la vía pública? A) 0,4382; B) 0,6652; **C) 0,3348.**
17. Las calificaciones de la asignatura de matemáticas de un grupo de 200 alumnos se distribuye normalmente con media 5,8 y desviación típica 0,85. ¿Cuántos alumnos han sacado más de un 6? **A) 81**; B) 135; C) 119.
18. En una variable aleatoria Y que sigue la distribución ji-cuadrado con 27 grados de libertad, la probabilidad de que Y sea menor de 43,1945 es: A) 0,950; **B) 0,975;** C) 0,025.
19. Con 13 grados de libertad, ¿qué valor del estadístico *t* de Student corresponde al percentil 25 de la distribución? A) 0,394; B) -2,160; **C) -0,694.**
20. El método de muestreo estratificado es: A) convencional; **B) probabilístico**; C) no probabilístico.
21. Se mide una variable X en una muestra aleatoria de 350 sujetos obteniendo una cuasivarianza de 36. El error típico de la media de la variable X vale: A) 0,11; **B) 0,32;** C) 0,23
22. La media de la distribución muestral de una proporción X es 0,25. ¿Cuál será el valor de la desviación típica de la distribución muestral de esta proporción en una muestra de 120 personas? A) 0,124; **B) 0,040**; C) 0,002.
23. ¿Cuál de estas propiedades <u>NO</u> corresponde a un buen estimador? A) Eficiencia; B) Suficiencia; **C) Eficacia.**
24. Manteniendo el tamaño de la muestra constante, ¿qué ocurre cuando se aumenta el nivel de confianza de un intervalo de confianza? A) La amplitud del intervalo se reduce; B) El nivel de riesgo no está relacionado con el nivel de confianza; **C) La amplitud del intervalo aumenta.**
25. Una escala de memoria se distribuye normalmente con media μ y varianza σ^2 = 24. Dado que es una población infinita, ¿qué tamaño muestral se necesita para que el error cometido en la estimación de la media de la variable memoria sea como máximo 0,85, con un nivel de significación de 0,10? **A) 89**; B) 18; C) 52.

PSICOTEST UNED

SOLUCIONES:

1. **B)**
   Páginas 8 y 9 del manual

2. **B)**

   Tabla 1.1. de la página 18

3. **C)**
   El percentil 75 nos dice el número de segundos a partir de cuál ha respondido el 75% más rápido de la muestra:

$$P_{75} = L_i + \left(\frac{\frac{n \cdot k}{100} - n_d}{n_c}\right) \cdot I = 29{,}5 + \frac{\left(\frac{120 \cdot 75}{100} - 83\right)}{29} \cdot 10 = 29{,}5 + 2{,}41 = 31{,}91$$

Tiempo de reacción (segundos)	p_i	n_i	n_a
0-9	0,042	5	5
10-19	0,291	35	40
20-29	0,358	43	83
30-39	0,242	29	112
40-49	0,067	8	120

4. **A)**
   Calcula media y mediana

Tiempo de reacción (segundos)	p_i	n_i	n_a	X_i	$n_i X_i$
0-9	0,042	5	5	4,5	22,5
10-19	0,291	35	40	14,5	507,5
20-29	0,358	43	83	24,5	1053,5
30-39	0,242	29	112	34,5	1000,5
40-49	0,067	8	120	44,5	356

$$P_{50} = L_i + \left(\frac{\frac{n \cdot k}{100} - n_d}{n_c}\right) \cdot I = 19{,}5 + \frac{\left(\frac{120 \cdot 50}{100} - 40\right)}{43} \cdot 10 = 19{,}5 + 4{,}65 = 24{,}15$$

Media:

$$\bar{X} = \frac{\sum n_i X_i}{n} = \frac{2940}{120} = 24{,}5$$

La media y la mediana son muy similares, así que no parece que haya datos atípicos. (pág. 74).

5. **A)**
   La amplitud intercuartil se calcula restando el cuartil 3 del cuartil 1 ($Q_3 - Q_1$). En la figura 1, el cuartil 3 es 200 (donde termina la caja) y el cuartil 1 es 100 (donde comienza la caja), por lo que la amplitud intercuartil es $Q_3 - Q_1$ = 200 - 100 = 100

PSICOTEST UNED

6. **C)**
Un índice de asimetría de Fisher positivo indicaría una asimetría positiva, es decir, habría muchas puntuaciones bajas y sólo unas pocas altas. En el diagrama de cajas, el bigote de arriba es más largo que el bigote de abajo, y además el 75% de las observaciones están en valores bajos, mientras que parece que hay unos pocos participantes con puntuaciones más altas, lo que estaría indicando una asimetría positiva (la asimetría positiva se explica en la página 118, y la asimetría del diagrama de cajas en la página 125 y 126).

7. **B)**
Página 112 "Debido a que, por lo general, las variables objeto de estudio se miden en unidades distintas, no tiene sentido compararlas en base a los valores de sus varianzas o de sus desviaciones típicas".

8. **B)**

		Grupo experimental		
		Tratamiento	Control	
Resultado	Se recupera	29 (23)	17 (23)	*46*
	No se recupera	8 (14)	20 (14)	*28*
		37	*37*	**74**

$$\chi^2 = \sum\sum \frac{(n_e - n_t)^2}{n_t} = \frac{(29-23)^2}{23} + \frac{(17-23)^2}{23} + \frac{(8-14)^2}{14} + \frac{(20-14)^2}{14}$$
$$= 1{,}5652 + 1{,}5652 + 2{,}5714 + 2{,}5714 = 8{,}2732$$

$$V = \sqrt{\frac{\chi^2}{n(m-1)}} = \sqrt{\frac{8{,}2732}{74(2-1)}} = 0{,}334$$

9. **A)**

Página 152: explicación de las frecuencias marginales

10. **B)**

$$S_{XY} = \frac{\sum X_i Y_i}{n} - \bar{X}\bar{Y} = \frac{4259{,}73}{87} - (9{,}96 \cdot 4{,}80) = 1{,}154$$
$$r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{1{,}154}{1{,}54 \cdot 2{,}37} = 0{,}316$$

11. **A)**

$$Y = a + b \cdot X$$
$$b = r_{XY} \cdot \frac{S_Y}{S_X} = 0{,}316 \cdot \frac{2{,}37}{1{,}54} = 0{,}486$$
$$a = \bar{Y} - b\bar{X} = 4{,}80 - 0{,}486 \cdot 9{,}96 = -0{,}041$$
$$Y = -0{,}041 + X \cdot 0{,}489$$
$$Y = -0{,}041 + 14{,}15 \cdot 0{,}489 = 6{,}88$$

12. **C)** Página 213: correlación biserial-puntual

PSICOTEST UNED

13. **C)**

P(recuperase/tratamiento)= $\frac{P(\text{recuperase} \cap \text{tratamiento})}{P(\text{tratamiento})} = \frac{\frac{29}{74}}{\frac{37}{74}} = \frac{0{,}392}{0{,}50} = 0{,}784$

14. **B)**

Según el teorema del producto, si los sucesos A y B son independientes, la probabilidad de su intersección será igual a la multiplicación de la P(A) y P(B) (página 269).

$P(A \cap B) = 0{,}063$

$P(A) \cdot P(B) = 0{,}14 \cdot 0{,}45 = 0{,}063$

Por lo tanto, estos sucesos son independientes.

15. **A)**

Página 308: La suma de todas las probabilidades correspondientes a cada valor de X es igual a 1.

16. **C)**

La probabilidad de haber experimentado abuso sexual en la vía pública es 0,40 $\left(p = \frac{44}{110} = 0{,}40\right)$.

n = 12

X > 5

P(X > 5) = 1 – P(X ≤ 5) = 1 – 0,6652 = 0,3348 (ver en Tabla II)

17. **A)**

Tabla III

$Z = \frac{6 - 5{,}8}{0{,}85} = 0{,}235$

P(X > 6) = P(Z > 0,235) ≅ P(Z > 0,24) = 1 - P(Z < 0,24) = 1- 0,5948 = 0,4052

200 * 0,4052 ≅ 81 alumnos han sacado más de un 6 en matemáticas.

18. **B)**

Tabla V

Con 27 grados de libertad, P(Y < 43,19) = 0,975.

19. **C)**

Tabla VI.

La *t*-Student que corresponde a un percentil 0,75 (con 13 grados de libertad) es 0,694, por lo tanto, la t-Student que corresponde a un percentil 0,25 es la misma, pero en negativo: -0,694 (debido a la simetría de la distribución *t* de Student).

20. **B)**

Tabla 9.1. de la página 379

21. **B)**

$$S_{\bar{X}} = \sqrt{\frac{S_{n-1}^2}{n}} = \sqrt{\frac{36}{350}} = 0{,}32$$

PSICOTEST UNED

22. **B)**

$$\hat{\sigma}_p = \sqrt{\frac{p\,(1-p)}{n}} = \sqrt{\frac{0{,}25 \cdot (1-0{,}25)}{120}} = 0{,}0395 \;\approx 0{,}040$$

23. **C)**
página 434: propiedades de un buen estimador

24. **C)**
Páginas 441 y 442.

25. **A)**
Hay un ejercicio similar en la página 461-462

$$n = \frac{\sigma^2 \cdot Z^2_{\alpha/2}}{E^2_{max}} = \frac{24 \;\cdot\; 1{,}64^2}{0{,}85^2} = 89{,}34$$

PSICOTEST UNED

Compartida por Bea E.
Antes de imprimir, piensa en tu responsabilidad con la naturaleza. Quizás no pueda salvar el planeta, pero sí puedes dejar de destruirlo.

INTRODUCCIÓN AL ANÁLISIS DE DATOS
Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Material: **Adenda (Formulario y Tablas) y calculadora (no programable)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Tabla 1. Puntuaciones de 100 adultos en una escala de *ansiedad* (X) agrupadas en intervalos y sus frecuencias absolutas (n_i).

X	n_i
17-20	10
13-16	20
9-12	42
5-8	21
1-4	7

Tabla 2. Resultados obtenidos en un estudio sobre la relación entre tipo de titulación y el género de 400 estudiantes de la UNED.

		Titulación	
		Psicología (P)	Derecho (D)
Género	Mujeres (M)	140	100
	Hombres (H)	60	100

Tabla 3. Datos obtenidos para pronosticar las puntuaciones en un examen de educación física (Y) a partir de las puntuaciones en un test de habilidad motora (X).

	Media	Desviación típica	Correlación X e Y
X	100	10	$r_{XY} = 0{,}7$
Y	8	2	

Tabla 4. Variables y características de su distribución.

Variable	Distribución	
X	χ^2_7	Ji-cuadrado con 7 grados de libertad
Y	N(100,15)	Normal con media 100 y desviación típica 15

1. La variable "nivel socioeconómico de la familia" (operativizada como *bajo*, *medio*, *alto*) es: A) cualitativa nominal; **B) cuasicuantitativa ordinal;** C) cuantitativa de intervalo.
2. La amplitud del intervalo de los datos de la Tabla 1 es: **A) 4;** B) 3; C) 3,5.
3. Un adulto de la Tabla 1 tiene una puntuación X = 9. Su nivel de ansiedad con respecto a la media de su grupo es: **A) inferior;** B) igual; C) superior.
4. Según los datos de la Tabla 1, ¿cuál es la puntuación de ansiedad X que deja por debajo al 25% de los adultos de la muestra? A) 10,7; B) 13,5; **C) 7,9.**
5. La varianza de las puntuaciones agrupadas en intervalos de la Tabla 1 es: A) 16,1; **B) 17,4;** C) 18,3.
6. El índice de asimetría de Pearson: A) es un índice adimensional; B) se aplica a distribuciones unimodales; **C) A y B son correctas.**
7. Teniendo en cuenta los datos de la Tabla 2, el valor del estadístico χ^2 (ji-cuadrado) entre las variables *género* y *titulación* se encuentra entre: **A) 16 y 18**; B) 10 y 12; C) 13 y 15.
8. Atendiendo a los datos de la Tabla 2, ¿cuál es el valor del coeficiente *V* de Cramer? **A) 0,20;** B) 0,32; C) 0,40.
9. Con los datos de la Tabla 3, la covarianza entre ambas variables es igual a: A) 9; B) 7; **C) 14.**

PSICOTEST UNED

10. Teniendo en cuenta los datos de la Tabla 3, la recta de regresión que permite pronosticar la variable *Y* en función de la variable X es: **A)** $\boldsymbol{Y' = -6 + 0,14X}$**;** B) $Y' = -3 + 1,4X$; C) $Y' = -14 + 6X$
11. En la recta de regresión, la proporción de varianza de la variable pronosticada Y explicada por la varianza de la variable predictora X es igual a: A) el coeficiente de correlación entre ambas variables; **B) el coeficiente de correlación entre ambas variables al cuadrado;** C) la raíz cuadrada del coeficiente de correlación entre ambas variables.
12. Con los datos de la Tabla 2, si elegimos al azar a un estudiante, ¿cuál es la probabilidad de que, habiendo estudiado psicología, sea mujer?: **A) 0,70**; B) 0,35; C) 0,60.
13. La probabilidad de ser fumador es 0,20, y la probabilidad de fumar dado que se ha estudiado el grado de Veterinaria es 0,20. Con esta información, se puede concluir que estos sucesos son: A) dependientes; **B) independientes**; C) mutuamente excluyentes.
14. Una variable aleatoria discreta X toma los valores 1 y 2. Sabemos que *P(X* = 1) = 0,4 y *P(X* = 2) = 0,6. ¿Cuánto vale la media de X?: **A) 1,6;** B) 1,4; C) 1,3.
15. Tenemos los siguientes datos de una variable discreta X: *f*(0) = 0,10; *f*(2) = 0,38; *f*(4) = 0,40; *f*(5) = 0,12. Podemos decir que: A) se trata de una función de distribución; **B) se trata de una función de probabilidad;** C) no se trata de una función de probabilidad porque no cumple una de las propiedades.
16. Se sabe que un 30% de la población adulta tiene alto el colesterol. Si elegimos aleatoriamente 7 personas, la probabilidad de que solo 2 de ellas tengan colesterol es: A) 0,2753; B) 0,2471; **C) 0,3177.**
17. Teniendo en cuenta los datos de la Tabla 4 para la variable X, ¿qué valor de χ^2 deja por debajo al 90% de los valores de la distribución? A) 10,6446; **B) 12,0170**; C) 14,0671**.**
18. Con los datos de la Tabla 4, ¿cuál es la probabilidad de que Y esté comprendida entre los valores 85 y 105?: **A) 0,4706;** B) 0,1587; C) 0,9332.
19. La distribución *F* de Fisher: A) es simétrica con media igual a 0; B) es asimétrica negativa y nunca toma valores menores que 0; **C) es asimétrica positiva y nunca toma valores menores que 0.**
20. Indica cuál de estas afirmaciones sobre muestreo es correcta: A) el muestreo aleatorio simple no considera que todos los elementos de la población sean extraídos de forma equiprobable; **B) el método de muestreo por cuotas es no probabilístico;** C) el muestreo por conglomerados es no probabilístico.
21. La variable *inteligencia emocional* se distribuye normalmente con media 25 y varianza desconocida. En una muestra de 41 estudiantes, se observa una cuasidesviación típica de 4,92. ¿Cuál es la probabilidad de que la media de inteligencia emocional esté entre 24 y 27? A) 0,750; **B) 0,895;** C) 0,200.
22. Sea X una variable con distribución normal en la población y con varianza poblacional igual a 25. Si extraemos una muestra aleatoria de tamaño 49 con reemplazamiento y obtenemos una media muestral de 20, ¿cuánto vale la desviación típica de la distribución muestral de la media?: **A) 0,71;** B) 0,65; C) 4,51.
23. Con los datos de la pregunta anterior, con un nivel de confianza del 95% y asumiendo población infinita, ¿entre qué valores estaría comprendida la media poblacional de X?: **A) (18,6 y 21,40);** B) (19,82 y 20,18); C) (17,07 y 20,94).
24. Dada una prevalencia poblacional de 0,50, se quiere conocer cuál es la prevalencia muestral de la ansiedad en adolescentes con un error máximo del 2% y una confianza del 95%. ¿Qué tamaño muestral es necesario asumiendo población infinita? A) 4201; B) 1240; **C) 2401.**
25. En un estudio se quiere averiguar el porcentaje de personas que están a favor del aborto. Para ello se ha extraído una muestra de 200 personas de las que 140 están a favor. Con un nivel de confianza de 0,90, los límites del intervalo de confianza para la proporción son: **A) (0,6465 y 0,7535);** B) (0,6250 y 0,7650); C) (0,6950 y 0,7050).

PSICOTEST UNED

SOLUCIONES:

1. B (página 18)
2. A (página 32)
3. A

X	X_i	n_i	$n_i X_i$
17-20	18,5	10	185
13-16	14,5	20	290
9-12	10,5	42	441
5-8	6,5	21	136,5
1-4	2,5	7	17,5
	$\sum$	100	1070

$$\bar{X} = \frac{\sum n_i X_i}{n} = \frac{1070}{100} = 10{,}7$$

Como X = 9 es inferior que el valor de la media 10,7 el adulto tiene un nivel de ansiedad inferior a la media del grupo.

4. C

X	n_i	n_a
17-20	10	100
13-16	20	90
9-12	42	70
5-8	21	28
1-4	7	7

$\frac{n \cdot k}{100} = \frac{100 \times 25}{100} = 25$ El intervalo crítico es 5-8

$$P_{25} = L_i + \left(\frac{\frac{n \cdot k}{100} - n_d}{n_c}\right) \cdot I = 4{,}5 + \left(\frac{\frac{100 \cdot 25}{100} - 7}{21}\right) \cdot 4 = 4{,}5 + \left(\frac{18}{21}\right) \cdot 4 = 7{,}93 \approx 7{,}9$$

5. B $\bar{X} = 10{,}7$

X	X_i	n_i	X_i^2	$n_i X_i^2$
17-20	18,5	10	342,25	3422,5
13-16	14,5	20	210,25	4205
9-12	10,5	42	110,25	4630,5
5-8	6,5	21	42,25	887,25
1-4	2,5	7	6,25	43,75
	$\sum$	100		13189

$$S_X^2 = \frac{\sum n_i X_i^2}{n} - \bar{X}^2 = \frac{13189}{100} - 10{,}7^2 = 131{,}89 - 114{,}49 = 17{,}4$$

6. C (página 117)

PSICOTEST UNED

7. A

		Titulación		
		Psicología (P)	Derecho (D)	
Género	Mujeres (M)	140 (120)	100 (120)	240
	Hombres (H)	60 (80)	100 (80)	160
		200	200	400

Cálculo de las frecuencias teóricas (entre paréntesis en la tabla)

$$\frac{(240 \cdot 200)}{400} = 120 \quad \frac{(240 \cdot 200)}{400} = 120 \quad \frac{(160 \cdot 200)}{400} = 80 \quad \frac{(160 \cdot 200)}{400} = 80$$

Cálculo del estadístico χ^2

$$\chi^2 = \frac{(140-120)^2}{120} + \frac{(100-120)^2}{120} + \frac{(60-80)^2}{80} + \frac{(100-80)^2}{80} =$$
$$= \frac{400}{120} + \frac{400}{120} + \frac{400}{80} + \frac{400}{80} \cong 16{,}67$$

8. A

$\chi^2 = 16{,}67$, n=400, m=2.

$$V = \sqrt{\frac{\chi^2}{n(m-1)}} = \sqrt{\frac{16{,}67}{400 \times 1}} = \sqrt{0{,}0417} = 0{,}204 \cong 0{,}20$$

9. C

$$r_{XY} = \frac{\text{Cov (X, Y)}}{S_X S_Y} = \frac{S_{XY}}{S_X S_Y} \quad \Rightarrow \quad S_{XY} = r_{XY} S_X S_Y = 0{,}7 \cdot 10 \cdot 2 = 14$$

10. A

11. $Y' = a + bX$

$$a = \bar{Y} - b\bar{X} = 8 - (0{,}14 \cdot 100) = -6; \;\; b = r_{XY}\frac{S_Y}{S_X} = 0{,}7 \cdot \frac{2}{10} = 0{,}14$$

12. $Y' = -6 + 0{,}14X$

11. B (página 229).

PSICOTEST UNED

12. A

		Titulación		
		Psicología (P)	Derecho (D)	
Género	Mujeres (M)	140	100	240
	Hombres (H)	60	100	160
		200	200	400

$$P(\text{mujer/psicología}) = \frac{\frac{140}{400}}{\frac{200}{400}} = 0{,}70$$

13. B

Cuando la P(A) es igual que la P(A/B), se concluye que los sucesos son independientes:

P(A/B) = P(A)

En este caso, ser fumador es el suceso A, con probabilidad 0,20. El suceso B es haber estudiado veterinaria. La probabilidad de P(A/B), o en otras palabras, la probabilidad de ser fumador dado haber estudiado veterinaria. es 0,20. Por lo tanto, se cumple la siguiente igualdad P(A/B) = P(A), y se concluye que los sucesos son independientes.

14. A

X	*f(x)*	*x·f(x)*
1	0,4	0,4
2	0,6	1,2
		1,6

$$\mu_X = \sum x \cdot f(x) = 1{,}6$$

15. B

16. C

p = 0,30; *n* = 7; *x* = 2

Buscamos en la Tabla I de la función de probabilidad binomial con estos datos y la solución correcta es 0,3177

17. B

Buscar la información en la Tabla V:

P($\chi^2 < x) = 0{,}90$, mirando en la Tabla V, con 7 grados de libertad, el valor de χ^2 que deja por debajo el 90% de la distribución es 12,0170.

PSICOTEST UNED

18. A

$$Z = \frac{X - \bar{X}}{S_X} = \frac{85 - 100}{15} = \frac{-15}{15} = -1$$
$$Z = \frac{X - \bar{X}}{S_X} = \frac{105 - 100}{15} = \frac{5}{15} = 0{,}33$$

$P(85 < X < 105) = P(-1 < Z < 0{,}33) = P(Z < 0{,}33) - P(Z < -1)$ = 0,6293 – 0,1587 = 0,4706

(mirar Tabla IV)

19. C

20. B (página 379)

21. B

$$t = \frac{\bar{X} - \mu}{\frac{S_{n-1}}{\sqrt{n}}} = \frac{24 - 25}{\frac{4{,}92}{\sqrt{41}}} = \frac{-1}{0{,}768} = -1{,}30$$
$$t = \frac{\bar{X} - \mu}{\frac{S_{n-1}}{\sqrt{n}}} = \frac{27 - 25}{\frac{4{,}92}{\sqrt{41}}} = \frac{2}{0{,}768} = 2{,}60$$

P(-1,30 ≤ t ≤ 2,60) = P(t ≤ 2,60) – P(t ≤ -1,30) = 0,995 – 0,100 = **0,895**

22. A

X se distribuye normalmente con $\sigma^2 = 25$ por lo que: $\sigma = 5$

Muestra de *n* = 49 con $\bar{X} = 20$.
Nos piden la desviación típica de la distribución muestral de la media, también llamada error típico de media.

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{49}} = \frac{5}{7} = 0{,}714 \approx 0{,}71$$

23. A

Variable con distribución normal y varianza poblacional conocida: $X \rightarrow N(\mu, 5)$
Muestra de *n* = 49 y $\bar{X} = 20$. Nivel de confianza del 95%.

$\alpha = 0{,}05 \rightarrow \left|\alpha/2\right| = |1{,}96|$ (Tabla III y IV);

$$L_{inf} = \bar{X} - \left|Z_{\frac{\alpha}{2}}\right| \sigma/\sqrt{n} = 20 - 1{,}96\frac{5}{\sqrt{49}} = 20 - 1{,}4 = 18{,}6$$

$$L_{sup} = \bar{X} + \left|Z_{\frac{\alpha}{2}}\right| \sigma/\sqrt{n} = 20 + 1{,}96\frac{5}{\sqrt{49}} = 20 + 1{,}4 = 21{,}4$$

Por tanto, los límites del intervalo son: (18,6; 21,4)

PSICOTEST UNED

24. C Página 472

$$E_{max} = Z_{\alpha/2}\sqrt{\frac{P\,(1-P)}{n}}$$

$$n = \frac{Z_{\alpha/2}^2\,P\,(1-P)}{E_{max}^2} = \frac{(-1{,}96)^2 \cdot 0{,}5 \cdot (1-0{,}5)}{0.02^2} = \frac{0{,}9604}{0{,}0004} = 2401$$

25. A

$n.c. = 0{,}90 \rightarrow z_{1-\alpha/2} = z_{0{,}95} = 1{,}65\ (Tabla\ IV)$

$$L_{inf} = 0{,}70 - 1{,}65\sqrt{\frac{0{,}70 \times 0{,}30}{200}} = 0{,}70 - 0{,}0535 = 0{,}6465$$

$$L_{sup} = 0{,}70 + 1{,}65\sqrt{\frac{0{,}70 \times 0{,}30}{200}} = 0{,}70 + 0{,}0535 = 0{,}7535$$

$$P = \frac{140}{200} = 0{,}70$$

PSICOTEST UNED

INTRODUCCIÓN AL ANÁLISIS DE DATOS
SEPTIEMBRE 2022 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (no programable)**
Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No es necesario entregar los enunciados

Tabla 1. X: Latencia en responder expresada en milisegundos. n_i: Frecuencia absoluta

X	n_i
39 - 41	9
36 - 38	18
33 - 35	26
30 - 32	23
27 – 29	16
24 - 26	8

Tabla 2. Puntuaciones de 5 sujetos en dos variables cuantitativas X e Y.

Sujeto	X	Y
1	8	6
2	16	4
3	20	7
4	24	8
5	32	10

Tabla 3. Valores de dos variables X e Y.

X	Y
1	5,5
2	6,2
3	8,7
4	9,5

Supuesto 1. Un test de laboratorio es efectivo al 95% detectando en una muestra de sangre cierta enfermedad cuando esta se padece ("verdadero positivo"). Sin embargo, el test también arroja un resultado "falso positivo" en el 1% de personas sanas sobre las que se realiza el test.

1. Un histograma para una variable cuantitativa agrupada en intervalos se construye: A) promediando los puntos medios de cada intervalo; **B) sobre los límites exactos de los intervalos**; C) restando a los límites exactos de cada intervalo el valor de sus límites aparentes.
2. La amplitud del intervalo de la distribución de la Tabla 1 es: **A) 3**; B) 2; C) 1.
3. El percentil 47 de la distribución de la Tabla 1 es: A) 29; B) 34; **C) 32,5**.
4. La media de la variable Y de la Tabla 2 es igual a: **A) 7;** B) 7,5; C) 6,75.
5. La desviación típica de la variable X de la Tabla 2 es igual a: A) 32; B) 64; **C) 8**.
6. El índice de asimetría de Pearson de la distribución de la Tabla 1 indica que: **A) la mayor parte de las puntuaciones se sitúan en la parte alta de la escala**; B) la mayor concentración de puntuaciones se produce en la parte baja de la escala; C) es una distribución simétrica.
7. Un índice de apuntamiento o curtosis igual a cero indica que la distribución es: **A) mesocúrtica**; B) platicúrtica; C) leptocúrtica.
8. De 100 alumnos que se examinan de dos asignaturas X e Y sabemos que: 70 alumnos aprueban ambas asignaturas, 10 suspenden las dos asignaturas, y 10 aprueban X pero suspenden Y. El valor absoluto del coeficiente ϕ (phi) es igual a: A) 0,50; B) 1; **C) 0,38**.
9. En la Tabla 2 se recogen las puntuaciones obtenidas por 5 sujetos en dos variables cuantitativas (X e Y). Si la covarianza entre X e Y vale 12,8. Entonces, el coeficiente de correlación de Pearson entre X e Y vale: A) 0,7; **B) 0,8**; C) -0,6.
10. En la Tabla 3 se muestran los valores de dos variables (X e Y). La recta de regresión de Y sobre X es: A) X' = 1,45 – 3,85 Y; B) Y' = 1,45 + 3,85 X; **C) Y' = 3,85 + 1,45 X**.

11. La proporción de la variabilidad de Y explicada por X según los datos de la Tabla 3 es igual a: A) 0,97; **B) 0,94**; C) 0,03.

12. El coeficiente de correlación más adecuado para relacionar dos variables cuantitativas que no tengan una distribución normal es el**: A) coeficiente de correlación de Spearman**; B) coeficiente de correlación de Pearson; C) coeficiente de correlación Biserial Puntual.

13. En el supuesto 1, y sabiendo que un 0,5% de la población mundial padece esta enfermedad. ¿Cuál es la probabilidad de que una persona padezca la enfermedad dado que el resultado de su análisis de sangre resultó positivo? Utilice el teorema de Bayes. A) 0,950; **B) 0,323**; C) 0,995.

14. En el supuesto 1, la especificidad de la prueba de laboratorio es: **A) 0,99**; B) 0,95; C) 0,01.

15. Los valores de una variable aleatoria discreta X son 0, 1, 2, 3, 4, 5. Si se sabe que $F(X \leq 4) = 0,974$ y que $F(X \leq 3)= 0,963$, entonces $P(X = 4)$ será: **A) 0,011**; B) 0,026; C) 0,001.

16. Sea Y una variable aleatoria discreta con valores 0, 1, 2 y 3. Si los cuatro valores de Y son equiprobables, su media es: A) 1,2; **B) 1,5**; C) 2.

17. En una distribución *F* de Snedecor con 8 grados de libertad en el numerador y 20 en el denominador, ¿cuál será el valor que deje por debajo al 90% de los casos?: A) 4,090; B) 2,425; **C) 1,999**.

18. Se ha convocado a 150 candidatos a una prueba de selección. La distribución de las calificaciones es normal y todos los candidatos conocen su puntuación tipificada. Si el evaluador invitase a salir del aula de exámenes a los candidatos con una puntuación típica inferior a -1,2 o superior a 2,45, ¿cuántos candidatos permanecerían, aproximadamente, en el aula?: A) 19; **B) 132**; C) 100.

19. En una variable que se distribuye normalmente, sabemos que la puntuación típica z = -2,65 supera las puntuaciones de 5 sujetos. ¿Cuál es el número total de sujetos? A) 400; **B) 1250**; C) 12500.

20. En una población con desviación típica igual a 20, extraemos muestras de tamaño 100. ¿Cuánto vale la varianza de la distribución muestral de la media? A) 2; **B) 4**; C) 0,02.

21. Cuando aplicamos el muestreo aleatorio sistemático: A) todos los elementos tienen la misma probabilidad de ser extraídos, y las extracciones son independientes; B) no todos los elementos tienen la misma probabilidad de ser extraídos, y las extracciones son independientes; **C) no todos los elementos tienen la misma probabilidad de ser extraídos, y las extracciones no son independientes.**

22. La distribución de las puntuaciones de un test de propensión a la Bulimia es normal con varianza 64. ¿Cuál es la probabilidad aproximada de que las varianzas muestrales se encuentren entre 41 y 91 si se extraen muestras de tamaño 30?: A) 0,60; **B) 0,85**; C) 0,95.

23. Se elige por muestreo aleatorio simple un grupo de 100 sujetos y se les administra un test de tolerancia a la frustración ante situaciones emocionalmente intensas. La media obtenida en la prueba fue 90. Se sabe que la varianza poblacional es 81. Si determinamos un intervalo de confianza del 99%, ¿entre qué valores se encontrará la verdadera media de la población? A) (88,236 ; 91,764); B) (86,687 ; 91,313); **C) (87,678 ; 92,322).**

24. Una muestra seleccionada al azar de profesores de secundaria indica que el 30% no está de acuerdo con las nuevas propuestas educativas. ¿Cuál es el tamaño muestral necesario para que el error máximo de estimación asociado al intervalo de confianza al 95% sea 0,06? Se asume población infinita. A) 200; **B) 225**; C) 250.

25. Cuando se realiza una estimación por intervalos, al estudiar la relación entre la amplitud de un intervalo y el nivel de confianza se puede comprobar que: A) a mayor confianza en la estimación mayor precisión; **B) a mayor confianza en la estimación menor precisión**; C) la precisión de la estimación no está relacionada con el nivel de confianza.

SOLUCIONES

1. B. páginas 33-34.

2. A

   Se toma cualquiera de los intervalos restando al límite superior exacto el límite inferior exacto. Ej, 41,5-38,5=3.

3. C.

Desarrollo de la Tabla 1

X	n_i	n_a
39 – 41	9	100
36 – 38	18	91
33 – 35	26	73
30 – 32	23	47
27 – 29	16	24
24 – 26	8	8
	$\sum n_i = 100$	

$$P_{47} = L_i + \left(\frac{\frac{n \cdot k}{100} - n_d}{n_c}\right) \cdot I = 29{,}5 + \left(\frac{\frac{100 \text{ x } 47}{100} - 24}{23}\right) \text{x } 3 = 32{,}5$$

4. A.

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{6+4+7+8+10}{5} = \frac{35}{5} = 7$$

5. C.

$$\overline{X} = \frac{\sum X_i}{n} = \frac{8+16+20+24+32}{5} = \frac{100}{5} = 20$$

$$S_x = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n}} = \sqrt{\frac{(8-20)^2 + (16-20)^2 + (20-20)^2 + (24-20)^2 + (32-20)^2}{5}};$$

$$S_x = \sqrt{\frac{144+16+0+16+144}{5}} = \sqrt{\frac{320}{5}} = \sqrt{64} = 8$$

6. A.

Desarrollo de la Tabla 1

X	n_i	Xi	n_i*X_i	$(n_i*X_i)^2$
39 – 41	9	40	360	14400
36 – 38	18	37	666	24642
33 – 35	26	34	884	30056
30 – 32	23	31	713	22103
27 – 29	16	28	448	12544
24 – 26	8	25	200	5000
∑	100		3271	108745

$$A_P = \frac{\overline{X} - Mo}{S_X};$$

$$\overline{X} = \frac{\sum n_i X_i}{n} = \frac{3271}{100} = 32{,}71;$$

$$S_x = \sqrt{\frac{\sum n_i X_i^2}{n} - \overline{X}^2} = \sqrt{\frac{108745}{100} - (32{,}71)^2} = 4{,}18;$$

Mo (datos agrupados) = punto medio del intervalo con mayor frecuencia

$A_P = \frac{32{,}71-34}{4{,}18} = -0{,}31$; Asimetría negativa

Páginas 38 y 118: Explicación de las asimetrías con un histograma.

7. A. Páginas 39 y 122

8. C

		Asignatura X		
		Suspende	Aprueba	
Asignatura Y	Suspende	10	10	20
	Aprueba	10	70	80
		20	80	100

$$\varphi = \frac{n_{11} \cdot n_{22} - n_{21} \cdot n_{12}}{\sqrt{n_{1+} \cdot n_{2+} \cdot n_{+1} \cdot n_{+2}}}$$

$$\varphi = \frac{10 \cdot 70 - 10 \cdot 10}{\sqrt{20 \cdot 80 \cdot 20 \cdot 80}} = \frac{600}{1600} = 0{,}375 \approx 0{,}38$$

9. B.

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

$$\overline{X} = \frac{\sum X_i}{n} = \frac{8 + 16 + 20 + 24 + 32}{5} = \frac{100}{5} = 20$$

$$S_x = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n}} = \sqrt{\frac{(8-20)^2 + (16-20)^2 + (20-20)^2 + (24-20)^2 + (32-20)^2}{5}};$$

$$S_x = \sqrt{\frac{144 + 16 + 0 + 16 + 144}{5}} = \sqrt{\frac{320}{5}} = \sqrt{64} = 8$$

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{6 + 4 + 7 + 8 + 10}{5} = \frac{35}{5} = 7$$

$$S_y = \sqrt{\frac{\sum (Y_i - \overline{Y})^2}{n}} = \sqrt{\frac{(6-7)^2 + (4-7)^2 + (7-7)^2 + (8-7)^2 + (10-7)^2}{5}};$$

$$S_y = \sqrt{\frac{1 + 9 + 0 + 1 + 9}{5}} = \sqrt{\frac{20}{5}} = \sqrt{4} = 2$$

$$r_{xy} = \frac{12{,}8}{8 * 2} = 0{,}8$$

10. A

Desarrollo de la tabla 3

X	Y	X^2	Y^2	XY
1	5,5	1	30,25	5,5
2	6,2	4	38,44	12,4
3	8,7	9	75,69	26,1
4	9,5	16	90,25	38
10	29,9	30	234,63	82

$$Y' = a + bX$$

$$b = \frac{n \cdot \sum XY - \sum X \sum Y}{n \cdot \sum X^2 - (\sum X)^2} = \frac{4 \cdot 82 - 10 \cdot 29{,}9}{4 \cdot 30 - 100} = \frac{29}{20} = 1{,}45$$

$$a = \overline{Y} - b \cdot \overline{X} = 7{,}48 - 1{,}45 \cdot 2{,}5 = 3{,}85$$

11. B.

$$r_{xy}^2$$

$$r_{xy} = \frac{n \cdot \sum XY - \sum X \sum Y}{\sqrt{n \cdot \sum X^2 - (\sum X)^2}\sqrt{n \cdot \sum Y^2 - (\sum Y)^2}} = \frac{4 \cdot 82 - 10 \cdot 29{,}9}{\sqrt{4 \cdot 30 - 100} \cdot \sqrt{4 \cdot 234{,}63 - 894{,}01}} = \frac{29}{29{,}81} = 0{,}97$$

$$r_{xy}^2 = 0{,}94$$

12. A. Página 170.

13. B.

Probabilidades asociadas a la Situación 1:

Definimos *E* como el evento que denota que la persona sometida al test sufra la enfermedad, y *P* como el evento que denota que el test ha resultado positivo:

Probabilidad de obtener un resultado positivo cuando se tiene la enfermedad;

$P(P\backslash E) = 0{,}95;$

Probabilidad de obtener un resultado positivo cuando no se tiene la enfermedad;

$P(P\backslash \bar{E}) = 0{,}01$

Probabilidad de estar enfermo y de no estar enfermo;

$P(E) = 0{,}005;\ P(\bar{E}) = 0{,}995.$

La probabilidad de que alguien que sufre la enfermedad obtenga un test con resultado positivo es: Teorema de Bayes

$$P(E\backslash P) = \frac{P(P \cap E)}{P(P)} = \frac{P(P\backslash E)P(E)}{P(P\backslash E)P(E) + P(P\backslash \bar{E})P(\bar{E})}$$

$$P(E\backslash P) = \frac{(0{,}95)(0{,}005)}{(0{,}95)(0{,}005) + (0{,}01)(0{,}995)} = 0{,}323$$

14. A. La especificidad hace referencia a 1-“falsos positivos”; es decir 1- 0,01= 0,99.

15. A. La función de distribución nos da el valor de probabilidad acumulado igual o menor. Por tanto, para obtener la porción que existe entre F(3) y F(4) restaremos la primera a la segunda obteniendo exclusivamente con qué probabilidad la variable aleatoria ha tomado el valor 4

F(4) = 0,974 y F(3)= 0,963;

P(x=4)= F(4)-F(3)= 0,974-0,963 = 0.011

16. B.

La función de probabilidad es:

Y	0	1	2	3
f(y)	0,25	0,25	0,25	0,25

$$\mu_Y = \sum y \cdot f(y) = 0 \cdot 0{,}25 + 1 \cdot 0{,}25 + 2 \cdot 0{,}25 + 3 \cdot 0{,}25 = 0 + 0{,}25 + 0{,}5 + 0{,}75 = 1{,}5$$

17. C

Para hallar el valor que deja por debajo al 90% de casos utilizaremos la tabla VII y el resultado es 1,999.

18. B. Obtenemos los porcentajes de candidatos que dejan por debajo las puntuaciones típicas que se aportan para así poder aislar el porcentaje de sujetos que nos interesa.

Así, Z = 2,45 deja por debajo una proporción de 0,9929 (99,29% de los sujetos); y Z = -1,2 deja por debajo una proporción de 0,1151 (11,51% de los sujetos)

Siendo esto así, 0,9929-0,1151= 0,8778 (87,78% de los sujetos), que de un total de 150 candidatos serán, aproximadamente, 132 los que se quedan en el aula de exámenes.

19. B. La puntuación típica z= -2,65 supera la puntuación de 5 sujetos, que es lo mismo que decir que el porcentaje de sujetos sobre el total que deja por debajo esa puntuación es el 0,4% (0,0040 en la tabla)

A su vez, si el 0,4% corresponde a 5 sujetos, el total de sujetos será igual a 1250.

20. B. Conocemos la varianza poblacional, ya que la desviación típica es 20, la varianza será 400. Aunque no utilizaremos esta información.
Para calcular la varianza de la distribución muestral se aplicará. $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{10} = 2$
Siendo entonces 2 la desviación típica de la distribución muestral de la media, la varianza será su cuadrado. Esto es, 4.

21. C. Página 381.

22. B. Sabemos que:

$$\chi^2 = \frac{nS_x^2}{\sigma^2} se\ distribuye\ \chi^2_{n-1}$$

Por tanto, calculamos los valores de χ^2_{n-1} para S_x^2=41 y S_x^2=91

$$\chi^2_{n-1} = \frac{nS_x^2}{\sigma^2} = \frac{30 \cdot 41}{64} = 19{,}22$$

$$\chi^2_{n-1} = \frac{nS_x^2}{\sigma^2} = \frac{30 \cdot 91}{64} = 42{,}66$$

Para χ^2_{29} =19,22 la probabilidad aproximada según la tabla χ^2 es 0,100

Para χ^2_{29} =42,66 la probabilidad aproximada según la tabla χ^2 es 0,950

Por tanto, $P(19{,}22 \leq S_x^2 \leq 42{,}66) = 0{,}950 - 0{,}100 = 0{,}85$

23. C. Intervalo de confianza para el parámetro media con varianza conocida

$$L_{inf} = \overline{X} - \left|Z\alpha_{/2}\right| \frac{\sigma}{\sqrt{n}} = 90 - 2{,}58 \cdot \frac{9}{10} = 87{,}678$$

$$L_{sup} = \overline{X} + \left|Z\alpha_{/2}\right| \frac{\sigma}{\sqrt{n}} = 90 + 2{,}58 \cdot \frac{9}{10} = 92{,}322$$

24. B.

$$n = \frac{z^2_{\alpha/2}\, P\,(1\ - P)}{E^2_{max}} = \frac{1{,}96^2 \text{ x } 0{,}30 \text{ x } 0{,}70}{0{,}06^2} = 224{,}17 \simeq 225$$

25. B. Páginas 441- 442.

Código asignatura	Nombre asignatura
62011037	**INTRODUCCIÓN AL ANÁLISIS DE DATOS**
Fecha alta y origen	Convocatoria
15/02/2022	**FEBRERO 2022 – TIPO B**
Curso virtual	

PSICOTEST UNED

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2022 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO B DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (no programable)**
Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Gráfico 1. Resultados obtenidos por un grupo de 50 varones en una prueba de orientación espacial (X).

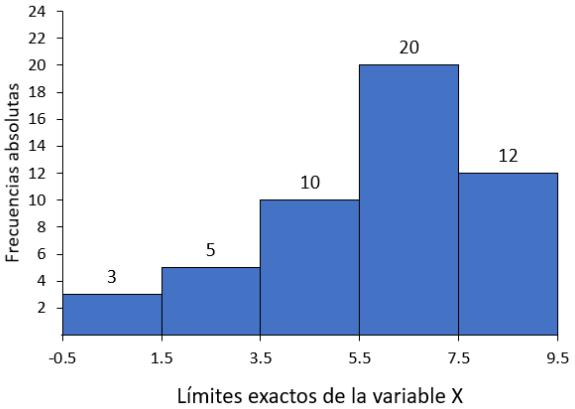


Tabla 1. Datos de 5 alumnos de Secundaria en un estudio para pronosticar la nota en Matemáticas (Y) a partir de las puntuaciones en un test de aptitud numérica (X).

Alumno	Test Aptitud Numérica (X)	Nota en Matemáticas (Y)
1	30	4
2	40	5
3	80	9
4	60	7
5	50	6

Tabla 2. Datos de 100 mujeres empresarias en un estudio para pronosticar el éxito profesional (Y) a partir de las puntuaciones en un test de habilidades sociales (X).

	Media	Desviación típica	Pendiente de la recta de regresión de Y sobre X
X	7	2	b= 1,35
Y	8	3	

Tabla 3. Variables y características de su distribución.

Variable	Distribución	
X	N(40,2)	Normal con media 40 y desviación típica 2
Y	t_{18}	t de Student con 18 grados de libertad
V	$F_{20,6}$	F con 20 grados de libertad en el numerador y 6 grados de libertad en el denominador

1. La moda es: A) uno de los índices estadísticos apropiados para las variables cualitativas; **B) el único índice apropiado para variables cualitativas**; C) un índice de dispersión.
2. La variable X del Gráfico 1 sería: A) de razón; **B) de intervalos**; C) discreta.
3. Con los datos Gráfico 1, la moda para los varones es: A) 20; B) 2,5; **C) 6,5**.
4. Siguiendo el Gráfico 1, si un varón obtiene una puntuación de 7 en orientación espacial, el percentil que le corresponde es: **A) 66**; B) 54; C) 32.
5. Con los datos del Gráfico 1, el Índice de Asimetría de Pearson calculado para la distribución de varones es, aproximadamente: A) -0,545; B) 0,302; **C) -0,302.**
6. La puntuación típica para un varón que ha obtenido una puntuación directa en orientación espacial de 8, con los datos del Gráfico 1 es: **A) 0,97**; B) 1,95; C) -1,95.
7. El estudio de la relación entre dos variables cualitativas se puede llevar a cabo observando: A) las frecuencias marginales de cada una de las dos variables por separado; B) las frecuencias conjuntas de ambas variables; **C) las distribuciones condicionadas de una de las dos variables agrupadas en función de los valores de la otra.**
8. El coeficiente de correlación por rangos de Spearman: A) es apropiado para variables nominales; B) es apropiado para variables cuantitativas que tengan una distribución normal; **C) oscila entre -1 y +1.**

PSICOTEST UNED

9. Según los datos de la Tabla 1, el valor de la asociación entre las variables, evaluada a través del coeficiente de correlación de Spearman, es: A) 0; **B) 1**; C) 0,95.
10. Según la Tabla 2, la covarianza entre las variables X e Y es**: A) 5,4**; B) 0,90; C) 4,5.
11. Con los datos de la Tabla 2, el coeficiente de correlación de Pearson entre X e Y es: A) 0,85; **B) 0,90**; C) 0,95.
12. Siguiendo los datos de la Tabla 2, ¿qué puntuación en éxito profesional le pronosticaremos a una mujer empresaria que ha obtenido una puntuación de 4 en habilidades sociales? A) 4,55; B) 2,45; **C) 3,95**.
13. Si P(A U B) = P(A) + P(B) diremos que los sucesos A y B son: **A) mutuamente excluyentes**; B) compatibles; **C) dependientes.** * Dado que las alternativas A y C son correctas, esta pregunta se dará por válida a todos los estudiantes, la hayan respondido o no.
14. Tráfico ha estimado que la probabilidad de multa por "exceso de velocidad", P(ev), es 0,50, por "saltarse un semáforo", P(ss), es de 0,10 y por el "resto de infracciones", P(r), es de 0,40. También se conoce que la probabilidad de "accidente por exceso de velocidad" es de 0,80; la de "accidente por saltarse un semáforo" es de 0,70 y la probabilidad de "accidente por otras infracciones" es de 0,30. La probabilidad total de tener un accidente por las tres causas es: A) 0,40; **B) 0,59**; C) 0,12.
15. Siguiendo con la pregunta anterior, ha ocurrido un accidente, la probabilidad de que haya sido por saltarse un semáforo es: **A) 0,1186**; B) 0,8814; C) 0,6780.
16. En Psicología de la Salud, se entiende por "prevalencia": **A) la proporción de personas que tienen una enfermedad en un momento determinado respecto de la población**; B) la proporción de casos nuevos de una enfermedad en una población, durante un período determinado; C) la "incidencia" del trastorno.
17. Si la función de probabilidad de una variable discreta X es f(1) = 0,25; f(2) = 0,35; f(3) = 0,20; f(4) = 0,20. La media de X es: A) 4,22; B) 6,45; **C) 2,35**.
18. Se sabe que el 20% de la población de estudiantes sufre ansiedad previa ante los exámenes. Si elegimos al azar una muestra de 8 estudiantes de un centro, ¿cuál es la probabilidad de que más de cuatro estudiantes sufran de ansiedad previa a los exámenes? A) 0,9896; B) 0,0459; **C) 0,0104**.
19. Con los datos de la Tabla 3, la probabilidad de que X sea menor o igual que 42 es aproximadamente**: A) 0,8413**; B) 0,1587; C) 0,9772.
20. Teniendo en cuenta la Tabla 3, para la variable Y el valor de probabilidad 0,688 corresponde al percentil: A) 25; **B) 75**; C) 50. *Dado que hay una errata en el enunciado (sobra la palabra probabilidad), esta pregunta se dará por válida a todos los estudiantes, la hayan respondido o no.
21. Atendiendo a los datos de la Tabla 3, ¿cuál es el percentil 10 de la variable V? A) 2,091; B) 0,353; **C) 0,478.**
22. ¿Cuál de estas afirmaciones es la correcta? A) La media muestral es un estimador insesgado y consistente, pero no suficiente de la media poblacional; B) La proporción muestral no se considera un buen estimador de la proporción poblacional; **C) La cuasivarianza muestral es un estimador insesgado de la varianza poblacional.**
23. Se sabe que el 22% de la población adulta española fuma. Si se selecciona una muestra aleatoria de 250 adultos españoles, ¿cuál es la probabilidad de que menos del 20% de la muestra sean fumadores? **A) 0,2005**; B) 0,7995; C) 0,2018.
24. En una muestra seleccionada al azar de 400 adolescentes se observó que la mitad jugaba a diario con consolas de videojuegos. ¿Cuál es el intervalo de confianza al 99% relativo a la proporción de adolescentes que juegan a diario con este tipo de consolas? A) 0,36 y 0,48; B) 0,44 y 0,65; **C) 0,44 y 0,56**.

PSICOTEST UNED

25. Sabiendo que P= 0,20, ¿cuál es el tamaño muestral necesario para que el $E_{máx}$ asociado al intervalo de confianza al 95% sea 0,08? Se asume población infinita. **A) 96**; B) 45; C) 84.

SOLUCIONES:

1. B (p.75)
2. B
3. A

   Mo (varones) = 6,5 (media del intervalo más frecuente, que es (6-7), con n_i = 20).

4. A (p. 83)

X	n_i	n_a
8-9	12	50
6-7	20	38
4-5	10	18
2-3	5	8
0-1	3	3
Σ	50	

Si X_i= 7, entonces el IC es (6-7) y la I= 7,5 – 5,5 = 2.

$$k = \left[\frac{\frac{(P_k - L_i)n_i}{I} + n_d}{n}\right]100 = \left[\frac{\frac{(7-5{,}5)20}{2} + 18}{50}\right]100 = 66$$

5. C (p. 117)

X	X_i	n (varones)	nX_i	X^2_i	nX^2_i
8-9	8,5	12	102	72,25	867
6-7	6,5	20	130	42,25	845
4-5	4,5	10	45	20,25	202,5
2-3	2,5	5	12,5	6,25	31,25
0-1	0,5	3	1,5	0,25	0,75
Σ		50	291		1946,5

$\bar{X} = \frac{\sum n_i X_i}{\sum n_i} = \frac{\sum n_i X_i}{n}$= 291/50= 5,82

$S_x^2 = \frac{\sum n_i X_i^2}{\sum n_i} - \bar{X}^2 = \frac{\sum n_i X_i^2}{n} - \bar{X}^2$= 1946,5/50 – $(5{,}82)^2$= 5,0576; S_x= $\sqrt{5{,}0576}$= 2,25

$$Ap = \frac{\bar{X} - Mo}{Sx} = \frac{5{,}82 - 6{,}5}{2{,}25} = -0{,}302$$

6. A
   (p. 128)

PSICOTEST UNED

$$z = \frac{X - \bar{X}}{S_X} = \frac{8 - 5{,}82}{2{,}25} = 0{,}97$$

7. C

8. C (170, 171)

9. B

Rangos(X)	Rangos (Y)	d	d_i^2
1	1	0	0
2	2	0	0
5	5	0	0
4	4	0	0
3	3	0	0

Spearman (p. 171) → $r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$ = 1 - 0 = 1

Existe una correlación perfecta entre ambas variables. A mayor aptitud numérica, mayor nota en Matemáticas.

10. A (pp. 197, 198, 220)

$r_{xy} = b\frac{S_x}{S_y}$; $r_{xy} = 1{,}35\frac{2}{3} = 0{,}90$

$r_{xy} = \frac{S_{xy}}{S_x S_y}$; $S_{xy} = r_{xy} S_x S_y = 0{,}90 \times 2 \times 3 = 5{,}4$

11. B

12. C (p. 221)

a= $\bar{Y}$– b($\bar{X}$)= 8- 1,35(7)= -1,45

Y´= a + bX= -1,45+ 1,35(4)= 3,95

.

13. A y C (p. 265) * Dado que las alternativas A y C son correctas, esta pregunta se dará por válida a todos los estudiantes, la hayan respondido o no

14. B

Teorema de la Probabilidad Total: $P(B) = \sum_{i=1}^{k} P(A_i) \cdot P(B|A_i)$

P(ev)= 0,50; P(A|ev)= 0,80

P(ss)= 0,10; P(A|ss)= 0,70

PSICOTEST UNED

P(r)= 0,40; P(A|r)= 0,30

P(A)= 0,50 x 0,80 + 0,10 x 0,70 + 0,40 x 0,30= 0,59

15. A

Teorema de Bayes: $P(A_i|B) = \frac{P(A_i)\cdot P(B|A_i)}{P(B)}$

P(ss|A)= (P(ss)P(A|ss))/P(A)= (0,10 x 0,70)/0,59= 0,1186

16. A
17. C

f(1)= 0,25; f(2)= 0,35; f(3)= 0,20; f(4)= 0,20

μ = E(X) = ∑xf(x) = 1(0,25) + 2(0,35) + 3(0,20) + 4(0,20)= 2,35

18. C
Tabla II de la Binomial con valores: p= 0,20; n=8.
P(X> 4)= 1 – P(X≤ $4) = 1 - F(4) =$ 1- 0,9896= 0,0104

19. A
Z= (42-40)/2= 1
P(X≤ 42)= P(Z≤ 1)= 0,8413 (Tabla IV)

20. B *Dado que hay una errata en el enunciado (sobra la palabra probabilidad), esta pregunta se dará por válida a todos los estudiantes, la hayan respondido o no

Tabla VI (*t* de Student)
Para gl=18, el valor 0,688 se corresponde con p=0,75
P(Y≤ 0,688) = 0,75. Por tanto, Y= 0,688 es el percentil 75.

21. C

$_{0,10}F_{20,6}$=1 / $_{0,90}F_{6,20}$= 1/ 2,091= 0,478

22. C

23. A

El 20% de 250 es 50 [(20 x 250)/100], por lo que en esta pregunta hay que calcular la probabilidad de que menos de 50 personas sean fumadoras: P(X < 50)

Como el tamaño de la muestra es grande (n > 20), podemos utilizar la aproximación de la binomial a la distribución normal para solucionar este ejercicio. Para ello, primero necesitamos hallar la media, μ, y la desviación típica, σ: (ver fórmula en la página 353 del manual):

μ = *np* = 250 x 0,22 = 55

σ = $\sqrt{npq} = \sqrt{250\ x\ 0{,}22\ x\ 0{,}78} = 6{,}55$

PSICOTEST UNED

Para hallar la probabilidad que se pide aproximando la distribución binominal a la distribución normal, tenemos que aplicar una pequeña corrección, restando 0,5 a nuestro valor: P(X < 49,5)

Ahora se procede a calcular la puntuación Z correspondiente, utilizando para ello la media y desviación típica que hemos calculado previamente:

P(X < 49,5) = P(Z ≤ [(49,5 - 55)/6,55]) = P(Z ≤ -0,84)

En la Tabla III podemos ver que la probabilidad que queda por debajo de una puntuación Z de -0,84 es 0,2005: P(X < 49,5) = P(Z ≤ [(49,5-55)/6,55]) = P(Z ≤ -0,84) = 0,2005

24. C

Intervalo de confianza la para el parámetro π (aproximación a la Normal); α= 0,01

P= 200/400= 0,5; $z_{\alpha/2} = z_{0,005} = -2,58$ y $z_{1-\alpha/2} = z_{0,995}$= 2,58

L_i= P- $z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}}$= 0,5-2,58 $\sqrt{\frac{0,5(0,5)}{400}}$= 0,44

L_s= P+ $z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}}$= 0,5+2,58 $\sqrt{\frac{0,5(0,5)}{400}}$= 0,56

25. A (p. 465)

Tamaño muestral para π ; $Z_{0,025}$= -1,96

$$n= \frac{\boldsymbol{Z^2_{\alpha/2}} P(1-P)}{\boldsymbol{E^2_{max}}} = \frac{(-1,96)^2(0,20)(0,80)}{(0,08)^2} \approx 96$$

PSICOTEST UNED

Código asignatura	Nombre asignatura
62011037	**INTRODUCCIÓN AL ANÁLISIS DE DATOS**
Fecha alta y origen	Convocatoria
15/02/2022	**FEBRERO 2022 – TIPO E**
Curso virtual	

PSICOTEST UNED

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2022 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO E DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (no programable)**
Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Tabla 1. Resultados obtenidos por un grupo de 40 niñas y 40 niños de 7 años en una prueba de lectoescritura (X).

X	Niñas	Niños
8-9	7	9
6-7	11	9
4-5	9	13
2-3	7	5
0-1	6	4

Tabla 2. Información obtenida en un estudio sobre la depresión y el cuidado de mayores realizado con 68 mujeres.

		Cuidado de mayores	
		Sí	No
Depresión	Sí	20	10
	No	14	24

Tabla 3. Datos de 5 alumnos de la UNED en un estudio para pronosticar la nota en Fundamentos de Investigación (Y) a partir de las puntuaciones en un test de inteligencia (X).

Alumno	**Test Inteligencia (X)**	**Notas Fundamentos (Y)**
1	100	4
2	110	5
3	130	9
4	120	7
5	115	6

Tabla 4. Variables y características de su distribución.

Variable	**Distribución**	
X	N(30,3)	Normal con media 30 y desviación típica 3
Y	t_{16}	t de Student con 16 grados de libertad
V	X^2_{21}	Ji-cuadrado con 21 grados de libertad

1. El gráfico más apropiado para representar la variable X de la Tabla 1 sería: A) el diagrama de barras; B) el diagrama de sectores; **C) el histograma**.
2. Con los datos de la Tabla 1, la media en X para las niñas es: A) igual a la media para los niños; B) mayor que la media para los niños; **C) menor que la media para los niños**.
3. Uno de los índices estadísticos apropiados para las variables ordinales o cuasicuantitativas es: A) la media; **B) la moda**; C) la desviación típica.
4. La media ponderada para el total de la muestra de personas (niños y niñas) que han realizado la prueba de lectoescritura de la Tabla 1 es: **A) 5**; B) 10; C) 2,5.
5. Sabiendo que las varianzas de las niñas y los niños de la Tabla 1 son, respectivamente, 3,2 y 10,37, ¿qué grupo presenta mayor variabilidad en su distribución? Utilice los coeficientes de variación. A) Las niñas con CV_1= 66,67; B) Los niños con CV_2=199,42; **C) Los niños con CV_2=61,92**.
6. Según la Tabla 1, la amplitud intercuartil para los niños es: A) 2; B) 7,28; **C) 3,63**.
7. Con los datos de la Tabla 2, el valor aproximado de Ji-cuadrado es: **A) 5,965**; B) 7,525; C) 9,362.
8. Siguiendo con los datos de la Tabla 2, el valor de la asociación entre las variables, evaluada a través del Coeficiente C de Contingencia y su $C_{máx}$ son: **A) 0,284 y 0,707**; B) 0,316 y 0,707; C) 0,284 y 0,816.
9. Según los datos de la Tabla 2, los porcentajes de la "Depresión" condicionados al "Sí Cuidado Mayores" es: A) 66,67% y 33,33%; **B) 58,82% y 41,18%;** C) 30% y 38%.
10. Atendiendo los datos de la Tabla 2, si hemos elegido una mujer al azar, la probabilidad de que no padezca depresión es: **A) 0,56**; B) 0,37; C) 0,21.

PSICOTEST UNED

11. Se sabe que las medias de X e Y son, respectivamente, 115 y 6,2, y que S_x = 10 y S_y = 1,72. Utilizando los datos de la Tabla 3, la covarianza entre las variables X e Y es: A) 25; **B) 17**; C) 14.
12. Con la información de la pregunta anterior y los datos de la Tabla 3, el coeficiente de correlación entre X e Y es igual a: A) 0,814; B) 0,777; **C) 0,988.**
13. Siguiendo los datos de la Tabla 3, ¿qué nota aproximada de Fundamentos de Investigación le pronosticaremos a un alumno que ha obtenido una puntuación de 115 en el Test de Inteligencia? A) 3,9; B) 2,8; **C) 6,2**.
14. Según los datos de la Tabla 2, si elegimos al azar una mujer, la probabilidad de que cuide de mayores y tenga depresión es, aproximadamente: A) 0,67; B) 0,59; **C) 0,29**.
15. En un experimento de condicionamiento animal se sitúa a una rata en el centro de un laberinto en forma de T. En cada ensayo la rata elige siempre una de las tres salidas A, B o C, con igual probabilidad. El suelo de cada una de estas salidas tiene una rejilla eléctrica que dispensa una descarga (D) de 5 voltios con distintas probabilidades: 1/4 para la salida A, ½ para la salida B y 0 para la C. En un determinado ensayo la rata no recibió descarga. La probabilidad total de no recibir descarga es, aproximadamente: A) 0,3333; **B) 0,7425**; C) 0,2475.
16. Siguiendo con la pregunta anterior, la probabilidad de que la rata haya salido por A dado que no recibió la descarga es: A) 0,75; **B) 0,3333**; C) 1.
17. Si la función de probabilidad de una variable discreta X es: f(0) = 0,25; f(1) = 0,35; f(2) = 0,40. La varianza de X es, aproximadamente: A) 1,95; B) 1,15; **C) 0,63**.
18. Se sabe que el 85% de la población de deportistas de élite siente ansiedad de forma previa a la competición. Si elegimos al azar una muestra de 18 deportistas de élite, ¿cuál es la probabilidad de que 10 de ellos tengan ansiedad precompetitiva? A) 0,9995; B) 0,0001; **C) 0,0022.**
19. Con los datos de la Tabla 4, si a una persona le corresponde el percentil 67, su puntuación en la variable X será igual a: A) 40,20; B) 50,12; **C) 31,32**.
20. Teniendo en cuenta la Tabla 4, para la variable Y el valor de probabilidad 0,865 corresponde al percentil: A) 20; **B) 80**; C) 50.
21. Atendiendo a los datos de la Tabla 4, ¿cuál es la probabilidad de que la variable V adopte valores menores o iguales a 11,5913? A) 0,95; B) 0,50; **C) 0,05**.
22. El Teorema del Límite Central: A) permite calcular las probabilidades asociadas a los valores de las medias siempre que se conozca la forma de la distribución de las variables; **B) es útil para calcular probabilidades cuando las variables no se ajustan a la normal, siempre que las muestras tengan tamaño suficiente (n≥30)**; C) se utiliza en los casos en que la variable se distribuye normalmente con varianza poblacional desconocida.
23. Se sabe que el 17% de la población adulta española tiene obesidad. Si se selecciona una muestra aleatoria de 300 adultos españoles, ¿cuál es la probabilidad de que menos del 15% de la muestra sean obesos? A) 0,2543; **B) 0,1788**; C) 0,8212.
24. A partir de los datos en extraversión de una muestra de 91 personas, cuya media es 15 y cuasivarianza de 4,25, ¿entre qué límites se encuentra la varianza poblacional para un nivel de confianza del 90%? A) 1,37 y 3,53; B) 2,37 y 4,53; **C) 3,38 y 5,53**.
25. Se sabe que la subescala de Depresión de la escala GADS de Goldberg es una variable que se distribuye $N(\mu, 3)$. Se extrae por m.a.s. una muestra de 55 adolescentes con autismo. ¿Qué error máximo cometeremos con ese tamaño muestral (asumiendo población finita, N=50000 y α=0,01) en la estimación de la media en Depresión de la población de adolescentes con autismo de España? **A) 1,04**; B) 1,57; C) 2,16.

PSICOTEST UNED

Código asignatura	Nombre asignatura
62011037	**INTRODUCCIÓN AL ANÁLISIS DE DATOS**
Fecha alta y origen	Convocatoria
15/02/2022	**FEBRERO 2022 – TIPO E**
Curso virtual	

PSICOTEST UNED

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2022 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO E DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (no programable)**
Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Tabla 1. Resultados obtenidos por un grupo de 40 niñas y 40 niños de 7 años en una prueba de lectoescritura (X).

X	Niñas	Niños
8-9	7	9
6-7	11	9
4-5	9	13
2-3	7	5
0-1	6	4

Tabla 2. Información obtenida en un estudio sobre la depresión y el cuidado de mayores realizado con 68 mujeres.

		Cuidado de mayores	
		Sí	No
Depresión	Sí	20	10
	No	14	24

Tabla 3. Datos de 5 alumnos de la UNED en un estudio para pronosticar la nota en Fundamentos de Investigación (Y) a partir de las puntuaciones en un test de inteligencia (X).

Alumno	**Test Inteligencia (X)**	**Notas Fundamentos (Y)**
1	100	4
2	110	5
3	130	9
4	120	7
5	115	6

Tabla 4. Variables y características de su distribución.

Variable	**Distribución**	
X	N(30,3)	Normal con media 30 y desviación típica 3
Y	t_{16}	t de Student con 16 grados de libertad
V	X^2_{21}	Ji-cuadrado con 21 grados de libertad

1. El gráfico más apropiado para representar la variable X de la Tabla 1 sería: A) el diagrama de barras; B) el diagrama de sectores; **C) el histograma**.
2. Con los datos de la Tabla 1, la media en X para las niñas es: A) igual a la media para los niños; B) mayor que la media para los niños; **C) menor que la media para los niños**.
3. Uno de los índices estadísticos apropiados para las variables ordinales o cuasicuantitativas es: A) la media; **B) la moda**; C) la desviación típica.
4. La media ponderada para el total de la muestra de personas (niños y niñas) que han realizado la prueba de lectoescritura de la Tabla 1 es: **A) 5**; B) 10; C) 2,5.
5. Sabiendo que las varianzas de las niñas y los niños de la Tabla 1 son, respectivamente, 3,2 y 10,37, ¿qué grupo presenta mayor variabilidad en su distribución? Utilice los coeficientes de variación. A) Las niñas con CV_1= 66,67; B) Los niños con CV_2=199,42; **C) Los niños con CV_2=61,92**.
6. Según la Tabla 1, la amplitud intercuartil para los niños es: A) 2; B) 7,28; **C) 3,63**.
7. Con los datos de la Tabla 2, el valor aproximado de Ji-cuadrado es: **A) 5,965**; B) 7,525; C) 9,362.
8. Siguiendo con los datos de la Tabla 2, el valor de la asociación entre las variables, evaluada a través del Coeficiente C de Contingencia y su $C_{máx}$ son: **A) 0,284 y 0,707**; B) 0,316 y 0,707; C) 0,284 y 0,816.
9. Según los datos de la Tabla 2, los porcentajes de la "Depresión" condicionados al "Sí Cuidado Mayores" es: A) 66,67% y 33,33%; **B) 58,82% y 41,18%;** C) 30% y 38%.
10. Atendiendo los datos de la Tabla 2, si hemos elegido una mujer al azar, la probabilidad de que no padezca depresión es: **A) 0,56**; B) 0,37; C) 0,21.

PSICOTEST UNED

11. Se sabe que las medias de X e Y son, respectivamente, 115 y 6,2, y que S_x = 10 y S_y = 1,72. Utilizando los datos de la Tabla 3, la covarianza entre las variables X e Y es: A) 25; **B) 17**; C) 14.
12. Con la información de la pregunta anterior y los datos de la Tabla 3, el coeficiente de correlación entre X e Y es igual a: A) 0,814; B) 0,777; **C) 0,988.**
13. Siguiendo los datos de la Tabla 3, ¿qué nota aproximada de Fundamentos de Investigación le pronosticaremos a un alumno que ha obtenido una puntuación de 115 en el Test de Inteligencia? A) 3,9; B) 2,8; **C) 6,2**.
14. Según los datos de la Tabla 2, si elegimos al azar una mujer, la probabilidad de que cuide de mayores y tenga depresión es, aproximadamente: A) 0,67; B) 0,59; **C) 0,29**.
15. En un experimento de condicionamiento animal se sitúa a una rata en el centro de un laberinto en forma de T. En cada ensayo la rata elige siempre una de las tres salidas A, B o C, con igual probabilidad. El suelo de cada una de estas salidas tiene una rejilla eléctrica que dispensa una descarga (D) de 5 voltios con distintas probabilidades: 1/4 para la salida A, ½ para la salida B y 0 para la C. En un determinado ensayo la rata no recibió descarga. La probabilidad total de no recibir descarga es, aproximadamente: A) 0,3333; **B) 0,7425**; C) 0,2475.
16. Siguiendo con la pregunta anterior, la probabilidad de que la rata haya salido por A dado que no recibió la descarga es: A) 0,75; **B) 0,3333**; C) 1.
17. Si la función de probabilidad de una variable discreta X es: f(0) = 0,25; f(1) = 0,35; f(2) = 0,40. La varianza de X es, aproximadamente: A) 1,95; B) 1,15; **C) 0,63**.
18. Se sabe que el 85% de la población de deportistas de élite siente ansiedad de forma previa a la competición. Si elegimos al azar una muestra de 18 deportistas de élite, ¿cuál es la probabilidad de que 10 de ellos tengan ansiedad precompetitiva? A) 0,9995; B) 0,0001; **C) 0,0022.**
19. Con los datos de la Tabla 4, si a una persona le corresponde el percentil 67, su puntuación en la variable X será igual a: A) 40,20; B) 50,12; **C) 31,32**.
20. Teniendo en cuenta la Tabla 4, para la variable Y el valor de probabilidad 0,865 corresponde al percentil: A) 20; **B) 80**; C) 50.
21. Atendiendo a los datos de la Tabla 4, ¿cuál es la probabilidad de que la variable V adopte valores menores o iguales a 11,5913? A) 0,95; B) 0,50; **C) 0,05**.
22. El Teorema del Límite Central: A) permite calcular las probabilidades asociadas a los valores de las medias siempre que se conozca la forma de la distribución de las variables; **B) es útil para calcular probabilidades cuando las variables no se ajustan a la normal, siempre que las muestras tengan tamaño suficiente (n≥30)**; C) se utiliza en los casos en que la variable se distribuye normalmente con varianza poblacional desconocida.
23. Se sabe que el 17% de la población adulta española tiene obesidad. Si se selecciona una muestra aleatoria de 300 adultos españoles, ¿cuál es la probabilidad de que menos del 15% de la muestra sean obesos? A) 0,2543; **B) 0,1788**; C) 0,8212.
24. A partir de los datos en extraversión de una muestra de 91 personas, cuya media es 15 y cuasivarianza de 4,25, ¿entre qué límites se encuentra la varianza poblacional para un nivel de confianza del 90%? A) 1,37 y 3,53; B) 2,37 y 4,53; **C) 3,38 y 5,53**.
25. Se sabe que la subescala de Depresión de la escala GADS de Goldberg es una variable que se distribuye $N(\mu, 3)$. Se extrae por m.a.s. una muestra de 55 adolescentes con autismo. ¿Qué error máximo cometeremos con ese tamaño muestral (asumiendo población finita, N=50000 y α=0,01) en la estimación de la media en Depresión de la población de adolescentes con autismo de España? **A) 1,04**; B) 1,57; C) 2,16.

PSICOTEST UNED

Examen realizado

Asignatura: INTRODUCCIÓN AL ANÁLISIS DE DATOS 09/09/2021 16:00
Estudiante: [redacted]

El examen de Introducción al análisis de datos constará de 20 preguntas de tres alternativas de respuesta y tendrá una duración de 90 minutos. El examen contendrá tanto preguntas teóricas como prácticas, y para su evaluación se tendrá en cuenta la fórmula de corrección del azar, por lo que cada acierto sumará 0,50 y cada fallo restará 0,25 (las preguntas que se dejen en blanco ni suman ni restan en la fórmula). A la hora de realizar las operaciones del examen, es preciso redondear todas las operaciones de los cálculos intermedios a dos decimales (cuatro en el caso de los ejercicios referentes a probabilidad). Aun así, es posible que, por problemas de redondeo, ninguna de las tres opciones de respuesta coincida exactamente con el resultado calculado, en ese caso recomendamos seleccionar la opción más aproximada. Para realizar estas operaciones se podrá utilizar una hoja en blanco que no hay que entregar en la aplicación. El material autorizado para realizar el examen será el formulario de la asignatura (sin anotaciones, aunque sí se puede subrayar) y una calculadora no programable. Se puede utilizar la fe de erratas del formulario impresa.
El valor de cada pregunta de test acertada es 0,5
El valor de cada pregunta de test fallada es -0,25

Sus respuestas a preguntas de test de test aparecen marcadas en verde claro si son correctas, en rojo si son incorrectas. Las soluciones se muestran con un borde verde.

A continuación se muestra el examen

Pregunta 1

La variable número de hijos de la Tabla 12 es:

Tabla 12. Número de hijos, X, en un grupo de 200 mujeres:

X_i	n_i
4	10
3	40
2	50
1	60
0	40

A continua

B discreta

C dicotómica

Se ha seleccionado la opción B. La opción correcta es la B.

Pregunta 2

La propiedad $0 \leq P(A) \leq 1$ es válida:

A sólo para la definición estadística de la probabilidad

B sólo para la definición clásica de la probabilidad

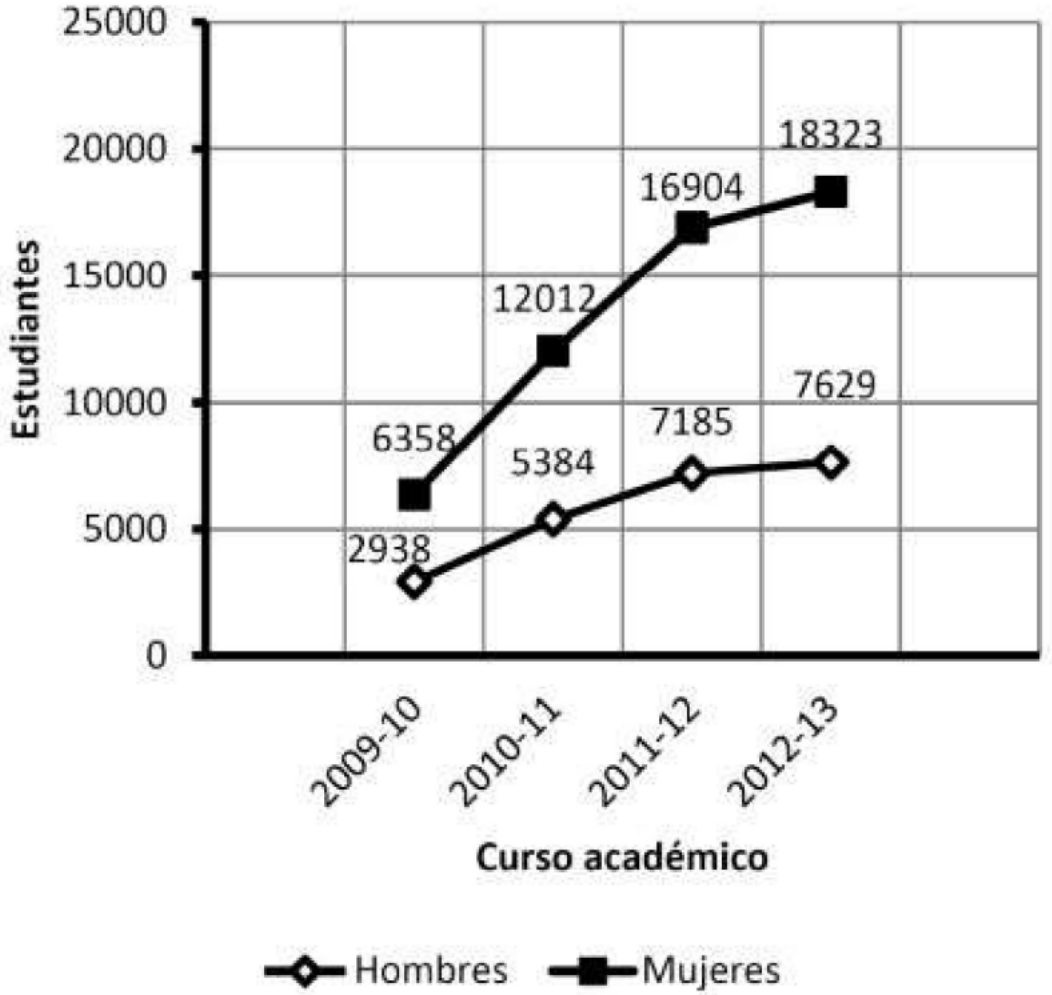
C para las definiciones clásica y estadística de la probabilidad

Se ha seleccionado la opción C. La opción correcta es la C.

Pregunta 3

Teniendo en cuenta el Gráfico 3, el número de mujeres matriculadas a lo largo de los cuatro cursos académicos:

Gráfico 3: Número de estudiantes (Hombres y Mujeres) matriculados en el Grado en Psicología durante cuatro cursos académicos.



A ha ido aumentando en menor medida que el número de hombres

B ha ido disminuyendo

C ha ido aumentando en mayor medida que el número de hombres

Se ha seleccionado la opción C. La opción correcta es la C.

Pregunta 4

Para calcular el rango o amplitud total de una variable necesitamos:

A las frecuencias absolutas

B que la variable sea cualitativa

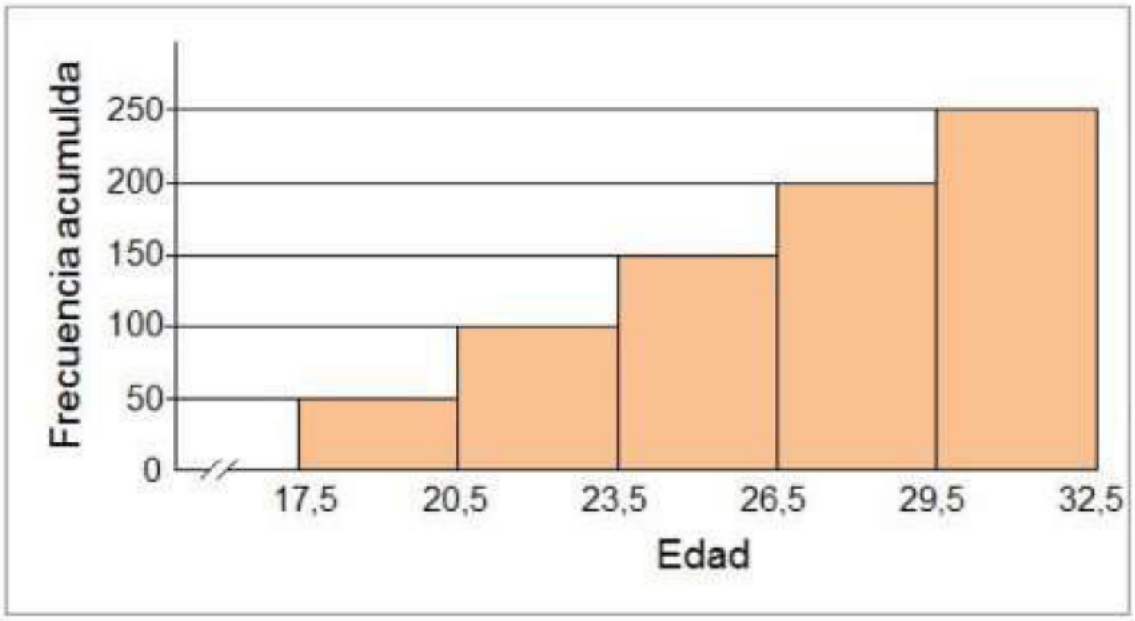
C las puntuaciones máxima y mínima

Se ha seleccionado la opción C. La opción correcta es la C.

Pregunta 5

En la situación 1, el número de sujetos con edades comprendidas entre 20,5 y 23,5 es:

Situación 1. La gráfica muestra la distribución de la edad (X) de los 250 sujetos de una investigación. En el eje horizontal, se recogen los límites exactos de los intervalos de X y en el eje vertical la frecuencia absoluta acumulada (n_a).



A 22

B 100

C 50

Se ha seleccionado la opción C. La opción correcta es la C.

Pregunta 6

Tenemos 10 puntuaciones cuya media es 15, si sumamos un 5 a cada una de las puntuaciones, la media de las nuevas puntuaciones es:

A 15

B 20

C 75

Se ha seleccionado la opción B. La opción correcta es la B.

Pregunta 7

Con los datos de la Tabla 10, ¿cuál es el valor del primer cuartil de la distribución?

Tabla 10. Puntuaciones de 36 personas en un test de razonamiento verbal:

Puntuación	n_i
50	1
49	2
48	6
47	7
46	8
45	4
44	3
43	2
42	3

A 45,25

B 44,75

C 46,25

Se ha seleccionado la opción B. La opción correcta es la B.

Pregunta 8

Se tienen las siguientes puntuaciones en un test de matemáticas: 5, 4, 2, 3, 7, 4, 6, 5, 8, 4. ¿Cuál es el valor de la Desviación Media de la distribución?

A 1,40

B 3,96

C 2,08

Se ha seleccionado la opción A. La opción correcta es la A.

Pregunta 9

En una muestra de 1000 alumnos universitarios, 254 hombres y 169 mujeres afirman usar Hangouts, mientras que 356 hombres y 221 mujeres usan Skype, ¿cuál es el valor de χ^2 para medir la asociación entre el género y la herramienta de comunicación utilizada?

A 0,28

B 0,66

C 0,16

Se ha seleccionado la opción A. La opción correcta es la A.

Pregunta 10

Con los datos de la Tabla 7, la ordenada en el origen de la recta de regresión de Y sobre X es:

Tabla 7. Resultados en una escala de perseverancia (X) y en una tarea de discriminación visual (Y) de 5 trabajadores de una empresa.

Trabajador	X	Y
1	4	3
2	6	5
3	9	10
4	12	12
5	14	9

A 1,05

B 5,55

C 3,3

Se ha seleccionado la opción A. La opción correcta es la A.

Pregunta 11

Teniendo en cuenta los datos de la Tabla 9, la varianza de los errores cometidos al pronosticar las puntuaciones en Y utilizando la recta de regresión es:

Tabla 9. Para dos variables cuantitativas, X e Y, disponemos de los siguientes datos:

X	Y	
$\overline{X} = 100$	$\overline{Y} = 8$	$Y' = 0{,}1X - 2$
$S_X = 16$	$S_Y = 2$	

A 2,56

B 1,44

C 5,42

Se ha seleccionado la opción B. La opción correcta es la B.

Pregunta 12

Se está analizando la eficacia de tres fármacos experimentales para la remisión de una determinada enfermedad en 10 pacientes, habiendo cuatro pacientes medicados con el fármaco 1, tres pacientes con el fármaco 2 y tres con el fármaco 3. Los resultados muestran que la probabilidad de curación de los pacientes que han tomado el fármaco 1 es 0,4, la de los que han tomado el fármaco 2 es 0,5 y la de los que han tomado el fármaco 3 es 0,6. ¿Cuál es la probabilidad de curación?

A 0,33

B 0,49

C 0,40

Se ha seleccionado la opción B. La opción correcta es la B.

Pregunta 13

Con los datos de la Tabla 2, ¿cuál es el porcentaje de riesgo de tener ansiedad si no se padece una enfermedad crónica?

Tabla 2. Datos de 200 personas:

Sí = Tienen enfermedad crónica
No = No tienen enfermedad crónica

	Sí	**No**
Ansiedad	64	20
No ansiedad	36	80

A 20%

B 64%

80%

C

Se ha seleccionado la opción A. La opción correcta es la A.

Pregunta 14

Teniendo en cuenta la Situación 3, si definimos la variable X como número de personas que padecen Trastorno por Estrés Postraumático, ¿cuánto vale la media de X?

Situación 3. Se sabe que después de una catástrofe la probabilidad, entre los supervivientes, de padecer un Trastorno por Estrés Postraumático es 0,45. Ocurrida una catástrofe hay 2000 supervivientes.

A 900

B 1000

C 800

Se ha seleccionado la opción A. La opción correcta es la A.

Pregunta 15

Una variable aleatoria discreta X toma los valores 0, 1, 2 y 3. Si sabemos que P(X > 2) = 0,125 ¿cuál es la probabilidad de que X sea igual a 3?

A 0,125

B 0,25

C 0,875

Se ha seleccionado la opción A. La opción correcta es la A.

Pregunta 16

Una variable X se distribuye según la *F* de Snedecor con 6 grados de libertad en el numerador y 8 grados de libertad en el denominador. El valor de la variable X que deja por encima de sí el 10 % de las observaciones es:

A 0,335

B 2,668

C 2,983

No se ha seleccionado ninguna opción. La opción correcta es la B.

Pregunta 17

Una variable X sigue la distribución χ^2 con 17 grados de libertad. El percentil 99 es:

A 0,01

B 0,99

C 33,4087

C

No se ha seleccionado ninguna opción. La opción correcta es la C.

Pregunta 18

Se sabe que el porcentaje de investigadores que se van de España al acabar el doctorado es del 25%. Si se extrae una muestra aleatoria de 170 investigadores, ¿cuál es el valor del error típico de estimación de la proporción?

A 0,0332

B 0,1433

C 0,2653

No se ha seleccionado ninguna opción. La opción correcta es la A.

Pregunta 19

Para estimar el intervalo de confianza para la media de una variable X, a un nivel de confianza del 99%, hemos extraído una muestra aleatoria de 144 personas y les hemos medido la variable X obteniendo una media de 12 y una cuasidesviación típica de 4. Los límites del intervalo de confianza son:

A 8 y 14

B 9,42 y 14,58

C 11,14 y 12,86

No se ha seleccionado ninguna opción. La opción correcta es la C.

Pregunta 20

Los límites superior e inferior del intervalo confidencial de la media poblacional de una variable X son 5,588 y 4,412 respectivamente. Suponiendo que la muestra seleccionada es de 100 personas, en las que se ha obtenido una media igual a 5 y que se ha trabajado con un nivel de confianza del 95%, ¿cuál es la desviación típica de la variable X en la población?

A 4

B 9

C 3

No se ha seleccionado ninguna opción. La opción correcta es la C.

Observaciones del estudiante:

<Sin observaciones>

Observaciones del docente:

<Sin observaciones>

VOLVER

Secretaría General - Centros Tecnológicos de la UNED - Vicerrectorado de Estudiantes - Vicerrectorado de Personal Docente e Investigador - Vicerrectorado de Tecnología - Vicerrectorado de Innovación y Digitalización - Vicerrectorado de Calidad - IUED - Centro de Prevención y Resolución de Conflictos.
Desarrollado en el Centro de la UNED Barbastro.

Soporte: soportePDI@csi.uned.es 91 398 68 00 Manual para docentes

INTRODUCCIÓN AL ANÁLISIS DE DATOS

FEBRERO 2021 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO B DURACION: 90 minutos

Material: **Addenda (Formulario y Tablas) y calculadora no programable**

Calificación= (0,4 x Aciertos) - (0,25 x Errores)

No debe entregar los enunciados

Tabla 1. Puntuaciones de 36 personas en un test de razonamiento verbal:

Puntuación	n_i
50	1
49	2
48	6
47	7
46	8
45	4
44	3
43	2
42	3

Tabla 2. Datos de 75 niños según las actividades extraescolares que realizan:

	Deporte	**Música**	**Idiomas**
Niño	20	4	6
Niña	15	11	19

Supuesto 1. Un 70% de los niños diagnosticados con Trastorno del Espectro Autista (TEA) realiza conductas autoestimulantes como aletear con las manos.

1. Se tienen las siguientes puntuaciones en un test de matemáticas: 5, 4, 2, 3, 7, 4, 6, 5, 8, 4. ¿Cuál es el valor de la Desviación Media de la distribución? **A) 1,40**; B) 2,08; C) 3,96.
2. La curtosis y la asimetría son medidas de: A) tendencia central; B) variabilidad; **C) forma.**
3. Con los datos de la tabla 1, ¿cuál es el valor del primer cuartil de la distribución? **A) 44,75**; B) 46,25; C) 45,25.
4. ¿Cuánto vale la varianza de las puntuaciones diferenciales de la distribución de la tabla 1? A) 2,02; **B) 4,08**; C) 6,07.
5. Las puntuaciones de una variable x son: 2, 4, 6, 8. Si multiplicamos sus valores por la constante b=-3, ¿cuánto vale su desviación típica? A) 5,52; B) 4,47; **C) 6,71**.
6. En el contexto de las tablas de contingencia, ¿qué es una distribución condicionada? **A) es una distribución que especifica cuantas observaciones hay de cada valor de una variable al imponer la condición de que la otra asuma un valor determinado**; B) es una distribución de frecuencias unidimensional que nos informa del número de observaciones para cada valor de una de las variables, prescindiendo de la información de los valores de las demás variables; C) es la distribución de probabilidad de la intersección de dos variables (X e Y) que se da cuando las variables ocurren de forma simultánea.
7. Con los datos de la tabla 2, ¿cuánto vale el coeficiente *C* de Contingencia entre las variables sexo y actividad extraescolar realizada? A) 0,145; **B) 0,311**; C) 0,279.
8. Para valorar la relación entre la posición de los siguientes ciclistas en la vuelta a España en 2016 y 2019, que se presenta a continuación, ¿qué coeficiente utilizaría y cuál es su valor? Año 2019: Quintana (4), Valverde (2), López (5), Pojacar (3), Roglic (1). Año 2016: Quintana (1), Valverde (2), Pojacar (3), Roglic (4), López (6). A) El coeficiente V de Cramer cuyo valor

es 0,1; **B) El coeficiente de correlación de Spearman cuyo valor es 0,1;** C) El coeficiente de correlación de Spearman cuyo valor es 0,2.

9. La relación entre atracción al riesgo (Y) y la experiencia emprendedora (X_1) es r_{yx1}=0,23; la relación entre atracción al riesgo (Y) y la intención emprendedora (X_2) es r_{yx2}=0,77 y la relación entre la experiencia e intención emprendedora r_{x1x2}=0,54. Si se utilizaran las variables experiencia e intención emprendedora (X_1 y X_2) para predecir la atracción al riesgo (Y), ¿qué porcentaje de la variabilidad de la atracción al riesgo se puede explicar por la experiencia e intención emprendedora? **A) 64,15%**; B) 28,22%; C) 40,16%.

10. ¿Cuál es la probabilidad de obtener una suma de 5 al lanzar dos dados? A) 0,2500; B) 0,1666; **C) 0,1111**.

11. Con los datos de la tabla 2, ¿cuál es la probabilidad de que siendo niña realice una actividad extraescolar de tipo deportivo? A) 0,6666; **B) 0,3333**; C) 0,2000.

12. La definición de probabilidad que no asume la equiprobabilidad y que se basa en la estabilidad de las frecuencias relativas cuando el número de repeticiones de un suceso aleatorio tiende a infinito es la definición: **A) estadística**; B) clásica; C) axiomática.

13. Siguiendo el supuesto 1, si en un centro hay 8 niños con TEA, ¿cuál es la probabilidad de que 3 de ellos presenten conductas autoestimulantes? A) 0,0317; **B) 0,0467**; C) 0,0345.

14. La siguiente tabla presenta la función f(x) asignada a una variable aleatoria discreta *X*.

X	0	1	2	3	4
f(x)	25/100	40/100	0	35/100	10/100

La función f(x): **A) no es una función de probabilidad porque incumple una propiedad**; B) no es una función de probabilidad porque f(2)=0; C) es una función de probabilidad porque f(x)>0.

15. Las puntuaciones de un test de inteligencia en un grupo de 1000 personas se distribuyen normalmente con una media de 56 y una varianza de 144. ¿Cuánto vale el percentil 85 de la distribución? A) **68,48**; B) 51,03; C) 64,22.

16. Sea X una variable que se distribuye según *t* con 7 grados de libertad, ¿cuál es la probabilidad de obtener valores menores o iguales a 0,711? A) 0,95; **B) 0,75**; C) 0,80.

17. Se sabe que el porcentaje de investigadores que se van de España al acabar el doctorado es del 25%, se extrae una muestra aleatoria de 170 investigadores, ¿cuál es el valor del error típico de estimación de la proporción? **A) 0,0332**; B) 0,1433; C) 0,2653.

18. ¿Cómo se llama el método de estimación de parámetros que considera los parámetros como variables aleatorias con una cierta distribución "a priori" cuyas probabilidades se transforman mediante los datos en probabilidades "a posteriori"? A) Estimación Bootstrap; **B) Estimación Bayesiana**; C) Estimación por intervalos.

19. Se quiere estimar la media de la variable optimismo en un grupo de personas que practican meditación. Se extrae una muestra de 74 personas, de una población de 2500. Sabiendo que el optimismo se distribuye *N* (μ; 17), ¿cuál es el valor del error máximo que se comete con este tamaño muestral con un nivel de significación α = 0,05? A) 1,273; **B) 3,814**; C) 5,016.

20. Una muestra aleatoria de pacientes con hipertensión indica que el 15% no tiene una disminución de su tensión a valores normales tomando Losartán, ¿cuál es el tamaño muestral necesario para que el error máximo de estimación asociado a un intervalo de confianza de 95% sea 0,07? A) 78; B) 66; **C) 100.**

SOLUCIONES

1. Solución: A.

   La fórmula de la desviación media (véase página 106) es: DM=$\frac{\sum|X_i-\bar{X}|}{n}$

   La media del ejercicio es 4,8 por lo cual la desviación media es:

   $DM = \frac{[|5-4{,}8|+|4-4{,}8|+|2-4{,}8|+|3-4{,}8|+|7-4{,}8|+|4-4{,}8|+|6-4{,}8|+|5-4{,}8|+|8-4{,}8|+|4-4{,}8|]}{10} = \frac{[0{,}2+0{,}8+2{,}8+1{,}8+2{,}2+0{,}8+1{,}2+0{,}2+3{,}2+0{,}8]}{10} = \frac{14}{10} = 1{,}40$

2. Solución: C. (Página 38).

3. Solución: A.

Puntuación	n_i	n_a
50	1	36
49	2	35
48	6	33
47	7	27
46	8	20
45	4	12
44	3	8
43	2	5
42	3	3

$Q_1=P_{25}$. Sabemos que $\frac{n\cdot 25}{100} = \frac{36\cdot 25}{100} = 9$ por lo que el intervalo crítico es [44,5-45,5] con n_a=12, n_c=4, n_d=8, *k*=25 y *n*=36, aplicando la fórmula,

$$P_{25} = L_i + \left[\frac{\frac{n\cdot k}{100} - n_d}{n_c}\right]\cdot I = 44{,}5 + \frac{9-8}{4}\cdot 1 = 44{,}75$$

4. Solución: B.

   La varianza de las puntuaciones diferenciales es la misma que la varianza de las puntuaciones directas (p. 127). Por lo cual calculamos la varianza de las puntuaciones directas:

Puntuación	n_i	$X_i n_i$	$(X_i-\bar{X})^2$	$n_i(X_i-\bar{X})^2$
50	1	50	15,784	15,784
49	2	98	8,838	17,677
48	6	288	3,892	23,352
47	7	329	0,947	6,629
46	8	368	0,0007	0,005
45	4	180	1,055	4,218
44	3	132	4,109	12,326
43	2	86	9,163	18,325
42	3	126	16,217	48,650
sumatorio	36	1657		146,966

$$\bar{X} = \frac{X_i n_i}{n} = \frac{1657}{36} = 46{,}027$$

La media es 46,027 por lo cual la varianza es (aplicando la fórmula 3.6 del libro):

$$S_x^2 = \frac{\sum n_i(X_i - \bar{X})^2}{n} = \frac{146{,}966}{36} = 4{,}082$$

5. Solución C.

$\bar{x} = 5; s_x = 2{,}236;\ s_v = |-3|.2{,}236 = 6{,}708\ \approx 6{,}71$

6. Solución: A. (Página 153).
7. Solución: B.

   Primero tenemos que hallar el valor de χ^2:

	Deporte	Música	Idiomas	suma
Niño	20 (14)	4 (6)	6 (10)	30
Niña	15 (21)	11 (9)	19 (15)	45
	35	15	25	75

$$\chi^2 = \frac{(20-14)^2}{14} + \frac{(4-6)^2}{6} + \frac{(6-10)^2}{10} + \frac{(15-21)^2}{21} + \frac{(11-9)^2}{9} + \frac{(19-15)^2}{15}$$
$$= 2{,}571 + 0.666 + 1{,}6 + 1{,}714 + 0{,}444 + 1{,}067 = 8{,}061$$

   Sustituyendo en la fórmula de C de Contingencia: $\sqrt{\frac{8{,}061}{8{,}061+75}} = \sqrt{0{,}097}$ = 0,3114

8. Solución: B.

   Poniendo los datos en una tabla:

Ciclista	2019	2016	d	d^2
Quintana	4	1	3	9
Valverde	2	2	0	0
Pojacar	3	5	-2	4
Roglic	1	3	-2	4
López	5	6	-1	1
Suma				18

   De acuerdo a la fórmula, de la correlación de Spearman (página 171):

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(18)}{5(5^2-1)} = 1 - \frac{108}{120} = 0{,}1$$

9. Solución: A.

   De forma similar al ejercicio 5.20 de la página 246:

	Experiencia emprendedora	Intención emprendedora	Atracción al riesgo
Experiencia emprendedora (X_1)	1	0,54	0,23

Intención emprendedora (X_2)		1	0,77
Atracción al riesgo (Y)			1

$$R^2_{Y.X1.X2} = \frac{r^2_{YX1}+r^2_{YX2}-2r^2_{YX1}r^2_{YX2}r^2_{X1X2}}{1-r^2_{X1X2}} = \frac{0{,}23^2+0{,}77^2-2\cdot 0{,}23\cdot 0{,}77\cdot 0{,}54}{1-0{,}54^2} = \frac{0{,}6459-0{,}1913}{0{,}708} = 0.6415$$

El 64,15% de la variabilidad en la atracción al riesgo se puede explicar por la experiencia y la intención emprendedora.

10. Solución: C.
    Son 6x6 = 36 combinaciones posibles y 4 casos favorables (4+1, 1+4, 3+2 y 2+3) por lo tanto, la probabilidad es 4/36 = 0,1111

11. Solución: B.
    De un total de 45 niñas solo 15 practican deporte, por lo tanto, la probabilidad es de 15/45 = 0,3333

12. Solución: A. (Página 261).

13. Solución: C.

    Página 323 ejemplo 7.10. Como p>0,5 se busca la probabilidad de que 5 de ellos no presenten conductas autoestimulantes con un valor p=0,3. En la tabla I podemos ver en la intersección de n=8, x=5 y p=0,3, la probabilidad es 0,0467

14. Solución: A.

    no es una función de probabilidad porque la suma de todas las probabilidades es más de 1 (página 308) incumpliendo una propiedad fundamental de estas funciones.

15. Solución: A.

    Buscamos en el interior de la tabla IV el valor más cercano a 0,85 que es 0,8485. Podemos ver que a este valor le corresponde una puntuación típica de 1,03; a partir de esta puntuación y de los datos dados en el problema calculamos:

$$1{,}03 = \frac{P_{85}-\bar{x}}{s_x} = \frac{P_{85}-56}{12}$$

$$P_{85} = (1{,}04)(12)+56 = 68{,}48$$

16. Solución: B. Consultamos la tabla VI, en la primera columna (grados de libertad) localizamos el valor 7, los valores correspondientes en su fila son los valores de *t*, localizamos 0,711 y se ve que corresponde a 0,75 en la primera fila.

17. Solución: A.

    $n.\pi = 170 \cdot 0{,}25 = 42{,}5 > 5$, por tanto, podemos emplear la fórmula $\sigma = \sqrt{\pi(1-\pi)}$ $\sigma = \sqrt{0{,}25(0{,}75)} = 0{,}433$, por tanto, $\sigma_p = \frac{\sigma}{\sqrt{n}} = \frac{0{,}433}{\sqrt{170}}$ = 0,0332

18. Solución: B. Página 433

19. Solución: B.

Emáx = $\frac{z_{1-\alpha/2}\sigma}{\sqrt{n}} = \frac{1{,}96 \cdot 17}{\sqrt{74}} = \frac{33{,}32}{8{,}602} = 3{,}873$

Como la población es finita hay que corregir multiplicando por el factor de corrección $\sqrt{\frac{N-n}{N-1}}$ entonces Emáx = $3{,}873 . \sqrt{\frac{2500-74}{2499}}$ = 3,814

20. Solución: C.

Se sabe que p = 0,15 y que Emáx = 0,07, por tanto,

$$n = \frac{1{,}96^2 \cdot 0{,}15 \cdot 0{,}85}{0{,}07^2} = \frac{0{,}4898}{0{,}0049} = 99{,}96 \approx 100$$

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2021 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 90 minutos

Material: **Addenda (Formulario y Tablas) y calculadora no programable**

Calificación= (0,5 x Aciertos) - (0,25 x Errores)

No debe entregar los enunciados

Tabla 1. Intervalo de edad en años de un grupo de empleados:

Intervalo de edad en años	Número de empleados
20-29	58
30-39	36
40-49	21
50-59	10
60-69	5

Tabla 2. Datos de 200 personas:

Sí = Tienen enfermedad crónica
No = No tienen enfermedad crónica

	Sí	**No**
Ansiedad	64	20
No ansiedad	36	80

Supuesto 1. Un 90% de las personas que tienen obsesiones tienen también comportamientos compulsivos

1. La medición es: A) **el proceso por el que se asignan números a objetos o sucesos según determinadas reglas**; B) el conjunto de valores resultantes de medir una característica de interés sobre cada elemento de una población o muestra; C) la asignación puramente arbitraria de números o símbolos a cada una de las variables.
2. Según la distribución de frecuencias de la tabla 1, los límites exactos del cuarto intervalo (comenzando por arriba) son: **A) 49,5 y 59,5**; B) 50,5 y 59,5; C) 49,5 y 60,5.
3. Según la tabla 1, ¿cuál es la media de la distribución? A) 35,54; B) 33,24; **C) 34,85.**
4. De acuerdo con la tabla 1, si m es la edad mediana de los empleados, ¿cuál de las siguientes desigualdades satisface m? A) $20 \leq m \leq 29$; B) $25 \leq m \leq 34$; **C) $30 \leq m \leq 39$**
5. De acuerdo con la tabla 1, si contratan 15 nuevos empleados en la franja de edad de 50-59 años, ¿qué porcentaje de empleados tendría la empresa en dicha franja de edad? A) 25,33%; **B) 17,24%;** C) 15,38%
6. Las puntuaciones de una variable X son: 2, 4, 6 y 8, ¿cuáles son sus puntuaciones tipificadas? A) -2, -1, 0 y 1; B) -1, 0, 1 y 2; **C) -1,34, -0,45, 0,45 y 1,34.**
7. En una muestra de 1000 alumnos universitarios, 254 hombres y 169 mujeres afirman usar Hangouts, mientras que 356 hombres y 221 mujeres usan Skype, ¿cuál es el valor de χ^2 para medir la asociación entre el género y la herramienta de comunicación utilizada? **A) 0,28**; B) 0,66; C) 0,16.
8. El peso de 5 niños (x) y sus puntuaciones en un test de psicomotricidad (y) se presentan a continuación en el formato (x, y): (32, 4) (48, 2), (25, 5), (28, 3), (26, 5), ¿cuál es la puntuación pronosticada para un niño que pesa 27 kg? A) 5,42; B) 6,13; **C) 4,36.**
9. Para los datos del ejercicio 8 ¿cuál de los siguientes es el valor que más se aproxima al coeficiente de correlación de Pearson? **A) -0,833**; B) -0,742; C) 0,264.
10. En el contexto de probabilidad, ¿qué es un suceso simple? **A) Todo resultado de un experimento aleatorio**; B) Un resultado que puede contener algún elemento del espacio muestral; C) Un resultado que ocurre siempre y contiene todos los elementos de un espacio muestral.

11. En una casa hay tres llaveros (A, B y C), el primero tiene 5 llaves, el segundo 7 y el tercero 8. Solo hay una llave para abrir la casa. Se selecciona al azar un llavero y de él una llave para abrir la casa, ¿cuál es la probabilidad de acertar con la llave? A) 0,2915; **B) 0,1559**; C) 0,1955.
12. Con los datos de la tabla 2, ¿cuál es el porcentaje de riesgo de tener ansiedad si no se padece una enfermedad crónica? **A) 20%**; B) 64%; C) 80%.
13. Una variable se distribuye según la t de Student con 20 grados de libertad, ¿cuál es la media de dicha variable? A) 20; **B) 0**; C) Hacen falta más datos para calcularla.
14. Sea X una variable aleatoria discreta con valores 0, 1 y 2. Siendo sus probabilidades respectivamente 0,4, 0,3, 0,3, ¿cuál será la desviación típica de la variable X? A) 0,57; **B) 0,83**; C) 1,05.
15. Una variable aleatoria sigue una distribución χ^2 con varianza igual a 4. Los grados de libertad de esta variable son: A) 8; **B) 2**; C) 4.
16. ¿Cuánto vale el error típico de una proporción sabiendo que las muestras extraídas (muestreo aleatorio simple) tienen un tamaño de 40 y el valor esperado es 0,5? **A) 0,079**; B) 0,145; C) 0,224.
17. Una variable X se distribuye según la distribución F con 7 y 20 grados de libertad del numerador y denominador, respectivamente. ¿Cuál es el percentil 10 de esa distribución? A) 2,595; **B) 0,385**; C) 2,040.
18. ¿Cuál de los siguientes métodos de muestreo es no probabilístico? A) Estratificado; B) Aleatorio simple; **C) Opinático.**
19. En un estudio se quiere averiguar el porcentaje de personas que están a favor del matrimonio homosexual. Para ello se ha extraído una muestra de 200 personas de las que 140 están a favor. Con un nivel de confianza de 0,95, los límites del intervalo de confianza para la proporción son: **A) 0,637 y 0,764**; B) 0,765 y 0,835; C) 0,595 y 0,605.
20. Se quiere estimar la media de la variable resiliencia en un grupo de pacientes con cáncer. Se extrae una muestra de 74 pacientes, de una población de 2500. Sabiendo que la resiliencia se distribuye *N* (μ; 17), ¿cuál es el valor del error máximo que se comete con este tamaño muestral con un nivel de significación α = 0,01? A) 4,098; B) 3,814; **C) 5,016**.

SOLUCIONES

1. Solución: A. (Página 11).
2. Solución: A. (Página 32).
3. Solución: C. Se calcula el punto medio de cada intervalo como la semisuma de los límites exactos y el producto de cada punto medio por su correspondiente frecuencia absoluta:

X_i	n_i	$n_i X_i$
25	58	1450
35	36	1260
45	21	945
55	10	550
65	5	325
	Media	4530/130 = 34,846

4. Solución: C.

X	n_i	n_a
20-29	58	58
30-39	36	94
40-49	21	115
50-59	10	125
60-69	5	130
	Intervalo Mediana	130/2= 65

5. Solución: B.

   Sumando 15 personas al intervalo 50-59 tendríamos 25 personas, por lo que el total sería 145. 25/145 = 0,172 que corresponde a un porcentaje de 17,241%

6. Solución: C.

   $\bar{X} = \frac{2+4+6+8}{4} = \frac{20}{4} = 5$; $s_X^2 = \frac{\sum X^2}{n} - \bar{X}^2 = \frac{2^2+4^2+6^2+8^2+}{4} - 5^2 = 5$; $S_X = \sqrt{5} = 2{,}24$

   $Z = \frac{X_i - \bar{X}}{S_x}$; por tanto $Z_1 = \frac{2-5}{2{,}24} = -1{,}34$; $Z_2 = \frac{4-5}{2{,}24} = -0{,}45$;

   $Z_3 = \frac{6-5}{2{,}24} = 0{,}45$; $Z_4 = \frac{8-5}{2{,}24} = 1{,}34$

7. Solución: A.

	Hangouts	Skype	Total
Hombres	254 (258,03)	356 (351,97)	610
Mujeres	169 (164,97)	221 (225,03)	390
Total	423	577	1000

Calculamos la frecuencia teórica de cada casilla:

$f_{t(1,1)} = \frac{610*423}{1000} = 258{,}03;\ f_{t(1,3)} = \frac{610*577}{1000} = 351{,}97$

$f_{t(2,1)} = \frac{390*423}{1000} = 164{,}97;\ f_{t(2,2)} = \frac{390*577}{1000} = 225{,}03;$

$\chi^2 = \frac{(254-258{,}03)^2}{258{,}03} + \frac{(356-351{,}97)^2}{351{,}97} + \frac{(169-164{,}97)^2}{164{,}97} + \frac{(221-225{,}03)^2}{225{,}03} = 0{,}0629 + 0{,}0461 + 0{,}0984 + 0{,}0721 = \mathbf{0{,}279}$ → **0,28**

8. Solución: C.

Niño	X	Y	XY	X^2	Y^2
1	32	4	128	1024	16
2	48	2	96	2304	4
3	25	5	125	625	25
4	28	3	84	784	9
5	26	5	130	676	25
∑	159	19	563	5413	79
medias	31,8	3,8			

$$B = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{5*563 - 159*19}{5*5413 - (159)^2} = \frac{-206}{1784} = -0{,}115$$

$$A = \bar{Y} - B\bar{X} = 3{,}8 - (-0{,}115)*31{,}8 = 7{,}46$$

$$Y' = -0{,}115X + 7{,}46 = (-0{,}115*27) + 7{,}46 = \mathbf{4{,}36}$$

9. Solución: A.

Calculamos S_X y S_Y para los datos del problema:

$$S_X = \sqrt{\frac{\sum X^2}{n} - \bar{X}^2} = \sqrt{\frac{5413}{5} - 31.8^2} = 8{,}447;$$

$$S_y = \sqrt{\frac{\sum Y^2}{n} - \bar{Y}^2} = \sqrt{\frac{79}{5} - 3{,}8^2} = \sqrt{1{,}36} = 1{,}166;$$

$$r_{xy} = b\frac{S_X}{S_Y} = -0{,}115\ x\ \frac{8{,}447}{1{,}116} = -0{,}833$$

10. Solución: A. (Página 254).

11. Solución: B.

Probabilidad de seleccionar cualquiera de los llaveros 1/3.

La probabilidad de que la llave abra si se selecciona el llavero A es 1/5 y de que no abra es 4/5;

La probabilidad de que abra con el llavero B es 1/7 y de que no abra 6/7

La probabilidad de que abra con el llavero C es 1/8 y de que no abra es 7/8.

Por tanto, la probabilidad de acertar es (1/3)(1/5)+ (1/3)(1/7)+ (1/3)(1/8) = 0,1559

12. Solución: A.

$P(A \cap No) = \frac{20}{200} = 0{,}1$; $P(No) = \frac{100}{200} = 0{,}5$

Por tanto $P(A \cap No) = \frac{P(A \cap No)}{P(No)} = \frac{0{,}1}{0{,}5} = 0{,}2$ → 20%

13. Solución: B. (Página 357).
14. Solución: B.

X	f(x)	xf(x)	x-μ	$(x-\mu)^2$	$(x-\mu)^2 f(x)$
0	0,4	0	-0,9	0,81	0,324
1	0,3	0,3	0,1	0,01	0,003
2	0,3	0,6	1,1	1,21	0,363

μ= 0+0,3+0,6=0,9

$$\sigma^2 = \sum (x-\mu)^2 . f(x) = 0{,}69;\ \sigma = \sqrt{0{,}69} = 0{,}83$$

15. Solución: B.

$\sigma^2 = 2n = 4 \rightarrow n = 2$

16. Solución: A. (Página 423).

Como $\pi = E(p) = 0{,}5$

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0{,}5(1-0{,}5)}{40}} = 0{,}079$$

17. Solución: B. (Página 361).

$${}_{0,10}F_{7,20} = \frac{1}{{}_{0,90}F_{20,7}} = \frac{1}{2{,}595} = 0{,}385$$

Tabla VII para p= 0,90, con n_1=20 y n_2 =7

18. Solución: C. (Página 385).

19. Solución: A.

$$n.c. = 0{,}95 \rightarrow z_{1-\alpha/2} = z_{0,975} = 1{,}96\ (Tabla\ IV)$$

$$L_i = 0{,}70 - 1{,}96\sqrt{\frac{0{,}70 \times 0{,}30}{200}} = 0{,}70 - 0{,}0635 = 0{,}6365 \rightarrow 0{,}637$$

$$L_s = 0{,}70 + 1{,}96\sqrt{\frac{0{,}70 \times 0{,}30}{200}} = 0{,}70 + 0{,}0635 = 0{,}7635 \rightarrow 0{,}764$$

20. Solución: C.

$E_{máx} = \frac{z_{1-\alpha/2}\sigma}{\sqrt{n}} = \frac{2{,}58\ x17}{\sqrt{74}} = \frac{43{,}86}{8{,}602} = 5{,}098$

Como la población es finita hay que corregir multiplicando por el factor de corrección $\sqrt{\frac{N-n}{N-1}}$

entonces $E_{máx.} = 5{,}098.\sqrt{\frac{2500-74}{2499}} = 5{,}016$

Código asignatura	Nombre asignatura
62011037	**INTRODUCCIÓN AL ANÁLISIS DE DATOS**
Fecha alta y origen	Convocatoria
20/02/2020	**FEBRERO 2020 – TIPO B**
Curso virtual	

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2020 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO B DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Tabla 1. Distribución de frecuencias en una prueba de memoria

X_i	n_i
7	3
8	4
9	6
10	15
11	12
12	10

Tabla 2. Datos de 20 alumnos donde:

Sí = Utilizan los cursos virtuales
No = No utilizan los cursos virtuales

	Sí	**No**
Aprobados	7	3
Suspensos	3	7

Supuesto 1. Un test de laboratorio es efectivo al 95% detectando en una muestra de sangre cierta enfermedad cuando esta se padece realmente. Sin embargo, el test también arroja un resultado "falso positivo" en el 1% de personas sanas sobre las que se realiza el test.

1. La variable estado civil tiene un nivel de medida: A) ordinal; **B) nominal**; C) de intervalo.
2. En una distribución de frecuencias relativas acumuladas, el valor para el último intervalo (el correspondiente al valor máximo) es: **A) 1**; B) 100; C) dependiente del tamaño de la muestra.
3. Con los datos de la Tabla 1. ¿Qué porcentaje de sujetos superan la puntuación 10,5?
   A) 74%; **B) 44%**; C) 56%.
4. ¿Cuál de las siguientes afirmaciones describe la distribución de frecuencias de la Tabla 1? A) es una distribución simétrica; **B) las mayores frecuencias corresponden a valores altos de X**; C) las menores frecuencias corresponden a valores altos de X.
5. ¿Cuál de las siguientes afirmaciones es FALSA? Los percentiles permiten comparar directamente entre sí puntuaciones obtenidas ...: A) …por sujetos diferentes dentro del mismo grupo y para la misma variable; B) …por el mismo sujeto en el mismo grupo y con diferentes variables; **C) …por el mismo grupo en diferentes variables**.
6. Dadas las siguientes puntuaciones: 5, 6, 3, 7, 4. ¿Cuánto vale su varianza? **A) 2**; B) 1,41; C) 6.
7. Si se quiere representar gráficamente la proporción de hombres y mujeres que votan a cada uno de los grandes partidos políticos que se presentan en tu ciudad. ¿Qué tipo de gráfica se utilizaría? A) El diagrama de sectores; B) El diagrama de tallo y hojas; **C) El diagrama de barras conjunto**.
8. Para los datos de la Tabla 2, el valor de χ^2 es: **A) 3,2**; B) 1,6; C) 0,37.
9. Siguiendo con los datos de la Tabla 2, el valor máximo del coeficiente de contingencia será: **A) 0,707**; B) 0,077; C) 0,770.
10. Dos psicólogos ordenan a cinco pacientes en función de su aptitud para un puesto de trabajo. Sabiendo que la suma de las diferencias de los dos órdenes al cuadrado vale 4, el coeficiente de correlación entre los dos órdenes toma el valor: **A) 0,8**; B) 0,2; C) 0,4.
11. ¿Cuál de los siguientes índices es el más adecuado para medir la relación entre dos variables cualitativas? **A)** χ^2; B) Spearman; C) Biserial puntual.

KUBBCIN

12. Respecto al modelo de regresión, si la varianza de los errores es 0,3 y la varianza de las puntuaciones observadas es 2,64, entonces la desviación típica de las puntuaciones pronosticadas será: A) 2,94; **B) 1,53**; C) 2,34.

13. ¿Cuál de las siguientes afirmaciones es VERDADERA? **A) Una correlación de -0,78 entre dos variables X e Y tiene la misma intensidad que otra correlación de +0,78 entre otras variables U y V**; B) Una correlación de 0,60 entre dos variables X e Y indica el doble de correlación que otra correlación de 0,30 entre otras dos variables U y V; C) Encontrar una relación entre dos variables cuantitativas significa que existe una relación de causa-efecto.

14. De la expresión $P(A \cap B) = P(A)P(B/A) = P(A)P(B)$ se deduce que: A) el suceso B está condicionado al suceso A; **B) los sucesos A y B son independientes**; C) el suceso A está condicionado al suceso B.

15. Se lanzan simultáneamente un dado y una moneda. ¿Cuál es la probabilidad de obtener un número par en el dado y una cara en la moneda? A) 0,5; **B) 0,25**; C) 0,75.

16. En el supuesto 1, la especificidad de la prueba de laboratorio es: **A) 0,99**; B) 0,95; C) 0,01.

17. Sea X una variable aleatoria discreta, cuya función de probabilidad asigna 1/4 a los valores 1, 2, 3 y 4. La función $F(x) = P(X \leq x)$ será: A) $P(x = 1) = 1/4$, $P(x = 2) = 3/4$, $P(x = 3) = 2/4$, $P(x = 4) = 1$; B) $P(x = 1) = P(x = 2) = P(x = 3) = P(x = 4) = 1/4$; **C) $P(x \leq 1) = 1/4$, $P(x \leq 2) = 2/4$, $P(x \leq 3) = 3/4$, $P(x \leq 4) = 1$**.

18. Sea X una variable aleatoria discreta con valores 0, 1, 2, 3. Siendo sus probabilidades respectivamente 0,2, 0,3, 0,3, 0,2. ¿Cuál será la varianza de la variable X? A) 1,50; B) 0,45; **C) 1,05**.

19. Sabiendo que X se distribuye normalmente, que la media de X es 60 y que la puntuación directa 40,8 es superada por el 89,97% de las observaciones. ¿Cuál es el valor aproximado de la desviación típica? **A) 15**; B) 1,28; C) 17,87.

20. En una distribución normal con varianza igual a 4. ¿Cuál es la proporción de observaciones entre la puntuación diferencial x = -1,5 y la media aritmética? A) 0,2266; **B) 0,2734**; C) 0,3520.

21. En una distribución F con 8 y 8 grados de libertad. ¿Cuál es el valor aproximado del percentil 1? **A) 0,166**; B) 6,029; C) 0,222.

22. Sabiendo que la media poblacional de una variable X vale 160 y su varianza poblacional vale 9. La distribución muestral de la media para un tamaño de muestra n = 81, tendrá: A) Una desviación típica de 1; **B) Una desviación típica de 0,33**; C) Los mismos valores de la población para la media y la desviación típica de la distribución muestral.

23. Una población se compone de los posibles resultados que se pueden obtener cada vez que se tira un dado no trucado. ¿Cuál es la media poblacional? **A) 3,5**; B) 4; C) 3.

24. Se ha extraído una muestra de 144 personas de una población. Se conoce la varianza poblacional que es igual a 9 y también que el límite superior en la estimación por intervalo de la media es 6,645, con un nivel de confianza del 99%. ¿Cuánto vale la media de la muestra seleccionada? A) 6,15; B) 6,06; **C) 6**.

25. En la construcción del intervalo de confianza para estimar, con una determinada probabilidad, entre qué valores se encontrará la media poblacional a partir de la información proporcionada por una muestra representativa, ¿cuál de las siguientes afirmaciones es verdadera? A) Al aumentar el nivel de confianza disminuye la amplitud del intervalo; **B) Al aumentar el tamaño de la muestra disminuye la amplitud del intervalo**; C) El tamaño de la muestra no afecta a la amplitud del intervalo de confianza.

EXAMEN

SOLUCIONES

1. **B**. Página 13.

2. **A**. Página 26.

3. **B**. Se ha de obtener el percentil correspondiente a la puntuación 10,5.

X_i	X	n_i	n_a
7	6,5 – 7,5	3	3
8	7,5 – 8,5	4	7
9	8,5 – 9,5	6	13
10	9,5 – 10,5	15	28
11	10,5 – 11,5	12	40
12	11,5 – 12,5	10	50

$$k = \left[\frac{\frac{(P_k - L_i)\cdot n_c}{I} + n_d}{n}\right]\cdot 100 = \left[\frac{\frac{(10{,}5 - 10{,}5)\cdot 12}{1} + 28}{50}\right]\cdot 100 = 56\%$$

56% corresponde con el porcentaje de observaciones por debajo de 10,5. Por tanto,

100% - 56% = 44% de observaciones quedarán por encima de la puntuación 10,5

4. **B**. Página 38.

5. **C**. Página 76.

6. **A**. Primero se calcula la media, para así poder calcular la varianza:

$$\bar{X} = \frac{\sum x_i}{n} = 5;\ S_x^2 = \frac{\sum X_i^2}{n} - \bar{X}^2 = \frac{135}{5} - 25 = 2$$

7. **C**. Página 155.

8. **A**. Se calculan las tablas de frecuencias observadas y esperadas y a continuación se halla χ^2:

Frecuencias observadas

	SI	NO	
Aprobado	7	3	10
Suspenso	3	7	10
	10	10	20

Frecuencias esperadas

	SI	NO
Aprobado	5	5
Suspenso	5	5

$$\chi^2 = \sum\sum \frac{(n_e - n_t)^2}{n_t} = \frac{4}{5} + \frac{4}{5} + \frac{4}{5} + \frac{4}{5} = \frac{16}{5} = 3{,}2$$

9. **A**. El valor máximo del coeficiente de contingencia depende del número de filas y columnas. Siendo k = número de filas = número de columnas. Se halla mediante la siguiente fórmula:

KIBBUTZ

$$Cmax = \sqrt{\frac{k-1}{k}} = \sqrt{\frac{2-1}{2}} = 0{,}707$$

10. **A**. Tenemos dos variables ordinales, luego hemos de aplicar el coeficiente de correlación de Spearman. Sabemos que la suma de los cuadrados de las diferencias entre los rangos es igual a 4, y aplicando la fórmula de Spearman.

$$r_s = 1 - \frac{6 \cdot \sum d^2}{n(n^2-1)} = 1 - \frac{6 \cdot 4}{5(5^2-1)} = 1 - \frac{24}{120} = 1 - 0{,}20 = 0{,}80$$

11. **A**. Página 217.

12. **B**. Partiendo de que la varianza de las puntuaciones en Y es igual a la varianza de los pronósticos más la varianza de los errores se obtiene que:

$$S_Y^2 = S_{Y'}^2 + S_{Y \cdot X}^2; 2{,}64 = S_{Y'}^2 + 0{,}3;\ S_{Y'} = \sqrt{2{,}64 - 0{,}3} = 1{,}53$$

13. **A**. Página 201.

14. **B**. Página 269.

15. **B**. Se definen los eventos Pr = resultado par en el dado y Cr = cara al tirar la moneda. Sabiendo que:

$$P(Pr) = \frac{3}{6} = 0{,}5; y\ P(Cr) = \frac{1}{2} = 0{,}5$$

Se obtiene que:

$$P(Pr \cap Cr) = P(Pr) \cdot P(Cr) = 0{,}5 \cdot 0{,}5 = 0{,}25$$

16. **A**. La especificidad hace referencia a la probabilidad de la prueba parar detectar a los verdaderos negativos. Esto es 1-“falsos positivos”; es decir 1- 0,01= 0,99.

17. **C**. A partir de los valores de la función de probabilidad que figura en la siguiente tabla:

x	1	2	3	4
f(x)	1/4	1/4	1/4	1/4

Se obtiene que:

$$F(1) = P(X \leq 1) = \frac{1}{4}$$

$$F(2) = P(X \leq 2) = P(X = 1) + P(X = 2) = \frac{1}{4} + \frac{1}{4} = \frac{2}{4}$$

KIBBUTZ

$$F(3) = P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$
$$F(4) = P(X \leq 4) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1$$

18. **C**. A partir de los valores de la función de probabilidad que figura en la siguiente tabla:

x	0	1	2	3
f(x)	0,2	0,3	0,3	0,2

En primer lugar se calcula la media:

$$\mu_x = \sum x \cdot f(x) = 0 \cdot 0{,}2 + 1 \cdot 0{,}3 + 2 \cdot 0{,}3 + 3 \cdot 0{,}2 = 0 + 0{,}3 + 0{,}6 + 0{,}6 = 1{,}5$$

A continuación se calcula la varianza: $\sigma^2 = V(x) = \sum (x - \mu)^2 \cdot f(x)$

x	***f*(x)**	**(x-μ)**	**(x-μ)^2**	**(x-μ)^2 . *f*(x)**
0	0,2	-1,5	2,25	0,45
1	0,3	-0,5	0,25	0,075
2	0,3	0,5	0,25	0,075
3	0,2	1,5	2,25	0,45
			∑	1,05

19. **A**. En primer lugar se busca la probabilidad superada por el 89,97% de las observaciones (que deja por debajo el 10,03% de la distribución), 1-0,8997 = 0,1003. A continuación se halla en la tabla de la curva normal su puntuación típica z = -1,28. A partir de ahí se aplica la formula correspondiente:

$$z = \frac{X - \bar{X}}{S_x}\ ;\ -1{,}28 = \frac{40{,}8 - 60}{S_x};\ S_x = \frac{40{,}8 - 60}{-1{,}28} = 15$$

20. **B**. Por debajo de la media se encuentra una proporción de observaciones igual a 0,5000. Y por debajo de una puntación diferencial igual a -1,5:

$$z = \frac{X - \bar{X}}{S_x} = \frac{x}{S_x} = \frac{-1{,}5}{\sqrt{4}} = -0{,}75;$$

Tras calcular la puntuación típica se busca en la tabla de la curva normal su proporción. z = -0,75 (0,2266). Por último se calcula la proporción de observaciones entre la media y x = -1,5.

Esto es: 0,5000-0,2266 = 0,2734.

KIBBUTZ

21. **A**. Se hace uso de la propiedad recíproca de la distribución F. De este modo, dado que los grados de libertad en el numerador y el denominador son los mismos, 8 y 8, bastará con calcular 1/ $F_{8,8}$ para un nivel de confianza 0,99. Esto es,

$$_{0,01}F_{8,8} = 1/_{(1-0,01)}F_{8,8} = 1/_{0,99}F_{8,8} = 1/6,029 = 0,166 \text{ (ver tabla VII)}$$

22. **B**. En este caso, la media de la distribución muestral de la media es igual a la media poblacional, es decir 160. El valor de la desviación típica de la distribución muestral, en cambio, será igual a:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\ ; \sigma_{\bar{X}} = \frac{3}{\sqrt{81}}\ ; \sigma_{\bar{X}} = \frac{3}{9} = 0{,}33$$

23. **A**. Se calcula el valor de la media poblacional:

$$\mu = \frac{\sum X_i}{N}\ ; \mu = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3{,}5$$

24. **C**. Para un nivel de confianza del 99% se utilizará z = 2,58. Por tanto, E_{max} será:

$$E_{max} = Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = 2{,}58\frac{3}{12} = 0{,}645$$

Y la media de la muestra seleccionada será:

$$L_{sup} = \bar{X} + E_{max};\ \bar{X} = L_{sup} - E_{max} = 6{,}645 - 0{,}645 = 6$$

25. **B**. Página 459.

KIBBUTZ

Código asignatura	Nombre asignatura
62011037	**INTRODUCCIÓN AL ANÁLISIS DE DATOS**
Fecha alta y origen	Convocatoria
20/02/2020	**FEBRERO 2020 – TIPO A**
Curso virtual	

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2020 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Tabla 1. Los datos de la tabla representan los puntos medios y las frecuencias absolutas en una prueba de reconocimiento facial.

X_i	n_i
6	4
9	11
12	17
15	7
18	1

Tabla 2. Puntuaciones de 6 sujetos en dos pruebas X e Y.

Sujeto	X	Y
1	2	8
2	3	9
3	5	7
4	7	12
5	9	12
6	10	12

Supuesto 1. Un problema se puede hacer de dos formas (A y B). El 80% de los alumnos utiliza el método A y de ellos resuelven correctamente el problema el 60%. De aquellos que siguen el método B lo resuelven correctamente la mitad.

1. ¿Cuál de los siguientes ejemplos corresponde a una variable cuantitativa continua? A) Número de votos para cada canción en el festival de Eurovisión; **B) Distancia entre la casa y la escuela para los niños de un Municipio**; C) Nivel de estudios de los trabajadores de una multinacional.
2. El polígono de frecuencias sólo se utiliza en el caso de representar gráficamente una variable: A) nominal; B) cualitativa; **C) cuantitativa continua**.
3. Una variable dicotómica tiene un nivel de medida: A) continuo; **B) nominal**; C) de intervalo.
4. De acuerdo con los datos de la Tabla 1, el valor de la puntuación que está dos desviaciones típicas por debajo de la media es: **A) 5,59**; B) 7,01; C) 8,42.
5. ¿Cuál es el percentil correspondiente a la puntuación 13,5 según los datos de la Tabla 1? A) 75; **B) 80**; C) 70.
6. ¿Cuál es el índice de tendencia central más apropiado para una variable nominal? **A) Moda**; B) Mediana; C) Media aritmética.
7. Dadas las siguientes puntuaciones: 5, -7, 0, 3, 12. ¿Cuál es el valor de la mediana? A) 5; **B) 3**; C) 0.
8. Dadas las siguientes puntuaciones: 3, 1, 2, 0, 4. ¿Cuál es el valor aproximado de la desviación típica? A) 2; **B) 1,41**; C) 6.
9. ¿Cuál de los siguientes índices NO es adecuado para calcular la relación entre dos variables cualitativas? A) χ^2; B) Coeficiente de contingencia; **C) Correlación biserial-puntual**.

KIBBUTZ

10. ¿Cuál es el índice de correlación que debe calcularse cuando se dispone de variables medidas con escala ordinal? A) El coeficiente φ; B) El coeficiente de correlación biserial-puntual; **C) El coeficiente de correlación de Spearman**.
11. Dos variables X e Y correlacionan entre sí (r_{xy}), y sus varianzas son $S_x^2 = 36$ y $S_y^2 = 25$. El valor de la covarianza estará comprendido entre: A) menos infinito y 0; B) 0 y más infinito; **C) -30 y +30**.
12. Considerando los datos de la Tabla 2. ¿Cuál es el valor aproximado de la pendiente de la ecuación de regresión para el pronóstico de Y a partir de X? **A) 0,58**; B) 0,47; C) 0,80.
13. Para los datos de la Tabla 2. ¿Qué porcentaje de la variabilidad de Y es explicada por la variabilidad de X? A) 54,6%; B) 58,6%; **C) 66,6%**.
14. En el experimento aleatorio consistente en lanzar dos monedas (una después de otra) y observar el resultado, se ha definido la variable aleatoria X = "número de caras obtenidas". Suponiendo monedas no trucadas, la probabilidad de f(X=0) vale: **A) 1/4**; B) 1/2; C) la variable X no puede adoptar el valor 0.
15. Siguiendo el Supuesto 1. ¿Cuál es la probabilidad de que un alumno resuelva correctamente el problema? A) 0,48; B) 0,10; **C) 0,58**.
16. En un espacio muestral hay dos sucesos A y B, tales que las siguientes afirmaciones son verdaderas: $P(\bar{A}) = 2/3$; $P(B) = 1/2$; $P(A \cap B) = 1/5$. ¿Cuál es la probabilidad de $(A \cup B)$? A) 12/25; **B) 19/30**; C) 8/13.
17. Si la probabilidad de nacer varón es 0,52. La probabilidad de que una pareja consiga 2 varones tras 5 nacimientos será: **A) 0,2987**; B) 0,2492; C) 0,0234.
18. Sea X una variable aleatoria discreta con valores 0, 2, 4, 6. Siendo todos los valores equiprobables, ¿cuál es su media? A) 0,25; B) 1,50; **C) 3**.
19. En una distribución normal con media 40 y desviación típica 10. ¿Cuánto vale el percentil 33 en puntuaciones directas? **A) 35,6**; B) -0,44; C) 44,4.
20. En la distribución de probabilidad de una variable aleatoria con distribución *t* de Student con 24 g.l. ¿Qué valor es superado por el 70% de las puntuaciones? A) 0,531; **B) -0,531**; C) 0,684.
21. En una variable que se distribuye normalmente, se sabe que la puntuación típica z = -2,65 supera a 5 sujetos. ¿Cuál es el número total de observaciones? A) 400; **B) 1250**; C) 125.
22. El método de muestreo de bola de nieve corresponde con un tipo de muestreo: A) Aleatorio estratificado; B) Probabilístico; **C) No probabilístico**.
23. En una población con desviación típica igual a 20 se extraen muestras de tamaño 100. ¿Cuánto vale la varianza de la distribución muestral? A) 2; **B) 4**; C) 0,02.
24. Las puntuaciones de un test de razonamiento numérico se distribuyen normalmente con una desviación típica poblacional de 9 puntos. Aplicado a una muestra de 196 personas se obtiene una media de 80 puntos. Basándose en los intervalos de confianza, ¿puede ser la media del test de 82 puntos para la población general? A) Sí, con un nivel de confianza del 95%; B) Sí, con un nivel de confianza del 90%; **C) No, con ninguno de los dos niveles de confianza anteriores**.
25. Los datos de una muestra seleccionada al azar de niños que utilizan videojuegos indican que el 75% ha incrementado sus habilidades visoespaciales. ¿Cuál será el tamaño muestral necesario para que el error máximo de estimación asociado a un intervalo de confianza del 99% sea 0,05? Se asume población infinita. **A) 500**; B) 288; C) 408.

EXPERTIA

SOLUCIONES

1. **B**. Página 18.

2. **C**. Página 34.

3. **B**. Página 13.

4. **A**. Primero se calculan la desviación típica y la media a partir de los datos de la siguiente tabla:

X_i	n_i	n_iX_i	X_i^2	$n_iX_i^2$
6	4	24	36	144
9	11	99	81	891
12	17	204	144	2448
15	7	105	225	1575
18	1	18	324	324
∑	**40**	**450**	∑	**5382**

$$S_x = \sqrt{\frac{\sum n_i \cdot X_i^2}{n} - \bar{X}^2} = \sqrt{\frac{5382}{40} - 11{,}25^2} = 2{,}83;\ \bar{X} = \frac{\sum n_i \cdot X_i}{n} = \frac{450}{40} = 11{,}25$$

A continuación se calcula la puntuación que se aparta dos desviaciones típicas por debajo de la media:

11,25 - 2 x 2,83 = 5,59

5. **B**. En la siguiente tabla se observan los puntos medios, y sabiendo que la distancia entre dos puntos medios adyacentes es igual a la amplitud de los intervalos, se pueden calcular los límites de cada intervalo:

X	ni	na
5 – 7	4	4
8 – 10	11	15
11 – 13	17	32
14 – 16	7	39
17 – 19	1	40
∑	40	-

13,5 es el límite superior del intervalo 11 – 13 (límite inferior del intervalo 14 – 16). Sabiendo esto, puede observarse que la columna de frecuencias acumuladas deja por debajo de sí a 32 observaciones, que en porcentajes corresponden a un 80%. Por tanto, 13,5 ocupa el percentil 80. También podría haberse aplicado la fórmula de los percentiles:

$$k = \left[\frac{\frac{(P_k - L_i) \cdot n_c}{I} + n_d}{n}\right] \cdot 100 = \left[\frac{\frac{(13{,}5 - 13{,}5) \cdot 7}{3} + 32}{40}\right] \cdot 100 = 80$$

6. **A**. Página 76.

KIBBUTZ

7. **B**. Primero se ordenan las puntuaciones -7; 0; 3; 5; 12. Tras observar que se dispone de un número impar de observaciones se aplicará la fórmula adecuada (página 2 del formulario). La mediana corresponderá con la observación que ocupa la posición central:

$$Md = X_{(n-1)/2} = 3$$

8. **B**. Primero se toman las puntuaciones 3; 1; 2; 0; 4. A partir de ahí se pueden calcular su media y su desviación típica aplicando las fórmulas correspondientes:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{10}{5} = 2 \ ; \ S_x = \sqrt{\frac{\sum X_i^2}{n} - \bar{X}^2} = \sqrt{\frac{30}{5} - 2^2} = 1{,}41$$

9. **C**. Página 217.

10. **C**. Página 217.

11. **C**. Sabiendo que:

$$r_{xy} = \frac{Cov\ (X,Y)}{S_x\ S_y} \ ; \ Cov\ (X,Y) = r_{xy}\ S_x\ S_y;$$

Siendo el valor mínimo que puede adoptar r_{xy} = -1, sería: Cov (X,Y) = -1 x 6 x 5 = -30;
y siendo el valor máximo que puede adoptar r_{xy} = 1, sería: Cov (X,Y) = 1 x 6 x 5 = 30.

12. **A**. El desarrollo de los datos en la siguiente tabla permite calcular la fórmula de la pendiente:

Sujeto	**X**	**Y**	**X^2**	**Y^2**	**XY**
1	2	8	4	64	16
2	3	9	9	81	27
3	5	7	25	49	35
4	7	12	49	144	84
5	9	12	81	144	108
6	10	12	100	144	120
∑			268	626	390

$$b = \frac{n\sum(XY) - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2} = 0{,}58$$

13. **C**. La proporción de varianza de Y explicada por la variabilidad de X viene dado por el cuadrado del coeficiente de correlación de Pearson. A partir de este valor se calculará el porcentaje:

$$r_{xy} = \frac{n \cdot \sum XY - \sum X \sum Y}{\sqrt{n \cdot \sum X^2 - (\sum X)^2}\sqrt{n \cdot \sum Y^2 - (\sum Y)^2}} = 0{,}816;$$

$$100 \cdot 0{,}816^2 = 100 \cdot 0{,}6658 = 66{,}6\%$$

KIBBUTZ

14. **A**. Si se define el evento Cr_1 = cara al tirar la primera moneda y Cr_2 = cara al tirar la segunda moneda. Siendo los dos lanzamientos sucesos independientes.

$$P(Cr_1) = \frac{1}{2} = 0{,}5; y\ P(Cr_2) = \frac{1}{2} = 0{,}5$$

Se obtendrá que:

$$P(Cr_1 \cap Cr_2) = P(Cr_1) \cdot P(Cr_2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

15. **C**. Se calculará la probabilidad de resolver el problema correctamente.

RC = Resolver Correctamente

P (*A*) = 0,80; *P* (*RC*/*A*) = 0,60

P (*B*) = 1 - *P* (*A*) = 1 - 0,80 = 0,20; *P* (*RC*/*B*) = 0,50

P (*RC*) = *P* (*A*∩*RC*) + *P* (*B*∩*RC*) = *P* (*A*) · *P* (*RC*/*A*) + *P* (*B*) · *P* (*RC*/*B*)

P (*RC*) = 0,80 · 0,60 + 0,20 · 0,50 = 0,48 + 0,10 = 0,58

16. **B**. Con los datos del enunciado puede calcularse que:

$$P\ (A \cup B) = P(A) + P(B) - P(A \cap B) = [1 - P(\bar{A})] + P(B) - P(A \cap B)$$

$$P\ (A \cup B) = \left(1 - \frac{2}{3}\right) + \frac{1}{2} - \frac{1}{5} = \frac{1}{3} + \frac{1}{2} - \frac{1}{5} = \frac{10 + 15 - 6}{30} = \frac{19}{30}$$

17. **A**. Se trata de un experimento binomial *B (n, p)* en el cual *n* = 5 y *p* = 0,52. Para calcular la probabilidad de obtener 2 varones en 5 nacimientos hay que obtener *f* (2)

$$f(2) = P(X = 2) = \binom{5}{2} \cdot 0{,}52^2 \cdot 0{,}48^3 = 10 \cdot 0{,}2704 \cdot 0{,}1105 = 0{,}2987$$

18. **C**. En primer lugar se calculará la función de probabilidad y a continuación puede calcularse la media.

x	0	2	4	6
f(x)	0,25	0,25	0,25	0,25

$$\mu_x = \sum x \cdot f(x) = 0 \cdot 0{,}25 + 2 \cdot 0{,}25 + 4 \cdot 0{,}25 + 6 \cdot 0{,}25 = 0 + 0{,}5 + 1 + 1{,}5 = 3$$

19. **A**. En primer lugar se acude a la tabla de la tabla de la curva normal para obtener la puntuación z que corresponde al percentil 33, es decir 0,3300, que es igual a z = -0,44. A continuación se despeja la puntuación directa correspondiente al percentil 33.

$$z = \frac{X - \bar{X}}{S_x}\ ;\ -0{,}44 = \frac{P_{33} - 40}{10};\ P_{33} = 35{,}6$$

KIBBUTZ

20. **B**. Dado que la distribución *t* de Student es simétrica, el valor superado por el 70% de las puntuaciones corresponderá con el percentil 30, es decir, el valor que deja por debajo el 30% de las puntuaciones. En primer lugar se acude a la tabla de la distribución *t* de Student con 24 g.l.
A continuación se hallará 0,531 (Percentil 70), para convertirlo después en su valor simétrico -0,531 (Percentil 30).

21. **B**. Sabiendo que la puntuación típica z = -2,65 supera a 5 sujetos. En primer lugar se halla la proporción de sujetos que está por debajo de z = -2,65, esto es 0,0040. A partir de ahí, si 5 sujetos suponen el 0,0040 del total (0,4%), el 100% de los sujetos será:

$$\text{Total de sujetos} = \frac{5 \cdot 100}{0{,}4} = 1250$$

22. **C**. Página 379.

23. **B**. Sabiendo que la desviación típica de la población es 20. Primero se hallará la desviación típica muestral correspondiente a una distribución normal de la variable X con varianza poblacional conocida. A continuación puede despejarse su varianza.

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\ ; \sigma_{\bar{X}} = \frac{20}{\sqrt{100}}\ ; \sigma_{\bar{X}} = 2;\ \sigma_{\bar{X}}^2 = 4$$

24. **C**. Se calcularán los intervalos confidenciales con un nivel de confianza del 90% y con un nivel de confianza del 95% para comprobar si la media de 82 cae o no dentro de alguno de estos intervalos. Cosa que no es así.

Con un nivel de confianza del 90%

$$L_{inf} = \bar{X} - \left|Z_{\alpha/2}\right| \frac{\sigma}{\sqrt{n}} = 80 - 1{,}64 \cdot \frac{9}{14} = 80 - 1{,}06 = 78{,}94$$

$$L_{sup} = \bar{X} + \left|Z_{\alpha/2}\right| \frac{\sigma}{\sqrt{n}} = 80 + 1{,}64 \cdot \frac{9}{14} = 80 + 1{,}06 = 81{,}06$$

Con un nivel de confianza del 95%

$$L_{inf} = \bar{X} - \left|Z_{\alpha/2}\right| \frac{\sigma}{\sqrt{n}} = 80 - 1{,}96 \cdot \frac{9}{14} = 80 - 1{,}26 = 78{,}74$$

$$L_{sup} = \bar{X} + \left|Z_{\alpha/2}\right| \frac{\sigma}{\sqrt{n}} = 80 + 1{,}96 \cdot \frac{9}{14} = 80 + 1{,}26 = 81{,}26$$

25. **A**. Se sabe que p = 0,75 y también que E_{max} = 0,05. A partir de estos datos puede calcularse el tamaño muestral.

$$n = \frac{Z_{\alpha/2}^2 \cdot P \cdot (1-P)}{E_{max}^2} = \frac{2{,}58^2 \cdot 0{,}75 \cdot 0{,}25}{0{,}05^2} = \frac{6{,}66 \cdot 0{,}75 \cdot 0{,}25}{0{,}0025} = \frac{1{,}25}{0{,}0025} = 500$$

KIBBUTZ

Código asignatura	Nombre asignatura
62011037	**Introducción al Análisis de Datos**
Fecha alta y origen	Convocatoria
19/09/2019	**SEPTIEMBRE 2019**
Curso virtual	

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
SEPTIEMBRE 2019 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO G DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Tabla 1. Frecuencia relativa de las puntuaciones en una prueba de recuerdo (X) para evaluar la comprensión lectora de 150 niños.

X	p_i
22-24	0,18
19-21	0,24
16-18	0,34
13-15	0,14
10-12	0,10

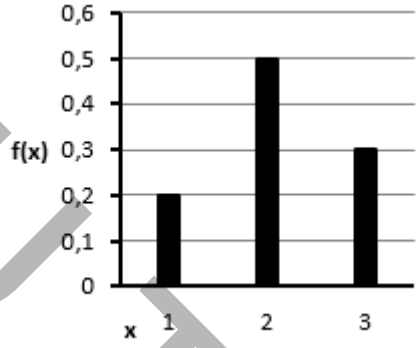
Tabla 2. Datos de autoestima (X) y ajuste emocional (Y) de seis adolescentes.

Adolescente	X	Y
A	20	50
B	80	60
C	120	100
D	30	70
E	10	30

Tabla 3. Participación de los jóvenes de un pueblo en un curso de informática gratuito, en función de su situación laboral.

		Participación curso (*Y*)	
		Sí	No
Situación laboral (*X*)	En paro	160	70
	En activo	60	110

Gráfica 1: Función de Probabilidad, f(x), de una variable aleatoria discreta X.



1. ¿Qué gráfico representaría de manera más adecuada los valores de la variable X que se indican en la Tabla 1? A) Diagrama de barras; **B) Histograma**; C) Diagrama de sectores.
2. ¿Cuánto vale la moda en la distribución de frecuencias de la Tabla 1? **A) 17**; B) 14; C) 20.
3. Para los datos recogidos en la Tabla 1 la frecuencia absoluta acumulada (n_a) del intervalo 19-21 de X es: A) 87; **B) 123**; C) 36.
4. Siguiendo con los datos de la Tabla 1, ¿cuál es el valor de la media aritmética? **A) 17,78**; B) 16,87; C) 15,33.
5. La frecuencia absoluta acumulada se refiere: A) al número de veces que se repite cada uno de los valores de una variable; B) al valor de la frecuencia relativa multiplicado por cien; **C) al número de veces que se repite cada valor o cualquiera de los valores inferiores.**
6. El extremo superior del bigote en un diagrama de caja viene determinado por: A) el valor del límite superior; B) el valor de la puntuación atípica; **C) el valor de la puntuación máxima que se encuentra dentro del límite superior.**
7. Según los datos de la Tabla 2, la amplitud total de la variable ajuste emocional es: A) 30; **B) 70**; C) 100.
8. Con la información recogida en la Tabla 1, ¿qué puntuación corresponde al percentil 40? A) 13,98; B) 19,80; **C) 16,91**.

KIBBUTZ

9. Siguiendo con los datos de la Tabla 2, la puntuación típica del sujeto B en autoestima es: A) 0,37; **B) 0,67**; C) 0,97.
10. La representación en un diagrama de barras conjunto es apropiada cuando: **A) al menos una de las dos variables es cualitativa;** B) las dos variables son cualitativas; C) las dos variables son cuantitativas.
11. Con los datos de la Tabla 3, el valor del estadístico χ^2 es: A) 33,27; B) 57,3; **C) 46,38.**
12. En el estudio sobre autoestima y ajuste emocional en adolescentes, cuya información aparece en la Tabla 2, la correlación entre ambas variables es: **A) 0,85**; B) -0,58; C) 0,38.
13. En una recta de regresión ($Y_i' = a + b\ X_i$) Y_i' recibe el nombre de: A) variable predictora; **B) variable pronosticada o criterio**; C) variable en el origen.
14. Supongamos que en una muestra de trabajadores de una empresa la variable empatía explica el 38% del bienestar laboral. Ese dato nos está indicando que: A) el coeficiente de correlación es 0,38; **B) el coeficiente de determinación es 0,38**; C) la pendiente de la recta de regresión es 0,38.
15. Con los datos de la Tabla 3, si se elige a un joven al azar, ¿cuál es la probabilidad de que haya participado en el curso de informática y esté en paro? A) 0,28; B) 0,58; **C) 0,40**.
16. ¿Cuál es la probabilidad del suceso complementario de A, $P(\bar{A})$ siendo ambos sucesos exhaustivos y excluyentes? **A) 1 – *P(A)*;** B) 1 + *P(A)*; C) $P(\bar{A}) = P(A)$.
17. Desde la Dirección General de Tráfico han elaborado un test para detectar el consumo de hachís en conductores. Para evaluar la sensibilidad y especificidad del test, lo han pasado a 300 personas que habían consumido previamente ese tipo de droga. Al aplicar el test, 25 personas han dado negativo en la detección del consumo. ¿Cuál es la sensibilidad de la prueba? A) 0,8125; B) 0,7880; **C) 0,9167**.
18. Según los datos de la Gráfica 1, ¿cuál es la esperanza matemática de X? A) 1,2; **B) 2,1**; C) 2,9.
19. Siguiendo los datos de la Gráfica 1, la probabilidad de sacar una puntuación igual o menor que 2 es: **A) 0,7**; B) 0,5; C) 0,9
20. Sabiendo que una variable X se distribuye normalmente, ¿cuál es la probabilidad de obtener valores comprendidos entre z = -0,72 y z = 2,45? A) 0,2358; **B) 0,7571**; C) 0,9929.
21. Sea X una variable que se distribuye según una distribución *t* con 8 grados de libertad. ¿Cuál es la probabilidad de obtener valores menores o iguales a 0,706? A) 0,700; B) 0,890; **C) 0,750.**
22. El muestreo aleatorio estratificado es un tipo de muestreo: A) no probabilístico; B) por cuotas; **C) probabilístico.**
23. Sea X una variable con distribución normal en la población con N(150; 8). Se extrae una muestra aleatoria simple de tamaño 15, ¿cuál es la probabilidad de que su media sea menor o igual a 145? **A) 0,0078;** B) 0,9474; C) 0,9922.
24. Se sabe que la variable percepción viso-espacial se distribuye normalmente en la población de adolescentes. Se extrae una muestra aleatoria de 150 adolescentes y se obtiene una cuasivarianza muestral de 5. ¿Entre qué valores se encuentra el intervalo de confianza de la varianza para un nivel de confianza del 95%? A) (2,62; 3,62); **B) (3,87; 6,13);** C) (7,24; 8,62).
25. Una variable X se distribuye normalmente en la población con una varianza poblacional igual a 49. Si se admite un error máximo de estimación de 0,2 y se fija un nivel de confianza del 95%, ¿qué tamaño muestral es necesario para estimar la media poblacional asumiendo población infinita?: A) 2750; B) 3210; **C) 4706.**

EXTRAOFICIAL

Soluciones:

1. B

2. A

El intervalo con la frecuencia más alta es 16-18, con ni = 51. Por tanto, la moda es su punto medio → Mo = 17

X	X_i	n_i
22-24	23	27
19-21	20	36
16-18	17	51
13-15	14	21
10-12	11	15
∑	85	150

3. B

X	p_i	n_i	n_a
22-24	0,18	27	150
19-21	0,24	36	123
16-18	0,34	51	87
13-15	0,14	21	36
10-12	0,10	15	15

$n_i = p_i \text{ x } n$ = 123

4. A El cálculo de la media con las frecuencias relativas:

X	X_i	p_i	$p_i\ X_i$
22-24	23	0,18	4.14
19-21	20	0,24	4.8
16-18	17	0,34	5,78
13-15	14	0,14	1,96
10-12	11	0,10	1.1
∑	85		17,78

$$\bar{X} = \sum p_i X_i = 17{,}78$$

También se puede calcular con las frecuencias absolutas

X	X_i	n_i	$n_i\ X_i$
22-24	23	27	621
19-21	20	36	720
16-18	17	51	867
13-15	14	21	294
10-12	11	15	165
∑	85	150	2667

$$\bar{X} = \frac{\sum n_i X_i}{n} = \frac{2667}{150} = 17{,}78$$

5. C

KUBUTZ

6. C (Página 124)

7. B

La amplitud total o rango del conjunto de puntuaciones de la variable Y (ajuste emocional) es:

$A_T = Y_{max} - Y_{min} = 100 - 30 = 70$

8. C

X	p_i	n_i	n_a
22-24	0,18	27	150
19-21	0,24	36	123
16-18	0,34	51	87
13-15	0,14	21	36
10-12	0,10	15	15

En este caso, dado que $\frac{nk}{100} = \frac{150x40}{100} = 60$ el intervalo crítico va a ser (16-18), que es el que contiene a la frecuencia absoluta acumulada de 60

$$P_{40} = L_i + \left(\frac{\frac{n \cdot k}{100} - n_d}{n_c}\right) \cdot I = 15{,}5 + \left(\frac{\frac{150 \times 40}{100} - 36}{51}\right) \times 3 = 15{,}5 + \left(\frac{60-36}{51}\right) \times 3 = 15{,}5 + 1{,}41 = 16{,}91$$

9. B

Adolescente	X_i	X_i^2
A	20	400
B	80	6400
C	120	14400
D	30	900
E	10	100
∑	260	22200

$$\bar{X} = \frac{\sum X_i}{n} = \frac{260}{5} = 52$$

$$S_x^2 = \frac{\sum X_i^2}{n} - \bar{X}^2 = \frac{22200}{5} - 52^2 = 4440 - 2704 = 1736$$

$$S_X = \sqrt{1736} = 41{,}665 \cong 41{,}67$$

$$Z = \frac{80 - \bar{X}}{S_x} = \frac{80 - 52}{41{,}67} = \frac{28}{41{,}67} = 0{,}67$$

10. A (Página 155)

KIBBUTZ

11. C

		Participación curso (*Y*)		
		Sí	**No**	
Situación laboral (*X*)	**En paro**	160 (126,5)	70 (103,5)	230
	En activo	60 (93,5)	110 (76,5)	170
		220	180	400

$n_{t11} = \frac{230\ x\ 220}{400} = 126{,}5$ $\quad n_{t12} = \frac{230\ x 180}{400} = 103{,}5$

$n_{t21} = \frac{170\ x\ 220}{400} = 93{,}5$ $\quad n_{t22} = \frac{170\ x\ 180}{400} = 76{,}5$

$\chi^2 = \sum\sum \frac{(n_e - n_t)^2}{n_t} = \frac{(160-126{,}5)^2}{126{,}5} + \frac{(70-103{,}5)^2}{103{,}5} + \frac{(60-93{,}5)^2}{93{,}5} + \frac{(110-76{,}5)^2}{76{,}5} = \frac{33{,}5^2}{126{,}5} + \frac{(-33{,}5)^2}{103{,}5} + \frac{(-33{,}5)^2}{93{,}5} + \frac{33{,}5^2}{76{,}5} =$
$8{,}87 + 10{,}84 + 12 + 14{,}67 = 46{,}38$

12. A

Adolescente	X_i	Y_i	XY	X_i^2	Y_i^2
A	20	50	1000	400	2500
B	80	60	4800	6400	3600
C	120	100	12000	14400	10000
D	30	70	2100	900	4900
E	10	30	300	100	900
∑	260	310	20200	22200	21900

$r_{XY} = \frac{S_{XY}}{S_X S_Y}$

Del ejercicio 8 tenemos calculada la media y desviación típica de X:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{260}{5} = 52$$

$$S_x^2 = \frac{\sum X_i^2}{n} - \bar{X}^2 = \frac{22200}{5} - 52^2 = 4440 - 2704 = 1736$$

$$S_X = \sqrt{1736} = 41{,}665 \cong 41{,}67$$

La media y desviación típica de Y son:

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{310}{5} = 62$$

$$S_Y^2 = \frac{\sum Y_i^2}{n} - \bar{Y}^2 = \frac{21900}{5} - 62^2 = 4380 - 3844 = 536$$

$$S_Y = \sqrt{536} = 23{,}15$$

KIBBUTZ

$$S_{XY} = \frac{\sum_{i=1}^{n} X_i Y_i}{n} - \bar{X}\,\bar{Y} = \frac{20200}{5} - 52\ x\ 62 = 4040 - 3224 = 816$$

Por tanto, $r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{816}{41{,}67\ x\ 23{,}15} = \frac{816}{964{,}66} = 0{,}85$

13. B (Página 220)
14. B (Página 221)
15. C En este caso, nos están pidiendo P (sí ∩ paro) = n $_{si \cap paro}$ /n $_{total}$ = $\frac{160}{400} = 0{,}40$

		Participación curso (*Y*)		
		Sí	**No**	
Situación laboral (*X*)	**En paro**	160	70	230
	En activo	60	110	170
		220	180	400

16. A página 264
17. C

La sensibilidad, en este caso, es la probabilidad de detectar los casos positivos (que han consumido) del total de la muestra que ha consumido (T)

P (+ \ T) = $\frac{n_{+\cap T}}{n_T} = \frac{275}{300} = 0{,}9167$ → 92%

18. B

x	*f(x)*	*x f(x)*
1	0,2	0,2
2	0,5	1,0
3	0,3	0,9
	1	2,1

$\mu = \sum x \cdot f(x) = 2{,}1$

19. A

F(2) = P (X ≤ 2) = P (X = 1) + P (X = 2) = f(1) + f(2) = 0,2 + 0,5 = 0,7

20. B

Para z = - 072, buscamos en la Tabla III. La probabilidad que deja por debajo de sí esa puntuación es 0,2358

Para z = 2,45 la probabilidad es 0,9929

Por tanto, la probabilidad de obtener valores comprendidos entre ambos valores de z es 0,9929 - 0,2358 = 0,7571

21. C Ver Tabla VI
22. C Página 379
23. A

$$P(\ \bar{X}\ \leq 145) = P\ (z \leq \frac{145 - 150}{8/\sqrt{15}}) = P\ (z\ \leq \frac{-5}{2{,}0656}) = P(z \leq -2{,}42)$$

Ver Tabla III es 0,0078

KIBBUTZ

24. B

$$L_{\inf} = S_{n-1}^2 - |Z_{\alpha/2}| S_{n-1}^2 \sqrt{\frac{2}{n}} = 5 - 1{,}96 \times 5 \times \sqrt{\frac{2}{150}} = 5 - 1{,}1316 = 3{,}8684 \quad \rightarrow 3{,}87$$

$$L_{\sup} = S_{n-1}^2 + |Z_{\alpha/2}| S_{n-1}^2 \sqrt{\frac{2}{n}} = 5 + 1{,}96 \times 5 \times \sqrt{\frac{2}{150}} = 5 + 1{,}1316 = 6{,}1316 \quad \rightarrow 6{,}13$$

25. C

Conocida la varianza poblacional asumiendo población infinita.

$n.c. = 0{,}95 \rightarrow \alpha = 0{,}05 \rightarrow Z_{\alpha/2} = Z_{0{,}025} = 1{,}96$ (Tabla III)

$\sigma^2 = 49$, $E_{máx} = 0{,}2$

$$n = \frac{\sigma^2 Z_{\alpha/2}^2}{E_{máx}^2} = \frac{49 \times 1{,}96^2}{0{,}2^2} = \frac{188{,}2384}{0{,}04} = 4705{,}96 \approx 4706$$

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2019 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO B DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Tabla 1. Distribución, en proporciones, de las puntuaciones de un grupo de 150 sujetos en un test de deseabilidad social.

X	p_i
7 – 9	0,10
10 – 12	0,14
13 – 15	0,34
16 – 18	0,24
19 – 21	0,18

Tabla 2. Datos de 80 aspirantes a un puesto de trabajo en un proceso de selección, en función de si han superado (1) o no (0) tanto un test como una entrevista.

	ENTREVISTA	
TEST	**0**	**1**
0	28	12
1	12	28

Tabla 3. Puntuaciones de 5 sujetos en dos pruebas X e Y.

Sujeto	X	Y
1	5	8
2	7	6
3	9	8
4	12	7
5	15	10

Supuesto 1. Una persona pide a su vecino que riegue una planta delicada mientras está de vacaciones. La probabilidad de que la planta sobreviva suponiendo que ha sido regada es 0,85 y la probabilidad de que la planta sobreviva suponiendo que NO ha sido regada es 0,20. La probabilidad de que el vecino se acuerde de regar la planta es 0,90.

1. Una variable politómica es un tipo de variable: A) ordinal con múltiples modalidades; B) cuasicuantitativa; **C) cualitativa.**
2. Supongamos que la inteligencia (CI) se mide en una escala de intervalo, y que para dos sujetos 1 y 2 tenemos: CI_1=70 y CI_2= 140. Podemos afirmar que: A) el sujeto 2 es el doble de inteligente que el sujeto 1; **B) el sujeto 2 supera al sujeto 1 en 70 unidades de medida**; C) el sujeto 1 es la mitad de inteligente que el sujeto 2.
3. En una distribución de frecuencias que recoge información sobre la edad a la que los jóvenes comienzan a consumir bebidas alcohólicas se quiere obtener la edad a la que el 50% de los jóvenes inician el consumo, ¿qué índice calcularía? A) Media; **B) Mediana**; C) Moda.
4. De acuerdo con los datos de la Tabla 1, ¿cuántos sujetos obtienen puntuaciones entre 9,5 y 12,5? **A) 21**; B) 23; C) 25.
5. ¿Cuál es la moda en los datos de la Tabla 1? A) 14; **B) 0,34**; C) al ser una variable cuantitativa no se puede calcular.
6. ¿A qué percentil corresponde el valor X = 11 en la distribución de la Tabla 1? **A) 17**; B) 20; C) 23.
7. ¿Cuál es la varianza de los datos de la Tabla 1? A) 10,20; **B) 12,89**; C) 11,23.

EXAMEN

8. A un grupo de niños de 1º de Bachillerato se les administra un test de razonamiento y otro de comprensión verbal. Para saber en cuál de estas pruebas existe mayor variabilidad, comparamos: A) las desviaciones medias de cada test; **B) los coeficientes de variación**; C) las varianzas de cada test en puntuaciones típicas.
9. La amplitud intercuartil siempre toma valores: A) entre -1 y +1; **B) mayores o iguales que cero**; C) puede tomar cualquier valor positivo o negativo.
10. Con los datos de la Tabla 2, la distribución marginal expresada en proporciones para la variable Entrevista toma los valores: A) 0,7 y 0,3; B) 0,3 y 0,7; **C) 0,5** y 0,5.
11. Con los datos de la Tabla 2, la proporción conjunta para superar el Test pero no la Entrevista vale: **A) 0,15**; B) 0,17; C) 0,3.
12. Con los datos de la Tabla 2, el grado de asociación entre el Test y la Entrevista expresado mediante el coeficiente φ es igual a: A) 12,80; **B) 0,40**; C) 0,71.
13. ¿Cuál es el valor del coeficiente de correlación de Pearson para los datos de la Tabla 3? **A) 0,58**; B) 0,65; C) 0,70.
14. Con los datos de la Tabla 3 realizamos las siguientes transformaciones: H = X + 3; V = Y – 4. ¿Cuál será la covarianza entre H y V? A) 4,36; **B) 2,72**; C) 3,35.
15. La pendiente de la recta de regresión de Y sobre X: A) es una constante que indica el valor que adopta la variable Y cuando la variable X vale 0; **B) es una constante que indica en qué medida cambian los valores de Y por cada incremento de una unidad en los valores de X**; C) es variable y depende del valor que adopta la variable X cuando la variable Y vale 0.
16. En el Supuesto 1, ¿cuál es la probabilidad de que la planta siga viva cuando la persona vuelve de vacaciones? A) 0,849; B) 0,745; **C) 0,785**.
17. En el Supuesto 1. Si la planta está muerta cuando la persona vuelve de vacaciones, ¿cuál es la probabilidad de que el vecino olvidase regarla? A) 0,10; **B) 0,372**; C) 0,145.
18. Los valores de una variable aleatoria discreta X son 0, 1, 2, 3, 4, 5. Si se sabe que $F(4) = P(X \leq 4) = 0{,}974$, determinar $P(X = 5)$: A) 0,037; **B) 0,026**; C) 0,012.
19. Una variable aleatoria discreta X puede tomar los valores 0, 1, 2 y 3, siendo estos valores equiprobables. La media, μ, de la variable X vale: **A) 1,5**; B) 2,5; C) 3,5.
20. El área de una distribución normal con $\mu = 15$ y $\sigma = 5$ que se encuentra entre los valores $X_s = 18$ y $X_i = 14$ se puede expresar como: A) $F(X \leq 18) – F(X \leq 14) = 0{,}105$; B) $P(X = 18) – P(X = 14) = 0{,}011$; **C) $F(X \leq 18) – F(X \leq 14) = 0{,}305$**.
21. Una variable aleatoria continua X se distribuye según una función de densidad de probabilidad normal tipificada. $P(X \leq 0)$ vale: **A) 0,5**; B) 0,691; C) 0,308.
22. ¿Cuál de las siguientes afirmaciones corresponde a un estadístico? A) En el Reino Unido, el 51,9% de los votantes ha rechazado la permanencia en la Unión Europea; B) En el referéndum del Reino Unido para la permanencia en la Unión Europea, la participación fue del 72,2%; **C) Las encuestas previas a la consulta electoral afirmaban que el NO a la permanencia en la Unión Europea alcanzaba cotas inferiores al 45%.**
23. En la siguiente definición: "Un ..... es una propiedad descriptiva de una población", el término adecuado es: A) estadístico; **B) parámetro**; C) índice.
24. ¿Cuál de las siguientes características no es una propiedad de los estimadores? A) Consistencia; B) Suficiencia; **C) Independencia.**
25. Se selecciona una muestra aleatoria de 90 sujetos de una población de pacientes con problemas alimenticios de entre 20 y 25 años de edad y un tercio de ellos sufre anorexia nerviosa. Determine con $\alpha = 0{,}01$ el intervalo de confianza para estimar la proporción de sujetos con anorexia nerviosa. A) (0,23 y 0,43); **B)** (0,20 y 0,46); C) (0,23 y 0,46).

KUBICIN

SOLUCIONES

1. C. Página 13.

2. B. Página 15.

3. B. Página 62.

4. A.

Desarrollo de la Tabla 1

X	p_i	n_i	n_a	X_i	X_i^2	n_iX_i	$n_iX_i^2$
7 – 9	0,10	15	15	8	64	120	960
10 – 12	0,14	21	36	11	121	231	2541
13 – 15	0,34	51	87	14	196	714	9996
16 – 18	0,24	36	123	17	289	612	10404
19 - 21	0,18	27	150	20	400	540	10800
∑	-	150	-	-	-	2217	34701

9,5 y 12,5 son los límites exactos del intervalo, y 0,14 la proporción de sujetos en ese intervalo. Por tanto 0,14 x 150 = 21 sujetos que obtienen puntuaciones en ese intervalo.

5. B. La moda es el punto medio del intervalo que contiene el mayor número de observaciones (n_i = 51), por lo que su valor es: Mo = 14. No hay que confundir la moda, que es la puntuación más repetida, con la proporción de veces (o el número de veces) que se repite dicha puntuación, que es en este caso 0,34 (opción B).

6. A.

$$k = \left[\frac{\frac{(P_k - L_i)\cdot n_c}{I} + n_d}{n}\right]\cdot\ 100 = \left[\frac{\frac{(11-9{,}5)\cdot 21}{3} + 15}{150}\right]\cdot 100 = 17$$

También se puede resolver esta cuestión tan sólo con la columna de proporciones. X = 11 es el punto medio del intervalo, luego deja por debajo a la mitad de las observaciones dentro de este intervalo (en proporciones 0,14 / 2 = 0,07). Se suman las observaciones del intervalo anterior (0,10), y en total supera a una proporción: 0,07 + 0,10 = 0,17, que expresado en porcentajes corresponde al 17%. Es decir: P_{17} = 11.

7. B. A partir del desarrollo de la Tabla 1, calculamos la media y posteriormente la varianza:

$$\bar{X} = \frac{\sum n_i \cdot X_i}{n} = \frac{2217}{150} = 14{,}78 \qquad S_x^2 = \frac{\sum n_i \cdot X_i^2}{n} - \bar{X}^2 = \frac{34701}{150} - 14{,}78^2 = 12{,}89$$

KIBBUTZ

8. B. La respuesta correcta es B, porque el coeficiente de variación se utiliza para comparar la variabilidad de dos distribuciones.

9. B. Página 115.

10. C. La distribución marginal expresada en proporciones para la variable Entrevista toma los valores 0,5 y 0,5.

	ENTREVISTA		
TEST	**0**	**1**	
0	28	12	
1	12	28	
	40(0,5)	40 (0,5)	80

11. A. El número de aspirantes que superan el Test pero no la Entrevista es 12, que sobre 80 aspirantes corresponde a una proporción 12/80 = 0,15.

12. B. Calculamos el coeficiente φ; dado por:

$$\varphi = \frac{n_{11} \cdot n_{22} - n_{12} \cdot n_{21}}{\sqrt{n_{1+} \cdot n_{2+} \cdot n_{+1} \cdot n_{+2}}} = \frac{28 \cdot 28 - 12 \cdot 12}{\sqrt{40 \cdot 40 \cdot 40 \cdot 40}} = \frac{640}{1600} = 0{,}4$$

Desarrollo de la Tabla 3

X	Y	X^2	Y^2	XY
5	8	25	64	40
7	6	49	36	42
9	8	81	64	72
12	7	144	49	84
15	10	225	100	150
48	39	524	313	388

13. A. Calculamos la fórmula del coeficiente de correlación de Pearson para los datos de la tabla 3:

$$r_{xy} = \frac{n \cdot \sum XY - \sum X \sum Y}{\sqrt{n \cdot \sum X^2 - (\sum X)^2}\sqrt{n \cdot \sum Y^2 - (\sum Y)^2}} = \frac{5 \cdot 388 - 48 \cdot 39}{\sqrt{5 \cdot 524 - 48^2} \cdot \sqrt{5 \cdot 313 - 39^2}} = 0{,}58$$

14. B. Dado que las transformaciones aplicadas a las variables X e Y no van a afectar al valor de la covarianza entre H y V, se calcula la covarianza entre X e Y

$$S_{xy} = \frac{\sum XY}{n} - \bar{X} \cdot \bar{Y} = \frac{388}{5} - \frac{48}{5} \cdot \frac{39}{5} = 2{,}72$$

KIBBUTZ

15. B. Página 220.

16. C.

Probabilidades asociadas a la Situación 1:

Definimos V como el evento que denota que la planta sigue viva, y R como el evento que denota que ha sido regada; Siguiendo la notación que se utiliza en el manual de la asignatura disponemos de:

La probabilidad de que la planta no siga viva dado que no ha sido regada, y la probabilidad de que siga viva a pesar de no haber sido regada.

$P(\bar{V}\backslash\bar{R}) = 0{,}80;\ P(V\backslash\bar{R}) = 0{,}20$;

La probabilidad de que la planta no siga viva dado que no fue regada, y la probabilidad de que siga viva habiendo sido regada.

$P(\bar{V}\backslash R) = 0{,}15;\ P(V\backslash R) = 0{,}85$;

La probabilidad de que la planta fuese regada y la probabilidad de que no fuese regada

$P(R) = 0{,}90;\ P(\bar{R}) = 0{,}10$

La probabilidad de que la planta siga viva es igual a la suma de las probabilidades de que la siga viva dado que se halla regado o no.

$$P(V) = P(V\backslash R)P(R) + P(V\backslash\bar{R})P(\bar{R})\ ;\ P(V) = (0{,}85)(0{,}90) + (0{,}20)(0{,}10) = 0{,}785$$

17. B. La probabilidad de que nuestro vecino olvidase regar la planta en el caso de que esta aparezca muerta es igual a la probabilidad de que la planta no fuese regada condicionada a que no sobreviviese.

$$P(\bar{R}\backslash\bar{V}) = \frac{P(\bar{V}\backslash\bar{R})P(\bar{R})}{P(\bar{V})}$$

Se podría utilizar el resultado de la pregunta 16, o de otro modo sería siempre necesario calcular $P(\bar{V})$.

$$P(\bar{R}\backslash\bar{V}) = \frac{P(0{,}80)P(0{,}10)}{P(0{,}215)} = \frac{16}{43} = 0{,}372$$

18. B. Si F(5) = P(X ≤ 5) = 1; Tan solo hay que restar F(4) para conocer P(X = 5). Así, 1- 0,974 = 0,026.

19. A. Siendo todos los valores equiprobables tendremos:

KIBBUTZ

x	f(x)	x f(x)
0	0,25	0
1	0,25	0,25
2	0,25	0,50
3	0,25	0,75
		1,50

20. C. Obtenemos las puntuaciones z para cada puntuación directa 18 y 14 utilizando la fórmula:

$$z = \frac{X-\mu}{\sigma} \begin{Bmatrix} z_s = z_{18} = \frac{18-15}{5} = 0{,}6; P(z \leq 0{,}6) = 0{,}7257 \\ z_i = z_{14} = \frac{14-15}{5} = -0{,}2; P(z \leq -0{,}2) = 0{,}4207 \end{Bmatrix}$$

La diferencia entre estas dos puntuaciones típicas corresponderá al área situada entre las puntuaciones 18 y 14, es decir F(X≤ 18) – F(X≤ 14)= 0,7257- 0,4207= 0,305.

21. A. La puntuación típica z= 0 corresponde con la media de la distribución normal, que deja por debajo la mitad de la distribución, es decir P(X≤ 0)= 0,5.

22. C. Página 377.

23. B. Página 376.

24. C. Página 434.

25. B. La variable aleatoria es una variable del tipo Bernoulli. Se conoce la proporción π en la población (o el valor *P* en la muestra). Respecto al tamaño de la muestra es suficientemente grande como para usar la aproximación a la normal, por tanto:

$$P \rightarrow N\left(\pi, \sqrt{\frac{\pi(1-\pi)}{n}}\right); \text{ P= 30/90= 0,33}$$

$$L_{inf} = P - \left|Z_{\alpha/2}\right| \pi, \sqrt{\frac{P(1-P)}{n}} = 0{,}33 - 2{,}58 \cdot \sqrt{\frac{0{,}33(1-0{,}33)}{90}} = 0{,}33 - 0{,}13 = 0{,}20$$

$$L_{sup} = P + \left|Z_{\alpha/2}\right| \pi, \sqrt{\frac{P(1-P)}{n}} = 0{,}33 + 2{,}58 \cdot \sqrt{\frac{0{,}33(1-0{,}33)}{90}} = 0{,}33 + 0{,}13 = 0{,}46$$

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2019 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Tabla 1. Distribución, expresada en proporciones acumuladas, de 150 sujetos en una prueba cognitiva.

X	Pa
22 – 24	1,00
19 – 21	0,82
16 – 18	0,58
13 – 15	0,24
10 – 12	0,10

Tabla 2. Distribución de dos variables X e Y.

	Y_1= 0	Y_2= 1
X_1= 0	2	2
X_2= 1	5	10
X_3= 2	3	8

Supuesto 1. Según una compañía de seguros las personas pueden dividirse en dos tipos: propensos y no propensos a sufrir accidentes. Sus estadísticas muestran que una persona propensa a sufrir accidentes tendrá un accidente durante el próximo año con una probabilidad 0,4, mientras que para una persona no propensa a sufrir accidentes esta probabilidad desciende hasta 0,2. También sabemos que el 30% de las personas son propensas a sufrir accidentes.

1. El punto medio de un intervalo es igual a la semisuma de: **A) los límites exactos**; B) la distancia entre dos puntos medios de dos intervalos adyacentes; C) los límites del intervalo en que se encuentra la mediana.
2. ¿Cuál de las siguientes escalas de medida posee un cero absoluto? A) de intervalo; B) ordinal; **C) de razón.**
3. Con los datos de la Tabla 1, ¿cuántos sujetos obtienen puntuaciones entre 12,5 y 15,5? **A) 21**; B) 23; C) 25.
4. Con los datos de la Tabla 1, ¿qué valor representa al intervalo de mayor frecuencia? A) no se puede calcular; **B) 17**; C) 20.
5. ¿A qué percentil corresponde la puntuación X = 20, en la distribución de la Tabla 1? **A) 70**; B) 41; C) 80.
6. ¿Cuál de los siguientes valores se aproxima más al valor de la $\overline{X}$ de los datos de la Tabla 1? A) 15,89; **B) 17,78**; C) 16,35.
7. La desviación típica de las puntuaciones típicas es: A) igual a cero; **B**) igual a 1; C) puede tomar cualquier valor mayor que cero.
8. Sabiendo que Q3 – Q2 = 3 y que Q2 – Q1 = 1. ¿Cuál es el valor de la amplitud intercuartil? A) 0,5; B) 2; **C) 4.**
9. Dadas las siguientes puntuaciones: 5, 6, 3, 7, 4, ¿cuánto vale su varianza? **A) 2**; B) 1,41; C) 6.
10. Con los datos de la Tabla 2, el valor más aproximado a Ji-cuadrado es: **A) 0,69**; B) 0,55; C) 0,15.

BORRADOR

11. Con los datos de la Tabla 2, la frecuencia teórica esperada en caso de independencia, para los valores $X_2 = 1$ e $Y_1 = 0$ vale: A) 8; **B) es igual a la frecuencia observada**; C) es igual a la frecuencia conjunta.
12. Dos psicólogos evalúan a un grupo de diez adolescentes en tratamiento. Los ordenan por sus características para escoger a uno que colaborará con ellos como coterapeuta. Las diferencias entre los rangos asignados han sido: { -9; -5; -3; 0; 4; 3; 2; 6; 1; 1 }. El valor más aproximado de la correlación entre las preferencias de los dos psicólogos es: A) -0,21; **B) -0,10**; C) 0,90.
13. El coeficiente de correlación entre dos variables X e Y es $r_{xy} = 0,70$, ¿cuál es el valor más aproximado de la varianza explicada por la regresión? **A) 0,49**; B) 0,51; C) 0,30.
14. La varianza de los errores cometidos al pronosticar la variable Y a partir de la variable X se corresponde con: A) la desviación típica de los errores de pronóstico; **B) el error cuadrático medio**; C) la varianza de la variable X.
15. El índice de correlación Biserial Puntual se utiliza con: A) dos variables dicotómicas; B) dos variables ordinales; **C) una variable dicotómica y otra cuantitativa**.
16. En el Supuesto 1, ¿cuál es la probabilidad de que una persona que va a firmar una póliza de seguro sufra un accidente durante el próximo año? A) 0,30; **B) 0,26**; C) 0,20.
17. En el Supuesto 1, dado el caso de que una persona haya tenido un accidente durante el último año, ¿cuál es la probabilidad de que esta persona sea propensa a sufrir accidentes? **A) 0,46**; B) 0,40; C) 0,30.
18. Los valores de una variable aleatoria discreta X son 0, 1, 2, 3, 4, 5. Si se sabe que $F(4) = P(X \leq 4) = 0,974$ y que $F(3) = P(X \leq 3) = 0,963$, ¿cuál es el valor más aproximado para $P(X = 4)$? **A) 0,011**; B) 0,022; C) 0,001.
19. Un experimento Bernoulli es: **A) un experimento aleatorio en el que sólo pueden ocurrir dos sucesos mutuamente excluyentes;** B) un experimento en el que pueden ocurrir tres o más sucesos; C) un experimento no aleatorio.
20. En una distribución normal con media 40 y desviación típica 10, ¿cuál es el valor más cercano que corresponde al percentil 33? **A) 35,6**; B) -0,44; C) 44,4.
21. En una distribución normal con varianza igual a 4, ¿cuál es la proporción de observaciones entre la puntuación diferencial x = -1,5 y la media aritmética? A) 0,23; **B) 0,27**; C) 0,35.
22. En la población de estudiantes de acceso a la universidad el 8% sufre dolores de cabeza en época de estudio. Si se extrae una muestra de 50 estudiantes, ¿cuál sería el valor más aproximado del error típico de medida de la proporción de estudiantes que sufre dolores de cabeza en época de estudio? A) 0,147; B) 0,014; **C) 0,038.**
23. El 25% de usuarios del transporte público informan de síntomas de ansiedad cuando se ven en medio de una aglomeración en el metro. ¿Cuál es la probabilidad de extraer una muestra de 100 en la que los síntomas de ansiedad se den en un 10% de los usuarios o menos? **A) 0,0003**; B) 0,003; C) 0,03.
24. En general, dado un mismo nivel de confianza, cuanto mayor es el tamaño de la muestra: **A) mayor es la precisión del intervalo**; B) menor es la precisión del intervalo; C) mayor es la amplitud del intervalo.
25. Se elige por muestreo aleatorio simple un grupo de 25 sujetos y se les administra una prueba de conservadurismo. La media obtenida en la prueba fue 90 y la cuasivarianza fue 81. Si determinamos un intervalo de confianza del 95%, ¿entre qué valores se encontrará la verdadera media de la población? **A) (86,28; 93,72)**; B) (88,24; 91,76); C) (86,92; 93,08).

KIBBUTZ

SOLUCIONES

1. A. Página 32.

2. C. Página 16.
3. A.

Desarrollo de la Tabla 1

X	p_a	p_i	n_i	X_i	n_a	$X_i\ n_i$
22 - 24	1,00	0,18	27	23	150	621
19– 21	0,82	0,24	36	20	123	720
16 – 18	0,58	0,34	51	17	87	867
13 – 15	0,24	0,14	21	14	36	294
10 - 12	0,10	0,10	15	11	15	165
∑	-	-	150	-	-	2667

Nos preguntan por los sujetos del intervalo 13 – 15, que son 21, que es el resultado de multiplicar la proporción de sujetos en ese intervalo (0,24 - 0,10 = 0,14) por el total de sujetos, es decir, 0,14 x 150 = 21.

4. B. Se pregunta por la Moda, en este caso el punto medio del intervalo con la mayor frecuencia (ni = 51) que se corresponde con X = 17.

5. A. Utilizamos la fórmula para averiguar el percentil:

$$k = \left[\frac{\frac{(P_k - L_i)\cdot n_c}{I} + n_d}{n}\right]\cdot\ 100 = \left[\frac{\frac{(20 - 18{,}5)\cdot 36}{3} + 87}{150}\right]\cdot 100 = 70$$

6. B. Utilizamos la fórmula de la media:

$$\bar{X} = \frac{\sum n_i \cdot X_i}{n} = \frac{2667}{150} = 17{,}78$$

7. B. Página 128.

8. C. Dado que sabemos: $(Q_3 – Q_2) = 3$ y $(Q_2 – Q_1) = 1$ podemos calcular $Q_3 – Q_1 = (Q_3 – Q_2) + (Q_2 – Q_1) = 4$.

9. A. Primero hallamos la media de las puntuaciones que es (5 + 6 + 3 + 7 + 4) / 5 = 5. A continuación elevamos al cuadrado las puntuaciones, 25, 34, 9, 49 y 16. Por último aplicamos la fórmula de la varianza.

$$S_x^2 = \frac{\sum n_i \cdot X_i^2}{n} - \bar{X}^2 = \frac{135}{5} - 5^2 = 2$$

KIBBUTZ

10. A. Calculamos la tabla de frecuencias observadas y esperadas y a continuación calculamos □^{2}

Frecuencias observadas (y esperadas)

	$Y_1 = 0$	$Y_2 = 1$	
$X_1 = 0$	2 (1,33)	2 (2,67)	4
$X_2 = 1$	5 (5)	10 (10)	15
$X_3 = 2$	3 (3,67)	8 (7,33)	11
	10	20	30

$$\chi^2 = \frac{\sum\sum\ (n_e - n_t)^2}{n_t} = 0{,}69$$

11. Esta pregunta se dará por válida a todos los estudiantes, independientemente de que la hayan contestado o no.
La respuesta correcta sería B, que puede extraerse de la tabla del ejercicio anterior o también calcular directamente las frecuencias observadas y esperadas que se requieren, es decir; $n_e = 5$ y $n_t = (15 \times 10) / 30 = 5$.
Sin embargo, el distractor de la respuesta C (es igual a la frecuencia conjunta) puede resultar ambiguo, ya que en el manual de la asignatura no se profundiza tanto como para definir las frecuencias absolutas conjuntas.

12. B. Tenemos dos variables ordinales, luego hemos de aplicar el coeficiente de correlación de Spearman. Calculamos la suma de los cuadrados de las diferencias entre los rangos:

$$\sum d^2 = (-9)^2 + (-5)^2 + (-3)^2 + (0)^2 + (4)^2 + (3)^2 + (2)^2 + (6)^2 + (1)^2 + (1)^2 = 182$$

Y aplicando la fórmula de Spearman.

$$r_s = 1 - \frac{6 \cdot \sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \cdot 182}{10(10^2 - 1)} = -0{,}10$$

13. A. $r_{xy}^2 = (0{,}79)^2 = 0{,}49$.

14. B. Página 226.

15. C. Página 217.

16. B.

Probabilidades asociadas a la Situación 1:

KIBBUTZ

Definimos *P* como el evento que denota que una persona es propensa a sufrir accidentes, y *NP* como el evento que denota que una persona no es propensa a sufrir accidentes. Definimos *A* como el evento que denota tener un accidente durante el próximo año:

Probabilidades de ser propenso o no a sufrir accidentes;

$P(P) = 0{,}3;\ P(NP) = 0{,}7.$

Probabilidad de tener un accidente durante el próximo año siendo propenso sufrir accidentes;

$P(A\backslash P) = 0{,}4.$

Probabilidad de tener un accidente durante el próximo año no siendo propenso a sufrir accidentes; $P(A\backslash NP) = 0{,}2.$

La probabilidad de tener un accidente durante el próximo año es igual a:

$$P(A) = P(A\backslash P)\cdot P(P) + P(A\backslash NP)P(NP) = (0{,}4)\cdot(0{,}3) + (0{,}2)\cdot(0{,}7) = 0{,}26$$

17. A. La probabilidad de que una persona propensa a sufrir accidentes haya tenido un accidente es igual a: Teorema de Bayes

$$P(P\backslash A) = \frac{P(P\cap A)}{P(A)} = \frac{P(A\backslash P)\cdot P(P)}{P(A\backslash P)\cdot P(P) + P(A\backslash NP)P(NP)}$$

$$P(P\backslash A) = \frac{(0{,}4)(0{,}3)}{(0{,}4)(0{,}3) + (0{,}2)(0{,}7)} = 0{,}46$$

18. A. En esta ocasión se trata de averiguar la probabilidad de X = 4, por tanto tendremos que restar F(4) - F(3) y así obtendremos 0,974 - 0,963 = 0,011.

19. A. Página 315.

20. A. En la tabla de la curva normal la puntuación típica que deja por debajo al 33% de las observaciones es -0,44. Por tanto, aplicando la fórmula de puntuación z tendremos:

$$z = \frac{X-\bar{X}}{S_x};\ -0{,}44 = \frac{X-40}{10}; -4{,}4 = X - 40; X = 35{,}6$$

21. B. Si $X = \bar{X} \Rightarrow z = 0$ y $P(z \geq 0) = 0{,}5$; si $x = -1{,}5 \Rightarrow z = -1{,}5/2 = -0{,}75$ y $P(z \geq -0{,}75) = 0{,}2226$

$$z = \frac{X-\bar{X}}{S_x} = \frac{x}{S_x}\ ;\ \frac{-1{,}5}{2} = -0{,}75$$

Para z = -0,75, P(z ≤ -0,75) = 0,2266. Luego entre la media y *x* = -1,5 tendremos una proporción de observaciones: 0,5 - 0,2266 = 0,2734.

KIBBUTZ

22. C. El error típico de medida de la proporción de estudiantes de la población que sufre dolores de cabeza en época de estudios es:

$$\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0{,}08 \cdot 0{,}92}{50}} = 0{,}038$$

23. A. Se trata de una variable de Bernoulli con dos posibilidades, éxito (síntomas de ansiedad) y fracaso (no síntomas de ansiedad). El tamaño muestral es n= 100, grande ya que:

(nπ) = 100 x 0,25 = 25 ≥ 5, y (n(1 - π)) = 100 x 0,75 = 75 ≥ 5

La distribución de la proporción se aproximará a la Normal con parámetros:

$$\Pi = 0{,}25 \;\; y \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0{,}25(0{,}75)}{100}} = 0{,}433;\ N\ (0{,}25;\ 0{,}433)$$

$$Z = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0{,}10-0{,}25}{0{,}433} = -0{,}346;\ P(Z \leq -0{,}346) = 0{,}0003,$$ que corresponde a un 0,03%

24. A. Página 459

25. A. n = 25, dado que la distribución muestral de la media sigue una distribución t de Student;

$$E_{max} = \left| t_{n-1;\frac{\alpha}{2}} \right| \frac{S_{n-1}}{\sqrt{n}} = 2{,}064 \cdot \frac{9}{5} = 3{,}72$$

$$L_{inf} = \bar{X} - E_{max} = 86{,}28;\ L_{sup} = \bar{X} + E_{max} = 93{,}72$$

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
SEPTIEMBRE 2018 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO D DURACIÓN: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

Calificación = (0,4 x Aciertos) - (0,2 x Errores)

Tabla 1. Distribución de la variable edad (X) de 250 participantes en un experimento sobre atención.

X	n_i
29,5 – 32,5	41
26,5 – 29,5	57
23,5 – 26,5	75
20,5 – 23,5	42
17,5 – 20,5	35

Tabla 2. Datos de un experimento en el que se desea conocer si la memoria de los mayores de 70 años sanos mejora con la relajación mental (entrenada mediante talleres de técnicas de meditación). El número de participantes fue de 300, de los cuales 155 asistieron al taller. La memoria se evaluó a todos los participantes mediante el test Rivermead-IIII.

		Mejora de la memoria	
		Sí	No
Asistencia a Taller de meditación	Sí	145	10
	No	45	100

Tabla 3. Resultados de la aplicación de dos test que miden respectivamente el trastorno dismórfico muscular (X) y el rendimiento deportivo (Y) a un grupo de 14 vigoréxicos.

	X	**Y**	**XY**	**X^2**	**Y^2**
∑	229	966	16381	3907	69322
Medias	16,36	69			
Desviaciones típicas	3,39	13,80			

1. Si en una escala de medida la única relación que se establece es “igualdad o desigualdad” diremos que se trata de una escala: **A) nominal;** B) de razón**;** C) de intervalo.
2. En la Tabla 1, la media y varianza de la distribución son respectivamente: **A) 25,324 y 14,4**; B) 50 y 5,03 C) 150 y 3,79.
3. En la Tabla 1, a la puntuación 24 le corresponde el percentil: A) 45; **B) 36;** C) 58.
4. La asimetría es una medida que se refiere al: **A) grado en que los datos se reparten de forma equilibrada por encima y por debajo de la tendencia central**; B) grado de apuntamiento de los datos; C) grado en que los datos se reparten de forma equilibrada por encima y por debajo de la dispersión.
5. En la Tabla 1, la amplitud intercuartil está comprendida entre: A) 2 y 3; **B) 5 y 6**; C) 4 y 5.
6. En la Tabla 2, el valor del coeficiente de contingencia C para cuantificar el grado de asociación entre asistir al taller de meditación y la mejora de la memoria se encuentra entre: A) 0,26 y 0,30; B) 0,33 y 0,36; **C) 0, 50 y 0,60.**
7. El diagrama de dispersión es un diagrama que utilizamos para representar: A) dos variables cualitativas; B) una variable cualitativa y otra cuantitativa; **C) dos variables cuantitativas.**
8. El coeficiente de correlación entre Burnout y el rendimiento académico tiene un valor de -0,98. A la vista de este valor podemos inferir que: A) el Burnout es causa del bajo rendimiento académico; B) el bajo rendimiento académico genera Burnout**; C) que existe una relación inversa entre el Burnout y el rendimiento académico.**
9. Para los datos de la Tabla 3, el valor de la pendiente de la recta de regresión de Y sobre X es**:**
   A) $3,60$; B) 5,6; C) 2,3.
10. El valor del coeficiente de determinación correspondiente a los datos de la Tabla 3 se encuentra entre los valores: **A) 0,76 y 0,79**; B) 0,86 y 0,89; C) -0,76 y 0,79.

KEY PUBLIC

11. La varianza de las puntuaciones pronosticadas para la Tabla 3 se encuentra entre los valores: A) 15 y 17; B) 125 y 127 **C) 148 y 150.**
12. Con los datos de la Tabla 2, si se elige al azar un participante y observamos que no ha mejorado su memoria, ¿entre qué valores se encuentra la probabilidad de que no haya asistido al taller de meditación? A) 0,1000 y 0,2000; **B) 0,9000 y 0,9100**; C) 0,1800 y 0,1900.
13. Si P(A∩B) = P(A)·P(B) diremos que los sucesos A y B son: A) complementarios; B) imposibles; **C) independientes.**
14. Si la función de probabilidad de una variable X es: f (0) = 0,4; f (1) = 0,6; y f (2) = 0,8. La media de X es: A) 0,3; **B) 2,2**; C) 3,5.
15. La expresión 0 ≤ P(A) ≤ 1 corresponde a una propiedad de la definición de probabilidad: A) clásica; B) estadística; **C) a ambas definiciones.**
16. Un test de inteligencia está compuesto por 10 preguntas de 5 alternativas cada una, siendo correcta sólo una de ellas. Si respondemos al azar, ¿cuál es la probabilidad de acertar 4 preguntas? A) 0,0264; B) 0,9672; **C) 0,0881.**
17. El 3% de los pacientes enfermos de corazón son alérgicos al paracetamol. En una muestra de 150 personas cardíacas que toman paracetamol, ¿cuál es la probabilidad de que la proporción de pacientes en la muestra que sean alérgicos al paracetamol sea mayor de 4%? A) 0,8315; **B) 0,2358**; C) 0,1214.
18. Sabiendo que una v.a. se distribuye según χ^2con 28 grados de libertad, ¿cuál es la probabilidad de que X sea mayor de 37,9159? A) 0,975; B) 0,0025; **C) 0,1000.**
19. La media de los pesos de 500 estudiantes de un colegio es 70 kg y la desviación típica 3 kg. Suponiendo que los pesos se distribuyen normalmente, ¿cuántos estudiantes pesan más de 90 kg? **A) Ninguno ya que 0,05% referido a sujetos carece de sentido;** B) 11; C) 23.
20. Aplicamos una terapia implosiva para perder el miedo a los ratones. Determinar el tamaño muestral necesario sabiendo que la probabilidad de éxito de la terapia es 0,80, el E_{max} admitido es 5,5% y el nivel de confianza 0,95 (se asume población infinita) **A) 204**; B) 306; C) 125.
21. ¿Qué son los límites de un intervalo de confianza? A) Es el nivel de riesgo repartido en las dos colas de la distribución cuando el contraste es bilateral; **B) Son los valores que acotan el rango en el que creemos se encuentra el verdadero valor del parámetro con una confianza determinada**; C) Es el contrario del nivel de riesgo.
22. Para estimar el intervalo confidencial de la media poblacional de una variable X con varianza poblacional teóricamente conocida y población infinita, hemos seleccionado una muestra de 100 personas, cuya media 10. Trabajando con un nivel de confianza del 95% se han obtenido para ese intervalo unos límites de 9,216 y 10,784. ¿Cuál es el valor de la desviación típica de esa variable X en la población? A) 16; **B) 4**; C) 2.
23. En la Comunidad de Madrid en los últimos años ha cambiado drásticamente la pirámide poblacional, de forma que ahora la distribución según las categorías es: jóvenes (de 18 a 35 años) 800000; adultos (36 a 65 años) 950000 y mayores (66 años en adelante) 1435000. Para realizar una encuesta sobre hábitos de vida saludables, con un tamaño muestral de 1200 personas (utilizando afijación proporcional), ¿cuantas personas de cada categoría deberemos seleccionar? (recordar que en muestreo hay que redondear siempre al valor superior): A) 400; 400 y 400; B) 327; 432 y 645; **C) 302; 358 y 541.**
24. El tiempo de espera de los pacientes en las consultas de la Seguridad Social sigue una distribución normal de media desconocida y desviación típica de 25 minutos. Se selecciona una muestra aleatoria de 144 pacientes, y se obtiene que la media del tiempo de espera es de 20 minutos. ¿Entre qué límites se estima que está la media de la población con un nivel de confianza del 95%? **A) (15,92; 24,08)**; B) (14,67; 25,33); C) (16,55; 23,45).
25. A partir de los datos de una muestra de 25 personas con neuroticismo, cuya media es 9 y cuasivarianza 3,84, ¿entre que límites se encuentra la varianza poblacional? ($\alpha = 0{,}05$).
A) (6,55; 2,06); B) (1,78; 8,22); **C) (2,34; 7,43).**

BORRADOR

SOLUCIONES

1. A

2. A

X	n_i	X_i	$n_i X_i$	$n_j X_i^2$
29,5 – 32,5	41	31	1271	39401
26,5 – 29,5	57	28	1596	44688
23,5 – 26,5	75	25	1875	46875
20,5 – 23,5	42	22	924	20328
17,5 – 20,5	35	19	665	12635
	250		6331	163927

$$\bar{X} = \frac{\sum n_i X_i}{n} = \frac{6331}{250} = 25{,}324 \qquad S^2 = \frac{\sum n_j X_i^2}{\sum n_j} - \bar{X}^2 = \frac{163927}{250} - 25{,}324^2 = 14{,}403$$

3. B

$$k = \left[\frac{\frac{(P_k - L_i).n_c}{I} + n_d}{n}\right] \times 100 = \left[\frac{\frac{(24-23{,}5).75}{3} + 77}{250}\right] \times 100 = 35{,}8 \sim 36$$

Por tanto, el percentil pedido es P_{36} (pág. 83)

4. A

5. B

X	n_i	n_a	X_i	$n_i X_i$
29,5 – 32,5	41	250	31	1271
26,5 – 29,5	57	209	28	1596
23,5 – 26,5	75	152	25	1875
20,5 – 23,5	42	77	22	924
17,5 – 20,5	35	35	19	665
	250			6631

$$P_{75} = L_{sup} + \left[\frac{\frac{n*75}{100} - n_d}{n_c}\right].I = 26{,}5 + \left(\frac{187{,}5-152}{57}\right) \text{x } 3 = 28{,}37$$

$$P_{25} = L_{inf} + \left[\frac{\frac{n*25}{100} - n_d}{n_c}\right].I = 20.5 + \left(\frac{62{,}5-35}{42}\right) \text{x } 3 = 22{,}46$$

Por tanto $A_{IQ} = Q_3 - Q_1 = P_{75} - P_{25} = 28{,}37 - 22{,}46 = 5{,}91$

KIBBUTZ

6. C

		Mejora en la memoria		
		Sí	No	
Asistencia al taller de memoria	Sí	145 (98,16)	10 (56,83)	155
	No	45 (91,83)	100 (53,16)	145
		190	110	300

Cálculo de las frecuencias teóricas :

$$n_t = \frac{tatal\ filas\ \times\ total\ coluna}{n}$$

$n_{1,1} = \frac{155\times190}{300} = 98{,}16$ $\quad n_{2,1} = \frac{145\times190}{300} = 91{,}83$

$n_{1,2}\ \frac{155\times110}{300} = 56{,}83$ $\quad n_{2,2}\ \frac{145\times110}{300} = 53{,}16$

:

Cálculo de χ^2

$\chi^2 = \frac{(145-98{,}16)^2}{98{,}16} + \frac{(10-56{,}83)^2}{56{,}83} + \frac{(45-91{,}83)^2}{91{,}83} + \frac{(100-53{,}16)^2}{53{,}16} =$ 22,35 + 38,59 + 23,88 + 41,27 = 126,09 $\cong 126$

Índice de contingencia: $C = \sqrt{\frac{\chi^2}{\chi^2+n}} = \sqrt{\frac{126{,}09}{426{,}09}} = \sqrt{0{,}296} = 0{,}54$

7. C
8. C

9. A

$$b = \frac{n\sum(XY) - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{14\times16381 - 229\times966}{14\times3907 - 229^2} = \frac{229334 - 221214}{54698 - 52441} = 3{,}60$$

10. A

$r_{xy} = b\frac{S_x}{S_y} = 3{,}60\ \times\frac{3{,}39}{13{,}80} = 0{,}88$ Por tanto, $r_{xy}^2 = 0{,}88^2 = 0{,}78$

11. C

$r_{xy}^2 = \frac{S_{y'}^2}{S_y^2}$ Por tanto, $S_{y'}^2 = r_{xy}^2 \times s_y^2 = 0{,}78 \times 190{,}44 = 148{,}54 \sim 149$

La varianza de las puntuaciones pronosticadas Y' es 148,54

También se puede calcular como: $S_{y'}^2 = b_{xy}^2 \times s_x^2 = 3{,}60^2 \times 3{,}39^2 = 148{,}9 \sim 149$

12. B

Se trata de calcular la probabilidad condicionada de no haber mejorado en memoria y no haber asistido al taller de meditación, es decir: P (No mejorar en memoria | No asistir al taller). Representamos por NoM no mejorar en memoria y por noT no haber asistido al taller de meditación. Tendremos:

$= P(NoM\ |NoT) = \frac{\mathrm{P(No\ M\cap NoT)}}{\mathrm{P(no\ T)}} = \frac{100/300}{110/300} = \frac{0{,}3333}{0{,}3667} = 0{,}9089$

13. C

XIBBUTZ

14. B

x	f(x)	x·f(x)
0	0,4	0
1	0,6	0,6
2	0,8	1,6
		2,2

15. C

Se trata de la propiedad de probabilidad enunciado como: 0 ≤ P(A) ≤ 1 que indica que la probabilidad es cuantificable numéricamente con un valor comprendido entre 0 y 1y es válido tanto para las definiciones clásica y estadística ya que corresponde a la definición axiomática.

16.

Es una binomial donde n = 10 p = 1/5 = 0,2, TABLA I del formulario.

P(x=4) = 0,0881

17. B

Dado que n>20 resolvemos por la aproximación a la normal

P=0,03; N=150; P (p>0,04)= 1-P (p≤ 0,04)

$$Z = \frac{0{,}04 - 0{,}03}{\sqrt{\frac{0{,}03(1-0{,}03)}{150}}} = \frac{0{,}01}{0{,}0135} = 0{,}7194$$

P (Z≤ 0,72) = 0,7642; P (p>0,04) = 1-0,7194= 0,2358

18. C

Tabla V P(X> 37,9159)= 1-P (37,9159)=1-0,900= 0,1000

19. A

z=$\frac{X_i - \bar{X}}{S_X} = \frac{90-70}{3} = \frac{10}{3} = 3.33$

Por tanto, P(X> 90)= 1- 0,9995 = 0,0005 Es decir, un 0.05% lo que en términos de individuos equivale a ninguno.

20. A

n.c. = 0,95 → $z_{1-\alpha/2} = z_{0,975}$ = 1,96 (Tabla III)

Se trata de calcular el error máximo para el parámetro proporción siendo n>30

$$N = \frac{Z^2 P(1-P)}{E_{max}^2} = 1{,}96^2 \frac{0{,}80(1-0{,}80)}{0{,}055^2} = 203{,}19 \cong 204$$

21. B

22. B

nc = 0,95⇒$Z_{1-\alpha/2}$ =1,96

XIBBUTN

$Lsup = \bar{X} + \left|Z_{1-\alpha/2}\right| \frac{\sigma}{\sqrt{n}}$; $10{,}784 = 10 + 1{,}96 \frac{\sigma}{\sqrt{100}}$ De donde, $10{,}784 = 10 + 1{,}96 \frac{\sigma}{10}$:

$10(10{,}784) - (10)(10) = 1{,}96\sigma$; $\sigma = \frac{10{,}784-100}{1{,}96} = 4$

De la misma manera se podría calcular empleando el límite inferior.

23. C

El muestreo correcto es aleatorio estratificado.

$n_j = \frac{800000}{3185000} \times 1200 = 301{,}4 \sim 302$; $n_A = \frac{950000}{3185000} \times 1200 = 357{,}9 \sim 358$

$$n_M = \frac{1435000}{3185000} \times 1200 = 540{,}6 \sim 541$$

El tamaño será: Jóvenes 302; adultos 358 y mayores 541

Muestra Total 1201

24. A

Dado que se conoce σ utilizaremos la distribución normal, y siendo el nivel de confianza del 95%, tendremos que Zα/2 = 1,96.

$L_{inf} = \bar{X} - \left|Z_{\frac{\alpha}{2}}\right| \sigma/\sqrt{n} = 20 - 1{,}96\, 25/12 = 15{,}92$

$L_{sup} = \bar{X} + \left|Z_{\frac{\alpha}{2}}\right| \sigma/\sqrt{n} = 20 + 1{,}96\, 25/12 = 24{,}08$

Por tanto el intervalo buscado es (15,92; 24,08)

25. C

Dado que n<100 la distribución sigue una χ^2cuyos gl, para este supuesto son 24.

Por tanto:

$L_{inf} = \frac{(n-1)\times S^2_{n-1}}{\chi^2_{n-1,1-\alpha/2}} = \frac{24\times 3{,}84}{39{,}3641} = \frac{92{,}16}{39{,}3641}$ = 2,34

$L_{sup} = \frac{(n-1)\times S^2_{n-1}}{\chi^2_{n-1,\alpha/2}} = \frac{24\times 3{,}84}{12{,}4012} = \frac{92{,}16}{12{,}4012} = 7{,}23$

Intervalo: (2,34; 7,43)

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2018 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO E DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

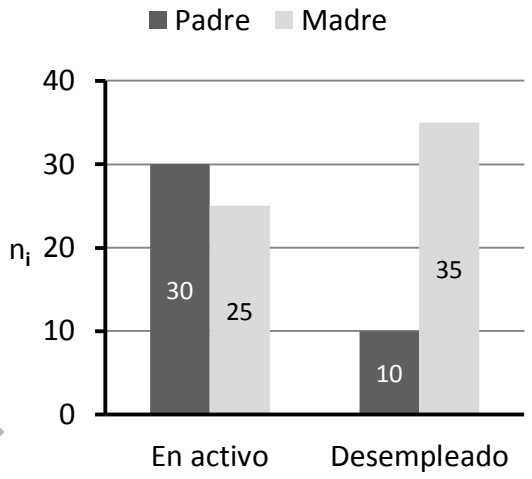
Tabla 1. Resultados obtenidos en un test de creatividad infantil (X) en una clase de 30 alumnos de primaria.

X	Alumnos
20-22	5
17-19	9
14-16	7
11-13	5
8-10	4

Tabla 2. Se realiza un estudio sobre percepción emocional y ajuste psicosocial en 1650 adolescentes. En la siguiente tabla se ofrece información descriptiva de ambas variables y su relación.

	Media	Desviación típica	Correlación
Percepción emocional (X)	15,20	3,94	$r_{XY} = 0{,}34$
Ajuste psicosocial (Y)	18,70	1,6	

Gráfica 1. Información obtenida en un estudio sobre la situación laboral de padres (n = 40) y madres (n = 60) que tienen hijos con autismo.



1. Con la información recogida en la Tabla 1, ¿qué percentil corresponde a un alumno que tiene una puntuación de 18? A) P_{80}; B) P_{90}; C) **P_{68}**.
2. El valor de la varianza de la variable creatividad, cuyos datos se representan en la Tabla 1 se encuentra entre: A) 10 y 12; B) **14 y 16**; C) 18 y 20.
3. Siguiendo con los datos de la Tabla 1, la distribución es: A) **asimétrica negativa**; B) simétrica; C) asimétrica positiva.
4. En la Estadística Inferencial, el *parámetro* es una propiedad descriptiva de: A) una muestra; B) **una población**; C) un estadístico.
5. Para los datos recogidos en la Tabla 1 la representación gráfica más habitual es un: A) **histograma**; B) diagrama de dispersión; C) diagrama de sectores.
6. Con los datos de la Tabla 1, la media en creatividad infantil para los alumnos de la clase es: A) 11,5; B) 16,2; C) **15,6.**
7. Según los datos de la Gráfica 1, la distribución de la situación laboral condicionada a ser madre es: A) 75% y 41,75%; B) **41,67% y 58,3%;** C) 25% y 58,3%.
8. Con los datos de la Gráfica 1, el valor de la asociación entre las variables, evaluada a través de χ^2, está comprendido entre: A) **9 y 12**; B) 13 y 16; C) 18 y 20.
9. Para que exista independencia entre dos variables cualitativas el índice χ^2 debe ser: A) mayor que 0; B) **igual a 0**; C) χ^2 no evalúa si existe o no independencia.
10. En el estudio sobre percepción emocional y ajuste psicosocial en adolescentes, cuya información aparece en la Tabla 2, la covarianza entre ambas variables es: A) 1,24; B) **2,14**; C) 4,12.

EBBUN

11. Con los datos de la Tabla 2, la ordenada en el origen y la pendiente de la ecuación de la recta de regresión para pronosticar el ajuste psicosocial (Y) a partir de la percepción emocional (X) es: A) 12,47 y 3,25; B) 10,33 y 3,94; C) **16,57 y 0,14.**
12. Con la recta de regresión obtenida en la pregunta anterior, ¿qué puntuación en ajuste psicosocial pronosticaremos a un adolescente que ha obtenido una puntuación de 12 en percepción emocional?: A) 16,27; B) 18,60; C) **18,25**.
13. En una escuela infantil 1/3 de los niños de 5 años está aprendiendo a leer con un método de lectura global (G) y el resto de niños con lectura fonológica y silábica (FS). Tras la aplicación de los dos métodos a final de curso, presenta una mejoría (M) significativa en la lectura el 70% de los niños que han recibido metodología (G) y el 85% de los niños con metodología (FS). Se elige un niño al azar y se observa que ha presentado mejoría (M), ¿cuál es la probabilidad de que haya recibido el método de lectura global (G)? A) **0,29**; B) 0,33; C) 0,70.
14. Se ha diseñado una nueva prueba para diagnosticar a personas con sordera moderada. Para ello, se ha pasado la prueba a 150 personas ya diagnosticadas previamente con ese tipo de sordera. Al aplicar la nueva prueba 25 personas han dado negativo en el diagnóstico. ¿Cuál es la sensibilidad de la prueba? A) **0,8333**; B) 0,7990; C) 0,8990.
15. Indica qué afirmación NO es correcta sobre el teorema del producto: A) se aplica a situaciones en las que se quiere calcular la probabilidad de que aparezcan dos sucesos de forma simultánea; B) sirve para calcular la probabilidad de intersección entre dos sucesos; C) **sirve para calcular la probabilidad de unión entre dos sucesos.**
16. La función de probabilidad de una variable *X* es: *f* (1) = 0,30; *f* (2) = 0,45 y *f* (3) = 0,25. ¿Entre qué valores se encuentra la varianza? A) 2 y 1,58; B) 2,50 y 1; C) **0,80 y 0,30.**
17. Se sabe que el 15% de la población de alumnos de un instituto tiene dificultades de aprendizaje. Si elegimos al azar una muestra de 10 estudiantes del centro, ¿cuál es la probabilidad de que tres estudiantes tengan dificultades de aprendizaje? A) 0,9500**;** B) **0,1298**; C) 0,5443.
18. Sabiendo que las puntuaciones en el test de percepción emocional de la Tabla 2 se distribuyen normalmente, ¿cuál es la proporción de sujetos con puntuaciones entre 13 y 17? A) **0,3895**; B) 0,6772; C) 0,2877.
19. Sea X una variable que se distribuye según una distribución *t* con 12 grados de libertad. ¿Cuál es la probabilidad de obtener valores menores o iguales a 2,681? A) 0,950; B) **0,990**; C) 0,995.
20. ¿Qué tipo de muestreo considera que todas las muestras posibles son equiprobables? A) Por cuotas; B) Intencional; C) **Aleatorio simple.**
21. En una población formada por los números del 1 al 10, ¿qué probabilidad hay de obtener una muestra de tamaño 3 en un muestreo aleatorio con reposición? A) **0,001**; B) 0,010; C) 0,030.
22. Sea X una variable con distribución normal en la población con N(μ; 4). Se extraen muestras de tamaño 15, ¿entre que valores de χ^2 se encuentra el 80% central de las varianzas muestrales? A) **(7,7895 y 21,0641)**; B) (5,6287 y 21,0641); C) (4,6604 y 23,6848).
23. La variable autoestima en la población de adolescentes se distribuye normalmente, con una desviación típica de 3,5. Se quiere estimar la media en la población y para ello se ha seleccionado mediante m. a. s. una muestra de 300 adolescentes, obteniéndose una media para la muestra de 18. Trabajando con un nivel de confianza del 95%, ¿cuáles son los límites del intervalo de confianza? A) (15,25 y 19,46); B) **(17,60 y 18,40)**; C) (17,33 y 19,14).
24. Un investigador trabaja con un nivel de confianza del 95%. Dada una variable X, se conocen los límites inferior y superior (5,70 y 8,30) del intervalo confidencial de la media poblacional. Se ha seleccionado una muestra de 130 personas, obteniéndose una media muestral de 7. ¿Cuál es la desviación típica de la variable X en la población? A**) 7,56;** B) 8,03; C) 7,05.
25. Se quiere estimar con un nivel de confianza del 99% y con un error máximo de estimación de 0,1 la proporción de niños de Educación Secundaria que tienen ordenador en casa. Estudios previos señalan que esa proporción es del 0,3. ¿Cuál es el tamaño de la muestra necesario para realizar dicha estimación asumiendo una población infinita? A) 100; B) 120; C) **140.**

KIBUCITY

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2018 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

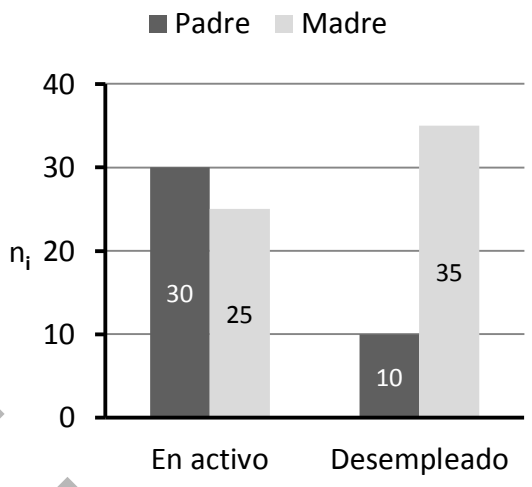
Tabla 1. Resultados obtenidos en un test de creatividad infantil (X) en una clase de 30 alumnos de primaria.

X	Alumnos
20-22	5
17-19	9
14-16	7
11-13	5
8-10	4

Tabla 2. Se realiza un estudio sobre percepción emocional y ajuste psicosocial en 1650 adolescentes. En la siguiente tabla se ofrece información descriptiva de ambas variables y su relación.

	Media	Desviación típica	Correlación
Percepción emocional (X)	15,20	3,94	$r_{XY} = 0{,}34$
Ajuste psicosocial (Y)	18,70	1,6	

Gráfica 1. Información obtenida en un estudio sobre la situación laboral de padres (n = 40) y madres (n = 60) que tienen hijos con autismo.



1. En la Estadística Inferencial, el *parámetro* es una propiedad descriptiva de: A) una muestra; B) **una población**; C) un estadístico.
2. Para los datos recogidos en la Tabla 1 la representación gráfica más habitual es un: A) **histograma**; B) diagrama de dispersión; C) diagrama de sectores.
3. Con los datos de la Tabla 1, la media en creatividad infantil para los alumnos de la clase es: A) 11,5; B) 16,2; C) **15,6.**
4. Siguiendo con la información recogida en la Tabla 1, ¿qué percentil corresponde a un alumno que tiene una puntuación de 18? A) P_{80}; B) P_{90}; C) $\mathbf{P_{68}}$.
5. El valor de la varianza de la variable creatividad, cuyos datos se representan en la Tabla 1 se encuentra entre: A) 10 y 12; B) **14 y 16**; C) 18 y 20.
6. Siguiendo con los datos de la Tabla 1, la distribución es: A) **asimétrica negativa**; B) simétrica; C) asimétrica positiva.
7. Según los datos de la Gráfica 1, la distribución de la situación laboral condicionada a ser madre es: A) 75% y 41,75%; B) **41,67% y 58,3%;** C) 25% y 58,3%.
8. Con los datos de la Gráfica 1, el valor de la asociación entre las variables, evaluada a través de χ^2, está comprendido entre: A) **9 y 12**; B) 13 y 16; C) 18 y 20.
9. Para que exista independencia entre dos variables cualitativas el índice χ^2 debe ser: A) mayor que 0; B) **igual a 0**; C) χ^2 no evalúa si existe o no independencia.
10. En el estudio sobre percepción emocional y ajuste psicosocial en adolescentes, cuya información aparece en la Tabla 2, la covarianza entre ambas variables es: A) 1,24; B) **2,14**; C) 4,12.

WEBBCIN

11. Con los datos de la Tabla 2, la ordenada en el origen y la pendiente de la ecuación de la recta de regresión para pronosticar el ajuste psicosocial (Y) a partir de la percepción emocional (X) es: A) 12,47 y 3,25; B) 10,33 y 3,94; C) **16,57 y 0,14.**
12. Con la recta de regresión obtenida en la pregunta anterior, ¿qué puntuación en ajuste psicosocial pronosticaremos a un adolescente que ha obtenido una puntuación de 12 en percepción emocional?: A) 16,27; B) 18,60; C) **18,25**.
13. Indica qué afirmación NO es correcta sobre el teorema del producto: A) se aplica a situaciones en las que se quiere calcular la probabilidad de que aparezcan dos sucesos de forma simultánea; B) sirve para calcular la probabilidad de intersección entre dos sucesos; C) **sirve para calcular la probabilidad de unión entre dos sucesos.**
14. En una escuela infantil 1/3 de los niños de 5 años está aprendiendo a leer con un método de lectura global (G) y el resto de niños con lectura fonológica y silábica (FS). Tras la aplicación de los dos métodos a final de curso, presenta una mejoría (M) significativa en la lectura el 70% de los niños que han recibido metodología (G) y el 85% de los niños con metodología (FS). Se elige un niño al azar y se observa que ha presentado mejoría (M), ¿cuál es la probabilidad de que haya recibido el método de lectura global (G)? A) **0,29**; B) 0,33; C) 0,70.
15. Se ha diseñado una nueva prueba para diagnosticar a personas con sordera moderada. Para ello, se ha pasado la prueba a 150 personas ya diagnosticadas previamente con ese tipo de sordera. Al aplicar la nueva prueba 25 personas han dado negativo en el diagnóstico. ¿Cuál es la sensibilidad de la prueba? A) **0,8333**; B) 0,7990; C) 0,8990.
16. La función de probabilidad de una variable *X* es: *f* (1) = 0,30; *f* (2) = 0,45 y *f* (3) = 0,25. ¿Entre qué valores se encuentra la varianza? A) 2 y 1,58; B) 2,50 y 1; C) **0,80 y 0,30.**
17. Se sabe que el 15% de la población de alumnos de un instituto tiene dificultades de aprendizaje. Si elegimos al azar una muestra de 10 estudiantes del centro, ¿cuál es la probabilidad de que tres estudiantes tengan dificultades de aprendizaje? A) 0,9500**;** B) **0,1298**; C) 0,5443.
18. Sabiendo que las puntuaciones en el test de percepción emocional de la Tabla 2 se distribuyen normalmente, ¿cuál es la proporción de sujetos con puntuaciones entre 13 y 17? A) **0,3895**; B) 0,6772; C) 0,2877.
19. Sea X una variable que se distribuye según una distribución *t* con 12 grados de libertad. ¿Cuál es la probabilidad de obtener valores menores o iguales a 2,681? A) 0,950; B) **0,990**; C) 0,995.
20. ¿Qué tipo de muestreo considera que todas las muestras posibles son equiprobables? A) Por cuotas; B) Intencional; C) **Aleatorio simple.**
21. En una población formada por los números del 1 al 10, ¿qué probabilidad hay de obtener una muestra de tamaño 3 en un muestreo aleatorio con reposición? A) **0,001**; B) 0,010; C) 0,030.
22. Sea X una variable con distribución normal en la población con N(μ; 4). Se extraen muestras de tamaño 15, ¿entre que valores de χ^2 se encuentra el 80% central de las varianzas muestrales? A) **(7,7895 y 21,0641)**; B) (5,6287 y 21,0641); C) (4,6604 y 23,6848).
23. La variable autoestima en la población de adolescentes se distribuye normalmente, con una desviación típica de 3,5. Se quiere estimar la media en la población y para ello se ha seleccionado mediante m. a. s. una muestra de 300 adolescentes, obteniéndose una media para la muestra de 18. Trabajando con un nivel de confianza del 95%, ¿cuáles son los límites del intervalo de confianza? A) (15,25 y 19,46); B) **(17,60 y 18,40)**; C) (17,33 y 19,14).
24. Un investigador trabaja con un nivel de confianza del 95%. Dada una variable X, se conocen los límites inferior y superior (5,70 y 8,30) del intervalo confidencial de la media poblacional. Se ha seleccionado una muestra de 130 personas, obteniéndose una media muestral de 7. ¿Cuál es la desviación típica de la variable X en la población? A**) 7,56;** B) 8,03; C) 7,05.
25. Se quiere estimar con un nivel de confianza del 99% y con un error máximo de estimación de 0,1 la proporción de niños de Educación Secundaria que tienen ordenador en casa. Estudios previos señalan que esa proporción es del 0,3. ¿Cuál es el tamaño de la muestra necesario para realizar dicha estimación asumiendo una población infinita? A) 100; B) 120; C) **140.**

EXAMEN

SOLUCIONES:

1. B
2. A
3. C

X	X_i	n_i	$n_i\ X_i$
20-22	21	5	105
17-19	18	9	162
14-16	15	7	105
11-13	12	5	60
8-10	9	4	36
$\sum$		30	468

$$\overline{X} = \frac{\sum n_i X_i}{n} = \frac{468}{30} = 15{,}6$$

4. C

X	X_i	n_i	n_a
20-22	21	5	30
17-19	18	9	25
14-16	15	7	16
11-13	12	5	9
8-10	9	4	4
$\sum$		30	

La puntuación X = 18 está en el intervalo [17-19] que va a ser, por tanto, el intervalo crítico.
Por lo tanto, P_k = 18; L_i = 16,5; n_c = 9, I = 3 y n_d = 16. Se aplica la fórmula y se obtiene lo siguiente:

$$k = \left[\frac{\frac{(P_k - L_i)\cdot n_c}{I} + n_d}{n}\right]\cdot 100 = \left[\frac{\frac{(18-16{,}5)\times 9}{3} + 16}{30}\right]\times 100 = \left[\frac{20{,}5}{30}\right]\times 100 =$$

$$= 0{,}683\times 100 = 68{,}33 \approx 68$$

5. B

De la pregunta 3 sabemos que la media es 15,6

X	X_i	n_i	$(X_i - \overline{X})$	$(X_i - \overline{X})^2$	$n_i(X_i - \overline{X})^2$
20-22	21	5	5,4	29,16	145,80
17-19	18	9	2,4	5,76	51,84
14-16	15	7	-0,6	0,36	2,52
11-13	12	5	-3,6	12,96	64,80
8-10	9	4	-6,6	43,56	174,24
$\sum$		30			439,20

KIBBUTZ

$$S_X^2 = \frac{\sum n_i (X_i - \overline{X})^2}{n} = \frac{439,2}{30} = 14,64$$

6. A

Conocemos los siguientes datos:

$\overline{X} = 15,6$; $Mo = 18$, ya que es el punto medio del intervalo con mayor frecuencia absoluta [17-19]

$$S_X = \sqrt{14,64} = 3,826$$

$$A_P = \frac{\overline{X} - Mo}{S_X} = \frac{15,6 - 18}{3,826} = -0,627$$

Como A_p es negativo, la distribución presenta una asimetría negativa.

7. B

Se trata de una distribución de Y (situación laboral) condicionada a uno de los valores de X (madre). Por tanto, se considera únicamente la fila en la que aparecen las frecuencias de las madres. Para calcular estos porcentajes, cada frecuencia de la fila de las madres se divide por el total de madres (60) y se multiplica por 100.

	En activo	Desempleado	
Padre	30	10	40
Madre	25 (41,67%)	35 (58,33%)	60
	55	45	Total 100

8. A

Primero hay que calcular las frecuencias teóricas (entre paréntesis en la tabla):

$$n_{t11} = \frac{40 \times 55}{100} = 22 \qquad n_{t12} = \frac{40 \times 45}{100} = 18$$

$$n_{t21} = \frac{60 \times 55}{100} = 33 \qquad n_{t22} = \frac{60 \times 45}{100} = 27$$

	Empleado	Desempleado	
Padre	30 (22)	10 (18)	40
Madre	25 (33)	35 (27)	60
	55	45	Total 100

$$\chi^2 = \frac{(30-22)^2}{22} + \frac{(10-18)^2}{18} + \frac{(25-33)^2}{33} + \frac{(35-27)^2}{27} = 2,91 + 3,56 + 1,94 + 2,37 = 10,78$$

9. B

KIBBUTZ

10. B

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} \rightarrow r_{XY} \times S_X S_Y = S_{XY} \rightarrow S_{XY} = 0{,}34 \times 3{,}94 \times 1{,}60 = 2{,}14$$

11. C

$$b = r_{XY} \frac{S_Y}{S_X} = 0{,}34 \times \frac{1{,}60}{3{,}94} = 0{,}1380 \approx 0{,}14$$

$$a = \overline{Y} - b\overline{X} = 18{,}70 - 0{,}14 \times 15{,}20 = 16{,}572 \approx 16{,}57$$

12. C

$$Y_i' = a + bX_i = 16{,}57 + 0{,}14 \times 12 = 16{,}57 + 1{,}68 = 18{,}25$$

13. C (Página 269)

14. A

La probabilidad de que el niño escogido que presenta mejoría en la lectura (*M)*, haya recibido la intervención en la lectura con el método (G) se obtiene mediante el teorema de Bayes.

Lo que se pide es:

$$P(G|M) = \frac{P(G) \cdot P(M|G)}{P(M)}$$

$$P(G) = 0{,}33$$

$$P(FS) = 0{,}67$$

$$P(M|G) = 0{,}70$$

$$P(M|FS) = 0{,}85$$

$$P(G/M) = \frac{P(G) \cdot P(M|G)}{P(G) \cdot P(M|G) + P(FS) \cdot P(M/FS)} = \frac{0{,}33 \times 0{,}70}{0{,}33 \times 0{,}70 + 0{,}67 \times 0{,}85} =$$

$$= \frac{0{,}231}{0{,}8} = 0{,}29$$

15. A

Sensibilidad = $P(+|T)$ Es decir, discrimina a los verdaderos positivos. De las 150 personas con sordera moderada ha detectado a 125. Por tanto:

$$P(+|T) = \frac{125}{150} = 0{,}8333$$

16. A

x	f(x)	$x \cdot f(x)$	$(x-\mu)$	$(x-\mu)^2$	$(x-\mu)^2 \cdot f(x)$
1	0,30	0,30	-0,95	0,9025	0,2708
2	0,45	0,90	0,05	0,0025	0,0011
3	0,25	0,75	1,05	1,1025	0,2756
		1,95			0,5475

$$\sigma^2 = \sum (x-\mu)^2 \cdot f(x) = 0{,}5475 \approx 0{,}55$$

$$\mu = \sum x \cdot f(x) = 1{,}95$$

KIBBUTZ

17. B

Debemos calcular $P(X=3)$ con *n* = 10 y *p* = 0,15; x = 3

Se resuelve por la función de probabilidad binomial, por lo que el valor se localiza en la Tabla I → 0,1298

18. A

Se pide:

$P(13 \leq X \leq 17) = P(X \leq 17) - P(X \leq 13)$

Para calcular estas proporciones, transformamos las puntuaciones directas a puntuaciones típicas y buscamos en las tablas de la distribución normal (Tablas III y IV).

$$z = \frac{X - \overline{X}}{S_X} = \frac{13 - 15{,}20}{3{,}94} = -0{,}558 \approx -0{,}56\,;\ P(z \leq -0{,}56) = 0{,}2877$$

$$z = \frac{X - \overline{X}}{S_X} = \frac{17 - 15{,}20}{3{,}94} = 0{,}4568 \approx 0{,}46\,;\ P(z \leq 0{,}46) = 0{,}6772$$

$$P(13 \leq X \leq 17) = P(X \leq 17) - P(X \leq 13) = 0{,}6772 - 0{,}2877 = 0{,}3895$$

19. B $t_{12} \leq 2{,}681) = 0{,}99$ (ver Tabla VI)

20. C (Página 379)

21. A

La probabilidad de una muestra concreta de tamaño 3 es: $\frac{1}{N^n} = \frac{1}{10^3} = 0{,}001$

22. A

Se trata de la distribución muestral de la varianza. Sabemos que dicha distribución es χ^2_{n-1}

$P(\chi^2 \leq \chi^2_{n-1}) = 0{,}10$

Acudiendo a la Tabla V con 14 grados de libertad, los valores de χ^2 correspondientes son:

$\chi^2_{14;0,10} = 7{,}7895$

$X^2_{14;0,90} = 21{,}0641$

23. B

Variable con distribución normal y varianza poblacional conocida, por tanto, la distribución muestral de la media es: $N\left(\mu;\ {}^{\sigma}/_{\sqrt{n}}\right)$

El intervalo confidencial viene definido por:

$$\boxed{P\left(\overline{X} - \left|Z_{\alpha/2}\right| {}^{\sigma}/_{\sqrt{n}} \leq \mu \leq \overline{X} + \left|Z_{\alpha/2}\right| {}^{\sigma}/_{\sqrt{n}}\right) = 1 - \alpha}$$

Siendo los límites:

$L_{inf} = \overline{X} - \left|Z_{\alpha/2}\right| {}^{\sigma}/_{\sqrt{n}}$	$L_{sup} = \overline{X} + \left|Z_{\alpha/2}\right| {}^{\sigma}/_{\sqrt{n}}$

Sustituyendo:

n.c. = $1 - \alpha$ = 0,95 → $\alpha = 0{,}05$ → $\left|{}^{\alpha}/_{2}\right| = |1{,}96|$ (Tabla III y IV)

KIBBUTZ

$$L_{inf} = 18 - 1{,}96\frac{3{,}5}{\sqrt{300}} = 18 - 0{,}40 = 17{,}60$$
$$L_{sup} = 18 + 1{,}96\frac{3{,}5}{\sqrt{300}} = 18 + 0{,}40 = 18{,}40$$

Los límites buscados son: (17,60; 18,40)

24. A

n.c. $= 0{,}95 \rightarrow Z_{\alpha/2} = Z_{0{,}975} = 1{,}96$ (Tabla III y IV)

$$L_{sup} = \bar{X} + \left|Z_{\alpha/2}\right| \sigma/\sqrt{n} \rightarrow 8{,}30 = 7 + 1{,}96 \times \sigma/\sqrt{130} \rightarrow 1{,}30 = 1{,}96 \times \sigma/11{,}40$$

$$\sigma = \frac{1{,}30 \times 11{,}40}{1{,}96} = \frac{14{,}82}{1{,}96} = 7{,}56$$

o, utilizando el límite inferior:

$$L_{inf} = \bar{X} - \left|Z_{\alpha/2}\right| \sigma/\sqrt{n} \rightarrow 5{,}70 = 7 - 1{,}96 \times \sigma/\sqrt{130} \rightarrow -1{,}30 = -1{,}96 \times \sigma/11{,}40$$

$$\sigma = \frac{1{,}30 \times 11{,}40}{1{,}96} = \frac{14{,}82}{1{,}96} = 7{,}56$$

25. C

Sabemos que p = 0,3, por lo que: (1- p) = 0,7.

$$n = \frac{Z^2_{\alpha/2} P(1-P)}{E^2_{max}} = \frac{2{,}58^2 \times 0{,}3 \times 0{,}7}{0{,}1^2} = \frac{1{,}398}{0{,}01} = 139{,}8 \approx 140$$

KIBBUTZ

Código asignatura	Nombre asignatura
62011037	**INTRODUCCIÓN AL ANÁLISIS DE DATOS**
Fecha alta y origen	Convocatoria
26/03/2020	**SEPTIEMBRE 2017**
Curso virtual	

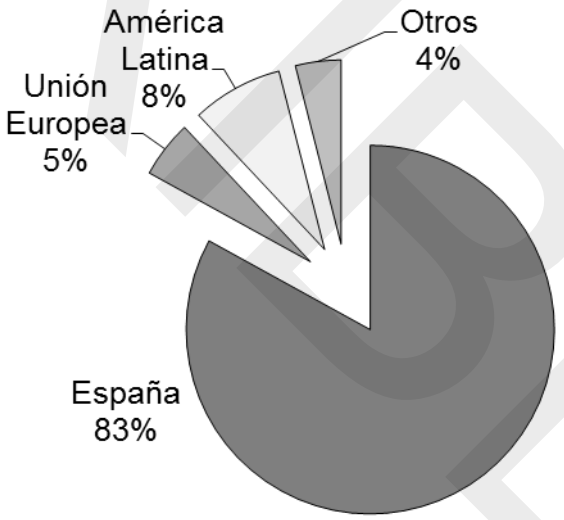
KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
SEPTIEMBRE 2017 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Material: **Adenda (Formulario y Tablas) y calculadora programable**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

SOLUCIONES



Gráfica 1. Lugar de nacimiento de los 2400 alumnos matriculados en una asignatura en la U.N.E.D.

CI	n_i
50-69	5
70-89	280
90-109	500
110-129	205
130-149	10
	1000

Tabla 1. Distribución de frecuencias de la variable Cociente Intelectual (CI) para un grupo de 1000 personas.

Tabla 2. Para dos variables cuantitativas, X e Y, disponemos de los siguientes datos:

X	Y	XY
$\overline{X}=8$	$\overline{Y}=100$	$Cov(X,Y)=S_{XY}=18$
$S_X=2$	$S_Y=10$	

Situación 1. El 80% de alumnos de Psicología escoge la especialidad de Psicología clínica en la carrera y de éstos el 50 % realiza el Máster de la especialidad. De los alumnos que no escogen la especialidad de clínica sólo el 20% realiza el Máster de la especialidad.

1. La variable "lugar de nacimiento" representada en la Gráfica 1 es: A) dicotómica; B) politómica ; C) discreta
2. El nivel de medida de la variable CI, en la Tabla 1, es: A) de intervalo ; B) ordinal; C) de razón
3. La Gráfica 1 se denomina diagrama de: A) barras; B) dispersión; C) sectores
4. Teniendo en cuenta la Gráfica 1, el número de alumnos nacidos en España es: A) 83; B) 830; C) 1992
5. Para los datos representados en la Gráfica 1 utilizaremos como índice de tendencia central: A) el porcentaje; B) la moda; C) la media
6. Con los datos de la Tabla 1, la media en CI es igual a: A) 78,2; B) 98,20; C) 108,2

7. Con los datos de la Tabla 1, la mediana es igual a: A) 98,10; B) 99,50; C) 102,30
8. Con los datos de la Tabla 1, el tercer cuartil es igual a: A) 98,20; B) 102,80; C) 108,10
9. Con los datos de la Tabla 1, la desviación típica es igual a: A) 7,41; B) 14,71; C) 216,31
10. Con los datos de la Tabla 1, el índice de asimetría de Pearson es igual a: A) -0,088; B) -8,8; C) 0,88
11. Para calcular el grado de asociación entre dos variables cualitativas utilizamos: A) la covarianza; B) el coeficiente de correlación de Pearson; C) el coeficiente de contingencia C
12. Con los datos de la Tabla 2, el coeficiente de correlación de Pearson es igual a: A) 0,6; B) 0,8; C) 0,9
13. Teniendo en cuenta la Tabla 2, la recta de regresión que permite pronosticar la variable *Y* en función de la *variable X* es: A) $Y' = 4{,}5X + 64$; B) $Y' = -0{,}5X + 70$; C) $Y' = 6X + 35$
14. Sabemos que, para dos sucesos A y B, $P(A|B) = P(A)$ y $P(B|A) = P(B)$. Los sucesos A y B son: A) complementarios**;** B) dependientes; C) independientes
15. Teniendo en cuenta la Situación 1, si elegimos un alumno al azar, ¿cuál es la probabilidad de que realice el Máster?: A) 0,44; B) 0,70; C) 0,80
16. Teniendo en cuenta la Situación 1, elegimos un alumno al azar y resulta que realiza el Máster, ¿cuál es la probabilidad de que en la carrera hiciera la especialidad de clínica?: A) 0,50; B) 0,91; C) 0,80
17. Una variable aleatoria discreta X toma los valores 0, 1 y 2. Sabemos que P(X=0)=0,64, P(X=1)=0,32 y P(X=2)=0,04 ¿cuánto vale la media o esperanza matemática de X?: A) 0,2; B) 0,4; C) 0,6
18. Teniendo en cuenta la situación 1, si elegimos al azar 2 alumnos, ¿cuál es la probabilidad de que ninguno estudie psicología clínica?: A) 0,04; B) 0,44; C) 0,64
19. Teniendo en cuenta la situación 1, si elegimos al azar 2 alumnos, ¿cuál es la probabilidad de que al menos uno estudie psicología clínica? : A) 0,96; B) 0,86; C) 0,46
20. Una variable X se distribuye normalmente con media 100 y desviación típica 10, ¿qué porcentaje de sujetos obtiene puntuaciones menores o iguales que 110?: A) 46,02% ; B) 64,30 %; C) 84,13%
21. En una distribución F de Snedecor con 10 grados de libertad en el numerador y 30 en el denominador, el percentil 95 es: A) 1,819; B) 2,165; C) 2,700
22. En una distribución chi-cuadrado con 30 grados de libertad, el valor 40,256 corresponde al percentil: A) 90; B) 9; C) 10
23. ¿Qué tamaño mínimo de la muestra debemos utilizar para estimar la media poblacional si la varianza poblacional es 100, el error de estimación máximo es 2 y el nivel de confianza es 0,95?: A) 69; B) 96; C) 126
24. En una muestra de 100 reclusos penitenciarios, hemos calculado la proporción de reclusos con problemas de *autoestima*. Los límites del intervalo de confianza para la proporción de reclusos con problemas de *autoestima* fueron 0,7 y 0,8. Podemos concluir que la proporción de reclusos penitenciarios con problemas de *autoestima* está entre 0,7 y 0,8: A) con probabilidad 1; B) con probabilidad $1-\alpha/2$; C) con una probabilidad igual al nivel de confianza utilizado
25. En un estudio donde la media de la muestra es 35, los límites del intervalo de confianza para un nivel de confianza de 0,95, son 30 y 40. ¿Cuánto vale la desviación típica de la distribución muestral de la media $\sigma_{\bar{X}}$?: A) 0,392; B) 2,55; C) 9,8

SOLUCIONES:

1. B
2. A
3. C
4. C

   2400·0,83=1992
5. B
6. B

X=CI	X_i	n_i	$n_i X_i$
50-69	59,5	5	297,50
70-89	79,5	280	22260,00
90-109	99,5	500	49750,00
110-129	119,5	205	24497,50
130-149	139,5	10	1395,00
		1000	98200,00

$$\overline{X} = \frac{98200}{1000} = 98{,}2$$

7. A

X	n_i	n_a
130-149	10	1000
110-129	205	990
90-109	500	785
70-89	280	285
50-69	5	5
	1000	

$$Md = 89{,}5 + \left(\frac{\frac{1000}{2} - 285}{500} \right) \cdot 20 = 89{,}5 + 8{,}6 = 98{,}10$$

8. C

X	n_i	n_a
130-149	10	1000
110-129	205	990
90-109	500	785
70-89	280	285
50-69	5	5
	1000	

KEBUTZ

$$Q_3 = P_{75} = L_i + \left(\frac{\frac{n \cdot k}{100} - n_d}{n_c} \right) \cdot I = 89{,}5 + \left(\frac{\frac{1000 \cdot 75}{100} - 285}{500} \right) \cdot 20 = 89{,}5 + 18{,}6 = 108{,}1$$

9. B

X=CI	X_i	n_i	$n_i X_i$	$n_i X_i^2$
50-69	59,5	5	297,50	17701,250
70-89	79,5	280	22260,00	1769670,000
90-109	99,5	500	49750,00	4950125,000
110-129	119,5	205	24497,50	2927451,250
130-149	139,5	10	1395,00	194602,500
		1000	98200,00	9859550,000

$$S_X^2 = \frac{9859550}{1000} - 98{,}2 = 9859 - 9643{,}24 = 216{,}31 \Rightarrow S_X = \sqrt{216{,}31} \cong 14{,}71$$

10. A

$$A_s = \frac{\overline{X} - Mo}{S_x} = \frac{98{,}2 - 99{,}5}{14{,}71} = -0{,}088$$

11. C

12. C

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{18}{2 \cdot 10} = \frac{18}{20} = 0{,}9$$

13. A

$$Y' = bX + a$$

$$b = r_{XY} \frac{S_Y}{S_X} = 0{,}9 \frac{10}{2} = 4{,}5 \qquad a = \overline{Y} - b\overline{X} = 100 - (4{,}5 \cdot 8) = 64$$

$$Y' = 4{,}5X + 64$$

14. C

15. A

$$\mathrm{C} = \text{"clínica"} \quad \overline{\mathrm{C}} = \text{"no clínica"}$$

$$\mathrm{P}(\mathrm{C}) = 0{,}80 \quad \mathrm{P}(\overline{\mathrm{C}}) = 1 - \mathrm{P}(\mathrm{C}) = 1 - 0{,}80 = 0{,}2$$

$$\mathrm{P}(\mathrm{M}|\mathrm{C}) = 0{,}50$$

$$\mathrm{P}(\mathrm{M}|\overline{\mathrm{C}}) = 0{,}20$$

$$\mathrm{P}(\mathrm{M}) = \mathrm{P}(\mathrm{M} \cap \mathrm{C}) + \mathrm{P}(\mathrm{M} \cap \overline{\mathrm{C}}) =$$

$$= \mathrm{P}(\mathrm{C}) \cdot \mathrm{P}(\mathrm{M}|\mathrm{C}) + \mathrm{P}(\overline{\mathrm{C}}) \cdot \mathrm{P}(\mathrm{M}|\overline{\mathrm{C}}) = 0{,}8 \cdot 0{,}50 + 0{,}20 \cdot 0{,}20 = 0{,}40 + 0{,}04 = 0{,}44$$

16. B

$$\mathrm{P}(\mathrm{C}|\mathrm{M}) = \frac{\mathrm{P}(\mathrm{C} \cap \mathrm{M})}{\mathrm{P}(\mathrm{M})} = \frac{\mathrm{P}(\mathrm{C}) \cdot \mathrm{P}(\mathrm{M}|\mathrm{C})}{\mathrm{P}(\mathrm{M})} = \frac{0{,}8 \cdot 0{,}5}{0{,}44} = \frac{0{,}40}{0{,}44} = 0{,}91$$

17. B

x	f(x)	x·f(x)
0	0,64	0
1	0,32	0,32
2	0,04	0,08
		0,40

$\overline{X} = \sum x \cdot f(x) = 0{,}40$

18. A

La probabilidad de estudiar psicología clínica es 0,8 y, por tanto, la probabilidad de estudiar otra especialidad es 0,2. La probabilidad de que ninguno de los dos estudie psicología clínica es lo mismo que la probabilidad de que los dos estudien otra especialidad.

Tabla I con n=2, p=0,2 y x=2 0,04.

19. A

La probabilidad de que al menos 1 estudie psicología clínica es lo mismo que la probabilidad de que ninguno ó 1 estudien otra especialidad

X="nº alumnos que estudian otra especialidad"

P(X=0)+P(X=1)=0,6400+0,3200=0,96

Tabla I con n=2, p=0,2 y x=0 x=1.

20. C

$z = \frac{110-100}{10} = 1 \Rightarrow$ Tabla IV $0{,}8413 \Rightarrow 84{,}13\%$

21. B

Tabla VII

22. A

Tabla V

23. B

$$n = \frac{z^2_{1-\alpha/2}\sigma^2}{E^2_{máx}} = \frac{(1{,}96)^2 \cdot 100}{2^2} = \frac{384{,}16}{4} = 96{,}04 \cong 96$$

24. C

25. B

$L_i = \overline{X} - z_{1-\alpha/2}\sigma_{\overline{X}} \quad \rightarrow \quad 30 = 35 - z_{1-\alpha/2}\sigma_{\overline{X}}$

$z_{1-\alpha/2}\sigma_{\overline{X}} = 35 - 30 = 5$

$z_{1-\alpha/2} = z_{0{,}975} = 1{,}96$

$\sigma_{\overline{X}} = \frac{5}{1{,}96} = 2{,}55$

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2017 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO F DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

SOLUCIONES

Gráfica 1: Número de personas detenidas por corrupción y número de delitos conocidos por corrupción en España en los últimos cinco años (según el diario El País de 19-2-2016).

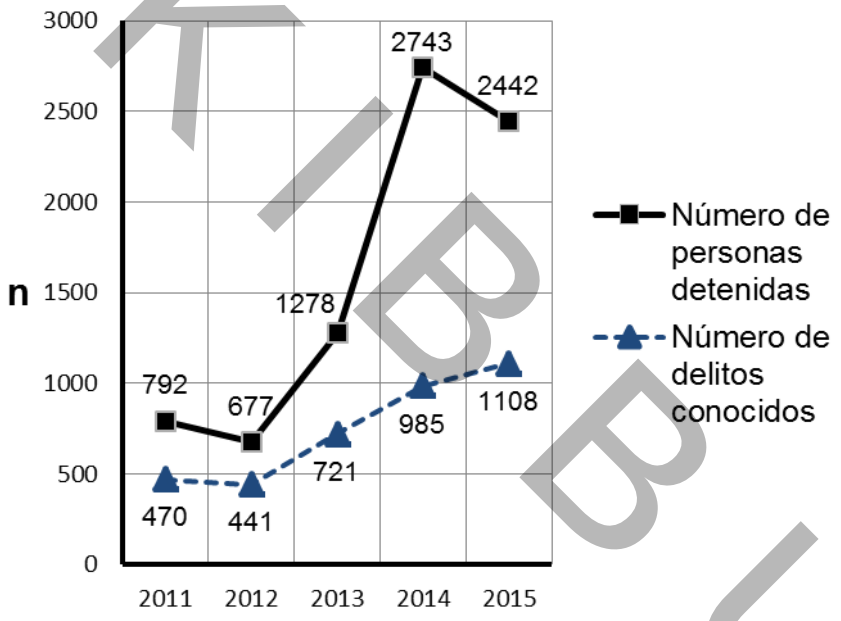


Tabla 1. Para pronosticar el rendimiento en un *Trabajo manual (Y)* a partir de las puntuaciones en un test de Impulsividad (X) disponemos de los siguientes datos de cinco escolares:

Escolar	***X***	***Y***
1	35	6
2	10	8
3	5	10
4	40	2
5	25	4

Se sabe que la pendiente de la recta de regresión de Y sobre X es igual a -0,18.

Gráfica 2: Número de chicos y chicas que sufren Acoso Físico y Ciberacoso en una muestra de 30 estudiantes de bachillerato acosados.

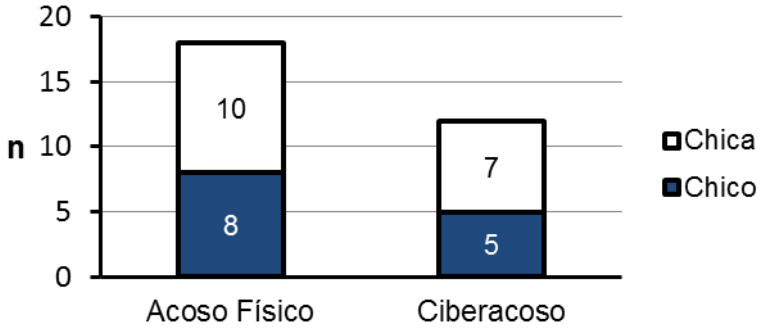


Tabla 2. Función de probabilidad, f(x), de la variable aleatoria discreta X.

x	***f(x)***
0	0,10
2	0,20
4	0,30
6	0,40

1. Con los datos de la Gráfica 1, el nivel de medida de la variable "número de personas detenidas por corrupción" es: A) de intervalo; B) ordinal; **C) de razón**.
2. Para los datos de la Tabla 1, ¿qué grafico se utiliza para representar la relación entre X e Y? A) Histograma; **B) Diagrama de dispersión**; C) Pictograma.
3. La variable "acoso" representada en el eje de abscisas de la Gráfica 2 es: A) politómica; **B) dicotómica**; C) cuantitativa discreta.
4. Para estudiar la tendencia central en variables cualitativas dicotómicas utilizamos: A) la media; B) la mediana; **C) la moda**.

KIBBUTZ

5. Con los datos de la Gráfica 1, la media de personas detenidas por corrupción a lo largo de los cinco años es: A) 970,76; B) 1278,54; **C) 1586,40.**
6. Según los datos de la Tabla 1, el percentil 50 de la variable "Impulsividad (X)" es: A) 17,5; **B) 25**; C) 35.
7. La cuasivarianza de la variable "Impulsividad (X)" de la Tabla 1 es: **A) 232,5**; B) 186,1; C) 10.
8. Con los datos de la Tabla 1, ¿qué puntuación típica le corresponde al escolar 3 en la variable "trabajo manual (Y)"?: A) 1,02; **B) 1,41**; C) 2,39.
9. Teniendo en cuenta los datos de la Tabla 1, ¿cuál es el índice de asimetría de Pearson del rendimiento en un trabajo manual (Y)? A) 0; B) 1; **C) No se puede calcular**.
10. Según los datos de la Tabla 1, en la recta de regresión de Y sobre X, la proporción de varianza de Y explicada por la varianza de X está entre: A) 0,50 y 0,55; **B) 0,75 y 0,80**; C) 0,85 y 0,90.
11. Con los datos de la Tabla 1, la ordenada en el origen de la ecuación de la recta de regresión para pronosticar la variable Y a partir de la variable X es: **A) 10,14**; B) 8,32; C) 7,51.
12. Considerando los datos de la Tabla 1, ¿qué puntuación en Y pronosticaremos a un escolar que ha obtenido una puntuación de X=23? **A) 6,0**; B) 4,2; C) 3,4.
13. Con los datos de la Gráfica 2, ¿podemos calcular el valor máximo que puede alcanzar el coeficiente C de contingencia entre las dos variables? **A) Sí**; B) No; C) Falta información.
14. En un espacio muestral hay dos sucesos A y B. Sabemos que P(A)=0,40, P(B)=0,30 y que $P(A \cap B) = 0{,}20$. ¿Cuál es la probabilidad de la unión de los sucesos A y B? A) 0,90; B) 0,70; **C) 0,50**.
15. Atendiendo a los datos de la Gráfica 2, si se ha elegido al azar un estudiante y resulta que es una chica, ¿cuál es la probabilidad de que sufra ciberacoso?: A) 0,23; **B) 0,41**;C) 0,57.
16. Con los datos de la Gráfica 2, si seleccionamos un estudiante al azar, ¿cuál es la probabilidad de que sea chico y sufra acoso físico?: **A) 0,27**; B) 0,43; C) 0,60.
17. Considerando un experimento aleatorio consistente en lanzar una moneda (no trucada) al aire tres veces y definiendo la variable aleatoria X como "número de cruces obtenidas", ¿cuántos resultados posibles tiene el espacio muestral?: **A) 8**; B) 4; C) 3.
18. Con los datos de la Tabla 2, la esperanza matemática de la variable aleatoria X es: A) 2; B) 3; **C) 4**.
19. Considerando los datos de la Tabla 2, la probabilidad de que la variable aleatoria X adopte valores menores o iguales a 4 es: A) 0,30; B) 0,40; **C) 0,60**.
20. En un colegio público, la probabilidad de que un alumno de 3º de Primaria suspenda la asignatura de matemáticas es de 0,3. Si seleccionamos al azar a 4 alumnos, ¿cuál es la probabilidad de que la mitad de esos alumnos suspendan la asignatura? **A) 0,2646**; B) 0,7354; C) 0,1500.
21. Las puntuaciones de una prueba de lectura siguen una distribución normal con media 50 y desviación típica 10. La proporción de sujetos en esta prueba con puntuaciones entre 40 y 60 es igual a: A) 0,1587; **B) 0,6826**; C) 0,8413.
22. Una variable aleatoria X se distribuye según la F de Snedecor con 8 grados de libertad en el numerador y 9 grados de libertad en el denominador. ¿Cuál es la probabilidad de que la variable X tome un valor menor o igual que 3,23?: **A) 0,95**; B) 0,80; C) 0,16.
23. En una variable aleatoria X que sigue la distribución Chi-cuadrado con varianza igual a 16, el percentil 90 es: A) 3,4895; **B) 13, 3616**; C) 23,5418.
24. Se ha aplicado un test de comprensión verbal a una muestra aleatoria de 324 niños de Educación Secundaria obteniendo una media igual a 52 y una cuasivarianza de 81. Para un nivel de confianza de 0,95, ¿cuál es el error de estimación máximo de la media?: **A) 0,98**; B) 1,29; C) 8,82.
25. En una muestra aleatoria de 300 estudiantes de Bachillerato se ha comprobado que 225 de ellos NO toman refrescos azucarados de forma habitual. Al nivel de confianza del 95%, ¿entre qué valores consideramos que se encuentra la proporción de la población de estudiantes de Bachillerato que SI toman refrescos azucarados de forma habitual? A) 0,152 y 0,348; **B) 0,201 y 0,299**; C) 0,701 y 0,799.

EXTRACCIÓN

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2017 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

SOLUCIONES

Situación 1. Una urna contiene 3 bolas blancas y 2 bolas negras. Se extraen 3 bolas con reposición y se define la variable aleatoria X "número de bolas negras extraídas".

Gráfico 1. Resultados, en cuanto a mejora de la memoria, de 150 ratones transgénicos, a los que se les ha transmitido la enfermedad de Alzheimer, y se les ha asignado aleatoriamente a recibir uno de los tres fármacos derivados de la proteína RGS14.

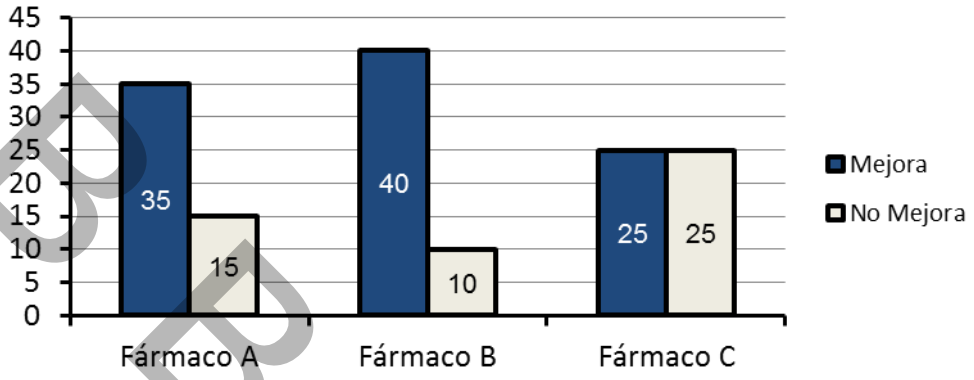


Tabla 2. Variables y características de su distribución

Variable	Distribución	Características
X	N(50,10)	Normal con media 50 y desviación típica 10
T	t_{40}	*t* de Student con 40 grados de libertad
V	$F_{10,5}$	F con 10 grados de libertad en el numerador y 5 grados de libertad en el denominador

1. La variable "autoritarismo legal" con las categorías "bajo", "medio" y "alto", está medida en una escala: A) ordinal; B) razón; C) intervalos.
2. Tres niños han obtenido las puntuaciones X=0, X=50, y X=100 en un test de inteligencia. Con esta escala es correcto afirmar que: A) el primer niño presenta carencia absoluta de inteligencia; B) el tercer niño con X=100 tiene el doble de inteligencia que el segundo niño con X=50; C) la diferencia en inteligencia entre X=0 y X=50 es la misma que entre X=50 y X=100.
3. La variable "Tipo de Fármaco", del Gráfico 1, es: A) dicotómica; B) cuantitativa; C) politómica.
4. Sea una variable X cuyos valores son 2, 3, 4, 6, 6 y 9. La media de X es: A) 4,5; B) 5; C) 6,5.
5. Si a todos los valores de la variable X del ejercicio anterior se les suma 4, ¿cuál es la media de esta nueva variable, que llamaremos Y?: A) $\overline{Y} = 5$; B) $\overline{Y} = 9$; C) $\overline{Y} = 34$.
6. Sea una variable X cuyos valores son 2, 3, 4, 6, 6 y 9. La moda de X es: A) 4,5; B) 6; C) 9.

EXAMEN

7. Con los datos del ejercicio anterior, ¿entre qué valores se encuentra la desviación típica? A) 2 y 3; B) 3,5 y 4,5; C) 5 y 6.
8. Si a todos los valores de la variable X del ejercicio anterior les sumamos 1, llamando Y a esta nueva variable, ¿cuál será la desviación típica de Y? A) $S_y = S_x + 1$; B) $S_y = S_x$; C) $S_y = 1-S_x$.
9. La amplitud semi-intercuartil es un estadístico utilizado para medir: A) tendencia central; B) variabilidad; C) asimetría.
10. Según los datos del Gráfico 1, asumiendo que ambas variables son independientes, ¿cuál es la frecuencia teórica de mejorar tras recibir el fármaco B? A) 16,7; B) 33,3; C) 40.
11. Con los datos del Gráfico 1, hemos obtenido un valor de X^2 igual a 10,5. El coeficiente C de Contingencia está comprendido entre: A) 0,1 y 0,4; B) 0,5 y 0,8; C) 0,9 y 1,1.
12. La ordenada en el origen de la recta de regresión Y' = 6X – 4 es: A) -4; B) 4; C) 6.
13. Se utiliza la regresión lineal para pronosticar el rendimiento en matemáticas de un grupo de estudiantes (Y) en función de su razonamiento abstracto (X). Sabiendo que la proporción de varianza explicada es 0,64 y que la pendiente de la recta de regresión utilizada es mayor que 0, ¿se puede calcular el coeficiente de correlación de Pearson entre ambas variables? A) No se puede calcular con estos datos; B) Sí se puede calcular y es 0,41; C) Sí se puede calcular y es 0,8.
14. Teniendo en cuenta la representación del Gráfico 1, elegida una rata al azar ha resultado que se le administró el fármaco A ¿cuál es la probabilidad de que mejore?: A) 0,23; B) 0,50; C) 0,70.
15. Teniendo en cuenta la representación del Gráfico 1, si elegimos una rata al azar ¿cuál es la probabilidad de que presente mejoría en su memoria?: A) 0,33; B) 0,67; C) 0,70.
16. Atendiendo a los datos de la situación 1, ¿cuál sería la probabilidad de que las tres bolas sean negras? A) 0,080; B) 0,064; C) 0,160.
17. La función de distribución de la variable aleatoria descrita en la Situación 1 para cada uno de los valores que adopta será: A) $[0{,}360; 0{,}840; 1]$; B) $[0{,}216; 0{,}648; 0{,}936; 1]$; C) $[0{,}216; 0{,}432; 0{,}288; 0{,}064]$.
18. Atendiendo a los datos de la Situación 1, ¿cuál es la probabilidad de extraer una bola negra ó más? A) 0,216; B) 0,784; C) 0,936.
19. Con los datos de la Situación 1, ¿cuál será la media de la variable aleatoria descrita? A) 0,50; B) 0,72; C) 1,20.
20. Atendiendo a los datos de la Tabla 2, si a una persona le corresponde el percentil 33, su puntuación en la variable X será igual a: A) 33; B) 45,6; C) 54,4.
21. Atendiendo a los datos de la Tabla 2, ¿cuál es la probabilidad de que la variable T adopte valores mayores que 1,05? A) 0,15; B) 0,25; C) 0,85**.**
22. Según los datos de la Tabla 2, ¿cuál es el percentil 5 de la variable V? A) 0,05; B) 0,21; C) 0,30.
23. El muestreo por cuotas es: A) probabilístico; B) aleatorio; C) no probabilístico.
24. ¿Cuál es el error máximo cometido para estimar la media en una muestra aleatoria de 100 personas, con una desviación típica poblacional de 3, y un nivel de confianza de 0,95? A) 0,588; B) 0,745; C) 1,746.
25. Sabiendo que el error típico de la media es 0,5, y la varianza de la población es 36, ¿con qué tamaño de muestra se realizó la investigación? A) 72; B) 144; C) 5184.

KIBBUTZ

SOLUCIONES:

1. A
2. C
3. C
4. B

$$\overline{X} = \frac{2+3+4+6+6+9}{6} = \frac{30}{6} = 5$$

5. B

$$\overline{Y} = b\overline{X} + a = 1 \times 5 + 4 = 9$$

6. B

La Mo es 6 puesto que es el valor que más se repite

7. A

$$S_x^2 = \frac{\sum X^2}{n} - \overline{X}^2 = \frac{4+9+16+36+36+81}{6} - 25 = 5,\widehat{3}$$

$$S_x = \sqrt{S_x^2} = \sqrt{5,\widehat{3}} = 2,31$$

8. B

Si a una variable X se le suma una constante, la desviación típica de la variable original no se ve afectada y sigue siendo la misma.

9. B
10. B

	Fármaco A	Fármaco B	Fármaco C	
Mejora	35	40	25	100
No mejora	15	10	25	50
	50	50	50	150

$$\textit{Frecuencia teórica} = n_t = \frac{\textit{total fila} \times \textit{total columna}}{n} = \frac{100 \times 50}{150} = 33,\widehat{3}$$

11. A

$$C = \sqrt{\frac{X^2}{X^2 + n}} = \sqrt{\frac{10,5}{10,5 + 150}} = \sqrt{0,06542} = 0,25$$

12. A

Y' = 6X – 4 → a = -4

13. C

La proporción de varianza explicada es $r_{XY}^2 = 0,64$

$r_{XY} = 0,8$

14. C

KIBBUTZ

	Fármaco A	Fármaco B	Fármaco C	
Mejora	35	40	25	100
No mejora	15	10	25	50
	50	50	50	150

$$P(M|A) = \frac{35}{50} = 0{,}7$$

15. B

	Fármaco A	Fármaco B	Fármaco C	
Mejora	35	40	25	100
No mejora	15	10	25	50
	50	50	50	150

$$P(M) = \frac{100}{150} = 0{,}67$$

16. B

$$\frac{2}{5}\cdot\frac{2}{5}\cdot\frac{2}{5} = \frac{8}{125} = 0{,}064$$

También utilizando la binomial con n=3, x=3 y p=0,4

17. B

Tabla II de la Binomial con n = 3 y p = 0,40.

18. B

$$P(X \geq 1) = 1 - P(X < 1) = 1 - F(0) = 1 - 0{,}216 = 0{,}784$$

19. C

$$np = 3 \times 0{,}40 = 1{,}20$$

20. B

Al Percentil 33 le corresponde una z = -0,44 (Tabla IV)

$$z = \frac{P_{33} - \mu}{\sigma} \Rightarrow -0{,}44 = \frac{P_{33} - 50}{10} \Rightarrow -4{,}4 = P_{33} - 50 \Rightarrow P_{33} = -4{,}4 + 50 = 45{,}6$$

21. A

$$P(X > 1{,}05) = 1 - P(X \leq 1{,}05) = 1 - 0{,}85 = 0{,}15$$

(Tabla VI: distribución t de Student con 40 g.l.)

22. C

$${}_{0,05}F_{10,5} = \frac{1}{{}_{0,95}F_{5,10}} = \frac{1}{3{,}326} = 0{,}30$$

23. C

24. A

$$n.c. = 0{,}95 \quad \rightarrow \quad Z_{1-\alpha/2} = Z_{0,975} = 1{,}96 \quad (\text{Tabla IV})$$

$$E_{max} = Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 1{,}96 \cdot \frac{3}{\sqrt{100}} = 1{,}96 \cdot \frac{3}{10} = 0{,}588$$

25. B

$$n = \frac{z^2_{1-\alpha/2}\sigma^2}{E^2_{máx}} = \frac{z^2_{1-\alpha/2}\sigma^2}{z^2_{1-\alpha/2}\sigma^2_{\bar{X}}} = \frac{36}{0{,}5^2} = \frac{36}{0{,}25} = 144$$

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
SEPTIEMBRE 2016 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO C DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

SOLUCIONES

KIBBUTZ

Tabla 1. Resultados de un test de autoestima en una muestra de adolescentes españoles.

Autoestima	n_i
80 - 99	60
60 – 79	120
40 – 59	180
20 – 39	90
0 – 19	50

Tabla 2. Nivel de ansiedad previo a la realización del examen de conducir y nota obtenida en el examen de 200 personas.

		Nota en el examen	
		Apto	No-apto
Nivel de ansiedad	Normal	100	10
	Alto	40	50

Tabla 3: Un psicólogo utilizó un test de Analogías Verbales (X) para predecir el rendimiento en Lengua de 4º de la ESO (Y). Obtuvo las puntuaciones de las dos variables en una muestra aleatoria de 1000 estudiantes. En la tabla se muestran las medias, las varianzas y la correlación entre ambas variables.

	X	Y
Media	30	15
Varianza	64	36
Correlación	$r_{XY} = 0{,}70$	

Tabla 4: Función de probabilidad de una variable aleatoria discreta X

x	f(x)
2	0,35
1	0,40
0	0,25

1. ¿Cuál es la amplitud de los intervalos de la variable *autoestima* de la Tabla 1?: A) 10; B) 19; **C) 20**
2. ¿Cuál es la media en autoestima de la muestra de adolescentes españoles de la Tabla 1?: A) 49,5; **B) 51,5;** C) 5150
3. Los percentiles son medidas de: A) tendencia central; **B) posición;** C) desviación
4. Con los datos de la Tabla 1, la Mediana de la variable Autoestima es: A) 50,5; B) 51,5; **C) 51,72**
5. Para calcular el rango o amplitud total de una variable necesitamos: A) las frecuencias absolutas; B) que la variable sea cualitativa; **C) las puntuaciones máxima y mínima**
6. ¿Cuál de los siguientes índices puede adoptar valores negativos? **A) El índice de asimetría de Pearson**; B) La varianza; C) La desviación típica.

7. Con los datos de la Tabla 3, ¿qué variable presenta un mayor grado de dispersión?: A) las puntuaciones en el test de analogías verbales; **B) las puntuaciones en lengua;** C) las dos variables presentan el mismo grado de dispersión
8. La puntuación típica de un sujeto en un test de razonamiento abstracto ha sido de 2. ¿Cuál fue su puntuación diferencial en el test si la media fue de 10 y la varianza de 9?: A) 0; B) 16; **C) 6**
9. Con los datos de la Tabla 2, el valor de X^2 está entre: A) 35 y 45; **B) 46 y 56**; C) 57 y 67
10. Cuando el coeficiente de contingencia es igual al C_{max}, podemos decir que: A) Existe una relación moderada entre las variables; **B) Existe una fuerte relación entre las variables**; C) No existe relación entre las variables
11. En la recta de regresión, la proporción de varianza de Y explicada por la varianza de X es igual a: **A) El coeficiente de correlación entre ambas variables al cuadrado**; B) El coeficiente de correlación entre ambas variables; C) La raíz cuadrada del coeficiente de correlación entre ambas variables
12. Con los datos de la Tabla 3, la covarianza entre ambas variables es igual a: A) 83,6; B) 25,3; **C) 33,6**
13. La recta de regresión de Y sobre X siempre pasa por el punto: **A)** $(\overline{X},\overline{Y})$; B) $(0,\overline{Y})$; C) $(\overline{X},0)$
14. Con los datos de la Tabla 3, la pendiente de la recta de regresión de Y sobre X en puntuaciones directas es: **A) 0,525**; B) 0,385; C) 0,495
15. Con los datos de la Tabla 2 si elegimos al azar una persona, ¿cuál es la probabilidad de que presente un nivel de ansiedad "Normal"?: A) 0,50; **B) 0,55**; C) 0,60
16. Con los datos de la Tabla 2 si elegimos al azar una persona, ¿cuál es la probabilidad de que presente un nivel de ansiedad "Normal" y esté "Apto" en el examen?: **A) 0,50**; B) 0,55; C) 0,60
17. Con los datos de la Tabla 2, elegimos al azar una persona y ha resultado que tiene un nivel de ansiedad "Alto", ¿cuál es la probabilidad de que presente una calificación de "Apto" en el examen?: **A) 0,44**; B) 0,55; C) 0,66
18. La probabilidad de que un paciente con esquizofrenia se recupere con un tratamiento determinado es de 0,40. Un psicólogo está tratando individual e independientemente a 10 pacientes con este trastorno. La probabilidad de que se recuperen al menos 7 pacientes es de: A) 0,9452; **B) 0,0548**; C) 0,0123
19. Con los datos de la Tabla 4, la función de distribución de la variable *X* es: **A) 0,25; 0,65 y 1**; B) 0,25; 0,40 y 0,35; C) 0,40; 0,60 y 1
20. Con los datos de la Tabla 4, la esperanza matemática de la variable *X* es igual a: A) 0,70; B) 1,0; **C) 1,1**
21. Las puntuaciones en un test de *memoria* se distribuyen normalmente en una población de 10000 adolescentes con media 100 y desviación típica 10. ¿Cuántos adolescentes tienen una puntuación superior a 115 en el test de *memoria*?: A) ninguno; B) 600; **C) 668**
22. En una distribución Chi-cuadrado con 20 grados de libertad, el percentil 10 vale: A) 8,2604; **B) 12,4426;** C) 20
23. En una distribución F de Snedecor con 10 grados de libertad en el numerador y 10 grados de libertad en el denominador, al valor 2,323 le corresponde el percentil: A) 9; B) 10; **C) 90**
24. Se mide la depresión en una muestra aleatoria de 225 sujetos obteniendo una cuasivarianza de 25. El error típico de la media vale: A) 0,11; **B) 0,33;** C) 3
25. Para estimar el intervalo de confianza para la media de una variable X, a un nivel de confianza del 99%, hemos extraído una muestra aleatoria de 144 personas y les hemos medido la variable X obteniendo una media de 12 y una cuasidesviación típica de 4. Los límites del intervalo de confianza son: A) 8 y 14; B) 9,42 y 14,58; **C) 11,14 y 12,86**

KIBBUTZ

SOLUCIONES:

1. C
2. B

Autoestima	X_i	n_i	n_iX_i
80 - 99	89.5	60	5370
60 – 79	69.5	120	8340
40 – 59	49.5	180	8910
20 – 39	29.5	90	2655
0 – 19	9.5	50	475
		500	25750

$$\overline{X} = \frac{\sum n_i X_i}{n} = \frac{25750}{500} = 51{,}5$$

3. B
4. C

Autoestima	X_i	n_i	n_a
80 - 99	89.5	60	500
60 – 79	69.5	120	440
40 – 59	49.5	180	320
20 – 39	29.5	90	140
0 – 19	9.5	50	50
		500	

$$Md = L_i + \left(\frac{\frac{n}{2} - n_d}{n_c}\right) \cdot I = 39{,}5 + \left(\frac{\frac{500}{2} - 140}{180}\right) \cdot 20 = 39{,}5 + 12{,}22 = 51{,}72$$

5. C
6. A
7. B

	X	Y
Media	30	15
Varianza	64	36

$$\left.\begin{array}{l} CV_X = \frac{S_X}{\overline{X}} \cdot 100 = \frac{8}{30} \cdot 100 = 26{,}67 \\ CV_Y = \frac{S_Y}{\overline{Y}} \cdot 100 = \frac{6}{15} \cdot 100 = 40 \end{array}\right\} \Rightarrow CV_Y > CV_X$$

8. C

KIBBUTZ

$$x = X - \overline{X}$$

$$z = \frac{x}{S_X} \rightarrow 2 = \frac{x}{3} \rightarrow x = 2 \times 3 = 6$$

9. B

		Nota en el examen		
		Apto	No-apto	
Nivel de Ansiedad	Normal	100 (77)	10 (33)	110
	Alto	40(63)	50(27)	90
		140	60	200

Las frecuencias teóricas aparecen entre paréntesis en la tabla.

$$X^2 = \frac{(100-77)^2}{77} + \frac{(40-63)^2}{63} + \frac{(10-33)^2}{33} + \frac{(50-27)^2}{27} = 6{,}8701 + 8.3968 + 16{,}0303 + 19{,}5926 = 50{,}89$$

10. B
11. A
12. C

	X	Y
Media	30	15
Varianza	64	36

$$r_{XY}^2 = 0{,}49 \Rightarrow r_{XY} = 0{,}70$$

$$S_{XY} = r_{XY} S_X S_Y = 33{,}6$$

13. A
14. A

$$b = r_{XY} \frac{S_Y}{S_X} = 0{,}70 \frac{6}{8} = 0{,}525$$

15. B

$$P(\text{Normal}) = \frac{110}{200} = 0{,}55$$

16. A

$$P(\text{Normal y Apto}) = \frac{100}{200} = 0{,}50$$

17. A

$$P(\text{Apto / Alto}) = \frac{40}{90} = 0{,}44$$

18. B

p=0,40 $P(X \geq 7) = 1 - P(X < 7) = 1 - P(X \leq 6) = 1 - F(6) = 1 - 0{,}9452 = 0{,}0348$

El valor de F(6) se busca en la tabla II, y es el valor de la intersección de la fila n=10, x=6 con la columna p=0,40.

19. A

KIBBUTZ

x	f(x)	F(x)
2	0,35	1
1	0,40	0,65
0	0,25	0,25

20. C

x	f(x)	xf(x)
2	0,35	0,70
1	0,40	0,40
0	0,25	0
		1,1

$\mu = 1{,}1$

21. C

$$Z = \frac{X - \mu}{\sigma} = \frac{115 - 100}{10} = 1{,}5$$

z = 1,5 $\rightarrow$ P(Z ≤ 1,5) = 0,9332 (Tabla IV)

P(Z > 1,5) = 1 – P(Z ≤ 1,5) = 1 - 0,9332 = 0,0668

10000 x 0,0668 = 668

22. B

P_{10} = 12,4426 (Tabla V para 20 g.l.)

23. C

Tabla VII para 10 g.l. en el numerador y 10 g.l. en el denominador

24. B

$$S_{\overline{X}} = \sqrt{\frac{S_{n-1}^2}{n}} = \sqrt{\frac{25}{225}} = 0{,}33$$

25. C

$$\overline{X} \pm z_{1-\alpha/2} \frac{S_{n-1}}{\sqrt{n}}$$

$$E_{máx} = z_{1-\alpha/2} \frac{S_{n-1}}{\sqrt{n}} = 2{,}58 \frac{4}{\sqrt{144}} = 0{,}86$$

$$\overline{X} - E_{máx} = 12 - 0{,}86 = 11{,}14$$

$$\overline{X} + E_{máx} = 12 + 0{,}86 = 12{,}86$$

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2016 Código asignatura: 62011037

EXAMEN TIPO TEST MODELO F DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

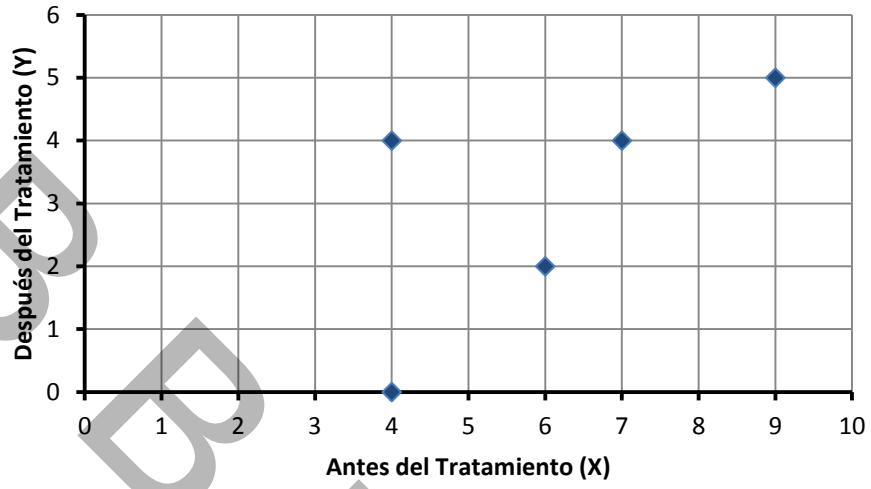
Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Tabla 1. Porcentaje de fallos obtenidos (X) por los 150 alumnos que se presentaron al examen teórico del carnet de conducir. Para aprobar el examen el nº máximo de fallos que se puede tener es 3.

X	***P_i***
4 o más	20
2-3	30
0-1	50

Gráfica 1. Número de conductas bulímicas registradas en una semana antes (X) y después (Y) de un tratamiento aplicado a 5 niñas con trastornos de la alimentación.



Gráfica 2: Representación del tipo de trastorno (disortografía o disgrafía) en función el sexo de un grupo de 200 escolares con trastornos en la escritura.

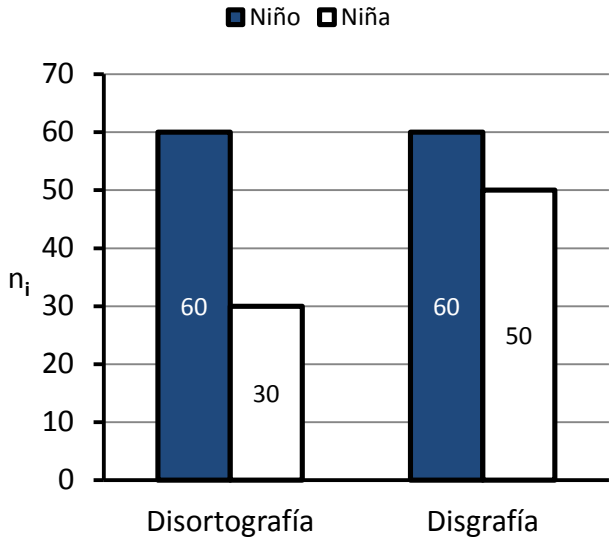


Tabla 2: Función de Distribución, F(x), de la variable aleatoria discreta X.

x	F(x)
0	0,25
1	0,50
2	0,75
3	1

1. Con los datos de la Tabla 1, el nivel de medida de la variable *número de fallos en el examen teórico del carnet de conducir* es: A) ordinal; B) de intervalo; **C) de razón**.
2. ¿En qué representación gráfica se utiliza el diagrama de dispersión? **A) en la Gráfica 1**; B) En la Gráfica 2; C) En ninguna gráfica.

EXTRABUTN

3. Según los datos de la Tabla 1, ¿cuántos alumnos han suspendido el carnet de conducir? A) 20; **B) 30**; C) 80.
4. Teniendo en cuenta los datos de la Tabla 1, la moda del nº de fallos obtenidos en el examen teórico de conducir es: **A) 0,5**; B) 1; C) 50.
5. Según los datos de la Tabla 1, ¿cuál es la mediana de los fallos obtenidos en el examen teórico de conducir? A) 0,75; **B) 1,5**; C) No se puede calcular.
6. Con los datos de la Tabla 1, el percentil 90: A) es 3,5; B) es 4,9; **C) no se puede calcular.**
7. Teniendo en cuenta la Gráfica 1, ¿cuál es la desviación típica del nº de conductas bulímicas registradas antes del tratamiento (X)? **A) 1,9**; B) 3,2; C) 3,6.
8. Con los datos de la Tabla 1, la distribución del nº de fallos cometidos en el examen teórico del carnet de conducir es: A) simétrica; **B) asimétrica positiva**; C) asimétrica negativa.
9. Atendiendo a los datos de la variable X de la Gráfica 1, ¿qué puntuación típica le corresponde a la niña que ha registrado 6 conductas bulímicas antes de recibir el tratamiento (X)? **A) 0**; B) 3; C) 6.
10. Teniendo en cuenta la Gráfica 2 y sabiendo que X^2 =3,03, el valor C de Contingencia está entre: **A) 0 y 0,20**; B) 0,21 y 0,40; C) 0,41 y 0,6.
11. Con los datos de la Gráfica 2, ¿cuál es el valor máximo que puede alcanzar el coeficiente C de contingencia? **A) 0,707**; B) 0,816; C) 1**.**
12. Atendiendo a los datos de la Gráfica 1, la pendiente de la recta de regresión que permite pronosticar el nº de conductas bulímicas después del tratamiento (Y) a partir del nº de conductas bulímicas registradas antes del tratamiento (X) es: A) 0,14; **B) 0,61**; C) 1,13.
13. Atendiendo a la Gráfica 1, ¿qué nº de conductas bulímicas pronosticaremos a una niña después del tratamiento, si antes de someterse al mismo su registro es de 10? **A) 5,44**; B) 6,1; C) 6,76.
14. Sean los sucesos A={1,3,4,6,8} y B={2,3,6,8,9}. La unión de A y B es igual a: A) {3,6,8}; B) {1,2,4,9}; **C) {1,2,3,4,6,8,9}**.
15. Con relación a la Gráfica 2, si elegimos al azar una persona, la probabilidad de que sea niña y padezca disortografía es: **A) 0,15**; B) 0,25; C) 0,375.
16. Teniendo en cuenta la Gráfica 2, si se elige al azar una persona y observamos que es un niño, ¿cuál es la probabilidad de que padezca disgrafía? A) 0,25; B) 0,35; **C) 0,50**.
17. Según los datos de la Tabla 2, ¿cuál es la probabilidad de que la variable aleatoria X adopte el valor 0? A) No se puede calcular; B) 0; **C) 0,25.**
18. Con los datos de la Tabla 2, ¿cuál es la esperanza matemática de la variable aleatoria X? A) 0,25; B) 1,25; **C) 1,50**.
19. En una determinada universidad, el 20% de los estudiantes tienen concedida una beca. Si elegimos aleatoriamente una muestra de 6 estudiantes de esa universidad, ¿cuál es la probabilidad de que más de la mitad disfruten de una beca? **A) 0,017**; B) 0,082; C) 0,983.
20. En una distribución normal tipificada, ¿cuál es el valor de la variable que deja por encima de sí el 16,6% de las observaciones? A) -0,97; **B) 0,97**; C) 1,66.
21. Las puntuaciones en un cuestionario de personalidad siguen una distribución normal con media 30 y desviación típica 5. El percentil 67 de la distribución es igual a: A) 0,44; **B) 32,2**; C) 44.
22. En una variable X que se distribuye según la Chi-cuadrado con media igual a 3, la probabilidad de que $(7{,}8147 \leq X \leq 9{,}3484)$ es: **A) 0,025**; B) 0,050; C) 0,075.
23. En una estimación por intervalos, el nivel de confianza es: A) α; B) 1-α/2; **C) 1-α**.
24. La capacidad abstracta se distribuye normalmente en la población de estudiantes de la ESO. En una muestra aleatoria de 25 estudiantes se obtuvo una media de 50 y una cuasivarianza de 16. Para un nivel de confianza de 0,95, ¿entre qué límites se estima esté la capacidad abstracta media de la población?: A) 40,5 y 50,5; B) 35,872 y 74,128; **C) 48,35 y 51,65**.
25. Si el error típico de la proporción es igual a 0,20 y el tamaño de la muestra es 144, ¿cuál es el valor de la varianza de la distribución muestral de la proporción? A) 0,02; **B) 0,04**; C) 0,20.

EXAMEN

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2016 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Figura 1. Número de respuestas correctas en un test de vocabulario de cinco ítems (X) aplicado a una muestra de 200 niños.

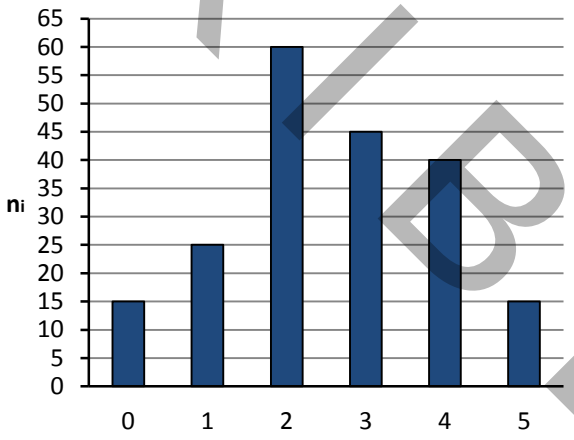


Tabla 1: Para pronosticar las puntuaciones en una asignatura (Y) a partir de las puntuaciones en un test de razonamiento (X) disponemos de los siguientes datos:

	Media	Varianza	Pendiente de la recta de regresión de Y sobre X
X	**50**	**36**	**b=0,20**
Y	**8**	**4**	

Tabla 2: Nivel de estrés laboral en función del sexo de los trabajadores de una determinada empresa.

	Nivel de estrés laboral		
	Bajo	**Medio**	**Alto**
Hombre	40	74	8
Mujer	40	28	10

Tabla 3: Conjunto de variables y su tipo de distribución.

Variable	**Distribución**
X	Normal con media 50 y varianza 100
Y	Chi-cuadrado con media igual a 5
V	t de Student con 25 grados de libertad

EBBUTN

1. El “número de respuestas correctas” de la Figura 1 es una variable: A) dicotómica; B) cuasicuantitativa; C) cuantitativa.
2. La Figura 1 es un diagrama de: A) barras; B) barras acumulativo; C) barras conjunto.
3. Con los datos de la Figura 1, ¿cuál es la frecuencia relativa para el valor X=4?: A) 0,20; B) 0,30; C) 0,08.
4. La media del número de respuestas correctas, según los resultados de la Figura 1 es: A)1,62; B) 2,06; C) 2,58.
5. Según los datos de la Figura 1, ¿qué percentil le corresponde a un niño con una puntuación de 4?: A) 76; B) 83; C) 93.
6. Con los datos de la Tabla 1, sabiendo que la variable X presenta una distribución simétrica y unimodal, ¿cuál es la moda de esta variable?: A) 44; B) 50; C) No se puede saber con la información dada.

7. La varianza de la variable “número de respuestas correctas” de la Figura 1 está entre: A) 1,65 y 2,10; B) 3,90 y 4,35; C) 5,65 y 6,10.
8. Considerando los datos de la Tabla 1, se sabe que en la variable Y la moda es mayor que la media. ¿Qué forma presenta la distribución de esa variable?: A) simétrica; B) asimétrica positiva; C) asimétrica negativa.
9. Atendiendo a los datos de la Tabla 1, ¿qué puntuación típica le corresponde a una persona que ha obtenido una puntuación de 59 en la variable X?: A) 0,4; B) 1,5; C) 2.
10. Con los datos de la Tabla 2, si el valor del estadístico X^2 para cuantificar la relación entre ambas variables es 11,86, ¿cuánto vale el coeficiente C de contingencia?: A) 0,24; B) 0,64; C) 0,84.
11. Según los datos de la Tabla 1, la covarianza entre ambas variables es igual a: A) 3,7; B) 7,2; C) 9,6.
12. Con los datos de la Tabla 1, la ordenada en el origen de la ecuación de la recta de regresión para pronosticar la variable Y a partir de la variable X es: A) -2; B) 0; C) 3.
13. Teniendo en cuenta los datos de la Tabla 1, ¿qué puntuación en Y pronosticaremos a una persona que ha obtenido un puntuación de X=56?: A) 9,2; B) 11,2; C) 14,2.
14. Sean los sucesos A={1, 3, 6, 8} y B={2, 3, 7, 8}. La intersección de A y B es igual a: A) {1, 2, 6, 7}; B) {1, 2, 3, 6, 7, 8}; C) {3, 8}.
15. Con los datos de la Tabla 2, si seleccionamos un trabajador al azar, ¿cuál es la probabilidad de que sea mujer y presente un nivel medio de estrés laboral?: A) 0,05; B) 0,14; C) 0,56.
16. Según los datos de la Tabla 2, si se ha elegido al azar un trabajador y resulta que es un hombre, ¿cuál es la probabilidad de que presente un nivel bajo de estrés laboral?: A) 0,20; B) 0,33; C) 0,50.
17. Una variable aleatoria es continua si: A) puede tomar infinitos valores; B) su función de probabilidad puede adoptar valores negativos; C) la suma de los valores de la función de distribución es igual a uno.
18. Una característica de la distribución binomial es que su esperanza matemática es igual: A) al producto de la probabilidad de éxito por la probabilidad de fracaso; B) a la varianza dividida entre la probabilidad de éxito; C) al número de ensayos multiplicado por la probabilidad de éxito.
19. En un hospital se sabe que de los pacientes diagnosticados de esquizofrenia que son tratados con un determinado antipsicótico el 35% presentan convulsiones como efecto adverso. Si elegimos al azar una muestra de 5 de estos pacientes, ¿cuál es la probabilidad de que tres o menos de estos pacientes padezcan convulsiones?: A) 0,1811; B) 0,8189; C) 0,9460.
20. Según los datos de la Tabla 3, la varianza de la variable Y es: A) 5; B) 10; C) 20.
21. Con los datos de la Tabla 3, la probabilidad de obtener valores menores o iguales a V=1,708 es: A) 0,05; B) 0,50; C) 0,95.
22. Atendiendo a los datos de la variable X de la Tabla 3, ¿cuál es la probabilidad de obtener valores comprendidos entre las puntuaciones 45 y 60?: A) 0,5328; B) 0,8413; C) 0,8621.
23. Las puntuaciones en una escala de asertividad se distribuyen según la normal con desviación típica poblacional igual a 12. Si queremos estimar la media de la población con un error de estimación máximo de 4 y con un nivel de confianza de 0,95, ¿qué tamaño debe tener la muestra seleccionada?: A) 18; B) 35; C) 76.
24. Se ha aplicado un test de creatividad a una muestra aleatoria de 196 niños de Educación Secundaria obteniendo una media igual a 12 y una cuasidesviación típica de 7. Para un nivel de confianza de 0,95, ¿entre qué límites se estima esté la media en creatividad de la población de niños de Educación Secundaria?: A) 10,04 y 13,96; B) 10,52 y 12,48; C) 11,02 y 12,98.
25. En un estudio sobre la proporción de fumadores en una determinada población los límites del intervalo de confianza para la proporción son 0,30 y 0,42. ¿Con qué error de estimación máximo se han calculado?: A) 0,06; B) 0,12; C) 0,24.

EXAMEN

SOLUCIONES:

1. C
2. A
3. A

$$p_i = \frac{n_i}{n} = \frac{40}{200} = 0{,}20$$

4. C

X_i	n_i	$X_i\, n_i$
5	15	75
4	40	160
3	45	135
2	60	120
1	25	25
0	15	0
	200	515

$$\overline{X} = \frac{\sum n_i X_i}{n} = \frac{515}{200} = 2{,}575 \approx 2{,}58$$

5. B

X_i	n_i	n_a
5	15	200
4	40	185
3	45	145
2	60	60
1	25	40
0	15	15
	200	

La puntuación X=4 está en el intervalo [3,5-4,5].

$$k = \left[\frac{\frac{(P_k - L_i)\cdot n_c}{I} + n_d}{n} \right] \cdot 100 = \left[\frac{\frac{(4-3{,}5)\cdot 40}{1} + 145}{200} \right] \cdot 100 = 82{,}5 \approx 83$$

Por lo tanto, a la puntuación X=4 le corresponde el percentil 83.

6. B
7. A

X_i	n_i	$X_i\, n_i$	$n_i\, X_i^2$
5	15	75	375
4	40	160	640
3	45	135	405
2	60	120	240
1	25	25	25
0	15	0	0
	200	515	1685

KIBBUTZ

$\overline{X} = 2{,}58$

$$S_x^2 = \frac{\sum n_i X_i^2}{n} - \overline{X}^2 = \frac{1685}{200} - 2{,}58^2 = 8{,}425 - 6{,}6564 = 1{,}7686 \approx 1{,}77$$

8. C
9. B

$$z = \frac{X - \overline{X}}{S_X} = \frac{59 - 50}{6} = \frac{9}{6} = 1{,}5$$

10. A

$$C = \sqrt{\frac{X^2}{X^2 + n}} = \sqrt{\frac{11{,}86}{11{,}86 + 200}} = 0{,}2366016999 \approx 0{,}24$$

11. B

$$b = r_{xy}\frac{S_Y}{S_X} \rightarrow r_{xy} = b\frac{S_X}{S_Y} = 0{,}20\frac{6}{2} = 0{,}6$$

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} \rightarrow S_{XY} = S_X S_Y r_{XY} = 6 \times 2 \times 0{,}60 = 7{,}2$$

12. A

$$a = \overline{Y} - b\overline{X} = 8 - 0{,}20 \times 50 = -2$$

13. A

$$Y_i' = a + bX = -2 + 0{,}20X = -2 + 0{,}20 \times 56 = 9{,}2$$

14. C

A={1, 3, 6, 8} y B={2, 3, 7, 8} $\qquad A \cap B = \{3, 8\}$

15. B

	Nivel de estrés laboral			
	Bajo	Medio	Alto	
Hombre	40	74	8	122
Mujer	40	28	10	78
	80	102	18	200

$$P(\text{Mujer} \cap \text{Medio}) = \frac{28}{200} = 0{,}14$$

16. B

	Nivel de estrés laboral			
	Bajo	Medio	Alto	
Hombre	40	74	8	122
Mujer	40	28	10	78
	80	102	18	200

KIBBUTZ

$$P(bajo/H) = \frac{P(Bajo \cap H)}{P(H)} = \frac{40/200}{122/200} = \frac{0{,}20}{0{,}61} = 0{,}3278688525 \approx 0{,}33$$

17. A

18. C

19. C

Función de distribución binomial (Tabla II) con n=5, p=0,35

$P(X \leq 3) = 0{,}9460$

20. B

La media y la varianza de una distribución Chi cuadrado son n y 2n, respectivamente.

$\mu = n = 5, \quad \sigma^2 = 2n = 2x5 = 10$

21. C

En la tabla VI para 25 g.l, la $P(T \leq 1{,}708) = 0{,}95$

22. A

Tablas III y IV

$$z = \frac{45-50}{10} = -0{,}5 \qquad z = \frac{60-50}{10} = 1$$

La probabilidad de obtener valores comprendidos entre las puntuaciones z = -0,5 y z = 1es:

$$P(-0{,}5 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq -0{,}5) = 0{,}8413 - 0{,}3085 = 0{,}5328$$

23. B

n.c. = 0,95 → $z_{1-\alpha/2} = z_{1-0{,}05/2} = z_{0{,}975} = 1{,}96$ (Tabla IV) $\quad E_{máx} = 4$

$$n = \frac{z^2_{1-\alpha/2}\sigma^2}{E^2_{máx}} = \frac{1{,}96^2 \times 12^2}{4^2} = 34{,}5744 \approx 35$$

24. C

n=196 (Aproximación de la distribución t de Student a la normal)

$$n.c. = 0{,}95 \quad \rightarrow \quad z_{1-\alpha/2} = z_{0{,}975} = 1{,}96 \quad (\text{Tabla IV})$$

$$E_{max} = z_{1-\alpha/2}\frac{S_{n-1}}{\sqrt{n}} \rightarrow E_{max} = 1{,}96\frac{7}{14} = 0{,}98$$

$$L_i = \overline{X} - E_{max} = 12 - 0{,}98 = 11{,}02 \qquad L_s = \overline{X} + E_{max} = 12 + 0{,}98 = 12{,}98$$

25. A

$$\left.\begin{array}{l} L_S = P + E_{max} \\ L_i = P - E_{max} \end{array}\right\} \Rightarrow L_S - L_i = E_{max} - (E_{max}) \Rightarrow L_S - L_i = 2E_{max} \Rightarrow$$

$$\Rightarrow E_{max} = \frac{L_S - L_i}{2} = \frac{0{,}42 - 0{,}30}{2} = 0{,}06$$

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2015 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Soluciones

Gráfica 1. *Antigüedad*, en años, de los trabajadores discapacitados de una empresa. La antigüedad está agrupada en intervalos y en el eje de abscisas están representados los puntos medios de esos intervalos.

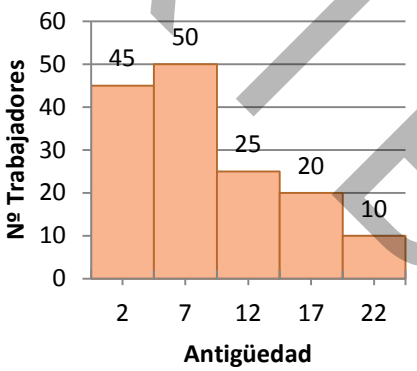
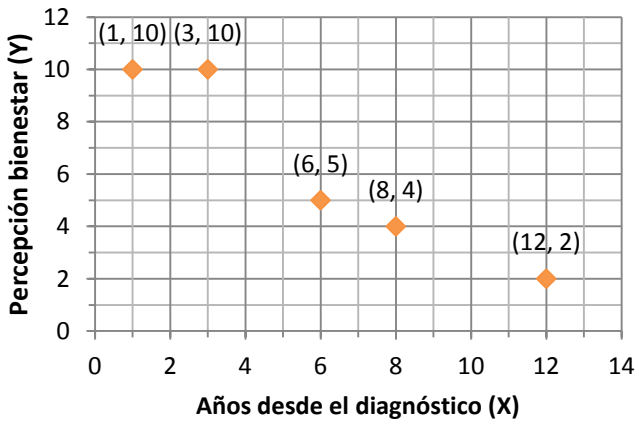


Tabla 1. Datos de la Encuesta de Población Activa en el último trimestre de 2013 por sexo y sector económico (unidades: miles de personas).

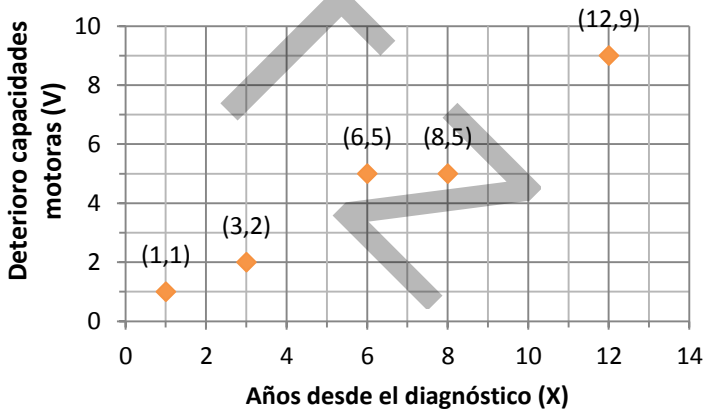
	Agricultura	*Industria*	*Construcción*	*Servicios*
Hombre	597,5	1.727,0	900,9	5.886,7
Mujer	193,4	547,2	77,4	6.828,0

En la Gráfica 2 se muestra la relación de los años transcurridos desde el diagnóstico de una enfermedad degenerativa con la percepción de bienestar percibido (**Gráfica 2.1**) y con el deterioro de las capacidades motoras (**Gráfica 2.2**).

Gráfica 2.1.



Gráfica 2.2.



1. La variable *antigüedad* de la Gráfica 1 es: A) ordinal; B) de intervalo; C) de razón.
2. Atendiendo a los datos de la Gráfica 1, ¿cuáles son los límites exactos del intervalo de mayor antigüedad? A) -0,5 y 4,5; B) 20 y 24; C) 19,5 y 24,5.
3. ¿Qué representación gráfica también se podría utilizar con los datos de la Gráfica 1? A) Diagrama de sectores; B) Polígono de frecuencias; C) Diagrama de barras conjunto.
4. La antigüedad media de los trabajadores discapacitados representados en la Gráfica 1 es: A) 8,67; B) 13; C) 260.

KIBBUTZ

5. Con los datos de la Gráfica 1 se ha seleccionado a un trabajador discapacitado que lleva 16 años trabajando en la empresa. ¿Qué porcentaje de trabajadores discapacitados tienen una antigüedad menor que él? A) 54%; B) 75%; C) 84%.
6. ¿Cuál es la moda de la variable antigüedad descrita en la Gráfica 1? A) 2; B) 7; C) 50.
7. Atendiendo a los datos de la Gráfica 1, la amplitud total es: A) 5; B) 20; C) 25.
8. Con los datos de la Gráfica 2.2, la varianza del deterioro de las capacidades motoras es: A) 2,80; B) 7,84; C) 14,8.
9. La distribución de la antigüedad de los trabajadores discapacitados representados en la Gráfica 1 es: A) simétrica; B) asimétrica positiva; C) asimétrica negativa.
10. Es apropiado analizar la relación entre dos variables utilizando el coeficiente C de contingencia en el caso de: A) la Gráfica 1; B) la Gráfica 2.2; C) la Tabla 1.
11. ¿Qué gráfica representa una relación lineal inversa entre dos variables? A) La Gráfica 1; B) La Gráfica 2.1; C) La Gráfica 2.2.
12. Con los datos de la Gráfica 2.2, el coeficiente de correlación lineal de Pearson entre las variables *años transcurridos desde el diagnóstico* (X) y *deterioro de las capacidades motoras* (V) es: A) -0,96; B) 0,735; C) 0,984.
13. Con los datos de la Gráfica 2.1, ¿qué puntuación le pronosticaremos en bienestar a un paciente diagnosticado hace 10 años? A) 2,96; B) 5,35; C) 19,16.
14. Con los datos de la Tabla 1, si elegimos una persona al azar ¿cuál es la probabilidad de que sea mujer? A) 0,456; B) 0,544; C) 0,750.
15. Con los datos de la Tabla 1, si elegimos una persona al azar, ¿cuál es la probabilidad de que sea mujer y trabaje en el sector de la industria? A) 0,03; B) 0,32; C) 0,56.
16. Con los datos de la Tabla 1, elegimos una persona al azar y resulta que trabaja en el sector Servicios, ¿cuál es la probabilidad de que sea una mujer? A) 0,407; B) 0,463; C) 0,537.
17. Considerando un experimento aleatorio consistente en lanzar una moneda (no trucada) al aire dos veces y definiendo la variable aleatoria X como “número de caras obtenidas”. ¿Cuántos resultados posibles tiene el espacio muestral? A) 2; B) 3; C) 4.
18. Sabiendo que la enfermedad de Alzheimer afecta a una de cada diez personas jubiladas, si seleccionamos a 10 jubilados al azar, ¿qué es más probable? A) Que ninguno sufra Alzheimer; B) Que uno de ellos sufra Alzheimer; C) Que dos de ellos sufran Alzheimer.
19. Con los datos de prevalencia de la enfermedad de Alzheimer del ejercicio anterior, ¿cuál es la probabilidad de que 3 ó más jubilados padezcan esta enfermedad? A) 0,0702; B) 0,720; C) 0,8063.
20. Las puntuaciones en motivación de un grupo de 100 estudiantes se distribuyen normalmente con media 8 y desviación típica 2. ¿Cuántos estudiantes han obtenido una puntuación en motivación por debajo de la media del grupo? A) 10; B) 50; C) 75.
21. Si una variable sigue una distribución *t* de Student con 16 grados de libertad, ¿cuál es la probabilidad de obtener valores menores o iguales que -2,12? A) 0,025; B) 0,050; C) 0,975.
22. En una distribución F con 5 grados de libertad en el numerador y 5 grados de libertad en el denominador, ¿cuál es el valor del percentil 10? A) 0,2896; B) 0,3453; C) 3,453.
23. Una distribución muestral de la media es: A) la distribución de probabilidad de una característica de la muestra; B) la distribución de probabilidad del estadístico media calculado en todas las posibles muestras de igual tamaño extraídas de una población; C) la distribución de probabilidad del parámetro media calculado en las muestras representativas.
24. Los límites inferior y superior del intervalo de confianza del 99%, construido sobre la media de 100 observaciones son 52 y 58. ¿Cuál será la media obtenida en la muestra? A) 50; B) 55; C) 60.
25. En una muestra aleatoria de 300 parados de una determinada comunidad autónoma, 6 de cada 10 son mujeres. Para un nivel de confianza de 0,95, ¿entre qué límites se estima que esté la proporción de mujeres paradas de toda la comunidad autónoma? A) 0,45 y 0,55; B) 0,36 y 0,72; C) 0, 55 y 0,66.

TUBEGITN

SOLUCIONES:

1. C
2. C
3. B
4. A

Antigüedad	n_i	n_iX_i
22	10	220
17	20	340
12	25	300
7	50	350
2	45	90
	150	1300

$$\overline{X} = \frac{\sum n_i X_i}{n} = \frac{1300}{150} = 8{,}67$$

5. C

Antigüedad	n_i	n_a
20-24	10	150
15-19	20	140
10-14	25	120
5-9	50	95
0-4	45	45

$$k = \left[\frac{\frac{(P_k - L_i)\cdot n_c}{I} + n_d}{n}\right]\cdot 100 = \left[\frac{\frac{(16-14{,}5)\cdot 20}{5} + 120}{150}\right]\cdot 100 = 84$$

6. B

El intervalo modal es 5-9, por lo que la moda es $M_o = 7$.

7. C

$A_T = X_{máx} - X_{min} = 24{,}5 - (-0{,}5) = 25$

8. B

Años desde el diagnóstico (X)	Deterioro cap. motoras (V)	V^2
1	1	1
3	2	4
6	5	25
8	5	25
12	9	81
	22	136

$$\overline{V} = \frac{22}{5} = 4{,}4$$

$$S_V^2 = \frac{136}{5} - 4{,}4^2 = 27{,}2 - 19{,}36 = 7{,}84$$

9. B

Observando la Gráfica 1 se aprecia que la mayoría de las puntuaciones (las mayores frecuencias) se encuentran en los valores bajos de antigüedad, por lo que se trata de una distribución asimétrica positiva. Utilizando el índice de asimetría de Pearson llegamos a esta misma conclusión:

$$A_S = \frac{\overline{X} - Mo}{S_X} = \frac{8{,}67 - 7}{S_x} = \frac{1{,}37}{S_x}$$

Dado que la desviación típica siempre es positiva (o nula) el resultado de esta división siempre será un número positivo.

10. C

KIBBUTZ

11. B

Observando la Gráfica 2.1 se observa que a mayores valores en X le corresponden menores valores en Y y viceversa.

12. C

Paciente	Años desde el diagnóstico (X)	Deterioro cap. motoras (V)	XV	X^2	V^2
1	1	1	1	1	1
2	3	2	6	9	4
3	6	5	30	36	25
4	8	5	40	64	25
5	12	9	108	144	81
∑	30	22	185	254	136

$$r_{XV} = \frac{n\sum XV - \sum X \sum V}{\sqrt{n\sum X^2 - \left(\sum X\right)^2}\sqrt{n\sum V^2 - \left(\sum V\right)^2}} = \frac{5\times 185 - 30\times 22}{\sqrt{5\times 254 - 30^2}\sqrt{5\times 136 - 22^2}} =$$

$$= \frac{925-660}{19{,}235\times 14} = \frac{265}{269.29} = 0{,}984$$

13. A

Paciente	Años desde el diagnóstico (X)	Percepción bienestar (Y)	XY	X^2
1	1	10	10	1
2	3	10	30	9
3	6	5	30	36
4	8	4	32	64
5	12	2	24	144
∑	30	31	126	254

$$Y' = a + bX$$

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2} = \frac{5\times 126 - 30\times 31}{5\times 254 - 30^2} = \frac{630-930}{1270-900} = \frac{-300}{370} = -0{,}81$$

$$a = \overline{Y} - b\overline{X} = \frac{31}{5} - (-0{,}81)\frac{30}{5} = 6{,}2 + 4.86 = 11.06$$

$$Y' = 11{,}06 - 0{,}81X = 11{,}06 - 0{,}81\times 10 = 2{,}96$$

14. A

	Agricultura	Industria	Construcción	Servicios	
Hombres	597500	1727000	900900	5886700	9112100
Mujeres	193400	547200	77400	6828000	7646000
	790900	2274200	978300	12714700	16758100

$$P(\text{Mujer}) = \frac{7646000}{16758100} = 0{,}456$$

15. A

	Agricultura	Industria	Construcción	Servicios	
Hombres	597500	1727000	900900	5886700	9112100
Mujeres	193400	547200	77400	6828000	7646000
	790900	2274200	978300	12714700	16758100

KIBBUTZ

$$P(\text{Mujer} \cap \text{Industria}) = \frac{547200}{16758100} = 0{,}03$$

16. C

	Agricultura	Industria	Construcción	Servicios	
Hombres	597500	1727000	900900	5886700	9112100
Mujeres	193400	547200	77400	6828000	7646000
	790900	2274200	978300	12714700	16758100

$$P(\text{mujer} \mid \text{servicios}) = \frac{6828000}{12714700} = 0{,}537$$

17. C

Resultados posibles: cara y cara, cara y cruz, cruz y cara, cruz y cruz.

18. B

P(X=0) = 0,3487 (Tabla I con n=10, p=0,1 y x=0)

P(X=1) = 0,3874 (Tabla I con n=10, p=0,1 y x=1)

P(X=2) = 0,1937 (Tabla I con n=10, p=0,1 y x=2)

Es más probable que uno de ellos padezca la enfermedad de Alzheimer.

19. A

P(X ≥ 3) = 1 – P(X ≤ 2) = 1 - 0,9298 = 0,0702. (Tabla II con n=10, p=0,10 y x=2).

20. B

Dado que la puntuación 8 es la media del grupo, la puntuación típica tiene que ser 0.

$$Z = \frac{8-8}{2} = 0$$

P(Z≤0) = 0,5000 (Tabla III o IV)

0,500×100 = 50.

21. A

P(X ≤ -2,12) = P(X > 2,12) = 1 – P(X ≤ 2,12) = 1 - 0,975 = 0,025 (Tabla VI)

22. A

$${}_{0,10}F_{5,5} = \frac{1}{{}_{0,90}F_{5,5}} = \frac{1}{3{,}453} = 0{,}2896 \text{ (Tabla VII)}$$

23. B

24. B

$$\frac{52+58}{2} = 55$$

25. C

$$L_i = 0{,}60 - 1{,}96\sqrt{\frac{0{,}60\cdot(1-0{,}60)}{300}} = 0{,}60 - 1{,}96\cdot 0{,}028 = 0{,}60 - 0{,}055 = 0{,}545 \approx 0{,}55$$

$$L_s = 0{,}60 + 1{,}96\sqrt{\frac{0{,}60\cdot(1-0{,}60)}{300}} = 0{,}60 + 1{,}96\cdot 0{,}028 = 0{,}60 + 0{,}055 = 0{,}655 \approx 0{,}66$$

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2015 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO C DURACION: 2 HORAS

Soluciones

Gráfica 1: Distribución de un grupo de 800 profesionales sanitarios colegiados en 2012 por titulación (200 en cada una de las titulaciones) y sexo.

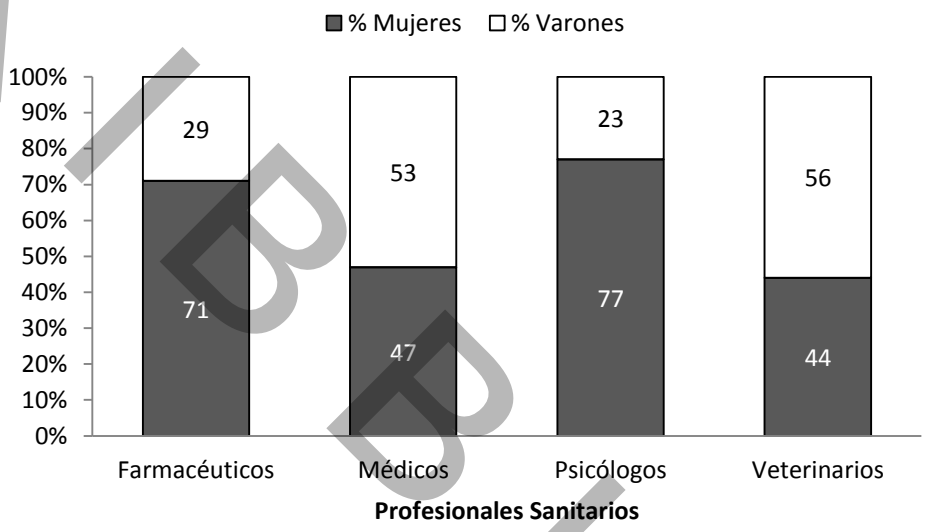


Tabla 1. Edad (X) en un grupo de 1500 profesores de universidad

X	n_i
30 ó menos	85
31-40	150
41-50	280
51-60	610
61-70	332
71 ó más	43
	1500

Tabla 2. Para dos variables cuantitativas, X e Y, disponemos de los siguientes datos:

X	Y	
$\overline{X} = 100$	$\overline{Y} = 8$	$Y' = 0{,}1X - 2$
$S_X = 16$	$S_Y = 2$	

1. En el eje de ordenadas de la Gráfica 1 aparece: A) la frecuencia absoluta; B) la frecuencia absoluta acumulada; C) el porcentaje.
2. Teniendo en cuenta la Gráfica 1, el número de colegiados varones es mayor que el número de colegiados mujeres en: A) psicólogos; B) farmacéuticos; C) médicos y veterinarios
3. La variable "Profesionales Sanitarios", de la Gráfica 1, es: A) dicotómica; B) cuantitativa; C) politómica.
4. La representación de la Gráfica 1 se denomina: A) histograma; B) diagrama de sectores; C) diagrama de barras conjunto.
5. Con los datos de la Tabla 1, la moda es igual a: A) 30,5; B) 55,5; C) 680.
6. Con los datos de la Tabla 1, ¿a qué percentil corresponde el valor 45,5?: A) 25; B) 50; C) 75.

KIBUCH

7. Con los datos de la Tabla 1, el tercer cuartil es igual a: A) 50; B) 75; C) 60,5.
8. Con los datos de la Tabla 1, ¿qué índice utilizaremos para estudiar la variabilidad?: A) el coeficiente de variación; B) la amplitud semi-intercuartil; C) la varianza.
9. Teniendo en cuenta la Tabla 2, ¿cuál de las dos variables (X e Y) presenta mayor variabilidad?: A) X; B) Y; C) Ambas tienen la misma variabilidad.
10. Señale una de las propiedades de las puntuaciones típicas: A) su media es 1; B) su desviación típica es 0; C) su varianza es 1.
11. Para calcular el grado de asociación entre las variables "profesión sanitaria" y "género", recogidas en la Gráfica 1, utilizaremos: A) el coeficiente de contingencia C; B) el coeficiente de correlación lineal de Pearson; C) la covarianza.
12. Teniendo en cuenta los datos de la Tabla 2, ¿qué puntuación directa pronosticaremos a una persona con una puntuación en X igual a 100?: A) 0,1; B) 10; C) 8.
13. Teniendo en cuenta los datos de la Tabla 2, ¿cuánto vale el coeficiente de correlación de Pearson entre X e Y?: A) 0,1; B) 0,8; C) 0,9.
14. Teniendo en cuenta los datos de la Tabla 2, la varianza de los errores cometidos al pronosticar las puntuaciones en Y utilizando la recta de regresión es: A) 1,44; B) 2,56; C) 5,42.
15. Teniendo en cuenta la representación de la Gráfica 1, elegida una persona al azar ha resultado ser profesional de Farmacia ¿cuál es la probabilidad de que sea mujer?: A) 0,29; B) 0,50; C) 0,71.
16. Teniendo en cuenta la representación de la Gráfica 1, si elegimos una persona al azar ¿cuál es la probabilidad de que sea profesional sanitario de la Psicología?: A) 0,125; B) 0,25; C) 0,50.
17. Se está analizando la eficacia de tres fármacos experimentales para la remisión de una determinada enfermedad en 10 pacientes, habiendo cuatro pacientes medicados con el fármaco 1, tres pacientes con el fármaco 2 y tres con el fármaco 3. Los resultados muestran que la probabilidad de curación de los pacientes que han tomado el fármaco 1 es 0,4, la de los que han tomado el fármaco 2 es 0,5 y la de los que han tomado el fármaco 3 es 0,6. ¿Cuál es la probabilidad de curación? A) 0,33; B) 0,40; C) 0,49.
18. Una variable aleatoria discreta X toma los valores 0, 1 y 2. Sabemos que P(X=0)=0,2 y P(X=1)=0,3, ¿cuánto vale la media o esperanza matemática de X?: A) 0,5; B) 1,3; C) 2.
19. En una Facultad de Psicología, el 70% de los estudiantes son mujeres. Si elegimos al azar dos estudiantes, ¿cuál es la probabilidad de que uno de ellos sea mujer?: A) 0,30; B) 0,42; C) 0,70.
20. Teniendo en cuenta los datos de la pregunta anterior, elegimos al azar 10 estudiantes ¿cuál es la probabilidad de que más de seis sean mujeres?: A) 0,6496; B) 0,5328; C) 0,2637.
21. En una distribución normal tipificada, si $P(Z \leq z) = 0{,}5$, entonces z es igual a: A) -1; B) 1; C) 0.
22. Las puntuaciones en CI (Cociente Intelectual) se distribuyen normalmente con $\mu = 100$ y $\sigma = 15$. Si a una persona le corresponde el Percentil 67, su puntuación en CI es igual a: A) 67; B) 106,6; C) 115.
23. Si una variable X se ajusta a la distribución *t* de Student con 30 grados de libertad, su percentil 20 es igual a: A) 0,854; B) 0,146; C) -0,854
24. Para una muestra de 100 sujetos, una varianza poblacional igual a 9 y a un nivel de confianza de 0,95, el error máximo cometido para estimar la media poblacional es: A) 0,588; B) 0,95; C) 1,746.
25. Para estimar la proporción de mujeres en la universidad hemos obtenido una muestra de 100 estudiantes universitarios y hemos comprobado que 60 de ellos son mujeres. A un nivel de confianza del 99% ¿entre que valores estimamos que se encuentra la proporción de mujeres universitarias?: A) 0,504 y 0,696; B) 0,474 y 0,726; C) 0,54 y 0,66.

SOLUCIONES:

1. C
2. C
3. C
4. C
5. B

$$Mo = \frac{51+60}{2} = 55{,}5$$

6. A

X	n_i	n_a
71 ó más	43	1500
61-70	332	1457
51-60	610	1125
41-50	280	515
31-40	150	235
30 ó menos	85	85
	1500	

$$k = \left[\frac{\frac{(P_k - L_i)\cdot n_c}{I} + n_d}{n}\right]\cdot 100 = \left[\frac{\frac{(45{,}5 - 40{,}5)\cdot 280}{10} + 235}{1500}\right]\cdot 100 = \frac{140+235}{1500}\cdot 100 = 25$$

7. C

X	n_i	n_a
71 ó más	43	1500
61-70	332	1457
51-60	610	1125
41-50	280	515
31-40	150	235
30 ó menos	85	85
	1500	

$$\frac{nk}{100} = \frac{1500\cdot 75}{100} = 1125$$

$$P_{75} = 50{,}5 + \left(\frac{\frac{1500\cdot 75}{100} - 515}{610}\right)\cdot 10 = 50{,}5 + 10 = 60{,}5$$

8. B
9. B

KIBBUTZ

$$CV_X = \frac{16}{100} \cdot 100 = 16$$

$$CV_Y = \frac{2}{8} \cdot 100 = 25$$

10. C

11. A

12. C

$$Y' = 0{,}1X - 2 = 0{,}1 \cdot 100 - 2 = 10 - 2 = 8$$

13. B

$$Y' = bX + a$$

$$b = r_{XY} \frac{S_Y}{S_X} \Rightarrow 0{,}1 = r_{XY} \frac{2}{16} \Rightarrow 0{,}1 = r_{XY} \cdot 0{,}125 \Rightarrow r_{XY} = \frac{0{,}1}{0{,}125} = 0{,}8$$

14. A

$$r_{XY}^2 = 1 - \frac{S_{YX}^2}{S_Y^2} \Rightarrow 0{,}64 = 1 - \frac{S_{YX}^2}{4} \Rightarrow 2{,}56 = 4 - S_{YX}^2 \Rightarrow S_{YX}^2 = 4 - 2{,}56 = 1{,}44$$

15. C

	Hombre (H)	Mujer (M)	
Farmacéuticos (F)	58	142	200
Médicos (M)	106	94	200
Psicólogos (Ps)	46	154	200
Veterinarios (V)	112	88	200
	322	478	800

$$P(M/F) = \frac{142}{200} = 0{,}71$$

16. B

	Hombre (H)	Mujer (M)	
Farmacéuticos (F)	58	142	200
Médicos (M)	106	94	200
Psicólogos (Ps)	46	154	200
Veterinarios (V)	112	88	200
	322	478	800

$$P(Ps) = \frac{200}{800} = \frac{1}{4} = 0{,}25$$

17. C

$$P(F_1) = \frac{4}{10} = 0{,}4 \qquad P(\text{curación}|F_1) = 0{,}4$$

$$P(F_2) = \frac{3}{10} = 0{,}3 \qquad P(\text{curación}|F_2) = 0{,}5$$

$$P(F_3) = \frac{3}{10} = 0{,}3 \qquad P(\text{curación}|F_3) = 0{,}6$$

$$P(\text{curación}) = P(F_1 \cap \text{curación}) + P(F_2 \cap \text{curación}) + P(F_3 \cap \text{curación}) =$$

$$= P(F_1) \cdot P(\text{curación}|F_1) + P(F_2) \cdot P(\text{curación}|F_2) + P(F_3) \cdot P(\text{curación}|F_3) =$$

$$= 0{,}4 \times 0{,}4 + 0{,}3 \times 0{,}5 + 0{,}3 \times 0{,}6 = 0{,}16 + 0{,}15 + 0{,}18 = 0{,}49$$

18. B

KIBBUTZ

x	f(x)	x·f(x)
0	0, 2	0
1	0,3	0,3
2	0,5	1
		1,3

$$\overline{X} = \sum x \cdot f(x) = 1{,}3$$

19. B

Tabla I con n=2, p=0,3 y x=1

20. A

La probabilidad de que más de 6 sean mujeres es igual a la probabilidad de que 3 ó menos sean hombres. Por tanto: 0,6496

Tabla II con n=10, p=0,3 y x=3

21. C

P(Z ≤ z) = 0,5 → z = 0 (Tabla IV)

22. B

Al Percentil 67 le corresponde una z = 0,44 (Tabla IV)

$$z = \frac{P_{67} - \mu}{\sigma} \Rightarrow 0{,}44 = \frac{P_{67} - 100}{15} \Rightarrow 6{,}6 = P_{67} - 100 \Rightarrow P_{67} = 6{,}6 + 100 = 106{,}6$$

23. C

Tabla VI: distribución t de Student

24. A

$$\text{n.c.} = 0{,}95 \quad \rightarrow \quad Z_{1-\alpha/2} = Z_{0{,}975} = 1{,}96 \quad (\text{Tabla IV})$$

$$E_{max} = Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 1{,}96 \cdot \frac{3}{\sqrt{100}} = 1{,}96 \cdot \frac{3}{10} = 0{,}588$$

25. B

$$P \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{P \cdot (1-P)}{n}}. = 0{,}6 \pm 2{,}58 \cdot \sqrt{\frac{0{,}6 \cdot 0{,}4}{100}} = 0{,}6 \pm 0{,}126 = \begin{cases} 0{,}726 \\ 0{,}474 \end{cases}$$

KIBBUTZ

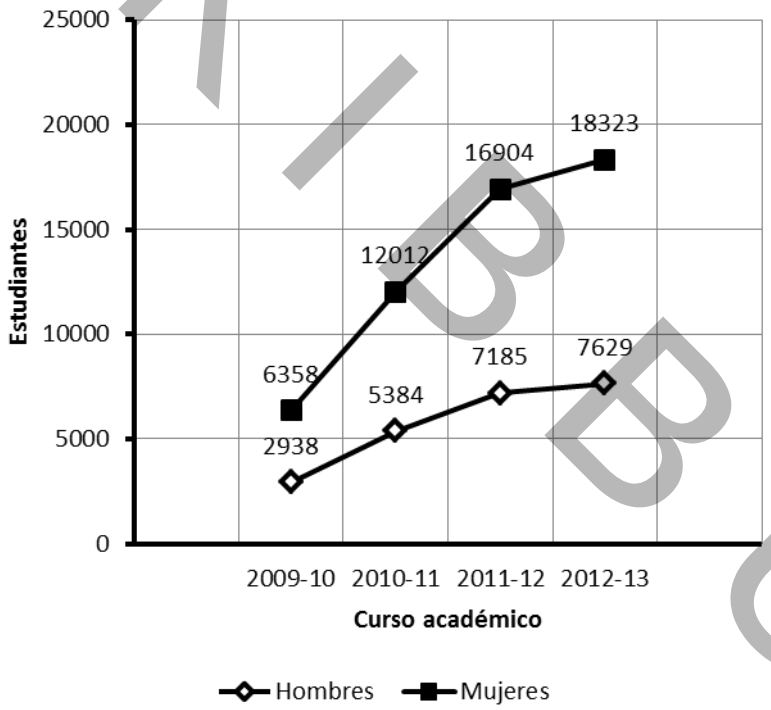
INTRODUCCIÓN AL ANÁLISIS DE DATOS
SEPTIEMBRE 2015 Reserva Código asignatura: 62011037
EXAMEN TIPO TEST MODELO D DURACION: 2 HORAS

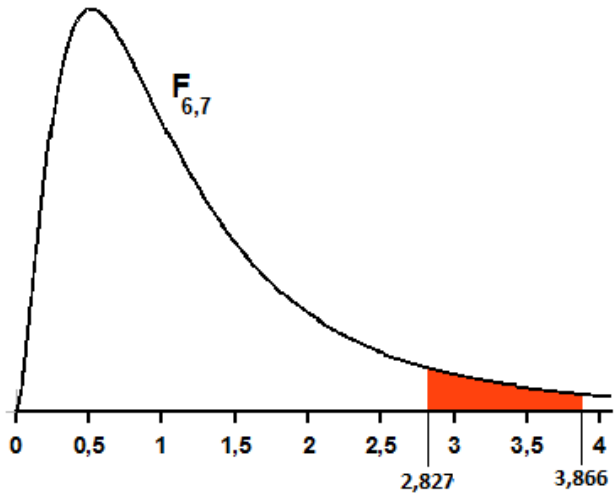
Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

SOLUCIONES

Gráfica 1: Número de estudiantes (Hombres y Mujeres) matriculados en el Grado en Psicología durante cuatro cursos académicos.





Gráfica 2. Distribución *F de Snedecor* con 6 grados de libertad en el numerador y 7 grados de libertad en el denominador.

Tabla 1. Número de hijos, X, en un grupo de 200 mujeres:

X_i	n_i
4	10
3	40
2	50
1	60
0	40

Tabla 2. Para dos variables cuantitativas, X e Y, disponemos de los siguientes datos para una muestra con n=200:

X	Y	XY
$\overline{X}=8$	$\overline{Y}=100$	$r_{XY}=0{,}9$
$S_X=2$	$S_Y=10$	

1. En el eje de ordenadas de la Gráfica 1 aparece: A) la frecuencia relativa; **B) la frecuencia absoluta**; C) la frecuencia absoluta acumulada.
2. El nivel de medida de la variable "número de hijos", de la Tabla 1, es: **A) de razón**; B) de intervalo; C) ordinal.
3. Teniendo en cuenta la Gráfica 1, el número de mujeres matriculadas a lo largo de los cuatro cursos académicos: A) ha ido aumentando en menor medida que el número de hombres; B) ha ido disminuyendo; **C) ha ido aumentando en mayor medida que el número de hombres.**
4. La variable "número de hijos", de la Tabla 1, es: A) dicotómica; **B) discreta**; C) continua.

KIBBUTZ

5. Teniendo en cuenta la Gráfica 1, la media de hombres matriculados en los cuatro cursos es: A) 13399; B) 5384; **C) 5784.**
6. Con los datos de la Tabla 1, la media es igual a: A) 1; **B) 1,6**; C) 2.
7. Con los datos de la Tabla 1, ¿a qué percentil corresponde el valor 1,5?: A) 40; **B) 50**; C) 60.
8. Con los datos de la Tabla 1, el tercer cuartil es igual a: A) 1,25; B) 1,75**; C) 2,5.**
9. Con los datos de la Tabla 1, la varianza es igual a: A) 1,15; **B) 1,34**; C) 2,47.
10. Con los datos de la Tabla 1, la distribución de la variable "número de hijos" es: **A) asimétrica positiva**; B) simétrica; C) asimétrica negativa.
11. Para calcular el grado de asociación entre dos variables cualitativas utilizamos: A) la covarianza; B) el coeficiente de correlación lineal de Pearson; **C) el coeficiente de contingencia C.**
12. En la Tabla 2, la covarianza entre X e Y es igual a: A) 0,9; B) 8; **C) 18.**
13. Teniendo en cuenta la Tabla 2, la recta de regresión que permite pronosticar la variable *Y* en función de la variable X es: **A)** $Y' = 4{,}5X + 64$; B) $Y' = -0{,}5X + 70$; C) $Y' = 64X + 4{,}5$.
14. Sabemos que, para dos sucesos independientes A y B, P(A)=0,4 y P(B)=0,6. Entonces $P(A \cap B)$ es igual a: A) 0,2; B) 1; **C) 0,24.**
15. En el Curso de Acceso de un Centro Asociado hay 20 alumnos y 20 alumnas. De ellos, tanto en hombres como en mujeres, el 75% han elegido la asignatura de inglés y el 25% restante otro idioma. Si elegimos un estudiante al azar ¿cuál es la probabilidad de que sea mujer y estudie inglés?: **A) 0,375**; B) 0,50; C) 0,75.
16. Teniendo en cuenta los datos de la pregunta anterior, elegimos un estudiante al azar y resulta que estudia inglés, ¿cuál es la probabilidad de que sea mujer?: A) 0,375; **B) 0,50**; C) 0,75.
17. Una variable aleatoria discreta X toma los valores 1 y 2. Sabemos que P(X=1)=0,3 y P(X=2)=0,7. ¿cuánto vale la media o esperanza matemática de X?: A) 0,5; B) 1,5; **C) 1,7.**
18. Uno de cada cuatro niños madrileños van a un colegio privado o concertado. Si elegimos al azar dos niños, ¿cuál es la probabilidad de que uno de ellos asista a un colegio privado o concertado?: A) 0,25; **B) 0,375**; C) 0,50.
19. Teniendo en cuenta los datos de la pregunta anterior, elegimos al azar 5 niños madrileños ¿cuál es la probabilidad de que más de dos asistan a un colegio público?: **A) 0,8965**; B) 0,6328; C) 0,2637.
20. En una distribución normal tipificada, si $P(Z \le z) = 0{,}9761$, entonces z es igual a: A) 0,08; B) 0,1035; **C) 1,98.**
21. Una asociación de intelectuales jóvenes sólo admite a jóvenes cuya puntuación en un test de *inteligencia* X sea superior al percentil 67 (P_{67}) de la distribución de las puntuaciones en el test. Sabiendo que las puntuaciones en el test se distribuyen *normalmente* con μ = 100 y σ = 15, para que un joven sea admitido debe obtener necesariamente una puntuación: **A) mayor que 106,60**; B) mayor que 110,20; C) mayor que 115.
22. Según la distribución de la Gráfica 2, la probabilidad de obtener un valor entre 2,827 y 3,866 (zona coloreada de la gráfica) es igual a: A) 0,01; B) 0,025; **C) 0,05.**
23. Para un nivel de confianza de 0,95, σ = 3 y un error de estimación máximo de 1,2, el tamaño de muestra requerido para estimar la *autoestima* media de una población resulta ser n = 24. Sabiendo que la variable autoestima se distribuye normalmente en la población, ¿en cuántos sujetos necesitaríamos incrementar la muestra de n = 24 si queremos asumir un error de estimación máximo de 1?: A) 35; **B) 11;** C) 0.
24. Hemos aplicado a una muestra aleatoria de 400 toxicómanos severos un test para cuantificar el *deterioro cognitivo* de este colectivo, obteniendo una media igual a 7 y una cuasivarianza de 4. Para un nivel de confianza del 95%, ¿cuáles son los límites entre los cuales se espera esté el *deterioro cognitivo* medio de la población de toxicómanos severos?: A) 6,608 y 7,392; B) 6,742 y 7,258; **C) 6,804 y 7,196.**
25. Con un tamaño de muestra n=100, el error típico de la proporción tomará su valor máximo si la proporción P es igual a: A) 0; **B) 0,5**; C) 1.

KEY BCN

SOLUCIONES:

1. B
2. A
3. C
4. B
5. C

$$\frac{29388+5384+7185+7629}{4}=\frac{23136}{4}=5784$$

6. B

X_i	n_i	$n_i X_i$
4	10	40
3	40	120
2	50	100
1	60	60
0	40	0
	200	320

$$\overline{X}=\frac{320}{200}=1{,}6$$

7. B

X	n_i	n_a
4	10	200
3	40	190
2	50	150
1	60	100
0	40	40
	200	

$$k=\left[\frac{\frac{(P_k-L_i)\cdot n_c}{I}+n_d}{n}\right]\cdot 100=\left[\frac{\frac{(1{,}5-0{,}5)\cdot 60}{1}+40}{200}\right]\cdot 100=0{,}5\cdot 100=50$$

8. C

X	n_i	n_a
4	10	200
3	40	190
2	50	150
1	60	100
0	40	40
	200	

KIBBUTZ

$$P_{75} = 1{,}5 + \left(\frac{\frac{200 \cdot 75}{100} - 100}{50} \right) \cdot 1 = 1{,}5 + 1 = 2{,}5$$

9. B

X_i	n_i	$n_i X_i$	$n_i X_i^2$
4	10	40	160
3	40	120	360
2	50	100	200
1	60	60	60
0	40	0	0
	200	320	780

$$S_X^2 = \frac{780}{200} - 1{,}6^2 = 3{,}9 - 2{,}56 = 1{,}34$$

10. A

$$A_S = \frac{\overline{X} - Mo}{S_X} = \frac{1{,}6 - 1}{\sqrt{1{,}34}} = 0{,}52 \Rightarrow \text{Asimetría Positiva}$$

11. C

12. C

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} \quad \Rightarrow \quad S_{XY} = r_{XY} S_X S_Y = 0{,}9 \cdot 2 \cdot 10 = 18$$

13. A

$$Y' = bX + a$$

$$b = r_{XY} \frac{S_Y}{S_X} = 0{,}9 \frac{10}{2} = 4{,}5 \qquad a = \overline{Y} - b\overline{X} = 100 - (4{,}5 \cdot 8) = 64$$

$$Y' = 4{,}5X + 64$$

14. C

$$P(A \cap B) = P(A)P(B) = 0{,}4 \cdot 0{,}6 = 0{,}24$$

15. A

	Inglés	Otro	
Hombre	15	5	20
Mujer	15	5	20
	30	10	40

$$P(M \cap I) = \frac{15}{40} = 0{,}375$$

16. B

	Inglés	Otro	
Hombre	15	5	20
Mujer	15	5	20
	30	10	40

$$P(M/I) = \frac{15}{30} = 0{,}5$$

17. C

KIBBUTZ

x	f(x)	x·f(x)
1	0,3	0,3
2	0,7	1,4
		1,7

$$\mu_X = \sum x \cdot f(x) = 1{,}7$$

18. B

Tabla I con n=2, p=0,25 y x=1

19. A

La probabilidad de que más de 2 asistan a un colegio público es igual a la probabilidad de que 2 ó menos asistan a un colegio privado o concertado. Por tanto: 0,8965

Tabla II con n=5, p=0,25 y x=2

20. C

P(Z ≤ 1,98) = 0,9761 → z = 1,98 (Tabla IV)

21. A

A la probabilidad 0,67 le corresponde una z = 0,44 (Tabla IV)

$$X = z\sigma + \mu \rightarrow X = 0{,}44 \cdot (15) + 100 = 106{,}6$$

22. C

$$P(F \le 2{,}827) = 0{,}90 \quad (\text{Tabla VII})$$

$$P(F \le 3{,}866) = 0{,}95 \quad (\text{Tabla VII})$$

$$P(2{,}827 \le F \le 3{,}866) = P(F \le 3{,}866) - P(F \le 2{,}827) = 0{,}95 - 0{,}90 = 0{,}05$$

23. B

$$n.c. = 0{,}95 \quad \rightarrow \quad Z_{1-\alpha/2} = Z_{0{,}975} = 1{,}96 \quad (\text{Tabla IV})$$

$$n = \frac{z_{1-\alpha/2}^2 \sigma^2}{E_{max}^2} = \frac{1{,}96^2 \cdot 3^2}{1^2} = 34{,}57 = 35$$

35-24=11 sujetos

24. C

n=400 (Aproximación de la distribución t de Student a la normal)

$$n.c. = 0{,}95 \quad \rightarrow \quad z_{1-\alpha/2} = z_{0{,}975} = 1{,}96 \quad (\text{Tabla IV})$$

$$E_{max} = z_{1-\alpha/2} \frac{S_{n-1}}{\sqrt{n}} \rightarrow E_{max} = 1{,}96 \frac{2}{20} = 0{,}196$$

$$L_i = \overline{X} - E_{max} = 7 - 0{,}196 = 6{,}804 \qquad L_s = \overline{X} + E_{max} = 7 + 0{,}196 = 7{,}196$$

25. B

Para un n dado, el error típico de la proporción toma su valor máximo cuando P es igual a 0,5. Así, para n=100, el valor máximo que puede tomar el error típico de la proporción es:

$$\sqrt{\frac{P(1-P)}{n}} \rightarrow \sqrt{\frac{P(1-P)}{100}} \rightarrow \sqrt{\frac{0{,}5(1-0{,}5)}{100}} = 0{,}05$$

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2017 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO F DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

SOLUCIONES

Gráfica 1: Número de personas detenidas por corrupción y número de delitos conocidos por corrupción en España en los últimos cinco años (según el diario El País de 19-2-2016).

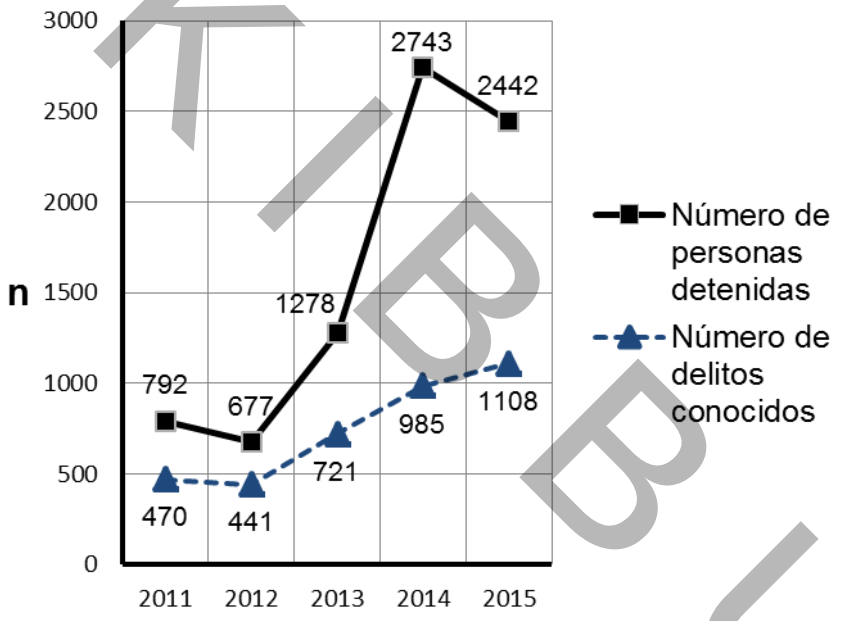


Tabla 1. Para pronosticar el rendimiento en un *Trabajo manual (Y)* a partir de las puntuaciones en un test de Impulsividad (X) disponemos de los siguientes datos de cinco escolares:

Escolar	***X***	***Y***
1	35	6
2	10	8
3	5	10
4	40	2
5	25	4

Se sabe que la pendiente de la recta de regresión de Y sobre X es igual a -0,18.

Gráfica 2: Número de chicos y chicas que sufren Acoso Físico y Ciberacoso en una muestra de 30 estudiantes de bachillerato acosados.

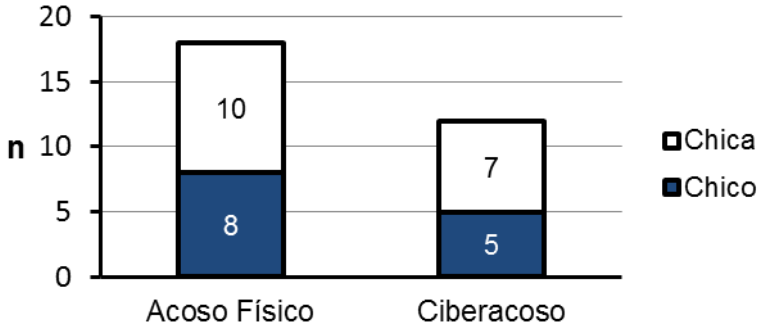


Tabla 2. Función de probabilidad, f(x), de la variable aleatoria discreta X.

x	***f(x)***
0	0,10
2	0,20
4	0,30
6	0,40

1. Con los datos de la Gráfica 1, el nivel de medida de la variable "número de personas detenidas por corrupción" es: A) de intervalo; B) ordinal; **C) de razón**.
2. Para los datos de la Tabla 1, ¿qué grafico se utiliza para representar la relación entre X e Y? A) Histograma; **B) Diagrama de dispersión**; C) Pictograma.
3. La variable "acoso" representada en el eje de abscisas de la Gráfica 2 es: A) politómica; **B) dicotómica**; C) cuantitativa discreta.
4. Para estudiar la tendencia central en variables cualitativas dicotómicas utilizamos: A) la media; B) la mediana; **C) la moda**.

KIBBUTZ

5. Con los datos de la Gráfica 1, la media de personas detenidas por corrupción a lo largo de los cinco años es: A) 970,76; B) 1278,54; **C) 1586,40.**
6. Según los datos de la Tabla 1, el percentil 50 de la variable "Impulsividad (X)" es: A) 17,5; **B) 25**; C) 35.
7. La cuasivarianza de la variable "Impulsividad (X)" de la Tabla 1 es: **A) 232,5**; B) 186,1; C) 10.
8. Con los datos de la Tabla 1, ¿qué puntuación típica le corresponde al escolar 3 en la variable "trabajo manual (Y)"?: A) 1,02; **B) 1,41**; C) 2,39.
9. Teniendo en cuenta los datos de la Tabla 1, ¿cuál es el índice de asimetría de Pearson del rendimiento en un trabajo manual (Y)? A) 0; B) 1; **C) No se puede calcular**.
10. Según los datos de la Tabla 1, en la recta de regresión de Y sobre X, la proporción de varianza de Y explicada por la varianza de X está entre: A) 0,50 y 0,55; **B) 0,75 y 0,80**; C) 0,85 y 0,90.
11. Con los datos de la Tabla 1, la ordenada en el origen de la ecuación de la recta de regresión para pronosticar la variable Y a partir de la variable X es: **A) 10,14**; B) 8,32; C) 7,51.
12. Considerando los datos de la Tabla 1, ¿qué puntuación en Y pronosticaremos a un escolar que ha obtenido una puntuación de X=23? **A) 6,0**; B) 4,2; C) 3,4.
13. Con los datos de la Gráfica 2, ¿podemos calcular el valor máximo que puede alcanzar el coeficiente C de contingencia entre las dos variables? **A) Sí**; B) No; C) Falta información.
14. En un espacio muestral hay dos sucesos A y B. Sabemos que P(A)=0,40, P(B)=0,30 y que $P(A \cap B) = 0{,}20$. ¿Cuál es la probabilidad de la unión de los sucesos A y B? A) 0,90; B) 0,70; **C) 0,50**.
15. Atendiendo a los datos de la Gráfica 2, si se ha elegido al azar un estudiante y resulta que es una chica, ¿cuál es la probabilidad de que sufra ciberacoso?: A) 0,23; **B) 0,41**;C) 0,57.
16. Con los datos de la Gráfica 2, si seleccionamos un estudiante al azar, ¿cuál es la probabilidad de que sea chico y sufra acoso físico?: **A) 0,27**; B) 0,43; C) 0,60.
17. Considerando un experimento aleatorio consistente en lanzar una moneda (no trucada) al aire tres veces y definiendo la variable aleatoria X como "número de cruces obtenidas", ¿cuántos resultados posibles tiene el espacio muestral?: **A) 8**; B) 4; C) 3.
18. Con los datos de la Tabla 2, la esperanza matemática de la variable aleatoria X es: A) 2; B) 3; **C) 4**.
19. Considerando los datos de la Tabla 2, la probabilidad de que la variable aleatoria X adopte valores menores o iguales a 4 es: A) 0,30; B) 0,40; **C) 0,60**.
20. En un colegio público, la probabilidad de que un alumno de 3º de Primaria suspenda la asignatura de matemáticas es de 0,3. Si seleccionamos al azar a 4 alumnos, ¿cuál es la probabilidad de que la mitad de esos alumnos suspendan la asignatura? **A) 0,2646**; B) 0,7354; C) 0,1500.
21. Las puntuaciones de una prueba de lectura siguen una distribución normal con media 50 y desviación típica 10. La proporción de sujetos en esta prueba con puntuaciones entre 40 y 60 es igual a: A) 0,1587; **B) 0,6826**; C) 0,8413.
22. Una variable aleatoria X se distribuye según la F de Snedecor con 8 grados de libertad en el numerador y 9 grados de libertad en el denominador. ¿Cuál es la probabilidad de que la variable X tome un valor menor o igual que 3,23?: **A) 0,95**; B) 0,80; C) 0,16.
23. En una variable aleatoria X que sigue la distribución Chi-cuadrado con varianza igual a 16, el percentil 90 es: A) 3,4895; **B) 13, 3616**; C) 23,5418.
24. Se ha aplicado un test de comprensión verbal a una muestra aleatoria de 324 niños de Educación Secundaria obteniendo una media igual a 52 y una cuasivarianza de 81. Para un nivel de confianza de 0,95, ¿cuál es el error de estimación máximo de la media?: **A) 0,98**; B) 1,29; C) 8,82.
25. En una muestra aleatoria de 300 estudiantes de Bachillerato se ha comprobado que 225 de ellos NO toman refrescos azucarados de forma habitual. Al nivel de confianza del 95%, ¿entre qué valores consideramos que se encuentra la proporción de la población de estudiantes de Bachillerato que SI toman refrescos azucarados de forma habitual? A) 0,152 y 0,348; **B) 0,201 y 0,299**; C) 0,701 y 0,799.

EXTRACCIÓN

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2015 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Soluciones

Gráfica 1. *Antigüedad*, en años, de los trabajadores discapacitados de una empresa. La antigüedad está agrupada en intervalos y en el eje de abscisas están representados los puntos medios de esos intervalos.

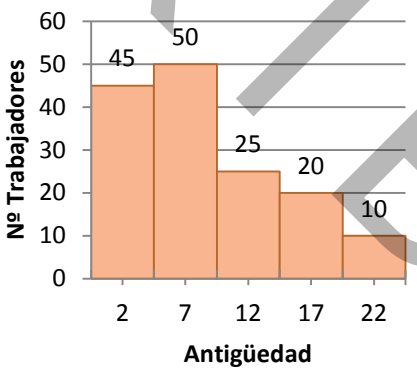
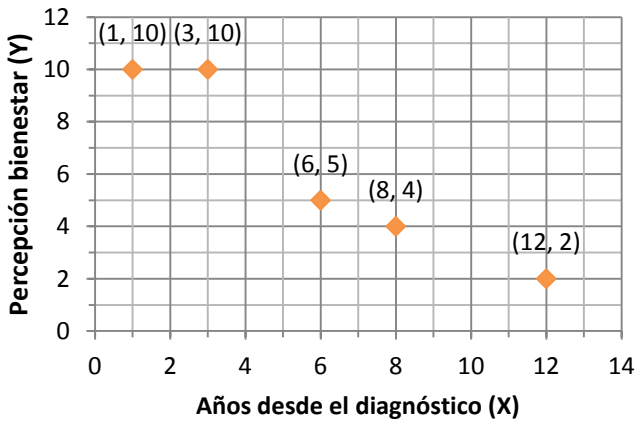


Tabla 1. Datos de la Encuesta de Población Activa en el último trimestre de 2013 por sexo y sector económico (unidades: miles de personas).

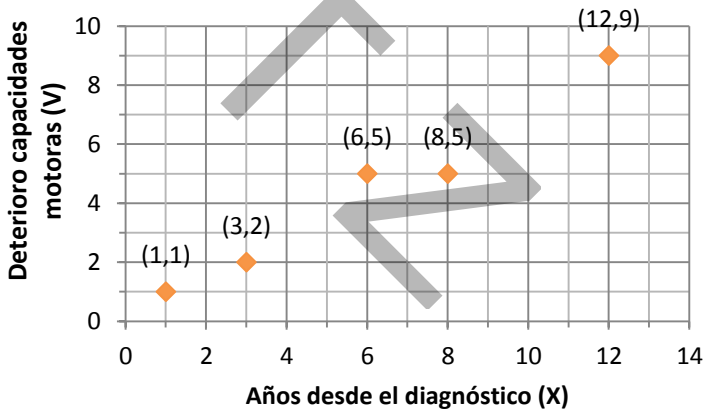
	Agricultura	*Industria*	*Construcción*	*Servicios*
Hombre	597,5	1.727,0	900,9	5.886,7
Mujer	193,4	547,2	77,4	6.828,0

En la Gráfica 2 se muestra la relación de los años transcurridos desde el diagnóstico de una enfermedad degenerativa con la percepción de bienestar percibido (**Gráfica 2.1**) y con el deterioro de las capacidades motoras (**Gráfica 2.2**).

Gráfica 2.1.



Gráfica 2.2.



1. La variable *antigüedad* de la Gráfica 1 es: A) ordinal; B) de intervalo; C) de razón.
2. Atendiendo a los datos de la Gráfica 1, ¿cuáles son los límites exactos del intervalo de mayor antigüedad? A) -0,5 y 4,5; B) 20 y 24; C) 19,5 y 24,5.
3. ¿Qué representación gráfica también se podría utilizar con los datos de la Gráfica 1? A) Diagrama de sectores; B) Polígono de frecuencias; C) Diagrama de barras conjunto.
4. La antigüedad media de los trabajadores discapacitados representados en la Gráfica 1 es: A) 8,67; B) 13; C) 260.

KIBBUTZ

5. Con los datos de la Gráfica 1 se ha seleccionado a un trabajador discapacitado que lleva 16 años trabajando en la empresa. ¿Qué porcentaje de trabajadores discapacitados tienen una antigüedad menor que él? A) 54%; B) 75%; C) 84%.
6. ¿Cuál es la moda de la variable antigüedad descrita en la Gráfica 1? A) 2; B) 7; C) 50.
7. Atendiendo a los datos de la Gráfica 1, la amplitud total es: A) 5; B) 20; C) 25.
8. Con los datos de la Gráfica 2.2, la varianza del deterioro de las capacidades motoras es: A) 2,80; B) 7,84; C) 14,8.
9. La distribución de la antigüedad de los trabajadores discapacitados representados en la Gráfica 1 es: A) simétrica; B) asimétrica positiva; C) asimétrica negativa.
10. Es apropiado analizar la relación entre dos variables utilizando el coeficiente C de contingencia en el caso de: A) la Gráfica 1; B) la Gráfica 2.2; C) la Tabla 1.
11. ¿Qué gráfica representa una relación lineal inversa entre dos variables? A) La Gráfica 1; B) La Gráfica 2.1; C) La Gráfica 2.2.
12. Con los datos de la Gráfica 2.2, el coeficiente de correlación lineal de Pearson entre las variables *años transcurridos desde el diagnóstico* (X) y *deterioro de las capacidades motoras* (V) es: A) -0,96; B) 0,735; C) 0,984.
13. Con los datos de la Gráfica 2.1, ¿qué puntuación le pronosticaremos en bienestar a un paciente diagnosticado hace 10 años? A) 2,96; B) 5,35; C) 19,16.
14. Con los datos de la Tabla 1, si elegimos una persona al azar ¿cuál es la probabilidad de que sea mujer? A) 0,456; B) 0,544; C) 0,750.
15. Con los datos de la Tabla 1, si elegimos una persona al azar, ¿cuál es la probabilidad de que sea mujer y trabaje en el sector de la industria? A) 0,03; B) 0,32; C) 0,56.
16. Con los datos de la Tabla 1, elegimos una persona al azar y resulta que trabaja en el sector Servicios, ¿cuál es la probabilidad de que sea una mujer? A) 0,407; B) 0,463; C) 0,537.
17. Considerando un experimento aleatorio consistente en lanzar una moneda (no trucada) al aire dos veces y definiendo la variable aleatoria X como “número de caras obtenidas”. ¿Cuántos resultados posibles tiene el espacio muestral? A) 2; B) 3; C) 4.
18. Sabiendo que la enfermedad de Alzheimer afecta a una de cada diez personas jubiladas, si seleccionamos a 10 jubilados al azar, ¿qué es más probable? A) Que ninguno sufra Alzheimer; B) Que uno de ellos sufra Alzheimer; C) Que dos de ellos sufran Alzheimer.
19. Con los datos de prevalencia de la enfermedad de Alzheimer del ejercicio anterior, ¿cuál es la probabilidad de que 3 ó más jubilados padezcan esta enfermedad? A) 0,0702; B) 0,720; C) 0,8063.
20. Las puntuaciones en motivación de un grupo de 100 estudiantes se distribuyen normalmente con media 8 y desviación típica 2. ¿Cuántos estudiantes han obtenido una puntuación en motivación por debajo de la media del grupo? A) 10; B) 50; C) 75.
21. Si una variable sigue una distribución *t* de Student con 16 grados de libertad, ¿cuál es la probabilidad de obtener valores menores o iguales que -2,12? A) 0,025; B) 0,050; C) 0,975.
22. En una distribución F con 5 grados de libertad en el numerador y 5 grados de libertad en el denominador, ¿cuál es el valor del percentil 10? A) 0,2896; B) 0,3453; C) 3,453.
23. Una distribución muestral de la media es: A) la distribución de probabilidad de una característica de la muestra; B) la distribución de probabilidad del estadístico media calculado en todas las posibles muestras de igual tamaño extraídas de una población; C) la distribución de probabilidad del parámetro media calculado en las muestras representativas.
24. Los límites inferior y superior del intervalo de confianza del 99%, construido sobre la media de 100 observaciones son 52 y 58. ¿Cuál será la media obtenida en la muestra? A) 50; B) 55; C) 60.
25. En una muestra aleatoria de 300 parados de una determinada comunidad autónoma, 6 de cada 10 son mujeres. Para un nivel de confianza de 0,95, ¿entre qué límites se estima que esté la proporción de mujeres paradas de toda la comunidad autónoma? A) 0,45 y 0,55; B) 0,36 y 0,72; C) 0, 55 y 0,66.

TUBEGITN

SOLUCIONES:

1. C
2. C
3. B
4. A

Antigüedad	n_i	n_iX_i
22	10	220
17	20	340
12	25	300
7	50	350
2	45	90
	150	1300

$$\overline{X} = \frac{\sum n_i X_i}{n} = \frac{1300}{150} = 8{,}67$$

5. C

Antigüedad	n_i	n_a
20-24	10	150
15-19	20	140
10-14	25	120
5-9	50	95
0-4	45	45

$$k = \left[\frac{\frac{(P_k - L_i)\cdot n_c}{I} + n_d}{n}\right]\cdot 100 = \left[\frac{\frac{(16-14{,}5)\cdot 20}{5} + 120}{150}\right]\cdot 100 = 84$$

6. B

El intervalo modal es 5-9, por lo que la moda es $M_o = 7$.

7. C

$A_T = X_{máx} - X_{min} = 24{,}5 - (-0{,}5) = 25$

8. B

Años desde el diagnóstico (X)	Deterioro cap. motoras (V)	V^2
1	1	1
3	2	4
6	5	25
8	5	25
12	9	81
	22	136

$$\overline{V} = \frac{22}{5} = 4{,}4$$

$$S_V^2 = \frac{136}{5} - 4{,}4^2 = 27{,}2 - 19{,}36 = 7{,}84$$

9. B

Observando la Gráfica 1 se aprecia que la mayoría de las puntuaciones (las mayores frecuencias) se encuentran en los valores bajos de antigüedad, por lo que se trata de una distribución asimétrica positiva. Utilizando el índice de asimetría de Pearson llegamos a esta misma conclusión:

$$A_S = \frac{\overline{X} - Mo}{S_X} = \frac{8{,}67 - 7}{S_x} = \frac{1{,}37}{S_x}$$

Dado que la desviación típica siempre es positiva (o nula) el resultado de esta división siempre será un número positivo.

10. C

KIBBUTZ

11. B

Observando la Gráfica 2.1 se observa que a mayores valores en X le corresponden menores valores en Y y viceversa.

12. C

Paciente	Años desde el diagnóstico (X)	Deterioro cap. motoras (V)	XV	X^2	V^2
1	1	1	1	1	1
2	3	2	6	9	4
3	6	5	30	36	25
4	8	5	40	64	25
5	12	9	108	144	81
∑	30	22	185	254	136

$$r_{XV} = \frac{n\sum XV - \sum X \sum V}{\sqrt{n\sum X^2 - \left(\sum X\right)^2}\sqrt{n\sum V^2 - \left(\sum V\right)^2}} = \frac{5\times 185 - 30\times 22}{\sqrt{5\times 254 - 30^2}\sqrt{5\times 136 - 22^2}} =$$

$$= \frac{925-660}{19{,}235\times 14} = \frac{265}{269.29} = 0{,}984$$

13. A

Paciente	Años desde el diagnóstico (X)	Percepción bienestar (Y)	XY	X^2
1	1	10	10	1
2	3	10	30	9
3	6	5	30	36
4	8	4	32	64
5	12	2	24	144
∑	30	31	126	254

$$Y' = a + bX$$

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2} = \frac{5\times 126 - 30\times 31}{5\times 254 - 30^2} = \frac{630-930}{1270-900} = \frac{-300}{370} = -0{,}81$$

$$a = \overline{Y} - b\overline{X} = \frac{31}{5} - (-0{,}81)\frac{30}{5} = 6{,}2 + 4.86 = 11.06$$

$$Y' = 11{,}06 - 0{,}81X = 11{,}06 - 0{,}81\times 10 = 2{,}96$$

14. A

	Agricultura	Industria	Construcción	Servicios	
Hombres	597500	1727000	900900	5886700	9112100
Mujeres	193400	547200	77400	6828000	7646000
	790900	2274200	978300	12714700	16758100

$$P(\text{Mujer}) = \frac{7646000}{16758100} = 0{,}456$$

15. A

	Agricultura	Industria	Construcción	Servicios	
Hombres	597500	1727000	900900	5886700	9112100
Mujeres	193400	547200	77400	6828000	7646000
	790900	2274200	978300	12714700	16758100

KIBBUTZ

$$P(\text{Mujer} \cap \text{Industria}) = \frac{547200}{16758100} = 0{,}03$$

16. C

	Agricultura	Industria	Construcción	Servicios	
Hombres	597500	1727000	900900	5886700	9112100
Mujeres	193400	547200	77400	6828000	7646000
	790900	2274200	978300	12714700	16758100

$$P(\text{mujer} \mid \text{servicios}) = \frac{6828000}{12714700} = 0{,}537$$

17. C

Resultados posibles: cara y cara, cara y cruz, cruz y cara, cruz y cruz.

18. B

P(X=0) = 0,3487 (Tabla I con n=10, p=0,1 y x=0)

P(X=1) = 0,3874 (Tabla I con n=10, p=0,1 y x=1)

P(X=2) = 0,1937 (Tabla I con n=10, p=0,1 y x=2)

Es más probable que uno de ellos padezca la enfermedad de Alzheimer.

19. A

P(X ≥ 3) = 1 – P(X ≤ 2) = 1 - 0,9298 = 0,0702. (Tabla II con n=10, p=0,10 y x=2).

20. B

Dado que la puntuación 8 es la media del grupo, la puntuación típica tiene que ser 0.

$$Z = \frac{8-8}{2} = 0$$

P(Z≤0) = 0,5000 (Tabla III o IV)

0,500×100 = 50.

21. A

P(X ≤ -2,12) = P(X > 2,12) = 1 – P(X ≤ 2,12) = 1 - 0,975 = 0,025 (Tabla VI)

22. A

$${}_{0,10}F_{5,5} = \frac{1}{{}_{0,90}F_{5,5}} = \frac{1}{3{,}453} = 0{,}2896 \text{ (Tabla VII)}$$

23. B

24. B

$$\frac{52+58}{2} = 55$$

25. C

$$L_i = 0{,}60 - 1{,}96\sqrt{\frac{0{,}60\cdot(1-0{,}60)}{300}} = 0{,}60 - 1{,}96\cdot 0{,}028 = 0{,}60 - 0{,}055 = 0{,}545 \approx 0{,}55$$

$$L_s = 0{,}60 + 1{,}96\sqrt{\frac{0{,}60\cdot(1-0{,}60)}{300}} = 0{,}60 + 1{,}96\cdot 0{,}028 = 0{,}60 + 0{,}055 = 0{,}655 \approx 0{,}66$$

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
SEPTIEMBRE 2016 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO C DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

SOLUCIONES

KIBBUTZ

Tabla 1. Resultados de un test de autoestima en una muestra de adolescentes españoles.

Autoestima	n_i
80 - 99	60
60 – 79	120
40 – 59	180
20 – 39	90
0 – 19	50

Tabla 2. Nivel de ansiedad previo a la realización del examen de conducir y nota obtenida en el examen de 200 personas.

		Nota en el examen	
		Apto	No-apto
Nivel de ansiedad	Normal	100	10
	Alto	40	50

Tabla 3: Un psicólogo utilizó un test de Analogías Verbales (X) para predecir el rendimiento en Lengua de 4º de la ESO (Y). Obtuvo las puntuaciones de las dos variables en una muestra aleatoria de 1000 estudiantes. En la tabla se muestran las medias, las varianzas y la correlación entre ambas variables.

	X	Y
Media	30	15
Varianza	64	36
Correlación	$r_{XY} = 0{,}70$	

Tabla 4: Función de probabilidad de una variable aleatoria discreta X

x	f(x)
2	0,35
1	0,40
0	0,25

1. ¿Cuál es la amplitud de los intervalos de la variable *autoestima* de la Tabla 1?: A) 10; B) 19; **C) 20**
2. ¿Cuál es la media en autoestima de la muestra de adolescentes españoles de la Tabla 1?: A) 49,5; **B) 51,5;** C) 5150
3. Los percentiles son medidas de: A) tendencia central; **B) posición;** C) desviación
4. Con los datos de la Tabla 1, la Mediana de la variable Autoestima es: A) 50,5; B) 51,5; **C) 51,72**
5. Para calcular el rango o amplitud total de una variable necesitamos: A) las frecuencias absolutas; B) que la variable sea cualitativa; **C) las puntuaciones máxima y mínima**
6. ¿Cuál de los siguientes índices puede adoptar valores negativos? **A) El índice de asimetría de Pearson**; B) La varianza; C) La desviación típica.

7. Con los datos de la Tabla 3, ¿qué variable presenta un mayor grado de dispersión?: A) las puntuaciones en el test de analogías verbales; **B) las puntuaciones en lengua;** C) las dos variables presentan el mismo grado de dispersión
8. La puntuación típica de un sujeto en un test de razonamiento abstracto ha sido de 2. ¿Cuál fue su puntuación diferencial en el test si la media fue de 10 y la varianza de 9?: A) 0; B) 16; **C) 6**
9. Con los datos de la Tabla 2, el valor de X^2 está entre: A) 35 y 45; **B) 46 y 56**; C) 57 y 67
10. Cuando el coeficiente de contingencia es igual al C_{max}, podemos decir que: A) Existe una relación moderada entre las variables; **B) Existe una fuerte relación entre las variables**; C) No existe relación entre las variables
11. En la recta de regresión, la proporción de varianza de Y explicada por la varianza de X es igual a: **A) El coeficiente de correlación entre ambas variables al cuadrado**; B) El coeficiente de correlación entre ambas variables; C) La raíz cuadrada del coeficiente de correlación entre ambas variables
12. Con los datos de la Tabla 3, la covarianza entre ambas variables es igual a: A) 83,6; B) 25,3; **C) 33,6**
13. La recta de regresión de Y sobre X siempre pasa por el punto: **A)** $(\overline{X},\overline{Y})$; B) $(0,\overline{Y})$; C) $(\overline{X},0)$
14. Con los datos de la Tabla 3, la pendiente de la recta de regresión de Y sobre X en puntuaciones directas es: **A) 0,525**; B) 0,385; C) 0,495
15. Con los datos de la Tabla 2 si elegimos al azar una persona, ¿cuál es la probabilidad de que presente un nivel de ansiedad "Normal"?: A) 0,50; **B) 0,55**; C) 0,60
16. Con los datos de la Tabla 2 si elegimos al azar una persona, ¿cuál es la probabilidad de que presente un nivel de ansiedad "Normal" y esté "Apto" en el examen?: **A) 0,50**; B) 0,55; C) 0,60
17. Con los datos de la Tabla 2, elegimos al azar una persona y ha resultado que tiene un nivel de ansiedad "Alto", ¿cuál es la probabilidad de que presente una calificación de "Apto" en el examen?: **A) 0,44**; B) 0,55; C) 0,66
18. La probabilidad de que un paciente con esquizofrenia se recupere con un tratamiento determinado es de 0,40. Un psicólogo está tratando individual e independientemente a 10 pacientes con este trastorno. La probabilidad de que se recuperen al menos 7 pacientes es de: A) 0,9452; **B) 0,0548**; C) 0,0123
19. Con los datos de la Tabla 4, la función de distribución de la variable *X* es: **A) 0,25; 0,65 y 1**; B) 0,25; 0,40 y 0,35; C) 0,40; 0,60 y 1
20. Con los datos de la Tabla 4, la esperanza matemática de la variable *X* es igual a: A) 0,70; B) 1,0; **C) 1,1**
21. Las puntuaciones en un test de *memoria* se distribuyen normalmente en una población de 10000 adolescentes con media 100 y desviación típica 10. ¿Cuántos adolescentes tienen una puntuación superior a 115 en el test de *memoria*?: A) ninguno; B) 600; **C) 668**
22. En una distribución Chi-cuadrado con 20 grados de libertad, el percentil 10 vale: A) 8,2604; **B) 12,4426;** C) 20
23. En una distribución F de Snedecor con 10 grados de libertad en el numerador y 10 grados de libertad en el denominador, al valor 2,323 le corresponde el percentil: A) 9; B) 10; **C) 90**
24. Se mide la depresión en una muestra aleatoria de 225 sujetos obteniendo una cuasivarianza de 25. El error típico de la media vale: A) 0,11; **B) 0,33;** C) 3
25. Para estimar el intervalo de confianza para la media de una variable X, a un nivel de confianza del 99%, hemos extraído una muestra aleatoria de 144 personas y les hemos medido la variable X obteniendo una media de 12 y una cuasidesviación típica de 4. Los límites del intervalo de confianza son: A) 8 y 14; B) 9,42 y 14,58; **C) 11,14 y 12,86**

KIBBUTZ

SOLUCIONES:

1. C
2. B

Autoestima	X_i	n_i	n_iX_i
80 - 99	89.5	60	5370
60 – 79	69.5	120	8340
40 – 59	49.5	180	8910
20 – 39	29.5	90	2655
0 – 19	9.5	50	475
		500	25750

$$\overline{X} = \frac{\sum n_i X_i}{n} = \frac{25750}{500} = 51{,}5$$

3. B
4. C

Autoestima	X_i	n_i	n_a
80 - 99	89.5	60	500
60 – 79	69.5	120	440
40 – 59	49.5	180	320
20 – 39	29.5	90	140
0 – 19	9.5	50	50
		500	

$$Md = L_i + \left(\frac{\frac{n}{2} - n_d}{n_c}\right) \cdot I = 39{,}5 + \left(\frac{\frac{500}{2} - 140}{180}\right) \cdot 20 = 39{,}5 + 12{,}22 = 51{,}72$$

5. C
6. A
7. B

	X	Y
Media	30	15
Varianza	64	36

$$\left.\begin{array}{l} CV_X = \frac{S_X}{\overline{X}} \cdot 100 = \frac{8}{30} \cdot 100 = 26{,}67 \\ CV_Y = \frac{S_Y}{\overline{Y}} \cdot 100 = \frac{6}{15} \cdot 100 = 40 \end{array}\right\} \Rightarrow CV_Y > CV_X$$

8. C

KIBBUTZ

$$x = X - \overline{X}$$

$$z = \frac{x}{S_X} \rightarrow 2 = \frac{x}{3} \rightarrow x = 2 \times 3 = 6$$

9. B

		Nota en el examen		
		Apto	No-apto	
Nivel de Ansiedad	Normal	100 (77)	10 (33)	110
	Alto	40(63)	50(27)	90
		140	60	200

Las frecuencias teóricas aparecen entre paréntesis en la tabla.

$$X^2 = \frac{(100-77)^2}{77} + \frac{(40-63)^2}{63} + \frac{(10-33)^2}{33} + \frac{(50-27)^2}{27} = 6{,}8701 + 8.3968 + 16{,}0303 + 19{,}5926 = 50{,}89$$

10. B
11. A
12. C

	X	Y
Media	30	15
Varianza	64	36

$$r_{XY}^2 = 0{,}49 \Rightarrow r_{XY} = 0{,}70$$

$$S_{XY} = r_{XY} S_X S_Y = 33{,}6$$

13. A
14. A

$$b = r_{XY} \frac{S_Y}{S_X} = 0{,}70 \frac{6}{8} = 0{,}525$$

15. B

$$P(\text{Normal}) = \frac{110}{200} = 0{,}55$$

16. A

$$P(\text{Normal y Apto}) = \frac{100}{200} = 0{,}50$$

17. A

$$P(\text{Apto / Alto}) = \frac{40}{90} = 0{,}44$$

18. B

p=0,40 $P(X \geq 7) = 1 - P(X < 7) = 1 - P(X \leq 6) = 1 - F(6) = 1 - 0{,}9452 = 0{,}0348$

El valor de F(6) se busca en la tabla II, y es el valor de la intersección de la fila n=10, x=6 con la columna p=0,40.

19. A

KIBBUTZ

x	f(x)	F(x)
2	0,35	1
1	0,40	0,65
0	0,25	0,25

20. C

x	f(x)	xf(x)
2	0,35	0,70
1	0,40	0,40
0	0,25	0
		1,1

$\mu = 1{,}1$

21. C

$$Z = \frac{X - \mu}{\sigma} = \frac{115 - 100}{10} = 1{,}5$$

z = 1,5 $\rightarrow$ P(Z ≤ 1,5) = 0,9332 (Tabla IV)

P(Z > 1,5) = 1 – P(Z ≤ 1,5) = 1 - 0,9332 = 0,0668

10000 x 0,0668 = 668

22. B

P_{10} = 12,4426 (Tabla V para 20 g.l.)

23. C

Tabla VII para 10 g.l. en el numerador y 10 g.l. en el denominador

24. B

$$S_{\overline{X}} = \sqrt{\frac{S_{n-1}^2}{n}} = \sqrt{\frac{25}{225}} = 0{,}33$$

25. C

$$\overline{X} \pm z_{1-\alpha/2} \frac{S_{n-1}}{\sqrt{n}}$$

$$E_{máx} = z_{1-\alpha/2} \frac{S_{n-1}}{\sqrt{n}} = 2{,}58 \frac{4}{\sqrt{144}} = 0{,}86$$

$$\overline{X} - E_{máx} = 12 - 0{,}86 = 11{,}14$$

$$\overline{X} + E_{máx} = 12 + 0{,}86 = 12{,}86$$

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2014 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO F

Material: **Addenda (Formulario y Tablas) y calculadora**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

Gráfica 1: Número de mujeres asesinadas por la violencia de género y número de ellas que habían presentado denuncia por malos tratos entre los años 2004 y 2012 (Fuente: Ministerio de Sanidad, Servicios Sociales e Igualdad).

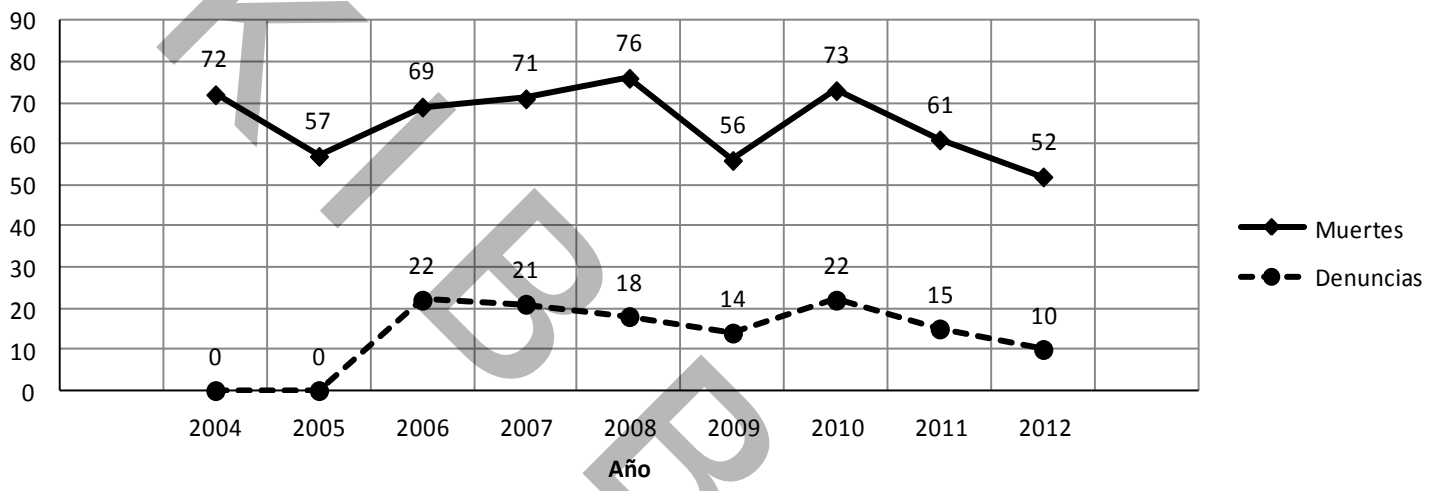


Tabla 1. Distribución de frecuencias de la variable Edad (X), de un grupo de 2000 personas.

X	n_i
21-30	400
31-40	600
41-50	500
51-60	400
61-70	100

Tabla 2. Para las puntuaciones de 200 personas en dos variables cuantitativas, X e Y, disponemos de los siguientes datos:

$\sum X = 1600$	$\sum Y = 20000$	$\sum XY = 163600$
$Y' = 4{,}5X + 64$		

Situación 1: Lanzamos al aire un dado en cuyas caras aparece un número del 1 al 6. Definimos el suceso A como "salir número par" y el suceso B como "salir múltiplo de 3".

1. En el eje de ordenadas de la Gráfica 1 aparece: A) la proporción; B) el porcentaje; **C) la frecuencia absoluta.**
2. Teniendo en cuenta la Gráfica 1, el número de mujeres asesinadas entre los años 2004 y 2012 ha sido: A) 128; **B) 587**; C) 600.
3. Teniendo en cuenta la Gráfica 1, en el año 2009 el porcentaje de mujeres asesinadas que habían presentado denuncia por malos tratos fué: A) 14%; **B) 25%;** C) 56%.
4. Teniendo en cuenta la Gráfica 1, la media de denuncias presentadas por mujeres que posteriormente fueron asesinadas es aproximadamente: **A) 13,6**; B) 18; C) 20.
5. La variable "Edad", de la Tabla 1, es: A) cualitativa; B) cuasicuantitativa; **C) cuantitativa**.

KIBBUTZ

6. Con los datos de la Tabla 1, la media es igual a: A) 35,5; **B) 41,5**; C) 45,5.
7. Con los datos de la Tabla 1, el tercer cuartil es igual a: A) 25,5; **B) 50,5**; C) 75,5.
8. Con los datos de la Tabla 1, ¿a qué percentil corresponde el valor 25,5?: **A) 10**; B) 20; C) 30.
9. Con los datos de la Tabla 1, la varianza es igual a: A) 50; **B) 134**; C) 247.
10. Con los datos de la Tabla 1, la distribución de la variable "Edad" es: **A) asimétrica positiva**; B) simétrica; C) asimétrica negativa.
11. Para calcular el grado de asociación entre dos variables cualitativas utilizamos: A) la covarianza; B) el coeficiente de correlación lineal de Pearson; **C) el coeficiente de contingencia C**.
12. En la Tabla 2, la covarianza entre X e Y es igual a: **A) 18**; B) 20; C) 25.
13. Teniendo en cuenta la Tabla 2, ¿qué puntuación pronosticaremos en Y a una persona que ha obtenido una puntuación de 8 en X?: A) 80; **B) 100**; C) 120**.**
14. Con los datos de la Situación 1, ¿cuánto vale la probabilidad de A?: A) 2/6; **B) 3/6**; C) 4/6
15. Teniendo en cuenta la Situación 1, ¿cuánto vale $P(A \cap B)$?: **A) 1/6**; B) 2/6; C) 3/6.
16. Teniendo en cuenta la Situación 1, ¿podemos afirmar que los sucesos A y B son independientes?: **A) Sí**; B) No; C) No podemos saberlo.
17. Una variable aleatoria discreta X toma los valores 0, 1 y 2. Sabemos que P(X=0)=0,2 ; P(X=1)=0,3 y P(X=2)=0,5. ¿cuánto vale la media o esperanza matemática de X?: A) 1; **B) 1,3**; C) 1,5.
18. Teniendo en cuenta los datos del ejercicio anterior, ¿cuánto vale la varianza de X?: A) 0,11; B) 0,41; **C) 0,61**.
19. El 75% de los alumnos de psicología escoge el itinerario de clínica. Si elegimos al azar 5 alumnos, ¿cuál es la probabilidad de que 3 ó más hayan escogido clínica?: A) 0,2637; B) 0,6985; **C) 0,8965**.
20. En una distribución normal tipificada, si $P(Z \leq z) = 0{,}0239$, entonces z es igual a: **A) -1,98**; B) 0,08; C) 1,98.
21. Si una variable X se distribuye según t de Student con 80 grados de libertad entonces $P(X \leq -0{,}678)$ es igual a: **A) 0,25**; B) 0,50; C) 0,75.
22. En una determinada facultad sólo se admiten estudiantes cuya puntuación en selectividad sea superior al percentil 67 (P_{67}). Sabiendo que las puntuaciones en selectividad se distribuyen *normalmente* con μ = 5,5 y σ = 1,5, para que un estudiante sea admitido debe obtener necesariamente una puntuación: A) mayor que 5,5; **B) mayor que 6,16**; C) mayor que 7,25.
23. Para conocer la forma de vida de los inmigrantes ilegales en España, que carecen de cualquier tipo de documentación y registro, ¿qué tipo de muestreo sería factible utilizar?: A) aleatorio simple; B) por conglomerados; **C) "bola de nieve**".
24. Sabemos que el CI (Cociente Intelectual) se distribuye normalmente con σ = 15. En una muestra, con n=144, obtenemos una media en CI de 100. Al nivel de confianza del 95% ¿entre qué valores estimaremos que se encuentra la media en CI en la población?: **A) 97,55 y 102,45**; B) 96,775 y 103,225; C) 85 y 115.
25. Para estimar las posibilidades de que la alumna PRS sea elegida por todos sus compañeros para delegada de curso elegimos una muestra de 100 estudiantes y les preguntamos si votarán o no a PRS obteniendo que 48 de ellos si lo harán. Con un nivel de confianza del 99%, y si para ser elegida necesita al menos el 51 % de los votos, podemos estimar que: A) no tiene posibilidades porque el intervalo confidencial estimado no incluye la proporción de 0,51**; B) tiene posibilidades porque el intervalo confidencial estimado incluye la proporción 0,51**; C) no podemos estimar si tiene posibilidades o no de ser elegida.

KUBBEN

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2014 Código asignatura: 62011037
EXAMEN MODELO A DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

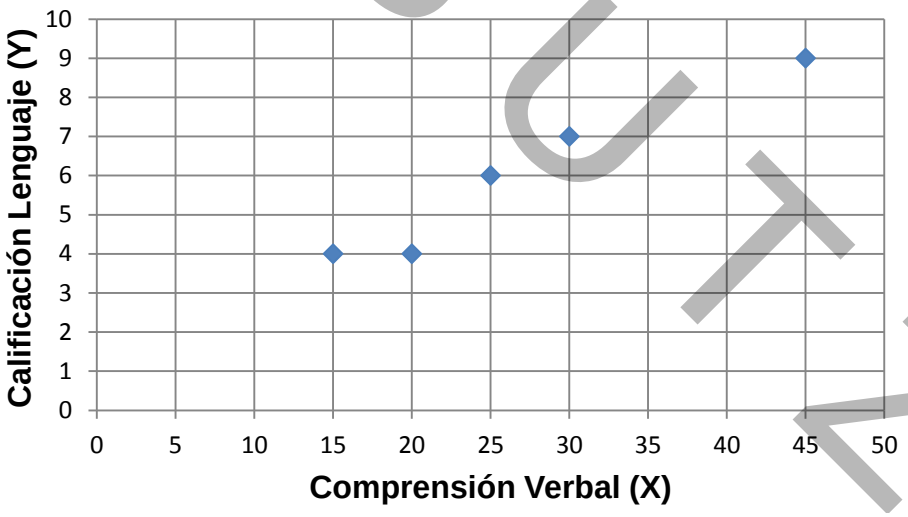
Calificación= (0,4 x Aciertos) - (0,2 x Errores)

Tabla 1. En una investigación se estudia la influencia de tres fármacos (A, B y C) en el tiempo (en segundos) en el que una rata encuentra la salida de un laberinto.

Tiempo en segundos	n_iA	n_iB	n_iC
120 ó más	3	9	2
90-119	10	18	9
60-89	24	18	19
30-59	16	10	18
0-29	7	5	12

Situación 1. En un barrio de las afueras de Madrid viven 10 trabajadores de una gran empresa cuya sede está en la otra punta de la ciudad. Hay tres posibles recorridos para realizar su trayecto y 5 de ellos escogen la ruta A, 3 de ellos la ruta B y 2 de ellos la ruta C. Se sabe además, que la probabilidad de encontrar atasco siguiendo la ruta A es 0,4, siguiendo la ruta B es 0,5 y siguiendo la ruta C es 0,65.

Figura 1. Representación gráfica conjunta de dos variables en una muestra de 5 niños de 2º de Primaria: la puntuación obtenida en un test de Comprensión Verbal (X) y la calificación obtenida en un examen de la asignatura Lenguaje (Y).



EXAMEN

1. En la investigación descrita en la Tabla 1, la variable *tiempo que tarda la rata en encontrar la salida del laberinto* actúa como variable: A) independiente; B) dependiente**;** C) extraña.
2. En la investigación descrita en la Tabla 1, la variable *tipo de fármaco empleado* es: A) politómica; B) cuasicuantitativa; C) cuantitativa.
3. ¿Cómo se denomina la representación gráfica utilizada en la Figura 1? A) Polígono de frecuencias; B) Polígono de frecuencias conjunto; C) Diagrama de dispersión.
4. Teniendo en cuenta los datos del fármaco A de la Tabla 1, ¿qué índice no podemos calcular? A) La media; B) La mediana; C) La moda.

5. Teniendo en cuenta los resultados del fármaco C de la Tabla 1, se puede afirmar que la distribución es: A) amodal; B) unimodal; C) bimodal.
6. Con los datos del fármaco C de la Tabla 1, ¿qué cuartil, aproximadamente, le corresponde a una rata que ha tardado 59,5 segundos en encontrar la salida del laberinto? A) Q_1; B) Q_2; C) Q_3.
7. Con los datos de la Figura 1, ¿cuál es la desviación típica de la calificación obtenida en Lenguaje? A) 1,9; B) 3,6; C) 10,3.
8. Para comparar la variabilidad de la puntuación obtenida en el test de Comprensión Verbal y la calificación en la asignatura de Lenguaje de la Figura 1, es recomendable utilizar: a) la varianza; B) el coeficiente de correlación lineal de Pearson; C) el coeficiente de variación.
9. Atendiendo a los datos de la Figura 1, ¿qué puntuación típica le corresponde a un niño que ha obtenido un 4 en el examen de Lenguaje? A) -2; B) -1,05; C) 0,55.
10. Se desea comprobar si hay relación entre la nacionalidad (inglés, alemán y español) y el destino preferido en vacaciones (playa, montaña, ciudad) en una muestra determinada. ¿Qué valor máximo puede adoptar el coeficiente de contingencia en este caso? A) 0,707; B) 0,816; C) 1.
11. Observando la Figura 1, podemos concluir que la relación entre la prueba de Comprensión Verbal y la calificación obtenida en la asignatura de Lenguaje es: A) lineal directa; B) lineal inversa; C) no lineal.
12. Atendiendo a los datos de la Figura 1, la recta de regresión que permite pronosticar la calificación en el examen (Y) en función de la puntuación obtenida en el test de Comprensión Verbal (X) es: A) Y' = 0,18 + 1,14X; B) Y' = 1,14 + 0,18X; C) Y' = 25,92 + 0,18X.
13. Con los datos de la Figura 1, ¿qué puntuación le pronosticaremos en el examen de Lenguaje a un niño que ha obtenido un 12 en el test de comprensión verbal? A) 3,3; B) 4,5; C) 6.
14. Sean los sucesos A={1,2,3,4,7} y B={1,2,4,6,10}. La unión de A y B es igual a: A) {1,2,4}; B) {1,2,3,4,5,6,7,8,9,10}; C) {1,2,3,4,6,7,10}.
15. Teniendo en cuenta los datos de la Situación 1, ¿cuál es la probabilidad de sufrir un atasco? A) 0,25; B) 0,48; C) 0,63.
16. Con los datos de la Situación 1, se ha elegido un trabajador al azar constatando que ha sufrido un atasco, ¿cuál es la probabilidad de que haya escogido la ruta B? A) 0,15; B) 0,31**;** C) 0,53.
17. Una variable aleatoria X tiene la misma probabilidad de presentar cualquiera de sus cuatro valores: 0, 1, 2 y 3. ¿Cuál es la media de X? A) 0; B) 0,25; C) 1,5.
18. ¿Cuál de las siguientes variables sigue una distribución binomial? A) Nº de caras obtenidas al lanzar una moneda tres veces; B) Nº veces que hay que lanzar un dado para obtener un 6; C) Tanto el lanzamiento de la moneda expuesto en la opción A como el lanzamiento del dado expuesto en la opción B se distribuyen según la binomial.
19. Un estudiante responde al azar las 10 preguntas de un examen de verdadero y falso. ¿Cuál es la probabilidad de que acierte todas las preguntas? A) 0,100; B) 0,010; C) 0,001.
20. Las puntuaciones en un examen de la asignatura de Conocimiento del Medio de un grupo de 100 alumnos se distribuyen normalmente con media 5 y desviación típica 1,5. ¿Cuántos estudiantes han obtenido una puntuación inferior a 5 en el examen? A) 30; B) 40; C) 50.
21. Una variable se distribuye según una t con 17 grados de libertad, ¿cuál es el valor que deja por encima de sí el 5% de la distribución? A) -1,740; B) 1,740; C) 2,898.
22. Una variable X se distribuye según la F de Snedecor con 7 grados de libertad en el numerador y 20 en el denominador. El percentil 90 es: A) 2,040; B) 2,595; C) 3,699.
23. Cuanto mayor es el error de estimación máximo: A) mayor es la amplitud del intervalo de confianza; B) menor es la amplitud del intervalo de confianza; C) mayor es la precisión de la estimación.
24. La variable autoestima se distribuye normalmente en la población adolescente. Se midió dicha variable en una muestra aleatoria de 25 adolescentes, obteniendo una media de 60 y una cuasivarianza de 100. ¿Cuánto vale el error típico de la media?: A) 0,4; B) 2; C) 4.
25. Si la amplitud de un intervalo de confianza para la proporción es 0,4, con un nivel de confianza de 0,95, ¿cuál es el error de estimación máximo? A) 0,2; B) 0,4; C) 0,8.

EXAMEN

SOLUCIONES:

1. B
2. A
3. C
4. A
5. B
6. B

Tiempo en segundos	n_i	n_a
más de 120	2	60
90-119	9	58
60-89	**19**	**49**
30-59	18	30
0-29	12	12

$$k = \left[\frac{\frac{(P_k - L_i)\cdot n_c}{I} + n_d}{n}\right]\times 100 = \left[\frac{\frac{(59{,}5 - 59{,}5)\cdot 19}{30} + 30}{60}\right]\times 100 = 50$$

$P_{50} = Q_2$

7. A

niño	Y	Y^2
1	4	16
2	4	16
3	6	36
4	7	49
5	9	81
$\sum$	30	198

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{30}{5} = 6$$

$$S_X^2 = \frac{\sum X_i^2}{n} - \overline{X}^2 = \frac{198}{5} - 6^2 = 3{,}6$$

$$S_X = \sqrt{S_X^2} = \sqrt{3{,}6} = 1{,}897 \approx 1{,}9$$

8. C
9. B

$$Z_x = \frac{X - \overline{X}}{S_x} = \frac{4-6}{1{,}9} = -1{,}05$$

10. B

$$C_{max} = \sqrt{\frac{k-1}{k}} = \sqrt{\frac{3-1}{3}} = 0{,}816$$

11. A
12. B

Niño	Comprensión Verbal (X)	Examen Lenguaje (Y)	XY	X^2
1	15	4	60	225
2	20	4	80	400
3	25	6	150	625
4	30	7	210	900
5	45	9	405	2025
$\sum$	135	30	905	4175

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2} = \frac{5\times 905 - 135\times 30}{5\times 4175 - 135^2} = \frac{4525 - 4050}{2650} = 0{,}18$$

$$\overline{X} = \frac{135}{5} = 27 \qquad \overline{Y} = \frac{30}{5} = 6$$

KIBBUTZ

$a = \overline{Y} - b\overline{X} = 6 - 0{,}18 \times 27 = 1{,}14$

$Y_i' = a + bX = 1{,}14 + 0{,}18X$

13. A

Y' = 1,14 + 0,18×12 = 3,3

14. C

15. B

$P(A) = \frac{5}{10} = 0{,}5 \qquad P(atasco|A) = 0{,}4$

$P(B) = \frac{3}{10} = 0{,}3 \qquad P(atasco|B) = 0{,}5$

$P(C) = \frac{2}{10} = 0{,}2 \qquad P(atasco|C) = 0{,}65$

$$P(atasco) = P(A \cap atasco) + P(B \cap atasco) + P(C \cap atasco) =$$
$$= P(A) \cdot P(atasco|A) + P(B) \cdot P(atasco|B) + P(C) \cdot P(atasco|C) =$$
$$= 0{,}5 \times 0{,}4 + 0{,}3 \times 0{,}5 + 0{,}2 \times 0{,}65 = 0{,}48$$

16. B

$$P(B|atasco) = \frac{P(B) \cdot P(atasco|B)}{P(atasco)} = \frac{0{,}3 \times 0{,}5}{0{,}48} = 0{,}31$$

17. C

X	f(x)	xf(x)
0	0,25	0
1	0,25	0,25
2	0,25	0,50
3	0,25	0,75
		1,5

$\mu = \sum x \cdot f(x) = 1{,}5$

18. A

19. C

Binomial con n = 10, p = 0,5 y x = 10 (Tabla I)

20. C

$$z = \frac{X - \overline{X}}{S_X} = \frac{5 - 5}{1{,}5} = 0$$

P(Z ≤ 0) = 0,500 → (Tabla IV)

0,5×100=50

21. B

Al ser simétrica con media 0, se busca directamente en la tabla t, la probabilidad 0,950, con 17 grados de libertad.

22. A

Se busca directamente en la Tabla VII para 7 g.l. en el numerador y 20 g.l. en el denominador)

KIBBUTZ

$$P(X \leq 2{,}040) = 0{,}90$$

23. A

24. B

$$S_{\overline{X}} = \frac{S_{n-1}}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

25. A

$$E_{max} = \frac{0{,}4}{2} = 0{,}2$$

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
SEPTIEMBRE 2013 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Tabla 1. Puntuaciones directas en un test de razonamiento de 400 niños de un colegio de Educación Primaria.

X	n_i
28-32	65
23-27	70
18-22	160
13-17	50
8-12	55

Gráfico 1. Representación gráfica de la función de probabilidad de la variable aleatoria X.

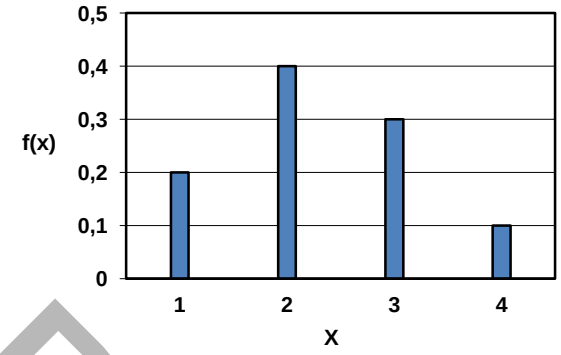


Tabla 2. Resultados obtenidos por 100 comerciales en un test de motivación de logro (X) y las ventas realizadas durante el último mes (Y).

	Media	Varianza	Covarianza entre X e Y
X	30	25	13,5
Y	6	9	

Tabla 3. Mejoría clínica de un grupo de 200 pacientes diagnosticados con depresión según el tipo de tratamiento recibido.

Tratamiento	Mejoría clínica	
	Sí	No
Cognitivo	40	25
Conductual	60	10
Farmacológico	50	15

1. La escala de medida de la variable *tipo de terapia recibida* de la Tabla 3 es: A) ordinal; B) de intervalo; C) nominal.
2. El gráfico adecuado para representar la relación entre las dos variables de la Tabla 2 es el diagrama de: A) dispersión; B) sectores; C) barras conjunto.
3. La variable *número de caras obtenidas al lanzar al aire dos monedas* es: A) dicotómica; B) discreta; C) continua.
4. Considerando los datos de la Tabla 1, la moda de las puntuaciones en el test es: A) 160; B) 20; C) 22,5.
5. Con los datos de la Tabla 1, ¿qué percentil le corresponde a un niño con una puntuación de 17 en el test?: A) 25; B) 27; C) 30.
6. Con los datos de la Tabla 1, el cuartil 3 es: A) 25; B) 27; C) 31.

LIBRE UNI

7. Respecto a la Tabla 2, ¿qué variable presenta un mayor grado de variabilidad?: A) el test de motivación de logro; B) las ventas realizadas durante el último mes; C) las dos muestran el mismo grado de variabilidad.
8. Atendiendo a la Tabla 1, la amplitud semi-intercuartil de las puntuaciones en el test es de: A) 4; B) 4,75; C) 8.
9. Con los datos de la Tabla 2, a un comercial que ha obtenido en el test de motivación de logro una puntuación típica de 1,5, ¿qué puntuación directa le corresponde?: A) 30; B) 31,5; C) 37,5.
10. Teniendo en cuenta los datos de la Tabla 2, el coeficiente de correlación de Pearson entre X e Y, vale: A) 0,6; B) 0,7 ; C) 0,9.
11. Con los datos de la Tabla 2, La recta de regresión para pronosticar las ventas mensuales (Y) a partir de la motivación de logro (X) es: A) Y' = -10,2+0,54X; B) Y´=-42,6+1,62X; C) Y´=-21+0,90X.
12. Con los datos de la Tabla 3, si ambas variables no estuviesen relacionadas entre sí, el número esperado de pacientes que han recibido terapia conductual y que no han mostrado mejoría clínica es de: A) 14,2; B) 17,5; C) 33,3.
13. Considerando los datos de la Tabla 3, el valor del Coeficiente de Contingencia Máximo ($C_{máx}$): A) es 0,707; B) es 0,816; C) no se puede calcular.
14. Con los datos de la Tabla 3, si elegimos un paciente al azar, ¿cuál es la probabilidad de que haya recibido terapia conductual y se observe mejoría clínica?: A) 0,30; B) 0,40; C) 0,86.
15. Considerando los datos de la Tabla 3, si suponemos que se ha seleccionado un paciente al azar y ha resultado que NO ha mejorado en su depresión, ¿cuál es la probabilidad de que haya recibido la terapia cognitiva?: A) 0,13; B) 0,38; C) 0,50.
16. Si $P(A\cap B)\neq P(A)P(B)$, entonces los sucesos A y B son: A) imposibles; B) dependientes; C) complementarios.
17. La esperanza matemática de la variable aleatoria representada en el Gráfico 1 es: A) 2,3; B) 2; C) 2,5.
18. Con los datos del Gráfico 1, el valor de la función de distribución para X=3, F(3), es: A) 0,30; B) 0,60;C) 0,90.
19. En un estudio sobre la implantación de nuevas tecnologías en el aula, concluyen que el 30% de los profesores utilizan la pizarra digital en sus clases. Si elegimos al azar una muestra de 10 profesores ¿cuál es la probabilidad de que al menos 5 de ellos utilicen la pizarra con sus alumnos?: A) 0,0473; B) 0,0781; C) 0,1503.
20. Las puntuaciones en un test de inteligencia siguen una distribución normal con $\mu = 100$ y $\sigma = 15$. El percentil 33 (P_{33}) de esta distribución vale: A) -0,44; B) 0,6293; C) 93,40
21. La probabilidad de que una variable X que se distribuye como una *Chi-cuadrado* con 14 grados de libertad, tome un valor superior a 21,0641 es: A) 0; B) 0,10; C) 0,90
22. Una variable que se distribuye según una *t de Student* con n grados de libertad, puede tomar: A) únicamente valores positivos; B) únicamente valores positivos y negativos; C) valores positivos y negativos y el cero
23. Para un nivel de confianza de 0,95, $\sigma = 3$ y un error de estimación máximo de 1,2, el tamaño de la muestra requerido para estimar *la autoestima* media de una población resulta ser n=24. Sabiendo que la variable *autoestima* se distribuye *normalmente* en la población, ¿qué tamaño debería tener la muestra si el nivel de confianza fuese 0,99?: A) 66; B) 42; C) 24
24. En una muestra aleatoria de 121 adolescentes, la media y la cuasidesviación típica de las puntuaciones en un test de inteligencia emocional valen 22 y 7,865 respectivamente. Sabiendo que el error de estimación máximo de la media vale 1,4, ¿cuánto vale $z_{1-\alpha/2}$?: A) 1,96; B) 2,33; C) 2,58
25. Teniendo en cuenta los datos del ejercicio anterior, ¿cuáles son los límites entre los cuales se espera esté la inteligencia emocional media de la población de adolescentes?: A) 19,42 y 24,58; B) 20,04 y 23,96; C) 20,6 y 23,4

EXTRACTO

SOLUCIONES:

1. C
2. A
3. B
4. B

   La moda es el punto medio del intervalo con mayor frecuencia absoluta. El intervalo es 18-22 y su punto medio es 20

5. A

   La puntuación 17 se encuentra en el intervalo 13-17.

X	n_i	n_a
28-32	65	400
23-27	70	335
18-22	160	265
13-17	50	105
8-12	55	55

$$k = \left[\frac{\frac{(P_k - L_i)\cdot n_c}{I} + n_d}{n} \right] \times 100 = \left[\frac{\frac{(17-12{,}5)50}{5} + 55}{400} \right] \times 100 = 25$$

6. A

X	n_i	n_a
28-32	65	400
23-27	70	335
18-22	160	265
13-17	50	105
8-12	55	55

$$P_{75} = L_i + \left(\frac{\frac{n \cdot k}{100} - n_d}{n_c} \right) \cdot I = 22{,}5 + \left(\frac{\frac{400 \cdot 75}{100} - 265}{70} \right) \cdot 5 = 25$$

7. B

$$CV_X = \frac{S_X}{\bar{X}} \cdot 100 = \frac{5}{30} \cdot 100 = 16{,}67$$

$$CV_Y = \frac{S_Y}{\bar{Y}} \cdot 100 = \frac{3}{6} \cdot 100 = 50$$

$$CV_Y > CV_X$$

KIBBUTZ

8. A

$$Q = \frac{P_{75} - P_{25}}{2} = \frac{25-17}{2} = 4$$

9. C

$$z = \frac{X - \overline{X}}{S_x} = 1{,}5 = \frac{X-30}{5} \rightarrow X = (1{,}5 \times 5) + 30 = 37{,}5$$

10. C

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{13{,}5}{5 \times 3} = 0{,}9$$

11. A

$$b = r_{XY}\frac{S_Y}{S_X} = 0{,}9\frac{3}{5} = 0{,}54$$

$$a = \overline{Y} - b\overline{X} = 6 - 0{,}54 \times 30 = -10{,}2$$

Y' = -10,2+0,54X

12. B

$$\frac{(50 \text{ x} 70)}{200} = 17{,}5$$

13. C

14. A

Tratamiento	Mejoría clínica		
	Sí	No	
Cognitivo	40	25	65
Conductual	60	10	70
Farmacológico	50	15	65
	150	50	200

$$P(\text{Tconductual} \cap \text{Sí}) = \frac{60}{200} = 0{,}3$$

15. C

$$P(\text{Tcognitivo} / \text{No}) = \frac{P(\text{Tcognitivo} \cap \text{No})}{P(\text{No})} = \frac{25/200}{50/200} = 0{,}50$$

16. B

17. A

X	f(x)	Xf(x)
4	0,1	0,4
3	0,3	0,9
2	0,4	0,8
1	0,2	0,2
		2,3

$$\mu = \sum x \cdot f(x) = 2{,}3$$

KIBBUTZ

18. C

X	f(x)	F(x)
4	0,1	1
3	0,3	0,9
2	0,4	0,6
1	0,2	0,2

19. C

p=0,30 $P(X \geq 5) = 1 - P(X < 5) = 1 - F(4) = 1 - 0{,}8497 = 0{,}1503$

El valor de F(4) se busca en la Tabla II, y es el valor de la intersección de la fila n=10, x=4 con la columna p=0,30.

20. C

P_{33} de una distribución normal tipificada vale -0,44 (Tablas III)

P_{33} de una distribución normal con $\overline{X} = 100$ y $\sigma = 15$ vale:

$X = z\sigma + \mu \rightarrow X = -0{,}44 \cdot (15) + 100 = 93{,}40$

21. B

$P(X \leq 21{,}0641) = 0{,}90$ (Tabla V para 14 g.l.)

$P(X > 21{,}0641) = 1 - P(X \leq 21{,}0641) = 1 - 0{,}90 = 0{,}10$

22. C

Una variable que sigue una distribución t de Student puede tomar, independientemente de los grados de libertad, valores positivos y negativos y el cero

23. B

$$n = \frac{z_{1-\alpha/2}^2 \sigma^2}{E_{max}^2} = \frac{2{,}58^2 \cdot 3^2}{1{,}2^2} = 41{,}60 = 42$$

24. A

n=121 (Aproximación de la distribución t de Student a la normal)

$$E_{max} = z_{1-\alpha/2} \frac{S_{n-1}}{\sqrt{n}} \rightarrow 1{,}4 = z_{1-\alpha/2} \frac{7{,}865}{\sqrt{121}} \rightarrow z_{1-\alpha/2} = 1{,}96$$

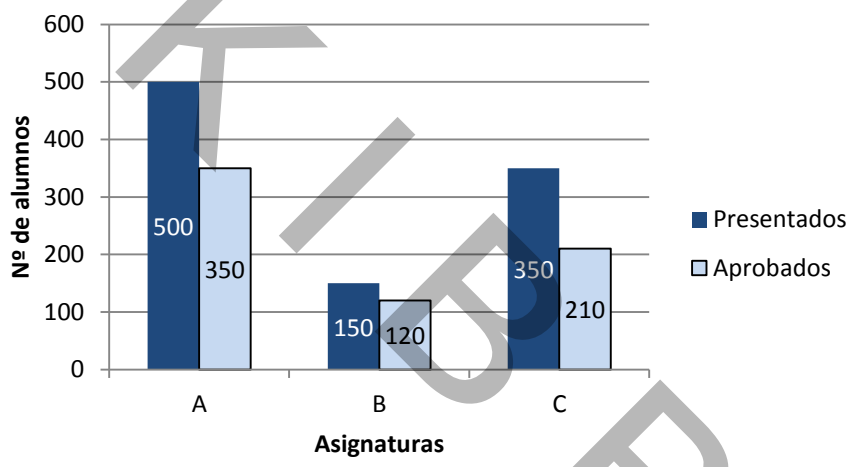
25. C

$L_i = \overline{X} - E_{max} = 22 - 1{,}4 = 20{,}6 \qquad L_s = \overline{X} + E_{max} = 22 + 1{,}4 = 23{,}4$

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2013 Código asignatura: 62011037
MODELO E DURACION: 2 HORAS

Gráfico 1. Número de alumnos matriculados y aprobados en tres asignaturas (A, B y C)



.

Tabla 1. Resultados, en un test X, de un grupo de estudiantes.

X	*niños*	*niñas*
0 - 9	10	8
10 – 19	15	20
20 – 29	20	40
30 – 39	45	12
40 - 49	15	80
50 - 59	35	40
	140	200

Tabla 2. Función de probabilidad de una variable aleatoria X.

x	f(x)
-2	0,1
-1	0,2
0	0,2
1	0,3
2	0,2

Situación 1. Población de tamaño N = 20

Gerardo, Alicia, Antonio, Ernesto, Virginia, Gustavo, Amalia, Teresa, Luis, Cristina, Rosa, Elena, Beatriz, Oscar, Francisco, Tomás, Ángela, Alejandra, Azucena, Belén

1. El Gráfico 1 es un diagrama de: A) barras acumulativo; **B) barras conjunto;** C) dispersión.
2. El "número de alumnos matriculados" del Gráfico 1 es una variable: **A) cuantitativa**; B) cualitativa; C) cuasicuantitativa.
3. ¿Qué escala/s de medida cuenta/n con un valor cero?: A) sólo la de intervalo; B) sólo la de razón**;** **C) la de intervalo y la de razón**.
4. Con los datos del Gráfico 1, ¿en qué asignatura se ha obtenido el mayor porcentaje de alumnos aprobados?: A) en la A; **B) en la B**; C) en la C.
5. Considerando el grupo de niños, de la Tabla 1, ¿cuál es la Moda de la variable X?: A) 40; **B) 34,5;** C) 80.
6. La media en X, para el grupo de niñas de la Tabla 1 es: **A) 37,3**; B) 35,5; C) 40,4 .
7. Para el grupo de niños de la Tabla 1, el Tercer Cuartil vale: A) 29,5; B) 45,5; **C) 49,5**.
8. La varianza para las puntuaciones de las niñas en la Tabla 1 es: A) 14,43; B) 147,68; **C) 208,16.**
9. En un conjunto de puntuaciones típicas el índice de asimetría de Pearson es 0,20, ¿cuánto vale la moda de esa distribución?: **A)-0,20**; B) 0,20; C) 0,80.
10. La amplitud semi-intercuartil para la distribución de los niños de la Tabla 1 es igual a: **A) 12,5**; B) 15,6; C) 20,4.

KIBBUTY

11. Se realizó un estudio con 200 alumnos sobre la relación entre la titularidad pública o privada de los centros escolares a los que pertenecían y el hecho de aprobar o suspender una prueba de rendimiento. Si el estadístico X^2 vale 4,22, el coeficiente de contingencia es igual a: A) 0,02; **B) 0,14**; C) 0,27.
12. Con los datos del gráfico 1, si se estudiara la relación entre las asignaturas y el hecho de aprobar o no mediante el estadístico X^2, ¿cuál es el valor de la frecuencia esperada o teórica de los que se matriculan en la asignatura A y no aprueban?: **A) 160**; B) 255; C) 60.
13. En una recta de regresión de Y sobre X la correlación entre ambas variables es de 0,80, ¿cuánto vale la proporción de la varianza no explicada en la regresión?: **A) 0,36**; B) 0,20; C) 0,64.
14. Con los datos del Gráfico 1, si escogemos al azar un alumno, ¿cuál es la probabilidad de que haya suspendido la asignatura B?: **A) 0,03**; B) 0,12; C) 0,30.
15. Escogiendo al azar un alumno de los que aparecen en el Gráfico 1 resultó haberse presentado a la asignatura A, ¿cuál es la probabilidad de que la haya aprobado?: A) 0,30; B) 0,35; **C) 0,70**.
16. La cuarta parte de los estudiantes, en una clase matemáticas, tiene una motivación baja mientras que el resto tiene una alta motivación. La probabilidad de un alumno de aprobar matemáticas, si tiene una alta motivación es 0,8 y si tiene una motivación baja es 0,4. Escogiendo al azar un alumno de la clase resultó haber aprobado matemáticas, ¿cuál es la probabilidad de que tuviera una motivación baja?: **A) 0,143**; B) 0,254; C) 0,303.
17. La probabilidad de que una variable aleatoria discreta X tome un valor concreto se denota como: A) P(X); **B) f(x);** C) F(x).
18. ¿Cuál es la esperanza matemática de la variable aleatoria X presentada en la Tabla 2?: **A) 0,3**; B) 0,5; C) 1,1.
19. ¿Cuál es la varianza de la variable aleatoria X presentada en la Tabla 2?: A) 0,50; **B) 1,61**; C) 0,98.
20. Las puntuaciones en una escala de *depresión* se distribuyen *normalmente* en una población de 20000 ancianos con $\mu = 10$ y $\sigma = 3$. Sabiendo que una puntuación mayor que 14 en esta escala indica *depresión severa*, ¿cuántos ancianos <u>no presentan</u> *depresión severa*?: A) 1836; B) 9082; **C) 18164.**
21. Una variable X sigue una *distribución Chi-cuadrado* con 6 grados de libertad. Sabemos que $P(X \leq 4{,}049) = 0{,}33$, entonces $P(X > 4{,}049)$: **A) vale 0,67;** B) vale 0; C) no se puede saber con la información dada.
22. Una variable X se distribuye según la *F de Snedecor* con 3 grados de libertad en el numerador y 4 grados de libertad en el denominador. ¿Cuál es la probabilidad de que la variable X tome un valor mayor que 24,259?: **A) 0,005;** B) 0,025; C) 0,05.
23. Hemos extraído una muestra aleatoria n = 5 de la población descrita en la Situación 1 mediante *muestreo estratificado*, con la misma proporción de mujeres y varones que en la población. ¿Cuál de las muestras siguientes podría ser la muestra obtenida?: A) Rosa, Cristina, Oscar, Tomás, Gerardo; B) Beatriz, Belén, Ángela, Elena. Oscar; **C) Teresa, Virginia, Elena, Gerardo, Ernesto**.
24. Hemos medido la variable *hiperactividad* en una muestra de 121 niñas, obteniendo una cuasidesviación típica $S_{n-1} = 3{,}99$. Por otra parte, sabemos que para un nivel de confianza de 0,95, el límite inferior del intervalo de confianza para la media vale 29,29. La media de la muestra vale: A) 26; B) 28; **C) 30**.
25. En una encuesta realizada a 150 personas elegidas aleatoriamente de una determinada comunidad, el 56% se mostró a favor de *incrementar la ayuda a los países emergentes*. Para un nivel de confianza de 0,99, ¿entre qué límites se estima esté la proporción de personas de toda la comunidad a favor del *incremento de la ayuda a los países emergentes*?: A) 0,44 y 0,56; **B) 0,46 y 0,66;** C) 0,01 y 0,99.

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2013 Código asignatura: 62011037
EXAMEN MODELO A

Gráfico 1. Nivel de estudios de los progenitores de una muestra aleatoria y representativa de 2000 estudiantes de Secundaria de una Comunidad Autónoma (2000 padres y 2000 madres).

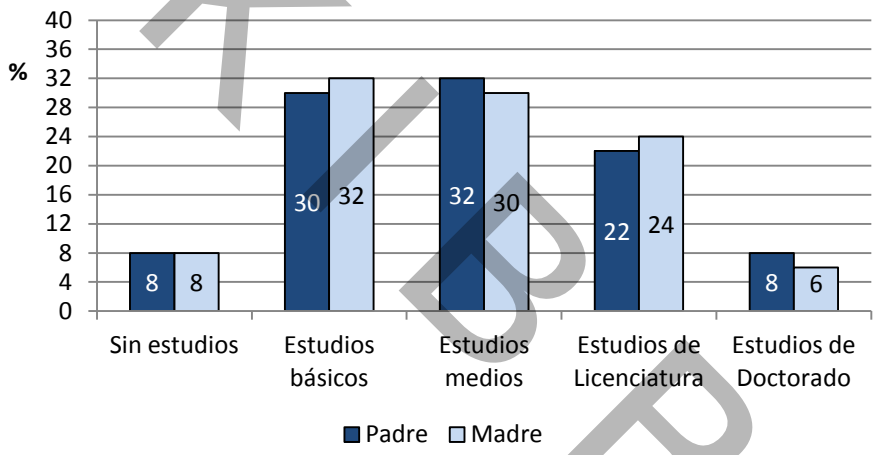


Tabla 1. Resultados de un test de comprensión lectora en un grupo de niños con dislexia.

Comprensión lectora	n_i
9 – 11	10
6 – 8	18
3 – 5	40
0 – 2	32

Tabla 2. Para pronosticar las puntuaciones en una tarea de atención (Y) a partir de las puntuaciones en un test de agudeza visual (X) disponemos de los siguientes datos obtenidos en un grupo de 100 personas:

X	Y	XY	
$\sum X = 1600$	$\sum Y = 2300$	$\sum XY = 39600$	$S_Y = 7$
$\sum X^2 = 28100$	$\sum Y^2 = 57800$		

1. Las conclusiones que se obtengan sobre el nivel de estudios del Gráfico 1: A) se pueden generalizar a la población de estudiantes de Secundaria de la Comunidad Autónoma; B) se pueden generalizar a la población española de estudiantes de Secundaria; C) no se pueden generalizar.
2. La variable "Nivel de estudios", de la Gráfica 1, ¿qué nivel de medida presenta?: A) nominal; B) ordinal; C) de intervalo.
3. El Gráfico 1 es un: A) diagrama de barras acumulativo; B) diagrama de barras conjunto; C) diagrama de dispersión.
4. La media del nivel de estudio de las madres, según los resultados del Gráfico 1 es: A) 2,5; B) 10,5; C) un estadístico que no tiene sentido calcular.
5. ¿Cuánto vale la media en la Comprensión lectora con los datos de la Tabla 1?: A) 55; B) 4,18; C) 25.
6. Con los resultados de la Tabla 1, ¿qué porcentaje de sujetos deja por debajo de sí la puntuación 4?: A) 4%; B) 40%; C) 52%.
7. Con los datos de la Tabla 2, el coeficiente de variación de la variable X es: A) 18,15; B) 24,75; C) 31,25.
8. Con los datos de la Tabla 2, sabiendo que la moda de las puntuaciones en la tarea de atención (Y) es 19, el índice de asimetría de Pearson de estas puntuaciones es: A) 0,26; B) 0,57; C) 0,62.

KIBBUTZ

9. Atendiendo a la Tabla 2, la puntuación típica de una persona que ha obtenido una puntuación directa X=24 en el test de agudeza visual es: A) 1,12; B) 1,60; C) 1,74.
10. Con los datos de la Tabla 2, la pendiente de la recta de regresión para predecir la puntuación en la tarea de atención a partir de la puntuación en el test de agudeza visual, en puntuaciones directas es: A) 0,83; B) 1,12; C) 1,54.
11. Con los datos de la Tabla 2, la ordenada en el origen de la recta de regresión de Y sobre X es: A) 5,08; B) 9,16; C) 15,34.
12. Considerando los datos de la Tabla 2, ¿qué puntuación en la tarea de atención se pronosticará a una persona que ha obtenido en el test de agudeza visual una puntuación de 24?: A) 31,96; B) 27,19; C) 24,78.
13. Con los datos de la Tabla 2, la correlación entre la tarea de atención (Y) y la agudeza visual (X) vale: A) 0,6; B) 0,55; C) 0,8.
14. La probabilidad de un suceso seguro: A) es mayor que 1; B) está comprendido entre 0 y 1; C) es igual a 1.
15. Atendiendo a los resultados del Gráfico 1, ¿cuál es la probabilidad de cursar estudios de Licenciatura?: A) 0,11; B) 0,23; C) no se puede calcular.
16. Suponiendo que se ha elegido una persona al azar, con los datos del Gráfico 1, y ha resultado ser madre, ¿cuál es la probabilidad de que haya cursado estudios básicos?: A) 0,16; B) 0,32; C) 0,62.
17. Si definimos la variable aleatoria X como “el número de veces que sale un 5” al lanzar un dado en 6 ocasiones, ¿cuántos valores diferentes puede adoptar esta variable aleatoria?: A) 5; B) 6; C) 7.
18. Una variable aleatoria discreta X puede adoptar, con la misma probabilidad, los valores 0, 1, 2 y 3. ¿Cuál es su esperanza matemática?: A) 0,75; B) 1,50; C) 1,75.
19. En un centro de desintoxicación, el 10% de los pacientes está ingresado por abuso de sustancias opiáceas. Si elegimos aleatoriamente una muestra de 5 pacientes del centro, ¿cuál es la posibilidad de que más de una esté ingresada en el centro por este motivo?: A) 0,0815; B) 0,3281; C) 0,9185.
20. La variable *grado de dependencia* se distribuye *normalmente* con $\mu = 5$ y $\sigma = 1{,}4$ en la población de personas dependientes de un determinado país. Sabiendo que el gobierno de ese país concede una subvención económica a las personas dependientes con *grado de dependencia* superior a 6, ¿cuál es la probabilidad de que una persona dependiente obtenga la subvención?: A) 0, 2389; B) 0,50; C) 0,7611
21. En el ejercicio anterior, la mediana de la distribución de la variable *grado de dependencia*: A) vale 4; B) vale 5; C) no se puede saber con la información dada
22. Una variable X se distribuye según la *t de Student* con 9 grados de libertad. $P(X \leq -1{,}1)$ es igual a: A) $P(X \leq 1{,}1)$; B) $P(X > 1{,}1)$; C) $1 - P(X > 1{,}1)$.
23. Si una variable X tiene de media $\mu = 80$ y desviación típica $\sigma = 9$ en la población, entonces la *distribución muestral de la media,* para un tamaño de muestra n = 36, tiene: A) de media 80 y desviación típica 2,25; B) de media 80 y desviación típica 1,5; C) de media 80 y desviación típica 0,25.
24. El error de estimación máximo de la media vale 0,98, siendo $\sigma_{\overline{X}} = 0{,}5$. El nivel de confianza es igual a: A) 0,95; B) 0,975; C) 0,99.
25. En una muestra aleatoria de 120 mujeres víctimas de malos tratos, 3 de cada 4 mujeres consiguieron alejarse de su maltratador y rehacer su vida. Para un nivel de confianza de 0,95, ¿entre qué límites se estima esté la proporción poblacional de *mujeres víctimas de malos tratos que consiguieron alejarse de su maltratador y rehacer su vida*?: A) 0,05 y 0,95; B) 0,25 y 0,75; C) 0,67 y 0,83.

EXAMEN

SOLUCIONES:

1. A
2. B
3. B
4. C
5. B

Comprensión lectora	n_i	X_i	n_iX_i
9 – 11	10	10	100
6 – 8	18	7	126
3 – 5	40	4	160
0 – 2	32	1	32
	100		418

$$\overline{X} = \frac{\sum n_i X_i}{n} = \frac{418}{100} = 4{,}18$$

6. C

Comprensión lectora	n_i	n_a
9 – 11	10	100
6 – 8	18	90
3 – 5	40	72
0 – 2	32	32

$$k = \left[\frac{\frac{(P_k - L_i)\cdot n_c}{I} + n_d}{n}\right] \times 100 = \left[\frac{\frac{(4-2{,}5)40}{3} + 32}{100}\right] \times 100 = 52$$

La puntuación 4 se encuentra en el intervalo 3-5.

7. C

X	**Y**	**XY**
$\sum X = 1600$	$\sum Y = 2300$	$\sum XY = 39600$
$\sum X^2 = 28100$	$\sum Y^2 = 55800$	

$$\overline{X} = \frac{1600}{100} = 16$$

$$S_X^2 = \frac{28100}{100} - 16^2 = 281 - 256 = 25 \rightarrow S_X = \sqrt{25} = 5$$

$$CV = \frac{S_X}{\overline{X}} \cdot 100 = \frac{5}{16} \times 100 = 31{,}25$$

8. B

$$\overline{Y} = \frac{2300}{100} = 23$$

$$A_S = \frac{23-19}{7} = 0{,}571428 \cong 0{,}57$$

9. B

$$\overline{X} = \frac{1600}{100} = 16$$

$$S_X = \sqrt{25} = 5$$

KIBBUTZ

$$z = \frac{X - \overline{X}}{S_x} = \frac{24 - 16}{5} = 1{,}6$$

10. B

$$b = \frac{100 \times 39600 - 1600 \times 2300}{100 \times 28100 - (1600)^2} = \frac{3960000 - 3680000}{250000} = 1{,}12$$

11. A

$$\overline{X} = \frac{1600}{100} = 16 \qquad \overline{Y} = \frac{2300}{100} = 23$$

$$a = 23 - 1{,}12 \times 16 = 23 - 17{,}92 = 5{,}08$$

12. A

$$a = 5{,}08$$
$$b = 1{,}12$$
$$Y' = 5{,}08 + 1{,}12X = 5{,}08 + 1{,}12 \times 24 = 31{,}96$$

13. C

$$b = r_{XY}\frac{S_Y}{S_X} \rightarrow r_{XY} = b\frac{S_X}{S_Y} = 1{,}12\frac{5}{7} = 0{,}8$$

14. C.

15. B.

$$\frac{920}{4000} = 0{,}23$$

16. B

$$\frac{640}{2000} = 0{,}32$$

17. C

$$X = \begin{cases} 0 & si\ no\ sale\ ningún\ cinco \\ 1 & si\ sale\ un\ cinco \\ 2 & si\ salen\ dos\ cincos \\ 3 & si\ salen\ tres\ cincos \\ 4 & si\ salen\ cuatro\ cincos \\ 5 & si\ salen\ cinco\ cincos \\ 6 & si\ siempre\ sale\ un\ cinco \end{cases}$$

Por tanto, son 7 los valores diferentes que puede adoptar la variable.

KIBBUTZ

18. B

X	f(x)	Xf(x)
0	0,25	0
1	0,25	0,25
2	0,25	0,50
3	0,25	0,75
		1,50

$$\mu = \sum x \cdot f(x) = 1{,}5$$

19. A

$$P(X > 1) = 1 - P(X \leq 1) = 1 - 0{,}9185 = 0{,}0815$$

El valor de P(X≤1) se busca en la Tabla II, con n=5, x=1 y p=0,10.

20. A

$$z = \frac{x - \mu}{\sigma} = \frac{6-5}{1{,}4} = 0{,}71$$

P(Z ≤ 0,71) = 0,7611 ((Tabla IV)

P(Z > 0,71) = 1 - P(Z ≤ 0,71) = 1- 0,7611 = 0,2389

Por lo tanto,

P(X > 6) = P(Z > 0,71) = 0,2389

21. B

En una distribución normal, la media, la mediana y la moda coinciden. Dado que la media μ = 5, la mediana también vale 5.

22. B

P(X ≤ 1,1) = 0,85 (Tabla VI para 9 grados de libertad)

P(X ≤ -1,1) = P(X > 1,1) = 1 - P(X ≤ 1,1) = 1 – 0,85 = 0,15

23. B

La media de la distribución muestral de la media $(\mu_{\overline{X}})$ es igual a μ y la desviación típica de la distribución muestral de la media $(\sigma_{\overline{X}})$ es igual a $\frac{\sigma}{\sqrt{n}}$

En este ejercicio, μ = 80, σ = 9 y n = 36, por lo que:

$$\mu_{\overline{X}} = \mu = 80$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{9}{6} = 1{,}5$$

24. A

Sabemos que:

$$E_{máx} = z_{1-\alpha/2}\sigma_{\overline{X}}$$

Por lo tanto,

KIBBUTZ

$$0{,}98 = z_{1-\alpha/2}\,0{,}5$$

Despejamos $z_{1-\alpha/2}$:

$$z_{1-\alpha/2} = \frac{0{,}98}{0{,}5} = 1{,}96$$

Vemos en la Tabla IV de la *distribución normal tipificada* que el valor 1,96 deja por debajo de sí el 97,5% % de las observaciones, por lo que $z_{1-\alpha/2} = z_{0,975} = 1{,}96$. Por tanto, el nivel de confianza es:

$1-\alpha/2 = 0{,}975 \rightarrow \alpha/2 = 1 - 0{,}975 = 0{,}025 \rightarrow \alpha = 0{,}025 \times 2 = 0{,}05 \rightarrow 1-\alpha = 1 - 0{,}05 = 0{,}95$

25. C

Proporción de *mujeres víctimas de malos tratos que consiguieron alejarse de su maltratador y rehacer su vida* en la muestra: P = 3/4 = 0,75

Si $1-\alpha = 0{,}95$, entonces $z_{1-\alpha/2} = z_{0,975} = 1{,}96$ (Tabla IV)

$$L_i = P - z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}} = 0{,}75 - 1{,}96\sqrt{\frac{0{,}75\cdot(1-0{,}75)}{120}} = 0{,}67$$

$$L_S = P + z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}} = 0{,}75 + 1{,}96\sqrt{\frac{0{,}75\cdot(1-0{,}75)}{120}} = 0{,}83$$

KIBBUTZ

Código asignatura	Nombre asignatura
62011037	**Introducción al Análisis de Datos**
Fecha alta y origen	Convocatoria
10/02/2015	Septiembre 2012 (Original – Tipo A)
Curso Virtual	

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
SEPTIEMBRE 2012
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

Gráfica 1. Resultados, en porcentajes, a la pregunta: "de los siguientes grupos e instituciones, ¿quiénes cree Ud. que tienen más respeto y preocupación por el medio ambiente? Fuente: Estudio Ciudadanía y conciencia medioambiental en España, publicado por el CIS en 2010.

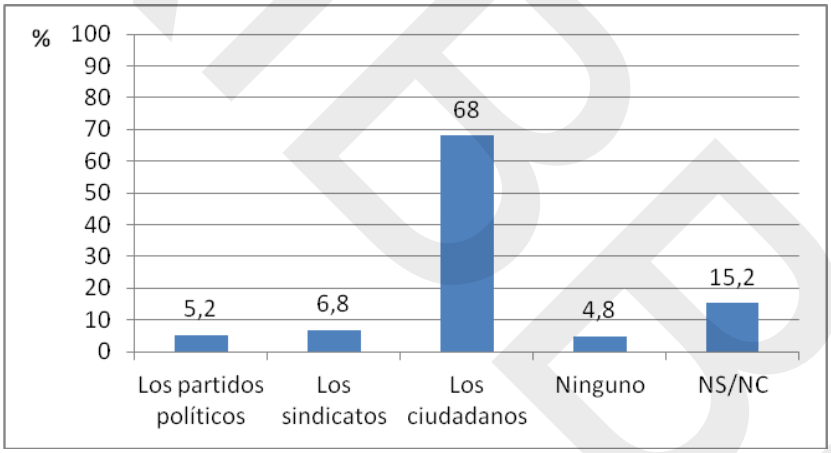


Tabla 1. Número de faltas de ortografía de un grupo de estudiantes de 3º de Primaria en una prueba de dictado.

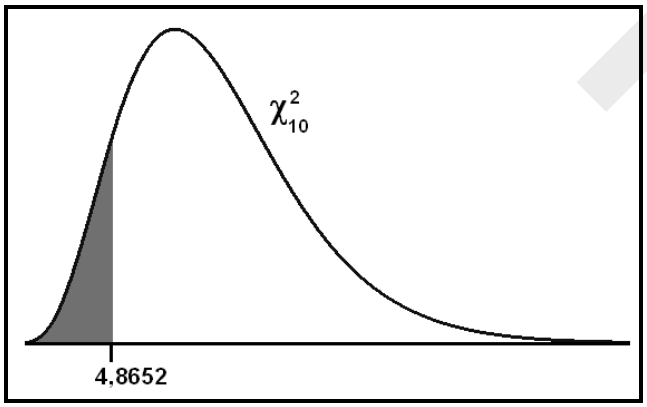
Faltas	n_i
8 ó más	1
6 – 7	4
4 – 5	8
2 – 3	12
0 – 1	5

Tabla 2. Resultados en un cuestionario de hábitos de estudio (X) y la calificación global en 4º de Secundaria (Y) de cinco alumnos de un centro educativo.

Alumno	Hábitos de estudio (X)	Calificación Global (Y)
A	2	6
B	8	5
C	12	10
D	3	7
E	1	3

Situación 1: El 20% de los alumnos de una determinada universidad cursan el Grado en Psicología (P). El 70% de ellos elige el itinerario de Psicología Clínica (C), el 20% el de Psicología Educativa (E) y el 10% el de Psicología del Trabajo (T).

Gráfica 2



1. Sabiendo que 2482 personas respondieron a la pregunta del CIS reflejada en la Gráfica 1, ¿cuántos consideran que los ciudadanos son los más respetuosos y preocupados por el medio ambiente?: A) 68; B) 1688; C) 1935
2. La Gráfica 1 refleja los resultados de una variable: A) nominal; B) ordinal; C) de razón
3. ¿Qué otro tipo de gráfico hubiera resultado apropiado para representar los datos de la Gráfica 1?: A) Diagrama de barras conjunto; B) Histograma; C) Diagrama de sectores

4. ¿Podemos calcular la media con los datos del Tabla 1?: A) Sí, su valor es 3,8; B) No, porque los datos están agrupados en intervalos; C) No, porque el intervalo máximo no tiene límite superior
5. Con los datos de la Tabla 1, ¿qué porcentaje de estudiantes superan las 5 faltas de ortografía?: A) 23%; B) 37%; C) 77%
6. ¿Qué índice NO es una medida de tendencia central?: A) La Mediana; B) La Desviación media; C) La Moda
7. Con los datos de la Tabla 2, la varianza de la calificación global (Y) es igual a: A) 2,31; B) 5,36; C) 4,62
8. En una distribución con media 6 y moda 5, la asimetría es: A) Positiva; B) Negativa; C) No tenemos suficientes datos determinar la asimetría
9. Con los datos de la Tabla 2, la puntuación típica del alumno D en el cuestionario de hábitos de estudio (X) está entre: A) -0,70 y -0,40; B) -0,10 y 0,20; C) 0,40 y 0,70
10. Con los datos de la Tabla 2, la covarianza entre X e Y vale: A) 3,66 ; B) 6,96; C) 9,66
11. Con los datos de la Tabla 2, la pendiente de la recta de regresión en puntuaciones directas para pronosticar las calificaciones globales (Y) a partir del cuestionario de hábitos de estudio (X) es: A) 0,16; B) 0,32; C) 0,40
12. Un coeficiente de correlación de Pearson alto y negativo implica que entre las variables: A) Existe una relación lineal inversa; B) No existe relación; C) Existe relación pero no es lineal
13. En una tabla de contingencia en la que se relacionan el género y otra variable dicotómica, las diferencias entre las frecuencias observadas y las teóricas es igual a cero en las cuatro celdas de la tabla. ¿Cuál sería el valor del coeficiente de contingencia máximo para esa tabla?: A) 1; B) 0,71; C) 0,50
14. Si dos sucesos, A y B, son independientes entonces: A) $P(A \cup B) = P(A) + P(B)$; B) $P(A \cap B) = P(A) + P(B)$; C) $P(A \cap B) = P(A) \cdot P(B)$
15. Teniendo en cuenta la Situación 1, elegimos un alumno al azar ¿cuál es la probabilidad de que no curse el Grado en Psicología?: A) 0,2 ; B) 0,8 ; C) 0,16
16. Teniendo en cuenta la Situación 1, elegimos un alumno al azar ¿cuál es la probabilidad de que curse el Grado en Psicología y haya elegido el itinerario de Psicología Clínica?: A) 0,14 ; B) 0,30 ; C) 0,70
17. Teniendo en cuenta la Situación 1, elegimos al azar 10 alumnos y definimos la variable aleatoria X como "número de alumnos que cursan el Grado en Psicología", la media de X es igual a: A) 2; B) 5; C) 8
18. Teniendo en cuenta la Situación 1, si elegimos al azar 10 alumnos ¿cuál es la probabilidad de que exactamente 2 cursen el Grado en Psicología?: A) 0,20; B) 0,3020; C) 0,5
19. Teniendo en cuenta la Situación 1, si elegimos al azar 10 alumnos ¿cuál es la probabilidad de que ninguno curse el Grado en Psicología?: A) 0,1074; B) 0,4020; C) 0,5328
20. En una distribución normal tipificada, el percentil 67 vale: A) 0,44; B) 0,67; C) 0,7486
21. En la Gráfica 2, el área de color gris vale: A) 0,10; B) 0,50; C) 4,8652
22. Respecto a la Gráfica 2, ¿cuánto vale la probabilidad de un valor igual o inferior a la media de la distribución?: A) 0; B) 0,05; C) las opciones A y B son incorrectas
23. Elija la opción correcta: A) En el muestreo aleatorio simple, todos los elementos de la población tienen la misma probabilidad de ser elegidos para formar parte de la muestra; B) El muestreo casual (o incidental) es un muestreo probabilístico; C) En el muestreo estratificado, se eligen los elementos en algunos estratos de la población
24. En una muestra aleatoria de 100 universitarios, 75 están a favor del movimiento 15-M. Dada la proporción de universitarios a favor del movimiento 15-M hallada en la muestra y para un nivel de confianza del 99%, el error de estimación máximo de la proporción a favor de este movimiento vale: A) 0,04; B) 0,11; C) 0,25
25. Con los datos del ejercicio anterior, ¿entre qué valores se estima se encuentra la proporción de la población universitaria a favor del movimiento 15-M?: A) 0,50 y 1; B) 0,64 y 0,86; C) 0,71 y 0,79

SOLUCIONES:

1. B
2. A
3. C
4. C
5. A

Faltas	n_i	n_a
8 o más	1	30
6 – 7	4	29
4 – 5	8	25
2 – 3	12	17
0 – 1	5	5

La puntuación 5 se encuentra en el intervalo 4-5.

$$k = \left[\frac{\frac{(P_k - L_i)\cdot n_c}{I} + n_d}{n} \right] \times 100 = \left[\frac{\frac{(5-3{,}5)8}{2} + 17}{30} \right] \times 100 = 76{,}67 \approx 77$$

$100 - 77 = 23$

6. B
7. B

Persona	Y_i	Y_i^2
A	6	36
B	5	25
C	10	100
D	7	49
E	3	9
	31	219

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{31}{5} = 6{,}2$$

$$S_Y^2 = \frac{\sum Y_i^2}{n} - \overline{Y}_i^2 = \frac{219}{5} - 6{,}2^2 = 5{,}36$$

8. A

Si la media es mayor que la moda, la distribución presenta asimetría positiva

9. A

Persona	X_i	X_i^2
A	2	4
B	8	64
C	12	144
D	3	9
E	1	1
	26	222

KEBBUTZ

$$\overline{X} = \frac{\sum X_i}{n} = \frac{26}{5} = 5{,}2 \qquad S_X^2 = \frac{\sum X_i^2}{n} - \overline{X}_i^2 = \frac{222}{5} - 5{,}2^2 = 17{,}36 \qquad S_X = \sqrt{S_X^2} = \sqrt{17{,}36} = 4{,}17$$

$$z = \frac{3 - 5{,}2}{4{,}17} = -0{,}53$$

10. B

Persona	Hábitos de estudio (X)	Calificación (Y)	XY
A	2	6	12
B	8	5	40
C	12	10	120
D	3	7	21
E	1	3	3
	26	31	196

$$S_{XY} = \frac{\sum XY}{n} - (\overline{X} \cdot \overline{Y}) = \frac{196}{5} - (5{,}2 \cdot 6{,}2) = 39{,}2 - 32{,}24 = 6{,}96$$

11. C

Persona	Hábitos de estudio (X)	Calificación (Y)	XY	X^2
A	2	6	12	4
B	8	5	40	64
C	12	10	120	144
D	3	7	21	9
E	1	3	3	1
	26	31	196	222

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{5 \times 196 - 26 \times 31}{5 \times 222 - 26^2} = \frac{174}{434} = 0{,}40$$

12. A

13. B

14. C

15. B

P(NoP)=1-P(P)=1-0,20=0,80

16. A

$$P(P \cap C) = P(P) \cdot P(C/P) = 0{,}20 \cdot 0{,}70 = 0{,}14$$

17. A

$$np = 10 \cdot 0{,}2 = 2$$

18. B

Tabla I con n=10, p=0,2 y x=2

19. A

Tabla I con n=10, p=0,2 y x=0

20. A

P_{67} = 0,44 (Tabla IV)

21. A

P(X ≤ 4, 8652) = 0,10 (Tabla V para 10 g.l.)

22. C

$\mu = n = 10$

Como se ve en la Tabla V para 10 g.l., la $P(X \leq 10)$ es un valor entre 0,10 y 0,90

23. A

24. B

$$E_{máx} = z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}} = 2{,}58\sqrt{\frac{0{,}75\cdot(0{,}25)}{100}} = 0{,}1117 \cong 0{,}11$$

25. B

$P - E_{máx} = 0{,}75 - 0{,}11 = 0{,}64$

$P + E_{máx} = 0{,}75 + 0{,}11 = 0{,}86$

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2012
EXAMEN TIPO TEST MODELO E

Tabla 1: Resultados en un test de agudeza visual (X) de siete personas en una investigación sobre la miopía.

Persona	X_i
1	20
2	6
3	9
4	6
5	12
6	8
7	16

Tabla 2: Puntuaciones de 200 universitarios en una escala de actitudes agrupadas en intervalos y las frecuencias absolutas (n_i) de cada intervalo. La varianza de esta distribución es igual a 132,84.

X	n_i
64-69	4
58-63	16
52-57	14
46-51	22
40-45	32
34-39	44
28-33	42
22-27	18
16-21	8

Tabla 3: En un estudio se investigó la relación entre la ansiedad de ejecución (Baja ó Alta) y la realización correcta de una tarea (Sí, No) en 400 personas. En la tabla se muestran algunas frecuencias observadas y, debajo y entre paréntesis, algunas frecuencias teóricas.

		Tarea realizada		
		Sí	**No**	
Ansiedad	**Baja**	68 (84,8)	--- (---)	160
	Alta	--- (---)	96 (112,8)	---
		212	---	400

1. En una escala de intervalo se han asignado a tres sujetos los valores X=0, X=3, y X=6. Con esta escala es correcto afirmar que: A) el tercer sujeto con X=6 tiene el doble de la variable medida que el segundo sujeto con X=3; B) el primer sujeto presenta carencia absoluta de la variable medida; **C) la diferencia entre X=0 y X=3 es la misma que entre X=3 y X=6 en la variable medida**.
2. Para una representación gráfica de la distribución conjunta de las dos variables de la tabla 3 debemos utilizar: **A) un diagrama de barras conjunto**; B) un diagrama de dispersión; C) un polígono de frecuencias
3. La amplitud del intervalo en la distribución de frecuencias de la tabla 2 es: A) 5; B)4; **C) 6**.
4. En los datos de la tabla 2, la media es igual a: A) 36,5; **B) 39,5**; C) 33,5.
5. Atendiendo a la tabla 2, el decil 2 es igual a: A) 28,5; B) 30,5; **C) 29,5.**
6. Con los datos de la tabla 1, ¿cuál de las siguientes medidas de tendencia central tendrá el valor más alto?: **A) la media;** B) la mediana; C) la moda.
7. Una característica del índice desviación media es que: A) puede ser negativo; **B) su valor siempre es igual o mayor que cero;** C) siempre va a valer cero.
8. Con los datos de la tabla 1, ¿cuál es la puntuación típica correspondiente al sujeto 2?: **A) -1,01**; B) -5; C) -0,41.
9. El índice de asimetría de Pearson de la distribución de la tabla 2 es: A) -0,26; B) 0; **C) 0,26.**

KIBBUTZ

10. En la tabla 3, ¿se puede calcular el coeficiente de contingencia máximo?: A) sí, porque el número de sujetos es grande; B) no porque el número de filas es distinto al número de columnas; **C) sí porque el número de filas es igual al número de columnas.**
11. Con los datos de la tabla 3, el valor de X^2 es igual a: A) 7,17; **B) 11,80** C) 16,8.
12. Hemos medido el grado de miopía (Y) de las siete personas de la tabla 1 y, a continuación, hemos calculado la recta de regresión de Y sobre X. Sabiendo que la correlación de Pearson entre el test de agudeza visual y el grado de miopía es de 0,68 y que la varianza de las puntuaciones pronosticadas, Y', es igual a 4, ¿cuál es la pendiente de la recta de regresión de Y sobre X?: A) 3,4; B) -0,28; **C) 0,41.**
13. Con los datos de la tabla 1 y del ejercicio anterior, ¿cuál es el valor de la varianza de los errores?: **A) 4,65**; B) 1,88; C) 1.
14. En un espacio muestral hay dos sucesos A y B. Sabemos que $P(A) = 0{,}4$, $P(\overline{B}) = 0{,}75$ y que $P(A \cup B) = 0{,}55$. ¿Cuál es la probabilidad de la intersección de los sucesos A y B?: **A) 0,10**; B) 0,60; C) 0,15.
15. Se realiza un estudio sobre el hábito de fumar en adultos de mediana edad con 200 hombres y 300 mujeres. Un 30% de los hombres reconocen que sí fuman habitualmente, mientras que 225 mujeres se declaran no fumadoras. Si seleccionamos una persona al azar, ¿cuál es la probabilidad de que sea hombre y no fume?: A) 0,70; **B) 0,28**; C) 0,73.
16. Con los datos del ejercicio anterior, si elegimos una persona al azar, ¿cuál es la probabilidad de que fume?: A) 0,32; B) 0,73; **C) 0,27.**
17. Dada una variable aleatoria discreta, X, la función que utilizamos para calcular la probabilidad de que la variable tome un valor menor o igual que un determinado valor se denomina: **A) función de distribución**; B) función de densidad de probabilidad; C) función de probabilidad.
18. Si lanzamos un dado al aire, cuyas caras están numeradas del 1 al 6, y definimos la variable aleatoria "número obtenido", ¿cuál es la esperanza matemática de esta variable?: A) 1,83; **B) 3,5**; C) 7,2.
19. En un centro educativo, la probabilidad de que un alumno suspenda una asignatura es de 0,2. Si seleccionamos al azar a 4 alumnos, ¿cuál es la probabilidad de que el 75% de esos alumnos aprueben la asignatura?: A) 0,0256; **B) 0,4096**; C) 0,25.
20. Las puntuaciones en una prueba de psicomotricidad siguen una distribución normal con varianza 9. Si el primer cuartil vale 4, ¿cuál es la media de las puntuaciones en psicomotricidad?: A) 0; B) 1,99; **C) 6,01.**
21. En una variable aleatoria X que sigue la distribución Chi-cuadrado con varianza igual a 40, la probabilidad de que $(10{,}8508 \leq X \leq 28{,}4120)$ es: A) 0,95; **B) 0,85**; C) 0,90.
22. Una variable aleatoria se distribuye según la *t* de Student con varianza igual a 1,20. ¿Cuál es el valor de la distribución que deja por encima al 20% de las observaciones?: **A) 0,873**; B) 0,849; C) 0,259.
23. En el muestreo aleatorio simple, la probabilidad de que cada individuo sea elegido de la población: **A) es la misma para todos los individuos**; B) es distinta para cada individuo; C) es la misma para algunos individuos
24. En una investigación sobre la inteligencia emocional, se ha extraído una muestra de 100 personas de una determinada población. Se sabe que la varianza poblacional es 16 y que el límite superior en la estimación por intervalo de la media es 3,784, con un nivel de confianza de 0,95. ¿Cuánto vale la media de la muestra seleccionada?: A) 1,7; **B) 3**; C) 2,5.
25. Se ha aplicado un test de rendimiento a una muestra de125 personas de las que un 30% han aprobado el test. Para un nivel de confianza de 0,95, ¿entre qué límites se estima está la proporción de aprobados de la población?: A) 0,20 y 0,40; B) 0,27 y 0,33; **C) 0,22 y 0,38.**

KUBUTY

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2012
EXAMEN TIPO TEST MODELO A

Tabla 1: Resultados en un test de agudeza visual (X) de siete personas en una investigación sobre la miopía.

Persona	X_i
1	20
2	6
3	9
4	6
5	12
6	8
7	16

Tabla 2: Puntuaciones de 200 universitarios en una escala de actitudes agrupadas en intervalos y las frecuencias absolutas (n_i) de cada intervalo. La varianza de esta distribución es igual a 132,84.

X	n_i
64-69	4
58-63	16
52-57	14
46-51	22
40-45	32
34-39	44
28-33	42
22-27	18
16-21	8

Tabla 3: En un estudio se investigó la relación entre la ansiedad de ejecución (Baja ó Alta) y la realización correcta de una tarea (Sí, No) en 400 personas. En la tabla se muestran algunas frecuencias observadas y, debajo y entre paréntesis, algunas frecuencias teóricas.

		Tarea realizada		
		Sí	**No**	
Ansiedad	**Baja**	68 (84,8)	--- (---)	160
	Alta	--- (---)	96 (112,8)	---
		212	---	400

1. En una escala de intervalo se han asignado a tres sujetos los valores X=0, X=3, y X=6. Con esta escala es correcto afirmar que: A) el tercer sujeto con X=6 tiene el doble de la variable medida que el segundo sujeto con X=3; B) el primer sujeto presenta carencia absoluta de la variable medida; C) la diferencia entre X=0 y X=3 es la misma que entre X=3 y X=6 en la variable medida.
2. La amplitud del intervalo en la distribución de frecuencias de la tabla 2 es: A) 5; B)4; C) 6.
3. Para una representación gráfica de la distribución conjunta de las dos variables de la tabla 3 debemos utilizar: A) un diagrama de barras conjunto; B) un diagrama de dispersión; C) un polígono de frecuencias.
4. Con los datos de la tabla 1, ¿cuál de las siguientes medidas de tendencia central tendrá el valor más alto?: A) la media; B) la mediana; C) la moda.
5. En los datos de la tabla 2, la media es igual a: A) 36,5; B) 39,5; C) 33,5.
6. Atendiendo a la tabla 2, el decil 2 es igual a: A) 28,5; B) 30,5; C) 29,5.
7. Una característica del índice desviación media es que: A) puede ser negativo; B) su valor siempre es igual o mayor que cero; C) siempre va a valer cero.
8. El índice de asimetría de Pearson de la distribución de la tabla 2 es: A) -0,26; B) 0; C) 0,26.
9. Con los datos de la tabla 1, ¿cuál es la puntuación típica correspondiente al sujeto 2?: A) -1,01; B) -5; C) -0,41.
10. Con los datos de la tabla 3, el valor de X^2 es igual a: A) 7,17; B) 11,80; C) 16,8.

KIBBUTZ

11. En la tabla 3, ¿se puede calcular el coeficiente de contingencia máximo?:A) sí, porque el número de sujetos es grande; B) no porque el número de filas es distinto al número de columnas; C) sí porque el número de filas es igual al número de columnas.
12. Hemos medido el grado de miopía (Y) de las siete personas de la tabla 1 y, a continuación, hemos calculado la recta de regresión de Y sobre X. Sabiendo que la correlación de Pearson entre el test de agudeza visual y el grado de miopía es de 0,68 y que la varianza de las puntuaciones pronosticadas, Y', es igual a 4, ¿cuál es la pendiente de la recta de regresión de Y sobre X?: A) 3,4; B) -0,28; C) 0,41.
13. Con los datos de la tabla 1 y del ejercicio anterior, ¿cuál es el valor de la varianza de los errores?: A) 4,65; B) 1,88; C) 1.
14. En un espacio muestral hay dos sucesos A y B. Sabemos que $P(A) = 0{,}4$, $P(\overline{B}) = 0{,}75$ y que $P(A \cup B) = 0{,}55$. ¿Cuál es la probabilidad de la intersección de los sucesos A y B?: A) 0,10; B) 0,60; C) 0,15.
15. Se realiza un estudio sobre el hábito de fumar en adultos de mediana edad con 200 hombres y 300 mujeres. Un 30% de los hombres reconocen que sí fuman habitualmente, mientras que 225 mujeres se declaran no fumadoras. Si seleccionamos una persona al azar, ¿cuál es la probabilidad de que sea hombre y no fume?: A) 0,70; B) 0,28; C) 0,73.
16. Con los datos del ejercicio anterior, si elegimos una persona al azar, ¿cuál es la probabilidad de que fume?: A) 0,32; B) 0,73; C) 0,27.
17. Dada una variable aleatoria discreta, X, la función que utilizamos para calcular la probabilidad de que la variable tome un valor menor o igual que un determinado valor se denomina: A) función de distribución; B) función de densidad de probabilidad; C) función de probabilidad.
18. Si lanzamos un dado al aire, cuyas caras están numeradas del 1 al 6, y definimos la variable aleatoria "número obtenido", ¿cuál es la esperanza matemática de esta variable?: A) 1,83; B) 3,5; C) 7,2.
19. En un centro educativo, la probabilidad de que un alumno suspenda una asignatura es de 0,2. Si seleccionamos al azar a 4 alumnos, ¿cuál es la probabilidad de que el 75% de esos alumnos aprueben la asignatura?: A) 0,0256; B) 0,4096; C) 0,25.
20. Las puntuaciones en una prueba de psicomotricidad siguen una distribución normal con varianza 9. Si el primer cuartil vale 4, ¿cuál es la media de las puntuaciones en psicomotricidad?: A) 0; B) 1,99; C) 6,01.
21. Una variable aleatoria se distribuye según la *t* de Student con varianza igual a 1,20. ¿Cuál es el valor de la distribución que deja por encima al 20% de las observaciones?: A) 0,873; B) 0,849; C) 0,259.
22. En una variable aleatoria X que sigue la distribución Chi-cuadrado con varianza igual a 40, la probabilidad de que $(10{,}8508 \leq X \leq 28{,}4120)$ es: A) 0,95; B) 0,85; C) 0,90.
23. En el muestreo aleatorio simple, la probabilidad de que cada individuo sea elegido de la población: A) es la misma para todos los individuos; B) es distinta para cada individuo; C) es la misma para algunos individuos
24. En una investigación sobre la inteligencia emocional, se ha extraído una muestra de 100 personas de una determinada población. Se sabe que la varianza poblacional es 16 y que el límite superior en la estimación por intervalo de la media es 3,784, con un nivel de confianza de 0,95. ¿Cuánto vale la media de la muestra seleccionada?: A) 1,7; B) 3; C) 2,5.
25. Se ha aplicado un test de rendimiento a una muestra de125 personas de las que un 30% han aprobado el test. Para un nivel de confianza de 0,95, ¿entre qué límites se estima está la proporción de aprobados de la población? A) 0,20 y 0,40; B) 0,27 y 0,33; C) 0,22 y 0,38.

INTERNET

SOLUCIONES:

1. C
2. C
3. A
4. A

$$\overline{X} = \frac{77}{7} = 11$$

Para calcular la mediana, ordenamos primero los datos de menor a mayor:

6, 6, 8, 9, 12, 16, 20

Dado que n=7 es impar, la mediana es el valor central, Md=9

La Mo=6

5. B

X	n_i	X_i	n_iX_i
64-69	4	66,5	266
58-63	16	60,5	968
52-57	14	54,5	763
46-51	22	48,5	1067
40-45	32	42,5	1360
34-39	44	36,5	1606
28-33	42	30,5	1281
22-27	18	24,5	441
16-21	8	18,5	148
	200		7900

$$\overline{X} = \frac{7900}{200} = 39{,}5$$

6. C

X	n_i	X_i	n_a
64-69	4	66,5	200
58-63	16	60,5	196
52-57	14	54,5	180
46-51	22	48,5	166
40-45	32	42,5	144
34-39	44	36,5	112
28-33	42	30,5	68
22-27	18	24,5	26
16-21	8	18,5	8
	200		

Dado que el decil 2 es el percentil 20, calculamos directamente el percentil 20 de la distribución.

$$P_{20} = L_i + \left(\frac{\frac{n \cdot k}{100} - n_d}{n_c} \right) \cdot I = 27{,}5 + \left(\frac{\frac{200 \cdot 20}{100} - 26}{42} \right) \cdot 6 = 29{,}5$$

KIBBUTZ

7. B
8. C

$$S_X^2 = 132{,}84 \rightarrow S_X = \sqrt{132{,}84} = 11{,}525 \approx 11{,}53$$

$$A_s = \frac{\overline{X} - Mo}{S_x} = \frac{39{,}5 - 36{,}5}{11{,}53} = 0{,}26$$

9. A

Persona	X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
1	20	9	81
2	6	-5	25
3	9	-2	4
4	6	-5	25
5	12	1	1
6	8	-3	9
7	16	5	25
			170

$$S_X^2 = \frac{170}{7} = 24{,}29 \qquad S_X = 4{,}93 \qquad \overline{X} = 11 \qquad z = \frac{-5}{4{,}93} = -1{,}01$$

10. B

		Tarea realizada		
		Sí	**No**	
Ansiedad	**Baja**	68 (84,8)	92 (75,2)	160
	Alta	144 (127,2)	96 (112,8)	240
		212	188	400

$$X^2 = \frac{(68-84{,}8)^2}{84{,}8} + \frac{(92-75{,}2)^2}{75{,}2} + \frac{(144-127{,}2)^2}{127{,}2} + \frac{(96-112{,}8)^2}{112{,}8} = 11{,}80$$

11. C

12. C

$$b = r_{XY}\frac{S_Y}{S_X}$$

$$S_Y^2 = \frac{S_{Y'}^2}{r_{XY}^2} = \frac{4}{0{,}68^2} = 8{,}65 \rightarrow b = r_{XY}\frac{S_Y}{S_X} = 0{,}68\frac{\sqrt{8{,}65}}{4{,}93} = 0{,}4056 \cong 0{,}41$$

13. A

$$S_E^2 = S_Y^2 - S_{Y'}^2 = 8{,}65 - 4 = 4{,}65$$

KABBUTZ

14. A

$$P(B) = 1 - P(\overline{B}) = 1 - 0{,}75 = 0{,}25$$
$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0{,}40 + 0{,}25 - 0{,}55 = 0{,}10$$

15. B

		Fumador		
		Sí	No	
Sexo	Hombre	60	140	200
	Mujer	75	225	300
		135	365	500

$$P(H \cap No) = \frac{140}{500} = 0{,}28$$

16. C

$$P(Sí) = \frac{135}{500} = 0{,}27$$

17. A

18. B

x	f(x)	$x \cdot f(x)$
1	1/6	1/6
2	1/6	2/6
3	1/6	3/6
4	1/6	4/6
5	1/6	5/6
6	1/6	6/6
		21/6

$$\mu = x \cdot f(x) = \frac{21}{6} = 3{,}5$$

19. B
Tabla I de la Binomial con n = 4, x=1 y p=0,20

20. C
P_{25}=4 y la puntuación típica que deja por debajo al 25% de los sujetos es z=-0,67

$$\frac{4-\mu}{3} = -0{,}67 \rightarrow \mu = 6{,}01$$

21. A

$$\sigma^2 = \frac{n}{n-2} = 1{,}20 \rightarrow n = 12 \text{ grados de libertad}$$
$$P(T > t_{12}) = 0{,}20 \rightarrow P(T \leq t_{12}) = 0{,}80$$
En la tabla VI, $t_{12} = 0{,}873$

KIBBUTZ

22. B

$$\sigma^2 = 2n = 40 \rightarrow n = 20$$

Buscamos en la tabla V de la Chi-cuadrado las probabilidades, por lo que:

$$P(10{,}8508 \le X \le 28{,}4120) = P(X \le 28{,}4120) - P(X \le 10{,}8508) = 0{,}90 - 0{,}05 = 0{,}85$$

23. A

En el muestreo aleatorio simple, cada individuo de la población tiene la misma probabilidad de ser elegido.

24. B

$$nc = 0{,}95 \Rightarrow z_{1-\alpha/2} = 1{,}96 \quad \text{(tabla IV)}$$

$$E_{máx} = z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = 1{,}96\frac{4}{\sqrt{100}} = 0{,}784$$

$$L_S = \overline{X} + E_{máx} \rightarrow \overline{X} = L_S - E_{máx} = 3{,}784 - 0{,}784 = 3$$

25. C

$$nc = 0{,}95 \Rightarrow z_{1-\alpha/2} = 1{,}96 \quad \text{(tabla IV)}$$

Proporción de aprobados en la muestra: P=0,30

$$L_i = P - z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}} = 0{,}30 - 1{,}96\sqrt{\frac{0{,}30\cdot(1-0{,}30)}{125}} = 0{,}2196 \cong 0{,}22$$

$$L_S = P + z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}} = 0{,}30 + 1{,}96\sqrt{\frac{0{,}30\cdot(1-0{,}30)}{125}} = 0{,}3803 \cong 0{,}38$$

KIBBUTZ

Código asignatura	Nombre asignatura
62011037	**Introducción al Análisis de Datos**
Fecha alta y origen	Convocatoria
10/02/2015	Septiembre 2012 (Original – Tipo A)
Curso Virtual	

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
SEPTIEMBRE 2012
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

Gráfica 1. Resultados, en porcentajes, a la pregunta: "de los siguientes grupos e instituciones, ¿quiénes cree Ud. que tienen más respeto y preocupación por el medio ambiente? Fuente: Estudio Ciudadanía y conciencia medioambiental en España, publicado por el CIS en 2010.

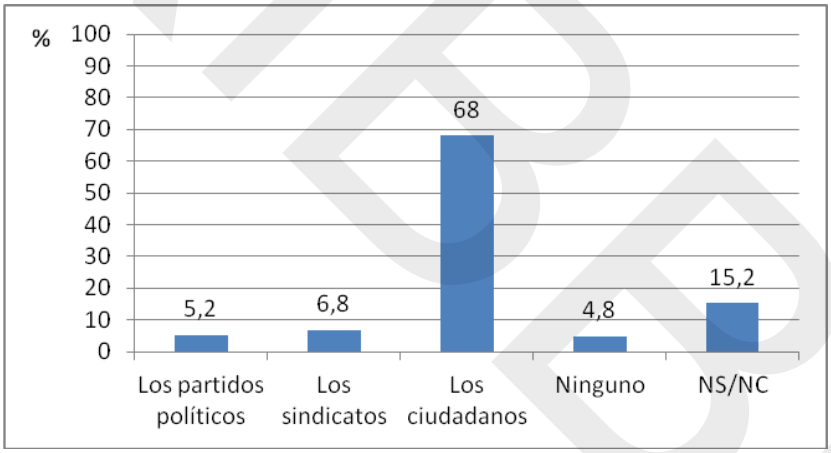


Tabla 1. Número de faltas de ortografía de un grupo de estudiantes de 3º de Primaria en una prueba de dictado.

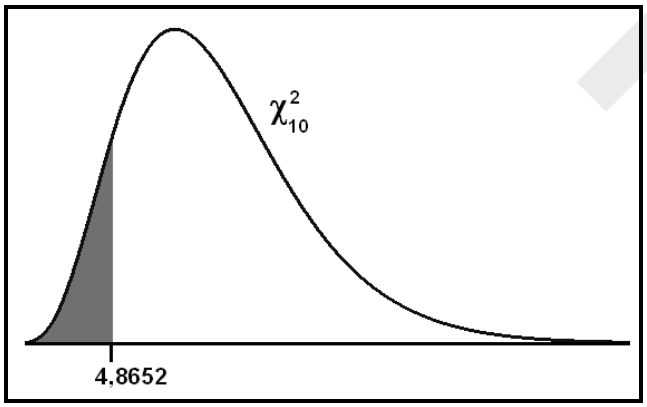
Faltas	n_i
8 ó más	1
6 – 7	4
4 – 5	8
2 – 3	12
0 – 1	5

Tabla 2. Resultados en un cuestionario de hábitos de estudio (X) y la calificación global en 4º de Secundaria (Y) de cinco alumnos de un centro educativo.

Alumno	Hábitos de estudio (X)	Calificación Global (Y)
A	2	6
B	8	5
C	12	10
D	3	7
E	1	3

Situación 1: El 20% de los alumnos de una determinada universidad cursan el Grado en Psicología (P). El 70% de ellos elige el itinerario de Psicología Clínica (C), el 20% el de Psicología Educativa (E) y el 10% el de Psicología del Trabajo (T).

Gráfica 2



1. Sabiendo que 2482 personas respondieron a la pregunta del CIS reflejada en la Gráfica 1, ¿cuántos consideran que los ciudadanos son los más respetuosos y preocupados por el medio ambiente?: A) 68; B) 1688; C) 1935
2. La Gráfica 1 refleja los resultados de una variable: A) nominal; B) ordinal; C) de razón
3. ¿Qué otro tipo de gráfico hubiera resultado apropiado para representar los datos de la Gráfica 1?: A) Diagrama de barras conjunto; B) Histograma; C) Diagrama de sectores

4. ¿Podemos calcular la media con los datos del Tabla 1?: A) Sí, su valor es 3,8; B) No, porque los datos están agrupados en intervalos; C) No, porque el intervalo máximo no tiene límite superior
5. Con los datos de la Tabla 1, ¿qué porcentaje de estudiantes superan las 5 faltas de ortografía?: A) 23%; B) 37%; C) 77%
6. ¿Qué índice NO es una medida de tendencia central?: A) La Mediana; B) La Desviación media; C) La Moda
7. Con los datos de la Tabla 2, la varianza de la calificación global (Y) es igual a: A) 2,31; B) 5,36; C) 4,62
8. En una distribución con media 6 y moda 5, la asimetría es: A) Positiva; B) Negativa; C) No tenemos suficientes datos determinar la asimetría
9. Con los datos de la Tabla 2, la puntuación típica del alumno D en el cuestionario de hábitos de estudio (X) está entre: A) -0,70 y -0,40; B) -0,10 y 0,20; C) 0,40 y 0,70
10. Con los datos de la Tabla 2, la covarianza entre X e Y vale: A) 3,66 ; B) 6,96; C) 9,66
11. Con los datos de la Tabla 2, la pendiente de la recta de regresión en puntuaciones directas para pronosticar las calificaciones globales (Y) a partir del cuestionario de hábitos de estudio (X) es: A) 0,16; B) 0,32; C) 0,40
12. Un coeficiente de correlación de Pearson alto y negativo implica que entre las variables: A) Existe una relación lineal inversa; B) No existe relación; C) Existe relación pero no es lineal
13. En una tabla de contingencia en la que se relacionan el género y otra variable dicotómica, las diferencias entre las frecuencias observadas y las teóricas es igual a cero en las cuatro celdas de la tabla. ¿Cuál sería el valor del coeficiente de contingencia máximo para esa tabla?: A) 1; B) 0,71; C) 0,50
14. Si dos sucesos, A y B, son independientes entonces: A) $P(A \cup B) = P(A) + P(B)$; B) $P(A \cap B) = P(A) + P(B)$; C) $P(A \cap B) = P(A) \cdot P(B)$
15. Teniendo en cuenta la Situación 1, elegimos un alumno al azar ¿cuál es la probabilidad de que no curse el Grado en Psicología?: A) 0,2 ; B) 0,8 ; C) 0,16
16. Teniendo en cuenta la Situación 1, elegimos un alumno al azar ¿cuál es la probabilidad de que curse el Grado en Psicología y haya elegido el itinerario de Psicología Clínica?: A) 0,14 ; B) 0,30 ; C) 0,70
17. Teniendo en cuenta la Situación 1, elegimos al azar 10 alumnos y definimos la variable aleatoria X como "número de alumnos que cursan el Grado en Psicología", la media de X es igual a: A) 2; B) 5; C) 8
18. Teniendo en cuenta la Situación 1, si elegimos al azar 10 alumnos ¿cuál es la probabilidad de que exactamente 2 cursen el Grado en Psicología?: A) 0,20; B) 0,3020; C) 0,5
19. Teniendo en cuenta la Situación 1, si elegimos al azar 10 alumnos ¿cuál es la probabilidad de que ninguno curse el Grado en Psicología?: A) 0,1074; B) 0,4020; C) 0,5328
20. En una distribución normal tipificada, el percentil 67 vale: A) 0,44; B) 0,67; C) 0,7486
21. En la Gráfica 2, el área de color gris vale: A) 0,10; B) 0,50; C) 4,8652
22. Respecto a la Gráfica 2, ¿cuánto vale la probabilidad de un valor igual o inferior a la media de la distribución?: A) 0; B) 0,05; C) las opciones A y B son incorrectas
23. Elija la opción correcta: A) En el muestreo aleatorio simple, todos los elementos de la población tienen la misma probabilidad de ser elegidos para formar parte de la muestra; B) El muestreo casual (o incidental) es un muestreo probabilístico; C) En el muestreo estratificado, se eligen los elementos en algunos estratos de la población
24. En una muestra aleatoria de 100 universitarios, 75 están a favor del movimiento 15-M. Dada la proporción de universitarios a favor del movimiento 15-M hallada en la muestra y para un nivel de confianza del 99%, el error de estimación máximo de la proporción a favor de este movimiento vale: A) 0,04; B) 0,11; C) 0,25
25. Con los datos del ejercicio anterior, ¿entre qué valores se estima se encuentra la proporción de la población universitaria a favor del movimiento 15-M?: A) 0,50 y 1; B) 0,64 y 0,86; C) 0,71 y 0,79

SOLUCIONES:

1. B
2. A
3. C
4. C
5. A

Faltas	n_i	n_a
8 o más	1	30
6 – 7	4	29
4 – 5	8	25
2 – 3	12	17
0 – 1	5	5

La puntuación 5 se encuentra en el intervalo 4-5.

$$k = \left[\frac{\frac{(P_k - L_i)\cdot n_c}{I} + n_d}{n} \right] \times 100 = \left[\frac{\frac{(5-3{,}5)8}{2} + 17}{30} \right] \times 100 = 76{,}67 \approx 77$$

$100 - 77 = 23$

6. B
7. B

Persona	Y_i	Y_i^2
A	6	36
B	5	25
C	10	100
D	7	49
E	3	9
	31	219

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{31}{5} = 6{,}2$$

$$S_Y^2 = \frac{\sum Y_i^2}{n} - \overline{Y}_i^2 = \frac{219}{5} - 6{,}2^2 = 5{,}36$$

8. A

Si la media es mayor que la moda, la distribución presenta asimetría positiva

9. A

Persona	X_i	X_i^2
A	2	4
B	8	64
C	12	144
D	3	9
E	1	1
	26	222

KEBBUTZ

$$\overline{X} = \frac{\sum X_i}{n} = \frac{26}{5} = 5{,}2 \qquad S_X^2 = \frac{\sum X_i^2}{n} - \overline{X}_i^2 = \frac{222}{5} - 5{,}2^2 = 17{,}36 \qquad S_X = \sqrt{S_X^2} = \sqrt{17{,}36} = 4{,}17$$

$$z = \frac{3 - 5{,}2}{4{,}17} = -0{,}53$$

10. B

Persona	Hábitos de estudio (X)	Calificación (Y)	XY
A	2	6	12
B	8	5	40
C	12	10	120
D	3	7	21
E	1	3	3
	26	31	196

$$S_{XY} = \frac{\sum XY}{n} - (\overline{X} \cdot \overline{Y}) = \frac{196}{5} - (5{,}2 \cdot 6{,}2) = 39{,}2 - 32{,}24 = 6{,}96$$

11. C

Persona	Hábitos de estudio (X)	Calificación (Y)	XY	X^2
A	2	6	12	4
B	8	5	40	64
C	12	10	120	144
D	3	7	21	9
E	1	3	3	1
	26	31	196	222

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{5 \times 196 - 26 \times 31}{5 \times 222 - 26^2} = \frac{174}{434} = 0{,}40$$

12. A

13. B

14. C

15. B

P(NoP)=1-P(P)=1-0,20=0,80

16. A

$$P(P \cap C) = P(P) \cdot P(C/P) = 0{,}20 \cdot 0{,}70 = 0{,}14$$

17. A

$$np = 10 \cdot 0{,}2 = 2$$

18. B

Tabla I con n=10, p=0,2 y x=2

19. A

Tabla I con n=10, p=0,2 y x=0

20. A

P_{67} = 0,44 (Tabla IV)

21. A

P(X ≤ 4, 8652) = 0,10 (Tabla V para 10 g.l.)

22. C

$\mu = n = 10$

Como se ve en la Tabla V para 10 g.l., la $P(X \leq 10)$ es un valor entre 0,10 y 0,90

23. A

24. B

$$E_{máx} = z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}} = 2{,}58\sqrt{\frac{0{,}75\cdot(0{,}25)}{100}} = 0{,}1117 \cong 0{,}11$$

25. B

$P - E_{máx} = 0{,}75 - 0{,}11 = 0{,}64$

$P + E_{máx} = 0{,}75 + 0{,}11 = 0{,}86$

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2014 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO F

Material: **Addenda (Formulario y Tablas) y calculadora**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

Gráfica 1: Número de mujeres asesinadas por la violencia de género y número de ellas que habían presentado denuncia por malos tratos entre los años 2004 y 2012 (Fuente: Ministerio de Sanidad, Servicios Sociales e Igualdad).

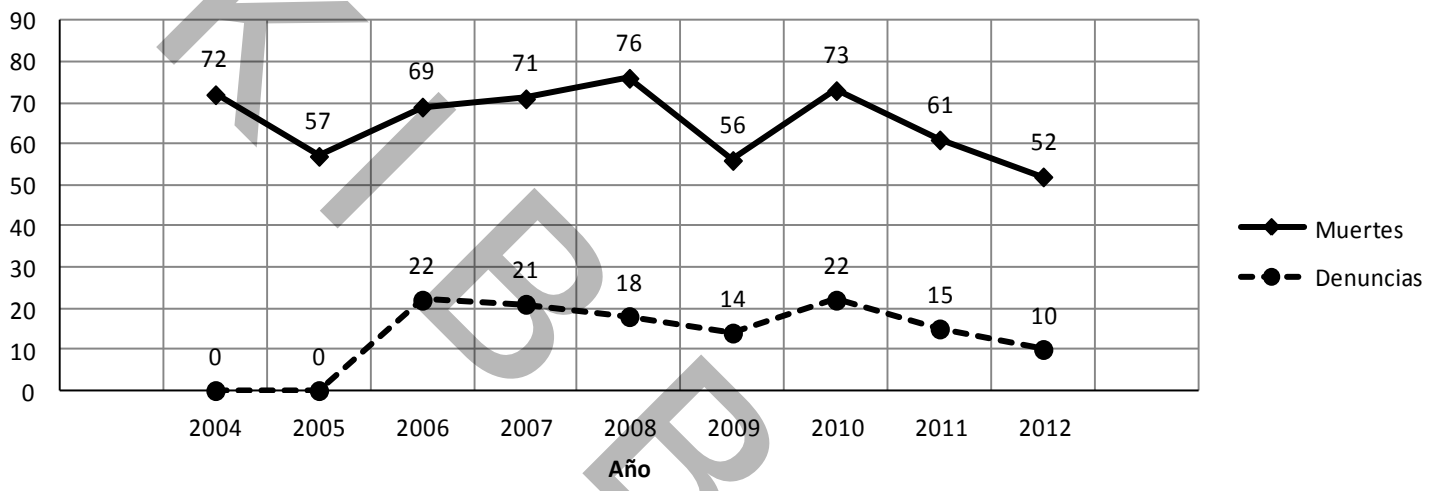


Tabla 1. Distribución de frecuencias de la variable Edad (X), de un grupo de 2000 personas.

X	n_i
21-30	400
31-40	600
41-50	500
51-60	400
61-70	100

Tabla 2. Para las puntuaciones de 200 personas en dos variables cuantitativas, X e Y, disponemos de los siguientes datos:

$\sum X = 1600$	$\sum Y = 20000$	$\sum XY = 163600$
$Y' = 4{,}5X + 64$		

Situación 1: Lanzamos al aire un dado en cuyas caras aparece un número del 1 al 6. Definimos el suceso A como "salir número par" y el suceso B como "salir múltiplo de 3".

1. En el eje de ordenadas de la Gráfica 1 aparece: A) la proporción; B) el porcentaje; **C) la frecuencia absoluta.**
2. Teniendo en cuenta la Gráfica 1, el número de mujeres asesinadas entre los años 2004 y 2012 ha sido: A) 128; **B) 587**; C) 600.
3. Teniendo en cuenta la Gráfica 1, en el año 2009 el porcentaje de mujeres asesinadas que habían presentado denuncia por malos tratos fué: A) 14%; **B) 25%;** C) 56%.
4. Teniendo en cuenta la Gráfica 1, la media de denuncias presentadas por mujeres que posteriormente fueron asesinadas es aproximadamente: **A) 13,6**; B) 18; C) 20.
5. La variable "Edad", de la Tabla 1, es: A) cualitativa; B) cuasicuantitativa; **C) cuantitativa**.

KIBBUTZ

6. Con los datos de la Tabla 1, la media es igual a: A) 35,5; **B) 41,5**; C) 45,5.
7. Con los datos de la Tabla 1, el tercer cuartil es igual a: A) 25,5; **B) 50,5**; C) 75,5.
8. Con los datos de la Tabla 1, ¿a qué percentil corresponde el valor 25,5?: **A) 10**; B) 20; C) 30.
9. Con los datos de la Tabla 1, la varianza es igual a: A) 50; **B) 134**; C) 247.
10. Con los datos de la Tabla 1, la distribución de la variable "Edad" es: **A) asimétrica positiva**; B) simétrica; C) asimétrica negativa.
11. Para calcular el grado de asociación entre dos variables cualitativas utilizamos: A) la covarianza; B) el coeficiente de correlación lineal de Pearson; **C) el coeficiente de contingencia C**.
12. En la Tabla 2, la covarianza entre X e Y es igual a: **A) 18**; B) 20; C) 25.
13. Teniendo en cuenta la Tabla 2, ¿qué puntuación pronosticaremos en Y a una persona que ha obtenido una puntuación de 8 en X?: A) 80; **B) 100**; C) 120**.**
14. Con los datos de la Situación 1, ¿cuánto vale la probabilidad de A?: A) 2/6; **B) 3/6**; C) 4/6
15. Teniendo en cuenta la Situación 1, ¿cuánto vale $P(A \cap B)$?: **A) 1/6**; B) 2/6; C) 3/6.
16. Teniendo en cuenta la Situación 1, ¿podemos afirmar que los sucesos A y B son independientes?: **A) Sí**; B) No; C) No podemos saberlo.
17. Una variable aleatoria discreta X toma los valores 0, 1 y 2. Sabemos que P(X=0)=0,2 ; P(X=1)=0,3 y P(X=2)=0,5. ¿cuánto vale la media o esperanza matemática de X?: A) 1; **B) 1,3**; C) 1,5.
18. Teniendo en cuenta los datos del ejercicio anterior, ¿cuánto vale la varianza de X?: A) 0,11; B) 0,41; **C) 0,61**.
19. El 75% de los alumnos de psicología escoge el itinerario de clínica. Si elegimos al azar 5 alumnos, ¿cuál es la probabilidad de que 3 ó más hayan escogido clínica?: A) 0,2637; B) 0,6985; **C) 0,8965**.
20. En una distribución normal tipificada, si $P(Z \leq z) = 0{,}0239$, entonces z es igual a: **A) -1,98**; B) 0,08; C) 1,98.
21. Si una variable X se distribuye según t de Student con 80 grados de libertad entonces $P(X \leq -0{,}678)$ es igual a: **A) 0,25**; B) 0,50; C) 0,75.
22. En una determinada facultad sólo se admiten estudiantes cuya puntuación en selectividad sea superior al percentil 67 (P_{67}). Sabiendo que las puntuaciones en selectividad se distribuyen *normalmente* con μ = 5,5 y σ = 1,5, para que un estudiante sea admitido debe obtener necesariamente una puntuación: A) mayor que 5,5; **B) mayor que 6,16**; C) mayor que 7,25.
23. Para conocer la forma de vida de los inmigrantes ilegales en España, que carecen de cualquier tipo de documentación y registro, ¿qué tipo de muestreo sería factible utilizar?: A) aleatorio simple; B) por conglomerados; **C) "bola de nieve**".
24. Sabemos que el CI (Cociente Intelectual) se distribuye normalmente con σ = 15. En una muestra, con n=144, obtenemos una media en CI de 100. Al nivel de confianza del 95% ¿entre qué valores estimaremos que se encuentra la media en CI en la población?: **A) 97,55 y 102,45**; B) 96,775 y 103,225; C) 85 y 115.
25. Para estimar las posibilidades de que la alumna PRS sea elegida por todos sus compañeros para delegada de curso elegimos una muestra de 100 estudiantes y les preguntamos si votarán o no a PRS obteniendo que 48 de ellos si lo harán. Con un nivel de confianza del 99%, y si para ser elegida necesita al menos el 51 % de los votos, podemos estimar que: A) no tiene posibilidades porque el intervalo confidencial estimado no incluye la proporción de 0,51**; B) tiene posibilidades porque el intervalo confidencial estimado incluye la proporción 0,51**; C) no podemos estimar si tiene posibilidades o no de ser elegida.

KUBBEN

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2012
EXAMEN TIPO TEST MODELO A

Tabla 1: Resultados en un test de agudeza visual (X) de siete personas en una investigación sobre la miopía.

Persona	X_i
1	20
2	6
3	9
4	6
5	12
6	8
7	16

Tabla 2: Puntuaciones de 200 universitarios en una escala de actitudes agrupadas en intervalos y las frecuencias absolutas (n_i) de cada intervalo. La varianza de esta distribución es igual a 132,84.

X	n_i
64-69	4
58-63	16
52-57	14
46-51	22
40-45	32
34-39	44
28-33	42
22-27	18
16-21	8

Tabla 3: En un estudio se investigó la relación entre la ansiedad de ejecución (Baja ó Alta) y la realización correcta de una tarea (Sí, No) en 400 personas. En la tabla se muestran algunas frecuencias observadas y, debajo y entre paréntesis, algunas frecuencias teóricas.

		Tarea realizada		
		Sí	**No**	
Ansiedad	**Baja**	68 (84,8)	--- (---)	160
	Alta	--- (---)	96 (112,8)	---
		212	---	400

1. En una escala de intervalo se han asignado a tres sujetos los valores X=0, X=3, y X=6. Con esta escala es correcto afirmar que: A) el tercer sujeto con X=6 tiene el doble de la variable medida que el segundo sujeto con X=3; B) el primer sujeto presenta carencia absoluta de la variable medida; C) la diferencia entre X=0 y X=3 es la misma que entre X=3 y X=6 en la variable medida.
2. La amplitud del intervalo en la distribución de frecuencias de la tabla 2 es: A) 5; B)4; C) 6.
3. Para una representación gráfica de la distribución conjunta de las dos variables de la tabla 3 debemos utilizar: A) un diagrama de barras conjunto; B) un diagrama de dispersión; C) un polígono de frecuencias.
4. Con los datos de la tabla 1, ¿cuál de las siguientes medidas de tendencia central tendrá el valor más alto?: A) la media; B) la mediana; C) la moda.
5. En los datos de la tabla 2, la media es igual a: A) 36,5; B) 39,5; C) 33,5.
6. Atendiendo a la tabla 2, el decil 2 es igual a: A) 28,5; B) 30,5; C) 29,5.
7. Una característica del índice desviación media es que: A) puede ser negativo; B) su valor siempre es igual o mayor que cero; C) siempre va a valer cero.
8. El índice de asimetría de Pearson de la distribución de la tabla 2 es: A) -0,26; B) 0; C) 0,26.
9. Con los datos de la tabla 1, ¿cuál es la puntuación típica correspondiente al sujeto 2?: A) -1,01; B) -5; C) -0,41.
10. Con los datos de la tabla 3, el valor de X^2 es igual a: A) 7,17; B) 11,80; C) 16,8.

KIBBUTZ

11. En la tabla 3, ¿se puede calcular el coeficiente de contingencia máximo?:A) sí, porque el número de sujetos es grande; B) no porque el número de filas es distinto al número de columnas; C) sí porque el número de filas es igual al número de columnas.
12. Hemos medido el grado de miopía (Y) de las siete personas de la tabla 1 y, a continuación, hemos calculado la recta de regresión de Y sobre X. Sabiendo que la correlación de Pearson entre el test de agudeza visual y el grado de miopía es de 0,68 y que la varianza de las puntuaciones pronosticadas, Y', es igual a 4, ¿cuál es la pendiente de la recta de regresión de Y sobre X?: A) 3,4; B) -0,28; C) 0,41.
13. Con los datos de la tabla 1 y del ejercicio anterior, ¿cuál es el valor de la varianza de los errores?: A) 4,65; B) 1,88; C) 1.
14. En un espacio muestral hay dos sucesos A y B. Sabemos que $P(A) = 0{,}4$, $P(\overline{B}) = 0{,}75$ y que $P(A \cup B) = 0{,}55$. ¿Cuál es la probabilidad de la intersección de los sucesos A y B?: A) 0,10; B) 0,60; C) 0,15.
15. Se realiza un estudio sobre el hábito de fumar en adultos de mediana edad con 200 hombres y 300 mujeres. Un 30% de los hombres reconocen que sí fuman habitualmente, mientras que 225 mujeres se declaran no fumadoras. Si seleccionamos una persona al azar, ¿cuál es la probabilidad de que sea hombre y no fume?: A) 0,70; B) 0,28; C) 0,73.
16. Con los datos del ejercicio anterior, si elegimos una persona al azar, ¿cuál es la probabilidad de que fume?: A) 0,32; B) 0,73; C) 0,27.
17. Dada una variable aleatoria discreta, X, la función que utilizamos para calcular la probabilidad de que la variable tome un valor menor o igual que un determinado valor se denomina: A) función de distribución; B) función de densidad de probabilidad; C) función de probabilidad.
18. Si lanzamos un dado al aire, cuyas caras están numeradas del 1 al 6, y definimos la variable aleatoria "número obtenido", ¿cuál es la esperanza matemática de esta variable?: A) 1,83; B) 3,5; C) 7,2.
19. En un centro educativo, la probabilidad de que un alumno suspenda una asignatura es de 0,2. Si seleccionamos al azar a 4 alumnos, ¿cuál es la probabilidad de que el 75% de esos alumnos aprueben la asignatura?: A) 0,0256; B) 0,4096; C) 0,25.
20. Las puntuaciones en una prueba de psicomotricidad siguen una distribución normal con varianza 9. Si el primer cuartil vale 4, ¿cuál es la media de las puntuaciones en psicomotricidad?: A) 0; B) 1,99; C) 6,01.
21. Una variable aleatoria se distribuye según la *t* de Student con varianza igual a 1,20. ¿Cuál es el valor de la distribución que deja por encima al 20% de las observaciones?: A) 0,873; B) 0,849; C) 0,259.
22. En una variable aleatoria X que sigue la distribución Chi-cuadrado con varianza igual a 40, la probabilidad de que $(10{,}8508 \leq X \leq 28{,}4120)$ es: A) 0,95; B) 0,85; C) 0,90.
23. En el muestreo aleatorio simple, la probabilidad de que cada individuo sea elegido de la población: A) es la misma para todos los individuos; B) es distinta para cada individuo; C) es la misma para algunos individuos
24. En una investigación sobre la inteligencia emocional, se ha extraído una muestra de 100 personas de una determinada población. Se sabe que la varianza poblacional es 16 y que el límite superior en la estimación por intervalo de la media es 3,784, con un nivel de confianza de 0,95. ¿Cuánto vale la media de la muestra seleccionada?: A) 1,7; B) 3; C) 2,5.
25. Se ha aplicado un test de rendimiento a una muestra de125 personas de las que un 30% han aprobado el test. Para un nivel de confianza de 0,95, ¿entre qué límites se estima está la proporción de aprobados de la población? A) 0,20 y 0,40; B) 0,27 y 0,33; C) 0,22 y 0,38.

INTERNET

SOLUCIONES:

1. C
2. C
3. A
4. A

$$\overline{X} = \frac{77}{7} = 11$$

Para calcular la mediana, ordenamos primero los datos de menor a mayor:

6, 6, 8, 9, 12, 16, 20

Dado que n=7 es impar, la mediana es el valor central, Md=9

La Mo=6

5. B

X	n_i	X_i	n_iX_i
64-69	4	66,5	266
58-63	16	60,5	968
52-57	14	54,5	763
46-51	22	48,5	1067
40-45	32	42,5	1360
34-39	44	36,5	1606
28-33	42	30,5	1281
22-27	18	24,5	441
16-21	8	18,5	148
	200		7900

$$\overline{X} = \frac{7900}{200} = 39{,}5$$

6. C

X	n_i	X_i	n_a
64-69	4	66,5	200
58-63	16	60,5	196
52-57	14	54,5	180
46-51	22	48,5	166
40-45	32	42,5	144
34-39	44	36,5	112
28-33	42	30,5	68
22-27	18	24,5	26
16-21	8	18,5	8
	200		

Dado que el decil 2 es el percentil 20, calculamos directamente el percentil 20 de la distribución.

$$P_{20} = L_i + \left(\frac{\frac{n \cdot k}{100} - n_d}{n_c} \right) \cdot I = 27{,}5 + \left(\frac{\frac{200 \cdot 20}{100} - 26}{42} \right) \cdot 6 = 29{,}5$$

KIBBUTZ

7. B
8. C

$$S_X^2 = 132{,}84 \rightarrow S_X = \sqrt{132{,}84} = 11{,}525 \approx 11{,}53$$

$$A_s = \frac{\overline{X} - Mo}{S_x} = \frac{39{,}5 - 36{,}5}{11{,}53} = 0{,}26$$

9. A

Persona	X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
1	20	9	81
2	6	-5	25
3	9	-2	4
4	6	-5	25
5	12	1	1
6	8	-3	9
7	16	5	25
			170

$$S_X^2 = \frac{170}{7} = 24{,}29 \qquad S_X = 4{,}93 \qquad \overline{X} = 11 \qquad z = \frac{-5}{4{,}93} = -1{,}01$$

10. B

		Tarea realizada		
		Sí	**No**	
Ansiedad	**Baja**	68 (84,8)	92 (75,2)	160
	Alta	144 (127,2)	96 (112,8)	240
		212	188	400

$$X^2 = \frac{(68-84{,}8)^2}{84{,}8} + \frac{(92-75{,}2)^2}{75{,}2} + \frac{(144-127{,}2)^2}{127{,}2} + \frac{(96-112{,}8)^2}{112{,}8} = 11{,}80$$

11. C

12. C

$$b = r_{XY}\frac{S_Y}{S_X}$$

$$S_Y^2 = \frac{S_{Y'}^2}{r_{XY}^2} = \frac{4}{0{,}68^2} = 8{,}65 \rightarrow b = r_{XY}\frac{S_Y}{S_X} = 0{,}68\frac{\sqrt{8{,}65}}{4{,}93} = 0{,}4056 \cong 0{,}41$$

13. A

$$S_E^2 = S_Y^2 - S_{Y'}^2 = 8{,}65 - 4 = 4{,}65$$

KABBUTZ

14. A

$$P(B) = 1 - P(\overline{B}) = 1 - 0{,}75 = 0{,}25$$
$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0{,}40 + 0{,}25 - 0{,}55 = 0{,}10$$

15. B

		Fumador		
		Sí	No	
Sexo	Hombre	60	140	200
	Mujer	75	225	300
		135	365	500

$$P(H \cap No) = \frac{140}{500} = 0{,}28$$

16. C

$$P(Sí) = \frac{135}{500} = 0{,}27$$

17. A

18. B

x	f(x)	$x \cdot f(x)$
1	1/6	1/6
2	1/6	2/6
3	1/6	3/6
4	1/6	4/6
5	1/6	5/6
6	1/6	6/6
		21/6

$$\mu = x \cdot f(x) = \frac{21}{6} = 3{,}5$$

19. B
Tabla I de la Binomial con n = 4, x=1 y p=0,20

20. C
P_{25}=4 y la puntuación típica que deja por debajo al 25% de los sujetos es z=-0,67

$$\frac{4-\mu}{3} = -0{,}67 \rightarrow \mu = 6{,}01$$

21. A

$$\sigma^2 = \frac{n}{n-2} = 1{,}20 \rightarrow n = 12 \text{ grados de libertad}$$
$$P(T > t_{12}) = 0{,}20 \rightarrow P(T \leq t_{12}) = 0{,}80$$
En la tabla VI, $t_{12} = 0{,}873$

KIBBUTZ

22. B

$$\sigma^2 = 2n = 40 \rightarrow n = 20$$

Buscamos en la tabla V de la Chi-cuadrado las probabilidades, por lo que:

$$P(10{,}8508 \le X \le 28{,}4120) = P(X \le 28{,}4120) - P(X \le 10{,}8508) = 0{,}90 - 0{,}05 = 0{,}85$$

23. A

En el muestreo aleatorio simple, cada individuo de la población tiene la misma probabilidad de ser elegido.

24. B

$$nc = 0{,}95 \Rightarrow z_{1-\alpha/2} = 1{,}96 \quad \text{(tabla IV)}$$

$$E_{máx} = z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = 1{,}96\frac{4}{\sqrt{100}} = 0{,}784$$

$$L_S = \overline{X} + E_{máx} \rightarrow \overline{X} = L_S - E_{máx} = 3{,}784 - 0{,}784 = 3$$

25. C

$$nc = 0{,}95 \Rightarrow z_{1-\alpha/2} = 1{,}96 \quad \text{(tabla IV)}$$

Proporción de aprobados en la muestra: P=0,30

$$L_i = P - z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}} = 0{,}30 - 1{,}96\sqrt{\frac{0{,}30\cdot(1-0{,}30)}{125}} = 0{,}2196 \cong 0{,}22$$

$$L_S = P + z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}} = 0{,}30 + 1{,}96\sqrt{\frac{0{,}30\cdot(1-0{,}30)}{125}} = 0{,}3803 \cong 0{,}38$$

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
SEPTIEMBRE 2013 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora (cualquier modelo)**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Tabla 1. Puntuaciones directas en un test de razonamiento de 400 niños de un colegio de Educación Primaria.

X	n_i
28-32	65
23-27	70
18-22	160
13-17	50
8-12	55

Gráfico 1. Representación gráfica de la función de probabilidad de la variable aleatoria X.

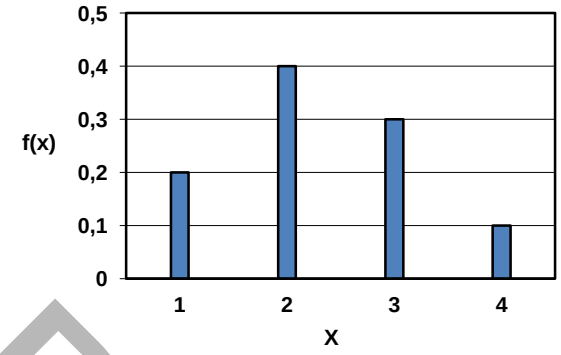


Tabla 2. Resultados obtenidos por 100 comerciales en un test de motivación de logro (X) y las ventas realizadas durante el último mes (Y).

	Media	Varianza	Covarianza entre X e Y
X	30	25	13,5
Y	6	9	

Tabla 3. Mejoría clínica de un grupo de 200 pacientes diagnosticados con depresión según el tipo de tratamiento recibido.

Tratamiento	Mejoría clínica	
	Sí	No
Cognitivo	40	25
Conductual	60	10
Farmacológico	50	15

1. La escala de medida de la variable *tipo de terapia recibida* de la Tabla 3 es: A) ordinal; B) de intervalo; C) nominal.
2. El gráfico adecuado para representar la relación entre las dos variables de la Tabla 2 es el diagrama de: A) dispersión; B) sectores; C) barras conjunto.
3. La variable *número de caras obtenidas al lanzar al aire dos monedas* es: A) dicotómica; B) discreta; C) continua.
4. Considerando los datos de la Tabla 1, la moda de las puntuaciones en el test es: A) 160; B) 20; C) 22,5.
5. Con los datos de la Tabla 1, ¿qué percentil le corresponde a un niño con una puntuación de 17 en el test?: A) 25; B) 27; C) 30.
6. Con los datos de la Tabla 1, el cuartil 3 es: A) 25; B) 27; C) 31.

LIBRE UNI

7. Respecto a la Tabla 2, ¿qué variable presenta un mayor grado de variabilidad?: A) el test de motivación de logro; B) las ventas realizadas durante el último mes; C) las dos muestran el mismo grado de variabilidad.
8. Atendiendo a la Tabla 1, la amplitud semi-intercuartil de las puntuaciones en el test es de: A) 4; B) 4,75; C) 8.
9. Con los datos de la Tabla 2, a un comercial que ha obtenido en el test de motivación de logro una puntuación típica de 1,5, ¿qué puntuación directa le corresponde?: A) 30; B) 31,5; C) 37,5.
10. Teniendo en cuenta los datos de la Tabla 2, el coeficiente de correlación de Pearson entre X e Y, vale: A) 0,6; B) 0,7 ; C) 0,9.
11. Con los datos de la Tabla 2, La recta de regresión para pronosticar las ventas mensuales (Y) a partir de la motivación de logro (X) es: A) Y' = -10,2+0,54X; B) Y´=-42,6+1,62X; C) Y´=-21+0,90X.
12. Con los datos de la Tabla 3, si ambas variables no estuviesen relacionadas entre sí, el número esperado de pacientes que han recibido terapia conductual y que no han mostrado mejoría clínica es de: A) 14,2; B) 17,5; C) 33,3.
13. Considerando los datos de la Tabla 3, el valor del Coeficiente de Contingencia Máximo ($C_{máx}$): A) es 0,707; B) es 0,816; C) no se puede calcular.
14. Con los datos de la Tabla 3, si elegimos un paciente al azar, ¿cuál es la probabilidad de que haya recibido terapia conductual y se observe mejoría clínica?: A) 0,30; B) 0,40; C) 0,86.
15. Considerando los datos de la Tabla 3, si suponemos que se ha seleccionado un paciente al azar y ha resultado que NO ha mejorado en su depresión, ¿cuál es la probabilidad de que haya recibido la terapia cognitiva?: A) 0,13; B) 0,38; C) 0,50.
16. Si $P(A\cap B)\neq P(A)P(B)$, entonces los sucesos A y B son: A) imposibles; B) dependientes; C) complementarios.
17. La esperanza matemática de la variable aleatoria representada en el Gráfico 1 es: A) 2,3; B) 2; C) 2,5.
18. Con los datos del Gráfico 1, el valor de la función de distribución para X=3, F(3), es: A) 0,30; B) 0,60;C) 0,90.
19. En un estudio sobre la implantación de nuevas tecnologías en el aula, concluyen que el 30% de los profesores utilizan la pizarra digital en sus clases. Si elegimos al azar una muestra de 10 profesores ¿cuál es la probabilidad de que al menos 5 de ellos utilicen la pizarra con sus alumnos?: A) 0,0473; B) 0,0781; C) 0,1503.
20. Las puntuaciones en un test de inteligencia siguen una distribución normal con $\mu = 100$ y $\sigma = 15$. El percentil 33 (P_{33}) de esta distribución vale: A) -0,44; B) 0,6293; C) 93,40
21. La probabilidad de que una variable X que se distribuye como una *Chi-cuadrado* con 14 grados de libertad, tome un valor superior a 21,0641 es: A) 0; B) 0,10; C) 0,90
22. Una variable que se distribuye según una *t de Student* con n grados de libertad, puede tomar: A) únicamente valores positivos; B) únicamente valores positivos y negativos; C) valores positivos y negativos y el cero
23. Para un nivel de confianza de 0,95, $\sigma = 3$ y un error de estimación máximo de 1,2, el tamaño de la muestra requerido para estimar *la autoestima* media de una población resulta ser n=24. Sabiendo que la variable *autoestima* se distribuye *normalmente* en la población, ¿qué tamaño debería tener la muestra si el nivel de confianza fuese 0,99?: A) 66; B) 42; C) 24
24. En una muestra aleatoria de 121 adolescentes, la media y la cuasidesviación típica de las puntuaciones en un test de inteligencia emocional valen 22 y 7,865 respectivamente. Sabiendo que el error de estimación máximo de la media vale 1,4, ¿cuánto vale $z_{1-\alpha/2}$?: A) 1,96; B) 2,33; C) 2,58
25. Teniendo en cuenta los datos del ejercicio anterior, ¿cuáles son los límites entre los cuales se espera esté la inteligencia emocional media de la población de adolescentes?: A) 19,42 y 24,58; B) 20,04 y 23,96; C) 20,6 y 23,4

EXTRACTO

SOLUCIONES:

1. C
2. A
3. B
4. B

   La moda es el punto medio del intervalo con mayor frecuencia absoluta. El intervalo es 18-22 y su punto medio es 20

5. A

   La puntuación 17 se encuentra en el intervalo 13-17.

X	n_i	n_a
28-32	65	400
23-27	70	335
18-22	160	265
13-17	50	105
8-12	55	55

$$k = \left[\frac{\frac{(P_k - L_i)\cdot n_c}{I} + n_d}{n} \right] \times 100 = \left[\frac{\frac{(17-12{,}5)50}{5} + 55}{400} \right] \times 100 = 25$$

6. A

X	n_i	n_a
28-32	65	400
23-27	70	335
18-22	160	265
13-17	50	105
8-12	55	55

$$P_{75} = L_i + \left(\frac{\frac{n \cdot k}{100} - n_d}{n_c} \right) \cdot I = 22{,}5 + \left(\frac{\frac{400 \cdot 75}{100} - 265}{70} \right) \cdot 5 = 25$$

7. B

$$CV_X = \frac{S_X}{\bar{X}} \cdot 100 = \frac{5}{30} \cdot 100 = 16{,}67$$

$$CV_Y = \frac{S_Y}{\bar{Y}} \cdot 100 = \frac{3}{6} \cdot 100 = 50$$

$$CV_Y > CV_X$$

KIBBUTZ

8. A

$$Q = \frac{P_{75} - P_{25}}{2} = \frac{25-17}{2} = 4$$

9. C

$$z = \frac{X - \overline{X}}{S_x} = 1{,}5 = \frac{X-30}{5} \rightarrow X = (1{,}5 \times 5) + 30 = 37{,}5$$

10. C

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{13{,}5}{5 \times 3} = 0{,}9$$

11. A

$$b = r_{XY}\frac{S_Y}{S_X} = 0{,}9\frac{3}{5} = 0{,}54$$

$$a = \overline{Y} - b\overline{X} = 6 - 0{,}54 \times 30 = -10{,}2$$

Y' = -10,2+0,54X

12. B

$$\frac{(50 \text{ x} 70)}{200} = 17{,}5$$

13. C

14. A

Tratamiento	Mejoría clínica		
	Sí	No	
Cognitivo	40	25	65
Conductual	60	10	70
Farmacológico	50	15	65
	150	50	200

$$P(\text{Tconductual} \cap \text{Sí}) = \frac{60}{200} = 0{,}3$$

15. C

$$P(\text{Tcognitivo} / \text{No}) = \frac{P(\text{Tcognitivo} \cap \text{No})}{P(\text{No})} = \frac{25/200}{50/200} = 0{,}50$$

16. B

17. A

X	f(x)	Xf(x)
4	0,1	0,4
3	0,3	0,9
2	0,4	0,8
1	0,2	0,2
		2,3

$$\mu = \sum x \cdot f(x) = 2{,}3$$

KIBBUTZ

18. C

X	f(x)	F(x)
4	0,1	1
3	0,3	0,9
2	0,4	0,6
1	0,2	0,2

19. C

p=0,30 $P(X \geq 5) = 1 - P(X < 5) = 1 - F(4) = 1 - 0{,}8497 = 0{,}1503$

El valor de F(4) se busca en la Tabla II, y es el valor de la intersección de la fila n=10, x=4 con la columna p=0,30.

20. C

P_{33} de una distribución normal tipificada vale -0,44 (Tablas III)

P_{33} de una distribución normal con $\overline{X} = 100$ y $\sigma = 15$ vale:

$X = z\sigma + \mu \rightarrow X = -0{,}44 \cdot (15) + 100 = 93{,}40$

21. B

$P(X \leq 21{,}0641) = 0{,}90$ (Tabla V para 14 g.l.)

$P(X > 21{,}0641) = 1 - P(X \leq 21{,}0641) = 1 - 0{,}90 = 0{,}10$

22. C

Una variable que sigue una distribución t de Student puede tomar, independientemente de los grados de libertad, valores positivos y negativos y el cero

23. B

$$n = \frac{z_{1-\alpha/2}^2 \sigma^2}{E_{max}^2} = \frac{2{,}58^2 \cdot 3^2}{1{,}2^2} = 41{,}60 = 42$$

24. A

n=121 (Aproximación de la distribución t de Student a la normal)

$$E_{max} = z_{1-\alpha/2} \frac{S_{n-1}}{\sqrt{n}} \rightarrow 1{,}4 = z_{1-\alpha/2} \frac{7{,}865}{\sqrt{121}} \rightarrow z_{1-\alpha/2} = 1{,}96$$

25. C

$L_i = \overline{X} - E_{max} = 22 - 1{,}4 = 20{,}6 \qquad L_s = \overline{X} + E_{max} = 22 + 1{,}4 = 23{,}4$

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
Septiembre 2011 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO B

Figura 1. Distribuciones de frecuencias obtenidas al aplicar una misma prueba de competencia lectora a alumnos de: (a) Primaria, (b) Secundaria y (c) Bachillerato.

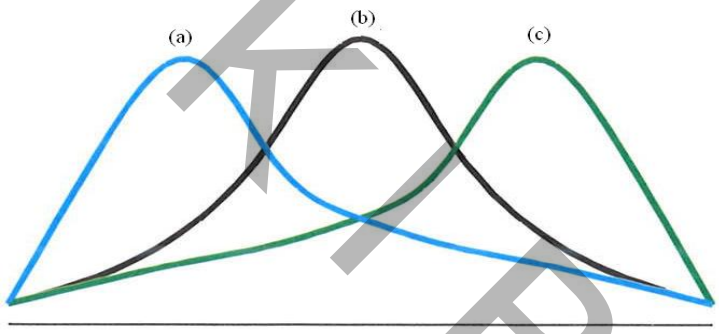


Tabla 1. Tiempo de reacción de 100 estudiantes en una tarea de atención visual focalizada. Se calcula que $\sum n_i X_i^2 = 12132725$

Tiempo de reacción	Frecuencia
381-400	10
361-380	20
341-360	30
321-340	25
301-320	15
	100

Figura 2. Diagrama de dispersión y recta de regresión de Y sobre X. La variable X representa la puntuación obtenida en un test de concentración, y la variable Y el número de errores cometidos al realizar una tarea monótona de atención.

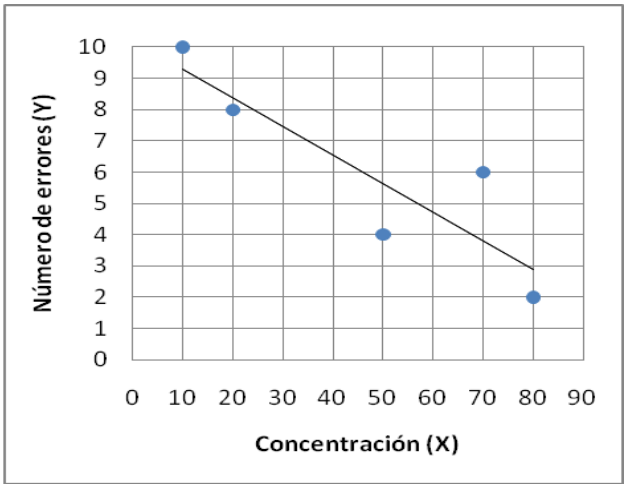


Tabla 2. Resultados obtenidos en un estudio sobre la discriminación laboral percibida de tres grupos distintos de inmigrantes.

		Discriminación laboral percibida			
		Baja	Moderada	Alta	
Inmigrantes	Rumanos	30	35	15	80
	Marroquíes	10	20	20	50
	Subsaharianos	10	30	30	70
		50	85	65	200

1. El proceso por el cual se asignan números a objetos o características según determinadas reglas se denomina: A) muestreo; B) estadística; C) medición.
2. Para que tenga sentido calcular las frecuencias acumuladas de una variable, ésta debe ser, como mínimo: A) nominal; B) ordinal; C) de intervalo.
3. Según la Figura 1, la distribución de frecuencias de los alumnos de bachillerato es: A) asimétrica positiva; B) asimétrica negativa; C) simétrica.
4. ¿Cuál es la amplitud de los intervalos de la variable Tiempo de reacción de la Tabla 1? A) 10; B) 20; C) 100.
5. Atendiendo a los datos de la Tabla 1, la mediana del tiempo de reacción es: A) 30; B) 347,2; C) 360,5.

KIBBUTZ

6. ¿Cuál es la moda de la variable tiempo de reacción según los datos de la Tabla 1? A) 30; B) 340,5; C) 350,5.
7. Atendiendo a las distribuciones de frecuencias de la Figura 1, ¿en cuál coincidirán los valores de media, mediana y moda? A) En la de alumnos de Primaria; B) En la de alumnos de secundaria; C) En la de alumnos de Bachillerato.
8. Sabiendo que el tiempo de reacción medio de los 100 estudiantes de la Tabla 1 es 347,5. ¿Cuál es su desviación típica? A) 23,9; B) 571; C) 1518,5.
9. ¿Cuál es el índice de asimetría de Pearson de la variable Tiempo de reacción de la Tabla 1? A) -0,13; B) 0; C) 0,13.
10. Según la Figura 2, la relación entre el nivel de concentración y el número de errores cometido es: A) lineal directa; B) lineal inversa; C) nula.
11. La recta de regresión representada en la Figura 2 es: A) Y' = 10,14 - 0,09X; B) Y' = 10,14 + 0,09X; C) Y' = 10,14X - 0,09.
12. Es apropiado analizar la relación entre dos variables utilizando el coeficiente C de contingencia en el caso de: A) la Figura 1; B) la Figura 2; C) la Tabla 2.
13. Según los resultados de la tabla 2, si seleccionamos al azar a una persona inmigrante, ¿cuál es la probabilidad de que haya percibido una discriminación laboral alta? A) 0,25; B) 0,325; C) 0,65.
14. Según los resultados de la tabla 2, ¿cuál es la probabilidad de tener la nacionalidad rumana y percibir un grado de discriminación laboral moderado? A) 0,175; B) 0,35; C) 0,425.
15. Suponiendo que se selecciona un inmigrante al azar de la Tabla 2 y ha resultado ser subsahariano, ¿cuál es la probabilidad de que perciba una baja discriminación laboral? A) 0,050; B) 0,143; C) 0,200.
16. Se llama función de probabilidad de una variable aleatoria discreta a aquella función que asocia, a cada valor de la variable, la probabilidad de que ésta adopte: A) ese valor; B) ese valor o cualquier otro inferior; C) ese valor o cualquier otro superior.
17. ¿Cuál es la media de una variable aleatoria X que toma los valores 0, 1 y 2, con una probabilidad de 0,6, 0,2 y 0,2 respectivamente? A) 0,2; B) 0,6; C) 1.
18. En un estudio sobre fobias concluyen que el 40% de la población adolescente con este problema presentan fobia social. Si elegimos al azar una muestra de 15 adolescentes, ¿cuál es la probabilidad de que cinco de ellos padezca este trastorno? A) 0,1859; B) 0,4032; C) 0,8290.
19. Suponiendo que las puntuaciones de los 4000 alumnos que se presentaron en la convocatoria de Febrero al examen de una asignatura se distribuyen normalmente con media igual a 4,7 y desviación típica igual a 3. ¿Cuántos alumnos obtuvieron una calificación igual o menor a 5? A) 1841; B) 2159; C) 2560.
20. Las puntuaciones en un test de inteligencia siguen una distribución normal de media 100 y desviación típica 15. Si queremos clasificar la población en cuatro grupos de igual tamaño ¿Qué puntuaciones delimitarán estos grupos? A) 80, 100 y 120; B) 85, 100 y 115; C) 90, 100 y 110.
21. En una distribución chi-cuadrado con 7 grados de libertad, el valor 1,239 se corresponde con el percentil: A) 1; B) 10; C) 99.
22. La distribución F: A) es simétrica; B) puede adoptar cualquier valor entre $-\infty$ y $+\infty$; C) posee la propiedad recíproca.
23. Extraemos con reposición todas las muestras posibles de una variable aleatoria X cuya distribución es uniforme. La distribución muestral de la media de dicha variable: A) se aproxima a la distribución normal; B) sigue la distribución uniforme; C) sigue la distribución chi-cuadrado.
24. El nivel de competencia matemática se distribuye normalmente en una población de estudiantes, con una desviación típica igual 6. Suponiendo que se desea un error de estimación máximo no superior a 2, con un nivel de confianza de 0,95. ¿Qué tamaño debe tener la muestra para estimar la media? A) 35; B) 58; C) 69.
25. Considerando los mismos datos de un intervalo de confianza, para tener un error de estimación menor, hay que aceptar un nivel de confianza: A) mayor; B) igual; C) menor.

EXAMEN

SOLUCIONES:

1. C
2. B
3. B
4. B
5. B

Tiempo reacción	n_i	n_a
381-400	10	100
361-380	20	90
341-360	30	70
321-340	25	40
301-320	15	15
	100	

$$Md = 340{,}5 + \left(\frac{\frac{100}{2} - 40}{30} \right) \cdot 20 = 340{,}5 + \left(\frac{1}{3} \right) \cdot 20 = 347{,}2$$

6. C.

$$Mo = \frac{341 + 360}{2} = 350{,}5$$

7. B
8. A

$$S_x^2 = \frac{\sum n_i X_i^2}{n} - \overline{X}^2 = \frac{1213275}{100} - 347{,}5^2 = 571 \qquad S_x = \sqrt{571} = 23{,}9$$

9. A

$$A_s = \frac{\overline{X} - Mo}{S_x} = \frac{347{,}5 - 350{,}5}{23{,}9} = -0{,}13$$

10. B
11. A

Concentración (X)	Nº errores (Y)	XY	X^2
10	10	100	100
50	4	200	2500
80	2	160	6400
70	6	420	4900
20	8	160	400
230	30	1040	14300

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2} = \frac{5 \times 1040 - 230 \times 30}{5 \times 14300 - 230^2} = \frac{-1700}{18600} = -0{,}09$$

$$\overline{X} = \frac{230}{5} = 46 \qquad \overline{Y} = \frac{30}{5} = 6$$

$$a = \overline{Y} - b\overline{X} = 6 + 0{,}09 \times 46 = 10{,}14$$

KIBBUTZ

$$Y_i^{'} = a + bX = 10{,}14 - 0{,}09X$$

12. C

13. B

$$P(alta) = \frac{65}{200} = 0{,}325$$

14. A

$$P(rumano \cap moderado) = \frac{35}{200} = 0{,}175$$

15. B

$$P(baja|subsahariano) = \frac{P(baja \cap subsahariano)}{P(subsahariano)} = \frac{\left(\frac{10}{200}\right)}{\left(\frac{70}{200}\right)} = \frac{10}{1700} = 0{,}143$$

16. A

17. B

x	f(x)	x·f(x)
0	0,6	0
1	0,2	0,2
2	0,2	0,4
		0,6

$$\mu = \sum x \cdot f(x) = 0{,}6$$

18. A

(Se busca directamente en la tabla I, con n = 15, p = 0,40 y x = 5).

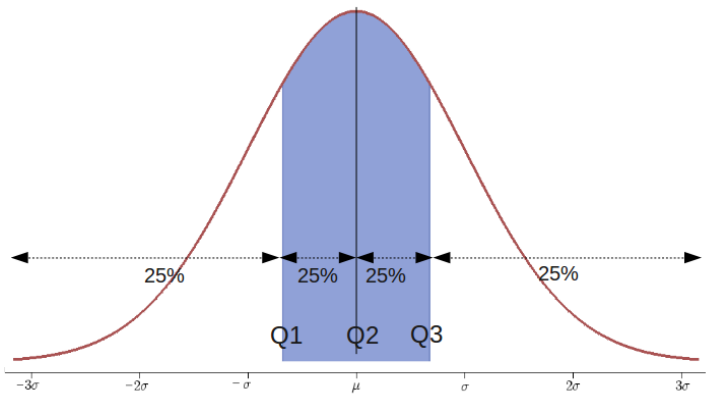
19. B

$$z = \frac{X - \overline{X}}{S_X} = \frac{5 - 4{,}7}{3} = 0{,}1$$

P(z<0,1) = 0,5398

0,5398×4000 =2159,2 ≈2159

20. C



Se piden los cuartiles. Buscamos en el interior de la tabla de la distribución normal las probabilidades 0, 25, 0,50 y 0,75, que se corresponden con las puntuaciones típicas -0,67, 0 y 0,67.

$$-0{,}67 = \frac{P_{25} - 100}{15} \Rightarrow P_{25} = 90$$

KIBBUTZ

$$0 = \frac{P_{50} - 100}{15} \Rightarrow P_{50} = 100$$

$$0{,}67 = \frac{P_{75} - 100}{15} \Rightarrow P_{75} = 110$$

21. A

Utilizando la Tabla V de la distribución Chi-cuadrado

22. C

23. A

24. A

$$\mathrm{n} = \frac{\mathrm{z}^2_{1-\alpha/2}\sigma^2}{E^2_{\max}} = \frac{1{,}96^2 \times 6^2}{2^2} = 34{,}57 \approx 35$$

25. C

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
Septiembre 2011 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO B

Figura 1. Distribuciones de frecuencias obtenidas al aplicar una misma prueba de competencia lectora a alumnos de: (a) Primaria, (b) Secundaria y (c) Bachillerato.

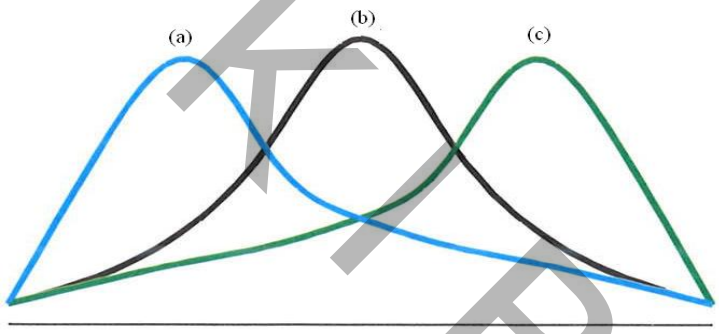


Tabla 1. Tiempo de reacción de 100 estudiantes en una tarea de atención visual focalizada. Se calcula que $\sum n_i X_i^2 = 12132725$

Tiempo de reacción	Frecuencia
381-400	10
361-380	20
341-360	30
321-340	25
301-320	15
	100

Figura 2. Diagrama de dispersión y recta de regresión de Y sobre X. La variable X representa la puntuación obtenida en un test de concentración, y la variable Y el número de errores cometidos al realizar una tarea monótona de atención.

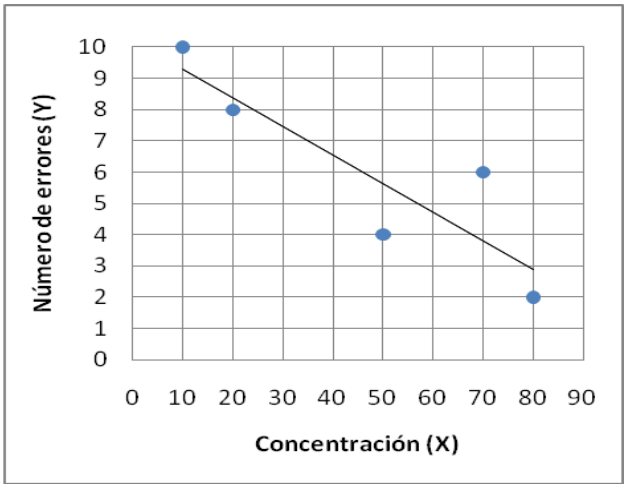


Tabla 2. Resultados obtenidos en un estudio sobre la discriminación laboral percibida de tres grupos distintos de inmigrantes.

		Discriminación laboral percibida			
		Baja	Moderada	Alta	
Inmigrantes	Rumanos	30	35	15	80
	Marroquíes	10	20	20	50
	Subsaharianos	10	30	30	70
		50	85	65	200

1. El proceso por el cual se asignan números a objetos o características según determinadas reglas se denomina: A) muestreo; B) estadística; C) medición.
2. Para que tenga sentido calcular las frecuencias acumuladas de una variable, ésta debe ser, como mínimo: A) nominal; B) ordinal; C) de intervalo.
3. Según la Figura 1, la distribución de frecuencias de los alumnos de bachillerato es: A) asimétrica positiva; B) asimétrica negativa; C) simétrica.
4. ¿Cuál es la amplitud de los intervalos de la variable Tiempo de reacción de la Tabla 1? A) 10; B) 20; C) 100.
5. Atendiendo a los datos de la Tabla 1, la mediana del tiempo de reacción es: A) 30; B) 347,2; C) 360,5.

KIBBUTZ

6. ¿Cuál es la moda de la variable tiempo de reacción según los datos de la Tabla 1? A) 30; B) 340,5; C) 350,5.
7. Atendiendo a las distribuciones de frecuencias de la Figura 1, ¿en cuál coincidirán los valores de media, mediana y moda? A) En la de alumnos de Primaria; B) En la de alumnos de secundaria; C) En la de alumnos de Bachillerato.
8. Sabiendo que el tiempo de reacción medio de los 100 estudiantes de la Tabla 1 es 347,5. ¿Cuál es su desviación típica? A) 23,9; B) 571; C) 1518,5.
9. ¿Cuál es el índice de asimetría de Pearson de la variable Tiempo de reacción de la Tabla 1? A) -0,13; B) 0; C) 0,13.
10. Según la Figura 2, la relación entre el nivel de concentración y el número de errores cometido es: A) lineal directa; B) lineal inversa; C) nula.
11. La recta de regresión representada en la Figura 2 es: A) Y' = 10,14 - 0,09X; B) Y' = 10,14 + 0,09X; C) Y' = 10,14X - 0,09.
12. Es apropiado analizar la relación entre dos variables utilizando el coeficiente C de contingencia en el caso de: A) la Figura 1; B) la Figura 2; C) la Tabla 2.
13. Según los resultados de la tabla 2, si seleccionamos al azar a una persona inmigrante, ¿cuál es la probabilidad de que haya percibido una discriminación laboral alta? A) 0,25; B) 0,325; C) 0,65.
14. Según los resultados de la tabla 2, ¿cuál es la probabilidad de tener la nacionalidad rumana y percibir un grado de discriminación laboral moderado? A) 0,175; B) 0,35; C) 0,425.
15. Suponiendo que se selecciona un inmigrante al azar de la Tabla 2 y ha resultado ser subsahariano, ¿cuál es la probabilidad de que perciba una baja discriminación laboral? A) 0,050; B) 0,143; C) 0,200.
16. Se llama función de probabilidad de una variable aleatoria discreta a aquella función que asocia, a cada valor de la variable, la probabilidad de que ésta adopte: A) ese valor; B) ese valor o cualquier otro inferior; C) ese valor o cualquier otro superior.
17. ¿Cuál es la media de una variable aleatoria X que toma los valores 0, 1 y 2, con una probabilidad de 0,6, 0,2 y 0,2 respectivamente? A) 0,2; B) 0,6; C) 1.
18. En un estudio sobre fobias concluyen que el 40% de la población adolescente con este problema presentan fobia social. Si elegimos al azar una muestra de 15 adolescentes, ¿cuál es la probabilidad de que cinco de ellos padezca este trastorno? A) 0,1859; B) 0,4032; C) 0,8290.
19. Suponiendo que las puntuaciones de los 4000 alumnos que se presentaron en la convocatoria de Febrero al examen de una asignatura se distribuyen normalmente con media igual a 4,7 y desviación típica igual a 3. ¿Cuántos alumnos obtuvieron una calificación igual o menor a 5? A) 1841; B) 2159; C) 2560.
20. Las puntuaciones en un test de inteligencia siguen una distribución normal de media 100 y desviación típica 15. Si queremos clasificar la población en cuatro grupos de igual tamaño ¿Qué puntuaciones delimitarán estos grupos? A) 80, 100 y 120; B) 85, 100 y 115; C) 90, 100 y 110.
21. En una distribución chi-cuadrado con 7 grados de libertad, el valor 1,239 se corresponde con el percentil: A) 1; B) 10; C) 99.
22. La distribución F: A) es simétrica; B) puede adoptar cualquier valor entre $-\infty$ y $+\infty$; C) posee la propiedad recíproca.
23. Extraemos con reposición todas las muestras posibles de una variable aleatoria X cuya distribución es uniforme. La distribución muestral de la media de dicha variable: A) se aproxima a la distribución normal; B) sigue la distribución uniforme; C) sigue la distribución chi-cuadrado.
24. El nivel de competencia matemática se distribuye normalmente en una población de estudiantes, con una desviación típica igual 6. Suponiendo que se desea un error de estimación máximo no superior a 2, con un nivel de confianza de 0,95. ¿Qué tamaño debe tener la muestra para estimar la media? A) 35; B) 58; C) 69.
25. Considerando los mismos datos de un intervalo de confianza, para tener un error de estimación menor, hay que aceptar un nivel de confianza: A) mayor; B) igual; C) menor.

EXAMEN

SOLUCIONES:

1. C
2. B
3. B
4. B
5. B

Tiempo reacción	n_i	n_a
381-400	10	100
361-380	20	90
341-360	30	70
321-340	25	40
301-320	15	15
	100	

$$Md = 340{,}5 + \left(\frac{\frac{100}{2} - 40}{30} \right) \cdot 20 = 340{,}5 + \left(\frac{1}{3} \right) \cdot 20 = 347{,}2$$

6. C.

$$Mo = \frac{341 + 360}{2} = 350{,}5$$

7. B
8. A

$$S_x^2 = \frac{\sum n_i X_i^2}{n} - \overline{X}^2 = \frac{1213275}{100} - 347{,}5^2 = 571 \qquad S_x = \sqrt{571} = 23{,}9$$

9. A

$$A_s = \frac{\overline{X} - Mo}{S_x} = \frac{347{,}5 - 350{,}5}{23{,}9} = -0{,}13$$

10. B
11. A

Concentración (X)	Nº errores (Y)	XY	X^2
10	10	100	100
50	4	200	2500
80	2	160	6400
70	6	420	4900
20	8	160	400
230	30	1040	14300

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2} = \frac{5 \times 1040 - 230 \times 30}{5 \times 14300 - 230^2} = \frac{-1700}{18600} = -0{,}09$$

$$\overline{X} = \frac{230}{5} = 46 \qquad \overline{Y} = \frac{30}{5} = 6$$

$$a = \overline{Y} - b\overline{X} = 6 + 0{,}09 \times 46 = 10{,}14$$

KIBBUTZ

$$Y_i^{'} = a + bX = 10{,}14 - 0{,}09X$$

12. C

13. B

$$P(alta) = \frac{65}{200} = 0{,}325$$

14. A

$$P(rumano \cap moderado) = \frac{35}{200} = 0{,}175$$

15. B

$$P(baja|subsahariano) = \frac{P(baja \cap subsahariano)}{P(subsahariano)} = \frac{\left(\frac{10}{200}\right)}{\left(\frac{70}{200}\right)} = \frac{10}{1700} = 0{,}143$$

16. A

17. B

x	f(x)	x·f(x)
0	0,6	0
1	0,2	0,2
2	0,2	0,4
		0,6

$$\mu = \sum x \cdot f(x) = 0{,}6$$

18. A

(Se busca directamente en la tabla I, con n = 15, p = 0,40 y x = 5).

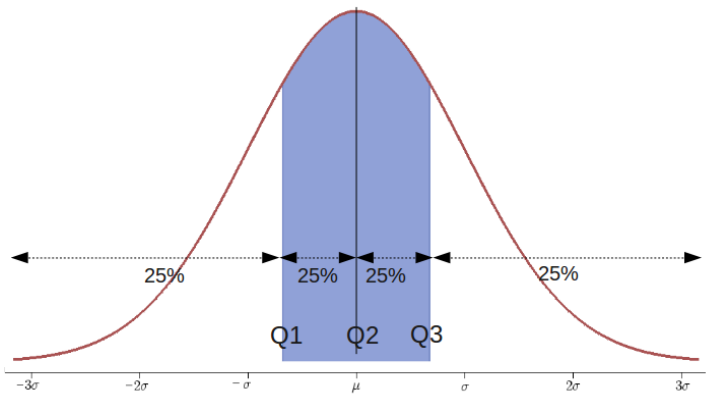
19. B

$$z = \frac{X - \overline{X}}{S_X} = \frac{5 - 4{,}7}{3} = 0{,}1$$

P(z<0,1) = 0,5398

0,5398×4000 =2159,2 ≈2159

20. C



Se piden los cuartiles. Buscamos en el interior de la tabla de la distribución normal las probabilidades 0, 25, 0,50 y 0,75, que se corresponden con las puntuaciones típicas -0,67, 0 y 0,67.

$$-0{,}67 = \frac{P_{25} - 100}{15} \Rightarrow P_{25} = 90$$

KIBBUTZ

$$0 = \frac{P_{50} - 100}{15} \Rightarrow P_{50} = 100$$

$$0{,}67 = \frac{P_{75} - 100}{15} \Rightarrow P_{75} = 110$$

21. A

Utilizando la Tabla V de la distribución Chi-cuadrado

22. C

23. A

24. A

$$\mathrm{n} = \frac{\mathrm{z}^2_{1-\alpha/2}\sigma^2}{E^2_{\max}} = \frac{1{,}96^2 \times 6^2}{2^2} = 34{,}57 \approx 35$$

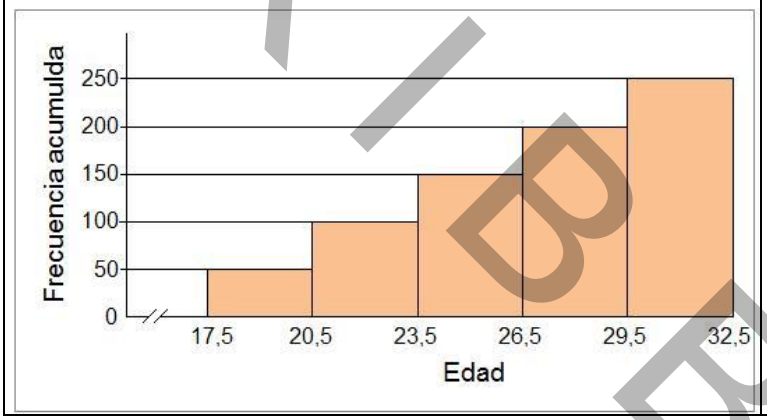
25. C

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2011 Código asignatura: 62011037
EXAMEN MODELO C

SOLUCIONES

Situación 1. La gráfica muestra la distribución de la edad (X) de los 250 sujetos de una investigación. En el eje horizontal, se recogen los límites exactos de los intervalos de X y en el eje vertical la frecuencia absoluta acumulada (n_a).



Situación 2. En una investigación para estudiar la relación entre siesta y mejora del aprendizaje, todos los sujetos realizaron una tarea de aprendizaje por la mañana. Después de comer, la mitad de los sujetos se echó la siesta. Finalmente, todos los sujetos volvieron a realizar la misma tarea de aprendizaje por la tarde. Los datos fueron los siguientes:

		Mejora del aprendizaje		
		Sí	No	
Siesta	Sí	170	30	200
	No	70	130	200
		240	160	400

1. El número de sujetos de una muestra que realizaron correctamente una tarea de discriminación en un experimento psicofísico es 80, lo que representa el 40% de la muestra. ¿Cuál el número de sujetos de la muestra? A) 200; B) 320 ; C) 500
2. La variable *mejora del aprendizaje* medida en la situación 2 es: A) dicotómica; B) cuantitativa discreta; C) cuasicuantitativa
3. En la situación 1, el número de sujetos con edades comprendidas entre 20,5 y 23,5 es A) 22; B) 50; C) 100
4. En la situación 1, la distribución de la edad de los sujetos: A) no tiene moda; B) tiene una moda; C) tiene dos modas
5. En la situación 1, el 80% de los sujetos tiene una edad menor que: A) 26,5; B) 28; C) 29,5
6. En la situación 1, la edad media de los sujetos es: A) 25; B) 50; C) 150
7. En la situación 1, la varianza de las edades de los sujetos es: A) 18; B) 22; C) 24
8. En la situación 1, el rango o amplitud total (A_T) del conjunto de las edades de los sujetos es: A) 3; B) 15; C) 50
9. Tenemos 10 puntuaciones cuya media es 15, si sumamos un 5 a cada una de las puntuaciones, la media de las nuevas puntuaciones es: A) 15; B) 20; C) 75
10. En la situación 2, el valor del estadístico X^2 para cuantificar el grado de asociación entre *siesta* y *mejora del aprendizaje* está entre: A) 30 y 40; B) 50 y 60; C) 100 y 110
11. En la situación 2, el valor del coeficiente de contingencia C para cuantificar el grado de asociación entre *siesta* y *mejora del aprendizaje* está entre: A) 0,26 y 0,30; B) 0,33 y 0,36; C) 0, 43 y 0,46
12. La recta de regresión que permite pronosticar el *riesgo de padecer una enfermedad coronaria* (Y) en función de la *hostilidad* (X) es Y' = 1,1 + 0,9X, ¿cuál es el *riesgo de padecer una enfermedad*

LIBRE

coronaria de una persona que ha obtenido una puntuación X = 8 en *hostilidad*?: A) 0,9; B) 1,1; C) 8,3

13. La propiedad $0 \leq P(A) \leq 1$ es válida: A) sólo para la definición clásica de la probabilidad**;** B) sólo para la definición estadística de la probabilidad; C) para las definiciones clásica y estadística de la probabilidad

14. En una clase, la mitad son chicos y la otra mitad son chicas. La mitad de los chicos y la mitad de las chicas han elegido inglés como optativa. Si elegimos una persona al azar de esta clase ¿cuál es la probabilidad de que sea chica y estudie inglés?: A) 0,25; B) 0,50; C) 0,75

15. Con los datos del ejercicio 14, ¿podemos decir que los sucesos "ser chica" y "estudiar inglés" son independientes: A) no**;** B) sí; C) no se puede saber con los datos disponibles

16. Una variable aleatoria discreta X toma los valores 0, 1, 2 y 3. Si sabemos que $P(X > 2) = 0,125$ ¿cuál es la probabilidad de que X sea igual a 3?: A) 0,125; B) 0,25; C) 0,875

17. Si lanzamos al aire 10 veces una moneda (no trucada), la varianza de la variable aleatoria "número de caras" es: A) 2,5; B) 4; C) 10

18. Lanzamos al aire 10 veces una moneda (no trucada), ¿cuál es la probabilidad de que salgan 4 caras? : A) 0,2051; B) 0,3770; C) 0,5000

19. Lanzamos al aire 100 veces una moneda (no trucada), ¿cuál es la probabilidad de que salgan 60 caras?: A) 0,0015; B) 0,0108; C) 0,0225

20. En una población de 2500 personas, las puntuaciones de un test de inteligencia siguen una distribución normal con media 100 y desviación típica 15. ¿Cuántas personas tienen en el test una puntuación superior a 130?: A) 15; B) 25; C) 57

21. En una distribución F de Snedecor con 7 grados de libertad en el numerador y 30 en el denominador, el percentil 5 es: A) 0,296; B) 2,334; C) 3,376

22. La probabilidad de que una variable que sigue una distribución t de Student con 10 grados de libertad tome el valor -0,7 o uno menor es: A) 0,25; B) 0,50; C) 0,75

23. Para realizar inferencias sobre el parámetro μ con cierta probabilidad: A) necesitamos el error típico de la media; B) debemos conocer la desviación típica de la población; C) podemos aplicar el muestreo casual

24. Sabemos que el error típico de la media vale 1,5 y el tamaño de la muestra es 100, ¿cuál es la desviación típica de la población?: A) 1,5; B) 15; C) 150

25. La amplitud deseada de un intervalo de confianza para la media es 4 para un nivel de confianza igual a 0,95, ¿cuál debe ser el tamaño de la muestra si la desviación típica de la población es 10? A) 24; B) 67; C) 96

IES BUCHIN

SOLUCIONES:

1. A

   $n_i = 80 \quad p_i = 0{,}40$

   $pi = n_i / n \;\rightarrow\; n = n_i / p_i$

   $n = 80 / 0{,}40 = 200$

2. A

   La variable *mejora del aprendizaje* es una variable cualitativa con dos categorías: *sí* y *no*.

3. B

Límites exactos	n_i	n_a
29,5 – 32,5	50	250
26,5 – 29,5	50	200
23,5 – 26,5	50	150
20,5 – 23,5	50	100
17,5 – 20,5	50	50

4. A

   La distribución no tiene moda (es amodal) dado que todos los intervalos tienen la misma frecuencia absoluta.

5. C

   En la gráfica de la situación 1 se observa que el 80% de los sujetos tiene una edad menor que 29,5. Obtendríamos el mismo resultado aplicando la siguiente fórmula:

Límites exactos	n_i	n_a
29,5 – 32,5	50	250
26,5 – 29,5	50	200
23,5 – 26,5	50	150
20,5 – 23,5	50	100
17,5 – 20,5	50	50

$$P_{80} = L_i + \left(\frac{\frac{n \cdot k}{100} - n_d}{n_c} \right) \cdot I = 26{,}5 + \left(\frac{\frac{250 \cdot 80}{100} - 150}{50} \right) \cdot 3 = 29{,}5$$

6. A

X	n_i	X_i	$n_i X_i$
29,5 – 32,5	50	31	1550
26,5 – 29,5	50	28	1400
23,5 – 26,5	50	25	1250
20,5 – 23,5	50	22	1100
17,5 – 20,5	50	19	950
	250		6250

$$\overline{X} = \frac{\sum n_i X_i}{n} = \frac{6250}{250} = 25$$

IBBCN

7. A

X	n_i	X_i	$n_i X_i^2$
29,5 – 32,5	50	31	48050
26,5 – 29,5	50	28	39200
23,5 – 26,5	50	25	31250
20,5 – 23,5	50	22	24200
17,5 – 20,5	50	19	18050
			160750

$$S_X^2 = \frac{\sum n_i X_i^2}{n} - \overline{X}^2 = \frac{160750}{250} - 25^2 = 18$$

8. B

Para una variable continua, la amplitud del intervalo es la diferencia entre el límite exacto superior y el límite exacto inferior: A_T = 32,5 – 17,5 = 15

9. B

$$\overline{X} = 15 \quad a = 5 \quad \rightarrow \quad \overline{Y} = 15 + 5 = 20$$

10. C

		Mejora del aprendizaje		
		Sí	No	
Siesta	Sí	170 (120)	30 (80)	200
	No	70 (120)	130 (80)	200
		240	160	400

$$x^2 = \frac{(170-120)^2}{120} + \frac{(30-80)^2}{80} + \frac{(70-120)^2}{120} + \frac{(130-80)^2}{80} = 104{,}17$$

11. C

$$C = \sqrt{\frac{X^2}{X^2+n}} = \sqrt{\frac{104{,}17}{104{,}17+400}} = 0{,}45$$

12. C

Y' = 1,1 + 0,9(8) = 8,3

13. C

El axioma 0 ≤ P(A) ≤ 1 es válido tanto para las definiciones clásica y estadística.

14. A

P(chica) = 0,5 P(inglés/chica) = 0,5

P(chica $\cap$ inglés) = P(chica) P(inglés/chica) = $0{,}5 \cdot 0{,}5 = 0{,}25$

KIBBUTZ

15. B

P(inglés) = P(chica $\cap$ inglés) + P(chico $\cap$ inglés) = 0,25 + 0,25 = 0,50

Los sucesos "ser chica" y "estudiar inglés" son independientes dado que:

P(inglés/chica) = P(inglés)

P(chica $\cap$ inglés) = P(chica) P(inglés)

16. A

P(X > 2) = P(X = 3) = 0,125

17. A

Binomial con n = 10 y p = 0,5 $\rightarrow$ varianza = npq = $10 \cdot 0{,}5 \cdot 0{,}5 = 2{,}5$

18. A

Binomial con n = 10, p = 0,5 y x = 4 (Tabla I)

19. B

Binomial con n=100, p=0,5 y x=60

Aproximación de la binomial a la normal:

$\text{Media} = np = 100 \cdot 0{,}5 = 50 \qquad \text{Desviación típica} = \sqrt{npq} = \sqrt{100 \cdot 0{,}5 \cdot 0{,}5} = \sqrt{25} = 5$

$$P\left[\frac{(60-0{,}5)-50}{5} \le Z \le \frac{(60+0{,}5)-50}{5}\right] = P(1{,}9 \le Z \le 2{,}1) = 0{,}9821 - 0{,}9713 = 0{,}0108 \quad \text{(Tabla IV)}$$

P(X = 60) = P(1,9 ≤ Z ≤ 2,1) = 0,0108 por aproximación de la binomial a la normal

20. C

$$z = \frac{(130-100)}{15} = 2$$

$P(Z > 2) = 1 - P(Z \le 2) = 1 - 0{,}9772 = 0{,}0228$ (Tala IV)

Nº de personas con inteligencia mayor que 130 $\rightarrow$ $2500 \cdot 0{,}0228 = 57$

21. A

$F_{0,05;7,30} = 1/F_{0,95;30,7} = 1/3{,}376 = 0{,}296$ (Tabla VII)

22. A

P(X ≤ -0,7) = 1-P(X ≤ 0,7) = 1-0,75 = 0,25 (Tabla VI con 10 g.l)

23. A

Necesitamos conocer la desviación típica de la distribución muestral de la media para realizar inferencias.

24. B

$\sigma_{\bar{X}} = 1{,}5 \quad n = 100$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \quad \rightarrow \quad \sigma = \sqrt{n} \cdot \sigma_{\bar{X}}$$

$$\sigma = \sqrt{100} \cdot 1{,}5 = 10 \cdot 1{,}5 = 15$$

KIBBUTZ

25. C

n.c. = 0,95 → $z_{1-\alpha/2} = z_{1-0,05/2} = z_{0,975} = 1,96$ (Tabla IV)

$E_{máx} = 4/2 = 2$

$$n = \frac{z^2_{1-\alpha/2}\sigma^2}{E^2_{máx}} = \frac{1,96^2.10^2}{2^2} = 96,04 \rightarrow 96$$

KIBBUTZ

INTRODUCCIÓN AL ANÁLISIS DE DATOS
FEBRERO 2011 Código asignatura: 62011037
EXAMEN TIPO TEST MODELO A DURACION: 2 HORAS

Material: **Addenda (Formulario y Tablas) y calculadora no programable**

Calificación= (0,4 x Aciertos) - (0,2 x Errores)

No debe entregar los enunciados

Tabla 1. Estatura en centímetros de 100 niños de 12 meses de edad.

Estatura	Frecuencia
79-81	**10**
76-78	**25**
73-75	**45**
70-72	**20**

Tabla 2. Resultados en un test de fluidez verbal de un grupo de vendedores de enciclopedias y número de ventas diarias realizadas.

Vendedor	Fluidez verbal (X)	Ventas diarias (Y)
1	**10**	**2**
2	**50**	**4**
3	**50**	**5**
4	**60**	**3**
5	**20**	**1**

Situación 1. Lanzamos al aire una vez un dado, definiendo dos sucesos: A = "obtener un número menor que tres" y B = "obtener un número impar". En el diagrama de Venn se representa una operación entre ambos sucesos.

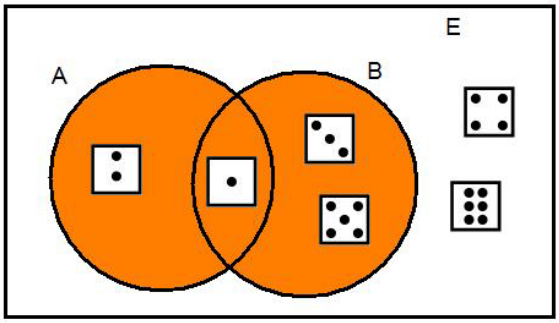
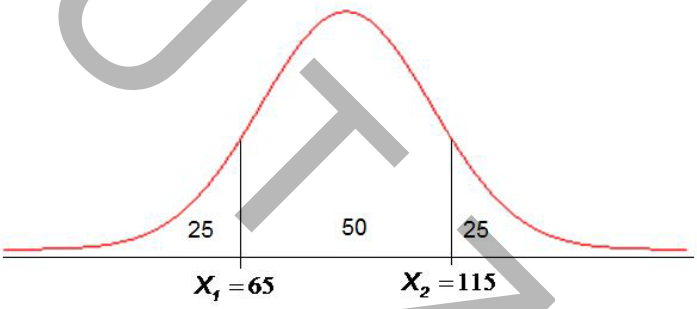


Gráfico 1. Puntuaciones obtenidas en un test de razonamiento abstracto (X) por un grupo de 100 estudiantes. Se distribuyen normalmente, con una desviación típica de 37,3. Hay 25 estudiantes que no alcanzan la puntuación 65 y 25 que superan la puntuación 115.



1. Mediante la Estadística Descriptiva se organizan y resumen conjuntos de observaciones procedentes de: A) muestras exclusivamente; B) muestras aleatorias exclusivamente; **C) muestras o poblaciones totales.**
2. La variable "ventas diarias realizadas" de la tabla 2 presenta un nivel de medida: A) ordinal; B) de intervalo; **C) de razón**.
3. El P_{50} de una distribución se corresponde con el: A) Q_1; **B) D_5**; C) Q_5.
4. ¿Qué porcentaje de niños de 12 meses de la tabla 1 tienen menor estatura que un niño de esa edad que mide 80 centímetros? A) 50; B) 90; **C) 95**.
5. Con los datos de la tabla 1, ¿cuál es la moda de la distribución? A) 45; **B) 74**; C) 80.
6. La amplitud total de la distribución de frecuencias de la tabla 1 es: A) 11; **B) 12**; C) 100.
7. La desviación típica de la variable estatura de la tabla 1 es un valor entre: **A) 2 y 3**; B) 4 y 5; C) 7 y 8.
8. La puntuación típica en Fluidez Verbal del vendedor 5 de la tabla 2 necesariamente será: **A) negativa;** B) igual a cero; C) positiva.

EXIBIDO

9. ¿Qué diagrama de dispersión corresponde a los datos presentados en la tabla 2? A) El 1; **B) El 2**; C) El 3.

Diagrama 1

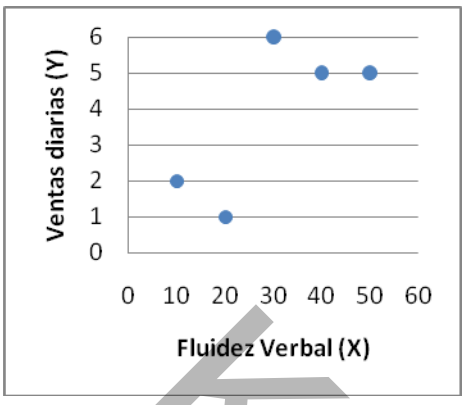


Diagrama 2

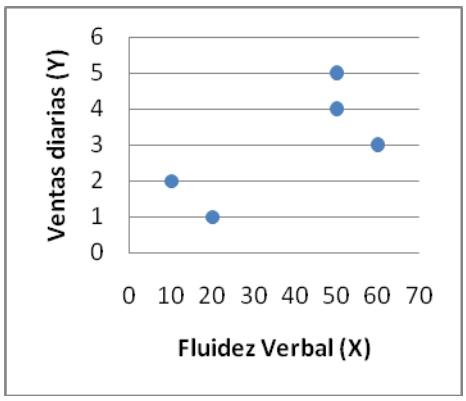
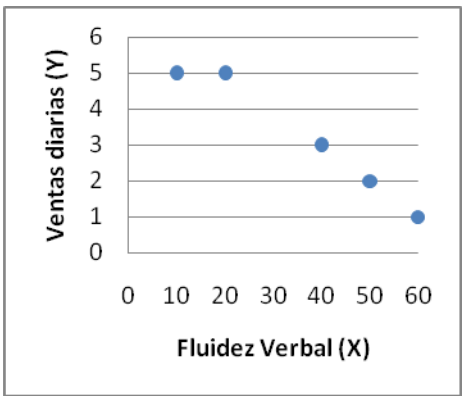


Diagrama 3



10. Entre las variables Fluidez Verbal y Nº de Ventas Diarias de la tabla 2 existe una relación lineal: **A) directa**; B) inversa; C) nula.

11. Con los datos de la tabla 2, la pendiente de la recta de regresión que permite pronosticar el número de ventas diarias (Y) a partir de la fluidez verbal de los vendedores (X) es: A) -0,053; B) 0; **C) 0,053**.

12. ¿Cuántas ventas diarias se pronosticará a un nuevo aspirante al puesto de vendedor de enciclopedias que ha obtenido en el test de fluidez verbal una puntuación de 40? A) Entre 0 y 1; B) Entre 2 y 3; **C) Entre 3 y 4.**

13. La zona sombreada del diagrama de Venn de la Situación 1 representa: **A)** $A \cup B$**;** B) A∩B; C) $\overline{A \cup B}$.

14. El espacio muestral descrito en la situación 1 está formado por: **A)** E={ ⚀ , ⚁ , ⚂ , ⚃ , ⚄ , ⚅ } B) E={ ⚀ , ⚁ , ⚂ , ⚄ }; C) E={ ⚀ , ⚂ , ⚄ }

15. Con los datos de la situación 1 se define un nuevo suceso C = "obtener un número par". ¿Cuál es $P(A \cup C)$? A) 1/6; B) 3/6; **C) 4/6**.

16. La Dirección General de Tráfico ha estimado que la probabilidad de infracción por "no respetar una señal de Stop" es 0,2, por "adelantamiento indebido" es 0,3 y por el "resto de infracciones" es 0,5. Además, la probabilidad de "accidente mortal supuesto no haber respetado el stop" es 0,5, la probabilidad de "accidente mortal supuesto adelantamiento indebido es 0,4" y la probabilidad de "accidente mortal supuesto otra infracción es 0,2. ¿Cuál es la probabilidad de que suceda un accidente mortal? A) 0,20; **B) 0,32;** C) 0,60.

17. Una variable aleatoria discreta X puede adoptar, con la misma probabilidad, los valores 1, 2, 3 y 4. ¿Cuál es su esperanza matemática? A) 0,25; B) 1; **C) 2,5**.

18. La función que asocia a cada valor de la variable la probabilidad de que ésta adopte ese valor o cualquier otro inferior es la función: A) aleatoria; B) de probabilidad; **C) de distribución**.

19. ¿Cuál es la probabilidad de que un estudiante que responde al azar las 20 preguntas de un examen de verdadero o falso acierte más de 15? **A) 0,0059**; B) 0,5900; C) 0,9941.

20. Con los datos del gráfico 1, ¿cuál es la media del test de razonamiento abstracto? A) 50; **B) 90;** C) 100.

21. Con los datos del gráfico 1, ¿cuál será el percentil 79? A) 81,30; B) 100; **C) 120,21**.

22. Atendiendo al gráfico 1, ¿Cuántas personas han obtenido una puntuación menor de 100 en el test de razonamiento abstracto? A) Entre 10 y 30; **B) Entre 50 y 70**; C) Entre 80 y 100.

23. El muestreo por cuotas es: A) aleatorio; B) probabilístico; **C) no probabilístico**.

24. Una muestra aleatoria de 16 estudiantes de ESO responde a una prueba de comprensión verbal que se distribuye normalmente, obteniendo una media de 80 y una varianza insesgada de 100. ¿Entre qué límites se hallará la verdadera comprensión verbal media de los estudiantes de secundaria, con un nivel de confianza de 0,99? **A) 72,63 y 87,37;** B); B) 75,62 y 84,38; C) 62,5 y 97,5.

25. Se ha aplicado una nueva terapia de afrontamiento de fobias a 100 pacientes obteniendo un resultado positivo en 70 de ellos. ¿Cuál es el error de estimación máximo para la proporción de pacientes curados con un nivel de confianza de 0,95? **A) 0,09;** B) 0,19; C) 0,30.

KUBBIN

SOLUCIONES:

1. C
2. C
3. B
4. C

Estatura	Frecuencia	na
79-81	10	100
76-78	25	90
73-75	45	65
70-72	20	20

La puntuación 80 se encuentra en el intervalo 79-81.

$$k = \left[\frac{\frac{(P_k - L_i)\cdot n_c}{I} + n_d}{n} \right] \times 100 = \left[\frac{\frac{(80-78{,}5)10}{3} + 90}{100} \right] \times 100 = 95$$

5. B

La moda es el punto medio del intervalo con mayor frecuencia $\left[\frac{73+75}{2}\right] = 74$

6. B

$$A_T = X_{máx} - X_{mín} = 81{,}5 - 69{,}5 = 12$$

7. A

Estatura	Xi	Frecuencia	n_iX_i	Xi^2	$n_iX_i^2$
79-81	80	10	800	6400	64000
76-78	77	25	1925	5929	148225
73-75	74	45	3330	5476	246420
70-72	71	20	1420	5041	100820
		100	7475		559465

$$\overline{X} = \frac{\sum n_i X_i}{n} = \frac{7475}{100} = 74{,}75$$

$$S_x^2 = \frac{\sum n_i X_i^2}{n} - \overline{X}^2 = \frac{559465}{100} - 74{,}75^2 = 7{,}09$$

$$S_x = \sqrt{S_x^2} = \sqrt{7{,}09} = 2{,}66$$

8. A

El vendedor 5 tiene una puntuación en fluidez verbal de 20, que es menor que la media $\overline{X} = \frac{\sum X_i}{n} = \frac{190}{5} = 38$ Por tanto, al pasar su puntuación a típica, $Z = \frac{X - \overline{X}}{S_x} = \frac{20-38}{S_x}$ el resultado será un valor negativo, ya que la desviación típica es siempre positiva.

KIBBUTZ

9. B

10. A

11. C

Vendedor	X	Y	XY	X^2
1	10	2	20	100
2	50	4	200	2500
3	50	5	250	2500
4	60	3	180	3600
5	20	1	20	400
	190	15	670	9100

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2} = \frac{5\times 670 - 190\times 15}{5\times 9100 - 190^2} = \frac{500}{9400} = 0{,}053$$

12. C

b = 0,053

$$\overline{X} = \frac{\sum X_i}{n} = \frac{190}{5} = 38 \qquad \overline{Y} = \frac{\sum Y_i}{n} = \frac{15}{5} = 3$$

$$a = \overline{Y} - b\overline{X} = 3\text{-}0{,}053\times 38 = 0{,}986$$

$$Y_i' = a + bX = 0{,}986 + 0{,}053X = 0{,}986 + 0{,}053\times 40 = 3{,}106$$

13. A

14. A

15. C

$$P(A\cup C) = P(A) + P(C) - P(A\cap C) = \frac{2}{6} + \frac{3}{6} - \frac{1}{6} = \frac{4}{6}$$

16. B

P(no stop) = 0,2

P(AI) = 0,3

P(resto) = 0,5

P(M│no stop) = 0,5

P(M│AI) = 0,4

P(M│resto) = 0,2

P(M) = P(no stop∩M) + P(AI∩M) + P(resto∩M) = P(no stop)×P(M│no stop) + P(AI)×P(M│AI) +

+ P(resto)× P(M│resto) = 0,2×0,5 + 0,3×0,4 + 0,5×0,2 = 0,1 + 0,12 + 0,1 = 0,32

17. C

X	f(x)	Xf(x)
1	0,25	0,25
2	0,25	0,50
3	0,25	0,75
4	0,25	1
		2,5

$$\mu = \sum x\cdot f(x) = 2{,}5$$

KIBBUTZ

18. C

19. A

$P(X > 15) = 1 - P(X \leq 15)$

Utilizando la Tabla II comprobamos (para n=20, x=15 y p=0,5) que $P(X \leq 15) = 0{,}9941$. Por tanto, $P(X > 15) = 1 - 0{,}9941 = 0{,}0059$

20. B

$$\overline{X} = \frac{65+115}{2} = \frac{180}{2} = 90$$

21. C

Se busca en la tabla IV de la curva normal la probabilidad 0,79.

$$0{,}81 = \frac{P_{79} - 90}{37{,}3} \Rightarrow P_{33} = 90 + 30{,}21 = 120{,}21$$

22. B

$$z = \frac{X - \overline{X}}{S_X} = \frac{100 - 90}{37{,}3} = 0{,}27$$

P(z<0,27) = 0,6064

0,6064 × 100 = 60,64 ≈ 61

23. C

24. A

$$L_i = \overline{X} - t_{15;0,995}\frac{S_{n-1}}{\sqrt{n}} = 80 - 2{,}947\frac{10}{\sqrt{16}} = 80 - 7{,}37 = 72{,}63$$

$$L_s = \overline{X} + t_{15;0,995}\frac{S_{n-1}}{\sqrt{n}} = 80 + 2{,}947\frac{10}{\sqrt{16}} = 80 + 7{,}37 = 87{,}37$$

25. A

n.c. = 0,95 → $z_{1-\alpha/2} = z_{0,975} = 1{,}96$ (Tabla IV)

Probabilidad de curar fobia = 70/100=0,70

$$E_{\text{máx.}} = 1{,}96\sqrt{\frac{0{,}70(1-0{,}70)}{100}} = 0{,}09$$

KIBBUTZ

X	Mujeres	Varones
8-9	20	12
6-7	16	13
4-5	10	17
2-3	8	10
0-1	6	8
∑	60	60

Tabla 1: Resultados obtenidos por un grupo de 60 mujeres y 60 hombres en una prueba de fluidez verbal (X)

Tabla 2: Para pronosticar las puntuaciones en una asignatura (Y) a partir de las puntuaciones en un test de razonamiento (X) disponemos de los siguientes datos obtenidos en un grupo de 500 niños:

	Media	Desviación típica	Recta de regresión
X	**100**	**10**	**Y´= - 8 + 0,16 X**
Y	**8**	**2**	

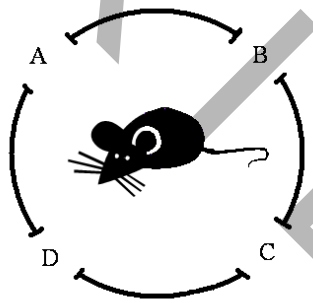


Figura 1. Rata situada en un laberinto con cuatro salidas (A, B, C y D) equiprobables

Variable	Distribución	
X	**N(20,5)**	Normal con media 20 y desviación típica 5
Y	χ^2_{40}	Chi-cuadrado con 40 grados de libertad
V	$F_{20,10}$	F con 20 grados de libertad en el numerador y 10 grados de libertad en el denominador

Tabla 3: Conjunto de variables y su tipo de distribución
.

1. La variable Género, con las categorías Hombre y Mujer, está medida en una escala: A) de razón ; B) ordinal ; C) nominal
2. La variable X, puntuaciones en una prueba de fluidez verbal, recogida en la Tabla 1, es: A) dicotómica; B) cualitativa; C) cuantitativa
3. Los datos recogidos en la Tabla 1, en fluidez verbal (X), para el grupo de mujeres pueden representarse mediante un: A) histograma; B) diagrama de sectores; C) diagrama de dispersión
4. La Moda, para el grupo de Varones de la Tabla 1, es: A) 4,5; B) 17; C) 60
5. Con los datos de la Tabla 1, la media en X para las Mujeres es: A) igual a la media para los Varones; B) mayor que la media para los Varones; C) menor que la media para los Varones
6. El Percentil 30, para el grupo de Mujeres en la Tabla 1, es: A) 3; B) 4,3; C) 7,5
7. La varianza de las puntuaciones en X, en la Tabla 1, para el grupo de varones es aproximadamente: A) 3,7 ; B) 5,7; C) 6,7
8. En la Tabla 1, si queremos saber en cuál de los dos grupos (mujeres o varones) es mayor la variabilidad en la variable X utilizaremos: A) las desviaciones típicas ; B) las desviaciones medias ; C) los coeficientes de variación
9. Si queremos estudiar la relación entre dos variables, X e Y, cada una de ellas con tres categorías utilizaremos el coeficiente: A) C de Contingencia; B) r_{XY} de Pearson ; C) A_S de Asimetría

KIBBUTZ

10. Si queremos estudiar la relación entre dos variables dicotómicas, X e Y, el valor del Coeficiente de Contingencia máximo que podemos obtener es: A) 0,20 ; B) 0,71 ; C) 0,9
11. Utilizando la recta de regresión, recogida en la Tabla 2, ¿qué puntuación en Y pronosticaremos a un alumno que ha obtenido una puntuación de 100 en X?: A) 4 ; B) 6; C) 8
12. Teniendo en cuenta los datos de la Tabla 2, el coeficiente de correlación de Pearson entre X e Y, vale: A) 0,7 ; B) 0,8 ; C) 0,9
13. En la definición clásica, la probabilidad es: A) el número de veces que se repite un suceso; B) el cociente entre el número de casos favorables y posibles de aparición de un suceso; C) la suma de las probabilidades de sucesos mutuamente excluyentes.
14. Si colocamos una rata en un laberinto, como el recogido en la Figura 1, ¿cuál es la probabilidad de que escoja la salida C?: A) 0,10; B) 0,20; C) 0,25
15. Si colocamos una rata en dos ocasiones en un laberinto, como el recogido en la Figura 1, y suponemos que no hay aprendizaje (los ensayos son independientes) ¿cuál es la probabilidad de que escoja la misma salida en las dos ocasiones?: A) 0,06; B) 0,25; C) 0,50
16. Si colocamos una rata en un laberinto, como el recogido en la Figura 1, y la variable aleatoria X toma el valor 1 cuando la rata escoge la salida A y 0 en otro caso. ¿cuánto vale la media de X?: A) 0,25; B) 0,50; C) 2,50
17. Si en un laberinto como el de la Figura 1 colocamos sucesivamente 20 ratas diferentes, ¿Cuál es la probabilidad de que 5 de ellas escojan la salida A?: A) 0,1686; B) 0,2023 ; C) 0,6172
18. Si en un laberinto como el de la Figura 1 colocamos sucesivamente 20 ratas diferentes, ¿Cuál es la probabilidad de que 5 ó menos escojan la salida A?: A) 0,2023 ; B) 0,3456; C) 0,6172
19. Teniendo en cuenta la Tabla 3, la $P(X \leq 25)$ es: A) 0,2500; B) 0,8413; C) 0,9681
20. Teniendo en cuenta la Tabla 3, el percentil 67 para la variable X vale: A) 22,2; B) 67; C) 76,2
21. Teniendo en cuenta la Tabla 3, para la variable Y, el valor 51,8051 es el percentil: A) 10 ; B) 50 ; C) 90.
22. Teniendo en cuenta la Tabla 3, para la variable V ¿cuál es el percentil 95?: A) 2,200 ; B) 2,774 ; C) 3,123
23. El procedimiento que “consiste en estimar, con cierta probabilidad, un parámetro desconocido a partir de una muestra aleatoria extraída de la población” se denomina : A) parametrización estadística; B) aleatorización estadística; C) inferencia estadística
24. La “desviación típica de la distribución muestral de la media” se denomina: A) desviación típica poblacional; B) variabilidad muestral; C) error típico de la media
25. Para estimar el intervalo confidencial de la media poblacional de una variable X con desviación típica poblacional igual a 4, hemos seleccionado una muestra de 100 personas y en ella hemos obtenido una media de 10. Trabajando con un nivel de confianza del 95%, los límites del intervalo confidencial son: A) 8,968 y 11,032; B) 9,216 y 10,784 ; C) 8 y 12

KIBBUTZ

SOLUCIONES:

1. C
2. C
3. A
4. A

   Mo: Punto medio del intervalo con mayor frecuencia

5. B

X	Mujeres n_M	Varones n_V	X_i	$X_i\ n_M$	$X_i\ n_V$
8 - 9	20	12	8,5	170	102
6 - 7	16	13	6,5	104	84,5
4 - 5	10	17	4,5	45	76,5
2 - 3	8	10	2,5	20	25
0 - 1	6	8	0,5	3	4
	60	60		342	292

$$\bar{X}_M = \frac{342}{60} = 5{,}7$$

$$\bar{X}_V = \frac{292}{60} = 4{,}87$$

6. B

X	$\mathbf{n_M}$	$\mathbf{n_a}$
8 - 9	20	60
6 - 7	16	40
4 - 5	10	24
2 - 3	8	14
0 - 1	6	6
	60	

$$P_{30} = 3{,}5 + \left(\frac{\frac{60\cdot 30}{100} - 14}{10}\right)\cdot 2 = 3{,}5 + \left(\frac{4}{10}\right)\cdot 2 = 3{,}5 + 0{,}8 = 4{,}3$$

7. C

X	n_V	X_i	$n_V X_i$	$(X_i - \bar{X})$	$n_V (X_i - \bar{X})^2$
8-9	12	8,5	102	3,63	158,1228
6-7	13	6,5	84,5	1,63	34,5397
4-5	17	4,5	76,5	-0,37	2,3273
2-3	10	2,5	25	-2,37	56,169
0-1	8	0,5	4	-4,37	152,7752
	60		292		403,934

$$\bar{X} = \frac{292}{60} = 4{,}87$$

$$S_V^2 = \frac{403{,}934}{60} \cong 6{,}7$$

8. C

   Los coeficientes de variación porque sus medias son distintas

9. A
10. B

$$C_{max} = \sqrt{\frac{k-1}{k}} = \sqrt{\frac{2-1}{2}} = \sqrt{0{,}5} = 0{,}71$$

11. C

$$Y' = -8 + 0{,}16X = -8 + 0{,}16\cdot 100 = -8 + 16 = 8$$

12. B

XIBBUTZ

$$r_{xy}\frac{S_y}{S_x} = 0{,}16 \rightarrow r_{xy} = 0{,}16\frac{S_x}{S_y} = 0{,}16\frac{10}{2} = 0{,}8$$

13. B

14. C

15. B

$$P(AA) = \frac{1}{4}\cdot\frac{1}{4} = \frac{1}{16} \qquad P(BB) = \frac{1}{4}\cdot\frac{1}{4} = \frac{1}{16} \qquad P(CC) = \frac{1}{4}\cdot\frac{1}{4} = \frac{1}{16} \qquad P(DD) = \frac{1}{4}\cdot\frac{1}{4} = \frac{1}{16}$$

$$4\cdot\frac{1}{16} = \frac{4}{16} = 0{,}25$$

16. A

x	f(x)	x·f(x)
0	3/4	0
1	1/4	1/4
	1	1/4

$$\overline{X} = \sum x\cdot f(x) = \frac{1}{4} = 0{,}25$$

17. B

Utilizando las tablas de la binomial (Tabla I) con n=20, p=0,25 y x=5, obtenemos 0,2023

18. C

Función de distribución binomial (Tabla II)

$P(X \leq 5) = 0{,}6172$

19. B

$$P(X \leq 25) = P\left(z \leq \frac{25-20}{5}\right) = P(z \leq 1) = 0{,}8413 \qquad \text{(Tabla IV)}$$

20. A

$$0{,}44 = \frac{P_{67} - 20}{5} \rightarrow P_{67} = (0{,}44 \cdot 5) + 20 = 22{,}2$$

21. C

Directamente en la Tabla V (chi-cuadrado)

22. B

Mirando directamente en la Tabla VII

23. C

24. C

25. B

$$nc = 0{,}95 \Rightarrow z_{1-\alpha/2} = 1{,}96$$

$$L_i = \overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = 10 - 1{,}96\frac{4}{\sqrt{100}} = 10 - 0{,}784 = 9{,}216$$

$$L_S = \overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = 10 + 1{,}96\frac{4}{\sqrt{100}} = 10 + 0{,}784 = 10{,}784$$

KIBBUTZ

GRADO EN PSICOLOGIA

INTRODUCCIÓN AL ANÁLISIS DE DATOS Código Asignatura: 62011037

FEBRERO 2010

EXAMEN MODELO C

X_i	n_i
1	80
2	52
3	28
4	12
5	6
6	2
7	1

$\overline{X} = 2{,}02$

$S_X^2 = 1{,}47$

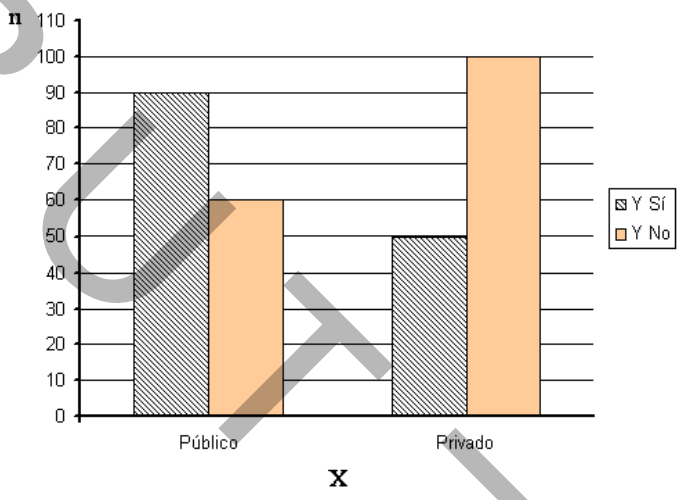
Tabla 1: Número de cigarrillos fumados en la última hora por 181 jóvenes fumadores que siguen un tratamiento para dejar de fumar.

X	n_i
38-40	18
35-37	36
32-34	52
29-31	46
26-28	32
23-25	16
	200

Tabla 2: Puntuaciones en un test de memoria de una muestra de 200 personas.

X	Y	XY
$\sum X = 200$	$\sum Y = 70$	$\sum XY = 1528$
$\sum X^2 = 4640$	$\sum Y^2 = 530$	

Tabla 3: Datos de las puntuaciones de 10 alumnos en un test de autoestima (X) y la calificación final del curso (Y).



Gráfica 1: Diagrama de barras en el que se representan conjuntamente la titularidad del centro de primaria en el que estudian los alumnos (X) y si realizan deberes o no en casa (Y).

1. Un parámetro es un valor numérico que: A) puede adoptar diferentes valores en una población; B) adopta un único valor en una población; C) adopta un valor diferente en cada muestra.
2. La variable *número de cigarrillos fumados* de la Tabla 1 presenta un nivel de medida: A) de intervalo; B) ordinal; C) de razón.
3. El diagrama de barras acumulados NO se puede utilizar en variables: A) nominales; B) ordinales C) cuantitativas discretas.
4. En la distribución de frecuencias de la Tabla 1, el valor de la mediana está comprendido entre: A) 1,40 y 1,60; B) 1,90 y 2,10; C) 1,65 y 1,75.
5. Con los datos de la Tabla 2, la moda de la distribución es: A) 52; B) 34; C) 33.

6. La media en el test de memoria de la distribución de la Tabla 2 es igual a: A) 28,50; B) 36,62; C) 31,71.
7. Si comparamos la variabilidad de las distribuciones de la Tabla 1 y la Tabla 2, ¿qué conjunto de puntuaciones presenta un mayor grado de dispersión?: A) el de la Tabla 2; B) el de la Tabla 1; C) las dos distribuciones presentan una variabilidad similar.
8. Con los datos de la Tabla 1, el índice de Asimetría de Pearson indica que la distribución es: A) asimétrica negativa; B) asimétrica positiva; C) simétrica.
9. Si tenemos en cuenta la forma de la distribución de la Tabla 1, la medida de variabilidad recomendada es: A) la amplitud semi-intercuartil; B) la varianza; C) el coeficiente de variación.
10. La varianza en el test de memoria de la distribución de la Tabla 2 es igual a: A) 10,78; B) 17,51; C) 13,24.
11. Con los datos de la Tabla 3, ¿cuál es la covarianza entre X e Y?: A) 7,3; B) 9,6; C) 12,8.
12. Con los datos de la Tabla 3, la proporción de la varianza de la calificación final del curso explicada por el test de autoestima vale: A) 0,80; B) 0,64; C) 0,89.
13. Con los datos de la Tabla 3, la pendiente de la recta de regresión que permite pronosticar la calificación final (Y) a partir del test de autoestima (X) es: A) 0,50; B) 1,05; C) 0,20.
14. Si en una tabla de contingencia las frecuencias observadas coinciden con las teóricas, el valor de X^2 es: A) 0; B) 1; C) -1.
15. Atendiendo a la Gráfica 1, si seleccionamos al azar a un niño, ¿cuál es la probabilidad de que estudie en un centro público y que realice deberes en casa?: A) 0,64; B) 0,50; C) 0,30.
16. Con los datos de la gráfica 1, si se elige al azar un niño y ha resultado ser de un centro privado, ¿cuál es la probabilidad de que no haga los deberes en casa?: A) 2/3; B) 1/3; C) 1/6.
17. Si A y B son dos sucesos dependientes, entonces la probabilidad de que ocurran conjuntamente ambos sucesos es igual a: A) $P(A)\cdot P(B\mid A)$; B) $P(A)\cdot P(B)$; C) $P(A)+P(B)-P(A\cap B)$.
18. La función de distribución de la variable aleatoria X *número de horas diarias de un adolescente conectado a internet* es F(0)=0,05, F(1)=0,28, F(2)=0,66; F(3)=0,92; F(4)=1. La media de X es: A) 1,56; B) 2,09; C) 1,67.
19. Se sabe que el 20 % de los españoles no ha acudido nunca a terapia con un psicólogo clínico. Si seleccionamos aleatoriamente una muestra de 10 personas, ¿cuál es la probabilidad de que tres personas de la muestra no hayan acudido a terapia? : A) 0,8791; B) 0,1209; C) 0,2013.
20. Las puntuaciones en una prueba de rendimiento en matemáticas siguen la distribución normal con media 500 y desviación típica 100. ¿qué proporción de sujetos obtienen una puntuación superior a 650?: A) 0,9332; B) 0,3224; C) 0,0668.
21. En una distribución t de Student, a medida que aumentan los grados de libertad, la distribución se aproxima más y más a la distribución: A) chi-cuadrado con pocos grados de libertad; B) normal; C) binomial.
22. Sea X una variable que sigue la distribución chi-cuadrado con 8 grados de libertad, ¿cuál es la desviación típica de esa variable?: A) 4; B) 16; C) 8.
23. En la distribución muestral de la media, el grado de variabilidad entre los valores de las medias muestrales se mide con: A) la desviación típica de la población; B) la cuasidesviación típica de la muestra; C) el error típico de la media.
24. Cuando NO existe homogeneidad en la población, es recomendable utilizar un muestreo: A) estratificado; B) aleatorio simple; C) sistemático.
25. Se sospecha que los padres con hijos que padecen el trastorno por déficit atencional con hiperactividad (TDAH) pueden manifestar también dicho trastorno. Para estudiar este aspecto se ha extraído una muestra de 200 padres y se ha obtenido que el 30% padecen el TDAH. Para un nivel de confianza del 95%, la amplitud del intervalo de confianza de la proporción de padres con TDAH es: A) 0,064; B) 0,127; C) 0,032.

KUBICHN

SOLUCIONES:

1. B
2. C
3. A
4. C

$\frac{n}{2} = \frac{181}{2} = 90{,}5$, por lo que el intervalo crítico es [1,5-2,5], con n_a=132.

X_i	n_i	n_a
7	1	181
6	2	180
5	6	178
4	12	172
3	28	160
2	52	132
1	80	80
	181	

$$Md = L_i + \left(\frac{\frac{n}{2} - n_d}{n_c}\right) \cdot I = 1{,}5 + \left(\frac{\frac{181}{2} - 80}{52}\right) \cdot 1 = 1{,}701923 \approx 1{,}70$$

5. C
6. C

X	n_i	Xi	n_iXi
38-40	18	39	702
35-37	36	36	1296
32-34	52	33	1716
29-31	46	30	1380
26-28	32	27	864
23-25	16	24	384
	200		6342

$$\overline{X} = \frac{\sum n_i X_i}{n} = \frac{6342}{200} = 31{,}71$$

7. B

$$CV_{X1} = \frac{S_X}{\overline{X}} \cdot 100 = \frac{1{,}21}{2{,}02} \cdot 100 = 59{,}90$$

$$CV_{X2} = \frac{S_X}{\overline{X}} \cdot 100 = \frac{4{,}18}{31{,}71} \cdot 100 = 13{,}18$$

$$CV_{X1} > CV_{X2}$$

8. B

$$S_X^2 = 1{,}47 \quad S_x = \sqrt{1{,}47} = 1{,}21$$

$$A_s = \frac{\overline{X} - M_O}{S_x} = \frac{2{,}02 - 1}{1{,}21} = 0{,}84$$

Asimetría positiva

9. A

KIBBUTZ

10. B

X	X_i	n_i	X_i^2	$n_i X_i^2$
38-40	39	18	1521	27378
35-37	36	36	1296	46656
32-34	33	52	1089	56628
29-31	30	46	900	41400
26-28	27	32	729	23328
23-25	24	16	576	9216
		200		204606

$$S_X^2 = \frac{\sum n_i X_i^2}{200} - (31{,}71)^2 = \frac{204606}{200} - 1005.5241 = 17{,}5059 \approx 17{,}51$$

11. C

X	Y	XY
$\sum X = 200$	$\sum Y = 70$	$\sum XY = 1528$
$\sum X^2 = 4640$	$\sum Y^2 = 530$	

$$S_{XY} = \frac{\sum XY}{n} - \overline{X}\overline{Y} = \frac{1528}{10} - 20 \cdot 7 = 12{,}8$$

12. B

$$S_X^2 = \frac{\sum X^2}{n} - \overline{X}^2 = 64 \rightarrow S_X = 8 \quad S_Y^2 = \frac{\sum Y^2}{n} - \overline{Y}^2 = 4 \rightarrow S_Y = 2$$

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{12{,}8}{8 \cdot 2} = 0{,}80 \rightarrow r_{XY}^2 = 0{,}64$$

13. C

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{10 \cdot 1528 - 200 \cdot 70}{10 \cdot 4640 - (200)^2} = \frac{1280}{6400} = 0{,}20$$

14. A

15. C

		Y		
		Sí	No	
X	Público	90	60	150
	Privado	50	100	150
		140	160	300

$$\mathrm{P}(\text{Público} \cap \text{Sí}) = \frac{90}{300} = 0{,}3$$

16. A

$$\mathrm{P}(\text{No} / \text{Privado}) = \frac{\mathrm{P}(\text{No} \cap \text{Priv})}{\mathrm{P}(\text{Priv})} = \frac{100/300}{150/300} = \frac{1/3}{1/2} = 2/3$$

17. A

KIBBUTZ

18. B

x	F(x)	f(x)	xf(x)
4	1	0,08	0,32
3	0,92	0,26	0,78
2	0,66	0,38	0,76
1	0,28	0,23	0,23
0	0,05	0,05	0
			2,09

$\mu = 2{,}09$

19. C

f(3)=P(X=3)=0,2013. Tabla 1, es el valor en la intersección de la fila n=10, x=3 con la columna p=0,20.

20. C

$$z = \frac{X - \overline{X}}{S_X} = \frac{650 - 500}{100} = \frac{150}{100} = 1{,}5$$

P(z>1,5)=1-P(z≤1,5)=1-0,9332=0,0668

21. B

22. A

$$\sigma = \sqrt{2n} = \sqrt{2 \cdot 8} = 4$$

23. C

24. A

25. B

$$E_{\max} = 1{,}96\sqrt{\frac{0{,}30 \cdot 0{,}70}{200}} = 1{,}96\sqrt{\frac{0{,}21}{200}} = 1{,}96\sqrt{0{,}00105} = 1{,}96 \cdot 0{,}0324 = 0{,}0635$$

La amplitud del intervalo es $2 \cdot E_{\max} = 2 \cdot 0{,}0635 = 0{,}127$

KIBBUTZ

GRADO EN PSICOLOGIA

INTRODUCCIÓN AL ANÁLISIS DE DATOS Código Asignatura: 62011037

FEBRERO 2010

EXAMEN MODELO A

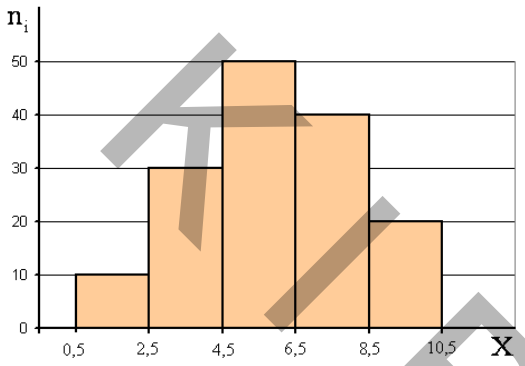


Figura 1: Representación gráfica de las calificaciones de 150 alumnos en una asignatura (X)

Tabla 1: Para estudiar la relación entre las puntuaciones en un test (X) y el rendimiento obtenido en una asignatura (Y) se utiliza una muestra de 500 niños (n=500), obteniéndose los siguientes resultados:

	Media	Desviación típica	Covarianza
X	100	10	24
Y	7	3	

Situación 1:

El 30 % de los niños padece algún problema de aprendizaje y de ellos el 80% acude al psicólogo. De los que no padecen problemas de aprendizaje sólo el 10% acude al psicólogo.

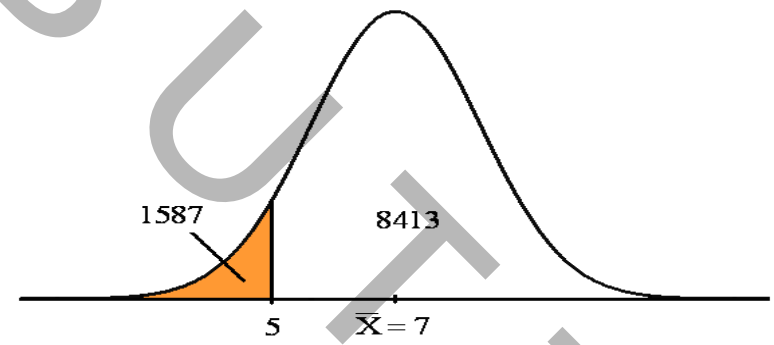


Figura 2: Distribución normal de las puntuaciones en la prueba de selectividad (X) de un grupo de 10000 alumnos con $\overline{X} = 7$

KIBBUTZ

1. La variable X, representada en la Figura 1, es: A) politómica; B) cualitativa; C) cuantitativa
2. La representación gráfica de la Figura 1 se denomina: A) diagrama de dispersión; B) histograma; C) polígono de frecuencias
3. En el eje de ordenadas de la Figura 1 se ha representado: A) la frecuencia absoluta; B) la frecuencia relativa; C) el porcentaje
4. Considerando la Figura 1, la Moda de la variable X es: A) 5,5; B) 6,5; C) 50
5. En la Figura 1, la calificación 6,5 corresponde al percentil: A) 50; B) 60; C) 65
6. El Percentil 30, para los datos de la Figura 1, es: A) 3; B) 4,7; C) 7,5
7. La varianza de las puntuaciones en X, de la Figura 1, es: A) 3,52 ; B) 4,91; C) 6,28

8. En la Tabla 1, ¿cuál de variables X e Y presenta mayor variabilidad?: A) X, porque su coeficiente de variación es mayor que el de Y ; B) Y, porque su coeficiente de variación es mayor que el de X ; C) No se puede determinar porque son variables distintas.
9. El coeficiente X^2 toma valores: A) iguales o superiores a cero; B) negativos ; C) comprendidos entre -1 y 1.
10. El coeficiente de correlación de Pearson entre X e Y, a partir de los datos de la Tabla 1, vale: A) 0,1; B) 0,8; C) 0,9
11. El signo de la pendiente de la recta de regresión de Y sobre X, en puntuaciones directas, depende de: A) las medias de X e Y; B) el cociente entre las desviaciones típicas de Y y X; C) el coeficiente de correlación de Pearson entre X e Y.
12. La recta de regresión para pronosticar las puntuaciones en la asignatura a partir de las puntuaciones en el test, teniendo en cuenta los datos de la Tabla 1 es: A) Y' = -3+0,8X; B) Y´=-17+0,24X; C) Y´=0,24X-10
13. En la definición clásica, la probabilidad es: A) el número de veces que se repite un suceso; B) el cociente entre el número de casos favorables y posibles de aparición de un suceso; C) la suma de las probabilidades de sucesos mutuamente excluyentes.
14. Si tenemos en cuenta los datos de la Situación 1, elegido un niño al azar ¿cuál es la probabilidad de que acuda al psicólogo?: A) 0,10; B) 0,24; C) 0,31
15. Continuando con la situación 1, elegido un niño al azar ha resultado que acude al psicólogo ¿cuál es la probabilidad de que padezca algún problema de aprendizaje?: A) 0,77; B) 0,66; C) 0,88
16. La función de probabilidad de una variable X es: f(0)=0,2, f(1)=0,3 y f(2)=0,5. La media de X es: A) 0,3; B) 1,3; C) 2,5
17. Se lanza una moneda al aire en 20 ocasiones. Sabiendo que P(Cara)=P(Cruz)=0,5 en cada ensayo, ¿Cuál es la probabilidad de obtener 10 Caras?: A) 0,0500; B) 0,1762 ; C) 0,2550 .
18. En un Centro de la UNED el 60% de los alumnos son mujeres. Si elegimos, al azar, una muestra de 5 alumnos ¿cuál es la probabilidad de que 2 sean varones?: A) 0,2350 ; B) 0,3456; C) 0,6544
19. En la Figura 2, ¿cuánto vale la desviación típica de X?: A) 3; B) 2; C) 4
20. Teniendo en cuenta los datos representados en la Figura 2, ¿cuántos alumnos han obtenido, en selectividad, una puntuación superior a 8?: A) 3085; B) 3830; C) 6915
21. En una distribución Chi-cuadrado con 60 grados de libertad, el valor 79,0819 es: A) el percentil 5 ; B) el percentil 90; C) el percentil 95.
22. En una distribución F con 40 y 20 grados de libertad en el numerador y en el denominador, respectivamente ¿cuál es el percentil 95?: A) 1,708 ; B) 1,994 ; C) 2,287
23. ¿Cuál de los siguientes tipos de muestreo es probabilístico?: A) por cuotas; B) opinático; C) por conglomerados
24. La media de la distribución muestral de la media es igual a: A) la desviación típica poblacional; B) la media poblacional; C) la desviación típica poblacional partido por la raíz cuadrada de n (siendo n el número de sujetos de la muestra) .
25. Para estimar el intervalo confidencial de la media poblacional de una variable X, hemos seleccionado una muestra de 100 personas y en ella hemos obtenido una media de 10. Trabajando con un nivel de confianza del 95% se han obtenido para ese intervalo unos límites de 9,216 y 10,784 ¿cuál es el valor de la desviación típica de esa variable X en la población?: A) 16; B) 4; C) 2

EXAMEN

SOLUCIONES:

1. C
2. B
3. A
4. A

   Mo: Punto medio del intervalo con mayor frecuencia: $\frac{4,5+6,5}{2}=\frac{11}{2}=5,5$

5. B

   Si observamos la Figura 1, podemos comprobar que la puntuación 6,5 deja por debajo de sí: 10+30+50=90 observaciones

$$\left.\begin{array}{l}150\rightarrow 100\%\\ 90\rightarrow X\end{array}\right\}\Rightarrow X=\frac{90\cdot 100}{150}=60\%$$

   Por tanto, 6,5 corresponde al P_{60}.

   También puede calcularse a partir de la distribución de frecuencias obtenida a partir de la Figura 1.

X	n_i	n_a
9 -10	20	150
7 - 8	40	130
5 - 6	50	90
3 - 4	30	40
1 - 2	10	10
	150	

$$k=\left[\frac{\frac{(P_k-L_i)\cdot n_c}{I}+n_d}{n}\right]\cdot 100=\left[\frac{\frac{(6,5-4,5)\cdot 50}{2}+40}{150}\right]\cdot 100=\frac{90}{150}\cdot 100=60$$

6. B

   Si tenemos 150 alumnos, el 30% son 45 alumnos. Es decir el P30 nos dejará por debajo de sí 45 alumnos.

   La puntuación 4,5 nos deja por debajo de sí 40 alumnos. La puntuación 6,5 nos deja por debajo de sí 90 alumnos. Por tanto:

$$\left.\begin{array}{l}2\rightarrow 50\\ X\rightarrow 5\end{array}\right\}\Rightarrow X=\frac{2\cdot 5}{50}=0,2$$

   Por tanto:

$$P_{30}=4,5+0,2=4,7$$

   También puede calcularse a partir de la distribución de frecuencias obtenida para la Figura 1.

X	n_i	n_a
9 -10	20	150
7 - 8	40	130
5 - 6	50	90
3 - 4	30	40
1 - 2	10	10
	150	

$$P_{30}=4,5+\left(\frac{\frac{150\cdot 30}{100}-40}{50}\right)\cdot 2=4,5+\left(\frac{5}{50}\right)\cdot 2=4,5+0,2=4,7$$

KIBBUTZ

7. B

X	X_i	n_i	$n_i X_i$	$(X_i - \overline{X})$	$n_i (X_i - \overline{X})^2$
9 -10	9,5	20	190	3,6	259,2
7 – 8	7,5	40	300	1,6	102,4
5 – 6	5,5	50	275	-0,4	8,0
3 – 4	3,5	30	105	-2,4	172,8
1 - 2	1,5	10	15	-4,4	193,6
		150	885		736

$$\overline{X} = \frac{885}{150} = 5{,}9$$

$$S_X^2 = \frac{736}{150} = 4{,}91$$

8. B

$$CV_X = \frac{S_X}{\overline{X}} \cdot 100 = \frac{10}{100} \cdot 100 = 10$$

$$CV_Y = \frac{S_Y}{\overline{Y}} \cdot 100 = \frac{3}{7} \cdot 100 = 42{,}86$$

$$CV_Y > CV_X$$

9. A

10. B

$$r_{XY} = \frac{S_{XY}}{S_x S_y} = \frac{24}{10 \cdot 3} = 0{,}8$$

11. C

La fórmula de la ecuación de regresión de Y sobre X es: $Y' = a + bX$, donde "b" (la pendiente) es:

$$b = r_{XY} \frac{S_Y}{S_x}$$

Puesto que la desviación típica siempre es un valor positivo (sólo toma el valor cero cuando las puntuaciones son iguales), el cociente:

$$\frac{S_Y}{S_X}$$

será siempre positivo. Por tanto el signo de la pendiente dependerá del signo del coeficiente de correlación de Pearson entre X e Y (r_{XY}).

12. B

$$Y' = \left[\overline{Y} - r_{xy} \frac{S_y}{S_x} \overline{X}\right] + \left(r_{XY} \frac{S_Y}{S_X}\right) \cdot X = \left[7 - 0{,}8 \frac{3}{10} 100\right] + \left(0{,}8 \frac{3}{10}\right) X = -17 + 0{,}24X$$

13. B

14. C

Llamemos:

PA =problemas de aprendizaje $\overline{PA}$ =sin problemas de aprendizaje

AP =acudir al psicólogo

$$P(PA) = 0{,}30 \quad P(\overline{PA}) = 1 - 0{,}30 = 0{,}70$$

$$P(AP \mid PA) = 0{,}80 \qquad P(AP \mid \overline{PA}) = 0{,}10$$

KIBBUTZ

$$P(AP) = P(AP \cap PA) + P(AP \cap \overline{PA}) = P(PA)\cdot P(AP|PA) + P(\overline{PA})\cdot P(AP|\overline{PA}) =$$
$$= 0{,}30\cdot 0{,}80 + 0{,}70\cdot 0{,}10 = 0{,}24 + 0{,}07 = 0{,}31$$

15. A

$$P(PA|AP) = \frac{P(PA \cap AP)}{P(AP)} = \frac{P(PA)\cdot P(AP|PA)}{P(AP)} = \frac{0{,}30\cdot 0{,}80}{0{,}31} = \frac{0{,}24}{0{,}31} = 0{,}7742 \cong 0{,}77$$

16. B

x	f(x)	x·f(x)
0	0,2	0
1	0,3	0,3
2	0,5	1
		1,3

17. B

Utilizando las tablas de la binomial (Tabla I) con n=20, p=0,5 y x=10, obtenemos 0,1762

18. B

Utilizando las tablas de la binomial (Tabla I) con n=5, p=0,4 y x=2, obtenemos 0,3456

19. B

En la Figura 2 se observa, además de que $\overline{X} = 7$, que 1587 alumnos de los 10000 no alcanzan la puntuación 5. Es decir, una proporción de 0,1587 no alcanza la puntuación 5. Si utilizamos lla Tabla III comprobamos que esa proporción se corresponde con una puntuación típica z=-1. Por tanto:

$$-1 = \frac{5-7}{S_x} \Rightarrow -S_x = 5 - 7 = -2 \Rightarrow S_x = 2$$

20. A

$$\frac{8-7}{2} = 0{,}5 \quad \Rightarrow \quad \text{tablas } 0{,}6915$$
$$1 - 0{,}6915 = 0{,}3085$$
$$0{,}3085\cdot 10000 = 3085$$

21. C

Directamente en la Tabla de chi-cuadrado.

22. B

Mirando directamente en la Tabla F

23. C

24. B

25. B

$$nc = 0{,}95 \Rightarrow z_{1-\alpha/2} = 1{,}96$$

Para resolver este ejercicio puede utilizarse tanto el límite superior como el límite inferior.

$$L_S = \overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = 10 + 1{,}96\frac{\sigma}{\sqrt{100}} = 10{,}784 \Rightarrow$$
$$\Rightarrow 10 + 1{,}96\frac{\sigma}{10} = 10{,}784 \Rightarrow 100 + 1{,}96\sigma = 107{,}84 \Rightarrow 1{,}96\sigma = 7{,}84 \Rightarrow \sigma = \frac{7{,}84}{1{,}96} = 4$$

KIBBUTZ

$$L_i = \overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = 10 - 1{,}96\frac{\sigma}{\sqrt{100}} = 9{,}216 \Rightarrow$$

$$\Rightarrow 10 - 1{,}96\frac{\sigma}{10} = 9{,}216 \Rightarrow 100 - 1{,}96\sigma = 92{,}16 \Rightarrow 1{,}96\sigma = 7{,}84 \Rightarrow \sigma = \frac{7{,}84}{1{,}96} = 4$$

KIBBUTZ

Analisis-de-datos.pdf

Albialbita

Introducción Al Análisis de Datos

1º Grado en Psicología

Facultad de Psicología
Universidad Nacional de Educación a Distancia

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

INTRODUCCIÓN al análisis de datos

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal

WUOLAH
Want to earn $1.236 extra per year?

Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

ÍNDICE

T1: CONCEPTOS BÁSICOS Y ORGANIZACIÓN DE DATOS 2
1. La investigación en psicología 2
2. Concepto y funciones de la estadística: descripción e inferencia 3
3. Variables: medición y clasificación 3
4. Descripción de V: distribución de frecuencias y representación gráfica 5
Descripción de variables cualitativas 5
Descripción de variables ordinales o cuasicuantitativas 6
Descripción de variables cuantitativas 6
Tendencia central, variabilidad y forma de una V: aproximación gráfica 8

T2: ÍNDICES DE TENDENCIA CENTRAL Y POSICIÓN 9
1. Índices de tendencia central 9
Media aritmética / promedio 9
La mediana (Md) 11
La moda (Mo) 12
Elección de un índice de tendencia central 12
2. Índices de posición 12
Percentiles (o centiles) 12
Cuartiles y deciles 13

T3: MEDIDAS DE VARIABILIDAD Y FORMA 14
1. Medidas de variabilidad 14
Amplitud total o rango / recorrido de las observaciones 14
Varianza y desviación típica 14
Coeficiente de variación 15
Amplitud intercuartil 15
2. Medidas de forma 16
Asimetría de una distribución 16
Apuntamiento o curtosis de una distribución 16
3. Diagrama de caja 16
4. Puntuaciones típicas 17

T4: RELACIÓN ENTRE VARIABLES I 18
1. Asociación entre 2 variables cualitativas 18
Tabla de contingencia 18
Representación gráfica: diagrama de barras conjunto 19
Medidas globales de asociación entre variables cualitativas 20
2. Relación entre variables ordinales 21
Coeficiente de correlación por rangos de Spearman 21

T5: RELACIÓN ENTRE VARIABLES II 22
1. Relación entre variables cuantitativas 22
Representación gráfica de la relación: el diagrama de dispersión 22
2. Regresión lineal simple 24
Cálculo de los coeficientes de regresión 24
Valoración del modelo 25
Características del modelo de regresión 26
3. Regresión lineal múltiple 26

T6: NOCIONES BÁSICAS DE PROBABILIDAD 27
1. Conceptos previos 27
2. Definición de probabilidad 28
3. Teorema de la suma 29
4. Probabilidad condicionada 29
5. Teorema del producto 29
6. Teorema de la probabilidad total 30
7. Teorema de Bayes 30
8. Aplicaciones de la probabilidad condicionada en psicología de la salud 31

T7: VARIABLES ALEATORIAS Y MODELOS DISCRETOS DE PROBAILIDAD 32
1. Concepto de variable aleatoria 32
2. Tipos de variables aleatorias 32
Variable aleatorias discretas 32
3. Distribuciones discretas de probabilidad 34
La distribución de Bernoulli 34
La distribución binomial 34

T8: MODELOS CONTINUOS DE PROBABILIDAD 36
1. Características de las variables aleatorias continuas 36
Función de densidad función de distribución 36
Media y varianza de una variable aleatoria continua 36
2. Distribución normal - Campana de Gauss / curva normal 37
Utilización de las tablas 37
Histograma y distribución normal 38
Aproximación de la binomial a la normal 38
3. La distribución de Pearson 38
4. La distribución t de Student 39
5. La distribución de F de Fisher-Snedecor 39

T9: MUESTREO Y DISTRIBUCIÓN MUESTRAL DE UN ESTADÍSTICO 40
1. Muestreo 40
Conceptos básicos en el muestreo 40
Tipos de muestreo 40
2. Distribución muestral de un estadístico 42
3. Distribución muestral del estadístico media 44
4. Distribución muestral del estadístico proporción 45
Distribución muestral del estadístico P para muestras pequeñas 46
Distribución muestral del estadístico P para muestras suficientemente grandes 46
5. Distribución muestral del estadístico varianza 46

T10: ESTIMACIÓN DE PARÁMETROS Y CÁLCULO DEL TAMAÑO MUESTRAL 48
1. Estimación de parámetros 48
Propiedades de los estimadores 49
Métodos de obtención de estimadores 50
Estimación puntual 50
Estimación por intervalos 50
2. Cálculo del intervalo de confianza 51
Intervalo de confianza para el parámetro μ con δ2 conocida 51
Intervalo de confianza para el parámetroμ con δ2 desconocida 52
Intervalo de confianza para el parámetro π (aproximación a la normal) 52
Intervalo de confianza para el parámetro δ2 53
3. Significado del nivel de confianza 53
4. Generalización de la construcción de intervalos 54
5. Factores que afectan al intervalo de confianza 54
6. Cálculo del tamaño muestral 54
Tamaño muestral para el parámetro media 54
Tamaño muestral para el parámetro proporción 55

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

T1: CONCEPTOS BÁSICOS Y ORGANIZACIÓN DE DATOS

Algunos usos, aplicada en casi todas las disciplinas (CCSS y salud, ↑ exponencialmente los últimos 30 años):

- Estudios epidemiológicos (medicina).
- Estudios toxicológicos relacionados con la eficacia de medicamentos (farmacia).
- Estudios genéticos y de impacto ambiental (biología).
- Muestreos en prospecciones petrolíferas o hidráulicas (geología).
- Censos de población e inf. demográfica (sociología).
- Estudios sobre la optimización del coste-beneficio (economía).
- **Medición de V y evaluación diagnóstica de tratamientos, programas educativos, sociales… (Psicología).**

- **Estadística teórica**: ocupa aspectos matemáticos formales y normativos.
- **Estadística aplicada**: aplicación a un campo concreto. *Bioestadística, psicoestadística, socioestadística…* Algunos, le han puesto el nombre de Análisis de datos.

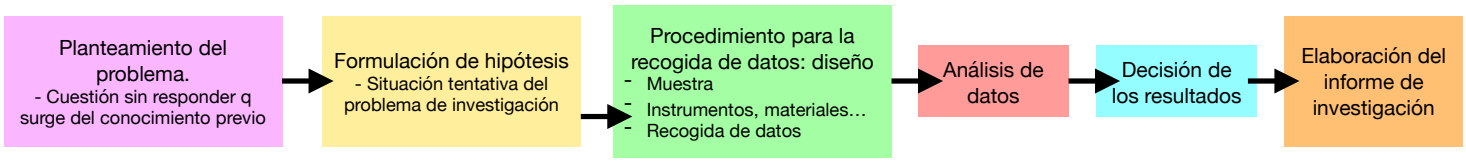
Se tiene la imagen como una rama de las matemáticas de difícil comprensión y ajena a nuestro día a día. Sin embargo, diariamente estamos sometidos a un bombardeo de datos estadísticos. El no ser capaz de distinguir una interpretación rigurosa de unos datos de una defectuosa, hace q se sea vulnerable a la manipulación. En ocasiones, las estadísticas presentadas en distintos medios (de com., políticos, publicidad, entorno laboral...) son incorrectas o engañosas por falta de preparación o por voluntad de «maquillar». Benjamín Disraeli (1er ministro del RU): *hay 3 tipos de mentiras: las mentiras, las grandes mentiras y las estadísticas*.

Nos proporciona las herramientas necesarias para valorar de manera crítica la inf. q recibimos.

1. LA INVESTIGACIÓN EN PSICOLOGÍA

Con la aparición de la ciencia moderna en el siglo XVII, el método científico pasó a ser la fuente de conocimiento más utilizada (hasta entonces eran el mito, sentido común o folclore popular). Las ciencias se distinguen entre sí por su objeto de estudio, pero tienen en común el método científico. La Psicología se sirve para acercarse a la conducta.

Método científico: procedimiento estructurado q utiliza la ciencia para la ampliación de sus conocimientos. Sistemático (tiene unas etapas definidas) y replicable (los datos obtenidos tienen q poder ser refutados -en las mismas circunstancias- por cualquier investigador interesado). Proporciona una manera de actuar para afrontar una investigación, a través de las sig. fases interdependientes:



1. Se define un problema, q puede surgir de teorías ya establecidas, de la lectura de la bibliografía o de la exp. directa con los hechos. En la mayoría de casos de lagunas o contradicciones en investigaciones anteriores.
2. Se plantea una hipótesis, q no es más que una solución tentativa al problema.
3. Las siguientes 3 fases contrastan si la hipótesis planteada es compatible con los hechos. Necesario establecer un procedimiento adecuado de recogida de inf., analizar los datos obtenidos y discutir los resultados en busca de conclusiones.
4. Elaborar un informe para dar a conocer los resultados.

Nos ocuparemos de la 4ª y 5ª fase; aprender a procesar los datos recogidos en una investigación con el fin de obtener la inf. q se precisa para contrastar la hipótesis formulada, y poder dar R al problema.

Diversos estudios ponen de manifiesto el efecto q la ansiedad ante los exámenes puede tener en la calificación PAU. Equipo investigador diseña un programa de tratamiento para paliar este efecto, q combina técnicas de estudio y relajación. Se ha selecciona a 40 estudiantes con este problema, voluntarios. 1/2 se ha asignado aleatoriamente al G1 (sin tratamiento) y 1/2 al G2 (tratamiento). Al finalizar el curso, se recogieron datos sobre las V relevantes de la investigación, además de algunas sociodemográficas, como sexo, nivel de estudios de la madre, opción de bachillerato elegido y h de estudio semanales.
¿cómo relacionaría los datos de este ejemplo con las fases de una investigación?

El problema objeto de estudio es se comprobar si el tratamiento influye en la calificación obtenida en la prueba. Hipótesis; el grupo investigador espera q sea eficaz, q el G2 tenga ↑ rendimiento en el examen q el G1. En la sig. fase se encontraría la det. de un plan de trabajo o procedimiento para la recogida de datos; decide escoger como muestra a 40 estudiantes con problemas de ansiedad ante los exámenes asignándolos de manera aleatoria a los grupos 1 y 2 comparando después sus resultados = análisis de los datos obtenidos y la discusión de dichos resultados. Se analizarían las calificaciones por ambos G. Se calcularía y compararía la media de ambos. Además, en otras investigaciones con otros objetivos, podría ser interesante plantear otro tipo de análisis, como cuantificar la

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

relación entre el n° de h estudiadas y la calificación en el examen de Lengua, o realizar pronósticos en el rendimiento en el examen en función de la ansiedad ante los exámenes y el n° de horas estudiadas (ambos procedimientos se estudiarán en el Tema 5 del programa). Por último, para difundir los resultados se elabora un informe.

El análisis de datos constituye una parte integral no solo de la activ. investigadora, tb en la práctica profesional. Resulta crucial tener unos conocimientos básicos para evaluar los resultados de una investigación, y leer de forma crítica las publicaciones de carácter psicológico *(artículos científicos, libros, informes de investigación o notas de prensa)*.

2. CONCEPTO Y FUNCIONES DE LA ESTADÍSTICA: DESCRIPCIÓN E INFERENCIA

Es la rama de las matemáticas q se encarga del estudio de det. carac. en una población, recogiendo los datos, agrupándolos, organizándolos en tablas, representándolos gráficamente y analizándolos para sacar conclusiones de dicha población.

2 grandes áreas: la Estadística Descriptiva y la Estadística Inferencial.

- **Estadística Descriptiva:** organiza y resumen conjuntos de observaciones cuantificadas procedentes de una muestra o de la población total. Mediante tablas, gráficos o valores numéricos. Se dispone de distintos procedimientos q nos permiten estudiar las cara. de 1 o + V:
  - **1V**: estadísticos q nos indicarán cuáles son los valores + habituales de esa V (de tendencia central), hasta q punto esos valores son similares o dif. entre sí (de variabilidad), en q grado las observaciones se reparten equilibradamente por encima y por debajo de la tendencia central (de asimetría) y cómo de apuntada es la distribución de las puntuaciones de la V (de curtosis).
  - **2V:** índices q nos indiquen hasta q punto están ambas V relacionadas entre sí (de asociación), procedimientos para predecir el valor de una V en función de otra (ecuaciones de regresión).
- **Estadística Inferencial**: se realizan inferencias acerca de una población basándose en los datos obtenidos a partir de una muestra. Estas generalizaciones a la población se basan en el cálculo de probabilidades.

En una investigación cualquiera, lo habitual es q se desee conocer un parámetro o carac. de los elementos de una población; sin embargo, suele ser demasiado extensa para estudiarla al completo *(conllevaría un coste inabordable)*. Por esto, se realiza un muestreo con el q se obtiene un conjunto de elementos q la representan y se estudia la carac. deseada en la muestra mediante estadísticos q se utilizarán para estimar los parámetros de la población.

- **Población**: conjunto de todos los elementos q cumplen una det. carac. objeto de estudio. *Niños con Trastorno por Déficit de Atención e Hiperactividad (TDAH) de la CCAA de Madrid.*
- **Muestra**: subconjunto cualquiera de una población. Personas, animales o cosas q cumplan una definición compartida por la población.

¿Por qué elegir un subconjunto y no trabajar con la población? Cuestión de viabilidad, no es posible trabajar con la población completa. El nº puede ser demasiado grande, puede haber familias (elementos de la población) quno deseen participar… Se trata de caracterizarla población. *Nos puede interesar conocer la inteligencia de los niños con TDAH. Utilizaríamos un test de inteligencia, el WISC-IV; al no tener acceso a la población completa de niños con TDAH se extrae una muestra de dicha población para obtener el nivel de inteligencia de cada niño de la muestra.*

- **Parámetro**: índice medido en una población q la describe de alguna manera.
  Propiedad descriptiva (una medida) de una población. Se denota con letras griegas; μ para la media, o^2x para la varianza y n para la proporción)
- **Estadístico**: índice medido en una muestra. Utilizando la estadística inferencial se pronostica el valor de los parámetros poblacionales a partir de los estadísticos muestrales. *Niños con TDAH se calcula la media en inteligencia de los de la muestra, q es el estadístico X, para pronosticar el valor medio en inteligencia de la población, parámetro μ (el valor q nos interesa)*.
  Con letras latinas; X para la media, S^2x para la varianza y P para la proporción.

3. VARIABLES: MEDICIÓN Y CLASIFICACIÓN

Variable: conjunto de valores resultantes de medir una carac. de interés sobre cada elemento indiv. de una población o muestra. Para representarlas se utilizan letras latinas mayúsculas. Para referirnos a un valor cualquiera de la variable X se utiliza el subíndice i (X_i); *n* el nº de elementos q componen la muestra. De manera genérica, se la designa como:

$$X_i \text{ siendo } i = 1,2,3\ldots, n$$

Si se trata de objetos físicos, el proceso de medición es directo y generalmente sencillo. Es cuestión de seguir unas reglas prescritas expresadas mediante det. escalas. *Medir la estatura de una persona asignando el nº correspondiente de la cinta métrica a la distancia q hay desde sus pies hasta su cabeza.* Cuando se trata de medir la timidez de un estudiante en una situación de interacción social, ya no es tan sencillo. El reto de la Psicología es su necesidad de medir en muchas ocasiones variables q no son directamente observables.

Medición: proceso por el cual se asignan nº a objetos o sucesos según det. reglas. Previo al análisis de datos, especifica el procedimiento de asignación de nº a los valores de la variable. *2 valores de la variable sexo (hombre y mujer) —> nº 1 y 2, y al peso de una rata el nº en gramos q da la balanza.* Para medir V psicológicas en muchas ocasiones se utilizan test psicológicos diseñados para ese fin. Su aplicación proporciona una puntuación para cada persona en esa variable.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Valoración de la calidad de vida, medida a través de una pregunta q forma parte de un test amplio y q se incluye en bastantes investigaciones sobre salud:
¿cómo calificarías tu calidad de vida? A) Muy mala. B) Regular. C) Normal. D) Bastante buena. E) Muy buena.
Asignar un nº a cada una de las opciones de R.

Psicología utiliza dif. escalas de medida (conjunto de reglas o modelos desarrollados para la asignación de números a las V) en función de la V a medir. *Escala centígrada de temperatura, q se basa en asignar 0º a la temperatura de congelación del agua y 100º a la de ebullición.*

En función de las relaciones matemáticas q puedan verificarse empíricamente entre los distintos valores de una V y, siguiendo la clasificación de Stevens (1946), pueden distinguirse 4 tipos de niveles o escalas de medida:

Escala nomial	Solo distinguiremos la igualdad o desigualdad entre 2 valores. Consiste en la asignación, puramente arbitraria de nº o símbolos a cada uno de los valores de la V. La única relación q se tiene en cuenta es la de igualdad (y desigualdad), q implica la pertenencia o no a una categoría det. Los valores de la variable se denominan **categorías**. Podemos decidir si un sujeto es igual o dif. a otro, pero no podemos establecer relaciones de orden respecto a esa V, ni de cantidad. *V enfermedad, distinguiendo entre: (1) «sanos» y (2) «enfermos», carece de sentido establecer relaciones entre estos 2 nº del tipo 1 + 1 = 2, sería considerar algo como q dos personas «sanas» = 1 «enferma».* Se puede asignar a cada valor de la V cualquier tipo de símbolo. **V cualitativas o categóricas**: Presentan un nivel de medida nominal. Se clasifican además, en función del nº de categorías q presentan. 2 = dicotómica *(sexo)*; +2 = politómica *(estado civil)*. En ocasiones se categorizan V q podrían medirse a un nivel superior; se ha dicotomizado si se han establecido 2 categorías, y politomizado si se han establecido +2 categorías. *V peso del roedor de un exp.: aunq podríamos medir exactamente su peso en gramos, puede resultar útil en una investigación dicotomizarla clasificando a las ratas en peso alto y bajo, o politomizarla, estableciendo 3/+ niveles de peso.*
Escala ordinal	Añade la posibilidad de establecer un orden. Se asignan nº a objetos para indicar la extensión relativa en q se posee una carac. sólo para indicar el orden de las posiciones. Los datos pueden utilizarse para jerarquizar u ordenar las observaciones, pero sin indicar la distancia q hay entre las posiciones. • Permite la identificación y diferenciación de los sujetos. • Permite establecer relaciones del tipo «mayor q» o «menor q», aunq no se plantea una distancia entre unas medidas y otras. La asignación de nº a las distintas categorías no puede ser completamente arbitraria, debe hacerse atendiendo al orden existente entre éstas. *V severidad de la enfermedad: 1 leve, 2 moderado y 3 grave. No es lo mismo padecer una con una intensidad leve o grave, y q la intensidad de la enfermedad en el caso de grave es mayor.* **V cuasicuantitativas** .
Escala de intervalo	Se usa 1 unidad y tienen sentido las diferencias. Ordenan los objetos según la magnitud del atributo q representan y proveen intervalos iguales entre las unidades de medida. Existencia de una unidad de medición común y constante. El origen es arbitrario, y no refleja en ningún momento ausencia de la magnitud que estamos midiendo. Los nº asignados: • Permiten decidir si un objeto es igual o dif. a otro. • Si posee en mayor o menor grado la carac. de interés. • La distancia entre los distintos valores consecutivos de la V es la misma. *La inteligencia medida con un test. 4 personas (A, B, C y D) han obtenido 80, 90, 150 y 160 puntos, podemos decir q la dif. en inteligencia entre A y B es la misma q entre C y D (90-80 = 160-150), el test proporciona una unidad de medida estable. Sin embargo, no se puede afirmar q D sea el doble de inteligente q A aunq tenga el doble de puntuación en el test, para realizar una afirmación de ese tipo sería necesario q el cero de la escala fuera absoluto*. En este caso es arbitrario pq obtener un 0 no refleja ausencia de la carac. medida, no significa q no se posea ni un ápice de inteligencia.
Escala de razón	Se pueden comparar 2 medidas mediante un cociente. Los nº asignados admiten como válidas las relaciones de igualdad-desigualdad, orden, suma, resta, multiplicación y división. Tiene todas las carac. de una medida de intervalo y se le puede asignar un punto de origen verdadero de valor 0 (= ausencia de la magnitud q estamos midiendo). El 0 ya no es arbitrario, sino un valor absoluto, se puede afirmar q A tiene 2, 3 o 4 veces la magnitud de la propiedad presente en B. Altura y peso. *Si una rata de laboratorio pesa 350 gramos y otra 175, la 1ª pesa el doble q la 2ª.*

En muchas ocasiones el nivel de medida de una V va a depender de cómo se haya definido. *La V calificación obtenida en el examen de Lengua de la PAU puede suscitar dudas razonables sobre su nivel de medida. Si se define como el nivel de conocimientos de Lengua necesarios para ingresar en la universidad se trataría de una de intervalo pq con esta interpretación el 0 sería arbitrario (0 ≠ ausencia total de los conocimientos necesarios sino q se ha obtenido un rendimiento nulo en las preguntas en concreto). Si en lugar del nivel de conocimientos, nos interesa simplemente contar el nº de aciertos se trataría de razón, aquí el 0 sí es absoluto e indicaría ausencia absoluta de preguntas acertadas.*

WUOLAH
Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Es muy importante la definición operativa de una V (cómo se define y se registra) pq puede det. su nivel de medida.
La mayoría de las V psicológicas se considera q están medidas en una escala de intervalo. *Perseverancia, rasgo de personalidad, se ha medido mediante una prueba psicológica o test. Sin embargo, si se define como el nº de intentos o ensayos q realiza una persona para conseguir un objetivo se trata de una escala de razón.*

Las V medidas en escala de intervalo y de razón son cuantitativas.
En función de los valores numéricos q pueden asignarse:

- **V continua**: para la q, dados 2 valores, siempre se puede encontrar un 3ero q esté incluido entre los 2 primeros.
  El peso, entre los valores 79 y 80 kg. se pueden considerar 1, 2, 3 o todos los decimales q se quiera.
- **V discreta**: adopta valores aislados. Fijados 2 valores consecutivos, no se puede tomar ninguno intermedio.
  Nº de hijos.

Tipo de variable	Escala de Medida	Características básicas	Relaciones válidas	Ejemplos
Cualitativa: ▪ Dicotómica ▪ Politómica	→ Nominal	Los números identifican y clasifican objetos	Relaciones del tipo «igual que» o «distinto que»	Sexo, estado civil, raza, diagnóstico clínico.
Cuasicuantitativa	→ Ordinal	Además, los números indican las posiciones relativas de los objetos	Además, relaciones del tipo «mayor que» o «menor que»	Dureza, posición en el ranking de la ATP, grado de satisfacción.
Cuantitativa: ▪ Discreta ▪ Continua	Intervalo	Además, hay una unidad de medición común	Además, igualdad o desigualdad de diferencias	Temperatura en grados centígrados, inteligencia.
	Razón	Además, el punto cero es absoluto	Además, igualdad o desigualdad de razones	Longitud, peso, altura, tiempo de reacción.

Todas las puntuaciones de test psicológicos se consideran de intervalo (V cuantitativa)

4. DESCRIPCIÓN DE V: DISTRIBUCIÓN DE FRECUENCIAS Y REPRESENTACIÓN GRÁFICA

Una vez q el investigador ha recabado la inf. a través del proceso de medida y recogido los datos correspondientes, dispone de un listado o base, comúnmente llamado **matriz de datos**. La generación de una base de datos supone la codificación previa de las observaciones, la introducción de estos en algún programa informático, depuración de los datos ya grabados (detección y tratamiento de los errores de grabación y valores perdidos), y eventualmente la realización de transformaciones de V q faciliten su posterior tratamiento estadístico. Hay muchos programas estadísticos para organizar y analizar los datos. En concreto, Excel para hacer distribuciones de frecuencia, gráficos y diversos análisis.

Codificar datos es asignar nº a las V cualitativas y cuasicuantitativas, y registrar los valores de las V cuantitativas q constituyen la BD, así como asignar un código (espacio en blanco o valor numérico) a los perdidos (no registrados u observados). En la matriz de datos, los casos se sitúan en las filas y las V en las columnas.

Una vez que los datos están codificados es preciso realizar una depuración de la BD, q conlleva el procesamiento de:

- Los datos perdidos; valores q no han sido registrados, habitualmente pq el participante no ha consignado ese dato. Existen procedimientos de imputación de datos, basados en los valores válidos de otros casos q se utilizan en ocasiones en V cuantitativas.
- Datos atípicos: valor muy dif. al resto de valores de la misma V. Suelen ser ocasionados por errores al introducir los datos o valores extremos. Distorsionan los resultados de los análisis, hay q identificarlos y tratarlos de manera adecuada, generalmente excluyéndolos.

Si los datos han sido registrados manualmente en un software es recomendable hacer un control de calidad de la grabación de los mismos, revisando la codificación de un porcentaje de los casos, habitualmente un 5%-10% del total.

Una vez depurada, se utiliza para extraer la inf. relevante. Si tenemos muy pocos datos es posible q la simple inspección visual sea suficiente para describir el fenómeno estudiado. Pero no es nada frecuente. Habitualmente el número de datos es elevado, por lo que se hace necesario organizar la inf. mediante una **distribución de frecuencias**: tabla en la q se resume la inf. disponible de una V. Se sitúan los valores de la V por filas y en las columnas se dispone el nº de ocurrencias por cada valor, porcentajes…

- Facilita la lectura de la inf. q contienen los datos, organización de estos.
- Ofrece la inf. necesaria para realizar representaciones gráficas.
- Facilita los cálculos para obtener los estadísticos q serán objeto de estudio en los próximos temas.

Descripción de variables cualitativas

V cualitativa: distribución de frecuencias y su representación gráfica mediante un diagrama de barras o de sectores.
En la 5ª columna aparece el Bachillerato elegido. Sin embargo, la simple inspección visual no es suficiente para hacerse una idea precisa de cuántos estudiantes han elegido cada una de las modalidades.

X	n_i	p_i	P_i
1. Ciencias y Tecnología	13	0,325	32,5
2. Humanidades y CC Sociales	21	0,525	52,5
3. Artes	6	0,15	15
Σ	40	1	100

$p_i = n_i / n$
$P_i = p_i \times 100$

En la distribución de frecuencias de variables cualitativas habitualmente se muestran las **frecuencias absolutas**, **relativas y** los **porcentajes**.

1. Se inspeccionan los valores q toma la V. *En este caso cualitativo (nominal) q puede adoptar 3 valores.*
2. En la 1ª columna se especifican los valores q adopta la variable X o el nº asignado a ese valor.
3. **Frecuencia absoluta** (n_i) q es el nº de observaciones en cada categoría.
   La suma de todas representa el total de la muestra (n)
4. **Frecuencia relativa o proporción** (p_i): q se obtiene dividiendo la frecuencia absoluta, n_i, entre el nº total de observaciones, q se representa por n**.**
5. La frecuencia relativa tb se expresa en términos de **porcentaje** (P_i); multiplicar cada una de las proporciones x100.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

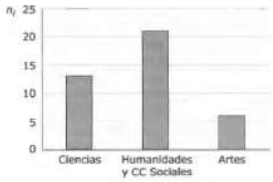
Stuvia.com - The Marketplace to Buy and Sell your Study Material

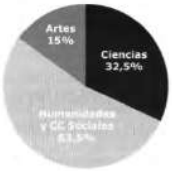
Ahora sí podemos hacernos una idea de la distribución de los estudiantes según el Bachillerato quhan elegido; el más demandado es el de Humanidades y CCSS (52,5%) y el menos demandado es el de Artes (15% del total).

Los 2 gráficos + habituales en la descripción de V cualitativas son los:

- **Gráficos de barra**: los distintos valores de la V se sitúan en el eje horizontal y las frecuencias/porcentajes en el eje de ordenadas. Cada barra representa una categoría, siendo su altura igual a su frecuencia (o porcentaje).
- **Gráficos de sectores**: cada sector representa una categoría de la V y su ángulo central debe ser proporcional a su frecuencia (o porcentaje).

El diagrama de barras se ha construido sobre las frecuencias absolutas de la variable y el diagrama de sectores sobre los porcentajes.





<u>El único índice apropiado para V cualitativas es la moda.</u>

Descripción de variables ordinales o cuasicuantitativas

En las V ordinales se procede de la misma manera, aunq con los valores situados en la tabla de acuerdo a un det. orden. *V nivel de estudios de la madre presenta los valores: Primarios, ESO, Bachillerato, Grado universitario y Posgrado universitario.*
En la distribución de frecuencias hay que preservar este orden, ya sea empezando por el valor + bajo o + alto de la V:

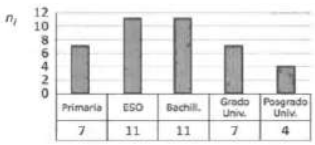
3 columnas +, para obtener estos valores, simplemente hay q ir acumulando (sumando), desde la cat. de menor valor a la de mayor. *Fecuencia absoluta acumulada en Bachillerato es 29, resultado de + las frecuencias de los valores anteriores (7 + 11 = 18) y la suya propia (18 + 11 = 29), 29 personas presentan un nivel de estudios de Bachillerato o inferior.* En las nominales carece de sentido el cálculo de las frecuencias acumuladas, ya q sus valores no establecen un orden det.

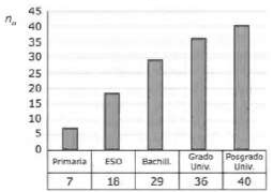
X	n_i	p_i	P_i	n_a	p_a	P_a
1. Primarios	7	0,175	17,5	7	0,175	17,5
2. ESO	11	0,275	27,5	18	0,450	45
3. Bachillerato	11	0,275	27,5	29	0,725	72,5
4. Grado universitario	7	0,175	17,5	36	0,900	90
5. Posgrado universitario	4	0,1	10	40	1	100
Σ	40	1	100			

- **Frecuencia absoluta acumulada** (n_a): nº de veces q se repite cada valor o cualquiera de los valores inferiores.
- **Proporción acumulada o frecuencia relativa acumulada** (p_a): cociente entre la frecuencia absoluta acumulada y el total de observaciones.
- **Porcentaje acumulado** (P_a) : valor de la frecuencia relativa acumulada multiplicado por cien.

$P_a = n_a / n$.
$P_a = P_a \times 100$

Generalmente tb representadas con un diagrama de barras o de sectores.
El diagrama de barras tb se puede realizar sobre las frecuencias, proporciones o porcentajes acumulados, respetando el orden.





<u>Algunos índices apropiados son la mediana y la moda y la amplitud intercuartil.</u>

Descripción de variables cuantitativas

Puede pasar q el nº de la variable sea:

1. Muy reducido; *nº de hijos*. Se ordenarán las frecuencias procediendo de forma indicada como para las ordinales.
2. Muy amplio; *V ansiedad ante los exámenes*. **Se agrupará la V en intervalos**: se forman grupos con valores consecutivos. Cada uno de estos grupos es situado en una fila y se calculan las frecuencias de cada uno o intervalo de valores y no de cada valor de la V.
   - <u>Decidir el nº de intervalos</u>. Siempre habrá muchas posibilidades; desde un nº pequeño a muchos intervalos de muy pequeña amplitud. Hay q tener presente q siempre se pierde inf. Dependerá del tratamiento q el investigador quiera dar, encontrando el equilibrio entre precisión del estudio y manejabilidad de datos.
   - Así, se formarán los **límites aparentes del intervalo**. *Entre 1 y 5, 5 y 10…* Para cada intervalo existe un límite superior e inferior.
     Tienen la misma unidad de medida q los valores de la V. Si los datos son enteros, los límites aparentes son enteros. Si contienen decimales, los tendrán el mismo nº de decimales q los datos recogidos.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Con los límites aparentes, existirá una discontinuidad entre un intervalo y el siguiente, ya q el límite superior de un intervalo no coincide con el inferior del siguiente intervalo. Con los límites exactos de una distribución no existe discontinuidad, ya q el superior exacto de un intervalo coincide con el inferior exacto del intervalo sig.

- **Límite Inferior Exacto** (LIE): restando al valor del límite inferior aparente media unidad de medida.
- **Límite Superior Exacto** (LSE): sumando al valor del límite superior aparente media unidad de medida.

Los límites exactos del intervalo 1-5 son 0,5-5,5, los del intervalo 6-10 son 5,5-10,5 y así sucesivamente.

A partir de los límites aparentes o exactos se calcula el **punto medio del intervalo**; la semi-suma del límite superior e inferior. Se utilizará para algunos cálculos de índices estadísticos con distribuciones agrupadas en intervalos.

$$PM = \frac{LIE + LSE}{2} \quad ó \quad PM = \frac{LIA + LSA}{2}$$

- Si los límites aparentes son enteros, la unidad de medida de la V es 1, y su mitad es 0,5, cantidad q habrá q restar al límite inferior y sumar al superior para calcular los límites exactos.
- Si los límites aparentes son nº con un decimal, la unidad de medida de la V es 0,1, por lo q la cantidad a sumar y restar para calcular los límites exactos será 0,05.
- Si los límites aparentes son nº con 2 decimales, la unidad de medida de la V es 0,01, por lo q la cantidad a sumar y restar para calcular los límites exactos será 0,005.
- Así sucesivamente...

Amplitud: dif. entre el LSE y el LIE

X Límites aparentes	X Límites exactos	X Punto medio	n_i	p_i	n_a	p_a
1 – 5	0,5 – 5,5	3	13	0,325	13	0,325
6 – 10	5,5 – 10,5	8	12	0,3	25	0,625
11 – 15	10,5 – 15,5	13	8	0,2	33	0,825
16 – 20	15,5 – 20,5	18	4	0,1	37	0,925
21 – 25	20,5 – 25,5	23	2	0,05	39	0,975
26 – 30	25,5 – 30,5	28	1	0,025	40	1
Σ			40	1		

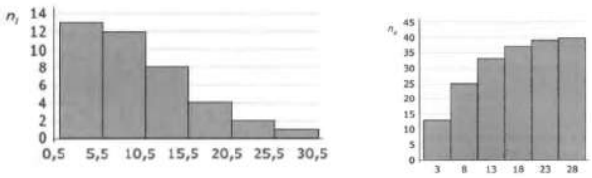
Si se mide el tiempo q se emplea en ejecutar una det. tarea, y los valores resultantes oscilan ente 3,01s y 3,30s, se podría establecer una distribución de frecuencias con 6 intervalos:
En ese caso nuestra unidad de medida es 0,01, ya q contienen 2 decimales. Para calcular los límites exactos hay q sumar y restar la mitad de esta unidad de medida que es 0,005. Así, los límites exactos serían:

X Límites aparentes	X Límites exactos
3,01 – 3,05	3,005 – 3,055
3,06 – 3,10	3,055 – 3,105
3,11 – 3,15	3,105 – 3,155
3,16 – 3,20	3,155 – 3,205
3,21 – 3,25	3,205 – 3,255
3,26 – 3,30	3,255 – 3,305

Intervalo abierto: q no tiene límite inferior o superior. *V ansiedad hubiera 2 sujetos con puntuación de 41 y 43, se puede optar por establecer el intervalo abierto «+ de 30», en lugar de añadir los 3 correspondientes, 2 con frecuencia nula.*

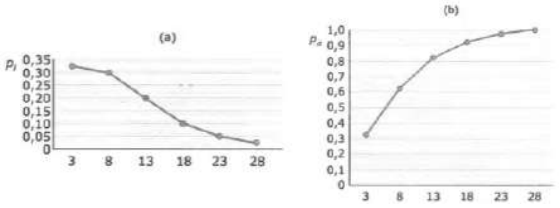
- **Histograma**: extensión del diagrama de barras. Dibuja los rectángulos unidos entre sí, indicando q existe continuidad en los valores. Es un gráfico de V continua dividida en intervalos en los q se eleva un rectángulo con área proporcional a su frecuencia. Puede construirse sobre frecuencias absolutas, relativas o porcentajes, ya sean o no acumulados. En el eje horizontal se sitúan los límites exactos de los intervalos o su punto medio.

(a) histograma acumulado sobre los límites exactos (b) histograma de la V ansiedad ante los exámenes sobre los puntos medios.



- **Diagrama de lineas**: sitúa un punto a una altura proporcional a la frecuencia en cada valor o en el punto medio de cada intervalo (si la V está agrupada en intervalos). Se unen los puntos para formar una línea. Tb polígono de frecuencias.

Diagrama de líneas de la V ansiedad ante los exámenes en proporciones (a) Diagrama de líneas acumulativo y en proporciones acumuladas (b)



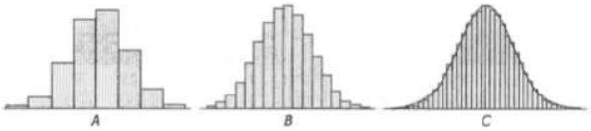
Los gráficos + habituales para representar a una V cuantitativa discreta son el diagrama de barras y de líneas.
Las V cuantitativas continuas agrupadas en intervalos en lugar del diagrama de barras se utiliza el histograma.
<u>Para describir una V cuantitativa se utilizan algunos índices estadísticos, + frecuentes, la media y la desviación típica.</u>

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

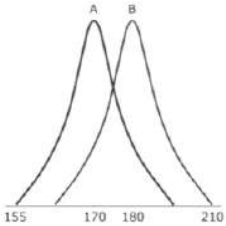
Tendencia central, variabilidad y forma de una V: aproximación gráfica

Cómo describir las V mediante los índices estadísticos adecuados, q se utilizan para medir la tendencia central, variabilidad y forma de la distribución. Pero, antes hay q ver gráficamente q carac. de la V pretenden evaluar utilizando curvas suavizadas, q son histogramas basados en un gran nº de observaciones, cuyos ángulos se han suavizado.
Así, si disponemos de los datos de una muestra en una variable X (A) y hacemos esos intervalos más pequeños (B), y más pequeños aún (C), al trazar un diagrama de líneas sobre los puntos medios, la línea resultante será una curva.



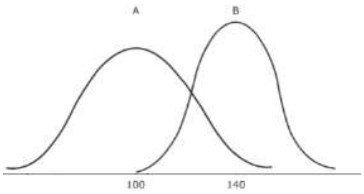
Tendencia central

Se refiere al lugar donde se centra una distribución particular en la escala de valores.
Estatura medida en un grupo de 1000 hombres. La tendencia central de los 2 grupos es distinta (las curvas no se solapan completamente) y el B (nacido en 1990) tiene una estatura promedio mayor q el A (pq la curva del grupo B está situada a la derecha, en puntuaciones + altas de estatura).
Esta puede cuantificarse mediante unos índices conocidos como estadísticos de tendencia central.



Variabilidad

Grado de concentración de los valores entre sí o con respecto a un valor central de la distribución. Una distribución de frecuencias es **homogénea** (tiene poca variabilidad) si los valores están cercanos al promedio y es **heterogénea** (tiene mucha variabilidad) si los valores se dispersan mucho con respecto al promedio.
G A representa las puntuaciones en inteligencia medidas en un G de niños de distintos colegios de la geografía española; el B las de un G con altas capacidades. Además de una tendencia central distinta (GB nivel mayor de inteligencia) podemos apreciar q las puntuaciones en el G con altas capacidades están + próximas entre sí q las del otro grupo. G A presenta una mayor variabilidad en inteligencia.

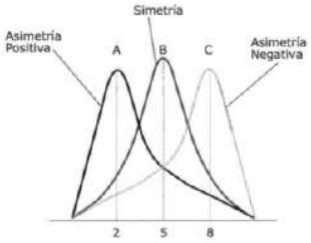


Forma

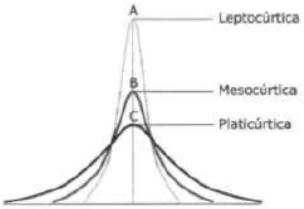
Para estudiar la forma de una variable se analiza su asimetría y su curtosis.

- **Asimetría**: grado en q los datos se reparten equilibradamente por encima y debajo de la tendencia central.
  - Asimetría positiva: mayor concentración de puntuaciones en la parte baja de la escala.
  - Asimetría negativa: mayor parte de las puntuaciones se sitúan en la parte alta de la escala.
  - Simétrica cuando al dividirla en 2 partes iguales, las 2 mitades se superponen.

Puntuación obtenida por un G de alumnos en un examen muy difícil (A), intermedio (B) y muy fácil (C). El conjunto de puntuaciones presenta una distribución asimétrica positiva si la mayoría de las puntuaciones obtenidas son bajas (A, difícil), simétrica cuando hay un número similar a ambos lados (B) y asimétrica negativa si la mayoría son altas (C, fácil).



- **Curtosis**: grado de apuntamiento de los datos. Si la distribución de frecuencias es muy apuntada se llama **leptocúrtica** (A), y si es muy aplastada se denomina **platicúrtica** (C) y si es intermedio, **mesocúrtica** (B).



Alumno	Xi (nº h)
1	7
2	11
3	16
4	5
5	14
6	10
7	12
8	10

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

T2: ÍNDICES DE TENDENCIA CENTRAL Y POSICIÓN

Como se ha mencionado en el tema anterior, una de las propiedades + importantes de una distribución de frecuencias es la **tendencia central**; valor o puntuación q represente al conjunto de observaciones. Con el fin de cuantificarla se han desarrollado una serie de medidas o estadísticos q indican sobre q puntuación se concentran las observaciones.

1. ÍNDICES DE TENDENCIA CENTRAL

En el análisis descriptivo de la distribución de frecuencias de una V, es habitual q el nº de observaciones sea grande y q nos planteemos resumir, mediante valores numéricos, sus principales propiedades. Nos interesa calcular un valor central q actúe como resumen numérico para representar al conjunto de datos; **medidas, índices o estadísticos** de tendencia central, q representan toda la distribución de frecuencias con un único valor y facilitan la comparación de dif. conjuntos de puntuaciones de una V. *Medimos el nivel de autoestima en 200 niños (100 ♂ y 100 ♀), además de estudiar la tendencia central en niños y niñas de forma conjunta, estos posibilitan la comparación de niños y niñas en su grado de autoestima. Para saber si es mayor en un sexo q en otro.*

3 medidas de tendencia central representativas de la distribució: media aritmética, mediana y moda.

Media aritmética / promedio

- El + conocido y usado. Por la sencillez de su cálculo y es el fundamento de un gran nº de técnicas estadísticas.
- Indica la tendencia general de una distribución de frecuencias de una V y es el valor central alrededor del cual están la mayoría de las observaciones.
- Desde una perspectiva geométrica: «centro de gravedad» de la distribución de frecuencias.
- **Sólo** puede calcularse para **V cuantitativas** (nivel de medida de intervalo o de razón).
- Suma de todos los valores observados de la V divididos por el nº total de observaciones.
- Utiliza toda la inf. disponible en los datos. Es necesario utilizar todas las puntuaciones de los participantes, no como los estadísticos.

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum X_i}{n}$$

X_i = valor de la V en el sujeto
n = nº total de observaciones

Cálculo en una tabla de distribución de frecuencias

Como el nº de observaciones suele ser muy grande, es usual q los datos se presenten aquí agrupados o no en intervalos. A partir de las frecuencias absolutas (n_i) o frecuencias relativas o proporciones (p_i).

Absolutas

$$\bar{X} = \frac{\sum n_i X_i}{\sum n_i} = \frac{\sum n_i X_i}{n}$$

X_i = valor *i* la variable X o punto medio del intervalo
n = nº total de observaciones
n_i = frecuencia absoluta del valor o intervalo *i*

Esta es la expresión general de la media a partir de las frecuencias absolutas. La fórmula anterior, definida para pocas observaciones, no es más q un caso particular en el e las frecuencias absolutas de cada valor es igual a uno. En efecto, si n_i= 1 para todos los valores de X:

$$\bar{X} = \frac{\sum n_i X_i}{n} = \frac{\sum 1 \cdot X_i}{n} = \frac{\sum X_i}{n}$$

Como se ve, es la fórmula previa.

Relativas

$$\bar{X} = \sum p_i X_i$$

P_i = frecuencia relativa o proporción de observaciones

Como es de esperar, con ambas fórmulas se obtiene el mismo resultado.

! Si se tienen las frecuencias absolutas acumuladas (n_a) y no las absolutas (ni), se deben obtener para poder calcular la media.

Frecuencias de las calificaciones en el examen de Lengua PAU				
X_i (nota)	n_i	n_iX_i	$p_i = n_i / n$	$p_i X_i$
1	1	1	0,025	0,025
2	1	2	0,025	0,05
3	1	3	0,025	0,075
4	8	32	0,2	0,800
5	5	25	0,125	0,625
6	7	42	0,175	1,05
7	8	56	0,2	1,400
8	7	56	0,175	1,400
9	1	9	0,025	0,225
10	1	10	0,025	0,250
Σ	40	236	1	5,9

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
WOOLAH Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

En una distribución de frecuencias agrupadas en intervalos se calcula igual teniendo en cuenta q los valores de X de la fórmula (Xi) serán los puntos medios de cada intervalo.

G1				
X	X_i	n_i	$p_i = n_i / n$	n_iX_i
1-5	3	5	0,25	15
6-10	8	4	0,2	32
11-15	13	6	0,3	78
16-20	18	3	0,15	54
21-25	23	1	0,05	23
26-30	28	1	0,05	28
Σ		20	1	230

G2				
X	X_i	n_i	$p_i = n_i / n$	n_iX_i
1-5	3	8	0,4	24
6-10	8	8	0,4	64
11-15	13	2	0,1	26
16-20	18	1	0,05	18
21-25	23	1	0,05	23
26-30	28	0	0	0
Σ		20	1	155

Puntuaciones agrupadas en intervalos de la V ansiedad antes de los exámenes. G1, sin tratamiento y G2 con tratamiento.

Con frecuencias absolutas (multiplicar columnas ni y Xi);
$X = \Sigma n_iX_i / n$
G1: 230/20 = 11,5
G2: 155/20 = 7,75

Con frecuencias relativas (multiplicar columnas pi y Xi):
$X = \Sigma p_iX_i$
G1: 11,5
G2: 7,75

Propiedades matemáticas de la media aritmética

1. En una distribución, la suma de las desviaciones de cada valor con respecto a su media es igual a 0.

$$(X_1 - \bar{X}) + (X_2 - \bar{X}) + \ldots + (X_n - \bar{X}) = \sum_{i=1}^{n}(X_i - \bar{X}) = 0 \qquad \sum_{i=1}^{n}(X_i - \bar{X}) = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n}\bar{X} = n\bar{X} - n\bar{X} = 0$$

Si si se dispone de un nº mayor de observaciones en el q se repiten valores y éstos se presentan mediante una distribución de frecuencias agrupados o no en intervalos, la expresión para comprobar la propiedad es:

$$\sum_{i=1}^{n} n_i (X_i - \bar{X}) = 0$$

Se resta el X_i a la media de distribución y esa diferencia se multiplica x la frecuencia absoluta de cada intervalo (n_i) obteniendo una nueva columna; $n_i(X_i - X)$. Hay q tener en cuenta la frecuencia absoluta de cada valor (n_i), -nº de veces q aparece cada puntuación o intervalo-.

2. Si a cada puntuación X_i de la variable X le sumamos una constante a (elegida arbitrariamente), la media de las nuevas puntuaciones es igual a la media de X más la constante. Si $Y_i = X_i + a$, entonces $\bar{Y} = \bar{X} + a$. —> Cuando se suma una constante cualquiera (*a*) a las puntuaciones de una distribución (X_i), la media de esa distribución ($\bar{X}$) se ve afectada, y para obtener la nueva media ($\bar{Y}$) tb se debe sumar esa constante a la media original.

3. Si cada puntuación X_i de la variable X se multiplica por una constante *b* (elegida arbitrariamente), la media de las nuevas puntuaciones es igual a la media de $\bar{X}$ multiplicada por la constante. Y; = b • X;, entonces $Y_i = b \cdot X_i$. Cando multiplicamos una constante cualquiera (b) por cada una de las puntuaciones de una distribución (X_i), la media de esa distribución ($\bar{X}$) se ve afectada y para obtener la nueva media ($\bar{Y}$) tb debemos multiplicar esa constante a la media original.

4. <u>La media de *J* muestras o media ponderada</u>:
Hasta ahora se ha hablado de la media de una V en una muestra con n casos u observaciones. Si se cuenta con la media de varios grupos e interesa conocer la media de todas las observaciones juntas, disponemos de las puntuaciones en la variable X en *J* muestras distintas o grupos con n_1, n_2, …, n_J observaciones y con medias $\bar{X}_1$, $\bar{X}_2$, …, $\bar{X}_J$ respectivamente. La media total de los *J* grupos, q incluye las puntuaciones de todas las muestras:

J = nº de grupos o muestras

$$\bar{X}_p = \frac{n_1 \cdot \bar{X}_1 + n_2 \cdot \bar{X}_2 + \ldots + n_J \cdot \bar{X}_J}{n_1 + n_2 + \ldots + n_J} = \frac{n_1 \cdot \bar{X}_1 + n_2 \cdot \bar{X}_2 + \ldots + n_J \cdot \bar{X}_J}{n}$$

Como se puede apreciar, la media de los *J* grupos no es más q una ponderación de las medias de cada grupo en base al nº de observaciones de dicho grupo (n_1). Es decir, la media de cada grupo tiene un peso en la media total q está en función del tamaño de la muestra o nº de casos de cada grupo. Por ello, a la media total (de todas las puntuaciones) se le denomina **media ponderada ($\bar{X}_p$)**.
No podemos calcular la media de una V medida en distintos grupos como la media de las medias de los grupos. Es necesario tener en cuenta el peso de la media de cada grupo a través del nº de casos de cada grupo ($n_J \cdot \bar{X}_J$).
En el caso particular en el q todos los grupos presentan el mismo nº de observaciones (k), es decir, n1 = n2 = … = n1 = k, la fórmula de la media ponderada se simplifica pq la media total se corresponde con la media de las medias de los grupos.

$$\bar{X} = \frac{\bar{X}_1 + \bar{X}_2 + \ldots + \bar{X}_J}{J}$$

WUOLAH
Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

La media ponderada tb se utiliza para obtener la media global en una puntuación q se basa en distintas pruebas con pesos diferentes en su importancia para la puntuación final. *A las prácticas se le ha otorgado un peso del 60% y al examen 40% en la calificación final. Si la media de las prácticas = 6 y en el examen = 5, la nota media final:*

$$\overline{X}_p = \frac{p_1 \cdot \overline{X}_1 + p_2 \cdot \overline{X}_2}{p_1 + p_2} = \frac{60 \times 6 + 40 \times 5}{100} = 5,6$$

P1 y P2: pesos asignados a las prácticas

Limitaciones de la media aritmética

- Cuando los datos están agrupados en intervalos, la media no se puede calcular si el intervalo máximo no tiene límite superior y/o el mínimo no tiene límite inferior.
- Es sensible a la existencia de unas pocas observaciones con valores extremos en la distribución de frecuencias. Como en distribuciones marcadamente asimétricas, por lo q afecta a su representatividad como valor central de la distribución. Estos valores extremos pueden ser producto de errores en la recogida (entonces se eliminan) o grabación de los datos, o valores q aportan inf. relevante de la V (entonces se recomienda aplicar otros índices menos sensibles a los valores extremos como la mediana).

La mediana (Md)

Valor q divide la distribución de frecuencias de la V en 2 partes iguales, cada una el 50% de las observaciones.

- Útil cuando la distribución es asimétrica una buena alternativa a la media aritmética para resumir la tendencia central de las puntuaciones. No se ve afectada por los valores extremos q pueda adoptar la V debido a q en su cálculo no intervienen todos los valores de la distribución, únicamente los q ocupan las posiciones centrales.
- Es un valor + apropiado para representar la tendencia central de la distribución.
- Se puede obtener en todo tipo de V excepto cualitativas.
- Su valor no tiene por qué coincidir con un valor real de la V (especialmente en cuantitativas discretas). Se trata de un valor q cuantifica la tendencia central de la distribución.

1. Se ordenan las *n* puntuaciones de menor a mayor.
2. Se observa si el nº de observaciones *n* es impar o par.
   - Impar; el valor de la mediana es el de la observación q ocupa la posición central dentro de ese conjunto de observaciones ya ordenadas. Coincide con la posición (n+1)/2.
   - Par; es la media aritmética de los 2 valores centrales de la distribución, los q ocupan las posiciones $X_{n/2}$ y $X_{(n/2)+1}$:

$$Md = \frac{X_{n/2} + X_{(n/2)+1}}{2}$$

$X_{n/2}$ = valor de la V en la posición n/2
$X_{(n/2)+1}$ = valor en la posición (n/2)+1

V: nº de h de estudio semanales.
La 4ta y la 5ta posición son los valores centrales de la distribución (de 8 = n).

4ta: (n/2 = 8/2 = 4)
5ta: ((n/2)+1) = 5

Md = 10+11/2 = 10,5

Aquí tb es normal es q el nº de observaciones no sea tan pequeño y aparezcan valores repetidos = datos presentedos en tablas de distribución de frecuencias agrupados o no en intervalos. En este caso, el intervalo en el q se encuentra la mediana se denomina **intervalo crítico** y se corresponde con aquél en el q la frecuencia absoluta acumulada n_a es igual o superior a n/2 o la proporción acumulada (p_a) es igual o mayor a 0,50:

$$Md = L_i + \left(\frac{\frac{n}{2} - n_d}{n_c}\right) \cdot I$$

Ej pág 65

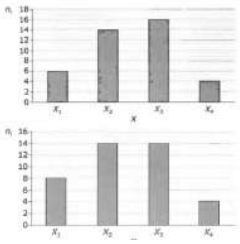
L_i = Límite inferior exacto del intervalo crítico
n = Nº de observaciones
n_d = Frecuencia absoluta acumulada por debajo del intervalo crítico
n_c = Frecuencia absoluta del intervalo crítico
I = Amplitud del intervalo crítico

- Siempre se empieza desde el valor más bajo de la V hasta el más alto, pq se define como el nº de veces q se repite cada valor o cualquiera de los valores inferiores.
- El intervalo crítico es el 1er intervalo (empezando por el intervalo de valores de X más pequeño) cuya frecuencia acumulada sea igual o mayor a n/2 es decir, al 50% de n. Debe ser igual o superior.
- El origen de la fórmula planteada se basa en el **método de interpolación**, en el q se asume la distribución homogénea de las puntuaciones dentro de cada intervalo.
- Cuando los datos no están agrupados en intervalos, el cálculo a es un caso particular de la fórmula anterior en la q la amplitud (I)= 1 y los límites exactos se obtienen sumando y restando 0,5 unidades a cada valor de la V.
- Se puede calcular en cualquier distribución de frecuencias de V, excepto cuando cualitativas o agrupadas en intervalos en la q existe uno abierto y éste es el intervalo crítico en el q se encuentra la mediana -pq necesitamos conocer su amplitud-.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

La moda (Mo)
Valor o categoría de la V con mayor frecuencia absoluta.

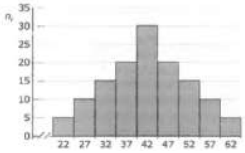


- Tanto en V cualitativas como en cuantitativas.
- Cuando en una V existe un único valor con la frecuencia absoluta máxima, la distribución presenta una única moda y es **unimodal**. Si son 2 los valores con la frecuencia más alta la distribución es **bimodal**, si son 3, **trimodal**...
- Puede ocurrir q una distribución no tenga moda, **distribución amodal**, todos los valores tienen la misma frecuencia absoluta.
- En V cualitativas: es la categoría con la máxima frecuencia (n_1)
  En V cuantitativas con datos no agrupados en intervalos, el valor con mayor frecuencia absoluta.
  En una distribución de una V cuantitativa con datos agrupados en intervalos, se localiza en el intervalo modal (intervalo con la frecuencia máxima) y es su punto medio (si el intervalo es 1-5, la moda = 3).

Principales carac. de la moda:
- Índice de cálculo e interpretación sencillos.
- Único índice q tb puede aplicarse a V cualitativas.
- Cuando los datos están agrupados en intervalos o existen abiertos, se puede calcular a no ser q el intervalo modal coincida con el abierto.

Elección de un índice de tendencia central
Como 1ª opción se recomienda la media aritmética, pq en ella están basadas un gran nº de técnicas estadísticas de gran importancia y uso frecuente (a no ser q tenga distribución asimétrica). Entonces, la siguiente opción es la mediana, q se puede obtener en V con nivel de medida ordinal, y se puede calcular en distribuciones con datos agrupados en intervalos con intervalos abiertos. Sin embargo, en ocasiones no se puede obtener
1. El nivel de medida de la V es nominal.
2. Con datos agrupados en intervalos, se encuentra en el intervalo abierto.

Entonces, la única alternativa es utilizar la moda, q no se puede calcular cuando la distribución sea amodal o el intervalo abierto coincide con el intervalo modal.



Hoy en día, con el uso de programas informáticos para el análisis estadístico de los datos, se recomienda, siempre y cuando sea pertinente, el cálculo de los 3. Es interesante resaltar q cuando la distribución de una V cuantitativa es simétrica y unimodal, coinciden los valores de la media, mediana y moda.

Nivel de medida	Tipo de variable	Índice estadístico de tendencia central aplicable
Nominal	Cualitativa	*Mo*
Ordinal	Cuasicuantitativa	*Mo, Md*
De intervalo De razón	Cuantitativa discreta Cuantitativa continua	*Mo, Md,* $\overline{X}$

2. ÍNDICES DE POSICIÓN
Hasta ahora hemos definido medidas q representaban al conjunto de datos para disponer de un indicador o resumen numérico de la tendencia central de todas las puntuaciones. Ahora la cuestión va dirigida a un sujeto o dato particular: en una distribución de frecuencias de una V, un sujeto «s» obtiene una puntuación X , ¿q posición ocupa este en la distribución con respecto al resto?, ¿q puntuación tendría q obtener para superar a un porcentaje det.?

Los índices estadísticos de posición informan acerca de la posición relativa de un sujeto con respecto a su grupo dentro de la distribución de frecuencias de la V. Debemos dividir la distribución en un nº de partes o secciones iguales entre sí en cuanto al nº de observaciones. Los 3 -percentiles, cuartiles y deciles- se usan con mucha frecuencia en la presentación de resultados estadísticos, especialmente los dos 1ºs. En Psicología, por ej para las normas de interpretación de las puntuaciones de los tests o baremos. Asimismo, su aplicación requiere al menos de un nivel de medida ordinal en la V objeto de estudio.

Percentiles (o centiles)
Los 99 valores q dividen en 100 partes iguales la distribución de frecuencias de la V. El **percentil k**, denotado por P_k, es deja por debajo de sí un porcentaje k de observaciones, donde k=1, 2...99 de la V de interés.
No son porcentajes, sino valores q dejan por debajo de sí un det. tanto por ciento o porcentaje de las observaciones o casos. Aunq el concepto es sencillo y fácil de entender, no hay una única manera de calcularlo. De hecho no es de extrañar q dependiendo del procedimiento aplicado y el software informático obtengamos percentiles ligeramente distintos, aunq todos válidos según la definición general.

El percentil 50, P_{50}, de una distribución deja por debajo de sí al 50% de las observaciones y por encima al otro 50% = la mediana o percentil 50. Por este motivo, el cálculo de los percentiles lo vamos a realizar utilizando una extensión del método de la mediana. La diferencia estriba en q, en la mediana se trataba de localizar la posición de n/2 en la columna

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

de las frecuencias absolutas acumuladas. En cambio, en los percentiles y de forma más general, se hace en base al nº n·k/100, el nº de casos q se correponden con el k% del percentil y se obtiene a partir de una sencilla regla de 3: si *n* observaciones son el 100% de los participantes, ¿cuántas observaciones serán el k%?

$$\left.\begin{array}{r} n \text{ observaciones} \rightarrow 100\% \text{ de los participantes} \\ x \rightarrow k\% \text{ de los participantes} \end{array}\right\} \Rightarrow x = \frac{n \times k}{100}$$

n = Nº de casos, suma de todos los n_i
K = Percentil a obtener

Este número, n·k/100, es igual a n/2 cuando calculamos el percentil 50.

Cálculo con datos agrupados en intervalos

1. Saber q nº de casos de todos (*n*) deja por debajo de sí el percentil k. Con n·k/100.
2. Localizar el intervalo en el q se encuentra el percentil *k*. Intervalo crítico, aquel donde la frecuencia absoluta acumulada (n_a) es igual o superior a n·k/100, es decir, al k% de *n*.
3. Obtenerlo utilizando la sig. fórmula:

$$P_k = L_i + \left(\frac{\frac{n \cdot k}{100} - n_d}{n_c} \right) \cdot I$$

n_d = Frecuencia absoluta acumulada por debajo del intervalo crítico
n_c = Frecuencia absoluta o del intervalo crítico
L_i = Límite inferior exacto del intervalo crítico
I = Amplitud del intervalo

- Cuando n·k/100 es exactamente igual a la frecuencia acumulada hasta un valor o intervalo el percentil se corresponde directamente con el límite superior exacto del intervalo crítico y no es necesario aplicar la fórmula.
- Cuando se tienen muy pocos datos no es habitual calcular percentiles pq tienen poca utilidad. Pero se les aplicaría tb la formula general asumiendo los intervalos con amplitud igual a 1.

Puede suceder q se tenga un valor o puntuación de la variable, X_i, y nos interese saber q percentil ocupa ese valor en la distribución. Realmente se está pidiendo el valor de k dado el valor de X_i, despejándo la k de la ecuación anterior.
Cálculo de K para X_i

$$k = \left[\frac{\frac{(P_k - L_i) \cdot n_c}{I} + n_d}{n} \right] \cdot 100$$

Podemos obtener un valor con decimales, y dado q los percentiles son 99 valores enteros, tomamos la cantidad entera + próxima. Si el primer decimal es =/mayor a 5, tomamos el nº entero superior; si es >5, el entero inferior.

Otra situación es q se pida el percentil de una puntuación q es, al mismo tiempo, el límite exacto superior de un intervalo y el inferior del siguiente. Se puede elegir cualquiera de los 2 como crítico y obtendríamos el mismo resultado.

Cuartiles y deciles

Las secciones o partes en las q se divide la distribución de frecuencias son muchas menos q en los percentiles.

Los **cuartiles** son 3 valores de la distribución q dividen en 4 partes de igual frecuencia a la distribución.

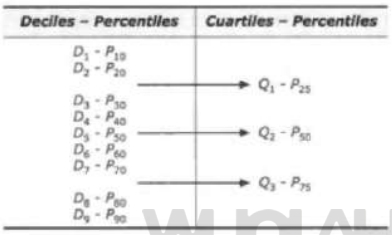
- **1er cuartil,** Q_1, deja por debajo al 25% de las observaciones y r encima al 75%. Se corresponde con el percentil 25 de la distribución; $Q_1 = P_{25}$.
- **2º cuartil**, Q_2, deja por debajo de sí al 50% y por encima al otro 50%. Equivalente al percentil 50 y a la mediana de la distribución, $Q_2 = P_{50}$ = Md.
- **3er cuartil,** Q_3, deja por debajo de sí al 75% y por encima al 25%. Se corresponde con el percentil 75 de la distribución, $Q_3 = P_{75}$.

Debido a la equivalencia con los percentiles, para el cálculo se utilizan los métodos para estos. Q_1 se calcula mediante P_{25}, Q_2 con P_{50}, y Q_3 con P_{75}. Se utilizan para construir índices para el estudio de la variabilidad de una distribución de frecuencias.

Los **deciles** son 9 valores q dividen en 10 partes iguales a la distribución. Se representan por D_i, donde i = 1,2,...,9.
El primer decil, D_1 deja por debajo de sí al 10% de las observaciones, el D_2 al 20%, el D_3 al 30% y así hasta el D_9.

D_1=P10, D_2=P20 ... D_5=P50=Md … D_9=P90.

Por lo tanto, tb se pueden calcular a partir de los percentiles correspondientes.



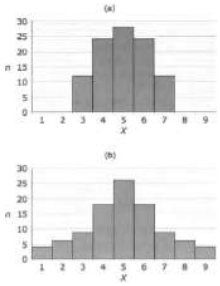
Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

T3: MEDIDAS DE VARIABILIDAD Y FORMA

2 nuevas propiedades de una distribución de puntuaciones: variabilidad o dispersión y forma de la distribución.

1. MEDIDAS DE VARIABILIDAD



El estudio de una distribución resultaría incompleto sin el análisis de la tendencia central; la variabilidad de los datos. Es el grado de variación en un conjunto de puntuaciones; en q se asemejan o diferencian entre sí o se aproximan o alejan de una medida de tendencia central.

Se han propuesto numerosos índices para medirla; la amplitud total, varianza y desviación típica, y amplitud intercuartil. Además, se presenta un índice, el coeficiente de variación, útil para comparar distintas distribuciones de frecuencias en términos de su variabilidad.

En la imagen, el gráfico a) presenta menor variabilidad = muestras más homogéneas. En el caso extremo y poco habitual de máxima homogeneidad, todos los valores de la V serían iguales entre sí y a la media, y no habría variabilidad. Si existe cierta dispersión, la muestra es más o menos heterogénea y las puntuaciones difieren entre sí.

2 tipos de índices:
- Miden el grado en el q las puntuaciones se asemejan o diferencian entre sí.
  - Amplitud total o rango.
  - Amplitud intercuartil.
- Se mide con respecto a alguna medida de tendencia central como la media aritmética.
  - Varianza.
  - Desviación típica.

Tanto unos como otros son útiles para el estudio de la variabilidad pero poco adecuados al comparar la dispersión de 2/+ distribuciones. Para realizar dicho análisis, un índice apropiado es el **coeficiente de variación**, q se basa en la relación entre la desviación típica y la media de cada distribución de frecuencias.

Amplitud total o rango / recorrido de las observaciones

Amplitud total = A_T.

$$A_T = X_{max} - X_{min}$$

Distancia en la escala numérica de un conjunto de puntuaciones entre los valores q representan la puntuación máxima y mínima. En V agrupadas en intervalos, la máx es el límite superior exacto del intervalo máximo y la mínima, el límite inferior exacto del intervalo mínimo.

Su principal inconveniente, es que por su facilidad y utilización de poca inf., es sensible únicamente a los valores extremos de la distribucion y no captura la poca o mucha dispersión entre el resto de valores.

Varianza y desviación típica

La medida de variabilidad tb se puede basar en la distancia entre las puntuaciones y un valor central de la distribución como la media aritmética. Un 1er índice q se puede plantear de forma lógica es el **promedio de las desviaciones o diferencias de cada puntuación** con respecto a su media.

$$\overline{X}_d = \frac{\sum d_i}{n} = \frac{\sum (X_i - \overline{X})}{n}$$

El problema de este índice es q el sumatorio del numerador siempre = 0, por lo q carece de sentido como índice. Para utilizar un índice con estas desviaciones, evitando q sea =0, 2 soluciones:

1. Calcular el valor absoluto de cada desviación antes de realizar la suma, obteniendo la **desviación media**:

$$DM = \frac{|X_1 - \overline{X}| + |X_2 - \overline{X}| + \cdots + |X_n - \overline{X}|}{n} = \frac{\sum |X_i - \overline{X}|}{n}$$

   Se emplea muy poco en la actualidad pq es poco manejable matemáticamente por el uso del valor absoluto.

2. Basarnos en el cuadrado de las diferencias y así obtenemos la **varianza**, de un conjunto de *n* puntuaciones en una variable X, denotada por S^2_X se define como el promedio de los cuadrados de las desviaciones de las puntuaciones con respecto a la media:

$$S_X^2 = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2}{n} = \frac{\sum (X_i - \overline{X})^2}{n}$$

   Primero se elevan al cuadrado las diferencias y después se obtiene el promedio de esas desviaciones al cuadrado. Se pueden simplificar los cálculos con la siguiente fórmula:

$$S_X^2 = \frac{\sum X_i^2}{n} - \overline{X}^2$$

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Cuando los datos se presentan en tablas de distribución de frecuencias es necesario tener en cuenta la frecuencia de cada intervalo:

$$S_X^2 = \frac{\sum n_i (x_i - \bar{x})^2}{\sum n_i} = \frac{\sum n_i (x_i - \bar{x})^2}{n}$$

$$S_X^2 = \frac{\sum n_i X_i^2}{\sum n_i} - \bar{X}^2 = \frac{\sum n_i X_i^2}{n} - \bar{X}^2$$

n = nº total de observaciones
X_i = valor *i* en la variable X o el punto medio del intervalo
n_i = frecuencia absoluta del valor o del intervalo *i*

Varianza a partir de una distribución de frecuencias relativas:

$$S_X^2 = \sum p_i X_i^2 - \bar{X}^2$$

p_i = frecuencia relativa o proporción de observaciones del valor o del intervalo *i*.

La varianza, al basarse en diferencias al cuadrado, es un nº positivo q se expresa en las unidades de la V al cuadrado. Si la variable X se mide en metros, las desviaciones con respecto a la media (xi - x̄), tb vendrán expresadas en metros, mientras q al elevarlas al cuadrado, son + fácilmente interpretables. Se calcula la raíz cuadrada de la varianza y se obtiene un índice q se denomina **desviación típica**.

La desviación típica de un conjunto de *n* puntuaciones, S_x, es la raíz cuadrada positiva de la varianza:

$$S_X = \sqrt{S_X^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

Tanto la varianza como la desviación típica son índices de dispersión muy útiles en el desarrollo posterior de la estadística inferencial, estando en la base de numerosas técnicas estadísticas. A la hora de cuantificar la variabilidad, la desviación típica se suele utilizar + debido a q se expresa en las mismas unidades de medida q la V objeto de estudio. Propiedades de ambos índices:

1. Su cálculo requiere el uso de todas las puntuaciones observadas en la distribución.
2. Miden la variabilidad de los datos con respecto a la media aritmética; únicamente deben aplicarse si es apropiado utilizar la media como medida de tendencia central.
3. Siempre son no negativas, pueden ser = o > q 0. Son iguales a cero únicamente si todas las puntuaciones son iguales entre sí = no variabilidad o dispersión. En el resto de los casos son positivas, siendo sus valores mayores a medida q aumenta la variabilidad de las puntuaciones.
4. Si a las puntuaciones de la variable X les aplicamos una transformación lineal: $Y_i = bX_i + a$ la varianza de las nuevas puntuaciones Y será $S^2_Y = b^2S^2_X$; y la desviación típica será $S_Y = |b|S_X$. Es decir, si a una variable X se le suma o resta una constante *a*, la varianza y desviación típica de la V original no se ven afectadas. En cambio, cuando multiplicamos los valores de X por una constante *b*, la varianza queda multiplicada por la constante al cuadrado y la desviación típica por el valor absoluto de dicha constante.

Por último, otro índice de variabilidad relacionado es la **cuasivarianza**, donde se divide por *n*-1. Y su **cuasidesviación típica** = su raíz cuadrada. Son medidas de dispersión para inferencia estadística. T10.

$$S_{n-1}^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

$$S_{n-1} = \sqrt{S_{n-1}^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Coeficiente de variación

Es frecuente q uno de los objetivos del análisis descriptivo de los datos sea la comparación del grado de variabilidad o dispersión entre 2 conjuntos de puntuaciones en una misma o distintas V. Como por lo general se miden en unidades distintas no tiene sentido compararlas en base a los valores de sus varianzas o desviaciones típicas. La solución es definir un índice de variabilidad relativa q no dependa de las unidades de medida; el **coeficiente de variación**.

Está definido para variables con X > O y es recomendable q su resultado se acompañe de la media y desviación típica de la distribución a partir de las cuales ha sido calculado. Cuando comparamos 2 conjuntos de puntuaciones de la misma V, tb es necesario para comparar la dispersión de ambas distribuciones. Únicamente es posible utilizar la desviación típica cuando la media de ambos grupos es la misma = mismas conclusiones con ambos índices.

$$CV = \frac{S_X}{\bar{X}} \cdot 100$$

Amplitud intercuartil

La varianza y desviación típica + media aritmética, son los estadísticos recomendados para estudiar la variabilidad y la tendencia central de una distribución de frecuencias. En ocasiones, por la asimetría de la distribución, no son aconsejables; necesitamos un índice resistente de dispersión adecuado, q se utilizaría junto con la mediana como medida de tendencia central, sería la **amplitud intercuartil, A_{IQ}**, o rango intercuartil = la diferencia entre el 3er y el 1er cuartil.

$$A_{IQ} = Q_3 - Q_1 = P_{75} - P_{25}$$

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Como se puede observar, este índice no informa de la variabilidad del conjunto de puntuaciones, sino del 50% de las mismas comprendidas entre el percentil 25 y el 75 de la distribución.

2. MEDIDAS DE FORMA

Otro aspecto importante es la forma q presenta la distribución, q se estudia a través de 2 propiedades, la asimetría (visto en el T1) y la curtosis. A continuación se van a describir 2 índices de asimetría y un coeficiente de curtosis, q aportan datos numéricos a ambas propiedades de la forma de la distribución.

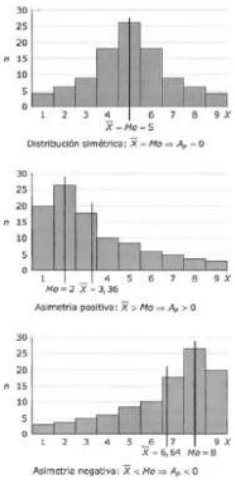
Asimetría de una distribución

La asimetría indica el grado en el q las puntuaciones se reparten por debajo y por encima de la medida de tendencia central. Existen dif. índices para cuantificarla. Vamos a describir los 2 q se utilizan con + frecuencia:

Índice de asimetría de Pearson: se basa en la relación entre la media y la moda, Es un índice **adimensional** (no tiene unidades de medida) q se aplica a **distribuciones unimodales** (con una única moda).

- Con distribución simétrica, media y moda coinciden = el numerador se anula y el valor de $A_p = 0$.
- Con asimetría positiva, la media es mayor q la moda, $A_p > 0$.
- Con asimetría negativa, el valor de la moda es superior al de la media, $A_p < 0$.

$$A_p = \frac{\overline{X} - Mo}{S_X}$$



Relación entre la representación gráfica de la asimetría de una distribución y el índice de asimetría de Pearson:

Índice de asimetría de Fisher: Se basa en las distancias de las puntuaciones respecto a su media elevadas al cubo, por lo q su valor puede ser positivo, negativo o 0.

$$A_F = \frac{\frac{\sum (X_i - \overline{X})^3}{n}}{S_X^3} = \frac{\sum (X_i - \overline{X})^3}{n \cdot S_X^3}$$

Si los datos se presentan en tablas de distribución de frecuencias —>

$$A_F = \frac{\sum n_i (X_i - \overline{X})^3}{n \cdot S_X^3}$$

Este tiene en cuenta todas y cada una de las puntuaciones de la muestra por lo q puede considerarse el mejor. Al igual q el de Pearson su valor es 0 si la distribución es simétrica ($A_F = 0$); menor q 0 si la distribución es asimétrica negativa ($A_F < 0$); y mayor q 0 si es asimétrica positiva ($A_F > 0$) .

Apuntamiento o curtosis de una distribución

La curtosis se refiere al grado de apuntamiento de los datos en la distribución de frecuencias. Tomando como referencia la curva normal, puede adoptar 3 formas:

A. **Leptocúrtica**: distribución muy apuntada, mayor q en la ''normal". Positiva ($C_r > 0$).
B. **Platicúrtica**: muy aplastada, menor q lo ''normal". índice negativo ($Cr < 0$).
C. **Mesocúrtica**: grado de apuntamiento intermedio."normal". índice 0 ($C_r = 0$).

Se basa en las distancias de cada puntuación respecto a la media elevadas a la 4ª potencia:

$$C_r = \frac{\sum (X_i - \overline{X})^4}{n \cdot S_X^4} - 3$$

Si los datos se presentan en tablas de distribución de frecuencias —>

$$C_r = \frac{\sum n_i (X_i - \overline{X})^4}{n \cdot S_X^4} - 3$$

3. DIAGRAMA DE CAJA

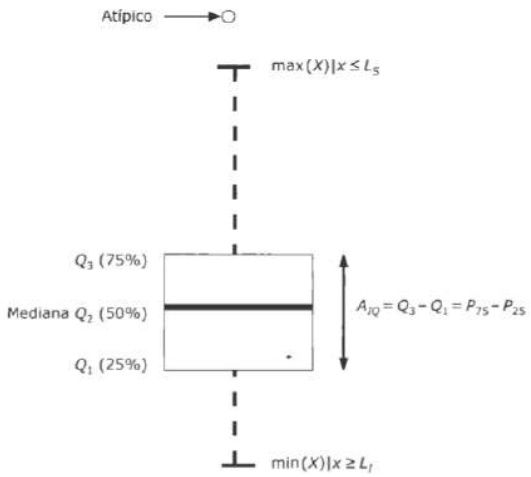
O gráfico de caja y bigotes *(boxplots o box and whiskers)* fue propuesto por Tukey (1977). Es una presentación visual útil para estudiar la asimetría de una V cuantitativa, y detectar si hay valores extremos o atípicos (outliers) en la distribución de frecuencias (sin agrupar en intervalos).

Se representa mediante una caja rectangular cuya altura = la amplitud o rango intercuartil $A_{IQ} = Q_3 - Q_1 = P_{75} - P_{25}$. Dentro se dibuja una línea para indicar dónde se sitúa la mediana -q coincide con el 2º cuartil-. Es atravesada por una línea vertical llamada **bigote**, en cuyos extremos se sitúan los valores mínimos y máximos de la V (sin considerar los valores atípicos en caso de q existan). Los límites q determinan si un valor es atípico se calculan multiplicando la amplitud intercuartil por 1,5 y restando este resultado al Q_1 (cálculo del límite inferior) o sumándolo al Q_3 (cálculo del límite superior).

$$L_S = Q_3 + A_{IQ} \times 1,5 \qquad L_I = Q_1 - A_{IQ} \times 1,5$$

Downloaded by: giselamartin | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
WUOLAH Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material



Cuando existen casos extremos o atípicos, aparecen como un círculo pequeño por encima o debajo de los bigotes del diagrama de caja. Para estudiar la asimetría se va a tener en cuenta la longitud de los bigotes y el nº de casos atípicos en ambas colas:

- Bigotes de la misma longitud y mismo nº de casos atípicos en ambos lados: aproximadamente simétrica.
- Bigotes de igual longitud pero + casos atípicos en un extremo = distribución con asimetría.
- Bigotes de dif. longitud = distribución asimétrica, como es el caso q se ha representado. La longitud del bigote superior es mayor que la del bigote inferior.

4. PUNTUACIONES TÍPICAS

Hasta ahora hemos tratado fundamentalmente con **puntuaciones directas** *(de un sujeto en un test)*. Estos son los 1ºs datos de los q habitualmente disponemos, pero la comparación de las puntuaciones directas de un mismo sujeto en 2 V distintas puede llevarnos a confusión pq nos ofrecen muy poca inf. De hecho, conocida una puntuación directa no sabemos si se trata de un valor alto o bajo pq esto depende del promedio del grupo. Una solución es trabajar con puntuaciones diferenciales.

$$x_i = X_i - \overline{X}$$

Si a una puntuación directa x_i le restamos la media de su grupo obtenemos una **puntuación diferencial** o de diferencia, q representamos por x_i (minúscula):
Estas aportan más inf.; si la puntuación coincide con la media de su grupo, es inferior o es superior a ella.
Estas puntuaciones presentan las siguientes propiedades:

A. su media es cero: $\bar{x} = 0$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{\Sigma\left(X_i - \overline{X}\right)}{n} = \frac{\Sigma X_i - \Sigma\overline{X}}{n} = \frac{\Sigma X_i}{n} - \frac{n\overline{X}}{n} = \overline{X} - \overline{X} = 0$$

B. La varianza de las puntuaciones diferenciales es igual a la varianza de las puntuaciones directas:

$$S_x^2 = \frac{\Sigma\left(x_i - \bar{x}\right)^2}{n} = \frac{\Sigma x_i^2}{n} = \frac{\Sigma\left(X_i - \overline{X}\right)^2}{n} = S_X^2$$

$$z_x = \frac{x_i}{S_x} = \frac{X_i - \overline{X}}{S_x}$$

Por tanto, al restar a las puntuaciones directas su media hemos obtenido una nueva escala con media 0 y con idéntica varianza a las puntuaciones directas. Sin embargo, 2 puntuaciones diferenciales idénticas pueden tener un significado muy dif. en función de la media y de la varianza de las distribuciones de las q proceden; entonces se utilizan las **puntuaciones típicas**, q van + allá y nos permiten no sólo comparar las puntuaciones de un sujeto en 2 V distintas, tb 2 sujetos distintos en 2 pruebas o V distintas. Reflejan las relaciones entre las puntuaciones con indep. de la unidad de medida = permiten hacer comparaciones entre distintos grupos e incluso entre distintas V.
Al proceso de obtener puntuaciones típicas se llama **tipificación**. Indican el nº de desviaciones típicas q se aparta de la media una det. puntuación. Tienen las siguientes propiedades:

A. Su media es 0

$$\bar{z}_x = \frac{\sum z_x}{n} = \frac{\Sigma\left(\frac{x_i}{S_x}\right)}{n} = \frac{\frac{1}{S_x}\Sigma x_i}{n} = \frac{\sum x_i}{nS_x} = \frac{0}{nS_x} = 0$$

B. Su varianza es igual a 1

$$S_{z_x}^2 = \frac{\sum\left(z_x - \bar{z}_x\right)^2}{n} = \frac{\sum\left(z_x\right)^2}{n} = \frac{\Sigma\left(\frac{x_i}{S_x}\right)^2}{n} = \frac{\frac{1}{S_x^2}\Sigma x_i^2}{n} = \frac{1}{S_x^2}\frac{\sum x_i^2}{n} = \frac{1}{S_x^2}S_x^2 = 1$$

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

T4: RELACIÓN ENTRE VARIABLES I

Estudiar conjuntamente +1 V nos va a permitir responder a preguntas. Al igual q en la descripción de 1 única V, el procedimiento a utilizar es dif. en función del tipo de V q se trate:

- **Ambas cualitativas**
  ¿Hay relación entre el sexo de los pacientes y el trastorno psicológico que padecen?
  Si existe o no cierta relación o asociación q hace que el valor q adopte una se asocie en alguna medida con det. valores de la otra. El estadístico q se utiliza para comprobar si existe esa relación o son Indep. es x^2, junto a algunos coeficientes q permiten valorar la fuerza de la asociación entre dichas V.
- **Ambas ordinales**
  Su tratamiento dependerá de la cantidad de valores q adopten ambas variables: si es un nº muy reducido, entonces se utilizarán los mismos procedimientos q para V cualitativas y si es amplio se utilizarán índices estadísticos adaptados como el coeficiente de correlación por rangos de Spearman.
- **Ambas cuantitativas**
  ¿Hay relación entre la motivación y el rendimiento académico de los estudiantes en una det. asignatura?
  Lo q se realmente se cuestiona es si al aumentar la motivación se incrementa tb su rendimiento y a la inversa. El coeficiente de correlación momento-producto de Pearson es el índice + utilizado.
  En el caso de encontrar relación se puede ir un paso más allá y predecir una V en función de la otra. Así, se podría pronosticar el rendimiento en función de la motivación, o mejor aún, en función de su motivación y Cociente Intelectual (CI) mediante el análisis de la regresión.
- **Una cualitativa y otra cuantitativa**
  ¿Hay relación entre la puntuación obtenida en una pregunta del examen y la obtenida en todo el examen?
  En q medida una predice a la otra. Se puede utilizar el coeficiente de correlación biserial puntual, muy directamente relacionado con el coeficiente q cuantifica la relación entre 2 V cuantitativas (el coeficiente de correlación momento-producto de Pearson) .

Para estudiar 2 V cualitativas, se utilizarán 3 estrategias relacionadas:

- **Tabla de contingencia**: forma resumida de representar los datos de las 2 V. informa sobre las frecuencias conjuntas (de ambas V) y marginales (de cada una de las 2 por separado), permite valorar la relación de ambas mediante el estudio de las distribuciones condicionadas de 1, agrupadas en función de los valores de la otra.
- **Diagramas de barras conjuntos**: representación gráfica apropiada. diagrama de barras adosadas y de barras apiladas.
- **Estadístico x2**: para comprobar la indep. entre 2 V cualitativas. Se proponen varias medidas globales para valorar la fuerza de la asociación: los coeficientes *C* de Contingencia, *V* de Cramer, y ϕ; índices, basados en el estadístico x^2, q tratan de superar algunas de sus limitaciones.

En V ordinales: el coeficiente de correlación por rangos de Spearman.

1. ASOCIACIÓN ENTRE 2 VARIABLES CUALITATIVAS

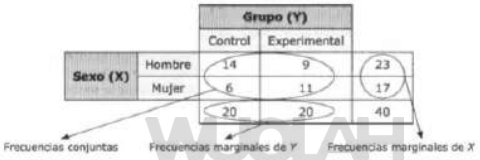
- A lo largo de ellas sólo es posible establecer categorías no ordenadas; q pueden ser intercambiadas.
- Pueden ser dicótómicas o politómicas.
- Tb se considerarán aquellas q, en un principio, tienen un mayor nivel de medida (ordinal, intervalos o razón) pero a posteriori, han sido categorizadas.
- Se dice q hay asociación si existe algún tipo de tendencia o patrón de emparejamiento entre sus distintos valores.

Tabla de contingencia

Forma de ordenar los datos para estudiar la relación entre V con pocas categorías. Es una distribución de frecuencias clasificada de acuerdo a los valores q pueden tomar las 2 V. Por eso, se sitúan los valores de una en las filas y los de la otra en las columnas.

Frecuencias conjuntas (n_{ij}): nº de individuos q toman el valor X_i en la variable X, e Y_j en la variable Y. La suma de todas representa el total de la muestra (*n*). Toma en consideración 1 de los valores de las 2 V.
Frecuencias marginales: Totales de cada valor de una única V. Hay de la variable X y de la variable Y. La suma de estas de cada V tiene q ser igual al total de la muestra.
Es muy frecuente tener +2 categorías en alguna de las V. El formato general es el mismo, añadiendo filas o columnas, y calculándose las distintas frecuencias de la forma indicada.

Distribución marginal: distribución de frecuencias unidimensional (marginal) q nos informan del nº de observaciones para cada valor de una de las V prescindiendo de la inf. sobre los valores de las demás. Hay una de la variable X (q contiene todas las frecuencias marginales de X) y una de la variable Y (q contiene todas las frecuencias marginales de Y).
Distribución condicionada: especifica las observaciones q hay de cada valor de una de las V al imponer la condición de q la otra tome un valor det. Hay una **de Y condicionada a un valor de X_i** q considera únicamente una fila y una **de X condicionada a un valor de Y_i,** q únicamente tiene en cuenta una columna.



Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Want to earn $1.236 extra per year?
Distribution of this document is illegal
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

		Variable Y				Total
		y_1	y_2	...	y_j	
Variable X	x_1	n_{11}	n_{12}	...	n_{1j}	n_{1+}
	x_2	n_{21}	n_{22}	...	n_{2j}	n_{2+}
	⋮	⋮	⋮	...	⋮	⋮
	x_i	n_{i1}	n_{i2}	...	n_{ij}	n_{i+}
Total		n_{+1}	n_{+2}	...	n_{+j}	n

Siendo $x_1,\dots,x_i$, los distintos valores de la variable X;
$y_1,\dots,y_j$, los distintos valores de la variable Y

			Grupo (Y)		Total
			Control	Experimental	
Sexo (X)	Hombre	n_{ij}	14	9	23
		P_{ij} del total	35%	22,5%	57,5%
		P_i por Sexo	60,87%	39,13%	100%
		P_j por Grupo	70%	45%	57,5%
	Mujer	n_{ij}	6	11	17
		P_{ij} del total	15%	27,5%	42,5%
		P_i por Sexo	35,3%	64,7%	100%
		P_j por Grupo	30%	55%	42,5%
Total		n_{ij}	20	20	40
		P_{ij} del total	50%	50%	100%
		P_i por Sexo	50%	50%	100%
		P_j por Grupo	100%	100%	100%

Aquí aparecen varios tipos de frecuencias absolutas:

- La de cada casilla o celda: surge de la distribución conjunta por combinación de 2 valores: nº de casos q comparten 2 características a la vez (n_{ij})·
- La total de cada valor o categoría de la V. El conjunto de estos valores = **distribución marginal absoluta**: nº de casos q tienen 1 carac., de la variable X o fila (n_{j+}) o de la variable Y o columna (n_{+j})·
- El total de casos analizados es *n* o n_{++}, de una muestra o del total.

Todas las de esta, son frecuencias absolutas (n_{ij}), pero de manera habitual, se presentan además con inf. de los porcentajes. Hay 3 tipos de porcentajes conjuntos q se pueden utilizar en ellas:

- **Porcentaje del total (P_{ij})**: nº de casos de cada celdilla dividido por el total de casos (*n*) y multiplicado por 100.
- **Porcentaje condicionado a X_i o porcentaje por fila**: nº de casos de cada celdilla dividido por el total de casos por fila y multiplicado por 100. El conjunto de estos valores = **distribución condicional de filas**.
  Hay q establecer la condición previa de pertenecer a uno de los 2 valores de X_i, *En la Tabla hay q fijarse en la 1ª fila, 23 hombres. Calculando solamente dentro de esta, la distribución de porcentajes para los 2 grupos son 60,87% hombres asignados al grupo control y 39,13% al experimental. Estos 2 porcentajes suman 100, ya q para su cálculo estamos considerando únicamente hombres.*
- **Porcentaje condicionado a Y_j** o por columna: nº de casos de cada celdilla dividido por el total de casos por columna y multiplicado por 100. El conjunto = **distribución condicional de columnas**.
  Hay q establecer la condición previa de pertenecer a uno de los 2 valores de Y_j. *En la 1ª columna hay 20 personas asignadas al grupo control. Calculando los porcentajes dentro de esta, el 70% son hombres y el 30% mujeres; Q suman 100.*

Al considerar los porcentajes para interpretar la relación hay q tener en cuenta si es simétrica o asimétrica.

- **Relación asimétrica**: una de las 2 se considera como factor explicativo de la distribución de la otra. Los porcentajes se calculan en el sentido de la V explicativa, por lo q la suma de los porcentajes en cada categoría referidos al total marginal de esa categoría será el 100%. Dicho de otra forma, si la se sitúa en las columnas de la tabla de contingencia, para hacer las comparaciones se calcularán los porcentajes por columna.
- **Simétrica**, no existe esa distinción. Se puede utilizar cualquiera de los porcentajes.

Cada uno de los 3 tipos de porcentajes pone el énfasis en una distribución diferente y ofrece comparaciones distintas, según el sentido de la predicción. Su utilización permite eliminar la influencia del tamaño de la muestra y de los marginales, por lo q se pueden realizar comparaciones entre valores de las distribuciones condicionadas, q indican la existencia de relación o no entre las V, así como la naturaleza de la relación.

Representación gráfica: diagrama de barras conjunto

Apropiado cuando al menos una de las 2 V es cualitativa. Se construye sobre los datos de la tabla de contingencia, situando una en el eje horizontal, y para identificar la otra se utilizan barras de distinto color o trama. Básicamente hay 2 formas de representarlo; barras adosadas y barras apiladas.

Diagrama de barras adosadas

Barras colocadas horizontalmente o verticalmente la frecuencia de cada casilla del interior de la tabla de contingencia. Para cada valor de la variable X se representa, una al lado de otra, la frecuencia con q se presenta cada valor de Y dentro de ese valor de X. Al estar situadas unas junto a otras permite la comparación rápida entre y dentro de cada V. Es habitual mostrarlo en porcentajes del total o condicionales (por fila o por columna).

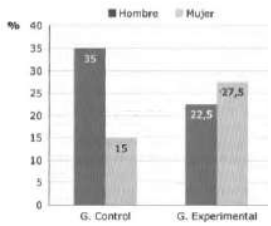
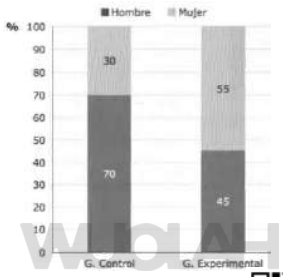


Diagrama de barras apiladas

Muestra una barra por cada valor q toma la variable Y, las cuales se dividen en distintos colores q representa a cada valor de la variable X. Indica la frecuencia con la q aparece cada valor de X en cada valor de Y, comparando entre categorías, la aportación de cada valor al total. Es la + adecuada para visualizar porcentajes condicionados.



Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Medidas globales de asociación entre variables cualitativas

Independencia: x^2

Asociación = la distribución de las frecuencias de los valores de 1 de las 2 V difiere en función de los valores de la otra.
Independencia = no existe tal patrón de relación entre sus valores.

Para saber si existe o no independencia se utiliza el estadístico x^2, q se basa en la comparación de las frecuencias conjuntas. Así, se comparan las frecuencias empíricas (u observadas) con las frecuencias teóricas (o esperadas) suponiendo q no hubiera asociación. Al compararlas, si no existen diferencias se concluye la ausencia de asociación o de relación de interdependencia = son independientes entre sí.

$$\chi^2 = \sum\sum \frac{(n_e - n_t)^2}{n_t}$$

n_e = frecuencia empírica (u observada)
n_t = frecuencia teórica (o esperada)

- **Frecuencia teórica**: aparece en cada casilla en caso de independencia; se calcula multiplicando las 2 frecuencias marginales y dividiendo por la frecuencia total *n*.
- **Frecuencia empírica**: la q se corresponde con los datos de la casilla.

$$n_t = \frac{Total\ fila \times Total\ columna}{n}$$

Una vez conocidas todas las frecuencias se puede calcular x_2. El sumatorio engloba toda la fracción, por lo q se van a sumar 4 fracciones, una por cada celdilla. En cada una, la frecuencia empírica - la teórica; el resultado se eleva al cuadrado y se divide entre la frecuencia teórica. El índice x^2 toma el valor 0 cuando 2 V son independientes, siendo mayor q 0 cuando exista asociación entre ellas, tanto mayor cuanto más intensa. Ahora bien, no tiene un límite máximo, una dificultad a nivel interpretativo. A veces no podemos saber (sin aplicar técnicas de estadística inferencial) si el valor está lo bastante próximo a 0 como para considerar q la relación es mínima o considerable.

Otro inconveniente es q al multiplicar las frecuencias de todas las casillas por una constante, aumenta, a pesar de q las proporciones de todas las casillas sean las mismas antes y después de dicha multiplicación. Esto hace q su valor solo pueda compararse para V en tablas de contingencia del mismo tamaño (I x J) y con el mismo n.

Características del estadístico x^2 :

- **Adopta valores entre 0 y +∞**. Dado q está definido por valores elevado al cuadrado y las frecuencias nunca son negativas, no puede tomar valores negativos.
- Únicamente adopta el valor 0 si la frecuencia empírica de la celdilla es igual a la teórica q le corresponde, en todas las celdillas de la tabla de contingencia.
- El tamaño de la muestra, n, debe ser relativamente grande. El criterio q se utiliza habitualmente es q la frecuencia esperada mínima por casilla sea al menos de 5 en aprox. el 80% de las casillas, considerando además q la frecuencia mínima esperada en cada casilla sea 1.
- Sirve para valorar la existencia o no de indep., pero no resulta apropiado para medir la intensidad de la relación, pues el tamaño de la muestra y el nº de categorías de las V influyen sobre sus valores.

La asociación entre variables no debe entenderse como una cuestión de todo o nada, sino como un continuo, que iría desde la ausencia de relación (independencia) al nivel máximo de relación entre las variables, que sería una relación determinista. Dado que x2 no resulta apropiado para evaluar el grado de relación entre variables, se han desarrollado va- rios índices que tratan de superar sus limitaciones. Aquí se verán algunos de ellos, que están basados en x2 y no tienen en cuenta si la relación es simétrica o no.
Estos coeficientes son índices globales del grado de intensidad de la relación, que si bien, tienen la ventaja de simplificar la información que proporcionan al resumir la tabla de contingencia en un único valor numé- rico, tienen la desventaja de no permitir ver el detalle de la relación entre las categorías de las variables (lo que sí se puede apreciar con el estudio de las distribuciones condicionadas ya vistas anteriormente).

Coeficiente C de Contingencia

Medida de asociación derivada de x^2 aplicable a tablas de contingencia de cualquier dimensionalidad (con indep. del nº de filas y columnas).

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

Puede asumir valores mayores o iguales a 0 y menores q 1.
Cuanto mayor es el valor de C, mayor es la relación; valores cercanos a 0 indican ausencia de relación.
C adopta el valor 0 cuando $x^2 = 0$ (si todas las frecuencias teóricas coinciden con las empíricas). Para adoptar el valor 1 el nº de observaciones (n) tendría q ser = 0, motivo por el q nunca llega a ese valor.

Especialmente útil cuando el nº de filas y de columnas coinciden pq se puede precisar más su valor máximo, lo q permite una interpretación mejor con la siguiente fórmula:

$$C_{max} = \sqrt{\frac{k-1}{k}}$$

k = nº de filas = nº de columnas

WUOLAH
Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Coeficiente V de Cramer

Modificación de x^2 q alcanza un valor máximo de 1 en caso de máxima asociación o asociación perfecta y un valor mínimo de 0 en una situación de independencia perfecta.

$$V = \sqrt{\frac{\chi^2}{n(m-1)}}$$

m = valor más pequeño entre el nº de filas y el nº de columnas

Es poco frecuente encontrar valores próximos a 1, de hecho pocas veces se alcanza 0,6. En términos empíricos se puede considerar al 0,6 prácticamente como un valor máximo habitual, por lo q 0,3, antes q considerarlo como bajo, es + bien un valor intermedio.

Coeficiente ϕ

Medida de asociación derivada de x^2 q se aplica a V dicotómicas (con tablas de contingencia 2x2). Al utilizarse en V q solo pueden adoptar 2 valores, la tabla de contingencia general queda reducida.

		Y		
		0	1	
X	0	n_{11}	n_{12}	n_{1+}
	1	n_{21}	n_{22}	n_{2+}
		n_{+1}	n_{+2}	n

ϕ puede adoptar valores entre -1 y 1:
- Positivo: el producto de n_{11} x n_{22} es mayor q el de n_{12} x n_{21}.
- Negativo: caso contrario.

$$\varphi = \frac{n_{11} \times n_{22} - n_{12} \times n_{21}}{\sqrt{n_{1+} \times n_{2+} \times n_{+1} \times n_{+2}}}$$

Así, para 2 V dicotómicas codificadas con 0 y 1, un valor positivo de ϕ indicará q los sujetos tienden a estar clasificados en 1 o en 0 en las dos variables; un coeficiente negativo quiere decir q la tendencia es a estar clasificado en 1 en una variable y en 0 en la otra. Esta fórmula es equivalente al coeficiente de correlación de Pearson aplicado a V dicotómicas. Se obtiene el mismo resultado (en valor absoluto) con la aplicación de V de Cramer.

Existe una variación de esta fórmula q puede aplicarse a V politómicas pero no tiene valor máximo, lo q dificulta su interpretación, motivo por el cual se desaconseja.

2. RELACIÓN ENTRE VARIABLES ORDINALES

Con estas V se pueden establecer relaciones de tipo mayor, menor, o igual. Sin embargo, no se pueden evaluar las distancias entre los distintos valores de la V *-en la V nivel socioeconómico, evaluada teniendo en cuenta 3 niveles (bajo, medio y alto), se puede afirmar q una persona con un nivel bajo tiene un nivel menor q una persona con un nivel medio, pero no se puede evaluar cuánto menor es-*.

Habitualmente, cuando se estudia la relación entre una V cualitativa y una ordinal se utilizan las mismas estrategias q en el estudio de 2 cualitativas. En el caso del estudio de 2 ordinales, la estrategia dependerá del nº de valores distintos q puedan adoptar:

- Si ambas adoptan un **nº reducido** de valores, se suelen utilizar tablas de contingencia, de manera similar a lo visto en V cualitativas. Cuando interesa estudiar la fuerza de la asociación, teniendo en cuenta el carácter ordinal, en lugar de los índices globales vistos, se utilizan otros desarrollados específicamente como la *d* de Sommers, o el coeficiente Gamma.
- Si alguna de las 2 V (o ambas) adoptan un **nº amplio** de valores, el estudio en tablas de contingencia deja de ser práctico por el elevado nº de filas y columnas. Se suele utilizar el coeficiente de correlación de Spearman o el coeficiente *tau-b* de Kendall.

Coeficiente de correlación por rangos de Spearman

Se basa en los rangos de los datos en lugar de hacerlo en los valores reales. Resulta apropiado en el caso de V ordinales o cuantitativas q no tengan una distribución normal.

Para calcularlo, 1º hay q ordenar todos los casos para cada una de las V y asignar un rango consecutivo a cada observación de cada por separado. Frecuentemente se producen empates o puntuaciones iguales = rangos empatados. En estos casos se asigna a las puntuaciones el rango promedio q ocuparían las observaciones empatadas.

Si la asociación fuera perfecta, esperaríamos q el rango q corresponde a cada caso de la variable X fuera exactamente igual al de la variable Y, por lo tanto el coeficiente se calcula en base a las diferencias registradas en los rangos entre ambas V, esperando q estas diferencias fueran 0. Conforme mayores son las dif. en las ordenaciones de ambas, + se aleja la relación de ser perfecta. Para evitar q las diferencias positivas anulen las negativas, el estadístico se calcula en función de la suma de las diferencias elevadas al cuadrado.

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

d_i = Rango(X_i) - Rango(Y_i)
n = n° de sujetos

Los valores oscilan de -1 a +1. El signo del coeficiente indica la dirección de la relación y el valor absoluto del coeficiente de correlación indica la fuerza de la relación.

- $r_s > 0$, existe relación directa. Cuánto + cercano a 1, más fuerte será la relación.
- $r_s < O$, relación inversa entre las V. Cuánto + se acerque a -1, más fuerte la relación.
- $r_s \approx 0$, apenas hay relación.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

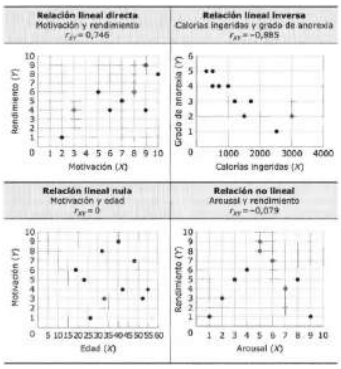
T5: RELACIÓN ENTRE VARIABLES II

1. RELACIÓN ENTRE VARIABLES CUANTITATIVAS

Las q están en un nivel de medida de intervalo o de razón. Poseen una unidad de medición común y constante. Su relación se estudiará mediante métodos gráficos y estadísticos.

Representación gráfica de la relación: el diagrama de dispersión

- O nube de puntos.
- Se utiliza con 2 V cuantitativas, ofreciendo una 1ª aproximación de la relación entre ambas.
- Para realizarlo se sitúa una de las V en el eje de abscisas y la otra en el de ordenadas. Para cada par de datos, se localiza la intersección de ambas V y se marca con un punto.
- Atendiendo, se puede observar q existe cierta relación lineal entre las V, correspondiendo, en mayor medida, calificaciones altas a mayor n° de h de estudio y viceversa. Hay, sin embargo, algunas excepciones *el estudiante 8, ha estudiado un n° de horas más bien alto (10) y ha obtenido un 1.*
- **Relación lineal directa**: los valores altos en Y tienden a emparejarse con valores altos en X, los valores intermedios en Y tienden a emparejarse con valores intermedios en X, y los valores bajos en Y tienden a emparejarse con valores bajos en X.
  Relación lineal inversa: los valores altos en Y tienden a emparejarse con valores bajos en X, los valores intermedios en Y tienden a emparejarse con valores intermedios en X, y los valores bajos en Y tienden a emparejarse con valores altos en X.
  Relación lineal nula: no hay un emparejamiento sistemático entre ellas en función de sus valores.
- Según la **Ley de Yerkes-Dodson**, la relación activación-rendimiento toma la forma de una U invertida. Para cada tipo de tarea se define un grado óptimo de activación en el cual el rendimiento para esa tarea es máximo. Por encima y por debajo de ese nivel óptimo, el rendimiento decrecerá tanto más cuanto más lejos se encuentre el nivel actual de activación del óptimo.



índices estadísticos para cuantificar la relación:

Covarianza

- Detecta la relación lineal entre X e Y; la variación conjunta de 2 V.
- Valor positivo = relación es directa.
  Valor negativo = relación inversa.
  Valor en torno a 0 = relación nula.
- Su valor absoluto será mayor cuanto + acusada sea la tendencia a la linealidad en el diagrama de dispersión. Se designa por S_{XY}, o Cov*(X, Y).*
- Su limitación, al igual q el coeficiente x^2, se desconocen los valores mín. y máx. q puede adoptar, lo q merma su capacidad para interpretar el grado de relación.

$$S_{XY} = Cov(X,Y) = \frac{\sum_{i=1}^{n} X_i Y_i}{n} - \bar{X}\,\bar{Y}$$

X_i = valor de la variable X en el caso i.
Y_i = valor de la variable Y en el caso i.
$\bar{X}$ =media de la variable X.
$\bar{Y}$ = media de la variable Y.

Coeficiente de correlación lineal de Pearson

- Índice q detecta la relación lineal entre X e Y, superando los límites de interpretación de la covarianza, al tener establecido un valor máximo (1) y mínimo (-1).
- Solo es apropiado para el estudio de las relaciones lineales.

$$r_{XY} = \frac{Cov(X,Y)}{S_X S_Y} = \frac{S_{XY}}{S_X S_Y}$$

$$r_{XY} = \frac{n\sum(XY) - \sum X \sum Y}{\sqrt{n\sum X^2 - (\sum X)^2}\sqrt{n\sum Y^2 - (\sum Y)^2}}$$

S_x = desviación típica de la variable X.
S_y = desviación típica de la variable Y.
S_{xy} = covarianza entre X e Y.

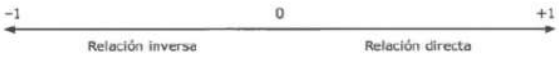
Atendiendo a la **1ª fórmula**;
El coeficiente es el cociente entre la covarianza entre X e Y y el producto de la desviación típica de X y la desviación típica de Y. Por tanto, será preferible, si ya se tienen alguno de los cálculos previos (la covarianza o las desviaciones típicas de X e Y). Cuando no haya ninguno, la fórmula 2 será + rápida.
En caso de utilizarla, hay q empezar calculando las medias y desviaciones típicas de X e Y, así como la covarianza. Una vez calculados ya se puede sustituir directamente.
Al utilizar la **fórmula 2ª**, se puede sustituir directamente, nada más añadir las columnas relativas a *XY*, X^2 e Y^2.

Obviamente el resultado es idéntico utilizando las 2.
Para interpretar los resultados hay q tener en cuenta :
- El valor absoluto. Cuanto mayor es, la relación tb es + fuerte.
- El signo del coeficiente de correlación de Pearson.
  + = valores mayores de la variable X tienden a corresponder a valores mayores de la variable Y, y a valores menores, a valores menores de la variable Y. Relación lineal directa.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

- = valores mayores de la variable X tienden a corresponder a valores menores de la variable Y, y a valores menores a valores mayores de la variable Y. Relación lineal inversa.
- El valor oscila siempre entre -1 y +1:
  $r_{xy} > 0$, relación lineal directa. Cuanto más se acerque a 1, más fuerte será. A mayores valores de la variable X nos encontraremos con valores altos de la variable Y y, a la inversa (bajos = bajos).
- $r_{xy} < 0$, relación lineal inversa. Cuanto más se acerque a -1, más fuerte será la relación, de forma q los valores altos de X se corresponderán con valores bajos de Y y viceversa.
- $r_{xy} \approx 0$, apenas hay relación lineal; una poco o nada tiene q ver con la otra.



Carac. del coeficiente de correlación lineal de Pearson:

- Índice simétrico, es igual la correlación de X con Y que la de Y con X ($r_{xy} = r_{yx}$).
- Los valores - 1 y 1 indican una correlación lineal perfecta y el valor 0, ausencia de correlación lineal.
- El valor absoluto no se ve afectado por transformaciones lineales de las V. Asimismo, $r_{xy} = \pm 1$, si una v es una transformación lineal de la otra.
- La correlación (por alta que sea) no implica q X sea la causa de Y, ni q Y sea la causa de X. Para poder hablar de causalidad se tienen que cumplir unos requisitos relativos al diseño de la investigación (q debe ser experimental).
- Puede verse afectado por 3ªs V *Si se mide la estatura y el razonamiento abstracto de los niños de Primaria, habrá una alta correlación; los + altos tendrán + edad. Si se limita a los niños de la misma edad, posiblemente desaparezca la correlación.*

Casos particulares del coeficiente de correlación lineal de Pearson

Hay varias fórmulas q se derivan y utilizan en algunos casos particulares, como el estudio de la relación entre 2 V ordinales, entre 2 V dicotómicas y entre 1 dicotómica y otra cuantitativa. Estas se desarrollaron pq su cálculo es más rápido, aunq con los programas informáticos existentes, no es un problema.

Relación entre variables ordinales

En el tema anterior, el coeficiente de correlación lineal de Spearman 2 V ordinales. Esta fórmula se deriva matemáticamente del coeficiente de correlación lineal de Pearson aplicado a rangos, por lo q su resultado es idéntico.
El único caso en el q no coinciden es en el de empates en los rangos, en cuyo caso hay q utilizar el coeficiente de correlación lineal de Pearson entre los rangos de las V. En caso de no haber empates se puede utilizar cualquiera teniendo en cuenta q el de Spearman simplifica bastante los cálculos.

Relación entre variables dicotómicas

La fórmula del coeficiente ϕ se deriva del coeficiente de correlación lineal de Pearson, por lo q el resultado de ambas es igual. Eso sí, el cálculo de ϕ se basa en la tabla de contingencia, por lo q es bastante + rápido q el de r_{xy} q precisa de las puntuaciones de cada sujeto en ambas V.

Relación entre una variable dicotómica y otra cuantitativa

En 1 V dicotómica los 2 valores se suelen representan por 0 y 1. El **coeficiente de correlación biserial puntual** se utiliza cuando una es dicotómica y la otra cuantitativa. Es muy utilizado en Psicometría . Se denota como r_{bp}.

$$r_{bp} = \frac{\overline{X}_p - \overline{X}_q}{S_x} \cdot \sqrt{p \cdot q}$$

$\overline{X}_p$: la media de las puntuaciones de la V cuantitativa X obtenidas por el grupo al q se le ha asignado un 1 en la V dicotómica.
$\overline{X}_q$: media de las puntuaciones de la V cuantitativa X obtenidas por el grupo al q se le ha asignado un 0 en la V dicotómica.
S_X: desviación típica de la V cuantitativa X.
***p*: proporción de casos del grupo al q se le ha asignado un 1 en la V dicotómica.**
***q*: proporción de casos del grupo al q se le ha asignado un 0 en la V dicotómica.**

Si la V en lugar de ser dicotómica es dicotomizada, el coeficiente a utilizar es el **coeficiente de correlación biserial**. 1 V se ha dicotomizado si hay un continuo latente entre las 2 categorías, pero se han establecido 2 únicos valores.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
WUOLAH Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

		Variable Y			
		Cualitativa dicotómica	**Cualitativa politómica**	**Ordinal**	**Cuantitativa**
Variable X	**Cualitativa dicotómica**	Estadístico x^2 Coef. C de Contingencia Coef. V de Cramer Coeficiente ϕ	Estadístico x^2 Coef. C de Contingencia Coef. V de Cramer	Estadístico x^2 Coef. C de Contingencia Coef. V de Cramer	Correlación biserial puntual
	Cualitativa politómica		Estadístico x^2 Coef. C de Contingencia Coef. V de Cramer	Estadístico x^2 Coef. C de Contingencia Coef. V de Cramer	
	Ordinal			Coef. de correlación de Spearman	Coef. de correlación de Spearman
	Cuantitativa				Covarianza Coef. de correlación de Pearson

2. REGRESIÓN LINEAL SIMPLE

Regresión - psicólogo inglés Sir Francis Galton (1822-1911): analizando la estatura de una muestra numerosa de padres e hijos, advirtió q los hijos de padres altos eran tb, en general, superiores en estatura al promedio, pero no tan altos como sus progenitores. Algo parecido ocurría con los hijos de padres bajos q, aún siendo + bajos q el promedio, no eran por término general, tan bajos como ellos. Dedujo q se producía en los hijos una **regresión a la media de la distribución**, q los valores tendían a regresar a la media de la V. Buscó una ecuación matemática para estimar los valores q adoptarían en una V sujetos para los q se conoce sus puntuaciones en otra V y la relación entre ambas. Esta ecuación resultaría ser la ecuación de una recta.

El modelo de regresión utiliza la inf. contenida en las relaciones lineales observadas entre las V. Si 2 variables X e Y se relacionan linealmente, la representación gráfica de su distribución conjunta se aproximará visualmente bastante a una línea recta = poder escribir una V en función de la otra con la ecuación de una recta: Y= *a* + bX.

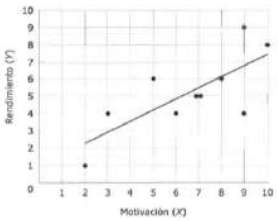
Correlación y regresión son 2 conceptos muy cercanos:

- **Regresión**, el interés se centra en predecir los valores de una V (Y) a partir de los valores conocidos en otra (X), más q en la variación conjunta de las 2. Relación asimétrica o direccional, los resultados serán distintos según se trate de la ecuación de regresión de Y sobre X, o de la de X sobre Y (q predice los valores de X a partir de Y).
- **Correlación**: la relación entre es simétrica. La forma de proceder implica 3 fases:
  1. Identificación del modelo de regresión = obtener los coeficientes de regresión q le caracterizan;
  2. Valoración del modelo = estudio de la capacidad predictiva del mismo;
  3. Aplicación del modelo para predecir V.

Cálculo de los coeficientes de regresión

Desde el punto de vista geométrico, la recta de regresión tiene la misma interpretación q cualquier otra recta, pero desde el estadístico, se ajusta a la nube de puntos del diagrama de dispersión con menos error.

- *a*: constante q se denomina **origen**, corresponde al valor q adopta la variable Y cuando la X vale 0. Indica el origen, el punto en el q la recta corta al eje de ordenadas.
- *b*: constante q se denomina **pendiente**, de ella depende la inclinación de la recta. Indica en q medida cambian los valores de Y por cada incremento de una unidad en los valores de X.
- *Y*: V cuyo valor se desea conocer y va a ser pronosticado a partir del valor de la variable X. Se suele denominar **V pronosticada o criterio** y denotarla como Y'.
- *X*: V cuyo valor se conoce y va a ser utilizado para pronosticar el valor del criterio. Se suele denominar **V predictora** o simplemente **predictor**.



Para calcular los valores de a y de b, conociendo los valores de X e Y, se utilizan:

$$b = \frac{n\sum(XY) - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} \qquad b = r_{xy}\frac{S_Y}{S_X} \qquad a = \overline{Y} - b\overline{X}$$

Se han presentado 2 fórmulas para calcular b. La 1ª es más rápida cuando tenemos los datos directos y la 2ª cuando ya tenemos realizados algunos cálculos previos. Conocido el valor de las dos constantes a y b de la ecuación anterior, tenemos ya completamente formulado el modelo de regresión. Habitualmente se suele poner Y' en lugar de Y para denotar q nos estamos refiriendo a los valores pronosticados en el criterio, no a los valores reales obtenidos por los sujetos.

$$Y'_i = a + bX_i$$

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Además, hay q tener en cuenta q, dado q es una predicción hay cierto nivel de error. De no haberlo, todos los puntos del diagrama de dispersión se encontrarían sobre la recta. Para cada uno de los sujetos se comete cierta cantidad de error al asignarle la puntuación pronosticada Y' en lugar de la Y. En este sentido, a la ecuación anterior habría q añadirle un término q reflejase este error:

$$Y_i = a + bX_i + E_i$$

Donde: $E_i = Y_i - Y_i'$

E_i es una medida del error individual cometido para cada una de las observaciones. Al utilizar un modelo de regresión se utiliza el modelo lineal con el q se comete un error lo más pequeño posible para todos los sujetos. Con el **criterio de mínimos cuadrados**, procedimiento q proporciona valores tales q la suma de los errores al cuadrado (SCE) para los *n* participantes sea mínimo. Formalmente, se establece obteniendo los valores *a* y *b* q minimizan la siguiente expresión:

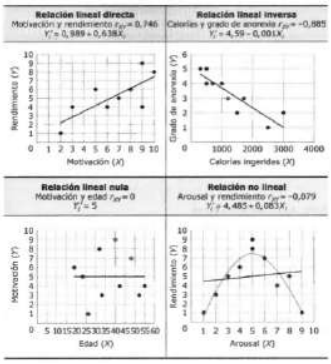
$$SCE = \sum (Y_i - Y_i')^2 = \sum [Y_i - (a + bX_i)]^2$$

Se puede demostrar (derivando parcialmente la función a minimizar respecto de cada uno de los parámetros, igualando a 0 y despejando) q este criterio proporciona las ecuaciones para estimar a y b con el menor error posible, consiguiendo la recta q mejor ajusta a la nube de puntos.

Valoración del modelo

¿Hasta q punto es un buen modelo para predecir la V criterio?

- Aunq se consiga el mejor ajuste posible a los datos disponibles, no es garantía de q ese sea óptimo para predecir la V criterio. Como es obvio, cuanta mayor relación, mejor pronóstico.
- Antes de realizar cálculos numéricos, lo ideal para valorar el ajuste es representar los datos mediante un diagrama de dispersión, para tener una 1ª aproximación de la posible relación.
- La línea continua representa la ecuación de regresión lineal q mejor ajusta a la nube de puntos. En algunos casos, este modelo lineal para predecir Y puede ser una mala opción.
- Para examinar la utilidad predictiva, además de la aproximación gráfica, se pueden utilizar 2 índices: varianza error y coeficiente de determinación.



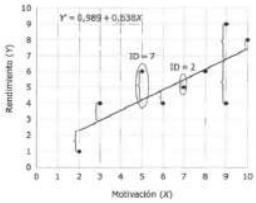
La varianza error

La varianza de los errores (o residuos) cometidos al pronosticar la variable Y a partir de X; la diferencia entre la puntuación q realmente obtendría y la q se le ha pronosticado .

$$E_i = Y_i + Y'_i$$

Se calcula aplicando la fórmula de la varianza a estas puntuaciones error, **error cuadrático medio**. Se puede denotar como S^2_E o S^2_{yx} y se interpreta como la varianza de los errores cometidos al pronosticar la variable Y a partir de X (o mediante la recta de regresión de Y sobre X).

$$S_E^2 = S_{Y \cdot X}^2 = \frac{\sum E_i^2}{n} - \bar{E}^2$$



Se ha marcado con una llave la distancia entre cada uno de los valores asumidos por la variable Y' representada en la recta de regresión (pronosticada) e Y (observada). Estas «distancias» son los errores cometidos al pronosticar el rendimiento en la asignatura (Y) a partir de los valores en motivación (X).
Se aprecian los errores cometidos. El estudiante n° 7 obtuvo en rendimiento académico (Y) 6, mientras la línea de la recta de regresión le pronostica una puntuación menor. El 2º estudiante el error es menor, hay menor distancia entre la puntuación obtenida en Y (5) y la q le pronosticaríamos, cercana a 5,5.

Cuanto menor valor de la varianza error, + similares serán las puntuaciones pronosticadas por el modelo y las q realmente obtendrían los sujetos; mejores predicciones por el modelo de regresión.

El coeficiente de determinación

Igual al coeficiente de correlación de Pearson elevado al cuadrado. Indica la proporción de varianza de la V pronosticada o criterio (Y) explicada por el modelo lineal; por la V predictora X.

r^2_{XY} → Coeficiente de determinación

No depende de las unidades en q se expresan los datos y toma valores entre 0 y 1. ↑ valor = + similares las puntuaciones pronosticadas y las q realmente obtendrían.

- 0 = la V predictora tiene nula capacidad predictiva.
- 1 = la V predictora explicaría toda la variación de Y, y las predicciones no tendrían error.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Características del modelo de regresión

- La pendiente siempre será del mismo signo q el coeficiente de correlación lineal de Pearson, informará sobre el tipo de relación lineal entre las V (directa o inversa). Dado q las desviaciones típicas siempre son positivas, *b* adopta el signo del coeficiente de correlación lineal de Pearson.

$$b = r_{XY}\frac{S_Y}{S_X}$$

- La media de los errores de predicción o residuos (E= Y- Y') es O. $\bar{E} = 0$
- La media de las puntuaciones pronosticadas coincide con la media de las verdaderas puntuaciones en Y. $\bar{Y}' = \bar{Y}$
- La varianza de las puntuaciones en Y = suma de la varianza de los pronósticos (hechos mediante la recta de regresión) + la varianza de los errores (o error cuadrático medio).

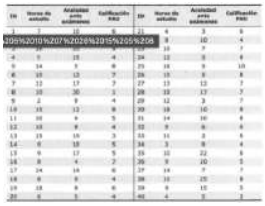
$$S_Y^2 = S_{Y'}^2 + S_{Y.X}^2$$

- El coeficiente de determinación = cociente entre la varianza de las puntuaciones pronosticadas y la varianza de las puntuaciones en Y. De ahí, q sea un indicador de la proporción de varianza del criterio q queda explicada con el modelo de regresión lineal. $r_{XY}^2 = \frac{S_{Y'}^2}{S_Y^2}$
- El complementario del coeficiente de determinación = cociente entre la varianza de los errores y la varianza de las puntuaciones en Y, e indica la proporción de la varianza del criterio q NO queda explicada por el modelo de regresión lineal. $1 - r_{XY}^2 = \frac{S_{Y.X}^2}{S_Y^2}$

3. REGRESIÓN LINEAL MÚLTIPLE

Si se utiliza más de una V predictora, la capacidad predictiva del modelo puede mejorar. Suelen ser + realistas, es raro encontrar criterios q se puedan predecir a partir de una única V predictora.

Dado q se suele realizar mediante software estadístico, no se expondrán las fórmulas. Simplemente se tratará de explicar con un ej., el cambio q se produce en la valoración del modelo a partir del coeficiente de determinación, cuando se introduce una 2ª V predictora.



En la tabla se muestran 3 V, cuantitativas, y tiene sentido pensar q tanto nº de horas de estudio semanales como el nivel de ansiedad presentado en un examen pueden influir (o predecir) la calificación obtenida.
Utilizando la regresión lineal simple, hay q considerar 2 rectas de regresión, 1 para la V predictora n° de horas de estudio y otra para el nivel de ansiedad:
1- la ecuación de la recta de regresión de la Calificación obtenida sobre el n° de h de estudio: Y'_i = 2,82 + 0,292X_i 0,584, con lo q el coeficiente de determinación = ese valor al cuadrado, 0,341. El 34,1% de la variabilidad se explica por el tiempo empleado para estudiar.
2- la ecuación de la recta de regresión de la calificación obtenida sobre el nivel de ansiedad es: Y'_i =8,36- O,239X Y_i = -0,793, con lo q el coeficiente de determinación será 0,629. 62,9% de la variabilidad de la calificación se explica por el nivel de ansiedad.

Al utilizar las 2 V predictoras (X_1 al n° de ho- ras estudiadas y X2 al nivel de ansiedad ante los exámenes), la ecuación de regresión sería: Yi = 5,714 + O, 226X1 - O, 214X2
En este caso, al haber 2 V predictoras implicadas, el coeficiente de determinación varía. Se denota como $R^2_{Y\cdot X1\cdot X2}$:

$$R^2_{Y.X_1X_2} = \frac{r^2_{YX_1} + r^2_{YX_2} - 2\cdot r_{YX_1}\cdot r_{YX_2}\cdot r_{X_1X_2}}{1 - r^2_{X_1X_2}}$$

Para facilitar la utilización se presentarán los datos en una **matriz de correlaciones**; una tabla con el mismo nº de filas y columnas q de V, en la q en cada casilla aparece las correlaciónes correspondientes a la fila y a la columna.

El valor predictivo es muy superior al de ambos modelos de regresión simples. Teniendo en cuenta ambas predictoras simultáneamente se explica el 82,7% de la variabilidad de la calificación, a partir del tiempo de estudio y del nivel de ansiedad ante los exámenes.

	Horas de estudio (X_1)	Ansiedad ante los exámenes (X_2)	Calificación PAU (Y)
Horas de estudio (X_1)	1	−0.185	0,584
Ansiedad ante los exámenes (X_2)		1	−0,793
Calificación PAU (Y)			1

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

T6: NOCIONES BÁSICAS DE PROBABILIDAD

En las CCSS en general, y en la Psicología y Ciencias de la Salud en particular, es habitual la imposibilidad de prever el resultado de un fenómeno. Según los casos, esto puede ser debido a diversas causas.

- Q una persona tenga fiebre o le duela la cabeza (variaciones de la homeostasis interna) el día en el q participa en un exp. sobre estrés. Su situación y R serán diferentes a las q daría.
- Q estemos empleando un instrumento de medida afectado por condiciones medioambientales, dando lugar a resultados de medida dif.

Estamos expuestos continuamente a sucesos sobre los q no tenemos la certeza de que vayan a ocurrir; pueden tener una mayor o menor probabilidad. Por tanto, puede haber un rango amplio de probabilidades. El resultado está influenciado por el azar o, de forma más correcta, estamos ante un fenómeno aleatorio. Ante estas variaciones = gran incertidumbre en los resultados, ¿cómo se puede actuar desde la perspectiva metodológica? Existen herramientas metodológicas q pueden trabajar con este tipo de datos: «La Estadística permite esbozar conclusiones válidas en situaciones de incertidumbre y variabilidad» y de la **Probabilidad**, ya q ésta es la teoría matemática q desarrolla modelos matemáticos adaptados al estudio de estas situaciones, mediante la asignación de probabilidades.

Estadística Inferencial: conjunto de métodos y técnicas q permiten inducir, a partir de la inf. empírica proporcionada por una muestra, el comportamiento de una det. población, con un riesgo de error medible en términos de probabilidad. Probabilidad y Estadística son ramas de las matemáticas q se complementan. El objeto de la Probabilidad es el estudio de **V aleatorias**, valores q dependen básicamente del azar o de la posibilidad de q puedan o no ocurrir. La Estadística, por su parte, es otra rama de las matemáticas cuyo objeto de estudio son los datos, entendidos como valores o atributos de los objetos de estudio de interés. Son disciplinas íntimamente relacionadas, ambas se refieren al estudio de un mismo tipo de situaciones, en las q hay incertidumbre. La Probabilidad aporta los modelos matemáticos (las distribuciones) para el estudio de la incertidumbre, y la Estadística los adapta a los datos reales (con incertidumbre).

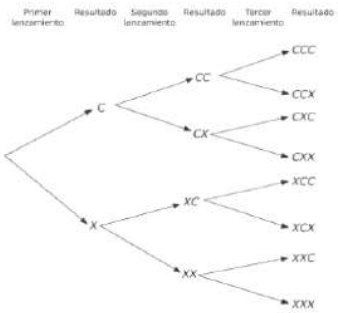
1. CONCEPTOS PREVIOS

Experimento aleatorio: proceso q se puede repetir indefinidamente en las mismas condiciones y cuyo resultado no se puede predecir con certeza. Conduce a la obtención de un resultado donde interviene el azar. 3 carac.:

- Todos los resultados posibles son conocidos con anterioridad a su realización. Tiene un conjunto definido.
- No se puede predecir con certeza el resultado concreto del exp.
- El exp. puede repetirse teóricamente un nº infinito de veces en idénticas condiciones.

Espacio muestral / espacio de resultados: conjunto de todos los resultados posibles de un exp. aleatorio. Habitualmente se representa por la letra mayúscula E.

- *Lanzar una moneda; C = cara y X = cruz. Espacio muestral E= {C; X}*
- *Lanzar un dado; E={1·2·3·4·5·6} • IIIII*
- *Lanzar 2 veces un dado: E= {(1,1); (1,2); (1,3); (1 ,4); (1,5); (1,6); (2,1); (2,2); (2,3); (2,4); (2,5); (2,6); (3,1); (3,2); (3,3); (3,4); (3,5); (3,6); (4,1); (4,2); (4,3); (4,4); (4,5); (4,6); (5,1); (5,2); (5,3); (5,4); (5,5); (5,6); (6,1); (6,2); (6,3); (6,4); (6,5); (6,6)}*
- *Introducir 3 ratas en un laberinto en forma de T (I = va hacia la izquierda; D = hacia la derecha): E = {(I,I,I); (I,I,D); (I,D,I); (D,I,I); (D,D,I); (D,I,D); (I,D,D); (D,D,D)}*



Una forma sistemática y didáctica de construirlos es mediante el **diagrama de árbol**, representación gráfica q muestra los resultados posibles de un exp. aleatorio.
En el caso de q tuviera solo 1 elemento no podríamos hablar de experimento aleatorio pq se puede predecir con certeza el resultado obtenido.

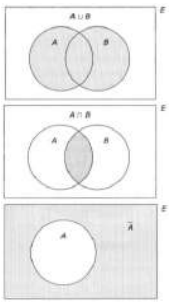
Sucesos: los resultados de un exp. aleatorio, o subconjuntos del espacio muestral. Se representan por letras mayúsculas: *A, B....* A su vez, pueden ser *elementales* o com*puestos.*

- **Suceso elemental**, suceso simple o punto muestra: cada uno de los resultados posibles del exp. aleatorio. 1 solo elemento del espacio muestral E.
- **Suceso compuesto:** 2 o + sucesos elementales.
- **Suceso seguro**: ocurre siempre. Consta de todos los sucesos elementales del espacio muestral y se identifica con el espacio muestral total E.
- **Suceso posible**: puede contener algún elemento del espacio muestral E = tiene posibilidad de q ocurra.
- **Suceso imposible**: no ocurre nunca. No contiene ningún elemento del espacio muestral y se identifica con 0 o conjunto vacío.

Operaciones con sucesos:
Entre los sucesos se establecen las mismas operaciones q en la **teoría de conjuntos**, ya que las operaciones son, en sí mismas, entre conjuntos teniendo en cuenta su definición.
Los **diagramas de Venn** se emplean para representarlos y estudiar visualmente propiedades y operaciones entre ellos.
El espacio muestral o de resultados E se representa mediante un rectángulo, y dentro de él se incluyen los sucesos mediante círculos.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material



• **Unión**: Subconjunto de E formado por los sucesos elementales que pertenecen a A, a B, o a ambos. Representado por A *u* B.
• **Intersección**: Representado por por A *n* B, al subconjunto de E formado solamente por los sucesos elementales q pertenecen a ambos simultáneamente. Cuando la intersección de 2 sucesos no contiene ningún elemento son **sucesos incompatibles o excluyentes**, q no pueden verificarse a la vez.
• **Complementario**: Representado por A, al subconjunto de E formado por los sucesos elementales q no pertenecen a A.

Las operaciones de unión e intersección pueden extenderse al caso de 2 o + sucesos *A u B u C*, y puede hablarse de complementario de la unión o de la intersección de 2 sucesos.

2. DEFINICIÓN DE PROBABILIDAD

En teoría de la probabilidad se toman todos los posibles resultados de un exp. aleatorio como elementos del espacio muestral *E* (espacio de resultados). Si E contiene un número finito de elementos, a cada uno se le puede asociar un nº no negativo, q es su probabilidad de ocurrencia, tal q la suma de todos los nº correspondientes a todos los elementos de *E* sea 1.

La **probabilidad de un suceso** es una medida numérica e cuantifica la posibilidad de q este ocurra. Los valores se encuentran comprendidos entre 0 y 1.
- Muy probables estarán próximos a 1 y menos, al 0.
- El valor 0 se asigna a los sucesos imposibles y 1 para los seguros.

En función del enfoque desde el q se estudie se ha propuesto un tipo dif. de definición. Cada una de ellas tiene una operativa distinta, pero todas un mismo objetivo: calcular la posibilidad de ocurrencia.

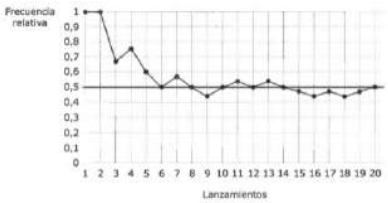
1. La **definición clásica** o a **priori**; ***Regla de Laplace***; la probabilidad de un suceso *A* es igual al cociente entre el nº de casos favorables de q ocurra y el nº de casos posibles en el supuesto de q todos los casos tengan la misma oportunidad de ocurrir (sean equiprobables).

$$P(A) = \frac{n_A}{n}$$

P(A) = Probabilidad de un suceso *A*
n_A = Nº de casos favorables
***n* = Nº de casos posibles**

A priori pq antes de realizar el exp. aleatorio se conocen de antemano los posibles resultados del espacio muestral E y sus probabilidades. Esta plantea algunos problemas a la h de aplicación; parte de la base de q los sucesos son **equiprobables** (tienen la misma probabilidad de ocurrencia).

Nº de lanzamientos	Resultado	Nº de caras	Frecuencia relativa
1	C	1	1/1 = 1
2	C	2	2/2 = 1
3	X	2	2/3 = 0,6666
4	C	3	3/4 = 0,7500
5	X	3	3/5 = 0,6000
6	X	3	3/6 = 0,5000
7	C	4	4/7 = 0,5714
8	X	4	4/8 = 0,5000
9	X	4	4/9 = 0,4444
10	C	5	5/10 = 0,5000
11	C	6	6/11 = 0,5454
12	X	6	6/12 = 0,5000
13	C	7	7/13 = 0,5384
14	X	7	7/14 = 0,5000
15	X	7	7/15 = 0,4666
16	X	7	7/16 = 0,4375
17	C	8	8/17 = 0,4706
18	X	8	8/18 = 0,4444
19	C	9	9/19 = 0,4736
20	C	10	10/20 = 0,5000



La probabilidad de salir cara en el lanzamiento de una moneda es de 1 caso favorable dividido por 2 casos posibles; 1/2. Supongamos q se realiza y anota si sale cara o cruz en cada tirada, así como la frecuencia relativa en cada caso. Los resultados podrían ser los q se presentan. Si se representan gráficamente, según ↑ el nº de lanzamientos, la línea quebrada q une las frecuencias se ajusta + a la horizontal trazada en la ordenada 1/2 (0,5) o valor teórico de la probabilidad definida por Laplace. Por tanto, la frecuencia relativa tiende a estabilizarse cuando el nº es muy elevado. **Ley del azar o ley de regularidad estadística**: fenómeno de estabilización de las frecuencias.

No obstante, no siempre es fácil aplicar este concepto, muchas veces no es posible repetir un experimento aleatorio un gran nº de veces, y si lo es, no es práctico.

* _____

Andréi Nicoláyevich Kolmogórov - **teoría axiomática de la probabilidad**: aplicación de la teoría de conjuntos a los sucesos q componen el espacio muestral. 2 ventajas:
- Recoge las definiciones de probabilidad anteriores; cumplen la axiomática.
- Permite el desarrollo matemático de la teoría de la probabilidad.

Definición axiomática de probabilidad: Dado un espacio muestral *E*, se denomina probabilidad de un suceso A_i, definido en el espacio muestral E y designado por $P(A_i)$, a un nº real asignado al suceso A_i, q cumple las sig. propiedades:

- $0 \le P(A_i) \le 1$
- $P(E) = 1$
- Si $A_1, A_2 ... A_k$ son sucesos incompatibles dos a dos, entonces:

$$P(A_1 \cup A_2 \cup ..... \cup A_k) = P(A_1) + P(A_2) + ... + P(A_k)$$

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
WUOLAH
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

$$P(A) = \frac{n_A}{n} = \frac{0}{6} = 0$$

$$P(A) = \frac{n_A}{n} = \frac{n}{n} = P(E) = 1$$

Las 2 primeras indican q la probabilidad es cuantificable numéricamente con un nº entre 0 y 1.

- Se asigna un 0 a un suceso imposible A:
- Se otorga un 1 a un suceso seguro A, q se corresponde con el espacio muestral E.

La 3ª propiedad indica q la unión de sucesos incompatibles (aquellos cuya intersección es el conjunto vacío: $A_i \cap A_j = \emptyset$) es igual a la suma de sus probabilidades individuales.

De estas propiedades se deriva q P(A) + P(Ā) = 1 siendo ambos sucesos exhaustivos y excluyentes (si no ocurre A, necesariamente lo hará Ā). Esto implica q P(Ā) = 1 - P(A), o lo q es lo mismo, la probabilidad del suceso complementario de A, (Ā), es igual a 1 menos la probabilidad de ocurrencia de A. A partir de la definición axiomática se deducen una serie de teoremas, de los cuales se van a examinar el de la suma (concepto de unión de sucesos) y el del producto (concepto de intersección).

2. La **definición estadística** o **a posteriori**; no asume la equiprobabilidad. Se basa en la estabilidad de las frecuencias relativas cuando el nº de repeticiones de un suceso aleatorio es muy elevado y tiende a infinito.
   Lanzamos un dado al aire muchas veces, y anotamos las frecuencias relativas de un suceso. Estas tienden a estabilizarse en un valor constante, entre 0 y 1 = probabilidad del suceso.
   P(A) o probabilidad de un suceso *A*: límite al q tiende la frecuencia relativa de aparición de un suceso *A* cuando el nº de ensayos *n* o repeticiones tiende a infinito.

$$\boxed{P(A) = \lim_{n \to +\infty} \frac{n_A}{n}}$$

   A posteriori pq las probabilidades se atribuyen a cada suceso después de un gran nº de repeticiones del exp. aleatorio.

3. TEOREMA DE LA SUMA

Partiendo del axioma 3, referido a la unión de sucesos, establece q la probabilidad de q ocurra el suceso A o el B es igual a la probabilidad de q ocurran A + la de q ocurra B - la de q ocurran A y B (intersección de ambos sucesos).

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

Si son mutuamente excluyentes (simultáneamente) o complementarios (ocurrencia de uno = no del otro), la regla se simplifica, resultando ser la suma de las probabilidades de cada suceso dado q $P(A \cap 8) = 0$

$$\boxed{P(A \cup B) = P(A) + P(B)}$$

4. PROBABILIDAD CONDICIONADA

Hasta ahora hemos estado tratando sucesos indepe., donde la probabilidad de uno no altera la del otro. Sin embargo, no siempre son tan simples; la aparición de u A puede depender de la aparición de B. En estos casos son dependientes, la probabilidad de A depende o está condicionada al suceso B: P (A|B)

Para dos sucesos cualesquiera A y B, la **probabilidad de A condicionada a B** (o de A supuesto B) es igual a la probabilidad de la intersección dividida por la probabilidad de B, siempre q P(B) ≠ 0

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}}$$

Siempre q P(A) ≠ 0

$$\boxed{P(B|A) = \frac{P(B \cap A)}{P(A)}}$$

Si los sucesos son indep.:

$$\boxed{P(A|B) = P(A) \quad y \quad P(B|A) = P(B)}$$

5. TEOREMA DEL PRODUCTO

Situaciones en las q se quiere calcular la probabilidad de q aparezcan 2 sucesos simultáneamente; calcular la probabilidad de intersección entre 2 sucesos.
Se ha visto qe la probabilidad condicionada se define: $P(B|A) = \frac{P(A \cap B)}{P(A)}$

Despejando $P(A \cap B)$: $P(A \cap B) = P(A) \cdot P(B|A)$ q corresponde al teorema del producto.

WUOLAH
Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Teorema del producto: probabilidad de ocurrencia de A y B es igual a la probabilidad de ocurrencia de A por la probabilidad de ocurrencia de B, dado q A ha ocurrido previamente (condición previa).

$$P(A \cap B) = P(A) \cdot P(B|A)$$

P(B|A): probabilidad de q ocurra B dado q ha ocurrido A.

Cuando son independientes P(B|A) = P(B): $P(A \cap B) = P(A) \cdot P(B)$

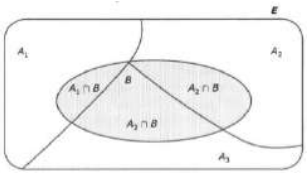
Cuando se extraen bolas o papeletas de una urna. Cuando se realiza +1 extracción, la probabilidad de q ocurra B dado q ha ocurrido A va a verse afectada por el hecho de q el elemento extraído vuelva a reponerse o no. **Extracción con reposición**; se mantiene siempre el mismo nº de bolas o papeletas; se devuelven a la urna. **Extracción sin reposición**: no se devuelven, las probabilidades de obtener una bola o papeleta concreta en esta 2ª extracción van a depender de lo obtenido en la 1ª.

6. TEOREMA DE LA PROBABILIDAD TOTAL

En un espacio muestral E, se dice q k sucesos A_1, A_2… A_k forman una **partición del espacio muestral** si se cumplen simultáneamente las siguientes condiciones:

- $A_i \cap A_j = \emptyset$ para cualquier par de sucesos A_i y A_j; son incompatibles y su intersección es el conjunto vacío.
- $A_1 \cup A_2 \cup \ldots \cup A_k = E$; la unión de todos los sucesos es igual al espacio muestral (son exhaustivos). En términos de probabilidad se cumple q: $P(A_1 \cup A_2 \cup \ldots \cup A_k) = P(A_1) + P(A_2) + \ldots + P(A_K) = 1$

Si sobre ese mismo espacio muestral se define un nuevo suceso *B*, el teorema de la probabilidad total permite calcular su probabilidad a partir de las $P(A_i)$ y de las $P(B|A_i)$.



Partición del espacio muestral en 3 sucesos A_1, A_2 y A_3 incompatibles entre sí, y otro suceso B en el mismo espacio muestral E. Para este caso concreto, la probabilidad de B se corresponde con la suma de las intersecciones de cada uno de los sucesos A_i con el suceso B:

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$$

Aplicando el teorema del producto:

$$P(B) = P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + P(A_3) \cdot P(B|A_3)$$

De manera general, el teorema de la probabilidad total, Proporciona el valor de la probabilidad de B en función de la probabilidad de los sucesos A; y de las probabilidades de B condicionadas a los A;

$$P(B) = \sum_{i=1}^{k} P(A_i) \cdot P(B|A_i)$$

.

- Otra forma de calcularlas es el **diagrama de árbol**. A la izq, las probabilidades de los sucesos A_1, A_2 y A_3 . Las probabilidades + a la derecha son las de los sucesos B y su complementario B condicionadas a los sucesos A_1, A_2 y A_3, respectivamente. Siempre la suma de las probabilidades del mismo debe ser 1. *$P(A_1)+P(A_2)+P(A_3)=1$; $P(B|A_1)+P(\bar{B}|A_1) = 1$*
  Las probabilidades condicionadas vienen dadas (derecha del diagrama) y las de intersección $P(A_1 \cap B)$, $P(A_2 \cap B)$ y $P(A_3 \cap B)$ se pueden calcular multiplicando las probabilidades de cada rama. Así, la probabilidad del suceso B se determina como:

$$P(B) = P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + P(A_3) \cdot P(B|A_3)$$

7. TEOREMA DE BAYES

En sucesos dependientes, la probabilidad condicionada tiene en cuenta inf. de un suceso para conocer la probabilidad de otro. El teorema de Bayes va más allá y permite calcular cómo se modifican las probabilidades de det. sucesos cuando se conoce alguna inf. adicional. El teorema de la probabilidad total permitía obtener la probabilidad de un suceso B y este, las probabilidades condicionadas de los sucesos A_i dado el suceso B.

Teniendo en cuenta la formula de probabilidad condicionada, se aplica el teorema del producto al numerador:

$$P(A_i|B) = \frac{P(A_i) \cdot P(B|A_i)}{P(B)}$$

Y aplicando el teorema de la probabilidad total al denominador se obtiene el teorema de Bayes:

$$P(A_i|B) = \frac{P(A_i) \cdot P(B|A_i)}{\sum_{i=1}^{k} P(A_i) \cdot P(B|A_i)}$$

$P(A_i)$: probabilidades de cada suceso A_i
$P(B|A_i)$: probabilidad del suceso B condicionada a los sucesos A_i

$$\sum_{i=1}^{k} P(A_i) - P(B|A_i) = P(B)$$

El teorema proporciona las probabilidades de A_i condicionadas por B en función de las probabilidades de los sucesos A_i y de las probabilidades de B condicionadas por dichos sucesos.

WUOLAH
Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

8. APLICACIONES DE LA PROBABILIDAD CONDICIONADA EN PSICOLOGÍA DE LA SALUD

- Epidemiología --> bastante frecuencia. Una situación habitual podría ser una decisión como diagnosticar a un paciente q presenta unos síntomas (suceso X), y q puede tener una enfermedad (suceso E). *Determinar la probabilidad de tener un infarto (E) si antes se ha tenido un dolor agudo del brazo derecho (síntoma).*

  Prevalencia: proporción de casos existentes de una enfermedad en un momento det. Indica la probabilidad de personas q tienen una carac. o enfermedad en relación a la población.
  Incidencia: proporción de casos nuevos de una enfermedad en una población durante un período det. Probabilidad de personas nuevas q pueden tener una carac. o enfermedad en un periodo concreto.

  Existe una relación entre ambos; si los casos nuevos (incidentes) no se resuelven, se hacen crónicos (prevalentes). Además, una ↓ en la incidencia = ↓ prevalencia y al revés.

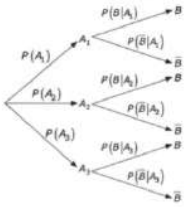
- Análisis de factores de riesgo o probabilidad de q ↑ un problema o enfermedad al estar expuesto a un riesgo—> parte de la base de q sujetos expuestos a un factor (X+) tienen más posibilidades de sufrir una enfermedad o tener un problema psicológico (E+) en comparación con el grupo no expuesto (E-) a dicho factor (X-).

- Valoración de la calidad de las pruebas diagnósticas. Tenemos una prueba para la evaluación diagnóstica de un trastorno (T), q nos va a permitir distinguir a las personas sanas o sin trastornos (NT) de las q lo tienen en función de un punto de corte establecido previamente. Se supone q la prueba dispone de 2 indicadores: (+) indica q tiene el trastorno y (-) q está sana y no tiene el trastorno.
  Los datos se presentan en una tabla de doble entrada:

  Una buena prueba diagnóstica presenta:
  - **Alta sensibilidad** P (+|T) o probabilidad de discriminar a los verdaderos positivos; la probabilidad de q los que tengan un trastorno den positivo (+).
  - **Alta especificidad** P (-|NT) o probabilidad de detectar a los verdaderos negativos; probabilidad de los q no tienen trastorno den negativo (-).

	Resultado de la prueba	
	+	-
Trastorno (T)	Verdaderos positivos	Falsos negativos
No Trastorno (NT)	Falsos positivos	Verdaderos negativos

  Sin embargo, las pruebas nunca son exactas y siempre se trabaja con un margen de error en la evaluación. A consecuencia, encontramos 2 tipos de valores predictivos:
  - **Valor predictivo positivo** P(T|+) o probabilidad de q todos los q den positivo tengan el trastorno. Pero hay sujetos q dan (+) y no lo tienen (falsos positivos), por lo q el valor predictivo positivo disminuye.
  - **Valor predictivo negativo** P(NT|-) o probabilidad de q todos los (-) estén sanos y no tengan trastorno. Sin embargo, hay sujetos q dan (-) y tienen el trastorno (falsos negativos) = valor predictivo negativo reducido.



De esta forma, si se determina el grado de sensibilidad, especificidad y valores predictivos de una prueba se puede conocer su calidad. Una prueba sería muy sensible si al aplicarse a un conjunto de personas q tienen el trastorno dan positivo (+) en un porcentaje muy alto; y muy específica si un porcentaje muy elevado de personas NT dan negativo (-). Lo mismo se puede deducir en relación a los valores predictivos positivos y negativos, cuanto más próximos a 100 (o a 1 en términos de probabilidad) más valor predictivo tienen y mejor es su calidad.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

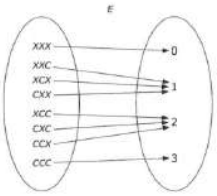
T7: VARIABLES ALEATORIAS Y MODELOS DISCRETOS DE PROBAILIDAD

En los primeros 5 temas se han estudiado las V estadísticas, el conjunto de valores resultantes de medir una carac. de interés sobre cada elemento individual de una población o muestra a través de un procedimiento de asignación numérica mediante la aplicación de det. reglas, dando lugar a los valores de la V estadística. Estos conjuntos de datos los hemos descrito mediante su correspondiente distribución de frecuencias.

Por otro lado, en el tema anterior se estudiaron los fundamentos de la teoría de probabilidades. Q un exp. se denomina aleatorio cuando su resultado no podemos predecirlo con certeza. Si se realiza una sola vez se obtendrá un único resultado del espacio muestral. Pero, a medida q aumenta el nº de ensayos irán apareciendo todos los resultados posibles, cada uno con su correspondiente probabilidad.

Con los conocimientos adquiridos ya es posible pasar del cálculo de probabilidades al estudio de las distribuciones de probabilidad. Éstas son las distribuciones teóricas correspondientes a la probabilidad de ocurrencia de un suceso si repitiésemos el exp. un número infinito (muy grande) de veces y q se refieren a la población.

1. CONCEPTO DE VARIABLE ALEATORIA



Esencial en la teoría de probabilidades. Surge ante la necesidad de cuantificar los resultados de los exp. aleatorios y poder realizar un estudio matemático de los mismos. ''Cualquier carac. medible q toma dif. valores con probabilidades det.
V aleatoria (v.a.), X, una función q asigna un nº real, y sólo uno, a cada resultado posible de un espacio muestral E de un exp. aleatorio. Se representan por letras mayúsculas del alfabeto latino, y se utilizan las letras minúsculas con subíndice para referirnos a los valores concretos q toman. Así X, Y... representan V aleatorias, X_1, X_2, ... y_1, y_2... representan los valores concretos q toman.

En el tema anterior definimos el espacio muestral E (espacio de resultados) como el conjunto de todos los resultados posibles de un exp., entendiendo por este un proceso q se puede repetir indefinidamente en las mismas condiciones y cuyo resultado no se puede predecir. Pues bien, cuando a cada suceso elemental o resultado posible del espacio muestral le asignamos un valor numérico se obtiene una V aleatoria denominada X, q toman valores numéricos, y se pueden definir dif. V sobre los resultados de un mismo exp. *Sobre el de «lanzar una moneda al aire en 3 ocasiones» podemos definir una V aleatoria como nº de caras obtenidas, nº de cruces obtenidas, o una V q toma el valor 1 cuando el nº de caras es mayor q el de cruces y toma el valor 0 en el otro caso. El azar interviene en el resultado q obtenemos y no en la V o función.*

2. TIPOS DE VARIABLES ALEATORIAS

- **Discretas**; adopta valores enteros; fijados 2 valores consecutivos, no puede tomar ninguno intermedio.
  Una V aleatoria X es discreta **(v.a.d.)** cuando sólo puede tomar un conjunto finito de valores o un conjunto infinito y numerable de valores. *Nº de caras q salen al lanzar 2 veces una moneda o el conjunto de los nº enteros, e puede adoptar un conjunto infinito y numerable de valores (los nº negativos, 0 y los nº positivos*).
- **Continuas**; dados dos valores, siempre se puede encontrar un 3er valor incluido entre los 2 primeros.
  Una V aleatoria X es continua **(v.a.c.)**: puede tomar infinitos valores o un conjunto de valores no numerable. *El tiempo de reacción ante un E, la estatura o el cociente intelectual.*

Variable aleatorias discretas

Función de probabilidad de una V.A. discreta

La descripción del comportamiento matemático se realizará similar a la 1ª parte del libro con las V estadísticas. En este caso, su distribución venía dada por los valores q toma la V y su correspondiente frecuencia; vendrá dada por los valores q la V puede tomar (X_1, X_2, …, X_n) y su correspondiente probabilidad .

Se llama función de probabilidad de una V.A. discreta X, y se representa por f(x), a aquella función q asocia a cada valor de la V la probabilidad de q ésta adopte ese valor:

$$f(x) = P(X = x)$$

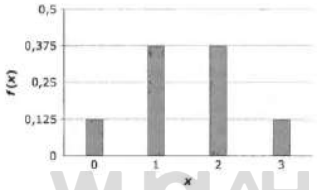
Exp. aleatorio consistente en lanzar una moneda al aire en 3 ocasiones. Si definimos una VA X como nº de caras, obtenemos la siguiente tabla:

Espacio muestral	X	P
⊕ ⊕ ⊕	$x_1 = 0$	1/8 = 0,125
⊕ ⊕ ☺		
⊕ ☺ ⊕	$x_2 = 1$	3/8 = 0,375
☺ ⊕ ⊕		
⊕ ☺ ⊕		
☺ ⊕ ☺	$x_3 = 2$	3/8 = 0,375
☺ ☺ ⊕		
☺ ☺ ☺	$x_4 = 3$	1/8 = 0,125

La 1ª columna recoge el espacio muestral del experimento E = {XXX, XXC, XCX, CXX, XCC, CXC, CCX, CCC}, siendo cada fila un suceso. El nº de sucesos o elementos del espacio muestral = 8. En la 2ª columna se muestran los valores q puede tomar la variable X y en la 3ª sus correspondientes probabilidades. Éstas se pueden calcular fácilmente teniendo en cuenta la definición clásica de probabilidad como la de obtener 3 cruces o ninguna cara (X_1 = 0) será 1/8, ya q hay un resultado favorable de ocho posibles.

Por tanto, la función de probabilidad de X es:

x	0	1	2	3
f(x)	0,125	0,375	0,375	0,125



La función de probabilidad de una V.A.D puede representarse mediante un diagrama de barras donde en el eje de abscisas se recogen los valores q toma la V y en el de ordenadas sus probabilidades.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Las 2 propiedades fundamentales q debe cumplir la función de probabilidad:

- Para cualquier valor de x, f(x) siempre toma valores positivos o nulos: $\forall x \in X \quad f(x) \geq 0$
- La suma de todas las probabilidades correspondientes a cada valor de X es =1. $\sum f(x) = f(x_1) + f(x_2) + \ldots + f(x_n)$ =1

Puede observarse q no son más q una adaptación de la definición axiomática de la probabilidad, aplicada al V.A..

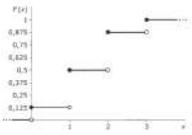
Función de distribución de una V.A. discreta

La **función de distribución** de probabilidad de una V.A. X se representa con la misma letra q su función de probabilidad, pero en mayúscula: F(x). indica la probabilidad de q la V tome un valor menor o igual q un valor concreto x. Asocia a cada valor de la V la probabilidad de q ésta adopte ese valor o cualquier otro inferior.

$$F(x) = P(X \leq x)$$

Si ordenamos de menor a mayor los valores x, esta función se obtiene acumulando o sumando los valores de la función de probabilidad:

$$F(x_k) = P(X \leq x_k) = f(x_1) + f(x_2) + \ldots + f(x_k)$$



En la función de probabilidad se le asigna a la probabilidad un valor concreto; esta es acumulativa, se le asigna la probabilidad a un valor concreto y todos los anteriores.
<— representación gráfica.
F (x) va «dando saltos» precisamente en los valores de la V (0, 1, 2 y 3).
El círculo blanco no incluye esos valores.
F(2) = 0,875 pero F(l,9999...) = F(1) = 0,5.

Se pueden deducir, sin demostraciones matemáticas, las propiedades fundamentales q debe cumplir:

- Todos los valores q toma la función son positivos o nulos.

$$\forall x \qquad F(x) \geq 0$$

- F(x) es nula o vale 0, para todo valor inferior al menor valor de la variable aleatoria, x_1:

$$F(x) = 0 \quad \text{si} \quad x < x_1$$

- F(x) es igual a uno para todo valor igual o superior al mayor valor de la V.A. Si llamamos x_n al mayor valor:

$$F(x) = 1 \qquad \text{si} \qquad x \geq x_n$$

- La función F(x) es no decreciente ya q es una acumulación o suma de probabilidades siempre positivas o nulas.

- La probabilidad P de q laV.A. X tome valores x mayores q x_1 y menores o iguales q x_2 ($x_1 < x \leq x_2$) es la diferencia entre los valores de la función de distribución correspondientes a su valor superior menos su valor inferior.

$$P(x_1 < x \leq x_2) = F(x_2) - F(x_1)$$

Media y varianza de una V.A. discreta

Aprendimos a describir una distribución de frecuencias de una V.Estadística a través de los índices de tendencia central y de dispersión. Lo mismo se puede hacer con una V.A.: calcular su media y su varianza.

Para una V.E.D. X se puede calcular su **media** obteniendo el sumatorio del producto de cada uno de los valores de la V por su frecuencia relativa o proporción. Pues para la media (letra griega «μ») de una V.A.D. X calcularemos el sumatorio de los productos de cada uno de los valores q toma la V por su correspondiente probabilidad:

$$\mu = E(X) = \sum x \cdot f(x)$$

Tb se denomina **esperanza matemática** o **valor esperado de X** y se representa por E(x). Este término tiene sus raíces en los juegos de azar y fue introducido con el fin de estimar las ganancias esperadas, si se repitiese el juego un elevado nº de veces. Referido a una V.A. representa el promedio teórico q tomaría si se repitiese el exp. infinitas veces. Por eso empleamos las letras griegas (μ), ya q se trata del parámetro correspondiente a la población de resultados del exp.

Para obtener la **varianza** de una V.A. X, δ_2 o V(X), debemos calcular el sumatorio del producto de cada uno de los valores q toma la V menos su media elevados al cuadrado multiplicados por su correspondiente valor de la función de probabilidad. Recordar la similitud con el índice estadístico correspondiente a la varianza de una V.E., T3.

$$\sigma^2 = V(X) = \sum (x - \mu)^2 \cdot f(x)$$

Una fórmula alternativa:

$$\sigma^2 = V(X) = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$$

$E(X^2) = \sum x^2 f(x)$

$[E(X)]^2$ es la media de la variable elevada al cuadrado, μ^2.

Por tanto, la varianza puede definirse tb como la esperanza de los cuadrados de X, $E(X^2)$, menos el cuadrado de la esperanza de X, $[E(x)]^2$.

De manera análoga a las V.E, la **desviación típica δ** de una V.A.D. X es la raíz cuadrada de la varianza:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum (x - \mu)^2 \cdot f(x)}$$

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
WUOLAH Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

3. DISTRIBUCIONES DISCRETAS DE PROBABILIDAD

No siempre es necesario construir las funciones de probabilidad y de distribución, en función de las condiciones de partida del exp. y de las carac. de la VA, podemos ajustar estas distribuciones a alguna (modelo teórico de probabilidad) ya conocida, q simplifica mucho el trabajo si se ajuste según sus propiedades.

Existen diversas distribuciones teóricas para VD, bien conocidas, por utilizarse frecuentemente como modelo, o por su interés como instrumento estadístico:

- De **Bernoulli** y **binomial**;
- **Poisson** o ''sucesos raros'' q se utiliza bajo o las mismas condiciones de la binomial para V dicotómicas, pero con un elevado nº de ensayos y un valor de p muy pequeño.
- **Multinominal**; ensayos q ofrecen +2 resultados posibles. Supone una generalización de la binomial o un caso particular de aquella.

De muchas se han elaborado una serie de tablas q facilitan su aplicación a problemas concretos. Por lo general, en Psicología y Ciencias de la Salud se trabaja con VAD q sólo pueden tomar 2 valores (dicotómicas) y q habitualmente representaremos por 1 y 0. En estos casos, resultan muy útiles la distribución de Bernoulli, y, especialmente, su generalización a *n* ensayos, q es la distribución binomial.

La distribución de Bernoulli

Lanzar una moneda al aire 1 admite sólo 2 resultados; cara o cruz. El acierto o fallo a una pregunta con 2 alternativas respondida al azar, el lado izq. o derecho de un laberinto en forma de T elegido por una rata no entrenada... son algunos de los múltiples ej. A una de ellas se le denomina «**éxito o acierto**» (q habitualmente, se codifica con 1) y a la otra «fracaso o error» (q se codifica como 0), sin q estos términos tengan connotaciones ni positivas ni negativas. Es el fundamento y la base de otras distribuciones discretas, entre las q destaca la distribución binomial.

De este modo, la VAD q sigue el modelo de Bernoulli se define como una VA dicotómica X, con 2 posibles valores mutuamente exclusivos: 1 (éxito) con probabilidad *p* y 0 (fracaso) con probabilidad *q*; la suma de ambas =1.

f (1) = P(X = 1) = p
f (0) = P (X = 0) = q
p + q = 1, por lo q q = 1 -p

Una VA X q sigue el modelo de Bernoulli con parámetro *p*, se denota como **X → Ber (P)** y presenta las carac.:

- Función de probabilidad: $f(x) = P(X = x) = p^x q^{1-x}$
- Función de distribución: $F(x) = P(X \leq x) = \sum p^x q^{1-x}$
- Media: μ = *p*
- Varianza: $\delta^2 = p(1 - p) = pq$
- Desviación típica: $\delta = \sqrt{pq}$

x puede adoptar el valor 0 (fracaso) o 1 (éxito)
***p* = probabilidad de éxito en el único ensayo del exp.**
***q* = probabilidad de fracaso (1-p) en el único ensayo del exp.**

La distribución binomial

Generalización de la distribución de Bernoulli en la q el exp. se repite +1 vez. Se repite *n* veces, y de forma indep., un ensayo Bernoulli en el q la probabilidad de «éxito», p, se mantiene constante en cada uno de los *n* ensayos.

Una VA X sigue una distribución binomial (con parámetros n y p) si expresa el nº de éxitos en n realizaciones indep. de un exp. con probabilidad *p* de obtener «éxito» y, por tanto, (1-p) de obtener «fracaso». Suele representarse por ***B*(n, p)**.

***B:* binomial,**
***n:* nº de ensayos o veces q se repite un exp. Bernoulli.**
***p:* probabilidad de «éxito».**

La distribución de Bernoulli descrita sería un caso particular de la binomial con parámetro *n* =1 (un único ensayo), el parámetro *p* sería la probabilidad de «éxito», y se representaría como una binomial B(1, p).

Carac. de una distribución B(n, p):

- Función de probabilidad: $f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}$
- Función de distribución: $F(x) = P(X \leq x) = \sum \binom{n}{x} p^x q^{n-x}$
- Media: μ = np.
- Varianza: δ^2 = npq
- Desviación típica: $\delta = \sqrt{npq}$
- El nº combinatorio «n sobre x» $\binom{n}{x}$ es igual a $\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$
  - el factorial de un nº *n* es: n ! = n •(n - 1) •(n - 2) • ... •(n - (n - 1))
  - el factorial 1! = 1 y el factorial 0!= 1

x = nº de aciertos
q = probabilidad de fracaso (1-p) en cada ensayo

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Aunq es relativamente fácil deducir las carac. anteriores no lo vamos a hacer aquí formalmente, recurriremos a su aplicación. Si se lanza una moneda 3 veces y definimos la VA X como «nº de caras obtenidas», esta seguirá el modelo de distribución binomial con parámetros *n*=3 y *p*=0,5. Diremos q X sigue un modelo B(3, 0,5) pq en cada lanzamiento sólo son posibles 2 resultados: «éxito» (cara) y «fracaso» (cruz); los ensayos son indep. entre sí (el resultado en un ensayo no depende de lo q haya salido o no en los anteriores) y la probabilidad de «éxito» se mantiene constante.
La utilización de funciones de probabilidad y distribución requiere cálculos tediosos. Las tablas de la función de probabilidad y de la función de distribución binomial, Tablas I y II del Formulario, nos evitan el cálculo de las probabilidades a partir de la ecuación de esas funciones, facilitando su obtención con elevado *n*.

Tabla I, función de probabilidad binomial:

- 1ª columna, *n*, incluye los valores desde 1 hasta 20.
- 2ª columna, nº de «éxitos» (x) q esperamos obtener para ese nº de ensayos y q abarcan desde 0 hasta ese nº.
- 1ª fila, algunos valores de la probabilidad de «éxito» (p) desde 0,01 a 0,5.
- En el interior se encuentran las probabilidades correspondientes. La probabilidad buscada, para unos valores concretos de n y x, se encuentra en la intersección de su fila con la correspondiente columna de p. *La probabilidad de obtener 2 éxitos en 3 ensayos con una probabilidad de éxito de 0,3 - Tabla en la posición q se indica y vale 0,1890.*

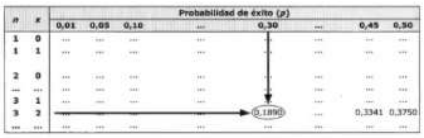


Tabla II:

- Idéntica a la anterior.
- Las probabilidades del interior son acumuladas.

Sin embargo, en ambas, sólo se contienen valores de p desde 0,1 hasta 0,5. ¿Q hacer cuando tengamos una p > 0,5? Hay q intercambiar las condiciones de «éxito» y «fracaso». Y, el nº de ensayos *n* sólo llega hasta 20. Esto no plantea ningún problema, para valores superiores podemos hacer una aproximación de la binomial a la distribución normal, como se verá en el próximo tema.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

T8: MODELOS CONTINUOS DE PROBABILIDAD

Se han estudiado las VAD, unificando conceptos como la distribución de frecuencias y la probabilidad, q nos ha permitido definir su función de probabilidad, esperanza matemática y varianza teórica. De forma análoga se pueden definir estos mismos conceptos para las VAContinuas. Sin embargo, el problema es q estas no toman un nº finito de valores, toma infinitos. Tenemos q acudir a un modelo probabilístico q permite determinar, mediante el cálculo integral, la probabilidad de un intervalo de la V y no de un valor concreto como en las discretas. El proceso del cálculo de integrales no se va a tratar en este libro ni es necesario. Para estas distribuciones se han elaborado tablas q contienen los valores de las probabilidades correspondientes.

1. CARACTERÍSTICAS DE LAS VARIABLES ALEATORIAS CONTINUAS

En el tema anterior se definió una VAC como aquella q puede adoptar infinitos valores o un conjunto de valores no numerables. Dado q dentro de cada intervalo de valores existen a su vez infinitos valores posibles, la probabilidad de q tome un valor det. es nula. Es decir, obtener un det. valor de X = 0, por lo q las probabilidades se van a asignar a un det. intervalo. Para ello, se acude al concepto de función de densidad de probabilidad en torno a un valor, en lugar de función de probabilidad de un valor de las VAD.

Función de densidad función de distribución

Función de densidad de probabilidad de una VAC, f(x): función q cumple las 2 condiciones siguientes:

1. Los valores de f(x) son siempre = 0 o positivos, nunca negativos.
2. Una integral, q en VC es el análogo al sumatorio en VDiscretas. El área total (desde -∞ hasta +∞ en la variable X) bajo la curva es igual a uno. De ahí q se aplique para la determinación de las probabilidades correspondientes a las variables continuas.

a) $f(x) \geq 0$

b) $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

Podemos calcular la probabilidad de q X se encuentre en un det. intervalo [a,b] mediante el cálculo integral:

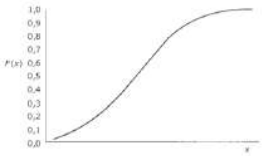
$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

f(x): función de densidad de probabilidad de X
La integral está definida para el intervalo [a,b]

f(x) no se corresponde con un valor puntual de probabilidad sino de densidad q, aplicándole el cálculo de integrales, permite obtener la probabilidad para un intervalo. Dicho de otro modo, f(x) no es una probabilidad, pero la integral de f(x) para un det. intervalo [a, b] de X si nos proporciona un valor de probabilidad .

La función de distribución, F(x), se describe como para las VD; la probabilidad acumulada hasta un cierto valor de la V. **Función de distribución acumulada** o función de distribución de probabilidad de una VAC, F(x), aquella función q asocia a cada valor de la variable X la probabilidad de obtener valores menores o iguales q un valor dado (= menor, ya q la probabilidad de ser igual al valor dado es 0).

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\,dx$$



Para la función de distribución en el caso continuo se mantienen las mismas propiedades q en las VAD:

- F(x) ≥ 0 —> todos los valores q toma la función de distribución son nulos o positivos F(x).
- F(-∞) =0 y F(+∞) =1 por lo q 0 ≤ F(x) ≤ 1 —> F(x), al ser una probabilidad, está acotada entre 0 y 1.
- ∀a ≤ b, P(a ≤ X ≤ b) = F(b) - F(*a*) —> la probabilidad de q X se encuentre en el intervalo [a,b] es la diferencia entre la función de distribución para X= b, F(b) y la función de distribución para X= a, F(a).

Media y varianza de una variable aleatoria continua

Como las VD, las VAC tb presentan una media o valor esperado y una varianza, q pueden obtenerse mediante procedimientos análogos a las VD, pero adaptados para el caso continuo.

Sea X una VAC, la **media** o **valor esperado**, μ o E(X):

$$\mu = E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)\,dx$$

La **varianza**:

$$\sigma^2 = V(X) = \int_{-\infty}^{+\infty} [x-\mu]^2 \cdot f(x)\,dx$$

VARIABLES ALEATORIAS DISCRETAS	VARIABLES ALEATORIAS CONTINUAS
Probabilidad para X = x $f(x) = P(X = x)$ $f(x)$ = función de probabilidad	***Probabilidad para el intervalo [a,b]*** $P(a \leq X \leq b) = \int_a^b f(x)\,dx$ $f(x)$ = función de densidad de probabilidad
Función de Distribución $F(x_k) = P(X \leq x_k) =$ $= f(x_1) + f(x_2) + \ldots + f(x_k)$	***Función de Distribución*** $F(x) = P(X \leq x_k) = \int_{-\infty}^{x_k} f(x)\,dx$
Media o Valor Esperado $\mu = E(X) = \sum x \cdot f(x)$	***Media o Valor Esperado*** $\mu = E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)\,dx$
Varianza $\sigma^2 = V(X) = \sum (x-\mu)^2 \cdot f(x)$	***Varianza*** $\sigma^2 = V(X) = \int_{-\infty}^{+\infty} [x-\mu]^2 \cdot f(x)\,dx$

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Las propiedades de la media y la varianza para VC son las mismas q las discretas. Por otra parte, para el cálculo de las q estudiaremos es necesario utilizar el cálculo integral, pq se han derivado fórmulas directas para la obtención de dichos parámetros.

2. DISTRIBUCIÓN NORMAL - CAMPANA DE GAUSS / CURVA NORMAL

Definida por De Moivre en un intento de encontrar las probabilidades acumuladas en una distribución binomial cuando n (nº de ensayos) es grande.

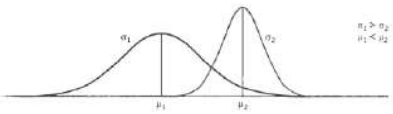
Función de densidad de probabilidad para una variable X q tiene distribución normal:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \quad \text{para} \quad -\infty < X < \infty$$

μ: parámetro media o valor esperado de la distribución.
δ: parámetro desviación típica de la distribución.
π = 3,1416
***e* = 2,718 (base de los logaritmos neperianos).**

Si una variable X tiene una función de densidad q se ajusta a la fórmula anterior, lo expresaremos por: X —> *N*(μ, δ), indicando q tiene una distribución Normal (N) con parámetros μ y δ.

No se trata de una única distribución, corresponde a toda una familia caracterizada por sus parámetros media, μ, y desviación típica, δ. Como puede observarse su forma de «campana» es más apuntada cuanta menor desviación típica. Su figura nos indica q la puntuación de la mayoría de los individuos, en una V q sigue esta distribución, se encuentra en torno a la media y, a medida q nos alejamos va disminuyendo la frecuencia.



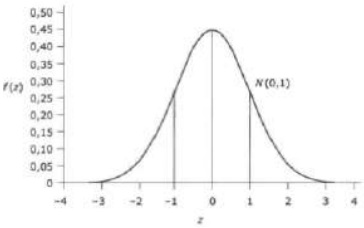
Según una de sus propiedades fundamentales, si a una variable X q se distribuye normalmente, con μ y δ le aplicamos una transformación lineal de la forma Y=bX+*a*, la nueva variable Y tb se distribuirá normalmente pero con media $\mu_Y = b\mu_x + a$ y desviación típica $\delta_Y = |b| \cdot \delta_x$

Por otra parte, si restamos la media y dividimos por la desviación típica obtenemos una nueva V q designamos por z. Esta se distribuirá normalmente con media =0 y desviación típica igual =1 **z → N(0,1)**. La demostración de μ_2=0 y δ_2=1 excede el presente curso.

$$z = \frac{X - \mu}{\sigma}$$

La **función de densidad de probabilidad de z**:

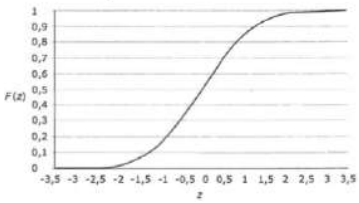
$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad \text{para} \quad -\infty < z < \infty$$



Esta distribución se denomina normal tipificada o normal estandarizada. Para la aplicación a problemas concretos recurriremos a las Tablas III y IV del Formulario. Si observamos la figura entre las propiedades fundamentales de una distribución normal podemos destacar:

- Es simétrica en torno a su media, μ, q coincide con su mediana y su moda.
- La curva normal tiene 2 puntos de inflexión, donde la curva pasa de ser cóncava a convexa. Estos están situados a una distancia de una desviación típica de la media.
- Es asintótica en el eje de abscisas; se extiende desde -∞ hasta +∞ sin llegar nunca a tocar el eje X.

Su función de distribución —>



Utilización de las tablas

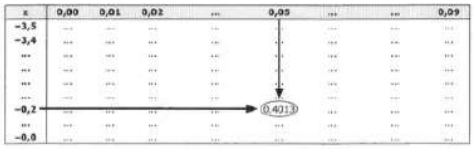
En las Tablas III y IV se recoge la función de distribución de la curva normal estándar. Se presentan todas las puntuaciones típicas desde -3,59 hasta +3,59 con intervalos de 0,01.

- 1ª columna, encabezada con la letra z, consta de un nº con un decimal, la puntuación típica.
- 1ª fila (derecha de la letra z), 2º decimal de la puntuación z.
- Todos los valores interiores representan probabilidades = llevan un cero delante de la coma.

Tabla III—> puntuaciones típicas negativas (por debajo de la media). Tabla IV —> las positivas (por encima de la media).

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

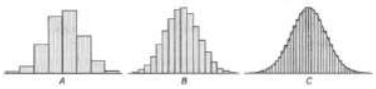


La puntuación típica z = -0,25 (Tabla III) deja por debajo de sí una probabilidad de 0,4013. Tabla III. La puntuación típica z=0,25 (Tabla IV) deja por debajo de sí una proporción de 0,5987. Al ser una distribución simétrica puede comprobarse que la proporción que queda por debajo de z =-0,25 es igual a la proporción q queda por encima de z =0,25 (1- 0,5987 =0,4013). Si la Tabla no recoge el valor exacto de z q se busca, se puede utilizar el + próximo. Algunos casos concretos son:

1. **Cálculo de la probabilidad para valores menores o iguales q una det. puntuación típica.**
   Se busca directamente en la Tabla.
2. **Cálculo de la probabilidad para valores mayores q una det. puntuación.**
   Se mira en la tabla la probabilidad q esa puntuación deja por debajo y se resta de 1.
3. **Cálculo de la probabilidad entre dos puntuaciones det.**
   Se restan las probabilidades q dejan por debajo de sí las 2 puntuaciones típicas.

Histograma y distribución normal

Datos de una muestra en una variable X (A). Si se hacen los intervalos + pequeños (B) y dibujamos el polígono de frecuencias (C) llegamos a una distribución similar a la normal. La puntuación de la mayoría de los casos se encuentra en torno a la media y, a medida q nos alejamos va disminuyendo. Esto va a permitir aplicar las propiedades de la curva normal a nuestros datos y utilizar las tablas de la misma forma q se ha visto.

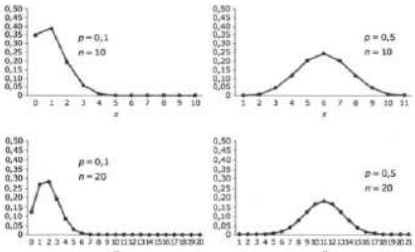


$$z_i = \frac{X_i - \overline{X}}{S_x}$$

Si se dispone de los datos originales de un grupo de sujetos en una det. variable X, y ésta se distribuye normalmente, para resolver det. cálculos utilizando Tablas III y IV, debemos transformar las puntuaciones directas en puntuaciones típicas mediante la fórmula de la izq.

Aproximación de la binomial a la normal

Cuando para la distribución binomial tenemos un n superior a 20, debemos aproximar la distribución binomial a la normal. Esta aproximación mejora a medida q *p* (probabilidad de éxito) se aproxima a 0,5 y *n* es grande:



Se sabe q una variable, X, q sigue una distribución binomial tiene una media μ =np y una desviación típica $\delta = \sqrt{npq}$. Por tanto, se puede transformar su función de probabilidad (q es discreta) a la normal:

La distribución normal es continua y, como para cualquier distribución continua, la probabilidad de q la variable X tome un valor concreto es cero: P (X= x) = O. Para aproximar la distribución binomial a la normal se establecerá un intervalo entre 0,5 unidades a la izq. y a la derecha de la puntuación:

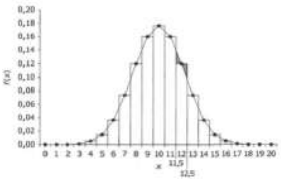
$$P(X = x) = P\left[(x - 0,5) \le x \le (x + 0,5)\right]$$

A continuación, transformamos las puntuaciones en típicas:

$$P(X = x) = P\left[\frac{(x - 0,5) - \mu}{\sigma} \le \frac{x - \mu}{\sigma} \le \frac{(x + 0,5) - \mu}{\sigma}\right]$$

Debido a q μ = *np* y $\delta = \sqrt{npq}$, la **aproximación de la binomial a la normal** se define como:

$$\boxed{P(X = x) = P\left[\frac{(x - 0,5) - np}{\sqrt{npq}} \le z \le \frac{(x + 0,5) - np}{\sqrt{npq}}\right]}$$



En el caso anterior, sumar y restar el valor 0,5 se llama **corrección por continuidad**, permitiendo utilizar las puntuaciones discretas, X, como si fuesen continuas. Para ello, se interpreta cada puntuación, X, como si fuesen los puntos medios de sus intervalos. Se intenta asegurar q el intervalo incluya los valores discretos de la binomial.

3. LA DISTRIBUCIÓN DE PEARSON

Ya se ha visto en el T4 el estadístico ji-cuadrado (x^2), para referirse a la correlación entre V cualitativas. A partir de ahora, x2 se va utilizar para hacer referencia a una distribución continua de probabilidad.

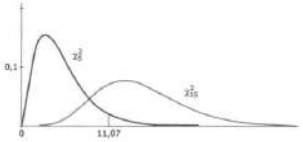
Sean $X_1, X_2, ..., X_n$ un conjunto de *n* variables aleatorias indep. con una distribución N(0,1), entonces una nueva V aleatoria $X = x^2_1 + x^2_2 + \ldots + X^2_n$ sigue una distribución X —> X^2_n (Ji- cuadrado con *n* grados de libertad)·

- Media o valor esperado μ =n
- Varianza $\delta^2 = 2n$

Los grados de libertad (n) indican q cada una de las *n* variables aleatorias puede tomar cualquier valor de sus posibles valores, sean cuales sean los tomados por las *n*-1 restantes.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material



Esta distribución se usa fundamentalmente en pruebas de bondad de ajuste (para contrastar si la distribución de una V se ajusta a una distribución det.l). Al igual q otras distribuciones, es una familia de curvas como las presentadas, q varían en función de los grados de libertad.

- Nunca adopta valores menores de 0.
- Es asimétrica positiva pero a medida q ↑ sus grados de libertad se va aproximando a la distribución normal.
- Para $n > 30$ la podemos aproximar a una distribución N (n, $\sqrt{2n}$).

La Tabla V del Formulario permite obtener las probabilidades acumuladas a algunos valores de toda la familia de distribuciones.

- 1ª fila —> probabilidades o proporciones.
- 1ª —> columna los grados de libertad correspondientes.
- En el interior —> los valores de la V. *Para una V q sigue una distribución x^2 con 5 grados de libertad, X ➔ X^2_5, el valor 11,07 deja por debajo una proporción de 0,95. Por tanto, P (X ≤ 11,07) = 0,95. Esta puntuación se corresponde con el percentil 95. Suele presentarse $_{0,95}X^2_5$ = 11,07.*

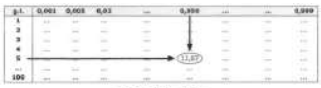


Tabla V del Formulario

Si lo q interesa es hallar P (X ≥ 11, 07):

$$P(X \geq 11{,}07) = 1 - P(X \leq 11{,}07) = 1 - 0{,}95 = 0{,}05$$

4. LA DISTRIBUCIÓN T DE STUDENT

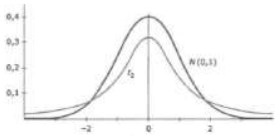
A la h de definir este tipo de distribución de probabilidad, al igual q se hizo anteriormente con x^2, se hará en función de otras distribuciones ya conocidas.

Sean X e Y dos V aleatorias independientes, donde X sigue una distribución N(0,1) e Y una distribución X^2_n. Entonces, la V aleatoria $T = \frac{X}{\sqrt{Y/n}}$ una distribución *t* con *n* grados de libertad y se expresa: T ➔ t_n Sus parámetros son :

- Media o valor esperado μ = 0
- Varianza δ^2 = n/n-2

Distribución t: el cociente entre una variable *N*(0,1) y la raíz cuadrada de una variable X^2_n dividida por sus grados de libertad. Su nombre se debe a su descubridor, el matemático Gosset, q publicó bajo el seudónimo de Student.

Distribución t con dos grados de libertad, junto a la distribución normal estándar —>



Características:

- Es simétrica, con μ = 0. Su forma es muy parecida a la *N*(0,1), aunq menos apuntada.
- Puede tomar cualquier valor entre -∞ y +∞.
- A medida q ↑ los grados de libertad, la distribución se aproxima más a una distribución normal.
- La curva es asintótica al eje de abscisas.

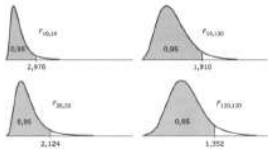
Fundamentalmente se utiliza en estadística inferencial. En la Tabla VI del Formulario se presentan los valores positivos.

- 1ª columna: grados de libertad.
- 1ª fila: distintas probabilidades o proporciones de valores menores o iguales q un valor positivo dado.
- Como es una distribución simétrica podemos hallar las probabilidades asociadas a valores negativos a partir de los valores positivos de la Tabla VI.

5. LA DISTRIBUCIÓN DE F DE FISHER-SNEDECOR

Si X_1 y X_2 son V aleatorias indep., con distribución x^2 con n_1 y n_2 grados de libertad respectivamente; una nueva variable F —>$F = \frac{X_1/n_1}{X_2/n_2}$ sigue una distribución F con n_1 y n_2 grados de libertad (F_{n1n2}). Siendo n_1 los grados de libertad del numerador y n_2 los del denominador.

- Su media o valor esperado viene definido por $\mu = \frac{n_2}{n_2 - 2}$ para $n_2 > 2$
- Su varianza por $\sigma^2 = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 4)(n_2 - 2)^2}$ para $n_2 > 4$



Fundamentalmente en el contraste de hipótesis (Análisis de Varianza...). Figura = representación según distintos grados delibertad.

Características:

- Es asimétrica positiva, nunca toma valores menores q 0.
- Propiedad recíproca; si X es una V con distribución F con n_1 y n_2 grados de libertad, entonces la variable Y = 1/X es tb una distribución F con n_2 y n_1 grados de libertad:
  p y 1 - *p*: probabilidades acumuladas asociadas al valor de la V. Esta propiedad es útil para obtener algunos percentiles o probabilidades q no aparecen en la tabla.
- La Tabla VII recoge solamente la probabilidad de q X sea menor o igual q 0,900; 0,950; 0,975; 0,990 y 0,995, los valores utilizados habitualmente.

$$_pF_{n_1,n_2} = \frac{1}{_{1-p}F_{n_2,n_1}}$$

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

T9: MUESTREO Y DISTRIBUCIÓN MUESTRAL DE UN ESTADÍSTICO

En temas anteriores se ha presentado la materia correspondiente a la Estadística Descriptiva. En los 2 siguientes temas se presentarán las bases sobre las q se asienta la **Inferencia Estadística** (disciplina q abarca las técnicas y métodos q permiten deducir las propiedades desconocidas de la población a partir de los datos obtenidos en la muestra). La inf. q obtenemos de las muestras permite estudiar el comportamiento de las VA (discretas o continuas) y de los índices estadísticos q las representan (q son tb VA). Sobre esta base, apoyándonos en la teoría de muestreo, podremos estimar los valores de los parámetros a partir de los valores de los correspondientes índices estadísticos.

ESTADÍSTICA DESCRIPTIVA
Recogida, organización y análisis de los datos.
PROBABILIDAD
Permite y legitima el salto de las carac. (conocidas) de la muestra hasta las desconocidas de la población.
ESTADÍSTICA INFERENCIAL
Obtener inf. acerca de la población a partir de la aportada por la muestra mediante la combinación de los modelos de probabilidad y de los estadísticos. Comprende los campos estimación de parámetros y contraste de hipótesis. La piedra angular de ambos es el concepto de distribución muestral de un estadístico, q establece la relación entre las carac. de la población y el comportamiento de los estadísticos de las muestras q las representan.

1. MUESTREO

Al plantearse la recogida de datos suele ocurrir q por razones de coste económico, tiempo o imposibilidad real, se hace muy difícil estudiar a todo el grupo sobre el q se quiere trabajar. Por lo general es imposible trabajar con el conjunto total población. Por tanto, tendremos q bus- car una muestra (subconjunto). Lo ideal es elegir las muestras de tal forma y modo q representen y sean fiel reflejo de las carac. relevantes a la investigación en la población de trabajo.

Es posible trabajar con «unos pocos» datos extraídos de un conjunto más amplio y q las conclusiones sean válidas para el conjunto total. Pero, ¿cuáles son los procedimientos de selección q lo garantizan? ¿es posible cuantificar el riesgo de equivocación/ error al realizar afirmaciones sobre las poblaciones?

Conceptos básicos en el muestreo

- **Población**: colección, finita o infinita, de elementos q comparten ciertas carac. *Todos los SH componen la población de hombres y mujeres; todas las personas que de forma habitual se expresan en francés componen la población de francófonos*. El concepto es indep. de la cantidad de elementos, puede estar compuesta por 1 solo; *población de satélites de la Tierra, la Luna*. 1 elemento pueda pertenecer a +1 población si cumple los criterios necesarios. Un elemento pertenece a una población solo durante el tiempo q se cumplen las propiedades. Atendiendo al nº de elementos, las poblaciones pueden ser finitas o infinitas. En general, son muy grandes y esto hace q sea prácticamente inviable trabajar con ellas y lo habitual es trabajar con muestras. Los índices q representan los valores q resumen las carac, de las poblaciones = **parámetros**. Son constantes, se calculan con todos los elementos de la población.
- **Censo**: en det. ocasiones resulta posible estudiar a todos y cada uno de los elementos q componen la población, realizándose el estudio de todos los elementos q componen la población. *Censo poblacional.*
- **Muestra**: cuando no es posible o conveniente trabajar con la población. Es representativa cuando reúne las mismas carac. q la población. La representatividadegarantiza q los resultados puedan ser generalizados. Los índices q representan los valores q resumen las carac. de las muestras = **estadísticos** y son VA cuyos valores varían en función de los elementos q compongan la muestra.
- **Muestreo**: conjunto de procedimientos y técnicas q permiten extraer muestras de una población garantizando la representatividad o de procedimientos y reglas para poblaciones. La generalización, desde la parte al todo, supone siempre un cierto error q se puede cuantificar y controlar a través de la Estadística mediante dos conceptos: error máximo (E_{max}) y nivel de confianza ($1-\alpha$). Pasos a seguir:
    A. Definir los casos (participantes u otros seres vivos, objetos, fenómenos o comunidades) sobre los cuales se habrán de recolectar los datos.
    B. Delimitar la población mediante una carac. q defina, de forma exhaustiva y excluyente, a los individuos.
    C. Elegir el método de selección de la muestra. (T9)
    D. Calcular el tamaño de la muestra. (T10)
    E. Aplicar el procedimiento de selección.
    F. Obtener la muestra.

Tipos de muestreo

- **Muestreo probabilístico**: se conoce la probabilidad q tiene cada elemento de la población de ser elegido para formar parte de la muestra y se conoce el **marco muestral** (listado de elementos q componen la población). Su ventaja es garantizar la representatividad = permitir hacer inferencias.
- **Muestreo no probabilístico**: no se conoce la probabilidad q tiene cada elemento de ser elegido. No garantiza la representatividad = conclusiones circunscritas a la situación sin posibilidad de generalizar.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Dentro de cada una de estas categorías hay dif. métodos según la forma en q se extraen los elementos:

Tipos de muestreo	Métodos de muestreo
Probabilistico	Aleatorio Simple Aleatorio Sistemático Aleatorio Estratificado Aleatorio por Conglomerados
No probabilistico	Por Cuotas Intencional (opinático) Incidental (casual) Bola de Nieve

Métodos de muestreo probabilístico

MUESTREO ALEATORIO SIMPLE

Consiste en tomar de una población de tamaño *N* una muestra de tamaño *n*, utilizando algún procedimiento q garantice q todos los elementos de la población tienen la misma probabilidad de ser elegidos.

- Se asigna un nº a cada elemento de la población.
- A través de algún medio mecánico o informático *(bolas dentro de una bolsa, tablas de nº aleatorios, nº aleatorios generados con una calculadora u ordenador…)* se eligen tantos sujetos como sea necesario para completar el tamaño de muestra.

Todas las muestras posibles son **equiprobables**, tienen la misma probabilidad de ser elegidas. Pero, la probabilidad de cada una de ellas y la probabilidad de pertenencia de los elementos será distinta en función de la forma en q se genere la muestra. Para formar la muestra se puede actuar de 2 modos:

- **Con reposición**: Tras elegir un elemento, éste se reincorpora a la población para q pueda ser elegido en la siguiente extracción, así la población siempre tiene el tamaño *N*. Cada elemento de la población (formada por N elementos) tiene la misma probabilidad (1/N) en cualquiera de las extracciones de pertenecer a la muestra (formada por n elementos). La probabilidad de obtener una muestra concreta de *n* elementos es: (1/N)x(1/N)x(1/N)x ... x(1/N) =(1/N^n). Es decir, q de las N^n muestras posibles de tamaño *n*, todas tienen la misma probabilidad de ser elegidas.
- **Sin reposición**. Una vez seleccionado un elemento de la población no se reintegra, la población va perdiendo tamaño. En la 1ª extracción, el tamaño es N, en la 2ª es N-1, en la 3ª es N-2… El tamaño de la población cambia con cada extracción. Aunq todos los elementos tienen la misma probabilidad de ser elegidos, esa va cambiando y el resultado de una extracción no es indep. del resultado obtenido en las demás. En este caso, la función de probabilidad conjunta de las V es diferente del producto de sus funciones de probabilidad individuales. Existen $\binom{N}{n}$ muestras posibles de tamaño *n*, y la probabilidad de una muestra concreta es $\frac{1}{\binom{N}{n}}$.

La gran mayoría de procedimientos de la Estadística Inferencial exigen el principio de indep. en la obtención de muestras, lo cual no se cumple en el muestreo aleatorio sin reposición (muy habitual en investigación). Esto se resuelve considerando q, cuando el tamaño de la población (N) es grande con respecto al de la muestra (n), las probabilidades calculadas con ambos muestreos (con y sin reposición) son prácticamente iguales. Finalmente, este tipo de muestreo, aunq resulta atractivo por su sencillez, tiene poca o nula utilidad práctica cuando la población es muy grande.

MUESTREO ALEATORIO SISTEMÁTICO

Cuando los elementos de la población están ordenados o pueden ordenarse. Es necesario asignar un nº a todos los elementos de la población, pero, en lugar de extraer *n* números aleatorios sólo se extrae uno (*i*). El nº i del q se parte es elegido al azar, y los elementos q compondrán la muestra ocupan los lugares i, i+k, i+2k, i+3k ... i + (n - 1)k, siendo k el resultado de dividir el tamaño de la población entre el tamaño de la muestra (k = N / n). Se toman los individuos de k en k partiendo del sujeto en la posición i. El nº i q empleamos como punto de partida será al azar entre 1 y k. En este tipo de muestreo no todos los elementos tienen la misma probabilidad de ser extraídos, y las extracciones no son indep. El riesgo está en aquellos casos en q se dan periodicidades en la población, ya q al seleccionar una periodicidad constante, los elementos seleccionados para la muestra pueden no ser representativos.

MUESTREO ALEATORIO ESTRATIFICADO

Cuando la población no es homogénea, existen grupos o estratos heterogéneos entre sí con gran homogeneidad dentro del estrato (se puede estratificar *según la profesión, municipio de residencia, sexo, estado civil…*). Se pretende asegurar q todos los estratos de interés estén representados adecuadamente en la muestra. Cada uno funciona de forma indep., pudiendo aplicarse dentro de ellos el muestreo aleatorio simple o sistemático. El procedimiento de composición de la muestra en los dif. estratos se denomina **afijación**, y puede ser de dif. tipos:

- **Afijación simple**: a cada estrato le corresponde igual nº de elementos muestrales.
- **Afijación proporcional**: la distribución se hace de acuerdo con el peso (tamaño) de la población en cada estrato.

La muestra total se forma por la suma de las muestras de cada estrato. Cada submuestra es indep. Permite aplicar técnicas de selección dif. dentro de cada estrato y obtener estimaciones separadas en cada una.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

MUESTREO ALEATORIO POR CONGLOMERADOS
Los 3 métodos de muestreo presentados hasta ahora están diseñados para seleccionar directamente los elementos de la población; las unidades muestrales (los sujetos) son los elementos de la población. En el muestreo por conglomerados, la unidad muestral es un grupo de elementos de la población q conforman una unidad más amplia. *Áreas sanitarias, dpt. universitarios, una caja de det. producto…*

El procedimiento consiste en seleccionar aleatoriamente un cierto nº de conglomerados (el necesario para alcanzar el tamaño muestral) y trabajar con todos los elementos pertenecientes a los conglomerados elegidos. Si el nº de es muy amplio seleccionamos algunos al azar; **muestreo por conglomerados bietápico**. En general se habla de muestreo por etapas o polietápico cuando hay +2 etapas. Para aplicarlo en cada etapa se van seleccionando conglomerados de menor tamaño hasta q en la última etapa se trabaja con los *n* elementos q componen esos conglomerados. *En un estudio sobre la población universitaria española se seleccionan Universidades; dentro de ellas Facultades, dentro de ellas carreras específicas y dentro los cursos, el último conglomerado. La muestra estaría formada por todos los individuos de los cursos.*

Métodos de muestreo no probabilístico
Cuando no es posible realizar un muestreo probabilístico pq desconocemos la probabilidad de inclusión de cada elemento en la muestra y/o tiene un excesivo costo económico o de tiempo. Se acude a métodos no probabilísticos, aun siendo conscientes de q no sirven para realizar generalizaciones pq no se tiene certeza de q la muestra extraída es representativa, ya q no todos los sujetos de la población tienen la misma probabilidad de ser elegidos. En general, se selecciona siguiendo det. criterios, procurando q la muestra resultante sea lo más parecida posible a la población.

MUESTREO POR CUOTAS / ACCIDENTAL
Base; un buen conocimiento de los estratos de la población y/o de los individuos + adecuados para la investigación. Semejanzas con el muestreo aleatorio estratificado, pero no tiene su carácter de aleatoriedad.

Se fijan unas «**cuotas**» (nº de individuos q reúnen unas det. condiciones). Una vez determinadas se eligen los primeros q se encuentren q cumplan esas carac. Se utiliza mucho en las encuestas de opinión. *La Consejería de Sanidad de una CA desea estudiar la incidencia de las drogas en la adolescencia. A través de los informes de la Consejería de Educación, se conoce cuáles son los centros educativos + afectados, se fija un nº de sujetos a entrevistar proporcional a cada uno de los centros fijados y, finalmente, se deja en manos de los encuestadores a q sujetos concretos se entrevista.*

MUESTREO OPINÁTICO O INTENCIONAL
Se caracteriza por un esfuerzo deliberado de obtener muestras «representativas» mediante la inclusión en la muestra de grupos supuestamente típicos. Muy frecuente su utilización en sondeos preelectorales de zonas q en anteriores votaciones han marcado tendencias de voto. *Caso del Estado de Ohio (EEUU) en relación con las elecciones a presidente en los EEUU, es un estado cambiante (swing state), no tiene una tendencia del voto. Desde comienzos del sXX sólo en 2 ocasiones falló la predicción; no coincidió el presidente elegido según los resultados obtenidos en el estado con el q el q realmente salió elegido. ¿Por q coinciden tantas veces? La respuesta viene det. por la gran diversidad q presenta su población, hay representantes de todos los tipos de votantes: de áreas rurales, de grandes ciudades, cristianos conservadores, afroamericanos y muchos trabajadores. Por tanto, es un Estado muy representativo, la gran mayoría de los dif. tipos de votantes están presentes en él.*

MUESTREO CASUAL O INCIDENTAL
Proceso en el q el investigador selecciona directa e intencionadamente los individuos de la población. El caso más frecuente es utilizar como muestra los individuos a los q se tiene fácil acceso *(profesores de universidad a sus alumnos)*.

MUESTREO DE BOLA DE NIEVE
Las unidades muestrales van incorporándose paulatinamente a partir de la referencias de los sujetos q ya han participado. Se localizan algunos individuos, los cuales conducen a otros, y estos a otros. Muy frecuente cuando se hacen estudios con poblaciones marginales, difíciles de identificar y localizar.

2. DISTRIBUCIÓN MUESTRAL DE UN ESTADÍSTICO
A una población se le mide una carac., *la altura*. Con estos datos se podrá hacer una distribución de frecuencias (los resultados se referirán a toda la población). Se podrá calcular su media y varianza. Si se hace con una muestra se obtiene la distribución muestral; se construye su distribución de frecuencias y calcula la media y varianza, q son en este caso los estadísticos, $\bar{X}$ y S^2.

Por otra parte, en una población cualquiera es posible extraer +1 muestra dif. del mismo tamaño. Por tanto, el valor concreto de un estadístico dependerá de los valores concretos q tomen cada uno de los elementos de la muestra. El estadístico obtenido ya no será una constante sino una V, su valor concreto dependerá de la muestra en la q se haya calculado. La distribución de probabilidad de todos los posibles valores del estadístico en las dif. muestras = **distribución muestral del estadístico**. Dicho de otro modo, dada una población de la q se van a extraer varias muestras, y para cada una se calcula un estadístico *(por ej la X)* de una V aleatoria X cualquiera. Puesto q la media (X) toma dif, valores, dependiendo de cada muestra, el conjunto de las distintas medias forman a su vez una VA q tendrá su propia distribución de probabilidad con sus carac.: forma, media y varianza, los parámetros q la definen q se representarán por letras griegas con un subíndice, q nos indica a q estadístico nos estamos refiriendo.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Por tanto, la **distribución muestral de un estadístico** es la distribución de probabilidad teórica de los valores de un estadístico cuando estos se calculan sobre las k muestras (siendo k muy grande, teóricamente infinito) de tamaño *n*, extraídas de la población y obtenidas mediante muestreo aleatorio simple.

Cada vez q queramos estimar un parámetro, ¿debemos extraer tantas muestras como sea posible, calcular la media o la proporción en todas esas muestras y luego obtener la media de todas las medias o la media de todas las proporciones calculadas?¿Y lo mismo con los estadísticos varianza, mediana o coeficiente de correlación de Pearson, por ej.? Para responder se retomará el concepto de esperanza matemática de una VA, la mejor opción para estimar parámetros desconocidos utilizando los modelos de probabilidad.

Si de cualquier población con media μ y desviación típica δ, se toman todas las posibles muestras aleatorias con reposición, cada una de tamaño *n*, la distribución muestral del estadístico media tiene como parámetros:

- $\mu_* = \mu$ —> la media de las medias es igual a la media poblacional.
- Desviación típica o error típico de la media $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$

En el caso de muestras finitas y muestreo sin reemplazamiento, la desviación típica deberá multiplicarse por el factor de corrección definido por $\sqrt{\frac{N-n}{N-1}}$. En este caso:

- $\mu_* = \mu$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Análisis con 10 muestras aleatorias de tamaño n = 30

10 columnas = valores observados en la V para cada muestra, desde la Muestra 1 (M1) hasta la 10 (M10).
En las 2 últimas filas figuran la media y la desviación típica en cada muestra.

M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
5	10	10	5	5	15	10	5	5	10
5	15	25	10	10	5	15	5	10	15
25	5	20	15	15	15	15	25	15	20
10	20	15	25	25	25	5	10	25	25
5	20	5	15	15	15	15	5	15	15
5	25	25	25	15	25	5	5	15	20
25	15	15	25	25	5	25	5	25	25
10	15	25	15	15	15	15	10	15	15
15	10	10	25	25	25	5	15	25	5
15	5	15	20	10	20	20	15	10	15
20	15	15	25	25	25	25	20	25	25
10	25	10	20	15	5	20	15	15	15
10	5	25	20	20	20	25	10	20	20
20	25	10	25	10	25	5	20	10	10
10	10	15	20	20	20	20	10	20	20
5	15	5	5	5	5	10	5	5	5
15	5	10	10	10	15	10	15	10	15
5	10	15	25	10	25	25	5	10	20
5	25	20	20	15	20	20	5	15	15
10	10	20	15	15	15	15	10	15	15
25	25	10	20	5	15	20	25	5	5
20	25	15	5	5	5	5	20	5	15
15	20	5	20	10	20	20	15	10	10
25	5	15	20	10	20	20	25	10	10
20	20	15	5	5	5	5	10	5	25
20	25	25	10	10	5	10	20	10	10
25	5	20	5	5	5	5	25	5	25
15	10	20	15	15	10	15	25	15	15
25	25	5	25	15	25	25	5	15	15
20	10	25	25	10	15	25	20	5	15
$\overline{x}$ 14,67	15,17	15,5	17,17	13,17	15,5	15,17	13,5	13	15,67
S_x 7,30	7,47	6,50	7,15	6,28	7,46	7,24	7,32	6,40	5,88

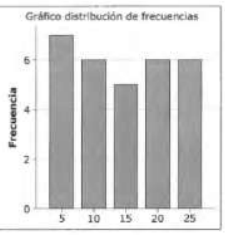
A continuación, se presentan los datos, la representación gráfica de la distribución de la variable X y los estadísticos muestrales de la MI.

Muestra 1

Distribución de frecuencias

X_i	n_i
5	7
10	6
15	5
20	6
25	6
Total	30

$\overline{X} = 14,67$
$S_X^2 = 53,22$
$S_x = 7,30$



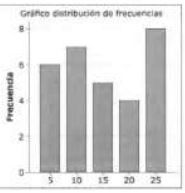
Tras repetir el exp. 9 veces +, se obtuvieron las 9 columnas siguientes de la tabla (en los gráficos se muestran las representaciones de cinco de estas muestras con los valores de sus estadísticos).

Muestra 2

Distribución de frecuencias

X_i	n_i
5	6
10	7
15	5
20	4
25	8
Total	30

$\overline{X} = 15,17$
$S_X^2 = 55,81$
$S_x = 7,47$

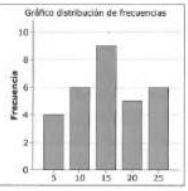


Muestra 3

Distribución de frecuencias

X_i	n_i
5	4
10	6
15	9
20	5
25	6
Total	30

$\overline{X} = 15,5$
$S_X^2 = 42,25$
$S_x = 6,5$

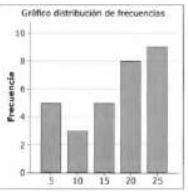


Muestra 4

Distribución de frecuencias

X_i	n_i
5	5
10	3
15	5
20	8
25	9
Total	30

$\overline{X} = 17,17$
$S_X^2 = 51,14$
$S_x = 7,15$



...

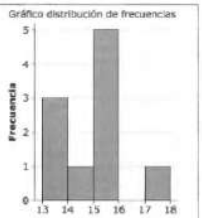
Distribución muestral de las 10 medias

Distribución de frecuencias, representación gráfica y cálculos de estadísticos correspondientes a la media y desviación típica de las 10 medias.

Distribución de frecuencias

$\overline{X}_i$	n_i
13,00	1
13,17	1
13,50	1
14,67	1
15,17	2
15,50	2
15,67	1
17,17	1
Total	10

$\mu_{\overline{X}} = 14,85$
$\sigma_{\overline{X}} = 1,23$



Como se puede observar, la media de todas las medias para 10 muestras ($\mu_* = 14,85$) se aproxima bastante a la media poblacional ($\mu = 15$). Sin embargo, la desviación típica de las medias no tiene, en principio, nada q ver con la desviación típica poblacional $1,23 \neq 7,07$ como era de esperar. No obstante, se puede comprobar q el error típico de la media de este estudio empírico es a x = 1, 23, siendo el teórico: $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{7,07}{\sqrt{30}} = 1,29$

La diferencia entre ambos se debe al nº de muestras tan reducido para construir esta distribución muestral.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Análisis con 500 muestras aleatorias de tamaño n = 30

11,00	11,50	11,90	12,00	12,00	12,10	12,20	12,20	12,20	12,30
12,30	12,50	12,50	12,50	12,70	12,90	12,90	13,00	12,50	13,17
13,17	13,17	13,17	11,50	11,50	13,17	13,20	13,20	13,20	13,20
13,20	13,20	13,20	13,20	12,50	13,20	13,20	13,20	13,20	13,20
13,20	13,20	13,20	13,20	13,20	13,20	13,20	13,20	13,20	13,20
13,20	13,20	13,23	13,23	13,23	13,30	13,30	13,30	13,44	13,50
13,50	13,50	13,50	13,50	13,50	13,50	13,50	13,50	13,50	13,50
13,50	13,50	13,50	13,50	13,50	13,50	13,50	13,50	13,50	13,50
13,55	13,60	13,60	13,60	13,67	13,67	13,70	12,50	13,70	13,70
13,70	12,50	13,70	13,70	13,70	13,70	13,70	13,70	13,80	13,80
13,90	13,90	13,90	13,90	13,90	13,90	14,00	14,00	14,00	14,00
14,00	14,00	14,00	14,00	14,00	14,00	14,00	14,10	14,10	14,10
14,11	14,17	14,17	14,17	14,17	14,20	14,20	14,20	14,20	14,20
14,20	14,20	14,20	14,20	14,20	14,20	14,20	14,20	14,20	14,20
14,20	14,20	14,20	14,20	14,20	14,20	14,20	14,21	14,30	14,30
14,30	14,30	14,30	14,30	14,30	14,30	14,40	14,40	14,44	14,50
14,50	14,50	14,50	14,50	14,55	14,60	14,60	14,60	14,60	14,60
14,60	14,66	14,66	14,66	14,67	14,67	14,67	14,70	14,70	14,70
14,70	14,70	14,70	14,70	14,70	14,70	14,70	14,70	14,70	14,70
14,70	14,70	14,70	14,70	14,70	14,70	14,70	14,70	14,70	14,70
14,70	14,80	14,80	14,80	14,80	14,80	14,80	14,80	14,80	14,80
14,80	14,80	14,80	14,80	14,80	14,80	14,80	14,80	14,80	14,90
14,90	14,90	14,90	14,90	14,90	14,90	14,90	14,90	14,90	14,90
14,90	14,90	14,90	14,90	15,00	15,00	15,00	15,00	12,50	15,00
15,00	15,00	15,00	15,10	15,10	15,10	15,10	15,10	15,10	15,10
15,10	15,10	15,10	15,10	15,10	15,10	15,10	15,10	15,10	15,12
15,17	15,17	15,17	15,17	15,17	15,17	15,20	15,20	15,20	15,20
15,20	15,20	15,20	15,20	15,20	15,20	15,20	15,20	15,20	15,20

… (faltan)

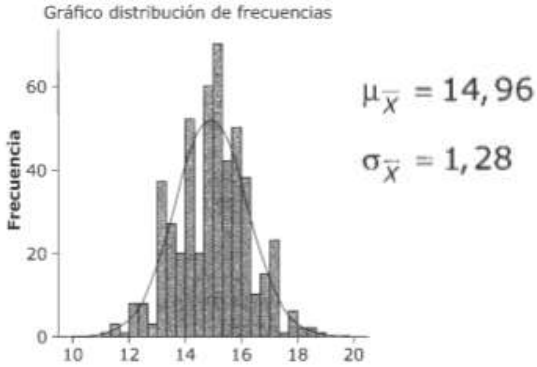


Tabla - medias de 500 muestras simuladas sobre la población antes definida (ahora sin los datos de cada muestra), cuyo tamaño muestral es de n= 30.

La media de todas las medias ($\mu_{\bar{x}}$ = 14, 96) se aproxima aún más a la media poblacional (μ = 15) y el error típico de la media ($\delta_{\bar{x}}$= 1,28) empírico tb se aproxima bastante al error típico teórico $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{7,07}{\sqrt{30}} = 1,29$

A medida q aumenta el nº de muestras, los valores de los parámetros de la distribución muestral de la media ($\mu_{\bar{x}}$ y $\delta_{\bar{x}}$) se aproximan + a la media poblacional (μ) y al error típico de la media antes definido $\left(\frac{\sigma}{\sqrt{n}}\right)$.

3. DISTRIBUCIÓN MUESTRAL DEL ESTADÍSTICO MEDIA

La **distribución muestral del estadístico media** se puede caracterizar mediante su valor esperado o media ($\mu_{\bar{x}}$), su error típico ($\delta_{\bar{x}}$) y la forma de la distribución. Para obtener esta inf., hay q tener en cuenta la distribución de la V de partida X en la población y si la varianza poblacional es conocida. 3 casos:

1. La variable X se distribuye según la curva normal con varianza conocida;
2. La variable X se distribuye de forma normal con varianza desconocida;
3. La variable X no sigue la distribución normal o no conocemos su distribución.

Distribución normal de la variable X con varianza poblacional conocida

Sea una variable aleatoria X con distribución normal y parámetros poblacionales μ y δ^2. Si se extrae un nº muy amplio de muestras de tamaño n, la distribución de sus medias tiende a una distribución normal definida como N (μ, $\delta/\sqrt{n}$) a medida que n tiende a infinito. Esto significa q:

$$\mu_{\bar{X}} = \mu$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

y q la distribución tb se ajusta a la normal. Para verificar q $\mu_{\bar{x}}$=μ y $\delta_{\bar{x}}=\frac{\sigma}{\sqrt{n}}$ se tendrán q recordar las propiedades de la media y varianza aplicándolas a la esperanza matemática (valor esperado) y varianza teórica.

La **esperanza matemática** de la VA media muestral será:

$$E(\bar{X}) = E\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}E(X_1 + X_2 + \ldots + X_n)\frac{1}{n}[E(X_1) + E(X_2) + \ldots E(X_n)] =$$

$$= \frac{1}{n}(\mu_1 + \mu_2 + \ldots + \mu_n) = \frac{1}{n}\sum \mu_i = \frac{1}{n}n\mu = \mu$$

$$\boxed{E(\bar{X}) = \mu}$$

La **varianza teórica** será:

$$\sigma^2_{\bar{X}} = \frac{1}{n^2}\sigma^2_{X_1+X_2+\ldots+X_n} = \frac{1}{n^2}\sigma^2_{X_1} + \frac{1}{n^2}\sigma^2_{X_2} + \ldots + \frac{1}{n^2}\sigma^2_{X_n} =$$

$$= \frac{1}{n^2}\sum \sigma^2_n = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$$

Desviación típica o error típico:

$$\boxed{\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}}$$

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Por tanto, la distribución muestral de la media $\bar{X}$ es: $\bar{X} \rightarrow N(\mu,\delta/\sqrt{n})$ q se lee: Normal con media μ y error típico $\frac{\sigma}{\sqrt{n}}$
Consecuentemente, ya q se trata de una distribución normal, podremos tipificar la variable X; calcular su puntuación Z_x y la distribución sigue siendo normal.

$$Z_{\bar{X}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$ q, como se sabe, se distribuye N(0,1).

Supuestos:
- La variable X; se distribuye según la normal.
- Se conoce δ^2
- Se extraen muestras de tamaño n mediante muestreo aleatorio simple.

Entonces la distribución muestral de las medias forma una VA q se distribuye $N(\mu,\delta/\sqrt{n})$ ¿Q ventaja aporta saberlo? Q permite aplicar todo lo q se conoce de la distribución normal = hacer todos los cálculos relativos a las probabilidades correspondientes a los valores de la media, aplicando el concepto de tipificación.

A medida q ↑ el tamaño de la muestra, el error típico de la media disminuye. En la distribución muestral de la media el error típico es inversamente proporcional al tamaño de la muestra (n), por lo tanto, a medida q ↑*n* la distribución muestral de las medias se hace más homogénea (presenta menor variabilidad).

Distribución normal de la variable X con varianza poblacional desconocida
En el caso anterior se parte del hecho de q se conoce la varianza poblacional, lo cual no es muy frecuente. En estos casos, se deberá estimar el valor del error típico de la media mediante la cuasidesviación típica muestral dada por:

$$\sigma_{\bar{X}} = \frac{S_{n-1}}{\sqrt{n}}$$

En estas condiciones la distribución muestral de la media ya no se ajusta a la normal, sigue la distribución t de Student. Al no conocer la varianza poblacional, la transformación viene dada ya por la distribución:

$$T = \frac{\bar{X} - \mu}{S_{n-1}/\sqrt{n}},$$ q sigue una distribución t de Student con n - 1 gl.

Supuestos:
- La variable X_i se distribuye $N(\mu, \delta)$
- Se desconoce el valor de δ^2
- Extraemos muestras mediante muestreo aleatorio simple de tamaño *n*.

Entonces, la distribución muestral de las medias forma una VA q se distribuye $t\left(\mu, S_{n-1}/\sqrt{n}\right)$ con n-1 grados de libertad.

La variable X no se distribuye normalmente
Por lo general, las situaciones anteriores no son habituales. Lo que ocurre generalmente es q las V no se ajustan al modelo de la normal o se desconoce su varianza poblacional. En estas situaciones, la Estadística aporta el ***Teorema del Límite Central***, q permite calcular las probabilidades asociadas a los valores de las medias sin necesidad de conocer la forma de la distribución de las V, siempre q las muestras tengan tamaño suficiente (n ~ 30).

Sea $X_1, X_2,..., X_n$ un conjunto de VA, indep. e idénticamente distribuidas con media μ_i y varianza $\delta^2_i \neq 0$ Si n es suficientemente grande ($n \geq 30$) la distribución muestral de la media de las X_i se aproxima a la distribución normal a medida q n aumenta, independientemente de las distribuciones q presenten $X_1, X_2,..., X_n$.

Este teorema establece q sí n es suficientemente grande, las V q se combinan son indep., tienen distribuciones idénticas y valor esperado y varianza finitas, entonces la distribución muestral del estadístico tiende a $N(\mu,\delta/\sqrt{n})$. En caso de no conocer δ, se utiliza como estimador la cuasidesvación típica con $X \rightarrow N(\mu, S_{n-1}/\sqrt{n})$

4. DISTRIBUCIÓN MUESTRAL DEL ESTADÍSTICO PROPORCIÓN

En Psicología y en Ciencias de la Salud son muy habituales estudios en los q están involucradas una o varias proporciones (o porcentajes) medidas en alguna V de interés *(proporción de universitarios varones frente a mujeres adictos a la cocaína)*. Para poder hacer los pertinentes estudios deberemos conocer la distribución muestral de este estadístico, distinguiendo entre muestras pequeñas y suficientemente grandes.

Sea una población en la q se mide una V q solo puede tomar 2 valores: éxito (1) o fracaso (0) (V q como sabemos se ajusta al modelo de Bernoulli). Definimos π como la proporción de aciertos en la población. Si extraemos todas las posibles muestras de tamaño *n* y medimos en cada una de ellas la variable aleatoria X= nº de éxitos en las n extracciones, y sea P = proporción de éxitos en las *n* extracciones, constante en todas las muestras, es decir, $P_1 = P_2 = ... = P_n = \pi$. Entonces, podemos definir las distribuciones muestrales de X y P según las muestras sean pequeñas o grandes.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Distribución muestral del estadístico P para muestras pequeñas

n < 30, la distribución muestral del estadístico X (nº de éxitos en *n* ensayos) es $X \rightarrow B\left(n\pi, \sqrt{n\pi(1-\pi)}\right)$

donde: $\mu_X = E(X) = n\pi$ y $\sigma_X = \sqrt{n\pi(1-\pi)}$. Dado q P es una mera transformación lineal de X(P=X/n), se demuestra q la distribución muestral del estadístico P (proporción de aciertos en n ensayos) es $P \rightarrow B\left(\pi, \sqrt{\frac{\pi(1-\pi)}{n}}\right)$ donde: $\mu_P = E(P) = \pi$ y $\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}}$

En efecto, siendo P = X/n, entonces:

$$E(P) = \frac{1}{n} n\pi = \pi$$

** se ha cambiado la notación y, en lugar de llamar P a la probabilidad de éxito en la población, la denominamos n (letra griega correspondiente a P) .*

La varianza: $\sigma_P^2 = \frac{\pi(1-\pi)}{n}$ cuya raíz o **error típico de medida** es:

$$\sigma_P = \sigma\left(\frac{1}{n}X\right) = \frac{1}{n}\sigma_X = \frac{1}{n}\sqrt{n\pi(1-\pi)} = \sqrt{\frac{\pi(1-\pi)}{n}}$$

En la tabla de la binomial se pueden determinar las probabilidades para diferentes tamaños muestrales y valores de π.
Supuestos:

- La VA X es una variable Bernoulli (solo 2 valores éxito o fracaso).
- Se conoce π proporción en la población.
- Las *n* observaciones son independientes.

Entonces: La distribución muestral de la variable P es Binomial, definida por $B\left(\pi, \sqrt{\frac{\pi(1-\pi)}{n}}\right)$

Distribución muestral del estadístico P para muestras suficientemente grandes

Por regla general, las muestras suelen ser grandes. Por el Teorema del Límite Central se sabe q, a medida q *n* crece, la distribución de las proporciones se aproxima a la distribución normal con parámetros:

$$\mu_P = E(P) = \pi$$

$$\sigma_P^2 = \frac{\pi(1-\pi)}{n}$$

Aplicando la tipificación tendremos:

$$Z = \frac{P-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

Supuestos:

- La VA X es una variable Bernoulli (solo dos valores éxito o fracaso).
- Se conocen π proporción en la población (o el valor P en la muestra)
- Las *n* observaciones son independientes.
- La muestra es suficientemente grande, es decir, n ≥ 30. Si n < 30 deberá cumplirse: n x π ≥ 5 y n x (1- π) ≥ 5

Entonces, los valores de las proporciones se aproximan a una normal definida por: $P \rightarrow N\left(\pi, \sqrt{\frac{\pi(1-\pi)}{n}}\right)$

5. DISTRIBUCIÓN MUESTRAL DEL ESTADÍSTICO VARIANZA

El razonamiento para obtener la distribución muestral de la varianza es el mismo q para la media y la proporción.
Si X es una variable que se distribuye en la población N(μ,δ), se extraen todas las muestras posibles de tamaño *n* con media X, varianza S^2_x y cuasivarianza S^2_{n-1}.

Entonces, las variables aleatorias $\frac{nS_X^2}{\sigma^2}$ y $\frac{(n-1)S_{n-1}^2}{\sigma^2}$ siguen una distribución x^2 con n-1 grados de libertad. Es decir, ambas V se distribuyen según X^2_{n-1}(n -1,), por lo q: μ =n- 1 y desviación típica, δ = $\sqrt{2(n-1)}$

Conociendo la distribución se deduce q las distribuciones muestrales de la varianza y cuasivarianza son:

La **distribución muestral del estadístico varianza (S^2_x)**, sigue una distribución X^2_{n-1} con parámetros: $\mu_{S_x^2} = E\left(S^2\right) = \frac{n-1}{n}\sigma^2$

Y error típico: $\sigma_{S_x^2} = \sigma^2 \frac{\sqrt{2(n-1)}}{n}$

La **distribución muestral del estadístico cuasivarianza (S^2_{n-1})** - distribución X^2_{n-1} con parámetros: $\mu_{S_{n-1}^2} = E\left(S_{n-1}^2\right) = \sigma^2$

Y error típico: $\sigma_{S_{n-1}^2} = \sigma^2\sqrt{\frac{2}{n-1}}$

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

En efecto, en una distribución x^2 con n-1 grados de libertad, la esperanza es $E(x^2_{n-1}) = n-1$ y la varianza igual a $V(X^2_{n-1}) = 2(n-1)$. Así podemos demostrar las fórmulas anteriores, recordando las propiedades de la media y la varianza.
En cuanto a la varianza muestral:

$$\frac{nS_X^2}{\sigma^2} = \chi^2_{n-1} \rightarrow S_X^2 = \frac{\sigma^2\chi^2_{n-1}}{n}$$

Calculamos la esperanza y la varianza:

$$E\left(S_X^2\right) = E\left(\frac{\sigma^2\chi^2_{n-1}}{n}\right) = \frac{\sigma^2}{n}E\left(\chi^2_{n-1}\right) = \frac{\sigma^2}{n}(n-1)$$

$$V\left(S_X^2\right) = V\left(\frac{\sigma^2\chi^2_{n-1}}{n}\right) = \frac{\sigma^4}{n^2}V\left(\chi^2_{n-1}\right) = \frac{\sigma^4}{n^2}2(n-1) \rightarrow \sigma_{S^2} = \sqrt{\frac{\sigma^4}{n^2}2(n-1)} =$$

$$= \sigma^2\frac{\sqrt{2(n-1)}}{n}$$

En cuanto a la cuasivarianza muestral:

$$\frac{(n-1)S^2_{n-1}}{\sigma^2} = \chi^2_{n-1} \rightarrow S^2_{n-1} = \frac{\sigma^2\chi^2_{n-1}}{n-1}$$

Calculamos la esperanza y la varianza:

$$E\left(S^2_{n-1}\right) = E\left(\frac{\sigma^2\chi^2_{n-1}}{n-1}\right) = \frac{\sigma^2}{n-1}E\left(\chi^2_{n-1}\right) = \frac{\sigma^2}{n-1}(n-1) = \sigma^2$$

$$V\left(S^2_{n-1}\right) = V\left(\frac{\sigma^2\chi^2_{n-1}}{n-1}\right) = \frac{\sigma^4}{(n-1)^2}V\left(\chi^2_{n-1}\right) = \frac{\sigma^4}{(n-1)^2}2(n-1) = \frac{2\sigma^4}{n-1} \rightarrow \sigma_{S^2_{n-1}} =$$

$$= \sqrt{\frac{2\sigma^4}{n-1}} = \sigma^2\sqrt{\frac{2}{n-1}}$$

Lo mostrado determina los valores de los parámetros de las distribuciones muestrales de la varianza y cuasivarianza. Para calcular las probabilidades asociadas a las varianzas, mediante la tabla de ji-cuadrado, se utilizan directamente las expresiones:

$$\boxed{\chi^2 = \frac{nS_X^2}{\sigma^2}} \quad y \quad \boxed{\chi^2 = \frac{(n-1)S^2_{n-1}}{\sigma^2}}$$

q se distribuyen $X^2_{n-1}(n-1, \sqrt{2(n-1)})$, siempre y cuando la variable X siga la distribución normal $N(\mu,\delta)$ en la población.

Aproximación a la normal de la distribución muestral de la varianza.
Cuando la muestra es suficientemente grande, n= 100, la distribución de las VA se aproxima a la normal con media $\mu = n-1$ y desviación típica, $\delta = \sqrt{2(n-1)}$

En este caso podemos tipificar las V, obteniendo:

$$\sigma = \sqrt{2(n-1)}.$$

Varianza:

$$\boxed{Z = \frac{\left(\frac{nS_X^2}{\sigma^2}\right) - (n-1)}{\sqrt{2(n-1)}}}$$

Cuasivarianza:

$$\boxed{Z = \frac{\left(\frac{(n-1)S^2_{n-1}}{\sigma^2}\right) - (n-1)}{\sqrt{2(n-1)}}}$$

Ambas con distribución aproximada N(0,1), q facilita la búsqueda de las probabilidades asociadas en la tabla de la Normal.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
WUOLAH Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

T10: ESTIMACIÓN DE PARÁMETROS Y CÁLCULO DEL TAMAÑO MUESTRAL

En el T9 se han estudiado el muestreo y el concepto de distribución muestraL de un estadístico; base de la Inferencia Estadística, q permite derivar a partir de los resultados de la muestra, los resultados q con una cierta probabilidad se pueden determinar para la población. Para llevar a cabo su proceso de análisis, la Inferencia combina los modelos de probabilidad con los estadísticos. La aplicación supone q el investigador se plantea preguntas con contenido teórico. Una vez formuladas las preguntas y evaluadas en relación con los conocimientos previos (a través del estudio y análisis crítico de la bibliografía), se traducen a términos estadísticos. Con estas formulaciones estadísticas se comprueba si la situación planteada se parece a algún modelo de los q nos ofrece la Estadística y la Probabilidad. Si es así, obtendremos una R estadística a la pregunta, q debe llenarse con el contenido teórico q suscitó la pregunta inicial.

Esta forma de trabajo exige garantizar q la realidad q se investiga y los modelos matemáticos (probabilístico y/o estadístico) q se aplican son similares. Es decir, q la realidad representada en el modelo teórico se ajusta al modelo matemático elegido para dar la R estadística. Esto significa q la comparación entre la pregunta estadística (expresión matemática del modelo teórico) y los modelos de probabilidad y/o estadísticos debe hacerse cumpliendo las exigencias matemáticas o supuestos referentes a :

- La métrica de las V o nivel de medida: cuantitativas, cuasi-cuantitativas, cualitativas y sus correspondientes escalas de medida, donde se pueden establecer det. relaciones matemáticas (igualdad en la escala nominal; orden de las posiciones en la escala ordinal; magnitud del intervalo, es decir, igualdad o desigualdad de diferencias en la escala de íntervalo; e igualdad o desigualdad de razones en la escala de razón). Estas operaciones matemáticas permiten cada una realizar unos det. cálculos.
- La forma de la distribución de la V: Bernoulli, Binomial, Nor- mal, Student…

Para entender mejor lo descrito anteriormente, se va a desarrollar el siguiente ejemplo.

La **Inferencia Estadística** tiene 2 ramas:

- Estadística **paramétrica**: la distribución de las V en la población es conocida (normal, binomial…), la muestra se selecciona por muestreo aleatorio simple y los datos están medidos al menos en escala de intervalos. Para cumplir con sus objetivos emplea 2 procedimientos -se basan en el conocimiento teórico de la distribución muestral del estadístico correspondiente al parámetro o parámetros q se quieren estimar-:
  - Estimación de parámetros: asignar un valor numérico o determinar un intervalo de valores numéricos al parámetro q deseamos conocer. Permite hacer conjeturas del tipo: *si en una muestra seleccionada al azar de una población en la q se ha medido una variable X_i, siendo la media, $\bar{X}$, ¿cuál será el valor más próximo o el intervalo de valores entre los q se encuentre el valor de media de la población μ con un cierto grado de confianza?*
  - Contraste de hipótesis: tiene como objetivo comprobar si un det. supuesto, referido a un parámetro o pará- metros poblacionales, es compatible con la evidencia empírica q nos proporciona la muestra. Responde a preguntas del tipo: *¿es el valor de la media poblacional de una variable X; un valor determinado según el grado de confianza q consideramos suficiente?*

    * Ambos se basan en los mismos modelos probabilísticos y estadísticos. Sin embargo, en la estimación de parámetros se parte de los datos muestrales para responder a una pregunta sobre la población; en el contraste se hace una afirmación sobre la población q luego se contrasta.
- Estadística **no paramétrica**: la distribución de las V no se ajusta a ninguna distribución conocida o los datos están medidos en una escala inferior a la escala de intervalo.

1. ESTIMACIÓN DE PARÁMETROS

Básicamente, inferir el valor desconocido de un parámetro. 4 tipos de estimaciones:

- **Estimación puntual**. Procedimiento mediante el cual asignamos un único valor al parámetro desconocido, a partir del resultado obtenido en una muestra. *Tras la aplicación de un programa de intervención dental para niños, encontramos q el 60% (P =0,60) se lavan los dientes 3 veces al día. Una estimación puntual nos llevaría a suponer q la proporción π en la población de niños q se lavarían los dientes 3 veces al día si participasen en el programa sería: $\boldsymbol{\pi} = 0{,}60$.*
- **Estimación por intervalos**. Daremos un rango de posibles valores, dentro del cual estimamos se encuentra el verdadero valor del parámetro con un det. grado de confianza. *Podríamos afirmar q después de participar en el programa de intervención, la proporción se encuentra entre 0,50 y 0,70; $0{,}50 < \pi < 0{,}70$ con un cierto margen de confianza.*
- **Estimación Bayesiana**. En lugar de considerar a los parámetros como constantes, se presentan como VA con una cierta distribución a priori. Las observaciones o datos aportan inf. q transforman las probabilidades a priori en probabilidades a posteriori.
- **Estimación Bootstrap**. Se basan en el remuestreo y técnicas de simulación = requieren el uso de ordenadores. Extraer de una misma muestra varias (muchas) muestras y estudiar el conjunto de las obtenidas. Se puede asimilar a un muestro aleatorio simple con reposición q se realizase en una población de un tamaño pequeño. De cada muestra extraída se calcula el estadístico de interés y se estudia su distribución.

Antes de estudiarlas se plantean una serie de cuestiones sobre si un estadístico es un buen estimador de un parámetro o no, como: ¿sirve cualquier estadístico para estimar un parámetro? ¿cualquier estadístico es un buen estimador? *¿La media es un buen estimador de la media poblacional?, ¿La proporción muestral es un buen estimador de la proporción poblacional?, ¿La varianza muestral es un buen estimador de la varianza poblacional?*

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Para q un estadístico pueda considerarse un buen estimador de un parámetro deberá; Carecer de sesgo, tener eficiencia, consistencia y suficiencia.

Propiedades de los estimadores

Carencia de sesgo

Sea θ el parámetro a estimar y $\hat{\theta}$ el valor del estimador (valor obtenido en la muestra), diremos q $\hat{\theta}$ es un **estimador insesgado** o carente de sesgo si E($\hat{\theta}$) = θ para cualquier valor de θ. Es decir, se cumple q la media de la distribución muestral (esperanza matemática de la distribución) coincide con el parámetro estimado.

$$E(\hat{\theta}) = \theta$$

Comprobación de si los principales estimadores (media, desviación típica, varianza y cuasivarianza) la cumplen:

- ¿Es la media un estimador insesgado de la media poblacional?

El estadístico media sigue una distribución normal o una t de Student (según conozcamos o no la varianza poblacional).

Conocida σ

$$\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

No conocida σ

$$\bar{X} \to t\left(\mu, \frac{S_{n-1}}{\sqrt{n}}\right)$$

En cualquiera de las 2 distribuciones E($\bar{x}$) = μ. Por lo tanto, es un estimador insesgado de la media poblacional.

- ¿Es la proporción es un estimador insesgado de la proporción poblacional? Es decir, ¿E(P) = π?

La distribución muestral de la proporción se define como:

$$P \to B\left(\pi, \sqrt{\frac{\pi(1-\pi)}{n}}\right)$$

Como se observa, E(P) = π por lo q P es un estimador insesgado de *n*.

- ¿Son la varianza y la cuasivarianza estimadores insesgados de la varianza poblacional?

Según se desprende de su distribución muestral: $E\left(S_X^2\right) = \frac{n-1}{n}\sigma^2$

Es obvio q la esperanza matemática de la varianza de la muestra, S^2_x no es exactamente su correspondiente valor poblacional δ^2, por lo q decimos que se trata de un estimador sesgado, siendo precisamente su sesgo el factor n-1/n.

Por el contrario, la cuasivarianza sí es un estimador insesgado, $E(S^2_{n-1}) = \delta^2$. Si queremos realizar una estimación puntual de δ^2, es preferible utilizar la cuasivarianza.

Eficiencia

Dados dos estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$ del mismo θ (parámetro), diremos q. $\hat{\theta}_1$ es más eficiente si su varianza (la de su distribución muestral) es menor. Es decir: $\sigma^2_{\hat{\theta}_1} < \sigma^2_{\hat{\theta}_2}$

Por tanto, entre 2 estimadores insesgados será preferible seleccionar el q presente una menor varianza (menor error típico de la distribución muestral del estadístico). El error típico refleja el mayor o menor alejamiento de los posibles valores del estadístico a su esperanza matemática (media de la distribución muestral). Un estimador es tanto mejor cuanto su distribución muestral esté más concentrada = varianza más pequeña.

Por ej, en relación con la varianza y la cuasivarianza, los errores típicos de ambos son: $\sigma_{S_X^2} = \sigma^2\frac{\sqrt{2(n-1)}}{n}$ y $\sigma_{S^2_{n-1}} = \sigma^2\sqrt{\frac{2}{n-1}}$

El de la cuasivarianza es mayor, por lo q diremos q la varianza es un estimador + eficiente q la cuasivarianza.

La eficiencia de un estimador siempre es relativa, ninguno puede ser perfectamente eficiente, dado q el error típico acompaña a cualquier distribución muestral. Se define la **eficiencia relativa, ER**, de un estimador $\hat{\theta}_1$ con respecto a otro $\hat{\theta}_2$, como la razón:

$$ER = \frac{\sigma_{\hat{\theta}_1}}{\sigma_{\hat{\theta}_2}}$$

- Cociente =1, ambos estimadores son igualmente eficientes.
- ER > 1, el estimador del denominador es más eficiente.
- ER < 1, el estimador del numerador es más eficiente.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
WUOLAH Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Consistencia

$$\lim_{n\to\infty} P\left(\left|\hat{\theta}-\theta\right|<\delta\right)=1$$

Indica q a medida q el tamaño muestral se hace grande (q tiende a infinito), el valor del estadístico se aproxima al valor del parámetro. Por tanto, un estimador es consistente cuando la probabilidad de q su valor se acerque al del parámetro es mayor a medida q aumenta el tamaño de la muestra. En otras palabras, si *n* tiende a infinito, la probabilidad de q $\left|\hat{\theta}-\theta\right|$ sea menor q cualquier valor **δ**, por pequeño q sea éste, tiende a 1.

Suficiencia

Se refiere a la capacidad del estimador de utilizar toda la inf. existente en la muestra en relación al parámetro; q el estimador emplee todos los valores de los datos (inf.) para estimar el parámetro.
Sabemos q la media de la muestra, $\bar{X}$, es un buen estimador de la media poblacional, μ. Tb podríamos utilizar otros estimadores como: la mediana, el promedio de los valores extremos de la distribución, la media de los cuartiles primero y tercero… Sin embargo, basta con observar las fórmulas para darse cuenta de q la media es un estimador suficiente.

Así, si utilizamos el promedio de los extremos de la distribución sería: $\frac{X_{i_{sup}}+X_{i_{inf}}}{2}$ sólo empleamos en el cálculo el valor de la puntuación más alta y + baja. Mientras q en la media $\bar{X}=\frac{\sum X_i}{n}$ es evidente q empleamos todos los valores de X_i. Así pues, el estimador suficiente de μ es $\bar{X}$.

Del mismo modo, la varianza y la varianza insesgada, así como la proporción son estimadores suficientes de δ^2 y *π*, respectivamente.

Como resumen, las **propiedades de los estimadores media**, **proporción** y **cuasivarianza** son:

1. La media muestral se considera buen estimador de la media poblacional ($\bar{X}$= μ). Cumple las propiedades de ser insesgado, consistente y suficiente.
2. La proporción muestral (*P*) se considera buen estimador de la proporción poblacional (P=*π*), ya q cumple las propiedades de ser insesgado, consistente y suficiente.
3. La varianza muestral (S^2_x) cumple las propiedades de ser consistente y suficiente pero no es insesgado. Por esta razón, la cuasivarianza se considera un buen estimador de la varianza poblacional $\left(\hat{\sigma}^2=S^2_{n-1}\right)$. Cumple las propiedades de ser insesgada, consistente y suficiente, aunq su eficiencia (q compara dos estimadores) es menor en relación con la varianza muestral.

Métodos de obtención de estimadores

Una vez ya se sabe q propiedades han de cumplir los estadísticos para ser empleados como estimadores, se estudia cómo se estima el parámetro = cómo se determina su valor o los valores entre los q se encuentra.

Existen varios métodos q garantizan las propiedades antes enunciadas y se aplican en circunstancias específicas; 2 los + empleados para los q habitualmente utilizamos en Psicología;

- **Método de mínimos cuadrados**: trata de obtener aquel estimador q minimice las distancias (al cuadrado) entre el valor estimado del parámetro y los resultados muestrales observados. No siempre es el mejor pero resulta muy útil para estimar los parámetros de la regresión, por ej.

$$\sum\left(X_i-\hat{\theta}\right)^2 \text{ sea mínimo, donde } i=1,\ 2,\ \ldots,\ n$$

- **Método de máxima verosimilitud**: obtiene como estimador de un parámetro aquel valor del estadístico q hace lo más verosímil posible la muestra obtenida; se trata de elegir, de entre todos los posibles valores del parámetro, aquel q maximice la probabilidad de obtener el resultado particular observado en la muestra.

Estimación puntual

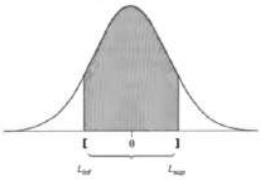
Independientemente del método escogido y partiendo de q el estimador seleccionado cumple las propiedades, la estimación puntual consiste en dar un valor numérico único al parámetro desconocido; utilizar el valor del estadístico para estimar el parámetro.
No deja de tener inconvenientes. Teniendo en cuenta el elevado nº de muestras q podemos extraer de la población y q de cada una de ellas podemos realizar una estimación, el nº de estimaciones podría resultar excesivo.
Además, aun cuando la muestra sea representativa, no se puede establecer ni la fiabilidad de la estimación ni el error q se comete. En estos casos, lo único q podemos afirmar es q el error cometido en la estimación se hará menor a medida q aumente el tamaño de la muestra. En definitiva, la estimación puntual no siempre es la más aconsejable y útil.

Estimación por intervalos

Obtener una medida del error (diferencia entre el estimador y el parámetro) q se comete al realizar la estimación con una det. probabilidad; es atribuir al parámetro un rango de valores posibles dentro del cual estará incluido el parámetro con una det. probabilidad.
Mediante la estimación por intervalos, en lugar de un solo valor (como en la estimación puntual), obtenemos un rango de posibles valores del parámetro, **intervalo de confianza** y sus límites, **límites del intervalo de confianza**.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material



Los corchetes indican los límites del intervalo y la llave los posibles valores estimados del parámetro θ. A la zona sombreada se le denomina **nivel de confianza (n.c.)** = probabilidad asociada al intervalo q contiene todos los posibles valores q puede tomar el parámetro θ. Se le llama de confianza y no de probabilidad ya q una vez extraída la muestra, contendrá al verdadero valor del parámetro o no. Lo q sabemos es q si repitiésemos el proceso con muchas muestras podríamos afirmar q el (1-α)% de los intervalos así construidos contendría al verdadero valor del parámetro. α es el nivel de significación y hace referencia a la cuantía del margen de error q se asume a priori.

Principal ventaja; se puede valorar la seguridad con la q se realizan las estimaciones mediante el nivel de confianza, el cual se expresa en términos de probabilidad.

Algunas carac.:

- **Relación entre amplitud del intervalo y nivel de confianza**

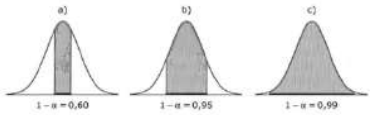


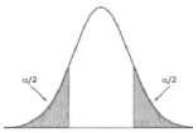
Figura 10.2. Distribución con nivel de confianza: a) 0,60; b) 0,95; c) 0,99.

<— 3 distribuciones en las q se han establecido 3 intervalos de confianza (zona sombreada) q van aumentado a medida q aumenta 1-α. Cuanto mayor es el intervalo de valores mayor es la probabilidad de q se encuentre dentro de él el verdadero valor de θ y la estimación es menos precisa = la precisión de la estimación se relaciona de forma inversa con el nivel de confianza, a mayor confianza, mayor intervalo (el rango de posibles valores del parámetro) = estimación - precisa. A mayor confianza menor precisión.

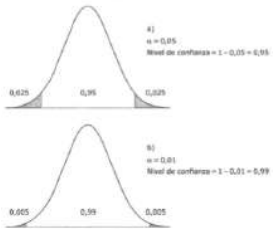
- **Fijación del nivel de confianza**

El investigador es quien decide y fija el valor del nivel de confianza en función de la valoración personal q hace sobre diversos aspectos: diseño de su trabajo, definición y obtención de la muestra, recogida de inf… Por convenio, en general se adoptan los niveles del confianza de 1-α = 0,95 ó 1-α = 0,99.

- **Nivel de riesgo o significación α**



El opuesto al nivel de confianza = **nivel de riesgo**, **margen de error** o **nivel de significación** y se representa por α. Indica la probabilidad de q el valor del parámetro no se encuentre dentro de los límites. α se reparte entre los 2 extremos de la curva q delimitan el intervalo de confianza (1-α). Es decir, q el margen de error se divide en 2 partes iguales siendo el área correspondiente a cada una α/2 (zonas sombreadas). Entre el nivel de confianza (1-α) y el de riesgo o significación a existe una relación inversa.



2. CÁLCULO DEL INTERVALO DE CONFIANZA

Para calcular el intervalo de confianza de un parámetro cualquiera es necesario conocer la distribución muestral del estadístico correspondiente y los parámetros q la definen. Es decir, la esperanza matemática y el error típico.

En general, para construir un intervalo de confianza para un parámetro se suma y se resta al estimador, θ el error máximo de estimación, E_{max}. Para comprender bien como se define el intervalo de confianza se retoma el concepto de puntuación Z (T3) ya q, junto al error de típico (T9) son los elementos necesarios para entender cómo se construyen los intervalos de confianza y su significado .

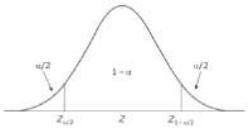
Intervalo de confianza para el parámetro μ con δ^2 conocida

Si la VA X sigue una distribución normal en la población, y se conoce la varianza poblacional, la distribución muestral de la media $\overline{X}$ es $\overline{X} \rightarrow N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$, siendo $E(\overline{X}) = \mu$ y $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$.

Ya q la distribución muestral de la media es normal, se tipifica sin más que aplicar la transformación a Z:

$$Z = \frac{\overline{X} - E(\overline{X})}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$ donde Z es N(0,1).

Acudiendo a las tablas de la distribución normal se puede calcular la probabilidad de q la variable Z se encuentre entre 2 valores concretos. Si a se corresponde al margen de error (o nivel de significación q se ha fijado) tendremos la representación —>



$$P\left(Z_{\alpha/2} \le Z \le Z_{1-\alpha/2}\right) = 1 - \alpha$$

Sustituyendo Z por su valor: $$P\left(Z_{\alpha/2} \le \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \le Z_{1-\alpha/2}\right) = 1 - \alpha$$

A partir de esta expresión, el objetivo es determinar los intervalos para μ: $$P\left(Z_{\alpha/2}\, \sigma/\sqrt{n} \le \overline{X} - \mu \le Z_{1-\alpha/2}\, \sigma/\sqrt{n}\right) = 1 - \alpha$$

Restando la media X en todos los términos: $$P\left(-\overline{X} + Z_{\alpha/2}\, \sigma/\sqrt{n} \le -\mu \le -\overline{X} + Z_{1-\alpha/2}\, \sigma/\sqrt{n}\right) = 1 - \alpha$$

multiplicando por -1: $$P\left(\overline{X} - Z_{\alpha/2}\, \sigma/\sqrt{n} \le \mu \le \overline{X} - Z_{1-\alpha/2}\, \sigma/\sqrt{n}\right) = 1 - \alpha$$

teniendo en cuenta q $Z_{\alpha/2} = -Z_{1-\alpha/2}$: $$\boxed{P\left(\overline{X} - \left|Z_{\alpha/2}\right| \sigma/\sqrt{n} \le \mu \le \overline{X} + \left|Z_{\alpha/2}\right| \sigma/\sqrt{n}\right) = 1 - \alpha}$$

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

donde los límites del intervalo son: $L_{inf} = \overline{X} - |Z_{\alpha/2}| \sigma/\sqrt{n}$ $L_{inf} = P - |Z_{\alpha/2}| \sqrt{\frac{P(1-P)}{n}}$

La **semi-amplitud** (mitad de la amplitud) del intervalo de confianza se denomina **Error máximo de estimación**, siendo su valor $E_{max} = |Z_{\alpha/2}| \sigma/\sqrt{n}$. Indica q el investigador asume con un nivel de confianza del (1-α)% q la diferencia máxima entre el valor estimado a partir de la muestra y el valor real del parámetro es igual a:

$$E_{max} = |Z_{\alpha/2}| \frac{\sigma}{\sqrt{n}}$$

La obtención de los intervalos correspondientes a la media, conocida la varianza poblacional y siendo normal la distribución de la V es aplicable tb al caso en el q la V no siga la distribución normal siempre y cuando el tamaño de la muestra sea grande ($n > 30$). Como se vio en el T9, el Teorema del Límite Central establece q la distribución muestral de la media se aproxima a la distribución normal a medida q el tamaño de la muestra va aumentando (en la práctica con n > 30 el ajuste es bastante bueno) sin q necesariamente la VA X tenga una distribución normal. De este modo, cuando el tamaño muestral sea grande y queramos estimar la media poblacional, podemos utilizar el intervalo antes definido.

Intervalo de confianza para el parámetroμ con δ² desconocida

Si la VA X tiene distribución normal en la población, pero la varianza es desconocida, sabemos del T9 q la distribución muestral de la media X sigue la distribución t de Student definida por:

$$\overline{X} \to t\left(\mu; \frac{S_{n-1}}{\sqrt{n}}\right) \text{ con } n-1 \text{ grados de libertad.}$$

Esto implica q. $T = \frac{\overline{X} - \mu}{S_{n-1}/\sqrt{n}}$, tb sigue una distribución t con n-1 g.l.

El intervalo de confianza para T sería: $P\left(t_{n-1;\alpha/2} \le T \le t_{n-1;1-\alpha/2}\right) = 1 - \alpha$

y sustituyendo en *T*: $P\left(t_{n-1;\alpha/2} \le \frac{\overline{X} - \mu}{\frac{S_{n-1}}{\sqrt{n}}} \le t_{n-1;1-\alpha/2}\right) = 1 - \alpha$

Siguiendo el mismo razonamiento q en el caso anterior (conocida) se tiene:

$$P\left(\overline{X} - |t_{n-1;\alpha/2}| \frac{S_{n-1}}{\sqrt{n}} \le \mu \le \overline{X} + |t_{n-1;\alpha/2}| \frac{S_{n-1}}{\sqrt{n}}\right) = 1 - \alpha$$

Los límites del intervalo de confianza son:

$$L_{inf} = \overline{X} - |t_{n-1;\alpha/2}| \frac{S_{n-1}}{\sqrt{n}} \qquad L_{sup} = \overline{X} + |t_{n-1;\alpha/2}| \frac{S_{n-1}}{\sqrt{n}}$$

En caso de q la variable X tenga una distribución desconocida y el tamaño muestral sea n2 30, según el Teorema del Límite Central, la distribución muestral de la media es normal $\overline{X} \to Z\left(\mu; \frac{S_{n-1}}{\sqrt{n}}\right)$

Por tanto, los límites de los intervalos en esta aproximación a la normal:

$$L_{inf} = \overline{X} - |Z_{\alpha/2}| \frac{S_{n-1}}{\sqrt{n}} \qquad L_{sup} = \overline{X} + |Z_{\alpha/2}| \frac{S_{n-1}}{\sqrt{n}}$$

Intervalo de confianza para el parámetro *π* (aproximación a la normal)

La distribución muestral del estadístico P cuando se cumple q la muestra es grande ($n \ge 30$; o $nP \ge 5$ y $n(1-P) \ge 5$) es *N*(0,1) aprox., por lo q $Z = \frac{P - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$. A partir de la tipificación de *P* podemos construir el **intervalo de confianza** de la siguiente manera:

$$P\left(Z_{\alpha/2} \le \frac{P - \pi}{\sqrt{\frac{P(1-P)}{n}}} \le Z_{1-\alpha/2}\right) = 1 - \alpha$$

Aplicando el mismo razonamiento q en los demás estadísticos y, haciendo las correspondientes transformaciones, se tiene q la probabilidad de obtener un intervalo de confianza q contenga el parámetro es:

$$P\left(P - |Z_{\alpha/2}| \sqrt{\frac{P(1-P)}{n}} \le \pi \le P + |Z_{\alpha/2}| \sqrt{\frac{P(1-P)}{n}}\right) = 1 - \alpha$$

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
WUOLAH Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

Siendo sus correspondientes límites:

$$L_{inf} = P - |Z_{\alpha/2}|\sqrt{\frac{P(1-P)}{n}} \qquad L_{sup} = P + |Z_{\alpha/2}|\sqrt{\frac{P(1-P)}{n}}$$

Intervalo de confianza para el parámetro δ^2

Al presentar la estimación puntual en el apartado (T9) se vio q la varianza S^2_x es un estimador sesgado de δ^2, siendo la cuasivarianza el estimador insesgado. Por ello vamos a utilizar la cuasivarianza muestral S^2_{n-1} como estimador de δ^2.

Además, sabemos q $\frac{(n-1)S^2_{n-1}}{\sigma^2}$ se distrubuye X^2_{n-1}

Podemos construir el intervalo de confianza con un nivel de confianza de 1-α:

$$P\left(\chi^2_{n-1,\alpha/2} \le \frac{(n-1)S^2_{n-1}}{\sigma^2} \le \chi^2_{n-1,1-\alpha/2}\right)$$

dividiendo por (n-1) S^2_{n-1} los términos de la desigualdad:

$$P\left(\frac{\chi^2_{n-1,\alpha/2}}{(n-1)S^2_{n-1}} \le \frac{1}{\sigma^2} \le \frac{\chi^2_{n-1,1-\alpha/2}}{(n-1)S^2_{n-1}}\right) = 1-\alpha$$

Por tanto, el intervalo de confianza para la varianza queda definido por la expresión:

$$P\left(\frac{(n-1)S^2_{n-1}}{\chi^2_{n-1,\alpha/2}} \ge \sigma^2 \ge \frac{(n-1)S^2_{n-1}}{\chi^2_{n-1,1-\alpha/2}}\right) = 1-\alpha$$

Sus límites inferior y superior son:

$$L_{inf} = \frac{(n-1)S^2_{n-1}}{\chi^2_{n-1,1-\alpha/2}} \qquad L_{sup} = \frac{(n-1)S^2_{n-1}}{\chi^2_{n-1,\alpha/2}}$$

El límite inferior se refiere a 1-(α/2) ya q se trata de una desigualdad dada por el límite superior del parámetro *π (apartado anterior)* q nos indica q $\sigma^2 \ge \frac{(n-1)S^2_{n-1}}{\chi^2_{n-1,1-\alpha/2}}$ o, lo q es lo mismo $\frac{(n-1)S^2_{n-1}}{\chi^2_{n-1,1-\alpha/2}} \le \sigma^2$, es decir, el límite inferior.

Cuando las muestras son grandes (*n* > 100), la distribución muestral de la varianza insesgada se puede aproximar a la normal. $N\left(\sigma^2; S^2_{n-1}\sqrt{\frac{2}{n}}\right)$. Por lo tanto, se puede construir el intervalo de confianza para la varianza definido como:

$$P\left(S^2_{n-1} - |Z_{\alpha/2}|S^2_{n-1}\sqrt{\frac{2}{n}} \le \sigma^2 \le S^2_{n-1} + |Z_{\alpha/2}|S^2_{n-1}\sqrt{\frac{2}{n}}\right) = 1-\alpha$$

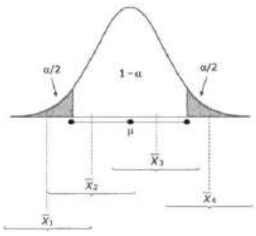
Los límites del intervalo son:

$$L_{inf} = S^2_{n-1} - |Z_{\alpha/2}|S^2_{n-1}\sqrt{\frac{2}{n}} \qquad L_{sup} = S^2_{n-1} + |Z_{\alpha/2}|S^2_{n-1}\sqrt{\frac{2}{n}}$$

3. SIGNIFICADO DEL NIVEL DE CONFIANZA

Hasta ahora; probabilidad de q el parámetro desconocido se encuentre entre los límites del intervalo. Sin embargo, no es del todo correcto, el concepto de probabilidad solo es aplicable a V y los valores de los límites del intervalo una vez calculados son valores constantes (no son variables).

Para interpretar correctamente el nivel de confianza asociado al intervalo se debe pensar: si se extraen un nº elevado de muestras (todas del mismo tamaño) y calculamos la media en cada una, obtendremos tantos intervalos de confianza como medias hayamos calculado. Pues bien, el 95% de todos los intervalos calculados tienen dentro al parámetro y el 5% no; de cada 100 intervalos, cabe esperar q 95 capten el valor del parámetro (intervalos correctos) y 5 no lo capten (incorrectos). Por tanto, una proporción de 1-α , de todos los intervalos de confianza contendrá al parámetro poblacional y una proporción α no los contendrá.



Los intervalos X_1 y X_4 no cubren el valor del parámetro; X_2 y X_3 sí. Las medias X_1 y X_4 están dentro de la zona sombreada y el intervalo de confianza NO contiene a la media poblacional. Sin embargo, X_2 y X_3 están en la zona no sombreada y contienen al parámetro. Por tanto, cualquier valor de la media en las zonas sombreadas da lugar a intervalos q NO contienen al parámetro, siendo la probabilidad de q esto ocurra de 0,05 (0,025 + 0,025). Por el contrario, el valor de la media en la zona no sombreada contendrá al parámetro y la probabilidad es de 0,95.
Se habla de probabilidad cuando se hace alusión a la V media, por eso al referirnos al intervalo hablaremos de confianza y no de probabilidad.

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Stuvia.com - The Marketplace to Buy and Sell your Study Material

4. GENERALIZACIÓN DE LA CONSTRUCCIÓN DE INTERVALOS

Se puede generalizar el procedimiento de construcción de intervalos representados por:

$$IC = [\text{Estimador} \pm \text{Error máximo de estimación}] = [\hat{\theta} \pm E_{max}]$$

Para **estimadores con distribución muestral conocida**, los pasos para construir el intervalo son:

1. Determinar el parámetro q queremos estimar y el estadístico (estimador) q, cumpliendo con las propiedades q debe tener un buen estimador, lo estima.
2. Conocer la distribución muestral del estadístico (estimador) y los parámetros q la definen (media y error típico). La distribución muestral nos da las probabilidades asociadas a cada uno de los valores.
3. Fijar el nivel de significación α (el investigador por valoración personal de la seguridad de sus datos y del empleado por otros investigadores en situaciones similares) o el nivel de confianza 1-α. Suele ser α= 0,05 o 0,01.
4. Determinar el error máximo de estimación (E_{max}) definido por el producto del error típico de la distribución muestral del estadístico q estima al parámetro por el valor del estadístico (Z, T, F…) correspondiente al nivel de significación prefijado.

5. FACTORES QUE AFECTAN AL INTERVALO DE CONFIANZA

- Nivel de confianza: la mayor o menor amplitud (tb precisión) depende fundamentalmente del nivel de confianza con el q se decide trabajar. Con 1- α= 0,95 tendremos intervalos menos amplios (o más precisos) q =0,99.
- Error típico: no es más q una medida de la variabilidad de la distribución muestral del estadístico = depende del tamaño muestral *n* y de la homogeneidad de la muestra, afectando ambos factores al intervalo de confianza.
- Tamaño muestral: Inversamente proporcional al error típico; a mayor *n,* menor error típico, menor amplitud del intervalo y mayor precisión.
- Homogeneidad de la muestra; Sí lo son = la varianza es pequeña = tb la desviación típica (poblacional o muestral), = error típico será pequeño = amplitud del intervalo menor = precisión mayor.

6. CÁLCULO DEL TAMAÑO MUESTRAL

Todo estudio en el q se trabaja con muestras, además de garantizar la representatividad es imprescindible determinar el tamaño q ha de tener; suficiente para garantizar la precisión deseada y/o detectar de forma correcta diferencias entre los grupos en el caso de q existiesen, valorar la intensidad de la relación…

La Estadística Paramétrica (dentro de la Inferencia Estadística) tiene 2 procedimientos de trabajo: la estimación de parámetros y el contraste de hipótesis. En ambos se parte de los datos muestrales. Sin embargo, en la 1ª se responde a una pregunta sobre la población; la 2ª hace una afirmación sobre la población q luego se comprueba. Por tanto, hay 2 situaciones q se deben considerar a la hora de det. el tamaño muestral.

En la **estimación de parámetros**, los factores q influyen en la determinación del tamaño muestral son:

- El parámetro q se va a estimar.
- El error máximo (E_{max}) q el investigador está dispuesto a admitir.
- El nivel de confianza (1-α) con el q se trabaja.
- La precisión q se desea para el estudio.
- La variabilidad q presenta la población en relación a la V en estudio.

En el **contraste de hipótesis** los factores son:

- El error tipo I (α) y error tipo II (β) y la potencia estadística.
- Magnitud de la diferencia (o tamaño del efecto).
- Direccionalidad de la hipótesis.
- Variabilidad de la población respecto a la V en estudio.

En ambos casos, el tamaño muestral debe ser un nº entero y, cuando su cálculo de lugar a un nº decimal debe redondearse siempre al inmediato superior.

Tamaño muestral para el parámetro media

La determinación del tamaño muestral gira en torno a los conceptos de error típico y error máximo de estimación. Al igual q en la determinación del intervalo de confianza, hay 2 posibles situaciones: conocer o desconocer la varianza poblacional.

Conocida la varianza poblacional

El error máximo de estimación viene dado por $E_{max} = \frac{|Z_{\alpha/2}|\sigma}{\sqrt{n}}$. Si elevamos al cuadrado y despejamos *n* tendremos:

$$n = \left(\frac{|Z_{\alpha/2}|\sigma}{E_{max}}\right)^2 = \frac{\sigma^2 Z_{\alpha/2}^2}{E_{max}^2}$$

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
WUOLAH Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

En el caso de poblaciones finitas y muestreo sin reposición debe multiplicarse por el factor de corrección: $\sqrt{\frac{N-n}{N-1}}$ donde N es el tamaño de la población. Entonces:

$$E_{max} = \frac{|Z_{\alpha/2}|\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Y, por tanto:

$$n = \frac{\sigma^2 Z^2_{\alpha/2} N}{E^2_{max}(N-1) + \sigma^2 Z^2_{\alpha/2}}$$

Desconocida la varianza poblacional

Lo habitual en la investigación es q no se conozcan la media y varianza poblacional. En estos casos, la varianza tb debe ser estimada al mismo tiempo q la media, mediante su estimador insesgado, la cuasivarianza.

La distribución muestral del estadístico media (el q queremos estimar) se ajusta a una distribución t de Student con n-1 grados de libertad. En la varianza, el error máximo de estimación, E_{max} vendrá dado por: siendo $t_{n-1;\alpha/2}$ el valor de la distribución t de Student con n-1 grados de libertad .

$$E_{max} = |t_{n-1;\alpha/2}|\frac{S_{n-1}}{\sqrt{n}}$$

$E^2_{max} = \frac{t^2_{n-1;\alpha/2} S^2_{n-1}}{n}$, y despejando *n*:

$$n = \frac{t^2_{n-1;\alpha/2} S^2_{n-1}}{E^2_{max}}$$

En el caso de poblaciones finitas o muestreo sin reposición habrá q multiplicar por el factor corrector siendo, en consecuencia, el valor n:

$$n = \frac{t^2_{n-1;\alpha/2} S^2_{n-1} N}{E^2_{max}(N-1) + t^2_{n-1;\alpha/2} S^2_{n-1}}$$

Como se ha dicho, $t_{1-\alpha/2;n-1}$ sería el valor de t de Student en la Tabla VI (Formulario) con n-1 grados de libertad para la probabilidad (1- α) especificada.

si no se conoce todavía *n*, ¿cómo buscar en las tablas t de Student con n-1 grados de libertad? 2 soluciones posibles:

A. Trabajar mediante aproximaciones sucesivas por un procedimiento iterativo;
B. Aproximar la distribución t de Student mediante la curva normal.

Otra cuestión es q, si aún no hemos obtenido la muestra (de hecho estamos determinando su tamaño), ¿Cómo podemos saber cuál es la varianza insesgada en nuestra muestra? Existen dif. posibilidades pero el procedimiento + cómodo y efectivo es obtener un valor aprox. para la desviación típica insesgada (S_{n-1}) . A partir de estudios previos, o de la realización de uno piloto, partiendo de este valor de S_{n-1} se calcula el tamaño muestral (*n*) y se procede a la selección de la muestra, medida de las V…

Tamaño muestral para el parámetro proporción

Como ya se sabe la distribución muestral de P es: $P \rightarrow N\left(\pi; \sqrt{\frac{\pi(1-\pi)}{n}}\right)$

Cuando n > 30 el error máximo es:

$$E_{max} = Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}$$

Por tanto:

$$n = \frac{Z^2_{\alpha/2} P(1-P)}{E^2_{max}}$$

Para la determinación del error típico de *P* surge ahora un problema añadido a los comentados en la estimación de la media, puesto que *p* y *q* = (1-p) dependen directamente de *P*, y es precisamente este parámetro el q hay q estimar. Lo q se suele hacer es suponer q la varianza de la distribución muestral es máxima (p=q=1-p=0,5), con lo q la muestra será casi con toda seguridad superior a lo estrictamente necesario pero, por contra, no habrá q hacer suposiciones arriesgadas sobre el valor de *p*.

Si la población es finita, o el muestreo es sin reposición, habrá q corregir el tamaño muestral multiplicado la expresión por el factor de corrección: $\sqrt{\frac{N-n}{N-1}}$

$$n = \frac{Z^2_{\alpha/2} P(1-P)}{E^2_{max}}\sqrt{\frac{N-n}{N-1}} = \frac{Z^2_{\alpha/2} P(1-P) N}{E^2_{max}(N-1) + Z^2_{\alpha/2} P(1-P)}$$

Downloaded by: giselamartn | giselamartinalcala0000@gmail.com
Distribution of this document is illegal
WUOLAH Want to earn $1.236 extra per year?
Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.