Universidad de Alcalá

Estadística-780004. Ejercicios de ejemplo Tema 1.

1. Urano tiene un total de 20 satélites, cuyos radios ecuatoriales oscilan entre los 13 y los 788 km, tal y como puede verse en el listado que se presenta más abajo. Se deben realizar los estudios estadísticos sobre los satélites menores (que son aquellos de radio menor de 50 km) que se piden en los apartados a) y b).

(Nombre satélite, radio en km): Cordelia, 13; Ofelia, 16; Bianca, 22; Crésida, 33; Desdémona, 29; Julieta, 42; Porcia, 55; Rosalinda, 27; Belinda, 34; Luna-1986U10, 20; Puck, 77; Miranda, 235; Ariel, 578; Umbriel, 584; Titania, 788; Oberón, 761; Calíbano, 30; Luna-999U1, 20; Sycorax, 60; Luna 1999U2, 15.

a) Con los datos sin agrupar calcular la (1) media aritmética, la mediana, la moda, (2) la desviación típica y la varianza y los (3) cuartiles. Realizar el (4) diagrama de caja y bigotes.

b) Representar los datos mediante (1) un diagrama de tallo y hojas. Agrupar los datos en 4 clases de equivalencia, incluyendo en cada clase los valores en la misma decena, dar (2) los límites y amplitud de cada clase, la marca de la clase, su frecuencia absoluta, relativa y acumulada, tanto relativa como absoluta; (3) la media aritmética, (4) la desviación típica. (6) Representar el histograma de frecuencias relativas.

2. Moore estableció en 1965 que el número de transistores en un circuito integrado se duplicaría cada año. En 1975 modificó su propia ley y estableció que se duplicaría cada dos años ¿Por qué modificó su ley? Lo que Moore sabía sobre la evolución de los procesadores ente 1971 y 1973 era: (Nombre del procesador, fecha de creación, número de transistores): 4004, 1971, 2300; 8008, 1972, 3.500; 8080, 1973, 4500; 8086,

3. ¿Cuál es la medida de promedio que mejor representa los datos de los salarios mensuales de una empresa sabiendo que dichos salarios son los que de describen en la siguiente lista? ¿Se podría decir que los salarios pagados por la empresa son altos?

Presidente: 22500; Gerente: 7500; Directivos (2): 5000; Jefe de Contabilidad: 2850; Jefes de Área (3): 2500; Jefes de Sección (4): 1850; Encargado: 1500; Empleados sin cargo (12): 1000

En el mismo caso analizar la representatividad de la media mediante la utilización de la desviación típica y el teorema de Tchebychev: a) para todos los empleados y b) para el encargado y los empleados sin cargo. Comparar los resultados.

Estadística-780004-GII-Grupo de mañana. Ejercicios de ejemplo Tema 2.

1. Teniendo en cuenta que los datos de radio ecuatorial, en miles de km, y densidad, en gr/cm3, de los planetas interiores del sistema solar son los siguientes:

(Nombre del planeta, radio ecuatorial, densidad): Mercurio, 2.4, 5.4; Venus, 6.1, 5.2; Tierra, 6.4, 5.5; Marte, 3.4, 3.9.

a) Calcular (1) la correlación existente entre el radio ecuatorial y la densidad para dichos planetas, establecer (2) una regresión lineal entre ambas magnitudes, (3) el ANOVA y (4) calcular el error estándar o la desviación típica residual

b) Sin tener en cuenta los decimales realizar (1) una tabla de frecuencias conjuntas de ambas variables, calcular sus frecuencias marginales, (2) su covarianza y (3) su matriz de covarianza.

Estadística-780004-GII-Grupo de mañana. Ejercicios de ejemplo Tema 3.

1. Se ha formado un equipo de investigación y análisis estadístico formado por 10 estudiantes procedentes de las titulaciones de Informática y de Sistemas de Información. De Informática se han elegido 4 chicas y un chico; y de Sistemas de Información 2 chicas y tres chicos.

Determinar el espacio muestral del experimento escoger al azar de entre los miembros del equipo de investigación (a) un chico o una chica, (b) escoger un estudiante de una titulación y (c) escoger un chico o una chica teniendo en cuenta la titulación.

Calcular la probabilidad de escoger al azar de entre los miembros del equipo de investigación (a) un estudiante que sea de informática, (b) un estudiante que sea de sistemas de información, (c) un chico, (d) una chica.

2. Sabiendo como es la constitución del equipo de investigación en estadística formado por estudiantes:

a) Verificar que la probabilidad de elegir al azar de entre los miembros del equipo un estudiante que sea de I, de SI, chico, chica, son menores que 1.

b) Verificar que las probabilidades de elegir al azar de entre los miembros del equipo un estudiante que sea de l y SI; chico o chica, son complementarias ¿lo son que sea de la SI y chico?

3. Se va a elegir un estudiante como responsable del equipo.

(a) ¿Qué probabilidad tiene el elegido de ser de Informática y chico?

(b) ¿Qué probabilidad tiene una chica de cualquiera de las dos titulaciones de ser elegida como representante?

4. ¿Cuántos números de 3 cifras diferentes se pueden formar con los ocho dígitos 2, 3, 4, 5, 6, 7, 8, 9?

5. ¿De cuántas formas diferentes pueden ponerse los nombres de los vecinos en los buzones de correos de un portal de 6 vecinos?

6. Se consideran n puntos en un plano, no alineados ¿cuántas rectas determinan dichos puntos? ¿Cuántos triángulos?

7. En una estantería hay 10 libros, de los cuales 4 son de estadística. Se toman 3 libros al azar ¿Cuál es la probabilidad de que al menos uno sea de estadística?

8. ¿Cuál es la probabilidad de que entre 4 personas escogidas al azar dos de ellas cumplan años el mismo día del año (no el mismo año)? ¿y si son k? ¿y si k=60? ¿se cumple en clase el resultado?

9. De un grupo de estudiantes se sabe que el 80% va a la universidad en transporte público y que el 25% no tienen carnet de conducir y que el 85% vienen en transporte público o no tienen carnet de conducir. Analizar para estos estudiantes si son independientes los sucesos venir en transporte público y no tener carnet de conducir.

Estadística-780004-GII-Grupo de mañana. Ejercicios de ejemplo Tema 4.

1. Se tiene: $f(x) = \begin{cases} m.(1 - x^4)\ si\ 0 < x < 1 \\ 0\ con\ cualquier\ otro\ valor \end{cases}$

a. Calcular (1) m y la (2) Función de distribución acumulativa.

b. Calcular (1) la media y (2) la varianza de la variable.

2. En un control de calidad, se observan 20 componentes fabricados por una máquina, asignándoles a cada uno un 0 si está correcto y un 1 si está defectuoso. Los resultados obtenidos son: 0, 0, 0, 1, 0; 0, 0, 0, 0, 0; 0, 0, 0, 0, 0; 0, 0, 0, 1, 0. Estudiar el problema desde la perspectiva de la Estadística Descriptiva y de la Inferencia Estadística.

3. Las ecuaciones de cálculo de la esperanza y la varianza de Bernuilli son: μ_x=p // σ_x^2=pq. Deducirlas a partir de la función de probabilidad.

4. ¿Cuál es la función de probabilidad de Bernuilli de éxito para el lanzamiento de un dado si solo se considera éxito sacar un 6?

5. Se lanza al aire ocho veces un dado ¿Cuál es la probabilidad de que salgan 2 números 6?

6. Una empresa aplica un descuento sobre cualquier factura que se pague en un periodo de 30 días desde su emisión. De todas las facturas, el 10% recibió el descuento. En una auditoría de la compañía se seleccionaron aleatoriamente 12 facturas ¿Cuál es la probabilidad de que, de las 12 facturas, tengan descuento menos de 4?

7. En un control de calidad, se observan 20 componentes fabricados por una máquina, asignándoles a cada uno un 0 si está correcto y un 1 si está defectuoso. Los resultados obtenidos son: 0, 0, 0, 1, 0; 0, 0, 0, 0, 0; 0, 0, 0, 0, 0; 0, 0, 0, 1, 0. Si esto se extiende a toda la población:

a. ¿Cuál es su función de probabilidad Geométrica?

b. ¿Cuál es la probabilidad de obtener 9 componentes correctos antes del primer defectuoso?

c. ¿y 19?

8. A una gasolinera llegan, de media, 3 coches por minuto. Determinar:

a. Cuál es la función de probabilidad.

b. Probabilidad de que en un minuto lleguen 2 coches.

c. Probabilidad de que en cinco minutos lleguen 12 coches.

9. Para hacer galletas con trozos de chocolate, a una masa de galletas suficiente para hacer 100 galletas, se le añaden 300 trozos de chocolate ¿Cuál es la probabilidad de que haya una galleta sin trozos de chocolate?

10. El número medio de visitas a un sitio web es de 5 cada minuto, determinar:

a. La función de probabilidad.

b. La media y la desviación típica.

c. La probabilidad que haya 17 visitas en 3 minutos.

d. La probabilidad que no haya ninguna visita en 1 segundo.

11. Se sabe que, debido a los procesos de llenado, el contenido de una lata de bebida de 33 cl. No es exactamente de 33 cl. en todas las latas, sino que se distribuye normalmente con una media de 33cl y una desviación típica de 2cl. Determinar:

a. La función de densidad.
b. ¿Cuál es la probabilidad de que el contenido de una lata sea superior a 33 cl?
c. Si un pack consta de 6 latas ¿cuál es la probabilidad de que el contenido sea inferior a 192 cl?

12. El número de horas de funcionamiento normal si fallos de un software es de 750 horas con una desviación típica de 8 horas. Calcular la probabilidad que:

a. La función de densidad.

b. Funcione al menos 760 horas sin fallos.

c. Funcione a lo sumo 748 horas sin fallos.

d. Funcione exactamente 755 horas sin fallos (tomar solo un decimal).

13. El porcentaje de tiros libres encestados por los jugadores de un equipo de baloncesto sigue una distribución normal. Un determinado jugador encesta menos del 80% con una probabilidad del 92,36%. Sabiendo que la media del equipo es de 67,5%, ¿Con qué probabilidad acertaría más del 90%?

14. Si una variable se distribuye en χ^2con 7 grados de libertad. Hallar $\chi^2_{1,7}$ y $\chi^2_{2,7}$ tales que:

a. $p\left(u > \chi^2_{2,7}\right) = 0.025$

b. $p\left(u \leq \chi^2_{1,7}\right) = 0.5$

c. $p\left(\chi^2_{1,7} \leq u \leq \chi^2_{2,7}\right) = 0.9$

15. Si una variable se distribuye en t de Student con 9 grados de libertad. Hallar t tal que:

a. $p(t \leq t_1) = 0.95$

b. $p(t > t_1) = 0.025$

c. $p(t \leq t_1) = 0.995$

d. $p(t \leq t_1) = 0.9$

16. Si una variable se distribuye en Fde Fisher con 3 grados de libertad en el numerador y cuatro grados de libertad en el denominador. Hallar $F_{3,4}$ tal que:

a. $p\left(F \leq F_{3,4}\right) = 0.975$

b. $p\left(F \leq F_{3,4}\right) = 0.25$

T6. Contraste de Hipótesis

Juan J. Cuadrado Gallego

Se quiere contrastar la hipótesis de que la duración de la vida laboral media de los trabajadores de un determinado país es de 40 años. Se sabe que la varianza es de 100. Para realizar el contraste se toma una muestra de 25 trabajadores y se obtiene que su vida laboral media es de 41 años ¿es aceptable la hipótesis con un nivel de significación del 10%?:

La población sigue una distribución normal y conocemos la varianza σ^2 poblacional

Las Hipótesis son: $H_0: \mu = \mu_0 = 40 \qquad H_1: \mu \neq 40$

Hacemos el contraste sobre la variable $d(\bar{x},\mu) = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{40-41}{\frac{10}{\sqrt{25}}} = 0.5$

(que sigue una distribución de probabilidad $N(0,1)$)

Si H_0 es cierta: 0.1 es la probabilidad de rechazarla y 90% es la probabilidad de aceptarla. Siendo 10% el nivel de significación del contraste.

Sabemos que: $p\left(-\lambda_{\alpha/2} < \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} < \lambda_{\alpha/2}\right) = (-1.64 < 0.5 < 1.64) = 90\%$

Se acepta la hipótesis nula ya que la región de aceptación con probabilidad *90%* es:

-1.64$< d_0 = 0.5 < 1.64$ siendo $d_c = 1.64$

T6. Contraste de Hipótesis

Juan J. Cuadrado Gallego

Contraste de la media aritmética:

La población sigue una distribución normal, no conocemos la varianza σ^2 poblacional, pero la muestra no es suficientemente grande.

Las Hipótesis son: $H_0 : \mu = \mu_0 \qquad H_1 : \mu \neq \mu_0$

Hacemos el contraste sobre la variable $d(\bar{x}, \mu) = \frac{\bar{x}-\mu}{s}\sqrt{n-1}$

(sigue una distribución de probabilidad t de student con n-1 grados de libertad)

Si H_0 es cierta: α es la probabilidad de rechazarla y $1-\alpha$ es la probabilidad de aceptarla. Siendo α el nivel de significación del contraste.

Sabemos que: $p\left(-t_{\alpha/2} < \frac{\bar{x}-\mu}{s}\sqrt{n-1} < t_{\alpha/2}\right) = 1-\alpha$

Por lo que la región de aceptación de H_0 con probabilidad p es:

$-d_c < d_0 < d_c$ siendo $d_c = t_{\alpha/2}$

T6. Contraste de Hipótesis

Juan J. Cuadrado Gallego

En la Universidad A, 400 alumnos con una media de 16 errores por examen, Universidad B, 200 alumnos, con una media de 15 errores. Desviaciones típicas poblacionales son de 40 para la Universidad A y 20 para la Universidad B ¿Se puede decir que el nivel de aprendizaje de estadística en ambas Universidades es el mismo con $\alpha = 4\%$?:

Las poblaciones siguen una distribución normal y conocemos las varianzas poblacionales. Las Hipótesis son: $H_0\colon \mu_x - \mu_y = 0 \quad H_1\colon \mu_x - \mu_y \neq 0$

Contraste sobre la variable$N(0,1)$]: $d(\bar{x}, \mu_0) = \dfrac{\bar{x}-\bar{y}-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} = \dfrac{16-\overline{15}-(0)}{\sqrt{\frac{40^2}{400}+\frac{20^2}{200}}} = 0{,}408$

Si H_0 es cierta: 0.04 es la probabilidad de rechazarla y 0.96 es la probabilidad de aceptarla. Siendo 4% el nivel de significación del contraste.

Sabemos que: $p\left(-\lambda_{\alpha/2} < \dfrac{\bar{x}-\bar{y}-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} < \lambda_{\alpha/2}\right) = (-2{,}055 < 0{,}408 < 2{,}055) = 96\%$

Se acepta la hipótesis nula ya que la región de aceptación con probabilidad *96%* es:

$-2{,}055 < 0{,}408 < 2{,}055$ siendo $d_c = 2{,}055$

EJERCICIOS DE PROBABILIDAD

1. En una clase de COU el 45% de los estudiantes suspende Matemáticas, el 60% suspende física y el 30% suspende ambas. Se selecciona al azar un alumno:
   a) Si suspendió Física ¿Cuál es la probabilidad de que suspendiera Matemáticas?
   P(M/F)=0.5
   b) Si suspendió Matemáticas " " Física?
   P(F/M)=0.67

2. En una determinada población, el 70% son aficionados al fútbol, el 60% al tenis y el 65% al baloncesto. El 45% lo son al fútbol y al tenis, el 40% al tenis y al baloncesto y el 50% al futbol y al baloncesto, mientras que el 30% lo son a los tres deportes. ¿Cuál es la probabilidad de que un individuo escogido al azar no sea aficionado a ninguno de los tres deportes?

   P(No se aficionado a ningún deporte)=0.1

3. Se lanza una moneda y si sale cara se ponen 7 bolas blancas en una urna y si sale cruz se ponen 4 blancas. Se vuelve a lanzar la moneda y se ponen 5 o 2 bolas negras, según se saque cara o cruz. Después se saca una bola de urna así compuesta.
   a. Calcular la probabilidad de que la bola sea blanca
      P(B)=0.62

   b. Si la bola resultó ser negra, determine la probabilidad de que hayan salido dos caras
      P(CaraCara/N)=0.27

4. Un libro tiene 3 capítulos. El 85% de las páginas del 1er capítulo no tiene ningún error. El 90% del segundo y el 95% del tercero tampoco tienen ningún error.
   El primer capítulo tiene 125 páginas, el 2º 150 y el 3º 175.
   1º ¿Cuál es la probabilidad de que al elegir una página al azar no tenga ningún error?
   P(No error)=0.905
   2º Supongamos que elegimos una página al azar y observamos que no tiene ningún error. ¿Cuál es la probabilidad de que sea del 2º capítulo?
   P(C2/No error)=0.333

5. Un jugado de baloncesto suele acertar el 75% de los tiros libres que lanza, si acierta el primer tiro puede realizar un segundo. Calcular la probabilidad de
   a. Hacer dos puntos
      P(dos puntos)=0.56
   b. Hacer uno
      P(Un punto)=0.19
   c. No hacer ningún punto
      P(Ningún punto)=0.25

6. De una baraja de 48 cartas se extrae simultáneamente dos de ellas. Calcular la probabilidad de que:
   a) Las dos sean copas.
      P(Dos copas)=0.059
   b) Al menos una sea copas.
      P(Al menos una copa)=0.441
   c) Una sea copa y la otra espada.
      P(Una copa y una espada)=0.128

7. Un taller sabe que por término medio acuden: por la mañana tres automóviles con problemas eléctricos, ocho con problemas mecánicos y tres con problemas de chapa, y por la tarde dos con problemas eléctricos, tres con problemas mecánicos y uno con problemas de chapa.
   a. Hacer una tabla ordenando los datos anteriores.
   b. Calcular el porcentaje de los que acuden por la tarde.
      P(tarde)=.3
   c. Calcular el porcentaje de los que acuden por problemas mecánicos.
      P(problemas mecánicos)=0.55
   d. Calcular la probabilidad de que un automóvil con problemas eléctricos acuda por la mañana.
      P(mañana/prob eléctricos)=0.6

8. Una clase consta de seis niñas y 10 niños. Si se escoge un comité de tres al azar, hallar la probabilidad de:
   a) Seleccionar tres niños.
      P(3 niños)=0.214
   b) Seleccionar exactamente dos niños y una niña.
      P(2 niños y 1 niña)=0.482
   c) Seleccionar por lo menos un niño.
      P(Al menos un niño)=0.964
   d) Seleccionar exactamente dos niñas y un niño.
      P(2 niñas y 1 niño)=0.258

EJERCICIOS DE PROBABILIDAD

1. En una clase de COU el 45% de los estudiantes suspende Matemáticas, el 60% suspende física y el 30% suspende ambas. Se selecciona al azar un alumno:
   a) Si suspendió Física ¿Cuál es la probabilidad de que suspendiera Matemáticas?
   b) Si suspendió Matemáticas " " Física?

2. En una determinada población, el 70% son aficionados al fútbol, el 60% al tenis y el 65% al baloncesto. El 45% lo son al fútbol y al tenis, el 40% al tenis y al baloncesto y el 50% al futbol y al baloncesto, mientras que el 30% lo son a los tres deportes. ¿Cuál es la probabilidad de que un individuo escogido al azar no sea aficionado a ninguno de los tres deportes?

3. Se lanza una moneda y si sale cara se ponen 7 bolas blancas en una urna y si sale cruz se ponen 4 blancas. Se vuelve a lanzar la moneda y se ponen 5 o 2 bolas negras, según se saque cara o cruz. Después se saca una bola de urna así compuesta.
   a. Calcular la probabilidad de que la bola sea roja
   b. Si la bola resultó ser negra, determine la probabilidad de que hayan salido dos caras

4. Un libro tiene 3 capítulos. El 85% de las páginas del 1er capítulo no tiene ningún error. El 90% del segundo y el 95% del tercero tampoco tienen ningún error.
   El primer capítulo tiene 125 páginas, el 2º 150 y el 3º 175.
   1º ¿Cuál es la probabilidad de que al elegir una página al azar no tenga ningún error?
   2º Supongamos que elegimos una página al azar y observamos que no tiene ningún error. ¿Cuál es la probabilidad de que sea del 2º capítulo?

5. Un jugado de baloncesto suele acertar el 75% de los tiros libres que lanza, si acierta el primer tiro puede realizar un segundo. Calcular la probabilidad de
   a. Hacer dos puntos
   b. Hacer uno
   c. No hacer ningún punto

6. De una baraja de 48 cartas se extrae simultáneamente dos de ellas. Calcular la probabilidad de que:
   a) Las dos sean copas.
   b) Al menos una sea copas.
   c) Una sea copa y la otra espada.

7. Un taller sabe que por término medio acuden: por la mañana tres automóviles con problemas eléctricos, ocho con problemas mecánicos y tres con problemas de chapa, y por la tarde dos con problemas eléctricos, tres con problemas mecánicos y uno con problemas de chapa.
   a. Hacer una tabla ordenando los datos anteriores.
   b. Calcular el porcentaje de los que acuden por la tarde.
   c. Calcular el porcentaje de los que acuden por problemas mecánicos.

d. Calcular la probabilidad de que un automóvil con problemas eléctricos acuda por la mañana.

8. Una clase consta de seis niñas y 10 niños. Si se escoge un comité de tres al azar, hallar la probabilidad de:
   a) Seleccionar tres niños.
   b) Seleccionar exactamente dos niños y una niña.
   c) Seleccionar por lo menos un niño.
   d) Seleccionar exactamente dos niñas y un niño.

Estadística

Profesor: **Marçal Mora-Cantallops**

Curso: 2018/19

Ejercicios distribución normal

1. La mayoría de escuelas de negocios piden, como requerimiento de acceso, que sus aplicantes realicen el GMAT (Graduate Management Admission Test), un examen adaptativo que busca cuantificar las habilidades analíticas, escritas y verbales de los candidatos. Las notas se pueden considerar normalmente distribuidas, y la media de este año es de 527 con una desviación estándar de 112.

   (a) ¿Cuál es la probabilidad de que un candidato saque más de 500 puntos en el examen?

   (b) Si la escuela a la que quiero acceder me dice que la nota no le importa pero que tengo que estar en el top 5% de candidatos, ¿qué nota debería haber sacado como mínimo?

   (c) ¿Qué porcentaje de estudiantes ha sacado entre 520 y 600?

2. La duración de un embarazo humano sigue una distribución normal centrada en 266 días y una desviación estándar de 16 días.

   (a) ¿Qué proporción de embarazos dura entre 240 y 270 días (que es aproximadamente el equivalente a 8-9 meses)?

   (b) ¿Qué número de días separa el 70% de los embarazos de menor duración del 30% de mayor duración?

Ejercicios de las otras distribuciones

1. Encuentra los valores del estadístico χ^2 para los casos siguientes:

   (a) $\chi^2_{0.01}$ para $df = 5$.

   (b) $\chi^2_{0.01}$ para $df = 25$.

   (c) $\chi^2_{0.05}$ para $df = 6$.

   (d) $\chi^2_{0.05}$ para $df = 12$.

   (e) $\chi^2_{0.10}$ para $df = 15$.

   (f) $\chi^2_{0.01}$ para $df = 100$.

2. Busca, para $df = 11$ y $\alpha = 0.05$:

   (a) χ^2_{α}.

   (b) $\chi^2_{\frac{\alpha}{2}}$.

3. Encuentra los valores del estadístico t_1 tal que:
   (a) $P(t > t_1)$ para $df = 5$ y $\alpha = 0.05$.
   (b) $P(t < t_1)$ para $df = 5$ y $\alpha = 0.95$.
   (c) $P(t > t_1)$ para $df = 13$ y $\alpha = 0.01$.
   (d) $P(t > t_1)$ para $df = 23$ y $\alpha = 0.01$.
   (e) $P(t > t_1)$ para $df = 25$ y $\alpha = 0.025$.
   (f) $P(t > t_1)$ para $df = 5$ y $\alpha = 0.95$.
4. Encuentra los valores del estadístico F para los casos siguientes:
   (a) $P(F > F_{1,1})$ con $p = 0.025$.
   (b) $P(F > F_{8,7})$ con $p = 0.1$.
   (c) $P(F > F_{8,4})$ con $p = 0.01$.
   (d) $P(F < F_{8,7})$ con $p = 0.9$.
   (e) $P(F > F_{120,16})$ con $p = 0.1$.

Ejercicios autoevaluación (Entrega 2ª)
TEMA 2: Descripción conjunta de varias variables

T2.1. La información estadística obtenida de una muestra de tamaño 12 sobre la relación existente entre la inversión realizada y el rendimiento obtenido en cientos de miles de euros para explotaciones agrícolas, se muestra en el siguiente cuadro:

Inversión (X)	11	14	16	15	16	18	20	21	14	20	19	11
Rendimiento (Y)	2	3	5	6	5	3	7	10	6	10	5	6

Calcular:

a) La recta de regresión del rendimiento respecto de la inversión.

b) La previsión de inversión que se obtendrá con un rendimiento de 1 250 000 €.

T2.2. El número de horas dedicadas al estudio de una asignatura y la calificación obtenida en el examen correspondiente, de ocho personas es:

Horas (X)	20	16	34	23	27	32	18	22
Calificación (Y)	6,5	6	8,5	7	9	9,5	7,5	8

Se pide:

a) Recta de regresión de Y sobre X.

b) Calificación estimada para una persona que hubiese estudiado 28 horas.

T2.3. En la tabla siguiente se indica la edad (en años) y la conducta agresiva (medida en una escala de cero a 10) de 10 niños.

Edad	6	6	6,7	7	7,4	7,9	8	8,2	8,5	8,9
Conducta agresiva	9	6	7	8	7	4	2	3	3	1

a) Obtener la recta de regresión de la conducta agresiva en función de la edad.

b) A partir de dicha recta, obtener el valor de la conducta agresiva que correspondería a un niño de 7,2 años.

Ejercicios autoevaluación (Entrega 2ª)
TEMA 2: Descripción conjunta de varias variables

T2.4. Los valores de dos variables X e Y se distribuyen según la tabla siguiente:

Y/X	100	50	25
14	1	1	0
18	2	3	0
22	0	1	2

Se pide:

a) Calcular la covarianza.
b) Obtener e interpretar el coeficiente de correlación lineal.
c) Ecuación de la recta de regresión de Y sobre X.
d) Si X=67 cual sería el valor de Y en la recta de regresión determinada en el apartado c).

T2.5. Las puntuaciones obtenidas por un grupo de alumnos en una batería de test que mide la habilidad verbal (X) y el razonamiento abstracto (Y) son las siguientes:

Y/X	>20	>30	>40	>50
(25-35)	6	4	0	0
(35-45)	3	6	1	0
(45-55)	0	2	5	3
(55-65)	0	1	2	7

Se pide:

a) ¿Existe correlación entre ambas variables?
b) Según los datos de la tabla, si uno de estos alumnos obtiene una puntuación de 70 puntos en razonamiento abstracto, ¿en cuánto se estimará su habilidad verbal?

T2.6. En una empresa de transportes trabajan cuatro conductores. Los años de antigüedad de permisos de conducir y el número de infracciones cometidas en el último año por cada uno de ellos son los siguientes:

Años (X)	3	4	5	6
Infracciones (Y)	4	3	2	1

Calcular el coeficiente de correlación lineal e interpretarlo.

Estadística

Profesor: **Marçal Mora-Cantallops**
Curso: 2018/19

Ejercicios distribuciones

1. Stephen Curry es un jugador de baloncesto, clave en los Golden State Warriors, de la NBA. Sus estadísticas de las últimas 10 temporadas muestran su acierto en el tiro desde la línea de tres puntos: encesta un 43,8% de los triples que intenta.

   (a) Si en un partido realiza ocho intentos de media, ¿cuántos triples esperaríamos que metiese cada partido?

   (b) ¿Cuál es la probabilidad de que no acierte ninguno? ¿Y que acierte 3?

   (c) ¿Qué cantidad de triples no superará en el 75% de sus peores actuaciones?

   (d) ¿Cuántos intentos debería realizar para asegurar, al 99,5%, que encesta al menos un triple en cada partido?

2. Un estadista colchonero revisa las estadísticas de Griezmann y se da cuenta que el año pasado anotó una media de 0,59 goles por partido. Otro culé se fija en las de Messi, que son de 0,94. Reflexiona primero sobre las preguntas y luego respóndelas analíticamente.

   (a) ¿Qué es más probable, que Griezmann marque al menos un gol en un partido o que Messi no marque ninguno?

   (b) ¿Alguno se quedará más de la mitad de los partidos sin marcar?

   (c) ¿Y si juegan cinco partidos, es más probable que Griezmann marque sólo un gol o que Messi marque cinco goles?

3. La mayoría de escuelas de negocios piden, como requerimiento de acceso, que sus aplicantes realicen el GMAT (Graduate Management Admission Test), un examen adaptativo que busca cuantificar las habilidades analíticas, escritas y verbales de los candidatos. Las notas se pueden considerar normalmente distribuidas, y la media de este año es de 527 con una desviación estándar de 112.

   (a) ¿Cuál es la probabilidad de que un candidato saque más de 500 puntos en el examen?

   (b) Si la escuela a la que quiero acceder me dice que la nota no le importa pero que tengo que estar en el top 5% de candidatos, ¿qué nota debería haber sacado como mínimo?

   (c) ¿Qué porcentaje de estudiantes ha sacado entre 520 y 600?

4. La duración de un embarazo humano sigue una distribución normal centrada en 266 días y una desviación estándar de 16 días.

(a) ¿Qué proporción de embarazos dura entre 240 y 270 días (que es aproximadamente el equivalente a 8-9 meses)?

(b) ¿Qué número de días separa el 70% de los embarazos de menor duración del 30% de mayor duración?

Ejercicios intervalos de confianza

5. Un estudio de 1916 obtuvo que la media de altura de 22 habitantes de Girona era de 168cm. Sabiendo que la varianza poblacional es 36, calcula el intervalo de confianza del 90% para la media de altura de la población de Girona.

6. El mismo estudio de 1916 obtuvo que la media de altura de 119 habitantes de Barcelona era de la misma que los de Girona, con la misma varianza poblacional. Calcula el intervalo de confianza del 90% para la media de altura de la población de Barcelona.

7. El total de alturas del estudio para España incluía 1177 individuos, con una media de 167cm y una desviación estándar de la población de 6, siguiendo una distribución normal. ¿Cuál es el intervalo de confianza del 95% para la altura de los españoles de hace un siglo?

8. ¿Cómo variarían los resultados anteriores si la varianza de la que disponemos fuese muestral en vez de poblacional?

9. Hemos realizado el mismo examen a un grupo de 15 alumnos y a otro de 50. Ambos grupos han obtenido una media de 5. Las notas siguen una distribución normal y las varianzas muestrales son 9 y 4, respectivamente. ¿Cuáles son los intervalos de confianza para un 95% en cada uno de los casos?

10. Fabricamos yogurts de 500gr, pero al ser tan denso es difícil asegurar que todos sean exactamente de 500gr (aunque sabemos que la media lo es). Tenemos programada una auditoría que verificará si cumplimos la restricción de +/- 2% en el 95% de los packs de la producción. Si tomamos 10 medidas al azar y obtenemos $\{500, 510, 480, 505, 495, 500, 490, 515, 505, 510\}$, ¿podemos confiar que estamos cumpliendo con la restricción?

11. Suponiendo que la varianza poblacional es el valor más grande del intervalo anterior, ¿cuántas unidades fabricadas deberá seleccionar la auditoría en su muestra para asegurar que calcula la media con un 95% de confianza y un error menor a los 10gr?

Estadística

Profesor: **Marçal Mora-Cantallops**
Curso: 2018/19

Ejercicios distribuciones

1. Stephen Curry es un jugador de baloncesto, clave en los Golden State Warriors, de la NBA. Sus estadísticas de las últimas 10 temporadas muestran su acierto en el tiro desde la línea de tres puntos: encesta un 43,8% de los triples que intenta.
   (a) Si en un partido realiza ocho intentos de media, ¿cuántos triples esperaríamos que metiese cada partido?
   (b) ¿Cuál es la probabilidad de que no acierte ninguno? ¿Y que acierte 3?
   (c) ¿Qué cantidad de triples no superará en el 75% de sus peores actuaciones?
   (d) ¿Cuántos intentos debería realizar para asegurar, al 99,5%, que encesta al menos un triple en cada partido?
2. Un estadista colchonero revisa las estadísticas de Griezmann y se da cuenta que el año pasado anotó una media de 0,59 goles por partido. Otro culé se fija en las de Messi, que son de 0,94. Reflexiona primero sobre las preguntas y luego respóndelas analíticamente.
   (a) ¿Qué es más probable, que Griezmann marque al menos un gol en un partido o que Messi no marque ninguno?
   (b) ¿Alguno se quedará más de la mitad de los partidos sin marcar?
   (c) ¿Y si juegan cinco partidos, es más probable que Griezmann marque sólo un gol o que Messi marque cinco goles?
3. La mayoría de escuelas de negocios piden, como requerimiento de acceso, que sus aplicantes realicen el GMAT (Graduate Management Admission Test), un examen adaptativo que busca cuantificar las habilidades analíticas, escritas y verbales de los candidatos. Las notas se pueden considerar normalmente distribuidas, y la media de este año es de 527 con una desviación estándar de 112.
   (a) ¿Cuál es la probabilidad de que un candidato saque más de 500 puntos en el examen?
   (b) Si la escuela a la que quiero acceder me dice que la nota no le importa pero que tengo que estar en el top 5% de candidatos, ¿qué nota debería haber sacado como mínimo?
   (c) ¿Qué porcentaje de estudiantes ha sacado entre 520 y 600?
4. La duración de un embarazo humano sigue una distribución normal centrada en 266 días y una desviación estándar de 16 días.

(a) ¿Qué proporción de embarazos dura entre 240 y 270 días (que es aproximadamente el equivalente a 8-9 meses)?

(b) ¿Qué número de días separa el 70% de los embarazos de menor duración del 30% de mayor duración?

Ejercicios intervalos de confianza

5. Un estudio de 1916 obtuvo que la media de altura de 22 habitantes de Girona era de 168cm. Sabiendo que la varianza poblacional es 36, calcula el intervalo de confianza del 90% para la media de altura de la población de Girona.

6. El mismo estudio de 1916 obtuvo que la media de altura de 119 habitantes de Barcelona era de la misma que los de Girona, con la misma varianza poblacional. Calcula el intervalo de confianza del 90% para la media de altura de la población de Barcelona.

7. El total de alturas del estudio para España incluía 1177 individuos, con una media de 167cm y una desviación estándar de la población de 6, siguiendo una distribución normal. ¿Cuál es el intervalo de confianza del 95% para la altura de los españoles de hace un siglo?

8. ¿Cómo variarían los resultados anteriores si la varianza de la que disponemos fuese muestral en vez de poblacional?

9. Hemos realizado el mismo examen a un grupo de 15 alumnos y a otro de 50. Ambos grupos han obtenido una media de 5. Las notas siguen una distribución normal y las varianzas muestrales son 9 y 4, respectivamente. ¿Cuáles son los intervalos de confianza para un 95% en cada uno de los casos?

10. Fabricamos yogurts de 500gr, pero al ser tan denso es difícil asegurar que todos sean exactamente de 500gr (aunque sabemos que la media lo es). Tenemos programada una auditoría que verificará si cumplimos la restricción de +/- 2% en el 95% de los packs de la producción. Si tomamos 10 medidas al azar y obtenemos {500, 510, 480, 505, 495, 500, 490, 515, 505, 510}, ¿podemos confiar que estamos cumpliendo con la restricción?

11. Suponiendo que la varianza poblacional es el valor más grande del intervalo anterior, ¿cuántas unidades fabricadas deberá seleccionar la auditoría en su muestra para asegurar que calcula la media con un 95% de confianza y un error menor a los 10gr?

Soluciones

1. (a) Siendo una binomial, la esperanza es $\mu = np = 8 \cdot 0.438 = 3.5$

   (b) $P(0) = \binom{8}{0} \cdot 0.438^0 \cdot (1 - 0.438)^8 = 0.009952$

   $P(3) = \binom{8}{3} \cdot 0.438^3 \cdot (1 - 0.438)^5 = 0.2638104$

   (c) $P(0) = \binom{8}{0} \cdot 0.438^0 \cdot (1 - 0.438)^8 = 0.009952$

   $P(1) = \binom{8}{1} \cdot 0.438^1 \cdot (1 - 0.438)^7 = 0.06204665$

   $P(2) = \binom{8}{2} \cdot 0.438^2 \cdot (1 - 0.438)^6 = 0.1692482$

   $P(3) = \binom{8}{3} \cdot 0.438^3 \cdot (1 - 0.438)^5 = 0.2638104$

   $P(4) = \binom{8}{4} \cdot 0.438^4 \cdot (1 - 0.438)^4 = 0.2570039$

   $P(x \leq 4) = 0.7620608$. **En su 75% peor no pasa de los 4 triples.**

   (d) Tenemos que jugar con el número de intentos:

   Con 8 intentos, $1 - P(0) = 99\%$.

   Con 9 intentos, $1 - P(0) = 99,44\%$.

   Con 10 intentos, $1 - P(0) = 99,68\%$.

2. (a) Es Poisson, puesto que es un ratio. $P(X) = e^{-\lambda} \left(\frac{\lambda^k}{k!}\right)$

   Para Griezmann, $1 - P(0) = 1 - e^{-0,59} \left(\frac{0,59^0}{0!}\right) = 0.446$

   Para Messi, $P(0) = e^{-0,94} \left(\frac{0,94^0}{0!}\right) = 0.391$

   Es más probable que Griezmann marque más de un gol que ver a Messi en blanco.

   (b) Para Griezmann, $P(0) = e^{-0,59} \left(\frac{0,59^0}{0!}\right) = 0.554$. **Se quedará sin marcar más de la mitad de los partidos.** Messi no (39,1%, del apartado anterior).

   (c) Al ser 5 partidos, hay que multiplicar el ratio de uno por cinco.

   Para Griezmann, $P(1) = e^{-0,59 \cdot 5} \left(\frac{(0,59 \cdot 5)^1}{1!}\right) = 0.154$

   Para Messi, $P(5) = e^{-0,94 \cdot 5} \left(\frac{(0,94 \cdot 5)^5}{5!}\right) = 0.174$

   Así que es más probable que Messi anote cinco goles que ver a Griezmann con sólo uno en cinco partidos.

3. Es una N(527, 112).

   (a) Para la $P(X > 500)$ calculamos primero el estadístico z.

   $z = \frac{500-527}{112} = -0.24$

   Como la tabla nos da los valores a la izquierda para los que están a la derecha de la media, en realidad tenemos que hacer el simétrico $(1 - P(z > 0.24))$ o lo que es lo mismo, directamente $P(z < 0.24)$. *Dibujarlo para verlo.*

Buscando en la tabla, $P(z < 0.24) = 0.5948$. **Hay un 59,5% de posibilidades de que un estudiante saque más de un 500.**

(b) Hay que entrar en la tabla del revés. El top 5% deja al 95% a su izquierda, entonces buscamos el valor de z que deja al 95% de los valores a la izquierda. Es $z = 1.65$, aunque si nos queremos poner finos podríamos incluso decir $z = 1.645$, que sería lo más correcto.

$z = 1.645 = \frac{x-527}{112}$, entonces $x = 711.24$. **Si los puntos son enteros, debería haber sacado 712 o más.**

(c) Lo tenemos que hacer en dos partes, porque tenemos que obtener la $P(x < 600)$ y restarle la $P(z < 520)$.

$z = \frac{600-527}{112} = 0.65$ -> $P(x < 600) = P(z < 0.65) = 0.7422$
$z = \frac{520-527}{112} = -0.06$ -> $P(x < 520) = P(z < -0.06) = P(z > 0.06) = 1 - P(z < 0.06) = 1 - 0.5239 = 0.4761$

La probabilidad de sacar entre 520 y 600 es, pues, $P(520 < x < 600) = 0.7422 - 0.4761 = 0.2661$**, o lo que es lo mismo, un 26,6%.**

4. Es una N(266, 16).

(a) Similar al ejercicio anterior, tenemos que obtener la $P(x < 270)$ y restarle la $P(z < 240)$.

$z = \frac{270-266}{16} = 0.25$ -> $P(x < 270) = P(z < 0.25) = 0.5987$
$z = \frac{240-266}{16} = -1.625$ -> $P(x < 240) = P(z < -1.625) = P(z > 1.625) = 1 - P(z < 1.625) = 1 - 0.9479 = 0.0521$ (Nota: me da igual si cogéis el punto medio o si redondeáis a 1.62 o 1.63, el resultado varía poco).

La probabilidad de un embarazo entre 240 y 270 días es, pues, $P(240 < x < 270) = 0.5987 - 0.0521 = 0.5466$**, o lo que es lo mismo, un 54,66%.**

(b) Hay que buscar en la tabla la z correspondiente a dejar el 70% a un lado.

Es aproximadamente $z = 0.525$; de nuevo, también son válidos 0.52 y 0.53 para el cálculo.

(c) $z = 0.525 = \frac{x-266}{16}$, entonces $x = 274.4$. **El 70% de los embarazos duran menos de 274,4 días**.

5. Sólo conocemos la varianza poblacional (la desviación es, por lo tanto, 6) y como $1 - \alpha = 0.9$, entonces $\alpha = 0.1$. No tenemos más información y con un número tan pequeño en la muestra tampoco podemos asegurar nada de la distribución. Así, el intervalo para la media μ:

$$\left(x - \frac{\sigma}{\sqrt{\alpha n}}, x + \frac{\sigma}{\sqrt{\alpha n}}\right) = \left(168 - \frac{6}{\sqrt{0.1 \cdot 22}}, 168 + \frac{6}{\sqrt{0.1 \cdot 22}}\right) = (168 - 4.05, 168 + 4.05) = (163.955, 172.045)$$

6. Ahora con $n > 30$ podemos suponer que sigue una distribución normal. Con $\alpha = 0.1$, $\frac{\alpha}{2} = 0.05$ y $\lambda_{\frac{\alpha}{2}} = 1.645$ (de la tabla).

$\left(x - \lambda_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, x + \lambda_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = \left(168 - 1.645\frac{6}{\sqrt{119}}, 168 + 1.645\frac{6}{\sqrt{119}}\right) = (168 - 0.9, 168 + 0.9) = (167.1, 168.9)$. **Es importante notar como es mucho más pequeño que en el caso anterior con menos muestras y sin poder suponer la distribución.**

7. Aquí ya nos dicen que es normal, además de ser una muestra grande, con $n >> 30$. Con $\alpha = 0.05$, $\frac{\alpha}{2} = 0.025$ y $\lambda_{\frac{\alpha}{2}} = 1.96$ (de la tabla).

   $\left(x - \lambda_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, x + \lambda_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = \left(167 - 1.96\frac{6}{\sqrt{1177}}, 167 + 1.96\frac{6}{\sqrt{1177}}\right) = (167 - 0.34, 167 + 0.34) = (166.66, 167.34)$. **Notad como el incremento de confianza incrementa el tamaño del intervalo (de multiplicar por 1.645 a 1.96) pero como el incremento del tamaño de la muestra lo reduce (y, en global, acaba resultando menor en este caso).**

8. Si la varianza fuese muestral en vez de poblacional, ya no sería σ, sino que tendríamos $s = 6$. No obstante, tanto el caso 6 como 7 tienen muestras suficientemente grandes, así que se puede suponer que $\sigma = s$. En el caso del ejercicio 5, si el enunciado no cambia **no podemos hacer nada**, porque ponía específicamente que no se conocía la distribución. Pero para poner un ejemplo didáctico vamos a suponer que sabemos que las alturas siguen una distribución normal (que es así, como tantas cosas en la naturaleza).

   Como la muestra es pequeña, hay que cambiar la normal por la t-student y n por n-1. Para $\alpha = 0.1$, $t_{\frac{\alpha}{2}} = 1.721$ (tiene $n - 1 = 21$ grados de libertad).

   $\left(x - t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n-1}}, x + t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n-1}}\right) = \left(168 - 1.721\frac{6}{\sqrt{21}}, 168 + 1.721\frac{6}{\sqrt{21}}\right) = (168 - 2.25, 168 + 2.25) = (165.75, 170.25)$.

   Fijaros que sabiendo información adicional sobre la distribución el intervalo es más pequeño que en (5), pero también como la t-student es un poco más gruesa (1.721) que la normal (1.645) para el mismo nivel de confianza.

9. En el primer caso es una t-student, porque la muestra es pequeña, sabemos que sigue una distribución normal y la varianza muestral. En el segundo caso, con 50 podemos suponer que la varianza muestral es igual a la poblacional y fijarnos en la normal, pero aunque no lo hiciéramos obtendríamos un valor parecido (fijaros en la fila ∞ de la tabla de la t-student).

   Para el grupo de 15 alumnos, $\alpha = 0.05$, $t_{\frac{\alpha}{2}} = 2.145$ (tiene $n - 1 = 14$ grados de libertad).

   $\left(x - t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n-1}}, x + t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n-1}}\right) = \left(5 - 2.145\frac{3}{\sqrt{14}}, 5 + 2.145\frac{3}{\sqrt{14}}\right) = (5 - 1.72, 5 + 1.72) = (3.28, 6.72)$.

   Para el grupo de 50 alumnos, hecho con la t-student, $\alpha = 0.05$, $t_{\frac{\alpha}{2}} = 1.960$ (tiene $n - 1 = 49$ grados de libertad).

$\left(x - t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n-1}}, x + t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n-1}}\right) = \left(5 - 1.960\frac{2}{\sqrt{49}}, 5 + 1.960\frac{2}{\sqrt{49}}\right) = (5 - 0.56, 5 + 0.56) = (4.44, 5.56)$.

Para el grupo de 50 alumnos, hecho con la normal, $\alpha = 0.05$, $\lambda_{\frac{\alpha}{2}} = 1.96$.

$\left(x - \lambda_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, x + \lambda_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = \left(5 - 1.96\frac{2}{\sqrt{50}}, 5 + 1.96\frac{2}{\sqrt{50}}\right) = (5 - 0.554, 5 + 0.554) = (4.446, 5.554)$.
Fijaros como el efecto es únicamente debido a n vs n-1 y es básicamente despreciable.

10. Aquí lo que no conocemos (y nos interesa) es la varianza poblacional. En un caso así podemos estimarla siguiendo la muestra, es decir, $\hat{\sigma} = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n} = \frac{1000}{10} = 100$. (También se puede dejar como $n\hat{\sigma} = 1000$).

    El intervalo es, entonces:

    $\left(\frac{n\hat{\sigma}}{\chi^2_{10,\frac{\alpha}{2}}}, \frac{n\hat{\sigma}}{\chi^2_{10,1-\frac{\alpha}{2}}}\right) = \left(\frac{1000}{\chi^2_{10,0.025}}, \frac{1000}{\chi^2_{10,0.975}}\right) = \left(\frac{1000}{20.483}, \frac{1000}{3.247}\right) = (48.821, 307.977)$

    Si lo pasamos a desviación estándar,$(6.99, 17.55)$**. Queríamos una desviación menor al 2% de 500gr, que son 10gr. Como el intervalo de confianza del 95% llega hasta los 17.55gr, no podemos asegurar que cumplamos. (Nota: podríamos asegurar, por ejemplo, una desviación menor al 4%, que serían 20gr).**

11. $\sigma^2 = 307.977$.

    Sabiendo que $n = 4(\lambda_{\frac{\alpha}{2}})^2\frac{\sigma^2}{error^2} = 4 \cdot 1.96^2 \cdot \frac{307.977}{10^2} = 47.33$

    En el peor de los casos, bastará con coger unos 48 yogurts (debidamente aleatorizados).