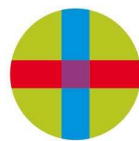


EJERCICIOS DE ESTADÍSTICA

Asignatura: Estadística Aplicada a la Psicología
Curso: 1º de Psicología

Santiago Angulo Díaz-Parreño (sangulo@ceu.es)
José Miguel Cárdenas Rebollo (cardenas@ceu.es)
Anselmo Romero Limón (arlimon@ceu.es)
Alfredo Sánchez Alberca (asalber@ceu.es)

Curso 2009-2010



CEU

*Universidad
San Pablo*

Índice

1. Estadística Descriptiva	2
2. Regresión y Correlación	8
3. Probabilidad	15
4. Variables Aleatorias	19
5. Estimación de parámetros	28
6. Contraste de hipótesis	32

1. Estadística Descriptiva

- ★ 1. El número de lesiones padecidas durante una temporada por cada jugador de un equipo de fútbol fue el siguiente:

0 – 1 – 2 – 1 – 3 – 0 – 1 – 0 – 1 – 2 – 0 – 1
1 – 1 – 2 – 0 – 1 – 3 – 2 – 1 – 2 – 1 – 0 – 1

Se pide:

- Construir la tabla de frecuencias.
- Dibujar el polígono de frecuencias.
- Calcular los cuartiles y el rango intercuartílico e interpretarlo.
- Calcular el coeficiente de asimetría e interpretarlo.

SOLUCIÓN

- $C_1 = 1$ lesión, $C_2 = 2$ lesiones y $C_3 = 3$ lesiones. $RI = 2$ lesiones, lo que indica que hay bastante dispersión central.
 - $\bar{x} = 1,125$ lesiones, $s^2 = 0,776$ lesiones², $s = 0,88$ lesiones y $g_1 = 0,49$ lo que indica que la distribución es un poco asimétrica a la derecha.
-

2. En un grupo de alumnos universitarios se realiza un estudio sobre la velocidad de lectura. Para ellos se les dio a leer un texto de un periódico y se contó el número de palabras que leyeron en un minuto, obteniendo los siguientes resultados:

286 – 242 – 185 – 244 – 296 – 211 – 233 – 179 – 190 – 274
255 – 222 – 260 – 216 – 204 – 247 – 232 – 257 – 196 – 261

Se pide:

- Construir la tabla de frecuencias agrupando en intervalos de amplitud 25 comenzando en 175.
- Calcular las medidas de tendencia central. ¿Es representativa la media?
- Calcular el coeficiente de apuntamiento e interpretarlo.

SOLUCIÓN

- $\bar{x} = 233,75$, $Med = 235$, $Mod = [225, 250]$ y $[250 – 275]$. $s^2 = 1017,19$, $s = 31,89$ y $cv = 0,14$ lo que indica que hay poca dispersión y la media es muy representativa.
 - $g_2 = -1,1$ de manera que la distribución es bastante platicúrtica.
-

3. En un cuestionario se ha preguntado a un grupo de individuos por su nivel de estudios (SE: sin estudios, EB: estudios básicos, ES: estudios secundarios, EU: estudios universitarios, ED: estudios de doctorado) obteniendo los siguientes resultados:

EU, ES, ES, EU, EB, EB, ED, SE, EB, SE, EU, ES, ES, ES, EB,
EB, SE, ED, EB, ES, ES, SE, EU, EU, EB, ES, EU, ES, EB, ES

Se pide:

- Construir la tabla de frecuencias y los diagramas asociados.

- b) Dar una medida de representatividad.
 c) Calcular los cuartiles y el percentil 90.

SOLUCIÓN

- b) $Me = ES$ y $Mo = ES$.
 c) $C_1 = EB$, $C_2 = ES$, $C_3 = EU$ y $P_{90} = EU$.
-

- ★ 4. Las pruebas de determinación del cociente intelectual realizadas en un colectivo de alumnos universitarios reflejan los siguientes resultados:

C.I.	Nº de alumnos
[80,90)	3
[90,100)	12
[100,110)	21
[110,120)	24
[120,130)	13
[130,140)	2

- a) Calcular el coeficiente de variación del cociente intelectual. ¿Es representativa la media?
 b) Si se considera “muy inteligente” a una persona cuyo cociente intelectual se encuentra en el 10 % de los más inteligentes, ¿cuál será el cociente que delimita la categoría “muy inteligente” en el colectivo de alumnos?

SOLUCIÓN

- a) $cv = 0,1$ lo que indica que hay poca dispersión y la media es muy representativa.
 b) $P_{90} = 125,77$.
-

- ★ 5. En un estudio sobre el crecimiento se tomaron dos muestras, una de niños recién nacidos y otra de niños con un año de edad. Las estaturas observadas en cada muestra fueron:

Recién nacidos: 51-50-51-53-49-50-53-50-47-50.

Niños de un año: 62-65-69-71-65-66-68-69.

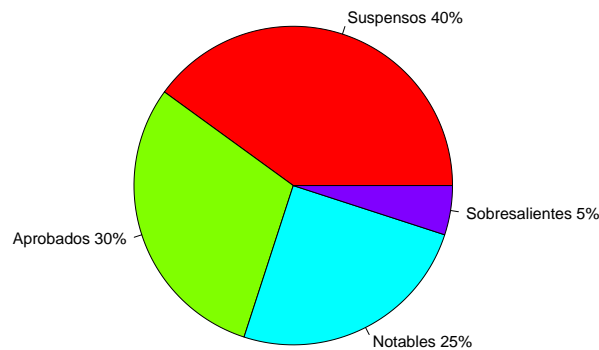
¿Según el coeficiente de variación, en cuál de las dos muestras es más representativa la media?

SOLUCIÓN

Llamando X a las estaturas de los niños recién nacidos e Y a las estaturas de los niños de 1 año: $\bar{x} = 50,4$ cm, $s_x = 1,685$ cm, $cv_x = 0,034$, $\bar{y} = 66,875$ cm, $s_y = 2,713$ cm, $cv_y = 0,041$, lo que indica que ambas medias son muy representativas pero un poco más la de los niños recién nacidos.

- ★ 6. El siguiente diagrama refleja el porcentaje de calificaciones obtenidas en un examen realizado a 80

alumnos:



Se pide:

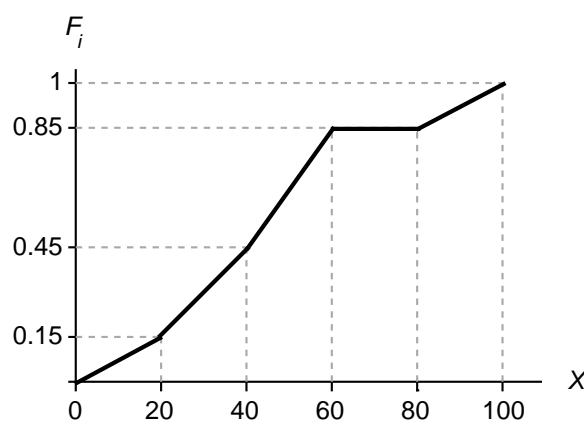
- Construir la tabla de frecuencias para las calificaciones.
- Dibujar el polígono de frecuencias acumuladas.
- Calcular todos los estadísticos de tendencia central que sean posibles.
- A partir de la variable calificación, construir la variable nota con los siguientes intervalos: Suspenso $[0, 5)$, Aprobado $[5, 7)$, Notable $[7, 9)$ y Sobresaliente $[9, 10]$, y calcular la nota media y estudiar su representatividad.

Nota: En los tres primeros apartados se debe trabajar con la variable calificación, mientras que en el último debe utilizarse la variable nota.

SOLUCIÓN

- $Me = \text{Aprobado}$ y $Mo = \text{Suspenso}$.
 - $\bar{x} = 5,275$ puntos, $s = 2,447$ puntos y $cv = 0,464$, de manera que la media es moderadamente representativa.
-

7. Dada la gráfica correspondiente a un polígono acumulativo de frecuencias relativas de una variable estadística agrupada en intervalos de una muestra de tamaño 20:



se pide:

- Construir la tabla de frecuencias.
- Dibujar el histograma correspondiente.
- Calcular la mediana y la moda.

d) Calcular la media aritmética y la desviación típica.

SOLUCIÓN

c) $Med = 42,5$ y $Mod = (40, 60)$.

d) $\hat{=44}$, $s^2 = 564$, $s = 23,75$.

- ★ 8. Sea la variable estadística agrupada en intervalos cuya distribución de frecuencias viene dada por la siguiente tabla:

Intervalos	n_i	f_i	N_i	F_i
[0, 10)	10	0.25		
[10, 20)			22	
[20, 30)		0.30		
[30, 40)				

a) Completar la tabla y hallar la desviación típica.

b) Calcular la mediana y el rango intercuartílico e interpretarlos.

SOLUCIÓN

a) $\bar{x} = 18,5$, $s^2 = 102,75$ y $s = 10,14$.

b) $Med = 18,33$, $C_1 = 10$, $C_3 = 26,27$ y $RI = 16,17$ lo que indica, teniendo en cuenta que el rango de toda la distribución es 40, que la dispersión es moderada.

9. Para determinar la eficacia de un nuevo método para la medición del hematocrito en sangre, se repitió la medida 8 veces sobre una misma muestra de sangre, obteniéndose los siguientes resultados (en porcentaje de hematocrito sobre volumen de plasma sanguíneo):

42,2 42,1 41,9 41,8 42 42,1 41,9 42.

¿Se puede afirmar que se trata de un buen método de medición?

SOLUCIÓN

$\bar{x} = 42\%$, $s^2 = 0,015\%^2$, $s = 0,1225\%$ y $cv = 0,003$ lo que indica que la variabilidad entre las mediciones es ínfima y por tanto se trata de un buen método de medición.

- ★ 10. Como parte de un proyecto de investigación, los investigadores obtuvieron los siguientes datos respecto a los niveles (en partes por millón) de cierto contaminante químico en 30 suelos diferentes de la Comunidad de Madrid:

3,09 6,06 7,34 5,32 4,29 5,36 6,01 7,84 3,87 5,23
 4,67 7,89 5,16 6,32 6,45 3,21 5,98 6,45 7,12 4,13
 5,16 3,04 4,56 5,67 5,98 6,23 7,34 5,32 4,21 7,13

Agrupar los datos en 5 clases de amplitud unidad, comenzando en 3, y sobre la distribución obtenida:

a) Calcular media, desviación típica y coeficiente de variación de los niveles de contaminante. Interpretar el coeficiente de variación.

- b) Calcular cuartiles de la distribución e interpretarlos.
- c) Dibujar el diagrama de caja y bigotes y comprobar si hay o no datos atípicos.

SOLUCIÓN

- a) $\bar{x} = 5,667$ ppm, $s^2 = 1,668$ ppm², $s = 1,292$ ppm y $cv = 0,228$, lo que indica que hay poca dispersión y la media es bastante representativa.
 - b) $C_1 = 4,7$ ppm, $C_2 = 5,667$ ppm y $C_3 = 6,75$ ppm.
 - c) Las vallas son $v_1 = 1,625$ y $v_2 = 9,825$. Todos los datos están entre las vallas y no hay datos atípicos. Los bigotes son $b_1 = 3,04$ ppm y $b_2 = 7,89$ ppm.
-

11. En cuestionario sobre la dependencia de las personas mayores de 75 años se preguntaba sobre la necesidad de ayuda en el desarrollo normal de su vida. Las posibles respuestas eran:

- a) Ninguna ayuda.
- b) Ayuda al subir las escaleras.
- c) Ayuda al subir las escaleras y al incorporarse de una posición sentada o tumbada.
- d) Ayuda al subir las escaleras, al incorporarse, y al vestirse.
- e) Ayuda para prácticamente todo.

El cuestionario lo respondieron 20 personas, y los resultados obtenidos fueron

b - d - a - b - c - c - b - c - d - e - a - b - c - e - a - b - c - d - b - b

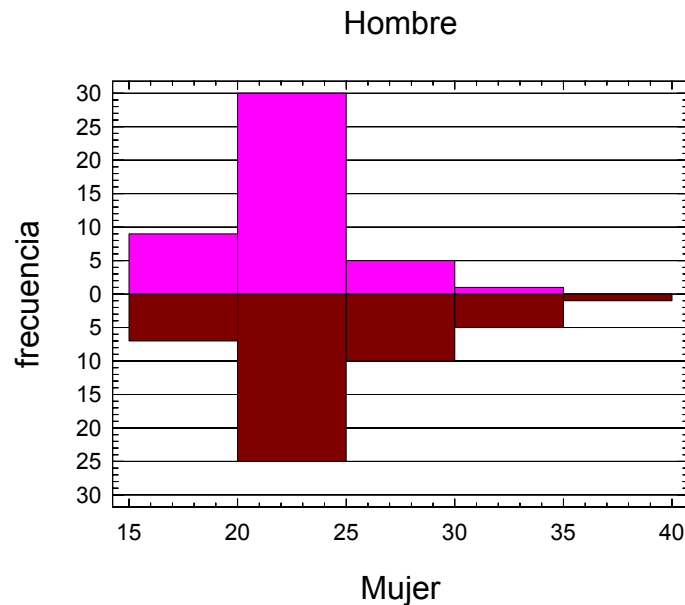
Se pide:

- a) Representar gráficamente la distribución de frecuencias.
- b) Calcular las medidas de tendencia central.
- c) Calcular los cuartiles y el decil 8.
- d) ¿Qué se puede decir sobre la dispersión?

SOLUCIÓN

- b) Me entre b y c y $Mo = b$.
 - c) $C_1 = b$, C_2 entre b y c, $C_3 = d$.
 - d) Suponiendo que hay la misma distancia entre categorías y asignando rangos a cada valor ($a = 1, b = 2, c = 3, d = 4, e = 5$), tenemos $\bar{x} = 2,7$, $s^2 = 1,41$, $s = 1,187$ y $cv = 0,44$, lo que indica una dispersión moderada.
-

- ★ 12. El siguiente histograma refleja la distribución del índice de masa corporal en una muestra de hombres y mujeres.



Se pide:

- Construir la tabla de frecuencias para hombres y mujeres por separado.
- Dibujar el diagrama de sectores para el sexo.
- ¿En qué grupo es más representativa la media? Justificar la respuesta.
- ¿Cómo calcularías la media de toda la muestra a partir de las medias de hombres y mujeres?
¿Cuánto vale?
- Calcular el rango intercuartílico del índice de masa corporal en los hombres.

SOLUCIÓN

- $\bar{m} = 24,17 \text{ Kg/m}^2$, $s_m^2 = 21,1806 \text{ (Kg/m}^2\text{)}^2$, $s_m = 4,6 \text{ Kg/m}^2$ y $cv_m = 0,19$.
 $\bar{h} = 22,18 \text{ Kg/m}^2$, $s_h^2 = 9,9506 \text{ (Kg/m}^2\text{)}^2$, $s_h = 3,15 \text{ Kg/m}^2$ y $cv_h = 0,14$, luego es más representativa la media en los hombres.
- $\bar{x} = 23,25 \text{ Kg/m}^2$.
- $RI = 3,75 \text{ Kg/m}^2$.

13. En un experimento se ha medido el tiempo de respuesta a un determinado estímulo visual en grupo de 16 individuos, obteniendo los siguientes resultados (en centésimas de segundo):

73, 42, 67, 55, 62, 52, 59, 82, 46, 61, 72, 51, 64, 55, 61.

Comprobar si hay datos atípicos en la muestra. En tal caso, sustituirlos por máximos o mínimos valores normales y hacer un resumen descriptivo de la muestra.

SOLUCIÓN

Los cuartiles son $C_1 = 52$ cs, $C_3 = 67$ cs y $RI = 15$ cs. Las vallas son $v_1 = 29,5$ y $v_2 = 89,5$, luego no hay datos atípicos. $\bar{x} = 60,13$ cs, $s^2 = 105,58 \text{ cs}^2$, $s = 10,28$ cs, $cv = 0,17$, $g_1 = 0,26$ y $g_2 = -0,37$.

14. En un grupo de operarios se ha medido porcentaje de tareas bien realizadas de dos tipos A y B, obteniendo los siguientes resultados:

% A	58	55	36	67	60	31	53	42
% B	76	81	82	84	80	75	87	79

Se pide:

- Calcular las puntuaciones típicas de cada individuo en ambos tipos de tareas.
- Teniendo en cuenta la distribuciones de puntuaciones de las tareas, ¿en qué tipo de tareas funciona mejor el primer individuo? ¿Y el último? Justificar la respuesta.

SOLUCIÓN

$\bar{x}_A = 50,25\%$, $\bar{x}_B = 80,5\%$, $s_A^2 = 138,44\%^2$, $s_B^2 = 13,75\%^2$, $s_A = 11,7\%$ y $s_B = 3,71\%$. Las puntuaciones típicas son

% A	0.66	0.4	-1.21	1.42	0.83	-1.64	0.23	-0.7
% B	-1.21	0.13	0.4	0.94	-0.13	-1.48	1.75	-0.4

2. Regresión y Correlación

15. Se ha realizado un estudio comparativo de las puntuaciones obtenidas por los alumnos en un test de ingreso en la universidad (X), y el número de asignaturas aprobadas en el primer curso (Y). Los resultados obtenidos se expresan en la siguiente tabla:

$X \setminus Y$	0	1	2	3	4
[0, 10)	2	2	1	0	0
[10, 20)	1	1	2	2	0
[20, 30)	0	1	3	4	1
[30, 40)	0	0	2	2	6

Se desea calcular:

- Recta de regresión de X sobre Y .
- Coefficiente de correlación e interpretación del mismo.
- Si la universidad en cuestión sólo contara con alumnos que al menos logren aprobar dos asignaturas, ¿qué número de preguntas respondidas correctamente exigirá en el test?

SOLUCIÓN

- $\bar{x} = 23$ puntos, $\bar{y} = 2,4$ asignaturas, $s_x^2 = 116$ puntos², $s_y^2 = 1,5733$ asignaturas², $s_x = 10,7703$ puntos, $s_y = 1,2453$ asignaturas y $s_{xy} = 9,8$ puntos·asignaturas.
Recta de regresión de X sobre Y : $x = 6,2288y + 8,0508$.
 - $r = 0,73$, lo que quiere decir que hay buena relación lineal entre las puntuaciones y las asignaturas aprobadas y además es creciente (a mayor puntuación en el test, más asignaturas aprobadas).
-

- ★ 16. Al realizar un estudio sobre la dosificación de un cierto medicamento, se trataron 6 pacientes con dosis diarias de 2 mg, 7 pacientes con 3 mg y otros 7 pacientes con 4 mg. De los pacientes tratados con 2 mg, 2 curaron al cabo de 5 días, y 4 al cabo de 6 días. De los pacientes tratados con 3 mg diarios, 2 curaron al cabo de 3 días, 4 al cabo de 5 días y 1 al cabo de 6 días. Y de los pacientes tratados con 4 mg diarios, 5 curaron al cabo de 3 días y 2 al cabo de 5 días.

Se pide:

- a) Dar el coeficiente de correlación e interpretación.
- b) Determinar el tiempo esperado de curación para una dosis de 5 mg diarios.

SOLUCIÓN

Llamando X a la dosis e Y al tiempo de curación:

- a) $\bar{x} = 3,05$ mg, $\bar{y} = 4,55$ días, $s_x^2 = 0,648$ mg², $s_y^2 = 1,448$ días², $s_x = 0,805$ mg, $s_y = 1,203$ días y $s_{xy} = -0,678$ mg·días.
 $r = -0,7$, que quiere decir que hay buena relación lineal entre la dosis y el tiempo de curación, y además es decreciente (a mayor dosis, menor tiempo de curación).
 - b) Recta de regresión del tiempo de curación sobre la dosis: $y = -1,046x + 7,741$.
 $y(5) = 2,511$ días.
-

17. Se determina la pérdida de actividad que experimenta un medicamento desde el momento de su fabricación a lo largo del tiempo, obteniéndose el siguiente resultado:

Tiempo (en años)	1	2	3	4	5
Actividad restante (%)	96	84	70	58	52

Se desea calcular:

- a) La relación fundamental (recta de regresión) entre actividad y tiempo transcurrido.
- b) El tiempo en meses que corresponde al 80 % de actividad.
- c) ¿Cuándo será nula la actividad?

SOLUCIÓN

Llamando T al tiempo y A a la actividad del fármaco:

- a) $\bar{t} = 3$ años, $\bar{a} = 72$ %, $s_t^2 = 2$ años², $s_a^2 = 264$ %², $s_{ta} = -22,8$ años·%.
Recta de regresión de actividad sobre tiempo: $a = -11,4t + 106,2$.
 - b) Recta de regresión de tiempo sobre actividad: $t = -0,086a + 9,2182$.
 $t(80) = 2,3091$ años.
 - c) $t(0) = 9,2182$ años.
-

18. Para comprobar el efecto de la herencia genética sobre la inteligencia se desarrolló un estudio en el que se midió el coeficiente intelectual de varias parejas de gemelos, obteniendo los siguientes resultados:

(128, 132) (116, 112) (86, 98) (65, 81) (104, 96) (111, 111) (101, 105) (72, 75)

Calcular el coeficiente de determinación lineal e interpretarlo. ¿Tiene sentido calcular el coeficiente de correlación?

SOLUCIÓN

Llamando X al coeficiente intelectual del primer hermano e Y al del segundo: $\bar{x} = 97,875$, $\bar{y} = 101,25$, $s_x^2 = 418,3594$, $s_y^2 = 288,4375$, $s_{xy} = 326,5313$ y $r^2 = 0,8836$, lo que indica que existe bastante relación lineal entre el coeficiente intelectual de los gemelos. No tiene sentido el coeficiente de correlación lineal porque es indiferente el orden en que tomemos a los gemelos.

19. Se consideran dos variables aleatorias X e Y tales que:

- La recta de regresión de Y sobre X viene dada por la ecuación: $y - x - 2 = 0$.
- La recta de regresión de X sobre Y viene dada por la ecuación: $y - 4x + 22 = 0$.

Calcular:

- a) Valores de \bar{x} e \bar{y} .
- b) Coeficiente de correlación lineal.

SOLUCIÓN

- a) $\bar{x} = 8$ y $\bar{y} = 10$.
 - b) $r = 0,5$.
-

20. En el ajuste rectilíneo a una distribución bidimensional se sabe que $\bar{x} = 2$, $\bar{y} = 1$, y el coeficiente de correlación lineal es 0 ($r = 0$).

- a) Si $x = 10$, ¿cuál será el valor interpolado para y ?
- b) Si $y = 5$, ¿cuál será el valor interpolado para x ?
- c) Dibuja las rectas de regresión de Y sobre X , y la de X sobre Y .

SOLUCIÓN

- a) $y(10) = 1$.
 - b) $x(5) = 2$.
-

- ★ 21. En un centro dietético se está probando una nueva dieta de adelgazamiento en una muestra de 12 individuos. Para cada uno de ellos se ha medido el número de días que lleva con la dieta y el número de kilos perdidos desde entonces, obteniéndose los siguientes resultados:

(33 , 3.9), (51 , 5.9), (30 , 3.2), (55 , 6.0), (38 , 4.9), (62 , 6.2),
 (35 , 4.5), (60 , 6.1), (44 , 5.6), (69 , 6.2), (47 , 5.8), (40 , 5.3)

Se pide:

- a) Dibujar el diagrama de dispersión. Según la nube de puntos, ¿qué tipo de modelo explicaría mejor la relación entre los días de dieta y los kilos perdidos?
- b) Dibujar el diagrama de dispersión tomando una escala logarítmica para los días de dieta.
- c) Calcular el modelo lineal y el logarítmico de los kilos perdidos con respecto a los días de dieta. Nota: Utilizar los datos muestrales sin agrupar.
- d) Utilizar el mejor de los modelos anteriores para predecir en número de kilos perdidos tras 40 días de dieta y tras 100 días. ¿Son fiables estas predicciones?

SOLUCIÓN

Llamando X a los días de dieta, Y a los Kg perdidos y $Z = \log X$.

- c) $\bar{x} = 47$ días, $\bar{y} = 5,3$ Kg, $s_x^2 = 143,833$ días², $s_y^2 = 0,885$ Kg², $s_{xy} = 9,942$ días·Kg. Modelo lineal: $y = 0,069x + 2,051$.
 $\bar{z} = 3,82$ logdías, $s_z^2 = 0,07$ log²días, $s_{yz} = 0,22$ logdías·Kg.
 Modelo logarítmico: $y = 3,4 \log y - 7,67$.

- d) Modelo lineal: $r^2 = 0,78$, modelo logarítmico: $r^2 = 0,86$.
Predicciones con el modelo logarítmico: $y(40) = 4,86$ Kg y $y(100) = 7,98$ Kg. Las predicciones son fiables ya que el coeficiente de determinación es alto.

- ★ 22. Se han medido dos variables S y T en 10 individuos, obteniéndose los siguientes resultados:

$(-1.5, 2.25), (0.8, 0.64), (-0.2, 0.04), (-0.8, 0.64), (0.4, 0.16),$
 $(0.2, 0.04), (-2.1, 4.41), (-0.4, 0.16), (1.5, 2.25), (2.1, 4.41).$

Se pide:

- Calcular la covarianza de S y T .
- ¿Se puede afirmar que S y T son independientes? Justificar la respuesta.
- ¿Qué valor predice la correspondiente recta de regresión para $t = 2$?

SOLUCIÓN

- $\bar{s} = 0, \bar{t} = 1,5$ y $s_{st} = 0$.
- No podemos afirmar que S y T son independientes, sólo se puede afirmar que no hay relación lineal.
- $s(2) = 0$.

23. En un experimento se ha medido el número de bacterias por unidad de volumen en un cultivo, cada hora transcurrida, obteniendo los siguientes resultados:

Horas	0	1	2	3	4	5	6	7	8
Nº Bacterias	25	28	47	65	86	121	190	290	362

Se pide:

- Dibujar el diagrama de dispersión. Según este diagrama, ¿qué tipo de modelo explicaría mejor la relación entre el número de bacterias y las horas transcurridas?
- Dibujar el diagrama de dispersión tomando una escala logarítmica para el número de bacterias.
- Según el modelo anterior, ¿Cuántas bacterias tendríamos al cabo de 3 horas y media? ¿Y al cabo de 10 horas? ¿Son fiables estas predicciones?
- ¿Cuánto tiempo tendría que transcurrir para que en el cultivo hubiese 100 bacterias?

SOLUCIÓN

Llamando X a las horas, Y a las bacterias y Z al logaritmo neperiano de las bacterias:

- $\bar{x} = 4$ horas, $\bar{z} = 4,5149 \log(\text{bacterias})$, $s_x^2 = 6,6667$ horas², $s_z^2 = 0,8361 \log^2(\text{bacterias})$ y $s_{xz} = 2,3466$ horas $\cdot\log(\text{bacterias})$.
Modelo lineal del logaritmo de las bacterias sobre las horas: $z = 0,3520x + 3,1070$.
Modelo exponencial de las bacterias sobre las horas: $y = e^{0,3520x+3,1070}$.
 $y(3,5) = 76,6254$ bacterias y $y(10) = 755,0986$ bacterias.
- Modelo lineal de las horas sobre el logaritmo de las bacterias: $x = 2,8218z - 8,7403$.
Modelo logarítmico de las horas sobre las bacterias: $x = 2,8218 \log y - 8,7403$.
 $x(100) = 4,25$ horas.

- ★ 24. La concentración de un fármaco en sangre, C en mg/dl, es función del tiempo, t en horas, y viene dada por la siguiente tabla:

t	2	3	4	5	6	7	8
C	25	36	48	64	86	114	168

- a) Según el modelo exponencial, ¿qué concentración de fármaco habría a las 4,8 horas? ¿Es fiable la predicción? Justificar adecuadamente la respuesta.
- b) Según el modelo lineal, ¿qué tiempo tendría que transcurrir para que la concentración de fármaco fuese de 100 mg/dl? ¿Es fiable la predicción? Justificar adecuadamente la respuesta.

SOLUCIÓN

Llamando T al tiempo, C a la concentración y Z al logaritmo de la concentración:

- a) $\bar{t} = 5$ horas, $\bar{z} = 4,1639 \log(\text{mg/dl})$, $s_t^2 = 4$ horas², $s_z^2 = 0,3785 \log^2(\text{mg/dl})$, $s_{tz} = 1,2291$ horas $\cdot\log(\text{mg/dl})$.
 Modelo exponencial de C sobre T : $c = e^{0,3073x+2,6275}$.
 $c(4,8) = 60,498$ mg/dl y es bastante fiable ya que $r^2 = 0,999$.
- b) $\bar{c} = 77,2857$ mg/dl, $s_c^2 = 2160,7755$ (mg/dl)², $s_{tc} = 89$ horas(mg/dl).
 Modelo lineal de T sobre C : $t = 0,0412c + 1,8167$.
 $t(100) = 5,9356$ y también es fiable ya que $r^2 = 0,9165$.
-

25. Se han recogido por medio de unos cuestionarios los niveles de estrés y energía de 14 mujeres durante un año. A partir de las respuestas del cuestionario se han asignado puntuaciones a cada una de ellas de manera que a mayor puntuación mayor grado de estrés y energía. Los datos recogidos son:

Edad	21	31	19	21	30	20	22	23	45	24	26	19	25	21
Estrés	25	19	20	19	24	6	29	25	49	0	10	25	13	23
Energía	25	20	45	60	50	50	10	60	40	60	50	60	85	50

Se pide:

- a) Dibujar un diagrama de dispersión que refleje la relación entre el estrés y la energía.
- b) ¿Existe relación lineal entre el estrés y la energía? ¿Y entre el estrés y la edad? Justificar la respuesta.
- c) ¿Qué efecto tendría sobre el coeficiente de correlación lineal la eliminación del individuo de 45 años? Justificar la respuesta.
- d) Calcular el coeficiente de correlación de Spearman entre estrés y energía e interpretarlo. ¿Coinciden las conclusiones con las que se deducen del coeficiente de correlación lineal?
26. En un estudio sobre la búsqueda visual se realiza un prueba que consiste en presentarle a un sujeto una matriz de n símbolos y pedirle que pulse rápidamente un botón si entre los símbolos se encuentra uno concreto, u otro botón diferente si no aparece dicho símbolo. El tiempo de respuesta de cada participante (en centésimas de segundo) y el número de símbolos de cada matriz aparecen en la siguiente tabla:

Matrices con el símbolo	n	4	5	6	7	8	9	10	11	12
	T	22	24	23	31	33	45	42	46	50
Matrices sin el símbolo	n	4	5	6	7	8	9	10	11	
	T	25	24	32	35	43	49	52	56	

Se pide:

- a) Construir la recta de regresión del tiempo de respuesta sobre el número de símbolos para las matrices con el símbolo y también para las matrices sin el símbolo.

- b) ¿En qué matrices, las que tienen el símbolo o las que no, explica mejor el tiempo de respuesta el número de símbolos? Justificar la respuesta.
- c) Según los modelos anteriores, ¿cuánto tiempo tardará en responder una persona elegida al azar en una matriz de 20 símbolos que contenga al símbolo? ¿Y si no lo contuviese?

SOLUCIÓN

Llamando X al número de símbolos e Y al tiempo de respuesta:

- a) Matrices con el símbolo: $\bar{x} = 8$ símbolos, $\bar{y} = 35,1111$ seg, $s_x^2 = 6,6667$ símbolos², $s_y^2 = 104,4321$ seg², $s_{xy} = 25,4446$ símbolos·seg.
 Recta de regresión del tiempo sobre el número de símbolos: $y = 3,8333x + 4,4444$. Matrices sin el símbolo: $\bar{x} = 7,5$ símbolos, $\bar{y} = 39,5$ seg, $s_x^2 = 5,25$ símbolos², $s_y^2 = 132,25$ seg², $s_{xy} = 26$ símbolos·seg.
 Recta de regresión del tiempo sobre el número de símbolos: $y = 4,9525x + 2,3571$.
- b) $r^2 = 0,9292$ en las matrices con el símbolo y $r^2 = 0,9736$ en las matrices sin el símbolo, así que el número de símbolos explica un poco mejor el tiempo de respuesta en las matrices sin el símbolo.
- c) $y(20) = 81,11$ seg si la matriz contiene el símbolo y $y(20) = 101,4$ seg si la matriz no contiene el símbolo.
-

27. Se ha realizado un estudio para averiguar la relación entre la edad y la fuerza física. Para ello se ha medido la edad de 16 participantes y el máximo peso (en Kg) que eran capaces de levantar. Los resultados obtenidos fueron:

Edad	10	12	15	18	21	23	24	26	28	30	31	34	36	38	41	44
Peso	12	25	36	46	54	60	61	60	59	56	54	52	50	50	48	46

Construir un modelo de regresión que explique la relación entre la fuerza física y el peso e interpretarlo.

SOLUCIÓN

Llamando X a la edad e Y al peso levantado, se construyen dos rectas de regresión, una para edades menores de 25 y otra para mayores:

- Menores de 25: $\bar{x} = 15,5714$ años, $\bar{y} = 42$ Kg, $s_x^2 = 25,3878$ años², $s_y^2 = 295,7143$ Kg², $s_{xy} = 85,7143$ años·Kg.
 Recta de regresión del peso sobre la edad: $y = 3,3762x - 17,3248$.
 - Mayores de 25: $\bar{x} = 35,2222$ años, $\bar{y} = 52,7778$ Kg, $s_x^2 = 32,6173$ años², $s_y^2 = 20,8395$ Kg², $s_{xy} = -25,5062$ años·Kg.
 Recta de regresión del peso sobre la edad: $y = -0,7820x + 79,5390$.
-

28. Para evaluar la capacidad de aprendizaje en la realización de una tarea, se ha medido el tiempo que tarda en realizarse una tarea en sucesivas repeticiones de la misma. Los resultados obtenidos son:

Repetición	1	2	3	4	5	6	7	8	9	10
Tiempo (min)	80	65	56	50	48	43	41	38	37	35

Se pide:

- a) Dibujar el diagrama de dispersión.
- b) En vista del diagrama de dispersión, construir el modelo de regresión más adecuado del tiempo en función de las repeticiones.

- c) ¿Qué porcentaje de la variabilidad del tiempo explican las repeticiones?
 d) ¿Cuanto tiempo tardará por término medio en la 5 repetición de la tarea?

SOLUCIÓN

Llamando X a las repeticiones, Y al tiempo y Z al logaritmo neperiano del tiempo, se tien:

- b) $\bar{x} = 5,5$ repeticiones, $\bar{z} = 3,8644 \ln(\text{min})$, $s_x^2 = 8,25$ repeticiones², $s_z^2 = 0,0637 \ln^2(\text{min})$ y $s_{xz} = -0,7014$ repeticiones $\cdot\ln(\text{min})$.

Recta de regresión del logaritmo del tiempo sobre las repeticiones: $z = -0,085x + 4,3320$.

Modelo exponencial del tiempo sobre las repeticiones: $y = e^{-0,085x+4,3320}$.

- c) $R^2 = 0,9364$, es decir, un 93,64 %.
 d) $y(5) = 49,74$ min.
-

29. En un estudio se ha preguntado a un grupo de personas sobre su ideología política X (izquierda, centro o derecha) y su opinión sobre la subida o bajada de impuestos Y , obteniendo la siguiente tabla de frecuencias:

$X \backslash Y$	Bajada	Mantenimiento	Subida
Izquierda	2	6	8
Centro	3	4	3
Derecha	6	5	3

¿Se puede concluir que existe relación entre la ideología y la opinión sobre la subida o bajada de impuestos? Justificar la respuesta.

SOLUCIÓN

$\chi^2 = 4,4$ y $C = 0,49$ lo que indica que existe bastante relación entre las variables.

30. Un estudio sobre 100 personas concluye que 26 personas son fumadores y bebedores habituales, 12 son bebedores pero no fumadores, 18 son fumadores pero no bebedores y 44 no beben ni fuman habitualmente. Según estos datos, ¿podemos decir que existe relación entre el tabaco y la bebida? Justificar la respuesta.

SOLUCIÓN

$\chi^2 = 14,83$ y $C = 0,36$ lo que indica que hay una relación moderada entre los hábitos de fumar y beber.

31. En un estudio en el que participaron las 8 universidades de una región se ha valorado la excelencia docente e investigadora, estableciendo los siguientes rankings (de mejor a peor):

Ranking docencia	3	4	8	5	2	1	6	7
Ranking investigación	6	5	4	3	7	8	1	2

¿Se puede decir que existe relación entre la excelencia docente e investigadora? Justificar la respuesta.

SOLUCIÓN

$r_s = -0,83$, lo que indica una fuerte relación inversa entre la excelencia docente y la excelencia investigadora.

3. Probabilidad

32. Se dispone de dos urnas, la primera con 10 bolas blancas y 6 bolas negras, y la segunda con 5 bolas rojas, 8 bolas azules y 3 bolas verdes. Construir el espacio muestral del experimento que consiste en sacar una bola de cada urna, y del experimento que consistiría en sacar dos bolas de cada urna.
33. Un experimento consiste en seleccionar a una pareja de personas y medir su grupo sanguíneo y si son fumadores o no. Expresar el espacio muestral en forma de árbol.
34. En una estantería en la que hay 3 cajas de un medicamento A y 2 de un medicamento B , se eligen 3 al azar. ¿Cuál es la probabilidad de que se hayan elegido 2 cajas del medicamento A y 1 del B ?

SOLUCIÓN

36/60.

35. En un laboratorio hay 4 frascos de ácido sulfúrico y 2 de ácido nítrico, y en otro hay 1 frascos de ácido sulfúrico y 3 de ácido nítrico. Se saca al azar un frasco de cada laboratorio. Hallar la probabilidad de que:
- a) Los dos frascos sean de ácido sulfúrico.
 - b) Los dos sean de ácido nítrico.
 - c) Uno sea de ácido sulfúrico y otro de ácido nítrico.
 - d) Calcular la probabilidad de estos mismos sucesos si el frasco elegido en el primer laboratorio se introduce en el segundo antes de sacar el frasco de este.

SOLUCIÓN

- a) 4/24.
- b) 6/24.
- c) 14/24.
- d) 8/30, 8/30 y 14/30 respectivamente.

36. Sean A y B sucesos de un mismo espacio muestral tales que: $P(A) = 3/8$, $P(B) = 1/2$, $P(A \cap B) = 1/4$. Calcular:
- a) $P(A \cup B)$.
 - b) $P(\overline{A})$ y $P(\overline{B})$.
 - c) $P(\overline{A} \cap \overline{B})$.
 - d) $P(A \cap \overline{B})$.
 - e) $P(A/B)$.
 - f) $P(A/\overline{B})$.

SOLUCIÓN

- a) $P(A \cup B) = 5/8$.
- b) $P(\overline{A}) = 5/8$ y $P(\overline{B}) = 1/2$.
- c) $P(\overline{A} \cap \overline{B}) = 3/8$.
- d) $P(A \cap \overline{B}) = 1/8$.
- e) $P(A/B) = 1/2$.

$$f) P(A/\overline{B}) = 1/4.$$

37. La probabilidad de contraer hepatitis a partir de una unidad de sangre es 0'01. Un paciente recibe dos unidades de sangre durante su estancia en el hospital. ¿Cuál es la probabilidad de que contraiga hepatitis como consecuencia de ello?

SOLUCIÓN

0,0199.

38. Sean A y B sucesos de un mismo espacio muestral, tales que $P(A) = 0,6$ y $P(A \cup B) = 0,9$. Calcular $P(B)$ si:

- a) A y B son independientes.
b) A y B son incompatibles.

SOLUCIÓN

- a) $P(B) = 0,75$.
b) $P(B) = 0,3$.
-

39. El tétanos es mortal en el 70 % de los casos. Si tres personas contraen el tétanos, ¿Cuál es la probabilidad de que mueran al menos dos de los tres?

SOLUCIÓN

0,784.

- ★ 40. En un estudio sobre el tabaco, se informa que el 40 % de los fumadores tiene un padre fumador, el 25 % tiene una madre fumadora, y el 52 % tiene al menos uno de los dos padres fumadores. Se elige una persona fumadora al azar. Calcular:

- a) Probabilidad de que la madre sea fumadora si lo es el padre.
b) Probabilidad de que la madre sea fumadora si no lo es el padre.
c) ¿Son independientes el tener padre fumador y el tener madre fumadora?

SOLUCIÓN

Llamando PF al susceso que consiste en tener un padre fumador y MF a tener una madre fumadora:

- a) $P(MF/PF) = 0,33$.
b) $P(MF/\overline{PF}) = 0,2$.
c) No son independientes.
-

- ★ 41. Tras observar los resultados de la prueba de selectividad se sabe que el 40 % de los alumnos aprueba el examen de Matemáticas, el 30 % el examen de Física y el 55 % suspenden los dos. Si se elige un alumno al azar, calcular:

- a) Probabilidad de que haya aprobado al menos uno de los dos exámenes.
- b) Probabilidad de que haya aprobado Matemáticas si ha aprobado Física.
- c) Probabilidad de que haya aprobado Física si ha suspendido Matemáticas.
- d) ¿Son independientes aprobar Matemáticas y aprobar Física?

SOLUCIÓN

Llamando M al suceso correspondiente a aprobar Matemáticas y F a aprobar Física:

- a) $P(M \cup F) = 0,45$.
 - b) $P(M/F) = 0,83$.
 - c) $P(F/\overline{M}) = 0,08$.
 - d) No son independientes.
-

- ★ 42. En un servicio clínico digestivo se sabe que, de cada 1000 pacientes con dolor de estómago, 700 presentan gastritis, 200 presentan úlcera y 100 presentan cáncer. En el análisis de la sintomatología gástrica, se ha comprobado que las probabilidades de presentar vómitos son 0,3 en el caso de gastritis, 0,6 en el caso de úlcera y 0,9 en el caso de cáncer. Llega un nuevo paciente con dolor de estómago que, además, presenta vómitos. ¿Qué diagnosticaríamos?

SOLUCIÓN

Llamando G a tener gastritis, U a tener úlcera, C a tener cáncer y V a tener vómitos, $P(G/V) = 0,5$, $P(U/V) = 0,286$ y $P(C/V) = 0,214$, de modo que se diagnosticaría gastritis.

- ★ 43. Un estudiante se somete a un examen de tipo test en el que cada pregunta tiene 3 respuestas posibles. El estudiante se sabe el 40 % de las preguntas, y el resto las contesta al azar. Se elige al azar una pregunta. ¿Qué probabilidad hay de que no la supiera si la contestó correctamente?

SOLUCIÓN

1/3.

- ★ 44. Se ha desarrollado un nuevo test diagnóstico para detectar el síndrome de Down en niños recién nacidos, con un sensibilidad del 80 % y una especificidad del 90 %. Si en una determinada población en la que hay un 1 % de recién nacidos con el síndrome, al aplicarle el test a un niño, da positivo, ¿cuál es la probabilidad de que tenga el síndrome? ¿le diagnosticarías la enfermedad? ¿Cuál debería ser la especificidad mínima del test para diagnosticar el síndrome en el caso de dar positivo?

Nota: La *sensibilidad* de un test diagnóstico es la proporción de personas con la enfermedad que tienen un resultado positivo en el test, mientras que la *especificidad* del test es la proporción de personas sin la enfermedad que tienen un resultado negativo en el test.

SOLUCIÓN

Llamando S a tener el síndrome de Down y $+$ a que el test de positivo, $P(S/+) = 0,0748$ y $P(\overline{S}/+) = 0,9252$, de modo que no se diagnosticaría el síndrome al ser más probable que no lo tenga. La especificidad mínima para que el test diagnostique el síndrome es $P(-/\overline{S}) = 0,9919$.

- ★ 45. En un estudio se han probado tres tipos de tratamientos A , B y C contra una determinada enfermedad. De los pacientes participantes en el estudio, el 50 % fueron tratados con el tratamiento A , el 30 % con el B y el 20 % con el C . Posteriormente se observaron los pacientes que sanaron y los que tuvieron algún efecto secundario, según se muestra en la siguiente tabla:

Tratamiento	Sanados	Con efectos secundarios
A	86 %	12 %
B	92 %	14 %
C	81 %	6 %

Se pide:

- Si se selecciona un enfermo al azar, ¿cuál es la probabilidad de que haya sanado? ¿Y de que haya tenido algún efecto secundario?
- Si un enfermo ha sanado, ¿qué tratamiento es más probable que haya recibido? ¿Y si en vez de decirnos que ha sanado nos dicen que no ha tenido efectos secundarios?
- Si en total hay un 8 % pacientes que no sanaron pero que tampoco tuvieron efectos secundarios, ¿cuál es la probabilidad de que un enfermo se haya curado sin tener efectos secundarios?

SOLUCIÓN

Llamado S a sanar y E a tener efectos secundarios:

- $P(S) = 0,868$ y $P(E) = 0,114$.
 - $P(A/S) = 0,495$, $P(B/S) = 0,318$ y $P(C/S) = 0,187$.
 $P(A/\bar{E}) = 0,497$, $P(B/\bar{E}) = 0,291$ y $P(C/\bar{E}) = 0,212$.
 En ambos casos el tratamiento más probable es el A .
 - $P(S \cap \bar{E}) = 0,806$.
-

- ★ 46. En una población se ha vacunado a la tercera parte de los individuos contra la gripe. Trascurrido el invierno, se comprueba que la probabilidad de estar vacunado si se tiene la gripe es 0,2, y que el 10 % de los vacunados tuvieron gripe.

- ¿Cuál fue la incidencia de la epidemia de gripe?
(Nota: La incidencia de una epidemia es la probabilidad de personas infectadas).
- ¿Cuál es la probabilidad de que una persona no vacunada contraiga la gripe?
- ¿Se puede afirmar que la vacuna tiene alguna eficacia?

SOLUCIÓN

Llamando G al suceso consistente en tener la gripe y V a estar vacunado:

- $P(G) = 1/6$.
 - $P(G/\bar{V}) = 0,2$.
 - Si resulta eficaz, aunque poco.
-

47. Se sabe que los grupos sanguíneos en una determinada población se distribuyen con las siguientes frecuencias:

$$0 : 30 \% \quad A : 45 \% \quad B : 18 \% \quad AB : 7 \%$$

Por otro lado, también se sabe que la octava parte de los individuos del grupo 0 tienen RH negativo, así como la cuarta parte del grupo A , la mitad del grupo B , y la tercera parte del grupo AB .

Se pide:

- a) ¿Cuál es la probabilidad de que un individuo elegido al azar sea del tipo A y tenga RH positivo?
- b) ¿Cuál es la probabilidad de que un individuo elegido al azar tenga RH negativo o sea del grupo universal?
- c) Si un individuo elegido al azar tiene RH positivo, ¿cuál es la probabilidad de que pertenezca al grupo B ?

SOLUCIÓN

Llamando 0 a tener grupo 0 , A a tener grupo A , B a tener grupo B , AB a tener grupo AB , $+$ a tener RH positivo y $-$ a tener RH negativo:

- a) $P(A \cap +) = 0,34$.
- b) $P(- \cup 0) = 0,53$.
- c) $P(B/+) = 0,12$.
-

48. Sabemos que el test de Bender para detectar alteraciones cerebrales tiene una sensibilidad del 88 % y una especificidad del 96 %. Por otro lado, sabemos que en una población la probabilidad de que un individuo elegido al azar tenga alteraciones cerebrales y además de positivo en el test es 0,08. Se pide:

- a) Calcular el porcentaje de personas que tienen alteraciones cerebrales en la población.
- b) ¿Es efectivo el test en esta población para detectar la ausencia de alteraciones cerebrales?

SOLUCIÓN

Llamando E a presentar alteraciones cerebrales y $+$ y $-$ a que el test de positivo y negativo respectivamente:

- a) $P(E) = 0,0901$, es decir, un 9,09 %.
- b) $P(\overline{E}/-) = 0,99$, lo cual indica que es muy efectivo para detectar la ausencia de alteraciones cerebrales.
-

49. Los estudios epidemiológicos indican que el 20 % de las personas mayores sufren un deterioro neuropsicológico. También se sabe que la tomografía axial computerizada (TAC) puede detectar ese trastorno en el 90 % de los que lo padecen, pero que también puede diagnosticarlo en el 5 % de las personas que no lo tienen. Si se toma una persona al azar y el TAC da positivo, ¿cuál es la probabilidad de que realmente esté enfermo?

SOLUCIÓN

Llamando E a sufrir el deterioro neuropsicológico y $+$ a que el TAC de positivo: $P(E/+) = 0,82$.

4. Variables Aleatorias

50. Sea X una variable aleatoria discreta cuya ley de probabilidad es

X	4	5	6	7	8
$f(x)$	0,15	0,35	0,10	0,25	0,15

- a) Calcular y representar gráficamente la función de distribución.
- b) Obtener:

- 1) $P(X < 7,5)$.
- 2) $P(X > 8)$.
- 3) $P(4 \leq X \leq 6,5)$.
- 4) $P(5 < X < 6)$.

SOLUCIÓN

a)

$$F(x) = \begin{cases} 0 & \text{si } x < 4, \\ 0,15 & \text{si } 4 \leq x < 5, \\ 0,5 & \text{si } 5 \leq x < 6, \\ 0,6 & \text{si } 6 \leq x < 7, \\ 0,85 & \text{si } 7 \leq x < 8, \\ 1 & \text{si } 8 \leq x. \end{cases}$$

b) $P(X < 7,5) = 0,85$, $P(X > 8) = 0$, $P(4 \leq x \leq 6,5) = 0,6$ y $P(5 < X < 6) = 0$.

51. Sea la variable aleatoria X con la siguiente función de distribución:

$$F(x) = \begin{cases} 0 & \text{si } x < 1, \\ 1/5 & \text{si } 1 \leq x < 4, \\ 3/4 & \text{si } 4 \leq x < 6, \\ 1 & \text{si } 6 \leq x. \end{cases}$$

Se pide:

- a) Distribución de probabilidad.
- b) Obtener:
 - 1) $P(X = 6)$.
 - 2) $P(X = 5)$.
 - 3) $P(2 < X < 5,5)$.
 - 4) $P(0 \leq X < 4)$.

SOLUCIÓN

a)

X	1	4	6
$f(x)$	1/5	11/20	1/4

b) $P(X = 6) = 1/4$, $P(X = 5) = 0$, $P(2 < X < 5,5) = 11/20$ y $P(0 \leq X < 4) = 1/5$.

★ 52. Se realiza un experimento aleatorio consistente en inyectar un virus a tres tipos de ratas y observar si sobreviven o no. Se sabe que la probabilidad de que viva el primer tipo de rata es 0,5, la de que viva el segundo es 0,4 y la de que viva el tercero 0,3. Se pide:

- a) Construir la variable aleatoria que mida el número de ratas vivas y su función de probabilidad.
- b) Calcular la función de distribución.
- c) Calcular $P(X \leq 1)$, $P(X \geq 2)$ y $P(X = 1,5)$.
- d) Calcular la media y la desviación típica.

SOLUCIÓN

a)

X	0	1	2	3
$f(x)$	0,21	0,44	0,29	0,06

b)

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ 0,21 & \text{si } 0 \leq x < 1, \\ 0,65 & \text{si } 1 \leq x < 2, \\ 0,94 & \text{si } 2 \leq x < 3, \\ 1 & \text{si } 3 \leq x. \end{cases}$$

c) $P(X \leq 1) = 0,65$, $P(X \geq 2) = 0,35$ y $P(X = 1,5) = 0$.d) $\mu = 1,2$ ratas, $\sigma^2 = 0,7$ ratas² y $\sigma = 0,84$ ratas.

-
53. La probabilidad de curación de un paciente al ser sometido a un determinado tratamiento es 0,85. Calcular la probabilidad de que en un grupo de 6 enfermos sometidos a tratamiento:

a) se curen la mitad.

b) se curen al menos 4.

SOLUCIÓN

Llamando X al número de pacientes curados de los 6 sometidos al tratamiento, se tiene que $X \sim B(6, 0,85)$.

a) $P(X = 3) = 0,041$.b) $P(X \geq 4) = 0,9526$.

-
54. Se sabe que la probabilidad de que aparezca una bacteria en un mm³ de cierta disolución es de 0,002. Si en cada mm³ a los sumo puede aparecer una bacteria, determinar la probabilidad de que en un cm³ haya como máximo 5 bacterias.

SOLUCIÓN

Llamando X al número de bacterias en 1 cm³ de disolución, se tiene $X \sim B(1000, 0,002) \approx P(2)$. $P(X \leq 5) = 0,9834$.

-
55. El trastorno de pánico aparece en 1 de cada 75 personas. ¿Cuál es la probabilidad de que en un grupo de 100 personas aparezca alguna con trastorno de pánico? ¿Cuál es el número esperado de personas con trastorno de pánico en ese grupo?

SOLUCIÓN

Llamando X al número de personas que sufren trastorno del pánico en el grupo de 100 personas, se tiene que $X \sim B(100, 1/75)$ y $P(X \geq 1) = 0,7379$.

-
- ★ 56. El número medio de llamadas por minuto que llegan a una centralita telefónica es igual a 120. Hallar las probabilidades de los sucesos siguientes:

- a) $A = \{\text{durante 2 segundos lleguen a la centralita menos de 4 llamadas}\}.$
 b) $B = \{\text{durante 3 segundos lleguen a la centralita 3 llamadas como mínimo}\}.$

SOLUCIÓN

- a) Si X es el número de llamadas en 2 segundos, entonces $X \sim P(4)$ y $P(X < 4) = 0,4335.$
 b) Si Y es el número de llamadas en 3 segundos, entonces $Y \sim P(6)$ y $P(Y \geq 3) = 0,938.$
-

57. Una mecanógrafa comete, en promedio, una errata cada 2000 caracteres que escribe. Suponiendo que escribe un folio con treinta líneas y setenta caracteres por línea, ¿cuál es la probabilidad de que cometa más de un error en dicho folio?

SOLUCIÓN

Llamando X al número de errores en un folio, se tiene que $X \sim B(2100, 1/2000) \approx P(1,05)$, y $P(X > 1) = 0,2826.$

58. Un examen de tipo test consta de 10 preguntas con tres respuestas posibles para cada una de ellas. Se obtiene un punto por cada respuesta acertada y se pierde medio punto por cada pregunta fallada. Un alumno sabe tres de las preguntas del test y las contesta correctamente, pero no sabe las otras siete y las contesta al azar. ¿Qué probabilidad tiene de aprobar el examen?

SOLUCIÓN

Llamando X al número de preguntas acertadas de las 7 contestadas al azar, se tiene $X \sim B(7, 1/3)$ y $P(X \geq 4) = 0,1733.$

- ★ 59. En un estudio sobre un determinado tipo de parásito que ataca el riñón de las ratas, se sabe que el número medio de parásitos en cada riñón es 3. Se pide:

- a) Calcular la probabilidad de que una rata tenga más de 8 parásitos.
 (Nota: se supone que una rata normal tiene dos riñones).
 b) Si se tienen 10 ratas, ¿cuál es la probabilidad de que haya al menos 9 con parásitos?

SOLUCIÓN

- a) Si X es el número de parásitos en una rata, $X \sim P(6)$ y $P(X > 8) = 0,1528.$
 b) Si Y es el número de ratas con parásitos en un grupo de 10 ratas, entonces $Y \sim B(10, 0,9975)$ y $P(Y \geq 9) = 0,9997.$
-

- ★ 60. En un servicio de urgencias de cierto hospital se sabe que, en media, llegan 2 pacientes a la hora. Calcular:

- a) Si los turnos en urgencias son de 8 horas, ¿cuál será la probabilidad de que en un turno lleguen más de 5 pacientes?
 b) Si el servicio de urgencias tiene capacidad para atender adecuadamente como mucho a 4 pacientes a la hora, ¿cuál es la probabilidad de que a lo largo de un turno de 8 horas el servicio de urgencias se vea desbordado en alguna de las horas del turno?

SOLUCIÓN

- a) Llamando X al número de pacientes en un turno, se tiene que $X \sim P(16)$ y $P(X > 5) = 0,9986$.
- b) Llamando Y al número de horas en el que el servicio se vea desbordado porque lleguen más de 4 pacientes, se tiene que $Y \sim B(8, 0,0527)$ y $P(Y \geq 1) = 0,3515$.

- ★ 61. El síndrome de Turner es una anomalía genética que se caracteriza porque las mujeres tienen sólo un cromosoma X . Afecta aproximadamente a 1 de cada 2000 mujeres. Además, aproximadamente 1 de cada 10 mujeres con síndrome de Turner, como consecuencia, también sufren un estrechamiento anormal de la aorta. Se pide:

- a) En un grupo de 4000 mujeres, ¿cuál es la probabilidad de que haya más de 3 afectadas por el síndrome de Turner? ¿Y de que haya alguna con estrechamiento de aorta como consecuencia de padecer el síndrome de Turner?
- b) En un grupo de 20 chicas afectadas por el síndrome de Turner, ¿cuál es la probabilidad de que menos de 3 sufran un estrechamiento anormal de la aorta?

SOLUCIÓN

- a) Si X es el número de mujeres afectadas por el síndrome de Turner en el grupo de 4000 mujeres, entonces $X \sim B(4000, 1/2000) \approx P(2)$ y $P(X > 3) = 0,1429$.
Si Y es el número de mujeres con estrechamiento de la aorta en el grupo de 4000 mujeres, entonces $Y \sim B(4000, 1/20000) \approx P(0,2)$ y $P(Y > 0) = 0,1813$.
- b) Si Z es el número de mujeres con estrechamiento de la aorta en el grupo de 20 mujeres con el síndrome de Turner, entonces $Z \sim B(20, 1/10)$ y $P(Z < 3) = 0,6769$.

62. Se sabe que 2 de cada 1000 pacientes son alérgicos a un fármaco A , y que 6 de cada 1000 lo son a un fármaco B . Además, el 30 % de los alérgicos a B , también lo son a A . Si se aplican los dos fármacos a 500 personas,

- a) ¿Cuál es la probabilidad de que no haya ninguna con alergia a A ?
- b) ¿Cuál es la probabilidad de que haya al menos 2 con alergia a B ?
- c) ¿Cuál es la probabilidad de que haya menos de 2 con las dos alergias?
- d) ¿Cuál es la probabilidad de que haya alguna con alergia?

SOLUCIÓN

- a) Llamando X_A al número de personas alérgicas al fármaco A en el grupo de 500 personas, se tiene que $X_A \sim B(500, 0,002) \approx P(1)$ y $P(X_A = 0) = 0,3678$.
- b) Llamando X_B al número de personas alérgicas al fármaco B en el grupo de 500 personas, se tiene que $X_B \sim B(500, 0,006) \approx P(3)$ y $P(X_B \geq 2) = 0,8009$.
- c) Llamando $X_{A \cap B}$ al número de personas alérgicas a ambos fármacos $A \cap B$ en el grupo de 500 personas, se tiene que $X_{A \cap B} \sim B(500, 0,0018) \approx P(0,9)$ y $P(X_{A \cap B} < 2) = 0,7725$.
- d) Llamando $X_{A \cup B}$ al número de personas alérgicas a alguno de los fármacos $A \cup B$ en el grupo de 500 personas, se tiene que $X_{A \cup B} \sim B(500, 0,0062) \approx P(3,1)$ y $P(X_{A \cup B} \geq 1) = 0,9550$.

63. En una clase hay 40 alumnos de los cuales el 35 % son fumadores. Si se toma una muestra aleatoria con reemplazamiento de 4 alumnos, ¿cuál es la probabilidad de que haya al menos 1 fumador? ¿Cuál sería dicha probabilidad si la muestra se hubiese tomado sin reemplazamiento?

SOLUCIÓN

Si X es el número de fumadores en una muestra aleatoria con reemplazamiento de tamaño 4, entonces $X \sim B(4, 0,35)$ y $P(X \geq 1) = 0,8215$.

Si la muestra es sin reemplazamiento $P(X \geq 1) = 0,8364$.

- ★ 64. Se sabe que por término medio 2 de cada 10000 niños que nacen son albinos.

- Si en una región nacen cada año 22000 niños ¿cuál es la probabilidad de que un año nazcan al menos 4 albinos?
- ¿Cuál es la probabilidad de que en esa región, en un periodo de 10 años no nazca ningún niño albino?

SOLUCIÓN

a) Llamando X al número de niños albinos que nacen en un año, se tiene que $X \sim B(22000, 2/10000) \approx P(4,4)$ y $P(X \geq 4) = 0,6406$.

b) Llamando Y al número de niños albinos que nacen en 10 años, se tiene que $Y \sim B(220000, 2/10000) \approx P(44)$ y $P(Y = 0) = 0$.

- ★ 65. Suponiendo una facultad en la que hay un 60 % de chicas y un 40 % de chicos:

- Si un año van 6 alumnos a hacer prácticas en un hospital, ¿qué probabilidad hay de que vayan más chicos que chicas?
- En un período de 5 años, ¿cuál es la probabilidad de que más de 1 año no haya ido ningún chico?

SOLUCIÓN

a) Si X es el número de chicos, $X \sim B(6, 0,4)$ y $P(X \geq 4) = 0,1792$.

b) Si Y es el número de años que no ha ido ningún chico, $Y \sim B(5, 0,0467)$ y $P(Y > 1) = 0,0199$.

- ★ 66. ¿Cuánto habría que restar a cada pregunta errada en un examen de tipo test de 5 preguntas con cuatro opciones y sólo una correcta, para que un individuo que responda al azar tenga una puntuación esperada de 0?

SOLUCIÓN

1/3.

67. Se sabe que el 6,8 % las personas presentan a lo largo de su adolescencia un trastorno de hiperactividad, de los cuales tres cuartas partes son mujeres. Si en la población hay el mismo número de hombres y mujeres, se pide:

- Calcular la probabilidad de que en una muestra de tres hombres, haya alguno que haya tenido hiperactividad en su adolescencia.

- b) Calcular la probabilidad de que en una muestra de 2 hombres y 2 mujeres, haya alguno que haya tenido hiperactividad en su adolescencia.

SOLUCIÓN

- a) Si llamamos X al número de hombres que han tenido hiperactividad en su adolescencia en una muestra de 3 hombres, se tiene que $X \sim B(3, 0,034)$ y $P(X \geq 1) = 0,0986$.
- b) Si llamamos X_H al número de hombres que han tenido hiperactividad en su adolescencia en una muestra de 2 hombres y X_M al número de mujeres que han tenido hiperactividad en su adolescencia en una muestra de 2 mujeres, entonces $X_H \sim B(2, 0,034)$ y $X_M \sim B(2, 0,102)$. Entonces $P(X_H \geq 1 \cup X_M \geq 1) = 0,2475$.
-

68. Un empleado suele acudir al trabajo en cualquier instante entre las 6 y las 7 con igual probabilidad. Se pide:

- a) Calcular la función de densidad de la variable que mide el instante en que acude a trabajar y dibujarla.
- b) Calcular la función de distribución y dibujarla.
- c) Calcular la probabilidad de que llegue entre las 6 y cuarto y las 6 y media.
- d) Calcular la hora media a la que se espera que llegue.

SOLUCIÓN

a)

$$f(x) = \begin{cases} 0 & \text{si } x < 6, \\ 1 & \text{si } 6 \leq x \leq 7, \\ 0 & \text{si } 7 < x. \end{cases}$$

b)

$$F(x) = \begin{cases} 0 & \text{si } x < 6, \\ x - 6 & \text{si } 6 \leq x \leq 7, \\ 1 & \text{si } 7 < x. \end{cases}$$

c) $P(6,25 < X < 6,5) = 0,25$.

d) $\mu = 6,5$, es decir, a las 6 horas y media.

69. Sea Z una variable aleatoria que sigue una distribución $N(0, 1)$. Determinar el valor de t en cada uno de los siguientes casos:

- a) El área entre 0 y t es 0,4783.
- b) El área a la izquierda de t es 0,6406.
- c) El área entre $-1,5$ y t es 0,2313.

SOLUCIÓN

- a) $t = 2,02$.
- b) $t = 0,36$.
- c) $t = -0,53$.

70. Entre los diabéticos, el nivel de glucosa en la sangre en ayunas, puede suponerse de distribución aproximadamente normal, con media 106 mg/100 ml y desviación típica 8 mg/100 ml.

- a) Hallar $P(X \leq 120 \text{ mg/100 ml})$.
- b) ¿Qué porcentaje de diabéticos tendrá niveles entre 90 y 120 mg/100 ml?
- c) Encontrar un valor que tenga la propiedad de que el 25 % de los diabéticos tenga un nivel de glucosa X por debajo de dicho valor.

SOLUCIÓN

- a) $P(X \leq 120) = 0,9599$.
 - b) $P(90 < X < 120) = 0,9371$, es decir, un 93,71 %.
 - c) 100,64 mg/100 ml.
-

★ 71. En un examen, el 63 % de los alumnos ha obtenido una nota superior a 5, y el 44 % entre 5 y 7. Suponiendo que las notas siguen una distribución normal:

- a) Calcular la media y la desviación típica de las notas.
- b) Calcular el porcentaje de alumnos con nota superior a 8.
- c) ¿Cuál es la nota por encima de la cual está el 5 % de los alumnos?

SOLUCIÓN

Llamando X a la nota obtenida en el examen, se tiene que $X \sim N(\mu, \sigma)$.

- a) $\mu = 5,55$ puntos y $\sigma = 1,65$ puntos.
 - b) $P(X > 8) = 0,0648$, es decir, el 6,48 % de los alumnos han tenido una nota superior a 8.
 - c) 8,21 puntos.
-

★ 72. Se supone que la tensión arterial de los habitantes de una población de 20000 habitantes sigue una distribución normal, cuya media es 13 y su rango intercuartílico 4. Se pide:

- a) ¿Cuántas personas tienen una tensión por encima de 16?
- b) ¿Cuánto tendrá que disminuir la tensión de una persona que tiene 16 para situarse en el 40 % de la población con tensión más baja?

SOLUCIÓN

- a) $P(X > 16) = 0,1587$ y el número de personas con tensión por encima de 16 mmHg es $0,1587 \cdot 20000 = 3174$.
 - b) Debe disminuir 3,75 mmHg.
-

- ★ 73. En una población de 30000 individuos se está interesado en medir el volumen sanguíneo de sus individuos. Se sabe que la desviación típica de la población es 0,4 litros y que el 50 % de los individuos tienen un volumen superior a 4,8 litros. ¿Cuántos individuos presentarán un volumen menor de 4,3 litros?

SOLUCIÓN

El número de individuos con menos de 4,3 litros de sangre es 3168.

74. Se consideran las variables aleatorias X_1 y X_2 . La variable X_1 sigue una distribución normal de media μ y desviación típica σ , mientras que la variable X_2 sigue también una distribución normal de media $\mu + 1$ y desviación típica σ . Si la probabilidad de que X_1 tome valores superiores a 14,2 es 0,5636, y la de que X_2 tome valores inferiores a 17,4 es 0,6103, se pide:
- a) Hallar los valores de μ y σ .
 - b) Si se rechazan los individuos que están fuera del intervalo (12,18), hallar los porcentajes de rechazo correspondientes a X_1 y X_2 .
 - c) Si se desea seleccionar el 20 % de individuos que tengan los valores más altos de X_1 , ¿cuál será el valor de X_1 a partir del cuál se seleccionarán?

SOLUCIÓN

- a) $\mu = 15$ y $\sigma = 5$.
 - b) $P(12 < X_1 < 18) = 0,4515$, luego el porcentaje de rechazos es $100 \% - 45,15 \% = 54,85 \%$.
 $P(12 < X_2 < 18) = 0,4436$, luego el porcentaje de rechazos es $100 \% - 44,36 \% = 55,64 \%$.
 - c) 19,21.
-

- ★ 75. El peso de los recién nacidos no prematuros en una ciudad sigue una distribución normal de media y desviación típica desconocidas. Teniendo en cuenta que, de un total de 200 recién nacidos no prematuros, 15 han pesado más de 4 kg y 25 menos de 2,5 kg:
- a) ¿Cuáles son la media y la desviación típica del peso?
 - b) ¿Cuántos niños no prematuros habrán nacido con un peso entre 3 y 3,5 kg?
 - c) Si los médicos consideran peligrosos los pesos por debajo del percentil 10, ¿cuál será dicho peso?, ¿cuántos niños habrán nacido con un peso por debajo de dicho percentil?

SOLUCIÓN

Llamando X al peso de los recién nacidos no prematuros, se tiene que $X \sim N(\mu, \sigma)$.

- a) $\mu = 3,17$ kg y $\sigma = 0,58$ kg.
 - b) 66 niños.
 - c) $P_{10} = 2,43$ kg y habrán nacido 20 niños por debajo de este peso.
-

76. El coeficiente intelectual es una puntuación derivada de los test de inteligencia que tiene media 100 puntos y desviación típica 15. Si se considera que una persona por encima de 145 es una superdotada, ¿qué porcentaje de superdotados habrá en la población? Si se considera que el 1 % de las personas con menor coeficiente intelectual son deficientes, ¿por debajo de qué coeficiente estarán dichas personas?

SOLUCIÓN

Llamando X al coeficiente intelectual, se tiene que $X \sim N(100, 15)$.
 $P(X > 145) = 0,0013$, luego el 0,13 % de las personas son superdotadas.
 $P_1 = 65,10$, por debajo de este coeficiente las personas son deficientes.

77. Un test diagnóstico para niños de 10 años con problemas de lectura da puntuaciones que se distribuyen normalmente con una media de 80 y una varianza de 100. Se pide:

- a) Dar la probabilidad de que un niño seleccionado al azar tenga una puntuación de:
 - 1) menos de 68;
 - 2) entre 75 y 90;
- b) El test indica que el niño tiene un problema del lenguaje si su puntuación está por debajo del 10 % de la población que realizó el test. ¿Por debajo de qué puntuación el test diagnosticará problemas de lenguaje?
- c) Si se selecciona una muestra de 16 niños:
 - 1) ¿Cuántos se espera que tengan una puntuación por encima de 68?
 - 2) ¿Cuál es la probabilidad de que su puntuación media supere los 84 puntos?

SOLUCIÓN

Llamando X a la puntuación del test diagnóstico, se tiene que $X \sim N(80, 10)$.

- a) $P(X < 68) = 0,1151$ y $P(75 < X < 90) = 0,5328$.
 - b) $P_{10} = 67,18$ puntos.
 - c) $P(X > 68) = 0,8849$ y el número esperado niños con puntuaciones por encima de 68 es $0,8849 \cdot 16 = 14,16$.
 $\bar{x} \sim N(80, 2,5)$ y $P(\bar{x} > 84) = 0,0548$.
-

5. Estimación de parámetros

78. Las notas en Estadística de una muestra de 10 alumnos han sido:

6,3, 5,4, 4,1, 5,0, 8,2, 7,6, 6,4, 5,6, 4,3, 5,2

Dar una estimación puntual de la nota media, de la varianza y del porcentaje de aprobados en la clase.

SOLUCIÓN

$\bar{x} = 5,81$ puntos, $\hat{s}^2 = 1,7721$ puntos² y $\hat{p} = 0,8$.

79. Una muestra aleatoria de tamaño 81 extraída de una población normal con $\sigma^2 = 64$, tiene una $\bar{x} = 78$. Calcular el intervalo de confianza del 95 % para μ .

SOLUCIÓN

$\mu \in 79 \pm 1,742 = (76,258, 79,742)$.

80. Se obtuvieron cinco determinaciones del pH de una solución con los siguientes resultados:

7,90, 7,85, 7,89, 7,86, 7,87.

Hallar unos límites de confianza de la media de todas las determinaciones del pH de la misma solución, al nivel de significación $\alpha = 0,01$.

SOLUCIÓN

$\mu \in 7,874 \pm 0,0426 = (7,8314, 7,9166)$.

81. El tiempo que tarda en hacer efecto un analgésico sigue una distribución aproximadamente normal. En una muestra de 20 pacientes se obtuvo una media de 25,4 minutos y una desviación típica de 5,8 minutos. Calcular el intervalo de confianza del 90 % e interpretarlo. ¿Cuántos pacientes habría que tomar para poder estimar la media con una precisión de ± 1 minuto?

SOLUCIÓN

Intervalo de confianza del 90 % para μ : (23,0992, 27,7008).

Tamaño muestral para una precisión de ± 1 minuto: $n = 96$.

82. Para realizar una determinada tarea se necesitan personas con un cierto nivel de pericia. Una empresa está interesada en contratar a un grupo de personas que tengan la pericia necesaria para realizar la tarea, pero que además sean bastante homogéneas en cuanto a la pericia, es decir, que su rendimiento sea parecido. Para ver si un grupo de alumnos que se han formado en dicha tarea cumplen los requisitos, se ha tomado una muestra y se les ha sometido una prueba para ver cuántas tareas son capaces de realizar con éxito en una hora. Los resultados obtenidos fueron:

16 - 12 - 14 - 21 - 11 - 15 - 17 - 15 - 18 - 24 - 10 - 14 - 17 - 14

Se pide:

- Si la empresa busca un grupo de empleados capaces de realizar una media de al menos 12 tareas por hora, ¿se puede afirmar con una confianza del 95 % que el grupo lo cumple?
- Si la empresa busca que entre los trabajadores haya una dispersión media de $\sigma < 3$ tareas, ¿se puede afirmar con una confianza del 95 % que el grupo lo cumple?

SOLUCIÓN

- Lo cumple ya que $\mu \in (13,4026, 17,7403)$ con un 95 % de confianza.
 - No lo cumple ya que $\sigma \in (2,7232, 6,0516)$ con un 95 % de confianza.
-

83. En un estudio para el estado de la salud oral de una ciudad, se tomó una muestra elegida al azar de 280 varones entre 35 y 44 años y se contó el número de piezas dentarias en la boca. Tras la revisión pertinente, los dentistas informaron que había 70 individuos con 28 o más dientes. Se desea realizar una estimación por intervalo de confianza de la proporción de individuos de esta ciudad con 28 dientes o más, con un nivel de confianza 0,95.

SOLUCIÓN

$p \in 0,25 \pm 0,101 = (0,149, 0,351)$.

84. En una muestra de 250 estudiantes de una universidad, 146 hablaban inglés. ¿Entre qué valores estará el porcentaje de individuos de la universidad que hablan inglés, con un nivel de confianza del 90 %?

SOLUCIÓN

$p \in (0,5327, 0,6353)$ con un 90 % de confianza.

85. Si el porcentaje de individuos daltónicos de una muestra aleatoria es 18 %, ¿cuál será el mínimo tamaño muestral necesario para conseguir una estimación del porcentaje de daltónicos con una confianza del 95 % y un error menor del 3 %?

SOLUCIÓN

$n = 2520$.

- ★ 86. Un país está siendo afectado por una epidemia de un virus. Para valorar la gravedad de la situación se tomaron 40 personas al azar y se comprobó que 12 de ellas tenían el virus. Determinar el intervalo de confianza para el porcentaje de infectados con un nivel de significación 0,05.

SOLUCIÓN

$p \in (0,1580, 0,4420)$ con un 95 % de confianza.

87. Se ha realizado un estudio para investigar el efecto del ejercicio físico en el nivel de colesterol en la sangre. En el estudio participaron once personas, a las que se les midió el nivel de colesterol antes y después de desarrollar un programa de ejercicios. Los resultados obtenidos fueron los siguientes

Persona	Nivel previo	Nivel posterior
1	182	198
2	232	210
3	191	194
4	200	220
5	148	138
6	249	220
7	276	219
8	213	161
9	241	210
10	280	213
11	262	226

Hallar un intervalo de confianza del 90 % para la diferencia del nivel medio de colesterol antes y después del ejercicio.

SOLUCIÓN

$\mu_{x_1 - x_2} \in 33,182 \pm 27,899 \text{ mg/dl} = (5,283\text{mg/dl}, 61,081\text{mg/dl})$.

88. Se está ensayando un nuevo procedimiento de rehabilitación para una cierta lesión. Para ello se trataron nueve pacientes con el procedimiento tradicional y otros nueve con el nuevo, y se midieron los días que tardaron en recuperarse, obteniéndose los siguientes resultados:

Método tradicional: 32-37-35-28-41-44-35-31-34
 Método nuevo: 35-31-29-25-34-40-27-32-31

Se desea obtener un intervalo de confianza del 95 % para la diferencia de las medias del tiempo de recuperación obtenido con ambos procedimientos. Se supone que los tiempos de recuperación siguen una distribución normal, y que las varianzas son aproximadamente iguales para los dos procedimientos.

SOLUCIÓN

$$\mu_1 - \mu_2 \in 3,667 \pm 4,712 \text{ días} = (-1,045 \text{ días}, 8,379 \text{ días}).$$

89. Para estudiar si la estación del año influye en el estado de ánimo de la gente, se ha tomado una muestra de 12 personas y se ha medido su nivel de depresión en verano e invierno mediante un cuestionario con puntuaciones de 0 a 100 (a mayor puntuación mayor depresión). Los resultados obtenidos fueron:

Invierno	65	72	84	31	80	61	75	52	73	79	85	71
Verano	60	51	81	45	62	53	70	52	64	51	67	62

¿Se puede afirmar que la estación del año influye en el estado de ánimo de la gente con un 99 % de confianza? ¿Cómo influye?

SOLUCIÓN

No se puede afirmar que la estación influya en el estado de ánimo ya que la media de la diferencia está en $(-0,7498, 19,0831)$ con un 99 % de confianza.

90. Un psicólogo está estudiando la concentración de una encima en la saliva como un posible indicador de la ansiedad crónica. En un experimento se tomó una muestra de 12 neuróticos por ansiedad y otra de 10 personas con bajos niveles de ansiedad. En ambas muestras se midió la concentración de la encima, obteniendo los siguientes resultados:

Con ansiedad:	2,60	2,90	2,60	2,70	3,91	3,15	3,94	2,46	2,91	3,88	3,55	3,96
Sin ansiedad:	2,37	1,10	2,55	2,64	2,20	2,12	2,47	2,90	1,66	2,72		

¿Se puede concluir a partir de estos datos que la población de neuróticos con ansiedad y la población de personas sin ansiedad son diferentes en el nivel medio de concentración de encimas? Justificar la respuesta.

SOLUCIÓN

$\frac{\sigma_1^2}{\sigma_2^2} \in (0,3383, 4,7486)$ con un nivel de confianza del 95 %, luego se puede suponer que las varianzas son iguales.

$\mu_1 - \mu_2 \in (0,4296, 1,4510)$ con un 95 % de confianza, luego se puede concluir que hay diferencias entre las medias.

- ★ 91. Para comparar la eficacia de dos tratamientos A y B en la prevención de repeticiones de infarto de miocardio, se aplicó el tratamiento A a 80 pacientes y el B a 60. Al cabo de dos años se observó que habían sufrido un nuevo infarto 14 pacientes de los sometidos al tratamiento A y 15 de los del B . Se pide:

- Construir un intervalo de confianza del 95 % para la diferencia entre las proporciones de personas sometidas a los tratamientos A y B que no vuelven a sufrir un infarto.
- A la vista del resultado obtenido, razonar si con ese nivel de confianza puede afirmarse que uno de los tratamientos es más eficaz que el otro.

SOLUCIÓN

- $p_A - p_B \in (-0,2126, 0,0626)$ con un nivel de confianza del 95 %.
 - No puede afirmarse que un tratamiento sea más eficaz que otro pues la diferencia de medias podría ser positiva, negativa o cero.
-

- ★ 92. En un análisis de obesidad dependiendo del hábitat en niños menores de 5 años, se obtienen los siguientes resultados:

	Casos analizados	Casos con sobrepeso
Hábitat rural	1150	480
Hábitat urbano	1460	660

Se pide:

- Construir un intervalo de confianza, con un nivel de significación 0,01, para la proporción de niños menores de 5 años con sobrepeso en el hábitat rural. Igualmente para el hábitat urbano.
- Construir un intervalo de confianza, con un nivel de confianza del 95 %, para la diferencia de proporciones de niños menores de 5 años con sobrepeso entre el hábitat rural y el urbano. A la vista del resultado obtenido, ¿se puede concluir, con un 95 % de confianza, que la proporción de niños menores de 5 años con sobrepeso depende del hábitat?

SOLUCIÓN

- $p_R \in (0,3799, 0,4548)$ y $p_U \in (0,4185, 0,4856)$ con un 99 % de confianza.
 - $p_R - p_U \in (-0,0729, 0,0036)$ con un nivel de confianza del 95 %, luego no se puede afirmar que haya diferencias en las proporciones de niños menores de 5 años con sobrepeso.
-

93. Se cree que el consumo de tabaco va ligado al consumo de alcohol y para corroborar esta hipótesis se ha realizado un estudio en el que se han obtenido los siguientes datos

	Bebedores	No bebedores
Fumadores	487	137
No fumadores	312	365

¿Se puede afirmar que existe relación entre el consumo de tabaco y el de alcohol? Justificar la respuesta.

6. Contraste de hipótesis

94. Se sabe que una vacuna que se está utilizando al cabo de dos años sólo protege al 60 % de las personas a las que se administró. Se desarrolla una nueva vacuna, y se quiere saber si al cabo de dos años protege a más personas que la primera. Para ello se seleccionan 10 personas al azar y se les inyecta la nueva vacuna. Establecemos que si más de 8 de los vacunados conservan la protección al cabo de dos años, entonces consideraremos la nueva vacuna mejor que la antigua. Se pide:

- Calcular la probabilidad de cometer un error de tipo I.
- Si la nueva vacuna protegiera a un 80 % de las personas vacunadas al cabo de 2 años, ¿Cuál será la probabilidad de cometer un error de tipo II?

Repetir los cálculos si se toma una muestra de 100 personas y se establece que la vacuna es mejor si más de 75 de los vacunados conservan la protección al cabo de 2 años.

Nota: Aproximar la distribución binomial mediante una distribución normal.

SOLUCIÓN

Contraste: $H_0 : p = 0,6$, $H_1 : p > 0,6$. Muestra de tamaño 10:

- $P(\text{Rechazar } H_0/H_0) = 0,0464$.
- $P(\text{Aceptar } H_0/H_1) = 0,6242$.

Muestra de tamaño 100:

- $P(\text{Rechazar } H_0/H_0) = 0,0011$.
 - $P(\text{Aceptar } H_0/H_1) = 0,1056$.
-

95. Se sabe que el tiempo de reacción ante un estímulo sigue una distribución normal de media 30 ms y desviación típica 10 ms. Se cree que la alcoholemia aumenta el tiempo de reacción de los sujetos, y para comprobar esta hipótesis se ha tomado una muestra aleatoria de 40 individuos a los que se les ha inducido una alcoholemia de 0,8 g/l y en los que se ha apreciado un tiempo medio de respuesta de 35 ms y una desviación típica de 12 ms. ¿Se puede afirmar que una alcoholemia de 0,8 gm/l influye en el tiempo medio de respuesta con un riesgo $\alpha = 0,05$? ¿Y con un riesgo $\alpha = 0,01$?

¿Cuál será la potencia del contraste para detectar una diferencia en la media del tiempo de reacción de 4 ms? ¿Cuál debería ser el tamaño muestral para aumentar la potencia hasta un 90 %?

SOLUCIÓN

Contraste para la media: $H_0 : \mu = 30$, $H_1 : \mu > 30$.

Región de aceptación para $\alpha = 0,01$: $z < 2,3263$.

Estadístico del contraste: $z = 2,6020$. Como cae fuera de la región de aceptación, se rechaza la hipótesis nula tanto para $\alpha = 0,01$ y con mayor motivo para $\alpha = 0,05$, de manera que se concluye que la alcoholemia influye en el tiempo de respuesta.

Potencia del contraste para $\delta = 4$: $1 - \beta = 1 - 0,5966 = 0,4034$.

El tamaño muestral para, $\alpha = 0,05$, $\delta = 4$ y una potencia del 90 % es $n = 80$.

96. Un fisioterapeuta afirma que con un nuevo procedimiento de rehabilitación que él aplica, determinada lesión tiene un tiempo de recuperación medio no mayor de 15 días. Se seleccionan al azar 36 personas que sufren dicho tipo de lesión para verificar su afirmación, y se obtiene un tiempo medio de recuperación de 13 días y una cuasivarianza de 9. ¿Contradice lo observado en la muestra la afirmación del fisioterapeuta para un $\alpha = 0,05$?

SOLUCIÓN

Contraste para la media: $H_0 : \mu = 15$, $H_1 : \mu < 15$.

Región de aceptación para $\alpha = 0,05$: $-1,6444 < z$.

Estadístico del contraste: $z = -4$. Como cae fuera de la región de aceptación se rechaza la hipótesis nula y se confirma la afirmación del fisioterapeuta.

97. Se cree que el nivel medio de protrombina en plasma de una población normal tiene una media de 19mg/100ml y una desviación típica de 4mg/100ml. Para contrastar estas hipótesis se tomó una muestra de 8 individuos en los que se obtuvieron los siguientes niveles de protrombina en plasma:

16,3 – 18,4 – 20,0 – 17,6 – 15,4 – 23,7 – 17,8 – 19,5

¿Se pueden aceptar ambas hipótesis con un riesgo $\alpha = 0,1$?

SOLUCIÓN

Contraste para la media: $H_0 : \mu = 19$, $H_1 : \mu \neq 19$.

Región de aceptación para $\alpha = 0,1$: $-1,8946 < t < 1,8946$.

Estadístico del contraste: $t = -0,4552$. Como cae dentro de la región de aceptación, se mantiene la hipótesis de que la media es 19mg/100ml.

Contraste para la varianza: $H_0 : \sigma = 4$, $H_1 : \sigma \neq 4$.

Región de aceptación para $\alpha = 0,1$: $2,1673 < j < 14,0671$.

Estadístico del contraste: $j = 2,8743$. Como cae dentro de la región de aceptación, también se mantiene la hipótesis de que la desviación típica es de 4mg/100ml.

98. Se decide retirar una cierta vacuna si produce más de un 10 % de reacciones alérgicas. Se consideran 100 pacientes sometidos a la vacuna y se observan 15 reacciones alérgicas. ¿Debe retirarse la vacuna? (Utilizar un $\alpha = 0,01$).

SOLUCIÓN

Contraste para la proporción: $H_0 : p = 0,1$, $H_1 : p > 0,1$.

Región de aceptación para $\alpha = 0,01$: $z > 2,3263$.

Estadístico del contraste: $z = 1,6667$. Como cae dentro de la región de aceptación, se acepta la hipótesis nula y se concluye que no hay pruebas suficientes para retirar la vacuna.

99. Un fabricante de baterías para automóvil asegura que la duración de sus baterías tiene una distribución aproximadamente normal con desviación típica no superior a 0,9 años. Si una muestra aleatoria de 10 de estas baterías tiene una cuasidesviación típica de 0,7 años, ¿qué se puede concluir sobre la afirmación del fabricante?

SOLUCIÓN

Contraste para la desviación típica: $H_0 : \sigma = 0,9$, $H_1 : \sigma < 0,9$.

Región de aceptación para $\alpha = 0,05$: $3,3251 < j$.

Estadístico del contraste: $j = 4$. Como cae dentro de la región de aceptación, no se puede rechazar la hipótesis nula y se concluye que no hay pruebas significativas de que sea cierta la afirmación del fabricante.

100. Un estudio afirma que el 70 % de los habitantes de la capital lee diariamente algún periódico. ¿Estaríamos de acuerdo con las conclusiones de dicho estudio si al preguntar a 15 personas elegidas aleatoriamente, 8 leen diariamente algún periódico?

SOLUCIÓN

Contraste para la proporción: $H_0 : p = 0,7$, $H_1 : p \neq 0,7$.

Región de aceptación para $\alpha = 0,05$: $-1,96 < z < 1,96$.

Estadístico del contraste: $z = -1,2939$. Como cae dentro de la región de aceptación, se acepta la hipótesis nula y se estaría de acuerdo con las afirmación del estudio.

101. Se realizó en dos hospitales una encuesta entre los pacientes sobre la satisfacción con la atención recibida, calificándola de 0 a 100. En el hospital A rellenaron la encuesta 12 pacientes, obteniéndose una media de 85 y una cuasivarianza de 16, mientras que en el hospital B rellenaron la encuesta 10 pacientes, obteniéndose una media de 81 y una cuasivarianza de 25. ¿Puede concluirse que el nivel de satisfacción en el hospital A es mayor que en el B?

Nota: Hacer previamente un contraste de igualdad de varianzas.

SOLUCIÓN

Contraste de comparación de varianzas: $H_0 : \sigma_A = \sigma_B$, $H_1 : \sigma_A \neq \sigma_B$.

Región de aceptación para $\alpha = 0,05$: $0,2787 < f < 3,9121$.

Estadístico del contraste: $f = 0,64$. Como cae dentro de la región de aceptación, se acepta la hipótesis de que las varianzas son iguales.

Contraste de comparación de medias: $H_0 : \mu_A = \mu_B$, $H_1 : \mu_A > \mu_B$.

Región de aceptación para $\alpha = 0,05$: $t > 1,7247$ Estadístico del contraste: $t = 2,0863$. Como cae fuera de la región de aceptación, se rechaza la hipótesis nula y se concluye que hay pruebas significativas de que el nivel de satisfacción en el hospital A es mayor que en el B.

102. Un fabricante de equipos de medida afirma que sus equipos pueden realizar al menos 12 mediciones más que los de la competencia sin necesidad de un nuevo ajuste. Para probar esta afirmación se realizan mediciones con 50 equipos de este fabricante y 50 de la competencia. En los suyos el número de mediciones hasta necesitar un nuevo ajuste tuvo de media 86,7 y cuasidesviación típica 6,28, mientras que en los de la competencia estos valores fueron 77,8 y 5,61 respectivamente. Verificar la afirmación del fabricante con $\alpha = 0,05$.

SOLUCIÓN

Contraste de comparación de medias: $H_0 : \mu_1 < \mu_2 + 12$, $H_1 : \mu_1 \geq \mu_2 + 12$.

Intervalo de confianza para la diferencia de medias: $\mu_1 - \mu_2 \in (6,5659, 11,2341)$ con un 95 % de confianza, luego hay diferencias significativas entre el número medio de mediciones, pero no se puede afirmar que sean mayores de 12 mediciones.

103. Un psicólogo está estudiando la concentración de una encima en la saliva como un posible indicador de la ansiedad crónica. En un experimento se tomó una muestra de 12 neuróticos por ansiedad y otra de 10 personas con bajos niveles de ansiedad. En ambas muestras se midió la concentración de la encima, obteniendo los siguientes resultados:

Con ansiedad:	2,60	2,90	2,60	2,70	3,91	3,15	3,94	2,46	2,91	3,88	3,55	3,96
Sin ansiedad:	2,37	1,10	2,55	2,64	2,20	2,12	2,47	2,90	1,66	2,72		

¿Se puede concluir a partir de estos datos que la población de neuróticos con ansiedad y la población de personas sin ansiedad son diferentes en el nivel medio de concentración de encimas? Justificar la respuesta.

SOLUCIÓN

Contraste de comparación de varianzas: $H_0 : \sigma_1 = \sigma_2$, $H_1 : \sigma_1 \neq \sigma_2$.

Región de aceptación para $\alpha = 0,05$: $0,2787 < f < 3,9121$.

Estadístico del contraste: $f = 1,2138$. Como cae dentro de la región de aceptación, se acepta la hipótesis nula y se concluye que las varianzas son iguales.

Contraste de comparación de medias: $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$.

Región de aceptación para $\alpha = 0,05$: $-2,0860 < t < 2,0860$.

Estadístico del contraste: $t = 3,8407$. Como cae fuera de la región de aceptación, se rechaza la hipótesis nula y se puede concluir que hay diferencia entre la concentración media de encimas para neuróticos con ansiedad y sin ansiedad.

104. En una investigación para ver la efectividad de una nueva droga antidepresiva, se ha tomado una muestra de 15 pacientes depresivos que han completado un cuestionario para detectar el nivel depresivo,

antes y después de recibir la droga. En la puntuación del cuestionario los valores menores indican una mayor depresión. Los resultados obtenidos han sido:

Antes:	18	21	16	19	14	23	16	14	21	18	17	14	16	14	20
Después:	23	20	17	20	16	22	18	18	21	16	19	20	15	15	21

Realizar un contraste para averiguar si la droga tiene un efecto positivo sobre la depresión. ¿Qué tamaño muestral sería necesario para detectar una diferencia en la puntuación como la que hay entre las medias de las muestras?

SOLUCIÓN

Contraste para la media de la diferencia entre antes y después: $H_0 : \mu = 0$, $H_1 : \mu < 0$.

Región de aceptación par $\alpha = 0,05$: $-1,7613 < t$.

Estadístico del contraste: $t = -2,2563$. Como cae fuera de la región de aceptación, se rechaza la hipótesis nula y se puede afirmar que la droga reduce la depresión.

El tamaño muestral para detectar una diferencia de $\delta = \bar{x}_1 - \bar{x}_2 = 17,4 - 18,7333 = -1,3333$ con una potencia del 90 % es $n = 26$ individuos.

105. Un experimento pretende contrastar la teoría de que la memoria a corto plazo se ve afectada por la similitud entre los estímulos. El experimento consiste en leer en voz alta una secuencia de letras a un sujeto, quien después de una breve pausa debe repetir la secuencia. Si la teoría es correcta, habrá más errores en la lista que contenga letras que suenan de forma similar que si contiene letras que se parecen. A cada sujeto se le presentan dos tipos de secuencias, una con letras que suenan de forma similar y otra con letras que se escriben de forma parecida. Los errores producidos en cada secuencia son:

Errores de letras que suenan parecidas:	7	5	6	11	3	8	4	10	9
Errores de letras con similar escritura:	8	2	5	9	5	4	4	7	4

¿Se puede validar la teoría?

SOLUCIÓN

Contraste para la media de la diferencia entre los errores de letras que suenan parecidas y los errores en letras con similar escritura: $H_0 : \mu = 0$, $H_1 : \mu > 0$.

Región de aceptación para $\alpha = 0,05$: $t > 1,8595$.

Estadístico del contraste: $t = 2,1320$. Como cae fuera de la región de aceptación, se rechaza la hipótesis nula y se concluye que la teoría es cierta.

106. Se utiliza un grupo de 150 pacientes para comprobar la teoría de que la vitamina C tiene alguna influencia en el tratamiento del cáncer. Los 150 pacientes fueron divididos en dos grupos de 75. Un grupo recibió 10 gramos de vitamina C y el otro un placebo cada día, además de la medicación habitual. De los que recibieron la vitamina C, 47 presentaban alguna mejoría al cabo de cuatro semanas, mientras que de los que recibieron el placebo, 43 experimentaron mejoría. Contrastar esta hipótesis.

SOLUCIÓN

Contraste para la comparación de proporciones: $H_0 : p_1 = p_2$, $H_1 : p_1 \neq p_2$.

Región de aceptación para $\alpha = 0,05$: $-1,96 < z < 1,96$.

Estadístico del contraste $z = 0,6677$. Como cae dentro de la región de aceptación, se acepta la hipótesis nula y no se pudo concluir que la vitamina C tenga influencia en el tratamiento del cáncer.

107. En un estudio sobre el consumo de alcohol entre los jóvenes durante los fines de semana, se preguntó a 100 chicos y a 125 chicas, de los que 63 chicos y 59 chicas contestaron que consumían. En vista de estos datos, ¿existe alguna diferencia significativa entre las respuestas de chicos y chicas? Utilizar $\alpha = 0,10$.

SOLUCIÓN

Contraste de comparación de proporciones: $H_0 : p_1 = p_2$, $H_1 : p_1 \neq p_2$.

Región de aceptación para $\alpha = 0,01$: $-1,6449 < z < 1,6449$.

Estadístico del contraste: $z = 2,4026$. Como cae fuera de la región de aceptación, se rechaza la hipótesis nula y se concluye que hay diferencias significativas entre el consumo de alcohol de chicos y chicas.

NOTA: Los problemas marcados con una estrella (★) son problemas de exámenes de otros años.