EJERCICIOS DE ESTADÍSTICA

Asignatura: Matemáticas

Curso: 1º de Grado en Farmacia

Eduardo López Ramírez (elopez@ceu.es)
José Rojo Montijano (jrojo.eps@ceu.es)
Anselmo Romero Limón (arlimon@ceu.es)
Alfredo Sánchez Alberca (asalber@ceu.es)

Curso 2014-2015

Índice

1. Estadística Descriptiva

1.

2.

★ 3. El siguiente diagrama refleja el porcentaje de calificaciones obtenidas en un examen realizado a 80 alumnos:

Se pide:

a) Construir la tabla de frecuencias para las calificaciones.

b) Dibujar el polígono de frecuencias acumuladas.

c) Calcular todos los estadísticos de tendencia central que sean posibles.

d) 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

c) $Me =$ Aprobado y $Mo =$ Suspenso.

d) $\bar{x} = 5{,}275$ puntos, $s = 2{,}447$ puntos y $cv = 0{,}464$, de manera que la media es moderadamente representativa.

---

4. Se ha llevado a cabo un estudio sobre el número de radiografías realizadas durante el último año a un grupo de 200 personas, y la información se presenta en la siguiente tabla incompleta:

Radiografías	Personas	f_i	F_i
0		0,20	
1	84		
2			0,72
3			
4	24		
5		0,02	

a) Completar tabla.

b) Calcular media, mediana, desviación típica y coeficiente de variación e interpretar los resultados.

SOLUCIÓN

b) $\bar{x} = 1{,}62$ radiografías, $s^2 = 1{,}875$ radiografías^2, $s = 1{,}37$ radiografías y $cv = 0{,}845$, lo que indica que hay bastante dispersión y la media no es muy representativa.

---

5. 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.

★ 6. Como parte de un proyecto de investigación, los investigadores obtuvieron los siguientes datos respecto a los niveles de peróxido lípido (en nmol/ml) en el suero de 30 individuos adultos bajo tratamiento de Diabetes Mellitus:

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 peróxido lípido. 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) $\bar{x} = 5{,}667$ nmol/ml, $s^2 = 1{,}668$ (nmol/ml)^2, $s = 1{,}292$ nmol/ml y $cv = 0{,}228$, lo que indica que hay poca dispersión y la media es bastante representativa.
   b) $C_1 = 4{,}7$ nmol/ml, $C_2 = 5{,}667$ nmol/ml y $C_3 = 6{,}75$ nmol/ml.
   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 $b_1 = 3{,}04$ nmol/ml y $b_2 = 7{,}89$ nmol/ml.

★ 7. A continuación figura la distribución de edades de una muestra de 65 individuos sujetos a rehabilitación tras un infarto de miocardio:

Edad	[40-50)	[50-60)	[60-70)	[70-80)	[80-90)
Personas	6	12	23	19	5

   ¿Se puede asumir que la muestra proviene de una población normal?

   Utilizar las siguientes sumas: $\sum x_i = 4275$ años, $\sum(x_i - \bar{x})^2 = 7462$ años^2, $\sum(x_i - \bar{x})^3 = -18249$ añós^3, $\sum(x_i - \bar{x})^4 = 2099636$ años^4.

   Solución

   $\bar{x} = 65{,}769$ años, $s^2 = 114{,}823$ años^2, $s = 10{,}716$ años. $g_1 = -0{,}228$ y $g_2 = -0{,}55$, como tanto el coeficiente de asimetría como el de apuntamiento están entre -2 y 2, podemos suponer que los datos provienen de una población normal.

8.

★ 9. Se desea realizar un estudio sobre los días necesarios para tratar una determinada lesión deportiva. Se utilizaron para ello dos tratamientos diferentes, y se observaron 50 pacientes con cada uno de los tratamientos, obteniendo los siguientes resultados:

Días	A	B
20-40	5	8
40-60	20	15
60-80	18	20
80-100	7	7

a) ¿En cuál de los dos tratamientos es más representativa la media del número de sesiones necesarias?

b) ¿Qué tratamiento presenta una distribución más asimétrica?

c) ¿Qué tratamiento presenta una distribución más apuntada?

Usar las siguientes sumas:
A: $\sum x_i = 3040$ días, $\sum(x_i - \bar{x})^2 = 14568$ días^2, $\sum(x_i - \bar{x})^3 = 17011{,}2$ días^3, $\sum(x_i - \bar{x})^4 = 9989603$ días^4
B: $\sum x_i = 3020$ días, $\sum(x_i - \bar{x})^2 = 16992$ días^2, $\sum(x_i - \bar{x})^3 = -42393{,}6$ días^3, $\sum(x_i - \bar{x})^4 = 12551516$ días^4

Solución

a) $\bar{x}_A = 60{,}8$ sesiones, $s_A^2 = 291{,}36$ sesiones^2, $s_A = 17{,}0693$ sesiones y $cv_A = 0{,}28$.
$\bar{x}_B = 60{,}4$ sesiones, $s_B^2 = 339{,}84$ sesiones^2, $s_B = 18{,}4348$ sesiones y $cv_B = 0{,}31$.
Así pues, como $cv_A < cv_B$ la media es un poco más representativa para el tratamiento A.

b) $g_{1_A} = 0{,}068$ y $g_{1_B} = -0{,}14$ lo que indica que la distribución para el tratamiento A es ligeramente asimétrica hacia la derecha y la del tratamiento B ligeramente asimétrica hacia la izquierda, aunque ambas son casi simétricas.

c) $g_{2_A} = -0{,}65$ y $g_{2_B} = -0{,}83$.

★ 10. Se ha realizado un estudio sobre la tensión arterial en dos ciudades A y B. Se tomó una muestra de 20 individuos, que arrojó los siguientes valores:

Tensión	135	128	137	110	154	142	121	127	114	103
Ciudad	A	B	A	B	A	A	A	B	B	B

Tensión	98	96	114	123	132	141	132	121	98	136
Ciudad	A	B	A	B	A	B	A	B	B	A

Se pide:

a) Construir la tabla de frecuencias para la tensión, agrupando en clases de amplitud 10, entre 95 y 155.

b) Dibujar el polígono de frecuencias acumuladas correspondiente a la tabla anterior.

c) Calcular el coeficiente de asimetría e interpretarlo.

d) Calcular el tercer decil e interpretarlo.

e) Considerando los datos sin agrupar, ¿en qué ciudad son más homogéneas las tensiones?

Solución

c) $\bar{x} = 123$ mmHg, $s^2 = 251$ mmHg^2, $s = 15{,}843$ mmHg y $g_1 = -0{,}12$, por lo que la distribución es un poco asimétrica hacia la izquierda.

d) $D_3 = 111,67$ mmHg.

e) $\bar{x}_A = 130{,}1$ mmHg, $s_A^2 = 219{,}89$ mmHg^2, $s_A = 14{,}8287$ mmHg y $cv_A = 0{,}1139$.
$\bar{x}_B = 116{,}1$ mmHg, $s_B^2 = 189{,}69$ mmHg^2, $s_B = 13{,}7728$ mmHg y y $cv_B = 0{,}1186$. Así pues, como el $cv_A < cv_B$ las tensiones son un poco más homogéneas en la ciudad A, aunque en ambos casos las muestras son muy homogéneas.

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

Se pide:

1. Dibujar el diagrama de sectores para el sexo.
2. ¿En qué grupo es más representativa la media? Justificar la respuesta.
3. ¿Cómo calcularías la media de toda la muestra a partir de las medias de hombres y mujeres? ¿Cuánto vale?

Solución

2. $\bar{m} = 24{,}17$ Kg/m^2, $s_m^2 = 21{,}1806$ (Kg/m^2)^2, $s_m = 4{,}6$ Kg/m^2 y $cv_m = 0{,}19$.
$\bar{h} = 22{,}28$ Kg/m^2, $s_m^2 = 9{,}9506$ (Kg/m^2)^2, $s_m = 3{,}15$ Kg/m^2 y $cv_h = 0{,}14$, luego es más representativa la media en los hombres.
3. $\bar{x} = 23{,}25$ Kg/m^2.

★ Se han medido las estaturas, X, en metros de 15 individuos, obteniéndose los siguientes sumatorios:

$$\sum_{i=1}^{15} x_i = 26{,}40 \text{ m}, \ \sum_{i=1}^{15} x_i^2 = 52{,}45 \text{ m}^2, \ \sum_{i=1}^{15} (x_i - \bar{x})^3 = -8{,}42 \text{ m}^3$$

Se pide:

1. Calcular media, desviación típica, coeficiente de variación y coeficiente de asimetría de la variable. Interpretar los dos últimos.
2. Si trabajamos con una nueva variable $Y = 1{,}1X - 0{,}2$, ¿cuánto valdrán los estadísticos anteriores? Justificar adecuadamente la respuesta.

Solución

1. $\bar{x} = 1{,}76$ m, $s^2 = 0{,}399$ m^2, $s = 0{,}631$ m, $CV = 0{,}359$ (dispersión moderada), $g_1 = -2{,}234$ (distribución muy asimétrica a la izquierda)
2. $\bar{y} = 1{,}736$, $s_y = 0{,}694$, $CV_y = 0{,}40$, $g_{1y} = -2{,}234$.

Dada la siguiente muestra del número de ingresos en urgencias en un determinado hospital,

$$5 - 7 - 11 - 7 - 4 - 7 - 24 - 7 - 8 - 8 - 9 - 8 - 13 - 7 - 8 - 17 - 7 - 6$$

estudiar la asimetría de la muestra y transformar los datos para conseguir una simetría más normal.

Solución

$\bar{x} = 9{,}055$ ingresos, $s^2 = 21{,}4969$ ingresos^2, $s = 4{,}6365$ ingresos y $g_1 = 2{,}02$. Para corregir asimetría tomamos $Y = \ln(X)$ y resulta $\bar{y} = 2{,}108$, $s_y^2 = 0{,}1668$, $s_y = 0{,}4084$ y $g_1 = 0{,}95$.

★ Si a todos los datos de una muestra se les suma una misma cantidad positiva, ¿cómo se ve afectada la representatividad de la media? ¿Y si se multiplican por un mismo número distinto de 0? Razonar la respuesta.

Solución

La representatividad de la media aumenta cuando se suma una constante a los datos y se mantiene igual cuando se multiplican por una constante.

★ En un laboratorio se ha medido el número de hijos que tuvieron unas ratas que habían seguido un tratamiento de fertilidad, obteniendo los siguientes resultados:

$$7-5-6-6-8-9-10-8-7-6-8-7-9-11-8-7-7-7-6-8$$

1. Existen datos atípicos en la muestra? Justificar la respuesta.
   Nota: En caso de detectarse algn dato atípico, no quitarlo.
2. Calcular los coeficientes de asimetría y apuntamiento e interpretarlos. Se puede concluir que la muestra proviene de una población normal?

Solución

1. El 11 es un dato atípico.
2. $g_1 = 0{,}61$ lo que indica que las distribución es asimétrica hacia la derecha, y $g_2 = 0{,}084$, lo que indica que la distribución es un poco platicúrtica. Como ambos valores están en el intervalo $[-2, 2]$ se puede suponer que la muestra viene de una población normal.

★ En un estudio sobre el número de kilocalorías diarias en la dieta de los españoles mayores de edad se han obtenido los siguientes datos, en una muestra de tamaño 200, separada por sexos:

Kilocaloras	Hombres	Mujeres
[1000, 1400)	1	15
[1400, 1800)	10	24
[1800, 2200)	25	28
[2200, 2600)	34	15
[2600, 3000)	26	10
[3000, 3400)	12	0

1. Qué media de consumo de kilocaloras es más representativa de su distribución, la de hombres o la de mujeres? Justificar adecuadamente la respuesta.
2. Cuánto vale el percentil 90 del consumo de kilocaloras en los hombres?
3. Cuánto vale el coeficiente de apuntamiento de la distribución global, considerando hombres y mujeres? Interpretarlo.

Solución

Llamando h y m al número de kilocalorías en hombres y mujeres respectivamente:

1. $\bar{h} = 2407{,}407$ kcal, $s_h = 468{,}1927$ kcal y $cv_h = 0{,}1945$.
   $\bar{m} = 1917{,}391$ kcal $s_m = 484{,}6802$ kcal y $cv_m = 0{,}2528$.
2. El percentil 90 en los hombres vale 3040 kcal.
3. $g_2 = -0{,}73$, lo que indica que la distribución es platicúrtica.

2. Regresión y Correlación

4. 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:

a) Recta de regresión de X sobre Y.

b) Coeficiente de correlación e interpretación del mismo.

c) 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

a) $\bar{x} = 23$ puntos, $\bar{y} = 2{,}4$ asignaturas, $s_x^2 = 116$ puntos^2, $s_y^2 = 1{,}5733$ asignaturas^2, $s_x = 10{,}7703$ puntos, $s_y = 1{,}2453$ asignaturas y $s_{xy} = 9{,}8$ puntos·asignaturas.
Recta de regresión de X sobre Y: $x = 6{,}2288y + 8{,}0508$.

b) $r = 0{,}73$, 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).

5.

6. 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 pide:

a) Calcular la recta de regresión de la actividad sobre el tiempo transcurrido.

b) Según el modelo lineal, ¿cuánto tiempo debe pasar para que la actividad del fármaco sea del 80 %? ¿Cuándo será nula la actividad?

Solución

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

a) $\bar{t} = 3$ años, $\bar{a} = 72\,\%$, $s_t^2 = 2$ años^2, $s_a^2 = 264\,\%^2$, $s_{ta} = -22{,}8$ 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 y $t(0) = 9{,}2182$ años.

★ 7. La tabla siguiente representa la distribución bidimensional de frecuencias de una muestra de 80 personas en un estudio sobre la relación entre el nivel de colesterol en sangre (X) en mg/dl y la tensión arterial máxima (Y) en mmHg.

$X \setminus Y$	$[110, 130)$	$[130, 150)$	$[150, 170)$	n_x
$[170, 190)$		4		12
$[190, 210)$	10	12	4	
$[210, 230)$	7		8	
$[230, 250)$	1			18
n_y		30	24	

Se pide:

a) Completar la tabla.

b) Calcular la recta de regresión del nivel de colesterol sobre la tensión.

c) Usar el modelo lineal para predecir el colesterol esperado para una persona con una tensión arterial de 160 mmHg.

d) Según el modelo lineal, ¿cuál es la tensión arterial máxima esperada para una persona cuyo nivel de colesterol es 270 mg/dl.

Usar las siguientes sumas: $\sum x_i = 16960$ mg/dl, $\sum y_j = 11160$ mmHg, $\sum x_i^2 = 3627200$ (mg/dl)^2, $\sum y_j^2 = 1576800$ mmHg^2 y $\sum x_i y_j = 2378800$ mg/dl·mmHg.

Solución

a) Tabla de frecuencias

$X \setminus Y$	$[110, 130)$	$[130, 150)$	$[150, 170)$	n_x
$[170, 190)$	8	4	0	12
$[190, 210)$	10	12	4	26
$[210, 230)$	7	9	8	24
$[230, 250)$	1	5	12	18
n_y	26	30	24	80

b) $\bar{x} = 212$ mg/dl, $\bar{y} = 139{,}5$ mmHg, $s_x^2 = 396$ (mg/dl)^2, $s_y^2 = 249{,}75$ mmHg^2 y $s_{xy} = 161$ mg/dl·mmHg. Recta de regresión del nivel de colesterol sobre la tensión arterial: $x = 122{,}0721 + 0{,}6446y$.

c) $x(160) = 225{,}2152$ mg/dl.

d) Recta de regresión de la tensión arterial sobre el colesterol: $y = 0{,}4066x + 53{,}3081$.
$y(270) = 163{,}0808$ mmHg.

★ 8. En un estudio para relacionar la longitud de la línea de la vida de la mano izquierda y la duración de la vida de una persona se han obtenido datos de 50 personas con los siguientes resultados (X=longitud de la línea en cm, Y=edad al morir en años):

$$\sum y = 3333 \quad \sum y^2 = 231933 \quad \sum x = 459{,}9 \quad \sum x^2 = 4308{,}57 \quad \sum xy = 30949.$$

A la vista de estos resultados, ¿cuanto vivirá, por termino medio, una persona con una línea de longitud 7.5 cm? ¿Es fiable esta estimación?

Solución

$\bar{x} = 9{,}198$ cm, $\bar{y} = 66{,}66$ años, $s_x^2 = 1{,}568$ cm^2, $s_y^2 = 195{,}104$ años^2 y $s_{xy} = 6{,}393$ cm·años.
Recta de regresión de la edad al morir sobre la longitud de la línea de la vida: $y = 4{,}077x + 29{,}158$.
$y(7{,}5) = 59{,}736$ años.
$r^2 = 0{,}13$, lo que quiere decir que casi no hay relación lineal entre las variables y la predicción anterior no es fiable.

---

9. 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) $r = 0{,}5$.

---

10. 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) $x(5) = 2$.

★ 11. En el estudio de regresión lineal con dos variables X e Y se sabe que $\overline{x} = 30$, $\overline{y} = 70$ y el coeficiente de correlación lineal es 0,8. También se sabe que para $x = 42$ el valor que predice la recta de regresión para y es 78.

Se pide:

a) Calcular el valor de x que se predice cuando $y = 74$.
b) Explicar razonadamente en cuál de las dos variables es más representativa la media.

Solución

a) Recta de regresión de X sobre Y: $x = 0{,}96x - 37{,}2$.
$x(74) = 33{,}84$.
b) $cv_x = 0{,}0408\sqrt{s_{xy}} > cv_y = 0{,}0146\sqrt{s_{xy}}$ y por tanto es más representativa la media de Y pues tiene menor dispersión relativa.

★ 12. 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) Calcular el modelo lineal y el logarítmico de los kilos perdidos con respecto a los días de dieta.

c) 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?

Usar las siguientes sumas (X=Días de dieta e Y=Peso perdido): $\sum x_i = 564$ días, $\sum \log(x_i) = 45{,}8086$ log(días), $\sum y_j = 63{,}6$ kg, $\sum x_i^2 = 28234$ días^2, $\sum \log(x_i)^2 = 175{,}6603$ log(días)^2, $\sum y_j^2 = 347{,}7$ kg^2, $\sum x_i y_j = 3108{,}5$ días·kg, $\sum \log(x_i) y_j = 245{,}4738$ log(días)·kg.

Solución

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

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

c) 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, aunque la de 100 días no lo es tanto por estar fuera del rango de valores observados en la muestra.

★ 13. 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:

a) Calcular la covarianza de S y T.

b) ¿Se puede afirmar que S y T son independientes? Justificar la respuesta.

c) ¿Qué valor predice la correspondiente recta de regresión para $t = 2$?

Solución

a) $\bar{s} = 0$, $\bar{t} = 1{,}5$ y $s_{st} = 0$.

b) No podemos afirmar que S y T son independientes, sólo se puede afirmar que no hay relación lineal.

c) $s(2) = 0$.

---

14. 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:

a) Dibujar el diagrama de dispersión. Según este diagrama, ¿qué tipo de modelo explicaría mejor la relación entre le número de bacterias y las horas transcurridas?

b) Dibujar el diagrama de dispersión tomando una escala logarítmica para el número de bacterias.

c) Según el modelo anterior, ¿Cuántas bacterias tendríamos al cabo de 3 horas y media? ¿Y al cabo de 10 horas? ¿Son fiables estas predicciones?

d) ¿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:

c) $\bar{x} = 4$ horas, $\bar{z} = 4{,}5149$ log(bacterias), $s_x^2 = 6{,}6667$ horas^2, $s_z^2 = 0{,}8361$ log^2(bacterias) y $s_{xz} = 2{,}3466$ horas·log(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.

d) 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.

★ 15. La tabla siguiente contiene los datos de las presiones sistólicas de 15 individuos en función de la edad de estos.

Edad(x)	20	30	40	50	60
	121	131	132	136	134
Sistólica(y)	130	125	129	128	142
	125	128	131	134	137

a) ¿Qué porcentaje de la varianza de la presión sistólica se explica, mediante un modelo de regresión lineal, por la varianza de la edad?

b) ¿Qué edad le correspondería a un individuo que presenta una presión sistólica de 133? ¿Es fiable esta predicción? Razona la respuesta.

Solución

a) $\bar{x} = 40$ años, $\bar{y} = 130{,}867$ mmHg, $s_x^2 = 200$ años^2, $s_y^2 = 26{,}295$ mmHg^2 y $s_{xy} = 58{,}667$ años·mmHg.
$r^2 = 0{,}654$, luego el modelo lineal explica el 65,4 % de la varianza de la presión sistólica.

b) Recta de regresión de la edad sobre la presión sistólica: $x = 2{,}231y - 251{,}978$.
$x(133) = 44{,}745$ años. La predicción es bastante fiable pues el coeficiente de determinación es alto.

★ 16. En un banco de sangre se mantiene el plasma a 0ºF. Cuando se necesita para una transfusión se calienta en un horno a una temperatura constante de 120ºF. En un experimento se ha medido la temperatura del plasma a distintos instantes desde el comienzo del calentamiento. Los resultados son:

Tiempo (min)	5	8	15	25	30	37	45	60
Temperatura (ºF)	25	50	86	102	110	114	118	120

Se pide:

a) Dibujar el diagrama de dispersión. ¿Qué modelo expliaría la relación entre la temperatura y el tiempo?

b) ¿Qué transformación de escala tendríamos que realizar en las variables para tener una nube de puntos con una tendencia lineal? Hacer la representación gráfica.

c) Construir el modelo de regresión logarítmico de la temperatura sobre el tiempo.

d) Según el modelo, ¿qué temperatura habrá a los 15 minutos? ¿Es fiable la predicción? Justificar la respuesta.

Usar las siguientes sumas (X=Tiempo e Y=Temperatura): $\sum x_i = 225$ min, $\sum \log(x_i) = 24{,}5289$ log(min), $\sum y_j = 725$ °F, $\sum \log(y_j) = 35{,}2051$ log(°F), $\sum x_i^2 = 8833$ min^2, $\sum \log(x_i)^2 = 80{,}4703$ log(min)^2, $\sum y_j^2 = 74345$ °F^2, $\sum \log(y_j)^2 = 157{,}1023$ log(°F)^2, $\sum x_i y_j = 24393$ min·°F, $\sum x_i \log(y_j) = 1048{,}0142$ min· log(°F), $\sum \log(x_i) y_j = 2431{,}7096$ log(min)°F, $\sum \log(x_i)\log(y_j) = 111{,}1165$ log(min) log(°F).

Solución

a) Un modelo logarítmico.

b) Aplicar una transformación logarítmica al tiempo, $z = \log(x)$.

c) $\bar{z} = 28{,}125$ log(min), $s_z^2 = 0{,}6577$ log^2(min), $\bar{y} = 90{,}625$ °F, $s_y^2 = 1080{,}2344$ °F^2 y $s_{zy} = 26{,}0969$ log(min)°F.
Modelo logarítmico de la temperatura sobre el tiempo: $y = -31{,}0325 + 39{,}6781 \log(x)$.

d) $y(15) = 76{,}4176$ °F. $r^2 = 0{,}9586$, que está muy cerca de 1, por lo que la predicción es fiable.

★ 17. 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:

Tiempo (horas)	2	3	4	5	6	7	8
Concentración de fármaco en sangre	25	36	48	64	86	114	168

Se pide:

a) Construir el modelo de regresión lineal de la concentración del fármaco sobre el tiempo.

b) Construir el modelo de regresión exponencial de la concentración del fármaco sobre el tiempo.

c) Usar el mejor de los dos modelos anteriores para predecir la concentración de fármaco en sangre que habrá a las 4,8 horas. ¿Es fiable la predicción?

Usar las siguientes sumas (C=Concentración del fármaco y T=tiempo): $\sum t_i = 35$ h, $\sum \log(t_i) = 10{,}6046$ log(h), $\sum c_j = 541$ mg/dl, $\sum \log(c_j) = 29{,}147$ log(mg/dl), $\sum t_i^2 = 203$ h^2, $\sum \log(t_i)^2 = 17{,}5206$ log(h)^2, $\sum c_j^2 = 56937$ (mg/dl)^2, $\sum \log(c_j)^2 = 124{,}0131$ log(mg/dl)^2, $\sum t_i c_j = 3328$ h·mg/dl, $\sum t_i \log(c_j) = 154{,}3387$ h· log(mg/dl), $\sum \log(t_i) c_j = 951{,}6961$ log(h)·mg/dl, $\sum \log(t_i)\log(c_j) = 46{,}08046$ log(h) · log(mg/dl).

Solución

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

a) $\bar{t} = 5$ h, $\bar{c} = 77{,}2857$ mg/dl, $s_t^2 = 4$ h^2, $s_c^2 = 2160{,}7755$ (mg/dl)^2, $s_{tc} = 89$ h(mg/dl).
Modelo lineal de C sobre T: $c = -33{,}9643 + 22{,}25t$.
$r^2 = 0{,}9165$.

b) $\bar{z} = 4{,}1639$ log(mg/dl), $s_z^2 = 0{,}3785$ log^2(mg/dl), $s_{tz} = 1{,}2291$ h· log(mg/dl).
Modelo exponencial de C sobre T: $c = e^{0{,}3073x+2{,}6275}$.
$r^2 = 0{,}9979$.

c) $c(4{,}8) = 60{,}498$ mg/dl y es bastante fiable ya que el coeficiente de determinación es muy alto.

---

18. La actividad de una sustancia radiactiva en función del tiempo (en número de desintegraciones por segundo) viene dada por la siguiente tabla:

t (horas)	0	10	20	30	40	50	60	70
A (10^7 desintegraciones/s)	25,9	8,16	2,57	0,81	0,25	0,08	0,03	0,01

a) Representar los datos de la actividad en función del tiempo. A la vista de la representación, ¿qué modelo de regresión explicaría mejor la relación entre la actividad y el tiempo transcurrido?

b) Representar los datos de la actividad en función del tiempo en papel semilogarítmico (con escala logarítmica en el eje de ordenadas).

c) Calcular la ecuación de la recta de regresión del logarítmo neperiaro de la actividad en función del tiempo.

d) Teniendo en cuenta que, en teoría, la actividad de una sustancia radiactiva en función del tiempo viene dada por la ecuación:

$$A(t) = A_0 e^{-\lambda t}$$

donde A_0 es la actividad inicial y λ es la llamada constante de desintegración, propia de cada sustancia radiactiva, utilizar la pendiente de la ecuación de la recta obtenida en el apartado anterior para calcular la constante de desintegración radiactiva de la sustancia con la que se han generado los datos.

Solución

Llamando X al tiempo e Y al logaritmo de la actividad

c) $\bar{x} = 35$, $\bar{y} = -0{,}7421$, $s_x^2 = 525$, $s_y^2 = 6{,}6664$ y $s_{xy} = -59{,}1434$.
Recta de regresión de Y sobre X: $y = -0{,}1127x + 3{,}2008$.

d) $\lambda = 0{,}1127$.

---

19. Para oscilaciones de pequeña amplitud, el periodo T de oscilación de un péndulo simple viene dado por:

$$T = 2\pi\sqrt{\frac{L}{g}}$$

donde L es la longitud del péndulo y g la aceleración de la gravedad. Para comprobar que dicha ley es cierta, se mide T para varias longitudes del péndulo, obteniéndose la siguiente tabla:

L (cm)	52,5	68,0	99,0	116,0	146,0
T (seg)	1,449	1,639	1,999	2,153	2,408

Se pide:

a) Representar los datos del periodo de oscilación frente a la longitud del péndulo. ¿Sería adecuado un modelo lineal para ajustar la nube de puntos?

b) Representar los datos del periodo de oscilación frente a la longitud en papel logarítmico (con escala logarítmica tanto en el eje de abcisas como en el de ordenadas). ¿Qué modelo de regresión sería adecuado para ajustar la nube de puntos obtenida?

c) Tomar logaritmos neperianos tanto del periodo de oscilación como de la longitud y representar en una gráfica los logaritmos obtenidos. ¿Qué modelo de regresión sería adecuado para ajustar la nube de puntos obtenida?

d) Calcular la ecuación de la recta de regresión que mejor ajusta la nube de puntos del apartado anterior.

e) Teniendo en cuenta el valor del término independiente de la recta obtenida en el apartado anterior, calcular el valor de g.

Solución

Llamando X al logaritmo del la longitud e Y al logaritmo del periodo:

d) $\bar{x} = 4{,}5025$ logcm, $\bar{y} = 0{,}6407$ logs, $s_x^2 = 0{,}1353$ log^{2}cm, $s_y^2 = 0{,}0339$ log^{2}s, $s_{xy} = 0{,}0677$ logcm·logs.
Recta de regresión de Y sobre X: $y = 0{,}5006x - 1{,}6132$.

e) $g = 994,4145$ cm/s^2.

---

20. En análisis colorimétrico, es frecuente utilizar la fracción de luz que absorbe una determinada sustancia disuelta como una medida de la concentración con la que dicha sustancia está presente en la disolución, siempre y cuando se utilice luz monocromática y la misma longitud recorrida por la luz en cada una de las mediciones. Si llamamos I_0 a la intensidad de luz incidente, I a la intensidad de luz transmitida y C a la concentración de la sustancia analizada, en un experimento de análisis colorimétrico realizado con Mn y una longitud de onda de 525 nm, se han obtenido los siguientes datos, donde la concentración de Mn viene dada en mg por cada 100 ml de disolución:

C	1,00	2,00	3,00	4,00
I/I_0	0,418	0,149	0,058	0,026

Se pide:

a) Representar los datos considerando I/I_0 en función de función de C. A la vista de la nube de puntos, ¿qué modelo de regresión sería el más adecuado para expresar la relación entre las variables?

b) Representar los datos pero en papel semilogarítmico.

c) Calcular la ecuación de la recta de regresión del logaritmo neperiano de I/I_0 frente a C.

Solución

Llamando C a la concentración e Z al logaritmo neperiano de I/I_0.

c) $\bar{c} = 2{,}5$ mg/100ml, $\bar{z} = -2{,}3183$, $s_c^2 = 1{,}25$ (mg/100ml)^2, $s_z^2 = 1{,}0788$ y $s_{cz} = -1{,}1595$.
Recta de regresión de Z sobre C: $z = -0{,}9276c + 0{,}0007$.

★ 21. Se quiere estudiar la relación entre las concentraciones de dos sustancias X e Y en la sangre. Para ello se han medido las concentraciones de estas sustancias en siete individuos, ambas en microgramos por decilitro de sangre, obteniendo los siguientes resultados

X	2,1	4,9	9,8	11,7	5,9	8,4	9,2
Y	1,3	1,5	1,7	1,8	1,5	1,7	1,7

Se pide:

a) ¿Existe relación lineal entre Y y X?

b) ¿Existe relación potencial entre Y y X?

c) Utilizar el mejor de los modelos anteriores para predecir la concentración de Y para $x = 8$ μgr/dl. ¿Es fiable la predicción? Justificar la respuesta.

Solución

a) Modelo lineal: $r^2 = 0{,}9696$, luego existe una relación lineal muy fuerte.

b) Modelo potencial: $r^2 = 0{,}9688$, luego también existe una relación potencial muy fuerte pero un poco menor que la lineal.

c) $y(8) = 1{,}6296$ μgr/dl.

---

3. Probabilidad

22. 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.

23. Construir el espacio muestral de los siguientes experimentos aleatorios:

    a) Seleccionar una persona al azar y medir su sexo y si es fumadora o no.

    b) Seleccionar una persona al azar y medir su grupo sanguíneo y si es fumadora o no.

    c) Seleccionar una persona al azar y medir su sexo, grupo sanguíneo y se es fumadora o no.

24. 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.

---

25. 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.

---

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

---

27. 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.

---

28. 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 incompatibles.

b) A y B son independientes.

Solución

a) $P(B)=0{,}3$.

b) $P(B)=0{,}75$.

---

29. 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.

★ 30. Un equipo de atención primaria de salud realiza un estudio de la población, para evaluar la incidencia de hipertensión e hipercolesterolemia. Para ello analizan a 1000 personas de dicha población, seleccionadas aleatoriamente, encontrándose que 180 presentan hipertensión, 140 hipercolesterolemia y 800 ninguna de ambas. Se pide calcular la probabilidad de que una persona tomada al azar.

a) Presente ambas enfermedades.

b) Presente hipertensión si no presenta hipercolesterolemia.

Solución

Llamando HT a tener hipertensión y HC a tener hipercolesterolemia:

a) $P(HT \cap HC) = 0{,}12$.

b) $P(HT/\overline{HC}) = 0{,}0698$.

★ 31. 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.

★ 32. En un experimento aleatorio se pueden dar dos sucesos A y B, y se sabe que $P(B) = 0{,}4$, $P(A/B) = 0{,}3$ y $P(A/\overline{B}) = 0{,}2$. Calcular las siguientes probabilidades:

a) $P(A)$.

b) $P(\overline{A} \cap \overline{B})$.

c) $P(\overline{A} \cup \overline{B})$.

Solución

a) $P(A) = 0{,}24$.

b) $P(\overline{A} \cap \overline{B}) = 0{,}52$.

c) $P(\overline{A} \cup \overline{B}) = 0{,}88$.

★ 33. 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.

★ 34. 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.

★ 35. 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 $P(-/\overline{S}) = 0{,}9919$.

★ 36. 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:

a) Si se selecciona un enfermo al azar, ¿cuál es la probabilidad de que haya sanado? ¿Y de que haya tenido algún efecto secundario?

b) 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?

c) 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:

a) $P(S) = 0{,}868$ y $P(E) = 0{,}114$.

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

c) $P(S \cap \overline{E}) = 0{,}806$.

★ 37. 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.

a) ¿Cuál fue la incidencia de la epidemia de gripe?
(Nota: La incidencia de una epidemia es la probabilidad de personas infectadas).

b) ¿Cuál es la probabilidad de que una persona no vacunada contraiga la gripe?

c) ¿Se puede afirmar que la vacuna tiene alguna eficacia?

Solución

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

a) $P(G) = 1/6$.

b) $P(G/\overline{V}) = 0{,}2$.

c) Si resulta eficaz, aunque poco.

★ 38. Una enfermedad se trata con 3 medicamentos diferentes: A en un 50 % de los casos, B en un 30 % y C en un 20 %, todo ello independientemente de si se es hombre o mujer. Si sabemos que el medicamento A produce efectos secundarios en un 5 % de los hombres y en un 10 % de las mujeres, el B en un 15 % de los hombres y en un 5 % de las mujeres, y el C en un 8 % de los hombres y en un 13 % de las mujeres, se pide:

a) ¿En qué colectivo resulta más probable que haya efectos secundarios, en hombres o en mujeres? Justificar adecuadamente la respuesta.

b) Calcular la probabilidad de que un hombre que presenta efectos secundarios haya sido tratado con C, y la de que una mujer que no los presenta haya sido tratada con A.

c) Si en total de enfermos hay un 65 % de hombres y un 35 % de mujeres, ¿qué probabilidad hay de que un enfermo que no presenta efectos secundarios sea mujer?

Solución

Llamando EH a que un hombre tenga efectos secundarios y EM a que los tenga una mujer:

a) $P(EH) = 0{,}086$ y $P(EM) = 0{,}091$.

b) $P(C/EH) = 0{,}186$, y $P(A/\overline{EM}) = 0{,}495$.

c) $P(M/\overline{E}) = 0{,}349$.

★ 39. Al aplicar un test diagnóstico a una población se obtuvieron un 1 % de personas enfermas en las que el test dio negativo, un 2 % de personas sanas en las que el test dio positivo, y un 90 % de personas sanas en las que el test dio negativo. Se pide:

a) Calcular la prevalencia de la enfermedad.

b) Calcular la sensibilidad del test.

c) Calcular la especificidad del test.

Solución

a) $P(E) = 0{,}08$.

b) $P(+/E) = 0{,}875$.

c) $P(-/\overline{E}) = 0{,}9783$.

★ 40. Según la clasficación de la New York Heart Association, el grado funcional de insuficiencia cardiaca se clasifica en 4 categorías dependiendo del esfuerzo físico para que se produzca disnea (dificultad respiratoria o falta de aire):

- A la categoría A pertenecen los pacientes en los que la disnea se produce sólo en niveles de esfuerzo altos.
- A la categoría B pertenecen los que la disnea se produce en niveles de esfuerzo medianos.
- A la categoría C pertenecen los que la disnea se produce en niveles de esfuerzo pequeños.
- A la categoría D pertenecen los que la disnea se produce incluso en reposo.

En un hospital se está investigando la evolución en el grado funcional de insuficiencia cardíaca como consecuencia de un tipo determinado de intervención en el corazón. Para los pacientes en los que se procedería a realizar la intervención, se observó que el 10 % pertenecían a la categoría A, el 20 % a la categoría B, el 30 % a la categoría C y el 40 % a la D. Después de la intervención todos los pacientes de la categoría A siguieron en A; el 50 % de los de B pasó a A y el resto siguió en B; el 30 % de los de C pasó a A, el 40 % pasó a B y el resto se quedó en C; el 10 % de los de D pasó a A, el 30 % pasó a B, el 40 % pasó a C y el resto siguió en D.

Se pide:

a) Si se toma al azar un paciente de dicho hospital que cumple los criterios para la intervención, ¿cuál es la probabilidad de que después de la misma esté en la categoría C?

b) Si sabemos que un paciente después de intervenido pertenece a la categoría B, ¿cuál es la categoría de la que resulta más probable que proceda?

c) Si el hospital trabaja con un total de 10000 pacientes intervenidos, ¿cuántos en ningún caso han pertenecido a la categoría C, ya sea antes o después de la intervención? ¿Y cuántos han pertenecido a la categoría A, ya sea antes o después de la intervención?

Solución

Llamando a, b, c y d a los sucesos consistentes en pertenecer a las categorías A, B, C, y D respectivamente antes de la intervención, y A, B, C, y D a los sucesos consistentes en pertenecer a esas categorías después:

a) $P(C) = 0{,}25$.

b) $P(b/B) = 0{,}2941$, $P(c/B) = 0{,}3529$, $P(d/B) = 0{,}3529$.

c) $P(\overline{c} \cap \overline{C}) = 0{,}54$ y $10000 \cdot 0{,}54 = 5400$ personas no han pertenecido nunca a la categoría C. $P(a \cup A) = 0{,}33$ y $10000 \cdot 0{,}33 = 3300$ personas que han estado en la categoría A antes o después de la intervención.

4. Variables Aleatorias

41. 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 $P(5 < X < 6) = 0$.

---

42. 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) Calcular la siguientes probabilidades:
1) $P(X = 6)$.
2) $P(X = 5)$.
3) $P(2 < X < 5{,}5)$.
4) $P(0 \leq X < 4)$.
c) Calcular la media.
d) Calcular la desviación típica.

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 $P(0 \leq X < 4) = 1/5$.
c) $\mu = 3{,}9$.
d) $\sigma = 1{,}6703$.

★ 43. 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. ¿Es representativa la media?

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, $\sigma^2 = 0{,}7$ ratas^2 y $\sigma = 0{,}84$ ratas.

---

44. 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) $P(X \geq 4) = 0{,}9526$.

---

45. Se sabe que la probabilidad de que aparezca una bacteria en un mm^3 de cierta disolución es de 0,002. Si en cada mm^3 a los sumo puede aparecer una bacteria, determinar la probabilidad de que en un cm^3 haya como máximo 5 bacterias.

Solución

Llamando X al número de bacterias en 1 cm^3 de disolución, se tiene $X \sim B(1000,\ 0{,}002) \approx P(2)$.
$P(X \leq 5) = 0{,}9834$.

★ 46. 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) $A = \{$durante 2 segundos lleguen a la centralita menos de 4 llamadas$\}$.

b) $B = \{$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 $P(Y \geq 3) = 0{,}938$.

---

47. Un proceso de fabricación de un fármaco produce por término medio 6 fármacos defectuosos por hora. ¿Cuál es probabilidad de que en un hora se produzcan menos de 3 fármacos defectuosos? ¿Y la de que en la próxima media hora se produzcan más de un fármaco defectuoso?

Solución

Llamando X al número de fármacos defectuosos en 1 hora, se tiene $X \sim P(6)$ y $P(X < 3) = 0{,}062$.
Llamando Y al número de fármacos defectuosos en 1/2 hora, se tiene $Y \sim P(3)$ y $P(Y > 1) = 0{,}8009$.

---

48. 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 $P(X \geq 4) = 0{,}1733$.

★ 49. 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 $P(Y \geq 9) = 0{,}9997$.

★ 50. 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 $P(Z < 3) = 0{,}6769$.

---

51. Por estudios previos se sabe que, en una comarca, hay dos tipos de larvas que parasitan, de forma completamente independiente, los chopos, y que producen su muerte. Si la larva de tipo A está parasitando un 15 % de los chopos, y la B un 30 %, y en una zona concreta de la comarca hay 10 chopos:

a) ¿Qué probabilidad hay que de estén siendo parasitados por A más de dos?

b) ¿Qué probabilidad hay de que estén libres de B más de 8?

c) ¿Qué probabilidad hay de que más de 1 tenga los dos tipos de larva?

d) ¿Qué probabilidad hay de que más de 3 tengan algún tipo de larva?

Solución

a) X_A es el número de chopos parasitados por larvas del tipo A, entonces $X_A \sim B(10,\ 0{,}15)$ y $P(X_A > 2) = 0{,}1798$.

b) Si $X_{\overline{B}}$ es el número de chopos no parasitados por larvas del tipo B, entonces $X_{\overline{B}} \sim B(10,\ 0{,}7)$ y $P(X_{\overline{B}} > 8) = 0{,}1493$.

c) Si llamamos $X_{A\cap B}$ al número de chopos parasitados por larvas de ambos tipos, entonces $X_{A\cap B} \sim B(10,\ 0{,}045)$ y $P(X_{A\cap B} > 1) = 0{,}0717$.

d) Si llamamos $X_{A\cup B}$ al número de chopos parasitados por algún tipo de larva, entonces $X_{A\cup B} \sim B(10,\ 0{,}405)$ y $P(X_{A\cup B} > 3) = 0{,}6302$.

---

52. 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 $X_A \sim B(500,\ 0{,}0018) \approx P(0{,}9)$ 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 $X_A \sim B(500,\ 0{,}0062) \approx P(3{,}1)$ y $P(X_{A\cup B} \geq 1) = 0{,}9550$.

---

53. 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 $P(X \geq 1) = 0{,}8364$.

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

a) Si un año van 6 alumnos a hacer prácticas en un hospital, ¿qué probabilidad hay de que vayan más chicos que chicas?

b) 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 $P(Y > 1) = 0{,}0199$.

---

55. Una variable aleatoria continua X tiene una función de densidad dada por:

$$f(x) = \begin{cases} k(6-3x) & \text{si } 0 \leq x \leq 2, \\ 0 & \text{si } x < 0 \text{ o } x > 2. \end{cases}$$

a) Determinar el valor de k.

b) Hallar $P(X \leq 1)$; $P(X > 2)$; $P(X = 1/4)$; $P(1/3 \leq X \leq 2/3)$.

c) Calcular μ y σ.

d) Hallar la función de distribución $F(x)$.

Solución

a) $k = 1/6$.

b) $P(X \leq 1) = 0{,}75$, $P(X > 2) = 0$, $P(X = 1/4) = 0$ y $P(1/3 \leq X \leq 2/3) = 1/4$.

c) $\mu = 2/3$, $\sigma^2 = 2/9$ y $\sigma = \sqrt{2}/3$.

d)

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ x - \frac{x^2}{4} & \text{si } 0 \leq x \leq 2, \\ 1 & \text{si } 2 < x. \end{cases}$$

★ 56. La función de densidad de la variable aleatoria continua X viene dada por la gráfica siguiente:

b) Comprobar que $f(x)$ es función de densidad.

Calcular la función de distribución.

Calcular la probabilidad de que la batería dure menos de 5 años, y de que dure entre 5 y 10 años.

Calcular la vida media de la batería.

Solución

a) $\int_{-\infty}^{\infty} f(x)\,dx = [-e^{-x/10}]_{-\infty}^{\infty} = 1.$

b)

$$F(x) = \begin{cases} 0 & \text{si } s < 0, \\ 1 - e^{-x/10} & \text{si } x \geq 0. \end{cases}$$

c) $P(X < 5) = 0{,}3935$ y $P(5 < X < 10) = 0{,}2387$.

d) $\mu = 10$ años.

---

57. 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.

---

58. 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) $t = -0{,}53$.

---

59. 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) Calcular la probabilidad de que una persona diabética elegida al azar tenga un nivel de glucosa inferior a 120mg/100ml.

    b) ¿Qué porcentaje de diabéticos tendrá niveles entre 90 y 120 mg/100 ml?

    c) Calcular e interpretar el primer cuartil del nivel de glucosa.

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.

---

60. En una población con 40000 personas, se sabe que 2276 tienen entre 0.80 y 0.84 miligramos de bilirrubina por decilitro de sangre, y que 11508 tienen más de 0.84. Suponiendo que la concentración de bilirrubina en sangre sigue una distribución normal, se pide:

    a) Calcular su media y su desviación típica.

    b) Calcular el número de personas con más de 1 miligramo de bilirrubina por decilitro de sangre.

Solución

a) $\mu = 0{,}7$ mg y $\sigma = 0{,}25$ mg.

b) $P(X > 1) = 0{,}1151$ y el número de personas con más de 1 mg de bilirrubina es $0{,}1151 \cdot 40000 = 4604$.

★ 61. 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.

★ 62. 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.

★ 63. Una solución contiene virus bacteriófagos T_4 en una concentración de $4 \cdot 10^6$ por mm^3. En la misma solución hay $2 \cdot 10^6$ bacterias por mm^3. Suponiendo que todos los virus infectan bacterias y que se distribuyen al azar entre las mismas, se pide:

a) ¿Cuál es el porcentaje de bacterias que no están infectadas por el virus?

b) ¿Qué porcentaje de bacterias tendrá al menos 2 virus fijados sobre ellas?

c) Si tomamos un volumen pequeño de dicha solución en el que hay 4 bacterias, ¿cuál es la probabilidad de que alguna esté infectada?

d) Si tomamos un volumen en el que hay 10000 bacterias, ¿cuál es la probabilidad de que estén infectadas al menos 8600?

★ 64. El gasto mensual en medicamentos de las familias españolas antes de la crisis seguía una distribución normal $X \sim N(160, \sigma)$, mientras que ahora sigue una distribución normal $Y \sim N(\mu, 2\sigma)$. Sabiendo que antes de la crisis el 10 % de las familias gastaba más de 200 euros y que después el 40 % gastaba menos de 100 euros, se pide:

a) ¿Qué porcentaje de familias gastarán ahora entre 110 y 120 euros?

b) ¿Qué percentil de la distribución actual se corresponde con el tercer decil de la distribución de antes de la crisis?

Solución

$X \sim N(160, 31{,}25)$ y $Y \sim N(115{,}625, 62{,}50)$.

a) $P(110 \leq Y \leq 120) = 0{,}0638$.

b) El decil tercero en X es $D_3 = 143{,}75$ euros y se corresponde aproximadamente con el percentil 67 de Y.

★ 65. Para el tratamiento de una enfermedad se utilizan 3 fármacos diferentes: A en un 20 % de los casos, B en un 30 %, y C en un 50 %. Los tratados con A curan en media al cabo de 6,2 días con una desviación típica de 1,1 días; los tratados con B en 7,1 días con una desviación típica de 1,3 y los tratados con C en 7,4 días con una desviación típica de 0,9. Con ello:

a) En qué fármaco es mayor el percentil 90, en B o en C?

b) Qué probabilidad total hay de que tomado un enfermo al azar haya curado antes de 8,1 días?

c) Sabiendo que un enfermo ha tardado en curar más de 7,2 días, con qué fármaco es más probable que haya sido tratado? Justificar adecuadamente la respuesta.

Solución

Llamando X al tiempo de curación y A, B y C a los sucesos consistentes en haber sido tratado respectivamente con cada uno de los fármacos:

a) El percentil 90 en el fármaco B es 8,7660 días y en C es 8,5534 días, así que es mayor en el fármaco B.

b) $P(X < 8{,}1) = 0{,}8161$.

c) $P(A/X > 7{,}2) = 0{,}0771$, $P(B/X > 7{,}2) = 0{,}2989$ y $P(C/X > 7{,}2) = 0{,}624$, así que es más probable que haya sido tratado con el fármaco C.

★ 66. Se sabe que la concentración de urea láctica, en mg/cm^3, en vacas sanas sigue una distribución normal de media 28 y desviación típica 1,5, mientras que en vacas con una determinada enfermedad sigue una distribución normal de media 34,5 y desviación típica 2. Para detectar la enfermedad se realiza un test diagnóstico que da positivo cuando el nivel de urea láctica está por encima de 32. Se pide:

a) ¿Qué sensibilidad $P(+/E)$ y qué especificidad $P(-/\overline{E})$ tiene el test diagnóstico?

b) Si el test da positivo en un 5 % de los casos analizados, ¿cuál será el porcentaje de vacas enfermas en la población?

c) Teniendo en cuenta el porcentaje de vacas enfermas anterior, ¿cuál será la probabilidad de diagnóstico acertado con este test?

Solución

a) $P(+/E) = 0{,}8944$ y $P(-/\overline{E}) = 0{,}9962$.

b) $P(E) = 0{,}0465$.

c) $P(E \cap +) + P(\overline{E} \cap -) = 0{,}0416 + 0{,}9498 = 0{,}99143$.

5. Estimación de parámetros

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

Solución

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

68. Para determinar si un pescado es o no apto para el consumo por su contenido en Hg (mercurio), se realizan 15 valoraciones obteniendo una media de 0,44 ppm (partes por millón) de Hg, y una desviación típica de 0,08 ppm. Calcular los límites de confianza para la media, a un nivel de significación $\alpha = 0{,}1$.

Solución

$\mu \in 0{,}44 \pm 0{,}0376$ ppm $= (0{,}4024\text{ppm},\ 0{,}4776\text{ppm})$.

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

$$7{,}90 \quad 7{,}85 \quad 7{,}89 \quad 7{,}86 \quad 7{,}87.$$

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

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

70. Se desea saber cuál debe ser el tamaño muestral mínimo de una muestra para poder realizar la estimación de la tasa media de glucosa plasmática de una determinada población, con un nivel de confianza 0,95 y pretendiendo una amplitud de 2,5 mg.

NOTA: En una muestra previa de tamaño 10 se obtuvo una desviación típica de 10 mg.

Solución

249 individuos.

71. En una explotación minera se mide el contenido en mercurio de las rocas extraídas. Tras analizar 20 rocas, se obtiene un contenido medio del 10,8 % y una desviación típica de 2,7 %. Se pide:

    a) Si, para que la explotación sea rentable, el porcentaje medio de contenido de mercurio debe ser superior al 10 %, ¿existen pruebas para afirmar que la explotación será rentable?

    b) ¿Y si para que la explotación sea rentable, el contenido de mineral debe tener cierta uniformidad ($\sigma < 3$)?

72. 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: $n = 96$.

73. 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

    $p \in 0{,}25 \pm 0{,}0507 = (0{,}1993,\ 0{,}3007)$.

★ 74. 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.

75. 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

$\mu_{x_1-x_2} \in 24{,}0909 \pm 19{,}4766$ mg/dl $= (4{,}6143\text{mg}, 43{,}5675\text{mg/dl})$.

---

76. 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 recuperase, 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

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

★ 77. Un equipo de investigación está interesado en ver si una droga reduce el colesterol en la sangre. Con tal fin se toma una muestra de 10 pacientes y determina el contenido de colesterol antes y después del tratamiento. Los resultados expresados en miligramos por cada 100 mililitros son los siguientes:

Paciente	1	2	3	4	5	6	7	8	9	10
Antes	217	252	229	200	209	213	215	260	232	216
Después	209	241	230	208	206	211	209	228	224	203

Se pide:

a) Construir la variable Diferencia que recoja la diferencia entre los niveles de colesterol antes y después del tratamiento, y calcular el intervalo de confianza con $1-\alpha = 0{,}95$ para la media de dicha variable.

b) A la vista del intervalo anterior, ¿hay pruebas significativas de que la droga disminuye el nivel de colesterol en sangre?

Solución

a) $\mu_{x_1-x_2} \in 7{,}4 \pm 7{,}572$ mg/100ml $= (-0{,}172\text{mg/100ml}, 14{,}972\text{mg/100ml})$.

b) No se puede afirmar que la droga disminuya el colesterol con un 95 % de confianza.

---

78. Dos químicos A y B realizan 14 y 16 determinaciones, respectivamente, de plutonio. Los resultados obtenidos se muestran en la siguiente tabla

A		B	
263.36	254.68	286.53	254.54
248.64	276.32	284.55	286.30
243.64	256.42	272.52	282.90
272.68	261.10	283.85	253.75
287.33	268.41	252.01	245.26
287.26	282.65	275.08	266.08
250.97	284.27	267.53	252.05
		253.82	269.81

Se pide:

a) Calcular intervalos de confianza del 95 % de confianza para cada caso.

b) ¿Se puede decir que existen diferencias significativas en la media?

Solución

a) Intervalo de confianza del 95 % para la media de A: $\mu_A \in 266{,}98 \pm 8{,}681 = (258{,}299,\ 275{,}661)$.
Intervalo de confianza del 95 % para la media de B: $\mu_B \in 267{,}91 \pm 7{,}677 = (260{,}233,\ 275{,}587)$.

b) $\mu_A - \mu_B \in -0{,}93 \pm 11{,}020 = (-11{,}95,\ 10{,}92)$, luego no hay diferencias significativas en las medias con un 95 % de confianza.

★ 79. 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:

a) 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.

b) 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.

★ 80. 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:

a) 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.

b) 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

a) $p_R \in (0{,}3799,\ 0{,}4548)$ y $p_U \in (0{,}4185,\ 0{,}4856)$ con un 99 % de confianza.

b) $p_R - p_U \in (-0{,}0729,\ 0{,}0036)$ 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.

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