EJERCICIOS DE ESTADÍSTICA

Asignatura: Matemáticas
Curso: Primero
Grado: Farmacia
Año: 2017-2018

Autores: Pablo Ares Gastesi (pablo.aresgastesi@ceu.es)
Eduardo López Ramírez (elopez@ceu.es)
Anselmo Romero Limón (arlimon@ceu.es)
Alfredo Sánchez Alberca (asalber@ceu.es)

CEU Universidad San Pablo

Índice

1. Estadística Descriptiva

★ 1. Clasificar las siguientes variables:

   a) Horas diarias de ejercicio.
   b) Nacionalidad.
   c) Tensión arterial.
   d) Gravedad de una enfermedad.
   e) Número de lesiones en una temporada.
   f) Ingesta diaria de calorías.
   g) Talla de ropa.
   h) Asignaturas aprobadas en un curso.

   Solución

   a) Variable cuantitativa continua.
   b) Atributo nominal.
   c) Variable cuantitativa continua.
   d) Atributo ordinal.
   e) Variable cuantitativa discreta.
   f) Variable cuantitativa continua.
   g) Atributo ordinal.
   h) Variable cuantitativa discreta.

★ 2. 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:

   a) Construir la tabla de distribución de frecuencias.
   b) Dibujar el diagrama de barras y el polígono de frecuencias asociado.
   c) Dibujar el diagrama de barras de frecuencias acumuladas el polígono de frecuencias asociado.

3. Se realizó una encuesta a 40 personas de más de 70 años sobre el número de medicamentos distintos que tomaban habitualmente. El resultado de dicha encuesta fue el siguiente:

3	1	2	2	0	1	4	2	3	5	1	3	2	3	1	4	2	4	3	2
3	5	0	1	2	0	2	3	0	1	1	5	3	4	2	3	0	1	2	3

   Se pide:

   a) Construir la tabla de distribución de frecuencias de la muestra.
   b) Dibujar el diagrama de barras de frecuencias relativas y el polígono de frecuencias asociado.
   c) Dibujar el diagrama de barras de frecuencias relativas acumuladas y el polígono de frecuencias asociado.

4. 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 resultado obtenidos fueron

B D A B C C B C D E A B C E A B C D B B A A B

Construir la tabla de distribución de frecuencias y el diagrama correspondiente.

5. En un hospital se realizó un estudio sobre el número de personas que ingresaron en urgencias en el mes de noviembre. Los datos observados fueron:

15	23	12	10	28	7	12	17	20	21	18	13	11	12	26
30	6	16	19	22	14	17	21	28	9	16	13	11	16	20

Se pide:

a) Construir la tabla de distribución de frecuencias agrupando en clases.
b) Dibujar el diagrama más adecuado para representar la distribución de frecuencias.
c) Dibujar el diagrama más apropiado para representar la distribución de frecuencias acumuladas.

★ 6. La siguiente tabla representa la distribución de frecuencias del tiempo de atención (en minutos) en una consulta médica:

Intervalos	n_i	f_i	N_i	F_i
$[0,5)$	2			
$[5,10)$			8	
$[10,15)$				0.7
$[15,20)$	6			

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

★ 7. 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:

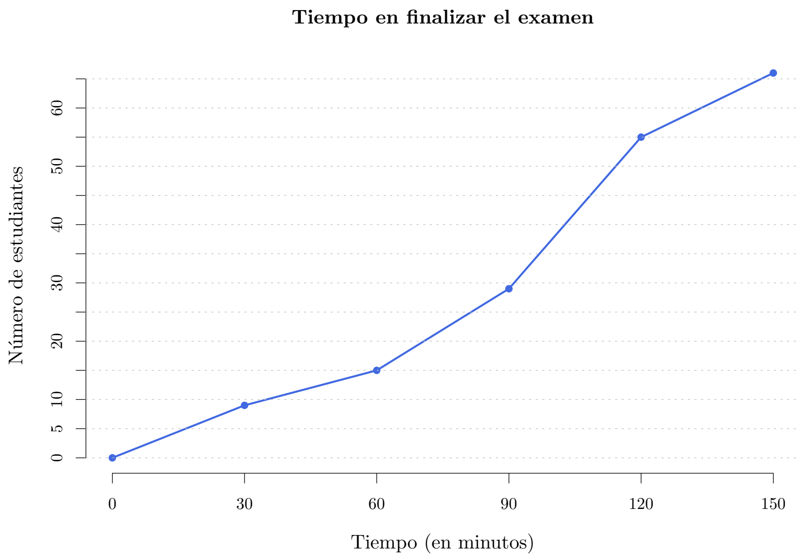
a) Media.
b) Mediana.
c) Moda.
d) Cuartiles.
e) Percentil 32.

Solución

a) $\bar{x} = 1{,}125$ lesiones.
b) $Me = 1$ lesiones.
c) $Mo = 1$ lesiones.
d) $C_1 = 0{,}5$ lesión, $C_2 = 1$ lesiones y $C_3 = 2$ lesiones.

e) $P_{32} = 1$ lesión.

★ 8. En un examen de estadística al que se han presentado 66 alumnos se ha contado el número de exámenes finalizados cada media hora, obteniendo el siguiente polígono:



Se pide:

a) ¿Cuánto tiempo tiene que pasar para que hayan finalizado el examen la mitad de los alumnos? ¿Y el 90 % de los estudiantes?

b) ¿Qué porcentaje de alumnos habrá terminado a los 100 minutos de examen?

c) ¿Cuál es el tiempo de duración del examen que mejor representa a los estudiantes de la muestra? ¿Es un valor representativo?

Solución

a) $Me = 94{,}6154$ minutos y $P_{90} = 132$ minutos.

b) 57,08 % de los estudiantes.

c) $\bar{x} = 85{,}9091$ minutos, $s^2 = 1408{,}2645$ minutos^2, $s = 37{,}5268$ minutos y $cv = 0{,}4368$, luego existe una dispersión moderada.

★ 9. 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
Un año:	62, 65, 69, 71, 65, 66, 68, 69

¿En cuál de las dos muestras es más representativa la media? Justificar la respuesta.

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.

10. 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):

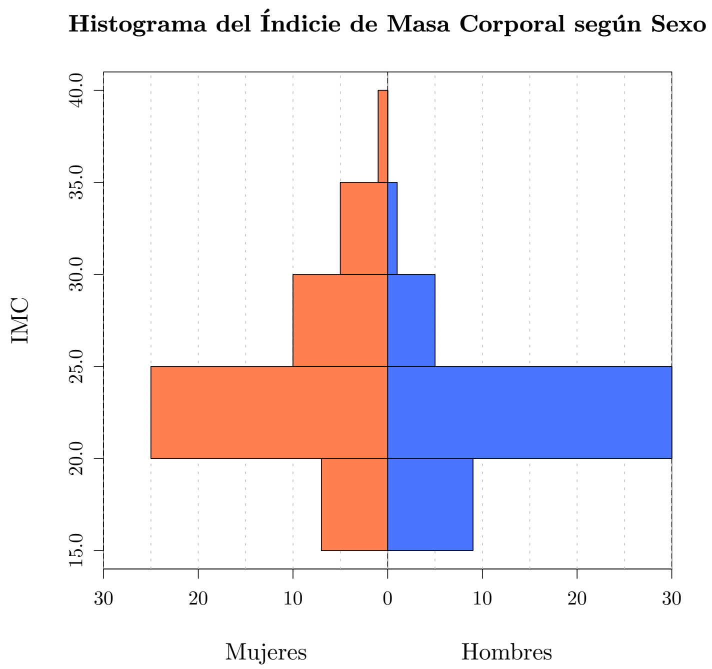
42,2 42,1 41,9 41,8 42 42,1 41,9 42

¿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.

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



Se pide:

a) Dibujar el diagrama de sectores para el sexo.

b) ¿En qué grupo es más representativa la media? Justificar la respuesta.

c) ¿Cómo calcularías la media de toda la muestra a partir de las medias de hombres y mujeres? ¿Cuánto vale?

Solución

b) $\bar{m} = 24{,}17$ 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.

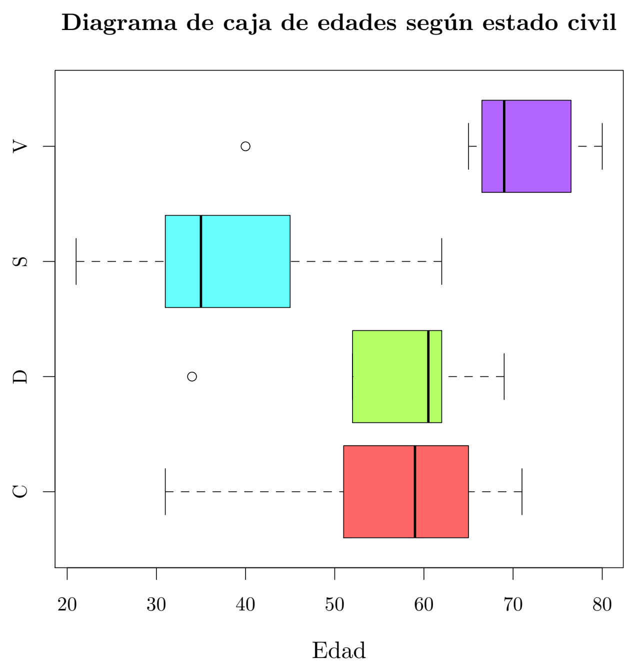
c) $\bar{x} = 23{,}25$ Kg/m^2.

★ 12. La siguiente tabla representa la distribución de frecuencias del número de usos del seguro de salud en un año de una muestra de clientes de una compañía de seguros.

Usos:	0	1	2	3	4	5	7
Clientes:	4	8	6	3	2	1	1

Dibujar el diagrama de caja. ¿Cómo es la simetría de la distribución?

★ 13. El diagrama de caja siguiente corresponde a la distribución de edades según el estado civil de una muestra de personas.



a) ¿Qué grupo tiene edades mayores?

b) ¿Qué grupo tiene una menor dispersión central en las edades?

c) ¿Qué grupos tienen datos atípicos?

d) ¿Qué grupo tienes una distribución de edad más asiméttrica?

Solución

a) Viudos.

b) Divorciados.

c) Viudos y divorciados.

d) Divorciados.

★ 14. 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.

★ 15. 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 $g_{2_B} = -0{,}83$.

★ 16. La presión arterial sistólica (en mmHg) de una muestra de personas es

$$135 \quad 128 \quad 137 \quad 110 \quad 154 \quad 142 \quad 121 \quad 127 \quad 114 \quad 103$$

a) Calcular los estadísticos de tendencia central.

b) ¿Cómo es la dispersión relativa con respecto a la media?

c) ¿Cómo es la asimetría de la distribución muestral?

d) ¿Cómo es la curtosis de la distribución muestral?

e) Si se sabe que el método utilizado para medir la presión está sesgado, y, para obtener los valores correctos, hay que aplicar la transformación $y = 1{,}2x - 5$, ¿cuáles serán los estadísticos anteriores para los valores correctos de la presión arterial sistólica?

Usar las siguientes sumas: $\sum x_i = 1271$ mmHg, $\sum (x_i - \bar{x})^2 = 2188{,}9$ mmHg^2, $\sum (x_i - \bar{x})^3 = 2764{,}32$ mmHg^3, $\sum (x_i - \bar{x})^4 = 1040080$ mmHg^4.

SOLUCIÓN

Llamando x a la presión arterial sistólica.

a) $\bar{x} = 127{,}1$ mmHg, $Me = 127{,}5$ mmHg, Mo todos los valores.

b) $s = 14{,}7949$ mmHg y $cv = 0{,}1164$.

c) $g_1 = 0{,}0854$ luego la distribución es casi simétrica.

d) $g_2 = -0{,}8292$ luego la distribución es menos apuntada de lo normal (platicúrtica).

e) $\bar{y} = 147{,}52$ mmHg, $Me = 148$ mmHg, $Mo = 157$ mmHg, $s = 17{,}7539$ mmHg, $cv = 0{,}1203$, $g_1 = 0{,}0854$ y $g_2 = -0{,}8292$.

★ 17. La siguiente tabla contiene las frecuencias de embarazos, abortos y nacimientos de una muestra de 999 mujeres de una ciudad.

Num	Embarazos	Abortos	Nacimientos
0	61	751	67
1	64	183	80
2	328	51	400
3	301	10	300
4	122	2	90
5	81	2	62
6	29	0	0
7	11	0	0
8	2	0	0

a) ¿Cuántos datos atípicos hay en el número de nacimientos de la muestra?

b) ¿Qué variable tiene menor dispersión con respecto a la media?

c) ¿Qué valor es relativamente mayor 7 embarazos o 4 abortos? Justificar la respuesta.

Usar las siguientes sumas:
Embarazos: $\sum x_i = 2783$, $\sum x_i^2 = 9773$.
Abortos: $\sum x_i = 333$, $\sum x_i^2 = 559$.
Nacimientos: $\sum x_i = 2450$, $\sum x_i^2 = 7370$.

SOLUCIÓN

Llamando x al número de embarazos, y al número de abortos y z al número de nacimientos.

a) 129 datos atípicos.

b) Embarazos: $\bar{x} = 2{,}7858$, $s_x = 1{,}422$ y $cv_x = 0{,}5105$.
Abortos: $\bar{y} = 0{,}3333$, $s_y = 0{,}6697$ y $cv_y = 2{,}009$.
Nacimientos: $\bar{z} = 2{,}4525$, $s_z = 1{,}1674$ y $cv_z = 0{,}476$.

c) La puntuación típica de 7 embarazos es 2,9635, y la puntuación típica de 4 abortos es 5,4754, luego es relativamente mayor el número de abortos.

---

2. Regresión y Correlación

★ 18. Dar ejemplos de:

a) Dos variables no relacionadas.

b) Dos variables relacionadas de manera creciente.

c) Dos variables relacionadas de manera decreciente.

★ 19. 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) Construir la tabla de la distribución conjunta de frecuencias.

b) Obtener las distribuciones de frecuencias marginales y calcular los principales estadísticos para cada variable.

c) Calcular la covarianza e interpretarla.

Solución

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

c) $\bar{x} = 3{,}05$ mg, $\bar{y} = 4{,}55$ días, $s_x^2 = 0{,}648$ mg^2, $s_y^2 = 1{,}448$ días^2, $s_x = 0{,}805$ mg, $s_y = 1{,}203$ días y $s_{xy} = -0{,}678$ mg·días, lo que indica que hay una relación lineal decreciente.

★ 20. 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.

---

21. 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.

★ 22. En un equipo de baloncesto se ha introducido un programa de estiramientos para ver si se consigue reducir el número de lesiones. Durante toda una temporada cada jugador realizó ejercicios de estiramiento durante un número fijo de minutos en cada entrenamiento. Al finalizar la temporada se midió el número de lesiones y se obtuvieron los resultados de la siguiente tabla:

Minutos de estiramiento	0	30	10	15	5	25	35	40
Lesiones	4	1	2	2	3	1	0	1

Se pide:

a) Calcular la recta de regresión del número de lesiones con respecto al tiempo de estiramiento.

b) ¿Cual es la disminución de lesiones esperada por cada minuto de estiramiento?

c) ¿Cuántos minutos de estiramiento debe realizar un jugador para no tener ninguna lesión? ¿Es fiable esta predicción?

Usar las siguientes sumas (X=Número de minutos de estiramiento, e Y=Número de lesiones): $\sum x_i =$ 160 minutos, $\sum y_j = 14$ lesiones, $\sum x_i^2 = 4700$ minutos^2, $\sum y_j^2 = 36$ lesiones^2 y $\sum x_i y_j = 160$ minutos·lesiones.

Solución

Llamando X a la variable que mide el tiempo de estiramiento, e Y a la que mide el número de lesiones en cada jugador:

a) Recta de regresión de Y sobre X: $y - 0{,}08x + 3{,}35$.
b) Por cada minuto más de estiramiento habrá 0,08 lesiones menos.
c) Para no tener ninguna lesión habrá que estirar al menos 38,26 minutos. $r = -0{,}91$, luego la predicción es bastante fiable.

---

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

---

24. 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 ($r = 0$).

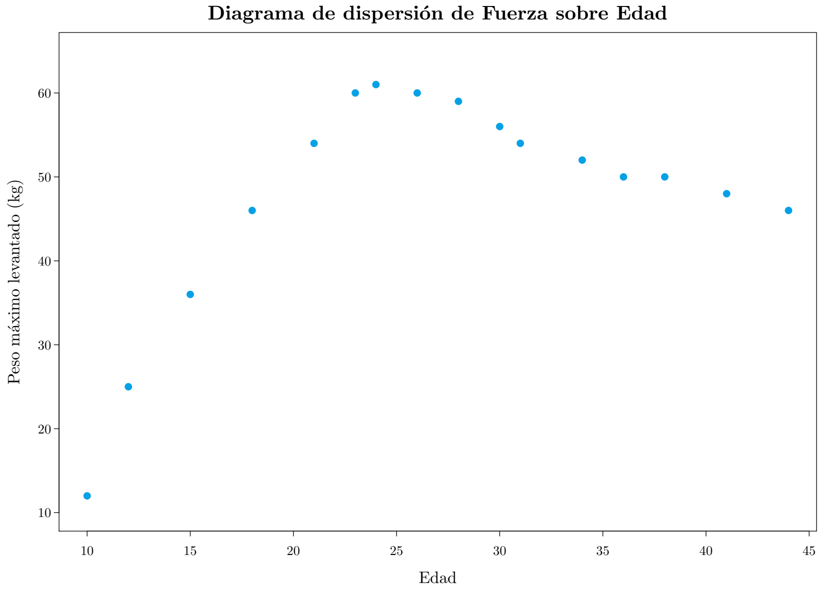
a) Si $x = 10$, ¿cuál será el valor interpolado para y?.
b) Si $y = 5$, ¿cuál será el valor interpolado para x?.
c) Dibuja las rectas de regresión de Y sobre$X$, y la de X sobre Y.

Solución

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

---

25. 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:



Se pide:

a) Calcular el coeficiente de determinación lineal de la muestra completa.
b) Calcular el coeficiente de determinación lineal de la muestra de personas menores de 25 años.
c) Calcular el coeficiente de determinación lineal de la muestra de personas mayores de 25 años.
d) ¿Para qué grupo de edades es más fuerte la relación lineal entre la fuerza física y la edad?

Usar las siguientes sumas (X=Edad e Y=Peso máximo levantado).

- Muestra total: $\sum x_i = 431$ años, $\sum y_j = 769$ kg, $\sum x_i^2 = 13173$ años^2, $\sum y_j^2 = 39675$ kg^2 y $\sum x_i y_j = 21792$ años·kg.
- Muestra menores de 25 años: $\sum x_i = 123$ años, $\sum y_j = 294$ kg, $\sum x_i^2 = 2339$ años^2, $\sum y_j^2 = 14418$ kg^2 and $\sum x_i y_j = 5766$ años·kg.
- Muestra mayores de 25 años: $\sum x_i = 308$ años, $\sum y_j = 475$ kg, $\sum x_i^2 = 10834$ años^2, $\sum y_j^2 = 25257$ kg^2 y $\sum x_i y_j = 16026$ años·kg.

Solución

Llamando X a la edad e Y al peso levantado.

a) $\bar{x} = 26{,}9375$ años, $s_x^2 = 97{,}6836$ años^2, $\bar{y} = 48{,}0625$ kg, $s_y^2 = 169{,}6836$ kg^2, $s_{xy} = 67{,}3164$ años·kg y $r^2 = 0{,}2734$.
b) $\bar{x} = 15{,}5714$ años, $\bar{y} = 42$ kg, $s_x^2 = 25{,}3878$ años^2, $s_y^2 = 295{,}7143$ kg^2, $s_{xy} = 85{,}7143$ años·kg y $r^2 = 0{,}9786$.
c) $\bar{x} = 35{,}2222$ años, $\bar{y} = 52{,}7778$ kg, $s_x^2 = 32{,}6173$ años^2, $s_y^2 = 20{,}8395$ kg^2, $s_{xy} = -25{,}5062$ años·kg y $r^2 = 0{,}9571$.
d) La relación lineal entre la edad y la fuerza física es un poco más fuerte en los menores de 25 años.

★ 26. 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) · 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.

★ 27. 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.

★ 28. Los investigadores están estudiando la correlación entre obesidad y la respuesta individual al dolor. La obesidad se mide como porcentaje sobre el peso ideal (x). La respuesta al dolor se mide utilizando el umbral de reflejo de flexión nociceptiva (y), que es una medida de sensación de punzada. Se obtuvieron los siguientes resultados:

x	89	90	75	30	51	75	62	45	90	20
y	10	12	4	4,5	5,5	7	9	8	15	3

a) Dibujar el diagrama de dispersión. Según la nube de puntos, ¿qué modelo explicaría mejor la relación entre el umbral de reflejo y el porcentaje sobre el peso ideal?

b) Según el modelo de regresión más adecuado, ¿qué respuesta al dolor se espera que tenga una persona con una obesidad del 50 %? ¿Es esta predicción fiable?

c) Según el modelo de regresión más adecuado, ¿cuál es la obesidad esperada de una persona con un umbral reflejo de flexión nociceptiva de 10? ¿Es esta predicción fiable?

Usar las siguientes sumas (X=Obesidad e Y=Respuesta al dolor): $\sum x_i = 629$, $\sum \log(x_i) = 40{,}4121$, $\sum y_j = 92{,}2$, $\sum \log(y_j) = 21{,}339$, $\sum x_i^2 = 45445$, $\sum \log(x_i)^2 = 165{,}6795$, $\sum y_j^2 = 960{,}14$, $\sum \log(y_j)^2 = 47{,}6231$, $\sum x_i y_j = 6537{,}7$, $\sum x_i \log(y_j) = 1443{,}1275$, $\sum \log(x_i) y_j = 387{,}5728$, $\sum \log(x_i)\log(y_j) = 88{,}3696$.

Solución

b) $\bar{x} = 62{,}9$, $s_x^2 = 588{,}09$, $\bar{y} = 9{,}22$, $s_y^2 = 11{,}0056$ and $s_{xy} = 82{,}0356$.
Recta de regresión de la respuesta al dolor sobre la obesidad: $y = 1{,}3232 + 0{,}1255x$. $r^2 = 0{,}8422$.
$\overline{\log(x)} = 4{,}0412$, $s^2_{\log(x)} = 0{,}2366$ and $s_{\log(x)y} = 1{,}4973$. Modelo logarítmico de la respuesta al dolor sobre la obesidad: $y = -16{,}3578 + 6{,}3293\log(x)$. $r^2 = 0{,}8611$.
$y(50) = 8{,}4023$.

c) Modelo exponencial de la obesidad sobre la respuesta al dolor: $x = e^{2{,}7868+0{,}1361y}$.
$x(10) = 63{,}2648$.

★ 29. 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), $\sum y_j = 725$ ºF, $\sum \log(y_j) = 35{,}2051$ log(ºF), $\sum x_i^2 = 8833$ min^2, $\sum \log(x_i)^2 = 80{,}4703$ log(min)^2, $\sum y_j^2 = 74345$ ºF^2, $\sum \log(y_j)^2 = 157{,}1023$ log(ºF)^2, $\sum x_i y_j = 24393$ min·ºF,

$\sum x_i \log(y_j) = 1048{,}0142$ min$\cdot \log(^{\text{o}}\text{F})$, $\sum \log(x_i) y_j = 2431{,}7096 \log(\text{min})^{\text{o}}\text{F}$, $\sum \log(x_i)\log(y_j) = 111{,}1165 \log(\text{min})\log(^{\text{o}}\text{F})$.

Solución

a) Un modelo logarítmico.

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

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

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

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

31. 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.

★ 32. 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

33. 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.

★ 34. Dos cajas contienen bolas con diferentes colores. La primera caja contiene 3 bolas blancas y 2 negras, y la segunda caja contiene 2 bolas azules, 1 bola roja y 1 bola verde. Construir el espacio muestral de los siguientes experimentos aleatorios:

a) Tomar una bola al azar de cada caja.
b) Tomar dos bolas al azar de cada caja.

★ 35. Las leyes de Morgan establecen que dados dos sucesos aleatorios A y B de un mimo espacio muestral, $\overline{A \cup B} = \bar{A} \cap \bar{B}$ y $\overline{A \cap B} = \bar{A} \cup \bar{B}$. Demostrar ambas igualdades usando diagramas de Venn.

★ 36. Calcular la probabilidad de los siguientes sucesos del experimento aleatorio consistente en lanzar 3 monedas:

a) Obtener exactamente una cara.
b) Obtener exactamente dos cruces.
c) Obtener dos o más caras.
d) Obtener alguna cruz.

Solución

a) $P(1 \text{ cara}) = 0{,}375$.
b) $P(2 \text{ cruces}) = 0{,}375$.
c) $P(2 \text{ o más caras}) = 0{,}5$.
d) $P(\text{alguna cruz}) = 0{,}875$.

37. 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.

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

39. 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.

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

★ 41. 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 suceso que consiste en tener un padre fumador y MF a tener una madre fumadora:

a) $P(MF/PF) = 0{,}325$.
b) $P(MF/\overline{PF}) = 0{,}2$.
c) No son independientes.

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

★ 43. La probabilidad de que una lesión A se reproduzca es $4/5$, la de que se reproduzca otra lesión B es $1/2$, y la de que ambas se reproduzcan $1/3$. Hallar la probabilidad de que:

a) Sólo se reproduzca la lesión B.

b) Al menos una se reproduzca.

c) Se reproduzca la lesión B si se ha reproducido la A.

d) Se reproduzca la lesión B si no se reproduce la lesión A.

SOLUCIÓN

Llamando A y B a los sucesos consistentes en que se reproduzcan las respectivas lesiones, se tiene:

a) $P(B \cap \overline{A}) = 1/6$.

b) $P(A \cup B) = 29/30$.

c) $P(B/A) = 5/12$.

d) $P(B/\overline{A}) = 5/6$.

★ 44. 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

$1/3$.

★ 45. 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.

★ 46. Para evaluar la efectividad de un test diagnóstico se aplicó el test a una muestra de personas con los siguientes resultados:

	Test +	Test −
Enfermos	2020	140
Sanos	80	7760

Calcular para este test:

a) La sensibilidad y la especificidad.

b) Los valores predictivo positivo y negativo.

c) La probabilidad de un diagnóstico acertado.

Solución

Llamando E y $\bar{E}$ a los eventos consistentes en estar enfermo y sano respectivamente.

a) Sensibilidad $P(+|E) = 0{,}9352$ y especificidad $P(-|\bar{E}) = 0{,}9898$.

b) VPP $P(E|+) = 0{,}9619$ y VPN $P(\bar{E}|-) = 0{,}9823$.

c) $P((E \cap +) \cup (\bar{E} \cap -)) = 0{,}978$.

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

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

★ 49. 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.

★ 50. Se dispone de dos test A y B para diagnosticar una enfermedad. El test A tiene una sensibilidad del 98 % y una especificidad del 80 %, mientras que el test B tiene una sensibilidad del 75 % y una especificidad del 99 %.

a) ¿Qué test es mejor para confirmar la enfermedad?

b) ¿Qué test es mejor para descartar la enfermedad?

c) A menudo un test se utiliza para descartar una enfermedad rara un gran numero de personas aparentemente sanas. Este tipo de test se conocen como *test de cribado*. ¿Cuál de los dos test funcionaría mejor como test de cribado? ¿Cuál es el valor predictivo positivo (VPP) de este test si la prevalencia de la enfermedad es 0,01? ¿Y si la prevalencia de la enfermedad es 0,2?

d) El valor predictivo positivo de un test de cribado no suele ser alto. ¿Cómo se pueden combinar ambos test para tener una confianza mayor en el diagnóstico de la enfermedad? Calcular la probabilidad a posteriori de tener la enfermedad con la combinación de ambos test si el resultado de ambos test es positivo y la prevalencia de la enfermedad es 0,01.

Solución

a) El test B ya que tiene una mayor especificidad.

b) El test A ya que tiene una mayor sensibilidad.

c) El test A funcionará mejor como un test de cribado.
Para una prevalencia de 0,01 el VVP es $P(E|+) = 0{,}0472$ y el VPN es $P(\bar{E}|-) = 0{,}9997$.
Para una prevalencia de 0,2 el VVP es $P(E|+) = 0{,}5506$ y el VPN es $P(\bar{E}|-) = 0{,}9938$.

d) Aplicando primero el test A a todo el mundo y luego el test B a los resultados positivos de A.
$P(E| + A \cap +B) = 0{,}7878$.

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

★ 52. 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 Discretas

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

54. 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) $\sigma = 1{,}6703$.

★ 55. 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.

---

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

---

57. Diez individuos entran en contacto con un portador de tuberculosis. La probabilidad de que la enfermedad se contagie del portador a un sujeto cualquiera es 0,10.

a) ¿Qué probabilidad hay de que ninguno se contagie?
b) ¿Qué probabilidad hay de que al menos dos se contagien?
c) ¿Cuántos se espera que contraigan la enfermedad?

Solución

Llamando X al número de personas contagiadas.

a) $P(X=0)=0{,}3487$.
b) $P(X\geq 2)=0{,}2639$.
c) $\mu=1$.

58. 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 $X\sim B(1000,\,0{,}002)\approx P(2)$. $P(X\leq 5)=0{,}9834$.

59. La probabilidad de que al administrar una vacuna dé una determinada reacción es 0,001. Si se vacunan 2000 personas, ¿Cuál es la probabilidad de que aparezca alguna reacción adversa?

Solución

Llamando X al número de reacciones adversas, $P(X\geq 1)=0{,}8648$.

★ 60. 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 $P(Y\geq 3)=0{,}938$.

61. 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 $P(Y>1)=0{,}8009$.

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

★ 63. Se ha comprobado experimentalmente que una de cada 20 billones de células expuestas a un determinado tipo de radiación muta volviéndose cancerígena. Sabiendo que el cuerpo humano tiene aproximadamente 1 billón de células por kilogramo de tejido, calcular la probabilidad de que una persona de 60 kg expuesta a dicha radiación desarrolle cáncer. Si la radiación ha afectado a 3 personas de 60 kg, ¿cuál es la probabilidad de que desarrolle el cáncer más de una?

Solución

Llamando X al número de mutaciones, $P(X > 0) = 0{,}9502$.
Llamando Y al número de personas que desarrollan el cáncer, $P(Y \geq 1) = 0{,}9999$.

★ 64. Un test diagnóstico para una enfermedad devuelve un 1 % de resultados positivos, y sus valores predictivos positivo y negativo valen, respectivamente, 0,95 y 0,98. Se pide:

a) ¿Cuál es la prevalencia de la enfermedad?
b) ¿Cuánto valen la sensibilidad y la especificidad del test?
c) Si aplicamos el test a 12 individuos enfermos, ¿qué probabilidad hay de que se equivoque en alguno de ellos?
d) Si aplicamos el test a 12 individuos, ¿que probabilidad hay de que acierte en todos?

Solución

a) $P(E) = 0{,}0293$.
b) Sensibilidad $P(+|E) = 0{,}3242$ y especificidad $P(-|\bar{E}) = 0{,}9995$.
c) Llamando X al número de diagnósticos erróneos en 12 individuos enfermos, $P(X \geq 1) = 1$.
d) Llamando Y al número de diagnósticos acertados en 12 individuos, $P(X = 12) = 0{,}7818$.

★ 65. 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 $P(Y \geq 9) = 0{,}9997$.

★ 66. 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 $P(Z < 3) = 0{,}6769$.

67. 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 $P(X_{A\cup B} \geq 1) = 0{,}9550$.

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

5. Variables Aleatorias Continuas

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

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

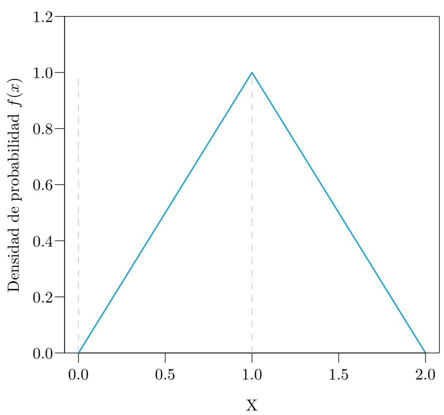
a) Determinar el valor de k.
b) Hallar $P(X \leq 1)$; $P(X > 2)$; $P(X = 1/4)$; $P(1/3 \leq X \leq 2/3)$.
c) Calcular μ y σ.
d) Hallar la función de distribución $F(x)$.

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)

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

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



Calcular:

a) $P(\frac{1}{2} \leq X \leq \frac{5}{4})$.
b) Función de distribución.
c) Media y desviación típica.

Solución

a) $P(\frac{1}{2} \leq X \leq \frac{5}{4}) = 19/32$.
b)

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

c) $\mu = 1$, $\sigma^2 = 1/6$ y $\sigma = 1/\sqrt{6}$.

★ 71. Se sabe que la duración de las baterías de un tipo de marcapasos (expresada en años) es una variable aleatoria cuya función de densidad es

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

Se pide:

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

b) Calcular la función de distribución.

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

d) 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.

---

72. 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.

★ 73. Sea Z una variable aleatoria continua que sigue un modelo de distribución normal estándar $N(0,1)$. Calcular las siguientes probabilidades utilizando la tabla de la función de distribución.

a) $P(Z < 1{,}24)$

b) $P(Z > -0{,}68)$

c) $P(-1{,}35 \leq Z \leq 0{,}44)$

SOLUCIÓN

a) $P(Z < 1{,}24) = 0{,}8925$.

b) $P(Z > -0{,}68) = 0{,}7517$.

c) $P(-1{,}35 \leq Z \leq 0{,}44) = 0{,}5815$.

★ 74. Sea Z una variable aleatoria continua con distribución normal estándar. Determinar el valor de x en los siguientes casos utilizando la tabla de la función de distribución.

a) $P(Z < x) = 0{,}6406$.

b) $P(Z > x) = 0{,}0606$.

c) $P(0 \leq Z \leq x) = 0{,}4783$.

d) $P(-1{,}5 \leq Z \leq x) = 0{,}2313$.

e) $P(-x \leq Z \leq x) = 0{,}5467$.

SOLUCIÓN

a) $x = 0{,}3601$.

b) $x = 1{,}5498$.

c) $x = 2{,}0198$.

d) $x = -0{,}5299$.

e) $x = 0{,}7499$.

★ 75. Sea X una variable aleatoria continua con distribución normal $N(10,2)$.

a) Calcular $P(X \leq 10)$.

b) Calcular $P(8 \leq X \leq 14)$.

c) Calcular el rango intercuartílico.

d) Calcular el tercer decil.

SOLUCIÓN

a) $P(X \leq 10) = 0{,}5$.

b) $P(8 \leq X \leq 14) = 0{,}8186$.

c) $RI = 2{,}698$.

d) $D_3 = 8{,}9512$.

---

76. 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.

77. Se sabe que el nivel de colesterol en varones de más de 30 años sigue una distribución normal, de media 220 mg/dl y desviación típica 30 mg/dl. Realizando un estudio sobre 20000 varones mayores de 30 años,

    a) ¿Cuántos se espera que tengan su nivel de colesterol entre 210 y 240 mg/dl?

    b) Si un nivel de colesterol por encima de 250 mg/dl puede provocar trombosis, ¿cuántos tienen riesgo de trombosis?

    c) ¿Cuál será el nivel de colesterol por encima del cual está el 20 % de la población?

Solución

a) $P(210 \leq X \leq 240) = 0{,}3781 \rightarrow 7561{,}3$ varones.

b) $P(X > 250) = 0{,}1587 \rightarrow 3173{,}1$ varones.

c) $P_{80} = 245{,}2486$ mg/dl.

★ 78. Un grupo de 100 alumnos realiza un examen. Una vez corregido, el profesor observa que la nota media es 4,2 y que 32 alumnos han tenido una calificación superior a 5. Suponiendo que las notas siguen una distribución normal, ¿cuántos alumnos habrán obtenido una calificación superior a 7?

Solución

$P(X > 7) = 0{,}0508 \rightarrow 5{,}1$ alumnos.

79. 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 $0{,}1151 \cdot 40000 = 4604$.

★ 80. 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.

★ 81. Un estudio trata de determinar el efecto de una dieta baja en grasas sobre el tiempo de vida de un tipo de ratas. Las ratas se dividieron en dos grupos, uno con una dieta normal y el otro con una dieta baja en grasas. Se supone que el tiempo de vida sigue un modelo de distribución normal con igual varianza pero con distintas medias. Si el 20 % de las ratas con dieta normal vivieron más de 12 meses, el 5 % vivieron menos de 8 meses y el 85 % de las ratas con dieta baja en grasas vivieron más de 11 meses,

a) ¿Cuál es la media y la desviación típica del tiempo de vida de las ratas con una dieta baja en grasas?

b) Si en el experimento había un 40 % de ratas con dieta normal y un 60 % de ratas con dieta baja en grasas, ¿cuál es la probabilidad de que una rata elegida al azar muera antes de 9 meses?

Solución

Llamando X al tiempo de vida de las ratas.

a) $\mu = 12{,}6673$ meses y $s = 1{,}6087$ meses.

b) $P(X < 9) = 0{,}068$.

★ 82. Un test diagnóstico para detectar el dopaje en atletas da un resultado positivo cuando la concentración de una determinada sustancia en sangre es mayor de 4 μg/ml. Si la distribución de la concentración de a sustancia en sangre en atletas dopados sigue un modelo de distribución normal con media 4,5 μg/ml y desviación típica 0,2 μg/ml, y en atletas no dopados sigue un modelo de distribución normal con media 3 μg/ml y desviación típica 0,3 μg/ml,

a) ¿cuál es la sensibilidad y la especificidad del test?

b) Si hay un 10 % de atelas dopados en una competición, ¿cuál es el valor predictivo positivo del test?

Solución

Llamando D al suceso aleatorio consistente en estar dopado, X a la concentración de la sustancia en atletas dopados e Y a la concentración de la sustancia en atletas no dopados.

a) Sensibilidad $P(+|D) = P(X > 4) = 0{,}9938$ y especificidad $P(-|\bar{D}) = P(Y < 4) = 0{,}9996$.

b) VPP $P(D|+) = 0{,}9961$.

★ 83. Según el teorema central del límite, se sabe que para muestras grandes ($n \geq 30$) la media muestral $\bar{x}$ sigue una distribución normal $N(\mu, \sigma/\sqrt{n})$, donde μ es la media de la población y σ su desviación típica.

Se sabe que en una población la elongación del triceps sural tiene una media de 60 cm y una desviación típica de 15 cm. Si se toma una muestra de tamaño 30, ¿cuál es la probabilidad de obtener una media muestral mayor de 62? Si se considera que una muestra es atípica si su media muestral está por debajo del percentil 5 de su distribución, ¿se puede considerar atípica una muestra de 60 individuos con $\bar{x} = 57$?

Solución

a) Llamando $\bar{X}$ a la variable que mide la elongación media del triceps sural en muestras de tamaño 30, $P(\bar{X} > 62) = 0{,}2327$.

b) Llamando $\bar{Y}$ a la variable que mide la elongación media del triceps sural en muestras de tamaño 60, $P_5 = 56{,}8$ cm, luego la muestra no es atípica.

6. Estimación de parámetros

84. 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

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

85. 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 $= (0{,}4024\text{ppm},\ 0{,}4776\text{ppm})$.

86. 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

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

87. 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.

---

88. 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 ($\sigma < 3$)?

89. 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: $n = 96$.

---

90. 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

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

★ 91. 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.

---

92. 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 $= (4{,}6143\text{mg}, 43{,}5675\text{mg/dl})$.

93. 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$)$.

★ 94. 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.

---

95. 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.

★ 96. 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.

★ 97. 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.