Capítulo 1: Estadstica descriptiva

1. Para los datos siguientes: 28, 22, 35, 42, 44, 53, 58, 41, 40, 32, 31, 38, 37, 61, 25, 35.
   (a) Calcula la mediana y los cuartiles.
   (b) Dibuja el box-plot
   (c) Construye un diagrama de tallo y hojas
   (d) Calcula la media y la desviación típica
   (e) Llamando x_i a los datos anteriores, calcula la media y la desviación típica para los datos y_i obtenidos como $y_i = 5 + 10x_i$.

2. A continuación se proporciona el consumo en *millas por galón* de 68 coches fabricados en Europa.

29.5	26.0	21.5	25.0	17.0	26.0	24.0	28.0	29.0	20.0
31.9	30.5	29.0	33.0	24.0	26.0	18.0	30.0	22.0	29.0
41.5	37.3	31.5	23.0	35.0	36.0	27.0	29.8	28.0	26.0
26.0	24.0	19.0	24.0	20.0	21.6	16.2	34.3	20.3	31.0
25.0	29.0	16.5	25.0	23.0	25.0	25.0	26.0	22.0	36.0
27.0	30.0	21.0	29.0	30.7	36.4	25.4	28.1	26.0	26.0
43.1	44.3	30.0	19.0	23.0	43.4	44.0	27.2		

   (a) Construye el histograma con 6 clases
   (b) Realiza el diagrama de tallos y hojas
   (c) Obtén la mediana y los cuartiles
   (d) Dibuja el box-plot. Indica las observaciones atípicas.
   (e) Transforma los datos con la función logarítmo: calcula la mediana y los cuartiles y dibuja el *box-plot* de los datos transformados.

3. En un departamento cuatro profesores imparten clases en grupos con 10, 18, 22 y 150 alumnos respectivamente. Si se pregunta a los profesores por el tamaño de su clase ¿cuál sería el valor medio y la desviación típica obtenida? ¿Y si se pregunta a todos los alumnos del departamento?

4. En la figura se presenta el diagrama de tallos y hojas de los residuos obtenidos de un diseño factorial. Representa el diagrama de caja (box plot) de los datos. (Nota.- La rama -6|91 representa los valores -0.69 y -0.61).

2	-6	|	91
2	-5	|	
4	-4	|	00
10	-3	|	766320
18	-2	|	98754310
29	-1	|	98654321100
(16)	-0	|	9977666554433211
36	0	|	015566677
27	1	|	2333478
20	2	|	134789
14	3	|	23455699
6	4	|	011355

5. Para n piezas se han medido dos dimensiones x_i e y_i. Es posible que la varianza de la variable x_i sea 4, la de y_i sea 9 y la de $z_i = x_i + y_i$ sea igual a 2? Justificar la respuesta.

6. Se toman 2 medidas x_i e y_i de n individuos. Dada dos constantes a y b positivas, demostrar que al multiplicar x_i por a e y_i por b, el coeficiente de correlación entre ambas no varía.

7. Demostrar que si entre dos variables x_i e y_i existe una relación lineal exacta $y_i = a + bx_i$, con $b > 0$, el coeficiente de correlación es uno.

8. Demostrar que el coeficiente de correlación es siempre en valor absoluto menor que uno.

9. En un proceso de fabricación se han medido tres variables y calculado la matriz de varianzas con el resultado siguiente:

$$\begin{pmatrix} 2 & 3 & 1 \\ 3 & 4 & 2 \\ 1 & 2 & 2 \end{pmatrix}$$

   (a) ¿Podemos afirmar que hay un error en los cálculos?

   (b) ¿Por qué?

10. A la variable x de media $\overline{x} = 100$ se le ha aplicado una transformación con el logaritmo decimal obteniéndose la nueva variable $y = \log_{10}(x)$. La media de la nueva variable es $\overline{y} = 2.5$. ¿Es posible este resultado?

Capítulo 2: Probabilidad Elemental

1. Sean A y B dos sucesos no disjuntos cualesquiera de un experimento aleatorio. Indicar cual de las siguientes probabilidades es mayor $P_1 = P(A \cap B \mid A)$ ó $P_2 = P(A \cap B \mid A \cup B)$. Justificar la respuesta.

2. Dado dos sucesos A y B, tal que $P(A) > 0$ y $P(B) > 0$ indicar, justificando la respuesta, si son ciertas o no las afirmaciones siguientes:

   a) Si A y B son sucesos mutuamente excluyentes entonces no pueden ser independientes.

   b) Si A y B son independientes entonces no pueden ser mutuamente excluyentes.

3. Demostrar que si dos sucesos A y B son independientes, también los son A y $\overline{B}$, y $\overline{A}$ y $\overline{B}$, donde $\overline{A}$ y $\overline{B}$ representan los sucesos complementarios de A y B respectivamente.

4. Sean dos sucesos equiprobables e independientes tal que la probabilidad de su unión es 0.75. Obtener la probabilidad de cada uno de ellos.

5. Se ha realizado una encuesta entre los estudiantes de grado del MIT (Massachusetts Institute of Technology) para conocer sus preferencias tecnológicas. El 35 % de los entrevistados tienen un iPhone y un IPad, el 80 % tienen al menos uno de estos dispositivos y el 60 % no tiene iPad. Se elige un estudiante al azar, calcula la probabilidad de que:

   a) Disponga de iPhone y no de iPad.

   b) Tenga un iPad pero no un Iphone.

   c) Tenga únicamente uno de los dos dispositivos.

   d) No Disponga de ninguno de los dos dispositivos.

6. Una encuesta realizada entre inmigrantes proporciona la siguiente información: el 80 % de los jóvenes entre 18 y 25 años no tiene trabajo, el 75 % de los jóvenes en esa edad no están matriculados en ningún centro educativo (no estudian). Si se toma un joven al azar de la muestra,¿cuál es la probabilidad de que ni estudie ni trabaje? (Otra forma de hacer la misma pregunta, ¿cuál es la proporción de jóvenes de la encuesta que ni estudian ni trabajan?

7. Demostrar que para cualquier par de sucesos, A_1, A_2 se cumple que

   $$P(A_1 \cap A_2) \geq P(A_1) + P(A_2) - 1,$$

   demostrar que la extensión a n sucesos es $P(A_1 \cap A_2 \cap \cdots \cap A_n) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$.

8. La probabilidad de que un componente de una máquina se averíe antes de 100 horas es 0.01. La máquina tiene 50 componentes; calcular la probabilidad de avería de la máquina antes de 100 horas en los casos siguientes:

   a) La máquina se avería cuando lo hace uno o más componentes.

   b) La máquina se avería cuando fallan dos o más componentes.

   c) La máquina sólo se avería cuando lo hacen todos los componentes.

9. Tres fichas están marcadas con las letras A, B, C, respectivamente, y una cuarta con las tres ABC. Se meten en una urna y se toma una de ellas al azar. Se pregunta si los tres sucesos consistentes en la presencia de la letra A, la letra B o la C sobre la ficha son o no independientes.

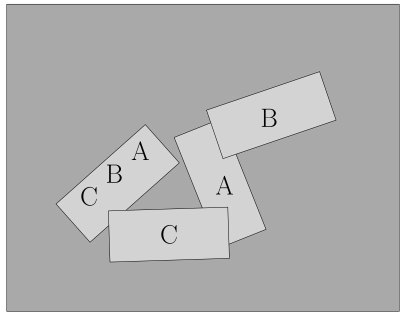


Figura 1: Cuatro fichas con letras A, B, C y ABC

10. En un campeonato de tenis usted tiene la opción de escoger la secuencia de partidos $A - B - A$ o la $B - A - B$, donde A y B indican sus oponentes. Para clasificarse debe usted ganar dos partidos consecutivos. El jugador A es mejor que el B. ¿ Qué secuencia será preferida?

11. Un jurado de tres miembros que decide por mayoría tiene dos miembros que deciden independientemente el veredicto correcto con probabilidad p y el tercero lanza una moneda. Si un juez tiene probabilidad p de acertar, ¿cuál de los dos sistemas tiene mayor probabilidad de acertar?

12. Una caja contiene r bolas rojas y b bolas blancas. Se elige una bola al azar de la caja (es decir todas las bolas tienen las misma probabilidad de ser elegidas), y después se extrae otra al azar entre las restantes de la caja. Calcular la probabilidad de

    a) Las dos bolas sean rojas
    b) Las dos bolas sean blancas
    c) La primera bola sea roja y la segunda blanca
    d) Las dos bolas sean de distinto color

13. Una caja contiene r bolas rojas y b bolas blancas. Se elige una bola al azar de la caja (es decir todas las bolas tienen las misma probabilidad de ser elegidas), después se devuelve a la caja, se mezcla y se extrae otra al azar. Calcular la probabilidad de

    a) Las dos bolas sean rojas
    b) Las dos bolas sean blancas
    c) La primera bola sea roja y la segunda blanca
    d) Las dos bolas sean de distinto color

14. Una urna contiene 3 bolas azules, 2 blancas y 5 rojas. Si se extraen 6 bolas con reposición, calcular la probabilidad de obtener 2 bolas de cada color.

15. El sorteo de la lotería primitiva consiste en acertar los seis números extraídos al azar y sin reposicin de una bombo que contiene 49 bolas numeradas del 1 al 49. Un participante ha realizado una apuesta múltiple y ha tachado 8 números. Calcular la probabilidad de acertar 6, 5, 4, 3, 2, 1 o ninguno de los números premiados. Si juega todos los días del año (365 sorteos) la misma apuesta ¿cuál es la probabilidad de que no acierte más de cuatro en ningún sorteo del año?

16. Si la probabilidad de que un interruptor cualquiera de la figura 2 se encuentre cerrado (es decir, pase corriente) es p, calcular la probabilidad de que pase corriente de A a B para las tres configuraciones de la figura (2).

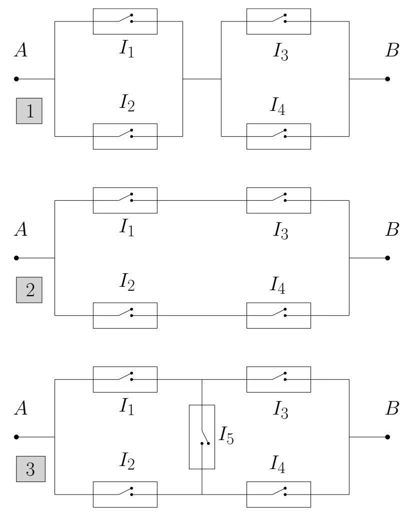
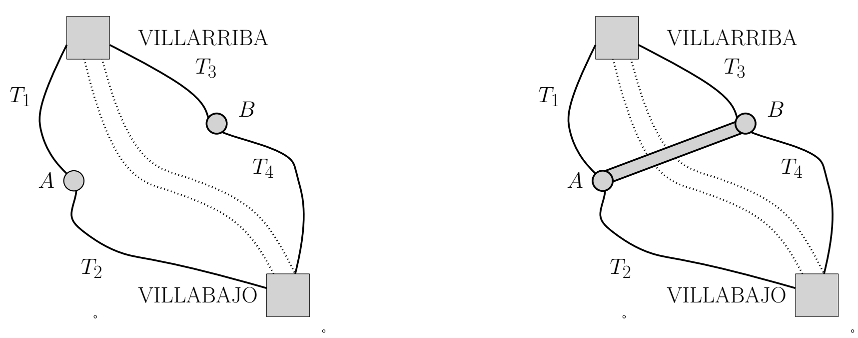


Figura 2: Tres circuitos

17. Un matrimonio tiene dos hijos y se sabe que uno de ellos es varón. Aceptando que la probabilidad de varón y hembra es la misma. ¿ Cuál es la probabilidad de que los dos sean varones?

18. Dos pueblos de alta montaña están comunicados por cuatro tramos de carretera según se muestra en la Figura 3a. Los días de invierno debido a la nieve es frecuente que las carreteras estén cortadas. Cuando nieva, la probabilidad de que cualquier tramo sea transitable es 0.8 con independencia de lo que ocurra con el resto. ¿Cuál es la probabilidad de que un vehículo pueda viajar de Villarriba a Villabajo un día con nieve? Para mejorar las comunicaciones se piensa construir un nuevo tramo de carretera entre los puntos A y B. Cuando nieva la probabilidad de accesibilidad es también 0.8 con independencia del resto. ¿Cuál es la probabilidad de que una persona pueda hacer el viaje anterior con esta nueva configuración 3b?



(a) Sin puente (b) Con puente

Figura 3: Red de carreteras entre Villarriba y Villabajo

19. Dos estudiantes A y B están matriculados en un curso. A asiste a las clases el 80 % de los días y B el 60 %, siendo las asistencias de ambos independientes. Si exactamente uno de los dos está en clase un día concreto, ¿cuál es la probabilidad de que sea A?

20. Cuatro estudiantes comparten piso, después de una comida deciden el siguiente procedimiento para elegir quién lava los platos: en un sombrero meten cuatro papeles doblados e indistinguibles, en

uno de ellos pone ‘Sí’ y en los otros tres ‘No’. El primero coge un papel y lo abre, si el resultado es ‘Sí’ le toca fregar, si el resultado es que ‘No’ se libra. No se devuelve el papel al sombrero. A continuacin el segundo coge el papel, y se aplica el mismo criterio, y as sucesivamente hasta que uno de ellos obtiene el papel con el ‘Sí’. ¿Quién tiene más probabilidad de lavar los platos, el primero o el último?

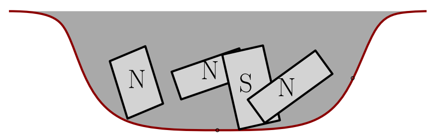


Figura 4: Sombrero con cuatro papeletas

21. Una comunidad de vecinos ha contratado un sistema de alarma para evitar robos en sus hogares. En caso de robo la alarma se pone en funcionamiento con seguridad. La probabilidad de que se produzca un robo en el vecindario es 0.001. Existen además otras causas (viento, golpes bruscos, ...) que ponen en funcionamiento el sistema de alarma, la probabilidad de que en una noche se produzca una falsa alarma es 0.01. Si se declara una señal de alarma, ¿cuál es la probabilidad de que sea falsa?

22. Una pareja que espera un hijo está preocupada porque un test practicado al feto ha dado positivo a una enfermedad muy rara que solo la padecen uno de cada 10000 individuos. Sin embargo, el test es bastante seguro: de acuerdo con el laboratorio acierta el 99 por ciento de los casos, tanto para los bebés con la enfermedad como para los sanos. ¿Cuál es la probabilidad de que el feto tenga la enfermedad teniendo en cuenta que el resultado del test ha sido positivo?

23. La probabilidad de que un componente se averíe en un período de tiempo dado es 0.01. Su estado (averiado o funcionando) se comprueba con un ensayo que cumple que cuando el componente funciona la probabilidad de que el ensayo diga lo contrario es 0.05, pero si el componente está averiado el ensayo no se equivoca. Si el ensayo indica que el componente está averiado, ¿cuál es la probabilidad de que realmente lo esté?

24. Un concursante debe elegir entre tres puertas, detrás de una de las cuales se encuentra un magnífico regalo. Hecha la elección, el presentador que sabe donde se encuentra el premio le abre una de las dos puertas no escogidas donde (lógicamente) no está el premio y a continuación le da al concursante la posibilidad de cambiar la puerta elegida. ¿Qué debe hacer el concursante?

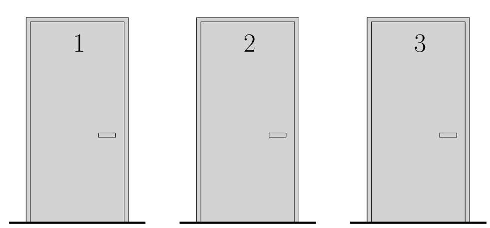


Figura 5: El concursante debe elegir una de las tres puertas

Capítulo 3: Variable Aleatoria

1. Dada la variable aleatoria discreta X, cuya función de probabilidad viene definida por

$$P(X = x) = kx, \quad x = 1, 2, ..., 5$$

   a) Calcular el valor de la constante k
   b) Calcular $P(X > 2)$
   c) Calcular $E[X]$ y $Var[X]$.
   d) Calcular $E[Y]$ si $Y = 2X + 5$

2. Dada la variable aleatoria X, cuya función de densidad es

$$f(x) = \begin{cases} k(1 - x^2), & \text{si } 0 < x < 1 \\ 0, & \text{en el resto} \end{cases}$$

   a) Obtener k.
   b) Dibujar la función de densidad.
   c) Calcular la probabilidad $P(X < 0.3)$.
   d) Obtener la media y la varianza de X.
   e) Obtener la media y la varianza de la variable $Y = 3X - 1$.
   f) Obtener la media y la varianza de la variable $Z = 3X^2$.

3. Un gran almacén guarda cajas que contienen piezas de distinto tipo. La proporción X de piezas de tipo A en una caja se puede considerar una variable aleatoria con función de densidad:

$$f(x) = kx(1 - x) \text{ con } 0 \leq x \leq 1$$

   a) Calcular el valor de k, la media y la varianza de la variable aleatoria X.
   b) Si se toman 10 cajas al azar.¿Cuál es la probabilidad de que ninguna de ellas contenga una proporción de piezas de tipo A igual o superior al 75 % ?

4. Un modelo que habitualmente se utiliza en balística para comprobar la correcta calibración de las armas es

$$f(x) = \frac{x}{\sigma^2} \exp\left[-\frac{x^2}{2\sigma^2}\right], \qquad x \geq 0, \sigma \geq 0,$$

   donde la variable aleatoria X es la distancia del punto de impacto del proyectil al centro del blanco al que iba dirigido y σ es el parámetro que mide la precisión. Si para una distancia determinada de disparo la precisión del arma es $\sigma = 10$ cm,

   a) ¿Cuál es la probabilidad de que un impacto esté a una distancia menor o igual de 5 cm del centro?
   b) Calcular la función de distribución.
   c) ¿Cuál es la probabilidad de que al lanzar 10 proyectiles, ninguno haya impactado a una distacia menor de 5 cm del centro del blanco?

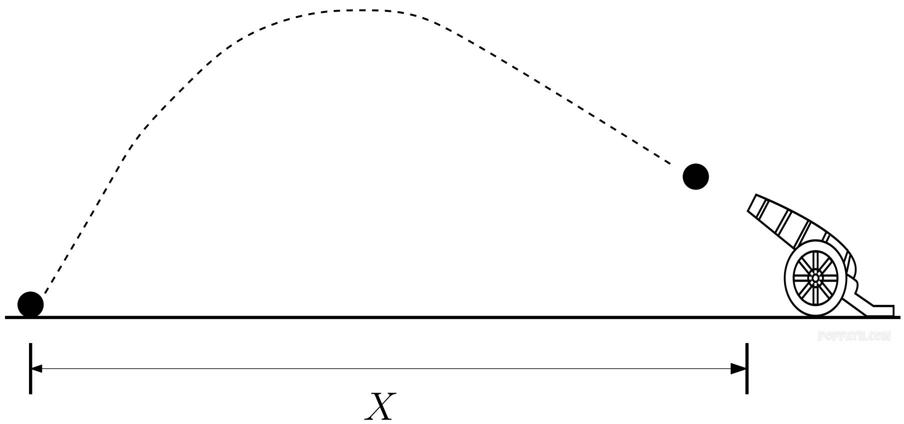


Figura 1: Disparo de un cañón

d) ¿Cuál es la probabilidad de que al lanzar 10 proyectiles, todos hayan impactado a una distacia menor de 5 cm del centro del blanco?

5. Una variable aleatoria tiene distribución uniforme en el intervalo $[-a, a]$. Calcular la media, la varianza, el coeficiente de asimetría y el coeficiente de apuntamiento (o curtosis).

6. Si X es una variable aleatoria con media μ. Demostrar que cuando $m = \mu$, $E[(X-m)^2]$ es mínima.

7. Demostrar la desigualdad de Markov

$$P(X \geq a) \leq \frac{1}{a} E\left[X\right]$$

donde X es una variable aleatoria positiva ($P(X > 0) = 1$).

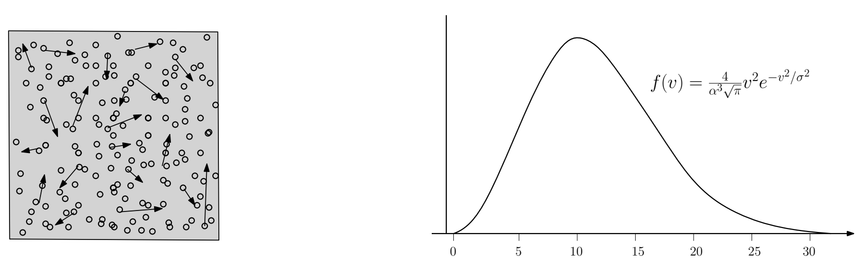


Figura 2: Distribución de Maxwell

8. De acuerdo con la teoría cinética de los gases, la velocidad V de una molécula de masa m de un gas a la temperatura (absoluta) T es una variable aleatoria con la siguiente función de densidad (distribución de Maxwell):

$$f(v) = \frac{4}{\alpha^3\sqrt{\pi}} v^2 e^{-v^2/\alpha^2}, \quad v \geq 0$$

donde $\alpha = \sqrt{2kT/m}$, siendo k la constante de Boltzmann. Además,

$$E(V) = 2\alpha/\sqrt{\pi}, \quad \mathrm{Var}(V) = (3/2 - 4/\pi)\alpha^2$$

a) Calcular el valor medio de la energía cinética, $mV^2/2$, de una molécula. ¿ A una misma temperatura T, qué gas tiene mayor valor medio de energía cinética, uno ligero u otro más pesado?

b) Obtener la función de densidad de la energía cinética de una molécula. Indicar si depende de la masa molecular.

9. Sea X una variable aleatoria con distribución uniforme en $(0, 1)$. Calcular la probabilidad de que $Y > 0.8$ si $Y = e^{-X^2}$.

10. La función de distribución de la variable aleatoria X es $F_X(x)$. Obtener la función de densidad de la variable aleatoria $Y = F_X(X)$.

11. La variable aleatoria Z tiene como función de densidad

$$f_Z(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{z^2}{2\sigma^2}\right), \quad -\infty < z < \infty, \quad \sigma > 0$$

Obtener la función de densidad de $Y = |Z|$ y su media.

12. Si X es una variable aleatoria con distribución uniforme entre 0 y θ, obtener la función de densidad de la variable aleatoria $Y = +\sqrt{X}$.

13. Se elige un punto al azar interior a la circunferencia de ecuación $x^2 + y^2 = r^2$. Llamando Z a la variable aleatoria definida por la distancia entre el punto elegido y el centro de la circunferencia, calcular las funciones de densidad y distribución de Z.

14. Supóngase una diana circular con centro en el origen de coordenadas de radio r y X, Y las coordenadas de un punto elegido al azar (por ejemplo, el lanzamiento de un dardo). Supóngase que cualquier otro punto de la diana tiene la misma probabilidad de ser elegido. Calcule $f_{XY}(x, y)$, las funciones de densidad marginales y condicionadas.

15. Obtén la distribución de probabilidad del máximo, del mínimo y de la suma de los resultados obtenidos al lanzar dos dados equilibrados. Se acepta que los resultados de los dados son variables aleatorias independientes.

16. La función de densidad de una variable aleatoria bidimensional viene dada por la expresión:

$$f_{XY}(x, y) = \begin{cases} xy + ce^x, & \text{cuando } 0 < x < 1 \text{ y } 0 < y < 1 \\ 0, & \text{en el resto} \end{cases}$$

Indicar si son independientes las variables aleatorias X e Y.

17. La cantidad en miligramos de dos componentes contenidos en un producto es una variable aleatoria bidemensional, cuya función de densidad viene dada por la expresión

$$f_{XY}(x, y) = \begin{cases} 4xy, & \text{cuando } 0 \leq x \leq 1 \text{ y } 0 \leq y \leq 1 \\ 0, & \text{en el resto} \end{cases}$$

a) Obtener las funciones de densidad marginales

b) Calcular la covarianza

c) Calcular $P(X + Y \leq 0.5)$

d) Calcular la probabilidad de que la cantidad del primer componentes sea menor que 0.3 miligramos cuando la del segundo es 0.8 miligramos.

18. La función de densidad de la variable aleatoria bidimensional (X, Y), bien dada por la expresión:

$$f_{XY}(x, y) = \begin{cases} kxy, & \text{cuando } 0 < x < y < 1 \\ 0, & \text{en el resto} \end{cases}$$

a) Calcular el valor de k y las distribuciones marginales.

b) Calcular $P(X < 0.5|Y = 0.5)$.

c) ¿Son independientes las variables aleatorias X e Y?

19. X e Y son variables aleatorias con coeficiente de correlación lineal $\rho = -1$. Si las varianzas son iguales, calcular la varianza de $Z = X + Y - 1$.

20. La distribución de probabilidad conjunta de las variables aleatorias Y_1 e Y_2 es la siguiente:

		Y_1		
		-1	0	1
	-1	1/16	3/16	1/16
Y_2	0	3/16	0	3/16
	1	1/16	3/16	1/16

Calcular su coeficiente de correlación e indicar si son independientes.

21. La función de densidad conjunta de X e Y viene dada por

$$f(x, y) = xy, \qquad 0 < x < 1, \;\; 0 < y < 2$$

a) Obtener las funciones de densidad marginales y decir si X e Y son independientes.

b) Calcular $\mathrm{P}(X + Y < 1)$.

c) Calcular $\mathrm{P}(X + Y < 2)$.

22. Un ordenador tarda un total de Y segundos en procesar un mensaje de correo electrónico, esta cantidad incluye el tiempo X durante el cual el mensaje está en la cola esperando a ser procesado $(Y \geq X)$. La función de densidad conjunta de las variables aleatorias X, Y es

$$f_{X,Y}(x, y) = e^{-y}, \quad 0 \leq x \leq y < \infty$$

Calcular la probabilidad de que un mensaje haya estado menos de un segundo en la cola si el tiempo total que ha durado su procesamiento ha sido mayor que dos segundos.

23. Sea X un valor elegido al azar de la distribución uniforme en el intervalo [0,1]. Se observa $X = x$ y a continuación se toma al azar otro valor Y de la distribución uniforme en el intervalo $[x, 1]$. Calcular la función de densidad marginal de Y.

24. Sean X, Y, U y V variables aleatorias, demostrar que si $Y = U + V$, entonces

$$\mathrm{Cov}(X, Y) = \mathrm{Cov}(X, U) + \mathrm{Cov}(X, V).$$

25. Sean dos variables aleatorias independientes X_1 y X_2 con la misma función de densidad $f(x) = e^{-x}, x \geq 0$

a) Obtener la función de densidad conjunta.

b) Obtener la función de densidad de $Y = X_1 + X_2$.

c) Añadimos otra variable X_3 independiente e idénticamente distribuida que las anteriores, obtener la función de densidad de $Y_3 = X_1 + X_2 + X_3$.

d) Dada n variables aleatorias independientes con la distribución anterior, se define $Y_n = X_1 + X_2 + \cdots + X_n$. Aplicando el procedimiento anterior de manera recurrente se obtiene que la función de densidad de Y_n es

$$f_{Y_n}(y) = \frac{y^{n-1}}{k(n)} e^{-y}, y \geq 0.$$

Obtener $E[Y_n]$ y deducir la expresión de la función $k(n)$.

Capítulo 4: Modelos de Probabilidad

1. Calcular la probabilidad de obtener al menos un seis al lanzar cuatro veces un dado.

2. Se lanza una pareja de dados 24 veces, calcular la probabilidad de por lo menos obtener una pareja de seises. (Este problema tiene interés histórico, fue resuelto por Pascal en el siglo XVIII para ayudar a un jugador llamado De Meré).

3. Una urna contiene 6 bolas, cuatro blancas y dos negras. Se extraen seis bolas con reposición, ¿Cuál es la probabilidad de que se obtengan 4 bolas blancas y dos negras? ¿Cuál es la probabilidad de que hayan extraído más bolas negras que blancas? ¿Cuál es la probabilidad de que hayan extraído más bolas blancas?

4. De un lote que contiene 200.000 piezas se extraen al azar 250. Si el número de piezas defectuosas en la muestra es mayor que c se rechaza el lote y se devuelve al proveedor. Calcular c si se desea que la probabilidad de rechazar un lote con $3\,\%$ de piezas defectuosas sea 0.05. (Nota: suponer que se toma la muestra con remplazamiento para hacer los cálculos).

5. Si las llamadas telefónicas a una centralita siguen una distribución de Poisson de parámetro $\lambda = 3$ llamadas/cinco minutos, calcular la probabilidad de:

    a) Seis llamadas en cinco minutos.
    b) Tres llamadas en diez minutos.
    c) Más de 15 en un cuarto de hora.
    d) Dos en un minuto.

6. La variable aleatoria X tiene distribución exponencial con media 1. Obtener la función de distribución y la función de densidad de

$$W = aX^{1/b}, \quad a > 0, b > 0$$

7. El número de averías diarias de una máquina sigue una distribución de Poisson de media 0.4 averías. Calcular la probabilidad de que haya tres días sucesivos sin averías.

8. A un puesto de servicio llegan de manera independiente, por término medio, 10 clientes/hora. Calcular la probabilidad de que lleguen 8 clientes en la próxima media hora sabiendo que en la última hora llegaron 14 clientes, y que la variable aleatoria *número de clientes que llegan en un hora* siguen una distribución de Poisson.

9. Sean X e Y dos variables independientes con distribución geométrica de parámetros 0.4 y 0.6 respectivamente. Calcular $P(X + Y = 3)$.

10. En una planta industrial (ver **??**) dos bombas B_1 y B_2 en paralelo conducen agua desde un pozo a una depuradora D, y posteriormente otras dos bombas B_3 y B_4, también en paralelo, la trasladan a un depósito como indica la figura.

    Los tiempos de vida de la depuradora y de las bombas son variables aleatorias independientes con distribución exponencial, siendo 20 mil horas la vida media de la depuradora y 30 mil horas la de cada bomba.

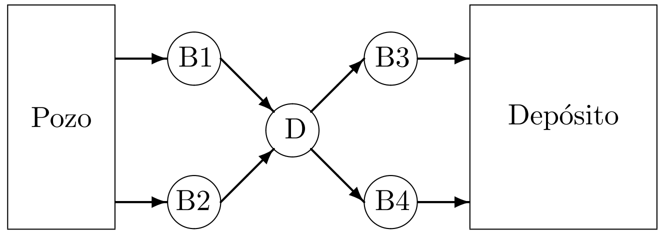


Figura 1: Planta industrial de depuración

a) Calcular la probabilidad de que llegue agua al depósito después de 20 mil horas de funcionamiento.

b) Calcular la probabilidad de que una depuradora que ha trabajado T horas falle antes de las mil horas siguientes. ¿Es razonable que para evitar fallos de la depuradora se renueve ésta cada 20 mil horas? ¿Por qué?

11. Un laboratorio de análisis realiza pruebas de sangre para detectar la presencia de un tipo de virus. Se sabe que una de cada 100 personas es portadora del virus. Se va a realizar un estudio en un colegio, para abaratar las pruebas se realiza un análisis combinado que consiste en: En lugar de analizar la sangre de cada individuo, se toman las muestras de 50 y se analiza la mezcla. Si el resultado del análisis es negativo, se concluye que los 50 individuos están sanos. Si el análisis es positivo, se repite a cada persona de manera individual. El análisis es infalible.

    a) Determinar el número esperado de pruebas (análisis) que se tendrá que realizar si se sigue este tipo de estrategia.

    b) ¿Cuál es la probabilidad de que un individuo determinado sea portador del virus, si el resultado del análisis realizado a su grupo de 50 ha resultado positivo?

12. De un lote con una proporción de piezas defectuosas p, se extraen piezas con reposición hasta que se observa la $k-$ésima defectuosa. Obtener la distribución de probabilidad de la variable aleatoria X *número total de piezas observadas.*

13. La función de densidad de una variable aleatoria X viene dada por la expresión

    $$f(x) = \begin{cases} x/8, & \text{si } 0 \leq x \leq 4 \\ 0, & \text{en el resto} \end{cases}$$

    Se generan secuencialmente valores de esta variable. ¿Cuántos valores de X habrá que generar por término medio hasta obtener un valor mayor que 3?

14. Una pareja decide tener hijos hasta el nacimiento de la primera niña. Calcular la probabilidad de que tengan más de 4 hijos. (Supóngase $P(niño) = P(niña) = 0.5$)

15. La distancia D entre dos vehículos consecutivos en una autopista sigue una distribución exponencial con media 200 metros. ¿Cuál es la probabilidad de que en un tramo de 1 km haya exactamente 5 vehículos?

16. Ricardo es un pescador experto que ha comprobado, después de una larga experiencia practicando su deporte favorito, que el número de peces capturados por la mañana puede ser representado por una variable aleatoria de Poisson de media 3 peces a la hora. Quiere ir a pescar el sábado próximo, si empieza a las 7 de la mañana, ¿cuál es la probabilidad de que capture el primer pez antes de las 7 h. 15 min.? ¿Cuál es la probabilidad de que capture 5 peces durante dos horas de pesca?

17. La variable aleatoria T representa la duración de vida de un componente electrónico. En teoría de la fiabilidad la probabilidad de que un componente falle en el instante t sabiendo que ha durado hasta t se denomina *tasa de fallo* y se representa por $\lambda(t)$, siendo su valor en función de t

$$\lambda(t) = \frac{f(t)}{1 - F(t)},$$

donde f y F son, respectivamente, las funciones de densidad y de distribución de la variable aleatoria T. Obtener la tasa de fallo en caso que T sea una variable aleatoria exponencial de media 1000 horas e interpretar el resultado.

18. Un examen consiste en 25 cuestiones. En cada cuestión, el alumno debe elegir entre 5 soluciones propuestas, de las que una (y sólo una) es cierta. El número mínimo de respuestas correctas que debe tener un alumno para aprobar es a. El profesor decide fijar a con el siguiente criterio: que la probabilidad de aprobar para un alumno que conteste todas las cuestiones al azar sea menor de 0.05. Obtener a. (Una cuestión es respondida al azar si cada uno de los cinco resultados propuestos tiene la misma probabilidad de ser escogido).

19. Obtener la función de densidad de una variable aleatoria χ^2 con un grado de libertad. (Si $X \rightsquigarrow N(0,1), Y = X^2$ es una χ_1^2.)

20. Dada una variable aleatoria X, cuya distribución es $N(0, \sigma^2)$, calcular la mediana de la variable $Y = |X|$.

21. La longitud L en milímetros de las piezas fabricadas en un proceso es una variable aleatoria que se distribuye según una $N(32, 0.3)$, considerándose aceptables aquellas cuya medida se encuentra dentro del intervalo (31.1, 32.6).

    a) Calcular la probabilidad de que una pieza elegida al azar sea aceptable.

    b) Si se toma al azar una muestra de tres piezas, ¿cuál es la probabilidad de que la primera y la tercera sean aceptables y la segunda no lo sea?

    c) ¿Cuál es la probabilidad de que en una muestra de tamaño 3 al menos una sea aceptable?

    d) Las piezas se embalan en lotes de 500. Calcular la probabilidad de que un lote tenga más de 15 defectuosas.

22. Un concesionario de automóviles recibe pedidos de un modelo según un proceso de Poisson de media 2 vehículos por semana. Los pedidos al fabricante se deben realizar con una antelación mínima de un mes, de forma que el concesionario pide en cada mes los vehículos que necesita para el mes siguiente. ¿Cuántos automóviles disponibles ha de tener a principios de un mes para satisfacer con probabilidad igual o mayor que 0.95 la demanda mensual? (Se considera que el mes tiene cuatro semanas).

23. Si la probabilidad de que un disparo impacte una diana es 0.0001, ¿cuál es la probabilidad de impactar en la diana 4 o más veces en 50000 disparos? Da un resultado numérico empleando la aproximación que consideres más adecuada. Se supone independencia.

24. En una urna hay 20 bolas, 5 son negras y 15 blancas. Se extraen 4 sin reposición y se define la variable aleatoria Y como el número de bolas negras.

a) Demostrar que la distribución de probabilidad de Y es

$$P(Y = k) = \frac{\binom{5}{k}\binom{15}{4-k}}{\binom{20}{4}}, \quad k = 0, 1, 2, 3, 4$$

b) Generalizar el resultado anterior si en la urna hay M bolas blancas, N bolas negras, se extraen K bolas e interesa conocer la distribución de probabilidad de Y número de bolas negras. La extracción es sin reposición. (Esta distribución se denomina *Hipergeométrica* y se utiliza en los problemas de muestreo sin reposición)

c) Plantear el problema del número de aciertos en la lotería primitiva como un caso particular de la distribución anterior. Aplicarlo para el caso en el que el participante tacha (elige) 6 números y en el caso en el que tacha 8 números.

25. Para controlar la calidad de un proceso textil se cuenta el número de defectos que aparecen en la tela fabricada. Según el fabricante, cuando el proceso funciona correctamente el número de defectos en una bobina de 100 metros cuadrados es una variable aleatoria de Poisson con media 4. Se ha instalado un equipo de visión artificial para realizar el recuento que permite inspeccionar 900 m^2 de tela cada hora. ¿Cuál es la probabilidad de que aparezcan más de 50 defectos en una hora si el proceso funciona bien?

26. Un compañía compra *chips* para montar en placas de ordenadores clónicos. Una empresa de reciclado le ofrece lotes de 10.000 chips a precios muy ventajosos pero con un porcentaje de defectuosos alto, alrededor del 10 %. Para realizar el control de calidad de los lotes recibidos está considerando dos alternativas: (a) Tomar 100 unidades al azar y rechazar el lote si existen más de 15 defectuosas. (b) Tomar 100 unidades al azar, dividirlas en 10 grupos y si en algún grupo hay más de una pieza defectuosa rechazar el lote.¿Cuál es la probabilidad de rechazar un lote con el 10 % de chips defectuosos para cada uno de los métodos? ¿Cuál es la probabilidad de aceptar un lote con el 20 % de chips defectuoso para cada método?

27. Una compañía para comprobar la calidad de ciertos lotes de 30000 piezas realiza el siguiente control: toma una muestra al azar de 300 piezas y si tiene 15 o más piezas defectuosas rechaza el lote, aceptándolo en caso contrario. La compañía cada mes aplica este control a 200 lotes, ¿cuál es el número esperado de lotes rechazados si todos los lotes de un mes tienen exáctamente un 4 % de piezas defectuosas?

28. Un servicio telefónico de urgencias recibe por término medio 10 llamadas cada minuto, ¿cuál es la probabilidad de recibir más de 550 llamadas en una hora? Se ha diseñado un "*call center*" con capacidad de respuesta de 650 llamadas a la hora, ¿cuál es el número esperado de horas al año con número de llamadas superior a su capacidad? (Se supone que las llamadas son independientes y todas las horas son similares)

29. A un congreso de medicina acuden 500 personas. Un laboratorio farmaceútico va a regalar corbatas a los hombres y pañuelos a las mujeres. Desgraciadamente no conocen el número exacto de cada sexo, aunque saben de otros congresos que la proporción es similar. Calcula el número mínimo de corbatas y de pañuelos que deben tener disponibles los organizadores para que todos los asistentes tengan el regalo que les corresponde con probabilidad de 0.99 (es decir ninguna mujer se quede sin pañuelo y ningún hombre sin corbata). Se supone que la probabilidad de hombre o mujer es igual a 0.5 y que la probabilidad de que un asistente sea de un determinado sexo es independiente del sexo de los restantes.

30. Federer y Nadal se encuentran empatados, 40-40 en un juego en el que está sacando Nadal. Según las estadísticas la probabilidad de que Nadal gane un punto determinado cuando tiene el saque es 0.6. ¿Cuál es la probabilidad de que el juego lo termine ganando Nadal? (Nota. Piensa en el desempate de la siguiente forma: se juegan dos puntos: si los gana un jugador ese jugador ha ganado el juego, si cada jugador gana un punto se juegan otros dos puntos y se vuelve a aplicar la misma regla).

31. *"Cibeles in Concert"* es una empresa que organiza viajes en autobus para asistir a actuaciones musicales. Para un concierto de *Bruce Springsteen* en Paris ha ofertado 300 plazas que salen de Madrid y Sevilla. Las reservas se hacen por Internet, el precio del viaje para los de Sevilla es de 60 euros y para los de Madrid 50 euros . Las 300 plazas se cubren con seguridad. Si la probabilidad de que un asistente salga de Sevilla es 1/3 y de Madrid 2/3, y se acepta independencia entre las 300 reservas, calcula los ingresos esperados por la compañía y la varianza de estos ingresos.

32. Una empresa de celulosa tiene dos líneas para fabricar pasta de papel en planchas de 1m $\times$ 1m. Una medida de su calidad es la limpieza, que se mide en número de impurezas (partículas) por m^2. La línea I, fabrica con una tasa media de 5 impurezas por m^2 y la línea II con 3 impurezas por m^2. El número de impurezas por plancha es una variable aleatoria con distribución de Poisson. Las planchas de pasta se empaquetan en balas de 2000 unidades y se almacenan clasificadas según la línea de procedencia. Cuando por algún motivo se encuentra en el almacén una bala sin clasificar se adopta el siguiente criterio: tomar una muestra aleatoria de 10 planchas y determinar el número medio de impurezas. Asignarla al grupo de la línea I si el número medio de impurezas es mayor que 4 y a la línea II en caso contrario. (Se supone que la probabilidad inicial de pertenecer a una u otra linea es la misma).

    a) Calcular la probabilidad de clasificar erróneamente una bala.

    b) En un caso concreto, el número de impurezas en cada una de las diez planchas han sido $1, 2, 3, 3, 4, 4, 5, 5, 6, 8$. Calcular la probabilidad de que pertenezca a cada una de las líneas.

33. Los billetes de banco son fabricados en pliegos. La impresión se realiza por dos máquinas iguales, una de ellas imprime el anverso y la otra el reverso. Sea X e Y, respectivamente, el número de defectos de impresión en el anverso y reverso de un pliego. Ambas variables son independientes con distribución de Poisson de parámetros λ_1 y λ_2.

    a) Demostrar que el número total de defectos en un pliego $Z = X + Y$ tiene distribución de Poisson. (Nota.- Utilizar que

    $$\mathrm{P}\{Z = n\} = \sum_{k=0}^{n} \mathrm{P}\{X = k\}\mathrm{P}\{Y = n - k\}$$

    y el desarrollo del binomio de Newton para $(\lambda_1 + \lambda_2)^n$.)

    b) Si el número total de defectos en un pliego es $Z = n$, ¿ cuál es la probabilidad de que haya exáctamente $X = k$ defectos en el anverso? (Obtener la expresión en función de λ_1, λ_2, n y k). ¿ De qué distribución de probabilidad se trata?

34. La llegada de los clientes a un banco se considera un proceso Poisson con parámetro λ. Sabiendo que en la última hora han llegado 2 clientes, ¿cuál es la probabilidad de que los dos entraran en los primeros 15 minutos?

35. Un equipo de radio tiene dos partes, el receptor y el amplificador. La duración del receptor es una variable aleatoria exponencial de media 500 horas y la duración del amplificador una variable exponencial de media 1000 horas. ¿Cuál es la probabilidad de que el fallo del equipo (cuando se produzca) sea debido a un fallo del receptor? (Se supone que las variables son independientes)

36. Sea X_1 una variable aleatoria N(10,1), X_2 una variable aleatoria N(20,1), y X_3 una variable aleatoria N(30,4). Se define

$$Z_1 = X_1 + X_2 - X_3$$
$$Z_2 = X_1 + X_2 + X_3$$
$$Z_3 = X_1 - X_2 - X_3$$

Si X_1, X_2, X_3 son independientes, calcular la matriz de varianzas de (Z_1, Z_2, Z_3).

37. Una oficina de correos tiene dos ventanillas de atención al público. Tres personas A,B y C llegan en el mismo instante a la oficina de correos y encuentran las dos ventanillas desocupadas. Los tiempos de servicio requeridos por las tres personas son variables aleatorias independientes con distribución exponencial de parámetro λ. Los tiempos de servicio de A y B comienzan de inmediato, mientras que C debe esperar a que termine el primero de los dos. ¿Cuál es la probabilidad de que C no sea el último en salir de la oficina de correos?

38. En cierta fabricación mecánica el 96 % de las piezas resultan con longitudes admisibles (dentro de tolerancias), un 3 % son piezas defectuosas cortas y un 1 % son defectuosas largas. Calcular la probabilidad de:

    a) En un lote de 250 piezas sean admisibles 242 o más.
    b) En un lote de 500 sean cortas 10 o menos.
    c) En 1000 piezas haya entre 6 y 12 largas. Todas las aproximaciones se calculan la distribución normal.

39. La estatura de los ciudadanos varones de un país sigue una distribución normal:

$$N(\mu = 175cm, \sigma = 5cm);$$

si se seleccionan al azar 100 ciudadanos, ¿cuál es la probabilidad de que al menos 30 superen los $180cm$?

40. En un proceso de fabricación de película fotográfica aparecen por término medio 1 defecto de cada 20 metros de película. Si la distribución de defectos es Poisson, calcular la probabilidad de 6 defectos en un rollo de 200 metros de pelicula (a) directamente; (b) utilizando la aproximación normal.

41. De un lote que contiene 200.000 piezas se extraen al azar 250. Si el número de piezas defectuosas en la muestra es mayor que c se rechaza el lote y se devuelve al proveedor. Calcular c si se desea que la probabilidad de rechazar un lote con 3 % de piezas defectuosas sea 0.05. (Nota: suponer que se toma la muestra con remplazamiento para hacer los cálculos y utilizar la aproximación normal para obtener c).

42. Para controlar la recepción de lotes de 10000 unidades de discos compactos se toma una muestra al azar de $n = 200$ discos clasificándolos como aceptables y defectuosos. Si el número de discos defectuosos es igual o inferior a $c = 15$ se acepta el lote, en caso contrario se rechaza. ¿Cuál es la probabilidad de aceptar un lote con el 12 % de discos defectuosos? ¿Cuál es la probabilidad de rechazar un lote con el 5 % de discos defectuosos? ¿Que valores de n y c deben utilizarse si se desea que las probabilidades anteriores sean iguales a 0.05?

43. La función de densidad conjunta de las variables aleatorias X, Y es la siguiente normal bidimensional:
$$f(x,y) = \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left(\frac{-1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$
Si $\rho = 0.3$, calcular $\Pr(Y \leq X + 1)$.

44. Un inversor tiene su dinero repartido en acciones de dos compañías. La rentabilidad (% de beneficio) de las compañías pueden ser consideradas como dos variables aleatorias, con media anual igual para las dos del 10 %, aunque el riesgo es muy diferente. Una tiene una desviación típica de 2.5 % y la otra del 1 %. Además se sabe que la correlación entre ellas es -0.50. ¿Qué proporción debe invertir en cada una para que el riesgo sea mínimo? (Nota: Mínimo riesgo es lo mismo que mínima varianza. $X \to N(10, 2.5)$ y $Y \to N(10, 1)$, $corr(X, Y) = -0.5$).

45. Cierto compuesto metálico se ha sometido durante una hora a una atmósfera de oxígeno a 200 grados centígrados. Una medida de su corrosión es la ganancia de peso durante este tiempo. Se ha comprobado que para un determinado compuesto esta ganancia X_1, en una hora, se distribuye segúun una normal $N(100, 5)$.

    a) Si se realizan los ensayos de manera secuencial, ¿cuántos se tendrán que hacer por término medio para encontrar una probeta con una ganancia mayor que 105?

    b) Se comprueba que sometiendo la probeta al ensayo durante dos horas, la ganancia en la segunda hora X_2 tiene una distribución normal $N(60, 5)$, siendo el coeficiente de correlación entre las variables X_1 y X_2, $\rho = -0.28$. Calcular la probabilidad de que una probeta tenga mayor ganancia de peso en la segunda hora que en la primera.

    c) Se forman los índices de oxidación $Z_1 = X_1 + X_2$ y $Z_1 = X_1 - X_2$. Calcular la función de densidad conjunta de estas dos nuevas variables. ¿Son independientes?

    d) Si la ganancia durante las dos horas ha sido 170, ¿cuál es el valor medio de la ganancia en la primera hora?

46. El abastecimiento de energía eléctrica de una comarca depende de tres centrales: una nuclear, una térmica de carbón y una hidráulica con potencias instaladas de 500 MW, 300 MW y 200 MW, respectivamente. Desde el punto de vista de fiabilidad, cada central sólo puede estar en uno de estos dos estados: disponible (con toda su potencia) o averiada (con potencia cero). En un día, la probabilidad de avería de la central nuclear es 0.10, de la térmica es 0.12 y de la hidráulica 0.05. Las averías son independientes y supondremos como hipótesis simplificadora que a lo largo de un día la central no cambia de estado.

    a) Para un día, calcular la función de distribución de la variable aleatoria X = “Potencia disponible en la comarca”.

b) Un pais tiene 30 centrales de cada uno de los tipos del apartado 1. Calcular la probabilidad de que en un día la potencia disponible en el pais sea menor que 24000 MW. (Utilizar la aproximación normal).

c) La potencia máxima diaria demandada en el pais es una variable aleatoria con distribución normal de media 23000 MW y desviación típica 1000 MW. ¿Cuál es la probabilidad de que en un día la demanda sea superior a la potencia disponible?

47. Una compañía desea aplicar un plan de muestreo para controlar la compra de lotes de 10000 unidades. La capacidad de inspección máxima que tienen es de 200 piezas. Determinar c, "*el número máximo de piezas defectuosas en la muestra que debe tener un lote aceptado*"si se desea que la probabilidad de rechazar un lote con el 10 % de piezas defectuosas sea igual 0.02. Un proveedor estima que sus lotes tienen un 5 % de piezas defectuosas, ¿qué probabilidad tiene de que un lote suyo sea aceptado?

48. Se tienen dos variables aleatorias independientes U y V, ambas con distribución uniforme en [0,1]. A partir de ellas se definen las variables aleatorias Z y W:

$$\begin{cases} Z = \sqrt{-2\log U} \\ W = 2\pi V \end{cases}$$

a) Calcular las probabilidades $P(Z < z_0)$ y $P(W < w_0)$. Utilizar estos resultados para deducir la expresión de la función de densidad de probabilidad conjunta $f_{Z,W}(z, w)$ para las variables aleatorias Z y W, así como el domino donde están definidas.

b) Una vez caracterizadas las variables aleatorias Z y W por su función de densidad de probabilidad conjunta $f_{Z,W}(z, w)$ se vuelve a definir 2 nuevas variables aleatorias X e Y mediante la siguiente transformación: $\begin{cases} X = Z \cdot \cos W \\ Y = Z \cdot senW \end{cases}$

Para estas nuevas variables se comprueba que la función de densidad de probabilidad conjunta es:

$$f_{X,Y}(x, y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left(x^2 + y^2\right)\right), \text{ con } (x, y) \in \mathbb{R}^2.$$

Demostrar que las variables aleatorias X e Y son independientes. Calcular la media y la varianza de las variables y sus funciones de densidad marginal identificando el tipo de variables de las que se trata.

c) Con la información del apartado (a) y (b) obtener el valor numérico de las siguientes probabilidades:

- $P(-1 < X < 1, -1 < Y < 1)$
- $P(X^2 + Y^2 < 1)$
- $P(Y < X)$

d) Finalmente, calcular los coeficientes de la siguiente tansformación lineal

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} c & 0 \\ d & e \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

con $c, e > 0$, para que las nuevas variables aleatorias $(X', Y')^T$ tengan una distribución de probabilidad con vector de medias y matriz de varianzas dadas por:

$$\mu = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \qquad S = \begin{pmatrix} 4 & 2 \\ 2 & 5 \end{pmatrix}.$$

49. Hay un juego muy popular en EEUU que se juega en carnavales en los que un jugador paga 1 dólar, elige un número del 1 al 6 y lanza tres dados. La banca le paga tantos dólares como veces aparece el número elegido. En un día de carnaval participan en el juego 500 personas. Obtén la distribución de probabilidad de las ganancias de la banca para ese día, indicando su media y desviación típica.

50. Se lanzan 100 dados. Obtén la distribución de la suma de los números pares. Calcula la media y la varianza.

Capítulo 5: Ejercicios de Inferencia

1. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria simple de una variable aleatoria exponencial con función de densidad, $f_X(x) = \frac{1}{\mu} e^{-x/\mu}, \quad x \geq 0, \ \mu > 0;$. Obtener el estimador de μ por el método de los momentos y por el método de máxima verosimilitud.

2. La función de densidad de Rayleigh de parámetro b es

$$f(x) = \frac{2x}{b^2} \exp\left(-\frac{x^2}{b^2}\right), \qquad x \geq 0, b \geq 0,$$

   a) Sabiendo que $E[X] = \frac{\sqrt{\pi}}{2} b$, estimar b por el método de los momentos.

   b) Estimar b por máxima verosimilitud.

3. La variable aleatoria X tiene distribución binomial con parámetros n y p, ambos desconocidos. Si {16,18,22,25,27} es una muestra aleatoria simple de la distribución anterior, estimar por el método de los momentos n y p.

4. Los taxis en servicio de una ciudad están numerados del 1 al N. Se observa una muestra de 5 taxis y se apuntan sus números. Obtener un estimador de N por el método de los momentos. Imagina que los números obtenidos son 7,23,34,36,84; a la vista de los resultados explica si es un buen estimador. (El ejercicio 5 sugiere un estimador más adecuado en un problema similar).

5. La variable aleatoria X sigue una distribución uniforme en el intervalo $[0, \theta]$. Dada una muestra aleatoria simple $x_1, x_2, ..., x_n$ se pide:

   a) Obtener el estimador de θ por el método de los momentos. Calcular su sesgo y su varianza

   b) Obtener el estimador máximo verosímil. Calcular su sesgo y su varianza.

   c) Compara el Error Cuadrático Medio de los dos estimadores.

   d) A partir del estimador máximo verosímil propón un estimador centrado, calcula su varianza y su ECM.

6. Una variable aleatoria discreta puede tomar los valores 0, 1 y 2 con probabilidades $1.5/\theta$, $2.5/\theta$ y $(\theta - 4)/\theta$ respectivamente, con $\theta > 4$. Se toma una muestra de tamaño 25 con los resultados siguientes (la segunda fila corresponde a la fracción observada O_i para 0, 1 y 2).

x	0	1	2
O_i	5	17	3

   Estimar θ por máxima verosimilitud.

7. Se ha tomado una muestra de tamaño 10 del tiempo, en minutos, entre el paso de dos autobuses T en una parada con los siguientes resultados: 9,10,6,4,15,6,1,5,4,10. Si la función de distribución del tiempo de paso es $F_T(t) = 1 - \exp(-\alpha t)$, estima el parámetro α por máxima verosimilitud y con este valor, la probabilidad estimada de esperar al autobús más de 8 minutos.

8. El club de tiro de una determinada ciudad está estudiando la distancia X del punto de impacto del proyectil al centro de la diana de sus 10 mejores tiradores.

   Sabiendo que la función de densidad de la variable aleatoria presentada es

$$f_X(x) = \frac{2x}{\theta^2} \exp[-\frac{x^2}{\theta^2}], \qquad x \geq 0, \theta \geq 0,$$

   Estimar θ si la distancia en cm al blanco de 10 tiradores fue

   2,1 3,2 6,3 5,4 2,2 6,9 7,1 6,6 2,5 9,1

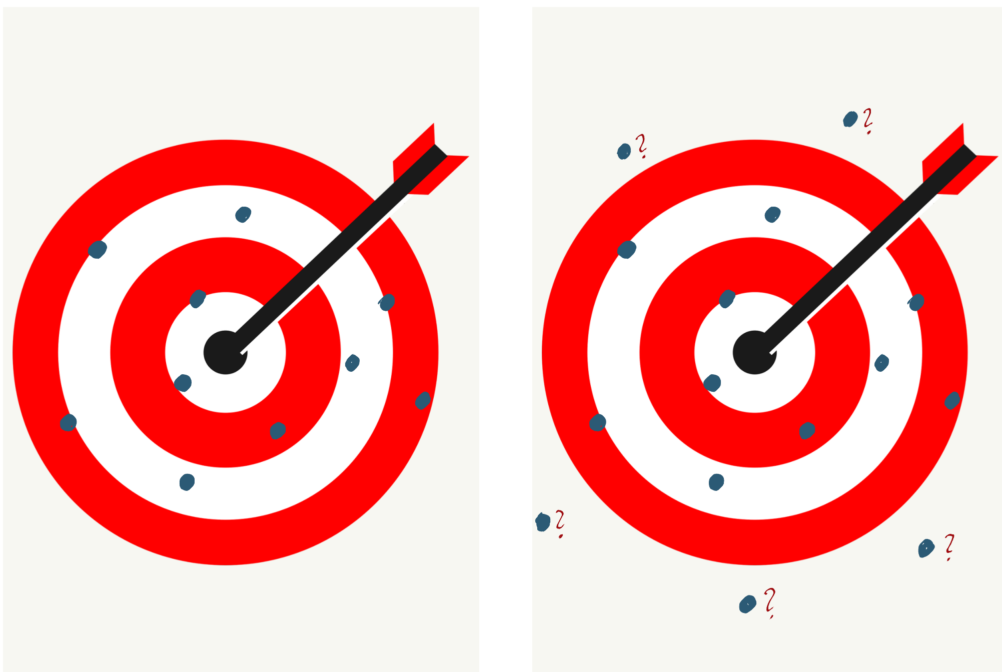


Figura 1: Resultados de disparos a una diana de los ejercicios 8 y 9

9. Repite el ejercicio 8 del club de tiro, suponiendo que en el test han participado 15 tiradores, 10 de ellos obtuvieron los valores indicados y los otros 5 no impactaron en la diana, por tanto sus valores son desconocidos y cumplen $x_i > 10$.

10. La velocidad de una molécula en un gas ideal según el modelo de Maxwell-Boltzmann es una variable aleatoria con función de densidad

$$f(x) = \frac{4}{\alpha^3\sqrt{\pi}}\, x^2\, e^{-(x/\alpha)^2}, \quad x \geq 0$$

   donde $\alpha > 0$, es un parámetro de la distribución. Teniendo en cuenta que

$$E[X] = \frac{2\alpha}{\sqrt{\pi}} \quad \text{y} \quad \text{Var}[X] = (\frac{3}{2} - \frac{4}{\pi})\alpha^2.$$

   a) Calcular el estimador máximo verosímil de α y obtén la varianza asíntótica del estimador.

b) Calcular el estimador por momentos de α y la varianza de dicho estimador.

11. La función de distribución de una variable aleatoria es

$$F_X(x) = \begin{cases} 0 & x < 0, \\ (x/\beta)^\alpha & 0 \le x \le \beta, \\ 1 & x > \beta. \end{cases}$$

donde los parámetros α y β son positivos.

a) Estimar los parámetros de la distribución por el método de máxima verosimilitud.

b) La longitud en milímetros de los huevos de una especie de pájaros se modelna con dicha distribución. Obtener los estimadores de α y β si se ha obtenido la siguiente muestra: 22.0, 23.9, 20.9, 23.8, 25.0, 24.0, 21.7, 23.8, 22.8, 23.1,23.1, 23.5, 23.0, 23.0

12. Los tiempos de funcionamiento de dos componentes electrónicos distintos siguen distribuciones exponenciales con esperanzas μ y 2μ. Se han obtenido los tiempos de fallo de una muestra de cada tipo de componente, en ambos casos de tamaño n. Obtener el estimador de máxima verosimilitud de μ, calcular su media y su varianza.

13. Una compañía, para determinar el número de consumidores de un determinado producto en Madrid, ha encuestado a personas elegidas al azar hasta encontrar a 20 que utilicen el producto. Estimar por máxima verosimilitud la proporción de consumidores en la ciudad si el número total de entrevistados ha sido 115

14. Sea $\overline{X}$ la media aritmética de una muestra aleatoria simple de una distribución N(μ, σ). Se define $\tilde{X} = c\overline{X}$ como nuevo estimador para μ. Determinar c (en función de μ y σ) para que el nuevo estimador tenga Error Cuadrático Medio (ECM) mínimo. Calcular c si se sabe que el coeficiente de variación $\sigma/\mu = 2$.

15. Para estimar la varianza σ^2 de una distribución normal $N(\mu, \sigma)$ se suelen utilizan tres estimadores

$$\hat{S}_1^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}, \quad \hat{S}_2^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n}, \quad \hat{S}_3^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n+1}$$

Compara los tres estimadores utilizando sesgo, varianza y ECM.

16. Para estimar la varianza σ^2 de una población normal se utiliza el estimador $\widehat{\sigma}^2 = k\widehat{S}^2$, siendo $\widehat{S}^2$ la varianza muestral corregida y k una constante. Calcular el valor de k que minimiza el error cuadrático medio. (Utilizar $Var[\chi_g^2] = 2g$, siendo g el número de grados de libertad).

17. Un sistema de lectura telemática de consumo de energía eléctrica emplea un mensaje de 128-bit. Ocasionalmente las interferencias aleatorias provocan que un bit se invierta produciéndose un error de transmisión. Se acepta que la probabilidad de que cada bit cambie en una transmisión es constante e igual a p, y que los cambios son independientes. Estima el valor de p si se ha comprobado que de las últimas 10000 lecturas efectuadas (todas de 128-bit) 340 eran erróneas.

18. Un estudiante ha realizado el siguiente ejercicio de simulación. Ha generado con R $n = 100$ observaciones de una normal de media 10 y desviación típica 1. Para estos 100 datos ha obtenido la mediana y la ha denominado M_1. Ha repetido el mismo proceso $N = 10000$ veces. De forma que ha obtenido $M_1, M_2, ..., M_N$. En el gráfico se muestra el histograma de estos 10000 valores, que tienen media 10.002 y desviación típica 0.1266.

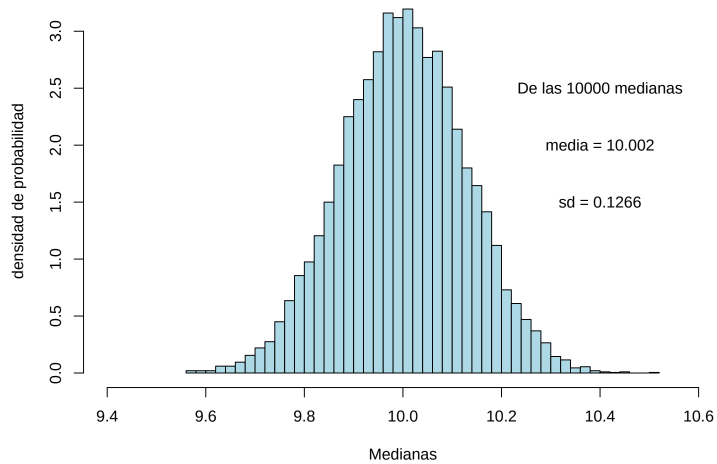


Figura 2: Histograma de las 10000 medianas de muestras de tamaño 100 de una distribución norma N(10,1)

A partir de esta simulación, explica qué propiedades tiene la mediana como estimador de la media μ de una distribución normal, sus ventajas e inconvenientes respecto a utilizar la media aritmética como estimador.

19. El tiempo que tarda un paciente en recuperarse de una dolencia es una variable aleatoria con distribución exponencial. Si al observar a 20 pacientes, al cabo de 20 días 15 no se han recuperado, mientras que los 5 restantes tardaron $17, 15, 19, 18$ y 17 días en hacerlo, estime por máxima verosimilitud el tiempo medio de recuperación.

20. El tiempo de duración de ciertos componentes electrónicos es una variable aleatoria con distribución exponencial. Se ha realizado un ensayo con 10 componentes cuyos tiempos de duración han sido: 37,45,92,104,109,200,295. Despues de 400 horas, tres componentes seguían funcionando. Con esta información, estimar por máxima verosimilitud el parámetro de la distribución exponencial.

21. Se han tomado 12 valores de una variable física X, que se supone sigue una distribución Normal, resultando

    30.2, 30.8, 29.3, 29, 30.9, 30.8, 29.7, 28.9, 30.5, 31.2, 31.3, 28.5.

    a) Construir un intervalo de confianza para la media de la población al 95 % de confianza.

    b) Construir un intervalo de confianza para la varianza de la población con el mismo nivel de confianza del apartado anterior.

22. Una muestra de 12 estaciones de servicio de una cadena de gasolineras proporciona un ingreso medio por persona al mes de 2340 euros con una desviación típica de 815 euros. Calcular un intervalo de confianza para el ingreso medio por trabajador en esta empresa. Calcular el número de estaciones que debemos estudiar para que el intervalo tenga una amplitud máxima de 500 euros.

23. Se han escogido al azar 15 probetas de un determinado acero, cuya resistencia a la compresión se supone que se distribuye normalmente, y se ha medido ésta en las unidades adecuadas, habiéndose observado los resultados siguientes

$$40.15,\ 65.10,\ 49.5,\ 22.4,\ 38.2,\ 60.4,\ 43.4,\ 26.35,\ 31.2,\ 55.6,\ 47.25,\ 73.2,\ 35.9,\ 45.25,\ 52.4$$

    a) Estimar la resistencia media del acero y su varianza.

    b) Hallar un intervalo de confianza del 99 % para la resistencia media.

    c) Hallar un intervalo de confianza del 99 % para la varianza.

    d) Cuántas probetas deberían haberse utilizado en el estudio si se quisiera estimar la resistencia media del acero con una precisión de ± 6 unidades y una confianza del 95 %?.

24. Una compañía de comida precocinada desea lanzar al mercado un nuevo producto. Para conocer la aceptación del mismo realiza previamente una encuesta entre 200 personas elegidas al azar, de las que 37 manifiestan su disposición a comprarlo.

    a) Obtener un intervalo de confianza ($\alpha = 0.05$) para la proporción p de compradores potenciales de este nuevo producto.

    b) ¿Cúal debería ser el tamaño muestral si se quisiera reducir la longitud del intervalo a la mitad.

25. Se desea estimar la proporción de niños entre 0 y 14 años que se encuentran adecuadamente vacunados contra la poliomielitis. Si se quiere que la diferencia en valor absoluto entre la estimación final y el verdadero valor de la proporción sea menor que 0.05 con probabilidad 0.95, ¿ Cúal es el tamaño muestral mínimo requerido?.

26. Una roca lunar es enviada a un laboratorio para determinar su nivel de radiactividad θ, nivel que se mide por el número medio de partículas emitidas por hora. Después de 15 horas, el equipo Geiger ha contabilizado un total de 3.547 partículas emitidas. Aceptando que el número de partículas emitidas sigue una distribución de Poisson, dar un intervalo con 95 % de confianza para el nivel de radiactividad de la roca. (Nota.- Utilizar que si Z tiene distribución N(0,1), entonces $P(Z \leq 1.96) = 0.975$).

27. Teniendo en cuenta que si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria simple de una variable aleatoria exponencial con función de densidad, $f(x) = \frac{1}{\lambda} e^{-x/\lambda}, \quad x \geq 0,\ \lambda > 0$; el estadístico $U = 2n\overline{X}/\lambda$ tiene distribución χ^2_{2n}, donde $\overline{X} = (X_1 + X_2 + \cdots + X_n)/n$; resolver la cuestión siguiente: El tiempo de funcionamiento de un equipo electrónico es una variable aleatoria con distribución exponencial. Se han tomado los tiempos de funcionamiento hasta el fallo de 30 equipos elegidos al azar, obteniéndose 6.2×10^3 horas de media. Calcular un intervalo con 95 % de confianza para la vida media de un equipo.

28. En una centralita ha habido 180 llamadas durante las últimas dos horas. Obtenga un intervalo de confianza para la esperanza del número de llamadas por hora suponiendo que el número de llamadas durante un periodo de duración T cualquiera sigue una distribución de Poisson.

Capítulo 6: Contraste de hipótesis

1. Un proceso industrial fabrica piezas cuya longitud en mm se distribuye según una distribución normal de media $\mu = 190$. Para controlar el proceso y se toman muestras de cinco piezas cada hora. En la última hora se han obtenido los siguientes valores:

   $$183.34,\ 211.68,\ 197.77,\ 205.89,\ 197.32$$

   La desviación típica de la distribución es conocida y supondremos que no cambia, $\sigma = 10$ mm. Contrasta la hipótesis de que la media del proceso μ es efectivamente 190 con $\alpha = 0.05$. Calcula el pvalor del contraste.

2. Repetir el ejercicio 1 con σ desconocida: Un proceso industrial fabrica piezas cuya longitud en mm se distribuye según una distribución normal de media $\mu = 190$. Para controlar el proceso se toman cada hora una muestra de cinco piezas. En la última hora se han obtenido los siguientes valores:

   $$183.34,\ 211.68,\ 197.77,\ 205.89,\ 197.32$$

   Contrasta la hipótesis de que la media del proceso μ es efectivamente 190 con $\alpha = 0.05$. Calcula el pvalor del contraste.

3. Un proceso de fabricación cuando está bajo control produce piezas cuya longitud tiene distribución normal de media $\mu = 190$ y varianza $\sigma^2 = 100$. Para controlar que la media no cambia, se toman muestras de tamaño $n = 5$ y se realiza el contraste de hipótesis:

   $$H_0 : \mu = 190$$

   $$H_1 : \mu \neq 190$$

   con $\alpha = 0.05$. Si se acepta H_0 el proceso continua fabricando. Si se rechaza H_0 el operario detiene el proceso y comprueba si todo está funcionando correctamente.

   a) ¿Cuál es la probabilidad de que si el proceso funciona correctamente, al tomar una muestra de tamaño 5, se rechace H_0 y se revise el proceso?
   b) ¿Cuál es la probabilidad de que cuando el proceso se desajuste y fabrique piezas con media $\mu = 195$, al tomar una muestra de tamaño 5 y realizar el contraste, se acepte H_0 y se concluya que la media del proceso no ha cambiado?
   c) Finalmente, se pide calcular la curva de potencia del contraste de media anterior. En todo el ejercicio suponer que la varianza es igual a 100 y no cambia.

4. Un proceso de fabricación en condiciones ideales produce piezas cuya longitud tiene distribución normal de media $\mu = 190$ y varianza $\sigma^2 = 100$. Para controlar que la varianza no cambia, se toman muestras de tamaño $n = 5$ y se realiza el contraste de hipótesis:

   $$H_0 : \sigma^2 = 100$$

$$H_1 : \sigma^2 \neq 100$$

con $\alpha = 0.05$.

a) Se ha tomado una muestra obteniendo los siguientes valores:

$$183.34,\ 211.68,\ 197.77,\ 205.89,\ 197.32$$

Realizar el contraste con $\alpha = 0.05$.

b) Calcula el pvalor del contraste en el caso de que la hipótesis alternativa sea (A) $H_1 : \sigma^2 > 100$, (B) $H_1 : \sigma^2 > 100$ o (C) $H_1 : \sigma^2 \neq 100$.

c) ¿Cuál es la probabilidad de que si el proceso funciona correctamente, al tomar una muestra de tamaño 5, se rechace H_0 ?

d) ¿Cuál es la probabilidad de que cuando el proceso se desajuste y $\sigma^2 = 200$, al tomar una muestra de tamaño 5 y realizar el contraste, se acepte H_0 y se concluya que la varianza del proceso no ha cambiado?

5. Para contrastar unilateralmente que la esperanza μ de una variable aleatoria normal es 10, se toma una muestra de tamaño 16 y se rechaza la hipótesis en el caso en que la media muestral sea mayor que 11, aceptándose en el caso contrario. Sabiendo que la desviación típica de la población es $\sigma = 2$, ¿cúal es la probabilidad de error de tipo I de este contraste?. ¿Cúal sería la probabilidad de error de tipo II del contraste si el valor verdadero de la esperanza fuese 12?.

6. Una medicina estándar es efectiva en el 75 % de los casos en los que se aplica. Se ha comprobado un nuevo medicamento en 100 pacientes, observándose su efectividad en 85 de ellos. ¿ Es la nueva medicina más efectiva que la estándar ? (Contrastar con $\alpha = 0.05$).

7. Un empresario quiere comprar una empresa que fabrica cojinetes. Durante los 5 últimos años la proporción de cojinetes defectuosos se ha mantenido en un 3 %. Para verificar esto, se toma una muestra de 200 cojinetes y obtiene que 9 son defectuosos.

   a) ¿ Se puede concluir que la proporción de cojinetes defectuosos ha aumentado?

   b) Calcular la potencia del contraste planteado anteriormente en función de p. Calcular la probabilidad de error de tipo II cuando la hipótesis alternativa es $p = 0.06$, siendo p la proporción de defectuosos.(Nota: Utilícese la aproximación normal y $\alpha = 0,05.$).

   c) Utilizando R calcula y dibuja las curvas de potencia para tamaños muestrales $n = 200$, $n = 300$, $n = 500$ y $n = 1000$. Interpreta el resultado.

8. Teniendo en cuenta que si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria simple de una variable aleatoria exponencial con función de densidad,

$$f(x) = \frac{1}{\mu} e^{-x/\mu}, \quad x \geq 0,\ \mu > 0;$$

el estadístico $U = 2n\overline{X}/\mu$ tiene distribución χ^2_{2n}, donde $\overline{X} = (X_1 + X_2 + \cdots + X_n)/n$; resolver las cuestiones siguientes:

a) El tiempo de funcionamiento de un equipo electrónico es una variable aleatoria con distribución exponencial. Se han tomado los tiempos de funcionamiento hasta el fallo de 30 equipos elegidos al azar, obteniéndose 6.2×10^3 horas de media. Contrastar con nivel de significación igual a 0.05, $H_0 : \mu = 5 \times 10^3$ horas, frente a $H_1 : \mu > 5 \times 10^3$ horas; indicando el pvalor.

b) Calcula la probabilidad de error tipo II cuando $\mu = 7 \times 10^3$ horas.

9. En la tabla se reproducen los valores para la velocidad de la luz en el aire que obtuvieron Michelson y Newcomb en 1879 y 1882 respectivamente, utilizando métodos distintos . A los datos originales en km/s se les ha restado 299000 para facilitar su tratamiento matemático.

	1	2	3	4	5	6	7	8	9	10
Mic.	850	740	900	1070	930	850	950	980	980	880
New.	883	816	778	796	682	711	611	599	1051	781

	11	12	13	14	15	$\overline{y}$	$\widehat{s}$
Mic.	1000	980	930	650	760	896.6	111.91
New.	578	796	774	820	772	763.2	119.47

a) Realizar el contraste de igualdad de medias utilizando el modelo de comparación de dos tratamientos con $\alpha = 0.05$.

b) Contrastar la hipótesis de que la variabilidad es la misma en los experimentos de Michelson y Newcomb. ($\alpha = 0.05$)

c) El valor utilizado actualmente para la velocidad de la luz en el vacío es 299.792,5 km/s (792,5 si se le resta 299.000). Michelson y Newcomb midieron en sus experimentos la velocidad de la luz en el aire. Contrastar para cada grupo por separado si la media de las observaciones es igual a 792.5 con $\alpha = 0.05$.

10. Se estudian dos tipos de neumáticos con los resultados siguientes:

Tipo	n_i	$\overline{y}_i(Km)$	$\widehat{s}_i(Km)$
A	121	27465	2500
B	121	27572	3000

a) Calcular un intervalo de confianza al 99 % para $\frac{\sigma_1^2}{\sigma_2^2}$.

b) Obtener un intervalo de confianza al 95 % para $\mu_1 - \mu_2$.

11. Se dispone de rendimientos de dos máquinas. Los resultados de la máquina A son 137.5; 140.7; 106.9; 175.1; 177.3; 120.4; 77.9 y 104.2, mientras que los resultados para la B son: 103.3; 121.7; 98.4; 161.5; 167.8 y 67.3. Suponiendo que los rendimientos de cada máquina tienen distribuciones normales, contrastar si las dos máquinas tienen el mismo rendimiento medio? Resolverlo con R, comprobando previamente la hipótesis de homocedasticidad (nivel de significación 0.05 para todos los contrastes).

12. Un fabricante de automóviles debe elegir entre un determinado tipo de piezas de acero suministradas por un proveedor A y otras suministradas por otro proveedor B. Para proceder a la elección se ha analizado la resistencia a la tracción de las piezas suministradas por ambos proveedores, tomando una muestra de tamaño 10 de las piezas del primero, y otra de tamaño 12 del segundo. La resistencia media de la muestra de A es de 54000 unidades y la de la muestra de B es de 49000 unidades, siendo las desviaciones típicas muestrales corregidas $\widehat{s}_A = 2100$ y $\widehat{s}_B = 1900$. Las resistencias de las piezas de ambos proveedores se distribuyen normalmente. Las piezas del proveedor B son más baratas que las del proveedor A, por lo que estas últimas sólo son rentables si tienen una resistencia media al menos 2000 unidades mayor que las de B, y la misma variabilidad.

    a) ¿A qué proveedor habría que comprar las piezas a la vista de los resultados muestrales?

    b) Obtener un intervalo de confianza del 90\ %

13. La estatura de 60 niños de una escuela infantil se resume en la siguiente tabla de frecuencias, dónde la última columna muestra la frecuencia esperada bajo la hipótesis de normalidad.

Intervalo	Frecuencia Observada	Frecuencia Esperada
41.5-43.5	4	4.08
43.5-45.5	7	5.58
45.5-47.5	12	9.06
47.5-49.5	8	11.27
49.5-51.5	6	11.27
51.5-53.5	11	9.08
53.5-55.5	9	5.58
55.5-57.5	3	4.08
Total	60	60

    ¿Se puede aceptar la hipótesis de normalidad de los datos ($\alpha = 0.05$) ?

14. Se tira 120 veces un dado y se obtienen los resultados de la tabla

VALOR	1	2	3	4	5	6
FRECUENCIA	20	14	23	12	26	25

    Contrastar la hipótesis de que el dado está equilibrado y que, por tanto, sus caras son equiprobables. (Tómese $\alpha = 0.05$).

15. Un modelo sísmico indica que la distribución de los epicentros de sismos en una región debería seguir una distribución de Poisson en el plano. Un grupo de expertos pretende

contrastar si ese modelo se cumple, para ello ha representado un mapa de la región dividido en cuadrículas de tamaño 100 km^2, y ha señalado con puntos las posiciones de los epicentros (véase figura adjunta). Realizar el contraste χ^2 de bondad de ajuste con nivel de significación $\alpha = 0,05$ proporcionando el pvalor del contraste.