ESCUELA POLITÉCNICA NACIONAL
DEPARTAMENTO DE FORMACIÓN BÁSICA
PROBABILIDAD Y ESTADÍSTICA • PRUEBA 1A
FECHA: 30 DE ABRIL 2019

EJERCICIOS:

1. En una vía de la ciudad se ha recopilado la información del número de pasajeros que transportan 18 autobuses escolares y se han obtenido los siguientes datos:

$$13, 15, 17, 15, 17, 16, 20, 18, 19, 18, 10, 19, 14, 16, 17, 16, 17, 14$$

a) Calcule las siguientes medidas: media aritmética, moda, mediana, desviación estándar y rango intercuartil. (1 punto)

Solución. **Media aritmética:**

$$\begin{aligned}\bar{x} &= \sum_{i=1}^{18} x_i \\ &= \frac{13+15+\ldots+17+14}{18} \\ &= 16{,}1667\end{aligned}$$

Moda:
Para su cálculo se identifica la observación con más repeticiones:

$$\text{Moda} = 17$$

Mediana:
Se ordenan los datos en forma ascendente:

$$10, 13, 14, 14, 15, 15, 16, 16, 16, 17, 17, 17, 17, 18, 18, 19, 19, 20$$

para su cálculo debe tomarse en cuenta que existe un número par de observaciones, por tanto:

$$\text{Mediana} = \frac{x_9 + x_{10}}{2} = \frac{16+17}{2} = 16{,}5$$

Desviación estándar:

$$\begin{aligned}s &= \sqrt{\frac{1}{18-1}\sum_{i=1}^{18}(x_i-\bar{x})^2} \\ &= \sqrt{\frac{(10-16{,}1667)^2+(13-16{,}1667)^2+\ldots+(19-16{,}1667)^2+(20-16{,}1667)^2}{17}} \\ &= 2{,}4314\end{aligned}$$

Rango Intercuartil:
Se inician calculando los percentiles 25 y 75:

$$\frac{nk}{100} = \frac{18(25)}{100} = 4{,}5 \approx 5, \qquad \frac{nk}{100} = \frac{18(75)}{100} = 13{,}5 \approx 14$$

de donde:

$$P_{25} = x_5 = 15, \qquad P_{75} = x_{14} = 18$$

además:

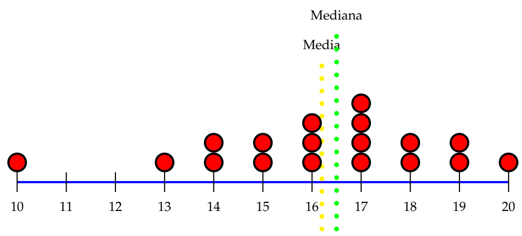
$$\text{Rango Intercuartil} = 18 - 15 = 3$$

□

b) Si el coeficiente de asimetría de la distribución es $-0{,}7299$ y el coeficiente de curtosis es 3,4997, interprete las características de la distribución de datos. (0.5 puntos)

Solución. Dado que el coeficiente de asimetría es menor a cero, la distribución de los datos será simétrica a la izquierda, esto implica que la cola izquierda de la distribución es más larga que la cola derecha. Por otro lado, un valor positivo de curtosis implica que la distribución de datos es más apuntada que la distribución normal, en tal virtud, las observaciones se encuentran muy concentradas en torno a su media y existe poca presencia de valores extremos. □

c) Represente los datos mediante un diagrama de puntos, en él ubique la media y la mediana, a continuación comente sobre la concentración de los datos. (0.5 puntos)



Solución. Los datos se encuentran concentrados alrededor de la media y mediana, pues ambas medidas de localización se encuentra cercanas. □

2. En una empresa de selección de personal aplicó un examen a 60 aspirantes a ocupar varios cargos en una agencia bancaria. Se anotó el tiempo (en minutos) que cada uno de los aspirantes se demoró en contestar el examen. Todos los tiempos se resumieron en la siguiente tabla de frecuencias. (Valor: 3 puntos)

Tiempo	**Punto medio** (x_i)	**Frecuencia absoluta** (n_i)	**Frecuencia relativa** (f_i)	**Frecuencia absoluta acumulada** (N_i)	**Frecuencia relativa acumulada** (F_i)
-	75		0.1		
-		14			
-					
-			0.2	52	
-	115				

a) Complete la tabla de frecuencias teniendo en cuenta que la amplitud de cada clase W es constante. (1 punto)

Solución.
Datos
n: Número total de personas
$f_1 = 0{,}1;\ \ f_4 = 0{,}2;\ \ n_2 = 14;\ N_4 = 52;\ \ F_1 = f_1 = 0{,}1;\ n = N_5 = 60;\ \ x_1 = 75;\ x_5 = 115;$
Planteamiento

$$f_4 = \frac{n_4}{n} \rightarrow n_4 = f_4 * n$$
$$n_4 = 0{,}2 * 60 = 12 \tag{1}$$

$$f_1 = \frac{n_1}{n} \rightarrow n_1 = f_1 * n$$
$$n_1 = 0{,}1 * 60 = 6 = N_2 \tag{2}$$

$$N_2 = n_1 + n_2 = 6 + 14$$
$$N_2 = 20 \tag{3}$$

$$f_2 = \frac{n_2}{n} = \frac{14}{60} = 0{,}23 \tag{4}$$

$$N_3 = N_4 - n_4 = 52 - 12 = 40 \tag{5}$$

$$N_2 = N_3 - n_3 \rightarrow n_3 = N_3 - N_2$$
$$n_3 = 40 - 20 = 20 \tag{6}$$

$$N_4 = N_5 - n_5 \rightarrow n_5 = N_5 - N_4$$
$$n_5 = 60 - 52 = 8 \tag{7}$$

$$f_3 = \frac{n_3}{n} = \frac{20}{60} = 0{,}33 \tag{8}$$

$$f_5 = \frac{n_5}{n} = \frac{8}{60} = 0{,}133 \tag{9}$$

Planteamiento de Ecuaciones para encontrar el ancho del intervalo

$$115 - 75 = 4W$$

$$W = 10 \tag{10}$$

$$\begin{aligned} x_2 &= x_1 + W \\ x_2 &= 75 + 10 = 85 \end{aligned} \tag{11}$$

$$\begin{aligned} L_1 &= x_1 - \frac{W}{2} \\ L_1 &= 75 - 5 = 70 \end{aligned} \tag{12}$$

$$\begin{aligned} S_1 &= x_1 + \frac{W}{2} \\ S_1 &= 75 + 5 = 80 \end{aligned} \tag{13}$$

TABLA DE FRECUENCIAS COMPLETA

Tiempo	Punto medio (x_i)	Frecuencia absoluta (n_i)	Frecuencia relativa (f_i)	Frecuencia absoluta acumulada (N_i)	Frecuencia relativa acumulada (F_i)
70-80	75	**6**	0.1	**6**	**0.1**
80-90	**85**	14	**0.23**	**20**	**0.33**
90-100	**95**	**20**	**0.33**	**40**	**0.66**
100-110	105	12	0.2	52	**0.86**
110-120	115	**8**	**0.13**	**60**	**1**

□

b) Indique el porcentaje de aspirantes que rindieron el examen en menos de 80 minutos. (0.5 puntos)

Demostración.

$$\text{Porcentaje} = \frac{6 \cdot 100\,\%}{60} = 10\,\% \tag{14}$$

□

c) Halle el promedio y la desviación estándar de la variable de Tiempo. (0.5 puntos)

Solución.
Media

$$\begin{aligned} \overline{x} &= \frac{\sum_{i=1}^{5} (n_i)(x_i)}{n} \\ \overline{x} &= \frac{(6)(75) + (14)(85) + (20)(95) + (12)(105) + (8)(115)}{60} \\ \overline{x} &= 95{,}33 \end{aligned} \tag{15}$$

Desviación estándar

$$\begin{aligned} \sigma &= \sqrt{\frac{\sum_{i=1}^{5} (n_i)(x_i)^2 - n\overline{x}^2}{n-1}} \\ \sigma &= \sqrt{\frac{553500 - (60)(95{,}33)^2}{59}} \\ \sigma &= 11{,}81 \end{aligned} \tag{16}$$

□

d) Dibuje el diagrama de caja, a partir de este gráfico responda lo siguiente: *i*) ¿En qué cuartil están más dispersos los datos?; *ii*) ¿Existen valores atípicos? (1 punto)

Solución.
Método Gráfico (Ojiva)

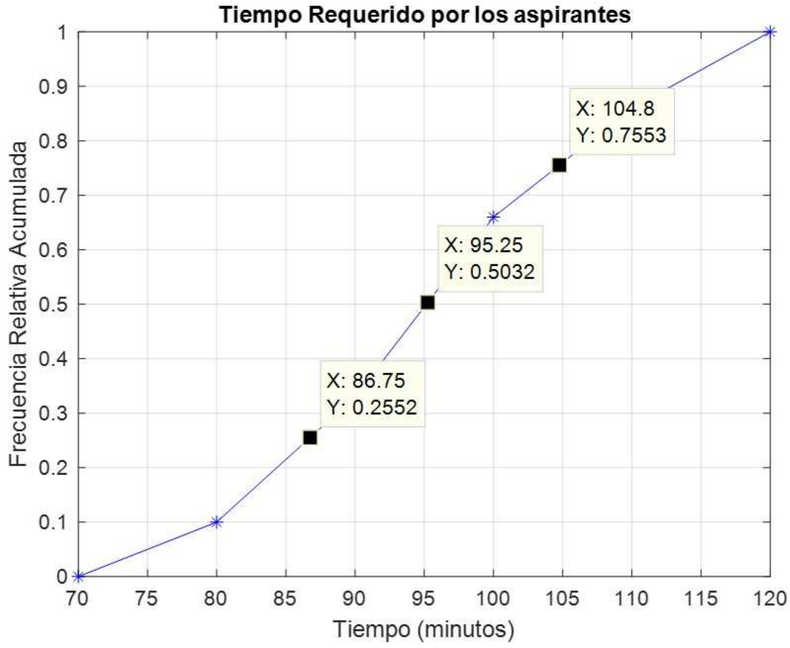


Figura 1: Ojiva.

Por lo que se tiene que:
$\tilde{x} = 95{,}25$
$Q_1 = 86{,}75$
$Q_3 = 104{,}8$
DIAGRAMA DE CAJA

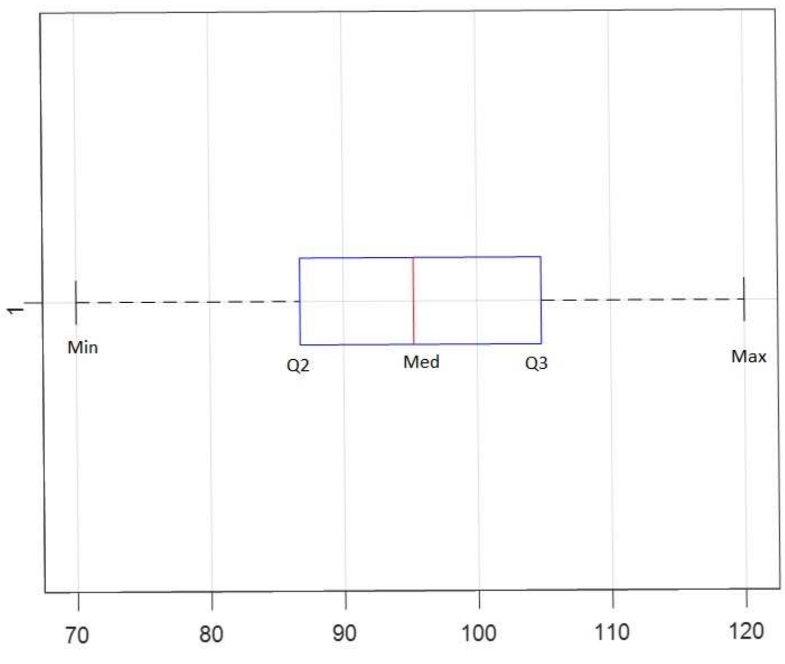


Figura 2: Diagrama de Caja.

Interpretación
La parte derecha de la caja es ligeramente mayor que la de la izquierda; ello quiere decir que los tiempos comprendidos entre el 50 % y el 75 % de la muestra están más dispersos que entre el 25 % y el 50 %. Además, ambos bigotes no presentan una gran diferencia entre ellos, esto implicaría la no existencia de datos atípicos. □

3. Una institución educativa de la ciudad cuenta con 1200 estudiantes: un 20 % pertenece a formación inicial, otro 45 % pertenece a formación básica y el resto a bachillerato. Además, se conoce que los estudiantes provienen de tres sectores: 400 estudiantes viven en el sector norte, 600 estudiantes viven en el sector centro y el resto viven en el sector sur. Por último, 210 estudiantes de formación básica viven en el sector norte, 320 estudiantes de bachillerato viven en el sector centro, 110

estudiantes de formación inicial viven en el sector sur y 70 estudiantes del bachillerato viven en el sector norte.

	Formación Inicial (I)	Formación Básica (B)	Bachillerato (A)	Total
Sector Norte (SN)				
Sector Centro (SC)				
Sector Sur (SS)				
Total				

Solución. A partir de la información proporcionada se obtiene que:

- El número de estudiantes de formación inicial es: $1200 \cdot 20\,\% = 240$.
- El número de estudiantes de formación básica es: $1200 \cdot 45\,\% = 540$.
- El número de estudiantes de bachillerato es: $1200 \cdot 35\,\% = 420$.

con lo cual se procede a completar la tabla:

	Formación Inicial (I)	Formación Básica (B)	Bachillerato (A)	Total
Sector Norte (SN)	120	210	70	400
Sector Centro (SC)	10	270	320	600
Sector Sur (SS)	110	60	30	200
Total	240	540	420	1200

□

Se escoge al azar a un estudiante de la institución educativa. Calcule la probabilidad de que dicho estudiante:

a) Sea de formación inicial o viva en el sector norte. (0.5 puntos)

Solución.

$$\begin{aligned} Pr(I \cup SN) &= Pr(I) + Pr(SN) - Pr(I \cap SN) \\ &= \frac{240}{1200} + \frac{400}{1200} - \frac{120}{1200} \\ &= \frac{520}{1200} \approx 0{,}4333 \end{aligned}$$

□

b) Sea de bachillerato y no viva en el sector sur. (0.5 puntos)

Solución.

$$\begin{aligned} Pr(A \cap SS^C) &= Pr(A) - Pr(A \cap SS) \\ &= \frac{420}{1200} - \frac{30}{1200} \\ &= \frac{390}{1200} \approx 0{,}3250 \end{aligned}$$

□

c) No viva en el sector centro o sea de formación básica. (0.5 puntos)

Solución.

$$\begin{aligned} Pr(SC^C \cup B) &= Pr(SC^C) + Pr(B) - Pr(SC^C \cap B) \\ &= 1 - Pr(SC) + Pr(B) - Pr(B) + Pr(SC \cap B) \\ &= 1 - Pr(SC) + Pr(SC \cap B) \\ &= 1 - \frac{600}{1200} + \frac{270}{1200} \\ &= \frac{870}{1200} \approx 0{,}7250 \end{aligned}$$

□

d) Viva en el sector sur y, sea de formación básica o bachillerato. (0.5 puntos)

Solución.

$$\begin{aligned}
Pr(SS \cap (B \cup A)) &= Pr((SS \cap B) \cup (SS \cap A)) \\
&= Pr(SS \cap B) + Pr(SS \cap A) - Pr((SS \cap B) \cap (SS \cap A)) \\
&= \frac{60}{1200} + \frac{30}{1200} - \frac{0}{1200} \\
&= \frac{90}{1200} \approx 0{,}0750
\end{aligned}$$

□

4. Se lanzan al aire dos dados distinguibles y se anotan los resultados de los dados.

a) Determinar el espacio muestral del experimento. (0.5 puntos)

b) Sean los sucesos A: "Salir al menos un 2", B: "Salir máximo un tres", determinar dichos sucesos como subconjuntos del espacio muestral y además encontrar:

1) $A \cup B$, (0.75 puntos)

2) $A^c \cap B^c$. (0.75 puntos)

c) Encontrar la probabilidad de que exactamente un dado haya salido 4. (0.5 puntos)

d) Encontrar la probabilidad de que haya salido un 4 y un múltiplo de 3. (0.5 puntos)

Solución. *a*) $S = \{(1,1), (1,2), (1,3), (1,4), \cdots, (5,6), (6,6)\}$, $|S| = 36$.

b) $A = \{(1,2), (2,1), (2,2), (2,3), (3,2) \cdots, (2,6)\}$, $|A| = 11$, $B^c = \{(3,3)\}$, $|B| = 35$

1) $A \cup B = B$, $|A \cup B| = 35$

2) $A^c \cap B^c = (A \cup B)^c = B^c = \{(1,4), (1,5), (1,6), \cdots, (6,6)\}$, $|A^c \cap B^c| = 1$.

c) $A = \{(1,4), (4,1), (2,4), (4,2), (4,5), (4,6)\}$, $P(A) = \dfrac{10}{36}$.

d) $A = \{(3,4), (4,3), (4,6), (6,4)\}$, $P(A) = \dfrac{4}{36}$.

□

EJERCICIOS:

1. El tratamiento de los niños con desórdenes de la conducta puede ser complejo. El tratamiento se puede proveer en una variedad de escenarios dependiendo de la severidad de los comportamientos. Además del reto que ofrece el tratamiento, se encuentran la falta de cooperación del niño/niña y el miedo y la falta de confianza de los adultos. Para poder diseñar un plan integral de tratamiento, el psiquiatra de niños y adolescentes puede utilizar la información del niño, la familia, los profesores y de otros especialistas médicos para entender las causas del desorden. Para ello, un psiquiatra local ha considerado una muestra aleatoria de 20 niños, anotando el tiempo necesario que requiere en cada niño para lograr un plan integral del tratamiento, obteniéndose lo siguiente (en horas):

6 7 7 8 8 8 8 9 9 9 9 9 9 9 10 10 10 10 10 11

a) Calcule las medidas de tendencia central y de dispersión de estos datos, indicando a qué tipo de medida pertenece. (1.5 pto.)

Solución. X: el tiempo necesario que requiere cada niño para lograr un plan integral del tratamiento (horas)
Medidas de tendencia central:

$$\begin{aligned}\bar{x} =& \frac{1}{n}\sum x_i \\ =& \frac{176}{20} \\ =& 8{,}8\end{aligned}$$

Calculo de la Mediana: En primera instancia se ordenan los datos

Posición	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Dato	6	7	7	8	8	8	8	9	9	9	9	9	9	9	10	10	10	10	10	11

Posición de la Mediana:

$$\frac{n+1}{2} = 10{,}5$$

por tanto, la mediana es el valor medio entre la décima y la undécima observación. En consecuencia,

$$\tilde{x} = 9$$

La **mediana** es igual a **9 horas**.
La **moda** es igual a **9 horas**, debido a que este es el valor que más se repite.
Medidas de dispersión: Desviación estándar:

$$\begin{aligned}s =& \sqrt{\frac{1}{n-1}\sum(x_i-\bar{x})^2} \\ &\vdots \\ =& 1{,}24\end{aligned}$$

La **desviación estándar** es de **1.24 horas**.
Se calcula el Rango

$$\begin{aligned}r =& x_{\text{máx}} - x_{\text{mín}} \\ =& 11-6 \\ =& 5\end{aligned}$$

horas. Se procede a calcular los cuartiles, para ello, se calcula P_{25}, P_{50} y P_{75}, a partir de:

$$P_{100p} = \begin{cases} \dfrac{x_{np}+x_{np+1}}{2} & \text{si } np \in \mathbb{Z} \\ x_{\lceil np \rceil} & \text{si } np \notin \mathbb{Z} \end{cases}$$

El cuartil 1 es la mediana de los primeros 10 datos, es decir, se encuentra entre la quinta y sexta observación:

$$Q_1 = 8$$

En consecuencia, **el cuartil 1** es **8 horas**.
El cuartil 3 es la mediana de los últimos 10 datos, es decir, se encuentra entre la 15ava y 16ava observación:

$$Q_3 = 10$$

El **cuartil 3** es 10 horas.
Rango entre cuartiles:

$$\begin{aligned} RIQ =& Q_3 - Q_1 \\ =& 10 - 8 \\ =& 2 \end{aligned}$$

El **rango intercuartil** es **2 horas**. □

b) Dibuje un diagrama de caja. Comente el resultado acerca de la distribución. (1.0 pto.)

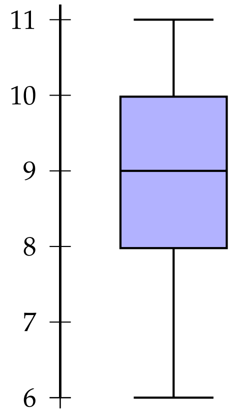
Solución. Para dibujar el gráfico de caja se necesita verificar si existen valores extremos:

$$\begin{aligned} x_i <& Q_1 - 1{,}5(Q_3 - Q_1) \\ <& 8 - 1{,}5(10 - 8) \\ <& 8 - 3 \\ <& 5 \end{aligned}$$

Si se toman el valor mínimo se tiene que 6 no es menor que 5, por lo tanto 6 no es un valor extremo. Ahora,

$$\begin{aligned} x_i <& Q_3 + 1{,}5(Q_3 - Q_1) \\ <& 10 + 1{,}5(10 - 8) \\ <& 10 + 3 \\ <& 13 \end{aligned}$$

se nota claramente que con el valor máximo se tiene que 11 no es mayor que 13, por lo tanto 11 no es un valor extremo. Distribución del tiempo necesario para diseñar un plan integral de un tratamiento que requiere un niño con desordenes de la conducta.



Se puede observar que debido a que el brazo izquierdo es ligeramente más alargado, la distribución de datos de los tiempos requeridos de tratamiento se hallan más dispersos entre 6 y 8 horas, mostrando un sesgo a la izquierda. □

2. Se supone que el diámetro de un cable eléctrico, digamos X es una variable aleatoria continua con función de densidad de probabilidad $f(x) = 6x(1-x), 0 \leq x \leq 1$.

a) Verificar que f es una función de densidad de probabilidad y graficarla. (0.5 pto)

Solución. Si esta es una función de densidad de probabilidad se tiene que verificar que:

- $f(x) \geq 0$: Dado que

$$f(x) = \begin{cases} 6x - 6x^2 & 0 \leq x \leq 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Se calcula el recorrido de la función:

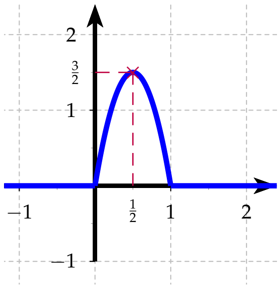
$$\begin{array}{ccc} & 0 \leq x \leq 1 & \\ 0 \leq x \leq \frac{1}{2} & \vee & \frac{1}{2} < x \leq 1 \\ -\frac{1}{2} \leq x - \frac{1}{2} \leq 0 & \vee & 0 < x - \frac{1}{2} \leq \frac{1}{2} \\ \frac{1}{4} \geq \left(x - \frac{1}{2}\right)^2 \geq 0 & \vee & 0 < \left(x - \frac{1}{2}\right)^2 - \frac{1}{4} \leq \frac{1}{4} \\ 0 \geq \left(x - \frac{1}{2}\right)^2 - \frac{1}{4} \geq -\frac{1}{4} & \vee & -\frac{1}{4} < \left(x - \frac{1}{2}\right)^2 - \frac{1}{4} \leq 0 \\ 0 \leq 6x(1-x) \leq \frac{3}{2} & \vee & \frac{3}{2} > 6x(1-x) \geq 0 \end{array}$$

De esta manera se observa que $f(x) \geq 0$

- $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

$$\begin{aligned}\int_0^1 6x - 6x^2\,dx =& 3x^2 - 2x^3\Big|_0^1\\ =& 3-2\\ =& 1\end{aligned}$$

Gráfico de la función



$\square$

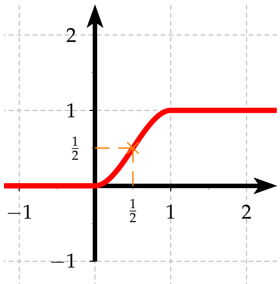
b) Obtener la función de distribución acumulada de probabilidad y graficarla. (0.5 pto)

Solución. Ahora para obtener la función de distribución acumulada se tiene que

$$\begin{aligned}F(x) &= \int_{-\infty}^{x} f(t)\,dt\\ &= \int_0^x 6t - 6t^2\,dt\\ &= 3t^2 - 2t^3\Big|_0^x\\ &= 3x^2 - 2x^3\end{aligned}$$

De donde

$$F(x) = \begin{cases} 0 & x < 0,\\ 3x^2 - 2x^3 & 0 \le x \le 1,\\ 1 & x > 1.\end{cases}$$



$\square$

c) Determinar un número b tal que $P(X < b) = 2P(X > b)$. (0.5 pto)

Solución. Se plantea la siguiente ecuación basado en la definición de la función de distribución acumulada

$$\begin{aligned}P(X < b) &= 2P(X > b)\\ &= 2(1 - P(X \le b))\\ &= 2(1 - P(X < b))\\ &= 2 - 2P(X < b)\\ 3P(X < b) &= 2\\ P(X < b) &= \frac{2}{3}\end{aligned}$$

$\square$

d) Calcular $P\left(X \le \frac{1}{2} \,\middle|\, \frac{1}{3} \le X < \frac{2}{3}\right)$. (1.0 pto.)

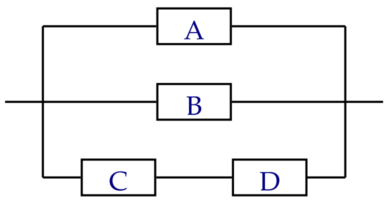
Solución. De la definición de probabilidad condicional se tiene:

$$\begin{aligned}
P\left(X \leq \frac{1}{2} \,\middle|\, \frac{1}{3} \leq X < \frac{2}{3}\right) =& \frac{P\left(X \leq \frac{1}{2} \wedge \frac{1}{3} \leq X < \frac{2}{3}\right)}{P\left(\frac{1}{3} \leq X < \frac{2}{3}\right)} \\
=& \frac{P\left(\frac{1}{3} \leq X < \frac{1}{2}\right)}{P\left(\frac{1}{3} \leq X < \frac{2}{3}\right)} \\
=& \frac{F\left(\frac{1}{2}\right) - F\left(\frac{1}{3}\right)}{F\left(\frac{2}{3}\right) - F\left(\frac{1}{3}\right)} \\
=& \frac{\frac{1}{2} - \frac{7}{27}}{\frac{20}{27} - \frac{7}{27}} \\
=& \frac{1}{2}
\end{aligned}$$

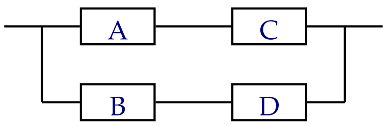
□

3. Se tienen 3 sistemas (I, II, III) con 4 dispositivos electrónicos interconectados que trabajan de manera independiente con una probabilidad de 0.7 cada uno. Calcule la probabilidad de funcionamiento de cada uno de estos tres sistemas e indique cuál es el menos confiable. (2 pto)

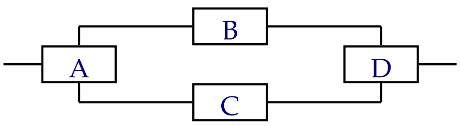
I)



II)



III)



Solución. Para el sistema I se tiene:

$$\begin{aligned}
P(I) =& P(A \cup B \cup (C \cap D)) \\
=& 1 - P\Big((A \cup B \cup (C \cap D))^c\Big) \\
=& 1 - P(A^c \cap B^c \cap (C^c \cup D^c)) \\
=& 1 - P(A^c)P(B^c)P(C^c \cup D^c) \\
=& 1 - P(A^c)P(B^c)\Big(P(C^c) + P(D^c) - P(C^c \cap D^c)\Big) \\
=& 1 - P(A^c)P(B^c)\Big(P(C^c) + P(D^c) - P(C^c)P(D^c)\Big) \\
=& 1 - (0{,}3)(0{,}3)(0{,}3 + 0{,}3 - (0{,}3)(0{,}3)) \\
=& 0{,}9541
\end{aligned}$$

Para el sistema II se tiene:

$$\begin{aligned}
P(II) =& P((A \cap C) \cup (B \cap D)) \\
=& P(A \cap C) + P(B \cap D) - P(A \cap C \cap B \cap D) \\
=& P(A)P(C) + P(B)P(D) - P(A)P(B)P(C)P(D) \\
=& (0{,}7)(0{,}7) + (0{,}7)(0{,}7) - (0{,}7)(0{,}7)(0{,}7)(0{,}7) \\
=& 0{,}7399
\end{aligned}$$

Para el sistema III se tiene:

$$\begin{aligned}
P(III) =& P(A \cap (B \cup C) \cap D)) \\
=& P(A)P(D)P(B \cup C) \\
=& P(A)P(D)\Big(P(B) + P(C) - P(B \cap C)\Big) \\
=& P(A)P(D)\Big(P(B) + P(C) - P(B)P(C)\Big) \\
=& (0{,}7)(0{,}7)\Big(0{,}7 + 0{,}7 - (0{,}7)(0{,}7)\Big) \\
=& 0{,}4459
\end{aligned}$$

De donde se tiene que el sistema menos confiable es: III □

4. Una entidad financiera, de acuerdo a estudios de riesgo para colocación de créditos, ha clasificado a los potenciales clientes en 5 grupos socio económicos: H_1, H_2, H_3, H_4, H_5 con una participación del 15 %, 25 %, 35 %, 15 % y 10 % respectivamente. De acuerdo al estudio, la posibilidad de **no** pago del crédito son: de 2 % para el grupo H_1, 3 % para el grupo H_2, 3 % para el grupo H_3, 5 % para el grupo H_4 y del 15 % para el grupo H_5.

   a) ¿Cuál es la probabilidad de que pague el crédito un cliente cualquiera? (1.5 pto.)

   Solución. Tomando al evento A como el no pago del crédito, por la regla de la probabilidad total, se tiene que:

$$
\begin{aligned}
P(A) =& \sum_{k=1}^{5} P(H_k)P(A|H_k) \\
=&P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + P(H_3)P(A|H_3) + P(H_4)P(A|H_4) + P(H_5)P(A|H_5) \\
=&(0{,}15)(0{,}02) + (0{,}25)(0{,}03) + (0{,}35)(0{,}03) + (0{,}15)(0{,}05) + (0{,}10)(0{,}15) \\
=&0{,}0435
\end{aligned}
$$

   La probabilidad de no pago es de 4.35 %. En consecuencia, la probabilidad de pago del crédito es de 95.65 % □

   b) Se toma al azar a un cliente y sabiendo que si pagó el crédito, ¿cuál es la probabilidad de que pertenezca al grupo socio económico H_3 (1.5 pto.)

   Solución. De la Regla de Bayes se tiene:

$$
\begin{aligned}
P(H_3|A^c) =& \frac{P(H_3 \cap A^c)}{P(A^c)} \\
=& \frac{P(A^c|H_3)P(H_3)}{P(A^c)} \\
=& \frac{(0{,}97)(0{,}35)}{0{,}9565} \\
=&0{,}3549
\end{aligned}
$$

   La probabilidad de que el cliente escogido al azar pertenezca al grupo H_3 es de 35.49 % □

EJERCICIOS:

1. Un andinista, por su experiencia, ha determinado que la probabilidad de escalar exitosamente el Cotopaxi es $\frac{1}{3}$. **(2.5 pto)**

   a) Cuál es la probabilidad de que el andinista llegue a la cumbre en su cuarto intento?.

   Demostración. Sea la variable aleatoria X: $\langle\langle$ número de intentos hasta alcanzar la cumbre $\rangle\rangle$.

$$X \sim G\left(\frac{1}{3}\right)$$

   por tanto,

$$\begin{aligned} P(X=4) &= \left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^3 \\ &= 0{,}09876 \end{aligned}$$

   □

   b) Cuántas excursiones esperará realizar para poder llegar a la cumbre?.

   Demostración.

$$\begin{aligned} E[X] &= \frac{1}{\frac{1}{3}} \\ &= 3 \end{aligned}$$

   □

   c) El próximo año el andinista tiene planificado ascender a 4 montañas que tienen la misma dificultad que el Cotopaxi. ¿Cuál es la probabilidad de que el andinista logre cumplir su objetivo en menos de 8 expediciones?.

   Demostración. Sea la variable aleatoria Y: $\langle\langle$ número de intentos hasta alcanzar la cumbre de 4 montañas$\rangle\rangle$.

$$Y \sim BN\left(4; \frac{1}{3}\right)$$

   por tanto,

$$\begin{aligned} P(Y<8) &= P(Y=4)+P(Y=5)+P(Y=6)+P(Y=7) \\ &= \binom{3}{3}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^0 + \binom{4}{3}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^1 + \binom{5}{3}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^2 + \binom{6}{3}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^3 \\ &= 0{,}173296 \end{aligned}$$

   □

   d) El andinista se impone la meta de que llegará al Cotopaxi por 6 ocasiones. ¿Cuántas expediciones tendrá que organizar para alcanzar la meta?.

   Demostración. Sea la variable aleatoria Z: $\langle\langle$ número de intentos hasta alcanzar la cumbre del Cotopaxi por 6 ocasiones $\rangle\rangle$.

$$Z \sim BN\left(6; \frac{1}{3}\right)$$

   por tanto,

$$\begin{aligned} E[Z] &= \frac{6}{\frac{1}{3}} \\ &= 18 \end{aligned}$$

   □

2. La vida útil de una torre de transmisión de potencia está distribuida exponencialmente, con una vida media de 25 años. Si se instalan tres torres al mismo tiempo las cuales operan independientemente: **(2.5 pto)**

   a) ¿Cuál es la probabilidad de que una torre trabaje correctamente (siga en pie) como máximo 35 años?

Demostración. Sea la variable aleatoria X: $\langle\langle$ vida útil de la torre $\rangle\rangle$
Datos $\theta = 25$

$$X \sim Exponencial$$
$$P(X < 35) = 1 - \exp\left(-\frac{35}{25}\right)$$
$$P(X < 35) = 0{,}7534$$

□

b) ¿Cuál es la probabilidad de que al menos 2 torres sigan en pie después de 35 años?

Demostración. **Datos** $\theta = 25$

$$P(X > 35) = 1 - P(X < 35)$$
$$P(X > 35) = 1 - 0{,}7534$$
$$P(X > 35) = 0{,}2466$$

Sea la variable aleatoria X_1: $\langle\langle$ número de torres que continuan en pie despúes de 35 años$\rangle\rangle$.
$n = 3$: Número total de torres
$p = 0{,}2466$: Probabilidad de que una torre siga en pie después de 35 años

$$X_1 \sim Bin(x_1; n, p)$$

por tanto,

$$\begin{aligned} P(X_1 \geq 2) &= P(x_1 = 2) + P(x_1 = 3) \\ &= \binom{3}{2}(0{,}2466)^2(0{,}7534)^1 + \binom{3}{3}(0{,}2466)^3(0{,}7534)^0 \\ &= 0{,}1524 \end{aligned}$$

□

3. Las puntuaciones en una prueba de Coeficiente Intelectual (CI) se distribuyen normalmente, con una media poblacional de 100. Se seleccionan $n = 20$ personas al azar, y se sabe que la desviación estándar en el grupo de muestra es 15. **(2.5 pto)**

a) ¿Cuál es la probabilidad de que el puntaje promedio de la prueba en el grupo de muestra se encuentre entre 95 y 108?

Demostración. Sea $\overline{X}$ la media de la muestra aleatoria de tamaño n, se tiene que

$$P(90 < \overline{X} < 110) = P(t_1 < t < t_2)$$

Sabiendo que $t = \dfrac{\overline{X} - \mu}{S/\sqrt{n}}$, se tiene que

$$t_1 = \frac{95 - 100}{15/\sqrt{20}} = -1{,}49$$
$$t_2 = \frac{108 - 100}{15/\sqrt{20}} = 2{,}38$$

por lo que,

$$P(t_1 < t < t_2) = P(-1{,}49 < t < 2{,}38)$$

Usando la propiedad de simetría de la distribución t-student, se sabe que

$$t_\alpha = -t_\alpha$$
$$1{,}49_\alpha = -1{,}49_\alpha$$

Ahora, usando la Tabla de t-student para $v = 20 - 1 = 19$ grados de libertad, se tiene que

$$1{,}49_\alpha \to \alpha \approx 0{,}080$$
$$2{,}38_\alpha \to \alpha \approx 0{,}015$$

por lo que,

$$\begin{aligned} P(-1{,}49 < t < 2{,}38) &= 1 - 0{,}015 - 0{,}080 \\ &= 0{,}905 \end{aligned}$$

□

b) ¿Cuál es el mínimo puntaje promedio que es necesario obtener en la prueba de CI para que una persona se encuentre en el 20 % de puntajes más altos?

Demostración. Usando la tabla de t-student para $v = 19$ grados de libertad, se tiene que

$$t_{0,20} = 0{,}8610$$

Ahora sabiendo que

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

despejando $\overline{X}$,

$$\overline{X} = t\left(\frac{S}{\sqrt{20}}\right) + \mu$$

reemplazando datos,

$$\overline{X} = (0{,}8610)\left(\frac{15}{\sqrt{20}}\right) + 100$$

$$\overline{X} = 102{,}88$$

□

4. El Examen Nacional para la Educación Superior (ENES) es un requisito establecido por el gobierno ecuatoriano para que los estudiantes de último año de bachillerato de las instituciones educativas del país accedan a la educación superior pública. Las puntuaciones (sobre 1000 puntos) de los bachilleres en el ENES siguen una distribución normal, con una media de 840 y una desviación estándar de 45. Determinar: **(2.5 pto)**

   a) ¿Cuál es la probabilidad de que un estudiante obtenga una nota superior a 940 puntos?

   b) Halle la probabilidad de que un estudiante obtenga una nota inferior a 800 puntos.

   c) Una universidad pública desea ofrecer una beca honorífica a aquellos estudiantes que obtengan puntuaciones que los coloquen en el 5 % más alto. ¿Cuál es la puntuación mínima que se requiere para obtener la beca?

   d) En cierta institución se sospecha que hubo fraude en la toma del examen, como medida para verificar este problema se tomaron los resultados de 40 estudiantes aleatoriamente y se obtuvo una media de 860 puntos. Los resultados obtenidos de esta muestra ayudan a afirmar esta sospecha o a refutarla?

Demostración. Sea $X :=$ "Nota de un estudiante seleccionado aleatoriamente"

a) $P(X > 940) = 1 - P(X \leq 940) = 1 - P\left(z \leq \dfrac{940 - 840}{45}\right) = 1 - P(z \leq 2{,}22) = 1 - 0{,}9868 = 0{,}0132$

b) $P(X < 800) = P\left(z < \dfrac{800 - 840}{45}\right) = P(z < -0{,}88) = 0{,}1894$

c) $P(X \geq t) = 0{,}05 \Longrightarrow 1 - P(X < t) = 0{,}05 \Longrightarrow P(X < t) = 0{,}95 \Longrightarrow P\left(z < \dfrac{t - 840}{45}\right) = 0{,}95 \Longrightarrow \dfrac{t - 840}{45} = 1{,}645 \Longrightarrow$ $t = 914{,}02$

d) Por el teorema del límite central tenemos que $\overline{X} \sim N(880, \frac{45^2}{40})$, entonces
$P(\overline{X} > 860) = 1 - P(\overline{X} \leq 860) = 1 - P\left(z \leq \dfrac{860 - 840}{45/\sqrt{40}}\right) = 1 - P(z \leq 2{,}81) = 1 - 0{,}9975 = 0{,}0025$
Así, la probabilidad de que el promedio sea mayor que 860 es casi nula por lo tanto se tendería a afirmar que en el colegio hubo fraude.

□

DEPARTAMENTO DE FORMACIÓN BÁSICA EPN

EJERCICIOS:

1. El tiempo de supervivencia (en minutos) de un microorganismo se puede modelar mediante una variable aleatoria con función de densidad

$$f(t) = \begin{cases} \frac{1}{40}e^{-\frac{t}{40}} & \text{si } t > 0, \\ 0 & \text{si } t \leq 0 \end{cases}$$

a) Determinar el tiempo de supervivencia media de la población de microorganismos? (0.5 puntos)

Solución. T: el tiempo de supervivencia(minutos) de un microorganismo.
$T \sim \varepsilon(\frac{1}{40})$

$$\begin{aligned} E(T) =& \frac{1}{\beta} \\ =& \beta \\ =& 40 \end{aligned}$$

□

b) Cuál es la probabilidad de que un microorganismo sobreviva por lo menos 50 minutos? (0.5 puntos).

Solución.

$$\begin{aligned} P(T \geq 50) =& 1 - P(T < 50) \\ =& 1 - F(50) \\ =& e^{-\frac{5}{4}} \\ =& 0{,}2865 \end{aligned}$$

□

c) Se seleccionan 125 microorganismos.¿Cuál es el número esperado de microorganismos que sobrevivirán por lo menos 50 minutos? (0.75 puntos).

Solución. Y: el número de microorganismos que sobreviven más de 50 minutos.
$Y \sim \beta(125, 0{,}2865)$

$$\begin{aligned} E(Y) =& np \\ =& 125(0{,}2865) \\ =& 35{,}8 \end{aligned}$$

□

d) Cuál es la probabilidad de que un microorganismo sobreviva solo 30 minutos, dado que ya ha sobrevivido 20 minutos? (0.75 puntos).

Solución.

$$\begin{aligned} P(T > 30 | T > 20) =& \frac{1 - F(30)}{1 - F(20)} \\ =& e^{-\frac{1}{4}} \\ =& 0{,}7788 \end{aligned}$$

□

2. La distribución de incendios forestales en el Distrito de Quito durante un periodo de verano por días de la semana se muestra en el cuadro siguiente. Construya un modelo de prueba de hipótesis para determinar con un 95 % de confianza si los incendios forestales ocurren o no de manera uniforme durante los días de la semana (2.5 puntos).

Días de la semana	Frecuencias observadas de incendios
Lunes	10
Martes	15
Miércoles	25
Jueves	20
Viernes	40
Sábado	55
Domingo	60

Solución. Se plantea un modelo de hipótesis con la prueba de χ^2 (chi-cuadrado)
Hipótesis nula H_0: La distribución de incendios sigue una ley uniforme
Hipótesis alternativa H_1: La distribución de incendios no sigue una ley uniforme

Se construye la frecuencia esperada

$$e_i = N \cdot P(X = i) = 225 \left(\frac{1}{7}\right) = 32{,}1$$

Días de la semana	F_O	F_E
Lunes	10	32.1
Martes	15	32.1
Miércoles	25	32.1
Jueves	20	32.1
Viernes	40	32.1
Sábado	55	32.1
Domingo	60	32.1

Calculamos el estadístico chi-cuadrado

$$\chi^2 = (10 - 32{,}1)2/32{,}1 + \ldots + (60 - 32{,}1)2/32{,}1 = 72{,}89$$

Se compara este valor con el cuantil de las tablas con 6 grados de libertad y 95 % confianza que es: $\chi^2_\alpha = 12{,}60$

Como $\chi^2 > X^2_\alpha$ se rechaza la hipótesis nula y por lo tanto se establece que los incendios forestales no ocurren de manera uniforme durante los días de la semana.

□

3. Una central lechera compra leche a varios proveedores. La central sospecha que algunos ganaderos añaden agua a la leche para aumentar sus beneficios. El exceso de agua se puede detectar midiendo el punto de congelación de la leche.La temperatura de congelación de la leche natural varía normalmente, con una media de $-0{,}545^oC$ y una desviación estándar de $0{,}008^oC$. La adición de agua aumenta la temperatura de congelación y la acerca a 0^oC (punto de congelación del agua).
El director del laboratorio de la central lechera determina la temperatura de congelación de 15 lotes consecutivos de leche procedentes de un mismo proveedor y encuentra que la media es igual a $-0,538^oC$. ¿Existe evidencia de que el proveedor está añadiendo agua a la leche? (2.5 puntos).

Solución. X: temperatura de congelación de la leche natural.

Datos: $\mu = -0{,}545^oC; \sigma = 0{,}008^oC; n = 15; \bar{x} = -0{,}538$

Planteando la prueba de hipótesis para la media de la temperatura de congelamiento de la leche natural,

$$H_0 : \mu = -0{,}545 \tag{1}$$
$$H_a : \mu > -0{,}545 \tag{2}$$
$$\alpha = 0{,}05 \tag{3}$$
$$Z_{obs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \tag{4}$$
$$Z_{obs} = \frac{-0{,}538 + 10{,}545}{0{,}008/\sqrt{15}} = 3{,}38 \tag{5}$$
$$Z_{0{,}05} = 1{,}645. \tag{6}$$

Para $\alpha = 0{,}05$, la REGIÓN DE RECHAZO ES:$Z_{obs} > Z_{0{,}05}$.

Como $3{,}37 > 1{,}645$, se rechaza H_0, por tanto la temperatura de congelación de la leche en promedio es mayor que $-0{,}545^oC$, es decir, existe evidencia de que el proveedor está añadiendo agua a la leche. □

4. Unos laboratorios médicos disponen de una máquina para envases con suero fisiológico y está regulada para colocar una cantidad media de 1000 ml con una desviación estándar de 10 ml. De forma periódica, la máquina es revisada tomando una muestra de 20 envases y determinando el contenido medio. En el último control realizado se encontró que la media de los 20 envases fue de 985 ml y se concluyó que la máquina no necesitaba ajuste. ¿Fue una decisión acertada o incorrecta? Utilice un nivel de confianza del 95 %. (2.5 puntos).

Solución. La máquina tiene como parámetros (regulada así): $\mu = 1000$ ml, $\sigma = 10$ ml
Se toma una muestra de $n = 20$ envases, y se determina una media de 985 ml
Calculamos un intervalo de confianza para la media poblacional y verificamos si la media con que viene regulada la máquina pertenece o no este intervalo (con lo cual podemos concluir si fue una decisión acertada o no): Utilizamos (varianza poblacional conocida):

$$\bar{X} - z_{1-\frac{\alpha}{2}}\sqrt{V(\bar{X})} < \mu < \bar{X} + z_{1-\frac{\alpha}{2}}\sqrt{V(\bar{X})}$$

$$V(\bar{X}) = \frac{\sigma^2}{n}$$

Entonces:

$$985 - 1{,}96\frac{10}{\sqrt{20}} < \mu < 985 + 1{,}96\frac{10}{\sqrt{20}}$$

$$980{,}6 < \mu < 989{,}4$$

Dado que la media $\mu = 1000$ está fuera del intervalo, se concluye que la máquina si necesitaba de un ajuste, y por lo tanto fue una decisión incorrecta.

También es válido utilizar en lugar de Z, los Cuantiles t-student (en este caso con 19 grados de libertad y 95 % confianza) $t_\alpha = 2,093$, y en esta situación el intervalo de confianza sería:

$$980{,}3 < \mu < 989{,}7$$

otra manera de abordar esto es planteando como una prueba de medias:

$$t = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Donde $\bar{x} = 985$, $\mu_0 = 1000$, $\sigma = 10$, $n = 20$.

En este caso $t = -6{,}7$, y considerando $t_\alpha = 2{,}093$ (por simetría), se tiene que esta media difiere significativamente del parámetro de contraste y por lo tanto fue una decisión equivocada.

□