Examen Extraordinario de Estadística I, 22 de Junio de 2012.
Grados en ADE, DER-ADE, ADE-INF, FICO, ECO, ECO-DER.

NORMAS: 1) Entregar cada problema en un cuadernillo distinto, aunque esté en blanco.
2) Realizar los cálculos con al menos dos cifras decimales significativas.
3) No se podrá abandonar el examen hasta transcurridos 30 minutos depués de haber empezado.
4) No está permitido salir del aula sin entregar el examen, aunque esté en blanco.

1. Con el objetivo de estudiar la temperatura del termostato de refrigeración de un cierto modelo de coche a los 100 km/h, se toma una muestra de treinta coches de dicho modelo y se mide la temperatura del termostato de cada coche a dicha velocidad. Los resultados obtenidos son los siguientes (medidos en grados Celsius):

65.8	69.4	69.4	69.7	71.5	72.2	74.1	75.4	75.8	76.3
77.2	77.6	77.6	77.9	78.3	78.7	78.9	81.2	81.2	81.7
82.3	82.3	82.4	84.5	84.7	85.2	85.4	88.2	102.5	105.5

(a) **(0.5 puntos)** Agrupa la muestra en intervalos de amplitud constante empezando por $[65, 69)$ y calcula la tabla de frecuencias absolutas y relativas.

Solución:

La tabla de frecuencias absolutas y relativas para los intervalos solicitados es:

Intervalo	Frec. abs.	Frec. rel.
$[65, 69)$	1	1/30
$[69, 73)$	5	5/30
$[73, 77)$	4	4/30
$[77, 81)$	7	7/30
$[81, 85)$	8	8/30
$[85, 89)$	3	3/30
$[89, 93)$	0	0
$[93, 97)$	0	0
$[97, 101)$	0	0
$[101, 105)$	1	1/30
$[105, 109)$	1	1/30
	30	1

(b) **(0.25 puntos)** ¿Qué porcentaje de observaciones se encuentra entre 78 y 81 grados Celsius?

Solución:

Puesto que hay tres observaciones, 78.3, 78.7 y 78.9, de 30 observaciones totales, tenemos que corresponden al 10%.

(c) **(1 punto)** Calcular los tres cuartiles de la muestra e interpretarlos.

Solución:

Los tres cuartiles muestrales son:

$$\begin{aligned}
x_{\left(\frac{31}{4}\right)} &= x_{(8)} = 75.4 \\
x_{\left(\frac{31}{2}\right)} &= \frac{x_{(15)} + x_{(16)}}{2} = \frac{78.3 + 78.7}{2} = 78.5 \\
x_{\left(\frac{93}{4}\right)} &= x_{(23)} = 82.4
\end{aligned}$$

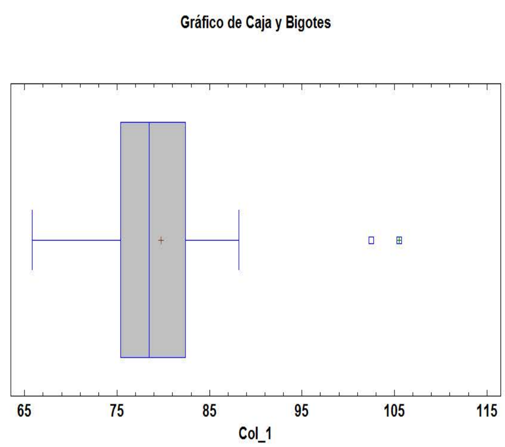
Esto implica que el 25% de las observaciones se encuentra por debajo de 75.4, el 50% de las observaciones se encuentra por debajo de 78.5 y el 75% de las observaciones se encuentra por debajo de 82.4. Por lo tanto, los tres cuartiles dividen la muestra de 30 observaciones en cuatro submuestras que recogen aproximadamente el mismo número de observaciones.

OBS: Los cuartiles también se pueden estimar de maneras alternativas. Por ejemplo:

$$\begin{aligned} x_{\left(\frac{31}{4}\right)} &= 0.25x_{(7)} + 0.75x_{(8)} = 75.075 \\ x_{\left(\frac{31}{2}\right)} &= \frac{x_{(15)} + x_{(16)}}{2} = \frac{78.3 + 78.7}{2} = 78.5 \\ x_{\left(\frac{93}{4}\right)} &= 0.75x_{(23)} + 0.25x_{(24)} = 82.925 \end{aligned}$$

(d) **(0.75 puntos)** Representar los datos mediante un diagrama de caja (boxplot) e identificar los posibles datos atípicos. Justificar la respuesta.

Solución:



En vista del diagrama de caja, existen dos datos atípicos ya que sus valores son mayores que $Q_3 + 1.5 \times RI$, donde RI es el rango intercuartilico.

2. Sea Y una variable aleatoria contínua definida en el intervalo $[0, 1]$ con función de densidad:

$$f_Y(y) = \begin{cases} 4y - 4y^3 & 0 \leq y \leq 1 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

Se pide:

(a) **(0.5 puntos)** Obtener la función de distribución de Y.

Solución:

$$F_Y(y) = \begin{cases} 0 & 0 < y \\ \int_0^y \left(4u - 4u^3\right) du = 2y^2 - y^4 = y^2\left(2 - y^2\right) & 0 \leq y \leq 1 \\ 1 & y > 1 \end{cases}$$

(b) **(0.5 puntos)** Calcular las probabilidades $P\left(0 < Y < \frac{1}{2}\right)$ y $P\left(\frac{1}{4} < Y < \frac{3}{4}\right)$.

Solución:

$$
\begin{aligned}
P\left(0 < Y < \frac{1}{2}\right) &= F_Y\left(\frac{1}{2}\right) - F_Y(0) = \frac{1}{2^2}\left(2 - \frac{1}{2^2}\right) - 0 = \frac{7}{16}. \\
P\left(\frac{1}{4} < Y < \frac{3}{4}\right) &= F_Y\left(\frac{3}{4}\right) - F_Y\left(\frac{1}{4}\right) = \frac{3^2}{4^2}\left(2 - \frac{3^2}{4^2}\right) - \frac{1}{4^2}\left(2 - \frac{1}{4^2}\right) = \frac{11}{16}.
\end{aligned}
$$

(c) **(1 punto)** Calcular la esperanza y la desviación típica de Y.

Solución:

$$
E[Y] = \int_0^1 y\left(4y - 4y^3\right) dy = \left(\frac{4}{3}y^3 - \frac{4}{5}y^5\right)\Big|_{y=0}^{y=1} = \frac{8}{15}.
$$

$$
V[Y] = E\left[Y^2\right] - E[Y]^2 = \int_0^1 y^2\left(4y - 4y^3\right) dy - \left(\frac{8}{15}\right)^2 = \left(y^4 - \frac{2}{3}y^5\right)\Big|_{y=0}^{y=1} - \frac{64}{225} = \frac{11}{225}.
$$

$$
DT[Y] = \sqrt{\frac{11}{225}} \simeq 0.2211.
$$

(d) **(0.5 puntos)** Calcular la esperanza y la desviación típica de $2Y + 3$.

Solución:

$$
\begin{aligned}
E[2Y+3] = 2E[Y] + 3 = 2\frac{8}{15} + 3 = \frac{17}{5} \\
DT[2Y+3] = 2DT[Y] \simeq 0.4422.
\end{aligned}
$$

3. Un alumno se presenta a un examen tipo test sin preparación. El examen consiste de 40 preguntas, y cada una de ellas tiene tres alternativas de las que sólo una es correcta. El examen se aprueba si se aciertan 20 preguntas o más. El alumno contesta a las preguntas totalmente al azar, de tal manera que la respuesta a una pregunta no afecta a la respuesta de las otras. Se pide:

   (a) **(1 punto)** Calcular la probabilidad de acertar las primeras 20 preguntas y fallar las 20 últimas.

   Solución:

   Introducimos para $i = 1, \ldots, 40$ las variables aleatorias:

$$
X_i = \begin{cases} 1 & \text{si acierta la pregunta } i\text{-ésima} \\ 0 & \text{si no acierta la pregunta } i\text{-ésima} \end{cases}
$$

   Cada X_i tiene una distribución $\text{Ber}\left(\frac{1}{3}\right)$. Por lo tanto, teniendo en cuenta la independencia entre las variables, tenemos que:

$$
\begin{aligned}
P(X_1 = 1, \ldots, X_{20} = 1, X_{21} = 0, \ldots, X_{40} = 0) = P(X_1 = 1) \cdots P(X_{20} = 1)\, P(X_{21} = 0) \cdots P(X_{40} = 0) = \\
= \left(\frac{1}{3}\right)^{20}\left(\frac{2}{3}\right)^{20} = 8.62 \times 10^{-4}.
\end{aligned}
$$

(b) **(1 punto)** Calcular de forma exacta la probabilidad de obtener más de un 9, es decir, de responder correctamente más de 36 preguntas.

Solución:

La variable aleatoria Y ="Número de respuestas correctas" es $Y = \sum_{i=1}^{40} X_i$ que tiene una distribución $\text{Bin}\left(40, \frac{1}{3}\right)$. Por lo tanto:

$$P(Y > 36) = P(Y = 37) + P(Y = 38) + P(Y = 39) + P(Y = 40) =$$
$$= \binom{40}{37}\left(\frac{1}{3}\right)^{37}\left(\frac{2}{3}\right)^{3} + \binom{40}{38}\left(\frac{1}{3}\right)^{38}\left(\frac{2}{3}\right)^{2} + \binom{40}{39}\left(\frac{1}{3}\right)^{39}\left(\frac{2}{3}\right) + \binom{40}{40}\left(\frac{1}{3}\right)^{40} =$$
$$= 6.5012 \times 10^{-15} + 2.5662 \times 10^{-16} + 6.5802 \times 10^{-18} + 8.2252 \times 10^{-20} =$$
$$= 6.7645 \times 10^{-15}.$$

(c) **(0.5 puntos)** Calcular de forma aproximada, utilizando el Teorema Central del Límite, la probabilidad de aprobar, es decir, de responder correctamente más de 20 preguntas.

Solución:

Puesto que $Y \sim \text{Bin}\left(40, \frac{1}{3}\right)$, entonces:

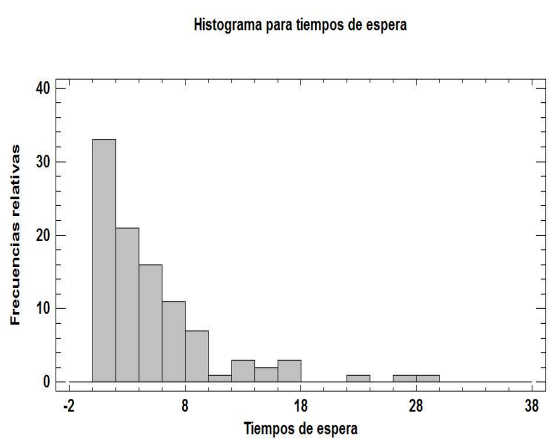
$$E[Y] = 40 \times \frac{1}{3} = \frac{40}{3}$$
$$DT[Y] = \left(40 \times \frac{1}{3} \times \frac{2}{3}\right)^{1/2} = \left(\frac{80}{9}\right)^{1/2} = \frac{80^{1/2}}{3}.$$

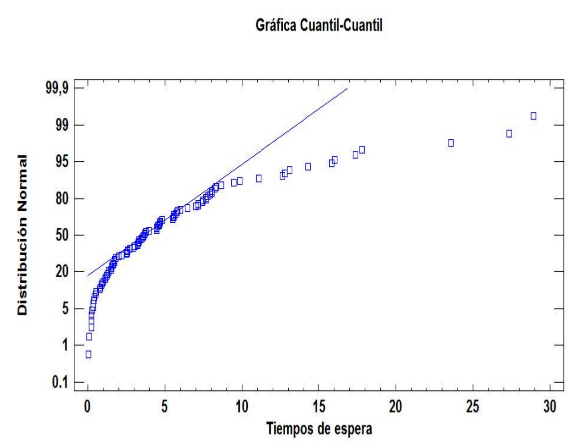
Entonces:

$$P(Y > 20) = P\left(\frac{Y - \frac{40}{3}}{\frac{80^{1/2}}{3}} > \frac{20 - \frac{40}{3}}{\frac{80^{1/2}}{3}}\right) \simeq P\left(Z > \frac{20}{80^{1/2}}\right) = P(Z > 2.2360) =$$
$$= 1 - P(Z \leq 2.2360) = 1 - 0.9871 = 0.0129,$$

donde $Z \sim N(0, 1)$.

4. En un centro comercial se está realizando un estudio acerca de la calidad del servicio que se está dando a los clientes. Concretamente se han recogido datos acerca del tiempo de espera (en minutos) para acceder a los ascensores del mismo. A continuación se muestran algunos de los análisis llevados a cabo:





Se pide:

(a) **(0.75 puntos)** Justificar si es cierto que el tiempo de espera para acceder a los ascensores de los clientes del centro comercial puede describirse mediante una ley de probabilidad Normal.

Solución:

En vista de los gráficos presentados, la distribución Normal no es adecuada para describir estos datos. En primer lugar, el histograma es claramente asimétrico, mientras que el gráfico cuantil-cuantil muestra claramente que los cuantiles muestrales no siguen aproximadamente una línea recta cuando son comparados con los cuantiles de la Normal.

(b) **(1 punto)** El centro comercial afirma que el tiempo de espera para acceder a los ascensores de los clientes del centro comercial es en media de 6 minutos con una desviación típica de 5 minutos. Si 50 personas toman el ascensor independientemente, cual es la probabilidad de que la suma de sus tiempos de espera esté entre 5.5 y 6 horas.

Solución:

Sea T la variable aleatoria "Tiempo de espera para acceder a los ascensores". Tenemos que $E[T] = 6$ y $DT[T] = 5$. Entonces, el TCL nos dice que:

$$Z = \frac{\overline{T} - 6}{\frac{5}{\sqrt{50}}} \underset{aprox.}{\sim} N(0,1).$$

Por lo tanto,

$$P\left(330 < \sum_{i=1}^{50} T_i < 360\right) = P\left(\frac{330}{50} < \overline{T} < \frac{360}{50}\right) = P\left(6.6 < \overline{T} < 7.2\right) =$$
$$= P\left(\frac{6.6-6}{\frac{5}{\sqrt{50}}} < Z < \frac{7.2-6}{\frac{5}{\sqrt{50}}}\right) = P(0.8485 < Z < 1.6970) =$$
$$= P(Z < 1.6970) - P(Z < 0.8485) = 0.9545 - 0.7995 = 0.1550.$$

(c) **(0.75 puntos)** Suponer que el tiempo espera para acceder a los ascensores de los clientes del centro comercial es en media de 6 minutos con una desviación típica de 5 minutos. Obtener una cota inferior de la probabilidad aproximada de que el tiempo total de espera para 25 personas esté entre 2 y 3 horas. (Utilizar la desigualdad de Chebyshev, es decir, para una variable aleatoria X con esperanza μ y varianza σ^2, entonces $P(|X-\mu| < k) \geq 1 - \frac{\sigma^2}{k}$, para cualquier constante positiva k).

Solución:

La desigualdad de Chebyshev nos dice que, para una variable aleatoria X con media μ y varianza σ^2, se verifica:

$$P(|X-\mu| < k) = P(\mu - k < X < \mu + k) \geq 1 - \frac{\sigma^2}{k^2}$$

Entonces, tenemos que la variable $\sum_{i=1}^{25} T_i$ tiene media $25 \times 6 = 150$ y varianza $25 \times 25 = 625$. Por lo tanto:

$$P\left(120 < \sum_{i=1}^{25} T_i < 180\right) = P\left(150 - 30 < \sum_{i=1}^{25} T_i < 150 + 30\right) \geq 1 - \frac{625}{900} = 0.3055.$$