Examen Final de Estadística I, 26 de Mayo de 2014.
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. En las siguientes tablas y gráficos se recoge información sobre las atenciones sanitarias no domiciliarias atendidas a través del teléfono 112 de Protección Civil durante los últimos 180 días en Pamplona. Para cada día se ha observado el tiempo medio de respuesta (en minutos) de las ambulancias para accidentes con heridos y el número de accidentes atendidos según su tipología (accidente de tráfico, laboral y otros).

Tabla 1: Tiempo medio de respuesta (en minutos)

(a) Global

Tiempo medio respuesta	
Media	13.9189854
Mediana	7.72569374
Desviación típica	14.1249749
Varianza de la muestra	199.514915
Curtosis	2.00834811
Coeficiente de asimetría	1.52135364
Rango	65.1620149
Mínimo	0.02300493
Máximo	65.1850198
Suma	2505.41738
Tamaño muestral	180

(b) de Lunes a Viernes

Tiempo medio respuesta (de lunes a viernes)	
Media	12.65007263
Mediana	7.43135166
Desviación típica	12.75266538
Varianza de la muestra	162.6304743
Curtosis	2.762768312
Coeficiente de asimetría	1.650832286
Rango	61.09465423
Mínimo	0.023004926
Máximo	61.11765915
Suma	1897.510894
Tamaño muestral	150

(c) Sábado y Domingo

Tiempo medio respuesta (sábado y domingo)	
Media	20.26354953
Mediana	16.45645042
Desviación típica	18.59931717
Varianza de la muestra	345.9345992
Curtosis	-0.193607802
Coeficiente de asimetría	0.858543921
Rango	64.99579226
Mínimo	0.189227592
Máximo	65.18501985
Suma	607.9064858
Tamaño muestral	30

Tabla 2: Número de accidentes atendidos

	Tipo de accidente		
Tipo día	tráfico	laboral	otros
de Lunes a Viernes	111	9	30
Sábado y Domingo	26	1	3

Gráfico 1: Tiempo medio de respuesta (global)

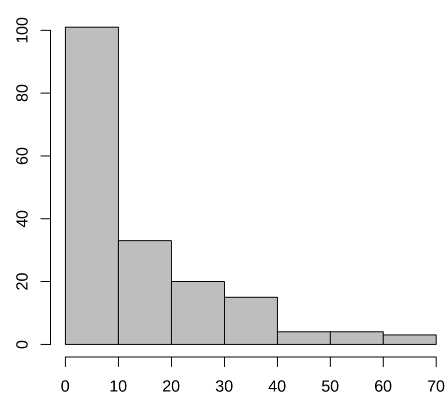
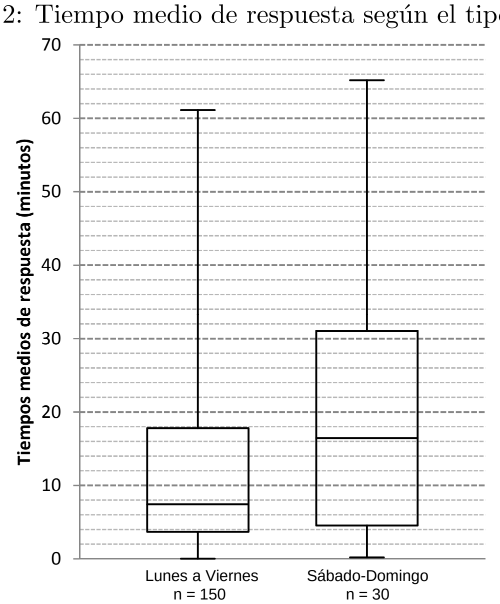


Gráfico 2: Tiempo medio de respuesta según el tipo de día



Responde justificadamente a las siguientes cuestiones, indicando claramente las tablas y/o gráficos que utilizas para tu razonamiento:

a) (0.5 puntos) Describe la forma de la distribución del tiempo medio de respuesta (global) e indica si hay datos atípicos. ¿Qué medida de centralización es más adecuada utilizar en este caso y por qué?

b) (0.75 puntos) Justifica si son ciertas o falsas las afirmaciones siguientes: **[1]** En más de la mitad de los casos el tiempo medio de respuesta de las ambulancias durante los fines de semana es más de dos veces superior al de los días laborables. **[2]** En el 75% de los casos el tiempo medio de respuesta de las ambulancias durante los fines de semana es superior a 30 minutos.

c) (0.5 puntos) Si tuvieras que dar un tiempo máximo de respuesta por debajo del cual pudieras garantizar que se ha atendido el 75% de las emergencias de lunes a viernes, ¿qué tiempo darías?

d) (0.75 puntos) Obtén las distribuciones de la variable *tipo de accidente* condicionadas al *tipo de día* y determina en qué tipo de día son más frecuentes los accidentes de tráfico.

Solución.

a) El Gráfico 1 es el histograma de la distribución de frecuencias absolutas de la variable *tiempo medio de respuesta (en minutos)* (sin distinguir por tipo de día). Se observa claramente que la distribución es asimétrica (con asimetría a la derecha, positiva).

   Al tratarse de una distribución asimétrica es mejor dar la Mediana como medida de centralización. En este caso, el tiempo mediano de respuesta (ver panel (a) de la Tabla 1) es de 7.7257 minutos.

   De forma aproximada (a partir del Gráfico 1) puede llegar a intuirse la presencia de datos atípicos. El razonamiento sería el siguiente: El tercer cuartil está en la posición 135, que corresponde a un valor situado entre 10 y 20, es decir, $Q_3 \in (10, 20)$. Por tanto, se deduce que $Q_3 < 20$, al igual que $RIC < 20$. Así, $Q_3 + 1.5\, RIC < 50$, con lo que cualquier dato superior a 50 puede ser considerado como atípico.

b) La afirmación **[1]** es cierta, puesto que la mediana del tiempo medio de respuesta de las ambulacias durante los fines de semana es de 16.46, mientras que de lunes a viernes vale 7.43 (ver paneles (b) y (c) de la Tabla 1 y también diagramas de caja del Gráfico 2). Sin embargo, la afirmación **[2]** es falsa, puesto que en el diagrama de caja del Gráfico 2 se observa que, para los fines de semana, el tercer cuartil del tiempo medio de respuesta es 31. Por tanto, el tiempo medio de repuesta es inferior a 31 minutos en el 75% de los casos.

c) Nos están pidiendo el tercer cuartil de la muestra: Q_3. En el diagrama de cajas correspondiente se aprecia que $Q_3 \approx 18$ minutos.

d) Para poder compararlas es más conveniente obtener las distribuciones condicionadas relativas (obtenidas a partir de la Tabla 2), que son:

Tipo de accidente más frecuente / tipo de día

	trafico	laboral	otros	
Lunes a Viernes	0,74	0,06	0,2	1
Sábado-Domingo	0,8666667	0,0333333	0,1	1

   De ella se observa que los accidentes de tráfico son más frecuentes durante el fin de semana, con un porcentaje del 86.67%.

2. Una empresa paga una cuota mensual de 175 euros en concepto de suscripción a un servicio de reparaciones. Además, cada visita de un técnico para realizar una reparación en la empresa cuesta 350 euros. Se sabe que la media de averías es de 9.5 al mes, con una desviación típica de 2.

   a) (0.5 puntos) Calcula el valor esperado y la varianza del coste mensual de las reparaciones (incluyendo la cuota mensual).

   b) (0.75 puntos) Utilizando la desigualdad de Chebyshev, da una cota para la probabilidad de que en un mes determinado el coste de las reparaciones sea menor o igual que 2000 o mayor o igual que 5000 euros.

   c) (0.75 puntos) Si se supone que el coste mensual sigue una distribución uniforme continua con la esperanza y la varianza obtenidas en *a)*, calcula de nuevo la probabilidad del apartado anterior.

   d) (0.5 puntos) A qué se debe la diferencia entre los resultados obtenidos en los apartados *b)* y *c)*?

Solución.

a) $X =$ número de averías al mes. $E[X] = 9.5$ y $Var[X] = 2^2 = 4$.
Sea $C =$ coste mensual de las reparaciones. Como $C = 350X+175$, entonces $E[C] = 350\cdot 9.5+175 = 3500$ y $Var[C] = 350^2 \cdot 4 = 490000$.

b)

$$P(C \leq 2000 \text{ ó } C \geq 5000) = P(C - 3500 \leq -1500 \text{ ó } C - 3500 \geq 1500) = P(|C - 3500| \geq 1500)$$

$$= P(|C - E[C]| \geq 1500) \overset{Des.Cheb.}{\leq} \frac{Var[C]}{1500^2} = \frac{490000}{1500^2} = 0.22.$$

c) Si $C \sim U(a, b)$, entonces $E[C] = \frac{a+b}{2} = 3500$ y $Var[C] = \frac{(b-a)^2}{12} = 490000$. Por tanto

$$\left.\begin{array}{rcl} \frac{a+b}{2} & = & 3500 \\ \frac{(b-a)^2}{12} & = & 490000 \end{array}\right\} \Leftrightarrow \left.\begin{array}{rcl} a & = & 7000 - b \\ b - a & = & \sqrt{12 \cdot 490000} \end{array}\right\} \Leftrightarrow \left.\begin{array}{rcl} a & = & \frac{7000-\sqrt{12\cdot 490000}}{2} = 2287.56 \\ b & = & \sqrt{12 \cdot 490000} + 2287.56 = 4712.44 \end{array}\right\}$$

Es decir, $C \sim U(2287.56, 4712.44)$ por lo que $P(C \leq 2000 \quad C \geq 5000) = 0$.

d) En el primer caso sólo contamos con los valores de la esperanza y la varianza, y la desigualdad de Chebyshev proporciona una cota muy gruesa de la probabilidad que se quiere calcular. En el segundo caso, al disponer de la distribución de la variable aleatoria podemos calcular la probabilidad de manera exacta.

3. La función de densidad asociada a la duración en días de un determinado producto es:

$$f(x) = \begin{cases} k\,(2 - x), & \text{si } 0 < x < 2, \\ 0, & \text{en otro caso,} \end{cases}$$

para un cierto valor de k.

a) (0.5 puntos) Determinar el valor de la constante k.

b) (0.5 puntos) Si el producto solamente está en condiciones óptimas para su venta entre 12 horas y 48 horas después de haber sido producido, calcular la probabilidad de que el producto esté en condiciones óptimas para su venta. (Considerar que un día son 24 horas).

c) (0.75 puntos) Una determinada empresa tiene capacidad para producir 4 de estos productos cada día. ¿Cuál es la probabilidad que no pueda vender todos los productos fabricados en el último día?

d) (0.75 puntos) La empresa quebrará si no puede vender todos los productos fabricados en al menos 320 días del año. Calcular la probabilidad de quiebra de la empresa. (Considerar que un año son 365 días).

Solución:

a) $\int_0^2 k\,(2 - x)\,dx = 1 \quad \Rightarrow \quad k\left(2x - \frac{x^2}{2}\right)\Big|_0^2 = 1 \quad \Rightarrow \quad k = \frac{1}{2}.$

b) Sea X la v.a. que representa la duración en días de cierto producto. Si el producto solamente está en condiciones óptimas para su venta entre 12 horas y 48 horas después de haber sido producido, entonces $P(0.5 < X < 2) = 0.5625$.

c) Sea Y la v.a. que cuenta el número de productos óptimos para la venta de un total de 4. $Y \sim B(4, 0.5625)$ y entonces, $P(Y < 4) = 1 - P(Y = 4) = 0.899$.

d) Sea T la v.a. que cuenta el número de días de un total de 365 en que no se venden todos los productos fabricados. $T \sim B(365, 0.899)$ que puede aproximarse por el TCL como $N(328.46, 5.983)$ y, por tanto, $P(\text{“}quiebra'') = P(T > 320) = P(Z > -1.41) = 1 - 0.0808 = 0.9192$, donde $Z \sim N(0, 1)$.

4. Los puntos anotados por un equipo de baloncesto a lo largo de un partido pueden modelarse según una variable aleatoria de media 90 y desviación típica 30.

(a) (0.5 puntos) Si tomamos una muestra de 35 partidos, calcular la probabilidad de que el puntuaje promedio de dicho equipo esté entre 80 y 100 puntos.

(b) (0.75 puntos) Calcular un intervalo de confianza al 95% para la media de puntos anotados por partido. Suponer que tomamos una muestra de 35 partidos y que la media muestral toma un valor de 85. Interpretar el resultado.

(c) (0.75 puntos) Suponer ahora que, para el intervalo de confianza del apartado *b)*, se quiere reducir su longitud en un 10%. Determinar el tamaño muestral (número de partidos) para que eso suceda.

(d) (0.5 puntos) Suponer ahora que el equipo contrata a un nuevo entrenador con el que se espera reducir la desviación típica hasta 15 (sin cambiar la media). Si una temporada tiene 50 partidos, dibujar (de forma aproximada) la función de densidad del puntuaje promedio de una temporada para ambos entrenadores y comparar ambas curvas. Comentar los resultados.

Solution:

a) Puesto que $n > 30$, puede aplicarse el TLC, que asegura que $\overline{X}$ tiene una ley aproximadamente normal. En concreto: $P\left(80 < \overline{X} < 100\right) = P\left(\frac{80-90}{30/\sqrt{35}} < Z < \frac{100-90}{30/\sqrt{35}}\right) = P\left(-1.97 < Z < 1.97\right) = 0.9512$, donde $Z = \frac{\bar{X}-90}{30/\sqrt{35}}$ sigue aproximadamente una distribucin $N(0,1)$.

b) En virtud del TLC, el intervalo de confianza es $I.C._{95\%}(\mu) = \overline{x} \pm 1.96\frac{\sigma}{\sqrt{n}} = 85 \pm 1.96\frac{30}{\sqrt{35}} = [75.06, 94.94]$. No sería correcto afirmar que con un 95% de probabilidad, el verdadero valor de la media está entre $[75.06, 94.94]$. Sin embargo, sí podríamos afirmar que el 95% de las veces que construimos un I.C. del modo anterior, encontraremos el verdadero valor de la media dentro de los mismos.

c) La longitud del intervalo es $2 \times 1.96 \times \frac{\sigma}{\sqrt{n_1}}$. Por tanto, queremos un n_2 tal que la nueva longitud del intervalo sea $2 \times 1.96 \times \frac{\sigma}{\sqrt{n_1}} \times 0.9$. Por tanto, tenemos la siguiente relación:

$$2 \times 1.96 \times \frac{\sigma}{\sqrt{n_1}} \times 0.9 = 2 \times 1.96 \times \frac{\sigma}{\sqrt{n_2}} \Rightarrow n_2 = \frac{n_1}{0.9^2} = \frac{35}{0.9^2} = 43.2 \rightarrow n_2 = 44.$$

d) Si dibujamos las funciones de densidad para el puntuaje promedio de un partido en una temporada, i.e. para $\overline{X}$, vemos que para ambos entrenadores la función está centrada en la misma media. Sin embargo, el segundo entrenador tiene un promedio de puntuaje más estable, por lo que dará resultados más homogéneos a lo largo de la temporada.