Estadística I Examen Extraordinario, 24 Junio 2014.
Grados en ADE, DER-ADE, ADE-INF, FICO, ECO, ECO-DER.

REGLAS DEL EXAMEN: 1) Usar cuadernillos separados para cada problema. 2) Hacer los cálculos con al menos dos decimales significativos. 3) No se puede abandonar el examen durante los primeros 30 minutos. 4) No está permitido salir de la clase sin entregar el examen.

1. El siguiente conjunto de datos muestra el número de coches vendidos en un mes en 23 concesionarios diferentes:

   9 8 14 23 23 27 18 18 8 21 20 18 14 20 29 18 33 19 23 24 24 14 19

   (a) (0.5 puntos) ¿Es la media de las ventas mensuales anteriores mayor de 20? ¿Y la mediana? ¿Es cierto que la venta más frecuente durante el mes fue de 20 coches? Justificar las respuestas.

   (b) (0.5 puntos) Calcular la cuasi-varianza muestral y el coeficiente de variación muestral para el conjunto de datos.

   (c) (1 punto) Dibujar un diagrama de caja para el conjunto de datos, obteniendo para ello las medidas numéricas necesarias. ¿Es cierto que la forma de la distribución presenta una fuerte asimetría a la derecha? Justificar la respuesta.

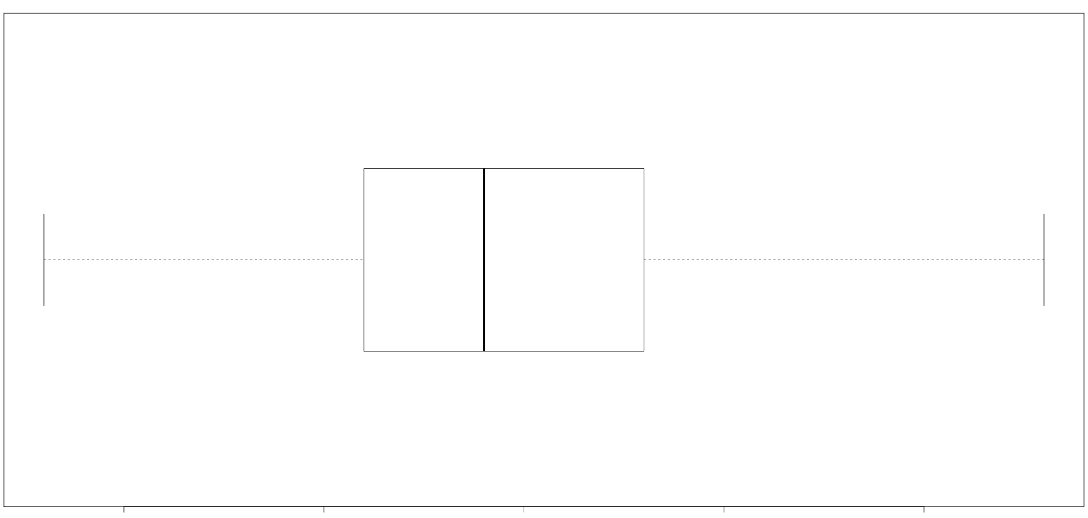
   (d) (0.5 puntos) Basándose en el diagrama de cajas realizado en el apartado anterior, ¿hay atípicos en el conjunto de datos? Justificar la respuesta.

   Solución.

   (a) La media muestral es $\overline{x} = 19.30$. La mediana muestral es $M = 19$. La moda muestral es 18. Entonces, la respuesta es NO para todas las preguntas.

   (b) La cuasi-varianza muestral es $\widehat{\sigma}^2 = 40.13043$. El coeficiente de variación muestral es 0.3281.

   (c) Los cuartiles muestrales son $Q_1 = 14$, $Q_2 = 19$ y $Q_3 = 23$. Entonces, el rango intercuartílico muestral es $IQR = 23 - 14 = 9$. El diagrama de caja es:



   La distribución es ligeramente asimétrica a la derecha. Por lo tanto, no es fuertemente asimétrica.

   (d) Por un lado, tenemos que $Q_1 - 1.5IQR = 14 - 1.5 \times 9 = 0.5$. Por otro lado, tenemos que $Q_3 + 1.5IQR = 23 + 1.5 \times 9 = 36.5$. Como los valores mínimo y máximo en la muestra son 8 y 33, respectivamente, y $0.5 < 8$ y $36.5 > 33$, concluimos que no hay datos atípicos.

2. Un joven emprendedor abre un negocio en una ciudad. Según cierta información fiable sobre negocios similares en dicha ciudad, se conoce que el 25% de ellos tienen ingresos anuales de hasta 1 millón de euros (incluido), el 50% de ellos tienen ingresos anules entre 1 (no incluido) y 2 millones de euros (incluido), y el resto tiene ingresos anules entre 2 (no incluido) y 3 millones de euros (incluido). Dada la información anterior, y suponiendo que los negocios en cada una de las franjas de beneficio consideradas se distribuyen uniformemente, se pide responder a las siguientes preguntas:

   (a) (0.75 puntos) Obtener la función de distribución acumulada de los ingresos anuales (en millones de euros).

   (b) (0.25 puntos) Obtener la probabilidad de que los ingresos anules del negocio abierto por el joven emprendedor estén entre 0.5 (no incluido) y 1.5 (incluido) millones de euros.

   (c) (0.75 puntos) Calcular el valor esperado y la desviación típica de los ingresos anuales (en millones de euros).

   (d) (0.75 puntos) Si el negocio tiene éxito, el joven emprendedor planea abrir otro negocio más pequeño que el primero. Se sabe que los ingresos anuales de los negocios más pequeños son un cuarto de los ingresos anuales de los negocios grandes. Entonces, ¿cuáles son los ingresos esperados del segundo negocio? ¿y la desviación típica?

Solución.

(a) Sea X los ingresos anuales del negocio (en millones de euros). Entonces, la función de densidad de X está dada por:

$$f(x) = \begin{cases} \frac{1}{4} & x \leq 1 \\ \frac{1}{2} & 1 < x \leq 2 \\ \frac{1}{4} & 2 < x \leq 3 \\ 0 & \text{en otro caso} \end{cases}$$

Por lo tanto, la función de distribución acumulada se puede calcular como:

- Para $x \leq 0$,
$$F(x) = 0$$

- Para $0 < x \leq 1$,
$$F(x) = \int_0^x \frac{1}{4} dx = \frac{x}{4}$$

- Para $1 < x \leq 2$,
$$F(x) = \int_0^1 \frac{1}{4} dx + \int_1^x \frac{1}{2} dx = \frac{1}{4} + \frac{1}{2}(x-1) = \frac{1}{2}\left(x - \frac{1}{2}\right)$$

- Para $2 < x \leq 3$,
$$F(x) = \int_0^1 \frac{1}{4} dx + \int_1^2 \frac{1}{2} dx + \int_2^x \frac{1}{4} dx =$$
$$= \frac{1}{4} + \frac{1}{2} + \frac{1}{4}(x-2) = \frac{1}{4}(x+1)$$

- Para $3 < x$,
$$F(x) = 1$$

En resumen,

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{4} & 0 < x \leq 1 \\ \frac{1}{2}\left(x - \frac{1}{2}\right) & 1 < x \leq 2 \\ \frac{1}{4}(x+1) & 2 < x \leq 3 \\ 1 & 3 \leq x \end{cases}$$

(b) La probabilidad de que los ingresos anuales del negocio (en millones de euros) estén entre 0.5 y 1.5 millones de euros está dada por:

$$\Pr(0.5 < x \leq 1.5) = F(1.5) - F(0.5) = \frac{1}{2} - \frac{1}{8} = \frac{3}{8} = 0.375$$

(c) La esperanza de X está dada por:

$$E[X] = \int_0^3 xf(x)\,dx = \int_0^1 \frac{x}{4}dx + \int_1^2 \frac{x}{2}dx + \int_2^3 \frac{x}{4}dx =$$
$$= \left.\frac{x^2}{8}\right|_{x=0}^{x=1} + \left.\frac{x^2}{4}\right|_{x=1}^{x=2} + \left.\frac{x^2}{8}\right|_{x=2}^{x=3} = \frac{1}{8} + \frac{3}{4} + \frac{5}{8} = \frac{3}{2} = 1.5 \text{ millones de euros}$$

La varianza de X está dada por:

$$V[X] = E\left[X^2\right] - E[X]^2$$

donde:

$$E\left[X^2\right] = \int_0^3 x^2 f(x)\,dx = \int_0^1 \frac{x^2}{4}dx + \int_1^2 \frac{x^2}{2}dx + \int_2^3 \frac{x^2}{4}dx =$$
$$= \left.\frac{x^3}{12}\right|_{x=0}^{x=1} + \left.\frac{x^3}{6}\right|_{x=1}^{x=2} + \left.\frac{x^3}{12}\right|_{x=2}^{x=3} = \frac{1}{12} + \frac{7}{6} + \frac{19}{12} = \frac{17}{6}$$

Entonces,

$$V[X] = \frac{17}{6} - \frac{9}{4} = \frac{7}{12} \simeq 0.5833$$

y la desviación típica es $\sqrt{\frac{7}{12}} \simeq 0.7637$ millones de euros.

(d) Los ingresos anuales esperados del segundo negocio y su desviación típica son un cuarto de las correspondientes al primer negocio. Entonces, el valor esperado es $\frac{3}{8} = 0.375$ millones de euros, mientras que la desviación típica es 0.1909.

3. Cada día un agente de ventas llama a 25 hogares independientes para intentar vender un nuevo producto, ganando una comisión de 45 euros por venta. Basado en su experiencia, el agente realiza, en promedio, una venta de cada seis llamadas en hogares con niños en edad escolar, y una venta de cada diez llamadas en hogares sin niños en edad escolar. El agente desconoce por completo el tipo de hogar al que va a llamar, pero sabe que el 30% de los hogares tienen niños en edad escolar. Se pide obtener:

   (a) (0.5 puntos) La probabilidad de que se produzca una venta en una llamada cualquiera.

   (b) (0.5 puntos) La probabilidad de que el número de ventas en un día cualquiera sea 3 o menos de 3.

   (c) (0.5 puntos) La media, varianza y desviación típica de la comisión total por ventas en un día.

   (d) (1 punto) La probabilidad (o una aproximación de ella) de que el agente gane en una semana (5 días de trabajo) más de 1000 euros en comisiones por ventas.

Solución.

(a) Por la ley de probabilidad total, escribiendo "niños en edad escolar" como NEE, la probabilidad de que una llamada resulte en una venta es:

$$\begin{aligned}\Pr(\text{Venta}) &= \Pr(\text{Venta}|\text{NEE})\Pr(\text{NEE}) + \Pr(\text{Venta}|\text{No NEE})\Pr(\text{No NEE}) = \\ &= \frac{1}{6} \times \frac{3}{10} + \frac{1}{10} \times \frac{7}{10} = 0.12\end{aligned}$$

(b) El número de ventas en un día es $N \sim Bin\,(25, 0.12)$. Por lo tanto, la probabilidad de que el número de ventas en un día sea 3 o menos de 3 está dado por:

$$\begin{aligned}
\Pr(N \leq 3) &= \Pr(N=0) + \Pr(N=1) + \Pr(N=2) + \Pr(N=3) = \\
&= \binom{25}{0} 0.12^0 (1-0.12)^{25} + \binom{25}{1} 0.12^1 (1-0.12)^{24} + \\
&+ \binom{25}{2} 0.12^2 (1-0.12)^{23} + \binom{25}{3} 0.12^3 (1-0.12)^{22} = \\
&= 1 \times 0.0409 + 25 \times 0.0056 + 300 \times 0.0008 + 2300 \times 0.0001 = \\
&\simeq 0.6475
\end{aligned}$$

(c) La media, varianza y desviación típica de N están dadas por:

$$\begin{aligned}
E[N] &= 25 \times 0.12 = 3 \\
V[N] &= 25 \times 0.12 \times 0.88 = 2.64 \\
DT[N] &= \sqrt{2.64} = 1.6248
\end{aligned}$$

Entonces, la media, varianza y desviación típica de la comisión de ventas totales en un día están dadas por:

$$\begin{aligned}
E[45N] &= 45 \times E[N] = 45 \times 3 = 135 \text{ euros} \\
V[45N] &= 45^2 \times V[N] = 5346 \text{ euros}^2 \\
DT[45N] &= 45 \times DT[N] = 45 \times 1.6248 = 73.116 \text{ euros}
\end{aligned}$$

(d) El número de ventas en una semana es $M \sim Bin\,(125, 0.12)$. Entonces, la comisión total ganada en una semana de trabajo es $W = 45M$. La media, varianza y desviación típica de M está dada por:

$$\begin{aligned}
E[M] &= 125 \times 0.12 = 15 \\
V[M] &= 125 \times 0.12 \times 0.88 = 13.2 \\
DT[M] &= \sqrt{13.2} = 3.6331
\end{aligned}$$

Usando el TCL, obtenemos que:

$$\begin{aligned}
\Pr(W > 1000) &= \Pr(45M > 1000) = \Pr(M > 22.22) = \\
&= \Pr\left(\frac{M-15}{3.6331} > \frac{22.22-15}{3.6331}\right) \simeq \Pr(Z > 1.9872) = \\
&= 1 - \Pr(Z < 1.9872) = 1 - 0.9761 = 0.0239
\end{aligned}$$

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

4. Los clientes de un banco pueden ingresar o retirar dinero. Sea X la cuantia de una operación realizada por un cliente donde el valor de X es positivo, si el dinero es ingresado, o negativo, si el dinero es retirado. Los analistas del banco asumen que la esperanza y la desviación típica de X son 0 y 100 euros, respectivamente.

   (a) (1 punto) Calcular la probabilidad de que la media de 100 operaciones realizadas por clientes independientes esté entre -3 and 3 euros.

   (b) (0.5 puntos) Sin realizar ningún tipo de cálculo adicional, decidir si la probabilidad en el apartado (a) será mayor o menor si el tamaño muestral aumenta. Justificar la respuesta.

   (c) (0.5 puntos) Supongamos que 100 clientes visitan el banco un cierto día. Calcular la probabilidad de que la suma de operaciones al final de dicho día sea positiva.

   (d) (0.5 puntos) Incluso si el banco asume que el valor esperado de X es 0, los analistas del banco sospechan que últimamente los clientes ingresan más dinero del que retiran. Para comprobar esta sospecha, los analistas toman una muestra de 25 operaciones realizadas por clientes independientes y obtienen una media muestral de 50 euros. Se pide obtener un intervalo de confianza al 90% para el valor esperado suponiendo que la distribución de X es Gaussiana con desviación típica 100. ¿Que nivel de confianza tienes de que el valor esperado sea realmente mayor que 0?

Solución:

(a) Como $n \geq 30$, por el TCL tenemos que $\overline{X} \sim N\left(0, \frac{100}{\sqrt{n}}\right) = N(0, 10)$. Entonces, la probabilidad requerida está dada por:

$$\Pr\left(-3 < \overline{X} < 3\right) = \Pr\left(\frac{-3}{10} < \frac{\overline{X}}{10} < \frac{3}{10}\right) \simeq \Pr(-0.3 < Z < 0.3) =$$
$$= \Pr(Z < 0.3) - \Pr(Z < -0.3) = 0.6179 - 0.3821 = 0.2358$$

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

(b) Si el tamaño muestral aumenta, la varianza de $\overline{X}$ decrece. Por lo tanto, la probabilidad de $\overline{X}$ de estar en el intervalo $(-3, 3)$ aumentará.

(c) Queremos calcular la siguiente probabilidad:

$$\Pr\left(\sum_{i=1}^{n} X_i > 0\right)$$

Esta probabilidad es equivalente a:

$$\Pr\left(\overline{X} > 0\right)$$

Entonces, como en a), por el TCL tenemos que $\overline{X} \sim N\left(0, \frac{100}{\sqrt{n}}\right) = N(0, 10)$. Entonces, la probabilidad requerida está dada por:

$$\Pr\left(\sum_{i=1}^{n} X_i > 0\right) = \Pr\left(\overline{X} > 0\right) = \Pr\left(\frac{\overline{X}}{10} > 0\right) \simeq \Pr(Z > 0) = 0.5$$

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

(d) El intervalo de confianza para la media al 90% está dado por:

$$\left(50 - 1.65\frac{100}{\sqrt{25}}, 50 + 1.65\frac{100}{\sqrt{25}}\right) = (17, 83)$$

Consecuentemente, estamos más de un 90% seguros de que la media verdadera sea positiva.