Examen Final de Estadística I, 18 de Mayo 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. La tabla siguiente contiene los resultados de una encuesta realizada en 1000 hogares españoles, donde se registraron los valores de (X, Y), siendo $X =$"número de coches en propiedad en 2011", con posibles valores $0, 1, 2$, e $Y =$"renta neta del hogar en 2011 (en miles de euros)".

		Y	
	$[0, 50)$	$[50, 100)$	$[100, \infty)$
0	324	105	37
X 1	112	234	6
2	1	4	177

(a) **(0.25 puntos)** ¿De qué tipo son las variables X e Y?

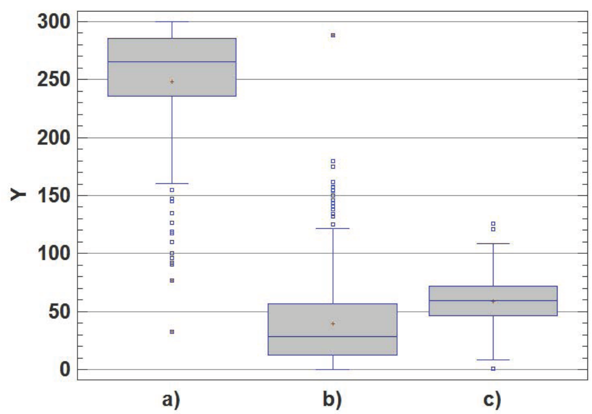
(b) **(0.5 puntos)** Obtener la frecuencia marginal absoluta de X. Calcular la media y cuasi-desviación típica de X.

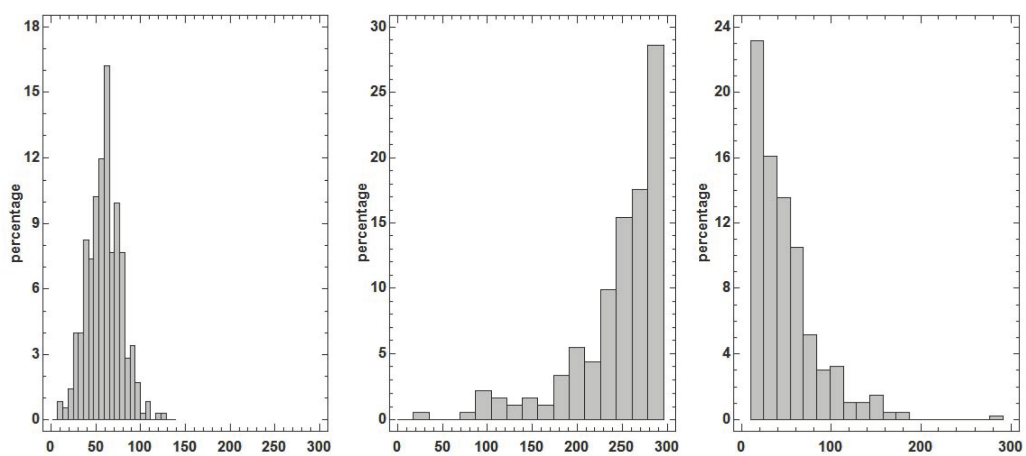
(c) **(0.5 puntos)** Para los hogares con renta neta inferior a 50 mil euros, determinar el promedio y el número más frecuente de coches en propiedad.

Se dispone de la siguiente información de la variable Y, respecto de los tres grupos que define X:

Resumen Estadístico para Y

	$X = 0$	$X = 1$	$X = 2$
Recuento	466	352	182
Promedio	39.70	58.93	248.12
Mediana	28.35	59.52	265.24
Varianza	1376.33	372.85	2672.49
Desv. Estándar	37.10	19.31	51.70
Coef. de Variación	93.44%	32.77%	20.84%
Mínimo	0.03	0.94	32.56
Máximo	288.44	126.09	299.81
Rango	288.41	125.15	267.26
Primer Cuartil	12.39	46.16	235.33
Tercer Cuartil	57.07	71.96	285.62
Rango Intercuartílico	44.67	25.80	50.29





(d) **(0.5 puntos)** Identificar cada diagrama de caja con cada uno de los grupos $X = 0$, $X = 1$, $X = 2$. Justificar la respuesta.

(e) **(0.25 puntos)** Relacionar los histogramas I), II), III) con los diagramas de caja a), b), c) del apartado anterior. Justificar la respuesta.

(f) **(0.5 puntos)** Escogemos un hogar de cada uno de los tres grupos y observamos su renta (en miles de euros). Los valores son: 51 para $X = 0$; 62 para $X = 1$ y 75 para $X = 2$. Decidir y justificar cuál de estos tres hogares es más pobre en relación con los tres grupos que define la variable X. (Indicación: estandarizar).

Solution:

(a) X is numerical discrete and Y is numerical continuous.

(b) The marginal frequency of X is

x_i	0	1	2
$n_{i\bullet}$	466	352	182

The mean of X is

$$\bar{x} = \frac{(0)466 + 1(352) + 2(182)}{1000} = 0.716$$

and its quasi-variance is

$$s_x^2 = \frac{(0)^2 \cdot 466 + 1^2 \cdot 352 + 2^2 \cdot 182 - 1000(0.716)^2}{1000 - 1} = \frac{567.344}{999} = 0.5679$$

and thus $s_x = \sqrt{s_x^2} = 0.7536$

(c) First we need to find the conditional distribution of $X|Y < 50$. It is given by

x	0	1	2
$f(x\|Y < 50)$	324/437=0.7414	112/437=0.2563	1/437=0.0023

The corresponding mean is

$$\overline{x|Y < 50} = 0\,\frac{324}{437} + 1\,\frac{112}{437} + 2\,\frac{1}{437} = \frac{114}{437} = 0.2609$$

The mean number is 0.2609. The mode is 0.

(d) Boxplot a) corresponds to those households for which $X = 2$, b) to those where $X = 0$ and c) to those where $X = 1$. Possible justifications: shape of the distributions (compare mean with the median), five-number summaries, positions of the quartiles...

(e) Histogram I corresponds to c) ($X = 1$), II to a) ($X = 2$), III to b) ($X = 0$). Possible justification: shape of the distributions, ...

(f) We use the standardized incomes to make the comparison. The standardized incomes are calculated in this table

	$X = 0$	$X = 1$	$X = 2$
mean	39.7021	58.9252	248.115
quasi-SD	37.0989	19.3092	51.6961
given incomes	51	62	75
their z-scores	0.3045	0.1592	-3.3487

The household with the income of 75,000 ($X = 2$) is the poorest in relative terms.
Alternatively, the last income is the only one below the group-average (negative z-score) whilst the other two are above (positive z-scores), so the last one is the poorest (no calculations needed).

2. En una fábrica hay dos líneas de montaje L1 y L2. En la línea L1 un 5% de los productos son defectuosos, mientras que en la línea L2, un 10% lo son. Un inspector de calidad escoge los productos que va a analizar de la siguiente manera: Primero escoge una de las líneas (con probabilidad 0.4 elige la línea L1 y con probabilidad 0.6 la L2) y luego dentro de esa línea escoge un producto al azar.

   (a) **(0.5 puntos)** Calcular la probabilidad de que el inspector analice un producto que sea defectuoso.

   En una inspección regular el inspector debe analizar 3 productos. Para ello, decide repetir el proceso anterior de forma independiente para cada uno de los productos que va a analizar.

   (b) **(0.5 puntos)** Considerar la variable aleatoria $X =$"número de productos defectuosos en una inspección regular". Describir el soporte de X (conjunto donde X toma valores) y determinar si X es discreta o continua.

   (c) **(0.5 puntos)** Obtener la función de probabilidad de X.

   (d) **(0.5 puntos)** Calcular la esperanza de X.

   (e) **(0.5 puntos)** Por cada producto defectuoso encontrado en una inspección regular, el inspector impone a la fábrica una multa de 30 euros. Considerar la v.a. $Y =$"importe total de la multa en una inspección regular". Describir el soporte de Y (conjunto donde Y toma valores) y calcular el importe medio con que puede ser multada la fábrica después de una inspección regular.

   Solución:

   (a) El primer apartado se hace por el teorema de la probabilidad total. Sea D el suceso: "producto defectuoso". Consideramos los sucesos $L1$ y (respectivamente $L2$): "el producto se elige de la línea L1 (respectivamente L2)".

   $$P(D) = P(D|L1)P(L1) + P(D|L2)P(L2) = 0.4 \cdot 0.05 + 0.6 \cdot 0.10 = 0.08$$

   (b)-(d) Los apartados del (b) al (d) se pueden hacer de dos formas:
   forma corta: La variable aleatoria $X =$"número de productos defectuosos en una inspección regular" sigue una distribución Binomial de parámetros $n = 3$ y $p = 0.08$. El soporte de X es $\{0, 1, 2, 3\}$ y, por tanto, es una variable aleatoria discreta. Su función de probabilidad es:

   $$P(X = k) = \binom{3}{k} 0.08^k \, 0.92^{3-k}, \text{ para } k = 0, 1, 2, 3.$$

   Su esperanza es $E(X) = n\,p = 3 \cdot 0.08 = 0.24$.
   forma larga:

   (b) El soporte de X es $\{0, 1, 2, 3\}$ y, por tanto, es una variable aleatoria discreta.

   (c) Considerar los sucesos $D_i =$"el producto i es defectuoso", $N_i =$"el producto i no es defectuoso". La función de probabilidad es:

   $$P(X = 0) = P(N_1)P(N_2)P(N_3) = 0.92^3 = 0,778688$$

   $$P(X = 1) = P(N_1)P(N_2)P(D_3) + P(D_1)P(N_2)P(N_3) + P(N_1)P(D_2)P(N_3) = 3 \cdot 0.08 \cdot 0.92 \cdot 0.92 = 0,203136$$

   $$P(X = 2) = P(N_1)P(D_2)P(D_3) + P(D_1)P(N_2)P(D_3) + P(D_1)P(D_2)P(N_3) = 3 \cdot 0.08 \cdot 0.08 \cdot 0.92 = 0,017664$$

   $$P(X = 3) = P(D_1)P(D_2)P(D_3) = 0.08^3 = 0,000512$$

   (d) La esperanza es:

   $$E(X) = 0 \cdot 0.778688 + 1 \cdot 0.203136 + 2 \cdot 0.017664 + 3 \cdot 0.000512 = 0.24$$

   (e) El soporte de la v.a. $Y =$"importe total de la multa en una inspección regular" es $\{0, 30, 60, 90\}$ y, puesto que $Y = 30X$, su esperanza es $E(Y) = 30E(X) = 30 \cdot 0.24 = 7.2$ euros, que es el importe medio con que puede ser multada la fábrica.

3. Una empresa de fabricación de componentes electrónicos recibe una media de 3 pedidos al día. Suponiendo que el número de pedidos recibidos por día sigue una distribución de Poisson, responde a las siguientes preguntas:

   (a) **(0.75 puntos)** ¿Cuál es la probabilidad de que en un día determinado se reciban más de 5 pedidos?

   (b) **(0.75 puntos)** ¿Cuál es la probabilidad de que lleguen dos pedidos en una hora? Suponer que un día laborable tiene 8 horas.

   (c) **(1 punto)** La empresa tiene la política de atender los pedidos el mismo día que se reciben, aunque esto implique que los empleados tengan que hacer horas extra. De manera general, esto ocurre cuando se reciben más de cinco pedidos en un mismo día. ¿Cuál es la probabilidad de que los empleados tengan que hacer horas extra al menos un día en una semana cualquiera? Suponer que una semana tiene 5 días laborables.

Solución:

(a) Sea X = "número de pedidos que se reciben al día". $X \sim Pois(3)$. La probabilidad de que en un día determinado se reciban más de 5 pedidos es:

$$P(X > 5) = 1 - P(X \leq 5) = 1 - \sum_{x=0}^{5} P(X = x) = 1 - \sum_{x=0}^{5} e^{-3}\frac{3^x}{x!} = 1 - e^{-3}\left(\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!} + \frac{3^3}{3!} + \frac{3^4}{4!} + \frac{3^5}{5!}\right) = 0.0839.$$

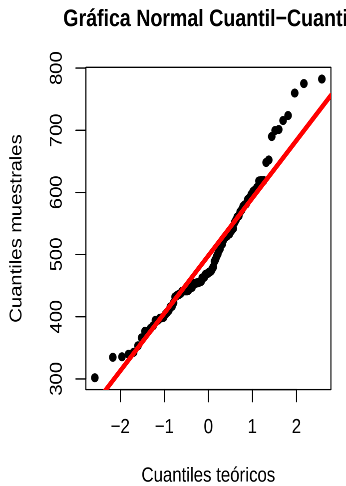
(b) Sea U = "número de pedidos que se reciben en una hora". Si $X \sim Pois(3)$, sabemos que $Y \sim Pois\left(\frac{1}{8}3\right)$. La probabilidad de que lleguen dos pedidos en una hora es:

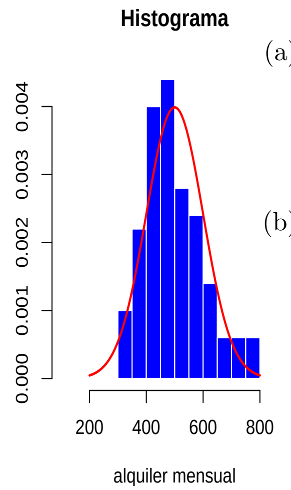
$$P(U = 2) = e^{-3/8}\frac{(3/8)^2}{2!} = 0.0483.$$

(c) Sea V = "número de días por semana en que los empleados tienen que hacer horas extra" = "número de días sobre 5 en que llegan más de 5 pedidos". Puesto que los pedidos llegados un día se atienden y despachan ese mismo día y no se acumulan al siguiente, y los pedidos llegan de forma independiente en los sucesivos días, tenemos que $V \sim Bin(5, p)$, donde $p = P(X > 5) = 0.0839$ es la probabilidad calculada en el apartado a). Entonces, la probabilidad de que los empleados tengan que hacer horas extra al menos un día en una semana determinada es:

$$P(V \geq 1) = 1 - P(V = 0) = 1 - \binom{5}{0} 0p^0(1-p)^{5-0} = 1 - \binom{5}{0} 00.9161^5 = 1 - 0.9161^5 = 0.3548.$$

4. Una agencia inmobiliaria está interesada en estudiar el precio del alquiler de los pisos de cierta localidad. Para ello, toma una muestra aleatoria simple (m.a.s.) de 80 familias, que considera representativas de la localidad, preguntándoles el alquiler mensual que pagan. Con la información obtenida la agencia realiza los siguientes gráficos:





(a) **(0.75 puntos)** Justificar si es cierto que los valores de los alquileres mensuales de las familias de esta localidad pueden describirse mediante una distribución de probabilidad Normal.

(b) **(1 punto)** La agencia inmobiliaria afirma que el alquiler mensual de las familias de esta localidad es en media de 500 euros con una desviación típica de 100 euros. Determinar la probabilidad (aproximada) de que la suma de los alquileres mensuales de 80 familias se encuentre entre 40000 y 42000 euros.

(c) **(0.75 puntos)** Si en lugar de tomar una m.a.s. de 80 familias sólo se hubiera encuestado a 25 familias, ¿se podría calcular la probabilidad (aproximada) de que la media de los alquileres mensuales de 25 familias se encontrara entre 500 y 600 euros? Justificar la respuesta.

Solución:

(a) No. En el QQ-plot los datos se alejan considerablemente de la línea recta. Además la forma del histograma no es simétrica.

(b) $P(40000 < \sum_{i=1}^{80} X_i < 42000) \approx P(500 < \overline{X}_n < 525) = P(0 < Z < 2.24) = 0.4875$

(c) En este caso no se podría justificar la utilización del Teorema Central del Límite.