ESTADÍSTICA 1: EXAMEN 2015-01-13

RECONSTRUCCIÓN CON SOLUCIONES

- Duración del examen: 2 horas.
- AVISO: Esto es una reconstrucción aproximada. Aunque está redactado con otras palabras, los ejercicios son equivalentes. Los datos numéricos difieren del examen original en los ejercicios 3 y 4, pero coinciden en el resto.

1º Esta es la distribución de la renta mensual de las familias de un pueblo:

Renta	n_i	f_i
450 – 900	31	0,1250
900 – 1050	44	0,2031
1050 – 1200	73	0,2188
1200 – 1500	80	0,2281
1500 – 2000	62	0,1313
2000 – 4500	30	0,0938
	N = 320	

a) Se quiere cobrar un impuesto al 10 % de población con mayor renta. ¿Cuál es el salario mínimo que deberíamos establecer para cobrar dicho impuesto? (1 pto.)
b) Se quiere dar una ayuda de 100 € a las familias que cobren menos de 1000 €. Estime cuánto dinero se deberá destinar a la ayuda. (Recuerde que el dinero de la ayuda resulta de multiplicar 100 € por el número de familias que reciben la ayuda.) (1 pto.)
c) Calcule el valor modal de la renta mensual. (0,75 ptos.)

2º En una clase de 20 alumnos, se registraron las notas que obtuvieron en un examen (A = aprobado; S = suspendido) junto con su sexo (M = masculino; F = femenino).

nº	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
M/F	M	F	M	M	F	M	M	F	F	F	M	F	M	F	F	M	F	M	F	F
A/S	S	A	S	A	A	S	A	S	A	A	S	A	S	S	A	A	A	S	A	S

Estudie la relación que existe entre las dos variables. (1,25 ptos.)

3º A partir de la siguiente serie temporal.

Cuatrimestre	Y_t
2010.1	31
2010.2	40
2010.3	36
2011.1	33
2011.2	42
2011.3	35
2012.1	34
2012.2	42
2012.3	37

a) Estudie la estacionalidad de la serie, calculando la tendencia por medias móviles y siguiendo un esquema multiplicativo. (1 pto.)
b) Obtenga la tendencia de la serie usando el método analítico. Calcule un estadístico para estudiar el grado de bondad de la recta obtenida. (1,5 ptos.)

4º En una empresa se ha recogido información sobre el salario de los trabajadores. Para cada categoría laboral, en esta tabla se indica el salario (precios) y el número de empleados (cantidades) de esa categoría.

	Año 2008		Año 2010	
Categoría	Salario	Empleados	Salario	Empleados
Técnicos	42	5	45	3
Administrativos	25	7	27	8
Subalternos	12	11	15	13

a) Calcule el índice de la media agregativa del año 2010 con base en 2008. (0,75 ptos.)
b) Calcule el índice de Laspeyres de precios (salarios) del año 2010 con base en 2008. (0,75 ptos.)
c) Calcule el índice de Paasche de cantidades (número de empleados) del año 2010 con base en 2008. (0,5 ptos.)
d) Si el año 2011 los salarios se incrementaron un 3,5 % con respecto al 2010, vuelva a calcular el índice del apartado b para el año 2011. (0,5 ptos.)

5º Se ha realizado un estudio sobre un conjunto de individuos, agrupándolos según su edad y su renta, de acuerdo a la siguiente tabla.

↓ Edad | Renta →	15 – 20	20 – 45	45 – 60
20 – 40	15	12	11
40 – 60	5	6	2
60 – 80	4	2	1

Calcule la mediana y la media de la renta de las personas de más de 40 años. Basándose en el cálculo obtenido (sin calcular nada más), ¿qué puede afirmar con respecto a la simetría de la distribución? (1 pto.)

SOLUCIONES

EJERCICIO 1

En este ejercicio necesitamos completar la tabla con las frecuencias acumuladas. También obtendremos la densidad de frecuencia ($d_i = n_i/A_i$) para la moda.

Renta	n_i	f_i	N_i	F_i	d_i
450 – 900	31	0,0969	31	0,0969	0,0689
900 – 1050	44	0,1375	75	0,2344	0,2933
1050 – 1200	73	0,2281	148	0,4625	0,4867
1200 – 1500	80	0,2500	228	0,7125	0,2667
1500 – 2000	62	0,1938	290	0,9063	0,1240
2000 – 4500	30	0,0938	320	1,0000	0,0120
	N = 320				

Apartado A

La renta mínima del 10 % más rico se corresponde con la renta máxima del 90 % más pobre, es decir, con el percentil 90.

Por tanto tenemos que calcular P_{90}, lo cual puede hacerse de dos maneras.

<u>Método 1: cálculo del percentil 90 usando frecuencias absolutas</u>

Primero determinamos el intervalo en que se encuentra este percentil:

$$N_{90\,\%} = \frac{90}{100} \cdot N = \frac{90}{100} \cdot 320 = 288$$

El percentil 90 se encuentra en el intervalo 1500 – 2000, ya que (N_i = 290) > 288.

Ahora aplicamos la fórmula del percentil por intervalos, usando n_i:

$$P_k = E_{i-1} + A_i \left(\frac{(k/100) \cdot N - N_{i-1}}{n_i} \right)$$

$$P_{90} = 1500 + 500 \left(\frac{(90/100) \cdot 320 - 228}{62} \right) = 1983,87\,€$$

<u>Método 2: cálculo del percentil 90 usando frecuencias relativas</u>

Tenemos que construir la columna F_i (frecuencias relativas acumuladas) en la tabla. El percentil 90 se encuentra en el intervalo 1500 – 2000, ya que (F_i = 0,906) > 0,9.

Ahora aplicamos la fórmula del percentil por intervalos, usando f_i:

$$P_k = E_{i-1} + A_i \left(\frac{(k/100) - F_{i-1}}{f_i} \right)$$

$$P_{90} = 1500 + 500 \left(\frac{(90/100) - 0,7125}{0,1938} \right) = 1983,87\,€$$

Apartado B

Hay que calcular el número de familias que cobran menos de 1000 € y luego multiplicar este valor por 100 €.

Aplicaremos la fórmula de la frecuencia acumulada para datos en intervalos, teniendo en cuenta que el valor de 1000 € pertenece al intervalo 900 – 1050.

$$N(x) = N_{i-1} + n_i\left(\frac{x - E_{i-1}}{A_i}\right)$$

$$N(x = 1000) = 31 + 44\left(\frac{1000 - 900}{150}\right) = 60,\widehat{3} \text{ familias}$$

Dinero destinado a la ayuda:

$$100\,€ \times 60,\widehat{3} \text{ familias} = 6033,33\,€$$

Apartado C

Como los datos están agrupados en <u>intervalos desiguales</u>, usaremos la fórmula correspondiente, pero primero hemos construido la columna de la densidad de frecuencia ($d_i = n_i/A_i$), así que nos fijaremos en d_i, en vez de en n_i directamente.

El intervalo modal es el de mayor densidad: 1050 – 1200 ($d_i = 0,4867$).

$$Mo = E_{i-1} + A_i\left(\frac{d_{i+1}}{d_{i-1} + d_{i+1}}\right)$$

$$Mo = 1050 + 150\left(\frac{0,2667}{0,2933 + 0,2667}\right) = 1121,44\,€$$

EJERCICIO 2

Nos piden estudiar la relación entre dos variables cualitativas, por lo que crearemos la tabla de contingencia para posteriormente calcular χ^2 y T^2.

	Femenino	Masculino	Total
Aprobado	8	3	11
Suspendido	3	6	9
Total	11	9	20

Construimos la <u>tabla de valores esperados</u>:

	Femenino	Masculino	Total
Aprobado	6,05	4,95	11
Suspendido	4,95	4,05	9
Total	11	9	20

Calculamos el coeficiente de contingencia χ^2 y el de Tschuprow:

$$\chi^2 = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{\left(n_{ij}-n'_{ij}\right)^2}{n'_{ij}}$$

$$\chi^2 = \frac{(8-6,05)^2}{6,05}+\frac{(3-4,95)^2}{4,95}+\frac{(3-4,95)^2}{4,95}+\frac{(6-4,05)^2}{4,05} = 3,1038$$

$$T^2 = \frac{\chi^2/N}{\sqrt{(h-1)(k-1)}}$$

$$T^2 = \frac{3,1038/20}{\sqrt{(2-1)(2-1)}} = \frac{3,1038/20}{1} = 0,1552$$

Como T^2 es cercano a 0, las variables están poco relacionadas.

EJERCICIO 3

Apartado A

En la siguiente tabla calculamos la tendencia por medias móviles (Tt) junto con la serie sin tendencia (Yt/Tt).

t	**Yt**	**Tendencia (Tt)**	**Yt/Tt**
2010.1	31		
2010.2	40	35,667	1,1215
2010.3	36	36,333	0,9908
2011.1	33	37,000	0,8919
2011.2	42	36,667	1,1455
2011.3	35	37,000	0,9459
2012.1	34	37,000	0,9189
2012.2	42	37,667	1,1150
2012.3	37		

Ahora obtendremos la <u>componente estacional</u> (Et). Para ello, hallamos la media de la serie sin tendencia de cada estación (Mi), y luego a cada estación le dividimos la media de las medias (media anual, MA).

↓ Cuat. | Año →	2010	2011	2012	Mi	Et=Mi/MA
1		0,8919	0,9189	0,9054	**0,9051**
2	1,1215	1,1455	1,1150	1,1273	**1,1269**
3	0,9908	0,9459		0,9684	**0,9680**

$$MA = \frac{0,9054+1,1273+0,9684}{3} = 1,0004$$

Apartado B

En la siguiente tabla calculamos la tendencia por el método analítico.

t	Yt	x	x^2	y^2	x*Yt	Tendencia
2010.1	31	1	1	961	31	34,800
2010.2	40	2	4	1600	80	35,267
2010.3	36	3	9	1296	108	35,733
2011.1	33	4	16	1089	132	36,200
2011.2	42	5	25	1764	210	36,667
2011.3	35	6	36	1225	210	37,133
2012.1	34	7	49	1156	238	37,600
2012.2	42	8	64	1764	336	38,067
2012.3	37	9	81	1369	333	38,533
	330	45	285	12224	1678	

La recta de regresión la obtenemos del siguiente modo:

$$\hat{y}_t = \alpha + \beta x_t \qquad \beta = \frac{\operatorname{Cov}(x,y)}{\operatorname{Var}(x)} \qquad \alpha = \overline{y} - \beta \overline{x}$$

$$\overline{x} = \frac{1+2+3+\ldots+9}{9} = \frac{1+9}{2} = 5 \qquad \overline{y} = \frac{31+40+36+\ldots+37}{9} = 36,\widehat{6}$$

$$\operatorname{Var}(x) = \left(\frac{1}{N}\sum_{i=1}^{j} x_i^2 n_i\right) - \overline{x}^2 = \frac{1}{9}(1+4+9+\ldots+81) - 5^2 = 6,\widehat{6}$$

$$\operatorname{Cov}(x,y) = \left(\frac{1}{N}\sum x_i y_j n_{ij}\right) - \overline{x}\,\overline{y} = \frac{1}{9}(31+80+108+\ldots+333) - 5\cdot 36,\widehat{6} = 3,\widehat{1}$$

$$\beta = \frac{\operatorname{Cov}(x,y)}{\operatorname{Var}(x)} = \frac{3,\widehat{1}}{6,\widehat{6}} = 0,4\widehat{6} \qquad \alpha = \overline{y} - \beta\overline{x} = 36,\widehat{6} - 0,4\widehat{6}\cdot 5 = 34,\widehat{3}$$

La recta de regresión es: $\hat{y}_t = 34,\widehat{3} + 0,4\widehat{6}x_t$

A partir de esta recta sustituimos los valores de *x* para obtener la tendencia. Esto lo hemos colocado en la última columna de la tabla.

Para medir la bondad del ajuste, calcularemos el coeficiente de correlación simple. (Necesitamos la varianza de *y* para después emplear la fórmula correspondiente.)

$$\operatorname{Var}(y) = \left(\frac{1}{N}\sum_{i=1}^{j} y_i^2 n_i\right) - \overline{y}^2 = \frac{1}{9}(961+1600+\ldots+1369) - 36,\widehat{6}^2 = 13,\widehat{7}$$

$$\rho = \frac{\operatorname{Cov}(x,y)}{\sigma_x \sigma_y} = \frac{3,\widehat{1}}{\sqrt{6,\widehat{6}}\cdot\sqrt{13,\widehat{7}}} = 0,3246$$

La recta no se ajusta muy bien a la serie, ya que el coeficiente de correlación es más cercano a 0 que a 1.

EJERCICIO 4

Apartado A

$$\text{Imag}_0^t = \frac{\sum_{i=1}^{M} p_{it}}{\sum_{i=1}^{M} p_{i0}} \cdot 100 = \frac{45+27+15}{42+25+12} \cdot 100 = 110,127$$

Apartado B

$$L_{p_t} = \frac{\sum_{i=1}^{M} p_{it} q_{i0}}{\sum_{i=1}^{M} p_{i0} q_{i0}} \cdot 100 = \frac{45 \cdot 5 + 27 \cdot 7 + 15 \cdot 11}{42 \cdot 5 + 25 \cdot 7 + 11 \cdot 11} \cdot 100 = 111,992$$

Apartado C

$$P_{q_t} = \frac{\sum_{i=1}^{M} p_{it} q_{it}}{\sum_{i=1}^{M} p_{it} q_{i0}} \cdot 100 = \frac{45 \cdot 3 + 27 \cdot 8 + 15 \cdot 13}{45 \cdot 5 + 27 \cdot 7 + 15 \cdot 11} \cdot 100 = 94,3$$

Apartado D

Los salarios en 2011, que aumentaron un 3,5 % respecto a 2010, son:

Técnicos	46,575
Administrativos	27,945
Subalternos	15,525

Calculamos el índice de Laspeyres de precios (salarios) del año 2011 con base en 2008.

$$L_{p_t} = \frac{\sum_{i=1}^{M} p_{it} q_{i0}}{\sum_{i=1}^{M} p_{i0} q_{i0}} \cdot 100 = \frac{46,575 \cdot 5 + 27,945 \cdot 7 + 15,525 \cdot 11}{42 \cdot 5 + 25 \cdot 7 + 11 \cdot 11} \cdot 100 = 115,912$$

EJERCICIO 5

La renta de las personas de 40 años se distribuye según esta tabla:

Renta	$\mathbf{x_i}$	$\mathbf{n_i}$	$\mathbf{N_i}$
15 – 20	17,5	5+4 = 11	11
20 – 45	32,5	6+2 = 8	19
45 – 60	52,5	2+1 = 3	22
		N = 22	

<u>Cálculo de la media:</u>

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{j} x_i n_i = \frac{1}{22}\left(17,5\cdot 11 + 32,5\cdot 8 + 52,5\cdot 3\right) = 27,\overset{\frown}{72}$$

<u>Cálculo de la mediana:</u>

Obtenemos el intervalo mediano, que es aquel cuya frecuencia acumulada alcance la mitad de la población, es decir, Ni>=11. En este caso el intervalo mediano es 15-20.

El valor exacto de la mediana lo calculamos con la fórmula correspondiente:

$$Me = E_{i-1} + A_i\left(\frac{0,5\cdot N - N_{i-1}}{n_i}\right) = 15 + 5\left(\frac{0,5\cdot 22 - 0}{11}\right) = 15 + 5\cdot 1 = 20$$

<u>Estudio de la simetría:</u>

La distribución tiene asimetría positiva, ya que la media es mayor que la mediana.