Solución examen de junio de 2004. ESTADISTICA (ITIG)

11 de junio de 2004

1) Un concesionario de coches hace un estudio sobre el tipo de coches que se compran los españoles en función de su salario (expresado en euros) cuyos resultados finales se resumen en: *(1,5 puntos)*

Coche/salario	0-1000	1000-2000	2000-4000	4000-5000
Utilitarios	0,05	0,12	0,12	0,02
Berlinas	0,02	0,2	0,25	0,02
Lujo	0,01	0,05	0,07	0,07

a) ¿Cuál es la distribución de los coches condicionada a que el salario sea inferior a 2000 euros?
b) ¿Cuál es el salario medio? ¿y el de los que compran coches de lujo?
c) Entre 190,18 y 4629,82 ¿al menos qué porcentaje de la muestra está incluido?
d) Si nos dieran la variable Y ≡ salario × 166 + 100 cual será la media y la desviación típica de y (calcúlela a partir de la media y desviación típica de la variable original)
e) ¿Son las dos variables independientes?

						Respuesta a	Respuesta b
marca de clas	500	1500	3500	4500			
Coche/salario	0-1000	1000-2000	2000-4000	4000-5000	Marginal	Distribución condionada a salario<2000€	Salario medio
Utilitario	5%	12%	12%	2%	31%	37,8%	2306,5
Berlinas	2%	20%	25%	2%	49%	48,9%	2602,0
Lujo	1%	5%	7%	7%	20%	13,3%	3200,0
Marginal	8%	37%	44%	11%	100%		2630,0
Distribución condionada coche de lujo	5,0%	25,0%	35,0%	35,0%			

c) Por Tchebychev se sabe que en la MEDIA +/- K VECES LA DESVIACIÓN TÍPICA SE ENCUENTRA AL MENOS EL $(1-1/K^2)$% DE LA MUESTRA. Como el salario medio es 2630 y la desviación típica es 1426. Por tanto la distancia k*desviación tipica es = 2630- 190,18 -→ k=1,71 → % incluido al menos un 65,8%

d) La media de y será = 166 x (media del salario) + 100 → media de y = 166 x 2630 + 100
La desviación típica de y será = 166 x (desviación típica del salario) → DT(y)=166x1426=236716.

e) Sí, como están calculadas las marginales basta coger cualquier cruce y comprobar que la frecuencia relativa no coincide con el producto de las marginales. Por ejemplo para berlinas de 2000-4000 €
$0,25 \neq 0,44 * 0,49$

2) Se ha contratado con tres compañías el suministro de ventiladores para la incorporación a los Ordenadores Personales que se venden de una determinada marca. A la primera compañía se le compran 10 unidades, a la segunda 40 y a la tercera 50. Si la probabilidad de fallo es 0.12, en el primer grupo 0.08 en el segundo y 0.07 del resto: *(1,5 puntos)*

a) ¿Cual es la probabilidad de que funcione un ventilador cualquiera?
b) Sabiendo que un ventilador funciona, ¿cuál es la probabilidad de que sea de la primera empresa?
c) ¿Y del resto?

El apartado a se resuelve aplicando el teorema de la probabilidad total, el b) el teorema de Bayes y el c) dándose cuenta que es el complementario del apartado b) por lo que no hay que volver a aplicar Bayes.

Los resultados son:

empresas	1	2	3	
ventas	10	40	50	
P fallo	0,12	0,08	0,07	
probabilidad de funcionar	0,088	0,368	0,465	0,921
probabilidad de ser de la primera empresa sabiendo que funciona	0,095548			
probabilidad de ser del resto	0,904452			

3) En el Centro de Atención Telefónica de una agencia de viajes se sabe que el número de llamadas perdidas se distribuye según una Poisson de media 2 a la hora. Así, si cada grupo tiene 100 operadores, cual es la probabilidad de que haya al menos 50 operadores que pierdan más de tres llamadas cada dos horas? *(1,5 puntos)*

λ = 2 llamadas/hora, pero como se pide la probabilidad de perder un número de llamadas cada dos horas, habrá que transformar a λ'= 4 llamadas/2 horas.

De ahí se puede calcular la probabilidad de que una persona pierda más de tres llamadas en dos horas:

$$P(x>3)=1-P(x\le 3)=1-\big(P(x=0)+P(x=1)+P(x=2)+P(x=3)\big)$$

Como $P(x=r)=\dfrac{\lambda^r e^{-\lambda}}{r!}$ ahora bien, utilizando λ', se tiene:

$$P(x>3)=1-\left(\frac{4^0e^{-4}}{1}+\frac{4^1e^{-4}}{1}+\frac{4^2e^{-4}}{2\times 1}+\frac{4^3e^{-4}}{3\times 2\times 1}\right)=0.5665$$

Ahora, como se quiere calcular la probabilidad de que, en un grupo de 100 operadores, haya más de 50 con esta probabilidad: es una distribución binomial: B(100,0.5665)

Como se pide calcular P(x>50) y npq>5, se puede aproximar a una Normal, cuya media es np=56.65 y varianza npq=24.557775, por tanto, la desviación típica será σ=4.956

De esta forma y utilizando también la corrección por continuidad:

$$P(x>50)=P\left(z>\frac{50.5-56.65}{4.956}\right)=P(z>-1.24)=P(z<1.24)=0.8925$$

El valor definitivo, lógicamente, se obtiene de las tablas de la Normal.

4) Una fábrica de teclados para ordenadores produce un modelo A cuya duración en horas sigue una distribución normal de media y desviaciones típicas desconocidas. Para estimar dichos parámetros, la empresa analiza una muestra de 20 teclados obteniéndose los siguientes resultados: *(2 puntos)*

   Media de la muestra = 5959 horas,

   Cuasidesviación típica = 100 h

   a) Construir un intervalo de confianza para la media de horas de funcionamiento con una confianza del 95%
   b) ¿Qué tamaño deberá tener la muestra para que, con una confianza del 99%, la proporción muestral resultante diste de la poblacional un máximo de 20 h?

a) El intervalo de confianza para la media, es: $\mu \in \left\{ \overline{x} \pm t_{n-1;\alpha/2} \cdot \frac{\hat{s}}{\sqrt{n}} \right\}$

Sustituyendo, resulta: $\mu \in [5912.27, 6005.73]$

b) La distancia entre la proporción muestral y la poblacional en este caso es $L = t_{n-1;\alpha/2} \frac{\hat{s}}{\sqrt{n}}$

Como la muestra tendrá que ser grande, $L = z_{\alpha/2} \frac{\hat{s}}{\sqrt{n}}$, siendo L=20, se despeja n y resulta n=166.4

Se comprueba que, efectivamente, es un número grande y por tanto, la solución es: la muestra deberá tener un tamaño de 167 para que la proporción muestral resultante diste de la poblacional un máximo de 20 h.

5) En una encuesta sobre la intención de voto en las elecciones europeas del 13 de junio se entrevistan a 1000 personas de los cuales 430 afirman que votarán al partido A. ¿Se puede asegurar con una confianza del 95%, que el partido A obtendrá más del 50% de los votos? *(1 punto)*

La proporción de la muestra es $\hat{p} = 0.43$

El contraste de hipótesis más sencillo para asegurar si el partido obtendrá más del 50% de los votos es:

H_0: $p \leq 0.5$ (p_0)
H_1: $p > 0.5$ (p_0)

Se calcula el estadístico de contraste:

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} = \frac{0.43 - 0.5}{\sqrt{0.5 \cdot 0.5 / 1000}} = -4.43$$

Como z_α = 1.64, la región de rechazo es $z_0 > 1.64$. Por tanto, $z_0 < 1.64$ y, entonces, se está la región de aceptación y no se puede asegurar con una confianza del 95%, que el partido A obtendrá más del 50% de los votos.

6) Un Estudiante de económicas quiere analiza la relación entre los gastos y los ingresos (medidos en €) de 30 personas. Para ello, dispone de la información que se resume en la siguiente tabla: *(2,5 puntos)*

Nº Casos	Ingresos	Gastos	Ingresos^2	Gastos^2	Ingresos·Gastos
30	5850	5580	1365500	1181720	1267900

a) *Estimar los parámetros de la recta de regresión, e interpretarlos.*

casos	ingresos	Gastos	ing*ing	Gas*gas	ing*gas
30	5850	5580	1365500	1181720	1267900
medias	**195**	**186**			
var	**7491,67**	**4794,67**			
covar	5993,33				
b	0,8				
a	30				
r	1				
R^2	1				

La pendiente será b y nos indica en cuánto variarán los gastos ante incrementos de los ingresos, al ser mayor que cero indica que variarán en el mismo sentido
La constante a indica cual será el gasto en ausencia de ingresos

b) *¿Cual es el coeficiente de correlación y el de determinación? ¿Ofrecen la misma información?*
Están puestos en la tabla superior.
El de correlación indica el grado de relación lineal y el sentido de la misa, el de determinación sólo indica el ajuste de la recta de regresión, sin ofrecer información sobre el sentido del ajuste.

c) *Según la recta calculada, ¿en cuantos miles de € han tenido que aumentar los ingresos si los gastos lo han hecho en 100€?*
Se sabe que $\Delta Y = b * \Delta x$, Si $\Delta x = 100 \rightarrow \Delta Y = 0{,}8 * 100 = 800€ = 0{,}8$ miles de €

d) *¿Como cambia la recta de regresión si la variable "y" se expresa en miles?*
Hay que demostrar cómo se alteran los coeficientes a y b ante el cambio de variables utilizando los coeficientes de origen, no intentando recalcular los datos de nuevo

e) *¿Y que ocurre con el grado de ajuste?*
Hay que demostrar que el ajuste no cambia ante transformaciones lineales

f) ¿Qué transformación de las variables conseguirá que el sumatorio de los residuos de la recta de regresión no sea cero?
Ninguna, pues la recta de regresión por su construcción, véase la primera condición de minimización para obtener los estimadores mínimos cuadráticos, la suma de residuos siempre será cero

g) *¿Coincide el gráfico representado a continuación con la recta de regresión planteada?*
No puede coincidir, pues el ajuste que indica la recta de regresión es perfecto, por lo que no puede existir ningún punto fuera de la recta.