UAM UNIVERSIDAD AUTONOMA DE MADRID

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

GRADO GESTIÓN AERONÁUTICA: EXAMEN ESTADÍSTICA TEÓRICA

9 de Enero de 2015. E-7. Aula 104

1.- La función de densidad de una variable aleatoria es:

$$f(x) = \begin{cases} a x^2 + b & 0 < x < 2 \\ 0 & \text{en el resto} \end{cases} \qquad \text{sabiendo que } P\left[\frac{1}{2} < x < 1\right] = 0{,}1666 .$$

Determinar a y b.

2.- El peso (en gramos) de las cajas de cereales de una determinada marca sigue una distribución $N(\mu, 5)$. Se han tomado los pesos de 16 cajas seleccionadas aleatoriamente, y los resultados obtenidos han sido:

506, 508, 499, 503, 504, 510, 497, 512, 514, 505, 493, 496, 506, 502, 509, 496.

a) Obtener el intervalo de confianza del 95% para la media poblacional.

b) Determinar cuál sería el tamaño muestral necesario para conseguir, con un 95% de confianza, un intervalo de longitud igual a 2 gramos.

c) Suponiendo ahora que σ es desconocida, calcular el intervalo de confianza para la media al 99%.

3.- La probabilidad de que un banco reciba un cheque sin fondos es 0.01

a) Si en una hora reciben 20 cheques, ¿cuál es la probabilidad de que tenga algún cheque sin fondos?

b) El banco dispone de 12 sucursales en la ciudad, ¿cuál es la probabilidad de que al menos cuatro sucursales reciban algún cheque sin fondos?

c) La media del valor de los cheques sin fondos es de 600 euros. Sabiendo que el banco trabaja 6 horas diarias, ¿qué cantidad no se espera pagar?

d) Si se computasen los 500 primeros cheques, ¿cuál es la probabilidad de recibir entre 3 y 6 (inclusive) cheques sin fondos?

UNIVERSIDAD AUTONOMA DE MADRID

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

UNICAMENTE PARA USO ESCOLAR (IMPRESO EN PAPEL RECICLADO)

4.- Con un nivel de significación del 4,72%, se desea contrastar la hipótesis nula de igualdad de medias de dos poblaciones $N(\mu_1, 4)$ y $N(\mu_2, 4{,}5)$. Para ello, se han tomado dos muestras aleatorias simples e independientes, respectivamente, obteniéndose los siguientes valores:

x_i	20,4	10,2	7,3	12,8	13,4	9,4
y_j	19,8	9,7	14,6	15.7	8,4	

5.- Una empresa multinacional desea conocer si con el 95% de fiabilidad existen diferencias significativas entre sus trabajadores en distintos países en el grado de satisfacción en el trabajo. Para ello se toman muestran aleatorias simples de trabajadores, obteniendo los siguientes resultados:

	Satisfacción en el trabajo			
	Muy satisfecho	Satisfecho	Insatisfecho	Muy insatisfecho
A	200 **(242,86)**	300 **(285,71)**	300	100 **(114,29)**
B	300	400	350 **(342,86)**	150 **(152,38)**
C	350 **(283,33)**	300 **(333,33)**	250	150

6.-

a) Si dos sucesos A y B tienen la misma probabilidad igual a 0,3 y son independientes, ¿cuánto vale la probabilidad de A condicionada a B?. ¿Y la probabilidad de $A \cap B$?

b) Dada una variable aleatoria X que se distribuye como una $B(30; 0,2)$ y otra variable aleatoria Y que se distribuye como una $B(150; 0,3)$. ¿Si X e Y son independientes, se puede afirmar que la variable $X + Y$ se distribuye como una $B(180; 0,5)$?. Justifica tu respuesta.

c) Una variable aleatoria χ^2 tiene 10 grados de libertad. Hallar la media, la varianza y la probabilidad de que dicha variable aleatoria sea mayor que 9,342

1.- La función de densidad de una variable aleatoria es:

$$f(x)=\begin{cases} a x^2+b & 0<x<2 \\ 0 & \text{en el resto}\end{cases} \qquad \text{sabiendo que } P\left[\frac{1}{2}<x<1\right]=0{,}1666\,.$$

Determinar a y b.

Solución:

Hay que calcular dos parámetros (a y b), por lo que se necesitan dos ecuaciones:

- Por ser función de densidad:

$$1=\int_0^2 f(x)\,dx=\int_0^2 (a x^2+b)\,dx=\left[a\frac{x^3}{3}+b x\right]_0^2=\frac{8a}{3}+2b \quad\mapsto\quad 8a+6b=3$$

- $$P\left[\frac{1}{2}\le x\le 1\right]=\int_{1/2}^1 f(x)\,dx=\int_{1/2}^1 (a x^2+b)\,dx=\left[a\frac{x^3}{3}+b x\right]_{1/2}^1=0{,}1666\,,\quad \text{con lo que:}$$

$$\left[a\frac{x^3}{3}+b x\right]_{1/2}^1=\left[\frac{a}{3}+b\right]-\left[\frac{a}{24}+\frac{b}{2}\right]=\frac{7a}{24}+\frac{b}{2}=0{,}1666 \quad\mapsto\quad 7a+12b\approx 4$$

en consecuencia,

$$\left.\begin{array}{r} 8a+6b=3 \\ 7a+12b=4 \end{array}\right\} \;\mapsto\; \left.\begin{array}{r} 16a+12b=6 \\ 7a+12b=4 \end{array}\right\} \;\to\; \boxed{a=\frac{2}{9}=0{,}22} \quad 6b=3-\frac{16}{9}=\frac{11}{9} \;\mapsto\; \boxed{b=\frac{11}{54}=0{,}20}$$

2.- El peso (en gramos) de las cajas de cereales de una determinada marca sigue una distribución $N(\mu,5)$. Se han tomado los pesos de 16 cajas seleccionadas aleatoriamente, y los resultados obtenidos han sido:

506, 508, 499, 503, 504, 510, 497, 512, 514, 505, 493, 496, 506, 502, 509, 496.

a) Obtener el intervalo de confianza del 95% para la media poblacional.

b) Determinar cuál sería el tamaño muestral necesario para conseguir, con un 95% de confianza, un intervalo de longitud igual a 2 gramos.

c) Suponiendo ahora que σ es desconocida, calcular el intervalo de confianza para la media al 99%.

Solución:

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

a) Se trata de construir un intervalo de confianza para la media poblacional μ de varianza conocida $\sigma^2 = 25$. El intervalo de confianza de nivel $1-\alpha$ viene dado por:

$$I_{1-\alpha}(\mu) = \Big[\underbrace{\bar{x}}_{\text{media muestral}} \pm \overbrace{z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}}^{\text{Error muestral}} \Big] \quad \begin{cases} L_{1-\alpha} = 2\, z_{\alpha/2}\,\dfrac{\sigma}{\sqrt{n}} \;\mapsto\; n = \left(\dfrac{2\, z_{\alpha/2}\,\sigma}{\text{longitud}}\right)^2 \quad L = \text{longitud o amplitud} \\ \text{Error muestral} = z_{\alpha/2}\,\dfrac{\sigma}{\sqrt{n}} \end{cases}$$

$$\bar{x} = \frac{\sum_{i=1}^{16} x_i}{16} = 503{,}75 \qquad \{1-\alpha = 0{,}95 \quad \alpha = 0{,}05 \quad \alpha/2 = 0{,}025 \qquad z_{\alpha/2} = 1{,}96$$

El intervalo de confianza solicitado será:

$$I_{0,95}(\mu) = \left[503{,}75 \pm 1{,}96\,\frac{5}{\sqrt{16}}\right] = \left[503{,}75 - 1{,}96\,\frac{5}{\sqrt{16}}\,,\, 503{,}75 + 1{,}96\,\frac{5}{\sqrt{16}}\right]$$

$$I_{0,95}(\mu) = [\,501{,}30\,;\; 506{,}20\,] \equiv P[\,501{,}30 \le \mu \le 506{,}20\,] = 0{,}95 = 1-\alpha$$

b) La amplitud o longitud vendrá dado por la fórmula: $I_{1-\alpha}(\mu) = \left[\bar{x} \pm z_{\alpha/2}\,\dfrac{\sigma}{\sqrt{n}}\right]$

$$\begin{pmatrix}\text{amplitud o}\\ \text{longitud}\end{pmatrix} = \left(\bar{x} + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\right) - \left(\bar{x} - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\right) = 2\, z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} \;\mapsto\; n = \left(\frac{2\, z_{\alpha/2}\,\sigma}{\text{amplitud}}\right)^2$$

siendo, $n = \left(\dfrac{2 . 1{,}96 . 5}{2}\right)^2 \approx 96$ cajas de cereales

c) Se trata de construir un intervalo de confianza para la media poblacional μ de varianza poblacional desconocida, con muestras pequeñas ($n \le 30$). El intervalo de confianza de nivel $(1-\alpha)$, viene dado por:

$$I_{1-\alpha}(\mu) = \left[\bar{x} \pm t_{(\alpha/2),n-1}\,\frac{s_x}{\sqrt{n}}\right] \qquad 1-\alpha = 0{,}99 \quad \alpha = 0{,}01 \quad t_{0,005;15} = 2{,}947$$

cuasivarianza muestral: $s_x^2 = \dfrac{\sum_{i=1}^{16}(x_i - \bar{x})^2}{15} = 36{,}037 \;\mapsto\; s_x \approx 6$ cuasidesviación típica

El intervalo de confianza solicitado será:

$$I_{0,99}(\mu) = \left[503,75 \pm 2,947\,\frac{6}{\sqrt{16}}\right] = \left[503,75 - 2,947\,\frac{6}{\sqrt{16}}\,,\,503,75 + 2,947\,\frac{6}{\sqrt{16}}\right]$$

$$I_{0,99}(\mu) = [\,499,33\,;\;508,17\,] \;\equiv\; P[\,499,33 \le \mu \le 508,17\,] = 0,99 = 1-\alpha$$

3.- La probabilidad de que un banco reciba un cheque sin fondos es 0.01

a) Si en una hora reciben 20 cheques, ¿cuál es la probabilidad de que tenga algún cheque sin fondos?

b) El banco dispone de 12 sucursales en la ciudad, ¿cuál es la probabilidad de que al menos cuatro sucursales reciban algún cheque sin fondos?

c) La media del valor de los cheques sin fondos es de 600 euros. Sabiendo que el banco trabaja 6 horas diarias, ¿qué cantidad no se espera pagar?

d) Si se computasen los 500 primeros cheques, ¿cuál es la probabilidad de recibir entre 3 y 6 (inclusive) cheques sin fondos?

Solución:

a) X = "número de cheques sin fondos" sigue una distribución binomial $X \sim B(20,\,0,01)$

$$P[X \ge 1] = 1 - P[X < 1] = 1 - P[X = 0] = 1 - \binom{20}{0}.0,01^0\,.0,99^{20} = 1 - 0,980 = 0,182$$

b) Y = "número de sucursales que reciben al menos 1 cheque sin fondos"
$Y \sim B(12,\,0,182)$

$$P[Y \ge 4] = 1 - P[Y < 4] = 1 - \big[P[X=0] + P[X=1] + P[X=2] + P[X=3]\big] =$$

$$= 1 - \left[\binom{12}{0}.0,182^0\,.0,818^{12} + \binom{12}{1}.0,182^1\,.0,818^{11} + \binom{12}{2}.0,182^2\,.0,818^{10} + \binom{12}{3}.0,182^3\,.0,818^9\right] =$$

$$= 1 - [0,0897 + 0,2396 + 0,2932 + 0,2174] = 0,16$$

c) $\dfrac{1\,\text{hora}}{20\text{ cheques}} = \dfrac{6\text{ horas}}{n\text{ cheques}} \;\mapsto\; n = 120\text{ cheques}$

Los cheques sin fondos esperados: $\mu = E(X) = n.\,p = 120\,.\,0,01 = 1,2$ cheques

En consecuencia, se espera no pagar $1,2\,.\,600 = 720$ euros

d) U = "número de cheques sin fondos computados" donde $U \sim b(500, 0{,}01)$, que al ser

$n.p = 500.0{,}01 = 5$ se aproxima a una distribución de Poisson de parámetro $P[\lambda = 5]$

$P[3 \le U \le 6] = P[U=3] + P[U=4] + P[U=5] + P[U=6] =$

$$= \left[\frac{5^3}{3!} + \frac{5^4}{4!} + \frac{5^5}{5!} + \frac{5^6}{6!}\right].e^{-5} = [20{,}833 + 26{,}042 + 26{,}042 + 21{,}701].e^{-5} = 0{,}6375$$

4.- Con un nivel de significación del 4,72%, se desea contrastar la hipótesis nula de igualdad de medias de dos poblaciones $N(\mu_1, 4)$ y $N(\mu_2, 4{,}5)$. Para ello, se han tomado dos muestras aleatorias simples e independientes, respectivamente, obteniéndose los siguientes valores:

x_i	20,4	10,2	7,3	12,8	13,4	9,4
y_j	19,8	9,7	14,6	15.7	8,4	

Solución:

En el contraste bilateral se establecen las hipótesis: $\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases}$

Regla de decisión $\begin{cases} \text{Si } |\overline{x} - \overline{y}| > k \text{ se rechaza } H_0 \\ \text{Si } |\overline{x} - \overline{y}| \le k \text{ se acepta } H_0 \end{cases}$

Para analizar el contraste, se realizan los cálculos muestrales:

$$\overline{x} = \frac{\sum_{i=1}^{6} x_i}{6} = 12{,}25 \qquad \overline{y} = \frac{\sum_{j=1}^{5} y_j}{5} = 13{,}64 \qquad n_1 = 6 \qquad n_2 = 5$$

La región crítica de dos colas $|\overline{x} - \overline{y}| > k$ es función de la diferencia de las medias muestrales. En esta línea, las distribuciones en el muestreo de las medias son:

$$\overline{x} \sim N\left[\mu_1, \frac{\sigma_1}{\sqrt{n_1}}\right], \ \overline{y} \sim N\left[\mu_2, \frac{\sigma_2}{\sqrt{n_2}}\right]$$

Bajo la hipótesis nula $H_0 : \mu_1 - \mu_2 = 0$, la diferencia de medias muestrales se distribuye:

$$(\overline{x}-\overline{y}) \sim N\left[0,\ \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right] \equiv N\left[0,\ \sqrt{\frac{4^2}{6}+\frac{4,5^2}{5}}\right] = N\left[0\,,\ 2,59\right]$$

Se determina el valor de k mediante el nivel de significación α:

$$\alpha = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ cierta}\right] = P\left[\,|\overline{x}-\overline{y}| > k \ / \ H_0 : \mu_1 - \mu_2 = 0\,\right] =$$

$$= P\left[\left|\frac{(\overline{x}-\overline{y})-0}{2,59}\right| > K\right] = P\left[\left(\frac{\overline{x}-\overline{y}}{2,59} < -K\right) \cup \left(\frac{\overline{x}-\overline{y}}{2,59} > K\right)\right] =$$

$$= P\left[\frac{\overline{x}-\overline{y}}{2,59} < -K\right] + P\left[\frac{\overline{x}-\overline{y}}{2,59} > K\right] = \frac{\alpha}{2}+\frac{\alpha}{2} = 0,05 \ \text{ por simetría } N(0,1)$$

La región crítica es $\left|\dfrac{\overline{x}-\overline{y}}{2,59}\right| > 1,995 = z_{0,0236} \quad \mapsto \quad \begin{cases} \overline{x}-\overline{y} > 5,17 \\ \overline{x}-\overline{y} < -5,17 \end{cases}$

En consecuencia, la región de aceptación: $-5,17 \le \overline{x}-\overline{y} \le 5,17$

La evidencia empírica $|\overline{x}-\overline{y}| = |12,25 - 13,64| = 1,39$, valor que se encuentra en la región de aceptación por lo que se acepta la hipótesis nula de igualdad de medias, con un nivel de significación del 4,72%.

Análogamente, se podría haber resuelto considerando la región de rechazo de la hipótesis nula H_0:

$$R = \left\{ |\overline{x}-\overline{y}| > z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \right\} \mapsto R = \left\{ |12,25-13,64| > (1,995)\sqrt{\frac{4^2}{6}+\frac{4,5^2}{5}} \right\}$$

La región de rechazo de la hipótesis nula no se cumple, $R \neq \{1,39 > 5,17\}$, concluyendo que existe igualdad entre las medias poblacionales.

Abscisas	Áreas
1,98 – 1,99	0,0239 – 0,0233
x – 1,99	0,0236 – 0,0233

$$x = 1,99 + \frac{0,01 \,.\, 0,0003}{0,0006} = 1,995$$

Asignatura.. Grupo......................................

Apellidos .... Nombre......................................

Ejercicio del día ....

5.- Una empresa multinacional desea conocer con un 95% de fiabilidad si existen diferencias significativas entre sus trabajadores en distintos países en el grado de satisfacción en el trabajo. Para ello se toman muestran aleatorias simples de trabajadores, obteniendo los siguientes resultados:

	Satisfacción en el trabajo			
	Muy satisfecho	Satisfecho	Insatisfecho	Muy insatisfecho
A	200 **(242,86)**	300 **(285,71)**	300	100 **(114,29)**
B	300	400	350 **(342,86)**	150 **(152,38)**
C	350 **(283,33)**	300 **(333,33)**	250	150

Solución:

La hipótesis nula H_0: 'Las proporciones de los trabajadores con los distintos grados de satisfacción son iguales en los tres países'

Se acepta H_0: $\chi^2_{(k-1)\,.\,(m-1)} = \sum_{i=1}^{k}\sum_{j=1}^{m} \frac{(n_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i=1}^{k}\sum_{j=1}^{m} \frac{n_{ij}^2}{e_{ij}} - n < \chi^2_{\alpha\,;\ (k-1)\,.\,(m-1)}$

Región de rechazo de la hipótesis nula: $R_{rechazo} = \left\{ \chi^2_{(k-1)\,.\,(m-1)} \geq \chi^2_{\alpha\,;\ (k-1)\,.\,(m-1)} \right\}$

Se forma la tabla de contingencia 3 x 4 donde cada frecuencia observada

$(n_{ij})_{i=1,\cdots,k\,;\ j=1,\cdots,m}$ tiene una frecuencia teórica o esperada $e_{ij} = \dfrac{n_{i\bullet} \times n_{\bullet j}}{n}$

Asignatura.. Grupo..................................

Apellidos .... Nombre..................................

Ejercicio del día ....

	Satisfacción en el trabajo				Total
	Muy satisfecho	Satisfecho	Insatisfecho	Muy insatisfecho	
A	200 **($e_{11} = 242,86$)**	300 **($e_{12} = 285,71$)**	300 **($e_{13} = 257,14$)**	100 **($e_{14} = 114,29$)**	900 **(900)**
B	300 **($e_{21} = 323,81$)**	400 **($e_{22} = 380,95$)**	350 **($e_{23} = 342,86$)**	150 **($e_{24} = 152,38$)**	1200 **(1200)**
C	350 **($e_{31} = 283,33$)**	300 **($e_{32} = 333,33$)**	250 **($e_{33} = 300$)**	150 **($e_{34} = 133,33$)**	1050 **(1050)**
Total	850	1000	900	400	3150

Estadístico observado: $\chi^2_{(3-1)\,.\,(4-1)} = \sum_{i=1}^{3}\sum_{j=1}^{4} \frac{(n_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i=1}^{3}\sum_{j=1}^{4} \frac{n_{ij}^2}{e_{ij}} - n =$

$$= \frac{200^2}{242,86} + \frac{300^2}{285,71} + \frac{300^2}{257,14} + \frac{100^2}{114,29} + \frac{300^2}{323,81} + \frac{400^2}{380,95} + \frac{350^2}{342,86} + \frac{150^2}{152,38} +$$

$$+ \frac{350^2}{283,33} + \frac{300^2}{333,33} + \frac{250^2}{300} + \frac{150^2}{133,33} - 3150 = 49,55$$

Estadístico teórico: $\chi^2_{0,05\ ;\ (3-1).(4-1)} = \chi^2_{0,05\ ;\ 6} = 12,592$

Como $\chi^2_6 = 49,55 > 12,592 = \chi^2_{0,05\ ;\ 6}$ se rechaza la hipótesis nula de homogeneidad de las tres muestras. Es decir, la satisfacción en el trabajo de los empleados de los tres países es significativamente distinta

UNIVERSIDAD AUTONOMA DE MADRID

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

(IMPRESO EN PAPEL RECICLADO) UNICAMENTE PARA USO ESCOLAR

6.-

a) Si dos sucesos A y B tienen la misma probabilidad igual a 0,3 y son independientes, ¿cuánto vale la probabilidad de A condicionada a B?. ¿Y la probabilidad de $A \cap B$?

b) Dada una variable aleatoria X que se distribuye como una $B(30;0,2)$ y otra variable aleatoria Y que se distribuye como una $B(150;0,3)$. ¿Si X e Y son independientes, se puede afirmar que la variable $X+Y$ se distribuye como una $B(180;0,5)$?. Justifica tu respuesta.

c) Una variable aleatoria χ^2 tiene 10 grados de libertad. Hallar la media, la varianza y la probabilidad de que dicha variable aleatoria sea mayor que 9,342

Solución

a) $P(A/B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(A).P(B)}{P(B)} = P(A) = 0,3 \qquad P(A \cap B) = P(A).P(B) = 0,3\,.\,0,3 = 0,09$

b) Dadas k variables aleatorias independientes $(X_1, X_2, \cdots, X_k)$ que se distribuyen según una binomial $B(n_i\,,\,p)$, la suma de las k variables es también una distribución binomial de parámetros $\left[n_1 + n_2 + \cdots + n_k\,,\,p\right]$, es decir:

$$Y = \sum_{i=1}^{k} X_i \sim B\left(\sum_{i=1}^{k} n_i\,,\,p\right) \qquad \text{(propiedad reproductiva)}$$

En esta línea, si las variables aleatorias dadas tuvieran la misma probabilidad, por ejemplo, $X \sim B(30;0,2)$ e $Y \sim B(150;0,2)$, entonces $X+Y \sim B(180;0,2)$.

La cuestión que se plantea es distinta, las variables X e Y tienen distinta probabilidad, con lo que se determina que $X+Y$ **no** siguen una distribución binomial $B(180;0,5)$

c) La media y varianza de la χ^2 de Pearson: $\mu = 10 \quad \sigma^2 = 2\,.\,10 = 20$

$$P\left[\chi^2_{10} > 9,342\right] = 0,5$$