Examen de Estadística II. 02-02-2009.

1. X, ventas mensuales por cliente en euros.
$\mu_x = 500$, $\sigma_x = 85$.

Asumiendo que las ventas por cliente son independientes y equidistribuidas podemos considerar $(x_1, \ldots, x_{55})$ como una m.a.s. Además, como el tamaño muestral $n = 55$ es suficientemente grande estamos en condiciones de aplicar el teorema central del límite para obtener que $\sum_{i=1}^{n} X_i \dot\sim N(n\mu, n\sigma^2) = N(27500, 397375)$.

Podemos considerar el salario mensual en euros como una variable aleatoria suma de la parte fija y la parte variable correspondiente a los 55 clientes, $Y = 800 + 0.05 \sum_{i=1}^{55} X_i$. Se trata de una combinación lineal de una variable aproximadamente Normal por lo que también $Y \dot\sim N(\mu_y = 800 + (0.05)(27500) = 2175, \sigma_y^2 = (0.05)^2(397375) = 993.4375)$.

$$p\,(Y > 2150) = p\left(Z > \frac{2150 - 2175}{\sqrt{993.4375}}\right) = p(Z > -0.7932) = \underline{0.7861},$$

donde $Z = \frac{Y - \mu_y}{\sigma_y} \dot\sim N(0, 1)$.

2. X, tiempo de fabricación de un producto en minutos.
$X \sim Ex(\theta)$, $f(x) = \theta \exp(-\theta x)$, $x > 0$, $\theta > 0$.
$(x_1, \ldots, x_{25})$ m.a.s. tal que $\overline{x} = 16$.

El estimador máximo verosímil de θ se obtiene maximizando la función de verosimilitud Exponencial,

$$l(\theta) = \prod_{i=1}^{25} f(x_i; \theta) = \prod_{i=1}^{25} \theta \exp(-\theta x_i) = \theta^{25} \exp\left(-\theta \sum_{i=1}^{25} x_i\right).$$

De modo que la log-verosimilitud es $\log l(\theta) = 25 \log \theta - \theta \left(\sum_{i=1}^{25} x_i\right)$ y su primera derivada resulta $\frac{d \log l(\theta)}{d\theta} = \frac{25}{\theta} - \sum_{i=1}^{25} x_i$, que igualada a cero proporciona la solución $\hat\theta = \frac{25}{\sum_{i=1}^{25} x_i} = \frac{1}{\overline{x}}$. Efectivamente, el inverso de la media muestral es el máximo de $\log l(\theta)$ (y por la inyectividad de la función logarítmica, también de $l(\theta)$), pues hace negativa su segunda derivada, $\frac{d^2 \log l(\theta)}{d\theta^2} = -\frac{25}{\theta^2} < 0$, $\forall \theta > 0$, luego la media muestral $\overline{X}$ es el estimador máximo verosímil de la media $\mu = \frac{1}{\theta}$ de una población Exponencial y, en este caso, $\underline{\overline{x} = 16}$ es la estimación máximo verosímil.

3. a) En el contexto de los contrastes de hipótesis.

b) Es el nivel de significación hasta el cual se acepta la hipótesis nula y, consecuentemente, a partir del cual se rechaza.

c) Para niveles de significación críticos muy pequeños la decisión correcta es el rechazo de la hipótesis nula, para niveles de significación críticos muy grandes la decisión correcta es la aceptación de la hipótesis nula y para el resto, en un entorno de los niveles habituales, la decisión depende del propio nivel de significación. La definición de dicho entorno es subjetiva. Un entorno razonable podría ser el intervalo $[0.005, 0.25]$ de modo que se considerarían niveles muy pequeños a aquellos inferiores a 0.005 y serían niveles de significación críticos muy grandes los superiores a 0.25.

4. X, gasto familiar de un hogar valenciano en las fiestas navideñas en euros.
Y, gasto familiar de un hogar alicantino en las fiestas navideñas en euros.
$(x_1, \ldots, x_{400})$ m.a.s. tal que $\overline{x} = 340$, $s_x = 50$.
$(y_1, \ldots, y_{300})$ m.a.s. tal que $\overline{y} = 350$, $s_y = 45$.

a) Hemos de obtener el nivel de significación crítico del siguiente contraste de hipótesis

$$\begin{aligned} H_0 &: \mu_x \geq \mu_y \\ H_1 &: \mu_x < \mu_y. \end{aligned}$$

Asumiendo independencia entre X e Y, esto es, que nada tiene que ver el gasto familiar de un hogar valenciano con el de un hogar alicantino y que los respectivos tamaños muestrales, $n_x = 400$ y $n_y = 300$, pueden ser considerados elevados, se tiene que $\frac{\overline{X}-\overline{Y}-(\mu_x-\mu_y)}{\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}} \dot{\sim} N(0,1)$.

El test con un nivel de significación α para resolver el contraste planteado es

$$\text{rechazar } H_0 \Leftrightarrow \overline{x} - \overline{y} \leq -z_\alpha \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}},$$

donde z_α es el cuantil de la distribución $N(0,1)$ que deja a su derecha una probabilidad igual a α.

El nivel de significación crítico se obtiene resolviendo para α la ecuación asociada a la inecuación anterior, esto es, $340 - 350 = -z_\alpha \sqrt{\frac{50^2}{400} + \frac{45^2}{300}} \Leftrightarrow z_\alpha = 2.7735 \Leftrightarrow \underline{\alpha_0 = 0.0028}$.

b) W, igual a 1 si la familia valenciana ha disminuido el gasto con respecto al año anterior y 0 en caso contrario.
$W \sim Br(\theta)$, $\mu_w = \theta$, $\sigma_w^2 = \theta(1-\theta)$.
$(w_1, \ldots, w_{400})$ m.a.s. tal que $\sum_{i=1}^{400} w_i = 120$.

Como el tamaño muestral es suficientemente grande y $(w_1, \ldots, w_{400})$ es una m.a.s. se tiene que $\frac{\overline{W}-\mu_w}{\sigma_w} \dot{\sim} N(0,1)$, aplicando el teorema central del límite. Por consiguiente, el intervalo de confianza $1-\alpha$ para θ es $\left[\overline{w} - z_{\frac{\alpha}{2}} \sqrt{\frac{\overline{w}(1-\overline{w})}{n}}, \overline{w} + z_{\frac{\alpha}{2}} \sqrt{\frac{\overline{w}(1-\overline{w})}{n}}\right]$, donde $z_{\frac{\alpha}{2}}$ es el cuantil de la distribución $N(0,1)$ que deja a su derecha una probabilidad igual a $\frac{\alpha}{2}$.

Sustituyendo $\overline{w} = \frac{\sum_{i=1}^{400} w_i}{400} = 0.3$ y $z_{0.025} = 1.96$, obtenemos el intervalo de confianza 0.95 para θ, $\left[0.3 - 1.96\sqrt{\frac{(0.3)(1-0.3)}{400}}, 0.3 + 1.96\sqrt{\frac{(0.3)(1-0.3)}{400}}\right] = \underline{[0.2551, 0.3449]}$.

5. X_1, preocupación de un estudiante de la Facultad de Economía en enero.
X_2, preocupación de un estudiante de la Facultad de Economía en noviembre.

a) $H_0: \ \pi_{11} = \pi_{11}^* = 0.46, \ \pi_{12} = \pi_{12}^* = 0.29, \ \pi_{13} = \pi_{13}^* = \theta$
$H_1: \ \text{no } H_0.$

Como $0.46+0.29+\theta = 1$, se tiene que $\theta = 0.25$ y no tenemos ningún parámetro desconocido en la hipótesis nula.

El estadístico χ^2 de Pearson cumple que

$$Q = \sum_{j=1}^{s} \frac{\left(n_{1j} - n\,\pi_{1j}^*\right)^2}{n\,\pi_{1j}^*} \dot{\sim} \chi^2_{s-1}$$

y el test que resuelve el contraste de bondad de ajuste es rechazar $H_0 \Leftrightarrow q \geq \chi^2_{s-1,\,\alpha}$, con $\chi^2_{s-1,\,\alpha}$ el cuantil de la distribución χ^2_{s-1} que deja a su derecha una probabilidad igual a α.

$$q = \frac{(150 - 300 \times 0.46)^2}{300 \times 0.46} + \frac{(85 - 300 \times 0.29)^2}{300 \times 0.29} + \frac{(65 - 300 \times 0.25)^2}{300 \times 0.25} = 2.4228,$$

y como se tiene que el cuantil $\chi^2_{2,0.298} \simeq 2.4228$, la decisión es <u>ACEPTAR H_0 para todo nivel de significación $\alpha \leq 0.298$</u> y, consecuentemente, los datos de la encuesta son compatibles con los porcentajes obtenidos a nivel nacional.

b) $H_0: \ X_1$ e X_2 homogéneas
$H_1: \ \text{no } H_0.$

Bajo ciertas condiciones genéricas y para tamaños muestrales grandes se tiene que

$$Q = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(n_{ij} - n_i \frac{n_{\cdot j}}{n}\right)^2}{n_i \frac{n_{\cdot j}}{n}} \dot{\sim} \chi^2_{(r-1)(s-1)}.$$

El test que resuelve el contraste de homogeneidad es rechazar $H_0 \Leftrightarrow q \geq \chi^2_{(r-1)(s-1),\alpha}$, donde $\chi^2_{(r-1)(s-1),\alpha}$ es el cuantil de la distribución $\chi^2_{(r-1)(s-1)}$ que deja a su derecha una probabilidad igual a α.

$$\begin{aligned} q \ &= \ \frac{\left(150 - 300\,\frac{220}{500}\right)^2}{300\,\frac{220}{500}} + \frac{\left(85 - 300\,\frac{165}{500}\right)^2}{300\,\frac{165}{500}} + \frac{\left(65 - 300\,\frac{115}{500}\right)^2}{300\,\frac{115}{500}} + \\ & \quad + \frac{\left(70 - 200\,\frac{220}{500}\right)^2}{200\,\frac{220}{500}} + \frac{\left(80 - 200\,\frac{165}{500}\right)^2}{200\,\frac{165}{500}} + \frac{\left(50 - 200\,\frac{115}{500}\right)^2}{200\,\frac{115}{500}} = 11.6656, \end{aligned}$$

y como se tiene que $q = 11.6656 > 5.991 = \chi^2_{2,0.05}$, la decisión es <u>RECHAZAR H_0 para un nivel de significación $\alpha = 0.05$</u> y, por consiguiente, se concluye que la opinión acerca de la mayor preocupación social de los estudiantes de Economía ha variado en los dos últimos meses.