Examen de Estadística II. 08-07-2002.

1. X, importe de las ventas diarias de televisores en euros.
Y, importe de las ventas diarias de DVD's en euros.
$\mu_x = 2500,\ \sigma_x = 1200,\ \mu_y = 3000,\ \sigma_y = 1000.$

$R = 800 + \sum_{i=1}^{45}(0.005\, X_i + 0.0075\, Y_i)$, salario del vendedor en euros.
Asumiendo que no existe relación entre las ventas diarias de televisores realizadas a lo largo de los 45 días, que todas ellas tienen la misma media y la misma varianza, que idéntica situación se tiene para las ventas de DVD's y que se pueden considerar independientes las variables X e Y; como el tamaño muestral $n = 45$ es suficientemente grande, estamos bajo las condiciones de aplicación del teorema central del límite y, consecuentemente, se tiene que $R \dot\sim N(\mu = 800 + 45(0.005 \cdot 2500 + 0.0075 \cdot 3000) = 2375, \sigma^2 = 45((0.005)^2 \cdot 1200^2 + (0.0075)^2 \cdot 1000^2) = 4151.25)$.
De este modo, calculamos de forma aproximada la probabilidad requerida como:

$$p(R > 2500) = p\left(Z > \frac{2500 - 2375}{\sqrt{4151.25}}\right) = p(Z > 1.9401) = \underline{0.0262},$$

con $Z = \frac{R-\mu}{\sigma} \dot\sim N(0,1)$.

2. X, igual a 1 si la familia está interesada en adquirir el producto y 0 en caso contrario.
Y, renta familiar en u.m.

a) El tamaño muestral necesario para la investigación de la media de una población dicotómica es función del nivel de confianza y error requeridos, es decir de la precisión deseada. Además, está evidentemente sujeto a las posibles restricciones presupuestarias del propio gabinete de estudios.
b) $X \sim Br(\theta),\ f(x) = \theta^x (1-\theta)^{1-x},\ x = 0, 1,\ \theta > 0.$
$(x_1, \ldots, x_n)$ m.a.s.
Como las variables aleatorias asociadas con la muestra son Bernouilli independientes, la función de verosimilitud resulta:

$$l(\theta) = \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} \theta^x (1-\theta)^{1-x} = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{\sum_{i=1}^n (1-x_i)}.$$

De modo que la log-verosimilitud es $\log l(\theta) = (\log\theta)\sum_{i=1}^n x_i + \log(1-\theta)\,(n - \sum_{i=1}^n x_i)$ y su primera derivada resulta $\frac{d \log l(\theta)}{d\theta} = \frac{\sum_{i=1}^n x_i}{\theta} - \frac{n - \sum_{i=1}^n x_i}{1-\theta}$, que igualada a cero proporciona la solución $\hat\theta = \frac{\sum_{i=1}^n x_i}{n} = \overline{x}$. Esta solución es efectivamente el máximo de $\log l(\theta)$ (y por la inyectividad de la función logarítmica, también de $l(\theta)$), pues hace negativa su segunda derivada, $\frac{d^2 \log l(\theta)}{d\theta^2} = -\frac{\sum_{i=1}^n x_i}{\theta^2} - \frac{n - \sum_{i=1}^n x_i}{(1-\theta)^2} < 0,\ \forall\theta > 0$, luego la media muestral $\overline{X}$ es el estimador máximo verosímil de la media θ de una población Bernouilli. Además, puede demostrarse que es un estimador insesgado, suficiente para θ, consistente y eficiente.

c) Podría plantearse un contraste de independencia entre X e Y. Para ello, previamente debería categorizarse la variable Y, de modo que consideráramos s categorías de renta familiar.

A continuación, calcularíamos el estadístico chi-cuadrado de Pearson,

$$Q = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(n_{ij} - \frac{n_{i\cdot}\, n_{\cdot j}}{n}\right)^2}{\frac{n_{i\cdot}\, n_{\cdot j}}{n}},$$

donde $r = 2$ es el número de categorías de X, n_{ij} es la frecuencia conjunta de la categoría i-ésima de X y j-ésima de Y, $n_{i\cdot}$ y $n_{\cdot j}$ son las correspondientes frecuencias marginales y n es el número total de datos.

Finalmente, el test que resuelve el contraste planteado es:

$$\text{rechazar } H_0 \Leftrightarrow q \geq \chi^2_{s-1,\alpha},$$

donde $\chi^2_{s-1,\alpha}$ es el cuantil de la distribución χ^2_{s-1} que deja a su derecha una probabilidad igual a α.

3. X, duración baja por accidente laboral en días.

Y, igual a 1 si el accidente es leve y 0 en caso contrario.

a) $X \sim N(\mu, \sigma^2)$.

$(x_1, \ldots, x_{10})$ m.a.s. tal que $\overline{x} = 5.8$, $s_x = 10$.

Como la población es Normal, se tiene que $\frac{\sqrt{n_x - 1}\,(\overline{X} - \mu)}{S_x} \sim t_{n_x - 1}$. A partir de este resultado, se obtiene el intervalo de confianza $1 - \alpha$ para μ como $\left[\overline{x} - t_{n_x-1,\frac{\alpha}{2}} \frac{s_x}{\sqrt{n_x - 1}}, \overline{x} + t_{n_x-1,\frac{\alpha}{2}} \frac{s_x}{\sqrt{n_x - 1}}\right]$, donde $t_{n_x-1,\frac{\alpha}{2}}$ es el cuantil de la distribución t_{n_x-1} que deja a su derecha una probabilidad igual a $\frac{\alpha}{2}$.

Utilizando la información muestral proporcionada y consultando en tablas de la distribución t de Student el cuantil $t_{9,0.025} = 2.262$, el intervalo numérico de confianza del 95% para μ resulta $\left[5.8 - (2.262)\frac{10}{\sqrt{10-1}}, 5.8 + (2.262)\frac{10}{\sqrt{10-1}}\right] = \underline{[-1.74, 13.34]}$.

b) $Y \sim Br(\theta)$.

$(y_1, \ldots, y_{250})$ m.a.s. tal que $\sum_{i=1}^{250} y_i = 130$.

Como la población es dicotómica, se ha de resolver el contraste de hipótesis bilateral:

$$\begin{aligned} H_0 &: \theta = \theta_0 = 0.5 \\ H_1 &: \theta \neq \theta_0 = 0.5\,. \end{aligned}$$

Como el tamaño muestral es suficientemente grande podemos aplicar el teorema central del límite y se tiene que $\overline{Y} \dot{\sim} N\left(\theta, \frac{\theta(1-\theta)}{n_y}\right)$, de modo que el test con un nivel de significación α que se propone para resolver el contraste planteado es:

$$\text{rechazar } H_0 \Leftrightarrow \overline{y} \notin \left[\theta_0 - z_{\frac{\alpha}{2}} \sqrt{\frac{\theta_0(1-\theta_0)}{n_y}}, \theta_0 + z_{\frac{\alpha}{2}} \sqrt{\frac{\theta_0(1-\theta_0)}{n_y}}\right],$$

donde $z_{\frac{\alpha}{2}}$ es el cuantil de la distribución $N(0,1)$ que deja a su derecha una probabilidad igual a $\frac{\alpha}{2}$.

Como $\overline{y} = \frac{\sum_{i=1}^{250} y_i}{n_y} = 0.52 > 0.5 = \theta_0$, hallamos el nivel de significación crítico, nivel hasta el cual se acepta la hipótesis nula y a partir del cual se rechaza, resolviendo la ecuación lineal para α, $\overline{y} = \theta_0 + z_{\frac{\alpha}{2}}\sqrt{\frac{\theta_0(1-\theta_0)}{n_y}}$. Esto es, $0.52 = 0.5 + z_{\frac{\alpha}{2}}\sqrt{\frac{0.5(1-0.5)}{250}} \leftrightarrow z_{\frac{\alpha}{2}} = 0.6325 \leftrightarrow \alpha = 0.5286$. Al obtener un nivel de significación crítico tan elevado, la decisión correcta pasa por <u>aceptar la igualdad de proporciones</u> para cualquier nivel de significación razonable.

4. X, tamaño de la vivienda.
Y, nivel de renta familiar.

a) Hemos de resolver el contraste de independencia:

$$\begin{aligned} H_0 &: X \text{ e } Y \text{ independientes} \\ H_1 &: \text{no } H_0 \,. \end{aligned}$$

Bajo ciertas condiciones genéricas y para un tamaño muestral grande se tiene que:

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

El test que resuelve el contraste planteado es:

$$\text{rechazar } H_0 \Leftrightarrow q \geq \chi^2_{(r-1)(s-1),\alpha},$$

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.

En nuestro caso, con $r = 3$ categorías para X, $s = 3$ categorías para Y y un nivel de significación $\alpha = 0.01$:

$$\begin{aligned} q \;=\; & \frac{\left(50 - \frac{(100)(130)}{400}\right)^2}{\frac{(100)(130)}{400}} + \frac{\left(40 - \frac{(100)(160)}{400}\right)^2}{\frac{(100)(160)}{400}} + \frac{\left(10 - \frac{(100)(110)}{400}\right)^2}{\frac{(100)(110)}{400}} + \\ & + \frac{\left(70 - \frac{(200)(130)}{400}\right)^2}{\frac{(200)(130)}{400}} + \frac{\left(85 - \frac{(200)(160)}{400}\right)^2}{\frac{(200)(160)}{400}} + \frac{\left(45 - \frac{(200)(110)}{400}\right)^2}{\frac{(200)(110)}{400}} + \\ & + \frac{\left(10 - \frac{(100)(130)}{400}\right)^2}{\frac{(100)(130)}{400}} + \frac{\left(35 - \frac{(100)(160)}{400}\right)^2}{\frac{(100)(160)}{400}} + \frac{\left(55 - \frac{(100)(110)}{400}\right)^2}{\frac{(100)(110)}{400}} = 66.7767, \end{aligned}$$

como se tiene que el estadístico es mayor que el cuantil, $q = 66.7767 > 13.28 = \chi^2_{4,0.01}$, la decisión correcta es <u>RECHAZAR H_0</u> para un nivel de significación $\alpha = 0.01$. Como el estadístico chi-cuadrado de Pearson es tan elevado, la conclusión es que para cualquier nivel de significación razonable el tamaño de la vivienda familiar sí que guarda relación con el nivel de renta.

b) En este caso se trata de un contraste de bondad de ajuste.

$$\begin{aligned} H_0 &: \pi_{13} = \pi_{13}^* = 0.1,\ \pi_{23} = \pi_{23}^* = \theta,\ \pi_{33} = \pi_{33}^* = \theta \\ H_1 &: \text{no } H_0\,, \end{aligned}$$

donde $0.1 + \theta + \theta = 1$, es decir, $\theta = 0.45$ y no tenemos ningún parámetro desconocido en la hipótesis nula.

El estadístico χ^2 de Pearson cumple que:

$$Q = \sum_{i=1}^{3} \frac{(n_{i3} - n\,\pi_{i3}^*)^2}{n\,\pi_{i3}^*} \,\dot{\sim}\, \chi^2_{r-1}\,,$$

y el test que resuelve el contraste de bondad de ajuste es:

$$\text{rechazar } H_0 \Leftrightarrow q \geq \chi^2_{r-1,\,\alpha},$$

con $\chi^2_{r-1,\,\alpha}$ el cuantil de la distribución χ^2_{r-1} que deja a su derecha una probabilidad igual a α.

$$q = \frac{(10 - 110\,(0.1))^2}{110\,(0.1)} + \frac{(45 - 110\,(0.45))^2}{110\,(0.45)} + \frac{(55 - 110\,(0.45))^2}{110\,(0.45)} = 1.\hat{1},$$

que es menor que el cuantil $\chi^2_{2,0.1} = 4.605$, luego la decisión es <u>ACEPTAR H_0</u> para un nivel de significación $\alpha = 0.1$.