Examen de Estadística II. 30-01-2006.

1. $X \sim Po(\lambda)$.
$(x_1, \ldots, x_{40})$ m.a.s. tal que $\sum_{i=1}^{40} x_i = 232$.

a) La media y la varianza de una distribución Poisson son iguales y coincidentes con el parámetro de la misma, consecuentemente ambas serán aproximadas por el mismo estimador. Por su parte, la mejor estimación posible de un parámetro poblacional es la proporcionada por el estimador máximo verosímil que, como las variables aleatorias asociadas con la muestra se suponen Poisson independientes, se obtiene maximizando la función de verosimilitud Poisson,

$$l(\lambda) = \prod_{i=1}^{40} f(x_i; \lambda) = \prod_{i=1}^{40} \exp(-\lambda) \frac{\lambda^{x_i}}{x_i!} = \exp(-40\lambda) \frac{\lambda^{\sum_{i=1}^{40} x_i}}{\prod_{i=1}^{40} x_i!}.$$

De modo que la log-verosimilitud es $\log l(\lambda) = -40\lambda + \left(\sum_{i=1}^{40} x_i\right) \log \lambda - \log \prod_{i=1}^{40} x_i!$ y su primera derivada resulta $\frac{d \log l(\lambda)}{d\lambda} = -40 + \frac{\sum_{i=1}^{40} x_i}{\lambda}$, que igualada a cero proporciona la solución $\hat{\lambda} = \frac{\sum_{i=1}^{40} x_i}{40} = \overline{x} = 5.8$. Esta solución es efectivamente el máximo de $\log l(\lambda)$ (y por la inyectividad de la función logarítmica, también de $l(\lambda)$), pues hace negativa su segunda derivada, $\frac{d^2 \log l(\lambda)}{d\lambda^2} = -\frac{\sum_{i=1}^{40} x_i}{\lambda^2} < 0$, $\forall \lambda > 0$, luego la media muestral $\overline{X}$ es el estimador máximo verosímil de la media y varianza λ de una población Poisson y, en este caso, $\underline{\overline{x} = 5.8}$ es la estimación máximo verosímil.

2. X, porcentaje de ingresos destinado a pagar el préstamo hipotecario.
$\mu = 34$, $\sigma = 20$.

Como disponemos de una m.a.s. $(x_1, \ldots, x_{200})$ y el tamaño muestral $n = 200$ es suficientemente grande, estamos en condiciones de aplicar el teorema central del límite para obtener que $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \stackrel{\cdot}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right) = N(34, 2)$.

De este modo, calculamos el valor de la constante k, el porcentaje medio de los ingresos destinados a pagar el préstamo hipotecario, mediante

$$p\left(\overline{X} > k\right) = 0.85 \Leftrightarrow p\left(Z > \frac{k-34}{\sqrt{2}}\right) = 0.85 \Leftrightarrow \frac{k-34}{\sqrt{2}} = -1.035 \Leftrightarrow \underline{k = 32.5363},$$

donde $Z = \frac{\overline{X}-34}{\sqrt{2}} \stackrel{\cdot}{\sim} N(0,1)$.

3. X, igual a 1 si el individuo encuestado es fumador y 0 en caso contrario.
$X \sim Br(\theta)$.
$(x_1, \ldots, x_{1000})$ m.a.s. tal que $\sum_{i=1}^{1000} x_i = 267$.

Se ha de resolver el siguiente contraste de hipótesis

$$\begin{aligned} H_0 &: \theta \leq \theta_0 = 0.25 \\ H_1 &: \theta > \theta_0 = 0.25. \end{aligned}$$

Como el tamaño muestral $n = 1000$ es suficientemente grande, podemos aplicar el teorema central del límite para obtener que $\overline{X} \sim N\left(\theta, \frac{\theta(1-\theta)}{n}\right)$, de modo que el test con un nivel de significación α que se propone para resolver el contraste planteado es

$$\text{rechazar } H_0 \Leftrightarrow \overline{x} \geq \theta_0 + z_\alpha \sqrt{\frac{\theta_0(1-\theta_0)}{n}}.$$

Como $0.267 < 0.25 + 1.645 \sqrt{\frac{(0.25)(1-0.25)}{1000}} = 0.2725$, la decisión es $\underline{\text{ACEPTAR } H_0 \text{ para}}$ $\underline{\alpha = 0.05}$, es decir, en base a los resultados muestrales debe aceptarse que el objetivo de las autoridades sanitarias se ha cumplido.

4. X, ingresos diarios del aparcamiento en euros.
$X \sim N(\mu, \sigma^2)$.
$(x_1, \ldots, x_{20})$ m.a.s. tal que $\overline{x} = 3000, s = 104$.

a) Como la población es Normal y la muestra no puede ser considerada de tamaño elevado se tiene que $\sqrt{n-1}\, \frac{\overline{X}-\mu}{S} \sim t_{n-1}$. Por consiguiente, el intervalo de confianza $1-\alpha$ para μ es $\left[\overline{x} - t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}}, \overline{x} + t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}}\right]$, donde $t_{n-1,\frac{\alpha}{2}}$ es el cuantil de la distribución t_{n-1} que deja a su derecha una probabilidad igual a $\frac{\alpha}{2}$.

De modo que el intervalo de confianza 0.95 para μ es $\left[3000 - 2.093\frac{104}{\sqrt{19}}, 3000 + 2.093\frac{104}{\sqrt{19}}\right] =$ $= \underline{[2950.0626, 3049.9374]}$, habiendo sustituido $t_{19,0.025} = 2.093$.

b) Hemos de obtener el nivel de significación crítico del contraste de hipótesis

$$\begin{aligned} H_0 &: \mu = \mu_0 = 3050 \\ H_1 &: \mu \neq \mu_0 = 3050. \end{aligned}$$

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

$$\text{rechazar } H_0 \Leftrightarrow \overline{x} \notin \left[\mu_0 - t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}}, \mu_0 + t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}}\right].$$

Resolviendo la ecuación para α, $\overline{x} = \mu_0 - t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}} \Leftrightarrow 3000 = 3050 - t_{19,\frac{\alpha}{2}} \frac{104}{\sqrt{19}}$, obtenemos $t_{19,\frac{\alpha}{2}} = 2.0956$ y el nivel de significación crítico $\underline{\alpha = 0.05}$.

5. X_1, libre de humos o se permite fumar.
X_2, zona de ubicación.

a) $H_0:\ X_1$ y X_2 independientes
$H_1:$ no H_0.
Bajo ciertas condiciones genéricas y para tamaños muestrales grandes se tiene que

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

El test que resuelve el contraste de independencia 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.

$$\begin{aligned} q \;&=\; \frac{\left(30 - \frac{(180)(70)}{300}\right)^2}{\frac{(180)(70)}{300}} + \frac{\left(70 - \frac{(180)(130)}{300}\right)^2}{\frac{(180)(130)}{300}} + \frac{\left(80 - \frac{(180)(100)}{300}\right)^2}{\frac{(180)(100)}{300}} + \\ &+ \frac{\left(40 - \frac{(120)(70)}{300}\right)^2}{\frac{(120)(70)}{300}} + \frac{\left(60 - \frac{(120)(130)}{300}\right)^2}{\frac{(120)(130)}{300}} + \frac{\left(20 - \frac{(120)(100)}{300}\right)^2}{\frac{(120)(100)}{300}} = 27.2894, \end{aligned}$$

y como se tiene que el cuantil, $\chi^2_{(2-1)(3-1),0.01} = \chi^2_{2,0.01} = 9.2103 < 27.2894 = q$, la decisión es RECHAZAR H_0 para un nivel de significación del 1% y, por consiguiente, se concluye que la catalogación de la cafetería como libre de humos o no sí que está relacionada con la ubicación de la misma.

b) $H_0:\ \pi_{\cdot 1} = \pi^*_{\cdot 1} = \theta,\ \pi_{\cdot 2} = \pi^*_{\cdot 2} = 2\theta,\ \pi_{\cdot 3} = \pi^*_{\cdot 3} = \theta$
$H_1:$ no H_0.
Como $\theta + 2\theta + \theta = 1$, se tiene que $\theta = \frac{1}{4}$ y no tenemos ningún parámetro desconocido en la hipótesis nula.
El estadístico χ^2 de Pearson cumple que

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

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.

$$q = \frac{(70 - 300 \times 0.25)^2}{300 \times 0.25} + \frac{(130 - 300 \times 0.5)^2}{300 \times 0.5} + \frac{(100 - 300 \times 0.25)^2}{300 \times 0.25} = 11.\widehat{3},$$

como el cuantil $\chi^2_{2,0.1} = 4.6052 < 11.\widehat{3} = q$, la decisión es RECHAZAR H_0 para el 10% de significación. Por consiguiente, no debe admitirse que en el casco antiguo y en la zona periférica el porcentaje de cafeterías es el mismo, mientras que en el distrito comercial es el doble.