Examen de Estadística II. 05-07-2006.

1. X, número de ejemplares diarios vendidos.
$X \sim Po(\lambda)$, $\mu_x = \lambda = 30$.
$Y = 1.3\,X$, ingresos diarios en euros.
$\mu_y = (1.3)\mu_x = (1.3)(30) = 39$, $\sigma_y^2 = (1.3)^2\sigma_x^2 = (1.3)^2(30) = 50.7$.

Como disponemos de una m.a.s. $(y_1, \ldots, y_{40})$ y el tamaño muestral $n = 40$ es suficientemente grande, estamos en condiciones de aplicar el teorema central del límite para obtener que $\sum_{i=1}^{n} Y_i \,\dot\sim\, N\left(n\,\mu_y, n\,\sigma_y^2\right) = N(1560, 2028)$.

Entonces, calulamos la probabilidad requerida mediante

$$p\left(\sum_{i=1}^{n} Y_i > 1625\right) = p\left(Z > \frac{1625 - 1560}{\sqrt{2028}}\right) = p(Z > 1.4434) = \underline{0.0742},$$

donde $Z = \frac{\sum_{i=1}^{n} Y_i - 1560}{\sqrt{2028}} \,\dot\sim\, N(0,1)$.

2. X_1, igual a 1 si el individuo encuestado está a favor de la iniciativa política A.
X_2, igual a 1 si el individuo encuestado está a favor de la iniciativa política B.
$X_1 \sim Br(\theta_1)$, $\epsilon_1 = 0.04$, $1 - \alpha_1 = 0.95$.
$X_2 \sim Br(\theta_2)$, $\epsilon_2 = 0.03$, $1 - \alpha_2 = 0.9$.

a) El tamaño muestral n requerido para la estimación de una proporción con una determinada confianza $1 - \alpha$ y un error máximo ϵ es $n = z_{\frac{\alpha}{2}}^2 \frac{0.25}{\epsilon^2}$, donde $z_{\frac{\alpha}{2}}$ es el cuantil de la distribución $N(0,1)$ que deja a su derecha una probabilidad igual a $\frac{\alpha}{2}$. De este modo, el tamaño necesario para la estimación de θ_1 con la precisión requerida es $n_1 = z_{0.025}^2 \frac{0.25}{(0.04)^2} = (1.96)^2 \frac{0.25}{(0.04)^2} = 600.25$ y el correspondiente a θ_2 resulta $n_2 = z_{0.05}^2 \frac{0.25}{(0.03)^2} = (1.645)^2 \frac{0.25}{(0.03)^2} = 751.6736$. Consecuentemente, y al existir independencia entre X_1 y X_2, deberán ser encuestadas $n = \underline{752}$ personas para asegurar la precisión indicada en ambas estimaciones.

b) El error ϵ asociado con la estimación de una proporción es $\epsilon = z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}$, donde $1-\alpha$ es la confianza requerida y n el tamaño muestral. Por lo tanto, el error asociado con la estimación de θ_1 es $\epsilon_1 = z_{0.025}\sqrt{\frac{(0.2)(0.8)}{752}} = \underline{0.0286}$ y el asociado a θ_2, $\epsilon_2 = z_{0.05}\sqrt{\frac{(0.85)(0.15)}{752}} = \underline{0.0214}$.

3. X, tiempo de retraso en minutos.
$X \sim Ex(\theta)$, $f(x) = \theta\exp(-\theta x)$, $x > 0$, $\theta > 0$.
$(x_1, \ldots, x_{150})$ m.a.s. tal que $\overline{x} = 15$.

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

$$l(\theta) = \prod_{i=1}^{150} f(x_i; \theta) = \prod_{i=1}^{150} \theta\exp(-\theta x_i) = \theta^{150}\exp\left(-\theta\sum_{i=1}^{150} x_i\right).$$

De modo que la log-verosimilitud es $\log l(\theta) = 150\log\theta - \theta\left(\sum_{i=1}^{150} x_i\right)$ y su primera derivada resulta $\frac{d\log l(\theta)}{d\theta} = \frac{150}{\theta} - \sum_{i=1}^{150} x_i$, que igualada a cero proporciona la solución $\hat{\theta} = \frac{150}{\sum_{i=1}^{150} x_i} = \frac{1}{\overline{x}}$. 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, $\frac{d^2\log l(\theta)}{d\theta^2} = -\frac{150}{\theta^2} < 0$, $\forall\theta > 0$, luego la <u>media muestral $\overline{X}$</u> es el estimador máximo verosímil de la media $\mu = \frac{1}{\theta}$ de una población Exponencial y, en este caso, $\overline{x} = 15$ es la estimación máximo verosímil.

b) $X \dot{\sim} Ex(\hat{\theta}) = Ex\left(\frac{1}{15}\right)$.

$p(X > 20) = \int_{20}^{+\infty} f(x)\,dx = \int_{20}^{+\infty} \frac{1}{15}\exp\left(-\frac{1}{15}x\right)\,dx = -\left[\exp\left(-\frac{1}{15}x\right)\right]_{20}^{+\infty} = \exp\left(-\frac{20}{15}\right) - \lim_{x\to+\infty}\exp\left(-\frac{1}{15}x\right) = \exp\left(-\frac{20}{15}\right) = \underline{0.2636}$

4. X, importe total de la cesta de la compra en un establecimiento de zona urbana, en euros.
Y, importe total de la cesta de la compra en un establecimiento de zona rural, en euros.
$(x_1,\ldots,x_{50})$ m.a.s. tal que $\overline{x} = 175, s_x = 55$.
$(y_1,\ldots,y_{45})$ m.a.s. tal que $\overline{y} = 143, s_y = 64$.

a) Asumiendo independencia entre X e Y, esto es, que nada tiene que ver la cesta de la compra en zona urbana con la de la zona rural y que los respectivos tamaños muestrales, $n_x = 50$ y $n_y = 45$, pueden ser considerados elevados, se tiene que $\frac{\overline{X}-\overline{Y}-(\mu_x-\mu_y)}{\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}} \dot{\sim} N(0,1)$. Por tanto, el intervalo de confianza $1-\alpha$ para $\mu_x - \mu_y$ es $\left[\overline{x}-\overline{y}-z_{\frac{\alpha}{2}}\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}, \overline{x}-\overline{y}+z_{\frac{\alpha}{2}}\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}\right]$, donde $z_{\frac{\alpha}{2}}$ es el cuantil de la distribución $N(0,1)$ que deja a su derecha una probabilidad igual a $\frac{\alpha}{2}$.

De modo que $\left[175-143-1.96\sqrt{\frac{55^2}{50}+\frac{64^2}{45}}, 175-143+1.96\sqrt{\frac{55^2}{50}+\frac{64^2}{45}}\right] = \underline{[7.8735, 56.1265]}$ es el intervalo de confianza 0.95 para $\mu_x - \mu_y$, habiendo sustituido $z_{0.025} = 1.96$.

b) Hemos de resolver el siguiente contraste de hipótesis

$$\begin{aligned} H_0 &: \mu_x - \mu_y = \mu_0 = 30 \\ H_1 &: \mu_x - \mu_y \neq \mu_0 = 30. \end{aligned}$$

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

$$\text{rechazar } H_0 \Leftrightarrow \mu_0 \notin \left[\overline{x}-\overline{y}-z_{\frac{\alpha}{2}}\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}, \overline{x}-\overline{y}+z_{\frac{\alpha}{2}}\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}\right].$$

Como $\mu_0 = 30 \in [7.8735, 56.1265]$, el intervalo de confianza 0.95 para $\mu_x - \mu_y$ obtenido en el apartado anterior, se tiene que la decisión es <u>ACEPTAR H_0 para $\alpha = 0.05$</u>, es decir, en base a los resultados muestrales debe aceptarse que en las zonas rurales el importe total medio de la cesta de la compra es 30 euros inferior al de las zonas urbanas.

5. X_1, intención de dejar de fumar o no.
X_2, categoría de edad.

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.

$$\begin{aligned} q \;&=\; \frac{\left(35 - \frac{(90)(65)}{145}\right)^2}{\frac{(90)(65)}{145}} + \frac{\left(50 - \frac{(90)(65)}{145}\right)^2}{\frac{(90)(65)}{145}} + \frac{\left(5 - \frac{(90)(15)}{145}\right)^2}{\frac{(90)(15)}{145}} + \\ &+ \frac{\left(30 - \frac{(55)(65)}{145}\right)^2}{\frac{(55)(65)}{145}} + \frac{\left(15 - \frac{(55)(65)}{145}\right)^2}{\frac{(55)(65)}{145}} + \frac{\left(10 - \frac{(55)(15)}{145}\right)^2}{\frac{(55)(15)}{145}} = 13.2194, \end{aligned}$$

y como se tiene que el cuantil, $\chi^2_{(2-1)(3-1),0.01} = \chi^2_{2,0.01} = 9.21 < 13.2194 = q$, la decisión es <u>RECHAZAR H_0 para un nivel de significación del 1%</u> y, por tanto, también para cualquiera de los niveles habituales. Por consiguiente, se concluye que la intención de dejar de fumar o no sí que está relacionada con la categoría de edad del trabajador.

b) $H_0: \ \pi_{11} = \pi^*_{11} = \theta, \ \pi_{12} = \pi^*_{12} = \theta, \ \pi_{13} = \pi^*_{13} = \frac{\theta}{2}$

$H_1:$ no H_0.

Como $\theta + \theta + \frac{\theta}{2} = 1$, se tiene que $\theta = \frac{2}{5}$ 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.

$$q = \frac{(35 - 90 \times 0.4)^2}{90 \times 0.4} + \frac{(50 - 90 \times 0.4)^2}{90 \times 0.4} + \frac{(5 - 90 \times 0.2)^2}{90 \times 0.2} = 14.86\widehat{1},$$

como el cuantil $\chi^2_{2,0.1} = 4.605 < 14.\widehat{3} = q$, la decisión es <u>RECHAZAR H_0 para el 10% de significación</u>. Por consiguiente, no debe admitirse que la distribución del total de trabajadores que sí tienen intención de dejar de fumar es tal que el porcentaje de fumadores de 20 a 35 años es el mismo que el de 35 a 50 años y el de mayores de 50 años es la mitad que cualquiera de los otros dos.