Examen de Estadística II. 09-07-2008.

1. X, igual a 1 si el estudiante es sancionado por el revisor y 0 en caso contrario.
$X \sim Br(\theta), \mu_x = \theta = 0.03,\ \sigma_x^2 = \theta\,(1-\theta) = 0.0291.$
$Y = 50\,X$, coste económico en euros del viaje sin billete.
$\mu_y = 50\,\mu_x = (50)(0.03) = 1.5,\ \sigma_y^2 = (50)^2\,\sigma_x^2 = (50)^2(0.0291) = 72.75.$

Como disponemos de una m.a.s. $(y_1, \ldots, y_{140})$ y el tamaño muestral $n = 140$ es suficientemente grande, estamos en condiciones de aplicar el teorema central del límite para obtener que $\sum_{i=1}^n Y_i \dot{\sim} N\left(n\,\mu_y, n\,\sigma_y^2\right) = N(210, 10185)$.

Para viajar todos los días del trimestre con billete el estudiante necesita 14 bonos, es decir, 85.4 euros. De este modo, calculamos la probabilidad requerida mediante

$$p\left(\sum_{i=1}^n Y_i > 85.4\right) = p\left(Z > \frac{85.4 - 210}{\sqrt{10185}}\right) = p(Z > -1.2346) = p(Z \leq 1.2346) = \underline{0.8916},$$

donde $Z = \frac{\sum_{i=1}^n Y_i - 210}{\sqrt{10185}} \dot{\sim} N(0,1)$.

2. X, número de botes de pintura vendidos diariamente.
$X \sim Po(\lambda)$.
$(x_1, \ldots, x_{100})$ m.a.s. tal que $\sum_{i=1}^{100} x_i = 1120$.

a) La media de una distribución Poisson resulta igual al parámetro de la misma, de modo que el estimador máximo verosímil de la media se obtiene maximizando la función de verosimilitud del parámetro λ de la distribución Poisson.

$$l(\lambda) = \prod_{i=1}^{100} f(x_i;\lambda) = \prod_{i=1}^{100} \exp(-\lambda)\frac{\lambda^{x_i}}{x_i!} = \exp(-100\lambda)\frac{\lambda^{\sum_{i=1}^{100} x_i}}{\prod_{i=1}^{100} x_i!}.$$

De modo que la log-verosimilitud es $\log l(\lambda) = -100\lambda + \left(\sum_{i=1}^{100} x_i\right)\log\lambda - \log\prod_{i=1}^{100} x_i!$ y su primera derivada resulta $\frac{d\log l(\lambda)}{d\lambda} = -100 + \frac{\sum_{i=1}^{100} x_i}{\lambda}$, que igualada a cero proporciona la solución $\hat{\lambda} = \frac{\sum_{i=1}^{100} x_i}{100} = \overline{x} = 11.2$. 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, $\frac{d^2\log l(\lambda)}{d\lambda^2} = -\frac{\sum_{i=1}^{100} x_i}{\lambda^2} < 0,\ \forall\lambda > 0$, luego la media muestral $\overline{X}$ es el estimador máximo verosímil de la media λ de una población Poisson y, en este caso, $\underline{\overline{x} = 11.2}$ es la estimación máximo verosímil.

b) $p(10 \leq X \leq 12) = p(X = 10) + p(X = 11) + p(X = 12) = \exp(-\lambda)\frac{\lambda^{10}}{10!} + \exp(-\lambda)\frac{\lambda^{11}}{11!} + \exp(-\lambda)\frac{\lambda^{12}}{12!} \simeq \exp(-11.2)\left(\frac{(11.2)^{10}}{10!} + \frac{(11.2)^{11}}{11!} + \frac{(11.2)^{12}}{12!}\right) = \underline{0.3474}$, donde hemos aproximado el parámetro λ desconocido mediante su estimación máximo verosímil.

3. X, gasto de una mujer en el período de rebajas de invierno, en euros.
$(x_1, \ldots, x_{300})$ m.a.s. tal que $\overline{x} = 130, s_x = 50$.

a) Como el tamaño muestral $n_x = 300$ es elevado podemos aplicar el teorema central del límite y obtener que $\frac{\overline{X}-\mu_x}{\frac{s_x}{\sqrt{n_x}}} \dot{\sim} N(0,1)$. Por tanto, el intervalo de confianza $1-\alpha$ para μ_x es $\left[\overline{x} - z_{\frac{\alpha}{2}} \frac{s_x}{\sqrt{n_x}}, \overline{x} + z_{\frac{\alpha}{2}} \frac{s_x}{\sqrt{n_x}}\right]$, 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 $\left[130 - 2.58\frac{50}{\sqrt{300}}, 130 + 2.58\frac{50}{\sqrt{300}}\right] = \underline{[122.5522, 137.4478]}$ es el intervalo de confianza 0.99 para μ_x, habiendo sustituido $z_{0.005} = 2.58$.

b) Y, gasto de un hombre en el período de rebajas de invierno, en euros.
$(y_1, \ldots, y_{200})$ m.a.s. tal que $\overline{y} = 124, s_y = 45$.
Hemos de resolver el siguiente contraste de hipótesis

$$\begin{aligned} H_0 &: \mu_x - \mu_y = \mu_0 = 0 \\ H_1 &: \mu_x - \mu_y \neq \mu_0 = 0. \end{aligned}$$

Asumiendo independencia entre X e Y, esto es, que nada tiene que ver el gasto en rebajas de una mujer con el de un hombre y que los respectivos tamaños muestrales, $n_x = 300$ y $n_y = 200$, pueden ser considerados elevados, en aplicación del teorema central del límite 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 consiguiente, se propone el siguiente test con un nivel de significación α para resolver el contraste planteado

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

$\mu_0 = 0 = 130 - 124 - z_{\frac{\alpha}{2}}\sqrt{\frac{50^2}{300} + \frac{45^2}{200}} = \overline{x} - \overline{y} - z_{\frac{\alpha}{2}} \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$, de donde $z_{\frac{\alpha}{2}} = 1.3965$ y el nivel de significación crítico es $\underline{\alpha = 0.163}$.

4. X_1, destino vacacional preferido por un madrileño.
X_2, destino vacacional preferido por un valenciano.

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

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

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.

$$\begin{aligned} q \;\; &= \;\; \frac{\left(130 - 300\,\frac{235}{500}\right)^2}{300\,\frac{235}{500}} + \frac{\left(110 - 300\,\frac{175}{500}\right)^2}{300\,\frac{175}{500}} + \frac{\left(60 - 300\,\frac{90}{500}\right)^2}{300\,\frac{90}{500}} + \\ & \quad + \frac{\left(105 - 200\,\frac{235}{500}\right)^2}{200\,\frac{235}{500}} + \frac{\left(65 - 200\,\frac{175}{500}\right)^2}{200\,\frac{175}{500}} + \frac{\left(30 - 200\,\frac{90}{500}\right)^2}{200\,\frac{90}{500}} = 4.4073, \end{aligned}$$

y como se tiene que el cuantil, $\chi^2_{(2-1)(3-1),0.1143} = \chi^2_{2,0.1143} \simeq 4.4073$, la decisión es ACEPTAR H_0 para todo nivel de significación $\alpha \leq 0.1143$ y, por consiguiente, se concluye que la distribución de las preferencias por el destino vacacional es la misma en madrileños y valencianos para cualquier nivel de significación habitual.

b) $H_0: \; \pi_{21} = \pi^*_{21} = 0.5, \; \pi_{22} = \pi^*_{22} = 0.3, \; \pi_{23} = \pi^*_{23} = 0.2$
$H_1: \;$ no H_0.
El estadístico χ^2 de Pearson cumple que

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

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.

$$q = \frac{(105 - 200 \times 0.5)^2}{200 \times 0.5} + \frac{(65 - 200 \times 0.3)^2}{200 \times 0.3} + \frac{(30 - 200 \times 0.2)^2}{200 \times 0.2} = 3.1\widehat{6},$$

el cuantil $\chi^2_{2,0.2059} \simeq 3.1\widehat{6}$, luego la decisión es ACEPTAR H_0 para todo nivel de significación $\alpha \leq 0.2059$. Por consiguiente, sí que pueden aceptarse como válidos los porcentajes propuestos en la Comunidad Valenciana para cualquier nivel de significación habitual.

5. a) El intervalo de confianza del 95% para la proporción de votos es $[37, 43]$ por lo que el valor 36% no estaría incluido. Si se considera que el estudio previo ha sido correctamente realizado, ello puede ser debido a que la muestra considerada es rara o atípica (esto sólo ocurriría con cinco muestras de cada cien) o porque realmente no estamos realizando inferencia sobre la proporción de votos a obtener por el partido y sí sobre la proporción de intenciones de voto. Esto es, como suele ocurrir, una cosa es la intención de voto y otra distinta el propio voto, de ahí que las estimaciones realizadas con el voto escrutado sí que son bastante más fiables.

b) Aceptar la hipótesis nula hasta el nivel $\alpha = 0.9$ y rechazarla a partir de él. Es decir, aceptar la hipótesis nula del contraste con casi total seguridad de no equivocarnos.