Examen de Estadística II. 24-01-2005.

1. X, número de jamones de bellota exportados semanalmente por Pardao.
$X \sim Po(\lambda)$.

a) $(x_1, \ldots, x_{52})$ m.a.s. tal que $\sum_{i=1}^{52} x_i = 792$.

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,

$$l(\lambda) = \prod_{i=1}^{52} f(x_i;\lambda) = \prod_{i=1}^{52} \exp(-\lambda)\frac{\lambda^{x_i}}{x_i!} = \exp(-52\lambda)\frac{\lambda^{\sum_{i=1}^{52} x_i}}{\prod_{i=1}^{52} x_i!}.$$

De modo que la log-verosimilitud es $\log l(\lambda) = -52\lambda + \left(\sum_{i=1}^{52} x_i\right)\log\lambda - \log\prod_{i=1}^{52} x_i!$ y su primera derivada resulta $\frac{d\log l(\lambda)}{d\lambda} = -52 + \frac{\sum_{i=1}^{52} x_i}{\lambda}$, que igualada a cero proporciona la solución $\hat{\lambda} = \frac{\sum_{i=1}^{52} x_i}{52} = \overline{x} = 15.2308$. 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, $\frac{d^2\log l(\lambda)}{d\lambda^2} = -\frac{\sum_{i=1}^{52} x_i}{\lambda^2} < 0$, $\forall\lambda > 0$, luego la media muestral $\overline{X}$ es el estimador máximo verosímil de la media λ de una población Poisson y, en este caso, $\underline{\overline{x} = 15.2308}$ es la estimación máximo verosímil.

b) La media y la varianza de una distribución Poisson son iguales y coincidentes con el parámetro de la misma, consecuentemente el estimador máximo verosímil de $\sigma^2 = \lambda$ también es la media muestral $\overline{X}$.

2. X, tiempo en horas que tarda una empresa de tratamiento de datos en realizar un encargo.
$X \sim Ex(\lambda = 2)$, $f(x) = \frac{1}{2}\exp\left(-\frac{1}{2}x\right)$, $\mu = \lambda = 2$, $\sigma^2 = \lambda^2 = 4$.

Asumiendo que no existe relación entre los diferentes servicios mensuales y que todos ellos tienen la misma distribución que X, como el tamaño muestral $n = 300$ es suficientemente grande, estamos en condiciones de aplicar el teorema central del límite para obtener que $\sum_{i=1}^{300} k\,X_i \stackrel{\cdot}{\sim} N(300\,k\,\mu, 300\,k^2\,\sigma^2) = N(600k, 1200k^2)$, donde k es el coste en euros por hora del servicio.

De este modo, calculamos de forma aproximada la constante k requerida como

$$p\left(\sum_{i=1}^{300} kX_i > 16000\right) = 0.95 \Leftrightarrow p\left(Z > \frac{16000 - 600k}{2k\sqrt{300}}\right) = 0.95 \Leftrightarrow$$

$$\Leftrightarrow \frac{8000 - 300k}{k\sqrt{300}} = -1.645 \Leftrightarrow \underline{k = 29.4651},$$

donde $Z = \frac{\sum_{i=1}^{300} k\,X_i - 600k}{2k\sqrt{300}} \stackrel{\cdot}{\sim} N(0,1)$.

3. X, diferencia porcentual entre los salarios de hombres y mujeres.
$(x_1, \ldots, x_{20})$ m.a.s. tal que $\overline{x} = 15,\ s = 5$.

Denotando por μ y σ^2 a la media y varianza poblacionales, se ha de resolver el siguiente contraste de hipótesis

$$\begin{aligned} H_0 &: \mu \leq \mu_0 = 12 \\ H_1 &: \mu > \mu_0 = 12. \end{aligned}$$

Como el tamaño muestral $n = 20$ no puede ser considerado elevado, hemos de asumir Normalidad para la población, $X \sim N(\mu, \sigma^2)$. En ese caso se tiene que $\sqrt{n-1}\,\frac{\overline{X}-\mu}{S} \sim t_{n-1}$, de modo que el test con un nivel de significación α que se propone para resolver el contraste planteado es

$$\text{rechazar } H_0 \Leftrightarrow \overline{x} \geq \mu_0 + t_{n-1,\alpha}\,\frac{s}{\sqrt{n-1}}.$$

Como $15 \geq 12 + 1.729\,\frac{5}{\sqrt{19}} = 13.9833$, la decisión es $\underline{\text{RECHAZAR } H_0 \text{ para } \alpha = 0.05}$, es decir, en base a los resultados muestrales no puede aceptarse como cierta la opinión del sindicato y por tanto se acepta que la diferencia entre los valores salariales medios es en términos porcentuales de más de un 12% a favor de los hombres.

4. La estimación de A será más fiable dado que está basada en una m.a.s. y, consecuentemente, es susceptible de ser medido su error. La muestra tomada por el estudiante B no es simple dado que no todos los individuos tienen la misma probabilidad de formar parte de la muestra.

5. X, igual a 1 si el espectador se muestra a favor de incrementar el tiempo dedicado a deportes y 0 en caso contrario.
$X \sim Br(\theta)$, $\mu = \theta$, $\sigma^2 = \theta(1-\theta)$.
$(x_1, \ldots, x_{150})$ m.a.s. tal que $\sum_{i=1}^{150} x_i = 90$.

a) Como el tamaño muestral es suficientemente grande y $(x_1, \ldots, x_{150})$ es una m.a.s. se tiene que $\frac{\overline{X}-\mu}{\sigma} \dot{\sim} N(0,1)$, aplicando el teorema central del límite. Por consiguiente, el intervalo de confianza $1-\alpha$ para θ es $\left[\overline{x} - z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}, \overline{x} + z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}\right]$, 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 $\overline{x} = \frac{\sum_{i=1}^{150} x_i}{150} = 0.6$ y $z_{0.025} = 1.96$, obtenemos el intervalo de confianza 0.95 para θ, $\left[0.6 - 1.96\sqrt{\frac{(0.6)(1-0.6)}{150}}, 0.6 + 1.96\sqrt{\frac{(0.6)(1-0.6)}{150}}\right] = \underline{[0.5216, 0.6784]}$.

b) $\epsilon = 0.04,\ 1-\alpha = 0.95$.

Para determinar el tamaño muestral necesario para alcanzar unos determinados requerimientos en cuanto a la precisión de una estimación de la proporción θ, utilizamos el hecho de que la media de una m.a.s. de tamaño n grande sigue aproximadamente una distribución Normal de media θ y varianza $\frac{\theta(1-\theta)}{n}$, en virtud del teorema central del límite.

Así, $n = \left(z_{\frac{\alpha}{2}}\right)^2 \frac{\theta(1-\theta)}{\epsilon^2} = (1.96)^2 \frac{0.25}{(0.04)^2} = 600.25$, donde se ha sustituido la varianza poblacional desconocida $\theta(1-\theta)$ por su valor máximo, alcanzado en $\theta = 0.5$.

De este modo, es necesaria una muestra de tamaño $\underline{601}$ para alcanzar los objetivos de precisión requeridos.

6. X_1, sexo.
X_2, juguete preferido por el niño.

a) $H_0: \ \pi_{11} = \pi_{11}^* = \theta, \ \pi_{12} = \pi_{12}^* = 2\theta, \ \pi_{13} = \pi_{13}^* = \theta$
$H_1: \ \text{no } H_0.$

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

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

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.

$$q = \frac{(9 - 50 \times 0.25)^2}{50 \times 0.25} + \frac{(26 - 50 \times 0.5)^2}{50 \times 0.5} + \frac{(15 - 50 \times 0.25)^2}{50 \times 0.25} = 1.52,$$

el cuantil $\chi^2_{2,0.45} \simeq 1.52$, luego la decisión es $\underline{\text{ACEPTAR } H_0}$ $\underline{\text{para todo nivel de significación } \alpha \leq 0.45}$. Por consiguiente, sí que puede aceptarse con un alto grado de fiabilidad que los niños prefieren por igual los juguetes educativos y los tradicionales, mientras que los que prefieren los electrónicos son el doble de los otros dos.

b) Y_1, juguete preferido por el niño varón.
Y_2, juguete preferido por la niña.

$H_0: \ Y_1 \text{ e } Y_2$ homogéneas
$H_1: \ \text{no } H_0.$

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

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

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.

$$\begin{aligned} q \;=\; & \frac{\left(9 - 50\,\frac{20}{90}\right)^2}{50\,\frac{20}{90}} + \frac{\left(26 - 50\,\frac{46}{90}\right)^2}{50\,\frac{46}{90}} + \frac{\left(15 - 50\,\frac{24}{90}\right)^2}{50\,\frac{24}{90}} + \\ & + \frac{\left(11 - 40\,\frac{20}{90}\right)^2}{40\,\frac{20}{90}} + \frac{\left(20 - 40\,\frac{46}{90}\right)^2}{40\,\frac{46}{90}} + \frac{\left(9 - 40\,\frac{24}{90}\right)^2}{40\,\frac{24}{90}} = 1.3886, \end{aligned}$$

y como se tiene que el cuantil, $\chi^2_{(2-1)(3-1),0.5} = \chi^2_{2,0.5} \simeq 1.3886$, la decisión es $\underline{\text{ACEPTAR } H_0}$ $\underline{\text{para todo nivel de significación}}$ $\underline{\alpha \leq 0.5}$ y, por consiguiente, se concluye que la distribución de las preferencias por el tipo de juguete de los niños es la misma que la de las niñas.