Examen de Estadística II. 06-07-2005.

1. Comparando los errores cuadráticos medios de los estimadores respectivos. Será preferible aquella estimación que provenga del estimador con menor error cuadrático medio.

2. X_1, tipo de cesta.
X_2, tamaño de la empresa.

a) $H_0: \ \pi_{1.} = \pi^*_{1.} = 0.35, \ \pi_{2.} = \pi^*_{2.} = 0.45, \ \pi_{3.} = \pi^*_{3.} = \theta$
$H_1: \$ no H_0.

Como $0.35 + 0.45 + \theta = 1$, se tiene que $\theta = 0.2$ y no tenemos ningún parámetro desconocido en la hipótesis nula.

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

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

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

$$q = \frac{(384 - 1200 \times 0.35)^2}{1200 \times 0.35} + \frac{(564 - 1200 \times 0.45)^2}{1200 \times 0.45} + \frac{(252 - 1200 \times 0.2)^2}{1200 \times 0.2} = 4.7524,$$

los cuantiles $\chi^2_{2,0.1} = 4.605$ y $\chi^2_{2,0.05} = 5.991$, luego la decisión es ACEPTAR H_0 para todo nivel de significación $\alpha \leq 0.075$. Por consiguiente, ha de ser tomada con cierta cautela la previsión del gerente acerca de los pedidos.

b) $H_0: \ X_1$ y X_2 independientes
$H_1: \$ no H_0.

El estadístico χ^2 de Pearson cumple que

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

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.

$$\begin{aligned} q \ = \ & \frac{\left(260 - \frac{(315)(350)}{500}\right)^2}{\frac{(315)(350)}{500}} + \frac{\left(25 - \frac{(65)(350)}{500}\right)^2}{\frac{(65)(350)}{500}} + \frac{\left(65 - \frac{(120)(350)}{500}\right)^2}{\frac{(120)(350)}{500}} + \\ & + \frac{\left(55 - \frac{(315)(150)}{500}\right)^2}{\frac{(315)(150)}{500}} + \frac{\left(40 - \frac{(65)(150)}{500}\right)^2}{\frac{(65)(150)}{500}} + \frac{\left(55 - \frac{(120)(150)}{500}\right)^2}{\frac{(120)(150)}{500}} = 68.6995 \end{aligned}$$

y como se tiene que el estadístico es mayor que el cuantil, $q = 68.6995 > 10.6 = \chi^2_{2,0.005} = \chi^2_{(3-1)(2-1),0.005}$, la decisión es RECHAZAR H_0 para todo nivel de significación $\alpha \geq 0.005$ y, por consiguiente, para cualquiera de los niveles habituales se tiene que el tipo de cesta pedida varía según el tamaño de la empresa.

3. X, igual a 1 si el cliente ha realizado aportaciones extraordinarias al plan de pensiones entidad y 0 en caso contrario.

$X \sim Br(\theta)$, $f(x) = \theta^x(1-\theta)^{1-x}$, $x = 0, 1$.

$(x_1, \ldots, x_{350})$ muestra tal que $\sum_{i=1}^{350} x_i = 225$.

El mejor estimador que se propone para una característica poblacional es el estimador máximo verosímil (e.m.v.). Suponiendo que los datos de la sucursal constituyen una muestra aleatoria y simple de la entidad bancaria, esto es, que las variables aleatorias asociadas a todos los clientes tienen la misma distribución Bernouilli y que puede considerarse independencia entre ellas, entonces el e.m.v. se obtiene de un modo sencillo maximizando la función de verosimilitud

$$l(\theta) = \prod_{i=1}^{350} f(x_i;\theta) = \prod_{i=1}^{350} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{350} x_i}(1-\theta)^{350-\sum_{i=1}^{350} x_i}.$$

De modo que la log-verosimilitud es $\log l(\theta) = \left(\sum_{i=1}^{350} x_i\right)\log\theta + \left(350 - \sum_{i=1}^{350} x_i\right)\log(1-\theta)$ y su primera derivada resulta $\frac{d\log l(\theta)}{d\theta} = \frac{\sum_{i=1}^{350} x_i}{\theta} - \frac{350-\sum_{i=1}^{350} x_i}{1-\theta}$, que igualada a cero proporciona la solución $\hat{\theta} = \frac{\sum_{i=1}^{350} x_i}{350} = \overline{x} = 0.6429$. 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, $\left.\frac{d^2\log l(\theta)}{d\theta^2}\right|_{\theta=\hat{\theta}} = -\frac{\sum_{i=1}^{350} x_i}{\hat{\theta}^2} - \frac{350-\sum_{i=1}^{350} x_i}{(1-\hat{\theta})^2} = -\frac{350\overline{x}}{\overline{x}^2} - \frac{350(1-\overline{x})}{(1-\overline{x})^2} = \frac{-350}{\overline{x}(1-\overline{x})} < 0$, luego la media muestral $\overline{X}$ es el estimador máximo verosímil de la media θ de una población Bernouilli y, en este caso, $\overline{x} = 0.6429$ es la estimación máximo verosímil.

4. X, rentabilidad de un plan de pensiones de renta variable, en tanto por ciento.

$(x_1, \ldots, x_{20})$ m.a.s. tal que $\overline{x} = 18.6$, $s = 3.77$.

Como el tamaño muestral es reducido, para obtener un intervalo de estimación para la media poblacional μ es necesario asumir la distribución Normal para la población. En ese caso, $\sqrt{n-1}\ \frac{\overline{X}-\mu}{S} \sim t_{n-1}$ y se obtiene el intervalo de confianza $1-\alpha$ para μ como $\left[\overline{x} - t_{n-1,\frac{\alpha}{2}}\frac{s}{\sqrt{n-1}}, \overline{x} + t_{n-1,\frac{\alpha}{2}}\frac{s}{\sqrt{n-1}}\right]$, 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 $\frac{\alpha}{2}$.

Sustituyendo los valores muestrales observados y el cuantil $t_{19,0.05} = 1.729$, obtenemos el intervalo de confianza 0.9 para μ, $\left[18.6 - 1.729\frac{3.77}{\sqrt{19}}, 18.6 + 1.729\frac{3.77}{\sqrt{19}}\right] = [17.1046, 20.0954]$.

5. X, número de botellas de vino vendidas por una bodega riojana.

$(x_1, \ldots, x_{100})$ m.a.s. tal que $\overline{x} = 508$, $s^2 = 600$.

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

$$\begin{aligned} H_0 &: \mu = \mu_0 = 500 \\ H_1 &: \mu \neq \mu_0 = 500. \end{aligned}$$

Aplicando el teorema central del límite, se tiene que $\overline{X} \dot{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$, de modo que el test con un nivel de significación aproximado α que se propone para resolver el contraste planteado es

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

donde, como el tamaño muestral es grande, se ha sustituido la varianza poblacional desconocida por su correspondiente estimación, s^2, la varianza muestral y $z_{\frac{\alpha}{2}}$ es el cuantil de la distribución Normal estándar que deja a su derecha una probabilidad igual a $\frac{\alpha}{2}$.

Hallamos el nivel de significación crítico α_0, nivel hasta el cual se acepta la hipótesis nula y a partir del cual se rechaza, resolviendo la ecuación $\overline{x} = 508 = 500 + z_{\alpha/2} \frac{\sqrt{600}}{\sqrt{100}} = \mu_0 + z_{\alpha/2} \frac{s}{\sqrt{n}}$, de donde, $z_{\alpha/2} = 3.266$ y $\alpha_0 \simeq \underline{0.0012}$.

6. X, número de turistas diarios.
$\mu = 300$, $\sigma = 100$.

Asumiendo que no existe relación entre el número de turistas diarios a lo largo del trimestre y que todos ellos tienen la misma distribución que X, como el tamaño muestral $n = 90$ es suficientemente grande, estamos en condiciones de aplicar el teorema central del límite para obtener una distribución aproximada para los ingresos totales en euros, $Y = 72 \sum_{i=1}^{90} X_i \dot{\sim} N((72)(90)\,\mu, (72)^2(90)\,\sigma^2) = N(1944000, 4665600000)$.

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

$$\begin{aligned} p(Y > k) &= 0.9 \Leftrightarrow p\left(Z > \frac{k - 1944000}{\sqrt{4665600000}}\right) = 0.9 \Leftrightarrow \frac{k - 1944000}{\sqrt{4665600000}} = -1.285 \Leftrightarrow \\ &\Leftrightarrow \underline{k = 1856227.8213}, \end{aligned}$$

donde $Z = \frac{Y - 1944000}{\sqrt{4665600000}} \dot{\sim} N(0,1)$.