Examen de Estadística II. 06-07-2010.

1. X_1, igual a 1 si el cliente está satisfecho con la línea 800, 0 en caso contrario.
X_2, igual a 1 si el cliente está satisfecho con la línea 902, 0 en caso contrario.
$X_1 \sim Br(\theta_1), \widehat{\theta}_1 = 0.15$.
$X_2 \sim Br(\theta_2), \widehat{\theta}_2 = 0.21$.
$(x_{11}, \ldots, x_{300\,1})$ m.a.
$(x_{12}, \ldots, x_{250\,2})$ m.a.

Asumiendo independencia entre los clientes de la línea 800, y también entre los de la línea 902, podemos considerar tanto $(x_{11}, \ldots, x_{300\,1})$ como $(x_{12}, \ldots, x_{250\,2})$, m.a.s. Además, como sendos tamaños muestrales, $n_1 = 300$ y $n_2 = 250$, son suficientemente grandes puede aplicarse el teorema central del límite para obtener

$$\sum_{i=1}^{n_1} X_{1i} \quad \dot{\sim} \quad N(n_1\mu_1, n_1\sigma_1^2) = N((300)(0.15), (300)(0.15)(0.85)) = N(45, 38.25)$$
$$\sum_{i=1}^{n_2} X_{2i} \quad \dot{\sim} \quad N(n_2\mu_2, n_2\sigma_2^2) = N((250)(0.21), (250)(0.21)(0.79)) = N(52.5, 41.475).$$

La proporción muestral global de clientes satisfechos es $Y = \frac{\sum_{i=1}^{n_1} X_{1i} + \sum_{i=1}^{n_2} X_{2i}}{n_1+n_2}$. Considerando independencia entre las opiniones acerca de la satisfacción con el servicio de los clientes de ambas líneas, al ser una combinación lineal de variables independientes aproximadamente Normales, se tiene que $Y \dot{\sim} N\left(\mu_y = \frac{45+52.5}{300+250} = 0.1773, \sigma_y^2 = \frac{38.25+41.475}{(550)^2} = 2635 \times 10^{-4}\right)$.

$$p\,(Y > 0.2) = p\left(Z > \frac{0.2 - 0.1773}{\sqrt{2635 \times 10^{-4}}}\right) = p(Z > 1.4001) = \underline{0.0808},$$

donde $Z = \frac{Y-\mu_y}{\sigma_y} \dot{\sim} N(0,1)$.

2. X, gasto de un joven valenciano en euros.
$(x_1, \ldots, x_{71})$ m.a.s. tal que $\overline{x} = 220, s = 50$.
$1 - \alpha = 0.95$.

Asumiendo que la población sigue una distribución $\underline{N(\mu, \sigma^2)}$ y dada una muestra aleatoria simple de ella de tamaño n, entonces el teorema de Cochran establece que $\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$. A partir de este resultado y suponiendo media poblacional μ desconocida, se obtiene el intervalo de confianza $1-\alpha$ para σ^2 como $\left[\frac{ns^2}{\chi^2_{n-1,\frac{\alpha}{2}}}, \frac{ns^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right]$, donde $\chi^2_{n-1,\frac{\alpha}{2}}$ y $\chi^2_{n-1,1-\frac{\alpha}{2}}$ son los cuantiles de la distribución χ^2_{n-1} que dejan a su derecha probabilidades iguales a $\frac{\alpha}{2}$ y $\left(1-\frac{\alpha}{2}\right)$, respectivamente. Para la confianza requerida los cuantiles resultan $\chi^2_{70,0.025} = 95.02$ y $\chi^2_{70,0.975} = 48.76$, de modo que sustituyendo obtenemos el intervalo de confianza 0.95 para σ^2, $\left[\frac{(71)(50)^2}{95.02}, \frac{(71)(50)^2}{48.76}\right] = \underline{[1868.0278, 3640.2789]}$.

3. X, tiempo de procesamiento de un producto en minutos.
$X \sim N(\mu, \sigma = 0.6)$.
$(x_1, \ldots, x_{20})$ m.a.s. tal que $\overline{x} = 10.2$.

Denotando por μ a la media poblacional se ha de resolver el siguiente contraste de hipótesis

$$\begin{aligned} H_0 &: \mu \leq \mu_0 = 10 \\ H_1 &: \mu > \mu_0 = 10. \end{aligned}$$

Como $(x_1, \ldots, x_{20})$ es una m.a.s. se tiene que $\sqrt{n}\,\frac{\overline{X}-\mu}{\sigma} \sim N(0,1)$, de modo que el test con un nivel de significación α que se propone para resolver el contraste planteado es

$$\text{rechazar } H_0 \Leftrightarrow \overline{x} \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}.$$

El nivel de significación crítico se obtiene resolviendo para α la ecuación asociada a la inecuación anterior, esto es, $10.2 = 10 + z_\alpha \frac{0.6}{\sqrt{20}} \Leftrightarrow z_\alpha = 1.4907 \Leftrightarrow \underline{\alpha_0 = 0.068}$, nivel que se encuentra entre los habituales por lo que no existen razones estadísticas que puedan inducir a afirmar que el tiempo medio no es superior a diez minutos.

4. X_1, nivel de confianza en la economía española.
X_2, intención de voto.

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

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

El test que resuelve el contraste de 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.

$$\begin{aligned} q \;=\; & \frac{\left(10 - \frac{(50)(55)}{155}\right)^2}{\frac{(50)(55)}{155}} + \frac{\left(20 - \frac{(50)(60)}{155}\right)^2}{\frac{(50)(60)}{155}} + \frac{\left(20 - \frac{(50)(40)}{155}\right)^2}{\frac{(50)(40)}{155}} + \\ & + \frac{\left(25 - \frac{(55)(55)}{155}\right)^2}{\frac{(55)(55)}{155}} + \frac{\left(20 - \frac{(55)(60)}{155}\right)^2}{\frac{(55)(60)}{155}} + \frac{\left(10 - \frac{(55)(40)}{155}\right)^2}{\frac{(55)(40)}{155}} + \\ & + \frac{\left(20 - \frac{(50)(55)}{155}\right)^2}{\frac{(50)(55)}{155}} + \frac{\left(20 - \frac{(50)(60)}{155}\right)^2}{\frac{(50)(60)}{155}} + \frac{\left(10 - \frac{(50)(40)}{155}\right)^2}{\frac{(50)(40)}{155}} = 11.1233, \end{aligned}$$

el cuantil $\chi^2_{(3-1)(3-1),0.0252} = \chi^2_{4,0.0252} \simeq 11.1233$, luego la decisión es <u>ACEPTAR H_0 para todo nivel de significación $\alpha \leq 0.0252$</u>.

b) $H_0:\ \pi_{31}=\pi^*_{31}=\theta,\ \pi_{32}=\pi^*_{32}=\theta,\ \pi_{33}=\pi^*_{33}=\theta$

$H_1:$ no H_0.

Como $\theta+\theta+\theta=1$, se tiene que $\theta=0.\widehat{3}$ y no tenemos ningún parámetro desconocido en la hipótesis nula.

El estadístico χ^2 de Pearson cumple que

$$Q=\sum_{j=1}^{s}\frac{\left(n_{3j}-n\,\pi^*_{3j}\right)^2}{n\,\pi^*_{3j}}\stackrel{\cdot}{\sim}\chi^2_{s-1}$$

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.

$$q=\frac{\left(20-50\times 0.\widehat{3}\right)^2}{50\times 0.\widehat{3}}+\frac{\left(20-50\times 0.\widehat{3}\right)^2}{50\times 0.\widehat{3}}+\frac{\left(10-50\times 0.\widehat{3}\right)^2}{50\times 0.\widehat{3}}=4,$$

el cuantil $\chi^2_{3-1,0.1437}=\chi^2_{2,0.1437}\simeq 4$, luego la decisión es <u>ACEPTAR H_0 para todo nivel de significación $\alpha\leq 0.1437$</u>. Por consiguiente, sí que sería admisible asumir que el colectivo de los votantes a otros partidos se reparte de manera uniforme entre los tres niveles de percepción para cualquier nivel de significación habitual.

5. X, tiempo de espera en minutos.

$X\sim Ex(\theta),\ f(x)=\theta\exp(-\theta x),\ x>0,\ \theta>0.$

$(10,12,14,16,18)$ m.a.s. tal que $\overline{x}=14$.

En la distribución exponencial coinciden la media y la desviación típica. Por su parte, la mejor estimación para la media es la proporcionada por el estimador máximo verosímil que se obtiene maximizando la correspondiente función de verosimilitud Exponencial,

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

De modo que la log-verosimilitud es $\log l(\theta)=n\log\theta-\theta\left(\sum_{i=1}^{n}x_i\right)$ y su primera derivada resulta $\frac{d\log l(\theta)}{d\theta}=\frac{n}{\theta}-\sum_{i=1}^{n}x_i$, que igualada a cero proporciona la solución $\hat{\theta}=\frac{n}{\sum_{i=1}^{n}x_i}=\frac{1}{\overline{x}}$. 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, $\frac{d^2\log l(\theta)}{d\theta^2}=-\frac{n}{\theta^2}<0,\ \forall\theta>0$, luego la media muestral $\overline{X}$ es el estimador máximo verosímil de la media (y de la desviación típica) $\mu=\sigma=\frac{1}{\theta}$ de una población Exponencial y, en este caso, $\overline{x}=\underline{14}$ es la estimación máximo verosímil para ambas características.