Examen de Estadística II. 07-07-2003.

1. X_1, sexo del paciente.
X_2, evolución de la enfermedad.

a) $H_0: \ X_1$ y X_2 independientes
$H_1: \ $ no H_0.
Bajo ciertas condiciones genéricas y para un tamaño muestral grande 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 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.

$$\begin{aligned} q &= \frac{\left(68 - \frac{(100)(103)}{150}\right)^2}{\frac{(100)(103)}{150}} + \frac{\left(32 - \frac{(100)(47)}{150}\right)^2}{\frac{(100)(47)}{150}} + \frac{\left(35 - \frac{(50)(103)}{150}\right)^2}{\frac{(50)(103)}{150}} + \\ &+ \frac{\left(15 - \frac{(50)(47)}{150}\right)^2}{\frac{(50)(47)}{150}} = 0.062 \end{aligned}$$

y como se tiene que el estadístico es menor que el cuantil, $q = 0.062 < 2.706 = \chi^2_{1,0.1} = \chi^2_{(2-1)(2-1),0.1}$, la decisión es <u>ACEPTAR H_0</u> para un <u>nivel de significación $\alpha = 0.1$</u> y, por consiguiente, para cualquiera de los niveles habituales se tiene que la efectividad del medicamento no está influenciada por el sexo del paciente.

b) $H_0: \ \pi_{.1} = \pi^*_{.1} = 2\theta, \ \pi_{.2} = \pi^*_{.2} = \theta$
$H_1: \ $ no H_0.
Como $2\theta + \theta = 1$, se tiene que $\theta = \frac{1}{3}$ y no tenemos ningún parámetro desconocido en la hipótesis nula.
El estadístico χ^2 de Pearson cumple que

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

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.

$$q = \frac{\left(103 - 150 \times 0.\hat{6}\right)^2}{150 \times 0.\hat{6}} + \frac{\left(47 - 150 \times 0.\hat{3}\right)^2}{150 \times 0.\hat{3}} = 0.27,$$

el cuantil $\chi^2_{1,0.6} = 0.275$, luego la decisión es ACEPTAR H_0 para todo nivel de significación $\alpha \leq 0.6$. Por consiguiente, sí que puede aceptarse con un alto grado de fiabilidad que los que mejoran son el doble de los que no mejoran.

2. X, diferencia del recorrido realizado por un coche con y sin el dispositivo, en kilómetros. $(x_1, \ldots, x_{49})$ m.a.s. tal que $\overline{x} = 1,\ s = 1.5$.

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

$$\begin{array}{lcl} H_0 & : & \mu \geq \mu_0 = 0 \\ H_1 & : & \mu < \mu_0 = 0. \end{array}$$

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

$$\text{rechazar } H_0 \Leftrightarrow \overline{x} \leq \mu_0 - z_\alpha \, \frac{s}{\sqrt{n}},$$

donde, como el tamaño muestral es grande, se ha sustituido la varianza poblacional desconocida por su correspondiente estimación, s^2, la varianza muestral.

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 $\overline{x} = 1 = 0 - z_\alpha \, \frac{1.5}{\sqrt{49}} = \mu_0 - z_\alpha \, \frac{s}{\sqrt{n}}$, de donde, $z_\alpha = -4.\hat{6}$ y $\alpha_0 \simeq 1$.

Por lo tanto, para casi cualquier nivel de significación α la decisión es ACEPTAR H_0, esto es, en base a los resultados muestrales se acepta como cierta la afirmación publicitaria de la empresa sobre la reducción del consumo de gasolina.

3.

a) Si el número de días es elevado, la distribución de las ventas es la misma para todos los días y se puede suponer independencia entre las mismas, entonces es de aplicación el teorema central del límite y la distribución del total de ventas es aproximadamente Normal.

En el caso de que pueda asumirse la distribución Normal para las ventas diarias, entonces no se precisa un número elevado de días para obtener la Normal como la distribución exacta del total.

b) La distribución Normal tiene un carácter reproductivo para combinaciones lineales, esto es, la combinación lineal de una Normal es también Normal. Por lo tanto, los beneficios obtenidos por el primer procedimiento sí que siguen una distribución Normal, pero no podemos asegurar lo mismo de los beneficios obtenidos mediante la transformación parabólica.

4. Para la comparación de estimadores se utiliza el error cuadrático medio, $E.C.M._{\hat{\theta}}(\theta) = Var(\hat{\theta}) + b^2(\theta)$, donde $b(\theta) = E(\hat{\theta}) - \theta$, es el sesgo del estimador. En el primer caso, el estimador insesgado A tiene un error cuadrático medio igual a su varianza, 5. En el segundo, el error del estimador B es $4.25 = 2 + (1.5)^2$, menor que el de A. En consecuencia, B es mejor estimador que A. El estimador C lo es de otro parámetro poblacional, por lo que no es comparable a los anteriores.

5. X, concentración de cinc en una galleta de chapapote, en mg/l.
$(x_1, \ldots, x_{17})$ m.a.s. tal que $\overline{x} = 140,\ s = 30$.

a) 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 $t_{16,0.025} = 2.12$, obtenemos el intervalo de confianza 0.95 para μ, $\left[140 - 2.12\frac{30}{\sqrt{16}}, 140 + 2.12\frac{30}{\sqrt{16}}\right] = \underline{[124.1, 155.9]}$.

b) $\epsilon = 4,\ 1-\alpha = 0.9$.

Si el tamaño muestral resultante es suficientemente grande, entonces podemos aproximar la distribución de la media muestral a una Normal y obtener el mismo como $n = z^2_{\frac{\alpha}{2}}\ \frac{\sigma^2}{\epsilon^2}$. Como la varianza poblacional σ^2 es desconocida, podemos sustituirla por la estimación proporcionada por la varianza de la muestra de tamaño 17 y entonces $n = (1.645)^2\ \frac{30^2}{4^2} = 152.2139 \simeq \underline{153}$, que como era preciso resulta suficientemente grande.