e-mail: imozas@elx.uned.es
www.telefonica.net/web/imm

ESTADÍSTICA TEÓRICA II. ENERO 2011 1ª SEMANA

PRIMERA PARTE: CUESTIONES

1.- Distribución de la diferencia de medias muestrales de poblaciones normales cuando se conocen las varianzas poblacionales.

Respuesta.-

Sean X e Y variables aleatorias normales e independientes $N(\mu_X, \sigma_X)$ y $N(\mu_Y, \sigma_Y)$ respectivamente, con σ_X y σ_Y conocidas y tomemos muestras de tamaño n para la variable X y de tamaño m para la variable Y. Entonces, para las medias muestrales se cumple que:

$$E(\overline{X}-\overline{Y}) = E(\overline{X}) - E(\overline{Y}) = \mu_X - \mu_Y$$

$$Var(\overline{X}-\overline{Y}) = Var(\overline{X}) + Var(\overline{Y}) = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m} \text{ de donde } DT(\overline{X}-\overline{Y}) = \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$$

Luego la variable $\overline{X}-\overline{Y}$ se distribuye normal $N\left(\mu_X - \mu_Y, \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right)$

2.- Se sabe que si se verifican las condiciones de regularidad de Wolfowitz, la varianza del estimador está acotada inferiormente. Con esta información, comente si dada la población definida por la función de densidad $f(x;\theta) = 2(\theta - x)\,\theta^{-2}$ en el intervalo $0 \le x \le \theta$ e igual a cero en el resto de la recta real, se puede asegurar que la varianza del estimador obtenido por el método de los momentos se puede acotar inferiormente. Razone la respuesta.

Respuesta.-

Una de las condiciones de regularidad de Wolfowitz es que el campo de variación de la variable aleatoria no dependa del parámetro θ. La variable aleatoria propuesta no cumple las condiciones de regularidad de Wolfowitz pues su campo de variación es $0 \le x \le \theta$. Luego no podemos asegurar que la varianza del estimador se pueda acotar inferiormente.

3.- Estimación del tamaño muestral para estimar la media μ de una población normal con desviación típica desconocida.

Respuesta.-

Se desea construir un intervalo del 100(1–α)% de confianza para la media μ, usando como estimador la media muestral $\overline{X}$, de forma que el error de estimación $|\overline{X}-\mu|$ no supere cierto valor $\varepsilon > 0$.

Si S es la desviación típica muestral, sabemos que $\frac{\overline{X}-\mu}{S}\sqrt{n}$ se distribuye como una t-Student con n–1 grados de libertad. Sea $t_{\frac{\alpha}{2}}$ tal que $1-\alpha = P\left(-t_{\frac{\alpha}{2}} \le \frac{\overline{X}-\mu}{S}\sqrt{n} \le t_{\frac{\alpha}{2}}\right) =$

$= P\left(\frac{|\overline{X}-\mu|}{S}\sqrt{n} \le t_{\frac{\alpha}{2}}\right) = P\left(|\overline{X}-\mu| \le \frac{S\cdot t_{\frac{\alpha}{2}}}{\sqrt{n}}\right)$. Para que $|\overline{X}-\mu| \le \varepsilon$, bastará tomar $\frac{S\cdot t_{\frac{\alpha}{2}}}{\sqrt{n}} \le \varepsilon$ de donde $\sqrt{n} \ge \frac{S\cdot t_{\frac{\alpha}{2}}}{\varepsilon}$ o bien $n \ge \frac{S^2\cdot t_{\frac{\alpha}{2}}^2}{\varepsilon^2}$. Si hacemos $2\varepsilon = L$ (longitud del intervalo), queda $n \ge 4\frac{S^2\cdot t_{\frac{\alpha}{2}}^2}{L^2}$.

e-mail: **imozas@elx.uned.es**
www.telefonica.net/web/imm

4.- Regla de decisión y región crítica para el contraste χ^2 de bondad de ajuste cuando los parámetros de la población son conocidos.

Respuesta.-

Supongamos una variable aleatoria X con función de distribución $F_0(x)$. Sea $S_1, S_2, ..., S_k$ una partición del campo de variación de X. tal que $p_i = P[X \in S_i]$.

Supongamos ahora una muestra aleatoria de tamaño n tal que n_i observaciones pertenezcan a S_i, i = 1, 2..., k; $\sum_{n=1}^{k} n_i = n$.

Planteamos el contraste:

H_0: la muestra pertenece a la población con función de distribución $F_0(x)$

H_1: la muestra no pertenece a la población con función de distribución $F_0(x)$.

Supuesta H_0, el valor esperado de observaciones en la clase S_i será $n \cdot p_i$. Se sabe que el estadístico $\sum_{I=1}^{k} \frac{(n_i - np_i)^2}{np_i}$ se distribuye como una χ^2 con k−1 grados de libertad. Entonces, para un nivel de significación α, la región de rechazo vendrá determinada por el valor $\chi^2_{1-\alpha}$ tal que $P(\chi^2_{k-1} > \chi^2_{1-\alpha})$, es decir, rechazaremos H_0 si $\sum_{I=1}^{k} \frac{(n_i - np_i)^2}{np_i} > \chi^2_{1-\alpha}$

PROBLEMAS

1.- Dada la población definida por la función de densidad

$$f(x; \theta) = \frac{1}{6} \theta^{-4} x^3 e^{\frac{-x}{\theta}} \text{ siendo } x \geq 0 ; \theta > 0$$

Para muestras de tamaño n, determine el estimador de θ por el método de máxima verosimilitud. Compruebe si es insesgado, consistente y suficiente.

Nota: $\int_0^{+\infty} x^{p-1} e^{-qx} dx = \frac{\Gamma(p)}{q^p} = \frac{(p-1)!}{q^p}$

Solución.-

La función de verosimilitud: $L = \left(\frac{1}{6}\right)^n \theta^{-4n} \left(\prod_{i=1}^{n} x_i^3\right) e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i}$. Tomando logaritmos:

$\ln L = \ln\left(\frac{1}{6}\right)^n - 4n \ln\theta + \ln\left(\prod_{i=1}^{n} x_i^3\right) - \frac{1}{\theta}\sum_{i=1}^{n} x_i$. Derivando respecto de θ e igualando a cero:

$\frac{\partial \ln L}{\partial \theta} = \frac{-4n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} = 0$, de donde $\hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{4n} = \frac{\overline{X}}{4}$.

e-mail: imozas@elx.uned.es
www.telefonica.net/web/imm

Veamos si $\hat{\theta}$ es insesgado. Se tiene que $E(\hat{\theta}) = E\left(\frac{\overline{X}}{4}\right) = \frac{E(\overline{X})}{4} = \frac{E(X)}{4}$. Por otra parte, $E(X) = \int_0^{+\infty} \frac{1}{6}\theta^{-4}x^4 e^{-\frac{x}{\theta}}dx = \frac{1}{6}\theta^{-4}\int_0^{+\infty} x^4 e^{-\frac{x}{\theta}}dx = \frac{1}{6}\theta^{-4}\frac{\Gamma(5)}{\left(\frac{1}{\theta}\right)^5} = 4\theta$, luego $E(\hat{\theta}) = \frac{4\theta}{4} = \theta$. Es decir, $\hat{\theta}$ es insesgado.

$\hat{\theta}$ es consistente porque los estimadores obtenidos por el método de la máxima verosimilitud lo son.

Para ver si $\hat{\theta}$ es suficiente, aplicaremos el teorema de factorización de Fisher-Neyman. Para ello, llamemos $t = T(x_1, x_2, ..., x_n) = \frac{\sum_{i=1}^{n} x_i}{4}$. Se tiene entonces que:

$$f(x_1, x_2, ..., x_n; \theta) = \left(\frac{1}{6}\right)^n \theta^{-4n}\left(\prod_{i=1}^{n} x_i^3\right) e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i} = \left(\frac{1}{6}\right)^n \theta^{-4n} e^{-\frac{4t}{\theta}}\left(\prod_{i=1}^{n} x_i^3\right).$$

Llamando $g(T(x_1, x_2, ..., x_n), \theta) = \left(\frac{1}{6}\right)^n \theta^{-4n} e^{-\frac{4t}{\theta}}$ y $h(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} x_i^3$, tenemos la factorización $f(x_1, x_2, ..., x_n; \theta) = g(T(x_1, x_2, ..., x_n), \theta)\cdot h(x_1, x_2, ..., x_n)$, luego $\hat{\theta}$ es suficiente.

2.- Después de un amplio debate interno la agrupación local del partido X de la ciudad Bellavista ha decidido celebrar primarias para la elección de su candidato a alcalde. Los dos aspirantes aseguran contar con el apoyo de los ciudadanos. Concretamente el primero de ellos, el Sr. P. dice que obtendría al menos el 70% de los votos. El segundo candidato, Sr. Q, afirma en privado que si esos datos son fiables estaría dispuesto a abandonar, para evitar el desgaste político que le supondría la derrota. La secretaría de la agrupación decide contrastar esta afirmación, realizando una encuesta a 800 personas empadronadas y mayores de 18 años, de la que obtiene que 496 votarán al Sr. P. Con esta información, y con un nivel de significación del 10%, ¿se retirará de las primarias el candidato Sr. Q?

Solución.-

Sea p la proporción de votos favorables al sr. P. Consideramos las hipótesis:

$H_0: p \geq 0,7$

$H_1: p < 0,7$

Supuesta p = 0,7, la proporción muestral $\overline{p}$ se distribuye aproximadamente normal $N\left(p, \sqrt{\frac{pq}{n}}\right) \cong N(0,7; 0,0162)$ o, equivalentemente, $\frac{\overline{p} - 0,7}{0,0162}$ se distribuirá aproximadamente normal N(0,1).

En las tablas de la normal N(0,1) encontramos que 0,1 = P(Z ≤ –1,28),. Por otra parte, la proporción muestral obtenida $\overline{p} = \frac{496}{800} = 0,62$ de donde $\frac{0,62 - 0,7}{0,0162} \cong -4,94$, luego rechazaremos H_0 y, consecuentemente, el sr. Q no se retirará de las primarias