CURSO 2006/2007.	JUNIO. 1ª semana
Código de la Carrera 42	Código de la Asignatura 209.

PRIMERA PARTE: CUESTIONES TEÓRICO-CONCEPTUALES

1. Relación entre la distribución normal y los cambios de orígen y escala.

Respuesta.-

Sea X una variable normal N(μ, σ). Si consideramos un número real cualquiera a y construimos la variable U = X + a, decimos que hemos efectuado un cambio de origen, de manera que si X = 0 (origen de la variable X) entonces U = a (origen de la variable U). Sabemos además que U es normal y que E(U) = E(X) + a = μ + a. También Var(U) = = Var(X) = σ^2, luego U es normal N(μ+a, σ).

Si consideramos ahora un número real b>0 y construimos la variable V = $\frac{1}{b}$X, decimos que hemos efectuado un cambio de escala, de forma que a un intervalo de longitud b de la variable X le corresponde un intervalo de longitud 1 de la variable V. También sabemos que V sigue siendo normal y como E(V) = $\frac{1}{b}$E(X) = $\frac{\mu}{b}$, y Var(V) = $\frac{1}{b^2}\text{var}(X) = \frac{\sigma^2}{b^2}$, se tiene que V es normal $N\left(\frac{\mu}{b}, \frac{\sigma}{b}\right)$.

Si hacemos simultáneamente un cambio de origen y de escala, es decir, si Y = $\frac{X+a}{b}$, se tiene que Y es normal $N\left(\frac{\mu+a}{b}, \frac{\sigma}{b}\right)$.

Un caso particular de cambio de origen y escala es la tipificación de la variable X, es decir, construir la variable Z = $\frac{X-\mu}{\sigma}$ que, teniendo en cuenta lo anterior, será normal N(0, 1).

2. Error Cuadrático Medio.

Si F(x,θ) es la función de distribución de una variable aleatoria X, siendo θ un parámetro desconocido y $\hat{\theta}$ un estimador, se denomina error cuadrático medio del estimador a $E\left[\left(\hat{\theta}-\theta\right)^2\right]$. Desarrollándolo:

$$E\left[\left(\hat{\theta}-\theta\right)^2\right] = E\left(\hat{\theta}^2 - 2\theta\hat{\theta} + \theta^2\right) = E\left(\hat{\theta}^2\right) - 2\theta E\left(\hat{\theta}\right) + \theta^2 = \left(E\left(\hat{\theta}\right) - \theta\right)^2 + E\left(\hat{\theta}^2\right) - \left(E\left(\hat{\theta}\right)\right)^2$$

La expresión $E\left(\hat{\theta}\right) - \theta$ se denomina sesgo del estimador, mientras que $E\left(\hat{\theta}^2\right) - \left(E\left(\hat{\theta}\right)\right)^2 =$ = $\text{Var}\left(\hat{\theta}\right)$, de forma que el error cuadrático medio ECM = $\text{sesgo}^2(\hat{\theta}) + \text{Var}\left(\hat{\theta}\right)$

3. ¿Por qué a veces es más conveniente utilizar la distribución de Poisson en vez de la Binomial para el cálculo de probabilidades?

Respuesta.-

Sabemos que $\lim_{n\to\infty}\binom{n}{x}p^x(1-p)^{n-x} = \frac{\lambda^x}{x!}e^{-\lambda}$, donde λ = np, es decir, si X es binomial B(N, p) , entonces, para valores grandes de n, la probabilidad P(X = x) puede calcularse considerando que X es de Poisson de parámetro λ = np. Es más conveniente usar la distribución de Poisson que la binomial por economía de cálculo. Por ejemplo, si n = 60,

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

p = 0,05 y x = 4, por la binomial habría que calcular $\binom{60}{4}0{,}05^4 0{,}95^{56}$, mientras que por la de Poisson se tendría $\frac{3^4}{4!}e^{-3}$

4. Pueden construirse intervalos de confianza en distribuciones que no sean normales? Razonar la respuesta.

Sí. Un procedimiento es el método pivotal que consiste en la obtención de una función del parámetro desconocido cuya distribución muestral no dependa del parámetro.

Un segundo procedimiento sería el método general de Neyman, con menos limitaciones que el anterior, basado en la distribución de un estimador puntual del parámetro.

En estos dos casos se supone conocida la distribución de la población, no siendo necesario que sea normal.

Pero, incluso cuando no se conoce la distribución de la población, podemos construir intervalos de confianza aplicando la desigualdad de Chebychev.

PROBLEMAS

Problema 1

La demanda diaria de cierta marca de refrescos es una variable aleatoria con esperanza 3.000 y varianza 8.100.

Se pide:

1. La distribución aproximada de la demanda total de refrescos en 100 días consecutivos.

2. Probabilidad de que el total de refrescos demandados en 100 días sea superior a 298.000.

3. ¿Cuál debe ser la producción de refrescos en 100 días para atender la demanda total con probabilidad 0,99?

Solución.-

Sea X_i la variable aleatoria "demanda del día i", i = 1, 2, …, 100 que supondremos independientes. Sea S = $\sum_{i=1}^{100} X_i$. Se tiene que E(S) = $\sum_{i=1}^{100} E(X_i)$ = 300000 y Var(S) = $= \sum_{i=1}^{100} Var(X_i)$ = 810000, de donde $\sigma_S = \sqrt{810000} = 900$. Entonces:

1) de acuerdo con el teorema central del límite, S se distribuye aproximadamente normal N(300000, 900);

2) P(S > 298000) = *(tipificando)* = $P\left(Z > \frac{-2000}{900}\right)$ = p(Z > –2,22) = *(tablas)* = 0,9868.

3) Sea n la producción de refrescos que se busca. Deberá ser P[S ≤ n] = 0,99 ↔ $P\left[Z \le \frac{n-300000}{900}\right]$ = 0,99. De las tablas se obtiene que $\frac{n-300000}{900}$ = 2,33 ↔ n = 302097.

Problema 2

Supongamos dos poblaciones N(60,40) y N(80,40), se selecciona una muestra aleatoria simple de tamaño n = 100.

Determinar:

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

1. La mejor región crítica para realizar el contraste:

H_0: $\mu = 60$

siendo la talla o tamaño del contraste $\alpha = 0{,}1$

H_1: $\mu = 80$

2. La potencia del contraste.

Solución.-

1.- Por el lema de Neyman-Pearson, la mejor región crítica será la que verifique:

$$\frac{\left(\frac{1}{40\sqrt{2\pi}}\right)^{100} e^{-\frac{\sum_{i=1}^{100}(x_i-60)^2}{3200}}}{\left(\frac{1}{40\sqrt{2\pi}}\right)^{100} e^{-\frac{\sum_{i=1}^{100}(x_i-80)^2}{3200}}} \le k \;\leftrightarrow\; e^{-\frac{40\sum_{i=1}^{100}(x_i-70)}{3200}} \le k \text{ ; tomando logaritmos } -\frac{40\sum_{i=1}^{100}(x_i-70)}{3200} \le \ln k$$

$\leftrightarrow \dfrac{\sum_{i=1}^{100}(x_i-70)}{100} \ge -\dfrac{4}{5}\ln k \leftrightarrow \dfrac{\sum_{i=1}^{100}x_i}{100} \ge 70 - \dfrac{4}{5}\ln k$. Llamando K = $70 - \dfrac{4}{5}\ln k$, la mejor región crítica es de la forma $\overline{X} \ge K$.

Como, bajo la hipótesis nula, $\dfrac{\overline{X}-60}{40}\cdot 10$ es normal N(0, 1) y de las tablas obtenemos que $0{,}1 = P\left[\dfrac{\overline{X}-60}{40}\cdot 10 \ge 1{,}28\right] \rightarrow \overline{X} \ge 65{,}12$.

2.- La probabilidad de error de tipo II, β = $P\left[\overline{X} < 65{,}12 / H_1 \text{ cierta}\right]$= (*tipificando, teniendo en cuenta que, bajo la hipótesis H_1, $\dfrac{\overline{X}-80}{40}\cdot 10$ es normal N(0,1)*) = $P\left[\dfrac{\overline{X}-80}{40}\cdot 10 < -3{,}72\right]$ = *(tablas)* = 0,0001. Luego la potencia del contraste:

$$1 - \beta = 1 - 0{,}0001 = 0{,}9999$$