Métodos Estadísticos en Ingeniería

19/06/2007

1) Responder razonadamente a las siguientes preguntas:

a) Si en una población normal $N(\mu, 7)$ extraemos una muestra de tamaño N con $\bar{x}$ =0.99, ¿hay que rechazar $H_0 : \ \mu = 1$ si N es suficientemente grande?

b) ¿Qué se puede decir si en una regresión lineal $\text{Cov}(X, Y) = 0$?

2) Utilizando la distribución normal aproximar la probabilidad de que al lanzar una moneda un millón de veces salgan más de 500200 caras.

3) Se sabe que en cierta población el número de personas que padecen la enfermedad E es del 1 %. Se ha investigado una prueba diagnóstica que ha resultado positiva en el 97 % de las personas que padecen la enfermedad E y en el 2 % de las personas sanas. Calcular la probabilidad de que una persona con prueba positiva padezca realmente la enfermedad.

4) Un peródico asegura que el partido A obtendrá el 20 % de los votos porque han hecho una encuesta a 100 personas y 20 votarán a A y el resto a otros partidos. ¿Qué rango de error se debe admitir si se quiere tener un nivel de confianza del 90 %? (Nota: el error es la semilongitud del intervalo de confianza).

5) Sea una población con función de densidad $f_\theta(x) = \theta^{-1} \, e^{-x/\theta}$ si $x > 0$ y cero en el resto, donde θ es un parámetro positivo. Hallar el estimador de máxima verosimilitud de θ cuando se extrae una muestra de tamaño N.

σ conocida $\Rightarrow \dfrac{\bar{x} - \mu}{\sigma/\sqrt{N}} \sim N(0, 1)$, $\qquad$ σ desconocida $\Rightarrow \dfrac{\bar{x} - \mu}{s/\sqrt{N}} \sim t_{N-1}$, $\qquad$ $\sigma^2 = np(1-p)$ para $B(n, p)$.

Soluciones

1) a) La región de rechazo viene dada por $\mathcal{R} = \{\frac{|\bar{x}-\mu|}{\sigma/\sqrt{N}} > z_2\}$ donde z_2 es el valor que corresponde al nivel de significación. Sustituyendo, se debe rechazar H_0 si 0.01> $7z_2/\sqrt{N}$ y tomando N suficientemente grande esto siempre se cumple, cualquiera que sea z_2, porque $1/\sqrt{N} \to 0$ cuando $N \to \infty$.

b) Supongamos σ_x y σ_y no nulas. Si $\mathrm{Cov}(X,Y) = 0$, el coeficiente de correlación lineal es cero y por tanto la regresión lineal inadecuada (variables incorreladas). Otra posible conclusión es que la recta de regresión es horizontal (pendiente nula).

Los casos $\sigma_x = 0$ y $\sigma_y = 0$ son extremos, corresponden a que los valores de las muestras de X e Y sean constantes.

2) Tenemos una binomial $B(n,p)$ con $n = 10^6$ (el número de tiradas) y $p = 1/2$ (la probabilidad de cara), la cual se puede aproximar por una normal $N(np, \sqrt{np(1-p)}) = N(500000, 500)$. Tipificando, $Z = (X - 500000)/500$ sigue una normal $N(0,1)$ y

$$P(X > 500200) = P\big(Z > (500200 - 500000)/500\big) = P(Z > 0{,}4) = 1 - P(Z \le 0{,}4) \approx 1 - (\frac{1}{2} + 0{,}1554)$$

y la aproximación buscada es 0,3446.

3) E=padecerla $\quad$ E^c=no padecerla $\quad$ M=dar positivo en la prueba.

Según el enunciado: $P(E) = 0.01$, $P(E^c) = 0.99$, $P(M/E) = 0.97$, $P(M/E^c) = 0.02$ y se nos pide $P(E/M)$. Por la fórmula de Bayes

$$P(E/M) = \frac{P(M/E)P(E)}{P(M/E)P(E) + P(M/E^c)P(E^c)} = \frac{0.97 \cdot 0.01}{0.97 \cdot 0.01 + 0.02 \cdot 0.99} = 0.3288\ldots$$

4) El intervalo de confianza para la proporción de los que votarán a A es $|p - 0{,}2|/(\sigma/\sqrt{100}) < z_2$ con $\sigma = \sqrt{\widehat{p}(1-\widehat{p})} = \sqrt{0{,}2 \cdot 0{,}8} = 0{,}4$ y z_2 es tal que $P(Z < z_2) = 1 - 0{,}1/2 = 0{,}95$ que según la tabla implica $z_2 = 1{,}645$.

Entonces el intervalo de confianza se escribe como $|p - 0{,}2| < 1{,}645 \cdot 0{,}4/10 = 0{,}0658$, dicho de otra forma el error es como mucho $\pm 6{,}58\,\%$ con este nivel de confianza.

5) La función de verosimilitud es

$$L(\theta) = f_\alpha(x_1)f_\alpha(x_2)\ldots f_\alpha(x_N) = \theta^{-1}\, e^{-x_1/\theta}\theta^{-1}\, e^{-x_2/\theta}\ldots\theta^{-1}\, e^{-x_N/\theta} = \theta^{-N}\, e^{-s/\theta}$$

donde $s = \sum_{i=1}^N x_i$. Entonces $\log L(\theta) = -N\log\theta - s/\theta$ y derivando $(\log L)'(\theta) = -N/\theta + s/\theta^2$. Por tanto la derivada se anula cuando $(-N\theta + s)/\theta^2 = 0$, es decir, cuando θ toma el valor $\widehat{\theta} = s/N = \bar{x}$. Como $(\log L)''(\widehat{\theta}) = N/\widehat{\theta}^2 - 2s/\widehat{\theta}^3 = -N^3/s^2 < 0$, entonces en $\widehat{\theta}$ se alcanza un máximo y da el estimador de máxima verosimilitud.