Probabilidad y Estadística I
Semana 13

Errores tipo I y tipo II en el tamaño de muestra

Profesor: Nicolás López MSc

Universidad del Rosario

Universidad del Rosario

Contenido

Introducción

Tamaño muestral mediante TLC

Tamaño de muestral mediante α y β

Introducción

Recordando las hipótesis a través de un ejemplo

Suponga que se va a procesar a un presunto ladrón mediante los siguientes pasos:

1. Hipótesis nula (H_0). 'Status quo' o valor predeterminado del parámetro de interés. *El ladrón es inocente hasta que se demuestre lo contrario*.
2. Hipótesis alterna (H_1). Complemento de H_0. *El ladrón no es inocente*.
3. Estadístico de prueba & valor p. Información que soporta la decisión adquirida a través de la muestra & probabilidad de haber sido observada bajo H_0. *Evidencia que incrimina al ladrón*. Similar a un estimador (función de la muestra).
4. Región de rechazo. Valores del estadístico de prueba para los cuales rechazo H_0. *Más de 5 evidencias incriminatorias rechazan la inocencia del implicado*.
5. Decisión.

Universidad del Rosario

Contenido

Introducción

Tamaño muestral mediante TLC

Tamaño de muestral mediante α y β

Universidad del Rosario

Tamaño muestral

Bajo muestreo aleatorio de una población determinada, se pregunta por el tamaño requerido de la muestra para obtener conclusiones confiables del experimento a realizar. Un tamaño pequeño de muestra no permite usar el TLC, pero esto no acota el tamaño poblacional.

Tamaño muestral

Para determinar n, se fija un margen de error admisible E y una confianza determinada $1-\alpha$ en la estimación. Se tiene entonces:

$$z_{\alpha/2}\sigma_{\hat{\theta}} < E$$

De la inecuación resultante se despeja n y se toma la cota inferior de dicho resultado. Suponga que quiere determinar n para estimar μ:

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < E \longrightarrow n > \left(\frac{z_{\alpha/2}}{E}\right)^2 \sigma^2 \longrightarrow n = \left(\frac{z_{\alpha/2}}{E}\right)^2 \sigma^2$$

Suponga ahora que quiere determinar n para estimar p:

$$z_{\alpha/2}\sqrt{\frac{pq}{n}} < E \longrightarrow n > \left(\frac{z_{\alpha/2}}{E}\right)^2 pq \longrightarrow n = \left(\frac{z_{\alpha/2}}{E}\right)^2 pq$$

Universidad del Rosario

Tamaño muestral

Note que en la estimación del tamaño muestral:

1. Para μ se desconoce σ. Puede aproximarse mediante s de un experimento previo o sabiendo el rango de valores de la variable R, ya que $R \approx 4s$.
2. Para p se desconoce p naturalmente. A menos que se tenga información auxiliar que soporte lo contrario, el valor conservador usado para p es 0.5.
3. Para $\mu_1 - \mu_2$ y $p_1 - p_2$ se sigue un proceso similar al descrito anteriormente y se tienen las mismas consideraciones.

Universidad del Rosario

Contenido

Introducción

Tamaño muestral mediante TLC

Tamaño de muestral mediante α y β

Universidad del Rosario

Probabilidad E2 y tamaño de muestra

Recuerde que para calcular β necesitamos establecer un valor fijo de θ en H_1. Si por ejemplo se tiene un valor determinado θ_a para el juego de hipótesis:

$$\begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta > \theta_0 \end{cases}$$

Como $RR = \{\hat{\theta} : \hat{\theta} > k\}$, se tiene β igual a

$$\beta = P(\hat{\theta} < k | \theta = \theta_a) = P\left(\frac{\hat{\theta} - \theta_a}{\sigma_{\hat{\theta}}} < \frac{k - \theta_a}{\sigma_{\hat{\theta}}} | \theta = \theta_a\right)$$

Donde $\sigma_{\hat{\theta}}$ decrece con el tamaño muestral para nuestros estimadores usuales. Entre mayor sea β, $\theta_0 - \theta_a$ es más difícil de distinguir con el tamaño de muestra dado n.

Probabilidad E1, E2 y tamaño de muestra

Suponga que se prueba

$$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu > \mu_0 \end{cases}$$

Y $\mu_a > \mu_0$ un valor de μ bajo H_1. Esto sugiere que para seleccionar el tamaño muestral, puedo fijar α y β (calculado cuando $\mu = \mu_a$) obteniendo un juego de ecuaciones en función de n y k

Probabilidad E1, E2 y tamaño de muestra

Se tiene para α

$$\alpha = P(\bar{Y} > k | \mu = \mu_0) = P\left(\frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}} > \frac{k - \mu_0}{\sigma/\sqrt{n}}\right)$$

Entonces

$$\frac{k - \mu_0}{\sigma/\sqrt{n}} = z_\alpha \longrightarrow k = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

Se tiene para β

$$\beta = P(\bar{Y} < k | \mu = \mu_a) = P\left(\frac{\bar{Y} - \mu_a}{\sigma/\sqrt{n}} > \frac{k - \mu_a}{\sigma/\sqrt{n}}\right)$$

Entonces

$$\frac{k - \mu_a}{\sigma/\sqrt{n}} = z_{1-\beta} = -z_\beta \longrightarrow k = \mu_a - z_\beta \frac{\sigma}{\sqrt{n}}$$

Probabilidad E1, E2 y tamaño de muestra

Igualando las dos ecuaciones para k y despejando para n se obtiene que:

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_a - \mu_0)^2}$$

El tamaño de muestra para una prueba de hipótesis a una cola superior con una confianza del 1-α.

Universidad del Rosario

Probabilidad E1, E2 y tamaño de muestra

Note que

1. Entre menor sea $\mu_a - \mu_0$, n será mayor.
2. La misma respuesta se obtiene para la hipótesis:

$$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu < \mu_0 \end{cases}$$

Y $\mu_a < \mu_0$ un valor de μ bajo H_1.