Estadística
Inferencia Estadística

Universidad Europea de Madrid

Laureate International Universities

Problemas en Inferencia Estadística

POBLACIÓN $\rightarrow$ $X \approx F(\theta)$ $\leftarrow$ **desconocido**

A partir de una M.A.S. $X_1, X_2, \ldots, X_n$ queremos estimar el valor de θ

Estimar : Asignar un valor a algo que desconocemos

¿Cómo?

A partir del modelo F y de la muestra

- **Estimación puntual: Un solo valor**
- **Estimación por intervalos: Un rango de valores**
- **Contraste de hipótesis: Rechazo o aceptación de ciertas hipótesis sobre el parámetro**

Universidad Europea de Madrid

Contraste de hipótesis

La finalidad del contraste es decidir si una determinada hipótesis sobre la distribución es confirmada o rechazada estadísticamente a partir de las observaciones de la muestra

Queremos contrastar:

$$\left.\begin{array}{l} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1 \end{array}\right\}$$

Hipótesis nula

Hipótesis alternativa

Universidad Europea de Madrid

Región de rechazo

Un test para contrastar la hipótesis nula H_0 frente a la hipótesis alternativa H_1, consiste en decidir, para cada posible muestra, si aceptamos o rechazamos la hipótesis nula. De esta manera, el test divide el espacio muestral en dos regiones:

$$\left.\begin{array}{l} \text{Región crítica} : R \\ \text{Región de aceptación} : A \end{array}\right\}$$

Rechazo la hipótesis nula

Acepto la hipótesis nula

Universidad Europea de Madrid

Errores

Podemos cometer 2 tipos de errores:

	H_0 cierta	H_0 falsa
Rechazar H_0	Error de tipo I	Decisión correcta
Aceptar H_0	Decisión correcta	Error de tipo II

Lo idoneo sería conseguir un test que hiciese mínima las probabilidades de ambos errores. Desgraciadamente esto no será habitualmente posible ya que la reducción de la probabilidad de cometer el Error de tipo I tendrá que hacerse a costa de aumentar la probabilidad de cometer el otro

Errores

Podemos pensar en el contraste de hipótesis en términos de detectores de humos.

Un Error de tipo I es una alarma sin fuego, por el contrario, un Error de tipo II es un fuego sin alarma.

Los cocineros saben cómo evitar un error de tipo I: quitando las pilas de la alarma. Por desgracia esto aumenta el error de tipo II.

Igualmente podemos disminuir el error de tipo II, haciendo que la alarma sea hipersensible, sin embargo esto aumenta las falsas alarmas.

Nivel de significación y Potencia de contraste

- Nivel de significación:

$$\alpha = P(\text{error de tipo 1}) = P(\text{rechazar la hipótesis nula siendo cierta})$$

- Potencia del contraste:

$$1-\beta = P(\text{rechazar la hipotesis nula siendo falsa})$$

donde $\beta = P(\text{error de tipo II})$

Universidad Europea de Madrid

Función de potencia

Llamamos función de potencia de un test con región crítica R a la función a la probabilidad de rechazar la hipótesis nula (para cada valor de θ) $P_\theta(R)$

Lo que interesa es que la función de potencia tome valores próximos a cero cuando $\theta \in \Theta_0$ (pequeña probabilidad de error de tipo I) y valores próximos a uno cuando $\theta \in \Theta_1$ (pequeña probabilidad de tipo II). Para ello se fija α, y se intenta maximizar la función de potencia para los valores $\theta \in \Theta_1$

Universidad Europea de Madrid

Observaciones

La hipótesis nula y la hipótesis alternativa NO desempeñan papeles simétricos. Los test de hipótesis tienden a ser conservadores con la hipótesis nula, es decir, es necesario una evidencia muy fuerte en contra de H_0 para que el test la rechace.

Elección de la hipótesis nula

➢ Si queremos contrastar $\theta = \theta_0$ frente a $\theta \neq \theta_0$, tomaremos como hipótesis nula $\theta = \theta_0$

➢ Si queremos contrastar $\theta < \theta_0$ y $\theta > \theta_0$, tenemos que recordar que los test son conservadores con la hipótesis nula, lo que nos sugiere que tomemos como alternativa aquella hipótesis que queremos probar estadísticamente.

Observaciones

Elección del nivel de significación

Normalmente se suele tomar como valor de α: 0.01, 0.05 ó 0.1.

El nivel elegido depende de lo serias que sean las consecuencias de rechazar equivocadamente H_0 , cuanto más desastrosas sean, más pequeño se ha de tomar.

P-valor o valor crítico

Una manera de evitar la decisión de elegir un nivel de significación es considerar el p-valor, que se define como el mínimo nivel de significación α necesario para rechazar H_0

- ❖ Un p-valor grande (> 0.2) confirma la aceptación de H_0 ya que eso significaría que la probabilidad de equivocarnos (rechazar H_0 siendo cierta) es grande, luego lo lógico, es aceptar.
- ❖ Por el contrario, un p-valor pequeño (< 0.01) confirma la decisión de rechazar H_0.
- ❖ En situaciones intermedias, el p-valor no indica nada en concreto, salvo tal vez, elegir otra muestra y volver a realizar el contraste.

Universidad Europea de Madrid

Contrastes más importantes

HIPÓTESIS NULA	REGIÓN DE RECHAZO
$H_0 : \mu = \mu_0$ σ conocida	$R = \{\lvert\overline{x} - \mu_0\rvert > z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\}$
$H_0 : \mu = \mu_0$ σ desconocida	$R = \{\lvert\overline{x} - \mu_0\rvert > t_{n-1;\alpha/2}\frac{s}{\sqrt{n}}\}$
$H_0 : \mu \leqslant \mu_0$ σ conocida	$R = \{\overline{x} - \mu_0 > z_{\alpha}\frac{\sigma}{\sqrt{n}}\}$
$H_0 : \mu \leqslant \mu_0$ σ desconocida	$R = \{\overline{x} - \mu_0 > t_{n-1;\alpha}\frac{s}{\sqrt{n}}\}$
$H_0 : \mu \geqslant \mu_0$ σ conocida	$R = \{\overline{x} - \mu_0 < z_{1-\alpha}\frac{\sigma}{\sqrt{n}}\}$
$H_0 : \mu \geqslant \mu_0$ σ desconocida	$R = \{\overline{x} - \mu_0 < t_{n-1;1-\alpha}\frac{s}{\sqrt{n}}\}$

Universidad Europea de Madrid

Contrastes más importantes

HIPÓTESIS NULA	REGIÓN DE RECHAZO
$H_0 : \sigma = \sigma_0$	$R = \{\frac{n-1}{\sigma_0^2} s^2 \notin [\chi^2_{n-1;1-\alpha/2}, \chi^2_{n-1;\alpha/2}]\}$
$H_0 : \sigma \leqslant \sigma_0$	$R = \{\frac{n-1}{\sigma_0^2} s^2 > \chi^2_{n-1;\alpha}\}$
$H_0 : \sigma \geqslant \sigma_0$	$R = \{\frac{n-1}{\sigma_0^2} s^2 < \chi^2_{n-1;1-\alpha}\}$

Dos poblaciones

HIPÓTESIS NULA	REGIÓN DE RECHAZO
$H_0 : \mu_1 = \mu_2$ $\sigma_1,\ \sigma_2$ conocidas	$R = \{\lvert\overline{x} - \overline{y}\rvert > z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\}$
$H_0 : \mu_1 = \mu_2$ $\sigma_1 = \sigma_2$ desconocidas	$R = \{\lvert\overline{x} - \overline{y}\rvert > t_{n_1+n_2-2;\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\}$
$H_0 : \mu_1 \leqslant \mu_2$ $\sigma_1,\ \sigma_2$ conocidas	$R = \{\overline{x} - \overline{y} > z_{\alpha}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\}$
$H_0 : \mu_1 \leqslant \mu_2$ $\sigma_1 = \sigma_2$ desconocidas	$R = \{\overline{x} - \overline{y} > t_{n_1+n_2-2;\alpha} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\}$
$H_0 : \mu_1 \geqslant \mu_2$ $\sigma_1,\ \sigma_2$ conocidas	$R = \{\overline{x} - \overline{y} < z_{1-\alpha}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\}$
$H_0 : \mu_1 \geqslant \mu_2$ $\sigma_1 = \sigma_2$ desconocidas	$R = \{\overline{x} - \overline{y} < t_{n_1+n_2-2;1-\alpha} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\}$

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

Dos poblaciones

HIPÓTESIS NULA	REGIÓN DE RECHAZO
$H_0 : \sigma_1 = \sigma_2$	$R = \{\frac{s_1^2}{s_2^2} \notin [F_{n_1-1,n_2-1;1-\alpha/2}, F_{n_1-1,n_2-1;\alpha/2}]\}$
$H_0 : \sigma_1 \leqslant \sigma_2$	$R = \{\frac{s_1^2}{s_2^2} > F_{n_1-1,n_2-1;\alpha}\}$
$H_0 : \sigma_1 \geqslant \sigma_2$	$R = \{\frac{s_1^2}{s_2^2} < F_{n_1-1;n_2-1;1-\alpha}\}$

Universidad Europea de Madrid