MAT-269: Sesión 10
Test de hipótesis I

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Definición 1

Una hipótesis es un enunciado con respecto a un parámetro poblacional.

Problema 1

El objetivo de un test de hipótesis es decidir, *basado en una muestra*, acerca de la población. Para este efecto, determinaremos cual de dos hipótesis complementarias es verdadera. Escribiremos este tipo de hipótesis, por ejemplo, como:

$$H_0 : \theta = \theta_0, \qquad \text{versus} \qquad H_1 : \theta = \theta_1,$$

aquí H_0 es llamada hipótesis nula, mientras que H_1 se denomina hipótesis alternativa.

EX UMBRA IN SOLEM

Preliminares: Test de hipótesis

En general, para un modelo estadístico $\mathcal{P} = \{\mathsf{P}_\theta : \theta \in \Theta\}$ basado en una muestra x_1, $\dots, x_n$, un test de hipótesis es formulado mediante introducir una partición en el espacio paramétrico de la forma:

$$\Theta = \Theta_0 \cup \Theta_1, \qquad \text{con} \qquad \Theta_0 \cap \Theta_1 = \varnothing$$

en cuyo caso, podemos escribir:

$$H_0 : \theta \in \Theta_0, \qquad \text{versus} \qquad H_1 : \theta \in \Theta_1,$$

se debe destacar que el modelo bajo H_0 asume la forma

$$\mathcal{P}_0 = \{\mathsf{P}_\theta : \theta \in \Theta_0\},$$

en cuyo caso decimos que la hipótesis H_0 es verdadera si la muestra se distribuye de acuerdo a $\mathsf{P}_\theta \in \mathcal{P}_0$.

Definición 2

Un test de hipótesis es un procedimiento o una regla para decidir si se debe aceptar o rechazar la hipótesis nula basado en un estadístico de prueba $T(\boldsymbol{X})$.[1]

Definición 3: Regla de rechazo

Aquél subconjunto del espacio muestral para el que H_0 es rechazada se denomina región de rechazo (o región crítica). La hipótesis nula será rechazada si $T(\boldsymbol{X})$ es "muy grande", esto es, se rechaza H_0 si:

$$T(\boldsymbol{X}) \geq C,$$

para algún valor crítico C.

[1] Evidentemente la aceptación/rechazo de H_0 es equivalente a la aceptación/rechazo de H_1, pero tradicionalmente esta es especificada en términos de H_0.

Problema 2

Deseamos probar hipótesis nolineales de la forma $H_0 : \boldsymbol{g}(\boldsymbol{\theta}) = \boldsymbol{0}$, donde $\boldsymbol{g} : \mathbb{R}^p \to \mathbb{R}^q$, tal que $\boldsymbol{G}(\boldsymbol{\theta}) = \partial \boldsymbol{g}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^\top$ es una matriz $q \times p$ con rango q. En otras palabras, deseamos resolver el problema restringido:

$$\max_{\theta \in \Theta} \ell_n(\boldsymbol{\theta}) \qquad \text{sujeto a:} \quad \boldsymbol{g}(\boldsymbol{\theta}) = \boldsymbol{0}.$$

Casos particulares:

1. Considere la hipótesis simple:

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0, \qquad \text{versus} \qquad H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0.$$

2. Suponga los contrastes lineales del tipo:

$$H_0 : \boldsymbol{R\theta} = \boldsymbol{r} \qquad \text{versus} \qquad H_1 : \boldsymbol{R\theta} \neq \boldsymbol{r},$$

donde $\boldsymbol{R} \in \mathbb{R}^{q \times p}$ con $\operatorname{rg}(\boldsymbol{R}) = q$ y $\boldsymbol{r} \in \mathbb{R}^q$.

Definición 4: Test de razón de verosimilitudes (Wilks, 1938)

El test de razón de verosimilitudes (LRT) para probar la hipótesis $H_0 : \boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}$ asume la forma

$$LR = 2\{\ell_n(\widehat{\boldsymbol{\theta}}) - \ell_n(\widetilde{\boldsymbol{\theta}})\},$$

donde $\widetilde{\boldsymbol{\theta}}$ y $\widehat{\boldsymbol{\theta}}$ representan el MLE bajo H_0 y H_1, respectivamente. Además, bajo H_0, LR tiene una distribución asintóticamente χ^2 con q grados de libertad.

Observación:

De este modo, debemos rechazar H_0 si:

$$RC = \{LR \geq \chi^2_{1-\alpha}(q)\},$$

donde $\chi^2_{1-\alpha}(q)$ denota un valor cuantil $1-\alpha$ de la distribución chi-cuadrado con q grados de libertad.

Definición 5: Test de Wald (1943)

El test de Wald es: rechazar H_0 cuando:

$$\{W \geq \chi^2_{1-\alpha}(q)\},$$

donde

$$W = n\boldsymbol{g}^\top(\widehat{\boldsymbol{\theta}})(\boldsymbol{G}\boldsymbol{\mathcal{F}}^{-1}(\widehat{\boldsymbol{\theta}})\boldsymbol{G}^\top)^{-1}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}),$$

con $\boldsymbol{G} = \partial\boldsymbol{g}(\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}^\top$. Tenemos que, bajo H_0, $W \xrightarrow{\mathsf{D}} \chi^2(q)$.

Definición 6: Test score o de multiplicadores de Lagrange (Rao, 1948)

El test de score es definido por el estadístico de prueba:

$$R = \frac{1}{n}\boldsymbol{U}_n^\top(\widetilde{\boldsymbol{\theta}})\boldsymbol{\mathcal{F}}^{-1}(\widetilde{\boldsymbol{\theta}})\boldsymbol{U}_n(\widetilde{\boldsymbol{\theta}}),$$

y rechazamos $H_0 : \boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}$, cuando:

$$\{R \geq \chi^2_{1-\alpha}(q)\}.$$

Definition 7: Test de forma bilineal (Crudu y Osorio, 2020)

El estadístico de forma bilineal para probar la hipótesis nula $H_0 : \boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}$, asume la forma

$$BF_1 = \widetilde{\boldsymbol{\lambda}}^\top \boldsymbol{g}(\widehat{\boldsymbol{\theta}}),$$

donde $\widetilde{\boldsymbol{\lambda}}$ denota el vector de multiplicadores de Lagrange[2]. Además, bajo H_0, BF_1 tiene distribución asintótica chi-cuadrado con q grados de libertad.

Oservación:

El estadístico BF_1 puede ser escrito de forma equivalente como

$$BF_2 = \boldsymbol{U}_n^\top(\widetilde{\boldsymbol{\theta}})\boldsymbol{G}^\top(\boldsymbol{G}\boldsymbol{G}^\top)^{-1}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}).$$

Además, bajo H_0 sigue que $BF_2 \xrightarrow{\mathsf{D}} \chi^2(q)$. De este modo, se rechaza H_0, si:

$$\{BF_k \geq \chi^2(q)\}, \qquad \text{para } k = 1, 2.$$

[2] estimados bajo $H_0 : \boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}$.

Anteriormente se determinó los MLE de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ para la distribución normal multivariada.

En efecto, sabemos que si $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ es una muestra aleatoria desde $\mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Entonces,

(a) $\overline{\boldsymbol{x}} \sim \mathsf{N}_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$ independiente de $\boldsymbol{Q} = (n-1)\boldsymbol{S} \sim \mathsf{W}_p(n-1, \boldsymbol{\Sigma})$.

(b) $T^2 = n(\overline{\boldsymbol{x}} - \boldsymbol{\mu})^\top \boldsymbol{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}) \sim \mathsf{T}^2(p, n-1)$ es la distribución T^2 de Hotelling con p y $n-1$ grados de libertad.

El estadístico T^2 puede ser usado para probar $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. De este modo, bajo H_0 tenemos

$$T_0^2 = n(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^\top \boldsymbol{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0) \sim \mathsf{T}^2(p, n-1).$$

Es decir, cuando H_0 es verdadera

$$F_0 = \frac{T_0^2}{n-1}\frac{n-p}{p} \sim F(p, n-p),$$

y rechazamos H_0 si:

$$F_0 \geq F_{1-\alpha}(p, n-p).$$

Resultado 1

El estadístico de prueba T_0^2 es equivalente al test de razón de verosimilitudes

$$\Lambda = \frac{\max_{\Sigma} L_n(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})}{\max_{\mu,\Sigma} L_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})},$$

de ahí que

$$LR = -2\log\Lambda = 2(\ell_n(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) - \ell_n(\boldsymbol{\mu}_0, \widetilde{\boldsymbol{\Sigma}})).$$

Demostración:

Sabemos que

$$\max_{\mu,\Sigma} L_n(\boldsymbol{\mu},\boldsymbol{\Sigma}) = L_n(\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\Sigma}}) = (2\pi)^{-np/2}|\widehat{\boldsymbol{\Sigma}}|^{-n/2}\, e^{-np/2},$$

ahora note que $L_n(\boldsymbol{\mu}_0,\boldsymbol{\Sigma})$ o bien $\ell_n(\boldsymbol{\mu}_0,\boldsymbol{\Sigma})$ es maximizada en $\boldsymbol{\Sigma} = \widetilde{\boldsymbol{\Sigma}}_0$, con

$$\widetilde{\boldsymbol{\Sigma}}_0 = \frac{1}{n}\sum_{i=1}^n (\boldsymbol{x}_i - \boldsymbol{\mu}_0)(\boldsymbol{x}_i - \boldsymbol{\mu}_0)^\top,$$

de ahí que

$$\max_{\Sigma} L_n(\boldsymbol{\mu}_0,\boldsymbol{\Sigma}) = L_n(\boldsymbol{\mu}_0,\widetilde{\boldsymbol{\Sigma}}_0) = (2\pi)^{-np/2}|\widetilde{\boldsymbol{\Sigma}}_0|^{-n/2}\, e^{-np/2}.$$

De este modo, el estadístico de razón de verosimilitudes es dado por:

$$\Lambda = \Big(\frac{|\widetilde{\boldsymbol{\Sigma}}_0|}{|\widehat{\boldsymbol{\Sigma}}|}\Big)^{-n/2}.$$

Es fácil notar que,

$$\begin{aligned}\widetilde{\boldsymbol{\Sigma}}_0 &= \frac{1}{n}\Big(\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top + n(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^\top\Big)\\ &= \widehat{\boldsymbol{\Sigma}} + (\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^\top,\end{aligned}$$

esto lleva a

$$\begin{aligned}|\widetilde{\boldsymbol{\Sigma}}_0| &= |\widehat{\boldsymbol{\Sigma}} + (\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^\top| = |\widehat{\boldsymbol{\Sigma}}\,(\boldsymbol{I} + \widehat{\boldsymbol{\Sigma}}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^\top)|\\ &= |\widehat{\boldsymbol{\Sigma}}||\boldsymbol{I} + \widehat{\boldsymbol{\Sigma}}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^\top| = \big(1 + (\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^\top\widehat{\boldsymbol{\Sigma}}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)\big)|\widehat{\boldsymbol{\Sigma}}|\end{aligned}$$

De este modo,

$$\frac{|\widetilde{\boldsymbol{\Sigma}}_0|}{|\widehat{\boldsymbol{\Sigma}}|} = 1 + (\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^\top\widehat{\boldsymbol{\Sigma}}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0).$$

Ahora,

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top = \frac{n-1}{n}\boldsymbol{S},$$

es decir, podemos escribir

$$\frac{|\widetilde{\boldsymbol{\Sigma}}_0|}{|\widehat{\boldsymbol{\Sigma}}|} = 1 + \frac{n}{n-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^\top \boldsymbol{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0) = 1 + \frac{T_0^2}{n-1}.$$

Es decir,

$$\Lambda = \Big(1 + \frac{T_0^2}{n-1}\Big)^{-n/2},$$

y

$$LR = -2\log\Lambda = n\log\Big(1 + \frac{T_0^2}{n-1}\Big).$$

Sabemos que, bajo $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$,

$$-2\log\Lambda \xrightarrow{\mathsf{D}} \chi^2(p),$$

lo que lleva a la región de rechazo (asintótica):

$$n\log\left(1 + \frac{T_0^2}{n-1}\right) \geq K,$$

donde $K = \chi^2_{1-\alpha}(p)$.

Note que LR es una función monótona de T_0^2 de ahí que rechazar H_0 si $LR \geq K$ es equivalente a rechazar H_0 si $T_0^2 \geq K'$, o bien, si:

$$F_0 = \left(\frac{n-p}{p}\right)\frac{T_0^2}{n-1} \geq F_{1-\alpha}(p, n-p).$$

Test para la media

Considere hipótesis del tipo $H_0 : \boldsymbol{A\mu} = \boldsymbol{a}$ donde $\boldsymbol{A} \in \mathbb{R}^{q\times p}$ con rango q y $\boldsymbol{a} \in \mathbb{R}^q$. Haciendo, $\boldsymbol{y}_i = \boldsymbol{A}\boldsymbol{x}_i$, $i = 1,\dots,n$, sigue que

$$\boldsymbol{y}_i \sim \mathsf{N}_q(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A}^\top), \qquad i = 1,\dots,n.$$

De este modo, es posible probar $H_0 : \boldsymbol{A\mu} = \boldsymbol{a}$ versus $H_1 : \boldsymbol{A\mu} \neq \boldsymbol{a}$ usando:

$$\begin{aligned} T_0^2 &= n(\overline{\boldsymbol{y}} - \boldsymbol{a})^\top \boldsymbol{S}_Y(\overline{\boldsymbol{y}} - \boldsymbol{a}) \\ &= n(\boldsymbol{A}\overline{\boldsymbol{x}} - \boldsymbol{a})^\top (\boldsymbol{ASA}^\top)^{-1}(\boldsymbol{A}\overline{\boldsymbol{x}} - \boldsymbol{a}), \end{aligned}$$

bajo H_0, sigue que $T_0^2 \sim \mathsf{T}^2(q, n-1)$ y rechazamos $H_0 : \boldsymbol{A\mu} = \boldsymbol{a}$, si:

$$\Big(\frac{n-q}{q}\Big)\frac{T_0^2}{n-1} \geq F_{1-\alpha}(q, n-q).$$

Referencias bibliográficas

Crudu, F., Osorio, F. (2020).
Bilinear form test statistics for extremum estimation.
Economics Letters **187**, 108885.

Rao, C.R. (1948).
Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation.
Proceedings of the Cambridge Philosophical Society **44**, 50-57.

Terrell, G.R. (2002).
The gradient statistic.
Computing Sciences and Statistics **34**, 206-215.

Wald, A. (1943).
Test of statistical hypothesis concerning several parameters when the number of observations is large.
Transactions of the American Mathematical Society **54**, 426-482.

Wilks, S.S. (1938).
The large-sample distribution of the likelihood ratio for testing composite hypothesis.
The Annals of Mathematical Statistics **9**, 60-62.