inferencia vec medias 2

Monday, February 21, 2022 7:37 AM

Inferencia sobre el vec de medias Pt 2

$\Lambda = \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|} \right)^{n/2}$ si $< C_{\Lambda,\alpha}$ rechazo Ho.

Para n grande
$-2 \ln(\Lambda) \sim \chi^2_p$

$\det(S) = |\hat{\Sigma}|$
$\det(\Sigma_0) = |\hat{\Sigma}_0|$

p-val pequeño es rechazo.
val-p $\leq \alpha$ rechazo Ho.

Ho: $\mu = \mu_0$.

$\Sigma_0 = \frac{1}{n} (X - 1\mu_0')'(X - 1\mu_0')$

Relaº entre T^2 y razón de verosimilitud
$(\Lambda)^{2/n} = \left(1 + \frac{T^2}{n-1}\right)^{-1}$ están inv. relacionados.

"Rechazo Ho para T^2 grande."
"Λ pequeño."

Λ es más potente q' el T^2.

n = # obs
p = # vars

$T^2 = n(\bar{x} - \mu_0)' S^{-1} (\bar{x} - \mu_0)$

X es $n \times p$.

I.C ⇒ Regiones de confianza

Sea θ un vec. de parámetros y Θ un conjunto de posibles vals. de θ.

Una región de confianza es una región en Θ determinada por los datos y se denota $R(X)$ con X una matriz de datos.

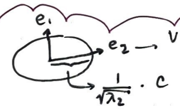
$R(X)$ es una región con confianza de $1-\alpha$ $(100(1-\alpha)\%)$ si
$P(R(X)$ contenga al verdadero $\theta) = 1-\alpha$

Si la poblaº es normal p-variada $R(X)$ para μ se deduce de:

$$\boxed{P\left(n(\bar{x}-\mu)' S^{-1} (\bar{x}-\mu) \leq \frac{(n-1)p}{n-p} \cdot F_{p,n-p}(\alpha)\right) = 1-\alpha} \quad (1)$$

Así, $\bar{x}$ está a una distancia de μ $\leq \underbrace{\left(\frac{p(n-1)}{n-p} F_{p,n-p}(\alpha)\right)^{1/2}}_{c}$ con proba. $1-\alpha$.

$x A x' \leq c^2$



(La distancia está medida en términos de nS^{-1})

Dada una muestra, se calculan $\bar{x}$ y S, para obtener la región de confianza a partir de la desigualdad.

- ↳ $R(X)$ será una hiperelipsoide centrada en $\bar{x}$
- ↳ Para determinar si μ_0 es un val. posible d. μ, hay q' hay q' evaluar la dist. (o la dist al 2) de $\bar{x}$.
- ↳ $R(\bar{x})$ es el conjunto de vals de μ_0 para los q' T^2 hace q' no se rechace Ho.

↳ Los ejes de la hiperelipsoide $R(X)$ son en dirección a los vecs. propios de nS^{-1} (mismos de S) y sus long. se determinan usando:

$n(\bar{x}-\mu)' S^{-1} (\bar{x}-\mu) \leq \frac{p(n-1)}{n-p} F_{p,n-p}(\alpha)$

$\sqrt{\frac{\lambda_i}{n}} \cdot c = \sqrt{\lambda_i} \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)}$

entonces, centrado en $\bar{x}$, los ejes de confianza son

$\bar{x} \pm \underbrace{\sqrt{\lambda_i}}_{\text{val propio}} \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)} \cdot \underbrace{e_i}_{\text{vect. propio de } S}$

Alternativas

1 - Intervalos de confianza uno a uno (simultáneos)
2 - Intervalos TC
3 - Bonferroni.

1 - Construir $R(X)$ para μ nos da info sobre las medias indiv.

Sea $X \sim N_p(\mu, \Sigma)$ $\quad z = a_1 x_1 + \dots + a_p x_p = a' X$.

uno-a-uno

$z \sim N_1(a'\mu, a'\Sigma a)$
$\mu_z = E[z] = a'\mu$
$\sigma^2 = var(z) = a'\Sigma a$

$a' = [\,]\cdots$

Si se tiene una m.a $X_1, \dots, X_n$ se puede crear la muestra $z_1, \dots, z_n$ usando:

$z_j = a_1 x_{j1} + \dots + a_p x_{jp}$

$\bar{z} = a'\bar{x} \rightarrow$ media muestral aleat.

$s^2_{z} = a' s a$. → varianza.

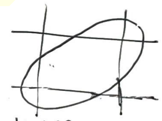
Si a fijo y σ_z^2 → desconocida.
el IC con confianza $1-\alpha$ de $\mu_z = a'\mu$ se construye:

$$z = \frac{\bar{z} - \mu_z}{s_z/\sqrt{n}} = \frac{\sqrt{n}(a'\bar{x} - a'\mu)}{\sqrt{a's a}} \sim t_{n-1} \text{ gl.}$$

y queda def. por los lims:

$$\bar{z} \pm t_{n-1}(\alpha/2) \frac{s_z}{\sqrt{n}}$$

$$a'\bar{x} \pm t_{n-1}(\alpha/2) \frac{\sqrt{a'sa}}{\sqrt{n}}$$



— uno a la vez
○ Región de confianza.

La región es mucho más precisa
↳ Es la única q' toma en cuenta toda la info (covarianzas, ...).

2. Intervalos T^2.
↳ Con $R(x)$ buscamos un IC "colectivo" de $1-\alpha$ pueda ser generado $\forall a$ contenga a μ_z de forma simult.

Problema → causa intervalos más grande → menos presi°.

Sean $x_1, \dots, x_n$ una m.a $N_p(\mu, \Sigma)$. Σ def pos, $\forall a$ el IC.

$$a'\bar{x} \pm \sqrt{\frac{p(n-1)}{n(n-p)} F_{n,n-p}(\alpha)\, a' s a}$$

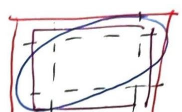
contiene $a'\mu$ con proba $1-\alpha$.

Dem:

$$T^2 \le c^2 \Rightarrow \frac{n(a'\bar{x} - a'\mu)^2}{a' s a} \le c^2.$$

$$\Rightarrow a'\bar{x} - c\sqrt{\frac{a'sa}{n}} \le a'\mu \le a'\bar{x} + c\sqrt{\frac{a'sa}{n}}$$

$$c^2 = \frac{p(n-1)}{n-p} F_{n,n-p}(\alpha)$$



T^2 $R(x)$ uno a uno
↓ El menos potente. ↳ El más!
Bonferroni
↳ punto medio.

3. Bonferroni → Recomendado para $p < 13$ (menos de 13 vars.)
T^2 > Bonferronis uno-a-uno
(q' tan estrecho)

Prob (todos los intervalos son válidos) $= 1 - P$ (por lo menos 1 intervalo es inválido)
$\ge 1 - \sum_{i=1}^{n} P$ (i-ésimo intervalo es falso)
$= 1 - \sum^{m} 1 - P$ (i-ésimo intervalo es válido)
$= 1 - \sum_{i=1}^{m} \alpha_i$

P (todos los intervalos son válidos) $\ge 1 - (\alpha_1 + \dots + \alpha_n)$.

$\alpha_i = \frac{\alpha}{m}$ (parto α en mis m medias de mis m vars.)

P(todos los intervalos son válidos) $\ge 1 - (\frac{\alpha}{m} + \dots \frac{\alpha}{m}) = 1 - \alpha$

$$a'\bar{x} \pm t_{n-1}\left(\frac{\alpha}{2m}\right) \sqrt{\frac{a'sa}{n}} \quad m = \# \text{ de vars.}$$

Hemos asumido $x_j \sim N_p(\mu, \Sigma)$
└ Cuando no es normal:
↳ si n grande: Sea $x_1, \dots, x_n$ una m.a con media μ y cov Σ.

$T^2 = n(\bar{x} - \mu_0) S^{-1} (\bar{x} - \mu_0) \underset{\text{aprox.}}{\approx} \chi^2_p$ → suma de ~Normales estándar al cuadrado

$$P\left(n(\bar{x} - \mu_0)' S^{-1} (\bar{x} - \mu_0) \le \chi^2_p(\alpha)\right) \approx 1 - \alpha.$$

$n \to \infty$: $\frac{(n-1)p}{n-p} F_{p,n-p} \to \chi^2_p$.
↓
$n - p \ge 30 \sim 50$

P.H: $H_0 = \mu = \mu_0$ rechazo H_0 si $T^2 > \chi^2_p(\alpha)$.

Intervalos T^2

$a'\bar{x} \pm \sqrt{\chi_p^2(\alpha)} \sqrt{\frac{a's a}{n}}$ → Ya no requiero q' sea normal! wiii n

↳ Puedo saber como se van
a comportar con una
alta confianza.