Índice general

0.1. Algunos teoremas útiles

Teorema: La función paramétrica $\Psi = c^T\beta$ es estimable si y solo si c es una combinación lineal de las filas de la matriz de diseño $\mathbf{X}$, o sea si existe $a \in \mathbb{R}^n$ tal que $c^T = a^T\mathbf{X}$.

Teorema de Gauss Markov: Supongamos que vale el modelo $\Omega : \mathbf{E}(Y) = \mathbf{X}\beta, \Sigma_Y = \sigma^2 I$. Toda función estimable $\Psi = c^T\beta$ tiene un únivo estimador lineal insesgado de mínima varianza $\hat{\Psi}$. Este estimador puede obtenerse reemplazando β por el estimador de mínimos cuadrados $\hat{\beta}$.

0.2. Un poco más de teoría

Distribución Normal Multivariada

Se dice que el vector aleatorio X tiene distribución normal multivariada de dimensión p ,$X \sim \mathcal{N}_p(\mu, \Sigma)$, de parametros $\mu \in \mathbb{R}^p$ y $\Sigma \in \mathbb{R}^{pxp}$ (simétrica y definida positiva) si su función de densidad conjunta está dada por

$$f_X(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-1/2(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

1. Si $X \sim \mathcal{N}_p(0, diag(\lambda_1, \ldots, \lambda_p))$, entonces $X_1, \ldots, X_p$ son independientes y tienen distribución $\mathcal{N}(0, \lambda_i)$.

2. Si $X \sim \mathcal{N}_p(\mu, \Sigma)$ y $A \in \mathbf{R}^{pxp}$ es no singular, entonces $AX + b \sim \mathcal{N}_p(A\mu + b, A\Sigma A^T)$

Def: Se dice que X es normal multivariada si y solo si $\forall \in \mathbb{R}^p$ se tiene que $t^T X$ es normal univariada.

Es decir:

$$X \sim \mathcal{N}_p(\mu, \Sigma) \iff a^t X \sim \mathcal{N}_p(a^t\mu, a^t\Sigma t), \forall a \in \mathbb{R}^p$$

Además, $X_1, \ldots, X_p$ son independientes si y solo si Σ es una matriz diagonal.

Propiedades: Sea $X \sim \mathcal{N}_p(\mu, \Sigma)$, $\Sigma > 0$, $X = (X^{(1)}, X^{(2)})^T$, $X^{(1)} \in \mathbb{R}^{p_1}$, $X^{(2)} \in \mathbb{R}^{p_2}$, $p_1+p_2 = p$

1. $X^{(1)} \sim \mathcal{N}_{p_1}(\mu^{(1)}, \Sigma_{11})$,$X^{(2)} \sim \mathcal{N}_{p_2}(\mu^{(2)}, \Sigma_{22})$

2. $X^{(1)}$ y $X^{(2)}$ son independientes sii $\Sigma_{12} = 0$

3. $A \in \mathbf{R}^{qxp}$, $rg(a) = q$, $AX \sim \mathcal{N}_q(A\mu, A\Sigma A^T)$

4. $X \sim \mathcal{N}_p(\mu, \Sigma) \iff X = AZ + \mu$, con $Z \sim \mathcal{N}_p(0, I_p)$ y $AA^T = \Sigma$

5. $(X - \mu)^t\Sigma^{-1}(X - \mu) \sim \chi^2_p$

IMPORTANTE: Complementar con el material bibliográfico recomendado.

0.3. Test y regiones de confianza

Ahora debemos agregar al modelo un supuesto adicional: Normalidad conjunta de los errores. Entonces, dadas X fijas tenemos que:

$$\Omega: \quad Y \sim \mathcal{N}_n(\mathbf{X}\beta, \sigma^2 I), \quad rg(\mathbf{X}) = r, \quad \beta \in \mathbb{R}^p$$

A partir de este modelo podremos deducir test de hipótesis, intervalos de confianza de nivel exacto para funciones paramétricas estimables, conjuntos o regiones de confianza simultáneos.

Sabemos que $\hat{\beta}$ y S^2 son funciones de estadísticos suficientes y completos, por lo tanto son IMVU.

Vamos a querer plantear hipótesis del tipo $C\beta = \delta$. Tendremos funciones estimables $\Psi_i = c^T\beta$, entonces usaremos como estimador a $\hat{\Psi} = C\hat{\beta}$. Observando su forma matricial, $\Psi = C\beta$ con $C \in \mathbb{R}^{qxp}$, con $c_1, \ldots, c_q$ linealmente independientes (es una propiedad necesaria para que Ψ sea estimable. Para más detalle ver Seber, Linear Analisys). Bajo estas condiciones, se tiene que $rg(C) = q$. En el caso de rango completo de la matriz de diseño, tendremos que $r = p$.

Teorema: Supongamos que se cumple el teorema

$$\Omega: \quad Y \sim \mathcal{N}_n(\mathbf{X}\beta, \sigma^2 I), \quad rg(\mathbf{X}) = r, \quad \beta \in \mathbb{R}^p$$

Entonces:

1. $\hat{\beta} \sim \mathcal{N}_p(\beta, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1})$
2. $\frac{(\hat{\beta}-\beta)^T(\mathbf{X}^T\mathbf{X})(\hat{\beta}-\beta)}{\sigma^2} \sim \chi^2_p$
3. $\hat{\beta}$ y $(n-p)S^2/\sigma^2$ son independientes
4. $(n-p)S^2/\sigma^2 \sim \chi^2_{n-p}$

Si tenemo que $\hat{\Psi} = C\hat{\beta}$, con $rg(C) = q$, entonces usando las propiedades de normal multivariada vistas en el capítulo anterior tenemos que

$$\hat{\Psi} \sim \mathcal{N}_q(C\beta, \sigma^2 C(\mathbf{X}^T\mathbf{X})C^T)$$

0.3.1. Test de hipótesis

En el anexo 1 se encuentran notas para repasar el tema de test de hipótesis visto en cursos de estadística.

Si buscamos realizar un test de hipótesis para un solo parámetro, por ejemplo queremos saber si la variable X_1 es significativa al momento de explicar la variabilidad de nuestra variable Y, podríamos plantear las siguientes hipótesis:

$$H_0: \beta_1 = 0 \quad Vs. \quad H_1: \beta_1 \neq 0$$

Entonces, un test de hipótesis con nivel de significación α será

$$\varphi(\underline{X}) = \begin{cases} 1 & \text{si} \quad |T| > k_\alpha \\ 0 & \text{en otro caso} \end{cases}$$

Con $T = \frac{\hat{\beta}}{S\sqrt{d_{ii}}}$, que bajo H_0 tiene distribución t de student con $n-p$ grados de libertad. (Llamamos d_{ii} al elemento diagonal de la matriz $D = (\mathbf{X}^T\mathbf{X})^{-1}$).

¿Porque elijo esa variable T? Dada la información que tenemos, podemos decir que $\frac{c^T\hat{\beta}-c^T\beta}{\sqrt{\sigma^2 c^T(\mathbf{X}^T\mathbf{X})^{-1})c}} \sim \mathcal{N}(0,1)$ y $(n-p)S^2/\sigma^2 \sim \chi^2_{n-p}$, entonces

$$T = \frac{c^T\hat{\beta} - c^T\beta}{\sqrt{S^2 c^T(\mathbf{X}^T\mathbf{X})^{-1})c}} \sim t_{n-p}$$

Bajo H_0, reemplazando resulta $T = \frac{\hat{\beta}}{S\sqrt{d_{11}}}$

Si llamamos T_{obs} al valor del estadístico T definido calculado en base a las observaciones $(x_i, y_i), i = 1, \ldots, n$, se puede calcular el *p-valor* de la siguiente forma

$$p - valor = 2\mathbf{P}(T \geq |T_{obs}|), T \sim t_{n-p}$$

ya que se trata de un test a dos colas. Reportar el p-valor cuando uno realiza un test sobre un conjunto de datos siempre permite al lector elegir su punto de corte respecto de rechazar o no una hipótesis.

Caso general

Un test exacto para

$$H_0 : C\beta = \delta \quad Vs. \quad H_1 : C\beta \neq \delta$$

será

$$\varphi(\underline{X}) = \begin{cases} 1 & \text{si} \quad F > \mathcal{F}_{q,n-p,1-\alpha} \\ 0 & \text{en otro caso} \end{cases}$$

ya que bajo H_0

$$F = \frac{(C\hat{\beta} - \delta)^T (C(\mathbf{X}^T\mathbf{X})^{-1}C^T)^{-1}(C\hat{\beta} - \delta)}{qS^2} \sim \mathcal{F}_{q,n-p}$$

En este caso, el p-valor se calcula como

$$p - valor = \mathbf{P}(F \geq F_{obs}), \quad F \sim \mathcal{F}_{q,n-p}$$

Para comparar con el caso particular, recordar que $\mathcal{F}_{1,n} = (t_n)^2$

El estadístico F se deduce a partir del test de cociente de verosimilitud mencionado en el *Anexo 1*, donde

$$f(y) = \frac{1}{(2\pi)^{n/2}} \frac{1}{(\sigma^2)^{n/2}} exp\{\frac{-1}{2\sigma^2}||y - X\beta||^2\}$$

Significación de la regresión

Supongamos un modelo con intercept, vamos a querer saber si la regresión es significativa, esto es si las variables que proponemos como predictoras realmente pueden tener una relación lineal con y bajo el modelo Ω. Entonces, las hipótesis a plantear serán:

$$H_0 : \beta_1 = \cdots = \beta_{p-1} = 0 \quad Vs. \quad H_1 : Algun \quad \beta_i \neq 0, \quad i = 1, \ldots, p-1$$

Si lo vemos como un caso particular del caso general presentado anteriormente, podemos ver que

$\Psi = (\beta_1, \ldots, \beta_{p-1})^T$, $\delta = (0, \ldots, 0)^T$,

$$\mathbf{C} = \begin{bmatrix} 0 & 1 \ldots & 0 \\ 0 & . & . \\ . & . & . \\ . & . & . \\ 0 & \ldots & 1 \end{bmatrix}$$

En este caso, $q = p - 1$, y el test resulta

$$\varphi(\underline{X}) = \begin{cases} 1 & \text{si} \quad F > \mathcal{F}_{p-1,n-p,1-\alpha} \\ 0 & \text{en otro caso} \end{cases}$$

Ejemplo: Boston Housing