Errores Normales

Si los errores además de ser independientes con media 0 y varianza σ^2, son normales, es decir

$$\epsilon_i \sim N(0, \sigma^2) \, i = 1, \ldots, n$$

el modelo Ω puede formularse de la siguiente forma:

$$\Omega \quad : \qquad \mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I})$$

En este caso puede demostrarse que el estimador de máxima verosimilitud de θ coincide con el de mínimos cuadrados.
(Ejercicio del TP)

Intermezzo

La Normal Multivariada

Normal Multivariada

Sean $\boldsymbol{\mu} \in \mathbb{R}^p$ y $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$ **simétrica y definida positiva**

Se dice que $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ si su densidad está dada por

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

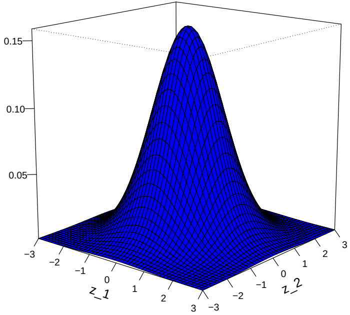
Normal Multivariada

Sean $\boldsymbol{\mu} \in \mathbb{R}^p$ y $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$ **simétrica y definida positiva**

Se dice que $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ si su densidad está dada por

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

Densidad de $\mathbf{Z} \sim N(\mathbf{0}_2, \mathbf{I}_2)$



Normal Multivariada

Sean $\boldsymbol{\mu} \in \mathbb{R}^p$ y $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ **simétrica y definida positiva**

Se dice que $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ si su densidad está dada por

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

<u>Nota</u>: la densidad conjunta depende de $\mathbf{x}$ a través de

$$D^2 = (\mathbf{x}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$$

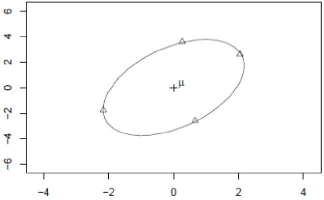
cuadrado de la distancia de Mahalanobis

Distancia de Mahalanobis

- La distancia de Mahalanobis de un punto $\mathbf{x}$ a la media $\boldsymbol{\mu}$ es D siendo

$$D^2 \quad = \quad (\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

- De esta forma, dos puntos tienen la misma distancia de Mahalanobis si están en el mismo elipsoide centrado en $\boldsymbol{\mu}$



Notemos que

- $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ si su densidad está dada por

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{\mathsf{t}} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \mu) \right\}$$

Notemos que

- $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ si su densidad está dada por

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{t}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mu)\right\}$$

- Si $\mathbf{X} \sim N(\mathbf{0}, \mathsf{diag}(\lambda_1, \ldots, \lambda_p))$

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{\prod_{j=1}^{p} \lambda_j^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{p}\frac{x_j^2}{\lambda_j}\right\}$$

Por lo tanto, $X_1, \ldots, X_p$ son independientes $X_j \sim N(0, \lambda_j)$.

Notemos que

- $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ si su densidad está dada por

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{t}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mu)\right\}$$

- Si $\mathbf{X} \sim N(\mathbf{0}, \mathsf{diag}(\lambda_1, \ldots, \lambda_p))$

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{\prod_{j=1}^{p} \lambda_j^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{p}\frac{x_j^2}{\lambda_j}\right\}$$

  Por lo tanto, $X_1, \ldots, X_p$ son independientes $X_j \sim N(0, \lambda_j)$.

- En particular, si $\mathbf{X} \sim N(\mathbf{0}, \mathsf{I}_p)$, $X_1, \ldots, X_p$ son i.i.d. $N(0, 1)$.

Notemos que

- $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ si su densidad está dada por

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{t}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mu)\right\}$$

- Si $\mathbf{X} \sim N(\mathbf{0}, \mathsf{diag}(\lambda_1, \ldots, \lambda_p))$

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{\prod_{j=1}^{p} \lambda_j^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{p}\frac{x_j^2}{\lambda_j}\right\}$$

  Por lo tanto, $X_1, \ldots, X_p$ son independientes $X_j \sim N(0, \lambda_j)$.

- En particular, si $\mathbf{X} \sim N(\mathbf{0}, \mathbf{I}_p)$, $X_1, \ldots, X_p$ son i.i.d. $N(0,1)$.
- Sea $\mathbf{X} = (X_1, \ldots, X_p)^{\mathsf{t}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces,
  $X_1, \ldots, X_p$ son independientes $\Longleftrightarrow \boldsymbol{\Sigma}$ es diagonal.

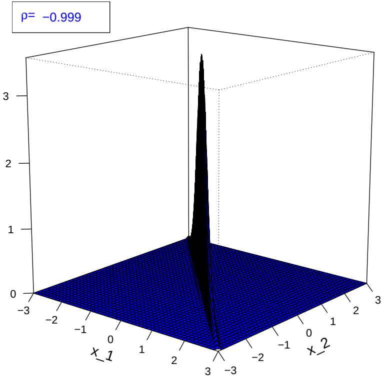
Caso $p = 2$

- Sea $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in \mathbb{R}^2$ y $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ definida positiva ($|\rho| < 1$)

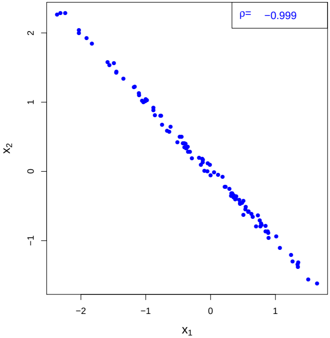
$$
\begin{aligned}
f(\mathbf{x}) = & \quad \frac{1}{2\pi}\,\frac{1}{\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right.\right. \\
+ & \quad \left.\left.\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right\}
\end{aligned}
$$

Caso $p = 2$

Densidad



Datos generados



Propiedades

- Si $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ entonces

$$E(\mathbf{X}) = \boldsymbol{\mu} \quad \text{y} \quad \boldsymbol{\Sigma}_{\mathbf{X}} = \boldsymbol{\Sigma}$$

Propiedades

- Si $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ entonces

$$E(\mathbf{X}) = \boldsymbol{\mu} \quad \text{y} \quad \boldsymbol{\Sigma}_{\mathbf{X}} = \boldsymbol{\Sigma}$$

- $\mathbf{X}$ es $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ si y sólo si $\mathbf{X} = \mathbf{B}\mathbf{Z} + \boldsymbol{\mu}$, donde $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I_p})$ y $\mathbf{B}$ es una matriz de $p \times p$ no singular tal que $\mathbf{B}\mathbf{B}^{\mathsf{t}} = \boldsymbol{\Sigma}$.

Propiedades

- Si $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ entonces

$$E(\mathbf{X}) = \boldsymbol{\mu} \quad \text{y} \quad \boldsymbol{\Sigma}_{\mathbf{X}} = \boldsymbol{\Sigma}$$

- $\mathbf{X}$ es $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ si y sólo si $\mathbf{X} = \mathbf{BZ} + \boldsymbol{\mu}$, donde $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I_p})$ y $\mathbf{B}$ es una matriz de $p \times p$ no singular tal que $\mathbf{BB}^{\mathsf{t}} = \boldsymbol{\Sigma}$.

- Sea $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Dada $\mathbf{A} \in \mathbb{R}^{q \times p}$, $rg(\mathbf{A}) = q \Longrightarrow$ $\mathbf{Y} = \mathbf{AX} + \mathbf{b} \sim N(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathsf{t}})$.

Casos Particulares

En particular, tenemos que

- Una combinación lineal de las componentes es normal:
  Si $\mathbf{X}$ es $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y $\mathbf{a} \in \mathbb{R}^p$, entonces $\mathbf{a}^{\mathsf{t}}\mathbf{X} \sim N(\mathbf{a}^{\mathsf{t}}\boldsymbol{\mu}, \mathbf{a}^{\mathsf{t}}\boldsymbol{\Sigma}\mathbf{a})$.

- Sea $\mathbf{X}^{\mathsf{t}} = (X_1, \ldots, X_p)$ un vector $p-$dimensional con distribución normal multivariada, luego la distribución marginal de cualquier subconjunto de componentes tiene distribución normal multivariada.

- En particular, cada componente es normal:
  Si $\mathbf{X}$ es $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y $\mathbf{a} = \mathbf{e}_i \in \mathbb{R}^p$ es el $i-$ésimo vector canónico en $\mathbb{R}^p$, entonces $\mathbf{e}_i^{\mathsf{t}}\mathbf{X} = X_i \sim N(\mu_i, \sigma_i^2)$, donde $\boldsymbol{\Sigma}_{ii} = \sigma_i^2$.

Relación con la χ^2

- Si $\mathbf{X}$ es $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces

$$(\mathbf{X} - \boldsymbol{\mu})^{\mathsf{t}} \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2_p$$

- ¿Cómo quedaría esto en el caso $N_p(\mathbf{0}, \mathbf{I})$?

Seguimos con el modelo Ω formulado de la siguiente forma:

$$\begin{aligned}\Omega \quad : \qquad & \mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\theta}, \sigma^2\mathbf{I}) \\ & \mathbf{X} \in \mathbb{R}^{n\times p}\,, \text{rango}(\mathbf{X}) = p\end{aligned}$$

En el caso de rango completo, es decir cuando $rg(\mathbf{X}) = p$, obtenemos el siguiente resultado:

Teorema: Supongamos que $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\theta}, \sigma^2\mathbf{I})$, rg$(\mathbf{X}) = p$, $\boldsymbol{\theta} \in \mathbb{R}^p$. Entonces, si $\widehat{\boldsymbol{\theta}}$ es el estimador de mínimos cuadrados de $\boldsymbol{\theta}$, tenemos que

i) $\widehat{\boldsymbol{\theta}} \sim N_p(\boldsymbol{\theta}, \sigma^2(\mathbf{X}^{\mathsf{t}}\mathbf{X})^{-1})$

ii) $\dfrac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\mathsf{t}}(\mathbf{X}^{\mathsf{t}}\mathbf{X})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})}{\sigma^2} \sim \chi^2_p$

iii) $\widehat{\boldsymbol{\theta}}$ y $\dfrac{(n-p)S^2}{\sigma^2}$ son independientes

iv) $\dfrac{(n-p)S^2}{\sigma^2} \sim \chi^2_{n-p}$

v) $\dfrac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\mathsf{t}}(\mathbf{X}^{\mathsf{t}}\mathbf{X})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})}{p\, S^2} \sim F_{p,\, n-p}$

Combinaciones Lineales

En el contexto del Modelo Lineal suelen interesarnos estimar o hacer inferencia sobre combinaciones lineales del vector de parámetros.

Así podría interesarnos, hacer un test o un intervalo sobre alguna componente o estimar el valor esperado de la respuesta en un nuevo $\mathbf{X}_o$.

Si nuestro modelo establece

$$E(Y_o) = \theta_1 X_{o1} + \theta_2 X_{o2} + \cdots + \theta_p X_{op}$$

podríamos tener interés en estimar a $\theta_1 X_{o1} + \theta_2 X_{o1} + \cdots + \theta_p X_{op}$ o testear hipótesis como las siguientes:

$$\begin{aligned} H_o : \theta_j = 0 \quad &\text{vs.} \quad H_1 : \theta_j \neq 0 \\ H_o : \theta_1 - \theta_2 = 0 \quad &\text{vs.} \quad H_1 : \theta_1 - \theta_2 \neq 0 \\ H_o : \theta_2 = \cdots = \theta_p = 0 \quad &\text{vs.} \quad H_1 : \text{existe } j : \theta_j \neq 0 \end{aligned}$$

Combinaciones Lineales

Esto involucra funciones de la forma

$$\varphi = \mathbf{a}^{\mathsf{t}}\boldsymbol{\theta}$$

que son combinaciones lineales del vector de parámetros.

$$\varphi = \mathbf{a}^{\mathsf{t}}\boldsymbol{\theta} \;\Rightarrow\; \widehat{\varphi} = \mathbf{a}^{\mathsf{t}}\widehat{\boldsymbol{\theta}}$$

Podría haber muchos estimadores de una función lineal $\mathbf{a}^{\mathsf{t}}\boldsymbol{\theta}$, si nos enfocamos en los estimadores lineales en $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathsf{t}}$ se puede probar el **Teorema de Gauss–Markov** , que establece que cuando el modelo $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\theta} \quad \boldsymbol{\Sigma}_{\mathbf{Y}} = \sigma^2\mathbf{I}$, rango$(\mathbf{X}) = p$, el estimador $\widehat{\psi}$ propuesto es el de menor varianza dentro de los estimadores insesgados lineales en $\mathbf{Y}$.

(Hay una versión más general para rango $r < p$ y funciones estimables)

Propiedades

Teorema Supongamos que $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\theta}, \sigma^2\mathbf{I})$, $\boldsymbol{\theta} \in \mathbb{R}^p$, rango$(\mathbf{X}) = p$ y definamos

$$\varphi = \mathbf{a}^{\mathsf{t}}\boldsymbol{\theta} \qquad \widehat{\varphi} = \mathbf{a}^{\mathsf{t}}\widehat{\boldsymbol{\theta}}$$

i) $\widehat{\varphi} = \mathbf{a}^{\mathsf{t}}\widehat{\boldsymbol{\theta}}$ es un estimador insesgado para $\varphi = \mathbf{a}^{\mathsf{t}}\boldsymbol{\theta}$ (Ejercicio)

ii) $\mathbb{V}(\widehat{\varphi}) = \sigma^2\mathbf{a}^{\mathsf{t}}(\mathbf{X}^{\mathsf{t}}\mathbf{X})^{-1}\mathbf{a} = \sigma^2_{\widehat{\varphi}}$ (Ejercicio)

Propiedades

Teorema Supongamos que $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\theta}, \sigma^2\mathbf{I})$, $\boldsymbol{\theta} \in \mathbb{R}^p$, rango$(\mathbf{X}) = p$ y definamos

$$\varphi = \mathbf{a}^{\mathsf{t}}\boldsymbol{\theta} \qquad \widehat{\varphi} = \mathbf{a}^{\mathsf{t}}\widehat{\boldsymbol{\theta}}$$

i) $\widehat{\varphi} = \mathbf{a}^{\mathsf{t}}\widehat{\boldsymbol{\theta}}$ es un estimador insesgado para $\varphi = \mathbf{a}^{\mathsf{t}}\boldsymbol{\theta}$ (Ejercicio)

ii) $\mathbb{V}(\widehat{\varphi}) = \sigma^2\mathbf{a}^{\mathsf{t}}(\mathbf{X}^{\mathsf{t}}\mathbf{X})^{-1}\mathbf{a} = \sigma^2_{\widehat{\varphi}}$ (Ejercicio)

iii) $\widehat{\varphi} \sim N(\varphi, \sigma^2\mathbf{a}^{\mathsf{t}}(\mathbf{X}^{\mathsf{t}}\mathbf{X})^{-1}\mathbf{a}) = N(\varphi, \sigma^2_{\widehat{\varphi}})$

Propiedades

Teorema Supongamos que $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\theta}, \sigma^2\mathbf{I})$, $\boldsymbol{\theta} \in \mathbb{R}^p$, rango$(\mathbf{X}) = p$ y definamos

$$\varphi = \mathbf{a}^{\mathsf{t}}\boldsymbol{\theta} \qquad \widehat{\varphi} = \mathbf{a}^{\mathsf{t}}\widehat{\boldsymbol{\theta}}$$

i) $\widehat{\varphi} = \mathbf{a}^{\mathsf{t}}\widehat{\boldsymbol{\theta}}$ es un estimador insesgado para $\varphi = \mathbf{a}^{\mathsf{t}}\boldsymbol{\theta}$ (Ejercicio)

ii) $\mathbb{V}(\widehat{\varphi}) = \sigma^2\mathbf{a}^{\mathsf{t}}(\mathbf{X}^{\mathsf{t}}\mathbf{X})^{-1}\mathbf{a} = \sigma^2_{\widehat{\varphi}}$ (Ejercicio)

iii) $\widehat{\varphi} \sim N(\varphi, \sigma^2\mathbf{a}^{\mathsf{t}}(\mathbf{X}^{\mathsf{t}}\mathbf{X})^{-1}\mathbf{a}) = N(\varphi, \sigma^2_{\widehat{\varphi}})$

iv) Sea $\widehat{\sigma}^2_{\widehat{\varphi}} = s^2\,\mathbf{a}^{\mathsf{t}}(\mathbf{X}^{\mathsf{t}}\mathbf{X})^{-1}\mathbf{a}$ entonces

$$\frac{(\widehat{\varphi} - \varphi)}{\widehat{\sigma}_{\widehat{\varphi}}} \sim t_{n-p}$$

Estimando cada componente

Estos resultados nos permiten deducir intervalos de confianza o tests para cada uno de los coeficientes del modelo lineal:

Como $\widehat{\boldsymbol{\theta}} \sim N_p(\boldsymbol{\theta}, \sigma^2(\mathbf{X}^{\mathsf{t}}\mathbf{X})^{-1})$, si $\mathbf{e}_i$ es el i-ésimo vector canónico, entonces

$$\widehat{\theta}_i = \mathbf{e}_i^{\mathsf{t}}\widehat{\boldsymbol{\theta}} \sim N(\theta_i, \sigma^2\mathbf{e}_i^{\mathsf{t}}(\mathbf{X}^{\mathsf{t}}\mathbf{X})^{-1}\mathbf{e}_i)$$

Si llamamos $\Sigma_{\widehat{\boldsymbol{\theta}}} = \sigma^2\mathbf{D}$

$$\boxed{\widehat{\theta}_i \sim N(\theta_i, \sigma^2 d_{ii})}$$

siendo d_{ii} el i-ésimo elemento diagonal de $\mathbf{D}$.

Si para un i fijo quisiéramos testear

$$\boxed{H_o : \theta_i = 0 \quad \text{vs.} \quad H_1 : \theta_i \neq 0}$$

tenemos que bajo H_o

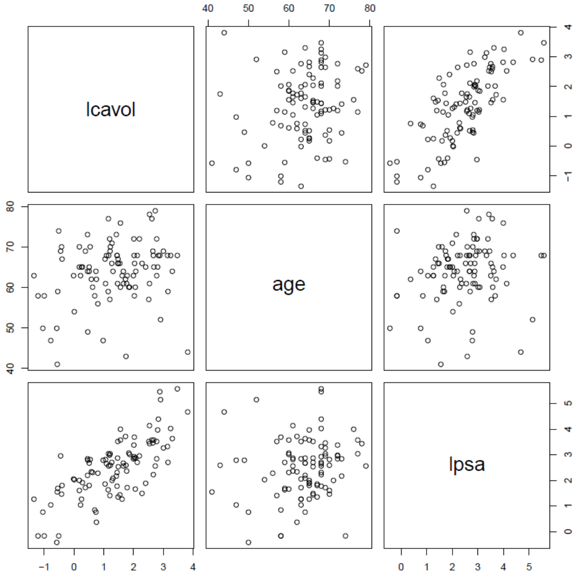
$$\frac{\widehat{\theta}_i}{s\sqrt{d_{ii}}} \sim t_{n-p}$$

Por lo tanto, rechazaremos H_o con nivel α si

$$\left| \frac{\widehat{\theta}_i}{s\sqrt{d_{ii}}} \right| > t_{n-p,\frac{\alpha}{2}}$$

Vayamos al ejemplo...

Ejemplo



$$lcavol_i = \theta_0 + \theta_1\, age_i + \theta_2\, lpsa_i + \varepsilon_i$$

```
salida.2<-lm(lcavol ~ age+lpsa)
summary(salida.2)
```

$$lcavol_i = \theta_0 + \theta_1\, age_i + \theta_2\, lpsa_i + \varepsilon_i$$

```
salida.2<-lm(lcavol ~ age+lpsa)
summary(salida.2)
```

Call:
lm(formula = lcavol ~ age + lpsa) Residuals:

Min	1Q	Median	3Q	Max
-2.23487	-0.62467	0.02114	0.54422	1.71757

Coefficients:

	Estimate	Std. Error	t value	$Pr(> \|t\|)$	
(Intercept)	-1.50978	0.70670	-2.136	0.0352	*
age	0.01637	0.01112	1.473	0.1442	
lpsa	0.73201	0.07170	10.210	<2e-16	***

Residual standard error: 0.7992 on 94 degrees of freedom

Multiple R-squared: 0.5498, Adjusted R-squared: 0.5402

F-statistic: 57.4 on 2 and 94 DF, p-value: < 2.2e-16

Propiedades

Teorema Supongamos que $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\theta}, \sigma^2\mathbf{I})$, rango$(\mathbf{X}) = p$, $\boldsymbol{\theta} \in \mathbb{R}^p$.

Sea $\mathbf{A} \in \mathbb{R}^{k\times p}$ con rango$(\mathbf{A}) = k \leq p$ y definamos

$$\boldsymbol{\psi} = \mathbf{A}\boldsymbol{\theta} \qquad \widehat{\boldsymbol{\psi}} = \mathbf{A}\widehat{\boldsymbol{\theta}}$$

i) $\widehat{\boldsymbol{\psi}}$ es un estimador insesgado para $\boldsymbol{\psi}$

ii) $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\psi}}} = \sigma^2\mathbf{A}(\mathbf{X}^{\mathsf{t}}\mathbf{X})^{-1}\mathbf{A}^{\mathsf{t}}$

iii) $\widehat{\boldsymbol{\psi}} \sim N\left(\boldsymbol{\psi}, \sigma^2\mathbf{A}(\mathbf{X}^{\mathsf{t}}\mathbf{X})^{-1}\mathbf{A}^{\mathsf{t}}\right) = N\left(\boldsymbol{\psi}, \boldsymbol{\Sigma}_{\widehat{\boldsymbol{\psi}}}\right)$

iv) Sea $\widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\psi}}} = s^2\,\mathbf{A}(\mathbf{X}^{\mathsf{t}}\mathbf{X})^{-1}\mathbf{A}^{\mathsf{t}}$ entonces

$$\frac{(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi})^{\mathsf{t}}\,\widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\psi}}}^{-1}\,(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi})}{k} \sim \mathcal{F}_{k,n-p}$$

Ejemplo: Bonus

Vamos a reproducir el cálculo del estadístico F reportado en el summary en el que se testea la significación de la regresión, o sea

$$H_0 : \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Ejemplo: Bonus

Vamos a reproducir el cálculo del estadístico F reportado en el summary en el que se testea la significación de la regresión, o sea

$$H_0 : \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

El vector reducido (es decir sin la primera componente) $\boldsymbol{\theta}^* = (\theta_1, \theta_2)^{\mathsf{t}}$ es el resultado de multiplicar a $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)^{\mathsf{t}}$ por la matriz

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Ejemplo: Bonus

Usando las propiedades de vectores aleatorios y de la distribución normal tenemos que

$$\begin{pmatrix} \widehat{\theta}_1 \\ \widehat{\theta}_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \sigma^2\,\mathbf{A}\,\left(\mathbf{X}^{\mathsf{t}}\mathbf{X}\right)^{-1}\,\mathbf{A}^{\mathsf{t}}\right)$$

Ejemplo: Bonus

Usando las propiedades de vectores aleatorios y de la distribución normal tenemos que

$$\begin{pmatrix} \widehat{\theta}_1 \\ \widehat{\theta}_2 \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \sigma^2 \, \mathbf{A} \, \left(\mathbf{X}^{\mathsf{t}} \mathbf{X} \right)^{-1} \mathbf{A}^{\mathsf{t}} \right)$$

que se puede estimar mediante los siguientes comandos

Ejemplo: Bonus

tita.est $<-$ salida.2$coefficients

ese$<-$ summary(salida.2)$sigma

XtX $=$ t(model.matrix(salida.2))%*%model.matrix(salida.2)
XtX.inv $=$ solve(XtX)
XtX.inv

Ejemplo: Bonus

tita.est < − salida.2$coefficients

ese< − summary(salida.2)$sigma

XtX = t(model.matrix(salida.2))%*%model.matrix(salida.2)
XtX.inv = solve(XtX)
XtX.inv

	(Intercept)	age	lpsa
(Intercept)	0.781970977	-0.0118329283	-0.0064316618
age	-0.011832928	0.0001934906	-0.0002116456
lpsa	-0.006431662	-0.0002116456	0.0080490331

Ejemplo: Bonus

tita.est $<-$ salida.2$coefficients

ese$<-$ summary(salida.2)$sigma

XtX $=$ t(model.matrix(salida.2))%*%model.matrix(salida.2)
XtX.inv $=$ solve(XtX)
XtX.inv

	(Intercept)	age	lpsa
(Intercept)	0.781970977	-0.0118329283	-0.0064316618
age	-0.011832928	0.0001934906	-0.0002116456
lpsa	-0.006431662	-0.0002116456	0.0080490331

A$<-$matrix(rep(0,6),ncol $=$ 3,nrow $=$ 2)
A$[1,2]$=A$[2,3]$=1
A

	$[,1]$	$[,2]$	$[,3]$
$[1,]$	0	1	0
$[2,]$	0	0	1

Ejemplo: Bonus

Calculemos $\widehat{\boldsymbol{\theta}}^*$

```
theta.star=A%*%tita.est
theta.star
```

Ejemplo: Bonus

\# Calculemos $\widehat{\boldsymbol{\theta}}^*$

```
theta.star=A%*%tita.est
theta.star
```

	[,1]
[1,]	0.01637148
[2,]	0.73201132

Ejemplo: Bonus

Calculemos $\widehat{\boldsymbol{\theta}}^*$

theta.star=A%*%tita.est
theta.star

	$[,1]$
$[1,]$	0.01637148
$[2,]$	0.73201132

CovMat$<-$ (ese$\wedge$2)*A%*%solve(XtX)%*%t(A)
CovMat

Ejemplo: Bonus

\# Calculemos $\widehat{\boldsymbol{\theta}}^*$

theta.star=A%*%tita.est
theta.star

	[,1]
[1,]	0.01637148
[2,]	0.73201132

CovMat< − (ese∧2)*A%*%solve(XtX)%*%t(A)
CovMat

	[,1]	[,2]
[1,]	0.0001235781	-0.0001351733
[2,]	-0.0001351733	0.0051407373

Ejemplo: Bonus

Dado que

$$V_1 = \frac{(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)^{\mathsf{t}} \left(\mathbf{A} \left(\mathbf{X}^{\mathsf{t}} \mathbf{X} \right)^{-1} \mathbf{A}^{\mathsf{t}} \right)^{-1} (\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)}{\sigma^2} \sim \chi^2_2$$

Ejemplo: Bonus

Dado que

$$V_1 = \frac{(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)^{\mathsf{t}} \left(\mathbf{A} \left(\mathbf{X}^{\mathsf{t}}\mathbf{X}\right)^{-1} \mathbf{A}^{\mathsf{t}}\right)^{-1} (\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)}{\sigma^2} \sim \chi^2_2$$

es independiente de

$$V_2 = \frac{(n-3)s^2}{\sigma^2} \sim \chi^2_{n-3}$$

Ejemplo: Bonus

Dado que

$$V_1 = \frac{(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)^{\mathsf{t}} \left(\mathbf{A} \left(\mathbf{X}^{\mathsf{t}}\mathbf{X}\right)^{-1} \mathbf{A}^{\mathsf{t}}\right)^{-1} (\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)}{\sigma^2} \sim \chi^2_2$$

es independiente de

$$V_2 = \frac{(n-3)s^2}{\sigma^2} \sim \chi^2_{n-3}$$

resulta que el cociente $(V_1/2)/(V_2/(n-3))$ tiene distribución $F_{2,n-3}$

$$\frac{(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)^{\mathsf{t}} \left(\mathbf{A} \left(\mathbf{X}^{\mathsf{t}}\mathbf{X}\right)^{-1} \mathbf{A}^{\mathsf{t}}\right)^{-1} (\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)}{2\ s^2} \sim F_{2,n-3}$$

Ejemplo: Bonus

Dado que

$$V_1 = \frac{(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)^{\mathsf{t}} \left(\mathbf{A}\left(\mathbf{X}^{\mathsf{t}}\mathbf{X}\right)^{-1}\mathbf{A}^{\mathsf{t}}\right)^{-1} (\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)}{\sigma^2} \sim \chi^2_2$$

es independiente de

$$V_2 = \frac{(n-3)s^2}{\sigma^2} \sim \chi^2_{n-3}$$

resulta que el cociente $(V_1/2)/(V_2/(n-3))$ tiene distribución $F_{2,n-3}$

$$\frac{(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)^{\mathsf{t}} \left(\mathbf{A}\left(\mathbf{X}^{\mathsf{t}}\mathbf{X}\right)^{-1}\mathbf{A}^{\mathsf{t}}\right)^{-1} (\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)}{2\ s^2} \sim F_{2,n-3}$$

Lo computamos como

Ejemplo: Bonus

```
t(theta.star)%*%(solve(CovMat))%*%(theta.star)/2
```

	[,1]
[1,]	57.40246

Ejemplo: Bonus

t(theta.star)%*%(solve(CovMat))%*%(theta.star)/2

$$\begin{array}{cc} & [,1] \\ [1,] & 57.40246 \end{array}$$

que coincide con el valor reportado en el summary en el F-statistic con 2 y 94 grados de libertad.