Modelo lineal General: introducción

Profesor: Alex de la Cruz H.

MODELOS LINEALES 1

EST631

Contenido

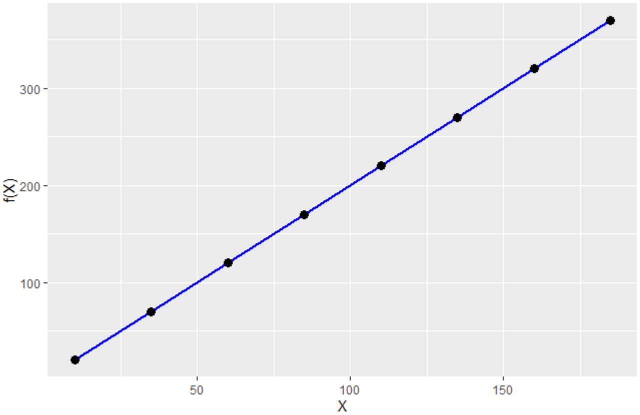
Introducción

Se aborda la relación entre dos variables, distinguiendo entre una relación funcional y una estadística.

- Una **relación funcional** para un valor de X, la función f determina el valor correspondiente de y de la forma $y = f(X)$.

  Ejemplo:

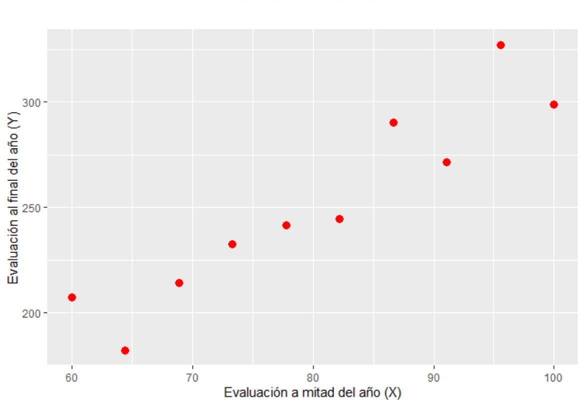
  relación entre las ventas en dólares (y) de un producto vendido a un precio fijo y el número de unidades vendidas (X). Si el precio de venta es de \$2 por unidad: $y = 2X$


300
200
100
f(X)
50
100
150
X

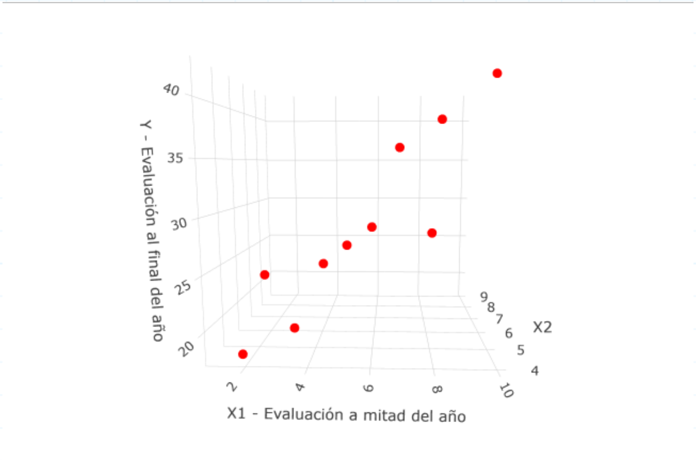
- Una **relación estadística**, no es perfecta. En general, las observaciones no caen directamente sobre la curva de la relación.

  Ejemplo:

  Evaluaciones de desempeño de 10 empleados a mitad (X) y al final del año (y).



- Por ejemplo, dos empleados con evaluaciones de mitad de año de $X = 80$ pueden obtener calificaciones muy diferentes en las evaluaciones de fin de año.



Este mismo caso se observa, si agregamos más variables.

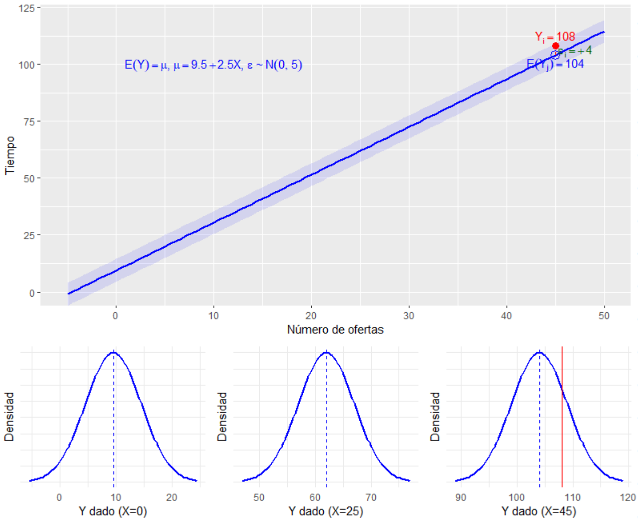
Modelo de regresión lineal

Dada una m.a. de y para un vector k de covariables les $(x_{11}, ..., x_{1k}, Y_1), \ldots, (x_{n1}, ..., x_{nk}, Y_n)$. El modelo de regresión lineal se expresa como:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i = \eta_i + \varepsilon_i \quad i = 1, 2, \ldots, n$$

- β_j es un coeficiente de regresión, $j = 0, 1, \ldots, k$
- Asumimos que $\varepsilon_i \sim N(0, \sigma^2)$ y $\text{Cov}(\varepsilon_i, \varepsilon_{i'}) = 0, \quad \forall i \neq i'$.

- Existe una distribución de probabilidad de Y_i para cada nivel de X.
- Las medias de estas distribuciones varían de manera sistemática con X.



$Y_i = 108$
$\varepsilon_i = +4$
$E(Y_i) = 104$
$E(Y) = \mu, \mu = 9.5 + 2.5X, \varepsilon \sim N(0, 5)$
Tiempo
Número de ofertas
Densidad
Y dado (X=0)
Y dado (X=25)
Y dado (X=45)

Dado el valor de $\boldsymbol{x}_i = (1, x_{i1}, \ldots, x_{ik})^\top$, queremos encontrar un modelo que describa la relación entre y y $\boldsymbol{x}_i$ mediante:

$$\mathbb{E}(Y_i \mid \boldsymbol{x}_i) = \mu_i = f(\boldsymbol{x}_i) = \eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}, \qquad i = 1, 2, \ldots, n$$

donde $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)^\top$ es el vector de coeficientes de regresión.

En general, es posible escribir el modelo en forma matricial:

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times p}{\boldsymbol{X}}\underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}}$$

$$\boldsymbol{y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Obs: En algunas partes consideraremos $p = k + 1$

Además de que la relación entre $\boldsymbol{y}$ y $\boldsymbol{X}$ es lineal, asumimos que:

1. Los errores ε_i son independientes con media nula.
2. La varianza del error ε_i es constante.
3. Los errores ε_i, se distribuyen normalmente.
4. $\text{Rank}(\boldsymbol{X}) = k + 1 = p$.

Esto significa:

1. $\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$.
2. $\text{Cov}(\boldsymbol{\varepsilon}) = \mathbb{E}\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top\right) = \sigma^2 I$.

$$\text{Var}(\boldsymbol{\varepsilon}) = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 \boldsymbol{I}$$

3. $\boldsymbol{\varepsilon} \sim N_n\left(0, \sigma^2 \boldsymbol{I}\right)$ (distribución normal multivariada n-dimensional).
4. La matriz $\boldsymbol{X}$ es de rango completo, $\text{rank}(\boldsymbol{X}) = p$. (Asumimos que $n >> p$.)

El método de MCO se basa en la función de pérdida: $S(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$. Luego, el estimador de $\boldsymbol{\beta}$ que minimiza esta función es:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$$

Esta solución coincide con el estimador de máxima verosimilitud (EMV).

Definición 1.

Una matriz simétrica $A \in \mathbb{R}^{n \times n}$ es semidefinida positiva si: $\forall z \in \mathbb{R}^n, \quad z^T A z \geq 0$. Esto se denota como:

$$A \succeq 0$$

Teorema 1.

*En el modelo de regresión lineal clásica, si se cumplen los supuestos, el estimador de mínimos cuadrados ordinarios (MCO), $\hat{\boldsymbol{\beta}}$ es el **mejor estimador lineal insesgado** (BLUE, por sus siglas en inglés).*

$$\forall \tilde{\beta} = Ay \text{ tal que } AX = I_k, \quad \mathrm{Var}(\tilde{\beta}) - \mathrm{Var}(\hat{\beta}) \succeq 0$$

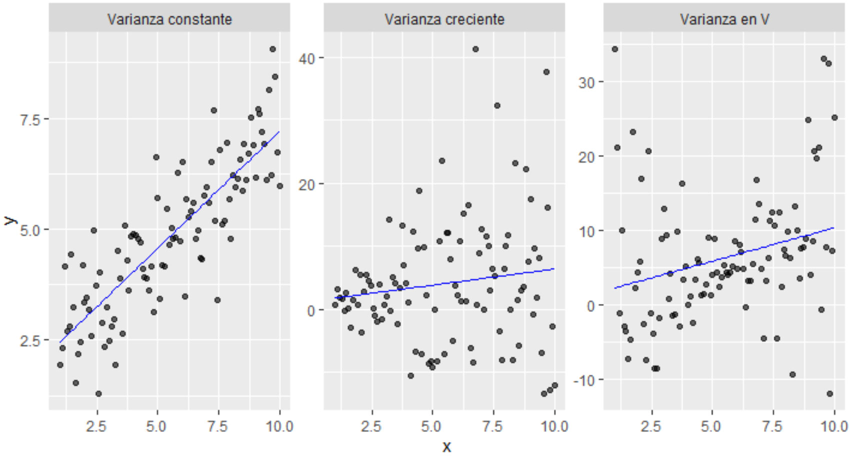
La solución a través del método de MCO coincide con el estimador de máxima verosimilitud, el cual también proporciona una solución para:

$$\hat{\sigma}^2 = \frac{(\boldsymbol{y} - \hat{\boldsymbol{y}})^\top (\boldsymbol{y} - \hat{\boldsymbol{y}})}{n - p}$$

Limitaciones del Modelo Clásico

1. Asume homocedasticidad (varianza constante σ^2)
2. Asume independencia entre observaciones ($\mathrm{Cov}(\epsilon_i, \epsilon_j) = 0$)
3. No permite estructuras más complejas de dependencia o varianza heterogénea

Limitaciones del Modelo Clásico



En muchos casos reales, los supuestos clásicos no se cumplen:

- Datos correlacionados: Medidas repetidas, diseños longitudinales, etc.
- Heterocedasticidad: La varianza puede cambiar entre grupos
- Estructuras de covarianza no diagonales: Autocorrelación en series de tiempo

Modelo de regresión lineal general

Modelo de regresión lineal general (RLG)

Sea una muestra aleatoria de tamaño n de una variable respuesta Y dada una observación de k covariables: $(Y_i, \boldsymbol{x}_i^\top)$, $i = 1, \ldots, n$. El modelo de RLG se expresa como:

$$Y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, 2, \ldots, n$$

donde $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)^\top$ es el vector de coeficientes de regresión.

- $\mathrm{Var}(\epsilon_i) = \sigma_i^2$ representan las varianzas (heterocedasticidad cuando difieren).
- $\mathrm{Cov}(\epsilon_i, \epsilon_j) = \sigma_{ij}$ representan las covarianzas entre errores.
- $\sigma_{ij} = \sigma_{ji}$ para todo $i \neq j$

En forma matricial:

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}_{n\times 1}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \sigma_{n3} & \cdots & \sigma_n^2 \end{pmatrix}_{n\times n}$$

donde $\boldsymbol{\Sigma}$ es definida positiva: $\boldsymbol{v}^T\boldsymbol{\Sigma}\boldsymbol{v} > 0$ para todo $\boldsymbol{v} \neq \mathbf{0} \in \mathbb{R}^n$

- Los elementos $\sigma_i^2 = \mathrm{Var}(\epsilon_i)$ representan las varianzas (heterocedasticidad cuando difieren)
- Los elementos $\sigma_{ij} = \mathrm{Cov}(\epsilon_i, \epsilon_j)$ representan las covarianzas entre errores
- La matriz es simétrica: $\sigma_{ij} = \sigma_{ji}$ para todo $i \neq j$
- $\boldsymbol{\Sigma}$ es definida positiva: $\boldsymbol{v}^T \boldsymbol{\Sigma} \boldsymbol{v} > 0$ para todo $\boldsymbol{v} \neq \mathbf{0}$

Supuestos clave:

1. Relación lineal: $\mathbb{E}(\boldsymbol{y}|\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\beta}$
2. Normalidad: $\boldsymbol{\epsilon}$ sigue una distribución normal multivariada
3. Estructura de covarianza general: $\mathrm{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}$

$\boldsymbol{\Sigma}$ captura posibles:

- Correlaciones entre observaciones ($\Sigma_{ij} \neq 0$)
- Heterocedasticidad ($\Sigma_{ii} \neq \Sigma_{jj}$)

Casos particulares:

- Si $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$: Regresión lineal clásica
- Si $\boldsymbol{\Sigma}$ es diagonal con σ_i^2 distintos: Regresión con heterocedasticidad
- Si $\boldsymbol{\Sigma}$ tiene patrones estructurados: Modelos de correlación serial

Estimador para $\boldsymbol{\beta}$

Puede ser obtenido a través de dos métodos:

- Mínimos Cuadrados Generalizados (GLS). Basado en la siguiente función de pérdida:

$$Q(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

- Estimador de máxima verosimilitud. Asumiendo que

$$\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma}) \Rightarrow \boldsymbol{y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$$

Método de mínimos Cuadrados Generalizados

El estimador de mínimos cuadrados generalizados (MCG o GLS) se obtiene minimizando:

$$Q(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

Esto es resolver:

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = \mathbf{0} \Rightarrow \frac{\partial Q}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \mathbf{0}$$

$$\boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{X} \hat{\boldsymbol{\beta}} = \boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{y}$$

Definición 2.

Sea $A \in \mathbb{R}^{n\times n}$ una matriz simétrica definida positiva. Entonces existe una única matriz triangular inferior $L \in \mathbb{R}^{n\times n}$ tal que: $A = LL^\top$.

O, equivalentemente, existe una triangular superior U tal que: $A = U^\top U$

Este estimador también puede obtenerse transformando el modelo mediante una matriz $\boldsymbol{C}$ tal que $\boldsymbol{C}^\top \boldsymbol{C} = \boldsymbol{\Sigma}^{-1}$ (descomposición de Cholesky), y aplicando MCO sobre el modelo transformado:

$$\boldsymbol{y}^* = \boldsymbol{C}\boldsymbol{y}, \quad \boldsymbol{X}^* = \boldsymbol{C}\boldsymbol{X} \quad \Rightarrow \hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{*\top}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\top}\boldsymbol{y}^*$$

Luego,

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{y}$$

Método de máxima verosimilitud

Los parámetros son: $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Sigma})^\top$

- La función de verosimilitud es:

$$L(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right)$$

- La log-verosimilitud es:

$$\ell(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

Maximizar la log-verosimilitud equivale a minimizar:

$$Q(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

Por tanto, la solución es equivalente al Método de mínimos cuadrados generalizados:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{y}$$

Se puede mostrar que:

$$\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$$

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{X})^{-1}$$

La estimación simultánea de $\boldsymbol{\beta}$ y $\boldsymbol{\Sigma}$ presenta un problema circular: Necesitamos $\boldsymbol{\Sigma}$ para estimar $\boldsymbol{\beta}$. Necesitamos $\boldsymbol{\beta}$ para estimar $\boldsymbol{\Sigma}$. Es necesario asumir una estructura para $\boldsymbol{\Sigma}$.

Algunos Casos particulares para $\boldsymbol{\Sigma}$

1. MCO clásico: errores iid

$$\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$$

- Varianzas constantes e independencia entre errores.
- El estimador de MCO es eficiente.

Algunos Casos particulares para Σ

2. Heterocedasticidad: varianzas distintas, no correlacionadas

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

- Las observaciones tienen diferentes niveles de variabilidad.
- No hay correlación entre errores.

Algunos Casos particulares para Σ

3. Autocorrelación: AR(1)

$$\mathbf{\Sigma} = \sigma^2 \mathbf{V} = \frac{\sigma^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix}, \qquad |\rho| < 1$$

- $\rho = \text{Corr}(\varepsilon_i, \varepsilon_{i-1}), \quad \varepsilon_i = \rho\varepsilon_{i-1} + u_i, \quad u_i \sim \mathcal{N}(0, \sigma^2)$
- $\text{Var}(\varepsilon_i) = \sigma_e^2 = \dfrac{\sigma^2}{1-\rho^2}$. Además, $\text{Cov}(\varepsilon_i, \varepsilon_j) = \rho^{|i-j|}\sigma_e^2$

Algunos Casos particulares para Σ

3. Autocorrelación: AR(1)

   - σ_e^2 es la varianza marginal de cada error ε_i.
   - ρ es el coeficiente de autocorrelación de primer orden (correlación entre errores consecutivos).
   - El elemento en la posición (i, j) de la matriz es $\sigma^2 \rho^{|i-j|}$. Así, los errores cercanos en el tiempo están más correlacionados que los distantes.
   - La matriz es simétrica y los valores disminuyen geométricamente desde la diagonal.

Algunos Casos particulares para $\boldsymbol{\Sigma}$

4. Correlación intragrupo: modelo con efectos agrupados

$$\boldsymbol{\Sigma} = \sigma_b^2 \boldsymbol{Z}\boldsymbol{Z}^\top + \sigma_e^2 \boldsymbol{I}_n$$

- La matriz $\boldsymbol{Z}$ representa la pertenencia a grupos.

$$(\boldsymbol{Z}\boldsymbol{Z}^\top)_{ij} = \begin{cases} 1 & \text{si } i \text{ y } j \text{ están en el mismo grupo,} \\ 0 & \text{en otro caso.} \end{cases}$$

- Modela correlación entre observaciones del mismo grupo.
- Común en modelos lineales mixtos.

Algunos Casos particulares para Σ

5. Covarianza general conocida

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 & 0.2 \\ 0.5 & 1 & 0.4 \\ 0.2 & 0.4 & 1 \end{bmatrix}$$

- Caso general: matriz simétrica y definida positiva.
- Puede no tener estructura interpretable simple.

Supuestos estructurales sobre $\boldsymbol{\Sigma}$

Debido a que $\boldsymbol{\Sigma}$ tiene demasiados parámetros, se imponen estructuras:

- Homoscedasticidad: $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$
- Heteroscedasticidad (diagonal): $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$
- Estructura parametrizada (por ejemplo, AR(1), MA(1), etc.)

Estimación por Mínimos Cuadrados Generalizados (GLS)

1. Estimar $\boldsymbol{\beta}$ por MCO:
$$\hat{\boldsymbol{\beta}}^{(0)} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

2. Calcular residuos:
$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{(0)}$$

3. Estimar $\boldsymbol{\Sigma}$:
   - Heteroscedasticidad: $\hat{\boldsymbol{\Sigma}} = \operatorname{diag}(\hat{\varepsilon}_1^2, \ldots, \hat{\varepsilon}_n^2)$
   - General: $\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}^\top$

4. Aplicar GLS:
$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\boldsymbol{X}^\top \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{y}$$

Estimación por Máxima Verosimilitud (MV).

Bajo normalidad, se maximiza la log-verosimilitud:

$$\ell(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})$$

Requiere:

- Asumir una estructura parametrizada para $\boldsymbol{\Sigma}$
- Optimización conjunta de $\boldsymbol{\beta}$ y parámetros de $\boldsymbol{\Sigma}$

Principales librerías R

1. nlme(Nonlinear mixed effects):

```r
library(nlme)
model <- gls(y ~ x1 + x2, data = datos,
          weights = varFixed(form = ~ x1),  # Estructura de varianza
          correlation = corAR1(form = ~ 1 | grupo))  # serial
summary(model)
```

- Heterocedasticidad con funciones como `varIdent`, `varPower`, `varFixed`, `etc`.
- Correlaciones temporales o espaciales en los residuos (e.g., corAR1).

2. glmmTMB (Generalized linear mixed models with template model builder):

```r
library(glmmTMB)
model <- glmmTMB(y ~ x1 + x2 + (1 | grupo),
                 dispformula = ~ x1,  # Varianza dependiente de x1
                 data = datos)
summary(model)
```

- Maneja heterocedasticidad y correlación en los residuos.
- Soporta correlaciones (AR1, etc.) en los residuos con argumento *time* o *ar1()*.

3. `car` (Companion to applied regression): No modela directamente, pero incluye pruebas para detectar heterocedasticidad (e.g., ncvTest, leveneTest), útiles para justificar el uso de los paquetes anteriores.

Ejercicio y aplicación

Ejercicio 1

Supón el siguiente modelo:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \text{con } \operatorname{Var}(\varepsilon_i) = \sigma^2 x_i^2$$

donde $x_i > 0$ para todo i.

1. Encuentre la forma adecuada de estimar $\boldsymbol{\beta}$
2. Aplicar MCO sobre una versión equivalente con errores homocedásticos.

Ejercicio 2

Plantee y demuestre que el siguiente estimador es el estimador lineal insesgado de varianza mínima (BLUE, por sus siglas en inglés) para $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{y}$$

1. Defina el espacio de estimadores lineales insesgados para $\boldsymbol{\beta}$ en el modelo con $\mathrm{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}$.
2. Verifique que $\widehat{\boldsymbol{\beta}}$ es lineal en $\boldsymbol{y}$ e insesgado.
3. Muestre que tiene menor varianza dentro de la clase de estimadores lineales insesgados.

Aplicación

Utilizaremos el conjunto de datos `Orange`, incluido en el programa `R`, que registra el crecimiento circunferencial (en mm) de 5 árboles de naranja (Tree) a lo largo del tiempo (age en días).

- Ajustar un modelo que permita varianzas distintas por árbol
- Comparar este modelo con un MCO tradicional
- Graficar los residuos de ambos modelos
- Tarea: ¿Por qué es importante modelar la heterocedasticidad en este caso?

Desarrollo en clase