Modelos Lineales y Modelos Lineales Generalizados

Plano de este parte.

- Modelos lineales.
- Familias Exponenciales de Distribuciones.
- Modelos Lineales Generalizados.

Modelos Lineales

Un modelo lineal tiene la forma:

$$\mathbf{Y} = \mathbf{x}\beta + \epsilon.$$

- $\mathbf{Y}$: $n \times 1$ vector de variables de respuesta.
- $\mathbf{x}$: $n \times p$ matriz de variables explicativos, tipicamente con la primera columna constante y igual a 1.
- β: $p \times 1$ vector de parámetros.

  $$\beta = (\beta_1, \ldots, \beta_p)'.)$$

- ϵ: $n \times 1$ vector de errores normales, identicos y independientes (i.i.d.). Luego

  $$\epsilon \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I}),$$

  donde $\mathbf{I}$ es el matriz de identidad de tamaño n.
- $\mathbb{E}(\mathbf{Y}) = \mathbf{x}\beta.$

Supuestos de Modelos Lineals

Linealdad: los observaciones se relacionan linealmente con los parámetros del modelo.

Normalidad: los errores son normales.

Independencia: los Errores son independientes.

Homocedástico: todos los errores tienen la misma varianza.

Ajustando un Modelo Lineal

- Se estima los parámetros β por el método de mínimos cuadrados. Es equivalente al método de estimación por máxima verosimilitud cuando los errores son normales.
- La densidad de un variable Y que es normal con media μ y varianza σ^2 es $f_Y(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$.
- La función de verosimilitud para $\mathbf{Y} = \mathbf{y}$ del modelo lineal es

$$
\begin{aligned}
L(\beta,\sigma^2 \mid \mathbf{y}) &= \prod_{i=1}^{n} f_Y\left(y_i; (\mathbf{x}\beta)_i, \sigma^2\right) \\
&= (2\pi\sigma^2)^{-n/2} \prod_{i=1}^{n} e^{-\left(y_i-(\mathbf{x}\beta)_i\right)^2 / 2\sigma^2} \\
&= (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i-(\mathbf{x}\beta)_i\right)^2} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-(1/2\sigma^2)(\mathbf{y}-\mathbf{x}\beta)'(\mathbf{y}-\mathbf{x}\beta)}.
\end{aligned}
$$

El Estimador $\hat{\beta}$

- El log de la verosimilitud es

$$\log L(\beta, \sigma^2 \mid \mathbf{y}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{x}\beta)'(\mathbf{y} - \mathbf{x}\beta).$$

- Elijir la $\hat{\beta}$ que Maximiza el log de la verosimilitud es equivalente a eligir la $\hat{\beta}$ que minimiza

$$(\mathbf{y} - \mathbf{x}\beta)'(\mathbf{y} - \mathbf{x}\beta).$$

$$\begin{array}{rrcl} & \frac{d}{d\beta}(\mathbf{y} - \mathbf{x}\beta)'(\mathbf{y} - \mathbf{x}\beta) & = & \mathbf{0} \\ \Longleftrightarrow & \frac{d}{d\beta}(\mathbf{y}'\mathbf{y} - 2\beta'\mathbf{x}'\mathbf{y} + \beta'\mathbf{x}'\mathbf{x}\beta) & = & \mathbf{0} \\ \Longleftrightarrow & \mathbf{x}'\mathbf{x}\beta - \mathbf{x}'\mathbf{y} & = & \mathbf{0} \\ \Longrightarrow & \hat{\beta} & = & (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}. \end{array}$$

- La media de $\hat{\beta}$ es

$$\begin{aligned} \mathbb{E}(\hat{\beta}) &= \mathbb{E}\left((\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{Y}\right) \\ &= (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{x}\beta = \beta. \end{aligned}$$

El Estimador $\hat{\beta}$ (2)

- $\hat{\beta}$ es un estimador sin sezgo: $\mathbb{E}(\hat{\beta}-\beta)=\beta-\beta=\mathbf{0}$.
- Matriz de covarianza:

$$\begin{aligned} \operatorname{cov}(\hat{\beta}) &= \mathbb{E}\left((\hat{\beta}-\mathbb{E}(\hat{\beta}))(\hat{\beta}-\mathbb{E}(\hat{\beta}))'\right) \\ &= \mathbb{E}\left((\hat{\beta}-\beta)(\hat{\beta}-\beta)'\right). \end{aligned}$$

  Pero

$$\begin{aligned} \hat{\beta}-\beta &= (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{Y}-(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{x}\beta \\ &= (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'(\mathbf{Y}-\mathbf{x}\beta). \end{aligned}$$

  Luego

$$\begin{aligned} \operatorname{cov}(\hat{\beta}) &= \mathbb{E}\left((\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'(\mathbf{Y}-\mathbf{x}\beta)(\mathbf{Y}-\mathbf{x}\beta)'\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\right) \\ &= (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbb{E}\left((\mathbf{Y}-\mathbf{x}\beta)(\mathbf{Y}-\mathbf{x}\beta)'\right)\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1} \\ &= (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\sigma^2\mathbf{I}\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1} \\ &= \sigma^2(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}=\sigma^2(\mathbf{x}'\mathbf{x})^{-1}. \end{aligned}$$

- $\hat{\beta}\sim \mathrm{N}\left(\beta,\sigma^2(\mathbf{x}'\mathbf{x})^{-1}\right)$.

Estimando σ^2

- Para obtener un estimador de σ^2, solucionamos

$$\frac{d}{d\sigma^2}\left(\frac{n}{2}\log(2\pi\sigma^2)\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{x}\beta)'(\mathbf{y}-\mathbf{x}\beta)\right)=0.$$

  El estimador de σ^2 que obtenemos es:

$$\hat{\sigma}^2=\frac{1}{n}(\mathbf{Y}-\mathbf{x}\hat{\beta})'(\mathbf{Y}-\mathbf{x}\hat{\beta}).$$

- Veremos que $\hat{\sigma}^2$ es sesgado.

Valores Ajustados

- Valores ajustados:
$$\hat{\mathbf{Y}} = \mathbf{x}\hat{\beta} = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{Y} = \mathbf{h}\mathbf{Y},$$
donde $\mathbf{h} = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'$ es el matriz de predición.
- Se lo denota $\mathbf{h}$ porque en Inglés se lamma "the <u>h</u>at matrix":
$$\mathbf{Y} \xrightarrow{\mathbf{h}} \hat{\mathbf{Y}}.$$
- $\hat{\mathbf{Y}}$ Representa los valores de los observaciones predicidos por el modelo ajustado.
- $$\begin{aligned} \mathbb{E}(\hat{\mathbf{Y}}) &= \mathbb{E}\left(\mathbf{h}\mathbf{Y}\right) = \mathbf{h}\mathbb{E}(\mathbf{Y}) = \mathbf{h}\mathbf{x}\beta \\ &= \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{x}\beta = \mathbf{x}\beta = \mathbb{E}(\mathbf{Y}). \end{aligned}$$
- $\operatorname{cov}(\hat{\mathbf{Y}}) = \operatorname{cov}(\mathbf{h}\mathbf{Y}) = \mathbf{h}\operatorname{cov}(\mathbf{Y})\mathbf{h}' = \sigma^2\mathbf{h}\mathbf{h}' = \sigma^2\mathbf{h}^2 = \sigma^2\mathbf{h}$ porque $\mathbf{h}$ es una proyección: $\mathbf{h} = \mathbf{h}'$ y
$$\mathbf{h}^2 = \mathbf{h}\mathbf{h} = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}' = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}' = \mathbf{h}.$$
- $\hat{\mathbf{Y}} \sim \mathrm{N}\left(\mathbf{x}\beta, \sigma^2\mathbf{h}\right).$

Los Residuos

- Residuos: $\mathbf{R} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{h})\mathbf{Y}$.
- $\mathbb{E}(\mathbf{R}) = \mathbb{E}(\mathbf{Y}) - \mathbb{E}(\hat{\mathbf{Y}}) = \mathbf{x}\beta - \mathbf{x}\beta = \mathbf{0}$.
- Matriz de covarianzas:

$$
\begin{aligned}
\operatorname{cov}(\mathbf{R}) &= \operatorname{cov}\left((\mathbf{I} - \mathbf{h})\mathbf{Y}\right) \\
&= (\mathbf{I} - \mathbf{h}) \operatorname{cov}(\mathbf{Y})(\mathbf{I} - \mathbf{h})' \\
&= (\mathbf{I} - \mathbf{h})\sigma^2 \mathbf{I}(\mathbf{I} - \mathbf{h})' \\
&= \sigma^2(\mathbf{I} - \mathbf{h})^2 = \sigma^2(\mathbf{I} - 2\mathbf{h} + \mathbf{h}) = \sigma^2(\mathbf{I} - \mathbf{h}).
\end{aligned}
$$

- Lo mismo que $\mathbf{h}$, $\mathbf{I} - \mathbf{h}$ es simétrico y es una proyección también.
- $\mathbf{R} \sim \mathrm{N}\left(\mathbf{0}, \sigma^2(\mathbf{I} - \mathbf{h})\right)$.

El Estimador $\hat{\sigma}_2$

- $\hat{\sigma}^2$ es un múltiplo del la suma de los residuos cuadráticos:

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n}(\mathbf{Y}-\mathbf{x}\hat{\beta})'(\mathbf{Y}-\mathbf{x}\hat{\beta}) \\ &= \frac{1}{n}(\mathbf{Y}-\hat{\mathbf{Y}})'(\mathbf{Y}-\hat{\mathbf{Y}}) = \frac{1}{n}\mathbf{R}\mathbf{R}'.\end{aligned}$$

- También,

$$\begin{aligned}&(\mathbf{Y}-\hat{\mathbf{Y}})'(\mathbf{Y}-\hat{\mathbf{Y}} \\ &\quad= \big(\mathbf{Y}-\mathbf{x}\beta-\mathbf{h}(\mathbf{Y}-\mathbf{x}\beta)\big)'\big(\mathbf{Y}-\mathbf{x}\beta-\mathbf{h}(\mathbf{Y}-\mathbf{x}\beta)\big) \\ &\quad= (\mathbf{Y}-\mathbf{x}\beta)'(\mathbf{I}-\mathbf{h})'(\mathbf{I}-\mathbf{h})(\mathbf{Y}-\mathbf{x}\beta) \\ &\quad= (\mathbf{Y}-\mathbf{x}\beta)'(\mathbf{I}-\mathbf{h})(\mathbf{Y}-\mathbf{x}\beta) \\ &\quad= \sum_{i=1}^{n}\sum_{j=1}^{n}(1-h_{ij})\big(Y_i-(\mathbf{x}\beta)_i\big)\big(Y_j-(\mathbf{x}\beta)_j\big).\end{aligned}$$

La Esperanza de $\hat{\sigma}^2$

- Tomando la esperanza, obtenemos

$$\begin{aligned}
&\mathbb{E}\left((\mathbf{Y}-\hat{\mathbf{Y}})'(\mathbf{Y}-\hat{\mathbf{Y}})\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}(1-h_{ij})\mathbb{E}\left(\left(Y_i-(\mathbf{x}\beta)_i\right)\left(Y_j-(\mathbf{x}\beta)_j\right)\right) \\
&= \sigma^2\sum_{i=1}^{n}(1-h_{ii}) = \sigma^2\left(n-\mathrm{tr}(\mathbf{h})\right),
\end{aligned}$$

donde $\mathrm{tr}(\mathbf{h}) = \sum_{i=1}^{n} h_{ii}$ es la traza del matriz $\mathbf{h}$.

Propriedades de la Traza.

Para dos matrices $\mathbf{a}$ y $\mathbf{b}$, y un constante c,

$$\begin{aligned}
\mathrm{tr}(\mathbf{a}+\mathbf{b}) &= \mathrm{tr}(\mathbf{a})+\mathrm{tr}(\mathbf{b}), \\
\mathrm{tr}(c\mathbf{a}) &= c\,\mathrm{tr}(\mathbf{a}), \\
\mathrm{tr}(\mathbf{a}') &= \mathrm{tr}(\mathbf{a}),\ \text{y}\,\mathrm{tr}(\mathbf{ab}) = \mathrm{tr}(\mathbf{ba}).
\end{aligned}$$

El Estimador S^2

- Luego

$$\text{tr}(\mathbf{h}) = \text{tr}\left(\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\right) = \text{tr}\left((\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{x}\right) = \text{tr}(\mathbf{I}_p) = p.$$

- La esperanza de $\hat{\sigma}^2$ es

$$\mathbb{E}(\hat{\sigma}^2) = \frac{n-p}{n}\sigma^2 < \sigma^2.$$

- Por eso, se define S^2, un estimador de σ^2 que es sin sesgo:

$$\begin{aligned} S^2 &= \frac{1}{n-p}(\mathbf{Y} - \mathbf{x}\beta)'(\mathbf{Y} - \mathbf{x}\beta) = \frac{n}{n-p}\hat{\sigma}^2. \\ \mathbb{E}(S^2) &= \mathbb{E}\left(\frac{n}{n-p}\hat{\sigma}^2\right) = \frac{n}{n-p} \cdot \frac{n-p}{n}\sigma^2 = \sigma^2. \end{aligned}$$

- $S^2(\mathbf{x}'\mathbf{x})^{-1}$, $S^2\mathbf{h}$ y $S^2(\mathbf{I} - \mathbf{h})$ son estimadores sin sesgo de $\text{cov}(\hat{\beta})$, $\text{cov}(\hat{\mathbf{Y}})$ y $\text{cov}(\mathbf{R})$ respectivamente.