CAPÍTULO 7. MODELOS LINEALES Y REGRESIÓN

Para leer

Box y Tiao (1973),
Broemeling (1984),
Gelman et al (1995), Capítulo 8,
Lee (1997), Capítulo 6.

Sea $\mathbf{Y}$ un vector de n observaciones. El modelo lineal para $\mathbf{Y}$ es

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}$$

donde $\boldsymbol{\theta}$ es un vector de parámetros de dimensión p, $\mathbf{X}$ es una matriz de diseño de dimensión $n \times p$ y

$$\boldsymbol{\epsilon}|\phi \sim \mathcal{N}\left(\mathbf{0}, \frac{1}{\phi}\mathbf{I}\right)$$

donde $\mathbf{I}$ es la matriz de identidad de dimensión $n \times n$.

La estructura incluye muchos modelos fundamentales como ADEVA, efectos aleatorios y fijos etc.

Ejemplo 90 *Observaciones univariables*

$$\mathbf{X}^T = (1, \ldots, 1), \quad \theta \textit{ escalar} \quad p = 1.$$

Luego, se tiene $Y_i \sim \mathcal{N}\left(\theta, \frac{1}{\phi}\right)$ *para* $i = 1, \ldots, n$.

Ejemplo 91 *Regresión lineal simple*

$$\mathbf{X}^T = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix} \quad \boldsymbol{\theta} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

$$E[y_i|\boldsymbol{\theta}] = \alpha + \beta x_i$$

Ejemplo 92 *ADEVA*

$$y_{ij} = \theta_i + \epsilon_{ij}$$

donde $\epsilon_{ij}|\phi \sim \mathcal{N}\left(0, \frac{1}{\phi}\right)$, *para* $i = 1, \ldots, p$ *y* $j = 1, \ldots, n_i$.

Entonces $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$,

$$\mathbf{y} = (y_{11}, \ldots, y_{1n_1}, y_{21}, \ldots, y_{2n_2}, \ldots, y_{pn_p})^T$$

$$\mathbf{X} = \begin{array}{cc} \left(\begin{array}{ccccc} 1 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{array}\right) & \begin{array}{l} \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} n_1 \\ \\ \\ \\ \\ \\ \\ \end{array} \end{array}$$

y se tiene $\mathbf{Y}|\boldsymbol{\theta}, \phi \sim \mathcal{N}\left(\mathbf{X}\boldsymbol{\theta}, \frac{1}{\phi}\mathbf{I}\right)$, *donde* $\mathbf{I}$ *es la identidad de dimensión* $n = \sum_{i=1}^{p} n_i$.

Inferencia para el modelo lineal dada la distribución a priori de referencia

Condicionando sobre los parámetros desconocidos, la distribución de los datos es normal multivariante.

$$\mathbf{Y}|\boldsymbol{\theta}, \phi \sim \mathcal{N}\left(\mathbf{X}\boldsymbol{\theta}, \frac{1}{\phi}\mathbf{I}\right).$$

Luego, se puede demostrar que bajo la distribución a priori no informativa

$$f(\boldsymbol{\theta}, \phi) \propto \frac{1}{\phi}$$

entonces los resultados de la inferencia bayesiana coinciden numéricamente con los resultados clásicos de siempre.

Teorema 13 *Sea* $\mathbf{Y}|\boldsymbol{\theta},\phi \sim \mathcal{N}\left(\mathbf{X}\boldsymbol{\theta}, \frac{1}{\phi}\mathbf{I}\right)$ *con distribución a priori* $f(\boldsymbol{\theta},\phi) \propto \frac{1}{\phi}$. *Luego la distribución a posteriori de* $\boldsymbol{\theta},\phi|\mathbf{y}$ *es*

$$\begin{aligned} \boldsymbol{\theta}|\phi,\mathbf{y} &\sim \mathcal{N}\left(\widehat{\boldsymbol{\theta}}, \frac{1}{\phi}(\mathbf{X}^T\mathbf{X})^{-1}\right) \\ \phi|\mathbf{y} &\sim \mathcal{G}\left(\frac{n-p}{2}, \frac{(n-p)s^2}{2}\right) \end{aligned}$$

donde

$$\begin{aligned} \widehat{\boldsymbol{\theta}} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \quad \text{y} \\ s^2 &= \frac{1}{n-p}\mathbf{y}^T\left(\mathbf{I}-\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\right)\mathbf{y} \end{aligned}$$

Además, la distribución marginal de θ_i *es*

$$\frac{\theta_i - \widehat{\theta}_i}{s\sqrt{(\mathbf{X}^T\mathbf{X})^{-1}_{ii}}} \mid \mathbf{y} \sim t_{n-p}.$$

Demostración

Utilizando el teorema de Bayes,

$$f(\boldsymbol{\theta}, \phi|\mathbf{y}) \propto \frac{1}{\phi}\phi^{\frac{n}{2}} \exp\left(-\frac{\phi}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})\right)$$

y luego

$$\propto \phi^{\frac{n}{2}-1} \exp\left(-\frac{\phi}{2}\left[\boldsymbol{\theta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\theta} - \boldsymbol{\theta}^T\mathbf{X}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\boldsymbol{\theta} + \mathbf{y}^T\mathbf{y}\right]\right)$$

$$\propto \phi^{\frac{n}{2}-1} \exp\left(-\frac{\phi}{2}\left[\boldsymbol{\theta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\theta} - \boldsymbol{\theta}^T\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} - \right.\right.$$
$$\left.\left. \mathbf{y}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\theta} + \mathbf{y}^T\mathbf{y}\right]\right)$$

$$\propto \phi^{\frac{n}{2}-1} \exp\left(-\frac{\phi}{2}\left[(\boldsymbol{\theta}-(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y})^T\mathbf{X}^T\mathbf{X}(\boldsymbol{\theta}-(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y})\right.\right.$$
$$\left.\left. +\mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}\right]\right)$$

$$\propto \phi^{\frac{n}{2}-1} \exp\left(-\frac{\phi}{2}\left[(\boldsymbol{\theta}-(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y})^T\mathbf{X}^T\mathbf{X}(\boldsymbol{\theta}-(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y})\right.\right.$$
$$\left.\left. +\mathbf{y}^T\left(\mathbf{I}-\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1})\mathbf{X}^T\right)\mathbf{y}\right]\right).$$

Luego, $\quad \boldsymbol{\theta}|\phi, \mathbf{y} \sim \mathcal{N}\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}, \frac{1}{\phi}(\mathbf{X}^T\mathbf{X})^{-1}\right).$

Además, integrando la distribución conjunta con respeto a $\boldsymbol{\theta}$*, se tiene*

$$f(\phi|\mathbf{y}) \propto \phi^{\frac{n-p}{2}-1} \exp\left(-\frac{\phi}{2}\left[\mathbf{y}^T(\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1})\mathbf{X}^T\mathbf{y}\right]\right)$$

que es el núcleo de una distribución gamma.

Ahora, la distribución de $\theta_i|\phi, \mathbf{y}$ *es*

$$\theta_i|\phi, \mathbf{y} \sim \mathcal{N}\left(\widehat{\theta}_i, \frac{1}{\phi}(\mathbf{X}^T\mathbf{X})^{-1}_{ii}\right)$$

y entonces, la distribución conjunta de θ_i, ϕ *es normal gamma*

$$\theta_i, \phi|\mathbf{y} \sim \mathcal{NG}\left(\widehat{\theta}_i, \frac{1}{(\mathbf{X}^T\mathbf{X})^{-1}_{ii}}, \frac{n-p}{2}, \frac{(n-p)s^2}{2}\right)$$

y aplicando el Teorema 5 para la distribución marginal de θ_i *proporciona el resultado.*

$\diamond$

Ejemplo 93 *Volvemos al Ejemplo 90.*

Se tiene $\mathbf{X}^T\mathbf{X} = n$ *y* $(\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{n}$. *Además,*

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$$

y entonces,

$$\theta|\phi, \mathbf{y} \sim \mathcal{N}\left(\bar{y}, \frac{1}{n\phi}\right).$$

El estimador de la varianza es

$$s^2 = \frac{1}{n-1}\mathbf{y}^T(\mathbf{I} - \frac{1}{n}\mathbf{1})\mathbf{y}$$

donde $\mathbf{1}$ *es una matriz de unos. Luego,* $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ *y se tiene*

$$\phi|\mathbf{y} \sim \mathcal{G}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right),$$

el mismo resultado que se ha visto en el Capítulo 3.

Ejemplo 94 *En el Ejemplo 91,*

$$
\begin{aligned}
\mathbf{X}^T\mathbf{X} &= \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix} \\
(\mathbf{X}^T\mathbf{X})^{-1} &= \frac{1}{n\sum_{i=1}^n (x_i-\bar{x})^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix} \\
\mathbf{X}^T\mathbf{y} &= \begin{pmatrix} n\bar{y} \\ \sum_{i=1}^n x_i y_i \end{pmatrix} \\
(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} &= \begin{pmatrix} \bar{y} - \bar{x}\frac{\sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^n (x_i-\bar{x})^2} \\ \frac{\sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^n (x_i-\bar{x})^2} \end{pmatrix} \\
&= \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix}
\end{aligned}
$$

donde $\widehat{\alpha}$ y $\widehat{\beta}$ son los estimadores clásicos de mínimos cuadrados.

$$\begin{array}{c}\alpha \\ \beta\end{array}\Bigg|\phi, \mathbf{y} \sim \mathcal{N}\left(\begin{array}{c}\widehat{\alpha} \\ \widehat{\beta}\end{array}, \frac{1}{\phi}(\mathbf{X}^T\mathbf{X})^{-1}\right)$$

Además,

$$\begin{aligned}\mathbf{y}^T(\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\mathbf{y} &= \left(\sum_{i=1}^{n}(y_i - \bar{y})^2 - \right. \\ &\qquad \left. \widehat{\beta}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right)\end{aligned}$$

y luego se tiene

$$\phi|\mathbf{y} \sim \mathcal{G}\left(\frac{n-2}{2}, \frac{(n-2)s^2}{2}\right)$$

donde s^2 es el estimador clásico del error residuo.

Se ve que los estimadores coinciden con los estimadores clásicos.

Ejemplo 95 *Retomando el Ejemplo 92,*

$$\begin{aligned}
\mathbf{X}^T\mathbf{X} &= \begin{pmatrix} n_1 & 0 & 0 & \dots & 0 \\ 0 & n_2 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & n_p \end{pmatrix} \\
(\mathbf{X}^T\mathbf{X})^{-1} &= \begin{pmatrix} \frac{1}{n_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{n_2} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{n_p} \end{pmatrix} \\
\mathbf{X}^T\mathbf{y} &= \begin{pmatrix} n_1\bar{y}_{1\cdot} \\ n_2\bar{y}_{2\cdot} \\ \vdots \\ n_p\bar{y}_{p\cdot} \end{pmatrix} \qquad \textit{donde } \bar{y}_{i\cdot} = \tfrac{1}{n_i}\textstyle\sum_{j=1}^{n_i} y_{ij} \\
(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} &= \begin{pmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \\ \vdots \\ \bar{y}_{p\cdot} \end{pmatrix}
\end{aligned}$$

Luego se tiene

$$\boldsymbol{\theta}|\mathbf{y} \sim \mathcal{N}\left(\begin{array}{c} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \\ \vdots \\ \bar{y}_{p\cdot} \end{array}, \frac{1}{\phi}\left(\begin{array}{ccccc} \frac{1}{n_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{n_2} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{n_p} \end{array}\right)\right)$$

Además,

$$\begin{aligned} \mathbf{y}^T(\mathbf{I}-\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\mathbf{y} &= \sum_{i=1}^{p}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{p} n_i \bar{y}_{i\cdot}^2 \\ &= \sum_{i=1}^{p}\sum_{j=1}^{n_i} (y_{ij}-\bar{y}_{i\cdot})^2 \end{aligned}$$

y entonces

$$\phi|\mathbf{y} \sim \mathcal{G}\left(\frac{n-p}{2}, \frac{(n-p)s^2}{2}\right)$$

donde $s^2 = \frac{1}{n-p}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\cdot})^2$.

Luego

$$\frac{\theta_i - \bar{y}_{i\cdot}}{s/\sqrt{n_i}} \sim t_{n-p}$$

y un intervalo de credibilidad coincide con el intervalo clásico.

Además, si se quiere estimar la diferencia entre dos efectos $\delta_{i_1 i_2} = \theta_{i_1} - \theta_{i_2}$ *se tiene*

$$\delta_{i_1 i_2}|\phi, \mathbf{y} \sim \mathcal{N}\left(\bar{y}_{i_1\cdot} - \bar{y}_{i_2\cdot}, \frac{1}{\phi}\left(\frac{1}{n_{i_1}} + \frac{1}{n_{i_2}}\right)\right)$$

y entonces un intervalo de credibilidad para $\delta_{i_1 i_2}$ *será*

$$\bar{y}_{i_1\cdot} - \bar{y}_{i_2\cdot} \pm s\sqrt{\frac{1}{n_{i_1}} + \frac{1}{n_{i_2}}} t_{n-p}(0 - 025)$$

igual al intervalo clásico de confianza.

La distribución marginal de θ

Integrando la distribución conjunta de $\boldsymbol{\theta}, \phi$ con respeto a ϕ, se tiene

$$
\begin{aligned}
f(\boldsymbol{\theta}|\mathbf{y}) &\propto \left((\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})^T\mathbf{X}^T\mathbf{X}(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})+(n-p)s^2\right)^{-\frac{n}{2}} \\
&\propto \left(1+\frac{(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})^T\mathbf{X}^T\mathbf{X}(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})}{(n-p)s^2}\right)^{-\frac{n-p+p}{2}}
\end{aligned}
$$

que es el núcleo de una distribución t multivariante, esclada y no centrada.

Es una densidad muy complicada pero fácil de muestrear con el siguiente algoritmo Monte Carlo:

`Para` $j = 1, \ldots, R$

1. `Muestrear` $\phi_j \sim \mathcal{G}\left(\frac{n-p}{2}, \frac{(n-p)s^2}{2}\right)$

2. `Muestrear` $\boldsymbol{\theta}_j \sim \mathcal{N}\left(\widehat{\boldsymbol{\theta}}, \frac{1}{\phi_j}(\mathbf{x}^T\mathbf{X})^{-1}\right)$

Predicción a través de modelos lineales

Sea $\tilde{\mathbf{X}}$ una nueva matriz de diseño. A menudo, se quiere predecir tanto la respuesta media $\tilde{\mathbf{X}}\boldsymbol{\theta}$ como nuevas observaciones $\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\boldsymbol{\theta} + \tilde{\boldsymbol{\epsilon}}$.

Luego se tiene

$$\begin{aligned}
\tilde{\mathbf{X}}\boldsymbol{\theta}|\phi, \mathbf{y} &\sim \mathcal{N}\left(\tilde{\mathbf{X}}\hat{\boldsymbol{\theta}}, \frac{1}{\phi}\tilde{\mathbf{X}}(\mathbf{X}^T\mathbf{X})^{-1}\tilde{\mathbf{X}}^T\right) \\
\tilde{\mathbf{Y}}|\phi, \mathbf{y} &\sim \mathcal{N}\left(\tilde{\mathbf{X}}\hat{\boldsymbol{\theta}}, \frac{1}{\phi}\left(\tilde{\mathbf{I}} + \tilde{\mathbf{X}}(\mathbf{X}^T\mathbf{X})^{-1}\tilde{\mathbf{X}}^T\right)\right)
\end{aligned}$$

y las distribuciones predictivas de $\tilde{\mathbf{X}}\boldsymbol{\theta}$ y $\tilde{\mathbf{Y}}$ son distribuciones t multivariantes.

No obstante, si $\tilde{\mathbf{x}}\theta$ es escalar e

$$\tilde{Y} = \tilde{\mathbf{x}}\boldsymbol{\theta} + \tilde{\epsilon}$$

entonces

$$\begin{aligned}
\tilde{\mathbf{x}}\theta|\phi, \mathbf{y} &\sim \mathcal{N}\left(\tilde{\mathbf{x}}\widehat{\boldsymbol{\theta}}, \frac{1}{\phi}\tilde{\mathbf{x}}(\mathbf{X}^T\mathbf{X})^{-1}\tilde{\mathbf{x}}^T\right) \\
\tilde{Y}|\phi, \mathbf{y} &\sim \mathcal{N}\left(\tilde{\mathbf{x}}\widehat{\boldsymbol{\theta}}, \frac{1}{\phi}\left(1 + \tilde{\mathbf{x}}(\mathbf{X}^T\mathbf{X})^{-1}\tilde{\mathbf{x}}^T\right)\right)
\end{aligned}$$

y, (integrando con respeto a ϕ,) las distribuciones predictivas de $\tilde{\mathbf{x}}\theta$ e $\tilde{Y}$ son

$$\begin{aligned}
\frac{\tilde{\mathbf{x}}\theta - \tilde{\mathbf{x}}\widehat{\boldsymbol{\theta}}}{s\sqrt{\tilde{\mathbf{x}}(\mathbf{X}^T\mathbf{X})^{-1}\tilde{\mathbf{x}}^T}} \mid \mathbf{y} &\sim t_{n-p} \\
\frac{\tilde{Y} - \tilde{\mathbf{x}}\widehat{\boldsymbol{\theta}}}{s\sqrt{1 + \tilde{\mathbf{x}}(\mathbf{X}^T\mathbf{X})^{-1}\tilde{\mathbf{x}}^T}} \mid \mathbf{y} &\sim t_{n-p}.
\end{aligned}$$

Entonces, los intervalos predictivos son iguales a los intervalos de predicción clásicos.

Ejemplo 96 *Volviendo a los Ejemplos 91 y 94, sea $\tilde{Y} = \alpha + \beta\tilde{x} + \tilde{\epsilon}$ una nueva observación.*

Entonces $\tilde{\mathrm{x}} = (1, \tilde{x})$ *y*

$$\tilde{x}(\mathbf{X}^T\mathbf{X})^{-1}\tilde{x}^T = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \tilde{x})^2}{n\sum_{i=1}^n (x_i - \bar{x})^2}$$

Luego se tiene

$$\frac{\tilde{Y} - (\hat{\alpha} + \hat{\beta}\tilde{x})}{s\sqrt{1 + \frac{1}{n} + \frac{(\tilde{x}-\bar{x})^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}} \sim t_{n-2}$$

y un intervalo predictivo para $\tilde{Y}$ es igual al intervalo clásico.

Un modelo lineal más general

En el modelo inicial se ha supuesto que la varianza de $\boldsymbol{\epsilon}$ es proporcional a la identidad. Ahora sea $\boldsymbol{\epsilon}|\boldsymbol{\Sigma} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ para una matriz de covarianzas más general $\boldsymbol{\Sigma}$ que implica el modelo

$$\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\Sigma} \sim \mathcal{N}\left(\mathbf{X}\boldsymbol{\theta}, \boldsymbol{\Sigma}\right).$$

Se consideran varias estructuras para la distribución a priori de $\boldsymbol{\theta}, \boldsymbol{\Sigma}$.

Inferencia conjugada con Σ conocida: el modelo lineal de dos etapas

Supongamos una distribución a priori normal, $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{m}, \mathbf{V})$ con $\mathbf{m}$ y $\mathbf{V}$ conocidos.

Recordamos el Ejemplo 30. Este modelo (Lindley y Smith 1972) se llama *el modelo lineal de 2 etapas*.

Teorema 14 *La distribución marginal de* $\mathbf{Y}$ *es*

$$\mathbf{Y} \sim \mathcal{N}\left(\mathbf{X}\mathbf{m}, \boldsymbol{\Sigma} + \mathbf{X}\mathbf{V}\mathbf{X}^T\right)$$

y la distribución a posteriori de $\boldsymbol{\theta}$ *dado* $\mathbf{y}$ *es*

$$\boldsymbol{\theta}|\mathbf{y} \sim \mathcal{N}(\mathbf{m}^*, \mathbf{V}^*) \qquad \textit{donde}$$

$$\begin{aligned}
\mathbf{V}^* &= \left(\mathbf{V}^{-1} + \mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1} \\
\mathbf{m}^* &= \mathbf{V}^*\left(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} + \mathbf{V}^{-1}\mathbf{m}\right) \\
&= \left(\mathbf{V}^{-1} + \mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} + \mathbf{V}^{-1}\mathbf{m}\right)
\end{aligned}$$

Demostración

En primer lugar hallamos la distribución marginal de $\mathbf{Y}$*. Se tiene*

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}$$

y notando que

$$\mathbf{X}\boldsymbol{\theta} \sim \mathcal{N}\left(\mathbf{X}\mathbf{m}, \mathbf{X}\mathbf{V}\mathbf{X}^T\right)$$

y que la distribución de marginal de $\mathbf{Y}$ *es la distribución de la suma de dos variables normales,* $\mathbf{X}\boldsymbol{\theta}$ *y* $\boldsymbol{\epsilon}$*, entonces se tiene*

$$\mathbf{Y} \sim \mathcal{N}\left(\mathbf{X}\mathbf{m}, \mathbf{X}\mathbf{V}\mathbf{X}^T + \boldsymbol{\Sigma}\right).$$

Ahora mediante el teorema de Bayes,

$$
\begin{aligned}
f(\boldsymbol{\theta}|\mathbf{y}) &\propto f(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta}) \\
&\propto \exp\left(-\frac{1}{2}(\boldsymbol{\theta}-\mathbf{m})^T\mathbf{V}^{-1}(\boldsymbol{\theta}-\mathbf{m})\right) \\
&\quad \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\boldsymbol{\theta}^T(\mathbf{V}^{-1}+\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})\boldsymbol{\theta}-\right.\right. \\
&\quad \left.\left. 2\boldsymbol{\theta}^T(\mathbf{V}^{-1}\mathbf{m}+\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y})\right]\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\boldsymbol{\theta}^T\mathbf{V}^{*-1}\boldsymbol{\theta}-\right.\right. \\
&\quad \left.\left. 2\boldsymbol{\theta}^T\mathbf{V}^{*-1}\mathbf{V}^*(\mathbf{V}^{-1}\mathbf{m}+\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y})\right]\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\boldsymbol{\theta}^T\mathbf{V}^{*-1}\boldsymbol{\theta}-2\boldsymbol{\theta}^T\mathbf{V}^{*-1}\mathbf{m}^*\right]\right) \\
&\propto \exp\left(-\frac{1}{2}(\boldsymbol{\theta}-\mathbf{m}^*)^T\mathbf{V}^{*-1}(\boldsymbol{\theta}-\mathbf{m}^*)\right)
\end{aligned}
$$

que es el núcleo de una distribución normal $\mathcal{N}(\mathbf{m}^*, \mathbf{V})$.

$\diamond$

Ejemplo 97 *Retomando el Ejemplo 90 con observaciones univariables y la distribución a priori* $\theta \sim \mathcal{N}\left(m, \frac{1}{\alpha\phi}\right)$ *se tiene*

$$\begin{aligned}
\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X} &= n\phi \\
\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{y} &= n\phi\bar{y} \\
\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X} + \mathbf{V}^{-1} &= (n+\alpha)\phi \\
\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{y} + \mathbf{V}^{-1}\mathbf{m} &= n\phi\bar{y} + \alpha\phi m \\
m^* &= \frac{n\bar{y} + \alpha m}{n+\alpha}
\end{aligned}$$

que es el resultado del Ejemplo 21.

Relación con el EMV

El estimador de mínimos cuadrados para este problema es

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}$$

y entonces

$$\begin{aligned}
\mathbf{m}^* &= \left(\mathbf{V}^{-1} + \mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} + \mathbf{V}^{-1}\mathbf{m}\right) \\
&= \left(\mathbf{V}^{-1} + \mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\left[\left(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)\hat{\boldsymbol{\theta}} + \mathbf{V}^{-1}\mathbf{m}\right]
\end{aligned}$$

y la media a posteriori es una media ponderada de la media a priori $\mathbf{m}$ y el EMV $\hat{\boldsymbol{\theta}}$ con ponderaciones proporcionales a las matrices de precisión.

Observación 58 *La media a posteriori de $\boldsymbol{\theta}$ corresponde en la estadística clásica a un estimador contraído. Se utilizan estimadores de este tipo si la matriz $\mathbf{X}^T\boldsymbol{\Upsilon}^{-1}\mathbf{X}$ es casi singular, es decir si hay colinealidad entre los datos.*

Ver, por ejemplo, Stein (1956).

Observación 59 *Se puede expresar la media a posteriori de otra manera.*

$$
\begin{aligned}
\mathbf{m}^* &= \mathbf{V}^* \left[\mathbf{X}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\mathbf{m}) + \left(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}+\mathbf{V}^{-1}\right)\mathbf{m}\right] \\
&= \mathbf{m} + \mathbf{V}^*\mathbf{X}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\mathbf{m})
\end{aligned}
$$

La cantidad $\mathbf{y}-\mathbf{X}\mathbf{m}$ *es la diferencia entre la observación* $\mathbf{y}$ *y su esperanza a priori* $\mathbf{X}\mathbf{m}$. *La cantidad* $\mathbf{V}^*\mathbf{X}^T\boldsymbol{\Sigma}^{-1}$ *se llama un filtro*

Observación 60 *Dejando que* $\mathbf{V}^{-1} \to \mathbf{0}$ *en las fórmulas del Teorema 14, proporciona el resultado de utilizar una distribución a priori uniforme para* $\boldsymbol{\theta}$. *En este caso, se tiene*

$$
\boldsymbol{\theta}|\mathbf{y} \sim \mathcal{N}\left((\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}, (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\right).
$$

Suponiendo que $\boldsymbol{\Sigma} = \frac{1}{\phi}\mathbf{I}$ *se tiene el resultado del Teorema 13.*

Resultados suponiendo la varianza Σ desconocida

Sea $\boldsymbol{\Sigma} = \frac{1}{\phi}\boldsymbol{\Upsilon}$ con $\boldsymbol{\Upsilon}$ conocida.

Dada la distribución a priori no informativa

$$f(\boldsymbol{\theta}, \phi) \propto \frac{1}{\phi}$$

se tiene el siguiente resultado:

Teorema 15 *La distribución a posteriori es*

$$\begin{aligned}
\boldsymbol{\theta}|\phi, \mathbf{y} &\sim \mathcal{N}\left((\mathbf{X}^T\boldsymbol{\Upsilon}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Upsilon}^{-1}\mathbf{y}, \frac{1}{\phi}(\mathbf{X}^T\boldsymbol{\Upsilon}^{-1}\mathbf{X})^{-1}\right) \\
\phi|\mathbf{y} &\sim \mathcal{G}\left(\frac{n-p}{2}, \frac{(n-p)s^2}{2}\right) \quad \textit{donde}
\end{aligned}$$

$$s^2 = \frac{1}{n-p}\left(\mathbf{y}^T\boldsymbol{\Upsilon}^{-1}\mathbf{y} - \mathbf{y}^T\boldsymbol{\Upsilon}^{-1}\mathbf{X}^T(\mathbf{X}^T\boldsymbol{\Upsilon}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Upsilon}^{-1}\mathbf{y}\right).$$

Demostración

El resultado sigue del Teorema 13.

Sea $\mathbf{\Upsilon}^{-1} = \mathbf{\Psi}^T\mathbf{\Psi}$ *donde* $\mathbf{\Psi}$ *es la descomposición de Cholesky. Luego, definiendo* $\mathbf{Z} = \mathbf{\Psi}\mathbf{Y}$*, se tiene*

$$\mathbf{Z}|\boldsymbol{\theta}, \phi \sim \mathcal{N}\left(\mathbf{\Psi}\mathbf{X}\boldsymbol{\theta}, \frac{1}{\phi}\mathbf{I}\right).$$

Además, se tiene

$$l(\boldsymbol{\theta}, \phi|\mathbf{z}) \propto l(\boldsymbol{\theta}, \phi|\mathbf{y})$$

y entonces, mediante el Teorema 13, definiendo $\mathbf{X}^* = \mathbf{\Psi}\mathbf{x}$*, se tiene*

$$\begin{aligned} f(\boldsymbol{\theta}, \phi|\mathbf{y}) &\propto f(\boldsymbol{\theta}, \phi|\mathbf{z}) \\ \boldsymbol{\theta}|\phi, \mathbf{z} &\sim \mathcal{N}\left((\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\mathbf{z}, \frac{1}{\phi}, (\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\right) \\ \phi|\mathbf{z} &\sim \mathcal{G}\left(\frac{n-p}{2}, \frac{(n-p)s^2}{2}\right) \end{aligned}$$

donde

$$\begin{aligned} s^2 &= \frac{1}{n-p}\mathbf{z}^T\left(\mathbf{I}-\mathbf{X}^*(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\right)\mathbf{z} \\ &= \frac{1}{n-p}\left(\mathbf{z}^T\mathbf{z}-\mathbf{z}^T\mathbf{X}^*(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\mathbf{z}\right). \end{aligned}$$

Ahora,

$$\begin{aligned} (\mathbf{X}^{*T}\mathbf{X}^*) &= \mathbf{X}^T\mathbf{\Psi}^T\mathbf{\Psi}\mathbf{X} \\ &= \mathbf{X}^T\mathbf{\Upsilon}^{-1}\mathbf{X} \\ \mathbf{X}^{*T}\mathbf{z} &= \mathbf{X}^T\mathbf{\Psi}^T\mathbf{\Psi}\mathbf{y} \\ &= \mathbf{X}^T\mathbf{\Upsilon}^{-1}\mathbf{y} \\ \mathbf{z}^T\mathbf{z} &= \mathbf{y}^T\mathbf{\Psi}^T\mathbf{\Psi}\mathbf{y} \\ &= \mathbf{y}^T\mathbf{\Upsilon}^{-1}\mathbf{y} \\ \mathbf{z}^T\mathbf{X}^*(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\mathbf{z} &= \mathbf{y}^T\mathbf{\Psi}^T\mathbf{\Psi}\mathbf{X}\left(\mathbf{X}^T\mathbf{\Upsilon}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{\Upsilon}^{-1}\mathbf{y} \\ &= \mathbf{y}^T\mathbf{\Upsilon}^{-1}\mathbf{y}\left(\mathbf{X}^T\mathbf{\Upsilon}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{\Upsilon}^{-1}\mathbf{y} \end{aligned}$$

$\diamond$

Observación 61 *Es posible extender los análisis al caso en que se utiliza una distribución a priori normal gamma.*

En este caso, la distribución a posteriori de $\boldsymbol{\theta}$ *es conjugada y la distribución marginal de* $\boldsymbol{\theta}$ *es t multivariante.*

Observación 62 *En el caso en lo cuál no se sabe la forma de* $\boldsymbol{\Sigma}$*, se puede utilizar el modelo normal Wishart introducido en el Capítulo 3, por ejemplo*

$$
\begin{aligned}
\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\Sigma} &\sim \mathcal{N}(\mathbf{X}\boldsymbol{\theta}, \boldsymbol{\Sigma}) \\
\boldsymbol{\theta}|\boldsymbol{\Sigma} &\sim \mathcal{N}\left(\mathbf{m}, \frac{1}{\alpha}\boldsymbol{\Sigma}\right) \\
\boldsymbol{\Sigma} &\sim \mathcal{WI}(\nu, \mathbf{W})
\end{aligned}
$$

La distribución a priori será conjugada. No obstante, su estructura es muy poco razonable en la mayoría de los problemas.

Métodos Numéricos para Modelos Lineales

En muchos problemas, no se conocen los valores de las matrices de varianza y las distribuciones conjugadas no representan bién las creencias a priora.

Ejemplo 98 *Volvemos al modelo de regresión lineal simple del Ejemplo 91, pero con varianza desconocida*

$$f(\phi) \propto \frac{1}{\phi}$$

y una distribución a priori para α, β *independiente de* ϕ*, es decir*

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \sim \mathcal{N}\left(\begin{matrix} m_a \\ m_b \end{matrix} ; \begin{pmatrix} v_a & 0 \\ 0 & v_b \end{pmatrix} \right).$$

En este caso, la inferencia no es conjugada pero, se pueden calcular las distribuciones acondicionadas a posteriori:

Sean $\boldsymbol{\theta} = (\alpha, \beta)^T$ *y* $\mathbf{X}$ *como en el Ejemplo 91. Entonces:*

$$
\begin{array}{rcl}
\phi|\boldsymbol{\theta}, \mathbf{y} & \sim & \mathcal{G}\left(\frac{n}{2}, \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\theta})\right) \\
\boldsymbol{\theta}|\phi, \mathbf{y} & \sim & \mathcal{N}(\mathbf{m}^*, \mathbf{V}^*) \qquad \textit{donde} \\
\mathbf{V}^* & = & \left(\phi\mathbf{X}^T\mathbf{X} + \mathbf{V}^{-1}\right)^{-1} \\
\mathbf{m}^* & = & \mathbf{V}^*\left(\phi(\mathbf{X}^T\mathbf{X})\hat{\boldsymbol{\theta}} + \mathbf{V}^{-1}\mathbf{m}\right)
\end{array}
$$

Entonces, es posible definir un algoritmo Gibbs para muestrear la distribución a posteriori conjunta de ϕ, α, β.

1. $t = 0$. Valores iniciales $\boldsymbol{\theta}^{(0)} = (\alpha^{(0)}, \beta^{(0)})$, $\phi^{(0)}$.

2. Muestrear
$\phi^{(t+1)} \sim \mathcal{G}\left(\frac{n}{2}, \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}^{(t)})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}^{(t)})\right)$.

3. Sean

   a) $\mathbf{V}^{*(t+1)} = \left(\phi^{(t+1)}\mathbf{X}^T\mathbf{X} + \mathbf{V}^{-1}\right)^{-1}$

   b) $\mathbf{m}^{*(t+1)} = \mathbf{V}^{*(t+1)}\left(\phi^{(t+1)}(\mathbf{X}^T\mathbf{X})\hat{\boldsymbol{\theta}} + \mathbf{V}^{-1}\mathbf{m}\right)$.

4. Muestrear
$\boldsymbol{\theta}^{(t+1)} \sim \mathcal{N}\left(\mathbf{m}^{*(t+1)}, \mathbf{V}^{*(t+1)}\right)$.

5. $t = t + 1$

6. Ir a 2.

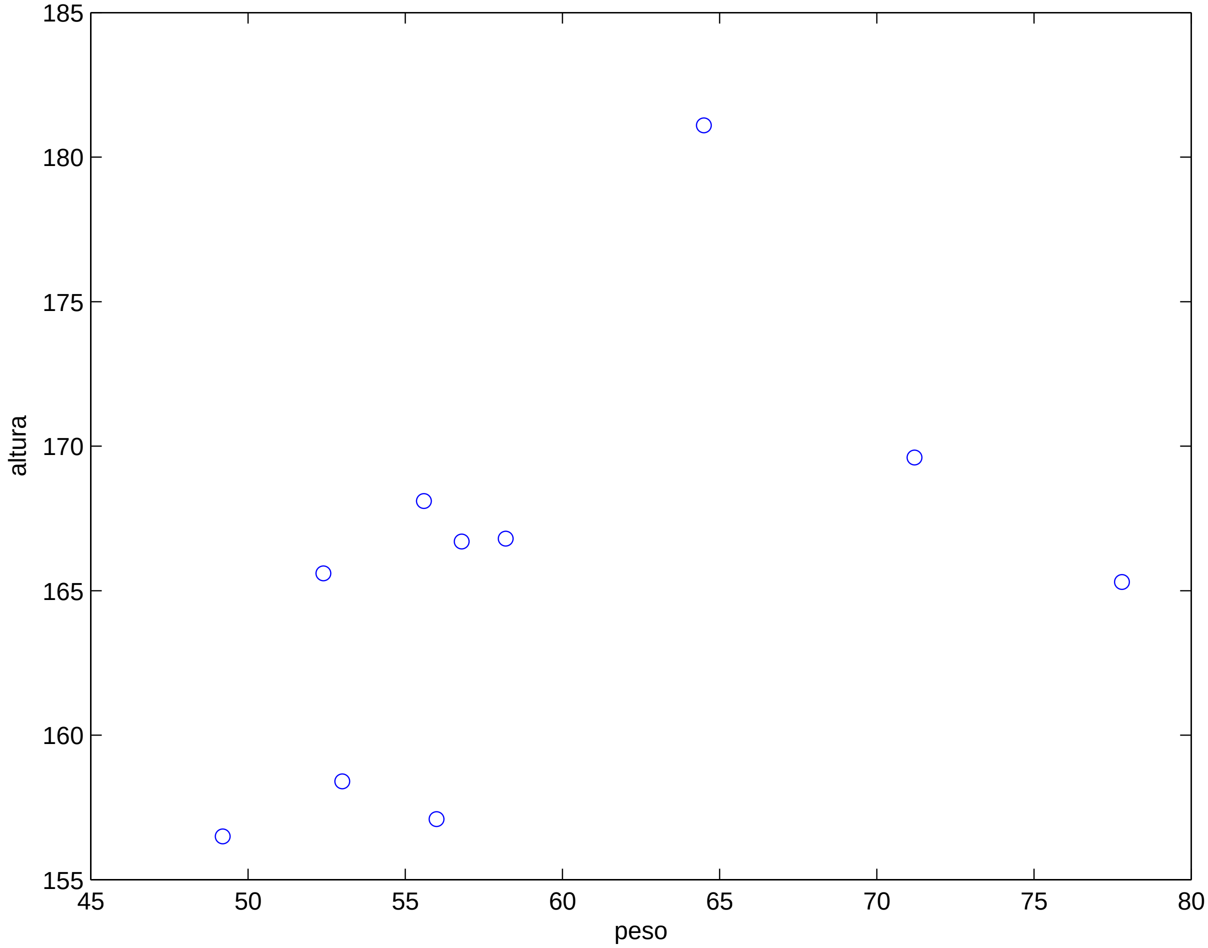
Despues de un tiempo suficiente, los valores muestreados simulan una muestra de la distribución a posteriori de $\alpha, \beta, \phi|\mathbf{y}$.

Ejemplo 99 *Se tiene los siguientes datos de peso* (x) *y altura* (y) *de 10 personas.*

x	71,2	58,2	56,0	64,5	53,0
y	169,6	166,8	157,1	181,1	158,4
x	52,4	56,8	49,2	55,6	77,8
y	165,6	166,7	156,5	168,1	165,3

Se ve en el siguiente gráfico que existe una relación positiva entre las dos variables. Supongamos el modelo lineal

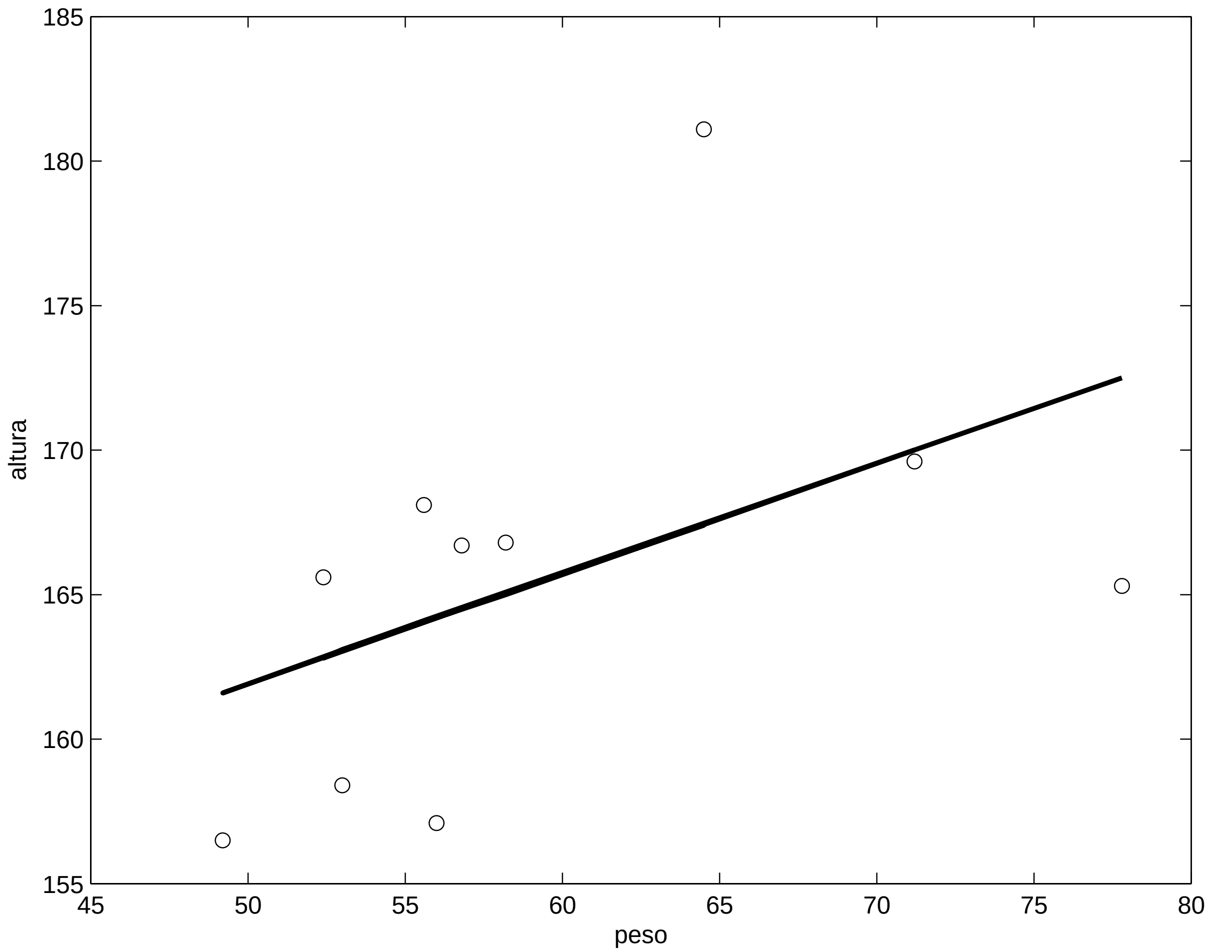
$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$



Supongamos un modelo con distribuciones independientes (poca informativas) para β_0. β_1 y ϕ. Se ve el codigo Winbugs para este modelo en la siguiente transparencia.

```
#----MODEL Definition----------------

model {

for(i in 1:N) {
    height[i] ~ dnorm(mu[i],tau)
    mu[i] <- beta0 + beta1* weight[i] }
beta0 ~ dnorm(0,1.0E-6)
beta1 ~ dnorm(0,1.0E-6)
#---------prior 1
tau ~ dgamma(1.0E-3,1.0E-3)
sigma2 <- 1.0/tau
#---------prior 2
#tau <- 1.0/(sigma*sigma)
#sigma ~ dunif(0,1000)
#sigma2 <-
sigma*sigma }

#----Initial values file----------------------------
list(beta0=0, beta1 = 0, tau = 1) #list(beta0 = 0, beta1
=0,sigma= 1)


#----Data File----------------------------------

list(N= 10,
height=c(169.6,166.8,157.1,181.1,158.4,165.6,166.7,156.5,168.1,165.3),
weight =c(71.2,58.2,56.0,64.5,53.0,52.4,56.8,49.2,55.6,77.8))
```

Las medias a posteriori de β_0 y β_1 son 143,1 *y* 0,377 *respectivamente. El siguiente dibujo ilustra la recta ajustada.*



Efectos fijos y aleatorios

Hasta ahora, se ha supuesto que la matriz de diseño $\mathbf{X}$ es conocida. Observamos que para hacer inferencia sobre $\boldsymbol{\theta}$ y $\boldsymbol{\Sigma}$, no importa si $\mathbf{X}$ es fija o aleatoria ya que

$$\begin{aligned} f(\boldsymbol{\theta}, \boldsymbol{\Sigma}|\mathbf{X}, \mathbf{y}) &\propto f(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{X}) f(\boldsymbol{\theta}, \boldsymbol{\Sigma}|\mathbf{X}) \\ &\propto f(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{X}) f(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \end{aligned}$$

suponiendo que los efectos $\mathbf{X}$ no dependen de $\boldsymbol{\theta}, \boldsymbol{\Sigma}$.

No obstante, para predicción se debe tomar en cuenta la incertidumbre sobre $\mathbf{X}$.

Comprobación del modelo y selección de variables

- Se puede estimar los residuos $\mathbf{y} - E[\mathbf{Y}|\boldsymbol{\theta}]$ o usar los residuos clásicos. Para otras ideas ver Gelman et al (1995).

- Para selección de variables se pueden usar factores Bayes, o el DIC o el BIC.

Aplicación 4: El modelo Lanchester de combate

Seguimos el análisis de Wiper et al (2000).

Las ecuaciones de Lanchester (1916) para la guerra moderna entre ejércitos de tamaños $x(t)$ e $y(t)$ son

$$\begin{aligned} \frac{\partial x}{\partial t} &= -\alpha y \\ \frac{\partial y}{\partial t} &= -\beta x \end{aligned} \qquad (1)$$

Estas ecuaciones implican la ley cuadrática de Lanchester:

$$\beta(x(0)^2 - x(t)^2) = \alpha(y(0)^2 - y(t)^2)$$

donde $x(0)$ e $y(0)$ son los tamaños iniciales de los dos ejércitos.

Un sistema más general

Se ha desarrollada un sistema más completa para otros tipos de guerra.

$$\begin{aligned}\frac{\partial x}{\partial t} &= -\beta x^{\phi_1} y^{\phi_2} \\ \frac{\partial y}{\partial t} &= -\alpha y^{\phi_1} x^{\phi_2}\end{aligned} \tag{2}$$

$\phi = (0, 0)$ proporciona una ley lineal (mano-a-mano),

$\phi = (0, 1)$ es la ley cuadrática (guerra moderna con fuego puntado),

$\phi = (1, 1)$ es una ley lineal (fuego aleatorio),

$\phi = (1, 0)$ es una ley logística (¿guerra de gran escala?).

Se tiene interés en

- clasificar los tipos de batalla (análisis histórica)
- medir la fuerza relativa de los dos ejércitos (es decir la razón α/β)
- predecir el ganador de la batalla

Pero los modelos son determinísticos.

¿Cómo incluimos la incertidumbre?

Discretizando el sistema.

Supongamos que se recuerdan las bajas al finalizar cada día. Luego, discretizando las ecuaciones de Lanchester en 2, se tiene

$$\begin{aligned} \Delta x_t &\approx -\beta x_t^{\phi_1} y_t^{\phi_2} \\ \Delta y_t &\approx -\alpha y_t^{\phi_1} x_t^{\phi_2} \end{aligned} \tag{3}$$

donde x_t and y_t son los tamaños de cada ejército al inicio del día y Δx_t y Δy_t son menos las bajas diarias de los dos ejércitos.

Linealización

Bracken (1995) utilizó el método de mínimos cuadrados para esimar los parámetros de 3 dados los datos de una batalla. Pero existen muchos problemas debidos a la no linealidad del sistema.

Tomando logarítmos en 3 y introduciendo un error tenemos un modelo lineal.

$$\mathbf{z}_t = \boldsymbol{\theta} + \mathbf{P}_t \boldsymbol{\phi} + \boldsymbol{\epsilon}_t$$

donde $\mathbf{z}_t$ son los logarítmos de las bajas en día t, $\mathbf{P}_t = \begin{pmatrix} \log y_t & \log x_t \\ \log x_t & \log y_t \end{pmatrix}$ y $\boldsymbol{\phi} = (\phi_1, \phi_2)^T$.

(Para paliar los problemas de colinealidad, también se transforma $\mathbf{P}_t \to \mathbf{P}_t - \bar{\mathbf{P}}$).

Supongamos errores normales $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$ y ajustamos el modelo mediante el enfoque bayesiano.

¿Porqué bayesiano?

- mucha información a priori
- estimadores clásicas inestables por los problemas de colinealidad
- fáciles de ajustar

Dadas las distribuciones a priori ($\boldsymbol{\theta}$, $\boldsymbol{\phi} \sim \mathcal{N}(\cdot, \cdot)$, $\mathbf{V}^{-1} \sim$ Wishart inversa) se usa el muestreo Gibbs para muestrear las distribuciones a posteriori.

Ejemplo: la campaña de las Árdenas

- Datos de 32 días de la campaña

- Una batalla de muy grande escala

- fuerzas son una medida compuesta (soldados, cañones, tanques)

- los alemanes atacaron durante los primeros cinco días y luego cambió el rumbo de la batalla y los americanos empezaron a ganar la batalla.

El ajuste del modelo básico no es adecuado. Hay una gran diferencia entre las dos partes de la batalla. Se mejora el ajuste incluyendo un nuevo parámetro ($\boldsymbol{\delta}$), donde el modelo nuevo es

$$\mathbf{z}_t = \boldsymbol{\theta} + \mathrm{P}_t\boldsymbol{\phi} + \boldsymbol{\delta} I_t + \boldsymbol{\epsilon}_t$$

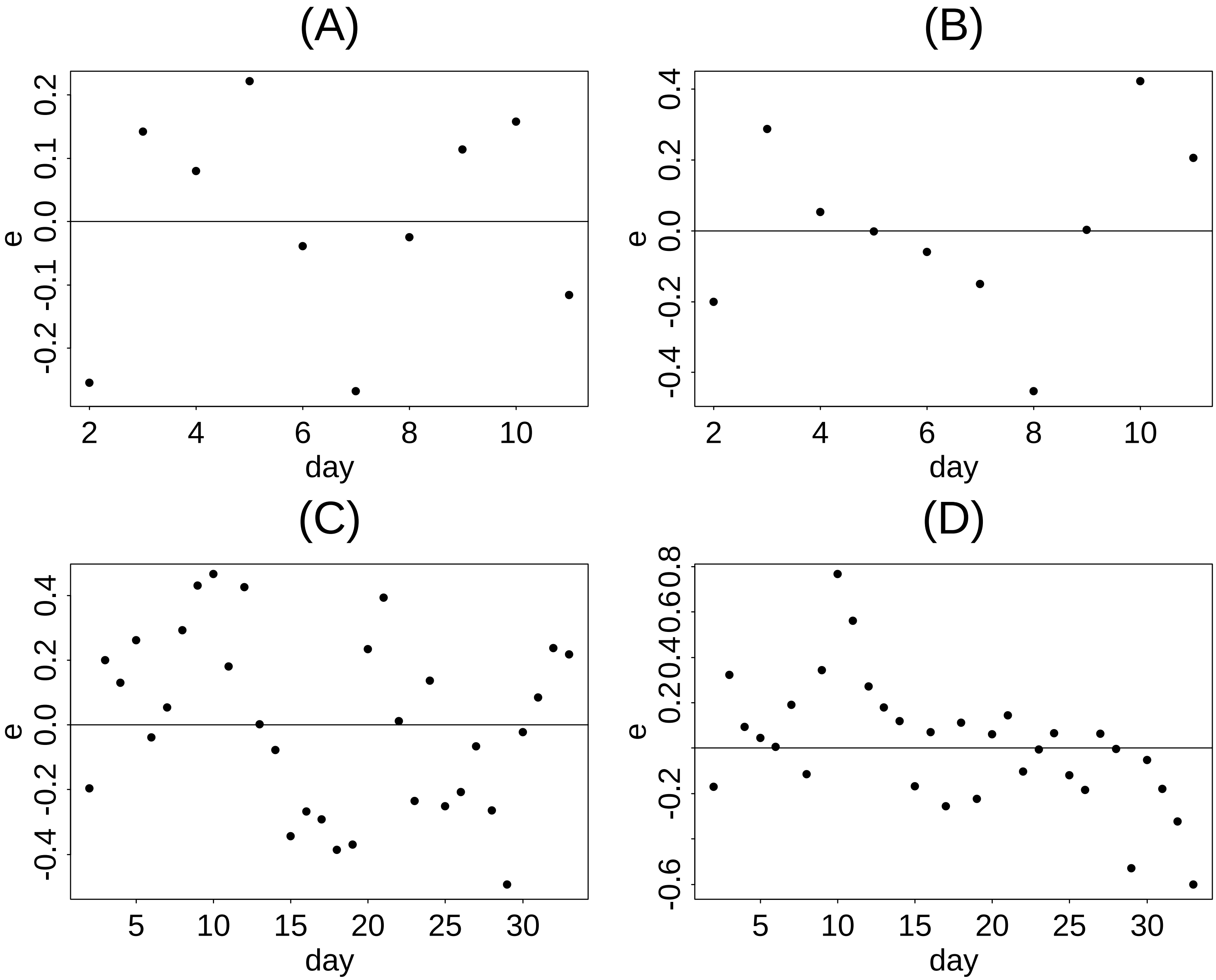
donde I_t es una función indicatriz del ejército atacante en el día t.

La tabla muestra las estimaciones a posteriori de los parámetros.

Parametro	Media (Desviación típica)	
$\boldsymbol{\theta}$	7,88 (,09)	8,04 (,08)
$\boldsymbol{\phi}$	0,84 (,41)	0,27 (,40)
$\boldsymbol{\delta}$	−0,38 (,09)	−0,01 (,09)

Errores estimados del ajuste

Se estiman los errores con $z_{ti} - E[z_{ti} | \text{parámetros}]$.



Conclusiones para esta batalla

1. Pocas diferencias en ajuste para varios modelos Lanchester (factores Bayes aprox. 1).

2. Menos bajas americanas en los días en que atacaron.

3. Pocas diferencias en las fuerzas.

4. Evidencia de falta de ajuste para los alemanes. (¿Efectos de desmoralización?).

5. Distribuciones a posteriori muy robustas a cambios en las a priori de $\boldsymbol{\theta}$, $\boldsymbol{\delta}$, bastante robustas a ϕ, y menos robustas a cambios en $\mathbf{V}$.

6. Modelos estocásticos pueden funcionar mejor. Ver Pettit et al (2003).