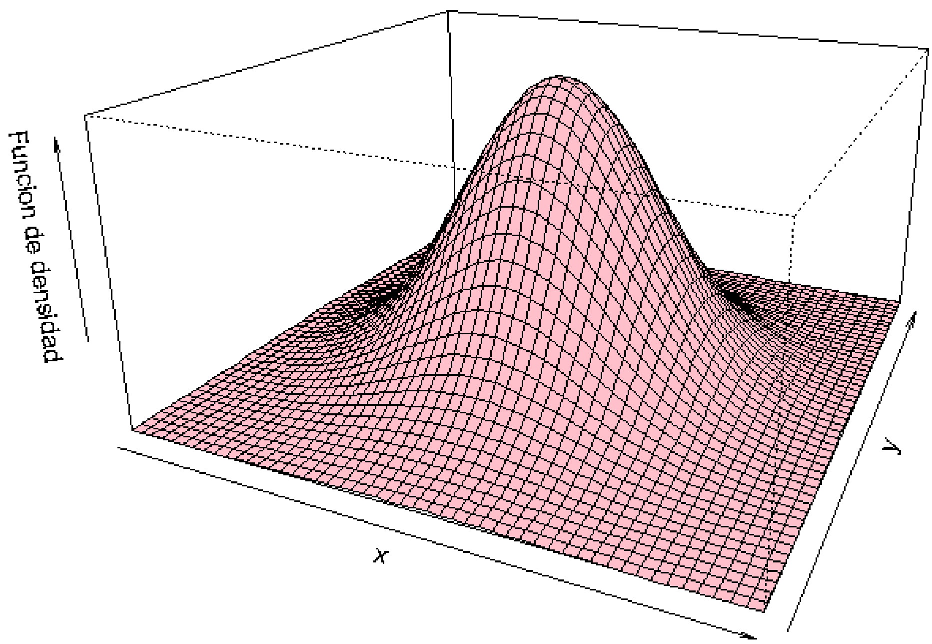
# Tema 4: Modelos Lineales

## Introducción

Un vector $\mathbf{y}$ de dimensión $k$ tiene una distribución normal multivariante con media $\mu$ y varianza $E[(\mathbf{y}-\mu)(\mathbf{y}-\mu)^t] = V$ si su función de densidad es

$$f(\mathbf{y}|\mu, V) = (2\pi)^{-\frac{k}{2}}|V|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y}-\mu)^t V^{-1}(\mathbf{y}-\mu)\right]$$



**NOTA**: La gráfica anterior está realizada con R, con la siguiente secuencia de comandos:

```
library(mvtnorm)
n = 50
x = seq(-3,3,length=n)
y = x
z = matrix(0,n,n)
sigma = diag(2)
for (i in 1:n)
    for (j in 1:n)
        z[i,j] = dmvnorm(c(x[i],y[j]),c(0,0),sigma)
    end
end
persp(x,y,z,theta=25,phi=20,zlab="Funcion de densidad",
expand=0.5,col="pink")
```

## Propiedades destacables de la distribución normal multivariante:

$(i)$ Cualquier subconjunto de $\mathbf{y}$ también se distribuye como normal. Si

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \sim N\left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} V_1 & V_{12} \\ V_{21} & V_2 \end{pmatrix} \right)$$

entonces $\mathbf{y}_1 \sim N(\mu_1, V_1)$ e $\mathbf{y}_2 \sim N(\mu_2, V_2)$.

$(ii)$ Si $\mathbf{y}_1 \sim N(\mu_1, V_1)$ e $\mathbf{y}_2 \sim N(\mu_2, V_2)$ son variables independientes con la misma dimensión entonces

$$\mathbf{y}_1 \pm \mathbf{y}_2 \sim N(\mu_1 \pm \mu_2, V_1 + V_2)$$

$(iii)$ Si $\mathbf{z} = D\mathbf{y}$ para una matriz $D$ entonces

$$\mathbf{z} \sim N\left(D\mu, DVD^T\right).$$

## Matrices, vectores y formas cuadráticas

Algunos resultados útiles son

$(i)$ Para matrices o vectores $A$ y $B$,

$$(A \cdot B)^t = B^t \cdot A^t.$$

$(ii)$ Si $\mathbf{x}_i$ $(i = 1, \ldots, n)$ y $\mu$ son vectores de dimensión $k$ y $V$ es una matriz simétrica de dimensión $k \times k$,

$$\sum_{i=1}^{n} \mathbf{x}_i^t V \mu = n\bar{\mathbf{x}}^t V \mu$$

donde $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$

(*iii*) Para vectores $\mathbf{x}$, $\mu$ de dimensión $k$ y una matriz simétrica $V$ se define una forma cuadrática como

$$(\mathbf{x}-\mu)^t V(\mathbf{x}-\mu).$$

Se tiene la expansión

$$\begin{aligned}(\mathbf{x}-\mu)^t V(\mathbf{x}-\mu) &= \mathbf{x}^t V\mathbf{x} - \mu^t V\mathbf{x} - \mathbf{x}^t V\mu + \mu^t V\mu \\ &= \mathbf{x}^t V\mathbf{x} - 2\mathbf{x}^t V\mu + \mu^t V\mu\end{aligned}$$

Se observa que el resultado es un **escalar**.

(*iv*) Se puede expresar una forma cuadrática de otra manera:

$$(\mathbf{x}-\mu)^t V(\mathbf{x}-\mu) = tr(V(\mathbf{x}-\mu)(\mathbf{x}-\mu)^t)$$

donde $tr(\cdot)$ es la traza de la matriz.

## Verosimilitud de la distribución normal

Sea $\mathbf{y}_1,\ldots,\mathbf{y}_n$ una muestra de una distribución normal multivariante $N(\mu,V)$. La verosimilitud correspondiente es

$$\begin{aligned}
L(\mu,V|\text{datos}) &\propto |V|^{-\frac{n}{2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i-\mu)^t V^{-1}(\mathbf{y}_i-\mu)\right)\\
&\propto |V|^{-\frac{n}{2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i-\bar{\mathbf{y}}+\bar{\mathbf{y}}-\mu)^t V^{-1}(\mathbf{y}_i-\bar{\mathbf{y}}+\bar{\mathbf{y}}-\mu)\right)\\
&\propto |V|^{-\frac{n}{2}}\exp\left(-\frac{1}{2}\left[\sum_{i=1}^{n}(\mathbf{y}_i-\bar{\mathbf{y}})^t V^{-1}(\mathbf{y}_i-\bar{\mathbf{y}})\right.\right.\\
&\qquad\left.\left.+2\sum_{i=1}^{n}(\mathbf{y}_i-\bar{\mathbf{y}})^t V^{-1}(\bar{\mathbf{y}}-\mu)+n(\bar{\mathbf{y}}-\mu)^t V^{-1}(\bar{\mathbf{y}}-\mu)\right]\right)\\
&\propto |V|^{-\frac{n}{2}}\exp\left(-\frac{1}{2}\left[\sum_{i=1}^{n}(\mathbf{y}_i-\bar{\mathbf{y}})^t V^{-1}(\mathbf{y}_i-\bar{\mathbf{y}})+n(\bar{\mathbf{y}}-\mu)^t V^{-1}(\bar{\mathbf{y}}-\mu)\right]\right)\\
&\propto |V|^{-\frac{n}{2}}\exp\left(-\frac{1}{2}\left[tr(V^{-1}S)\right]+n(\bar{\mathbf{y}}-\mu)^t V^{-1}(\bar{\mathbf{y}}-\mu)\right)
\end{aligned}$$

donde $S=\displaystyle\sum_{i=1}^{n}(\mathbf{y}_i-\bar{\mathbf{y}})(\mathbf{y}_i-\bar{\mathbf{y}})^t$ y $\sum_{i=1}^{n}(\mathbf{y}_i-\bar{\mathbf{y}})^t V^{-1}(\bar{\mathbf{y}}-\mu)=0$ aplicando la propiedad (*ii*).

La fórmula para la verosimilitud es, así, parecida al caso de la normal univariante.

# Modelo lineal

Un modelo lineal básico se puede escribir como

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon$$

donde $\mathbf{y}$ es un vector $n$-dimensional de observaciones, $\mathbf{X}$ es una matriz $n \times k$ de coeficientes conocidos, $\beta$ es un vector $k$-dimensional de parámetros y $\varepsilon$ es un vector $n$-dimensional de errores aleatorios. Los elementos de $\varepsilon$ se asume que tienen media cero, son incorrelados y tienen como varianza común $\sigma^2$.

Si se asume, además, que los componentes del vector de $\varepsilon$ se distribuyen de forma conjunta como una normal multivariante $N\left(\mathbf{0}, \sigma^2 I\right)$, donde $I$ es la matriz identidad de orden $(n \times n)$. Este modelo se denomina *modelo lineal normal.*

Así, la distribución condicionada de $\mathbf{y}$, dados los parámetros $\left(\beta, \sigma^2\right)$, es una distribución normal $N\left(\mathbf{X}\beta, \sigma^2 I\right)$,

La verosimilitud es

$$f\left(\mathbf{y}|\beta, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left(\mathbf{y} - \mathbf{X}\beta\right)^t\left(\mathbf{y} - \mathbf{X}\beta\right)\right].$$

Se puede escribir la forma cuadrática como

$$\left(\mathbf{y} - \mathbf{X}\beta\right)^t\left(\mathbf{y} - \mathbf{X}\beta\right) = \beta^t\mathbf{X}^t\mathbf{X}\beta - \beta^t\mathbf{X}^t\mathbf{y} - \mathbf{y}^t\mathbf{X}\beta + \mathbf{y}^t\mathbf{y}$$

si $\mathbf{X}'\mathbf{X}$ no es singular, entonces se puede reordenar la expresión anterior, después de hacer unas operaciones básicas, como

$$\left(\mathbf{y} - \mathbf{X}\beta\right)^t\left(\mathbf{y} - \mathbf{X}\beta\right) = \left(\beta - \hat{\beta}\right)^t\mathbf{X}^t\mathbf{X}\left(\beta - \hat{\beta}\right) + SCR$$

donde

$$\hat{\beta} = \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{y}$$

que es el estimador clásico por mínimos cuadrados (o de máxima verosimilitud) de $\beta$ y

$$SCR = \left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)^t\left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)$$

es la *suma de cuadrados residual*, que también se relaciona con el estimador de la varianza

$$\hat{\sigma}^2 = \frac{\left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)^t\left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)}{n - k} = \frac{SCR}{n - k}$$

donde $k$ es el número de parámetros que se estiman. Así,

$$\left(\mathbf{y} - \mathbf{X}\beta\right)^t\left(\mathbf{y} - \mathbf{X}\beta\right) = \left(\beta - \hat{\beta}\right)^t\mathbf{X}^t\mathbf{X}\left(\beta - \hat{\beta}\right) + (n - k)\hat{\sigma}^2$$

## Modelo bayesiano con distribuciones a priori no informativas

Se pueden definir distribuciones no informativas considerando independencia entre los parámetros,

$$\begin{aligned}\pi\left(\beta\right) &\propto 1\\ \pi\left(\sigma^2\right) &\propto \frac{1}{\sigma^2}=\sigma^{-2}\end{aligned}$$

De este modo, la distribución a posteriori conjunta queda como

$$\begin{aligned}&\pi\left(\beta,\sigma^2|\mathbf{y}\right)\propto\pi\left(\beta,\sigma^2\right)\cdot f\left(\mathbf{y}|\beta,\sigma^2\right)\propto\\ \propto\ &\left(\sigma^2\right)^{-\frac{n}{2}-1}\exp\left[-\frac{1}{2\sigma^2}\left(\left(n-k\right)\hat{\sigma}^2+\left(\beta-\hat{\beta}\right)^t\mathbf{X}^t\mathbf{X}\left(\beta-\hat{\beta}\right)\right)\right]\end{aligned}$$

### Distribución marginal de $\beta$

Primero se hace el siguiente cambio de variable:

$$\begin{aligned}s &= \sigma^{-2}\Longrightarrow\sigma^2=s^{-1}\\ J &= \left|\frac{\partial\sigma^2}{\partial s}\right|=\left|\frac{\partial}{\partial s}s^{-1}\right|=s^{-2}\end{aligned}$$

Así

$$\begin{aligned}\pi\left(\beta,s|\mathbf{y}\right) &\propto s^{\frac{n}{2}+1}\exp\left[-\frac{s}{2}\left(\left(n-k\right)\hat{\sigma}^2+\left(\beta-\hat{\beta}\right)^t\mathbf{X}^t\mathbf{X}\left(\beta-\hat{\beta}\right)\right)\right]\cdot s^{-2}\\ &\propto s^{\frac{n}{2}-1}\exp\left[-\frac{s}{2}\left(\left(n-k\right)\hat{\sigma}^2+\left(\beta-\hat{\beta}\right)^t\mathbf{X}^t\mathbf{X}\left(\beta-\hat{\beta}\right)\right)\right]\end{aligned}$$

La dsitribución marginal es, entonces,

$$\pi\left(\beta|\mathbf{y}\right)\propto\int_0^\infty s^{\frac{n}{2}-1}\exp\left[-\frac{s}{2}\left(\left(n-k\right)\hat{\sigma}^2+\left(\beta-\hat{\beta}\right)^t\mathbf{X}^t\mathbf{X}\left(\beta-\hat{\beta}\right)\right)\right]ds$$

Se observa que esta integral está relacionada con una función de densidad gamma, por lo que

$$\int_0^\infty s^p e^{-qs}ds=\frac{\Gamma(p+1)}{q^{p+1}}$$

e identificando

$$\begin{aligned}p &= \frac{n}{2}-1\\ q &= \frac{1}{2}\left(\left(n-k\right)\hat{\sigma}^2+\left(\beta-\hat{\beta}\right)^t\mathbf{X}^t\mathbf{X}\left(\beta-\hat{\beta}\right)\right)\end{aligned}$$

entonces

$$\pi\left(\beta|\mathbf{y}\right)\propto q^{-(p+1)}\propto q^{-\frac{n}{2}}$$

es decir,

$$\begin{aligned}\pi\left(\beta|\mathbf{y}\right) &\propto \left[(n-k)\,\hat{\sigma}^2+\left(\beta-\hat{\beta}\right)^t\mathbf{X}^t\mathbf{X}\left(\beta-\hat{\beta}\right)\right]^{-\frac{n}{2}}\\ &\propto \left[1+\frac{1}{n-k}\left(\beta-\hat{\beta}\right)^t\hat{\sigma}^{-2}\mathbf{X}^t\mathbf{X}\left(\beta-\hat{\beta}\right)\right]^{-\frac{n}{2}}\end{aligned}$$

lo cual es el *núcleo* de una distribución $t$ de *Student multivariante* tal que

$$\begin{aligned}E[\beta] &= \hat{\beta}\\ Var[\beta] &= \frac{(n-k)\,\hat{\sigma}^2\cdot\left(\mathbf{X}^t\mathbf{X}\right)^{-1}}{n-k-2}\end{aligned}$$

Mira, por ejemplo, en *Bayesian Data Analysis* de Gelman et al. (2003) pág. 577.

**NOTA:**

Un vector $d$-dimensional $\theta$ se distribuye como una $t$ de Student multivariante con $\nu>0$ grados de libertad, $\theta\sim t_\nu(\mu,\Sigma)$ donde el parámetro de localización es $\mu=(\mu_1,\ldots,\mu_d)'$ y el de escala es una matriz simétrica definida positiva de dimensión $(d\times d)$ denominada $\Sigma$, si su función de densidad es

$$p(\theta)=\frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\nu^{\frac{d}{2}}\pi^{\frac{d}{2}}}|\Sigma|^{-\frac{1}{2}}\left[1+\frac{1}{\nu}\left(\theta-\mu\right)^t\Sigma^{-1}\left(\theta-\mu\right)\right]^{-\frac{\nu+d}{2}},$$

de modo que

$$\begin{aligned}E(\theta)&=\mu, \qquad \text{para } \nu>1\\ Var(\theta)&=\frac{\nu}{\nu-2}\Sigma, \qquad \text{para } \nu>2\end{aligned}$$

### Distribución marginal de $\sigma^2$

En este caso, la integral es $\pi\left(\boldsymbol{\sigma^2}|\mathbf{y}\right)$, es decir,

$$\begin{aligned}\pi\left(\boldsymbol{\sigma^2}|\mathbf{y}\right) &\propto \int_{\mathbb{R}^k}\left(\sigma^2\right)^{-\frac{n}{2}-1}\exp\left[-\frac{1}{2\sigma^2}\left((n-k)\,\hat{\sigma}^2+\left(\beta-\hat{\beta}\right)^t\mathbf{X}^t\mathbf{X}\left(\beta-\hat{\beta}\right)\right)\right]d\beta\\ &\propto \left(\sigma^2\right)^{-\frac{n}{2}-1}\exp\left[-\frac{1}{2\sigma^2}\left(n-k\right)\hat{\sigma}^2\right]\\ &\times \int_{\mathbb{R}^k}\exp\left[-\frac{1}{2\sigma^2}\left(\beta-\hat{\beta}\right)^t\mathbf{X}^t\mathbf{X}\left(\beta-\hat{\beta}\right)\right]d\beta\end{aligned}$$

el segundo término es el *núcleo* de una distribución normal multivariante de dimensión $k$, por lo que

$$\begin{aligned}\pi\left(\boldsymbol{\sigma^2}|\mathbf{y}\right) &\propto \left(\sigma^2\right)^{-\frac{n}{2}-1}\exp\left[-\frac{1}{2\sigma^2}\left(n-k\right)\hat{\sigma}^2\right]\left(2\pi\sigma^2\right)^{\frac{k}{2}}\\ &\propto \left(\sigma^2\right)^{-\frac{1}{2}(n-k)-1}\exp\left[-\frac{1}{2\sigma^2}\left(n-k\right)\hat{\sigma}^2\right]\end{aligned}$$

que es una distribución gamma inversa de parámetros

$$\begin{aligned}\alpha &= \frac{1}{2}(n-k)\\ \beta &= \frac{1}{2}(n-k)\,\hat{\sigma}^2\end{aligned}$$

Se obtiene, entonces el mismo resultado que en el caso del modelo *normal-normal* visto en el tema previo.

## Modelo con distribuciones a priori conjugadas

Ahora consideramos distribuciones a priori informativas para $\beta$ y $\sigma^2$. En este caso es mejor considerar las distribuciones conjugadas como el producto de dos términos:

$$P\left(\beta,\sigma^2\right) = P\left(\beta|\sigma^2\right)\cdot P\left(\sigma^2\right)$$

donde ambos componentes a priori se se especifican mediante

$$\begin{aligned}P\left(\beta|\sigma^2\right) &= (2\pi)^{-\frac{k}{2}}\,|\Sigma|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\beta-\mathbb{B}\right)^t\Sigma^{-1}\left(\beta-\mathbb{B}\right)\right]\\ &= \left(2\pi\sigma^2\right)^{-\frac{k}{2}}\exp\left[-\frac{1}{2\sigma^2}\left(\beta-\mathbb{B}\right)^t\left(\beta-\mathbb{B}\right)\right]\\ P\left(\sigma^2\right) &\propto (\sigma^2)^{-\frac{1}{2}(a-k)}\exp\left[-\frac{B}{\sigma^2}\right]\end{aligned}$$

En el caso habitual $\Sigma=\sigma^2\mathbf{I}$

Así, la distribución de $\beta$ condicionada por $\sigma^2$ es una normal multivariante de media $\mathbb{B}$ y varianza $\sigma^2$.

La distribución a priori para $\sigma^2$ es una gamma inversa de parámetros $\left(\frac{1}{2}(a-k)+1\right)$ y $B$. La distribución gamma inversa de $\sigma^2$ no solo es conjugada sino que es la distribución a posteriori marginal para el modelo con distribución a priori no informativa.

Multiplicando la verosimilitud por la distribución a priori se tiene

$$\begin{aligned}&\pi\left(\beta,\sigma^2|\mathbf{X},y\right)\propto\\ &(\sigma^2)^{-\frac{n}{2}}\exp\left[-\frac{1}{2\sigma^2}\left((n-k)\,\hat{\sigma}^2+\left(\beta-\hat{\beta}\right)^t\mathbf{X}^t\mathbf{X}\left(\beta-\hat{\beta}\right)\right)\right]\times\\ &\times\left(\sigma^2\right)^{-\frac{k}{2}}\exp\left[-\frac{1}{2\sigma^2}\left(\beta-\mathbb{B}\right)^t\left(\beta-\mathbb{B}\right)\right](\sigma^2)^{-\frac{1}{2}(a-k)}\exp\left[-\frac{B}{\sigma^2}\right]\\ \propto\;&(\sigma^2)^{-\frac{1}{2}(n+a)}\exp\left[-\frac{1}{2\sigma^2}\left((n-k)\,\hat{\sigma}^2+\left(\beta-\hat{\beta}\right)^t\mathbf{X}^t\mathbf{X}\left(\beta-\hat{\beta}\right)\right)\right.\\ &\left.+2B+\left(\beta-\mathbb{B}\right)^t\left(\beta-\mathbb{B}\right)\right]\end{aligned}$$

Si se denomina

$$\begin{aligned}\widetilde{\beta} &= \left(\mathbf{I}+\mathbf{X}^t\mathbf{X}\right)^{-1}\left(\mathbb{B}+\mathbf{X}^t\mathbf{X}\hat{\beta}\right)\\ \widetilde{s} &= 2B+(n-k)\,\hat{\sigma}^2+\left(\mathbb{B}-\widetilde{\beta}\right)^t\mathbb{B}+\left(\hat{\beta}-\widetilde{\beta}\right)\mathbf{X}^t\mathbf{X}\hat{\beta}\end{aligned}$$

entonces se puede reexpresar la distribución a posteriori como

$$\begin{array}{l}\pi\left(\beta,\sigma^2|\mathbf{X},y\right)\propto\\ \left(\sigma^2\right)^{-\frac{1}{2}(n+a)}\exp\left[-\frac{1}{2\sigma^2}\left(\widetilde{s}+\left(\beta-\widetilde{\beta}\right)^t\left(\mathbf{I}+\mathbf{X}^t\mathbf{X}\right)\left(\beta-\widetilde{\beta}\right)\right)\right]\end{array}$$

Así tenemos la ventaja de que se puede calcular de una manera sencilla la distribución marginal de $\beta$ :

$$\pi\left(\beta|\mathbf{X},y\right)\propto\left[\widetilde{s}+\left(\beta-\widetilde{\beta}\right)^t\left(\mathbf{I}+\mathbf{X}^t\mathbf{X}\right)\left(\beta-\widetilde{\beta}\right)\right]^{-\frac{n+a}{2}+1}$$

que es una distribución $t$ de Student multivariante con $(n+a-k-2)$ grados de libertad, cuya media y matriz de covarianzas son

$$\begin{aligned}E\left(\beta|\mathbf{X},y\right)&=\widetilde{\beta}\\ \operatorname{Cov}\left(\beta|\mathbf{X},y\right)&=\frac{\widetilde{s}\left(\mathbf{I}+\mathbf{X}^t\mathbf{X}\right)^{-1}}{n+a-k-3}\end{aligned}$$

La distribución marginal de $\sigma^2$ es también similar a la obtenida usando una a priori no informativa

$$\pi\left(\sigma^2|\mathbf{X},y\right)\propto\left(\sigma^2\right)^{-\frac{1}{2}(n+a-k+1)}\exp\left[-\frac{1}{2\sigma^2}\hat{\sigma}^2\left(n+a-k\right)\right]$$

es decir, se obtiene una distribución gamma inversa.

## Distribución Predictiva

En los modelos de regresión es importante predecir los valores de una variable respuesta para futuras observaciones. Consideramos ahora la predicción de un modelo de regesión desde el punto de vista bayesiano. Supongamos que queremos aplicar el análisis de regresión en un nuevo conjunto de datos donde se ha observado una matriz de covariables (de diseño) $\widetilde{X}$ y se quiere predecir la variable dependiente $\tilde{y}$.

Si $\beta$ y $\sigma^2$ son conocidas entonces $\tilde{y}$ se distribuye como una normal $N\left(\widetilde{X}\beta,\sigma^2\mathbf{I}\right)$. Sin embargo estos parámetros no están fijados, de modo que se tiene que usar la información recogida por la distribución a posteriori de ellos. Por tanto se calcula la distribución predictiva a posteriori $P\left(\tilde{y}|y\right)$.

Hay dos fuentes de incertidumbre en la distribución predictiva:

1. La fuente de variabilidad debida a la varianza $\sigma^2$ que no está recogida por $\widetilde{X}\beta$.

2. La incertidumbre a posteriori de los parámetros como resultado de su estimación a partir de una muestra finita.

Cuando el tamaño muestral $n \to \infty$ la varianza debido a la incertidumbre a posteriori desaparece pero la variabilidad debida a la varianza original $\sigma^2$ permanece presente.

La distribución predictiva a posteriori es

$$P\left(\tilde{y}|y\right) = \int P\left(\tilde{y}|\beta, \sigma^2\right) P\left(\beta, \sigma^2|y\right) d\beta d\sigma^2.$$

Desde el punto de vista de la simulación es fácil muestrear de esta expresión. Para cada valor simulado de la distribución a posteriori de $(\beta, \sigma^2)$ se debe obtener un valor $\tilde{y}$ de $N\left(\widetilde{X}\beta, \sigma^2\mathbf{I}\right)$. La media y varianza de esta muestra son la media predictiva y varianza predictiva respectivamente.

Desde el punto de vista de los métodos de simulación es fácil muestrear de esta expresión. Se simulan muestras de la distribución a posteriori de $(\beta, \sigma^2)$ y para cada par de valores se obtiene otro valor simulado de $\tilde{y}$ a partir de $N\left(\widetilde{X}\beta, \sigma^2\mathbf{I}\right)$.

Entonces, la media y varianza de esta muestra son la media predictiva y varianza predictiva respectivamente.

Se pueden considerar también expresiones analíticas para la esperanza y la varianza de la distribución predictiva a posteriori. Para simplificar, se pueden calcular condicionando para $\sigma^2$, lo que equivale a asumir que $\sigma^2$ es conocido:

$$\begin{aligned}
E_{\tilde{y}}\left(\tilde{y}|\sigma^2, y\right) &= E_\beta\left[E_{\tilde{y}}\left(\tilde{y}|\beta, \sigma^2, y\right)|\sigma^2, y\right] \\
&= E_\beta\left[\widetilde{X}\beta|\sigma^2, y\right] \\
&= \widetilde{X}\hat{\beta}
\end{aligned}$$

Del mismo modo,

$$\begin{aligned}
\mathrm{Var}_{\tilde{y}}\left(\tilde{y}|\sigma^2, y\right) &= E_\beta\left[\mathrm{Var}_{\tilde{y}}\left(\tilde{y}|\beta, \sigma^2, y\right)|\sigma^2, y\right] + \mathrm{Var}_\beta\left[E_{\tilde{y}}\left(\tilde{y}|\beta, \sigma^2, y\right)|\sigma^2, y\right] \\
&= E_\beta\left[\sigma^2\mathbf{I}|\sigma^2, y\right] + \mathrm{Var}_\beta\left[\widetilde{X}\beta|\sigma^2, y\right] \\
&= \left(\mathbf{I} + \widetilde{X}(X^tX)^{-1}\widetilde{X}^t\right)\sigma^2
\end{aligned}$$

Así, condicionalmente a $\sigma^2$ la varianza a posteriori predictiva tiene dos componentes $\sigma^2\mathbf{I}$, que representa la variación muestral, y $\widetilde{X}(X^tX)^{-1}\widetilde{X}^t\sigma^2$ debido a la incertidumbre con respecto a $\beta$.

Para obtener la distribución predictiva marginal $P\left(\tilde{y}|y\right)$ se tiene que integrar sobre la distribución a posteriori marginal de $\sigma^2$. Se obtiene como resultado una distribución $t$ de Student multivariante centrada en $\widetilde{X}\hat{\beta}$ y matriz de escala $\left(\mathbf{I} + \widetilde{X}(X^tX)^{-1}\widetilde{X}^t\right)s^2$ con $n-k$ grados de libertad.

En la práctica rara vez se restringe uno a usar distribuciones no informativas como se ha hecho anteriormente y se usan distribuciones a priori más flexibles. En este caso es preferible usar métodos de simulación *MCMC*.

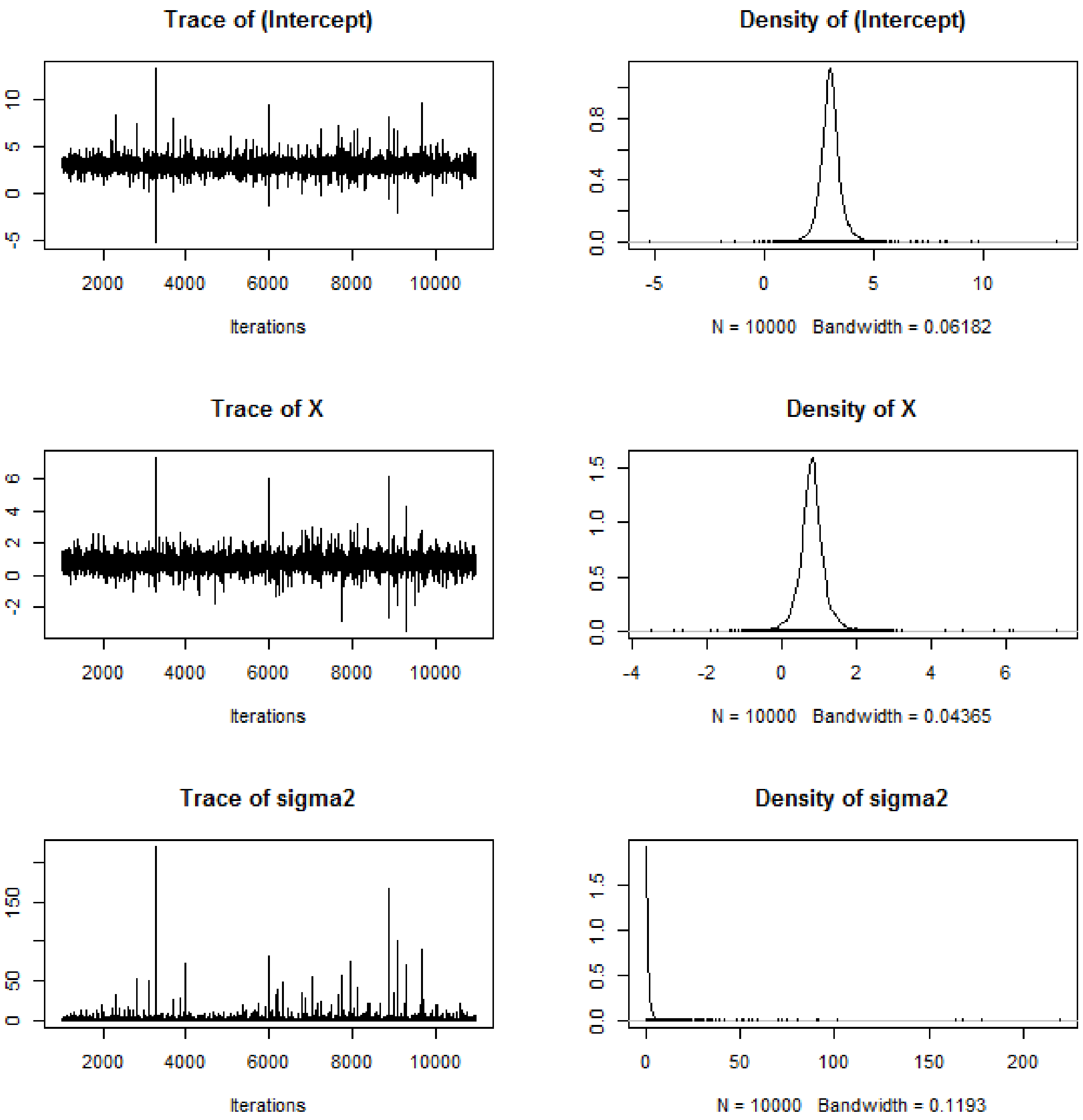
## Programación con R y con Bugs

### Programa de Regresión con la librería MCMCpack de R

Consideramos distribuciones a priori no informativas.

```r
library(MCMCpack)
help(MCMCregress)
cosa = list(X=c(-2,-1,0,1,2), Y=c(1,3,3,3,5))

# Se asume una distribucion no informativa
posterior = MCMCregress(Y~X, data=cosa)
summary(posterior)
plot(posterior)

# Comparamos con el modelo de regresion clasico
clasico = lm(Y~X, data=cosa)
summary(clasico)
```

```
Iterations = 1001:11000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

             Mean     SD Naive SE Time-series SE
(Intercept) 3.013 0.5459 0.005459       0.005439
X           0.807 0.3873 0.003873       0.003954
sigma2      1.510 5.0969 0.050969       0.091333

2. Quantiles for each variable:

               2.5%    25%    50%    75% 97.5%
(Intercept) 2.04394 2.7552 3.0043 3.2483 4.083
X           0.05681 0.6287 0.8047 0.9769 1.562
sigma2      0.17091 0.3876 0.6837 1.3395 7.155
```




En el caso de la regresión clásica

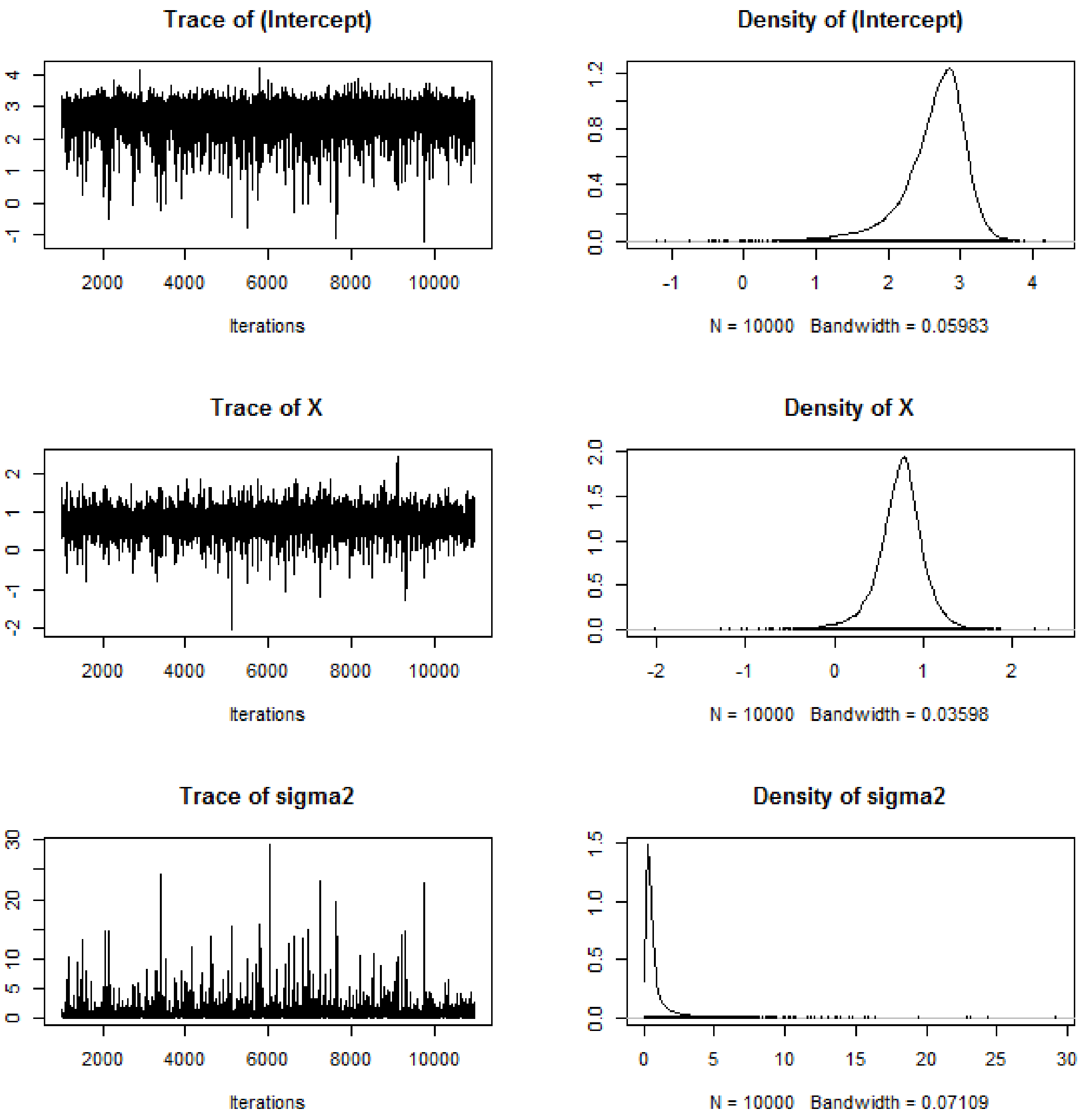
```
Call:
lm(formula = Y ~ X, data = cosa)

Residuals:
         1          2          3          4          5
-4.000e-01  8.000e-01 -1.249e-16 -8.000e-01  4.000e-01

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.0000     0.3266   9.186  0.00273 **
X             0.8000     0.2309   3.464  0.04052 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7303 on 3 degrees of freedom
Multiple R-squared:    0.8,     Adjusted R-squared:  0.7333
F-statistic:    12 on 1 and 3 DF,  p-value: 0.04052
```

Consideramos distribuciones a priori *informativas*. La distribución a priori de $\beta$ se especifica usando los argumentos `b0` and `B0`, donde `b0` es la media a priori y `B0` es la precisión a priori de $\beta$

La distribución a priori de $\sigma^2$ está determinada por dos parámetros `c0` and `d0`. En este caso se usa una distribución a priori débilmente informativa (`c0`=2 y `d0`=0.1)

```
library(MCMCpack)
help(MCMCregress)
cosa = list(X=c(-2,-1,0,1,2), Y=c(1,3,3,3,5))

# Se asume una distribucion informativa
posterior = MCMCregress(Y~X, b0=0, B0=1, c0=2, d0=0.1, data=cosa)
summary(posterior)
plot(posterior)
```

```
Iterations = 1001:11000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

              Mean     SD Naive SE Time-series SE
(Intercept) 2.6414 0.4611 0.004611       0.007920
X           0.7460 0.2726 0.002726       0.002996
sigma2      0.8258 1.2698 0.012698       0.022917

2. Quantiles for each variable:

              2.5%    25%    50%    75% 97.5%
(Intercept) 1.4495 2.4519 2.7255 2.9291 3.293
X           0.1392 0.6113 0.7635 0.8983 1.248
sigma2      0.1442 0.2900 0.4708 0.8570 3.906
```

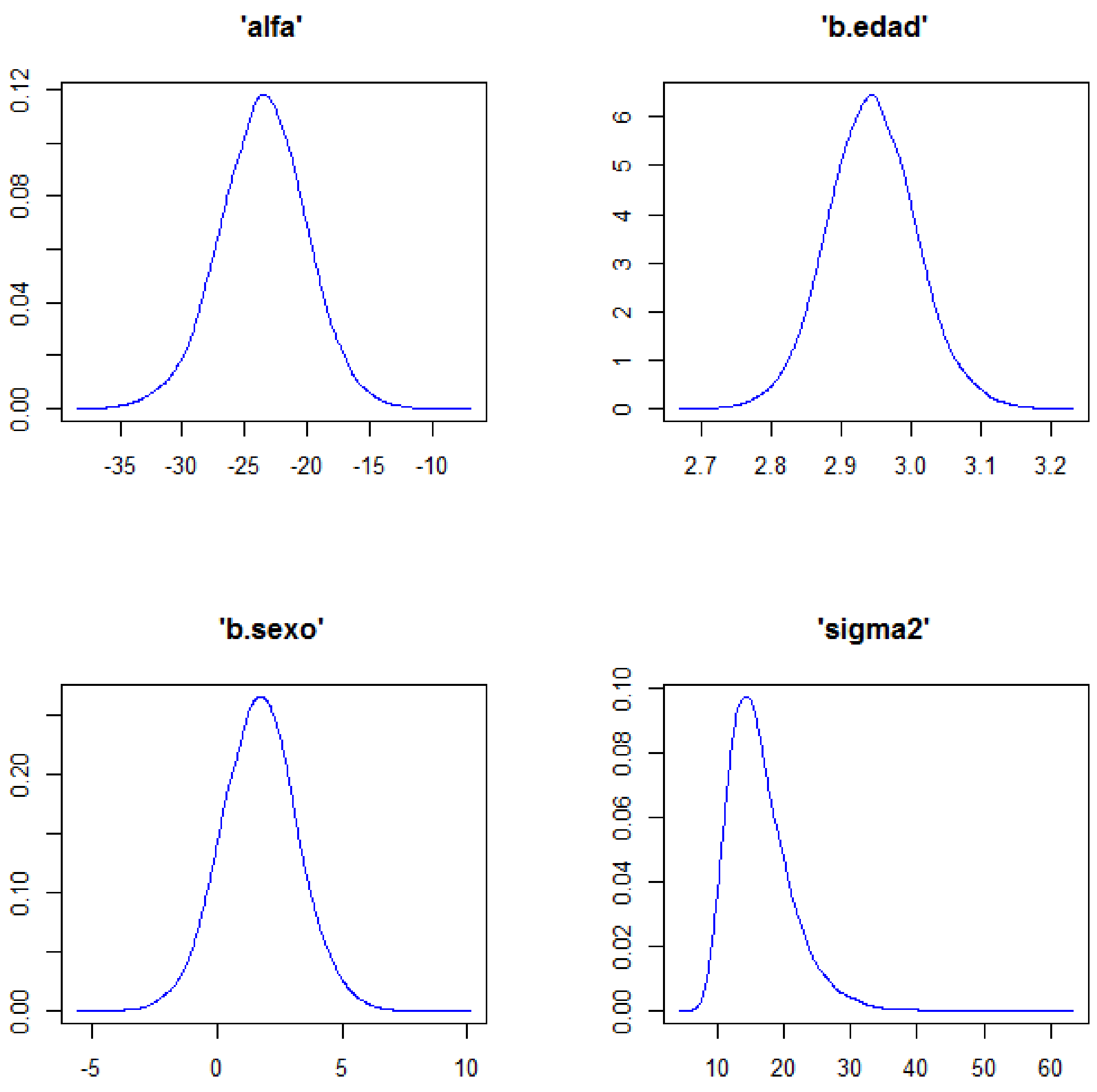

Trace of (Intercept)
Iterations
Density of (Intercept)
N = 10000 Bandwidth = 0.05983
Trace of X
Iterations
Density of X
N = 10000 Bandwidth = 0.03598
Trace of sigma2
Iterations
Density of sigma2
N = 10000 Bandwidth = 0.07109

### Programa de Regresión con Openbugs y BRugs

```
rm(list=ls(all=TRUE))
ruta = "C:/DondeSea"
setwd(ruta)

#.................................................................
# PROGRAMA de BUGS
#.................................................................
cat("
# Regresion Lineal
# Se considera el parametro de precision tao = 1/varianza
model
{
for (i in 1:n) {
mu[i] <- alfa + b.sexo*sexo[i] + b.edad*edad[i]
bp[i] ~ dnorm(mu[i],tao)
}

alfa ~ dnorm(0.0,1.0E-4)
b.sexo ~ dnorm(0.0,1.0E-4)
b.edad ~ dnorm(0.0,1.0E-4)
tao ~ dgamma(1.0E-3,1.0E-3)
sigma2 <- 1/tao
}
",
file="modelo.txt")
#.................................................................

library(BRugs)

# Introduzco los datos
sexo = c(0,1,0,0,0,1,0,1,1,0,1,0,0,0,0,1,0,1,0,0,1,0,1,0,0,
0,1,1,1,1)
edad = c(59,52,37,40,67,43,61,34,51,58,54,31,49,45,66,48,41,
47,53,62,60,33,44,70,56,69,35,36,68,38)
bp = c(143,132,88,98,177,102,154,83,131,150,131,69,111,114,
170,117,96,116,131,158,156,75,111,184,141,182,74,87,183,89)
n = length(sexo)
datos = list("sexo","edad","bp","n")
parametros = c("alfa","b.sexo","b.edad","sigma2")
iniciales = function() {list(alfa=rnorm(1), b.sexo=rnorm(1),
b.edad=rnorm(1), tao=rgamma(1,1,1))}

# Con BRugs
modelo = BRugsFit(data=datos, inits=iniciales, para=parametros,
nBurnin = 10000, nIter = 10000, modelFile="modelo.txt",
numChains=3,working.directory=ruta)
samplesStats("*") # Resultados finales
samplesDensity("*", mfrow=c(2,2), col=4)
```

```r
# Alternativamente con OpenBugs
library(R2OpenBUGS)
modelo = bugs(data=datos,inits=iniciales,
parameters.to.save=parametros,
model.file="modelo.txt",
n.chains=3, n.iter=20000, n.burnin=10000,
working.directory=ruta, clearWD=TRUE, debug=TRUE)

print(modelo)
```

```
          mean      sd  MC_error  val2.5pc  median  val97.5pc
alfa    -23.550 3.50900 0.113000   -30.680 -23.500    -16.750
b.edad    2.944 0.06385 0.002047     2.821   2.944      3.075
b.sexo    1.692 1.53700 0.020870    -1.339   1.697      4.749
sigma2   16.400 4.85800 0.037810     9.488  15.540     28.220
```



Se puede considerar la distribución predictiva de un modelo de regresión lineal de una manera sencilla, usando la librería `LearnBayes`:

```
library(LearnBayes)

gorjeos = c(20,16.0,19.8,18.4,17.1,15.5,14.7,17.1,15.4,16.2,15,
17.2, 16,17,14.1)

tiempo = c(88.6,71.6,93.3,84.3,80.6,75.2,69.7,82,69.4,83.3,78.6
,82.6,80.6, 83.5,76.3)

X = cbind(1,gorjeos)
m = 1000
teta.muestra = blinreg(tiempo, X, m)

nuevo1 = c(1,15)
nuevo2 = c(1,20)
X1 = rbind(nuevo1, nuevo2)

cadena = blinregpred(X1, teta.muestra)

# Medias predictivas
apply(cadena,2,mean)

# Desviacion estandar de las cadenas
apply(cadena,2,sd)

# Intervalo de probabilidad al 95%
apply(cadena,2,function(x)quantile(x,c(0.025,0.975)))
```

```
# Medias predictivas
[1] 74.51377 91.04300

# Desviacion estandar de las cadenas
[1] 4.313695 4.646278

# Intervalo de probabilidad al 95%
          [,1]     [,2]
2.5%  65.40498 81.78671
97.5% 82.48410 99.87782
```

Alternativamente se puede usar `BRugs`:

```
rm(list=ls(all=TRUE))
ruta = "C:/DondeSea"
setwd(ruta)

#.................................................................
# PROGRAMA de BUGS
#.................................................................
cat("
# Regresion Lineal con predicciones
# Se considera el parametro de precision tao = 1/varianza
model
{
for (i in 1:n) {
mu[i] <- alfa + b.sexo*sexo[i] + b.edad*edad[i]
bp[i] ~ dnorm(mu[i],tao)
}

# Valores a predecir
for (i in 1:3) {
mup[i] <- alfa + b.sexo*sexop[i] + b.edad*edadp[i]
bpp[i] ~ dnorm(mup[i],tao)
}

alfa ~ dnorm(0.0,1.0E-4)
b.sexo ~ dnorm(0.0,1.0E-4)
b.edad ~ dnorm(0.0,1.0E-4)
tao ~ dgamma(1.0E-3,1.0E-3)
sigma2 <- 1/tao
}
",
file="modelo.txt")
#.................................................................

library(BRugs)

# Introduzco los datos
sexo = c(0,1,0,0,0,1,0,1,1,0,1,0,0,0,0,1,0,1,0,0,1,0,1,0,0,
0,1,1,1,1)

edad = c(59,52,37,40,67,43,61,34,51,58,54,31,49,45,66,48,41,
47,53,62,60,33,44,70,56,69,35,36,68,38)

bp = c(143,132,88,98,177,102,154,83,131,150,131,69,111,114,
170,117,96,116,131,158,156,75,111,184,141,182,74,87,183,89)
```

```
# Covariables para prediccion
sexop = c(1,0,0)
edadp = c(75,78,80)

n = length(sexo)
datos = list("sexo","edad","bp","n","sexop","edadp")
parametros = c("alfa","b.sexo","b.edad","sigma2","bpp")
iniciales = function() {list(alfa=rnorm(1), b.sexo=rnorm(1),
b.edad=rnorm(1), tao=rgamma(1,1,1))}

# Con BRugs
modelo = BRugsFit(data=datos, inits=iniciales, para=parametros,
nBurnin = 10000, nIter = 10000, modelFile="modelo.txt",
numChains=3,working.directory=ruta)
samplesStats("*") # Resultados finales

# Alternativamente con OpenBugs
library(R2OpenBUGS)
modelo = bugs(data=datos,inits=iniciales,
parameters.to.save=parametros,
model.file="modelo.txt",
n.chains=3, n.iter=20000, n.burnin=10000,
working.directory=ruta, clearWD=TRUE, debug=TRUE)
```

```
            mean     sd MC_error val2.5pc  median val97.5pc
alfa     -23.450 3.4070 0.113900  -30.160 -23.410   -16.670
b.edad     2.943 0.0621 0.002055    2.820   2.942     3.066
b.sexo     1.687 1.5360 0.021430   -1.314   1.681     4.732
bpp[1]   198.900 4.5540 0.061280  189.900 198.900   208.000
bpp[2]   206.100 4.4640 0.052570  197.300 206.100   214.900
bpp[3]   211.900 4.4860 0.054270  203.100 212.000   220.800
sigma2    16.320 4.7790 0.037650    9.457  15.490    27.860
```

Alternativamente se puede usar SAS:

```
OPTIONS nodate;
TITLE 'Regresion Linear Simple';
ODS rtf file='/folders/myfolders/cosa.rtf' style=minimal
    startpage=no;

DATA Clase;
INPUT nombre $ altura peso @@;

/* Datos Originales */
DATALINES;
Alfredo 69.0 112.5   Alicia   56.5  84.0   Barbara 65.3  98.0
Carol   62.8 102.5   Enrique 63.5 102.5   Jaime   57.3  83.0
Lisa    59.8  84.5   Juana   62.5 112.5   Javier  62.5  84.0
Juan    59.0  99.5   manuel  51.3  50.5   Sara    64.3  90.0
Luisa   56.3  77.0   Maria   66.5 112.0   Felipe  72.0 150.0
Roberto 64.8 128.0   Ronaldo 67.0 133.0   Tomas   57.5  85.0
Guillermo 66.5 112.0
;

/* Datos para predecir */
DATA ClasePred;
INPUT nombre $ altura peso @@;

DATALINES;
Toni 79.0 82.5   Ali  66.5  74.0 Vanessa 55.3  88.0
Epi  69.8 105.5  Blas 68.5 102.5 Luis    47.3  89.0
;
```

```
PROC mcmc data=Clase outpost=sale nmc=50000 thin=5;

   /* Valores iniciales de los parametros */
   parms beta0 0 beta1 0;
   parms sigma2 1;

   /* Distribuciones a priori */
   prior beta0 beta1 ~ normal(mean=0, var=1e6);
   prior sigma2 ~ igamma(shape=3/10, scale=10/3);

   /* Modelo */
   mu = beta0 + beta1*altura;
   model peso ~ normal(mean=mu, var=sigma2);

   /* Predicciones */
   preddist outpred=salePredi covariates=ClasePred;

RUN;
ODS rtf close;
```

# Introducción a los Modelos Lineales Generalizados

Se pueden generalizar los modelos de regresión habituales para modelizar relaciones entre variables discretas y también para modelos de clasificación.

En este caso se trata simplemente de definir una función *link* que relacione la media de los datos y una combinación lineal.

Si se asume que

$$E[Y] = \mu$$

entonces, se puede definir una función *link* $g$ aplicada al parámetro $\mu$ tal que

$$g(\mu) = X\beta$$

es decir,

$$\mu = E[Y] = g^{-1}(X\beta)$$

Como ejemplos de modelos y funciones *link* se tiene

| | | **Link** | **Inversa link** |
|---|---|---|---|
| **Modelo regresión** | **Función** | $g(\mu)$ | $\mu = g^{-1}(X\beta)$ |
| Normal | Identidad | $\mu$ | $X\beta$ |
| Logístico | Logit | $\log\left(\frac{\mu}{1-\mu}\right)$ | $\frac{\exp(X\beta)}{1+\exp(X\beta)}$ |
| Probit | Normal inversa | $\Phi^{-1}(\mu)$ | $\Phi(X\beta)$ |
| Poisson | logaritmo | $\log(\mu)$ | $\exp(X\beta)$ |
| Gamma | Inversa | $-\frac{1}{\mu}$ | $-\frac{1}{X\beta}$ |
| Binomial negativo | Logaritmo | $\log(1-\mu)$ | $1-\exp(X\beta)$ |

**Ejemplo**

Consideramos un ejemplo de regresión logística. Supongamos una muestra de 100 personas en las que se considera si presentan o no un tipo de enfermedad degenerativa (se recoge en la variable *tiene*).

| tiene | sexo | edad | tiene | sexo | edad | tiene | sexo | edad | tiene | sexo | edad |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 69 | 1 | 1 | 76 | 1 | 0 | 72 | 0 | 0 | 51 |
| 1 | 1 | 57 | 1 | 1 | 72 | 1 | 1 | 71 | 1 | 1 | 76 |
| 1 | 1 | 61 | 0 | 0 | 53 | 0 | 0 | 54 | 1 | 1 | 75 |
| 0 | 0 | 60 | 1 | 1 | 69 | 0 | 1 | 52 | 0 | 0 | 66 |
| 1 | 1 | 69 | 0 | 1 | 59 | 0 | 1 | 54 | 1 | 0 | 75 |
| 1 | 1 | 74 | 1 | 1 | 73 | 0 | 1 | 50 | 1 | 0 | 78 |
| 0 | 0 | 63 | 1 | 0 | 77 | 1 | 0 | 75 | 1 | 0 | 70 |
| 1 | 0 | 68 | 0 | 0 | 55 | 1 | 0 | 59 | 0 | 1 | 67 |
| 1 | 0 | 64 | 1 | 0 | 77 | 1 | 1 | 65 | 0 | 0 | 51 |
| 0 | 0 | 53 | 1 | 1 | 68 | 1 | 1 | 60 | 1 | 0 | 70 |
| 1 | 1 | 60 | 1 | 1 | 62 | 1 | 1 | 60 | 1 | 1 | 71 |
| 0 | 1 | 58 | 0 | 0 | 56 | 0 | 1 | 57 | 1 | 1 | 71 |
| 1 | 1 | 79 | 1 | 0 | 68 | 0 | 0 | 51 | 1 | 1 | 74 |
| 1 | 1 | 56 | 1 | 1 | 70 | 0 | 0 | 51 | 1 | 0 | 74 |
| 0 | 1 | 53 | 0 | 1 | 60 | 1 | 0 | 63 | 1 | 1 | 60 |
| 1 | 0 | 74 | 1 | 0 | 65 | 0 | 1 | 57 | 1 | 0 | 58 |
| 0 | 1 | 56 | 0 | 1 | 55 | 1 | 1 | 80 | 0 | 1 | 55 |
| 1 | 0 | 76 | 0 | 0 | 64 | 0 | 1 | 52 | 1 | 1 | 61 |
| 1 | 0 | 72 | 1 | 0 | 75 | 0 | 0 | 65 | 1 | 1 | 65 |
| 0 | 0 | 56 | 0 | 0 | 60 | 1 | 0 | 72 | 0 | 0 | 52 |
| 1 | 1 | 66 | 1 | 1 | 67 | 1 | 1 | 80 | 0 | 1 | 68 |
| 0 | 1 | 52 | 0 | 0 | 61 | 1 | 1 | 73 | 1 | 1 | 75 |
| 1 | 1 | 77 | 1 | 0 | 69 | 1 | 1 | 76 | 0 | 1 | 52 |
| 1 | 0 | 70 | 1 | 0 | 75 | 1 | 0 | 79 | 0 | 1 | 53 |
| 1 | 0 | 69 | 1 | 1 | 68 | 0 | 0 | 66 | 1 | 1 | 70 |

Se trata de estudiar la relación entre la presencia de esta enfermedad, la edad y el sexo de las personas. Para ello se puede aplicar una regresión logística.

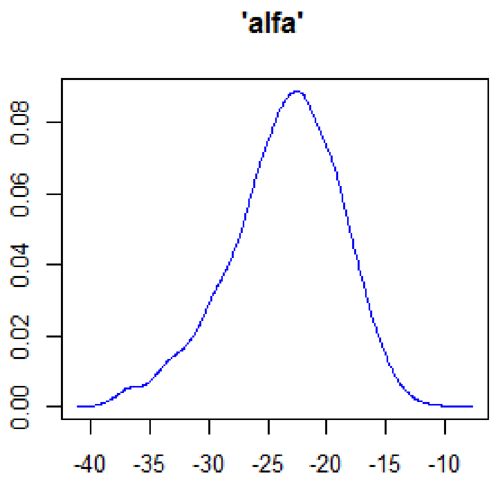
Ejemplo con `OpenBugs` o con `BRugs`:

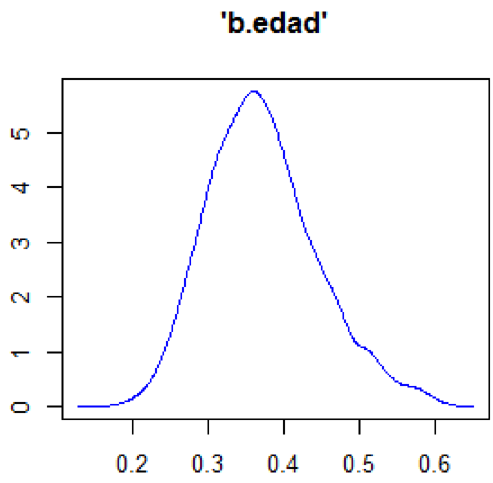
```
rm(list=ls(all=TRUE))
ruta = "C:/DondeSea"
setwd(ruta)

#..........................................................
# PROGRAMA de BUGS
#..........................................................

cat("
# Regresion Logistica con OpenBugs
model
{
    for (i in 1:n) {
          logit(p[i]) <- alfa + b.sexo*sexo[i] + b.edad*edad[i]
          tiene[i] ~ dbern(p[i])
    }
    alfa ~ dnorm(0.0, 1.0E-4)
    b.sexo ~ dnorm(0.0, 1.0E-4)
    b.edad ~ dnorm(0.0, 1.0E-4)
}
",
file="modelo.txt")
#.........................................................

library(BRugs)

# Datos
sexo = c(1,1,1,0,1,1,0,0,0,0,1,1,1,1,1,0,1,0,0,0,
1,1,1,0,0,1,1,0,1,1,1,0,0,0,1,1,0,0,1,1,0,1,0,0,
0,1,0,0,0,1,0,1,0,1,1,1,0,0,1,1,1,1,0,0,0,1,1,1,
0,0,1,1,1,0,0,0,1,1,0,0,0,0,1,0,0,1,1,1,0,1,0,1,
1,1,0,1,1,1,1,1)

edad = c(69,57,61,60,69,74,63,68,64,53,60,58,79,56,53,
74,56,76,72,56,66,52,77,70,69,76,72,53,69,59,73,77,55,
77,68,62,56,68,70,60,65,55,64,75,60,67,61,69,75,68,72,
71,54,52,54,50,75,59,65,60,60,57,51,51,63,57,80,52,65,
72,80,73,76,79,66,51,76,75,66,75,78,70,67,51,70,71,71,
74,74,60,58,55,61,65,52,68,75,52,53,70)
```

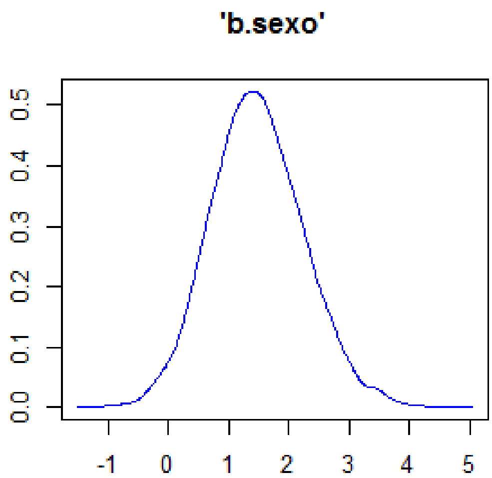
```
tiene = c(1,1,1,0,1,1,0,1,1,0,1,0,1,1,0,1,0,1,1,
0,1,0,1,1,1,1,1,0,1,0,1,1,0,1,1,1,0,1,1,0,1,0,0,
1,0,1,0,1,1,1,1,1,0,0,0,0,1,1,1,1,1,0,0,0,1,0,1,
0,0,1,1,1,1,1,0,0,1,1,0,1,1,1,0,0,1,1,1,1,1,1,1,
0,1,1,0,0,1,0,0,1)

n = length(sexo)

datos = list("sexo","edad","tiene","n")
parametros = c("alfa","b.sexo","b.edad")
```

```
iniciales = function(){list(alfa=rnorm(1),
b.sexo=rnorm(1), b.edad=rnorm(1))}

# Con BRugs
modelo = BRugsFit(data=datos, inits=iniciales,
para=parametros, nBurnin=10000, nIter=10000,
modelFile="modelo.txt",
numChains=3, working.directory=ruta)
samplesStats("*")           # Resultados finales
samplesDensity("*", mfrow=c(2,2), col=4)

# Con OpenBugs
library(R2OpenBUGS)
modelo = bugs(data=datos,inits=iniciales,
parameters.to.save=parametros,
model.file="modelo.txt",
n.chains=3, n.iter=20000, n.burnin=10000,
working.directory=ruta, clearWD=TRUE, debug=TRUE)

print(modelo)
```

```
            mean      sd  MC_error   val2.5pc    median  val97.5pc start  sample
alfa    -23.5500 4.73300 0.110600  -34.05000  -23.1200   -15.3800 10001   30000
b.edad    0.3714 0.07339 0.001717    0.24450    0.3651     0.5356 10001   30000
b.sexo    1.4780 0.77190 0.012420    0.02422    1.4510     3.0520 10001   30000
```


'alfa'
0.00
0.02
0.04
0.06
0.08
-40
-35
-30
-25
-20
-15
-10

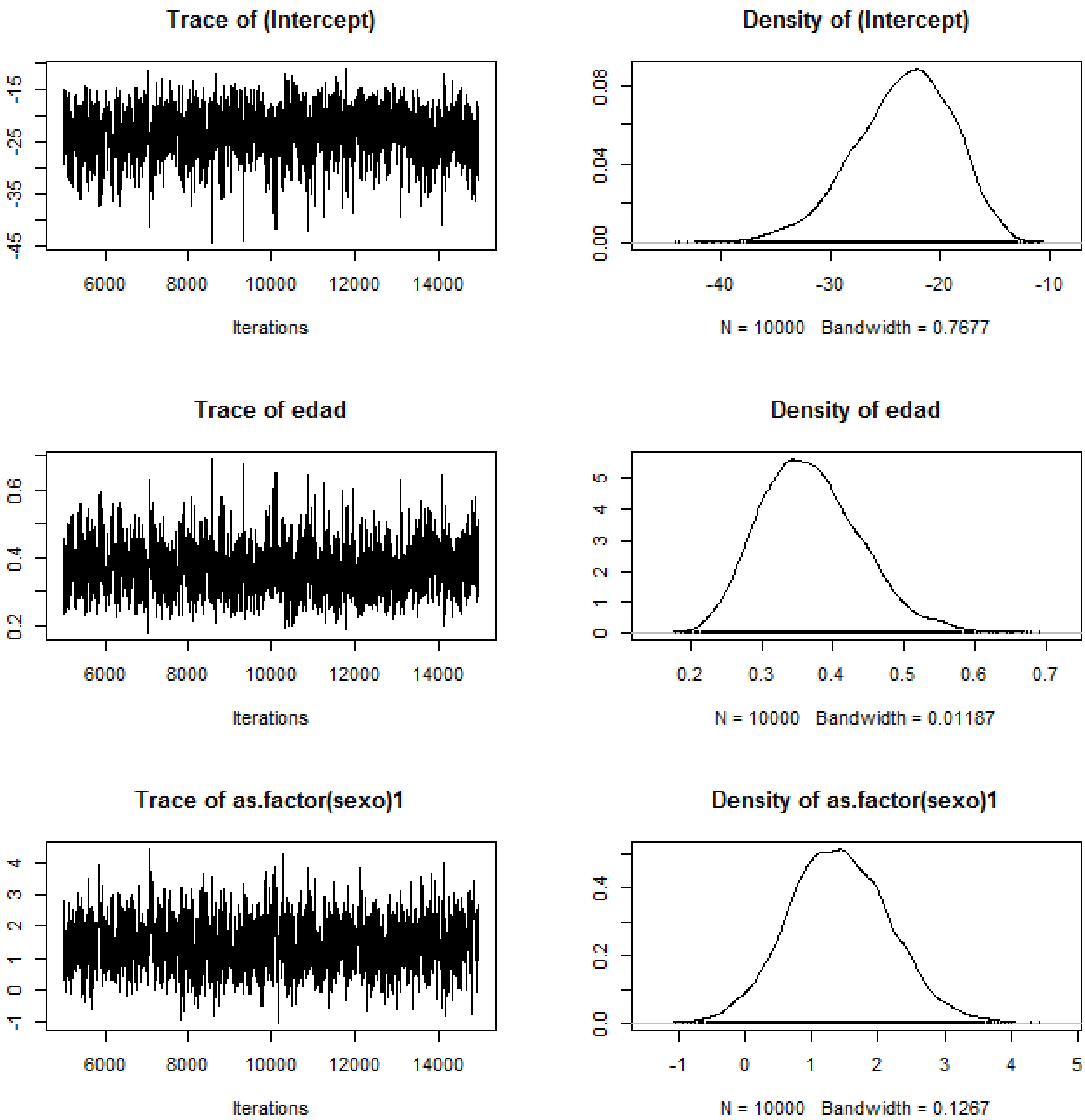

'b.edad'
0
1
2
3
4
5
0.2
0.3
0.4
0.5
0.6


'b.sexo'
0.0
0.1
0.2
0.3
0.4
0.5
-1
0
1
2
3
4
5

El mismo ejemplo con `MCMCpack`:

```
library(MCMCpack)

# Con una distribucion a priori impropia
posterior = MCMClogit(tiene ~ edad + as.factor(sexo),
burnin=5000, mcmc=10000)

# Con una distribucion a priori normal multivariante
posterior = MCMClogit(tiene ~ edad + as.factor(sexo),
burnin=5000, mcmc=10000, b0=0, B0=.001)

plot(posterior)
summary(posterior)
```

```
Iterations = 5001:15000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                     Mean      SD  Naive SE Time-series SE
(Intercept)      -23.3779 4.56941 0.0456941       0.154211
edad               0.3689 0.07065 0.0007065       0.002386
as.factor(sexo)1   1.4151 0.75397 0.0075397       0.025395

2. Quantiles for each variable:

                      2.5%      25%      50%      75%    97.5%
(Intercept)      -33.38697 -26.3090 -23.0174 -20.1153 -15.5434
edad               0.24739   0.3183   0.3635   0.4142   0.5228
as.factor(sexo)1  -0.05774   0.8973   1.3970   1.9223   2.9582
```


Trace of (Intercept)
Density of (Intercept)
Iterations
N = 10000 Bandwidth = 0.7677
Trace of edad
Density of edad
Iterations
N = 10000 Bandwidth = 0.01187
Trace of as.factor(sexo)1
Density of as.factor(sexo)1
Iterations
N = 10000 Bandwidth = 0.1267

## Ejemplo de Regresión de Poisson

Se considera una regresión de Poisson para los siguientes datos

| Diag | Cases | Pyears | Inc | Age |
|---|---|---|---|---|
| 1988 | 7 | 255771 | 0.3 | 6.0 |
| 1989 | 8 | 276644 | 0.3 | 5.6 |
| 1990 | 16 | 295901 | 0.5 | 5.0 |
| 1991 | 14 | 309682 | 0.5 | 4.4 |
| 1992 | 20 | 316457 | 0.6 | 4.0 |
| 1993 | 35 | 316802 | 1.1 | 5.8 |
| 1994 | 29 | 318305 | 0.9 | 4.6 |
| 1995 | 46 | 303544 | 1.5 | 4.3 |
| 1996 | 36 | 260644 | 1.4 | 4.7 |
| 1997 | 47 | 216826 | 2.2 | 4.3 |
| 1998 | 34 | 161664 | 2.1 | 5.4 |
| 1999 | 13 | 60502 | 2.1 | 5.9 |

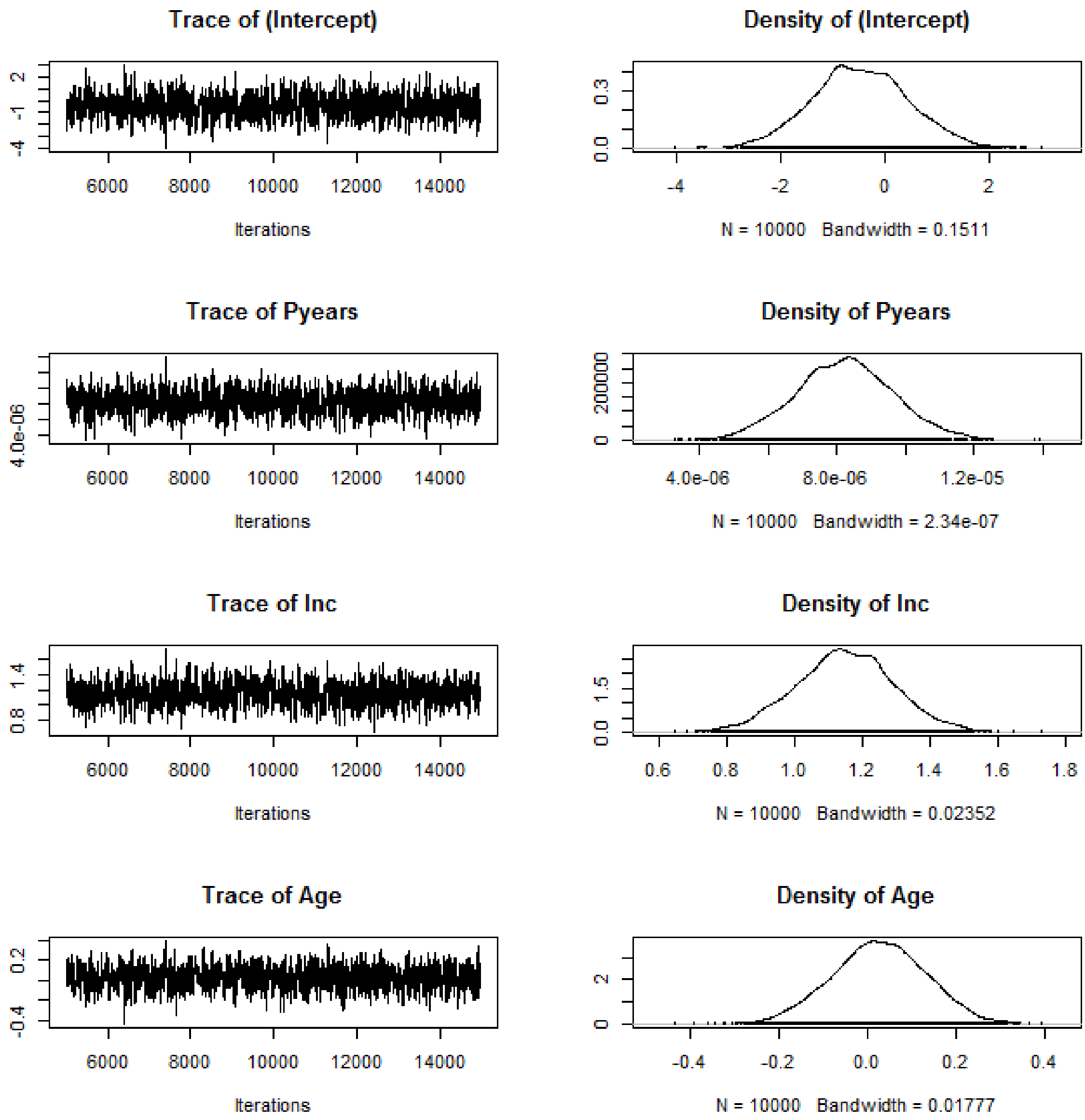
La definición de las variables es:

1. `Diag`: año de diagnóstico
2. `Cases`: casos detectados en ese año
3. `Pyears`: Numero de personas en situación de riesgo
4. `Inc`: Incidencia media para cada 10000 personas
5. `Age`: Mediana de edad de detección de la enfermedad

Solución usando el paquete `MCMCpack`:

```
library(MCMCpack)

Cases = c(7,8,16,14,20,35,29,46,36,47,34,13)

Pyears = c(255771,276644,295901,309682,316457,
316802,318305,303544,260644,216826,161664,60502)

Inc = c(0.3,0.3,0.5,0.5,0.6,1.1,0.9,1.5,1.4,2.2,2.1,2.1)

Age = c(6.0,5.6,5.0,4.4,4.0,5.8,4.6,4.3,4.7,4.3,5.4,5.9)

posterior = MCMCpoisson(Cases ~ Pyears + Inc + Age, burnin=5000,
mcmc=10000)

plot(posterior)
summary(posterior)
```

```
Iterations = 5001:15000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                  Mean        SD  Naive SE Time-series SE
(Intercept) -4.710e-01 9.232e-01 9.232e-03      3.374e-02
Pyears       8.237e-06 1.425e-06 1.425e-08      5.190e-08
Inc          1.148e+00 1.432e-01 1.432e-03      5.242e-03
Age          2.606e-02 1.058e-01 1.058e-03      3.827e-03

2. Quantiles for each variable:

                  2.5%        25%        50%       75%     97.5%
(Intercept) -2.250e+00 -1.067e+00 -4.879e-01 1.383e-01 1.363e+00
Pyears       5.428e-06  7.308e-06  8.241e-06 9.174e-06 1.107e-05
Inc          8.672e-01  1.055e+00  1.148e+00 1.243e+00 1.434e+00
Age         -1.804e-01 -4.369e-02  2.654e-02 9.831e-02 2.284e-01
```


Trace of (Intercept)
Density of (Intercept)
Iterations
N = 10000 Bandwidth = 0.1511
Trace of Pyears
Density of Pyears
Iterations
N = 10000 Bandwidth = 2.34e-07
Trace of Inc
Density of Inc
Iterations
N = 10000 Bandwidth = 0.02352
Trace of Age
Density of Age
Iterations
N = 10000 Bandwidth = 0.01777