

***ANALISIS DE REGRESION LINEAL CON
ERRORES DISTRIBUIDOS NORMAL SESGADOS***

*María Guadalupe Russell Noriega and Graciela González
Farías*

Comunicación Técnica No I-02-30/10-12-2002
(PE/CIMAT)
(Tesina)



Análisis de Regresión Lineal con Errores Distribuidos Normal Sesgados

María Guadalupe Russell Noriega, mgrussell@cimat.mx

Graciela González Farías, farias@cimat.mx

Centro de Investigación en Matemáticas, A.C.

Apartado Postal 402, C.P. 36000, Guanajuato, Gto., México.

December 10, 2002

Contenido

1	Regresión lineal simple con errores normales sesgados	1
	Introducción	1
	Distribución normal sesgada univariada	1
	Modelo de Regresión lineal simple con errores normales sesgados	3
	Estimación de parámetros	3
	Ecuaciones de verosimilitud	3
	Cero trivial	5
	Matriz de información de Fisher observada	7
	Cálculo de las segundas derivadas	8
	Momentos de los estimadores de mínimos cuadrados	12
	Ejemplo simulado	14
	Conclusiones	19
2	Inferencias para muestras no-aleatorias (Copas y Li 1997)	21
	Introducción	21
	Modelo básico	22
	Cálculos de la media y varianza	26
	Función de verosimilitud	28
	Ejemplo simulado	30
	Conclusiones	33
	Apéndice	35

Prólogo

Existe una tendencia general en la literatura estadística en dirección a métodos más flexibles para representar características de los datos en la medida de lo posible y reducir así supuestos poco realistas al momento de los análisis realizados. Por ejemplo en el área de análisis clínicos, específicamente en el estudio de datos biomédicos, comparación de distintos tratamientos, y observaciones faltantes, las distribuciones de los datos presentan de manera natural un comportamiento asimétrico.

Otro ejemplo es en el estudio de observaciones multivariadas continuas dentro de una aproximación paramétrica, donde un supuesto a vencer y que juega un papel fundamental es el supuesto de normalidad, el cual es la base de muchos métodos de análisis multivariado. Sin lugar a dudas la razón por la cual dicho supuesto es difícil de modificar es debido a la flexibilidad matemática de la distribución normal, en particular su simplicidad cuando trabajamos con operaciones fundamentales como combinaciones lineales, marginalización y condicionamiento, así como su cerradura bajo dichas operaciones.

Desde un punto de vista práctico, la aproximación más adoptada para lograr normalidad es la transformación de las variables, lo cual en muchos casos funciona muy bien. Sin embargo pueden presentarse algunos problemas como son:

a) Las transformaciones para cada una de las variables involucradas generalmente son por separado, lo que resulta más difícil de interpretar, especialmente cuando cada variable es transformada por una función diferente.

b) Cuando se asume el supuesto de homocedasticidad, algunas veces la transformación requerida difiere de la transformación para lograr normalidad.

Existen una gran cantidad de funciones de densidad que se aproximan a la normal cuando ciertos parámetros tienden a un valor apropiado. Sin embargo sólo un pequeño grupo de distribuciones paramétricas incluyen el caso normal como un miembro propio y no solamente como un caso límite. Más aún, entre la clase de distribuciones con la propiedad de inclusión estricta, resulta que algunas de ellas no son matemáticamente tratables, mientras que otras no aparecen de forma natural. De lo anterior algunas características deseables de dichas distribuciones son:

- a) Inclusión estricta de la densidad normal;
- b) Manejable matemáticamente;
- c) Rango amplio de los índices sesgo y curtosis.

Azzalini y Capitanio (1999) dan referencias explícitas en las cuales se estudian diferentes clases paramétricas de distribuciones con algunas características deseables. En nuestro caso particular trabajamos con la distribución normal sesgada definida en el Capítulo 1, así como con algunas propiedades de nuestro interés particular.

En el Capítulo 1 tratamos el problema de regresión lineal simple suponiendo que la distribución de los errores es normal sesgada (Azzalini y Capitanio, 1999). La Sección 2 nos sirve para mostrar una breve introducción a la distribución normal sesgada, su definición y algunas propiedades de interés. En las secciones 3 y 4 abordamos el problema de regresión con errores normal sesgados, estimamos los parámetros del modelo mediante la maximización de la función de verosimilitud. Como parte del material estudiado de dichas secciones, se presentan algunas observaciones en torno a los estimadores de mínimos cuadrados para el intercepto y la pendiente; calculamos los primeros momentos y mencionamos algunas propiedades distribucionales. Finalmente en la Sección 5 presentamos un conjunto de datos simulados e ilustramos el procedimiento de simulación, y estimación.

En el Capítulo 2 abordamos el problema de regresión usual en el cual la variable dependiente es explicada por ciertas covariables mediante un modelo lineal con errores normales, con la diferencia de que la variable dependiente es observada siempre que una cierta variable auxiliar sea positiva. Dicha variable auxiliar es modelada también por una relación del tipo lineal con errores normales. En la Sección 1 damos una pequeña introducción de este modelo analizado por Copas y Li (1997) para diferentes aplicaciones estadísticas en las cuales el proceso generador de los datos no es aleatorio. En la Sección 2 presentamos el modelo básico, así como la distribución condicional para la variable dependiente estudiada en Copas y Li (1997). Observamos que dicha distribución condicional resulta ser un caso particular de la distribución normal sesgada general dada en Domínguez-Molina *et al* (2001) y bajo ciertas restricciones de los parámetros la distribución resultante es la distribución de Azzalini (1985) después de agregarle parámetros de localización y escala de la manera usual, dicha expresión viene en Azzalini y Dalla Valle (1996). Presentamos también los cálculos de los primeros dos momentos. En la Sección 3 abordamos el problema de estimación para conjuntos de datos simulados; el proceso de estimación lo realizamos por medio de la maximización directa de la función de log verosimilitud.

Capítulo 1

Regresión lineal simple con errores normales sesgados

Introducción

El énfasis de este trabajo se basa en explorar el potencial de la distribución normal sesgada en aplicaciones a modelos de regresión lineal. En el contexto de regresión lineal existen una gran cantidad de aplicaciones reales, en las cuales los supuestos de normalidad y varianza constante no son factibles. En numerosas situaciones el comportamiento de los datos resulta ser asimétrico y de aquí la inquietud de estudiar el modelo de regresión lineal asumiendo una distribución con características matemáticas similares a la distribución normal y capaz de reproducir el fenómeno de asimetría presentado por los datos. Suponemos entonces que la distribución de los errores es normal sesgada.

Distribución normal sesgada univariada

Consideremos una variable aleatoria continua X con función de densidad de la forma:

$$f(x; \mu, \sigma, \delta) = 2 \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \Phi\left[\delta \left(\frac{x - \mu}{\sigma}\right)\right], \quad (1.1)$$

con μ el parámetro de localización y σ el parámetro de escala. δ es un número fijo arbitrario conocido como el parámetro de forma, ya que dicho parámetro regula la forma o sesgo de la función de densidad. Finalmente $\phi(x)$ y $\Phi(x)$ denotan la función de densidad y de distribución de una variable normal estándar, respectivamente. Diremos que $X \sim SN(\mu, \sigma, \delta)$

siempre que su función de densidad este dada por la ecuación (1.1).

La densidad $f(x; \mu, \sigma, \delta)$ posee algunas propiedades interesantes, como son:

1. Cuando $\delta = 0$, el sesgo desaparece y obtenemos la densidad normal con parámetros μ y σ^2 .
2. Cuando δ crece (en valor absoluto) el sesgo de la distribución crece.
3. Cuando $\delta \rightarrow \infty$, la densidad converge a la llamada densidad half-normal con parámetros μ y σ^2 .
4. Si el signo de δ cambia, la densidad se refleja en el lado opuesto del eje vertical.
5. Si X variable aleatoria que distribuye $SN(0, 1, \delta)$, entonces X^2 tiene una distribución χ_1^2 independientemente del valor de δ .

Usando los resultados de Azzalini (1985), se tiene que los primeros tres momentos, así como la varianza de la variable aleatoria X son:

$$\begin{aligned} E(X) &= \mu + \sqrt{\frac{2}{\pi}}\sigma\lambda; & E(X^2) &= \mu^2 + 2\sqrt{\frac{2}{\pi}}\mu\sigma\lambda + \sigma^2 \\ E(X^3) &= \mu^3 + 3\sqrt{\frac{2}{\pi}}\mu^2\sigma\lambda + 3\mu\sigma^2 + 3\sqrt{\frac{2}{\pi}}\sigma^3\lambda - \sqrt{\frac{2}{\pi}}\sigma^3\lambda^3 \\ \text{Var}(X) &= \sigma^2 \left(1 - \frac{2}{\pi}\lambda^2\right), \end{aligned} \tag{1.2}$$

donde $\lambda = \lambda(\delta) = \delta / (1 + \delta^2)^{1/2}$.

Para un estudio detallado de la distribución normal sesgada ver Azzalini (1985), Azzalini y Dalla Valle (1996) y Azzalini y Capitanio (1999).

A manera de notación decimos que

$$U_i \sim SNID(\mu, \sigma, \delta), \quad i = 1, 2, \dots, n,$$

lo cual se lee como, los U_i son normales sesgados e independientemente distribuidos con parámetros comunes μ , σ y δ .

Modelo de Regresión lineal simple con errores normales sesgados

Considere un modelo de regresión donde la distribución de los errores sigue una distribución normal sesgada y ε_i independiente de ε_j para $i \neq j$, tal y como lo proponen en Azzalini y Capitanio (1999). La forma del modelo para el caso univariado es:

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad \varepsilon_i \sim SNID(0, \sigma, \delta), i = 1, \dots, n, \quad (1.3)$$

por lo tanto se sigue que,

$$y_i \sim SNID(\alpha + \beta x_i, \sigma, \delta).$$

La función de verosimilitud para los parámetros $\alpha, \beta, \sigma, \delta$, está dada por el producto de las densidades de la forma (1.1) y se escribe como

$$\begin{aligned} L(\alpha, \beta, \sigma, \delta; y) &= \frac{1}{\sigma^n} \prod_{i=1}^n \phi\left(\frac{y_i - \alpha - \beta x_i}{\sigma}\right) \Phi\left(\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma}\right)\right) \\ &\propto \frac{1}{\sigma^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2\right) \prod_{i=1}^n \Phi\left(\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma}\right)\right), \end{aligned}$$

así mismo la correspondiente función de log verosimilitud es la siguiente,

$$\ell(\alpha, \beta, \sigma, \delta; y) = -n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 + \sum_{i=1}^n \ln \Phi\left(\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma}\right)\right). \quad (1.4)$$

Se da un ejemplo del cálculo de los estimadores máximo verosímil el cual se realiza numéricamente de la log verosimilitud dada en (1.4). Se maximiza $\ell(\alpha, \beta, \sigma, \delta; y)$, para un conjunto de datos simulados, mediante la función `nlmin` de S-PLUS.

Estimación de parámetros

Ecuaciones de verosimilitud

Sea $\Lambda(x) = \ln \Phi(x)$, y $\boldsymbol{\theta} = (\alpha, \beta, \sigma, \delta)$ de modo que la log verosimilitud dada en (1.4) se escribe como:

$$\ell(\boldsymbol{\theta}) = -n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 + \sum_{i=1}^n \Lambda\left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma}\right)\right].$$

Las derivadas parciales con respecto a cada uno de los parámetros α , β , σ , y δ , se presentan a continuación.

1. Notemos que si $h(\alpha) = \delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \Rightarrow h'(\alpha) = -\frac{\delta}{\sigma}$, de modo que la derivada parcial con respecto a α de $\ell(\boldsymbol{\theta})$ es,

$$\frac{\partial}{\partial \alpha} \ell(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i) - \frac{\delta}{\sigma} \sum_{i=1}^n \Lambda' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right]. \quad (1.5)$$

2. Así mismo considerando $h(\beta) = \delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \Rightarrow h'(\beta) = -\frac{\delta}{\sigma} x_i$ y de aquí que la derivada parcial de $\ell(\boldsymbol{\theta})$ con respecto a β es,

$$\frac{\partial}{\partial \beta} \ell(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \sum_{i=1}^n x_i (y_i - \alpha - \beta x_i) - \frac{\delta}{\sigma} \sum_{i=1}^n x_i \Lambda' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right]. \quad (1.6)$$

3. Si $h(\sigma) = \delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \Rightarrow h'(\sigma) = -\frac{\delta}{\sigma^2} (y_i - \alpha - \beta x_i)$, tenemos que la derivada parcial de $\ell(\boldsymbol{\theta})$ con respecto a σ esta dada por,

$$\begin{aligned} \frac{\partial}{\partial \sigma} \ell(\boldsymbol{\theta}) &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \\ &\quad - \frac{\delta}{\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i) \Lambda' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right]. \end{aligned} \quad (1.7)$$

4. Finalmente si $h(\delta) = \delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \Rightarrow h'(\delta) = \frac{1}{\sigma} (y_i - \alpha - \beta x_i)$ y la derivada parcial de $\ell(\boldsymbol{\theta})$ con respecto a δ es,

$$\frac{\partial}{\partial \delta} \ell(\boldsymbol{\theta}) = \frac{1}{\sigma} \sum_{i=1}^n (y_i - \alpha - \beta x_i) \Lambda' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right]. \quad (1.8)$$

Observación: Al resolver (1.8) obtenemos que el EMV de δ , $\hat{\delta}_{\alpha, \beta, \sigma}$, restringido a los valores de α, β y σ origina la siguiente ecuación

$$\sum_{i=1}^n (y_i - \alpha - \beta x_i) \Lambda' \left[\hat{\delta}_{\alpha, \beta, \sigma} \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right] = 0,$$

de la cual, al sustituir δ por $\hat{\delta}_{\alpha,\beta,\sigma}$ en la ecuación (1.7) y resolviendo dicha ecuación con respecto a σ obtenemos que,

$$-\frac{n}{\hat{\sigma}_{\alpha,\beta}} + \frac{1}{\hat{\sigma}_{\alpha,\beta}^3} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 = 0, \quad (1.9)$$

de donde se desprende que el EMV de σ , $\hat{\sigma}_{\alpha,\beta}$, restringido a valores dados de α y β está dado por:

$$\hat{\sigma}_{\alpha,\beta}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2. \quad (1.10)$$

Note que el estimador $\hat{\sigma}_{\alpha,\beta}^2$ tiene la misma forma que el EMV usual de σ cuando $\delta = 0$. Sin embargo, es importante recordar que la expresión (1.10) depende de $\hat{\delta}_{\alpha,\beta,\sigma}$ y aún cuando nos ahorra una dimensión en la estimación de los EMV de α, β, σ y δ , debido a que tenemos una expresión cerrada para $\hat{\sigma} = \hat{\sigma}_{\hat{\alpha},\hat{\beta}}$, no ayuda a resolver la verosimilitud de una manera sencilla.

Observemos que si δ es conocido y diferente de cero el EMV de σ debe ser aquel que satisfaga la ecuación (1.7), debido a que ya no se cumple (1.9). Azzalini (1985) comete el error de suponer que se sigue cumpliendo (1.9) cuando δ es conocido (ver Arnold, *et al* 1993, página 479).

Cero trivial

En esta sección mostramos que los estimadores de mínimos cuadrados (MC) satisfacen las ecuaciones de verosimilitud (1.5), (1.6), (1.7) y (1.8) cuando $\delta = 0$. Sin embargo dicho punto es de inflexión, *i.e.*, no es ni máximo ni mínimo lo cual se demuestra al probar que la matriz de información de Fisher no es positiva definida ni negativa definida.

Recordemos que los estimadores de MC para α, β , en el modelo de regresión lineal simple están dados por las siguientes expresiones:

$$\hat{\alpha}_{MC} = \bar{y} - \bar{x}\hat{\beta}_{MC}, \quad \hat{\beta}_{MC} = \frac{s_{xy}}{s_{xx}},$$

donde $s_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$, $s_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$ y si $\hat{\varepsilon}_{MC,i} = y_i - \hat{\alpha}_{MC} - \hat{\beta}_{MC}x_i$, se cumple que

$$\sum_{i=1}^n \hat{\varepsilon}_{MC,i} = 0, \quad \sum_{i=1}^n x_i \hat{\varepsilon}_{MC,i} = 0, \quad \sum_{i=1}^n \hat{\varepsilon}_{MC,i}^2 = n\hat{\sigma}_0^2,$$

donde $\hat{\sigma}_0$ es el EMV usual para σ cuando $\delta = 0$. Observemos que el punto $\hat{\boldsymbol{\theta}}_0 = (\hat{\alpha}_{MC}, \hat{\beta}_{MC}, \hat{\sigma}_0, 0)$ es una raíz de las ecuaciones de verosimilitud, es decir, las primeras derivadas parciales con respecto a cada uno de los parámetros evaluadas en dicho punto son cero.

Ya que $\Lambda(x) = \ln \Phi(x)$, se cumple que

$$\Lambda'(x) = \frac{d}{dx} \Lambda(x) = \frac{\phi(x)}{\Phi(x)} \quad \text{y} \quad \Lambda'(0) = \frac{\phi(0)}{\Phi(0)} = \sqrt{\frac{2}{\pi}}.$$

De esta forma al evaluar las primeras derivadas en el punto $\hat{\boldsymbol{\theta}}_0$ obtenemos que:

$$\begin{aligned} \left. \frac{\partial}{\partial \alpha} \ell(\boldsymbol{\theta}; y) \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} &= \frac{1}{\hat{\sigma}_0^2} \sum_{i=1}^n (y_i - \hat{\alpha}_{MC} - \hat{\beta}_{MC}x_i) \\ &= \frac{1}{\hat{\sigma}_0^2} n (\bar{y} - \hat{\alpha}_{MC} - \hat{\beta}_{MC}\bar{x}) = 0 \end{aligned}$$

$$\begin{aligned} \left. \frac{\partial}{\partial \beta} \ell(\boldsymbol{\theta}; y) \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} &= \frac{1}{\hat{\sigma}_0^2} \sum_{i=1}^n x_i (y_i - \hat{\alpha}_{MC} - \hat{\beta}_{MC}x_i) \\ &= \frac{1}{\hat{\sigma}_0^2} \left(-n\hat{\alpha}_{MC}\bar{x} + \sum_{i=1}^n x_i (y_i - \hat{\beta}_{MC}x_i) \right) = 0 \end{aligned}$$

$$\begin{aligned} \left. \frac{\partial}{\partial \sigma} \ell(\boldsymbol{\theta}; y) \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} &= -\frac{n}{\hat{\sigma}_0} + \frac{1}{\hat{\sigma}_0^3} \sum_{i=1}^n (y_i - \hat{\alpha}_{MC} - \hat{\beta}_{MC}x_i)^2 \\ &= -\frac{n}{\hat{\sigma}_0} + \frac{n}{\hat{\sigma}_0} = 0 \end{aligned}$$

$$\begin{aligned} \left. \frac{\partial}{\partial \delta} \ell(\boldsymbol{\theta}; y) \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} &= \frac{1}{\hat{\sigma}_0} \sum_{i=1}^n (y_i - \hat{\alpha}_{MC} - \hat{\beta}_{MC}x_i) \Lambda'(0) \\ &= \frac{1}{\hat{\sigma}_0} \frac{\phi(0)}{\Phi(0)} \sum_{i=1}^n (y_i - \hat{\alpha}_{MC} - \hat{\beta}_{MC}x_i) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\hat{\sigma}_0} \sqrt{\frac{2}{\pi}} \sum_{i=1}^n \left(y_i - \hat{\alpha}_{MC} - \hat{\beta}_{MC} x_i \right) \\
&= \frac{1}{\hat{\sigma}_0} \sqrt{\frac{2}{\pi}} \sum_{i=1}^n \hat{\varepsilon}_{MC,i} = 0.
\end{aligned}$$

Azzalini (1985) observa que el punto $\hat{\boldsymbol{\theta}}_0 = (\hat{\alpha}_{MC}, \hat{\beta}_{MC}, \hat{\sigma}_0, 0)$ es un punto de inflexión y de aquí que la matriz de información de Fisher esperada no es positiva definida.¹ Los detalles de la demostración de la afirmación anterior no vienen en el trabajo anterior. En la siguiente sección se prueba que la matriz de información de Fisher observada no es positiva definida. Mediante argumentos muy similares se observa que la matriz de información de Fisher esperada tampoco es positiva definida.

Matriz de información de Fisher observada

Para identificar si el punto $\hat{\boldsymbol{\theta}}_0$ es un punto de inflexión trabajamos con la matriz de Fisher observada.

La *matriz de información de Fisher observada* es menos el hessiano de la función de log verosimilitud, *i.e.*,

$$\begin{aligned}
j(\alpha, \beta, \sigma, \delta) &= -\nabla \nabla^T \ell(\alpha, \beta, \sigma, \delta) \\
&= -\frac{\partial^2 \ell(\alpha, \beta, \sigma, \delta)}{\partial(\alpha, \beta, \sigma, \delta) \partial(\alpha, \beta, \sigma, \delta)^T}.
\end{aligned}$$

En términos de $\boldsymbol{\theta} = (\alpha, \beta, \sigma, \delta)^T$ tenemos que,

$$j(\boldsymbol{\theta}) = -\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^T \ell(\boldsymbol{\theta}) = \left(-\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell(\boldsymbol{\theta}) \right)_{i,j=1}^4.$$

¹En realidad Azzalini (1985, página 175) dice que la matriz de información de Fisher es singular; sin embargo, cualquier libro de álgebra lineal define una matriz como singular si su determinante es cero. En este caso la matriz resultante no es de covarianzas porque no es positiva definida. No todos sus eigenvalores son del mismo signo.

Cálculo de las segundas derivadas

En el cálculo de las segundas derivadas utilizaremos las siguientes igualdades:

- $\Lambda'(x) = \frac{d}{dx}\Lambda(x) = \frac{d}{dx}\ln\Phi(x) = \frac{\phi(x)}{\Phi(x)}$
- $\frac{\partial}{\partial x}\Lambda(h(x)) = \frac{\phi(h(x))}{\Phi(h(x))}h'(x) = \Lambda'(h(x))h'(x)$
- $\frac{\partial}{\partial x}\Lambda'(h(x)) = \Lambda''(h(x))h'(x)$.

1. Recordemos que la derivada de $\ell(\alpha, \beta, \sigma, \delta)$ con respecto a α dada en (1.5) está dada por

$$\frac{\partial}{\partial \alpha}\ell(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i) - \frac{\delta}{\sigma} \sum_{i=1}^n \Lambda' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right],$$

y de aquí obtenemos las siguientes cuatro derivadas cruzadas

$$\frac{\partial^2}{\partial \alpha^2}\ell(\boldsymbol{\theta}) = -\frac{n}{\sigma^2} - \frac{\delta^2}{\sigma^2} \sum_{i=1}^n \Lambda'' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right],$$

$$\frac{\partial^2}{\partial \alpha \partial \beta}\ell(\boldsymbol{\theta}) = -\frac{1}{\sigma^2} \sum_{i=1}^n x_i - \frac{\delta^2}{\sigma^2} \sum_{i=1}^n x_i \Lambda'' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right],$$

$$\begin{aligned} \frac{\partial^2}{\partial \alpha \partial \sigma}\ell(\boldsymbol{\theta}) &= -\frac{2}{\sigma^3} \sum_{i=1}^n (y_i - \alpha - \beta x_i) + \frac{\delta}{\sigma^2} \sum_{i=1}^n \Lambda' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right] \\ &\quad + \frac{\delta^2}{\sigma^3} \sum_{i=1}^n (y_i - \alpha - \beta x_i) \Lambda'' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right], \end{aligned}$$

$$\begin{aligned} \frac{\partial^2}{\partial \alpha \partial \delta}\ell(\boldsymbol{\theta}) &= -\frac{1}{\sigma} \sum_{i=1}^n \Lambda' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right] \\ &\quad - \frac{\delta}{\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i) \Lambda'' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right]. \end{aligned}$$

2. Similarmente, reescribiendo (1.6) para facilitar los cálculos de las derivadas cruzadas de $\ell(\alpha, \beta, \sigma, \delta)$ con respecto a los parámetros correspondientes, obtenemos que

$$\frac{\partial}{\partial \beta}\ell(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \sum_{i=1}^n x_i (y_i - \alpha - \beta x_i) - \frac{\delta}{\sigma} \sum_{i=1}^n x_i \Lambda' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right],$$

de lo cual se sigue que,

$$\frac{\partial^2}{\partial \beta^2} \ell(\boldsymbol{\theta}) = -\frac{1}{\sigma^2} \sum_{i=1}^n x_i^2 + \frac{\delta^2}{\sigma^2} \sum_{i=1}^n x_i^2 \Lambda'' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right],$$

$$\begin{aligned} \frac{\partial^2}{\partial \beta \partial \sigma} \ell(\boldsymbol{\theta}) &= -\frac{2}{\sigma^3} \sum_{i=1}^n x_i (y_i - \alpha - \beta x_i) + \frac{\delta}{\sigma^2} \sum_{i=1}^n x_i \Lambda' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right] \\ &\quad + \frac{\delta^2}{\sigma^3} \sum_{i=1}^n x_i (y_i - \alpha - \beta x_i) \Lambda'' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right], \end{aligned}$$

$$\begin{aligned} \frac{\partial^2}{\partial \beta \partial \delta} \ell(\boldsymbol{\theta}) &= -\frac{1}{\sigma} \sum_{i=1}^n x_i \Lambda' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right] \\ &\quad - \frac{\delta}{\sigma^2} \sum_{i=1}^n x_i (y_i - \alpha - \beta x_i) \Lambda'' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right]. \end{aligned}$$

3. Así mismo ya que la derivada de $\ell(\boldsymbol{\theta})$ con respecto a σ es,

$$\begin{aligned} \frac{\partial}{\partial \sigma} \ell(\boldsymbol{\theta}) &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \\ &\quad - \frac{\delta}{\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i) \Lambda' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right] \end{aligned}$$

así

$$\begin{aligned} \frac{\partial^2}{\partial \sigma^2} \ell(\boldsymbol{\theta}) &= \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \\ &\quad + 2 \frac{\delta}{\sigma^3} \sum_{i=1}^n (y_i - \alpha - \beta x_i) \Lambda' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right] \\ &\quad + \frac{\delta^2}{\sigma^4} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \Lambda'' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right], \end{aligned}$$

y

$$\begin{aligned} \frac{\partial^2}{\partial \sigma \partial \delta} \ell(\boldsymbol{\theta}) &= -\frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i) \Lambda' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right] \\ &\quad - \frac{\delta}{\sigma^3} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \Lambda'' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right]. \end{aligned}$$

4. Finalmente, recordemos que la derivada de $\ell(\alpha, \beta, \sigma, \delta)$ con respecto a δ es,

$$\frac{\partial}{\partial \delta} \ell(\boldsymbol{\theta}) = \frac{1}{\sigma} \sum_{i=1}^n (y_i - \alpha - \beta x_i) \Lambda' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right],$$

y por lo tanto,

$$\frac{\partial^2}{\partial \delta^2} \ell(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \Lambda'' \left[\delta \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right].$$

El siguiente paso es evaluar las segundas derivadas parciales de $\ell(\boldsymbol{\theta})$ en el punto $\hat{\boldsymbol{\theta}}_0 = (\hat{\alpha}_{MC}, \hat{\beta}_{MC}, \hat{\sigma}_0, 0)$, usando además las siguientes igualdades: $\Lambda'(0) = \sqrt{\frac{2}{\pi}}$, y $\Lambda''(0) = -\frac{2}{\pi}$.

$$\left. \frac{\partial^2}{\partial \alpha^2} \ell(\boldsymbol{\theta}; y) \right|_{\hat{\boldsymbol{\theta}}_0} = -\frac{n}{2\hat{\sigma}_0^2},$$

$$\left. \frac{\partial^2}{\partial \alpha \partial \beta} \ell(\boldsymbol{\theta}; y) \right|_{(\alpha, \beta, \sigma, \delta) = \hat{\boldsymbol{\theta}}_0} = -\frac{n}{\hat{\sigma}_0^2} \bar{x},$$

$$\left. \frac{\partial^2}{\partial \alpha \partial \sigma} \ell(\boldsymbol{\theta}; y) \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_0} = -\frac{2}{\hat{\sigma}_0^3} \sum_{i=1}^n \hat{\varepsilon}_{MC,i} = 0,$$

$$\left. \frac{\partial^2}{\partial \alpha \partial \delta} \ell(\boldsymbol{\theta}; y) \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_0} = -\frac{n}{\hat{\sigma}_0} \sqrt{\frac{2}{\pi}},$$

$$\left. \frac{\partial^2}{\partial \beta^2} \ell(\boldsymbol{\theta}; y) \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_0} = -\frac{1}{\hat{\sigma}_0^2} \sum_{i=1}^n x_i^2,$$

$$\left. \frac{\partial^2}{\partial \beta \partial \sigma} \ell(\boldsymbol{\theta}; y) \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_0} = -\frac{2}{\hat{\sigma}_0^3} \sum_{i=1}^n x_i \hat{\varepsilon}_{MC,i} = 0,$$

$$\left. \frac{\partial^2}{\partial \beta \partial \delta} \ell(\boldsymbol{\theta}; y) \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_0} = -\sqrt{\frac{2}{\pi}} \frac{n}{\hat{\sigma}_0} \bar{x},$$

$$\begin{aligned} \left. \frac{\partial^2}{\partial \sigma^2} \ell(\boldsymbol{\theta}; y) \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_0} &= \frac{n}{\hat{\sigma}_0^2} - \frac{3}{\hat{\sigma}_0^4} \sum_{i=1}^n \hat{\varepsilon}_{MC,i}^2 \\ &= \frac{n}{\hat{\sigma}_0^2} - \frac{3n\hat{\sigma}_0^2}{\hat{\sigma}_0^4} \\ &= -2\frac{n}{\hat{\sigma}_0^2}, \end{aligned}$$

$$\frac{\partial^2}{\partial \delta \partial \sigma} \ell(\boldsymbol{\theta}; y) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} = -\sqrt{\frac{2}{\pi}} \frac{1}{\hat{\sigma}_0^2} \sum_{i=1}^n \hat{\varepsilon}_{MC,i} = 0.$$

Dado que $\sum_{i=1}^n \hat{\varepsilon}_{MC,i}^2 = n\hat{\sigma}_0^2$ obtenemos que

$$\frac{\partial^2}{\partial \delta^2} \ell(\boldsymbol{\theta}; y) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} = \frac{2}{\pi} \frac{1}{\hat{\sigma}_0^2} \sum_{i=1}^n \hat{\varepsilon}_{MC,i}^2 = \frac{2}{\pi} n.$$

De las ecuaciones anteriores se tiene que la matriz de información de Fisher observada para $\boldsymbol{\theta}$ evaluada en $\hat{\boldsymbol{\theta}}_0$ es,

$$j(\hat{\boldsymbol{\theta}}_0) = - \begin{pmatrix} -\frac{n}{\hat{\sigma}_0^2} & -\frac{n}{\hat{\sigma}_0^2} \bar{x} & 0 & -\frac{n}{\hat{\sigma}_0} \sqrt{\frac{2}{\pi}} \\ -\frac{n}{\hat{\sigma}_0^2} \bar{x} & -\frac{1}{\hat{\sigma}_0^2} \sum_{i=1}^n x_i^2 & 0 & -\sqrt{\frac{2}{\pi}} \frac{n}{\hat{\sigma}_0} \bar{x} \\ 0 & 0 & -2\frac{n}{\hat{\sigma}_0^2} & 0 \\ -\frac{n}{\hat{\sigma}_0} \sqrt{\frac{2}{\pi}} & -\sqrt{\frac{2}{\pi}} \frac{n}{\hat{\sigma}_0} \bar{x} & 0 & \frac{2}{\pi} n \end{pmatrix}.$$

El determinante de la información de Fisher observada es,

$$\det [j(\hat{\boldsymbol{\theta}}_0)] = (-1)^4 8 \frac{n^3}{\pi \hat{\sigma}_0^6} \left(n\bar{x}^2 - \sum_{i=1}^n x_i^2 \right),$$

utilizando la relación

$$\sum_{i=1}^n x_i^2 - n\bar{x}^2 = \sum_{i=1}^n (x_i - \bar{x})^2,$$

llegamos a

$$\det [j(\hat{\boldsymbol{\theta}}_0)] = -8 \frac{n^3}{\pi \hat{\sigma}_0^6} \sum_{i=1}^n (x_i - \bar{x})^2 < 0.$$

Como el determinante de la información de Fisher observada es negativo para $k = 4$ (par), el punto $\hat{\boldsymbol{\theta}}_0 = (\hat{\alpha}_{MC}, \hat{\beta}_{MC}, \hat{\sigma}_0, 0)$ no es un punto que maximiza ni minimiza la función de verosimilitud (ver Marsden y Hoffman (1998), página 365).

Observación: La *matriz de información de Fisher* (o *matriz de información de Fisher esperada*) se define como $i(\boldsymbol{\theta}) = E(j(\boldsymbol{\theta}))$. De manera similar al caso anterior y sin necesidad de calcular las esperanzas involucradas, es posible verificar que el punto $(\hat{\alpha}_{MC}, \hat{\beta}_{MC}, \hat{\sigma}_0, 0)$ es punto de inflexión, y concluir así que $i(\boldsymbol{\theta})$ no es positiva definida ni negativa definida.

Momentos de los estimadores de mínimos cuadrados

En esta sección calculamos la media y la varianza de los estimadores de MC bajo el supuesto de que los errores se distribuyen con la distribución $SN(0, \sigma, \delta)$.

Consideremos el modelo dado en la ecuación (1.3), bajo el supuesto de errores con distribución normal sesgada, se tiene que $y_i \sim SNID(\alpha + \beta x_i, \sigma, \delta)$, con

$$\begin{aligned} E(y_i) &= \alpha + \beta x_i + \lambda \sigma \sqrt{\frac{2}{\pi}} = \alpha^* + \beta x_i \\ \text{Var}(y_i) &= \sigma^2 \left(1 - \frac{2}{\pi} \lambda^2\right) = \sigma_y^2, \end{aligned} \quad (1.11)$$

donde $\alpha^* = \alpha + \lambda \sigma \sqrt{\frac{2}{\pi}}$ y $\lambda = \frac{\delta}{(1+\delta^2)^{1/2}}$, luego se sigue que

$$\begin{aligned} E(\widehat{\beta}_{MC}) &= \frac{1}{s_{xx}} E \left[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right] \\ &= \frac{1}{s_{xx}} \sum_{i=1}^n (x_i - \bar{x}) \left[\alpha^* + \beta x_i - \frac{1}{n} \sum_{i=1}^n (\alpha^* + \beta x_i) \right] \\ &= \frac{1}{s_{xx}} \sum_{i=1}^n (x_i - \bar{x}) (\alpha^* + \beta x_i - \alpha^* - \beta \bar{x}) \\ &= \frac{\beta}{s_{xx}} \sum_{i=1}^n (x_i - \bar{x})^2 = \beta. \end{aligned}$$

Similarmente el valor esperado del estimador del intercepto por MC es,

$$\begin{aligned} E(\widehat{\alpha}_{MC}) &= E(\bar{y} - \widehat{\beta}_{MC} \bar{x}) \\ &= E(\bar{y}) - E(\widehat{\beta}_{MC}) \bar{x} \\ &= \frac{1}{n} \sum_{i=1}^n (\alpha^* + \beta x_i) - \beta \bar{x} \\ &= \alpha^* + \beta \bar{x} - \beta \bar{x} \\ &= \alpha^* = \alpha + \lambda \sigma \sqrt{\frac{2}{\pi}}. \end{aligned}$$

El cálculo de las varianzas es como sigue

$$\text{Var}(\widehat{\beta}_{MC}) = \frac{1}{s_{xx}^2} \text{Var} \left[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right]$$

$$\begin{aligned}
&= \frac{1}{s_{xx}^2} \text{Var} \left[\sum_{i=1}^n (x_i - \bar{x}) y_i \right] \\
&= \frac{1}{s_{xx}^2} \sum_{i=1}^n (x_i - \bar{x})^2 \text{Var}(y_i) \\
&= \frac{1}{s_{xx}^2} \sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2 \left(1 - \frac{2}{\pi} \lambda^2 \right) \\
&= \frac{1}{s_{xx}} \sigma^2 \left(1 - \frac{2}{\pi} \lambda^2 \right) = \frac{\sigma_y^2}{s_{xx}};
\end{aligned}$$

$$\begin{aligned}
\text{Var}(\hat{\alpha}_{MC}) &= \text{Var}(\bar{y} - \hat{\beta}_{MC} \bar{x}) \\
&= \text{Var}(\bar{y}) + \bar{x}^2 \text{Var}(\hat{\beta}_{MC}) - 2 \text{Cov}(\bar{y}, \hat{\beta}_{MC}),
\end{aligned}$$

como veremos inmediatamente $\text{Cov}(\bar{y}, \hat{\beta}_{MC}) = 0$, de donde obtenemos que

$$\begin{aligned}
\text{Var}(\hat{\alpha}_{MC}) &= \frac{1}{n} \sigma^2 \left(1 - \frac{2}{\pi} \lambda^2 \right) + \bar{x}^2 \frac{1}{s_{xx}} \sigma^2 \left(1 - \frac{2}{\pi} \lambda^2 \right) \\
&= \sigma^2 \left(1 - \frac{2}{\pi} \lambda^2 \right) \left[\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}} \right] = \sigma_y^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}} \right].
\end{aligned}$$

Demostración de que $\text{Cov}(\bar{y}, \hat{\beta}_{MC}) = 0$.

$$\begin{aligned}
\text{Cov}(\bar{y}, \hat{\beta}_{MC}) &= E \left(\bar{y} - \alpha - \beta \bar{x} - \lambda \sigma \sqrt{\frac{2}{\pi}} \right) (\hat{\beta}_{MC} - \beta) \\
&= E \left[\bar{y} (\hat{\beta}_{MC} - \beta) \right] \\
&= E \left\{ \bar{y} \left[\frac{1}{s_{xx}} \sum_{i=1}^n (x_i - \bar{x}) (y_i - \bar{y}) - \beta \right] \right\} \\
&= \frac{1}{s_{xx}} \sum_{i=1}^n (x_i - \bar{x}) E(\bar{y} y_i) - \beta E(\bar{y}) \\
&= \frac{1}{s_{xx}} \sum_{i=1}^n (x_i - \bar{x}) E \left(\frac{1}{n} \sum_{j=1}^n y_j y_i \right) - \beta \left(\alpha + \beta \bar{x} + \lambda \sigma \sqrt{\frac{2}{\pi}} \right) \\
&= \frac{1}{s_{xx}} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x}) E(y_j y_i) - \beta \left(\alpha + \beta \bar{x} + \lambda \sigma \sqrt{\frac{2}{\pi}} \right),
\end{aligned}$$

por independencia entre y_i y y_j para $i \neq j$ tenemos que

$$\begin{aligned}
& \text{Cov} \left(\bar{y}, \widehat{\beta}_{MC} \right) \\
&= \frac{1}{s_{xx}} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x}) \left(\alpha + \beta x_i + \lambda \sigma \sqrt{\frac{2}{\pi}} \right) \left(\alpha + \beta x_j + \lambda \sigma \sqrt{\frac{2}{\pi}} \right) \\
&\quad - \beta \left(\alpha + \beta \bar{x} + \lambda \sigma \sqrt{\frac{2}{\pi}} \right) \\
&= \frac{1}{s_{xx}} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x}) \left(\alpha \beta x_j + \beta x_i \alpha + \beta^2 x_i x_j + \beta x_i \lambda \sigma \frac{\sqrt{2}}{\sqrt{\pi}} + \lambda \sigma \frac{\sqrt{2}}{\sqrt{\pi}} \beta x_j \right) \\
&\quad + \frac{1}{s_{xx}} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x}) \left(\alpha^2 + 2\alpha \lambda \sigma \frac{\sqrt{2}}{\sqrt{\pi}} + 2\lambda^2 \frac{\sigma^2}{\pi} \right) - \beta \left(\alpha + \beta \bar{x} + \lambda \sigma \sqrt{\frac{2}{\pi}} \right),
\end{aligned}$$

con ayuda de las relaciones

$$\sum_{i=1}^n (x_i - \bar{x}) = 0 \quad \text{y} \quad \sum_{i=1}^n (x_i - \bar{x}) x_i = s_{xx},$$

lo anterior se reduce a

$$\begin{aligned}
& \text{Cov} \left(\bar{y}, \widehat{\beta}_{MC} \right) \\
&= \frac{1}{s_{xx}} \frac{1}{n} \sum_{j=1}^n \left(\alpha \beta \sum_{i=1}^n (x_i - \bar{x}) x_i + \beta^2 \sum_{i=1}^n (x_i - \bar{x}) x_i x_j + \beta \lambda \sigma \frac{\sqrt{2}}{\sqrt{\pi}} \sum_{i=1}^n (x_i - \bar{x}) x_i \right) \\
&\quad - \beta \left(\alpha + \beta \bar{x} + \lambda \sigma \sqrt{\frac{2}{\pi}} \right) \\
&= \frac{1}{s_{xx}} \frac{1}{n} \sum_{j=1}^n \left(\alpha \beta s_{xx} + \beta^2 s_{xx} x_j + \beta \lambda \sigma \frac{\sqrt{2}}{\sqrt{\pi}} s_{xx} \right) - \beta \left(\alpha + \beta \bar{x} + \lambda \sigma \sqrt{\frac{2}{\pi}} \right) \\
&= \alpha \beta \frac{1}{s_{xx}} s_{xx} + \frac{1}{s_{xx}} \beta^2 s_{xx} \bar{x} + \beta \lambda \sigma \frac{\sqrt{2}}{\sqrt{\pi}} \frac{1}{s_{xx}} s_{xx} - \beta \left(\alpha + \beta \bar{x} + \lambda \sigma \sqrt{\frac{2}{\pi}} \right) \\
&= \alpha \beta + \beta^2 \bar{x} + \beta \lambda \sigma \frac{\sqrt{2}}{\sqrt{\pi}} - \beta \alpha - \beta^2 \bar{x} - \beta \lambda \sigma \sqrt{\frac{2}{\pi}} = 0.
\end{aligned}$$

Ejemplo simulado

Suponga el modelo de regresión dado en la ecuación (1.3), con la finalidad de ilustrar el procedimiento de estimación, para cada uno de los parámetros en el modelo, vía la maxi-

mización de la log verosimilitud; así como por el método de mínimos cuadrados, se simulan muestras de tamaño 200. El procedimiento de simulación se basa en la siguiente proposición, debida a Henze (1986).

Proposición: Si X_0 y X_1 son variables independientes $N(0, 1)$ entonces

$$Y = \frac{\delta}{\sqrt{1 + \delta^2}} |X_0| + \frac{1}{\sqrt{1 + \delta^2}} X_1$$

es una variable que distribuye $SN(0, 1, \delta)$.

Veamos la demostración de esta proposición.

Demostración. $X_0 \sim N(0, 1)$ la función generadora de momentos (f.g.m.) de $|X_0|$ es,

$$\begin{aligned} Ee^{t|X_0|} &= \int_0^\infty e^{tx} 2\phi(u) du \\ &= 2e^{\frac{1}{2}t^2} \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}(u-t)^2} du \\ &= 2e^{\frac{1}{2}t^2} \frac{1}{\sqrt{2\pi}} \int_{-t}^\infty e^{-\frac{1}{2}u^2} du \\ &= 2e^{\frac{1}{2}t^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{1}{2}u^2} du \\ &= 2e^{\frac{1}{2}t^2} \Phi(t), \end{aligned}$$

ahora, si $X_1 \sim N(0, 1)$ entonces su f.g.m. es

$$Ee^{tX_1} = e^{\frac{1}{2}t^2},$$

por lo tanto la f.g.m. de $Z = \frac{\delta}{\sqrt{1+\delta^2}} |X_0| + \frac{1}{\sqrt{1+\delta^2}} X_1$ está dada por

$$\begin{aligned} M_Z(t) &= M_{\frac{\delta}{\sqrt{1+\delta^2}} |X_0| + \frac{1}{\sqrt{1+\delta^2}} X_1}(t) \\ &= M_{|X_0|} \left(\frac{\delta t}{\sqrt{1 + \delta^2}} \right) M_{X_1} \left(\frac{t}{\sqrt{1 + \delta^2}} \right) \\ &= 2\Phi \left(\frac{\delta t}{\sqrt{1 + \delta^2}} \right) e^{\frac{1}{2}t^2}, \end{aligned}$$

la cual corresponde a la f.g.m. de una $SN(0, 1, \delta)$.

Para simular números aleatorios con distribución $SN(\mu, \sigma, \delta)$ tomamos $Z = \mu + \sigma Y$ con Y simulado a partir de la proposición anterior.

Una vez simuladas las muestras de tamaño 200 se calculan los estimadores de α , β , σ^2 y δ por medio de la maximización directa de la función de log verosimilitud mediante el procedimiento `nlmin` de SPLUS. Recordemos que en el análisis de regresión, mínimos cuadrados tratan de estimar la media y la varianza de la variable aleatoria Y , en nuestro caso particular tenemos que dicha media y varianza están dadas en las ecuaciones (1.11). En este sentido calculamos los estimadores de α^* , β y σ_y^2 por medio de mínimos cuadrados, con la finalidad de compararlos con los valores de los estimadores obtenidos por máxima verosimilitud.

El conjunto de datos simulados se generó bajo el siguiente escenario:

$$\alpha = 1; \beta = 0.3; \delta = 10; \sigma = 2; n = 200,$$

considerando los valores anteriores y de las ecuaciones (1.11) tenemos que,

$$\alpha^* = \alpha + \lambda\sigma\sqrt{\frac{2}{\pi}} = 2.588 \text{ y } \sigma_y^2 = \sigma^2 \left(1 - \frac{2}{\pi}\lambda^2\right) = 1.4787.$$

La tabla siguiente resume las estimaciones de los parámetros por MV y MC.

Parámetros	α	β	δ	σ	α^*	σ_y^2
Simulación	1	0.3	10	2	2.588	1.48
MC o MV (Dist. Normal)		0.299			2.62	1.53
MV (Dist. Normal Sesgada)	1.01	0.299	11.7	2.01	2.6	1.5

La figura 1.1 presenta el comportamiento de los errores en el modelo,

El comportamiento gráfico de los residuales obtenidos mediante mínimos cuadrados, así como los residuales obtenidos por máxima verosimilitud con respecto al parámetro de localización, se muestra en la figura 1.2.

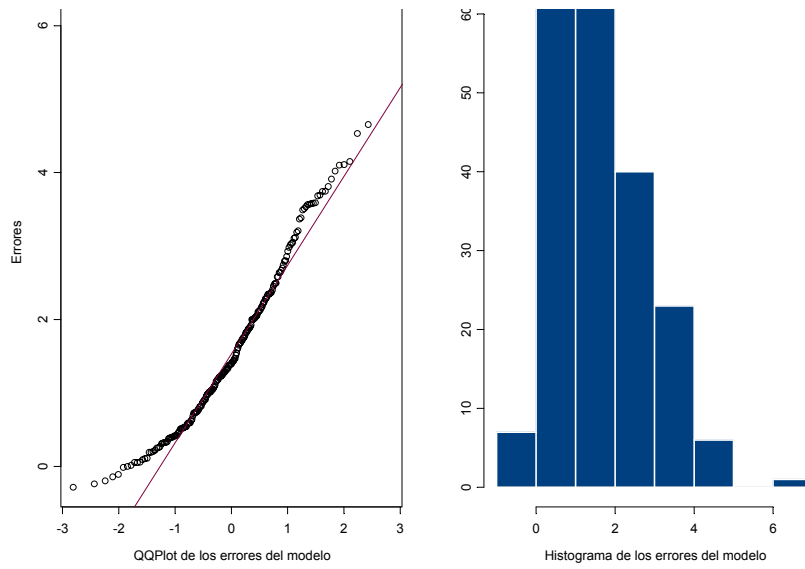


Figura 1.1:

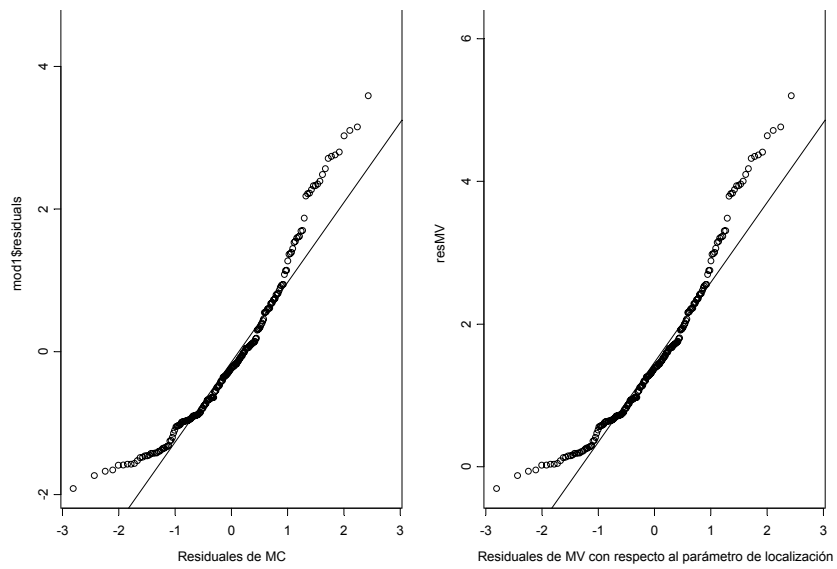


Figura 1.2:

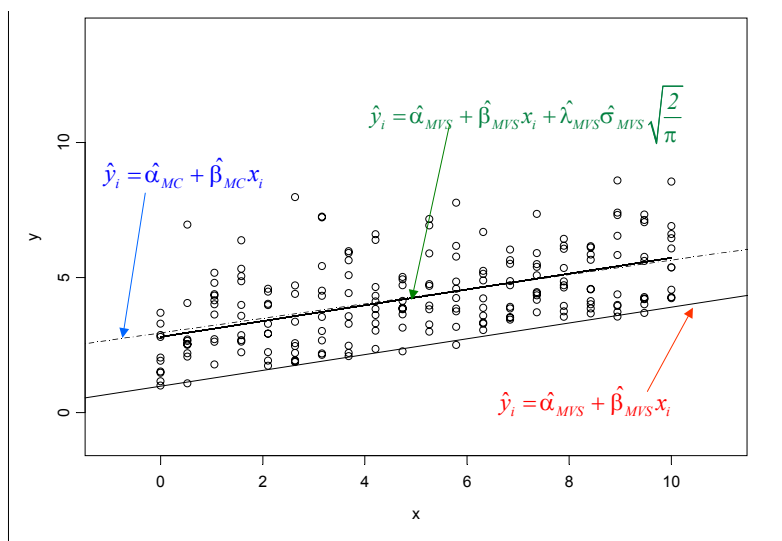


Figura 1.3:

La figura 1.3 presenta la gráfica de dispersión de los datos simulados, con las correspondientes ecuaciones de las rectas ajustadas por MC y MV; recordemos que el procedimiento de MV considera que la distribución de los errores es la normal sesgada.

En la mayoría de los escenarios simulados, los estimadores de la pendiente por el método de mínimos cuadrados y máxima verosimilitud resultaron muy parecidos numéricamente, prácticamente las ecuaciones de las rectas estimadas resultaban ser casi paralelas, lo cual se debe que el estimador de la pendiente por MC resulta ser un estimador insesgado y consistente del parámetro β en el modelo. También se observa que los valores estimados para la varianza son muy cercano al valor verdadero, independientemente del comportamiento del resto de los valores estimados.

El comportamiento de los residuos (por MV y MC) parece ser normal sesgada general (GNS), pero aún no se ha demostrado dicho resultado, ver González-Farías *et al* (2002).

Conclusiones

En este capítulo observamos que las propiedades distribucionales de los estimadores de mínimos cuadrados de la pendiente y el intercepto del modelo de regresión lineal con errores normales sesgados son similares a los del caso en que los errores son normales. Por ejemplo se observó lo siguiente:

1. El estimador de la pendiente, $\hat{\beta}_{MC}$, es un estimador insesgado de β . Esta propiedad del estimador de la pendiente puede ser de gran utilidad, por ejemplo dicho valor puede considerarse como valor inicial para el EMV de β , o bien como un valor dado en la función verosimilitud.
2. Los estimadores $\hat{\beta}_{MC}$ y $\hat{\beta}_{MV}$ son aproximadamente iguales (observación empírica).
3. El estimador $\hat{\beta}_{MC}$ es consistente al igual que $\hat{\beta}_{MV}$ (por esta razón no es sorprendente que $\hat{\beta}_{MC} \cong \hat{\beta}_{MV}$), sólo garantizamos consistencia para δ conocido.
4. La media muestral, \bar{y} , no está correlacionada con $\hat{\beta}_{MC}$, *i.e.*, $\text{Cov}(\bar{y}, \hat{\beta}_{MC}) = 0$. Cabe señalar que no son independientes ya que la independencia sólo se tiene en el caso normal, por lo que cuando $\delta = 0$ son independientes.
5. Los estimadores $\hat{\sigma}_{MV}^2$ y $\hat{\sigma}_{MC}^2$ presentan expresiones similares. Dichos estimadores son iguales en $\delta = 0$.
6. La distribución de $\hat{\alpha}_{MC}$, $\hat{\beta}_{MC}$ es normal sesgada general, como se menciona más adelante.

El estimador de mínimos cuadrados para el intercepto, $\hat{\alpha}_{MC}$, no es insesgado, en realidad

$$E(\hat{\alpha}_{MC}) = \alpha + \lambda\sigma\sqrt{\frac{2}{\pi}},$$

donde $\lambda = \frac{\delta}{\sqrt{1+\delta^2}}$, más aún

$$\hat{\alpha}_{MC} \xrightarrow{p} \alpha + \lambda\sigma\sqrt{\frac{2}{\pi}},$$

el estimador de mínimos cuadrados del intercepto nunca será bueno (si se desconoce los valores de δ y σ). Si δ y σ son conocidos se puede estimar α mediante

$$\bar{\alpha} = \hat{\alpha}_{MC} - \lambda\sigma\sqrt{\frac{2}{\pi}},$$

el cual es consistente.

El EMV, $\hat{\sigma}$, de σ tiene la misma forma que el EMV de $\hat{\sigma}_0$ cuando $\delta = 0$ (caso normal), como se ilustra en la siguiente expresión.

$$\hat{\sigma}_{MV}^2 = \frac{1}{n} \sum \left(y_i - \hat{\alpha}_{MV} - \hat{\beta}_{MV}x_i \right)^2 \text{ y } \hat{\sigma}_0^2 = \frac{1}{n} \sum \left(y_i - \hat{\alpha}_{MC} - \hat{\beta}_{MC}x_i \right)^2.$$

El MV de σ tiene la misma interpretación que en el caso de MC (mide las desviaciones alrededor de la media).

Dado de que los estimadores de mínimos cuadrados de α y β son combinaciones lineales de las observaciones, se desprende que su distribución es normal sesgada general. Lo anterior se debe a que la distribución normal sesgada es cerrada bajo combinaciones lineales de rango completo por renglón o por columna como se prueba en Gonzalez-Farías, *et al* (2002).

La función de verosimilitud perfil de δ es muy irregular, puede tener múltiples máximos incluyendo entre ellos a $\delta = \pm\infty$; $\delta = 0$ es siempre una raíz pero no necesariamente un punto que maximice la verosimilitud. En este caso es recomendable seguir las recomendaciones de Azzalini (1985), Arnold *et al* (1993), Copas y Li (1997). Dichas recomendaciones consisten en estudiar las propiedades de los parámetros α, β y σ , para valores de δ con altos niveles de verosimilitud perfil. Domínguez-Molina y González-Farías (2002) recomiendan niveles de verosimilitud perfil altos (*e.g.*, ≥ 0.5); también observan que la distribución de $R = -2 \ln \left(L_p \left(\hat{\delta} \right) - L_p \left(\delta \right) \right)$ dista mucho de la distribución *ji*-cuadrada, esto sucede incluso para muestras de tamaño 500. Azzalini (1985) recomienda muestras de tamaño grande y sugiere $n \geq 30$, pero Arnold *et al* (1993) dejan claro que n debe ser más grande, $n \geq 200$.

Capítulo 2

Inferencias para muestras no-aleatorias (Copas y Li 1997)

Introducción

Dentro de las contribuciones más importantes de R. Fisher se encuentran los estudios realizados sobre la importancia de aleatorización, no sólo en las áreas de diseños de experimentos y muestreo, sino también como soporte lógico de los distintos métodos de análisis estadístico. El argumento esencial de Fisher es que la aleatorización o equivalentemente el supuesto de muestreo de una población, justifica el uso de las distintas pruebas estándar de significancia, así como el correcto uso de los métodos de inferencia normal. Sin embargo los métodos diseñados para análisis de datos experimentales se aplican también de manera rutinaria a datos observacionales sin que estos reconozcan la ausencia de aleatoriedad desde el punto de vista de Fisher, eliminando así la base sólida para la correcta aplicación de estos métodos.

Una gran parte de la estadística moderna se desarrolla en torno a pruebas y verificaciones de los supuestos en los distintos modelos, sin embargo el argumento de aleatoriedad, que da soporte a los supuestos estándar del modelo no es reversible. Es decir, la verificación empírica de los supuestos en un modelo no implican necesariamente que se justifique la propiedad de aleatoriedad. Algunas características de interés de datos observacionales, tales como diferencia significativa entre respuestas de sujetos ante diferentes tratamientos, pueden ser bien justificadas o explicadas por la asimetría incurrida al momento de que dichos sujetos

son asignados a cada uno de los tratamientos.

Copas y Li (1997) extienden algunos modelos estadísticos simples incluyendo un parámetro adicional θ que modela los grados de no-aleatoriedad en el mecanismo generador de los datos. El valor de $\theta = 0$ es la hipótesis de aleatoriedad, de modo tal que se supone que los datos son el resultado de un experimento aleatorio. Pensando en θ como un parámetro desconocido paralelo a los parámetros ya existentes en el modelo, se podría proceder en la forma habitual de inferencia paramétrica o bien realizar una prueba de hipótesis de modo que si $\theta = 0$ la justificación de las inferencias estándar esta dada. La forma de proceder de Copas y Li (1997) es mediante inferencia condicional en un rango de valores diferentes de θ y analizar la sensibilidad de sus conclusiones dependiendo de las características de θ cerca del valor cero. También se estudia la log-verosimilitud para θ después de maximizar en función de los otros parámetros, corroborando que mucha de la información de los datos está dada por θ .

En nuestro caso particular estudiaremos el modelo propuesto por Copas y Li (1997) observando el hecho de que la distribución condicional resultante es un caso particular de la distribución *normal sesgada general* (GSN) dada en Domínguez-Molina *et al* (2001). También se tiene que para ciertos valores de los parámetros dicha distribución contiene a la distribución de Azzalini y Dalla Valle (1996). En la parte final de este capítulo simulamos conjuntos de datos de diferentes tamaños y posteriormente estimamos los parámetros mediante la maximización de la función de log verosimilitud.

Modelo básico

Sea Y la variable respuesta de interés y relacionada con la covariable X mediante el siguiente modelo de regresión

$$Y = \beta^T X + \sigma \varepsilon_1, \quad (2.1)$$

X es un vector de m componentes y $x_1 = 1$ de modo que la primer componente del vector β es el término intercepto y ε_1 es una variable aleatoria $N(0, 1)$. El modelo principal está

completamente especificado por la ecuación de selección

$$Z = \gamma^T X + \varepsilon_2, \quad (2.2)$$

asumiendo también que ε_2 es $N(0, 1)$, de modo que $(\varepsilon_1, \varepsilon_2)$ es normal estándar bivariada con coeficiente de correlación ρ .

Algunas aplicaciones principales de este modelo son:

1. Modelos para datos faltantes (Y es observada sólo si $Z > 0$), como por ejemplo en el estudio de técnicas de muestreo ante la no-respuesta.
2. Modelos para comparación de ensayos o tratamientos (un sujeto es designado al tratamiento A si $Z > 0$ o al tratamiento B si $Z \leq 0$), como por ejemplo en el estudios de pruebas t -pareadas, comparación de dos muestras, entre otras.

Se supone que la covariable X es fija y observada, mientras que los valores actuales de la variable Z nunca son observados, pero siempre se conoce si dicha variable es positiva o negativa.

La distribución de $(Y, Z)^T$ se obtiene fácilmente dado que la distribución de los errores es

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim N_2 \left[0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right],$$

y por (2.1) y (2.2) obtenemos que

$$\begin{pmatrix} Y \\ Z \end{pmatrix} \sim N_2 \left[\begin{pmatrix} \beta^T X \\ \gamma^T X \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma\rho \\ \sigma\rho & 1 \end{pmatrix} \right].$$

La función de densidad condicional de $y|x, z > 0$ se obtiene directamente utilizando la relación,

$$f(y|x, z > 0) = \frac{f_Y(y) \Pr(Z > 0|y)}{\Pr(Z > 0)}. \quad (2.3)$$

Es conocido que la variable aleatoria $Z|y$ tiene distribución normal dada por

$$Z|y \sim N\left(\gamma^T X + \frac{\rho}{\sigma}(y - \beta^T X), 1 - \rho^2\right),$$

de modo que,

$$\Pr(Z \leq z|y) = \Phi\left(\frac{z - \gamma^T X - \rho\sigma^{-1}(y - \beta^T X)}{\sqrt{1 - \rho^2}}\right),$$

o bien,

$$\Pr(Z > 0|y) = \Pr(-Z \leq 0|y) = \Phi\left(\frac{\gamma^T X + \rho\sigma^{-1}(y - \beta^T X)}{\sqrt{1 - \rho^2}}\right).$$

Regresando a la densidad $f(y|x, z > 0)$ dada en la ecuación (2.3) obtenemos que

$$f(y|x, z > 0) = \frac{\sigma^{-1}\phi\left(\frac{y - \beta^T X}{\sigma}\right)\Phi\left(\frac{\gamma^T X + \rho\sigma^{-1}(y - \beta^T X)}{\sqrt{1 - \rho^2}}\right)}{\Phi(\gamma^T X)}. \quad (2.4)$$

Con fines de aplicación Copas y Li (1997) proponen reparametrizar la densidad condicional $f(y|x, z > 0)$, obteniendo la siguiente representación equivalente,

$$f(y|x, z > 0) = \sigma^{-1}\Phi^{-1}(\gamma^T x)\phi\left\{\left(\frac{y - \beta^T x}{\sigma}\right)\right\}\Phi\left\{(1 + \theta^2)^{1/2}\gamma^T x + \theta\left(\frac{y - \beta^T x}{\sigma}\right)\right\}, \quad (2.5)$$

$$\theta = \frac{\rho}{(1 - \rho^2)^{1/2}}.$$

Copas y Li (1997) identifican los siguientes casos particulares del modelo dado en las ecuaciones (2.1) y (2.2).

1. Si $\rho = 0$ la expresión (2.4) corresponde a la densidad marginal normal de Y en el modelo dado en las ecuaciones (2.1) y (2.2).
2. El caso más simple de este modelo es $\beta = \gamma = 0$ y $\sigma = 1$, es decir el caso en que Y y Z son normales estándar bivariada con coeficiente de correlación ρ , de modo que $f(y|z > 0) = 2\phi(y)\Phi\left(\frac{\rho}{\sqrt{1 - \rho^2}}y\right)$ es la distribución normal sesgada de Azzalini (1985).

Sin embargo notemos que si $\gamma = 0$ y por notación $\phi(y; \beta^T X, \sigma^2) = \sigma^{-1} \phi\left(\frac{y - \beta^T X}{\sigma}\right)$, la densidad resultante corresponde a la distribución de Azzalini y Dalla Valle (1996),

$$\begin{aligned} f(y|x, z > 0) &= \frac{\sigma^{-1} \phi\left(\frac{y - \beta^T X}{\sigma}\right) \Phi\left(\frac{\rho}{\sqrt{1 - \rho^2}} \left(\frac{y - \beta^T X}{\sigma}\right)\right)}{\Phi(0)} \\ &= 2\phi(y; \beta^T X, \sigma^2) \Phi\left(\frac{\rho}{\sqrt{1 - \rho^2}} \left(\frac{y - \beta^T X}{\sigma}\right)\right). \end{aligned} \quad (2.6)$$

Mientras que si $\gamma \neq 0$ la densidad resultante es la densidad dada en Domínguez-Molina, *et al* (2001), ya que

$$\begin{aligned} f(y|x, z > 0) &= \Phi^{-1}(\gamma^T X) \sigma^{-1} \phi\left(\frac{y - \beta^T X}{\sigma}\right) \Phi\left(\frac{\gamma^T X + \frac{\rho}{\sigma}(y - \beta^T X)}{\sqrt{1 - \rho^2}}\right) \\ &= \Phi^{-1}(\gamma^T X) \phi(y; \beta^T X, \sigma^2) \Phi\left(\frac{\frac{\rho}{\sigma}y - \left(\frac{\rho}{\sigma}\beta^T + \gamma^T\right)X}{\sqrt{1 - \rho^2}}\right). \end{aligned} \quad (2.7)$$

Usando el hecho de que $\Phi_p(x; \mu, \Sigma) = \Pr(X \leq x)$, donde $X \sim N_p(\mu, \Sigma)$ y $\Phi_q(Dy; \nu, \Delta) = \Pr(X \leq Dy)$, con $X \sim N_p(\nu, \Delta)$, tenemos que

$$\begin{aligned} \Phi_1\left(\frac{\rho}{\sigma}\beta^T X; \left(\frac{\rho}{\sigma}\beta^T - \gamma^T\right)X, 1\right) &= \Pr\left(U \leq \frac{\rho}{\sigma}\beta^T X\right), U \sim N\left(\frac{\rho}{\sigma}\beta^T X - \gamma^T X, 1\right) \\ &= \Pr\left(U - \left(\frac{\rho}{\sigma}\beta^T X - \gamma^T X\right) \leq \frac{\rho}{\sigma}\beta^T X - \left(\frac{\rho}{\sigma}\beta^T X - \gamma^T X\right)\right) \\ &= \Pr\left(U - \frac{\rho}{\sigma}\beta^T X + \gamma^T X \leq \gamma^T X\right) \\ &= \Pr\left(Z \leq \gamma^T X\right), Z \sim N(0, 1) \\ &= \Phi(\gamma^T X). \end{aligned}$$

Así mismo,

$$\begin{aligned} \Phi_1\left(\frac{\rho}{\sigma}y; \left(\frac{\rho}{\sigma}\beta^T + \gamma^T\right)X, 1 - \rho^2\right) &= \Pr\left(U \leq \frac{\rho}{\sigma}y\right), U \sim N\left(\frac{\rho}{\sigma}\beta^T X - \gamma^T X, 1 - \rho^2\right) \\ &= \Pr\left(\frac{U - \left(\frac{\rho}{\sigma}\beta^T X - \gamma^T X\right)}{\sqrt{1 - \rho^2}} \leq \frac{\frac{\rho}{\sigma}y - \left(\frac{\rho}{\sigma}\beta^T X - \gamma^T X\right)}{\sqrt{1 - \rho^2}}\right) \\ &= \Pr\left(Z \leq \frac{\frac{\rho}{\sigma}y - \left(\frac{\rho}{\sigma}\beta^T X - \gamma^T X\right)}{\sqrt{1 - \rho^2}}\right), Z \sim N(0, 1) \\ &= \Phi\left(\frac{\frac{\rho}{\sigma}y - \left(\frac{\rho}{\sigma}\beta^T X - \gamma^T X\right)}{\sqrt{1 - \rho^2}}\right). \end{aligned}$$

De las expresiones anteriores podemos reexpresar la ecuación (2.7) como,

$$f(y|x, z > 0) = \Phi_1^{-1} \left(\frac{\rho}{\sigma} \beta^T X; \frac{\rho}{\sigma} \beta^T X - \gamma^T X, 1 \right) \\ \times \phi \left(y; \beta^T X, \sigma^2 \right) \Phi_1 \left(\frac{\rho}{\sigma} y; \left(\frac{\rho}{\sigma} \beta^T - \gamma^T \right) X, 1 - \rho^2 \right),$$

la cual corresponde a la distribución de una variable aleatoria con distribución GSN tal y como la proponen en Domínguez-Molina, *et al* (2001), y cuya notación es,

$$Y \sim GSN_{1,1} \left(\beta^T X, \sigma^2, \frac{\rho}{\sigma}, \left(\frac{\rho}{\sigma} \beta^T - \gamma^T \right) X, 1 - \rho^2 \right).$$

Cálculos de la media y varianza

Para calcular los momentos de la variable $Y|x, z > 0$ utilizamos la función generatriz de momentos de una variable aleatoria $W \sim GSN_{p,q}(\mu, \Sigma, D, \nu, \Delta)$ dada en Domínguez-Molina *et al* (2001), cuya expresión es la siguiente:

$$M_W(t) = \frac{\Phi_1(D(\mu + \Sigma t); \nu, \Delta + D\Sigma D^T)}{\Phi_1(D\mu; \nu, \Delta + D\Sigma D^T)} e^{\mu^T t + \frac{1}{2} t^T \Sigma t}, \quad t \in \mathfrak{R}^p,$$

donde $\nu \in \mathfrak{R}^p$, $\mu \in \mathfrak{R}^p$, D es una matriz de constantes de $q \times p$, Δ ($q \times q$) y Σ ($p \times p$) son matrices de covarianzas. En nuestro caso particular con $p = q = 1$, $\mu = \beta^T X$, $\Sigma = \sigma^2$, $D = \frac{\rho}{\sigma}$, $\nu = \left(\frac{\rho}{\sigma} \beta^T - \gamma^T \right) X$, $\Delta = 1 - \rho^2$ y $W = Y|x, z > 0$ podemos reexpresar la f.g.m. como

$$M_W(t) = \frac{\Phi_1 \left(\frac{\rho}{\sigma} (\beta^T X + \sigma^2 t); \left(\frac{\rho}{\sigma} \beta^T - \gamma^T \right) X, 1 \right)}{\Phi_1 \left(\frac{\rho}{\sigma} \beta^T X; \left(\frac{\rho}{\sigma} \beta^T - \gamma^T \right) X, 1 \right)} e^{(\beta^T X)^T t + \frac{1}{2} t^2 \sigma^2}, \quad t \in \mathfrak{R} \quad (2.8) \\ = \frac{\Phi \left(\frac{\rho}{\sigma} (\beta^T X + \sigma^2 t) - \left(\frac{\rho}{\sigma} \beta^T - \gamma^T \right) X \right)}{\Phi \left(\frac{\rho}{\sigma} \beta^T X - \left(\frac{\rho}{\sigma} \beta^T - \gamma^T \right) X \right)} e^{(\beta^T X)^T t + \frac{1}{2} t^2 \sigma^2} \\ = \frac{\Phi(\rho \sigma t + \gamma^T X)}{\Phi(\gamma^T X)} e^{(\beta^T X)^T t + \frac{1}{2} t^2 \sigma^2}.$$

La primer derivada de la f.g.m en (2.8) está dada por

$$\frac{\partial M_W(t)}{\partial t} = (\beta^T X + \sigma^2 t) \frac{\Phi(\rho \sigma t + \gamma^T X)}{\Phi(\gamma^T X)} e^{(\beta^T X)^T t + \frac{1}{2} t^2 \sigma^2} \\ + \frac{\Phi^*(\rho \sigma t + \gamma^T X)}{\Phi(\gamma^T X)} e^{(\beta^T X)^T t + \frac{1}{2} t^2 \sigma^2}, \quad (2.9)$$

donde,

$$\Phi^* (\rho\sigma t + \gamma^T X) = \frac{\partial \Phi (\rho\sigma t + \gamma^T X)}{\partial t} = \rho\sigma\phi (\rho\sigma t + \gamma^T X).$$

Evaluando la derivada anterior en $t = 0$ llegamos a que:

$$\Phi^* (\rho\sigma t + \gamma^T X)|_{t=0} = \left. \frac{\partial \Phi (\rho\sigma t + \gamma^T X)}{\partial t} \right|_{t=0} = \rho\sigma\phi (\gamma^T X),$$

de esta forma el valor esperado de W es,

$$\begin{aligned} E(W) &= E(Y|x, Z > 0) = \left. \frac{\partial M_W(t)}{\partial t} \right|_{t=0} \\ &= \beta^T X + \rho\sigma \frac{\phi(\gamma^T X)}{\Phi(\gamma^T X)}. \end{aligned} \quad (2.10)$$

El segundo momento de la f.g.m para W se obtiene derivando con respecto a t la ecuación (2.9), es decir,

$$\begin{aligned} \frac{\partial^2 M_W(t)}{\partial t^2} &= \frac{\partial}{\partial t} \left[\frac{(\beta^T X + \sigma^2 t) \Phi(\rho\sigma t + \gamma^T X) + \rho\sigma\phi(\rho\sigma t + \gamma^T X)}{\Phi(\gamma^T X)} e^{(\beta^T X)^T t + \frac{1}{2} t^2 \sigma^2} \right] \\ &= \frac{(\beta^T X + \sigma^2 t) \Phi(\rho\sigma t + \gamma^T X) + \rho\sigma\phi(\rho\sigma t + \gamma^T X)}{\Phi(\gamma^T X)} \\ &\quad \times (\beta^T X + \sigma^2 t) e^{(\beta^T X)^T t + \frac{1}{2} t^2 \sigma^2} \\ &\quad + \frac{\sigma^2 \Phi(\rho\sigma t + \gamma^T X) + (\beta^T X + \sigma^2 t) \rho\sigma\phi(\rho\sigma t + \gamma^T X) + \rho\sigma\phi^{\{1\}}(\rho\sigma t + \gamma^T X)}{\Phi(\gamma^T X)} \\ &\quad \times e^{(\beta^T X)^T t + \frac{1}{2} t^2 \sigma^2}, \end{aligned}$$

donde

$$\begin{aligned} \phi^{\{1\}}(\rho\sigma t + \gamma^T X) &= \frac{\partial \phi(\rho\sigma t + \gamma^T X)}{\partial t} = \frac{1}{\sqrt{2\pi}} \frac{\partial}{\partial t} \exp \left[-\frac{1}{2} (\rho\sigma t + \gamma^T X)^2 \right] \\ &= -\frac{2(\rho\sigma t + \gamma^T X)\rho\sigma}{2\sqrt{2\pi}} \exp \left[-\frac{1}{2} (\rho\sigma t + \gamma^T X)^2 \right] \\ &= -(\rho^2 \sigma^2 t + \rho\sigma \gamma^T X) \phi(\rho\sigma t + \gamma^T X). \end{aligned}$$

Finalmente el valor esperado de W^2 está dado por la segunda derivada de la f.g.m. evaluada en $t = 0$, es decir,

$$E(W^2) = \left. \frac{\partial^2 M_W(t)}{\partial t^2} \right|_{t=0}$$

$$\begin{aligned}
 &= (\beta^T X) \left(\frac{(\beta^T X) \Phi(\gamma^T X) + \rho\sigma\phi(\gamma^T X)}{\Phi(\gamma^T X)} \right) \\
 &\quad + \frac{\sigma^2\Phi(\gamma^T X) + (\beta^T X)\rho\sigma\phi(\gamma^T X) - (\rho^2\sigma^2\gamma^T X)\phi(\gamma^T X)}{\Phi(\gamma^T X)} \\
 &= (\beta^T X)^2 + \sigma^2 + \frac{\rho\sigma\phi(\gamma^T X)(2(\beta^T X) - \rho\sigma\gamma^T X)}{\Phi(\gamma^T X)}. \tag{2.11}
 \end{aligned}$$

De las ecuaciones (2.10) y (2.11) para los primeros dos momentos de la variable W se sigue que la varianza de W está dada por,

$$\begin{aligned}
 \text{Var}(W) &= EW^2 - (E(W))^2 \\
 &= \sigma^2 + (\beta^T X)^2 + \frac{\rho\sigma\phi(\gamma^T X)(2(\beta^T X) - \rho\sigma\gamma^T X)}{\Phi(\gamma^T X)} \\
 &\quad - \left(\beta^T X + \rho\sigma \frac{\phi(\gamma^T X)}{\Phi(\gamma^T X)} \right)^2 \\
 &= \sigma^2 + (\beta^T X)^2 + \frac{\rho\sigma\phi(\gamma^T X)(2(\beta^T X) - \rho\sigma\gamma^T X)}{\Phi(\gamma^T X)} \\
 &\quad - (\beta^T X)^2 - \rho^2\sigma^2 \left(\frac{\phi(\gamma^T X)}{\Phi(\gamma^T X)} \right)^2 - 2\beta^T X\rho\sigma \frac{\phi(\gamma^T X)}{\Phi(\gamma^T X)} \\
 &= \sigma^2 - \rho^2\sigma^2 \left(\frac{\phi(\gamma^T X)}{\Phi(\gamma^T X)}\gamma^T X + \left(\frac{\phi(\gamma^T X)}{\Phi(\gamma^T X)} \right)^2 \right) \\
 &= \sigma^2 \left(1 - \rho^2 \frac{\phi(\gamma^T X)}{\Phi(\gamma^T X)} \left(\gamma^T X + \frac{\phi(\gamma^T X)}{\Phi(\gamma^T X)} \right) \right). \tag{2.12}
 \end{aligned}$$

Las expresiones para el valor esperado y la varianza dadas en las ecuaciones (2.10) y (2.12) respectivamente, coinciden con las expresiones del valor esperado y la varianza dadas en Copas y Li (1997, páginas 10 y 11).

Función de verosimilitud

Considerando la función de densidad condicional dada en (2.5), se tiene que la función de verosimilitud esta dada por,

$$L(\beta, \gamma, \sigma, \theta) = \prod_{i=1}^n f(y_i | x_i, z_i > 0)$$

$$= \prod_{i=1}^n \left\{ \sigma^{-1} \Phi^{-1}(\gamma^T x_i) \phi\left(\frac{y_i - \beta^T x_i}{\sigma}\right) \Phi\left[(1 + \theta^2)^{1/2} \gamma^T x_i + \theta \left(\frac{y - \beta^T x_i}{\sigma}\right)\right] \right\},$$

y de aquí que la función de log verosimilitud toma la siguiente forma,

$$\ell(\beta, \gamma, \sigma, \theta) = \sum_{i=1}^n \log \left[\frac{\phi\left(\frac{y_i - \beta^T x_i}{\sigma}\right) \Phi\left[(1 + \theta^2)^{1/2} \gamma^T x_i + \theta \left(\frac{y - \beta^T x_i}{\sigma}\right)\right]}{\sigma \Phi(\gamma^T x_i)} \right]. \quad (2.13)$$

Es posible mostrar que si fijamos θ a un valor específico y restringimos al resto de los parámetros a ser soluciones de las ecuaciones de la log verosimilitud, la matriz Hessiano de la log verosimilitud es siempre negativa definida y las estimaciones $\hat{\beta}_\theta, \hat{\gamma}_\theta, \hat{\sigma}_\theta$ son definidos de manera única para todo θ . A partir de lo anterior se define la log verosimilitud perfil dada por:

$$L_p(\theta) = \max_{\beta, \gamma, \sigma} \ell(\beta, \gamma, \sigma, \theta) = \ell\left(\hat{\beta}_\theta, \hat{\gamma}_\theta, \hat{\sigma}_\theta, \theta\right). \quad (2.14)$$

Observación. Cabe mencionar que el procedimiento de estimación dado en Copas y Li (1997) no es términos de la función de verosimilitud, sino mediante un procedimiento en dos etapas. La primer etapa consiste en estimar el parámetro γ notando aquellos casos en que $z > 0$, lo cual se realiza mediante un análisis probit. Una vez estimado γ se calcula $\phi(\gamma^T x) / \Phi(\gamma^T x)$ para cada uno de los casos en que $z > 0$, luego se supone a está como una nueva covariable en la ecuación (2.10) y se hace un ajuste por mínimos cuadrados. El coeficiente de la covariable adicional da un estimador de $\rho\sigma$. De la expresión para la varianza (2.12) se toma el valor promedio del lado derecho de la ecuación y se iguala este con la varianza residual observada, para luego encontrar los estimadores de ρ y σ .

El procedimiento de estimación en nuestro caso consiste en maximizar la función de log verosimilitud mediante la función `nlmin` de S-PLUS.

Ejemplo simulado

A manera de ilustración a continuación presentamos un conjunto de 80 datos simulados a partir del modelo dado en las ecuaciones (2.1) y (2.2), el procedimiento de simulación se codificó en S-PLUS. Una vez simulados los datos implementamos el procedimiento de estimación de los parámetros por medio de la maximización de la función log de verosimilitud dada en la ecuación (2.13), mediante la función `nlmin` de S-PLUS. Los valores de cada uno de los parámetros en el modelo para el ejercicio de simulación fueron: $\beta^T = (3, 1)$, $\sigma = 1.5$, $\gamma = (-1, 0.3)$ y $\rho = 0.5$. Los valores de la variable x , fija, son tales que $x \in [0, 10]$. En el apéndice presentamos algunos comportamientos gráficos para diferentes tamaños de muestra.

Los valores estimados mediante máxima verosimilitud son: $\hat{\beta}^T = (4.7988745, 0.5997101)$, $\hat{\sigma} = 1.5118585$, $\hat{\gamma}^T = (3.1444591, -0.6116600)$ y $\hat{\rho} = 0.48$. En nuestro ejercicio de simulación observamos que los parámetros β y σ son más fáciles de identificar que el parámetro γ . En el caso del parámetro ρ tenemos en términos generales que para muestras ≥ 60 el valor estimado tiende a ser más cercano al valor verdadero.

La figura 2.1 muestra el diagrama de dispersión para una muestra simulada de tamaño 80, dicha gráfica contiene también la recta ajustada mediante máxima verosimilitud y la recta ajustada por mínimos cuadrados. Notamos que a diferencia del modelo de Azzalini simulado en el Capítulo 1, las rectas ajustadas por MV y MC no presentan el comportamiento de líneas paralelas. En la figura se presenta también el histograma de la variable simulada Y , la cual presenta un comportamiento asimétrico, como se esperaba.

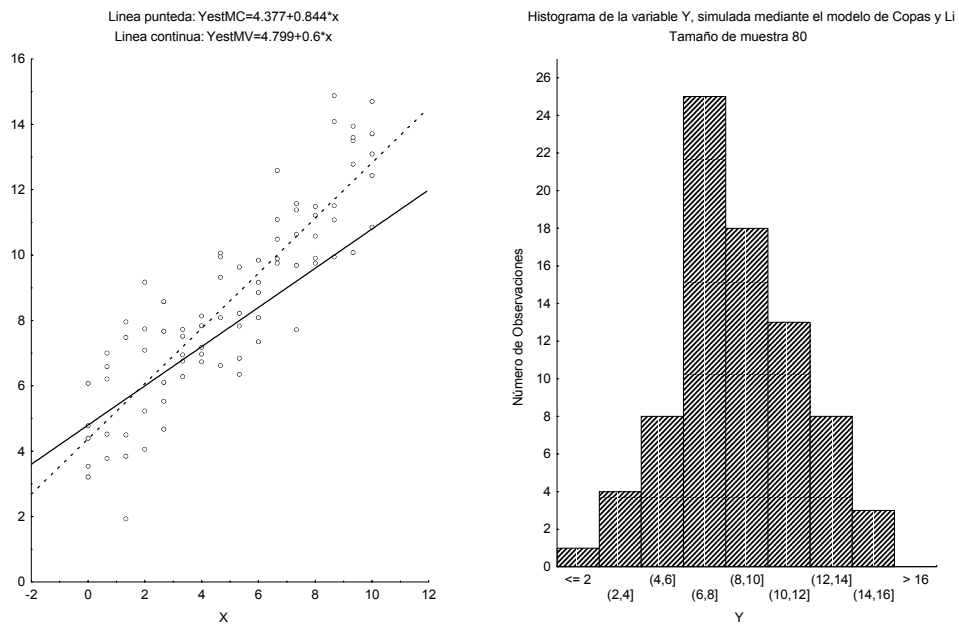


Figura 2.1:

Finalmente presentamos la tabla siguiente con los 80 valores simulados obtenidos para cada una de las variables en el modelo.

x	y	x	y	x	y
0.0000000	4.400677	7.3333333	7.722377	4.0000000	8.141384
0.6666667	6.591787	8.0000000	9.754369	4.6666667	8.082771
1.3333333	3.844622	8.6666667	11.081214	5.3333333	7.826751
2.0000000	5.229667	9.3333333	13.939849	6.0000000	7.345705
2.6666667	5.528669	10.000000	13.094434	6.6666667	11.093030
3.3333333	7.520036	0.0000000	3.212593	7.3333333	10.636816
4.0000000	7.176000	0.6666667	4.522847	8.0000000	11.216275
4.6666667	10.061868	1.3333333	7.483263	8.6666667	14.085211
5.3333333	9.636994	2.0000000	4.063920	9.3333333	10.078037
6.0000000	9.166127	2.6666667	7.668722	10.000000	12.435423
6.6666667	9.890149	3.3333333	6.949807	0.0000000	3.536965
7.3333333	9.686267	4.0000000	7.835338	0.6666667	7.002053
8.0000000	11.485838	4.6666667	6.620128	1.3333333	1.930482
8.6666667	11.511959	5.3333333	8.221621	2.0000000	7.745039
9.3333333	13.507558	6.0000000	9.844239	2.6666667	8.578018
10.00000	14.698816	6.6666667	10.481713	3.3333333	6.280803
0.0000000	6.073121	7.3333333	11.570599	4.0000000	6.970834
0.6666667	6.205289	8.0000000	9.908349	4.6666667	9.941781
1.3333333	4.496418	8.6666667	14.880520	5.3333333	6.845918
2.0000000	9.177568	9.3333333	13.601589	6.0000000	8.848887
2.6666667	4.676151	10.000000	13.714419	6.6666667	12.593834
3.3333333	7.725052	0.0000000	4.774932	7.3333333	11.385669
4.0000000	6.732898	0.6666667	3.774572	8.0000000	10.580797
4.6666667	9.327970	1.3333333	7.956836	8.6666667	9.944766
5.3333333	6.349635	2.0000000	7.087839	9.3333333	12.784097
6.0000000	8.088190	2.6666667	6.107136	10.000000	10.852067
6.6666667	9.753964	3.3333333	6.761192		

Conclusiones

Las distribuciones de Azzalini (1985) y la distribución normal sesgada general (GSN) de Domínguez-Molina, *et al* (2001) han sido poco exploradas en el terreno aplicado, en gran parte por el desconocimiento de posibles casos de estudio en los que dichas distribuciones se apliquen de manera natural. Sin embargo en el artículo de Copas y Li (1997) se presenta una gama amplia de posibles aplicaciones, principalmente en el campo de investigaciones biomédicas. El hecho de que la distribución condicional de Copas y Li (1997) sea un caso particular de la distribución GSN nos abre las puertas para la búsqueda de aplicaciones a fenómenos reales de dicha distribución. Por supuesto que hace falta un mayor refinamiento en el proceso de estimación de los parámetros ya que estos son muy inestable numéricamente hablando.

El procedimiento de estimación implementado en Copas y Li (1997) no es en términos de la función de verosimilitud, sino mediante un procedimiento en dos etapas. La primer etapa consiste en estimar el parámetro γ con el signo de z , lo cual se realiza mediante un análisis probit. Una vez estimado γ se calcula $\phi(\gamma^T x) / \Phi(\gamma^T x)$ para cada uno de los casos en que $z > 0$, luego se supone a está como una nueva covariable en la ecuación (2.10) y se hace un ajuste por mínimos cuadrados. El coeficiente de la covariable adicional da un estimador de $\rho\sigma$. De la expresión para la varianza (2.12) se toma el valor promedio del lado derecho de la ecuación y se iguala este con la varianza residual observada, para luego encontrar los estimadores de ρ y σ .

Los EMV de $\beta \in \mathfrak{R}^2$ y σ son estables numéricamente y el EMV de δ también es estable numéricamente pero para muestras grandes, en general muestras de tamaños mayor que 60. Sin embargo, el estimador de $\gamma \in \mathfrak{R}^2$ es muy inestable, casi siempre da valores alejados del verdadero.

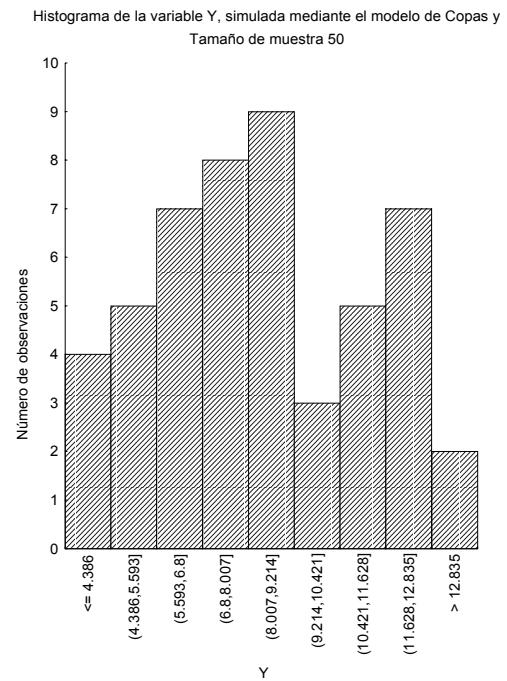
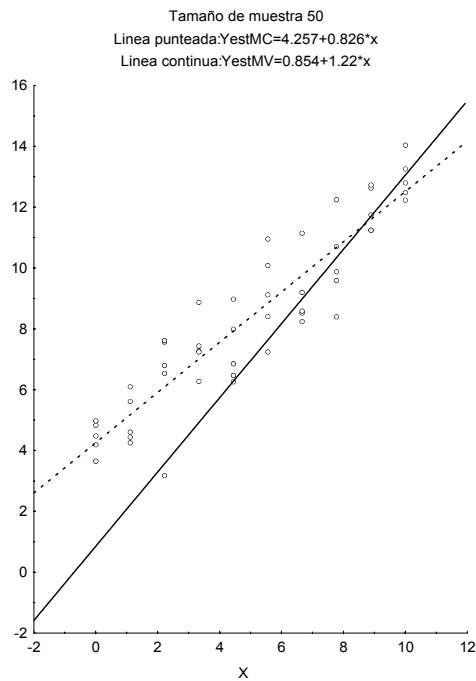
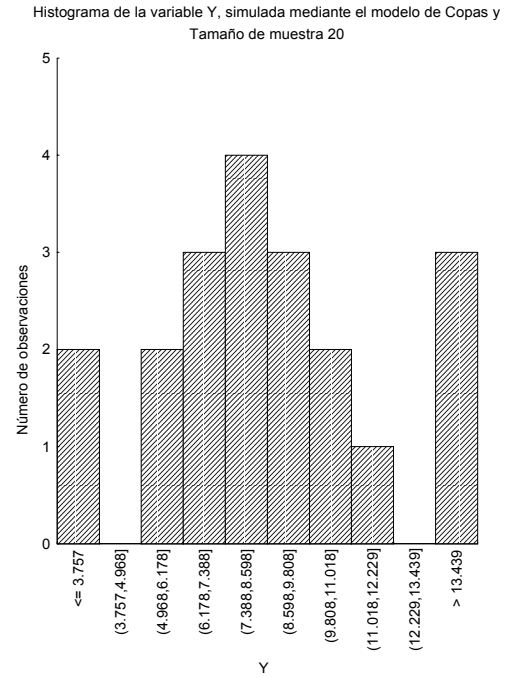
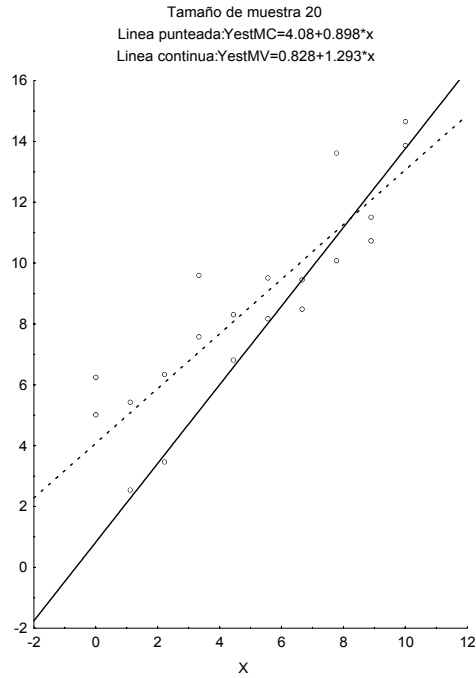
En este capítulo no se discutieron las propiedades de los estimadores de mínimos cuadrados de la recta con errores normales sesgados generales. Pero es fácil probar que la dis-

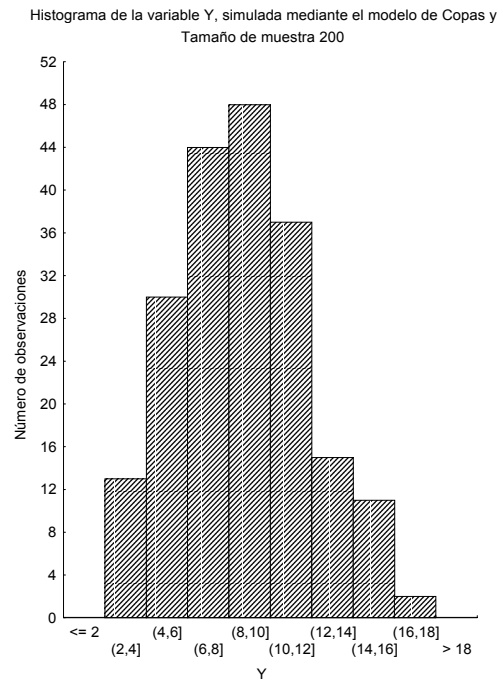
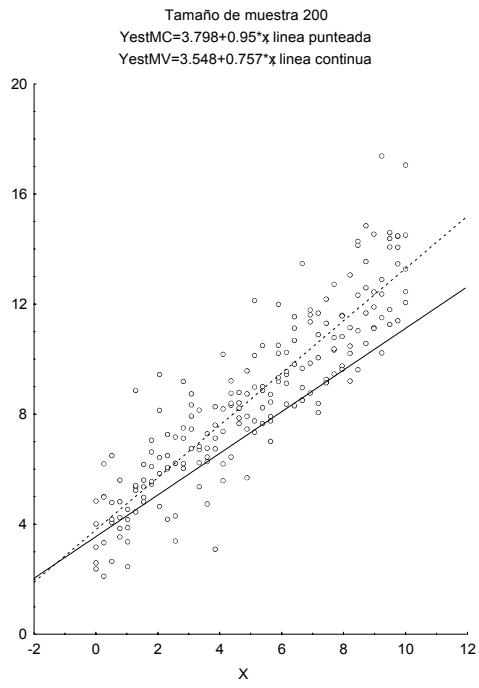
tribución de los estimadores de mínimos cuadrados es también normal sesgada general (ver González-Farías, *et al* 2002).

El modelo de Copas y Li (1997) asigna varianza igual a uno a la variable latente z , esta restricción es fácil de eliminar en aplicaciones posteriores.

Las rectas de mínimos cuadrados y la de máxima verosimilitud son notoriamente no paralelas, a diferencia con el modelo de la normal sesgada de Azzalini del capítulo anterior. En este caso podemos pensar que los estimadores de mínimos cuadrados no son útiles en general.

Apéndice





Bibliografía

- [1] Arnold, B.C., Beaver, R.J., Groeneveld, R.A. & Meeker, W.Q. (1993). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika*, **58**, 3, 471-488.

- [2] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*. **12**, 171-178.

- [3] Azzalini, A. and Capitanio, A.(1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society, B*, **61**, 579-602.

- [4] Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.

- [5] Copas, J.B. & Li, H.G. (1997). Inference for Non-random Samples with disjunction. *Journal of the Royal Statistical Society, B*, **59**, 1, 55-95.

- [6] Domínguez-Molina, J. A., González-Farías, G. and Gupta, A.K. (2001) A General Multivariate Skew Normal Distribution. *Department of Mathematics and Statistics, Bowling Green State University, Technical Report No. 01-09*.

- [7] Domínguez-Molina, J. A. and González-Farías, G. (2002). An Optimal procedure for estimation in the univariate skew normal distribution. *Lehmann Symposium, CIMAT, Guanajuato, Gto. 2002*.

- [8] González-Farías, G., Domínguez-Molina, J.A. and Gupta, A.K. (2002). Additive prop-

erties of skew normal random vectors. Por aparecer en *Journal of Statistical Planning and Inference (JSPI)*.

- [9] Henze, N. (1986). A probabilistic representation of the “skew-normal” distribution. *Scandinavian Journal of Statistics*, **13**, 271-275.
- [10] Marsden, J. E. y Hoffman, M. J. (1998). *Análisis Clásico Elemental*. Addison-Wesley Iberoamericana.
- [11] S-PLUS 2000 Professional Release 1. Copyright (c) 1988-1999 MathSoft, Inc.