

MODELO LINEAL CENSURADO CON DISTRIBUCIÓN NORMAL Y T-STUDENT

MARÍA CAMILA GONZÁLEZ REINA

Trabajo de grado

Profesor Asesor
Victor Hugo Lachos Dávila
Coordinador de curso graduación en Estadística, UNICAMP, SP-Brasil

UNIVERSIDAD EL TOLIMA
FACULTAD DE CIENCIAS
PROGRAMA EN MATEMATICAS CON ENFASIS EN ESTADISTICA
IBAGUE-TOLIMA
2013

Índice general

1. Preliminares	5
1.1. Introducción	5
1.2. Marco teórico	5
1.2.1. Modelo De Regresión Lineal Con Respuesta Censurada	6
1.2.2. Algoritmo Newton-Raphson	7
1.2.3. Algoritmo EM	7
2. Distribución Normal	10
2.1. Newton-Raphson	10
2.2. Algoritmo EM	12
3. Distribución t-student	17
3.1. Newton-Raphson	17
3.2. Algoritmo EM	19
4. Aplicación	22

Índice de cuadros

4.1. Datos Tormin 2000	22
4.2. Resultados obtenidos para la estimación de los parámetros vía N-R	23
4.3. Resultados obtenidos para la estimación de los parámetros vía algoritmo EM	23

Índice de figuras

4.1. Envelopes modelos utilizados	24
---	----

Capítulo 1

Preliminares

1.1. Introducción

En algunas ocasiones se encuentran observaciones cuyos valores no son todos conocidos, estos tipos de datos son considerados datos censurados.

Frecuentemente estos datos aparecen cuando el experimento bajo estudio involucra el tiempo de vida o supervivencia.

Por ejemplo, si estamos probando la vida útil de 20 bombillas, y se decide parar el experimento cuando ocurran las primeras 15 fallas, los restantes valores de la vida útil serán desconocidos y sólo se tendrá registrados que son mayores que la decima quinta falla registrada.

También, cuando medimos tiempos de falla, algunos artículos bajo estudio pueden no haber fallado cuando termino el estudio, resultando en una cota inferior para el tiempo de falla de estos artículos. Este es un ejemplo de censura a derecha.

Es interesante mencionar que existen otros tipos de censura como censura a izquierda e intervalar, sin embargo en este caso vamos a tratar solo la censura a derecha.

1.2. Marco teórico

Una variable es censurada si solo se sabe que el verdadero valor asumido por la variable esta contenido en un intervalo (a,b).

Existen tres tipos de variables censuradas.

Sea Y_i^* una variable respuesta de interes y Y_i la variable que es observada en el experimento.

- Censura por la derecha (o superior):

$$Y_i = \begin{cases} Y_i^* & \text{Si } Y_i^* < L_i \\ L_i & \text{Si } Y_i^* \geq L_i \end{cases}$$

Cuando $a \geq -\infty$ y $b = \infty$, entonces $Y_i \in (L_i, \infty)$, $i = 1, \dots, n$.

- Censura por la izquierda (o inferior):

$$Y_i = \begin{cases} Y_i^* & \text{Si } Y_i^* > U_i \\ U_i & \text{Si } Y_i^* \leq U_i \end{cases}$$

Cuando $a = -\infty$ y $b \leq \infty$ entonces $Y_i \in (-\infty, U_i)$.

- Tercer tipo de censura: $Y_i^* \in [L_i, U_i]$.

En este tipo de censura, la i -ésima observación no es censurada, si $L_i = U_i = Y_i^*$.

Donde L= Lower (limite inferior) y U= Upper (limite superior).

1.2.1. Modelo De Regresión Lineal Con Respuesta Censurada

El modelo de regresión lineal usual es def nido de la siguiente manera:

$$Y_i = \mathbf{x}_i^t \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n$$

donde, Y_i es la variable respuesta, $\mathbf{x}_i = [x_{i1}, \dots, x_{ip}]^t$ vector $p \times 1$ de variables explicativas, $\boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^t$ vector de parámetro de regresión y ε son los errores aleatorios considerados independientes e idénticamente distribuidos con $E[\varepsilon_i] = 0$ y $Var[\varepsilon_i] = \sigma^2$, con $i = 1, \dots, n$.

En el modelo de regresión lineal con censura a derecha, Y_i no es totalmente observado, por el contrario, se observa el vector (\mathbf{Q}, \mathbf{C}) en el que $\mathbf{Q} = [Q_1, \dots, Q_n]^t$ el cual es el vector con los valores observados de la variable respuesta Y y $\mathbf{C} = [C_1, \dots, C_n]^t$ el vector de indicadores de censura.

Expresado así,

$$Y_i = Q_i \text{ Si } C_i = 0 \quad (\text{No censurado})$$

$$Y_i \geq Q_i \text{ Si } C_i = 1 \quad \text{con } i = 1, \dots, n \quad (\text{Censurado})$$

Si f es una función de densidad de Y_i y $I_{\{A\}}$ es una función indicadora del conjunto A, para $i = 1, \dots, n$, entonces la función de verosimilitud es dada de la siguiente manera:

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^n [f(y_i | \boldsymbol{\theta})]^{I_{\{C_i=0\}}} (P(Y_i \geq \theta_i))^{I_{\{C_i=1\}}} \\ &= \prod_{i=1}^n [f(y_i | \boldsymbol{\theta})]^{I_{\{C_i=0\}}} (1 - P(Y_i \leq Q_i))^{I_{\{C_i=1\}}} \end{aligned}$$

Y la log-verosimilitud es expresada así:

$$l(\boldsymbol{\theta}) = \log[L(\boldsymbol{\theta})]$$

1.2.2. Algoritmo Newton-Raphson

La idea básica del Método de Newton Raphson es bastante sencilla, consiste en la linearización de una función derivable. Suponga $f : \mathbf{R} \mapsto \mathbf{R}$ diferenciable y se quiere resolver la ecuación $f(x) = 0$. Comenzando desde un punto inicial x_0 se puede construir una aproximación lineal de $f(x)$ en una vecindad de x_0 :

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0).$$

Por la ecuación anterior, se tiene que

$$0 = f(x) \approx f(x_0) + f'(x_0)(x - x_0).$$

Por lo tanto se puede suponer que

$$0 = f(x_0) + f'(x_0)(x - x_0).$$

Que tiene solución, digamos

$$x_1 = x_0 - f'(x_0)^{-1}f(x_0)$$

Ahora, si $f'(x_0)^{-1} \neq 0$ repetimos el proceso y encontramos

$$x_2 = x_1 - f'(x_1)^{-1}f(x_1)$$

Por lo tanto, en la k -ésima iteración, si $f'(x_k)^{-1} \neq 0$, encontramos

$$x_{k+1} = x_k - f'(x_k)^{-1}f(x_k).$$

La ecuación anterior motiva para la siguiente definición.

La secuencia del método de Newton Raphson para resolver $f(x) = 0$ donde f es una función derivable es la secuencia dada por la siguiente regla de recurrencia

$$x_{k+1} = x_k - f'(x_k)^{-1}f(x_k) \quad k = 0, 1, 2, \dots$$

Para que esta definición tenga sentido, se debe garantizar que $x_k \in I$ esto es, que x_k se encuentre en el dominio de f y que $f'(x_k) \neq 0$ para todo $k = 0, 1, 2, \dots$

La idea del método es exactamente esta, el uso de la expresión de una función derivable, en su forma lineal, en lugar de la propia función, visto que la expresión en cuestión es una buena aproximación local de la función ver Lima (1995).

Este método fue propuesto por Isaac Newton en 1669 para encontrar raíces de funciones polinomiales. Poco después, en 1690 J. Raphson extendió el método para cualquier tipo de funciones.

1.2.3. Algoritmo EM

El algoritmo EM, propuesto por Dempster, Laird y Rubin en 1977 es un método iterativo para estimar parámetros de modelos que envuelven variables latentes, es decir que no pueden ser completamente observadas, maximizando la esperanza de la función log-verisimilitud.

Este algoritmo consta de dos pasos, el primero es el paso E, que calcula la esperanza de la función de log-verosimilitud, y por ultimo el paso M, actualiza los valores de los parámetros del modelo maximizando la función calculada en el paso E, dados los valores de los parámetros en la iteración anterior del algoritmo.

Para implementar este algoritmo, consideramos usar la función de log-verosimilitud completa, tomamos q_i como una observación de la variable latente Y_i , cuando $C_i = 1$, con $i = 1, \dots, n$, teniendo así, los datos “aumentados” $(y_1, y_2, \dots, y_m, q_{m+1}, \dots, q_n)^T = y_{obs}$

$$L_c(\theta | y_i, c_i, Q_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}(y_i - x_i\beta)^2\right\}$$

$$\begin{aligned} l_c(\theta | y_i, c_i, Q_i) &= \text{constante} + \sum_{i=1}^n \log(\sigma^2)^{-1/2} - \frac{1}{2\sigma^2}(y_i - x_i\beta)^2 \\ &= \text{constante} - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - x_i\beta)^2 \\ &= \text{constante} - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i^2 - 2y_i x_i\beta + (x_i\beta)^2) \end{aligned}$$

Como se mencionó anteriormente, cada iteración de este algoritmo consiste en dos pasos:

• **Paso E:** *Expectation*. Calculo de la esperanza de la log-verosimilitud completa.

Es decir, desenvuelve el cálculo de la esperanza de funciones de \mathbf{Y} .

Esta etapa consiste en calcular la función $Q(\theta|\theta^{(t)})$, de la siguiente manera:

$$\begin{aligned} Q(\theta|\theta^{(t)}) &= E_Y[l(\theta|Y_{obs}, Y)/Y_{obs}\theta^{(t)}] \\ &= \int_Y l(\theta|Y_{obs}, Y)f(Y|Y_{obs}\theta^{(t)}) \end{aligned}$$

• **Paso M:** *Maximization*. Obtiene nuevos valores de los parámetros a partir de los datos proporcionados por el paso anterior.

Sea $\theta^{(t)}$ la estimativa actual θ ; consiste en maximizar $Q(\theta|\theta^{(t)})$ con relación a θ , es decir,

$$\theta^{(t+1)} = \text{argmax} Q(\theta | \theta^{(t)})$$

Estos pasos anteriores son repetidos hasta la convergencia, que es dada de la siguiente forma:

$$\|l(\hat{\theta}^{(t+1)}) - l(\hat{\theta}^{(t)})\| < \varepsilon. \varepsilon \downarrow 0$$

Análisis de Resíduos

El análisis de residuos tiene como objetivo principal identificar observaciones atípicas o la falta de ajuste del modelo, así los residuos son medidas de concordancia entre los datos y el modelo ajustado. Se puede decir que la mayoría de los residuos están basados en la diferencia entre las respuestas observadas y las medias observadas, en particular se definen los residuos estandarizados también llamados residuos de Pearson, como:

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{Var}(y_i)}} \quad i = 1, \dots, n,$$

donde, para los modelos aquí estudiados $\hat{\mu}_i$ corresponde al predictor lineal de la observación i , es decir, $\mathbf{x}_i^t \hat{\boldsymbol{\beta}}$ y $\widehat{Var}(y_i) = \hat{\sigma}^2$, siendo $\hat{\boldsymbol{\beta}}$ y $\hat{\sigma}^2$ los estimadores de máxima verosimilitud de $\boldsymbol{\beta}$ y σ^2 respectivamente.

Como es sugerido en Atkinson (1981), también se pueden generar envelopes para detectar observaciones atípicas o falta de ajuste de la distribución específica para el error.

Capítulo 2

Distribución Normal

Considerando el modelo de regresión lineal censurado descrito en la sección anterior, donde, $\epsilon_i \sim N(0, \sigma^2)$ con $i = 1, \dots, n$ y $Y_i \stackrel{iid}{\sim} N(x_i \boldsymbol{\beta}, \sigma^2)$.

Así, tratamos a Y_i como una variable latente y estimamos los parámetros del modelo a través del algoritmo EM.

La función de verosimilitud de una distribución normal esta dada por:

$$\begin{aligned} L(\boldsymbol{\phi}) &= \prod_{i=1}^n \left[\phi(q_i | x_i^t \boldsymbol{\beta}, \sigma^2) \right]^{I(C_i=0)} \left[1 - \Phi\left(\frac{q_i - x_i^t \boldsymbol{\beta}}{\sigma}\right) \right]^{I(C_i=1)} \\ &= \prod_{i=1}^n \left[\frac{1}{\sigma} \phi\left(\frac{q_i - x_i^t \boldsymbol{\beta}}{\sigma}\right) \right]^{I(C_i=0)} \left[1 - \Phi\left(\frac{q_i - x_i^t \boldsymbol{\beta}}{\sigma}\right) \right]^{I(C_i=1)} \end{aligned}$$

donde $\phi(\cdot)$ es la densidad de una normal estándar y $\Phi(\cdot)$ la función de distribución acumulada.

2.1. Newton-Raphson

Cuando se considera el modelo de regresión lineal

$$Y_i = \beta_1 + \beta_2 X_i + \epsilon_i, \quad \epsilon_i \stackrel{iid}{\sim} N(0; \sigma^2),$$

donde la respuesta $\mathbf{y} = (Y_1, \dots, Y_n)^T$ no es completamente observada para todos los individuos se tiene el caso censurado del modelo, recordando, sean (\mathbf{Q}, \mathbf{C}) los datos observados, donde $\mathbf{Q} = (Q_1, \dots, Q_n)^T$ es el vector de respuestas observadas y $\mathbf{C} = (C_1, \dots, C_n)^T$ el vector indicador de censura. Así se va a tener que:

$$Y_i = Q_i \text{ si } C_i = 0 \text{ y } Y_i \geq Q_i \text{ si } C_i = 1.$$

La función de versoimilitud es la siguiente

$$L(\boldsymbol{\theta} | \mathbf{Y}_{Obs}) = L(\boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^m f_Y(y_i) \prod_{i=m+1}^n \bar{F}_Y(y_i)$$

Siendo $f_Y(y_i)$ la función de densidad de una distribución normal con parámetros $(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2)$ calculada en el punto y_i ,

$$\bar{F}_Y(y_i) = \int_{y_i}^{\infty} f_Y(u) du,$$

por lo tanto, se tiene

$$L(\boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^m \left[\frac{1}{\sigma} \phi\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right) \right] \prod_{i=m+1}^n \left[1 - \Phi\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right) \right]$$

Con $\phi(v)$ siendo el valor de la densidad de la normal estándar calculada en el valor de v y $\Phi(v)$ es la función acumulada de la normal estándar calculada en v .

Aplicando el logaritmo, se obtiene:

$$\begin{aligned} l(\boldsymbol{\beta}, \sigma^2) &= \sum_{i=1}^m \log[\sigma^{-1} \phi(c_i)] + \sum_{i=m+1}^n \log[1 - \Phi(c_i)] \\ &= -\frac{m}{2} \log(\sigma^2) + \sum_{i=1}^m \log(\phi(c_i)) + \sum_{i=m+1}^n \log[1 - \Phi(c_i)] \end{aligned}$$

siendo

$$c_i = \frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}.$$

Para la implementación del algoritmo es necesario calcular el vector gradiente y la matriz Hessiana de la log-verosimilitud, así;

$$\nabla l(\boldsymbol{\beta}, \sigma^2) = \begin{pmatrix} \frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} \\ \frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} \end{pmatrix}, \quad \nabla^2 l(\boldsymbol{\beta}, \sigma^2) = \begin{pmatrix} \frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta} \partial \sigma^2} \\ \frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}^T \partial \sigma^2} & \frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2)}{\partial (\sigma^2)^2} \end{pmatrix}$$

A continuación se presentan las derivadas

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} = \sum_{i=1}^m \left[\frac{1}{\phi(c_i)} \frac{\partial \phi(c_i)}{\partial c_i} \frac{\partial c_i}{\partial \boldsymbol{\beta}} \right] + \sum_{i=m+1}^n \left[\frac{1}{1 - \Phi(c_i)} (-\phi(c_i)) \frac{\partial c_i}{\partial \boldsymbol{\beta}} \right]$$

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} = -\frac{m}{2\sigma^2} + \sum_{i=1}^m \left[\frac{1}{\phi(c_i)} \frac{\partial \phi(c_i)}{\partial c_i} \frac{\partial c_i}{\partial \sigma^2} \right] + \sum_{i=m+1}^n \left[\frac{1}{1 - \Phi(c_i)} (-\phi(c_i)) \frac{\partial c_i}{\partial \sigma^2} \right]$$

$$\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \sum_{i=1}^m \left\{ \left[\frac{-1}{\phi^2(c_i)} \left(\frac{\partial \phi(c_i)}{\partial c_i} \right)^2 \frac{\partial c_i}{\partial \boldsymbol{\beta}} \frac{\partial c_i}{\partial \boldsymbol{\beta}^T} \right] + \left[\frac{1}{\phi(c_i)} \frac{\partial^2 \phi(c_i)}{\partial c_i^2} \frac{\partial c_i}{\partial \boldsymbol{\beta}} \frac{\partial c_i}{\partial \boldsymbol{\beta}^T} \right] \right\}$$

$$+ \sum_{i=m+1}^n \left\{ \left[\frac{-1}{[1 - \Phi(c_i)]^2} (\phi^2(c_i)) \frac{\partial c_i}{\partial \boldsymbol{\beta}} \frac{\partial c_i}{\partial \boldsymbol{\beta}^T} \right] + \left[\frac{-1}{[1 - \Phi(c_i)]} \frac{\partial \phi(c_i)}{\partial c_i} \frac{\partial c_i}{\partial \boldsymbol{\beta}} \frac{\partial c_i}{\partial \boldsymbol{\beta}^T} \right] \right\}$$

$$\begin{aligned}
\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta} \partial \sigma^2} &= \sum_{i=1}^m \left\{ \left[\frac{-1}{\phi^2(c_i)} \left(\frac{\partial \phi(c_i)}{\partial c_i} \right)^2 \frac{\partial c_i}{\partial \boldsymbol{\beta}} \frac{\partial c_i}{\partial \sigma^2} \right] + \left[\frac{1}{\phi(c_i)} \frac{\partial^2 \phi(c_i)}{\partial c_i^2} \frac{\partial c_i}{\partial \boldsymbol{\beta}} \frac{\partial c_i}{\partial \sigma^2} \right] + \left[\frac{1}{\phi(c_i)} \frac{\partial \phi(c_i)}{\partial c_i} \frac{\partial c_i}{\partial \boldsymbol{\beta}} \frac{\partial c_i}{\partial \sigma^2} \right] \right\} \\
&+ \sum_{i=m+1}^n \left\{ \left[\frac{-1}{[1 - \Phi(c_i)]^2} (\phi^2(c_i)) \frac{\partial c_i}{\partial \sigma^2} \frac{\partial c_i}{\partial \boldsymbol{\beta}} \right] + \left[\frac{-1}{[1 - \Phi(c_i)]} \frac{\partial \phi(c_i)}{\partial c_i} \frac{\partial c_i}{\partial \sigma^2} \frac{\partial c_i}{\partial \boldsymbol{\beta}} \right] \right\} + \left[\frac{-1}{[1 - \Phi(c_i)]} \phi(c_i) \frac{\partial^2 c_i}{\partial \boldsymbol{\beta} \partial \sigma^2} \right] \Big\} \\
\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2)}{\partial (\sigma^2)^2} &= \frac{m}{2\sigma^4} + \sum_{i=1}^m \left\{ \left[\frac{-1}{\phi^2(c_i)} \left(\frac{\partial \phi(c_i)}{\partial c_i} \right)^2 \left(\frac{\partial c_i}{\partial \sigma^2} \right)^2 \right] + \left[\frac{1}{\phi(c_i)} \frac{\partial^2 \phi(c_i)}{\partial c_i^2} \left(\frac{\partial c_i}{\partial \sigma^2} \right)^2 \right] + \left[\frac{1}{\phi(c_i)} \frac{\partial \phi(c_i)}{\partial c_i} \frac{\partial^2 c_i}{\partial (\sigma^2)^4} \right] \right\} \\
&+ \sum_{i=m+1}^n \left\{ \left[\frac{-1}{[1 - \Phi(c_i)]} (\phi^2(c_i)) \left(\frac{\partial c_i}{\partial \sigma^2} \right)^2 \right] + \left[\frac{-1}{[1 - \Phi(c_i)]} \frac{\partial \phi(c_i)}{\partial c_i} \left(\frac{\partial c_i}{\partial \sigma^2} \right)^2 \right] + \left[\frac{-1}{[1 - \Phi(c_i)]} \phi(c_i) \frac{\partial^2 c_i}{\partial (\sigma^2)^2} \right] \right\}
\end{aligned}$$

Para que las ecuaciones anteriores queden de forma mas legibles, a continuación se explicita las funciones utilizadas

$$\begin{aligned}
\frac{\partial \phi(c_i)}{\partial c_i} &= \frac{-c_i}{\sqrt{2\pi}} \exp(-0,5c_i^2) \\
\frac{\partial^2 \phi(c_i)}{\partial c_i^2} &= -\frac{1}{\sqrt{2\pi}} \exp(-0,5c_i^2) + \frac{c_i^2}{\sqrt{2\pi}} \exp(-0,5c_i^2) \\
\frac{\partial c_i}{\partial \boldsymbol{\beta}} &= \frac{-\mathbf{x}_i}{\sigma} \quad \frac{\partial^2 c_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \mathbf{0}_{p \times p} \quad \frac{\partial c_i}{\partial \sigma^2} = -\frac{1}{2\sigma^3} (y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \\
\frac{\partial^2 c_i}{\partial (\sigma^2)^2} &= -\frac{3}{4\sigma^5} (y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \quad \frac{\partial^2 c_i}{\partial \boldsymbol{\beta} \partial \sigma^2} = \frac{\mathbf{x}_i}{2\sigma^2}
\end{aligned}$$

2.2. Algoritmo EM

Para el desembolvimiento del algoritmo EM es necesaria la introducción de la verosimilitud completa de los datos, más para esto se va a ser uso de la distribución Normal truncada univariada, a continuación se da la definición:

Definición

Sea X una variable aleatoria. Decimos que X tiene distribución normal truncada en el intervalo $A = (a, b)$, $b > a$, con parámetro de locación μ y de escala σ^2 , denotada por $NT(\mu, \sigma^2, (a, b))$ la función de densidad es:

$$f(x) = \frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{\sigma \left[\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \right]},$$

donde $\phi(\cdot)$ es la densidad de una normal estandar y $\Phi(\cdot)$ la respectiva función de distribución acumulada.

Los primeros momentos de la distribución normal truncada son dados de la siguiente forma:

Si $X \sim NT(\mu, \sigma^2, (a, \infty))$, tenemos que

$$E(X) = \mu + \frac{\phi\left(\frac{a-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{a-\mu}{\sigma}\right)}\sigma$$

y

$$E(X^2) = \mu^2 + \sigma^2 + \frac{\phi\left(\frac{a-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{a-\mu}{\sigma}\right)}(\mu + a)\sigma$$

Ahora se procede con la verosimilitud completa, con $\boldsymbol{\theta} = [\boldsymbol{\beta}^T, \sigma^2]^T$ la cual es dada por:

$$L_c(\boldsymbol{\theta} | \mathbf{y}, \mathbf{C}, \mathbf{Q}) = \prod_{i=1}^n (2\pi\sigma^2)^{1/2} \exp\left\{-\frac{1}{2\sigma^2}(q_i - x_i^T \boldsymbol{\beta})^2\right\}$$

teniendo también la log-verosimilitud completa que es dada de la siguiente forma,

$$\begin{aligned} l_c(\boldsymbol{\theta} | \mathbf{y}, \mathbf{C}, \mathbf{Q}) &= \sum_{i=1}^n \left(-\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (q_i - x_i^T \boldsymbol{\beta})^2 \right) \\ &= \sum_{i=1}^n \left(-\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (q_i - x_i^T \boldsymbol{\beta})^2 \right) \\ &\propto \sum_{i=1}^n \left(-\frac{1}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (q_i - x_i^T \boldsymbol{\beta})^2 \right) \\ &\propto -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (q_i^2 - 2q_i x_i^T \boldsymbol{\beta} + (x_i^T \boldsymbol{\beta})^2) \end{aligned}$$

La función Q es dada por:

$$\begin{aligned} Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)}) &= E_Y[l_c(\boldsymbol{\theta}) | C, Q, \hat{\boldsymbol{\theta}}^{(k)}] \\ &\propto -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (E[Y_i^2] - 2x_i^t \boldsymbol{\beta} E[Y_i] + (x_i^t \boldsymbol{\beta})^2) \\ &\propto -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n [\hat{y}_i^2 - 2\hat{y}_i x_i^t \boldsymbol{\beta} + (x_i^t \boldsymbol{\beta})^2], \end{aligned}$$

donde

$$\begin{aligned} \hat{y}_i^2 &= E[Y_i^2 | Q_i, C_i, \hat{\boldsymbol{\theta}}^{(k)}] \\ \hat{y}_i &= E[Y_i | Q_i, C_i, \hat{\boldsymbol{\theta}}^{(k)}] \end{aligned}$$

y $\widehat{\theta}^{(k)}$ es un valor actualizado de $\widehat{\theta}$.

Si $C_i = 0$, entonces $\widehat{y}_i^2 = y_i^2$ y $\widehat{y}_i = y_i$

Si $C_i = 1$, entonces $Y_i | \theta_i, C_i = 1$

$$Y_i | Q_i, C_i = 1 \stackrel{iid}{\sim} NT(\mathbf{x}_i^t \boldsymbol{\beta}, \sigma^2, (Q_i, \infty)), i = 1, \dots, n$$

Dando, un valor inicial $\boldsymbol{\theta}^{(0)}$ para el vector de parámetros, la iteración $(k + 1)$ del algoritmo EM seria la siguiente:

Passo E: Dado $\boldsymbol{\theta}^{(k)}$, calculamos para $i = 1, \dots, n$

■ Caso $C_i = 1$:

$$\widehat{y}_i^t = x_i^t \widehat{\beta}^{(k)} + \frac{\phi\left(\frac{Q_i - x_i^t \widehat{\beta}^{(k)}}{\widehat{\sigma}^{(k)}}\right)}{1 - \Phi\left(\frac{Q_i - x_i^t \widehat{\beta}^{(k)}}{\widehat{\sigma}^{(k)}}\right)} \widehat{\sigma}^{(k)},$$

$$\widehat{y}_i^{2(k)} = [x_i^t \widehat{\beta}^{(k)}]^2 + \widehat{\sigma}^{2(k)} + \frac{\phi\left(\frac{Q_i - x_i^t \widehat{\beta}^{(k)}}{\widehat{\sigma}^{(k)}}\right)}{1 - \Phi\left(\frac{Q_i - x_i^t \widehat{\beta}^{(k)}}{\widehat{\sigma}^{(k)}}\right)} (x_i^t \widehat{\beta}^{(k)} + Q_i) \widehat{\sigma}^{(k)}.$$

■ Caso $C_i = 0$:

$$\begin{aligned} \widehat{y}_i^{(k)} &= y_i, \\ \widehat{y}_i^{2(k)} &= y_i^2. \end{aligned}$$

Passo M: Actualizar $\widehat{\boldsymbol{\theta}}^{(k)}$ maximizando $Q(\boldsymbol{\theta} / \widehat{\boldsymbol{\theta}}^{(k)})$ con relación a $\boldsymbol{\theta}$, tomando las siguientes expresiones: x_i es la i -ésima línea de la matriz \mathbf{x} de dimensión $n \times p$ y $\widehat{y}_i^{(k)}$ es el i -ésimo elemento del vector $\widehat{\mathbf{y}}^{(k)}$ de dimensión $n \times 1$, para $i = 1, \dots, n$.

De la siguiente forma:

$$\begin{aligned} \frac{\partial Q(\boldsymbol{\theta} / \widehat{\boldsymbol{\theta}}^{(k)})}{\partial \boldsymbol{\beta}} &\propto \frac{1}{2\sigma^2} \sum_{i=1}^n [-2\widehat{y}_i^{(k)} x_i^t + 2x_i^t \beta x_i] \\ &\propto - \sum_{i=1}^n \widehat{y}_i^{(k)} x_i^t + \sum_{i=1}^n x_i^t \beta x_i, \end{aligned}$$

$$\begin{aligned}
\frac{\partial Q(\theta/\hat{\theta}^{(k)})}{\partial \beta} = 0 &\Rightarrow \sum_{i=1}^n x_i^t \hat{\beta}^{(k+1)} x_i = \sum_{i=1}^n \hat{y}_i^{(k)} x_i^t \\
&\Rightarrow \hat{\beta}^{(k+1)} = \left(\sum_{i=1}^n x_i^t x_i \right)^{-1} \left(\sum_{i=1}^n \hat{y}_i^{(k)} x_i^t \right) \\
&\Rightarrow \hat{\beta}^{(k+1)} = (\mathbf{x}^t \mathbf{x})^{-1} \mathbf{x}^t \hat{\mathbf{y}}^{(k)}
\end{aligned}$$

y

$$\frac{\partial Q(\theta/\hat{\sigma}^{(k)})}{\partial \sigma} = -\frac{n}{2\hat{\sigma}^{2(k+1)}} + \frac{1}{2(\hat{\sigma}^{2(k+1)})^2} \sum_{i=1}^n [\hat{y}_i^{2(k)} - 2\hat{y}_i^{(k)} x_i^t \hat{\beta}^{(k+1)} + (x_i^t \hat{\beta}^{(k+1)})^2].$$

$$\begin{aligned}
\frac{\partial Q(\theta/\hat{\sigma}^{(k)})}{\partial \sigma} = 0 &\Rightarrow \frac{1}{2(\hat{\sigma}^{2(k+1)})^2} \sum_{i=1}^n \left[\hat{y}_i^{2(k)} - 2\hat{y}_i^{(k)} x_i^t \hat{\beta}^{(k+1)} + (x_i^t \hat{\beta}^{(k+1)})^2 \right] = \frac{n}{2\hat{\sigma}^{2(k+1)}} \\
&\Rightarrow \hat{\sigma}^{2(k+1)} = \frac{1}{n} \sum_{i=1}^n \left[\hat{y}_i^{2(k)} - 2\hat{y}_i^{(k)} x_i^t \hat{\beta}^{(k+1)} + (x_i^t \hat{\beta}^{(k+1)})^2 \right] \\
&\Rightarrow \hat{\sigma}^{2(k+1)} = \frac{1}{n} \sum_{i=1}^n \left[\hat{y}_i^{2(k)} - 2\hat{y}_i^{(k)} \mathbf{x}_i^t \hat{\beta}^{(k+1)} + (\mathbf{x}_i^t \hat{\beta}^{(k+1)})^2 \right]
\end{aligned}$$

Dadas las estimativas iniciales $\hat{\beta}^{(0)}$ y $\hat{\sigma}^{2(0)}$, la iteración k del algoritmo consiste en:

1. Calcular $\hat{y}_i^{(k)}$
2. Calcular $\hat{y}_i^{2(k)}$
3. Calcular $\hat{\beta}^{(k+1)}$
4. Calcular $\hat{\sigma}^{2(k+1)}$

En que:

$$\hat{y}_i^{(k)} = \begin{cases} Q_i & \text{Si } C_i = 0 \\ X_i \hat{\beta}^{(k)} + \frac{\phi\left(\frac{Q_i - X_i \hat{\beta}^{(k)}}{\hat{\sigma}^{(k)}}\right) \hat{\sigma}^{(k)}}{1 - \Phi\left(\frac{Q_i - X_i \hat{\beta}^{(k)}}{\hat{\sigma}^{(k)}}\right)} & \text{Si } C_i = 1 \end{cases}$$

$$\hat{y}_i^2 = \begin{cases} Q_i^2 & \text{Si } C_i = 0 \\ (X_i \hat{\beta}^{(k)})^2 + \hat{\sigma}^{2(k)} + \frac{\phi\left(\frac{Q_i - X_i \hat{\beta}^{(k)}}{\hat{\sigma}^{(k)}}\right)}{1 - \Phi\left(\frac{Q_i - X_i \hat{\beta}^{(k)}}{\hat{\sigma}^{(k)}}\right)} (X_i \hat{\beta}^{(k)} + Q_i) \hat{\sigma}^{(k)} & \text{Si } C_i = 1 \end{cases}$$

- $\hat{\beta}^{(k+1)} = (X'X)^{-1}X'\hat{y}^{(k)}$
- $\hat{\sigma}^{2(k+1)} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i^{2(k)} - 2\hat{y}_i^{(k)}X_i\hat{\beta}^{(k+1)} + (X_i\hat{\beta}^{(k+1)})^2)$

Estos pasos deben ser repetidos hasta la convergencia, por ejemplo, podemos considerar el siguiente criterio:

$$\|l(\hat{\theta}^{(k+1)}) - l(\hat{\theta}^{(k)})\| < \varepsilon \quad \varepsilon \downarrow 0$$

Capítulo 3

Distribución t-student

El modelo de regresión lineal con errores distribuidos conforme a una t-Student es definida así:

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \stackrel{iid}{\sim} t_\nu(0, \sigma^2) \quad i = 1, \dots, n$$

3.1. Newton-Raphson

Cuando se considera el modelo de regresión lineal

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i, \quad \varepsilon_i \stackrel{iid}{\sim} t(0; \sigma^2, \nu),$$

donde la respuesta $\mathbf{y} = (Y_1, \dots, Y_n)^T$ no es completamente observada para todos los individuos se tiene el caso censurado del modelo, recordando, sean (\mathbf{Q}, \mathbf{C}) los datos observados, donde $\mathbf{Q} = (Q_1, \dots, Q_n)^T$ es el vector de respuestas observadas y $\mathbf{C} = (C_1, \dots, C_n)^T$ el vector indicador de censura. Así se va a tener que:

$$\begin{aligned} Y_i &= Q_i \text{ si } C_i = 0 \\ &\text{y} \\ Y_i &\geq Q_i \text{ si } C_i = 1. \end{aligned}$$

La función de verosimilitud es la siguiente,

$$L(\boldsymbol{\theta} | \mathbf{Y}_{Obs}) = L(\boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^m f_Y(y_i) \prod_{i=m+1}^n \bar{F}_Y(y_i)$$

Siendo $f_Y(y_i)$ la función de densidad de una distribución t de Student con parámetros $(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2, \nu)$ calculada en el punto y_i ,

$$\bar{F}_Y(y_i) = \int_{y_i}^{\infty} f_Y(u) du,$$

por lo tanto, se tiene

$$L(\boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^m \left[\frac{1}{\sigma} t\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right) \right] \prod_{i=m+1}^n \left[1 - t\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right) \right]$$

Con $t(v)$ siendo el valor de la densidad de la t de Student centralizada calculada en el valor de v y Tv es la función acumulada de la t de Student centralizada calculada en v .

Aplicando el logaritmo, se obtiene:

$$\begin{aligned} l(\boldsymbol{\beta}, \sigma^2) &= \sum_{i=1}^m \log[\sigma^{-1} \phi(c_i)] + \sum_{i=m+1}^n \log[1 - T(c_i)] \\ &= -\frac{m}{2} \log(\sigma^2) + \sum_{i=1}^m \log(t(c_i)) + \sum_{i=m+1}^n \log[1 - T(c_i)] \end{aligned}$$

siendo

$$c_i = \frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}.$$

Para la implementación del algoritmo es necesario calcular el vector gradiente y la matriz Hessiana de la log-verosimilitud, que es dada como:

$$\nabla l(\boldsymbol{\beta}, \sigma^2) = \begin{pmatrix} \frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} \\ \frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} \end{pmatrix}, \quad \nabla^2 l(\boldsymbol{\beta}, \sigma^2) = \begin{pmatrix} \frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta} \partial \sigma^2} \\ \frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}^T \partial \sigma^2} & \frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2)}{\partial (\sigma^2)^2} \end{pmatrix}$$

A continuación se presentan las derivadas

$$\begin{aligned} \frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^m \left[\frac{1}{t(c_i)} \frac{\partial t(c_i)}{\partial c_i} \frac{\partial c_i}{\partial \boldsymbol{\beta}} \right] + \sum_{i=m+1}^n \left[\frac{1}{1 - T(c_i)} (-t(c_i)) \frac{\partial c_i}{\partial \boldsymbol{\beta}} \right] \\ \frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} &= -\frac{m}{2\sigma^2} + \sum_{i=1}^m \left[\frac{1}{t(c_i)} \frac{\partial t(c_i)}{\partial c_i} \frac{\partial c_i}{\partial \sigma^2} \right] + \sum_{i=m+1}^n \left[\frac{1}{1 - T(c_i)} (-t(c_i)) \frac{\partial c_i}{\partial \sigma^2} \right] \\ \frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= \sum_{i=1}^m \left\{ \left[\frac{-1}{t^2(c_i)} \left(\frac{\partial t(c_i)}{\partial c_i} \right)^2 \frac{\partial c_i}{\partial \boldsymbol{\beta}} \frac{\partial c_i}{\partial \boldsymbol{\beta}^T} \right] + \left[\frac{1}{t(c_i)} \frac{\partial^2 t(c_i)}{\partial c_i^2} \frac{\partial c_i}{\partial \boldsymbol{\beta}} \frac{\partial c_i}{\partial \boldsymbol{\beta}^T} \right] \right\} \\ &+ \sum_{i=m+1}^n \left\{ \left[\frac{-1}{[1 - T(c_i)]^2} (t^2(c_i)) \frac{\partial c_i}{\partial \boldsymbol{\beta}} \frac{\partial c_i}{\partial \boldsymbol{\beta}^T} \right] + \left[\frac{-1}{[1 - T(c_i)]} \frac{\partial t(c_i)}{\partial c_i} \frac{\partial c_i}{\partial \boldsymbol{\beta}} \frac{\partial c_i}{\partial \boldsymbol{\beta}^T} \right] \right\} \\ \frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta} \partial \sigma^2} &= \sum_{i=1}^m \left\{ \left[\frac{-1}{t^2(c_i)} \left(\frac{\partial t(c_i)}{\partial c_i} \right)^2 \frac{\partial c_i}{\partial \boldsymbol{\beta}} \frac{\partial c_i}{\partial \sigma^2} \right] + \left[\frac{1}{t(c_i)} \frac{\partial^2 t(c_i)}{\partial c_i^2} \frac{\partial c_i}{\partial \boldsymbol{\beta}} \frac{\partial c_i}{\partial \sigma^2} \right] + \left[\frac{1}{t(c_i)} \frac{\partial t(c_i)}{\partial c_i} \frac{\partial c_i}{\partial \boldsymbol{\beta}} \frac{\partial c_i}{\partial \sigma^2} \right] \right\} \\ &+ \sum_{i=m+1}^n \left\{ \left[\frac{-1}{[1 - T(c_i)]^2} (t^2(c_i)) \frac{\partial c_i}{\partial \sigma^2} \frac{\partial c_i}{\partial \boldsymbol{\beta}} \right] + \left[\frac{-1}{[1 - T(c_i)]} \frac{\partial t(c_i)}{\partial c_i} \frac{\partial c_i}{\partial \sigma^2} \frac{\partial c_i}{\partial \boldsymbol{\beta}} \right] + \left[\frac{-1}{[1 - T(c_i)]} \phi(c_i) \frac{\partial^2 c_i}{\partial \boldsymbol{\beta} \partial \sigma^2} \right] \right\} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2)}{\partial(\sigma^2)^2} &= \frac{m}{2\sigma^4} + \sum_{i=1}^m \left\{ \left[\frac{-1}{t^2(c_i)} \left(\frac{\partial t(c_i)}{\partial c_i} \right)^2 \left(\frac{\partial c_i}{\partial \sigma^2} \right)^2 \right] + \left[\frac{1}{t(c_i)} \frac{\partial^2 t(c_i)}{\partial c_i^2} \left(\frac{\partial c_i}{\partial \sigma^2} \right)^2 \right] + \left[\frac{1}{t(c_i)} \frac{\partial t(c_i)}{\partial c_i} \frac{\partial^2 c_i}{\partial(\sigma^2)^4} \right] \right\} \\ &+ \sum_{i=m+1}^n \left\{ \left[\frac{-1}{[1-T(c_i)]} (t^2(c_i)) \left(\frac{\partial c_i}{\partial \sigma^2} \right)^2 \right] + \left[\frac{-1}{[1-T(c_i)]} \frac{\partial t(c_i)}{\partial c_i} \left(\frac{\partial c_i}{\partial \sigma^2} \right)^2 \right] + \left[\frac{-1}{[1-T(c_i)]} t(c_i) \frac{\partial^2 c_i}{\partial(\sigma^2)^2} \right] \right\} \end{aligned}$$

Para que las ecuaciones anteriores queden de forma mas legibles, a continuación se explicita las funciones utilizadas

$$\begin{aligned} \frac{\partial c_i}{\partial \boldsymbol{\beta}} &= \frac{-\mathbf{x}_i}{\sigma} & \frac{\partial^2 c_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= \mathbf{0}_{p \times p} & \frac{\partial c_i}{\partial \sigma^2} &= \frac{1}{2\sigma^3} (y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \\ \frac{\partial^2 c_i}{\partial(\sigma^2)^2} &= -\frac{3}{4\sigma^5} (y_i - \mathbf{x}_i^T \boldsymbol{\beta}) & \frac{\partial^2 c_i}{\partial \boldsymbol{\beta} \sigma^2} &= \frac{\mathbf{x}_i}{2\sigma^2} \end{aligned}$$

3.2. Algoritmo EM

Usando el modelo censurado, $Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i$, $i = 1, \dots, n$, $\varepsilon_i = U_i^{-1/2} Z_i U_i \stackrel{iid}{\sim} \text{Gamma}(\frac{\nu}{2}, \frac{\nu}{2})$, $Z_i \stackrel{iid}{\sim} N(0, \sigma^2)$, $i = 1, \dots, n$, la log-verosimilitud completa asociada con $(\mathbf{y}, \mathbf{u}, \mathbf{C}, \mathbf{Q})$, es:

$$l_c(\boldsymbol{\theta} | \mathbf{y}, \mathbf{u}, \mathbf{C}, \mathbf{Q}) = \text{constante} - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n u_i (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2$$

Donde c es una constante que es independiente de $\boldsymbol{\theta}$.

Sea,

$$\begin{aligned} \widehat{uy^0}_i &= E[U_i | \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, Q_i, C_i] \\ \widehat{uy^1}_i &= E[U_i Y_i | \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, Q_i, C_i] \\ \widehat{uy^2}_i &= E[U_i Y_i^2 | \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, Q_i, C_i] \end{aligned}$$

El siguiente resultado se encuentra en Kim (2008) y presenta los primeros dos momentos de la t-Student truncada,

TEOREMA*. Sea $X \sim Tt_\nu(0, 1, (\alpha, \beta))$, entonces:

$$\begin{aligned} E(X) &= G(\nu) [(\nu + \alpha^2)^{-(\frac{\nu-1}{2})} - (\nu - \beta^2)^{-(\frac{\nu-1}{2})}], \quad \nu > 1 \\ E(X^2) &= \frac{F_{PVII}(\nu, \nu-2)(\beta) - F_{PVII}(\nu, \nu-2)(\alpha)}{F_{T\nu}(\beta) F_{T\nu}(\alpha)} \left(\frac{\nu}{\nu-2} \right) \\ &\quad + G(\nu) \{ \alpha(\nu + \alpha^2)^{-\frac{\nu-1}{2}} - \beta(\nu + \beta^2)^{-\frac{\nu-1}{2}} \}, \quad \nu > 2 \end{aligned}$$

en que,

$$G(\nu) = \frac{\Gamma(\frac{\nu-1}{2}) \cdot \nu^{\nu/2}}{2(F_{T\nu}(\beta) - F_{T\nu}(\alpha))\Gamma(\nu/2)\Gamma(1/2)}$$

$F_{PVI}(\nu, \sigma)$ es la acumulada de Pearson.VII estándar con parámetros (ν, σ) y $F_{T\nu}$ es la acumulada de la t-Student estándar con ν grados de libertad.

COROLARIO. Si $Y = \mu + \sigma x$, con $X \sim Tt_{\nu}(0, 1, (\sigma, \beta))$, entonces, $Y \sim Tt_{\nu}(\mu, \sigma^2, (a, b))$ con $a = \mu + \sigma\alpha$, $b = \mu + \sigma\beta$ y

$$\begin{aligned} E(Y) &= \mu + \sigma E(X) \\ E(Y^2) &= \mu^2 + 2\mu\sigma E(X) + \sigma^2 E(X^2) \end{aligned}$$

LEMA. Sea $d^2 = (Y - \mu)^2 / \sigma^2$. Si $Y \sim Tt_{\nu}(\mu, \sigma^2; (a, b))$ y $U \sim Gamma(\nu/2, \nu/2)$ entonces:

$$E\left[\left(\frac{\nu + 1}{\nu + d^2}\right)^k Y^s\right] = E(Z^s) \frac{F_{\nu+2k}(b^*) - F_{\nu+2k}(a^*)}{F_{\nu}(b) - F_{\nu}(a)} \frac{\Gamma(\frac{\nu+1}{2})\Gamma(\frac{\nu+2k}{2})}{\Gamma(\frac{\nu}{2})\Gamma(\frac{\nu+2k+1}{2})} \frac{(\nu + 1)^k}{\nu^k}$$

con $k=0,1,2$ y $s=1,2$.

en que, $Z \sim Tt_{\nu+2k}(\mu, \sigma^{2*}, A = (\alpha, \beta))$ y $\sigma^{2*} = \frac{\nu\sigma^2}{\nu+2k}$. $F_{\nu} = (\cdot)$ es la acumulada de la t_{ν} estándar.

$$\begin{aligned} b^* &= \frac{\beta - \mu}{\sigma^*} & a^* &= \frac{\alpha - \mu}{\sigma^*} \\ b &= \frac{\beta - \mu}{\sigma} & a^* &= \frac{\alpha - \mu}{\sigma} \end{aligned}$$

Las anteriores definiciones son útiles en la implementación del algoritmo EM, en el calculo de la matriz de varianzas y covarianzas del estimador de máxima verosimilitud de β .

La log-verosimilitud completa en relación a los parámetros $\theta = [\beta^t, \sigma^2]^t$, es dada de la siguiente manera:

$$\begin{aligned} Q(\theta | \hat{\theta}^{(k)}) &= E[l_c(\theta) | \mathbf{C}, \mathbf{Q}, \hat{\theta}^{(k)}] \\ &= c - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (\widehat{uy_i^2}^{(k)} - 2\widehat{uy_i^1}^{(k)} \mathbf{x}_i^t \beta + \widehat{uy_i^0}^{(k)} (\mathbf{x}_i^t \beta)^2) \end{aligned}$$

Donde $\hat{\theta}^{(k)}$ es un valor actualizado de $\hat{\theta}$.

Para realizar el algoritmo EM, es necesario definir un valor inicial θ^0 para el vector de parámetros, la iteración $(k+1)$ del algoritmo es dada de la siguiente manera:

Passo E: Dada una estimativa actual $\hat{\boldsymbol{\theta}}^{(k)}$, se calcula $\widehat{uy}_i^s{}^{(k)}$, si $c_i = 0$,
 $\widehat{UY}_i^s = Y_i^s \cdot \left(\frac{\nu+1}{\nu+d^2} \right)$, si $c_i = 1$, se usan los teoremas anteriores, para $i = 1, \dots, n$ y $s=0,1,2$.

Passo M: Actualice $\hat{\boldsymbol{\theta}}^{(k)}$ maximizando $Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{(k)})$ en relación a los parámetros de $\boldsymbol{\theta}$, es decir,

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \left(\sum_{i=1}^n \widehat{uy}_i^0{}^{(k)} \mathbf{x}_i \mathbf{x}_i^t \right)^{-1} \sum_{i=1}^n \mathbf{x}_i \widehat{uy}_i^1{}^{(k)}$$

$$\widehat{\sigma}^2{}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \left(\widehat{uy}_i^2{}^{(k)} - 2\widehat{uy}_i^1{}^{(k)} \mathbf{x}_i \widehat{\boldsymbol{\beta}}^{(k+1)} + \widehat{uy}_i^0{}^{(k)} (\mathbf{x}_i^t \widehat{\boldsymbol{\beta}}^{(k+1)})^2 \right)$$

Estos pasos son repetidos hasta la convergencia, es decir,
 $\|l(\hat{\boldsymbol{\theta}}^{(k+1)}) - l(\hat{\boldsymbol{\theta}}^{(k)})$ ó $l(\hat{\boldsymbol{\theta}}^{(k+1)})/l(\hat{\boldsymbol{\theta}}^{(k)}) - 1\|$ sea pequeña.

Capítulo 4

Aplicación

Los datos disponibles en Tan et al. (2010) “Insulation life data with censoring times” tratan de pruebas de vida acelerados de 40 motores eléctricos en cuatro diferentes temperaturas: 150°C, 170°C, 190°C, y 220°C, fueron probados 10 motores en cada una de las temperaturas. 23 de las 40 observaciones del tiempo de vida fueron censuradas a la derecha. La covariable denotada por T representa $1000(\text{temperatura}+273,2)$ y la variable respuesta y representa \log_{10} es decir, el tiempo de falla, en la Tabla 4.1 se encuentran los datos.

Cuadro 4.1: Datos Tormin 2000

i	y_i	T_i	i	y_i	T	i	y_i	T_i	i	y_i	T_i
1	3.2464	2.2563	13	2.6107	2.0276	18	3.9066	2.3629	30	3.7362	2.2563
2	3.4428	2.2563	14	2.6107	2.0276	19	3.9066	2.3629	31	3.2253	2.1589
3	3.5371	2.2563	15	2.7024	2.0276	20	3.9066	2.3629	32	3.2253	2.1589
4	3.5492	2.2563	16	2.7024	2.0276	21	3.9066	2.3629	33	3.2253	2.1589
5	3.5775	2.2563	17	2.7024	2.0276	22	3.9066	2.3629	34	3.2253	2.1589
6	3.6866	2.2563				23	3.9066	2.3629	35	3.2253	2.1589
7	3.7157	2.2563				24	3.9066	2.3629	36	2.7226	2.0276
8	2.6107	2.1589				25	3.9066	2.3629	37	2.7226	2.0276
9	2.6107	2.1589				26	3.9066	2.3629	38	2.7226	2.0276
10	3.1284	2.1589				27	3.9066	2.3629	39	2.7226	2.0276
11	3.1284	2.1589				28	3.7362	2.2563	40	2.7226	2.0276
12	3.1584	2.1589				29	3.7362	2.2563			2.0276

Para este conjunto de datos se aplico el algoritmo EM para datos censurados, la variable respuesta es censurada a la derecha y corresponde al tiempo de vida observado. La matriz diseño o matriz de covariables fue estructurada de la siguiente forma: la primera columna esta compuesta por un vector de unos y la segunda columna es una función de la temperatura a la cual fueron expuesto los motores, de esta forma el vector β es de dimensión 2×1 y va a tener en su primera componente el intercepto del modelo y en la segunda la pendiente. Se tendra $n = 40$ y $p = 2$.

Usando el método de Newton-Raphson para los casos de los modelos Normal y t-Student, con criterio de parada la diferencia entre la norma de las estimativas en el paso i con las del paso $i - 1$ y con un error de 10^{-6} , se obtuvieron que par ambos modelos, despues de 8 iteraciones la convergencia del método fué alcanzada y con los resultados que son presentados en la Tabla 4.2.

Cuadro 4.2: Resultados obtenidos para la estimación de los parámetros vía N-R

	$\hat{\beta}_0$	EP ($\hat{\beta}_0$)	$\hat{\beta}_1$	EP ($\hat{\beta}_1$)	$\hat{\sigma}^2$	EP ($\hat{\sigma}^2$)
Normal	-6.271	1.4293	4.483	0.659	0.146	0.053
T-Student	-6.057	1.1362	4.3589	0.5267	0.0863	0.040

Como es discutido en Massuia, M. B., Cabral, C. R. B., Matos, L. A., Lachos, V. H. (2012) se asume que los errores son independientes e idénticamente distribuidos según una distribución $N(0, \sigma^2)$. Se hizo la estimación según el modelo propuesto en Massuia, M. B., Cabral, C. R. B., Matos, L. A., Lachos, V. H. (2012) y también según el modelo t-Student, con el fin de comparar sus resultados.

Se aplicó el algoritmo EM desarrollado usando como valores iniciales aquellos obtenidos a través del ajuste de mínimos cuadrados de una regresión simple. Los resultados de la aplicación del algoritmo EM se presentan en la Tabla 4.3.

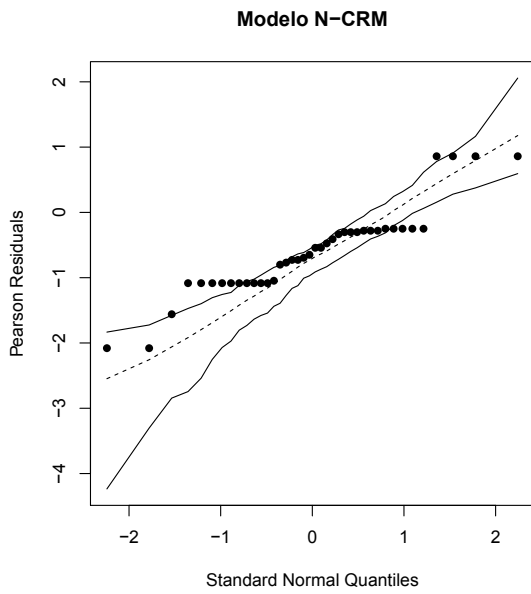
Cuadro 4.3: Resultados obtenidos para la estimación de los parámetros vía algoritmo EM

	$\hat{\beta}_0$	EP ($\hat{\beta}_0$)	$\hat{\beta}_1$	EP ($\hat{\beta}_1$)	$\hat{\sigma}^2$
Normal	-6.271	1.404	4.483	0.644	0.146
T-Student	-5.679	0.830	4.134	0.386	0.017

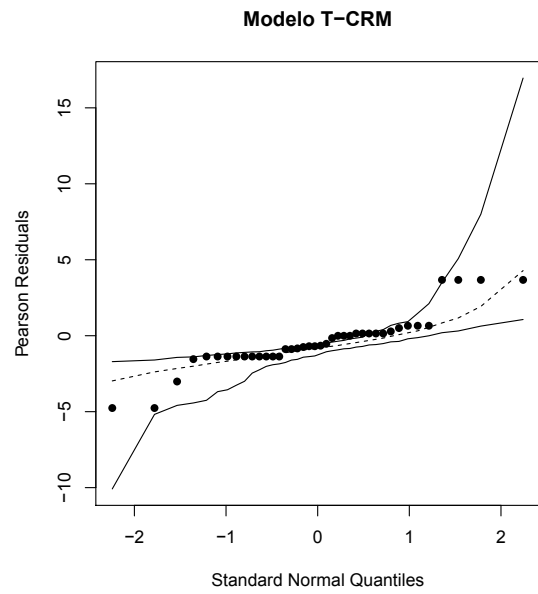
En la Figura 4.1 se presentan los gráficos de envelopes como fueron ya descritos anteriormente, en ellos podemos ver como el ajuste del modelo Normal 4.1a resultó no ser apropiado ya que hay una gran cantidad de puntos por fuera de las bandas (o envelopes) así mismo se puede apreciar la forma de S indicando que un modelo con estructura de colas más pesadas podría ser más adecuado. Este es el caso del ajuste obtenido por el modelo t-Student, (Figura 4.1b) en donde podemos ver como la totalidad de los puntos están dentro de las bandas y no se percibe presencia de asimetría ni de datos atípicos. Por estas razones se concluye que el modelo t-Student ajustó de forma más adecuada el conjunto de datos que el modelo Normal.

Es interesante ver en la Tabla 4.2 como las estimativas de los parámetros usando N-R son bastante próximas a las obtenidas mediante el algoritmo EM (Tabla 4.3), tanto para el modelo Normal como para el modelo t-Student, mostrando que estas estimativas son robustas respecto a la técnica de maximización implementada. También es interesante observar como los desvíos estándar de N-R son menores que los obtenidos con el algoritmo EM.

Como fue explicado arriba, el modelo t-Student pareció proveer un mejor ajuste de los datos que el modelo Normal, así que adoptando estas estimativas del modelo vía algoritmo EM podemos decir que para un incremento en una unidad de la variable T se va a obtener un incremento en 4,134 unidades de la variable \log_{10} (Tiempo de vida).



(a) Envelope Normal



(b) Envelope T-Student

Figura 4.1: Envelopes modelos utilizados

Bibliografía

- Atkinson, A. 1981. “Two Graphical Displays for Outlying and Influential Observation in Regression.” *BIOMETRIKA* 68, no. 13.
- Dempster, A. P. y Laird, N.M. y Rubin D. B. “Maximum likelihood from incomplete data via the EM algorithm.”
- Kim, H. J. 2008. “Moments of truncated Student-t distribution.” *Journal of the Korean Statistical Society* 37 (1): 81–87.
- Lima, E. L. 1995. *Curso de Analise Vol 2*. Rio de Janeiro: IMPA.
- Louis, T. 1982. “Finding the Observed Information Matrix When Using the EM Algorithm.” *Journal of the Royal Statistical Society, Series B*, vol. 44.
- Massuia, M. B., Cabral C. R. B. Matos L. A. Lachos V. H. 2012. “Influence diagnostics for Student-t censored linear regression models.” Technical Report, Instituto de Matemática e Estatística, Universidade Estadual de Campinas.
- Moreira, F. R. 2010. “Uma discussao sobre o metodo de Newton.” *Revista Eelectronica de Matematica* 1, no. 2.
- Tan, M.T. y Tian, G. y NG K.W. 2010. *Bayesian Missing Data Problems EM, Data Augmentation and Non Iterative Computation*.