



# Modelos de Regresión Skew Normal

Asignatura: Modelos Lineales

Bladimir Morales Torrez

Enero 2022

## Contents

<b>1</b>	<b>Introducción</b>	<b>3</b>
<b>2</b>	<b>Marco Teórico</b>	<b>3</b>
2.1	Modelo Normal . . . . .	3
2.2	Modelo Skew-Normal . . . . .	4
<b>3</b>	<b>Modelo de regresión lineal Skew Normal</b>	<b>8</b>
3.1	Supuestos . . . . .	8
3.2	Estimación del modelo . . . . .	8
3.3	Análisis de Residuos normalizados . . . . .	12
3.4	Prueba de hipótesis e Intervalos de confianza . . . . .	12
3.5	Modelo ajustado . . . . .	12

<b>4</b>	<b>Aplicación</b>	<b>13</b>
4.1	Datos e información . . . . .	13
4.2	Regresión lineal . . . . .	13
<b>5</b>	<b>Conclusiones</b>	<b>20</b>
	<b>Bibliografía</b>	<b>22</b>

# 1 Introducción

Las variables aleatorias son elementos fundamentales en la estadística y en muchas ocasiones posterior al estudio de su comportamiento se adaptan muy bien a funciones de probabilidad ya conocidas o en su defecto se empieza a construir nuevas funciones de probabilidad con sus correspondientes parámetros.

En el contexto actual existe bastante teoría consolidada sobre el modelo de distribución Normal y su implicancia en la teoría de la probabilidad, inferencia estadística (clásica y bayesiana), y así también en modelos de regresión, series de tiempo y otros. Mencionar también que muchas familias de funciones de probabilidad se aproximan a la Normal a medida que un determinado parámetro tiene a un valor apropiado.

Es así que (Azzalini, 1985) define la función de densidad Skew-Normal, propiedades y estadísticos principales, generando posteriormente bastantes investigaciones y artículos referente a esta distribución. En el prefacio del libro (Azzalini, 2013), menciona que la construcción de esta teoría que parte de una función de densidad simétrica y mediante modificaciones adecuadas de la misma conduce a generar un conjunto de distribuciones no simétricas, donde el objetivo es estudiar familias paramétricas flexibles de distribuciones continuas, permitir una posible desviación de la simetría las cuales generen familias de distribuciones más flexibles y realistas.

En el presente documento se mostrará por un lado la formulación del modelo Normal, luego su modificación de parámetros para generar el modelo asimétrica Skew-Normal (Azzalini, 1985) y su teoría en modelos de regresión lineal con errores Skew-normal, y para finalizar aplicaciones de este modelo de regresión.

## 2 Marco Teórico

### 2.1 Modelo Normal

El modelo normal es una de las más importantes y conocidos distribuciones de probabilidad continua en el campo estadístico.

#### 2.1.1 Definición

Sea  $X$  una variable aleatoria con distribución normal con media  $\mu$  y varianza  $\sigma^2$ , que se denota como  $X \sim N(\mu, \sigma^2)$ , con función de densidad dada por:

$$f(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\}, \forall x \in \mathbb{R} \quad (1)$$

donde los parámetros:  $-\infty < \mu < \infty$  y  $\sigma^2 > 0$ .

#### 2.1.2 Función generadora de momentos

Sea  $X$  una variable aleatoria con distribución  $N(\mu, \sigma^2)$ , entonces la función generadora de momentos es:

$$M_X(t) = \mathbb{E}(e^{tx}) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right) \quad (2)$$

#### 2.1.3 Distribución normal estandar

La distribución normal con  $\mu = 0$  y  $\sigma^2 = 1$  se denomina como la distribución normal estándar. Así la función de densidad es:

$$\varphi(z) = (2\pi)^{-\frac{1}{2}} \exp\left\{-\frac{z^2}{2}\right\}, \quad \forall z \in \mathbb{R} \quad (3)$$

y la función de distribución será:

$$\Phi(z) = \mathbb{P}(Z < z) = \int_{-\infty}^z \varphi(u) du, \quad \forall z \in \mathbb{R} \quad (4)$$

**Observación.** Hacer notar que cuando la tenemos una v.a.  $X \sim N(\mu, \sigma^2)$ , entonces  $Z$  es una variable aleatoria estandarizada  $Z = \frac{X-\mu}{\sigma}$  donde  $Z \sim N(0, 1)$ .

## 2.2 Modelo Skew-Normal

En principio para desarrollar el modelo Skew-Normal se debe tener en cuenta la siguiente proposición.

**Proposición 1.** Sea  $f_0$  una función de densidad de probabilidad en  $\mathbb{R}^d$ , sea  $G_0(\cdot)$  una función de distribución continua en la recta real y sea  $w(\cdot)$  una función de valor real en  $\mathbb{R}^d$ , tal que:

$$f_0(-x) = f_0(x), \quad w(-x) = -w(x), \quad G_0(-y) = 1 - G_0(y)$$

para todo  $x \in \mathbb{R}^d$ ,  $y \in \mathbb{R}$ . Entonces

$$f(x) = 2f_0(x)G_0\{w(x)\} \quad (5)$$

es una función de densidad en  $\mathbb{R}^d$ .

### 2.2.1 Definición

Si por (5) se selecciona a  $f_0 = \varphi$  y  $G_0 = \Phi$  que representan la función de densidad (3) y la función de distribución (4) respectivamente de la  $N(0, 1)$  y  $w(x) = \alpha x$ , para algún valor real  $\alpha$ , produce la función de densidad de la distribución Skew-Normal básica:

$$\varphi(x; \alpha) = 2\varphi(x)\Phi(\alpha x) \quad (-\infty < x < \infty) \quad (6)$$

Ahora se introducirá un parámetro de escala  $\omega$  y otro de localización  $\xi$ . Si  $Z$  es una variable aleatoria continua con función de densidad (6), entonces la variable  $Y$  queda expresada como:

$$Y = \xi + \omega Z \quad (\xi \in \mathbb{R}, \omega \in \mathbb{R}^+) \quad (7)$$

Así,  $Y$  es llamada variable Skew-Normal(SN) con parámetro de localización  $\xi$ , parámetro de escala  $\omega$  y parámetro de inclinación  $\alpha$ . La función de densidad en  $x \in \mathbb{R}$  de la forma (6) es:

$$\frac{1}{\omega} \varphi\left(\frac{x - \xi}{\omega}; \alpha\right) \equiv \frac{2}{\omega} \varphi\left(\frac{x - \xi}{\omega}\right) \Phi\left(\alpha \frac{x - \xi}{\omega}\right) \quad (8)$$

y se escribe:

$$Y \sim SN(\xi, \omega^2, \alpha)$$

donde el cuadrado de  $\omega$  por analogía corresponde a la notación de  $N(\mu, \sigma^2)$ .

*Observación.* Cuando  $\xi = 0$  y  $\omega = 1$  se vuelve a la definición básica de la Skew-Normal (6).

**Proposición 2.** Si  $Z$  denota una variable aleatoria  $SN(0, 1, \alpha)$ , con función de densidad  $\varphi(x; \alpha)$ , se cumplen las siguiente propiedades:

- a)  $\varphi(x; 0) = \varphi(x), \quad \forall x$
- b)  $\varphi(0; \alpha) = \varphi(0), \quad \forall \alpha$
- c)  $-Z \sim SN(0, 1, -\alpha)$ , entonces  $\varphi(-x; \alpha) = \varphi(x; -\alpha), \quad \forall x$
- d)  $\lim_{\alpha \rightarrow \infty} \varphi(x; \alpha) = 2\varphi(x)I_{[0, \infty)}(x), \quad \forall x$
- e)  $Z^2 \sim \chi_1^2$  independientemente de  $\alpha$
- f) Si  $Z' \sim SN(0, 1, \alpha')$  con  $\alpha' < \alpha$ , entonces  $Z' <_{st} Z$

### 2.2.2 Función generadora de momentos

La función generadora de momentos de  $Y$  es obtenida por:

$$\begin{aligned}
 M_Y(t) &= \mathbb{E}(e^{yt}) = \mathbb{E}(\exp\{(\xi + wz)t\}) \\
 &= 2\exp\{\xi t + \frac{1}{2}\omega^2 t^2\} \int_{-\infty}^{\infty} \varphi(z - \omega t)\Phi(\alpha z)dz \\
 &= 2\exp\{\xi t + \frac{1}{2}\omega^2 t^2\}\Phi(\delta\omega t)
 \end{aligned} \tag{9}$$

donde  $\delta = \delta(\alpha) = \frac{\alpha}{\sqrt{1+\alpha^2}}$

Luego se utiliza la función generadora cumulante:

$$K_Y(t) = \log M_Y(t) = \xi t + \frac{1}{2}\omega^2 t^2 + \zeta_0(\delta\omega t) \tag{10}$$

donde

$$\zeta_0(x) = \log\{2\Phi(x)\} \tag{11}$$

y sus derivada  $\zeta_r(x) = \frac{d^r}{dx^r} \zeta_0(x)$  con  $r = 1, 2, \dots$

así, se tiene:

$$\begin{aligned}
\zeta_1(x) &= \frac{d}{dx} \zeta_0(x) = \frac{\varphi(x)}{\Phi(x)} \\
\zeta_2(x) &= \frac{d^2}{dx^2} \zeta_0(x) = \frac{d}{dx} \zeta_1(x) \\
&= \frac{\varphi'(x)\Phi(x) - \varphi(x)^2}{\Phi(x)^2} = \frac{-x\varphi(x)}{\Phi(x)} - \frac{\varphi(x)^2}{\Phi(x)^2} = -x\zeta_1(x) - \zeta_1(x)^2 \\
\zeta_3(x) &= \frac{d^3}{dx^3} \zeta_0(x) = \frac{d}{dx} \zeta_2(x) \\
&= -\zeta_1(x) - x\zeta_1(x)' - 2\zeta_1(x)\zeta_1'(x) \\
&= -\zeta_1(x) - x[-x\zeta_1(x) - \zeta_1(x)^2] - 2[-x\zeta_1(x) - \zeta_1(x)^2]\zeta_1'(x) \\
&= -\zeta_1(x) + x^2\zeta_1(x) + x\zeta_1(x)^2 + 2x\zeta_1(x)^2 + 2\zeta_1(x)^3 \\
&= 2\zeta_1(x)^3 + 3x\zeta_1(x)^2 + x^2\zeta_1(x) - \zeta_1(x) \\
\zeta_4(x) &= \frac{d^4}{dx^4} \zeta_0(x) = \frac{d}{dx} \zeta_3(x) \\
&= -6\zeta_1(x)^2\zeta_1'(x) + 3[\zeta_1(x)^2 + 2x\zeta_1(x)\zeta_1'(x)] + [2x\zeta_1(x) + x^2\zeta_1'(x)] - \zeta_1(x)' \\
&= -6\zeta_1(x)^2[-x\zeta_1(x) - \zeta_1(x)^2] + 3\{\zeta_1(x)^2 + 2x\zeta_1(x)[-x\zeta_1(x) - \zeta_1(x)^2]\} + \\
&\quad \{2x\zeta_1(x) + x^2[-x\zeta_1(x) - \zeta_1(x)^2]\} - [-x\zeta_1(x) - \zeta_1(x)^2] \\
&= -6x\zeta_1(x)^3 - 6\zeta_1(x)^4 + 3\zeta_1(x)^2 - 6x^2\zeta_1(x)^2 - 6x\zeta_1(x)^3 + 2x\zeta_1(x) - x^3\zeta_1(x) - x^2\zeta_1(x)^2 + x\zeta_1(x) + \zeta_1(x)^2 \\
&= -6\zeta_1(x)^4 - 12x\zeta_1(x)^3 - 7x^2\zeta_1(x)^2 + 4\zeta_1(x)^2 - x^3\zeta_1(x) + 3x\zeta_1(x) \tag{12}
\end{aligned}$$

Para todo  $\zeta_r(x)$  con  $r > 1$  pueden escribirse como funciones de  $\zeta_1(x)$  y potencias de  $x$ . Notar que  $\zeta_1(x) > 0$ ,  $x + \zeta_1(x) > 0$  y  $\zeta_2(x) < 0$ .

Utilizando (10), las derivadas de  $K_Y(t)$  hasta el cuarto orden son:

$$\mathbb{E}\{Y\} = \xi + \omega\mu_z \tag{13}$$

$$Var\{Y\} = (\omega\sigma_z^2)^2 \tag{14}$$

$$\mathbb{E}\{(Y - \mathbb{E}(Y))^3\} = \frac{1}{2}(4 - \pi)(\omega\mu_z)^3 \tag{15}$$

$$\mathbb{E}\{(Y - \mathbb{E}(Y))^4\} = 2(\pi - 3)(\omega\mu_z)^4 \tag{16}$$

Donde  $\mu_z = \mathbb{E}(Z) = b\delta$ ,  $\sigma_z^2 = Var(Z) = 1 - \mu_z^2 = 1 - b^2\delta^2$  y  $b = \zeta_1(0) = \sqrt{(\frac{2}{\pi})}$

### 2.2.3 Asimetría

Para la asimetría estandarizando se tiene:

$$\gamma_1\{Y\} = \gamma_1\{Z\} = \frac{\mathbb{E}\{(Y - \mathbb{E}(Y))^3\}}{\omega^3\sigma_z^3} = \frac{4 - \pi}{2} \frac{\mu_z^3}{\sigma_z^3}$$

### 2.2.4 Kurtosis

Para la Kurtosis estandarizando se tiene:

$$\gamma_2\{Y\} = \gamma_2\{Z\} = \frac{\mathbb{E}\{(Y - \mathbb{E}(Y))^4\}}{\omega^4\sigma_z^4} = 2(\pi - 3) \frac{\mu_z^4}{\sigma_z^4}$$

### 2.2.5 Función log-verosimilitud

Si  $y$  solo un valor muestreado de una variable aleatoria  $Y \sim SN(\xi, \omega^2, \alpha)$  su función log-verosimilitud es:

$$\ell_1(\theta^{DP}; y) = cte - \log \omega - \frac{(y - \xi)^2}{2\omega^2} + \zeta_0\left(\alpha \frac{y - \xi}{\omega}\right) \quad (17)$$

donde  $\theta^{DP} = (\xi, \omega, \alpha)^t$  y  $\zeta_0(\cdot)$  esta definida como (11). El superíndice  $DP$  significa *parametros directos* (*direct parameters*).

Si  $z = \left(\frac{y - \xi}{\omega}\right)$  y  $\zeta_1(\cdot)$  es definida como (12) el componente del vector SCORE son:

$$\begin{aligned} \frac{\partial \ell_1}{\partial \xi} &= \frac{z}{\omega} - \frac{\alpha}{\omega} \zeta_1(\alpha z) \\ \frac{\partial \ell_1}{\partial \omega} &= -\frac{1}{\omega} + \frac{z^2}{\omega} - \frac{\alpha}{\omega} \zeta_1(\alpha z) z \\ \frac{\partial \ell_1}{\partial \alpha} &= \zeta_1(\alpha z) z \end{aligned} \quad (18)$$

Si tenemos una muestra aleatoria  $y_1, \dots, y_n$  de  $Y \sim SN(\xi, \omega^2, \alpha)$ , el log de verosimilitud es  $\ell(\theta)^{DP}$  se obtiene mediante la suma de  $n$  términos del tipo (17) y la correspondiente suma de términos (18) conduce a las ecuaciones de verosimilitud.

$$\begin{aligned} \sum_i z_i - \alpha \sum_i \zeta_1(\alpha z_i) &= 0 \\ \sum_i z_i^2 - \alpha \sum_i z_i \zeta_1(\alpha z_i) &= n \\ \sum_i z_i \zeta_1(\alpha z_i) &= 0 \end{aligned} \quad (19)$$

Donde  $z_i = \frac{y_i - \xi}{\omega}$ , para  $i = 1, \dots, n$ . La presencia de la función no lineal  $\zeta_1$  impide la solución de estas ecuaciones, por lo que se emplea métodos numéricos. De la tercera ecuación, la segunda requiere que  $\xi$  y  $\omega$  satisfagan:

$$\hat{\omega}^2 = \frac{1}{n} \sum_i (y_i - \hat{\xi})^2$$

que produce un hecho conocido para las variables normales.

*Observación.* Cuando  $\alpha = 0$  se reduce a la función de log-verosimilitud de la distribución normal alcanza su máximo cuando  $\xi = \bar{y}$  media muestral y  $\omega = s$  desviación estándar muestral no corregida, donde:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad s = \left( \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \right)^{-1/2}$$

Además, el punto  $\theta^{DP} = (\bar{y}, s, 0)^t$  es también una solución de la tercera ecuación (19), y por tanto es un punto estacionario de  $\ell(\theta^{DP})$ , para cualquier muestra.

Si  $\hat{\xi}(\alpha)$  y  $\hat{\omega}(\alpha)$  denotan la estimación de máxima verosimilitud (MLE) de  $\xi$  y  $\omega$ , para cualquier valor fijo de  $\alpha$  la función de log-verosimilitud del perfil es:

$$\ell^*(\alpha) = \ell(\hat{\theta}^{DP}(\alpha))$$

donde  $\hat{\theta}^{DP}(\alpha) = (\hat{\xi}(\alpha), \hat{\omega}(\alpha), \alpha)^t$  que se deriva de las ecuaciones anteriores y son eficientes para el SCORE de  $\ell^*(\alpha)$ , cuyas derivadas son:

$$\frac{d\ell^*(\alpha)}{d\alpha} = \frac{\partial \ell(\theta^{DP})}{\partial \xi} \frac{d\hat{\xi}(\alpha)}{d\alpha} + \frac{\partial \ell(\theta^{DP})}{\partial \omega} \frac{d\hat{\omega}(\alpha)}{d\alpha} + \frac{\partial \ell(\theta^{DP})}{\partial \alpha}$$

donde las derivadas parciales se evalúan en  $\theta^{DP} = \hat{\theta}^{DP}(\alpha)$ . A partir de (18), estas derivadas parciales desaparecen en  $\hat{\theta}^{DP}(0) = (\bar{y}, s, 0)^t$ , y se deduce que  $\ell^*(\alpha)$  siempre tiene un punto estacionario en  $\alpha = 0$ . Lo que implica que la prueba de Score para la  $H_0 : \alpha = 0$  no es válida al menos en su forma estándar.

### 3 Modelo de regresión lineal Skew Normal

Para el parámetro de localización  $\xi$ , que se expresa como una combinación lineal de un conjunto  $p$ -dimensional de covariables  $x$ , así,

$$\xi = x^t \beta, \quad \beta \in \mathbb{R}^p$$

Entonces se tiene el modelo de regresión lineal dado por:

$$Y_i = x_i^t \beta + \varepsilon_i, \quad i = 1, \dots, n$$

con  $\varepsilon_i \sim^{iid} SN(0, \omega^2, \alpha)$ ,  $\beta$  un vector de parámetros desconocidos.

#### 3.1 Supuestos

Para este modelo se tiene los siguientes supuestos:

- Sea una muestra aleatoria independiente de  $y = (y_1, \dots, y_n)^t$  con  $n > p$ .
- Matriz de diseño  $X = (x_1, \dots, x_n)^t$  de dimensión  $n \times p$  de rango  $p$ , donde la primera columna de  $X$  es el vector  $1_n$  (Multicolinealidad).
- Los residuos deben tener distribución Skew-Normal.
- Los residuos son independientes.
- Los residuos son homocedásticos.

#### 3.2 Estimación del modelo

En este caso, la expresión de la función del vector SCORE, será:

$$\frac{\partial \ell_1}{\partial \xi} = \left( \frac{z}{\omega} - \frac{\alpha}{\omega} \zeta_1(\alpha z) \right) x \quad (20)$$

Luego se tiene que del estimador:



$$\hat{\omega}^2 = \frac{1}{n} \sum_i (y_i - \hat{\xi}_i)^2; \quad \hat{\xi}_i = x_i^t \hat{\beta}; \quad i = 1, \dots, n$$

Cuando  $\alpha = 0$ , la suma de términos de (20) conduce a  $\sum_i z_i x_i = 0$ , las ecuaciones son las del modelos lineal clásico.

Por lo tanto un punto estacionario de la log-verosimilitud se produce en  $\hat{\theta}^{DP} = (\bar{\beta}^t, s, 0)^t$ , donde  $\bar{\beta} = (X^t X)^{-1} X^t y$  es la estimación por mínimos cuadrados y en este caso  $s$  viene dado por la desviación estándar no corregida de los residuos por mínimos cuadrados  $y - x\bar{\beta}$ .

### 3.2.1 Información de Fisher para Parámetros directos DP

Si se tiene un conjunto de  $n$  observaciones  $y = (y_1, \dots, y_n)^t$ , independientes. Las contribuciones de una sola observación a las funciones SCORE del parámetro directo  $\theta^{DP} = (\beta^t, \omega, \alpha)^t$  vienen dadas por (20) y las dos últimas expresiones de (18), respectivamente. Diferenciando la función SCORE con las sumatorias con  $n$  observaciones, se tiene:

$$\begin{aligned} -\frac{\partial^2 \ell}{\partial \beta \partial \beta^t} &= \omega^{-2} X^t (I_n + \alpha^2 Z_2) X \\ -\frac{\partial^2 \ell}{\partial \beta \partial \omega} &= \omega^{-2} X^t (2z - \alpha \zeta_1(\alpha z) + \alpha^2 Z_2 z) \\ -\frac{\partial^2 \ell}{\partial \beta \partial \alpha} &= \omega^{-1} X^t (\alpha \zeta_1(\alpha z) - \alpha Z_2 z) \\ -\frac{\partial^2 \ell}{\partial \omega^2} &= \omega^{-2} (-n + 3(1_n^t z^2) - 2\alpha \zeta_1(\alpha z)^t z + \alpha^2 \zeta_2(\alpha z)^t z^2) \\ -\frac{\partial^2 \ell}{\partial \omega \partial \alpha} &= \omega^{-1} (\zeta_1(\alpha z)^t z + \alpha \zeta_2(\alpha z)^t z^2) \\ -\frac{\partial^2 \ell}{\partial \alpha^2} &= \zeta_2(\alpha z)^t z^2 \end{aligned}$$

donde  $z = \frac{y - X\bar{\beta}}{\omega}$ ,  $Z_2 = \text{diag}(-\zeta_2(\alpha z)) > 0$

Al aplicar media a las segundas derivadas de la log-verosimilitud, se pueden determinar de manera simple los siguientes términos:

$$\mathbb{E}\{(Z^k \zeta_1(\alpha Z))\} = \frac{b}{(1 + \alpha^2)^{(k+1)/2}} \mathbb{E}(U^k)$$

Por un lado se tiene que:

$$\mathbb{E}\{(Z^k \zeta_1(\alpha Z))\} = \frac{b}{(1 + \alpha^2)^{(k+1)/2}} (1 \times 3 \times \dots \times (k-1)); \quad k = 0, 2, 4, \dots$$

y por otro:

$$\mathbb{E}\{(Z^k \zeta_1(\alpha Z))\} = \frac{b}{(1 + \alpha^2)^{(k+1)/2}} 0; \quad k = 1, 3, 5, \dots$$

donde  $U \sim N(0, 1)$  y  $b = \sqrt{2/\pi}$

Este término no es fácil de resolver:

$$a_k = a_k(\alpha) = \mathbb{E}\{Z^k \zeta_1(\alpha Z)^2\}, \quad k = 0, 1, \dots$$

que se debe calcular numéricamente para  $k = 0, 1, 2$ . Con estos elementos y recordando que  $\zeta_2(u) = -\zeta_1(u)\{u + \zeta_1(u)\}$  la matriz de información de Fisher es:

$$\mathcal{I}^{DP}(\theta^{DP}) = \begin{pmatrix} \frac{1+\alpha^2 a_0}{\omega^2} X^t X & \cdot & \cdot \\ \frac{1}{\omega} \left( \frac{b\alpha(1+2\alpha^2)}{(1+\alpha^2)^{3/2}} + \alpha^2 a_1 \right) 1_n^t X & n \frac{2+\alpha^2 a_2}{\omega^2} & \cdot \\ \frac{1}{\omega} \left( \frac{b}{(1+\alpha^2)^{3/2}} - \alpha a_1 \right) 1_n^t X & -n \frac{\alpha a_2}{\omega} & n a_2 \end{pmatrix}$$

que es una matriz simétrica.

cundo  $\alpha = 0$ , consideremos este caso con más detenimiento. La matriz de información esperada se reduce entonces a

$$\mathcal{I}^{DP}((\beta^t, \omega, 0)^t) = \begin{pmatrix} \frac{1}{\omega^2} X^t X & \cdot & \cdot \\ 0 & \frac{2n}{\omega^2} & \cdot \\ \frac{1}{\omega} b 1_n^t X & 0 & b^2 n \end{pmatrix}$$

con determinante

$$\frac{2nb^2}{\omega^{2(p+1)}} \det(X^t X) [n - 1_n^t X (X^t X)^{-1} X^t 1_n]$$

Será igual a 0, si el  $1_n$  pertenece al espacio de columnas de  $X$ .

Cuando  $\alpha = 0$ , la singularidad de la información esperada impide las aplicaciones de la teoría asintótica estándar de *MLE*. Aunque este comportamiento anómalo se limita al valor específico  $\alpha = 0$ , que corresponde al subconjunto de distribuciones normales. Por ejemplo, un problema natural a considerar es probar la hipótesis nula de que  $\alpha = 0$ , pero la metodología estándar no se aplica dada la singularidad anterior.

La singularidad de la matriz de información esperada de  $\tilde{\theta}^{DP} = (\tilde{\beta}^t, s, 0)^t$  es siempre una solución de las ecuaciones de verosimilitud. Evaluando de la segunda derivada se tiene:

$$-\frac{\partial^2 \ell}{\partial \theta^{DP} \partial (\theta^{DP} t)}|_{\theta^{DP} = \tilde{\theta}^{DP}} = \begin{pmatrix} \frac{1}{s^2} X^t X & \cdot & \cdot \\ 0 & \frac{2n}{s^2} & \cdot \\ \frac{1}{s} b 1_n^t X & 0 & b^2 n \end{pmatrix}$$

que también es singular.

### 3.2.2 Parametrización Centrada

Para resolver el problema presentado anteriormente, se introduce una reparametrización destinada a eliminar el problema de la singularidad cuando  $\alpha = 0$ , la cual es:

$$Y = \mu + \sigma Z_0, \quad Z_0 = \frac{Z - \mu_z}{\sigma_z} \sim SN\left(-\frac{\mu_z}{\sigma_z}, \frac{1}{\sigma_z^2}, \alpha\right)$$

donde

- $\sigma^2 = \text{Var}(Y) = (\omega\sigma_z)^2$
- $\mu_z = \mathbb{E}(Z) = b\delta$
- $\sigma_z^2 = \text{Var}(Z) = 1 - b^2\delta^2$
- $b = \sqrt{2/\pi}$
- $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$ .

Ahora se considera un nuevo vector de parámetros  $\theta^{CP} = (\mu, \sigma, \gamma_1)^t$ , donde CP quiere decir *parámetros centrados* (*centred parameters*), donde

$$\mu = \xi + b\omega \frac{\alpha}{\sqrt{1+\alpha^2}} = \xi + b\omega\delta(\alpha)$$

$$\sigma = \omega(1 - b^2\delta(\alpha)^2)^{1/2}$$

$$\gamma_1 = \frac{4-\pi}{2} \frac{b^3\delta(\alpha)^3}{[1 - b^2\delta(\alpha)^2]^{3/2}} = \frac{4-\pi}{2} \frac{b^3\alpha^3}{[1 + (1-b^2)\alpha^2]^{3/2}}$$

$$\omega = \frac{\delta}{(1-b^2\delta(\alpha)^2)^{1/2}} \text{ y } \xi = \mu - b\omega\delta(\alpha)$$

Una ventaja importante de la parametrización centrada, es que  $\mu$  es un parámetro de localización mucho más familiar, normalmente con una interpretación matemática más clara que  $\xi$ . Por razones similares,  $\sigma$  y  $\gamma_1$  serían a  $\omega$  y  $\alpha$ , respectivamente.

Entonces la formulación CP se establece como  $\theta^{CP} = (\beta^{CP}, \sigma, \gamma_1)$  Ahora para los parámetros se cumple  $\beta^{CP} = \beta^{DP}$ , excepto el primer parámetro intercepto  $\beta_0$ , de forma que:

$$\beta_0^{CP} = \beta_0^{DP} + \omega\mu_z$$

que coincide con  $\mathbb{E}(Y) = \xi + \omega\mu_z$ .

Luego las matrices de información esperada y observada para CP se obtienen a partir de:

$$\mathcal{I}^{CP}(\theta^{CP}) = D^t \mathcal{I}^{DP}(\theta^{DP}) D, \quad \mathcal{I}^{CP}(\hat{\theta}^{CP}) = \hat{D}^t \mathcal{I}^{DP}(\hat{\theta}^{DP}) \hat{D}$$

donde  $D$  denota la matriz Jacobiana:

$$D = D_{rs} = \left( \frac{\partial \theta_r^{DP}}{\partial \theta_s^{CP}} \right) = \begin{pmatrix} 1 & 0 & -\frac{\mu_z}{\sigma_z} & \frac{\partial \xi}{\partial \gamma_1} \\ 0 & I_{p-1} & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma_z} & \frac{\partial \omega}{\partial \gamma_1} \\ 0 & 0 & 0 & \frac{d\alpha}{d\gamma_1} \end{pmatrix}$$

y  $\hat{D}$  denota a  $D$  evaluado en el punto de MLE.

### 3.2.3 Distribución asintótica de MLE

Se ha visto que la matriz de información cuando  $\alpha = 0$  es singular, lo que viola una de las condiciones estándar para la normalidad asintótica de la MLE. Este tipo de situación cae bajo el paraguas de la teoría asintótica no estándar desarrollada por (Rotnitzky et al., 2000).

Es así que el  $CP$  se comporta regularmente en el punto  $\gamma_1 = 0$ , de dos maneras:

- (a) el perfil log-verosimilitud de  $\gamma_1$  no tiene ningún punto estacionario en  $\gamma_1 = 0$
- (b) si  $\theta^{CP} = (\mu, \sigma, 0)$ , entonces

$$\sqrt{n}(\hat{\theta}^{CP} - \theta^{CP}) \xrightarrow{d} N_3 \left( 0, \text{diag}(\sigma^2, \frac{1}{2}\sigma^2, 6) \right)$$

### 3.3 Análisis de Residuos normalizados

Los residuos de la regresión estan dados por:

$$\hat{z}_i = \frac{(y_i - \hat{\xi})}{\hat{\omega}}; \quad i = 1, \dots, n$$

tal que  $z_i$  debe ser muestreada aproximadamente como una distribución  $\chi_1^2$ , recordando que:

$$Z^2 = \frac{(Y - \xi)^2}{\omega^2} \sim \chi_1^2$$

### 3.4 Prueba de hipótesis e Intervalos de confianza

Recordemos que, en un problema de estimación regular de  $k$  parámetros, el conjunto de valores de los parámetros que tienen una deviance no mayor al cuantil  $p$ -ésimo de la distribución  $\chi_k^2$  delimitan una región de confianza con un nivel de confianza  $p$ . En consecuencia, la deviance puede utilizarse para la prueba de hipótesis donde se evaluará si un punto nominado pertenece a la región de confianza. En el presente escenario, la no regularidad de la deviance antes mencionada afecta a la construcción de conjuntos de confianza, que en el presente caso con  $k = 1$  suelen corresponder a intervalos.

El parámetro  $\alpha = 0$  se produce cuando  $D(\alpha)$  es bastante grande, por lo tanto es una región que no interfiere efectivamente en la construcción de intervalos de confianza y pruebas de hipótesis, al menos para los niveles de confianza habitualmente considerados. Sin embargo, es evidente que el problema sigue existiendo para otros conjuntos de datos con  $\hat{\alpha}$  más cercanos a 0, para los que  $D(0)$  no será tan grande.

En general la función de log-verosimilitud asociada a las variables de la  $SN$  es algo inusual.

### 3.5 Modelo ajustado

Pot lo visto anteriormente se tiene el modelo de regresión lineal estimado y ajustado tomando en cuenta los parámetros  $CP$  y esta dado por:

$$\hat{Y}_i = x_i^t \hat{\beta} + \frac{b\delta}{\sqrt{1 - b^2\delta^2}}, \quad i = 1, \dots, n$$

con  $\varepsilon_i \sim^{iid} SN(0, \sigma^2, \gamma_1)$  *iid*,  $\beta$

## 4 Aplicación

Para poder mostrar de manera práctica lo visto en secciones anteriores se realizará algunas simulaciones con el software R de las distribuciones (Azzallini, 2017) y así también un modelo de regresión lineal para errores skew-normal, con la librería `sn` que muestra bastantes aplicaciones (Azzallini, 2021).

### 4.1 Datos e información

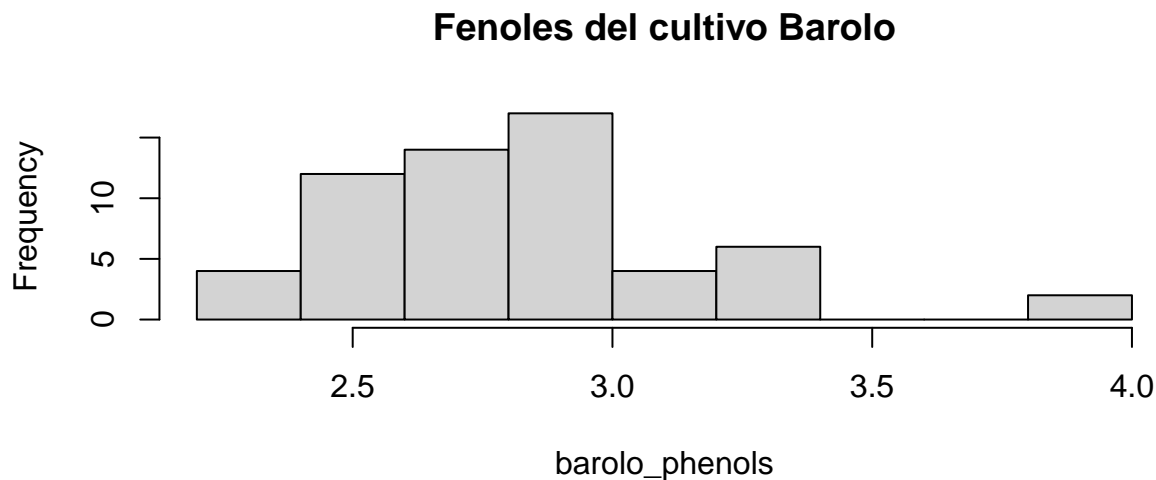
Se utiliza la base de datos que esta en R `data(wines)` que hace referencia a tres cultivos (Barolo, Barbera y Grignolino) de la región de Piedmont de Italia, que contiene 178 diferentes vinos italianos con 59, 48 y 71 datos respectivamente y se recogieron 28 medidas químicas de cada clase de vino.

### 4.2 Regresión lineal

#### 4.2.1 Datos 1

Para este primer paso de la modelación solo se trabajará con el cultivo Barolo y la variable fenoles (phenols) que será nuestra variable de respuesta y se pondrá como covariable a la constante 1 esto con el cometido de visualizar el comportamiento de la  $Y$ .

```
data(wines)
barolo_phenols <- wines[wines$wine=="Barolo", "phenols"]
hist(barolo_phenols, main="Fenoles del cultivo Barolo")
```



Se modelará en primera instancia con un modelo de regresión clásico y luego con un modelo de regresión Skew-Normal con tipo de parametrización  $DP$ .

```
fit <- lm(barolo_phenols ~ 1)
fit_sn <- selm(barolo_phenols ~ 1, family="SN")
summary(fit)
```

```
##
## Call:
## lm(formula = barolo_phenols ~ 1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.64017 -0.24017 -0.04017  0.15983  1.03983
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.84017    0.04413   64.36  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.339 on 58 degrees of freedom
```

```
summary(fit_sn,param.type="DP")
```

```
## Call: selm(formula = barolo_phenols ~ 1, family = "SN")
## Number of observations: 59
## Family: SN
## Estimation method: MLE
## Log-likelihood: -16.73059
## Parameter type: DP
##
## DP residuals:
##      Min       1Q   Median       3Q      Max
## -0.2426  0.1574  0.3574  0.5574  1.4374
##
## Regression coefficients
##      estimate std.err z-ratio Pr{>|z|}
## xi  2.44261  0.08074 30.25194      0
##
## Parameters of the SEC random component
##      estimate std.err
## omega  0.5206  0.078
## alpha  3.2481  1.926
```

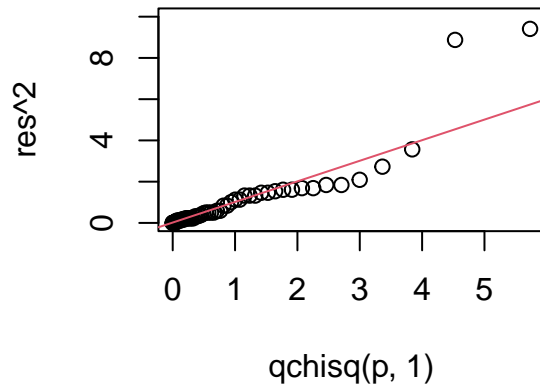
Ahora vemos el comportamiento que tiene sus residuos al cuadrado con el qq-plot.

```
res<-(barolo_phenols-fit$fitted.values)/0.339

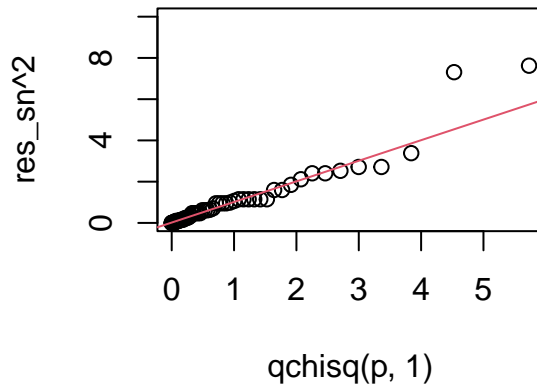
xi<-coef(fit_sn,param.type="DP")[1]
omega<-coef(fit_sn,param.type="DP")[2]
res_sn<-(barolo_phenols-xi)/omega
p<-(1:59)/60

par(mfrow=c(1,2))
qqplot(qchisq(p,1),res^2,ylim = c(0,10),main = "QQ-Plot residuos Normales")
qqline(res_sn^2,distribution = function(p) qchisq(p, df = 1),probs = c(0.215, 0.5), col = 2)
qqplot(qchisq(p,1),res_sn^2,ylim = c(0,10),main = "QQ-Plot residuos SN")
qqline(res_sn^2,distribution = function(p) qchisq(p, df = 1),probs = c(0.215, 0.5), col = 2)
```

**QQ-Plot residuos Normales**



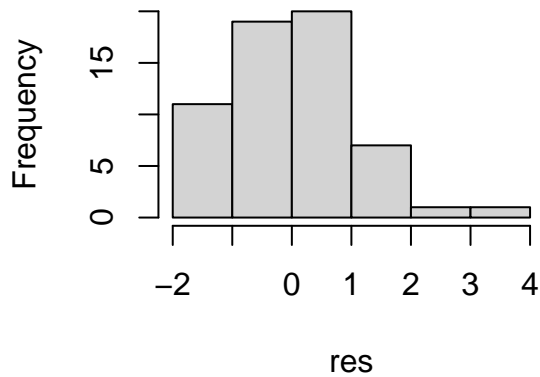
**QQ-Plot residuos SN**



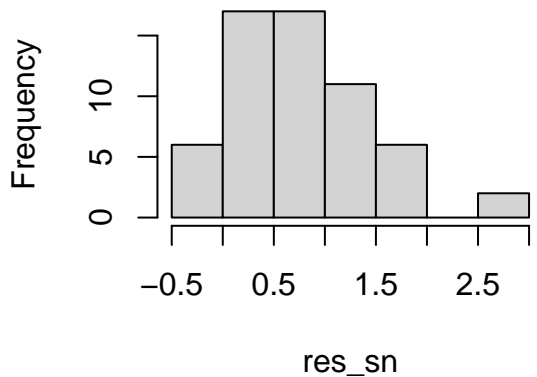
y sus respectivos histogramas.

```
par(mfrow=c(1,2))
hist(res,main = "Residuos Normales")
hist(res_sn ,main = "Residuos SN")
```

**Residuos Normales**



**Residuos SN**



#### 4.2.2 Datos 2

Para este segundo caso se trabajará con el cultivo Grignolino y la variable proantocianinas (proanthocyanins) que será nuestra variable de respuesta y se pondrá como covariable a la variable nitrógeno (nitrogen) esto con el cometido de visualizar el comportamiento de la Y.

```
vinos<-tibble(cultivo=wines$wine,nitrogeno=wines$nitrogen,proantocianinas=wines$proanthocyanins)
vinos<-vinos %>% filter(vinos=="Grignolino")
mod<-lm(proantocianinas~nitrogeno,data = vinos)
```

```
mod_sn<-selm(proantocianinas~nitrogeno,data = vinos,family = "SN")
summary(mod)
```

```
##
## Call:
## lm(formula = proantocianinas ~ nitrogeno, data = vinos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.12577 -0.28621 -0.06955  0.27961  1.88217
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.0253465  0.2089787   9.692 1.67e-14 ***
## nitrogeno   -0.0016710  0.0008329  -2.006  0.0488 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5895 on 69 degrees of freedom
## Multiple R-squared:  0.05512,    Adjusted R-squared:  0.04142
## F-statistic: 4.025 on 1 and 69 DF,  p-value: 0.04876
```

```
summary(mod_sn,param.type="DP")
```

```
## Call: selm(formula = proantocianinas ~ nitrogeno, family = "SN", data = vinos)
## Number of observations: 71
## Family: SN
## Estimation method: MLE
## Log-likelihood: -58.9809
## Parameter type: DP
##
## DP residuals:
##      Min       1Q   Median       3Q      Max
## -0.4723  0.3027  0.5388  0.8700  2.4800
##
## Regression coefficients
##              estimate    std.err    z-ratio Pr{>|z|}
## (Intercept.DP)  1.5473102  0.2031164  7.6178496   0.000
## nitrogeno      -0.0022822  0.0008962 -2.5466338   0.011
##
## Parameters of the SEC random component
##      estimate std.err
## omega  0.8532  0.110
## alpha  2.4925  0.963
```

```
summary(mod_sn,param.type="CP")
```

```
## Call: selm(formula = proantocianinas ~ nitrogeno, family = "SN", data = vinos)
## Number of observations: 71
## Family: SN
```

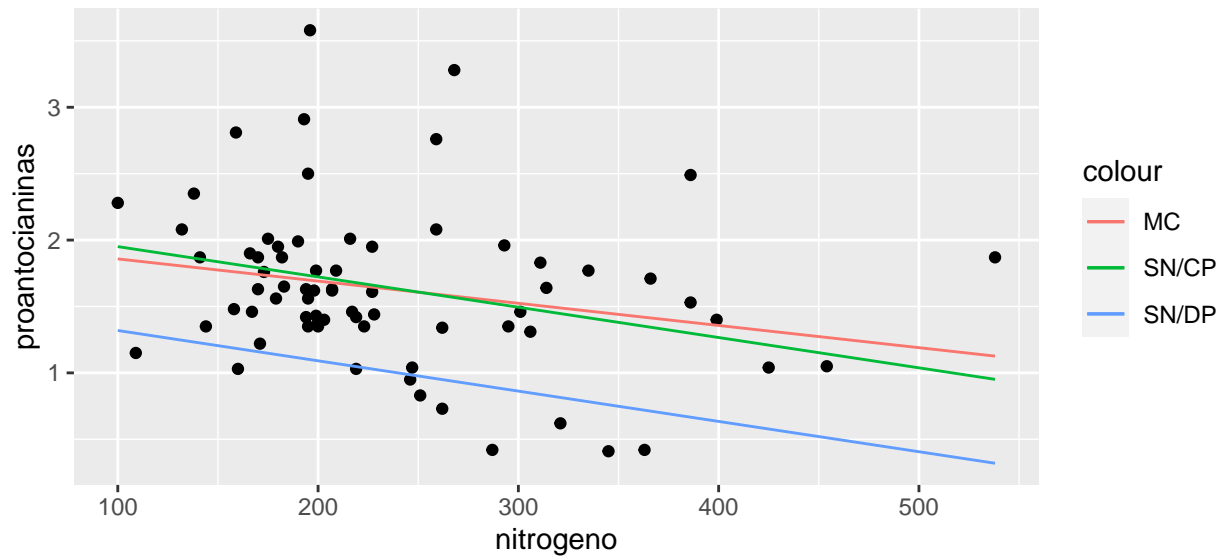


```
## Estimation method: MLE
## Log-likelihood: -58.9809
## Parameter type: CP
##
## CP residuals:
##      Min      1Q   Median      3Q      Max
## -1.10408 -0.32905 -0.09295  0.23820  1.84824
##
## Regression coefficients
##              estimate    std.err    z-ratio Pr{>|z|}
## (Intercept.CP)  2.1790802  0.2227445  9.7828671    0.000
## nitrogeno      -0.0022822  0.0008962 -2.5466338    0.011
##
## Parameters of the SEC random component
##      estimate std.err
## s.d.      0.5734  0.052
## gamma1    0.5742  0.204
```

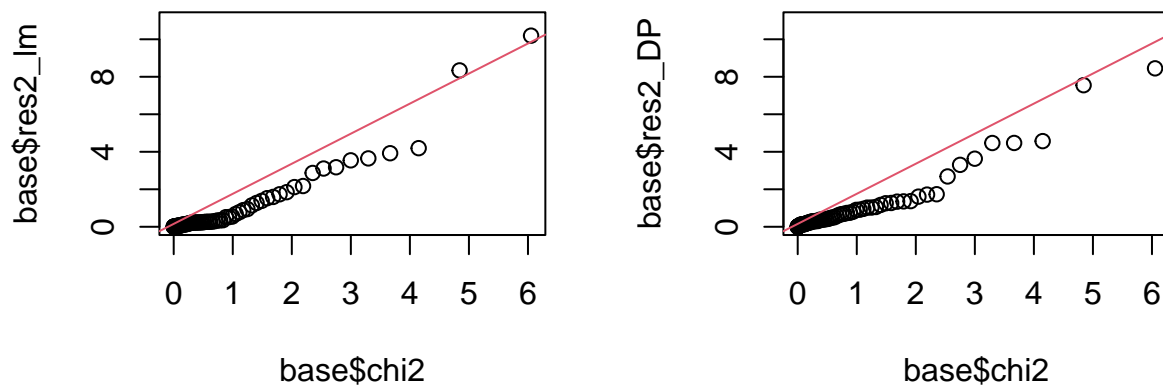
```
omega_dp<-coef(mod_sn,param.type="DP")[3]
sigma_cp<-coef(mod_sn,param.type="CP")[3]

base<-vinos %>%
  mutate(fit_lm=fitted(mod),
         SN_DP=fitted.selm(mod_sn,param.type = "DP"),
         SN_CP=fitted.selm(mod_sn,param.type = "CP"),
         res_lm=(proantocianinas-fit_lm)/0.5895,
         res_DP=(proantocianinas-SN_DP)/omega_dp,
         res_CP=(proantocianinas-SN_DP)/sigma_cp,
         res2_lm=res_lm^2,
         res2_DP=res_DP^2,
         res2_CP=res_CP^2,
         p=(1:nrow())/72,
         chi2=qchisq(p,1))
```

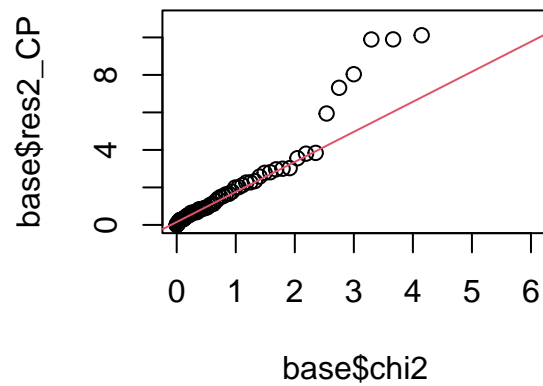
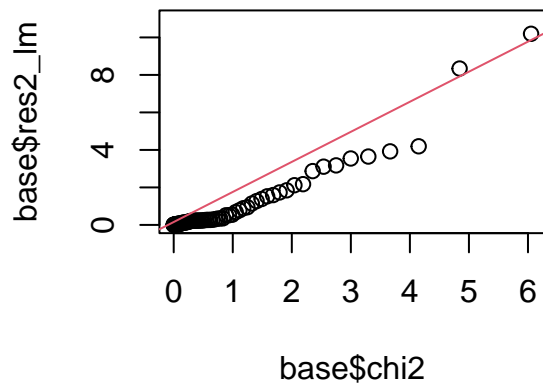
```
base %>%
  ggplot(aes(nitrogeno,proantocianinas))+
  geom_point()+
  geom_line(aes(nitrogeno,fit_lm,color="MC"))+
  geom_line(aes(nitrogeno,SN_DP,color="SN/DP"))+
  geom_line(aes(nitrogeno,SN_CP,color="SN/CP"))
```



```
par(mfrow=c(1,2))
qqplot(base$chi2,base$res2_lm,ylim=c(0,11))
qqline(base$res2_CP,distribution = function(p) qchisq(p, df = 1),probs = c(0.215, 0.5), col = 2)
qqplot(base$chi2,base$res2_DP,ylim=c(0,11))
qqline(base$res2_CP,distribution = function(p) qchisq(p, df = 1),probs = c(0.215, 0.5), col = 2)
```



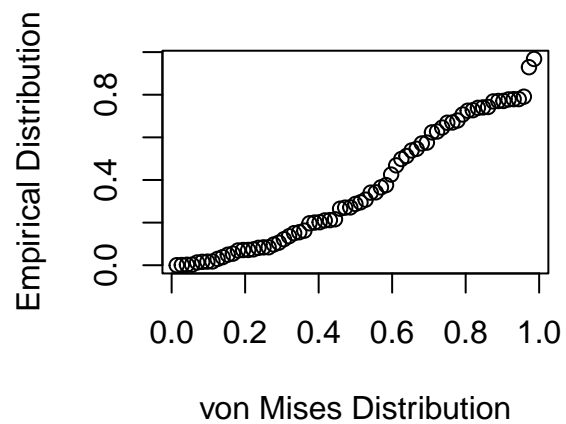
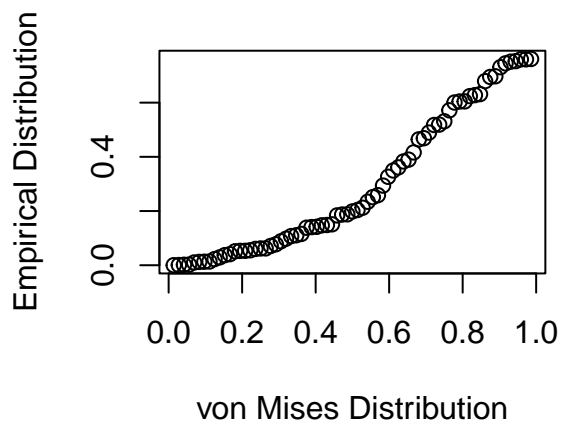
```
par(mfrow=c(1,2))
qqplot(base$chi2,base$res2_lm,ylim=c(0,11))
qqline(base$res2_CP,distribution = function(p) qchisq(p, df = 1),probs = c(0.215, 0.5), col = 2)
qqplot(base$chi2,base$res2_DP,ylim=c(0,11))
qqline(base$res2_CP,distribution = function(p) qchisq(p, df = 1),probs = c(0.215, 0.5), col = 2)
```



```
par(mfrow=c(1,2))
pp.plot(base$res2_DP,ref.line = F)
```

```
##          mu      kappa
## 1 0.4973523 2.270257
```

```
pp.plot(base$res2_CP,ref.line = F)
```

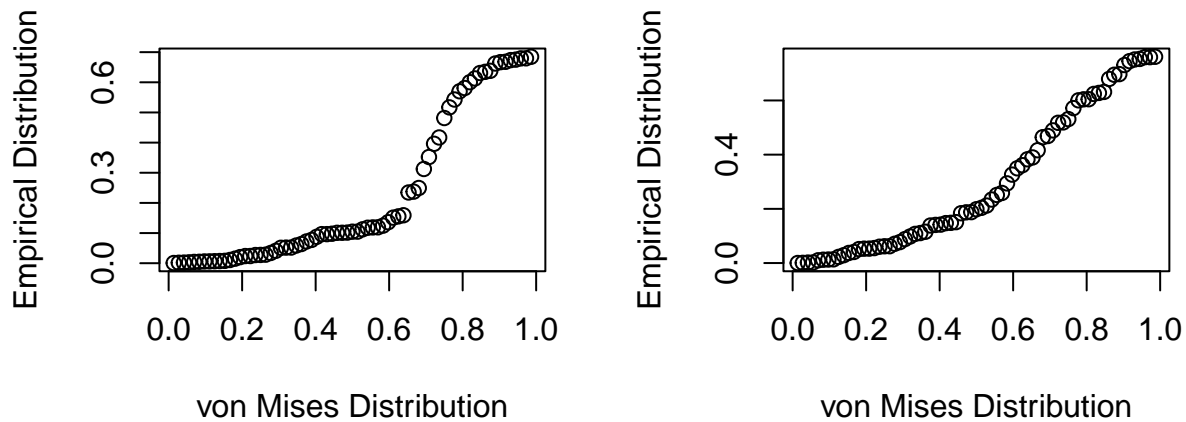


```
##          mu      kappa
## 1 0.8791769 1.167279
```

```
par(mfrow=c(1,2))
pp.plot(base$res2_lm,ref.line = F)
```

```
##          mu      kappa
## 1 0.3818201 1.734008
```

```
pp.plot(base$res2_DP,ref.line = F)
```



```
##          mu      kappa
## 1 0.4973523 2.270257
```

Estimadores.

	$\beta_0^{DP}$	$\beta_1$	$\omega$	$\alpha$	$\beta_0^{CP}$	$\sigma$	$\gamma_1$
estimate	1.547	-0.00228	0.853	2.49	2.179	0.573	0.57
std. err.	0.203	0.00089	0.110	0.96	0.223	0.052	0.20

Así se puede ver que en los estimadores los *CP* en su parámetro  $\beta_0$  cambia radicalmente con el de *DP* y las varianzas estimadas son mucho menores para el modelo estimado con *CP*. Concluyendo así que el modelo Skew-Normal con parametrización centrada produce mejores resultados que los otros modelos planteados.

## 5 Conclusiones

Se hizo una revisión bibliográfica bastante extensa de las distribuciones asimétricas (Skew), encontrando así diferentes investigaciones de diferentes autores, donde el principal precursor y creador es Azzalini, el cual desde 1985 presentó esta nueva familia de distribuciones asimétricas. Inicio con la distribución normal estándar, encontrando sus momentos, luego realizando inferencia encontró sus estimadores y tuvo aproximaciones bayesianas. De la misma forma trabajo con la distribución *t* de Student tomando en cuenta para ambos casos parámetros de centralidad. Pasó luego a la parte multivariable de las distribuciones e investigó sobre las distribuciones elípticas (colas pesadas).

En este trabajo se contextualizó las distribuciones Skew-Normal, se mostró los momentos de las distribuciones, la función log de verosimilitud y su respectiva teoría de modelación en regresión lineal.

Al finalizar se programó la simulación de la distribución skew-normal con el software R y se indagó respecto a la librería **sn** la cual cuenta con bastantes funciones relacionadas a las distribuciones asimétricas simulando

como ejemplos las distribuciones skew-normal. También se pudo comparar tres modelos de regresión lineal, el primero modelo lineal general, luego un modelo lineal skew-normal con parámetros  $CP$  y  $DP$  encontrando así un mejor ajuste a los datos con el modelo Skew-Normal con parámetros de estimación  $DP$ .

## Bibliografia

- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2), 171–178.
- Azzalini, A. (2013). *The skew-normal and related families*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139248891>
- Azzalini, A. (2017). *An introduction to the r package sn*.
- Azzalini, A. (2021). Package sn, the skew-normal and related distributions such as the skew-t and the SUN. *CRAN*.