Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Estatística

Relatório Técnico – RTP 04/2013

INFERÊNCIA EM MODELOS LINEARES ASSIMÉTRICOS

Bárbara da Costa C. Dias
Dr. Clécio da Silva Ferreira

Juiz de Fora
2013

RESUMO

Este trabalho tem por objetivo analisar o comportamento de combinações de parâmetros de uma distribuição Normal Assimétrica em relação aos resultados de testes de hipóteses feitos com os parâmetros λ e β_1.

Para isso, foram utilizados quatro testes assintóticos: o teste da Razão de Verossimilhanças, teste usando o Intervalo de Confiança e os testes de normalidade Shapiro-Wilk (1965) e Lilliefors (1967).

Com o intuito de estudar melhor a influência desses parâmetros em tais testes, estudamos os momentos da distribuição normal assimétrica, através dos valores encontrados para o valor esperado, variância, coeficiente de assimetria e curtose.

Foi verificado que o coeficiente de assimetria de uma distribuição Normal Assimétrica só envolve o parâmetro λ e a curtose envolve todos os parâmetros. Através do teste de hipótese, foi verificado que um baixo valor para β_1, leva a uma significância do teste.

1- INTRODUÇÃO

A distribuição Normal Assimétrica consegue modelar a assimetria dos dados, tendo como caso particular a distribuição Normal. Essa distribuição começou a ser discutida formalmente através de Azzalini (1985), estudando suas propriedades e mostrando os problemas que o parâmetro de assimetria causa para os métodos de estimação mais usados (métodos dos momentos e de máxima verossimilhança).

Posteriormente Azzalini (2005) apresentou uma discussão em distribuições normais assimétricas com aplicações em modelos de regressão. Generalizações para o caso multivariado dessas ideias têm sido propostas por vários autores, por exemplo, Azzalini e Dalla-Valle (1996), Azzalini e Capitanio (1999).

Testes de hipóteses paramétricos são utilizados para verificar, através de uma amostra, se os dados estudados são compatíveis com alguma hipótese de interesse.

Um dos métodos mais convencionais para se testar uma hipótese, é a utilização do Intervalo de Confiança, que visa estimar um intervalo de parâmetros, com um nível determinado de confiança.

Shapiro e Wilk (1965) propôs um teste de normalidade, ou seja, a hipótese nula é de que a distribuição testada tem uma distribuição normal. O teste recebeu uma atenção considerável na literatura, e sua distribuição assintótica foi abrangida pelos resultados de Wet e Wenter (1973), e foi recentemente estudada por Sen (2002). Outro teste de normalidade é o de Lilliefors (1967), que é uma adaptação do teste Kolmogorov-Smirnov.

O teste da Razão de Verossimilhanças, que foi desenvolvido por Fan et alii (2001) é um teste de hipótese bastante utilizado, sua estatística de teste é construída a partir da função log-verossimilhanças sob o modelo irrestrito e sob o modelo restrito. Para grandes amostras, tem-se que a estatística de teste da Razão de Verossimilhanças segue para uma distribuição qui-quadrado com 1 grau de liberdade.

O objetivo desse trabalho é apresentar um estudo que relaciona combinações de parâmetros, variâncias dos estimadores e tamanhos de amostras, obtendo assim, uma alta taxa de rejeição dos testes de hipóteses.

2- MODELO DE REGRESSÃO COM ERROS ASSIMÉTRICOS

2.1- DRITRIBUIÇÃO NORMAL ASSIMÉTRICA PADRÃO

Uma variável aleatória $Z \sim SN(\lambda)$, é dita ser normal assimétrica padrão, com parâmetro de assimetria λ, se apresentar a seguinte função de densidade de probabilidade:

$$f_Z(z) = 2\phi(z)\Phi(\lambda z), \qquad z \in \mathbb{R} \tag{1}$$

onde $\phi(.)\ e\ \Phi(.)$ são as funções de densidade de probabilidade e de distribuição de uma normal padrão, respectivamente.

A função de distribuição associada à densidade (1) é denotada por $\mathrm{F_Z(z;\ \lambda)}$ e dada por

$$F_Z(z;\ \lambda) = 2\Phi_2(z, 0|\Omega), \qquad com\ \ \Omega = \begin{bmatrix} 1 & -\delta \\ -\delta & 1 \end{bmatrix}, \qquad \delta = \frac{\lambda}{\sqrt{1+\lambda^2}}\ , \qquad z \in \mathbb{R}$$

sendo $\Phi_2(.|\Omega)$ a função de distribuição de uma normal bivariada com média zero e matriz de variância Ω.

Essa distribuição é uniparamétrica, e seu parâmetro λ representa a assimetria da sua função de densidade.

A densidade em (1) possui algumas propriedades interessantes, cujas provas podem ser obtidas em Azzalini (1985) e Azzalini (2004). Uma das propriedades mais importante é a seguinte:

$$\text{Se } Z \sim SN(\lambda) \text{ então } Z^2 \sim \chi_1^2 \tag{2}$$

2.2- DRITRIBUIÇÃO NORMAL ASSIMÉTRICA LOCAÇÃO-ESCALA

Estendendo o modelo citado anteriormente, é apresentada uma variável aleatória $Y \sim SN(\mu, \sigma^2, \lambda)$, com distribuição normal assimétrica locação-escala, que possui três parâmetros, de locação μ, de escala σ^2 e de assimetria λ. Sua função de densidade de probabilidade é:

$$f_Y(y) = \frac{2}{\sigma}\,\phi\left(\frac{y-\mu}{\sigma}\right)\Phi\left(\lambda\frac{y-\mu}{\sigma}\right),\ y \in \mathbb{R}. \tag{3}$$

A função de distribuição de (3), é denotada por $F_Y(y;\ \mu, \sigma^2, \lambda)$ e dada por

$$F_Y(y;\ \mu, \sigma^2, \lambda) = 2\Phi_2\left(\frac{y-\mu}{\sigma}, 0 \middle| \Omega\right), \text{com } \Omega = \begin{bmatrix} 1 & -\delta \\ -\delta & 1 \end{bmatrix}, \delta = \frac{\lambda}{\sqrt{1+\lambda^2}}, z \in \mathbb{R}$$

sendo $\Phi_2(.\,|\Omega)$ a função de distribuição de uma normal bivariada com média zero e matriz de variância Ω.

A media e a variância de uma variável aleatória $Y \sim SN\ (\mu, \sigma^2, \lambda)$, são expressas por,

$$E(Y) = \mu + \sigma c\rho \quad e \quad Var\,(Y) = \sigma^2(1 - c^2\rho^2).$$

Com $\rho = \frac{\lambda}{\sqrt{1+\lambda^2}}$ e $c = \sqrt{\frac{2}{\pi}}$.

Resultado 1:

Sejam X e Y duas variáveis aleatórias tais que $X \sim SN\ (\mu, \sigma^2, \lambda)$ e $Y = a + bZ, a\ e\ b \in \mathbb{R}$. Então

$$Y \sim SN\ (a + b\mu, b^2\sigma^2, sinal(b)\lambda), \tag{4}$$

Onde $sinal(x) = 1\ se\ x \geq 0\ e\ sinal(x) = -1\ se\ x < 0$. A prova desse resultado pode ser encontrado em Rodríguez(2005).

2.3- MODELO DE REGRESSÃO NORMAL ASSIMÉTRICO

Considere $y_1, \ldots, y_n$ um conjunto de *n* observações independentes. Associado à i-ésima observação, considere o preditor linear $\mu_i = \boldsymbol{x}_i^T\boldsymbol{\beta}$, onde β é um vetor *p*-dimensional de coeficientes de regressão desconhecidos e considerando um vetor *px1* de covariáveis x_i. Temos então o seguinte modelo de regressão

$$\begin{array}{ccc} y_i & = & \boldsymbol{x}_i^T\boldsymbol{\beta} + \varepsilon_i\,, \\ \varepsilon_i & \sim & SN\ (0, \sigma^2, \lambda), \end{array} \quad i = 1, \ldots, n, \tag{5}$$

onde ε_i são erros independentes. Logo $Y_i \sim SN\ (\boldsymbol{x}_i^T\boldsymbol{\beta}, \sigma^2, \lambda)\ \cdot, i = 1, \ldots, n$. Note que $E(\varepsilon_i) = c\rho\ \neq 0\,,\ para\ \lambda \neq 0$. Então para que Y_i seja não-viesado, basta considerar a $\varepsilon_i \sim SN\ (-c\sigma\rho, \sigma^2, \lambda)$.

2.4- MOMENTOS DE UMA NORMAL ASSIMÉTRICA

O k-ésimo momento de uma variável aleatória com distribuição normal assimétrica $Y \sim SN\ (\mu, \sigma^2, \lambda)$ é calculado através da seguinte integral

$$E[Y^K] = \int_{-\infty}^{+\infty} y^k\ \frac{2}{\sigma^2}\ \phi\left(\frac{y-\mu}{\sigma^2}\right)\ \Phi\left(\frac{y-\mu}{\sigma^2}\ \lambda\right) \mathrm{dy}$$

Como resolver essa integral é considerávelmente complicado, é recomendável lidar primeiramente com a distribuição normal assimétrica padrão $Z \sim SN(\lambda)$

O seguinte lema (dado em Lin et al., 2007) fornece uma forma simples de obter momentos superiores da normal assimétrica sem usar sua função geradora de momentos.

Lema 1. Se $Y \sim SN\ (\mu, \sigma^2, \lambda)$ e $X \sim N\ \left(\mu, \frac{\sigma^2}{1+\lambda^2}\right)$ então

(i) $E(X^{n+1}) = \mu\ E(X^n) + \frac{\sigma^2}{1+\lambda^2} \frac{d}{d\mu} E(X^n)$

(ii) $E(Y^{n+1}) = \mu\ E(Y^n) + \sigma^2 \frac{d}{d\mu} E(Y^n) + c\rho\sigma\ E(X^n)$

(iii) $E[\ Y - E(Y)\ ]^{n+1} = \sigma^2 \frac{d}{d\mu} E[\ Y - E(Y)\ ]^n + n\sigma^2\ E[\ Y - E(Y)\ ]^{n-1} - E[\ E(Y) - \mu\ ]\ E[\ Y - E(Y)\ ]^n + c\rho\sigma\ E[\ X - E(Y)\ ]^n$

A seguinte proposição (dado em Martínez *et al.*, 2008) fornece uma forma simples de obter os momentos ímpares da normal assimétrica.

Proposição 1 . Seja $Y \sim SN\ (0, 1, \lambda)$. Então

$$\mu_{2n+1} = \mathrm{E}\left(\mathrm{Y}^{2n+1}\right) 2\mathrm{n}\mu_{2n-1} + \sqrt{\frac{2}{\pi}}\ \frac{(2n)!}{2^n\ (n)!}\ \frac{\lambda}{1+\ \lambda^{2^{n+1/2}}}\ , \qquad \text{para n} = 1,2,\ldots$$

sendo, $\mu_1 = c\rho$.

Usando a Proposição 1 e o resultado dado em (2) de que $Z^2 \sim \chi_1^2$, com $Z \sim SN(\lambda)$, pode-se de forma simples encontrar os seguintes resultados para os primeiros quatro momentos de uma normal assimétrica padrão:

$$
\begin{aligned}
&\text{E(Z)}= c\rho \\
&\text{E(Z}^2\text{)}= 1 \\
&\text{E(Z}^3\text{)}= c\rho\left[2+\frac{1}{1+\lambda^2}\right] \\
&\text{E(Z}^4\text{)}= 3
\end{aligned} \tag{6}
$$

Usando a o resultado 1 dado em (4), pode-se dizer que $Y = \mu + \sigma Z$, com $Y \sim NA\ (\mu, \sigma^2, \lambda)$ e $Z \sim SN(\lambda)$, logo podemos usar os primeiros quatro momentos da normal assimétrica padrão, juntamente com o Lema 1 para encontrarmos os primeiros quatro momentos da normal assimétrica locação- escala, da seguinte forma:

E(Y)= E($\mu + \sigma$ Z)= $\mu + \sigma$ E(Z)

E(Y^2)= E(($\mu + \sigma$ Z)^2)

E(Y^3)= E(($\mu + \sigma$ Z)^3)

E(Y^4)= E(($\mu + \sigma$ Z)^4)

Resultado 2: Usando o resultado em (6), é possivel encontrar os seguintes resultados:

$$
\begin{gathered}
E(Y) = \mu + c\sigma\rho, \\
E(Y^2) = \mu^2 + 2c\mu\sigma\rho + \sigma^2, \\
E(Y^3) = \mu^3 + 3c\mu^2\sigma\rho + 3\mu\sigma^2 + 3c\sigma^3\rho + c\sigma^3\rho^3, \\
E(Y^4) = \mu^4 + 4c\mu^3\sigma\rho + 6\mu^2\sigma^2 + 12c\mu\sigma^3\rho - 4c\mu\sigma^3\rho^2 + 3\sigma^4.
\end{gathered} \tag{7}
$$

2.5- MATRIZ DE INFORMAÇÃO DE FISHER OBSERVADA

Considerando $Y_1, \ldots, Y_n$ um conjunto de *n* observações independentes, com distribuição $Y_i \sim NA\ (\mu_i, \sigma^2, \lambda)$, $\mu_i = x_i^T\beta,\ i = 1, \ldots, n$. Sendo β um vetor p-dimensional de coeficientes de regressão desconhecidos.

Seja $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \lambda)^T$, a função de log-verossimilhança $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta})$ é da forma $\ell_i(\boldsymbol{\theta}) = log2 + \ell_{1i}(\boldsymbol{\theta}) + \log\left[\Phi\left(\ell_{2i}(\boldsymbol{\theta})\right)\right]$, onde $\ell_{1i}(\boldsymbol{\theta}) = \phi(y_i|\boldsymbol{x}_i{}^T\boldsymbol{\beta}, \sigma^2)$ e $\ell_{2i}(\boldsymbol{\theta}) = \lambda\frac{y_i - \boldsymbol{x}_i{}^T\boldsymbol{\beta}}{\sigma}$.

Logo, a primeira derivada de $\ell_i(\boldsymbol{\theta})$é dada por

$$\frac{\partial\ \ell_i(\boldsymbol{\theta})}{\partial\psi} = \frac{\partial\ \ell_{1i}(\boldsymbol{\theta})}{\partial\psi} + W_\Phi\left(\ell_{2i}(\boldsymbol{\theta})\right)\frac{\partial\ \ell_{2i}(\boldsymbol{\theta})}{\partial\psi}, \psi = \boldsymbol{\beta}, \sigma^2, \lambda$$

E a segunda derivada é

$$\frac{\partial^2\ \ell_i(\boldsymbol{\theta})}{\partial\gamma\partial\psi^T} = \frac{\partial^2\ \ell_{1i}(\boldsymbol{\theta})}{\partial\gamma\partial\psi^T} + W_\Phi\left(\ell_{2i}(\boldsymbol{\theta})\right)\frac{\partial^2\ \ell_{2i}(\boldsymbol{\theta})}{\partial\gamma\partial\psi^T} +, W_\Phi^1\left(\ell_{2i}(\boldsymbol{\theta})\right)\frac{\partial\ \ell_{2i}(\boldsymbol{\theta})}{\partial\gamma}\frac{\partial\ \ell_{2i}(\boldsymbol{\theta})}{\partial\psi^T}$$

Onde $W_\Phi^1(x) = -W_\Phi(x)\big(x + W_\Phi(x)\big)$ é a derivada de $W_\Phi\ (x)$.

Logo, a matriz de informação observada para $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \lambda)^T$ é dada por

$$\mathbf{I}(\boldsymbol{\theta}) = \mathrm{I}_1(\boldsymbol{\theta}) + \mathrm{I}_2(\boldsymbol{\theta}) \tag{8}$$

$$\mathrm{I}_\mathrm{k}(\boldsymbol{\theta}) = \begin{pmatrix} \mathrm{I}^\mathrm{k}_{\beta\beta} & \mathrm{I}^\mathrm{k}_{\sigma^2\beta} & \mathrm{I}^\mathrm{k}_{\lambda\beta} \\ & \mathrm{I}^\mathrm{k}_{\sigma^2\sigma^2} & \mathrm{I}^\mathrm{k}_{\lambda\sigma^2} \\ & & \mathrm{I}^\mathrm{k}_{\lambda\lambda} \end{pmatrix}$$ para k = 1, 2 e $\gamma, \psi = \boldsymbol{\beta}, \sigma^2, \lambda$, onde

$\mathrm{I}^1_{\beta\beta} = \frac{1}{\sigma^2}\mathbf{X}^\mathrm{T}\mathbf{X}$,

$\mathrm{I}^1_{\sigma^2\beta} = \frac{1}{\sigma^4}\mathbf{X}^\mathrm{T}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$,

$\mathrm{I}^1_{\sigma^2\sigma^2} = -\frac{\mathrm{n}}{2\sigma^4} + \frac{1}{\sigma^6}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\mathrm{T}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$,

$\mathrm{I}^1_{\lambda\boldsymbol{\Gamma}} = \mathbf{0}, \boldsymbol{\Gamma} = \boldsymbol{\beta}, \sigma^2, \lambda$,

$\mathrm{I}^2_{\beta\beta} = -\frac{\lambda^2}{\sigma^2}\mathbf{X}^\mathrm{T}\mathbf{D}\left(W^1_\Phi\left[\lambda\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{\sigma}\right]\right)\mathbf{X}$ **,**

$\mathrm{I}^2_{\sigma^2\beta} = -\frac{\lambda}{2\sigma^3}\mathbf{X}^\mathrm{T}W_\Phi\left[\lambda\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{\sigma}\right] - \frac{\lambda^2}{2\sigma^4}\mathbf{X}^\mathrm{T}\mathbf{D}\left(W^1_\Phi\left[\lambda\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{\sigma}\right]\right)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ **,**

$\mathrm{I}^2_{\lambda\beta} = \frac{1}{\sigma}\mathbf{X}^\mathrm{T}W_\Phi\left[\lambda\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{\sigma}\right] + \frac{\lambda}{\sigma^2}\mathbf{X}^\mathrm{T}\mathbf{D}\left(W^1_\Phi\left[\lambda\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{\sigma}\right]\right)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ **,**

$\mathrm{I}^2_{\sigma^2\sigma^2} = -\frac{3\lambda}{4\sigma^5}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\mathrm{T}W_\Phi\left[\lambda\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{\sigma}\right] - \frac{\lambda^2}{4\sigma^6}\mathrm{Q}_\mathrm{w}(\boldsymbol{\beta})$,

$\mathrm{I}^2_{\lambda\sigma^2} = \frac{1}{2\sigma^3}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\mathrm{T}W_\Phi\left[\lambda\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{\sigma}\right] + \frac{\lambda}{2\sigma^4}\mathrm{Q}_\mathrm{w}(\boldsymbol{\beta})$,

$\mathrm{I}^2_{\lambda\lambda} = -\frac{1}{\sigma^2}\mathrm{Q}_\mathrm{w}(\boldsymbol{\beta})$,

onde $\mathbf{D}(\mathbf{a})$ é a matriz diagonal do vetor **a** e $\mathrm{Q}_\mathrm{w}(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\mathrm{T}\mathbf{D}\left(W^1_\Phi\left[\lambda\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{\sigma}\right]\right)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$**.**

2.5- ESTIMAÇÃO DOS PARÂMETROS VIA ALGORITMO EM

A função verossimilhança para o modelo (5), denotado por $\ell(\boldsymbol{\theta})$, pode ser escrito como

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{m} \log\left[2\phi(y_i|\boldsymbol{x}_i{}^T\boldsymbol{\beta},\quad \sigma^2)\Phi\left(\frac{\lambda(y_i - \boldsymbol{x}_i{}^T\boldsymbol{\beta})}{\sigma^2}\right)\right]$$

$$= \sum_{i=1}^{m} \log\left[2\int_0^{+\infty} \phi(y_i|\boldsymbol{x}_i{}^T\boldsymbol{\beta}, \sigma^2)\phi(t_i|\lambda(y_i - \boldsymbol{x}_i{}^T\boldsymbol{\beta}), \sigma^2)dt_i\right].$$

Uma forma de encontrar as estimativas de máxima verossimilhança dos parâmetros é maximizar a função acima, porem esse método é complicado para esse determinado caso, devido a presença de integrais. Por isso, neste caso a saída é usar o método do algoritmo EM.

Seja $\boldsymbol{y}$ o conjunto de dados observados e $\boldsymbol{s}$ denotando o conjunto de dados faltantes. O dado completo $\boldsymbol{y}_c = (\boldsymbol{y}, \boldsymbol{s})$ é $\boldsymbol{y}$ aumentado com $\boldsymbol{s}$. Denota-se por $\ell_c(\boldsymbol{\theta}|\boldsymbol{y}_c)$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, a função log-verossimilhança dos dados completos e por $Q(\boldsymbol{\theta}|\,\widehat{\boldsymbol{\theta}}) = E[\ell_c(\boldsymbol{\theta}|\boldsymbol{y}_c)|\,\boldsymbol{y}, \widehat{\boldsymbol{\theta}}]$, o valor esperado desta função. Cada iteração do algoritmo EM envolve dois passos, um passo E (esperança) e um passo M (maximização), definidos como:

- Passo E: Calcule $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$ como uma função de $\boldsymbol{\theta}$;
- Passo M: Encontre $\boldsymbol{\theta}^{(k+1)}$ que maximiza $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$.

Considere $\widehat{\boldsymbol{\theta}}^{(k)} = \left({\beta^{(k)}}^T, {\sigma^2}^{(k)}, \lambda^{(k)}, \gamma^{(k)} \right)^T$ a estimativa de $\boldsymbol{\theta}$ na k-ésima iteração.

Passo E: Dado $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, calcule $\hat{t}_j^{(k)}$ e $\widehat{t^2}_j^{(k)}$, para $j = 1, \dots, n$.

Passo M: Atualize $\widehat{\boldsymbol{\theta}}^{(k+1)}$ maximizando $Q(\boldsymbol{\theta}|\,\widehat{\boldsymbol{\theta}}^{(k)})$ sob $\boldsymbol{\theta}$, o que leva às seguintes soluções analíticas:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} - \frac{\lambda^{(k)}}{1 + {\lambda^{(k)}}^2}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\hat{\boldsymbol{t}}^{(k)} - \gamma^{(k)}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{x}_i\,,$$

$$\hat{\sigma}^{2(k+1)} = \frac{1}{2n}\left\{\left(1 + {\lambda^{(k)}}^2\right)\left[Q\left(\boldsymbol{\beta}^{(k)}\right) - 2\gamma^{(k)}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}^{(k)}\right) + {\gamma^{(k)}}^2\ \right] + \ {\widehat{\boldsymbol{t}^2}^{(k)}}^T \mathbf{1}_n \right.$$
$$\left. - 2\lambda^{(k)}{\hat{\boldsymbol{t}}^{(k)}}^T\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(k)}\right) + 2\lambda^{(k)}\gamma^{(k)}\hat{t}_i^{(k)}\right\},$$

$$\hat{\lambda}^{(k+1)} = \frac{{\hat{\boldsymbol{t}}^{(k)}}^T\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(k)}\right) - \gamma^{(k)}\hat{t}_i^{(k)}}{Q_\phi\left(\boldsymbol{\beta}^{(k)}\right)}$$

$$\hat{\gamma}^{(k+1)} = \left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}^{(k)}\right) - \frac{\lambda^{(k)}}{1 + {\lambda^{(k)}}^2}$$

onde $Q\left(\boldsymbol{\beta}^{(k)}\right) = \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(k)}\right)^T\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(k)}\right)$,

$$Q_\phi\left(\boldsymbol{\beta}^{(k)}\right) = Q\left(\boldsymbol{\beta}^{(k)}\right) - 2\gamma^{(k)}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}^{(k)}\right) + {\gamma^{(k)}}^2,$$

e

$$\hat{t}_j = \hat{\lambda}\hat{\eta}_j + \hat{\sigma}W_\phi\left(\frac{\hat{\lambda}\hat{\eta}_j}{\hat{\sigma}}\right) \text{ e } \widehat{t^2}_j = \hat{\lambda}^2\hat{\eta}_j^2 + \hat{\sigma}^2 + \hat{\lambda}\hat{\sigma}\hat{\eta}_j\, W_\phi\left(\frac{\hat{\lambda}\hat{\eta}_j}{\hat{\sigma}}\right), \;\; j = 1, \dots, n,$$

com $W_\phi(x) = \phi(x)/\Phi(x), \hat{\eta}_j = \hat{\lambda}\big(y_j - \boldsymbol{x}_j^T\widehat{\boldsymbol{\beta}}\big),\ j = 1, \dots, n\ , j \neq i \text{ e } \hat{\eta}_i = \hat{\lambda}\big(y_i - \boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}} - \hat{\gamma}\big)$, calculado no passo k.

Claramente, se $\gamma = 0$, as equações do Passo M se reduzem às equações obtidas em Ferreira (2008) para o modelo normal assimétrico. Note que, quando $\lambda = 0$ e $\gamma = 0$, $\widehat{\boldsymbol{\beta}}^{(k+1)} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ e $\hat{\sigma}^{2^{(k+1)}} = \frac{(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})}{n}$ são os EMV de β e σ^2, respectivamente, do modelo normal simétrico.

São usados como valores iniciais para $\boldsymbol{\theta}$ no algoritmo os estimadores de momentos (Rodríguez, 2005).

Os passos necessários para a implementação do algoritmo EM para encontrar os estimadores de máxima verossimilhança (EMV) dos parâmetros do modelo definido em (**5**) encontra-se em Ferreira (2008).

2.7- INFLUÊNCIA DOS PARÂMETROS NOS MOMENTOS DA DISTRIBUIÇÃO *SKEW-NORMAL*

2.7.1- VALOR ESPERADO

Por definição, o valor esperado de uma variável aleatória $Y \sim SN\ (\mu, \sigma^2, \lambda)$ é calculado da seguinte maneira:

$$E[Y] = \int_{-\infty}^{+\infty} y\frac{2}{\sigma^2}\,\phi\left(\frac{x-\mu}{\sigma^2}\right)\,\Phi\left(\frac{x-\mu}{\sigma^2}\,\lambda\right)\mathrm{dy}$$

E pode-se mostrar no (7) que $E(Y) = \mu + c\sigma\rho$.

2.7.2- VARIÂNCIA

Por definição, a variância de uma variável aleatória $Y \sim SN\ (\mu, \sigma^2, \lambda)$ é calculada da seguinte maneira:

$$Var(Y) = E(Y^2) - [E(Y)]^2$$

Pelo Resultado 2 (7), temos que $E(Y^2) = \mu^2 + 2c\mu\sigma\rho + \sigma^2$. Então

$$Var(Y) = \sigma^2(1 + c^2\rho^2) \tag{9}$$

Existe uma relação entre a $Var(Y)$ e os parâmetros σ^2 e λ. Pelo fórmula apresentada anteriormente é possivel notar que quanto maior o valor de σ^2, maior é a $Var(Y)$.Por outro lado, quanto menor o valor de λ, maior será a $Var(Y)$.

2.7.3- COEFICIENTE DE ASSIMETRIA

Por definição, o coeficiente de assimetria de uma variável aleatória $Y \sim SN\ (\mu, \sigma^2, \lambda)$ é calculado da seguinte maneira:

$$\gamma_1 = \frac{E[(Y - E(Y))^3]}{E[(Y - E(Y))^2]^{3/2}}$$

Usando os resultados dados no Resultado 2 (7), pode-se encontrar o seguinte resultado para a assimetria de uma variável normal assimétrica:

$$\gamma_1 = \frac{\rho^3 c\ (2c^2 - 1)}{(1 - c^2\rho^2)^{3/2}} \tag{10}$$

Note que γ_1 é somente função do parâmetro de assimetria λ, ou seja, quanto maior o valor do parâmetro λ, maior o valor da assimetria γ_1.

A interpretação de γ_1 é a seguinte:

Se $\gamma_1 = 0$ indica uma distribuição simétrica,

Se $\gamma_1 < 0$ indica uma distribuição assimétrica negativa,

Se $\gamma_1 > 0$ indica uma distribuição assimétrica positiva.

2.7.4- CURTOSE

Por definição, a curtose de uma variável aleatória $Y \sim SN\ (\mu, \sigma^2, \lambda)$ é calculada da seguinte maneira:

$$\gamma_2 = \frac{E[(Y - E(Y))^4]}{E[(Y - E(Y))^2]^2}$$

Usando os resultados dados no Resultado 2 (7), pode encontrar o seguinte resultado para a assimetria de uma variável normal assimétrica:

$$\gamma_2 = \frac{3\sigma^3 - 3c^4\sigma^3\rho^4 + 4c^2\sigma^3\rho^4 - 6c^2\sigma^3\rho^2 - c^3\mu\sigma^2\rho^3 + 12c^2\mu^2\sigma\rho^2 - c\mu^3\rho - 12c^2\mu^2\rho - 4c\mu\sigma^2\rho^2 - 4c\mu\sigma^2\rho^3}{\sigma^3(1 - c^2\rho^2)^2}. \tag{11}$$

A curtose é uma medida que caracteriza o grau de achatamento de uma distribuição, quanto menor o γ_2, maior é o achatamento da distribuição. Uma das interpretações para a curtose é a seguinte:

Se $\gamma_2 < 3$, a distribuição é muito achatada (Platicúrtica)

Se $\gamma_2 = 3$, a distribuição é aproximadamente normal (Mesocúrtica)

Se $\gamma_2 > 3$, a distribuição é pontiaguda (Leptocúrtica).

2.8- TESTES ASSINTÓTICOS

Teste da Razão de Verossimilhança:

A estatística de teste da Razão de Verossimilhanças é definida como

$$\xi_{RV} = 2\{\ell(\widehat{\boldsymbol{\theta}}; \boldsymbol{y}) - \ell(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}; \boldsymbol{y})\}, \qquad (12)$$

Onde $\boldsymbol{\ell}(\boldsymbol{\theta}; \boldsymbol{y})$ é a função log-verossimilhança sob o modelo irrestrito e $\boldsymbol{\ell}(\widehat{\boldsymbol{\theta}}_{\mathbf{0}}; \boldsymbol{y})$ a função log-verossimilhança sob o modelo restrito ($\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ é o estimador de máxima verossimilhança de $\boldsymbol{\theta}$ sob a hipótese nula)

Sob H_0 e para grandes amostras, tem-se que $\xi_{SR} \sim \chi_1^2$.

Intervalo de Confiança:

Seja $MI(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \frac{\partial l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$ a matriz de informação de Fisher, obtida em.Considerando que assintoticamente $\widehat{\boldsymbol{\theta}}_{MV} \approx N_p(\boldsymbol{\theta}, MI^{-1}(\boldsymbol{\theta}))$, temos um intervalo de confiança para θ_i igual a 13

$$IC(\theta_i) = \widehat{\theta}_i \pm Z_{\alpha/2} \sqrt{v_{ii}} \qquad (14)$$

,

onde v_{ii}= i-ésimo elemento da diagonal de $MI^{-1}(\boldsymbol{\theta})$.

Teste de Shapiro-Wilk e Teste de Lilliefors:

Shapiro-Wilk (1965) e Lilliefors (1967) são testes para testar normalidade, ou seja, a hipótese nula é de que a distribuição testada tem uma distribuição normal versus a hipótese alternativa de que a distribuição testada não é normal. Mais informação sobre esses testes são encontradas no W. J. Conover (1998).

3- ESTUDOS DE SIMULAÇÃO

3.1- TESTES DE HIPÓTESES PARA $\boldsymbol{\beta_1}$E RESULTADOS

Considerando o seguinte modelo

$$\begin{array}{rcl} Y_i & = & \beta_0 + \beta_1 x_i + \ \varepsilon_i \, , i = 1, \dots, n. \\ \varepsilon_i & \sim & SN\,(-b\sigma\rho, \sigma^2, \lambda). \end{array}$$

foram feitos dois testes assintóticos para β_1, o teste da Razão de Verossimilhanças (12) e o teste usando o Intervalo de Confiança (14). Considerando as seguintes hipóteses

$$\mathrm{H}_0\!: \beta_1 = 0 \ versus \ \mathrm{H}_1\!: \beta_1 \neq O$$

O objetivo desses testes foi detectar, a partir de valores simulados de $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \lambda)^T$, qual seria um valor mínimo para β_1 que levasse à rejeição da hipótese nula.

Primeiramente foram testadas várias combinações na geração do tamanho da amostra (n) e dos parâmetros β_1, $\sigma^2 \cdot$, β_0 e λ. Mostrando que tais combinações causam mudanças no resultado do teste.

Para um β_1 maior que 0.1, vimos que para todas as combinações, há rejeição de H_0, para um nível de 5% é praticamente de 100%, ou seja, um β_1 maior que 0.1 é quase sempre significativamente diferente de zero

Com isso fixamos o β_1=0.1, β_0=5, n=100, λ =2 e variamos o σ^2 nos testes propostos. Para amostras de tamanho n=100, replicadas 100 vezes, com σ^2=3, obtivemos taxas de rejeição parecidas (77% via Intervalo de Confiança, 76% via Razão de Verossimilhança e de 76% paras duas simultaneamente). Com σ^2=4, obtivemos menores taxas de rejeição (58% via Intervalo de Confiança, 57% via Razão de Verossimilhança e de 57% paras duas simultaneamente).

Logo, variação em σ^2 afeta o resultado dos testes assintóticos. Isso é coerente com a teoria, pois quanto maior σ^2, maior será a var(Y) (dada em (9)) de uma variável aleatória com distribuição normal assimétrica. Então, quanto maior o σ^2, menos eu rejeito minha hipótese nula ($\mathrm{H}_0\!: \beta_1 = 0$).

3.2- TESTES DE HIPÓTESES PARA λ E RESULTADOS

Considerando o seguinte modelo

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_i + \varepsilon_i \,, i = 1, \dots, n. \\ \varepsilon_i &\sim SN\left(-b\sigma\rho, \sigma^2, \lambda\right). \end{aligned}$$

Foram feitos quatro testes assintóticos para λ, o teste da Razão de Verossimilhanças (12), teste usando o Intervalo de Confiança (14) e os testes de normalidade Shapiro-Wilk (1965) e Lilliefors (1967). Considerando as seguintes hipóteses:

$$\mathrm{H}_0\colon \lambda = 0 \; versus \; \mathrm{H}_1\colon \lambda \neq O$$

O objetivo desses testes foi detectar, a partir de valores simulados de $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \lambda)^T$, qual seria um valor mínimo para $n(tamanho\ da\ amostra)$, que levasse à uma rejeição da hipótese nula igual ou superior a 95% nos quatro testes de hipóteses.

Primeiramente foram testados combinações na geração σ^2 e λ, deixando fixo o tamanho da amostra n e os parâmetros β_0 e β_1. Não variamos os parâmetros β_0 e β_1, pois eles não influenciam no resultado do teste para λ.

Note que $Y \sim SN\ (\mu, \sigma^2, 0)$ é igual a $Y \sim N\ (\mu, \sigma^2)$, ou seja, usar os testes Shapiro-Wilk (1965) e Lilliefors (1967) para testar normalidade é a mesma coisa que testar se $\lambda = 0$.

Sabemos que assintoticamente o estimador de máxima verossimilhança é não viesado, então quanto maior o tamanho da amostra mais próximo o parâmetro estimado será do parâmetro gerado.

A relação entre o λ e a variância da distribuição Skew Normal é direta, podemos perceber isso na fórmula da variância de Y (dada em (9)). Pode-se notar que quanto maior o λ, menor a var(Y), e com isso será mais provável a rejeição da hipótese nula.

Outra forma de analisar isso seria a de que quanto menor é o λ gerado, mais a distribuição vai se aproximar de uma distribuição Normal, logo é mais provável a não rejeição da hipótese nula. Isso pode ser notado pela formula da assimetria da normal assimétrica (dada em (10)).

Analisando novamente a variância de Y (9), podemos perceber que existe uma relação proporcional forte entre ela e o parâmetro σ^2. Quanto maior o σ^2, maior a var(Y), aumentando a chance de rejeitarmos com menor frequência a hipótese nula e quanto menor o σ^2, menor a var(Y), aumentando a chance de rejeitarmos a hipótese nula com maior frequência.

Nos testes foram feitas 100 simulações para cada combinação de parâmetros, obtendo assim a porcentagem de rejeição da hipótese nula.

Os resultados são apresentados no quadro a seguir:

Tabela 1: Tamanhos mínimos de amostras para diversas combinações do parâmetro $\boldsymbol{\lambda}$.

Parâmetros				Momentos			Taxa de Rejeição dos Testes				n^*
$\boldsymbol{\beta_0}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\sigma^2}$	$\boldsymbol{\lambda}$	**Variância de Y**	**Assimetria de Y**	**Curtose de Y**	**RV**	**IC**	**Shapiro**	**Lilliefors**	
5	1	1	5	0,387866	0,850965	0,3485569	100%	99%	96%	100%	200
5	1	1	4	0,400828	0,7844268	0,3741195	100%	100%	98%	100%	300
5	1	1	3	0,427042	0,6670236	0,4299403	100%	100%	100%	100%	500
5	1	1	2	0,490704	0,4538256	0,5894022	100%	100%	99%	100%	1000

n^*=Tamanho mínimo da amostra necessário para alcançar uma taxa de rejeição mínima de 95% (alta rejeição).

4- CONCLUSÕES

Com esse trabalho, podemos concluir que os momentos da distribuição normal assimétrica estão relacionados com o seu valor esperado, variância, coeficiente de assimetria e curtose, que influenciam muito no comportamento de seus parâmetros, e consequentemente nos resultados dos testes de hipóteses feitos com os parâmetros β_1 *e* λ.

No teste de hipótese para testar se o $\beta_1 = 0$, levando em conta as combinações de parâmetros e tamanhos de amostras, podemos perceber que o valor mínimo para β_1 que levasse à rejeição da hipótese nula era de 0.1 (para tamanhos de amostras mínimos). Ou seja, podemos concluir que para todas as combinações de valores simulados de $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \lambda)^T$, β_1 é quase sempre significativamente diferente de zero, para valores de β_1 maiores que 0.1, levando há rejeição da hipótese nula.

O teste de hipóteses para testar se $\lambda = 0$ teve o objetivo de detectar, a partir de valores simulados de $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \lambda)^T$, qual seria um valor mínimo para $n(tamanho\ da\ amostra)$, que levasse à uma rejeição da hipótese nula igual ou superior a 95% nos quatro testes assintóticos.

Os resultados encontrados variaram de acordo com o valor do parâmetro λ gerado, que quanto maior, menor é o valor mínimo encontrado para n, que leva a rejeição da hipótese nula igual ou superior a 95%.

Pode-se notar que os resultados obtidos nos testes de hipóteses estão de acordo com a interpretação teórica dos momentos de uma distribuição Normal Assimétrica.

Uma próxima etapa desse estudo seria dar continuidade ao teste de hipóteses para λ, através da variação do parâmetro σ^2. Outro fato importante a se estudar, é entender como esses testes de hipóteses funcionam para distribuições de misturas assimétricas.

REFERÊNCIAS BIBLIOGRÁFICAS

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

Azzalini, A. (2005). The skew-normal distribution and related multivariate families (with discussion). *Scandinavian Journal Statistics*, 32, 159-188.

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

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

Azzalini, A. (2004). The skew-normal distribution and related multivariate families. http://tango.stat.unipd.it/SN/review.ps.
CONOVER, W.J. (1998). Practical Nonparametric Statistics, John Wiley & Sons.

Fan, J., Zhang, C., & Zhang, J. (2001). Generalized likelihood ratio statistics and Wilksphenomenon. The Annals of Statistics, 29:153–193.

Ferreira. C.S. (2008). *Inferência e diagnóstico em modelos assimétricos.* Tese de Doutorado. Departamento de Estatística. IME-USP. São Paulo.

Lin, T.I., Lee, J.C. e Yen, S.Y. (2007). Finite mixture modeling using the skew normal distribution. *Statistica Sinica*, 17, 909-927.

Lilliefors, H.W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association, 62*, 399-402 (6.2, Appendix).

Martínez, E.H. , Varela, H., Gomez, H.W. e Bolfarine, H. (2008). A note on the likelihood and moments of the skew-normal distribution.

P.K. Sen. Shapiro-Wilk type goodness-of-fit tests for normality: Asymptotics revisited. In C. Huber-Carol, N. Balakrishnan, M.S. Nikulin, and M. Mesbah, editors, *Goodnessof-Fit Tests and Model Validity*, pages 73–88. Birkh¨auser, Boston, 2002.

Rodríguez. C.L.B. (2005). *Inferência bayesiana no modelo normal assimétrico*. Dissertação de mestrado. Departamento de Estatística. IME-USP. São Paulo.

Shapiro, S.S., and Wilk, M.B (1968). Approximations for the null distribution of the W statistic. *Technometrics, 10,* 861-866 (6.2).

T. de Wet and J.H. Wenter. Asymptotic distributions of quadratic forms with application to test of fit. *Ann. Statist.*, 31:276–295, 1973.