

Estimação robusta para a distribuição Skew-t

Mariane dos S. Bispo¹

UNICAMP, Campinas, SP

Camila B. Zeller²

UFJF, Juiz de Fora, MG

Resumo. Neste trabalho, consideramos métodos robustos na presença de outliers, propondo um novo algoritmo Esperança-Maximização (EM), para obter os estimadores de máxima verossimilhança (ML) para todos os parâmetros da Skew-t (ST), denotado por EM_q (algoritmo da Esperança-Maximização com a verossimilhança L_q), isto é, que aborda ML_q dentro da estrutura EM. Ao fazer isso, herdamos a robustez dos estimadores obtidos por ML_q e o tornamos disponível no contexto do modelo ST. Assim, obtemos estimadores duplamente reponderados para todos os parâmetros da distribuição ST.

Palavras-chave. Algoritmo EM, Distribuição Skew-t, Método de Máxima Verossimilhança L_q , Robustez

1 Introdução

Outliers e robustez são termos considerados com muita importância neste trabalho. O termo outlier, muitas vezes, é usado informalmente e [1] ressalta que outliers são definidos com respeito aos modelos. Dessa forma, observações podem ser outliers em um modelo e serem perfeitamente regulares para outro modelo. É bem conhecido na literatura que a inferência sob as distribuições elípticas é suscetível aos dados atípicos. Além da presença de dados atípicos, a distribuição dos dados pode ser assimétrica e neste contexto, recorreremos aos membros da classe de distribuições assimétricas, também conhecidas por skew-elípticas, ver [2].

O termo robustez está relacionada com a sensibilidade dos resultados de uma análise estatística aos desvios nas suposições de que esta análise se baseia; veja [3]. Dentro da classe de distribuições assimétricas, destaca-se a distribuição ST [4]. Dessa forma, o objetivo principal deste trabalho é obter os estimadores de ML_q para todos os parâmetros da distribuição ST como alternativa aos estimadores de ML. Em nossa abordagem, o papel das observações é alterado com uma mudança no modelo original ao se adicionar um parâmetro de distorção q . Então, podemos considerar a estimação por ML_q como a minimização da discrepância de um modelo mais geral em relação a aquele que diminui o papel das observações extremas gerando estimadores mais robustos que os de ML, em termos da redução dos pesos destas observações no processo de estimação.

2 Estimadores de ML_q

A estimação por ML_q , introduzida por [5], faz uso da q -entropia, proposta por [6] e [7]. Para uma variável aleatória Y com fdp (função de densidade de probabilidade) $f(y|\theta)$, a q -entropia é definida por

$$H_q(Y) = -E[L_q(f(Y|\theta))], \quad (1)$$

¹m235141@dac.unicamp.br

²camila.zeller@ufjf.br

onde L_q é a função q-logarítmica dada por

$$L_q(u) = \begin{cases} \log(u), & u \geq 0, q = 1, \\ \frac{u^{1-q}-1}{1-q}, & u \geq 0, q \neq 1. \end{cases} \quad (2)$$

Observe que a q-entropia pode ser considerada uma generalização da entropia de Shannon. Quando $q = 1$ em (1), obtemos a entropia de Shannon.

Usando a versão empírica da q-entropia, dada nas equações (1)-(2), [5] introduziu o método de estimação por ML_q . Considere $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ uma amostra aleatória com pdf $f(\mathbf{y}|\boldsymbol{\theta})$. O estimador de ML_q para $\boldsymbol{\theta}$ é dada por $\widehat{\boldsymbol{\theta}}_q = \arg \max_{\boldsymbol{\theta}} \sum_{i=1}^n L_q(f(\mathbf{y}_i|\boldsymbol{\theta}))$, com $0 < q < 1$. Note que quando $q = 1$, obtemos os estimadores de ML usuais. Além disso, após algumas manipulações algébricas, [5] mostra que o estimador de ML_q para $\boldsymbol{\theta}$ é solução da seguinte equação de estimação

$$\sum_{i=1}^n \mathbf{U}_q(\mathbf{y}_i|\boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{U}(\mathbf{y}_i|\boldsymbol{\theta}) f(\mathbf{y}_i|\boldsymbol{\theta})^{1-q} = 0, \quad (3)$$

onde $\mathbf{U}(\mathbf{y}_i|\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log(f(\mathbf{y}_i|\boldsymbol{\theta}))$ é o vetor escore. Podemos observar que a equação (3) oferece uma interpretação natural do ML_q , como uma solução de uma verossimilhança ponderada. Se $q = 1$, todas as observações recebem o mesmo peso. Porém, quando $q \neq 1$, o ML_q fornece uma reponderação relativa ao modelo e assim, as observações atípicas recebem menores pesos durante o processo de estimação.

2.1 Algoritmo EM_q para a distribuição ST

Seja $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\boldsymbol{\lambda}}, \widehat{\nu})$ as estimativas atuais dos parâmetros, vamos definir $W_{qi}(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\boldsymbol{\lambda}}, \widehat{\nu}) = f(\mathbf{y}_i|\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\boldsymbol{\lambda}}, \widehat{\nu})^{1-q}$. Assim, as novas estimativas ML_q serão obtidas ao maximizar a seguinte função

$$\sum_{i=1}^n W_{qi}(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\boldsymbol{\lambda}}, \widehat{\nu}) \log(f(\mathbf{y}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)), \quad (4)$$

com pdf $f(\mathbf{y}|\boldsymbol{\theta})$ da distribuição ST, definida como

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu) = 2t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) T \left(\sqrt{\frac{p+\nu}{d_{\boldsymbol{\lambda}}+\nu}} \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-\frac{1}{2}} (\mathbf{y} - \boldsymbol{\mu}); \nu + p \right), \quad \mathbf{y} \in \mathbb{R}^p, \quad (5)$$

tal que $t_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$ é a fdp da distribuição T-Student p -variada e $T(\cdot; \nu)$ é a acumulada da T-Student padrão univariada.

No contexto dos dados completos $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{u}^\top, \mathbf{t}^\top)^\top$, a equação de estimação (3) pode ser escrita como

$$\sum_{i=1}^n W_{qi}(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\boldsymbol{\lambda}}, \widehat{\nu}) U(\mathbf{y}_i, u_i, t_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu) = 0, \quad (6)$$

na qual $U(\mathbf{y}_i, u_i, t_i|\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log(f(\mathbf{y}_i, u_i, t_i|\boldsymbol{\theta}))$ é a derivada da log-verossimilhança dos dados completos com respeito aos parâmetros. A seguir, apresentamos as etapas do algoritmo EM_q proposto.

Passos do algoritmo EM_q para a distribuição ST

1. Dado $0 < q < 1$ fixo, forneça o chute inicial $\widehat{\boldsymbol{\theta}}_q^{(0)} = (\widehat{\boldsymbol{\mu}}_q^{(0)}, \widehat{\boldsymbol{\alpha}}_q^{(0)}, \widehat{\boldsymbol{\lambda}}_q^{(0)}, \widehat{\nu}_q^{(0)})$ e o valor de ε para o critério de parada.

2. Passo E: Dado $\theta = \widehat{\theta}_q^{(k)}$, calcule

$$\widehat{u}_{qi}^{(k)} = \frac{\widehat{\nu}_q^{(k)} + p}{\widehat{\nu}_q^{(k)} + \widehat{d}_{q\lambda}}, \quad \widehat{ut}_{qi}^{(k)} = \widehat{u}_{qi}^{(k)} \widehat{\mu}_{Tqi}^{(k)} + \widehat{M}_{Tq}^{(k)} \widehat{\eta}_{1qi}^{(k)},$$

$$\widehat{ut}_{qi}^{2(k)} = \widehat{u}_{qi}^{(k)} \widehat{\mu}_{Tqi}^{2(k)} + \widehat{M}_{Tq}^{2(k)} + \widehat{\mu}_{Tqi}^{(k)} \widehat{M}_{Tq}^{(k)} \widehat{\eta}_{1qi}^{(k)},$$

onde $\widehat{M}_{Tq}^2 = [1 + \widehat{\Delta}_q^\top \widehat{\Gamma}_q^{-1} \widehat{\Delta}_q]^{-1}$, $\widehat{\mu}_{Tq} = \widehat{M}_{Tq}^2 \widehat{\Delta}_q^\top \widehat{\Gamma}_q^{-1} (y - \widehat{\mu}_q)$, $\widehat{\Delta}_q = \widehat{\Sigma}_q^{\frac{1}{2}} \widehat{\delta}_q$, $\widehat{\Gamma}_q = \widehat{\Sigma}_q - \widehat{\Delta}_q \widehat{\Delta}_q^\top$, $\widehat{\delta}_q = \frac{\widehat{\lambda}_q}{\sqrt{1 + \widehat{\lambda}_q^\top \widehat{\lambda}_q}}$, $\widehat{d}_{q\lambda}$ é a distância de Mahalanobis, usada para a identificação de observações atípicas, avaliada nas estimativas de ML_q de μ e Σ . Para mais detalhes do cálculo de $\widehat{\eta}_{1qi}^{(k)}$, veja [4].

3. Passo-CM: Obtenha $\widehat{\mu}_q^{(k+1)}$, $\widehat{\Sigma}_q^{(k+1)}$ e $\widehat{\lambda}_q^{(k+1)}$, ao resolver a equação (6) em relação a cada um dos parâmetros μ , Σ e λ , respectivamente, e considerando $\nu = \widehat{\nu}_q^{(k)}$. Essa maximização resulta nas seguintes expressões fechadas

$$\widehat{\mu}_q^{(k+1)} = \frac{\sum_{i=1}^n W_{qi}^{(k)} \widehat{u}_{qi}^{(k)} \mathbf{y}_i - \sum_{i=1}^n W_{qi}^{(k)} \widehat{ut}_{qi}^{(k)} \widehat{\Delta}_q^{(k)}}{\sum_{i=1}^n W_{qi}^{(k)} \widehat{u}_{qi}^{(k)}},$$

$$\widehat{\Gamma}_q^{(k+1)} = \frac{1}{n \sum_{i=1}^n W_{qi}^{(k)}} \sum_{i=1}^n \left[W_{qi}^{(k)} \widehat{u}_{qi}^{(k)} (\mathbf{y}_i - \widehat{\mu}_q^{(k)})^\top (\mathbf{y}_i - \widehat{\mu}_q^{(k)}) - W_{qi}^{(k)} \widehat{ut}_{qi}^{(k)} (\mathbf{y}_i - \widehat{\mu}_q^{(k)})^\top \widehat{\Delta}_q^{(k)} - W_{qi}^{(k)} \widehat{ut}_{qi}^{(k)} \widehat{\Delta}_q^{(k)\top} (\mathbf{y}_i - \widehat{\mu}_q^{(k)}) + W_{qi}^{(k)} \widehat{ut}_{qi}^{2(k)} \widehat{\Delta}_q^{(k)\top} \widehat{\Delta}_q^{(k)} \right],$$

$$\widehat{\Delta}_q^{(k+1)} = \frac{\sum_{i=1}^n W_{qi}^{(k)} \widehat{ut}_{qi}^{(k)} (\mathbf{y}_i - \widehat{\mu}_q^{(k)})}{\sum_{i=1}^n W_{qi}^{(k)} \widehat{ut}_{qi}^{2(k)}}, \quad \widehat{\Sigma}_q^{(k+1)} = \widehat{\Gamma}_q^{(k+1)} - \widehat{\Delta}_q^{(k+1)} \widehat{\Delta}_q^{(k+1)\top}$$

$$\text{e } \widehat{\lambda}_q^{(k+1)} = \frac{(\widehat{\Sigma}_q^{(k+1)})^{-1/2} \widehat{\Delta}_q^{(k+1)}}{(1 - \widehat{\Delta}_q^{(k+1)\top} \widehat{\Delta}_q^{(k+1)})^{-1} (\widehat{\Sigma}_q^{(k+1)})^{-1} \widehat{\Delta}_q^{(k+1)}}^{1/2}, \text{ em que } W_{qi}^{(k)} = W_{qi}^{(k)}(\widehat{\mu}_q^{(k)}, \widehat{\Sigma}_q^{(k)}, \widehat{\lambda}_q^{(k)}, \widehat{\nu}_q^{(k)}).$$

3. Passo-CME: Atualize $\widehat{\nu}_q^{(k+1)}$, tal que

$$\widehat{\nu}_q^{(k+1)} = \arg \max_{\nu} \sum_{i=1}^n W_{qi}^{(k)} (\widehat{\mu}_q^{(k)}, \widehat{\Sigma}_q^{(k)}, \widehat{\lambda}_q^{(k)}, \widehat{\nu}_q^{(k)}) \log(f(\mathbf{y}_i | \widehat{\mu}_q^{(k)}, \widehat{\Sigma}_q^{(k)}, \widehat{\lambda}_q^{(k)}, \nu)).$$

4. Repita os passos anteriores até que um critério de parada seja alcançado. Como critério de parada, adotamos $|\frac{\ell(\widehat{\theta}^{(k+1)} | \mathbf{y}_{obs})}{\ell(\widehat{\theta}^{(k)} | \mathbf{y}_{obs})} - 1| < \varepsilon$, onde $\varepsilon = 10^{-4}$ e $\ell^{(k)}$ é a k -ésima log-verossimilhança da distribuição ST.

[8] mostra que quando o parametro de distorção q está no intervalo $(0,1)$ o estimador de ML_q tende a colocar pesos mais baixos em observações com baixa verossimilhança e pesos mais altos em observações com maior verossimilhança. Neste trabalho, consideramos métodos robustos na

presença de outliers. Assim, definimos uma malha para q de 0 a 1 com incrementos de 0.01, e escolhemos o valor de q que fornece a maior estimativa de probabilidade nas caudas da distribuição skew-t, ou seja, o valor de q que fornece o modelo de caudas mais pesadas.

Uma das vantagens do modelo ST é a sua capacidade de fornecer estimadores robustos ML_q contra outliers, no cenário em que ν é considerado parâmetro desconhecido e portanto é estimado.

Em [9], podemos ver que a robustez local dos estimadores M pode ser investigada ao se discutir sua função influência, isto é, os estimadores M com funções de influência (IF) limitada são considerados como os estimadores com boas propriedades de robustez local. A IF é, em certo sentido, proporcional à função de escore. Portanto, para os estimadores, o comportamento da função de influência é determinado pelas funções de escore de cada parâmetro. Se as funções de escore são limitadas, a função influência para o estimador correspondente será limitada.

Avaliamos a robustez dos estimadores ML e ML_q no modelo ST, no contexto multivariado, quando ν não é conhecido. No que diz respeito às propriedades de robustez local dos estimadores, as funções de escore devem ser limitadas quando $|y|$ tende ao ∞ . Sem perda de generalidade, a Figura (1) exibe a estimação do grau de liberdade ν de uma distribuição ST padrão considerando $\lambda = (0,5, 1)^T$ com $q = 0.85$.

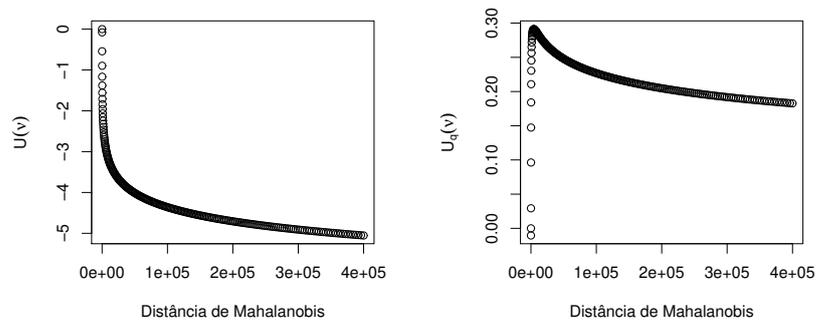


Figura 1: Função de escore para o parâmetro ν do modelo ST. Primeira coluna: caso ML e Segunda Coluna: caso ML_q .

Portanto, ao fazer a estimação simultânea dos graus de liberdade com os outros parâmetros da distribuição ST, observou-se que diferentemente dos outros parâmetros, a função de escore para ν é ilimitada sob o método de estimação ML. Assim, os estimadores de ML não são mais localmente robustos devido ao caráter ilimitado da função de escore relacionada ao parâmetro de grau de liberdade. Entretanto, ressaltamos que quando estimamos os graus de liberdade simultaneamente com μ, Σ e λ usando o método de estimação de ML_q , os estimadores resultantes terão função de influência limitada que não é o caso para os estimadores de ML. Finalmente, podemos concluir que os estimadores de ML_q no modelo ST são robustos contra outliers, no cenário em que ν é considerado parâmetro desconhecido, e portanto é estimado.

3 Aplicação em Dados Reais: AIS

Em [10] é apresentado o conjunto de dados AIS composto por variáveis biomédicas de 202 atletas da elite australiana que treinam no AIS (Australian Institute of Sport), sendo igualmente divididos entre homens e mulheres. Como em [4], consideramos quatro variáveis deste conjunto de dados, sendo elas $\{BMI, Bfat, ssf, LBM\}$ que representam o índice de massa corporal (BMI),

a porcentagem de gordura corporal (*Bfat*), a soma das dobras cutâneas (*ssf*) e a massa magra (*LBM*). A Tabela 1 apresenta as estimativas de ML e de ML_q e seus respectivos desvios padrões (SD), calculados a partir do procedimento paramétrico de bootstrap [11], sob o modelo ST ajustado aos dados AIS, e adicionalmente, estão as medidas amostrais dos dados reais, utilizadas como chute inicial dos algoritmos. Observa-se que as estimativa de ML e ML_q para ν foram 12.1074 e 2.4626, respectivamente. Percebe-se que ocorreu um mudança de sinal nas estimativas dos parâmetros σ_{24} e σ_{34} , sob os métodos de estimação considerados. Podemos observar também que, em relação ao desvio padrão, os estimadores de ML_q mostraram ser mais precisos em relação aos de ML em quase todos os parâmetros, exceto para o parâmetro de assimetria λ .

Tabela 1: AIS: Estimativas de ML e ML_q ($q = 0.47$) para os parâmetros do modelo ST ajustado com os correspondentes desvios padrões, SE e SE_q .

Parâmetro	Amostra	ML	ML_q	SE	SE_q
μ_1	22.9559	20.7937	20.7858	0.4793	0.3807
μ_2	13.5074	9.9519	5.6580	0.5812	0.1253
μ_3	69.0218	49.9101	29.5222	1.8971	0.5934
μ_4	64.8737	61.2288	66.5692	1.6992	1.0633
σ_{11}	8.2021	1.9398	0.7684	0.0822	0.0997
σ_{12}	3.3249	0.2787	0.3153	0.1104	0.0449
σ_{22}	38.3139	2.4004	0.4138	0.0856	0.0394
σ_{13}	29.9489	1.5749	1.2444	0.2026	0.1425
σ_{23}	194.1189	5.9912	2.0015	0.2000	0.0752
σ_{33}	1060.5009	32.9368	12.4912	0.3795	0.3484
σ_{14}	26.7213	2.1763	0.2782	0.1780	0.0974
σ_{24}	-29.2745	-0.9216	0.2874	0.1947	0.0618
σ_{34}	-88.4253	-0.4219	2.5499	0.4705	0.2480
σ_{44}	170.8301	12.7962	3.0054	0.2405	0.2022
λ_1	0.9465	1.0220	1.3026	0.4678	1.4445
λ_2	0.7595	0.5852	1.0511	0.5832	1.1229
λ_3	1.1747	1.1502	6.0762	0.6631	2.2710
λ_4	0.3585	0.6275	1.0025	0.7674	1.5905
ν	—	12.1074	2.6165	3.9704	2.1865

A Figura 2 ilustra a distância de Mahalanobis para o modelo ajustado ST sob o método de estimação ML. Note que as observações 11, 43, 51, 53, 56, 75, 133, 160, 163 e 178 se destacam como outliers.

Para ilustrar o impacto das observações atípicas detectadas nas estimativas de ML e ML_q , a Tabela 2 apresenta as medidas TRC (mudança relativa total) e MRC (máximo da mudança relativa), definidas por

$$TRC = \sum_{j=1}^{n_p} \left| \frac{\hat{\theta}_j - \hat{\theta}_{[i]j}}{\hat{\theta}_j} \right| \text{ e } MRC = \max_{j=1, \dots, n_p} \left| \frac{\hat{\theta}_j - \hat{\theta}_{[i]j}}{\hat{\theta}_j} \right|, \quad (7)$$

onde n_p é a dimensão de θ e o $\hat{\theta}_{[i]}$ é a estimativa de máxima verossimilhança de θ sem a i -ésima observação, para mais detalhes ver [12]. Note que as maiores mudanças ocorreram no modelo ST ajustado sob o método de estimação ML, quando excluímos todas as observações que se destacaram na Figura 2 conjuntamente e a observação 163 individualmente.

Na Tabela 3, temos os pesos atribuídos \hat{u}_i e \hat{u}_{qi} , provenientes dos algoritmos EM e EM_q , respectivamente, às observações atípicas identificadas na Figura 2. Note que os pesos \hat{u}_{qi} são os menores,

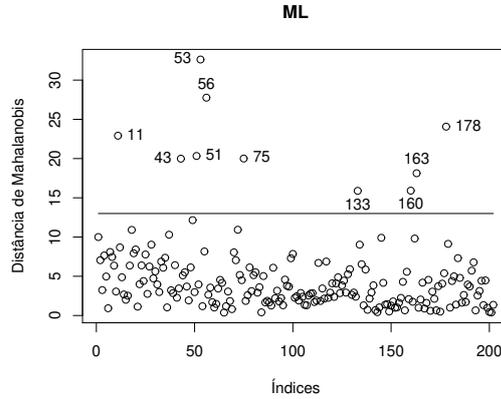


Figura 2: AIS: Gráfico da distância de Mahalanobis para o modelo ST ajustado, sob o método de estimação ML, considerando $\Pi = 0.95$.

Tabela 2: AIS: MCR e TRC para o modelo ST ajustado, sob ambos métodos de estimação, quando observações atípicas são excluídas. Índice q indica que as medidas foram obtidas sob o método de estimação ML_q .

Observações excluídas	TRC	TRC _q	MRC	MRC _q
163	0.8005	0.6676	0.3954	0.1242
11, 43, 51, 53, 56, 75, 133, 160, 163 e 178	9.1757	4.1131	7.2594	1.0954

confirmando a flexibilidade das estimativas de ML_q do modelo ST com caudas pesadas contra observações atípicas. O algoritmo EM_q proposto, neste trabalho, fornece estimadores duplamente reponderados para os todos os parâmetros da distribuição ST e assim, este processo de estimação de θ tende a atribuir pesos menores ainda para as observações anômalas no sentido da distância de Mahalanobis.

Tabela 3: Pesos atribuídos às observações atípicas.

Índice i	\hat{u}_i	\hat{u}_{qi}
11	0.4697	0.0122
43	0.5134	0.0131
51	0.5082	0.0132
53	0.3683	0.0099
56	0.4114	0.0038
75	0.5067	0.0613
133	0.5896	0.0706
160	0.5802	0.0093
163	0.5378	0.0254
178	0.4505	0.0059

4 Considerações Finais

Em resumo, neste trabalho, apresentamos o método de estimação via máxima verossimilhança L_q , que pode ser considerado uma generalização da estimação de máxima verossimilhança usual,

para todos os parâmetros da distribuição Skew-t. Fornecemos um algoritmo EM modificado para o cálculo dos estimadores ML_q , mostrando a robustez do método e um melhor ajuste do modelo ST às observações atípicas, ao se estimar de forma simultânea os graus de liberdade com os demais parâmetros do modelo de interesse. Os resultados obtidos neste trabalho são uma generalização do trabalho de [13] que estudou o método ML_q para a distribuição t-Student. A aplicação numérica foi realizada no programa R; veja [14] para mais detalhes. O procedimento de estimação ML usual no modelo Skew-t, via algoritmo EM, foi feito usando o pacote *mixsmsn*, veja [15]. E os demais procedimentos foram programados.

Referências

- [1] A. Lucas. “Robustness of the student t based M-estimator”. Em: **Communications in Statistics-Theory and Methods** 26.5 (1997), pp. 1165–1182.
- [2] M. D. Branco e D. K. Dey. “A general class of multivariate skew-elliptical distributions”. Em: **Journal of Multivariate Analysis** 79.1 (2001), pp. 99–113.
- [3] D. C. Montgomery, E. A. Peck e G. G. Vining. **Introduction to linear regression analysis**. John Wiley & Sons, 2021.
- [4] A. Azzalini e A. Capitanio. “Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution”. Em: **Journal of the Royal Statistical Society: Series B (Statistical Methodology)** 65.2 (2003), pp. 367–389.
- [5] D. Ferrari. **Maximum Lq-likelihood estimation**. University of Minnesota, 2008.
- [6] J. Havrda e F. Charvát. “Quantification method of classification processes. Concept of structural α -entropy”. Em: **Kybernetika** 3.1 (1967), pp. 30–35.
- [7] C. Tsallis. “Possible generalization of Boltzmann-Gibbs statistics”. Em: **Journal of statistical physics** 52 (1988), pp. 479–487.
- [8] Y. Qin e C. E. Priebe. “Maximum L q-Likelihood estimation via the Expectation-Maximization algorithm: a robust estimation of mixture models”. Em: **Journal of the American Statistical Association** 108.503 (2013), pp. 914–928.
- [9] F. R. Hampel, E. M. Ronchetti, P. Rousseeuw e W. A. Stahel. **Robust statistics: the approach based on influence functions**. Wiley-Interscience; New York, 1986.
- [10] R. D. Cook e S. Weisberg. **An introduction to regression graphics**. Vol. 405. John Wiley & Sons, 2009.
- [11] B. Efron. “Computers and the theory of statistics: thinking the unthinkable”. Em: **SIAM review** 21.4 (1979), pp. 460–480.
- [12] S. Y. Lee, B. Lu e X. Y. Song. “Assessing local influence for nonlinear structural equation models with ignorable missing data”. Em: **Computational statistics & data analysis** 50.5 (2006), pp. 1356–1377.
- [13] F. Z. Doğru, Y. M. Bulut e O. Arslan. “Doubly reweighted estimators for the parameters of the multivariate t-distribution”. Em: **Communications in Statistics-Theory and Methods** 47.19 (2018), pp. 4751–4771.
- [14] R Development Core Team. “A language and environment for statistical computing”. Em: **[http://www. R-project. org](http://www.R-project.org)** (2009).
- [15] M. O. Prates, V. H. Lachos e C. R. B. Cabral. “mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions”. Em: **Journal of Statistical Software** 54 (2013), pp. 1–20.