Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Computação Científica
Departamento de Estatística

Inferência no Modelo de Grubbs t-Elíptico

Nelson Silva Lopes
Orientador: Prof. Dr. Filidor Edilfonso Vilca-Labra

Campinas - SP
2004

INFERÊNCIA NO MODELO DE GRUBBS t-ELÍPTICO

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Nelson Silva Lopes e aprovada pela comissão julgadora.

Campinas, 11 de Fevereiro de 2004.

Prof. Dr. Filidor Edilfonso Vilca-Labra
Orientador

Banca Examinadora:

1. Prof. Dr. Filidor Edilfonso Vilca-Labra (Orientador) - IMECC/UNICAMP.
2. Prof. Dra. Reiko Aoki - ICMC/USP
3. Profa. Dr. Luiz Koodi Hotta - IMECC/UNICAMP.

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do Título de Mestre em Estatística.

Aos meus pais Nivaldo e Aldete.

Agradecimentos

Aos meus pais, Nivaldo Moreira Lopes e Aldete Silva Lopes, que dedicaram suas vidas para educar seus filhos e jamais mediram esforços na realização desta tarefa. Ao meu irmão, Nivan Silva Lopes, que sempre "elevou minha moral", nos momentos de angústia.

Ao meu orientador, Professor Dr. Filidor Edilfonso Vilca-Labra, cuja dedicação, disposição e paciência foram fundamentais na realização deste trabalho. Sua preocupação constante com minha formação acadêmica e seus vários conselhos o revelaram, acima de tudo, um grande amigo.

À minha amada, Kárita Rachel, que nunca se cansou da difícil tarefa de colorir meus dias cinzentos.

Ao meu amigo, Júlio Vasquez, que foi a primeira pessoa a me incentivar a fazer o mestrado, que me acolheu em minha chegada em Campinas e me ajudou com minhas dúvidas nos primeiros cursos. Ao meu amigo, Joelmir Feliciano, por todos momentos de alegria e de dor compartilhados.

Aos mais novos amigos: Alexandre Rübesam, Helder Palaro, Ricardo Takeyama e Tatiana Benaglia, sempre presentes me auxiliando principalmente com meus problemas computacionais. A amizade cultivada nestes dois anos, foi pra mim a melhor herança de Campinas, e será guardada com muito carinho onde quer que vocês estiverem.

Ao Benilton Carvalho, por todos esforços na manutenção do nosso laboratório e pelo apoio computacional. À Rossana Mendoza, por toda força com o Latex. Ao colega de orietação, Luiz Zerbinatti, pela troca de experiências. Ao Anderson Motta e Clécio Ferreira, pelas sugestões feitas em minhas decisões acadêmicas. Às colegas de turma: Fernanda, Izabella e Paula. Aos demais colegas de mestrado: Cézar, Daniel, Roberta, Rodrigo, Samara e Vitor Hugo.

À todos os familiares e amigos, em especial a turma do NPOR-Goiânia-1998: "Guararapes 350 anos".

Ao IMECC, pelas excelentes condições de trabalho e à CAPES, pelo suporte financeiro.

Resumo

Neste trabalho apresentamos um estudo de inferência estatística no modelo proposto por Grubbs (1948,1973), comumente usado para comparar instrumentos ou métodos de medição, considerando que as observações seguem uma distribuição t-Student com algumas extensões aos modelos elípticos.

O estudo é baseado no método de máxima verossimilhança. Testes de hipóteses utilizando as estatísticas de Wald, Escore e Razão de Verossimilhanças são considerados. As estimativas de máxima verossimilhança sob o modelo t-Student são obtidas via algoritmo EM.

Apresentamos ainda um estudo do modelo de Grubbs com ponto de mudança. Neste estudo, consideramos estimadores obtidos pelo método dos momentos quando existe a presença de um ponto de mudança.

Finalmente, apresentamos alguns estudos de simulação e ilustramos os resultados teóricos usando dados encontrados na literatura.

Abstract

This work presents a statistical inference study of the model proposed by Grubbs (1948, 1973), usually employed to compare devices or measure methods, assuming that the data follow a t-Student distribution with some extensions to the elliptical models.

The study is based on the maximum likelihood method. We consider tests of hypotheses using the Wald, the Score and the Likelihood Ratio statistics. The maximum likelihood estimates under the assumption of a t-Student model are obtained via the EM algorithm.

We also present a study of the Grubbs model with changepoints. In this study we consider estimators obtained by the method of moments in the presence of a changepoint.

Finally we present some simulation studies and illustrate the theoretical results in data from literature.

Sumário

Lista de Tabelas

Lista de Figuras

1 *Introdução*

1.1 Formulação do problema

Diversas áreas científicas como a física, engenharia, biologia, entre outras, têm apresentado significativo interesse no problema de comparararação de instrumentos. Neste problema, pretende-se determinar a relação entre diversos testes ou instrumentos que fornecem medidas indiretas similares. Um exemplo é a capacidade pulmonar humana, que pode ser medida, por exemplo, por dois instrumentos, um mais tradicional e outro mais moderno, portátil e de operação mais simples. Submetendo um grupo de indivíduos aos dois instrumentos, as medidas (ambas sujeitas a erros) dos dois aparelhos podem ser comparadas. Grubbs (1948, 1973) compara três cronômetros, Barnett (1969) dá um exemplo onde quatro combinações instrumento-operador concebidos para medir a capacidade vital num grupo de pacientes são avaliados e Jaech (1985) dá vários exemplos na área industrial. Mais recentemente, Christensen e Blackwood (1993) comparam cinco "thermocouples"; Galea-Rojas (1996) apresenta um estudo de inferência no modelo t-Student e Bedrick (2001) compara três métodos para medir sedimentos de solo.

Em muitas situações práticas, por exemplo, na área industrial e agricultura, é assumido que todos os instrumentos medem a característica de interesse na mesma escala. Uma classe particular do modelo utilizado para comparar instrumentos de medição foi introduzida por Grubbs (1948, 1973). Outros trabalhos são, por exemplo, Christensen e Blackwood (1993) e Bedrick (2001), este último apresentando um estudo de inferência baseado na estatística de "Escore", usando uma parametrização diferente da qual será considerada neste trabalho. Este modelo é definido como segue: suponha que dispomos de p ($p \geq 3$) instrumentos para medir uma característica ou resposta x num grupo de n unidades. Sejam

x_j : O verdadeiro valor da característica de interesse na unidade j;

y_{ij} : Medida fornecida pelo instrumento i para a característica da unidade j, $i = 1, \cdots, p$ e $j = 1, \cdots, n$.

No estudo destes problemas estatísticos, Grubbs (1948,1973) propõe um modelo definido pelas relações

$$y_{ij} = \alpha_i + x_j + \epsilon_{ij}, \tag{1.1}$$

em que α_i é o vício aditivo do instrumento i e os erros aleatórios de medição ϵ_{ij}, sob a hipótese de normalidade que geralmente é assumida, são independentes e para cada i, $\epsilon_{i1}, \cdots, \epsilon_{in}$ são tais que

$$E(\epsilon_{ij}) = 0, Var(\epsilon_{ij}) = \phi_i, \tag{1.2}$$

$i = 1, \cdots, p, j = 1, \cdots, n$. Na literatura o modelo definido acima é conhecido como modelo de Grubbs.

O modelo definido em (1.1) - (1.2) representa uma classe particular de modelos de regressão linear com erros nas variáveis, amplamente discutida em Fuller (1987). Deste ponto de vista, o modelo definido em (1.1) - (1.2) pode ser especificado segundo as quantidades x_j como :

I. Os x_j's são constantes desconhecidas, chamadas de parâmetros incidentais.

II. Os x_j's são variáveis aleatórias não-correlacionados com média μ_x e variância ϕ_x e são não-correlacionados com ϵ_{ij}.

III. Os x_j's são variáveis aleatórias não-correlacionados com média μ_{xj} e variância ϕ_x e são não-correlacionados com ϵ_{ij}.

O modelo descrito por (I) e (1.1) - (1.2) é denominado *modelo funcional* de Grubbs, neste caso o número de parâmetros cresce com o número de observações. Para fazer um estudo de inferência neste modelo é necessário considerar suposições adicionais de alguns parâmetros, especialmente sobre as quantidades desconhecidas $x_1, \cdots, x_n$; veja por exemplo, Vilca Labra *et al.* (1998). Neste caso os parâmetros do modelo são

$$(x_1, \cdots, x_n)^\top \text{ e } \boldsymbol{\theta} = (\alpha_1, \cdots, \alpha_p, \phi_1, \cdots, \phi_p)^\top.$$

O modelo descrito por (II) e (1.1) - (1.2) é denominado *modelo estrutural* de Grubbs, ao contrário do *modelo funcional*, não é possível encontrar uma solução de máxima verossimilhança para os parâmetros do modelo, pois o mesmo não é identificável no sentido de que dois diferentes vetores de parâmetros podem dar origem a uma mesma distribuição. Assim, faz-se necessário o conhecimento adicional de algum parâmetro do modelo para possibilitar a estimação dos parâmetros de maior interesse. Neste caso o vetor de parâmetros do modelo é

$$\boldsymbol{\theta} = (\mu_x, \alpha_1, \cdots, \alpha_p, \phi_x, \phi_1, \cdots, \phi_p)^\top.$$

Finalmente, o modelo descrito por (III) e (1.1) - (1.2) é denominado *modelo ultraestrutural*. Neste caso, os modelos *funcional* e *estrutural* podem ser vistos como casos especiais deste modelo. Os parâmetros do modelo ultraestrutural são

$$(\mu_{x1}, \cdots, \mu_{xn})^\top \text{ e } \boldsymbol{\theta} = (\alpha_1, \cdots, \alpha_p, \phi_x, \phi_1, \cdots, \phi_p)^\top.$$

Do ponto de vista de aplicação, o modelo funcional é adequado quando temos interesse apenas nos indivíduos participantes do estudo, enquanto que o estrutural é adequado quando o interesse é generalizar os resultados do estudo para população de onde vem os indivíduos envolvidos no estudo.

Neste trabalho vamos considerar o modelo estrutural definido em (1.1) - (1.2) e (II). Este modelo pode ser escrito em forma matricial como:

$$\mathbf{Y}_j = \mathbf{a} + \mathbf{1}_p x_j + \boldsymbol{\epsilon}_j, \tag{1.3}$$

em que $\mathbf{a} = (\alpha_1, \cdots, \alpha_p)^\top$, $\mathbf{Y}_j = (y_{1j}, \cdots, y_{pj})^\top$, $\boldsymbol{\epsilon}_j = (\epsilon_{1j}, \cdots, \epsilon_{pj})^\top$ são vetores $p \times 1$, e $\mathbf{1}_p$ denota o vetor p-dimensional em que todos os componentes são iguais a unidade. Note que

$$E(\mathbf{Y}_j) = \mathbf{a} + \mathbf{1}_p \mu_x \;\; e \;\; Var(\mathbf{Y}_j) = \phi_x \mathbf{1}_p \mathbf{1}_p^\top + D(\boldsymbol{\phi}), j = 1, \cdots, n, \tag{1.4}$$

em que $D(\boldsymbol{\phi}) = Diag(\phi_1, \cdots, \phi_p)$ é uma matriz diagonal $p \times p$, $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_p)$. Sob as condições do modelo acima, o vetor de parâmetros é dado por

$$\boldsymbol{\theta} = (\mu_x, \mathbf{a}^\top, \phi_x, \boldsymbol{\phi}^\top)^\top. \tag{1.5}$$

Sob as condições do modelo estrutural definido em (1.3), $\boldsymbol{\theta}$ não é identificável, isto é, existem diferentes $\boldsymbol{\theta}_*$ e $\boldsymbol{\theta}_{**}$ tais que $F(\mathbf{y}, \boldsymbol{\theta}_*) = F(\mathbf{y}, \boldsymbol{\theta}_{**})$, para todo $\mathbf{y} \in \mathbb{R}^p$, onde F é a função de distribuição de $\mathbf{Y}_j, j = 1, \cdots, n$. Por exemplo, para o caso de $p = 3$, sejam

$$\boldsymbol{\theta}_* = (1, 1, 1, 1, 1, 1, 1, 1)^\top$$

e

$$\boldsymbol{\theta}_{**} = (4, -2, -2, -2, 1, 1, 1, 1)^\top,$$

de (1.4) é fácil verificar que para ambos valores de $\boldsymbol{\theta}$ definidos acima, geram a mesma média e a mesma matriz de covariância de $\mathbf{Y}_j$, isto é

$$E(\mathbf{Y}_j) = (2, 2, 2)^\top$$

e

$$Var(\mathbf{Y}_j) = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

A falta de identificabilidade implica ausência de estimadores consistentes, o que traria sérias implicações na teoria assintótica de estimação e teste de hipóteses. Contudo, o fato do modelo ser identificável não implica que exista um estimador consistente para $\boldsymbol{\theta}$. Uma das formas de contornar o problema de identificabilidade é colocar restrições sobre alguns elementos do vetor de parâmetros $\boldsymbol{\theta}$. Similarmente como em Barnett (1969) e Bedrick (2001), vamos assumir que exista um método de medição, o qual chamamos de instrumento de referência ou de controle, que mede x sem vício. Especificamente, supondo que o instrumento de referência seja o número 1, vamos assumir que

$$\alpha_1 = 0. \tag{1.6}$$

Assim, o modelo proposto por Grubbs (1948,1973) considerando um instrumento de referência, é dado por

$$\mathbf{Y}_j = \mathbf{a} + \mathbf{1}_p x_j + \boldsymbol{\epsilon}_j, j = 1, \cdots, n, \tag{1.7}$$

em que $\mathbf{a} = (0, \boldsymbol{\alpha}^\top)^\top$, com $\boldsymbol{\alpha} = (\alpha_2, \cdots, \alpha_p)^\top$, $\mathbf{Y}_j$ e $\boldsymbol{\epsilon}_j$ como em (1.3). Agora, sob a condição de identificabilidade, o vetor de parâmetros dados em (1.5) reduz-se a um vetor de dimensão $2p + 1$ dado por

$$\boldsymbol{\theta} = (\mu_x, \boldsymbol{\alpha}^\top, \phi_x, \boldsymbol{\phi}^\top)^\top. \tag{1.8}$$

No contexto do modelo de Grubbs, a qualidade de um instrumento de medição é avaliada em termos do vício ou acurácia e da precisão (inversa da variância). Contudo é importante distinguir os conceitos de acurácia e precisão; precisão se refere a proximidade de cada observação à sua própria média, enquanto que a acurácia mede a proximidade das medidas ao verdadeiro valor. Assim, a acurácia está relacionada com os vícios, enquanto que a precisão está relacionada com a dimensão dos erros de medida, ou seja, um instrumento é preciso se ϕ_i é pequena.

Pelo exposto acima, o problema de comparar instrumentos ou métodos de medição pode ser estudado fazendo inferência sobre os vícios $\alpha_2, \cdots, \alpha_p$ dos instrumentos $2, \cdots, p$, respectivamente, e comparando-os com o instrumento de referência. A precisão de um instrumento que é baseada na variância é um índice ou medida para avaliar a qualidade de um instrumento, veja por exemplo, Shyr e Gleser (1986).

A precisão do i-ésimo instrumento denotado por π_i, é definido como sendo:

$$\pi_i = \frac{1}{\phi_i}, i = 1, \cdots, p. \tag{1.9}$$

Outra medida ou índice utilizado na literatura, no contexto de calibração comparativa, é a confiabilidade do instrumento, veja por exemplo, Carter (1981). A confiabilidade do i-ésimo instrumento, denotado por ρ_i, define-se por

$$\rho_i = \frac{\phi_x}{\phi_x + \phi_i}, i = 1, \cdots, p. \tag{1.10}$$

O coeficiente ρ_i é uma medida tipo coeficiente de determinação usada em regressão linear e pode ser interpretada como a proporção da variação observada "explicada"pelo i-ésimo instrumento, ou seja, $\rho_i = \rho^2_{y_{ij},x_j}$ e $0 \leq \rho_i \leq 1, i = 1, \cdots, p$.

Das relações em (1.9) e (1.10), pode-se mostrar facilmente que

$$\rho_k = \rho_l \Leftrightarrow \pi_k = \pi_l \Leftrightarrow \phi_k = \phi_l, l \neq k = 1, \cdots, p, \tag{1.11}$$

isto é, comparar p instrumentos no modelo de Grubbs é equivalente a comparar as precisões dos instrumentos ou as variâncias dos erros.

Assim, comparar p instrumentos de medição utilizando o modelo de Grubbs reduz-se a fazer inferências sobre os vícios $\alpha_2, \cdots, \alpha_p$ e/ou as variâncias $\phi_1, \cdots, \phi_p$. Considerar os vícios e as variâncias simultaneamente é importante para verificar se o instrumento i fornece medidas exatas da característica x, pois um valor de ϕ_i próximo de zero implica apenas que y_i (medida fornecida pelo instrumento i) é uma medida confiável ou de alta precisão de $\alpha_i + x$. O instrumento forneceria medidas exatas da característica x só se ϕ_i for próximo de zero e se ele é não viciado ($\alpha_i = 0$).

Deste modo, uma hipótese inicial de interesse usando o modelo (1.7) é

$$H_{01} : \alpha_2 = \alpha_3 = \cdots = \alpha_p = 0 \text{ e } \pi_1 = \pi_2 = \cdots = \pi_p, \tag{1.12}$$

que, de (1.11), é equivalente à hipótese

$$H_{01} : \alpha_2 = \alpha_3 = \cdots = \alpha_p = 0 \text{ e } \phi_1 = \phi_2 = \cdots = \phi_p, \tag{1.13}$$

ou seja, estamos interessados em testar se os p instrumentos medem a característica x sem vício aditivo (acurácia) e com mesma precisão.

Outras hipóteses que poderiam ser consideradas são:

$$H_{02} : \alpha_2 = \alpha_3 = \cdots = \alpha_p = 0, \tag{1.14}$$

isto é, estamos avaliando a acurácia das medições feitas pelos diferentes instrumentos, e

$$H_{03} : \phi_1 = \phi_2 = \cdots = \phi_p, \tag{1.15}$$

para avaliar se os instrumentos são igualmente precisos.

Logicamente, dependendo do problema, poderia ser de interesse testar outro tipo de hipótese, embora as mencionadas acima sejam de interesse geral.

Na literatura existente sobre o modelo de Grubbs, é assumido que as observações seguem a distribuição normal. Entretanto, em muitas situações o modelo normal é claramente impróprio, por exemplo, quando os dados provêm de uma distribuição com caudas mais pesadas que a distribuição normal. A vulnerabilidade da inferência baseada no modelo normal, com respeito a observações aberrantes ("outliers"), tem sido o motivo de vários estudos sobre métodos para a análise de dados. Um enfoque é através da identificação e eliminação de "outliers"seguido de inferência estatística (máxima verossimilhança) baseada na distribuição normal. De fato, uma ampla área de pesquisa é a detecção de dados atípicos, especialmente em modelos lineares e análise multivariada. Algumas referências básicas são Barnett e Lewis (1994), Belsley et al. (1980), Chatterjee e Hadi (1988), Cook e Weisberg (1982), Hawkins (1980) e Peña e Yohai (1995). Na área de modelos de regressão com erros nas variáveis, dois trabalhos pioneiros são Kelly (1984) e Wellman e Guest (1991). Uma referência básica para influência local é Cook (1986).

Um outro enfoque é a acomodação de observações aberrantes usando procedimentos robustos clássicos. Nesta direção, métodos robustos têm sido desenvolvidos, modificando o método de máxima verossimilhança, de modo que os "outliers"tenham menos influência sobre as estimativas finais; ver Huber (1981), Hampel et al. (1986), Rousseeau e Leroy (1987) e Staudte e Sheather (1990). Na área de modelos de regressão com erros nas variáveis, algumas referências são Kelly (1984), Zamar (1989) e Cheng e Van Ness (1990).

Uma outra possibilidade é usar distribuições simétricas, com caudas mais pesadas que a distribuição normal, as quais permitem reduzir a influência dos "outliers"sobre os estimadores de máxima verossimilhança (EMV). Uma classe de distribuições que oferece várias alternativas neste sentido é a classe de distribuições elípticas. Esta classe de distribuições, que contém as distribuições normal, t-Student e normal contaminada, entre outras, têm tido um interesse crescente nos últimos anos; ver por exemplo, Fang e Zhang (1990), Fang et al. (1990), Fang e Anderson (1990), Kano et al. (1993), Berkane et al. (1994) e Arellano (1994).

Dizemos que um vetor aleatório $\mathbf{Y} : n \times 1$ tem distribuição elíptica, com parâmetro de locação $\boldsymbol{\mu}$ e matriz de escala $\boldsymbol{\Sigma}$, definida positiva, se sua função densidade pode ser expressa como

$$f_{\mathbf{Y}}(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} f((\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})), \mathbf{y} \in \mathbb{R}^n, \tag{1.16}$$

em que $\mathbf{y} \in \mathbb{R}^n$ e a função $f : \mathbb{R} \longrightarrow [0, \infty)$ é tal que $\int_0^\infty u^{n-1} f(u^2) du < \infty$. Para um vetor $\mathbf{Y}$ assim distribuído, usamos a notação $\mathbf{Y} \sim El_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; f)$. A função f é conhecida como função geradora de densidade.

A Tabela (1.1), apresenta alguns exemplos de distribuições elípticas multivariadas. As constantes $c_1, \cdots, c_5$ são usadas para denotar constantes normalizadoras.

Tabela 1.1: Distribuições Elípticas Multivariadas

Distribuição	Notação	Função Geradora
Normal	$N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$f(u) = c_1 e^{-u/2}, u \geq 0$
t-Student	$t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$	$f(u) = c_2(\nu + u)^{-(\nu+n)/2}, u \geq 0$
Cauchy	$C_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$f(u) = c_3(1 + u)^{-(n+1)/2}, u \geq 0$
Normal Contaminada	$NC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \delta, \tau)$	$f(u) = c_1(1 - \delta)e^{-u/2} + \delta\tau^{-n/2} e^{-u/(2\tau)}, u \geq 0$
Logística	$L_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$f(u) = c_4 e^{-u}/(1 + e^{-u})^2, u \geq 0$
Exponencial Potência	$EP_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \alpha)$	$f(u) = c_5 e^{-u^\alpha/2}, u \geq 0$

Deste modo, este trabalho trata de um estudo de inferência no modelo de Grubbs supondo que a distribuição das observações pertence à classe das distribuições elípticas.

Para testar as hipóteses de interesse, vamos utilizar as estatísticas da Razão de Verossimilhança, Escore e de Wald. Estimadores de máxima verossimilhança sob H_0 e sob o modelo irrestrito são apresentados no decorrer deste trabalho.

Também consideramos neste trabalho, as situações onde as características de interesse tenham diferentes médias. Em muitas situações, por exemplo o modelo de regressão linear com variável resposta sendo um fenômeno fisiológico, em que a inclinação da reta de regressão não é esperada ser constante, mas muda repentinamente em um certo ponto, deve-se utilizar um modelo que considere tal informação. Neste caso, este "ponto de mudança" que é de fundamental interesse, pois ele pode ser um indicador de um fenômeno fisiológico, tal como a menopausa em um gráfico da densidade óssea contra a idade em um estudo com mulheres. Deste modo, o problema do ponto de mudança no contexto do modelo de Grubbs será considerado.

1.2 Objetivos do Trabalho

O objetivo deste trabalho é apresentar um estudo de inferência estatística no modelo de Grubbs estrutural sob a classe das distribuições elípticas, com ênfase nas distribuições t-Student e normal. Sob o modelo t-Student, é apresentado um estudo de testes de hipóteses, baseado no método de máxima verossimilhança, de modo que as estimativas de máxima verossimilhança serão obtidas via algoritmo EM (Dempster et al.,1977). Além disso, seguindo a metodologia de Chang e Huang (1997) e Alfaro (2000), vamos apresentar um estudo do modelo de Grubbs com ponto de mudança. Assim, os objetivos específicos deste trabalho podem ser resumidos como segue:

1) Apresentar um estudo de inferência, considerando o modelo de Grubbs dependente e independente. As matrizes de informação observada e esperada serão obtidas. Estimadores de momentos e de máxima verossimilhança serão considerados, sendo que no segundo caso, obteremos as estimativas dos parâmetros via o algoritmo EM para o modelo t-Student independente.

2) Testar as hipóteses H_{01}, H_{02} e H_{03} utilizando a estatística Wald, Razão de verossimilhanças e Escore, com especial atenção no modelo t-Student independente.

3) Apresentar um estudo do modelo de Grubbs com ponto de mudança, considerando o teste da Razão de Verossimilhanças para detectar a presença de pontos de mudança na média da característica de interesse. Sob a presença do ponto de mudança, apresentar estimadores dos parâmetros.

4) Aplicar os resultados obtidos a dados reais. Além disso, realizar um estudo de simulação para avaliar o desempenho dos testes propostos.

1.3 Organização do Trabalho

Este trabalho é constituído por 5 capítulos e três apêndices. No Capítulo 2 apresentamos um estudo de inferência no modelo de Grubbs supondo que as observações seguem uma distribuição elíptica. São obtidas as matrizes de informação de Fisher observada e esperada, bem como o algoritmo EM para o modelo t-Student. No modelo independente, testes de hipótese baseados nas estatísticas de Wald, Escore e Razão de Verossimilhanças são discutidos. Estimadores do tipo momentos e suas propriedades assintóticas são consideradas.

No Capítulo 3 é feito um estudo de inferência do modelo de Grubbs com ponto de mudança. O teste de Razão de Verossimilhanças para testar a igualdade de médias é considerado. Propriedades assintóticas dos estimadores de momentos, sob a classe das distribuições elípticas, são apresentadas.

No Capítulo 4 aplicamos os resultados obtidos do modelo independente a dados reais e simulados. Realizamos um estudo sobre a convergência do algoritmo EM, apresentamos uma ilustração numérica e um estudo de simulação dos resultados obtidos no Capítulo 2. Também realizamos um pequeno estudo de simulação sobre os resultados do terceiro capítulo.

O Capítulo 5 é dedicado a conclusões e alguns direcionamentos para estudos subsequentes.

Finalmente, nos apêndices, apresentamos definições e algumas propriedades das distribuições elípticas, em particular a distribuição t-Student multivariada. Ainda nos apêndices, apresentamos os detalhes algébricos para a obtenção do teste de Razão de Verossimilhanças do modelo com ponto de mudança.

2 Modelo de Grubbs Elíptico

2.1 Introdução

No modelo introduzido por Grubbs (1948, 1973), supõe-se que os instrumentos de medição no problema de comparação de instrumentos medem a característica de interesse na mesma escala. Essas suposições são frequentemente assumidas em muitas aplicações científicas, por exemplo, na área industrial e agricultura.

Na maioria dos trabalhos encontrados na literatura é assumido que os ϵ_{ij}, como no modelo definido em (1.1), têm distribuição normal, são independentes e identicamente distribuídos (iid) através dos p instrumentos com média zero e variância ϕ_i, que os verdadeiros valores x_j são variáveis aleatórias tendo distribuição normal com média μ_x e variância ϕ_x, e independentes dos ϵ_{ij}. Assim $y_{ij} \sim N(\alpha_i + \mu_x, \phi_x + \phi_i)$ e $Cov(y_{ij}, y_{kj}) = \phi_x, i \neq k, i, k = 1, \cdots, p$ e $j = 1, \cdots, n$, com $\alpha_1 = 0$. Alguns resultados sobre este modelo podem ser encontrados, por exemplo, nos trabalhos: Grubbs(1948, 1973), Brindley e Brandley (1985) e Jaech (1985).

Galea-Rojas (1996), apresenta um estudo introdutório sob a classe das distribuções elípticas considerando a parametrização de análise de fatores, a qual dificulta a interpretação em termos dos parâmetros α_i's.

Neste capítulo apresentamos um estudo estatístico sobre o modelo estrutural de Grubbs, considerando que as observações seguem uma distribuição elíptica.

2.2 Especificação do Modelo Elíptico

Considere o modelo de Grubbs, definido em forma matricial como

$$\mathbf{Y}_j = \mathbf{a} + \mathbf{1}_p x_j + \boldsymbol{\epsilon}_j, \tag{2.1}$$

em que $\mathbf{a} = (\alpha_1, \cdots, \alpha_p)^\top$, $\mathbf{Y}_j = (y_{1j}, \cdots, y_{pj})^\top$, $\boldsymbol{\epsilon}_j = (\epsilon_{1j}, \cdots, \epsilon_{pj})^\top$ são vetores $p \times 1$ com $j = 1, \cdots, n$ e $\mathbf{1}_p$ denota o vetor p-dimensional em que todos os componentes são iguais a unidade.

Para especificar o modelo de Grubbs elíptico, vamos escrever o modelo acima como

$$\mathbf{Y}_j = \mathbf{a} + \mathbf{Q}\mathbf{r}_j, \tag{2.2}$$

em que $\mathbf{Q} = [\mathbf{1}_p | \mathbf{I}_p]$, $\mathbf{r}_j = (x_j, \boldsymbol{\epsilon}_j^\top)^\top$, $j = 1, \cdots, n$, com $\mathbf{a}$ como em (2.1) e $\mathbf{I}_k$ é a matriz identidade de ordem k. Assim a distribuição do modelo elíptico é caracterizada pela distribuição dos $\mathbf{r}_j, j = 1, \cdots, n$.

Os modelos elípticos, dos quais o modelo normal é um membro particular, têm aparecido com muita frequência na literatura estatística mais recente. Por exemplo, no contexto dos modelos lineares, vários autores têm usado estes modelos para explorar as consequências de ter erros não-normais ou com distribuição tendo caudas mais pesadas (ou mais leves) que a distribuição normal. Seguindo as especificações dos modelos encontrados na literatura, por exemplo, Vilca Labra *et al.* (1998), vamos especificar o modelo elíptico de duas formas, a saber:

(i) **Modelo elíptico dependente:** assume que os erros (ou as características de interesse) têm distribuição conjunta elíptica, de modo que eles podem ter covariância nula sem serem independentes. Veja, por exemplo, Zellner (1976); Ghosh e Sinha (1980); Ullah e Walsh (1984); Chib *et al.* (1988) e Anderson *et al.* (1986);

(ii) **Modelo elíptico independente:** supõe que os erros (ou as características de interesse) são independentes e têm distribuições marginais elípticas. Veja, por exemplo, Lange *et al.* (1989); Taylor (1992) e Kano *et al.* (1993).

Note que em *(ii)* a distribuição conjunta é o produto de distribuições elípticas, mas ela não é necessariamente elíptica. Convém ressaltar aqui que a distribuição conjunta ser elíptica juntamente com independência é equivalente a assumir normalidade.

Em Arellano (1994), mostra-se que o problema de identificabilidade é preservado para modelos de regressão elípticos com erros nas variáveis. Seguindo o mesmo procedimento, temos que a condição $\alpha_1 = 0$, torna o modelo de Grubbs elíptico identificável e esta condição será considerada no decorrer deste trabalho.

2.3 Modelo de Grubbs Dependente

Consideremos novamente o modelo definido em (2.1)-(2.2) sob a condição de identificabilidade $\alpha_1 = 0$, isto é, $\mathbf{a} = (0, \alpha_2, \cdots, \alpha_p)^\top$. Assim, o vetor de todas as observações pode ser escrito como

$$\mathbf{Y} = (\mathbf{Y}_1^\top, \cdots, \mathbf{Y}_n^\top)^\top = (\mathbf{1}_n \otimes \mathbb{I}_{(p)}^\top)\mathbf{a} + (\mathbf{1}_n \otimes \mathbf{Q})\mathbf{r}, \tag{2.3}$$

em que $\mathbb{I}_{(p)} = [\mathbf{0}_q|\mathbf{I}_q]$ matriz $q \times p$, $\mathbf{r} = (\mathbf{r}_1^\top, \cdots, \mathbf{r}_n^\top)^\top$ vetor $n(p+1)$-dimensional, $\boldsymbol{\alpha} = (\alpha_2, \cdots, \alpha_p)$ vetor q-dimensional, com $q = p - 1$. O símbolo $\otimes$ denota o produto de Kronecker e $\mathbf{0}_k$ é um vetor k-dimensional com todas as entradas iguais a zero.

Deste modo, o modelo de Grubbs elíptico dependente é definido pela suposição

$$\mathbf{r} \sim El_{n(p+1)}(\boldsymbol{\eta}_*, \boldsymbol{\Psi}_*),$$

em que $\boldsymbol{\eta}_* = \mathbf{1}_n \otimes \boldsymbol{\eta}$, $\boldsymbol{\Psi}_* = \mathbf{I}_n \otimes \boldsymbol{\Psi}$, com $\boldsymbol{\eta} = (\mu_x, 0, \cdots, 0)^\top$ vetor $(p+1)$-dimensional e $\boldsymbol{\Psi} = Diag(\phi_x, \boldsymbol{\phi})$ matriz $(p+1) \times (p+1)$ com $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_p)$.

Pela propriedade de linearidade da distribuição elíptica (veja Apêndice A), temos que

$$\mathbf{Y} \sim El_{np}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*), \tag{2.4}$$

em que $\boldsymbol{\mu}_* = (\mathbf{1}_n \otimes \mathbf{1}_p)\mu_x + (\mathbf{1}_n \otimes \mathbb{I}_{(p)}^\top)\mathbf{a}$, e $\boldsymbol{\Sigma}_* = \mathbf{I}_n \otimes \boldsymbol{\Sigma}$, com $\boldsymbol{\Sigma} = \phi_x \mathbf{1}_p \mathbf{1}_p^\top + D(\boldsymbol{\phi})$.

2.3.1 Matriz de Informação Observada

Supondo sua existência, temos que a função densidade de probabilidade do modelo (2.4), agora denotado por $\mathbf{Y} \sim El_{np}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*; f)$, é dada por

$$f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) = |\boldsymbol{\Sigma}_*|^{-1/2} f\{(\mathbf{y}-\boldsymbol{\mu}_*)^\top \boldsymbol{\Sigma}_*^{-1}(\mathbf{y}-\boldsymbol{\mu}_*)\}, \tag{2.5}$$

em que

$$|\boldsymbol{\Sigma}_*|^{-1/2} = |\mathbf{I}_n \otimes \boldsymbol{\Sigma}|^{-1/2} = (|\mathbf{I}_n|^p |\boldsymbol{\Sigma}|^n)^{-1/2} = |\boldsymbol{\Sigma}|^{-n/2}$$

e

$$\begin{aligned}
(\mathbf{y}-\boldsymbol{\mu}_*)^\top \boldsymbol{\Sigma}_*^{-1}(\mathbf{y}-\boldsymbol{\mu}_*) &= \sum_{j=1}^n (\mathbf{y}_j - (\mathbf{1}_p\mu_x + \mathbb{I}_{(p)}^\top \mathbf{a}))^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_j - (\mathbf{1}_p\mu_x + \mathbb{I}_{(p)}^\top \mathbf{a})) \\
&= \sum_{j=1}^n ||\mathbf{T}_j||^2,
\end{aligned}$$

com $||\mathbf{T}_j||^2 = (\mathbf{y}_j - \mathbf{a} - \mathbf{1}_p\mu_x)^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_j - \mathbf{a} - \mathbf{1}_p\mu_x)$, $\boldsymbol{\Sigma}^{-1} = D^{-1}(\boldsymbol{\phi}) - c^{-1}\mathbf{M}$, $|\boldsymbol{\Sigma}| = c|D(\boldsymbol{\phi})|$, $c = 1 + \phi_x \mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})\mathbf{1}_p$ e $\mathbf{M} = \phi_x D^{-1}(\boldsymbol{\phi})\mathbf{1}_p\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})$. Segue que a função de verossimilhança é dada por

$$L(\boldsymbol{\theta};\mathbf{y}) = |\boldsymbol{\Sigma}|^{-n/2} f\left(\sum_{j=1}^n ||\mathbf{T}_j||^2\right).$$

Assim, a função de log-verossimilhança pode ser escrita como

$$\ell(\boldsymbol{\theta}) = -\frac{n}{2}\log|\boldsymbol{\Sigma}| + \log f\left(\sum_{j=1}^n ||\mathbf{T}_j||^2\right). \tag{2.6}$$

a) Derivadas parciais de primeira ordem de $\ell(\boldsymbol{\theta})$

Derivando $\ell(\boldsymbol{\theta})$ com respeito a $\boldsymbol{\theta} = (\mu_x, \boldsymbol{\alpha}^\top, \phi_x, \boldsymbol{\phi}^\top)^\top$, temos que

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{n}{2}\frac{\partial}{\partial \boldsymbol{\theta}}\log|\boldsymbol{\Sigma}| + W_f\left(\sum_{j=1}^n ||\mathbf{T}_j||^2\right)\left[\sum_{j=1}^n \frac{\partial ||\mathbf{T}_j||^2}{\partial \boldsymbol{\theta}}\right], \tag{2.7}$$

em que

$$W_f(u) = \frac{f'(u)}{f(u)} \quad e \quad f'(u) = \left(\frac{df(x)}{dx}\right)\Big|_{x=u}.$$

Assim, temos que as derivadas de primeira ordem de $\ell(\boldsymbol{\theta})$ dependem das derivadas de $\log|\boldsymbol{\Sigma}|$ e $||\mathbf{T}_j||^2$ com respeito a $\mu_x, \boldsymbol{\alpha}, \phi_x$ e $\boldsymbol{\phi}$.

Derivando em forma vetorial e após algumas manipulações algébricas, temos que

$$\begin{aligned}
\frac{\partial}{\partial \mu_x} \log|\boldsymbol{\Sigma}| &= 0,\\
\frac{\partial}{\partial \boldsymbol{\alpha}} \log|\boldsymbol{\Sigma}| &= \mathbf{0}_q,\\
\frac{\partial}{\partial \phi_x} \log|\boldsymbol{\Sigma}| &= \frac{c-1}{c\phi_x},\\
\frac{\partial}{\partial \boldsymbol{\phi}} \log|\boldsymbol{\Sigma}| &= -c^{-1}\phi_x D^{-2}(\boldsymbol{\phi})\mathbf{1}_p + D^{-1}(\boldsymbol{\phi})\mathbf{1}_p,\\
\frac{\partial}{\partial \mu_x}||\mathbf{T}_j||^2 &= -2\mathbf{1}_p^\top \boldsymbol{\Sigma}^{-1}\mathbf{W}_j,\\
\frac{\partial}{\partial \boldsymbol{\alpha}}||\mathbf{T}_j||^2 &= -2\mathbb{I}_{(p)}\boldsymbol{\Sigma}^{-1}\mathbf{W}_j,\\
\frac{\partial}{\partial \phi_x}||\mathbf{T}_j||^2 &= -\frac{c^{-2}}{\phi_x}\mathbf{W}_j^\top \mathbf{M}\mathbf{W}_j,\\
\frac{\partial}{\partial \boldsymbol{\phi}}||\mathbf{T}_j||^2 &= -D(\mathbf{W}_j)D^{-2}(\boldsymbol{\phi})\mathbf{W}_j - c^{-2}\phi_x\mathbf{W}_j^\top \mathbf{M}\mathbf{W}_j D^{-2}(\boldsymbol{\phi})\mathbf{1}_p\\
&\quad +2c^{-1}\phi_x\mathbf{W}_j^\top D^{-1}(\boldsymbol{\phi})\mathbf{1}_p D^{-2}(\boldsymbol{\phi})\mathbf{W}_j,
\end{aligned}$$

com $D^{-k}(\mathbf{v}) = Diag(v_1^{-k}, \cdots, v_p^{-k})$, quando $\mathbf{v} = (v_1, \cdots, v_p)^\top$ é um vetor em $\mathbb{R}^p$ (esta notação também é válida se $\mathbf{v}$ é um vetor aleatório), $\mathbf{W}_j = \mathbf{Y}_j - \boldsymbol{\mu}$ e $\boldsymbol{\mu} = \mathbf{a} + \mathbf{1}_p\boldsymbol{\mu}_x$.

b) Derivadas parciais de segunda ordem de $\ell(\boldsymbol{\theta})$

Seja $\ell(\boldsymbol{\theta})$ a função de log-verossimilhança definida em (2.6). Então, a matriz de derivadas parciais de segunda ordem com respeito a $\boldsymbol{\theta}$ é dada por

$$\begin{aligned}\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} &= -\frac{n}{2}\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \log|\boldsymbol{\Sigma}| + W_f'\left(\sum_{j=1}^{n} ||\mathbf{T}_j||^2\right)\left[\sum_{j=1}^{n} \frac{\partial ||\mathbf{T}_j||^2}{\partial \boldsymbol{\theta}}\right]\left[\sum_{j=1}^{n} \frac{\partial ||\mathbf{T}_j||^2}{\partial \boldsymbol{\theta}^\top}\right] \\ &+ W_f\left(\sum_{j=1}^{n} ||\mathbf{T}_j||^2\right)\left[\sum_{j=1}^{n} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} ||\mathbf{T}_j||^2\right]. \end{aligned} \tag{2.8}$$

Assim, temos que as derivadas de segunda ordem de $\ell(\boldsymbol{\theta})$ dependem das derivadas de segunda ordem de $\log|\boldsymbol{\Sigma}|$ e de $||\mathbf{T}_j||^2$, com respeito a $\mu_x, \boldsymbol{\alpha}, \phi_x$ e $\boldsymbol{\phi}$.

Considerando as propriedades de derivadas vetoriais e após algumas manipulações algébricas, temos que

$$
\begin{aligned}
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \mu_x \partial \lambda^\top} &= 0, \quad \lambda = \mu_x, \boldsymbol{\alpha}, \phi_x, \boldsymbol{\phi}, \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\alpha} \partial \lambda^\top} &= 0, \quad \lambda = \boldsymbol{\alpha}, \phi_x, \boldsymbol{\phi}, \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \phi_x \partial \phi_x} &= -\frac{1}{\phi_x^2}\left(\frac{c-1}{c}\right)^2, \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \phi_x \partial \boldsymbol{\phi}^\top} &= -c^2 \mathbf{1}_p^\top D^{-2}(\boldsymbol{\phi}), \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}^\top} &= -D^{-2}(\boldsymbol{\phi}) + 2\phi_x c^{-1} D^{-3}(\boldsymbol{\phi}) - \phi_x c^{-2} D^{-1}(\boldsymbol{\phi}) \mathbf{M} D^{-1}(\boldsymbol{\phi}), \\
\frac{\partial^2 ||\mathbf{T}_j||^2}{\partial \mu_x \partial \mu_x} &= 2\mathbf{1}_p^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}_p, \\
\frac{\partial^2 ||\mathbf{T}_j||^2}{\partial \mu_x \partial \boldsymbol{\alpha}^\top} &= 2\mathbf{1}_p^\top \boldsymbol{\Sigma}^{-1} \mathbb{I}_{(p)}^\top, \\
\frac{\partial^2 ||\mathbf{T}_j||^2}{\partial \mu_x \partial \phi_x} &= 2\frac{c^{-2}}{\phi_x}(c-1)\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi}) \mathbf{W}_j, \\
\frac{\partial^2 ||\mathbf{T}_j||^2}{\partial \mu_x \partial \boldsymbol{\phi}^\top} &= 2c^{-1} \mathbf{W}_j^\top \boldsymbol{\Sigma}^{-1} D^{-1}(\boldsymbol{\phi}),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2||\mathbf{T}_j||^2}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^\top} &= 2\mathbb{I}_{(p)}\boldsymbol{\Sigma}^{-1}\mathbb{I}_{(p)}^\top,\\
\frac{\partial^2||\mathbf{T}_j||^2}{\partial\boldsymbol{\alpha}\partial\phi_x} &= 2\frac{c^{-2}}{\phi_x}\mathbb{I}_{(p)}\mathbf{M}\mathbf{W}_j,\\
\frac{\partial^2||\mathbf{T}_j||^2}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\phi}^\top} &= 2\mathbb{I}_{(p)}[\boldsymbol{\Sigma}^{-1}D(\mathbf{W}_j)D^{-1}(\boldsymbol{\phi}) + c^{-2}\phi_x\mathbf{M}\mathbf{W}_j\mathbf{1}_p^\top D^{-2}(\boldsymbol{\phi})\\
&\quad -c^{-1}\phi_x\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})\mathbf{W}_jD^{-2}(\boldsymbol{\phi})],\\
\frac{\partial^2||\mathbf{T}_j||^2}{\partial\phi_x\partial\phi_x} &= 2\frac{c^{-3}}{\phi_x}\mathbf{W}_j^\top\mathbf{M}\mathbf{W}_j\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})\mathbf{1}_p,\\
\frac{\partial^2||\mathbf{T}_j||^2}{\partial\phi_x\partial\boldsymbol{\phi}^\top} &= -2c^{-3}\mathbf{W}_j^\top\mathbf{M}\mathbf{W}_j\mathbf{1}_p^\top D^{-2}(\boldsymbol{\phi}) + 2c^{-2}\mathbf{W}_j^\top D^{-1}(\boldsymbol{\phi})\mathbf{1}_p\mathbf{W}_j^\top D^{-2}(\boldsymbol{\phi}),\\
\frac{\partial^2||\mathbf{T}_j||^2}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^\top} &= 2D^2(\mathbf{W}_j)D^{-3}(\boldsymbol{\phi}) - 2c^{-3}\phi_x\mathbf{W}_j^\top\mathbf{M}\mathbf{W}_jD^{-1}(\boldsymbol{\phi})\mathbf{M}D^{-1}(\boldsymbol{\phi})\\
&\quad +2\phi_xc^{-2}D^{-2}(\boldsymbol{\phi})\mathbf{W}_j\mathbf{W}_j^\top\mathbf{M}D^{-1}(\boldsymbol{\phi}) + 2c^{-2}\phi_x\mathbf{W}_j^\top\mathbf{M}\mathbf{W}_jD^{-3}(\boldsymbol{\phi})\\
&\quad +2c^{-2}\phi_xD^{-1}(\boldsymbol{\phi})\mathbf{M}\mathbf{W}_j\mathbf{W}_j^\top D^{-2}(\boldsymbol{\phi}) - 2c^{-1}\phi_xD^{-2}(\boldsymbol{\phi})\mathbf{W}_j\mathbf{W}_j^\top D^{-2}(\boldsymbol{\phi})\\
&\quad -4c^{-1}\phi_x\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})\mathbf{W}_jD(\mathbf{W}_j)D^{-3}(\boldsymbol{\phi}).
\end{aligned}
$$

Portanto, dos resultados acima e a equação (2.8) temos a matriz das derivadas de segunda ordem. Ainda da equação (2.8), podemos notar que sob o modelo elíptico dependente, a matriz de informação observada depende das funções $W_f(u)$ e $W_f'(u)$.

Na Tabela (2.1), vamos exibir $W_f(u)$ e $W_f'(u)$ para as distribuições dadas na Tabela (1.1).

Vale lembrar que o n da Tabela (2.1) é a dimensão do vetor $\mathbf{Y}$ apresentado na Tabela (1.1). Para o modelo elíptico dependente, este valor é np.

Assim a matriz informação observada é dada por:

$$\mathbf{I}_F = -\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}.$$

2.3.2 Matriz de Informação Esperada

Tabela 2.1: Expressões de $W_f(u)$ e $W'_f(u)$ para as distribuições elípticas da Tabela (1.1)

Distribuição	$W_f(u)$	$W'_f(u)$
Normal	$-1/2$	0
t-Student	$-((\nu+n)/2)(\nu+u)^{-1}$	$((n+\nu)/2)(\nu+u)^{-2}$
Cauchy	$-((1+n)/2)(1+u)^{-1}$	$((1+n)/2)(1+u)^{-2}$
Normal Contaminada	$-\frac{1}{2}\frac{f_1(u)}{f_0(u)}$	$\frac{1}{4}\left\{\frac{f_2(u)}{f_0(u)} - \left[\frac{f_1(u)}{f_0(u)}\right]^2\right\}$
Logística	$-\tanh(u/2)$	$-2/(e^{u/2}+e^{-u/2})^2$
Exponencial Potência	$-\frac{\alpha}{2}u^{\alpha-1}$	$-\frac{\alpha(\alpha-1)}{2}u^{\alpha-2}$

Nota: $f_i(u) = (1-\delta)e^{-u/2} + \delta\tau^{-(n/2)-i}e^{-u/2\tau}$, $i = 0, 1, 2$.

Nesta sub-seção apresentamos a matriz de informação esperada sob o modelo elíptico de Grubbs dependente definido em (2.4).

Teorema 2.1. *Sob o modelo elíptico de Grubbs dependente (2.4) a matriz de informação esperada por elemento, $J_F(\boldsymbol{\theta})$, é dada por*

$$J_F(\boldsymbol{\theta}) = \begin{pmatrix} J_{\mu_x\mu_x} & J_{\mu_x\boldsymbol{\alpha}} & 0 & 0 \\ J_{\boldsymbol{\alpha}\mu_x} & J_{\boldsymbol{\alpha}\boldsymbol{\alpha}} & 0 & 0 \\ 0 & 0 & J_{\phi_x\phi_x} & J_{\phi_x\boldsymbol{\phi}} \\ 0 & 0 & J_{\boldsymbol{\phi}\phi_x} & J_{\boldsymbol{\phi}\boldsymbol{\phi}} \end{pmatrix},$$

em que

$$J_{\mu_x\mu_x} = \frac{4n}{d}a_f(1,2)\mathbf{1}_p^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}_p,$$

$$J_{\mu_x\boldsymbol{\alpha}} = \frac{4n}{d}a_f(1,2)\mathbf{1}_p\boldsymbol{\Sigma}^{-1}\mathbb{I}_{(p)}^\top,$$

$$J_{\boldsymbol{\alpha}\boldsymbol{\alpha}} = \frac{4n}{d}a_f(1,2)\mathbb{I}_{(p)}\boldsymbol{\Sigma}^{-1}\mathbb{I}_{(p)}^\top,$$

$$J_{\phi_x\phi_x} = c^{-2}\frac{(c-1)^2}{\phi_x}\left[n\left(\frac{n}{4} + \frac{na_f(1,1)}{d} + \frac{a_f(2,2)}{d(d+2)}\right) + \frac{2na_f(2,2)}{d(d+2)}\right],$$

$$
\begin{aligned}
J_{\phi_x \boldsymbol{\phi}} &= n\left(\frac{n}{4}+\frac{na_f(1,1)}{d}+\frac{a_f(2,2)}{d(d+2)}\right)\frac{1}{\phi_x}\frac{c-1}{c}\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi}) \\
&+\left(\frac{2na_f(2,2)}{d(d+2)}\frac{1}{c^2}-n\left(\frac{n}{4}+\frac{na_f(1,1)}{d}+\frac{a_f(2,2)}{d(d+2)}\right)\frac{c-1}{c^2}\right)\mathbf{1}_p D^{-2}(\boldsymbol{\phi}),
\end{aligned}
$$

$$
\begin{aligned}
J_{\boldsymbol{\phi}\boldsymbol{\phi}} &= n\left(\frac{n}{4}+\frac{na_f(1,1)}{d}+\frac{a_f(2,2)}{d(d+2)}\right)\left[D^{-1}(\boldsymbol{\phi})\mathbf{1}_p\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})-\frac{\phi_x}{c}D^{-1}(\boldsymbol{\phi})\mathbf{1}_p\mathbf{1}_p^\top D^{-2}(\boldsymbol{\phi})\right. \\
&\left.-\frac{\phi_x}{c}D^{-2}(\boldsymbol{\phi})\mathbf{1}_p\mathbf{1}^\top D^{-1}(\boldsymbol{\phi})+\frac{\phi_x^2}{c^2}D^{-2}(\boldsymbol{\phi})\mathbf{1}_p\mathbf{1}^\top D^{-2}(\boldsymbol{\phi})\right] \\
&+\frac{2na_f(2,2)}{d(d+2)}\left[D^{-2}(\boldsymbol{\phi})-\frac{2\phi_x}{c}D^{-3}(\boldsymbol{\phi})+\frac{\phi_x^2}{c^2}D^{-2}(\boldsymbol{\phi})\mathbf{1}_p\mathbf{1}_p^\top D^{-2}(\boldsymbol{\phi})\right],
\end{aligned}
$$

com

$$
a_f(i,j)=E\left\{(\sum_{k=1}^{n}||\mathbf{T}_k||^2)^i\left(W_f\left(\sum_{k=1}^{n}||\mathbf{T}_k||^2\right)\right)^j\right\},
$$

tal que $i=1,2,\ j=1,2\ e\ i\le j.$

Prova: Seja

$$
J_F(\boldsymbol{\theta})=[J_{i,j}]=\left[E\left(\frac{\partial\ell}{\partial\theta_i}\frac{\partial\ell}{\partial\theta_j}\right)\right],
$$

a matriz de informação do modelo de Grubbs dependente. Usando a seguinte fórmula dada por Arellano (1994),

$$
\begin{aligned}
J_{i,j} &= \frac{4n}{d}a_f(1,2)\frac{\partial\boldsymbol{\mu}^\top}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\theta_j} \\
&+n\left(\frac{n}{4}+\frac{n}{d}a_f(1,1)+\frac{1}{d(d+2)}a_f(2,2)\right)tr\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\right)tr\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\right) \\
&+\frac{2n}{d(d+2)}a_f(2,2)tr\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\right)
\end{aligned}
$$

em que d denota a dimensão de $\mathbf{Y}$ ($d=np$), o resultado segue depois de uma certa álgebra.

Em particular, para o modelo normal tem-se que (ver Arellano, 1994)

$$a_f(1,2) = \frac{d}{4},\ a_f(1,1) = -\frac{d}{2} \text{ e } a_f(2,2) = \frac{d(d+2)}{4}.$$

No caso da distribuição t-student com ν graus de liberdade tem-se

$$a_f(1,2) = \frac{d(\nu+d)}{4(\nu+d+2)},\ a_f(1,1) = -\frac{d}{2} \text{ e } a_f(2,2) = \frac{(\nu+d)d(d+2)}{4(\nu+d+2)}.$$

Notemos que se $\nu \longrightarrow \infty$, obtemos os resultados correspondentes ao modelo normal.

2.3.3 Estimação de Máxima Verossimilhança

Nesta subseção mostramos como obter os EMV para o modelo elíptico dependente (2.4)-(2.5).

Assim como em Anderson *et al.* (1986) e Vilca-Labra *et al.* (1998), vamos assumir que a função geradora de densidade f é decrescente e continuamente diferenciável no intervalo $(0,\infty)$. Essa suposição e as proposições que apresentaremos a seguir serão fundamentais para o objetivo, já citado, desta subseção.

Proposição 2.1. *Seja $(\hat{\boldsymbol{\mu}}_*, \hat{\boldsymbol{\Sigma}}_*)$ o EMV de $(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*)$. Então sob a condição de identificabilidade, o EMV $\hat{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$, é solução das equações*

$$\boldsymbol{\mu}_*(\hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\mu}}_* \text{ e } \boldsymbol{\Sigma}_*(\hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\Sigma}}_*.$$

Prova: Segue da propriedade de invariância dos EMV.

Proposição 2.2. *Seja $(\hat{\boldsymbol{\mu}}_{*N}, \hat{\boldsymbol{\Sigma}}_{*N})$ os EMV de $(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*)$ sob a hipótese de normalidade no modelo elíptico dado em (2.4). Então os EMV de $(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*)$ do modelo elíptico dependente são dadas por*

$$\hat{\boldsymbol{\mu}}_* = \hat{\boldsymbol{\mu}}_{*N} \text{ e } \boldsymbol{\Sigma}_* = \frac{np}{u_0}\hat{\boldsymbol{\Sigma}}_{*N},$$

em que u_0 é o máximo da função $u^{\frac{np}{2}} f(u)$, $u > 0$.

Prova: Segue imediatamente do Teorema 1 de Anderson et al. (1986).

Das proposições acima e os EMV sob o modelo normal temos que o EMV de $\boldsymbol{\theta}$ sob o modelo elíptico dependente é solução das equações da Proposição 2.2. O EMV sob a hipótese de normalidade é apresentado em Lachos-Dávila (2002). Este estimador é obtido através do algoritmo EM.

Note que no caso especial em que $f(u) = k(n,\nu)\nu^{\frac{1}{2}\nu}(\nu+u)^{-\frac{1}{2}(\nu+np)}$, com $k(n,\nu)$ como em (B.3) correspondente à distribuição t-Student com ν graus de liberdade, nós temos que $u_0 = np$, para todo $\nu > 0$. Ou seja, o EMV do modelo t-Student dependente é igual ao do modelo normal.

2.4 Modelo de Grubbs Independente

Se no modelo definido em (2.1)-(2.2), sob a condição de identificabilidade, considerarmos que $\mathbf{r_1}, \cdots, \mathbf{r_n}$ são iid tais que

$$\mathbf{r}_j \sim El_p(\boldsymbol{\eta}, \boldsymbol{\Psi}; f), \quad j = 1, \cdots, n, \tag{2.9}$$

em que $\boldsymbol{\eta} = (\mu_x, 0, \cdots, 0)^\top$ vetor $(p+1)$-dimensional e $\boldsymbol{\Psi} = Diag(\phi_x, \boldsymbol{\phi})$, então teremos o modelo elíptico de Grubbs independente

$$\mathbf{Y}_j \sim El_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; f), \quad j = 1, \cdots, n, \tag{2.10}$$

em que $\boldsymbol{\mu} = \mathbf{a} + \mathbf{1}_p\mu_x$ e $\boldsymbol{\Sigma} = \phi_x \mathbf{1}_p \mathbf{1}_p^\top + D(\boldsymbol{\phi})$, com $\mathbf{a} = (0, \alpha_1, \cdots, \alpha_p)^\top$ e $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_p)^\top$.

2.4.1 Matriz de Informação Observada

A função de densidade de probabilidade no modelo de Grubbs independente é dada por

$$f_{\mathbf{Y}_j}(\mathbf{y}_j; \boldsymbol{\theta}) = |\boldsymbol{\Sigma}|^{-1/2} f\left((\mathbf{y}_j - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y}_j - \boldsymbol{\mu})\right).$$

Logo, para uma amostra aleatória $Y_1, \cdots, Y_n$, a função de verossimilhança é dada por

$$L(\boldsymbol{\theta}; \mathbf{y}_1, \cdots, \mathbf{y}_n) = \prod_{j=1}^{n} |\boldsymbol{\Sigma}|^{-1/2} f\left((\mathbf{y}_j - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_j - \boldsymbol{\mu})\right),$$

e a função de log-verossimilhança pode ser escrita como

$$\ell(\boldsymbol{\theta}) = \sum_{j=1}^{n} \ell_j(\boldsymbol{\theta}) = \sum_{j=1}^{n} \log f_{\mathbf{Y}_j}(\mathbf{y}_j; \boldsymbol{\theta}), \tag{2.11}$$

em que

$$\ell_j(\boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}| + \log f\left((\mathbf{y}_j - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_j - \boldsymbol{\mu})\right), j = 1, \cdots, n.$$

Se tomarmos $\mathbf{T}_j = \boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_j - \boldsymbol{\mu})$, então

$$\ell_j(\boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}| + \log f(||\mathbf{T}_j||^2), j = 1, \cdots, n.$$

a) Derivadas parciais de primeira ordem de $\boldsymbol{\ell(\theta)}$

Como $\ell(\boldsymbol{\theta}) = \sum_{j=1}^{n} \ell_j(\boldsymbol{\theta})$, então $\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{j=1}^{n} \frac{\partial \ell_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, em que

$$\frac{\partial \ell_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\theta}}\log|\boldsymbol{\Sigma}| + W_f(||\mathbf{T}_j||^2)\frac{\partial ||T_j||^2}{\partial \boldsymbol{\theta}}.$$

Assim, temos que as derivadas de primeira ordem de $\ell(\boldsymbol{\theta})$ dependem, novamente, das derivadas de $\log|\boldsymbol{\Sigma}|$ e $||\mathbf{T}_j||^2$ com respeito a μ_x,$\boldsymbol{\alpha}$,ϕ_x e $\boldsymbol{\phi}$ que são dadas na Subseção 2.2.1.

b) Derivadas parciais de segunda ordem de $\boldsymbol{\ell(\theta)}$

A matriz de derivadas parciais de segunda ordem com respeito a $\boldsymbol{\theta}$, é dada por

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \sum_{j=1}^{n} \frac{\partial^2 \ell_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top},$$

em que

$$\frac{\partial^2 \ell_j(\theta)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = -\frac{1}{2}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\log|\boldsymbol{\Sigma}| + W_f'(||\mathbf{T}_j||^2)\frac{\partial}{\partial\boldsymbol{\theta}}||\mathbf{T}_j||^2\frac{\partial}{\partial\boldsymbol{\theta}^\top}||\mathbf{T}_j||^2 + W_f(||\mathbf{T}_j||^2)\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}||\mathbf{T}_j||^2.$$

Novamente, temos que as derivadas de segunda ordem de $\ell(\boldsymbol{\theta})$ dependem das derivadas de segunda ordem de $\log|\boldsymbol{\Sigma}|$ e de $||\mathbf{T}_j||^2$ com respeito a μ_x,$\boldsymbol{\alpha}$,ϕ_x e $\boldsymbol{\phi}$, que foram apresentadas na Subseção 2.2.1.

2.4.2 Matriz de Informação Esperada

Nesta subseção obtemos a matriz de informação esperada no modelo elíptico de Grubbs independente (2.10).

Teorema 2.2. *Sob o modelo elíptico de Grubbs independente (2.10), a matriz de informação esperada por elemento,* $K_F(\boldsymbol{\theta})$*, é dada por*

$$K_F(\boldsymbol{\theta}) = \begin{pmatrix} K_{\mu_x\mu_x} & K_{\mu_x\boldsymbol{\alpha}} & 0 & 0 \\ K_{\boldsymbol{\alpha}\mu_x} & K_{\boldsymbol{\alpha}\boldsymbol{\alpha}} & 0 & 0 \\ 0 & 0 & K_{\phi_x\phi_x} & K_{\phi_x\boldsymbol{\phi}} \\ 0 & 0 & K_{\boldsymbol{\phi}\phi_x} & K_{\boldsymbol{\phi}\boldsymbol{\phi}} \end{pmatrix},$$

em que

$$
\begin{aligned}
K_{\mu_x\mu_x} &= \frac{4n}{d} b_f(1,2)\mathbf{1}_p^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}_p,\\
K_{\mu_x\boldsymbol{\alpha}} &= \frac{4n}{d} b_f(1,2)\mathbf{1}_p^\top \boldsymbol{\Sigma}^{-1}\mathbb{I}_{(p)}^\top,\\
K_{\boldsymbol{\alpha}\boldsymbol{\alpha}} &= \frac{4n}{d} b_f(1,2)\mathbb{I}_{(p)}\boldsymbol{\Sigma}^{-1}\mathbb{I}_{(p)}^\top,\\
K_{\phi_x\phi_x} &= c^{-2}\frac{(c-1)^2}{\phi_x^2}\left[n\left(\frac{1}{4}+\frac{b_f(1,1)}{d}+\frac{b_f(2,2)}{d(d+2)}\right)+\frac{2nb_f(2,2)}{d(d+2)}\right],\\
K_{\phi_x\boldsymbol{\phi}} &= n\left(\frac{1}{4}+\frac{b_f(1,1)}{d}+\frac{b_f(2,2)}{d(d+2)}\right)\frac{1}{\phi_x}\frac{c-1}{c}\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})\\
&\quad+\left(\frac{2nb_f(2,2)}{d(d+2)}\frac{1}{c^2}-n\left(\frac{1}{4}+\frac{b_f(1,1)}{d}+\frac{b_f(2,2)}{d(d+2)}\right)\frac{c-1}{c^2}\right)\mathbf{1}_p^\top D^{-2}(\boldsymbol{\phi}),
\end{aligned}
$$

$$
\begin{aligned}
K_{\boldsymbol{\phi}\boldsymbol{\phi}} &= n\left(\frac{1}{4}+\frac{b_f(1,1)}{d}+\frac{b_f(2,2)}{d(d+2)}\right)\Big[D^{-1}(\boldsymbol{\phi})\mathbf{1}_p\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})-\frac{\phi_x}{c}D^{-1}(\boldsymbol{\phi})\mathbf{1}_p\mathbf{1}_p^\top D^{-2}(\boldsymbol{\phi})\\
&\quad-\frac{\phi_x}{c}D^{-2}(\boldsymbol{\phi})\mathbf{1}_p\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})+\frac{\phi_x^2}{c^2}D^{-2}(\boldsymbol{\phi})\mathbf{1}_p\mathbf{1}_p^\top D^{-2}(\boldsymbol{\phi})\Big]\\
&\quad+\frac{2nb_f(2,2)}{d(d+2)}\left[D^{-2}(\boldsymbol{\phi})-\frac{2\phi_x}{c}D^{-3}(\boldsymbol{\phi})+\frac{\phi_x^2}{c^2}D^{-2}(\boldsymbol{\phi})\mathbf{1}_p\mathbf{1}_p^\top D^{-2}(\boldsymbol{\phi})\right],
\end{aligned}
$$

com

$$b_f(i,j)=E\left\{||\mathbf{T}_j||^{2i}\left(W_f(||\mathbf{T}_j||^2)\right)^j\right\},$$

tal que $i=1,2,\ j=1,2$ *e* $i\leq j$.

Prova: Seja

$$K_F(\boldsymbol{\theta})=[K_{i,j}]=\left[E\left(\frac{\partial\ell}{\partial\theta_i}\frac{\partial\ell}{\partial\theta_j}\right)\right],$$

a matriz de informação do modelo de Grubbs independente. Usando a seguinte fórmula dada por Arellano (1994),

$$\begin{aligned} K_{i,j} &= \frac{4n}{d} b_f(1,2) \frac{\partial \boldsymbol{\mu}^\top}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_j} \\ &+ n \left(\frac{1}{4} + \frac{1}{d} b_f(1,1) + \frac{1}{d(d+2)} b_f(2,2) \right) tr \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \right) tr \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \right) \\ &+ \frac{2n}{d(d+2)} b_f(2,2) tr \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \right) \end{aligned}$$

em que d denota a dimensão de $\mathbf{Y}_j$ $(d = p)$, $j = 1, \cdots, n$, o resultado segue depois de uma certa álgebra.

Em particular, para o modelo normal tem-se que (ver Arellano, 1994)

$$b_f(1,2) = \frac{d}{4},\ b_f(1,1) = -\frac{d}{2} \text{ e } b_f(2,2) = \frac{d(d+2)}{4}.$$

No caso da distribuição t-student com ν graus de liberdade $(\nu > 2)$ tem-se

$$b_f(1,2) = \frac{d(\nu+d)}{4(\nu+d+2)},\ b_f(1,1) = -\frac{d}{2} \text{ e } b_f(2,2) = \frac{(\nu+d)d(d+2)}{4(\nu+d+2)}.$$

Notemos que se $\nu \longrightarrow \infty$ obtemos os resultados correspondentes ao modelo normal.

2.4.3 Estimação de Máxima Verossimilhança

Nesta subseção vamos obter os estimadores de máxima verossimilhança (EMV) para o modelo elíptico independente. Para efeito deste trabalho vamos considerar a distribuição t-student.

Recentemente, vários autores têm sugerido a distribuição t-Student multivariada como alternativa à normal multivariada por ter as caudas mais pesadas e, portanto, "acomoda"possíveis "outliers"presentes nos dados. Assim, o uso da distribuição t-Student forneceria um procedimento robusto de análise de dados. Por exemplo, Rubin (1983) obteve os EMV para os parâmetros da distribuição t-Student multivariada usando algoritmo EM; Little (1988) estende o resultado de Rubin (1983) a dados incompletos, Sutradhar e Ali (1986) usam máxima verossimilhança no modelo de regressão com distribuição t-Student. Lange *et al.* (1989) ilustram o uso da distribuição t-student em regressão e em algumas áreas de análise multivariada, ver também Taylor (1992). Sutradhar (1993) propôs um teste de Escore para testar que a matriz de covariância é igual a uma matriz dada, usando também a distribuição t-Student; Bolfarine e Arellano (1994) usam a distribuição t-Student no modelo linear simples com erros nas variáveis.

Devido à inferência estatística baseada na distribuição t-Student combinar simplicidade conceitual e computacional com generalidade, já que pode ser aplicada numa variedade de situações, é que propomos como alternativa à distribuição normal, a distribuição t-Student multivariada com segundos momentos finitos.

a) O algoritmo EM para o modelo t-Student

Vamos supor que em (2.9) temos

$$\mathbf{r}_j \sim t_{p+1}(\boldsymbol{\eta}, \boldsymbol{\Psi}; \nu), j = 1, \cdots, n, \tag{2.12}$$

em que $\boldsymbol{\eta}$ e $\boldsymbol{\Psi}$ dados em (2.9). Então teremos

$$\mathbf{Y}_j \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu), \quad j = 1, \cdots, n, \tag{2.13}$$

em que $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ são como (2.10). Ainda, sob o modelo acima, temos que a função densidade de probabilidade de $\mathbf{Y}_j$ é dada por

$$f_{\mathbf{Y}}(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} k(n, \nu) \nu^{\frac{1}{2}\nu} [\nu + (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})]^{-\frac{1}{2}(\nu+p)}, \mathbf{y} \in \mathbb{R}^p. \tag{2.14}$$

em que $k(n, \nu)$ é dado em (B.3).

Tal como é ilustrado em Lange *et al.* (1989) em diferentes contextos, a conveniência deste modelo vem dada pelo fato de que se ν é fixado em um valor adequado, então seus EMV são robustos com respeito a observações aberrantes. No Capítulo 4 faremos uma discussão sobre a escolha de ν.

Encontrar os EMV de $\boldsymbol{\theta}$ por métodos convencionais é complicado, dado que as equações resultantes não tem forma fechada. Embora existam muitas metodologias para encontrar os EMV, aqui mostramos como obter os EMV de $\boldsymbol{\theta}$ usando o algoritmo EM, o qual é fácil de implementar e computacionalmente conveniente; além disso, as estimativas das variâncias são todas não negativas.

Como é conhecido, o algoritmo EM (Dempster *et al.*, 1977) aumenta os dados $\mathbf{Y}$= $(\mathbf{Y}_1, \cdots, \mathbf{Y}_n)^\top$ por adicionamento de dados hipotéticos, ou não-observáveis, neste caso vamos considerar $\mathbf{x} = (x_1, \cdots, x_n)^\top$ de tal forma que o EMV de $\boldsymbol{\theta}$ baseado em $\mathbf{Z} = (\mathbf{Y}, \mathbf{x})$ seja fácil de calcular. Dada a estimativa $\boldsymbol{\theta}^{(m)}$, na iteração m, a iteração $m+1$ do algoritmo EM consiste de duas etapas: uma etapa **E** de esperança e uma etapa **M** de maximização. Na etapa **E** calculamos o valor esperado do logaritmo da função de verossimilhança completa $l(\boldsymbol{\theta}|\mathbf{Z})$ com respeito à distribuição condicional de $\mathbf{x}$ dado $\mathbf{Y}$ e $\boldsymbol{\theta}^{(m)}$. Na etapa **M** maximizamos a função resultante com respeito a $\boldsymbol{\theta}$, e obtemos assim uma nova estimativa $\boldsymbol{\theta}^{(m+1)}$. Cada iteração do EM incrementa o logaritmo da função de verossimilhança observada $\ell(\boldsymbol{\theta}|\mathbf{Y}); \ell(\boldsymbol{\theta}^{(m)}|\mathbf{Y}) \leq \ell(\boldsymbol{\theta}^{(m+1)}|\mathbf{Y})$.

Seja $\mathbf{Z}_j = \begin{pmatrix} x_j \\ \mathbf{Y}_j \end{pmatrix}$, $j = 1, \cdots, n$. Então, de (2.12) temos que

$$\mathbf{Z}_j \overset{iid}{\sim} t_{p+1}(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z; \nu), \quad j = 1, \cdots, n, \tag{2.15}$$

em que

$$\boldsymbol{\mu}_z = \begin{pmatrix} \mu_x \\ \boldsymbol{\mu} \end{pmatrix} \quad e \quad \boldsymbol{\Sigma}_z = \begin{pmatrix} \phi_x & \mathbf{1}_p^\top \phi_x \\ \mathbf{1}_p \phi_x & \phi_x \mathbf{1}_p \mathbf{1}_p^\top + D(\boldsymbol{\phi}) \end{pmatrix}.$$

Sabemos que a distribuição t-Student pode ser obtida como mistura de uma distribuição normal e uma χ^2 (ver Apêndice B). Assim, seja

$$\mathbf{Z}_j | Q_j = q_j \sim N_{p+1}(\boldsymbol{\mu}_z, q_j^{-1} \boldsymbol{\Sigma}_z), Q_j \sim \chi^2_{(\nu)}/\nu, \quad j = 1, \cdots, n \tag{2.16}$$

e $f_{\mathbf{Z}_j, Q_j}(\mathbf{Z}_j, q_j)$ a densidade conjunta de $(\mathbf{Z}_j, q_j)$, $j = 1, \cdots, n$, que pode ser escrita como

$$f_{\mathbf{Z}_j, Q_j}(\mathbf{z}_j, q_j) = f_1(\mathbf{z}_j | q_j) f_2(q_j), \quad j = 1, \cdots, n,$$

em que f_1 é a função de densidade condicional de $\mathbf{Z}_j$ dado $Q_j = q_j$ e f_2 é a função densidade de Q_j, $j = 1, \cdots, n$.

Logo, a log-verossimilhança completa, $\ell_{\mathbf{Z}}(\boldsymbol{\theta},\nu)$ é dada por

$$\begin{aligned}\ell_{\mathbf{Z}}(\boldsymbol{\theta},\nu) &= cte - \frac{n}{2}\log|\boldsymbol{\Sigma}_z| \\ &\quad -\frac{1}{2}\sum_{j=1}^{n} q_j(\mathbf{z}_j-\boldsymbol{\mu}_z)^\top\boldsymbol{\Sigma}_z^{-1}(\mathbf{z}_j-\boldsymbol{\mu}_z)+\sum_{j=1}^{n}\log f_2(q_j),\end{aligned} \tag{2.17}$$

em que

$$|\boldsymbol{\Sigma}_z| = \phi_x|D(\boldsymbol{\phi})| \quad e \quad \boldsymbol{\Sigma}_z^{-1} = \begin{pmatrix} \frac{c}{\phi_x} & -\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi}) \\ -D^{-1}(\boldsymbol{\phi})\mathbf{1}_p & D^{-1}(\boldsymbol{\phi}) \end{pmatrix}.$$

Observe que

$$\begin{aligned}\| \boldsymbol{\Sigma}_z^{-1/2}(\mathbf{Z}_j-\boldsymbol{\mu}_z) \|^2 &= \begin{pmatrix} x_j-\mu_x \\ \mathbf{Y}_j-\boldsymbol{\mu} \end{pmatrix}^\top \begin{pmatrix} \frac{c}{\phi_x} & -\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi}) \\ -D^{-1}(\boldsymbol{\phi})\mathbf{1}_p & D^{-1}(\boldsymbol{\phi}) \end{pmatrix} \begin{pmatrix} x_j-\mu_x \\ \mathbf{Y}_j-\boldsymbol{\mu} \end{pmatrix} \\ &= c\frac{(x_j-\mu_x)^2}{\phi_x} - 2\Bigg[\frac{(x_j-\mu_x)(y_{1j}-\mu_x)}{\phi_1} \\ &\quad +(x_j-\mu_x)\mathbf{1}_q^\top D_1^{-1}(\boldsymbol{\phi})(\mathbf{Y}_j^*-\boldsymbol{\alpha}-\mathbf{1}_q\mu_x)\Bigg] \\ &\quad +\frac{(y_{1j}-\mu_x)^2}{\phi_1} + (\mathbf{Y}_j^*-\boldsymbol{\alpha}-\mathbf{1}_q\mu_x)^\top D_1^{-1}(\boldsymbol{\phi})(\mathbf{Y}_j^*-\boldsymbol{\alpha}-\mathbf{1}_q\mu_x),\end{aligned}$$

em que $D_1(\boldsymbol{\phi}) = Diag(\phi_2,\cdots,\phi_p)$ e $\mathbf{Y}_j^* = (y_{2j},\cdots,y_{pj})^\top$.

Etapa E: Calculamos o valor esperado:

$$\begin{aligned}
E(\ell(\boldsymbol{\theta}|\mathbf{Z})|\mathbf{Y},\boldsymbol{\theta}^{(m)}) \;=\;& cte - \frac{n}{2}\log\phi_x^{(m)} - \frac{n}{2}\log|D(\boldsymbol{\phi}^{(m)})| \\
& -\frac{1}{2}\sum_{j=1}^{n} q_j^{(m)}\Bigg\{c^{(m)}\frac{(x_j^{(m)}-\mu_x^{(m)})^2}{\phi_x^{(m)}} - 2\Bigg[\frac{(x_j^{(m)}-\mu_x^{(m)})(y_{1j}-\mu_x^{(m)})}{\phi_1^{(m)}} \\
& +(x_j^{(m)}-\mu_x^{(m)})\mathbf{1}_q^\top D_1^{-1}(\boldsymbol{\phi}^{(m)})(\mathbf{Y}_j^* - \boldsymbol{\alpha}^{(m)} - \mathbf{1}_q\mu_x^{(m))}\Bigg] \\
& +\frac{(y_{1j}-\mu_x^{(m)})^2}{\phi_1^{(m)}} \\
& +(\mathbf{Y}_j^* - \boldsymbol{\alpha}^{(m)} - \mathbf{1}_q\mu_x^{(m)})^\top D_1^{-1}(\boldsymbol{\phi}^{(m)})(\mathbf{Y}_j^* - \boldsymbol{\alpha}^{(m)} - \mathbf{1}_q\mu_x^{(m)})\Bigg\}
\end{aligned} \tag{2.18}$$

em que

$$q_j^{(m)} \;=\; E(q_j|\mathbf{Y}_j,\boldsymbol{\theta}^{(m)}) = (\nu+p)(\nu+||\mathbf{T}_j^{(m)}||^2)^{-1}, \tag{2.19}$$

$$x_j^{(m)} \;=\; E(x_j|\mathbf{Y}_j,\boldsymbol{\theta}^{(m)}) = \mu_x + \frac{\phi_x^{(m)}}{c^{(m)}}\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi}^{(m)})(\mathbf{Y}_j - \boldsymbol{\mu}^{(m)}), \tag{2.20}$$

$$x_j^{2(m)} \;=\; E(x_j^2|\mathbf{Y}_j,\boldsymbol{\theta}^{(m)}) = \frac{\phi_x^{(m)}}{c^{(m)}}\frac{\nu+||\mathbf{T}_j^{(m)}||^2}{\nu+p-2} + (x_j^{(m)})^2, \tag{2.21}$$

com $c^{(m)} = 1 + \phi_x^{(m)}\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi}^{(m)})\mathbf{1}_p$ e $\mathbf{T}_j^{(m)} = (\phi_x^{(m)}\mathbf{1}_p\mathbf{1}_p^\top + D(\phi^{(m)}))^{-1/2}(\mathbf{Y}_j - \boldsymbol{\mu}^{(m)})$.

As equações (2.19)-(2.21) seguem das propriedades da distribuição t-student (ver Apêndice B).

Etapa M: Nesta etapa determinamos $\boldsymbol{\theta}^{(m+1)}$, maximizando (2.18). Derivando com respeito a $\boldsymbol{\theta} = \boldsymbol{\theta}^{(m)}$, obtemos as seguintes estimativas:

$$\begin{aligned}
\mu_x^{(m+1)} &= \frac{\sum_{j=1}^n q_j^{(m)} x_j^{(m)}}{\sum_{j=1}^n q_j^{(m)}},\\
\boldsymbol{\alpha}^{(m+1)} &= \frac{\sum_{j=1}^n q_j^{(m)} \mathbf{Y}_j^*}{\sum_{j=1}^n q_j^{(m)}} - \mathbf{1}_q \mu_x^{(m+1)},\\
\phi_x^{(m+1)} &= \frac{\sum_{j=1}^n q_j^{(m)} (x_j^{(m)} - \mu_x^{(m+1)})^2}{n},\\
\phi_1^{(m+1)} &= \frac{\sum_{j=1}^n q_j^{(m)} (y_{1j} - x_j^{(m)})^2}{n},\\
\phi_i^{(m+1)} &= \frac{\sum_{j=1}^n q_j^{(m)} (y_{ij} - \alpha_i^{(m+1)} - x_j^{(m)})^2}{n}, \quad i = 2, \cdots, p.
\end{aligned}$$

Como estimetivas iniciais ($\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}$) podemos usar os estimadores de momentos propostos por Grubbs (veja Seção 2.7).

b) Os algoritmos de Scoring e Newton-Raphson

Os métodos iterativos do tipo Newton consistem de técnicas para a obtenção de raízes de uma função. Eles são métodos alternativos muito utilizados na obtenção dos EMV, resolvendo as equações de verossimilhança. Com o objetivo de fornecer técnicas para a obtenção dos EMV, para modelos cujas observações têm distribuições diferentes da normal ou t-Student, apresentaremos dois algoritmos que pertecem à classe dos métodos do tipo Newton. São eles: o algoritmo de *Scoring* e o algoritmo de *Newton-Raphson*.

No método de *Scoring*, dada a estimativa $\hat{\boldsymbol{\theta}}^{(k)}$ correspondente a k-ésima iteração e definindo $U = U(\boldsymbol{\theta}) = \frac{\partial \ell}{\partial \boldsymbol{\theta}}$, temos que

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \lambda_k (K_F^{(k)})^{-1} U^{(k)} \tag{2.22}$$

em que $K_F^{(k)} = K_F(\boldsymbol{\theta}^{(k)})$ (a matriz de informação de Fisher esperada com $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$) , $U^{(k)} = U(\boldsymbol{\theta}^{(k)})$ e $\lambda_k > 0$ é o comprimento do passo k (usualmente escolhe-se $\lambda_k = 1$). Uma limitação deste algoritmo, sob o modelo elíptico, é a dependência da matriz de informação esperada, pois nem sempre é possível obter de forma explícita os elementos dessa matriz devido dificuldade de calcular os valores b_f do Teorema 2.2.

Um modo de contornar esta limitação é utilizar o algoritmo de *Newton-Raphson*. Este algoritmo é dado por (2.22) substituindo a matriz $K_F^{(k)}$ pela matriz $I_F^{(k)} = I_F(\boldsymbol{\theta}^{(k)})$, com $I_F(\boldsymbol{\theta}) = -\frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$. Quando a log-verossimilhança é uma função côncava e unimodal, então a sequência de iterações $[\boldsymbol{\theta}^{(k)}]$ converge para o EMV de $\boldsymbol{\theta}$. Quando estas condições não são satisfeitas, não se pode garantir a convergência a partir de um valor inicial arbitrário (veja McLachlan e Krishnan, 1993).

2.5 Estimação de Máxima Verossimilhança Restrita

Nesta seção, apresentamos expressões para os EMV quando alguma informação adicional é incluída no modelo de Grubbs. Para tal finalidade, usaremos o modelo t-elíptico definido em (2.12)-(2.13) por motivos já mencionados na Subseção 2.4.3. Consideramos aqui os seguintes casos:

H_{01}. Os vícios são todos iguais a zero e as variâncias dos instrumentos são iguais, isto é, $\alpha_2 = \cdots = \alpha_p = 0$ e $\phi_1 = \cdots = \phi_p$;

H_{02}. As variâncias dos instrumentos são iguais, isto é, $\phi_1 = \cdots = \phi_p$ e

H_{03}. Os vícios dos instrumentos são todos iguais a zero, isto é, $\alpha_1 = \cdots = \alpha_p = 0$.

Devido a complexidade das equações de verossimilhança, os EMV são obtidos por meio do algoritmo EM em todos os casos. Estes casos serão de grande utilidade na seção de teste de hipóteses, quando necessitaremos dos EMV sob as hipóteses nulas H_{01}, H_{02} e H_{03} definidas na Subseção 1.1.

Como iremos notar, em todos os casos as estimativas das variâncias são não negativas. Isto não ocorre com o estimador de momentos, conforme veremos na Seção 2.7. Segundo Jaech (1985), quando uma das variâncias é estimada como sendo zero (perto ou menor que zero), a conclusão pode ser que o correspondente instrumento meça a característica de interesse sem erro ou de forma precisa.

2.5.1 O caso $\boldsymbol{\alpha_2 = \cdots = \alpha_p = 0}$ e $\boldsymbol{\phi_1 = \cdots = \phi_p}$

Nesta Subseção, vamos considerar a possibilidade de que os instrumentos meçam a característica de interesse sem vício e com a mesma variabilidade, as quais consideramos como sendo iguais a ϕ, isto é, $\alpha_2 = \cdots = \alpha_p = 0$ e $\phi_1 = \cdots = \phi_p = \phi$. Logo, segue que

$$\mathbf{Y}_j \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu), j = 1, \cdots, n, \tag{2.23}$$

em que $\boldsymbol{\mu} = \mu_x \mathbf{1}_p$ e $\boldsymbol{\Sigma} = \phi_x \mathbf{1}_p \mathbf{1}_p^\top + \phi \mathbf{I}_p$.

Neste caso

$$\boldsymbol{\theta} = (\mu_x, \phi_x, \phi)^\top,$$

com $||\mathbf{T}_j||^2 = \mathbf{W}_j^\top \boldsymbol{\Sigma}^{-1} \mathbf{W}_j$, $\mathbf{W}_j = \mathbf{y}_j - \boldsymbol{\mu}$, $\boldsymbol{\Sigma}^{-1} = \frac{1}{\phi}\mathbf{I}_p - c^{-1}\mathbf{M}$, $c = 1 + p\frac{\phi_x}{\phi}$ e $\mathbf{M} = \frac{\phi_x}{\phi^2}\mathbf{1}_p\mathbf{1}_p^\top$.

Analogamente à Subseção 2.4.3, seja $\mathbf{Z}_j = \begin{pmatrix} x_j \\ \mathbf{Y}_j \end{pmatrix}$, $j = 1, \cdots, n$. Então, de (2.23) temos que

$$\mathbf{Z}_j \overset{iid}{\sim} t_{p+1}(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z; \nu), \quad j = 1, \cdots, n, \tag{2.24}$$

em que

$$\boldsymbol{\mu}_z = \begin{pmatrix} \mu_x \\ \mu_x \mathbf{1}_p \end{pmatrix} \quad e \quad \boldsymbol{\Sigma}_z = \begin{pmatrix} \phi_x & \mathbf{1}_p^\top \phi_x \\ \mathbf{1}_p \phi_x & \phi_x \mathbf{1}_p \mathbf{1}_p^\top + \phi \mathbf{I}_p) \end{pmatrix}.$$

Ainda

$$\mathbf{Z}_j|(Q_j = q_j) \sim N_{p+1}(\boldsymbol{\mu}_z, q_j^{-1}\boldsymbol{\Sigma}_z), Q_j \sim \chi^2_{(\nu)}/\nu, \quad j = 1, \cdots, n \tag{2.25}$$

e

$$f_{\mathbf{Z}_j, Q_j}(\mathbf{z}_j, q_j) = f_1(\mathbf{z}_j | q_j) f_2(q_j), \quad j = 1, \cdots, n,$$

em que f_1 é a função de densidade condicional de $\mathbf{Z}_j$ dado $Q_j = q_j$ e f_2 é a função densidade de Q_j, $j = 1, \cdots, n$. Portanto, a log-verossimilhança completa, $\ell_{\mathbf{Z}}(\boldsymbol{\theta}, \nu)$ é dada por

$$\begin{aligned}\ell_{\mathbf{Z}}(\boldsymbol{\theta},\nu) &= cte - \frac{n}{2}\log\phi_x - \frac{np}{2}\log\phi \\ &\quad -\frac{1}{2}\sum_{j=1}^{n} q_j(\mathbf{z}_j-\boldsymbol{\mu}_z)^\top\boldsymbol{\Sigma}_z^{-1}(\mathbf{z}_j-\boldsymbol{\mu}_z) + \sum_{j=1}^{n}\log f_2(q_j),\end{aligned} \tag{2.26}$$

em que

$$\boldsymbol{\Sigma}_z^{-1} = \begin{pmatrix} \frac{c}{\phi_x} & -\frac{1}{\phi}\mathbf{1}_p^\top \\ -\frac{1}{\phi}\mathbf{1}_p & \frac{1}{\phi}\mathbf{I}_p \end{pmatrix}.$$

Logo o algoritmo EM procede com as etapas:

Etapa E: Calculamos

$$q_j^{(m)} = E(q_j|\mathbf{Y}_j,\boldsymbol{\theta}^{(m)}) = (\nu+p)(\nu+||\mathbf{T}_j^{(m)}||^2)^{-1}, \tag{2.27}$$

$$x_j^{(m)} = E(x_j|\mathbf{Y}_j,\boldsymbol{\theta}^{(m)}) = \mu_x + \frac{\phi_x^{(m)}}{c^{(m)}\phi^{(m)}}\mathbf{1}_p^\top(\mathbf{Y}_j-\boldsymbol{\mu}^{(m)}), \tag{2.28}$$

$$x_j^{2(m)} = E(x_j^2|\mathbf{Y}_j,\boldsymbol{\theta}^{(m)}) = \frac{\phi_x^{(m)}}{c^{(m)}}\frac{\nu+||\mathbf{T}_j^{(m)}||^2}{\nu+p-2} + (x_j^{(m)})^2, \tag{2.29}$$

com $c^{(m)} = 1 + p\frac{\phi_x^{(m)}}{\phi^{(m)}}$ e $\mathbf{T}_j^{(m)} = (\phi_x^{(m)}\mathbf{1}_p\mathbf{1}_p^\top + \phi^{(m)}\mathbf{I}_p)^{-1/2}(\mathbf{Y}_j-\boldsymbol{\mu}^{(m)})$.

Etapa M: Determinamos $\boldsymbol{\theta}^{(m+1)}$, a partir de (2.27)-(2.29) de acordo com

$$
\begin{aligned}
\mu_x^{(m+1)} &= \frac{\sum_{j=1}^{n} q_j^{(m)} x_j^{(m)}}{\sum_{j=1}^{n} q_j^{(m)}}, \\
\phi_x^{(m+1)} &= \frac{\sum_{j=1}^{n} q_j^{(m)} (x_j^{(m)} - \mu_x^{(m+1)})^2}{n}, \\
\phi^{(m+1)} &= \frac{1}{np} \sum_{j=1}^{n} q_j^{(m)} \left[\left((x_j^{(m)} - \mu_x^{(m+1)}) \mathbf{1}_p - (\mathbf{Y}_j - \boldsymbol{\mu}^{(m+1)}) \right)^\top \right. \\
&\times \left. \left((x_j^{(m)} - \mu_x^{(m+1)}) \mathbf{1}_p - (\mathbf{Y}_j - \boldsymbol{\mu}^{(m+1)}) \right) \right],
\end{aligned}
$$

em que $\boldsymbol{\mu}^{(m+1)} = \boldsymbol{\mu}^{(m+1)}(\boldsymbol{\theta}) = \boldsymbol{\mu}(\boldsymbol{\theta}^{(m+1)})$.

2.5.2 O caso $\phi_1 = \cdots = \phi_p$

Vamos considerar que as variâncias são iguais a ϕ, isto é

$$\phi_1 = \cdots = \phi_p = \phi.$$

Então, a distribuição dos $\mathbf{Y}_j$ é dada por

$$\mathbf{Y}_j \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu), \tag{2.30}$$

em que $\boldsymbol{\mu} = \mathbf{a} + \mu_x \mathbf{1}_p$ e $\boldsymbol{\Sigma} = \phi_x \mathbf{1}_p \mathbf{1}_p^\top + \phi \mathbf{I}_p$, com $\mathbf{a}$ como em (2.10). Neste caso

$$\boldsymbol{\theta} = (\mu_x, \boldsymbol{\alpha}^\top, \phi_x, \phi)^\top,$$

com $||\mathbf{T}_j||^2 = \mathbf{W}_j^\top \boldsymbol{\Sigma}^{-1} \mathbf{W}_j$, $\mathbf{W}_j = \mathbf{y}_j - \boldsymbol{\mu}$ e $\boldsymbol{\Sigma}^{-1}$, c e $\mathbf{M}$ iguais aos da subseção anterior.

Com o objetivo de não sermos demasiadamente repetitivos, apresentaremos o algoritmo omitindo algumas passagens. Entretanto, gostaríamos de ressaltar que estas passagens são análogas as da Seção 2.4.3. Assim, segue o algoritmo EM:

Etapa E: Calculamos

$$q_j^{(m)} = E(q_j|\mathbf{Y}_j, \boldsymbol{\theta}^{(m)}) = (\nu + p)(\nu + ||\mathbf{T}_j^{(m)}||^2)^{-1}, \tag{2.31}$$

$$x_j^{(m)} = E(x_j|\mathbf{Y}_j, \boldsymbol{\theta}^{(m)}) = \mu_x + \frac{\phi_x^{(m)}}{c^{(m)}\phi^{(m)}}\mathbf{1}_p^\top(\mathbf{Y}_j - \boldsymbol{\mu}^{(m)}), \tag{2.32}$$

$$x_j^{2(m)} = E(x_j^2|\mathbf{Y}_j, \boldsymbol{\theta}^{(m)}) = \frac{\phi_x^{(m)}}{c^{(m)}}\frac{\nu + ||\mathbf{T}_j^{(m)}||^2}{\nu + p - 2} + (x_j^{(m)})^2, \tag{2.33}$$

com $c^{(m)} = 1 + p\frac{\phi_x^{(m)}}{\phi^{(m)}}$ e $\mathbf{T}_j^{(m)} = (\phi_x^{(m)}\mathbf{1}_p\mathbf{1}_p^\top + \phi^{(m)}\mathbf{I}_p)^{-1/2}(\mathbf{Y}_j - \boldsymbol{\mu}^{(m)})$.

Etapa M: A partir de (2.31)-(2.33), obtemos $\boldsymbol{\theta}^{(m+1)}$ pelas expressões

$$\begin{aligned}
\mu_x^{(m+1)} &= \frac{\sum_{j=1}^n q_j^{(m)}x_j^{(m)}}{\sum_{j=1}^n q_j^{(m)}}, \\
\boldsymbol{\alpha}^{(m+1)} &= \frac{\sum_{j=1}^n q_j^{(m)}\mathbf{Y}_j^*}{\sum_{j=1}^n q_j^{(m)}} - \mathbf{1}_q\mu_x^{(m+1)}, \\
\phi_x^{(m+1)} &= \frac{\sum_{j=1}^n q_j^{(m)}(x_j^{(m)} - \mu_x^{(m+1)})^2}{n}, \\
\phi^{(m+1)} &= \frac{1}{np}\sum_{j=1}^n q_j^{(m)}\left[\left((x_j^{(m)} - \mu_x^{(m+1)})\mathbf{1}_p - (\mathbf{Y}_j - \boldsymbol{\mu}^{(m+1)})\right)^\top\right. \\
&\times \left.\left((x_j^{(m)} - \mu_x^{(m+1)})\mathbf{1}_p - (\mathbf{Y}_j - \boldsymbol{\mu}^{(m+1)})\right)\right],
\end{aligned}$$

em que $\boldsymbol{\mu}^{(m+1)} = \boldsymbol{\mu}^{(m+1)}(\boldsymbol{\theta}) = \boldsymbol{\mu}(\boldsymbol{\theta}^{(m+1)})$.

2.5.3 O caso $\boldsymbol{\alpha_2 = \cdots = \alpha_p = 0}$

Considerando a possibilidade de que os diferentes instrumentos medem a característica de interesse sem vício, isto é, $\alpha_2 = \cdots = \alpha_p = 0$, segue que

$$\mathbf{Y}_j \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu), \tag{2.34}$$

em que

$$\boldsymbol{\mu} = \mu_x \mathbf{1}_p \text{ e } \boldsymbol{\Sigma} = \phi_x \mathbf{1}_p \mathbf{1}_p^\top + D(\phi).$$

Neste caso

$$\boldsymbol{\theta} = (\mu_x, \phi_x, \boldsymbol{\phi}^\top)^\top.$$

De maneira similar a discutida na Subseção 2.4.3, temos que o algoritmo EM procede como segue

Etapa E: Calculamos

$$\begin{aligned}
q_j^{(m)} &= E(q_j|\mathbf{Y}_j, \boldsymbol{\theta}^{(m)}) = (\nu + p)(\nu + ||\mathbf{T}_j^{(m)}||^2)^{-1}, && (2.35)\\
x_j^{(m)} &= E(x_j|\mathbf{Y}_j, \boldsymbol{\theta}^{(m)}) = \mu_x + \frac{\phi_x^{(m)}}{c^{(m)}} \mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi}^{(m)})(\mathbf{Y}_j - \boldsymbol{\mu}^{(m)}), && (2.36)\\
x_j^{2(m)} &= E(x_j^2|\mathbf{Y}_j, \boldsymbol{\theta}^{(m)}) = \frac{\phi_x^{(m)}}{c^{(m)}} \frac{\nu + ||\mathbf{T}_j^{(m)}||^2}{\nu + p - 2} + (x_j^{(m)})^2, && (2.37)
\end{aligned}$$

com $c^{(m)} = 1 + \phi_x^{(m)} \mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi}^{(m)}) \mathbf{1}_p$ e $\mathbf{T}_j^{(m)} = (\phi_x^{(m)} \mathbf{1}_p \mathbf{1}_p^\top + D(\phi^{(m)}))^{-1/2} (\mathbf{Y}_j - \boldsymbol{\mu}^{(m)})$.

Etapa M: Nesta etapa determinamos $\boldsymbol{\theta}^{(m+1)}$ a partir de (2.35)-(2.37)

$$\begin{aligned}
\mu_x^{(m+1)} &= \frac{\sum_{j=1}^n q_j^{(m)} x_j^{(m)}}{\sum_{j=1}^n q_j^{(m)}},\\
\phi_x^{(m+1)} &= \frac{\sum_{j=1}^n q_j^{(m)} (x_j^{(m)} - \mu_x^{(m+1)})^2}{n},\\
\phi_i^{(m+1)} &= \frac{\sum_{j=1}^n q_j^{(m)} (y_{ij} - x_j^{(m)})^2}{n}, \quad i = 1, \cdots, p.
\end{aligned}$$

2.6 Testes de Hipóteses

Nesta seção vamos apresentar um estudo de testes de hipóteses para avaliar as hipóteses definidas na Seção 1.1, para o modelo elíptico independente definido em (2.12)-(2.13), utilizando as estatísticas da Razão de Verosimilhança (Q), Escore (E) e de Wald (W).

Sejam $\hat{\boldsymbol{\theta}}$ e $\tilde{\boldsymbol{\theta}}$ os EMV de $\boldsymbol{\theta}$ sob o modelo irrestrito e sob a hipótese nula, respectivamente. As hipóteses definidas na Seção 1.1, podem ser escritas como $H_0 : \mathbf{A}\boldsymbol{\theta} = \boldsymbol{\gamma}_0$, onde a matriz $\mathbf{A}$ de dimensão $r \times (2p+1)$ tem posto $posto(\mathbf{A}) = r \leq 2p+1$ e $\boldsymbol{\gamma}_0 : r \times 1$ é um vetor conhecido. Logo as estatísticas Q, E e W podem ser escritas como

$$Q = 2[\ell(\hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\theta}})], \ E = \frac{1}{n}\left[\frac{\partial \ell(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right]^{\top} [\mathbf{K}_F^{-1}(\tilde{\boldsymbol{\theta}})] \left[\frac{\partial \ell(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right]$$

e

$$W = n[\mathbf{A}\hat{\boldsymbol{\theta}} - \boldsymbol{\gamma}_0]^{\top}[\mathbf{A}\mathbf{K}_F^{-1}(\hat{\boldsymbol{\theta}})\mathbf{A}^{\top}]^{-1}[\mathbf{A}\hat{\boldsymbol{\theta}} - \boldsymbol{\gamma}_0],$$

em que $\ell(\boldsymbol{\theta})$ é a função de log-verossimilhança definida em (2.11) e $\mathbf{K}_F$ é a matriz de informação de Fisher esperada definida no Teorema 2.2.

As três estatísticas definidas acima são assintoticamente equivalentes sob H_0. De fato as distribuições delas convergem para a distribuição qui-quadrado com r graus de liberdade ($\mathcal{X}_r^2$). Isto é, sob H_0, Q, E e W têm a mesma distribuição assintótica; veja por exemplo, Sen e Singer (1993).

2.6.1 Teste de Wald

Seja $\boldsymbol{\theta} = (\boldsymbol{\theta}_L^{\top}, \boldsymbol{\theta}_E^{\top})^{\top}$, onde $\boldsymbol{\theta}_L = (\mu_x, \alpha_2, \cdots, \alpha_p)^{\top}$ e $\boldsymbol{\theta}_E = (\phi_x, \phi_1, \cdots, \phi_p)^{\top}$. Então a matriz de informação de Fisher esperada, dada no Teorema 2.1, pode ser representada por $\mathbf{K}_F(\boldsymbol{\theta}) = \mathrm{Diag}(\mathbf{K}_L, \mathbf{K}_E)$, onde

$$\mathbf{K}_L = \begin{pmatrix} K_{\mu_x\mu_x} & K_{\mu_x\alpha} \\ K_{\alpha\mu_x} & K_{\alpha\alpha} \end{pmatrix} \quad e \quad \mathbf{K}_E = \begin{pmatrix} K_{\phi_x\phi_x} & K_{\phi_x\phi} \\ K_{\phi\phi_x} & K_{\phi\phi} \end{pmatrix}$$

são as submatrizes de informação dos vetores na locação $\boldsymbol{\theta}_L = (\mu_x, \boldsymbol{\alpha}^\top)^\top$ e na escala $\boldsymbol{\theta}_E = (\phi_x, \boldsymbol{\phi}^\top)^\top$, respectivamente.

Agora, se $\widehat{\boldsymbol{\theta}}$ é o EMV de $\boldsymbol{\theta}$, então utilizando o Teorema 5.2.1 dado em Sen e Singer (1993), temos que

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N_{2p+1}(\mathbf{0}, \mathbf{K}_F^{-1}(\boldsymbol{\theta})), \tag{2.38}$$

em que $\mathbf{K}_F^{-1}(\boldsymbol{\theta}) = \text{Diag}(\mathbf{K}_L^{-1}, \mathbf{K}_E^{-1})$.

Por outro lado, se $\boldsymbol{\theta}_* = g(\boldsymbol{\theta}) = (\alpha_2, \cdots, \alpha_p, \phi_1, \cdots, \phi_p)^\top = (g_1(\boldsymbol{\theta}),\ g_2(\boldsymbol{\theta}),\ \cdots,\ g_{2p-1}(\boldsymbol{\theta}))^\top$, então usando o método delta temos que

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_* - \boldsymbol{\theta}_*) \xrightarrow{d} N_{2p-1}(\mathbf{0}, \boldsymbol{\Omega}_*), \tag{2.39}$$

em que $\boldsymbol{\Omega}_* = \mathbf{D}\mathbf{K}_F^{-1}\mathbf{D}^\top$, com $\mathbf{K}_F$ a matriz de informação esperada, definida no Teorema 2.2 e $\mathbf{D}$ é uma matriz $(2p-1) \times (2p+1)$, tal que

$$\mathbf{D} = \left[\frac{\partial g_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j}\right] = \begin{pmatrix} \mathbb{I}_{(p)} & \mathbf{0} \\ \mathbf{0} & \mathbb{I}_{(p+1)} \end{pmatrix},$$

assim de (2.38) e (2.39), temos que

$$\boldsymbol{\Omega}_* = \begin{pmatrix} \mathbb{I}_{(p)}\mathbf{K}_L^{-1}\mathbb{I}_{(p)}^\top & \mathbf{0} \\ \mathbf{0} & \mathbb{I}_{(p+1)}\mathbf{K}_E^{-1}\mathbb{I}_{(p+1)}^\top \end{pmatrix}, \tag{2.40}$$

em que $\mathbb{I}_{(r)}, r = p, p+1$, como definido em (2.3).

Para testar se os instrumentos medem a característica de interesse sem vício e com a mesma precisão, a hipótese é

$$H_{01} : \alpha_2 = \alpha_3 = \cdots = \alpha_p = 0 \ \ e \ \ \phi_1 = \phi_2 = \cdots = \phi_p. \tag{2.41}$$

Note que, H_{01} pode ser escrita como

$$H_{01} : \mathbf{C}\boldsymbol{\theta}_* = \mathbf{q}_o, \tag{2.42}$$

em que $\mathbf{C} = \begin{pmatrix} \mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_1 \end{pmatrix}$ é uma matriz $(2p-2) \times (2p-1)$ de posto $2p-2$, com $q = p-1$

e

$$\mathbf{A}_1 = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix}, \tag{2.43}$$

matriz $(p-1) \times p$ e $\mathbf{q}_o$ vetor nulo de dimensão $2(p-1)$. Logo, a estatística de Wald para testar a hipótese H_{01} é dada por

$$W_{01} = n(\mathbf{C}\widehat{\boldsymbol{\theta}}_* - \mathbf{q}_o)^\top [\mathbf{C}\widehat{\boldsymbol{\Omega}}_*\mathbf{C}^\top]^{-1}(\mathbf{C}\widehat{\boldsymbol{\theta}}_* - \mathbf{q}_o) \tag{2.44}$$

onde $\widehat{\boldsymbol{\Omega}}_*$ denota a matriz $\boldsymbol{\Omega}_*$ com os parâmetros desconhecidos $\boldsymbol{\theta}_*$, substituídos pelo EMV $\widehat{\boldsymbol{\theta}}_*$ de $\boldsymbol{\theta}_*$. Portanto, H_{01} é rejeitada ao nível α se $W_{01} > \chi^2_{1-\alpha}$, onde $\chi^2_{1-\alpha}$ denota o percentil $100(1-\alpha)\%$ da distribuição qui-quadrado com $2(p-1)$ graus de liberdade.

Observe que de (2.40) e após as manipulações algébricas, $\mathbf{C}\boldsymbol{\Omega}_*\mathbf{C}^\top$ pode ser escrita por

$$\mathbf{C}\boldsymbol{\Omega}_*\mathbf{C}^\top = \begin{pmatrix} \mathbb{I}_{(p)}\mathbf{K}_L^{-1}\mathbb{I}_{(p)}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_1\mathbb{I}_{(p+1)}\mathbf{K}_E^{-1}\mathbb{I}_{(p+1)}^\top\mathbf{A}_1^\top \end{pmatrix},$$

e conseqüentemente a estatística W_{01} pode ser escrita como

$$W_{01} = n\widehat{\boldsymbol{\alpha}}^\top(\mathbb{I}_{(p)}\widehat{\mathbf{K}}_L^{-1}\mathbb{I}_{(p)}^\top)^{-1}\widehat{\boldsymbol{\alpha}} + n\widehat{\boldsymbol{\phi}}^\top\mathbf{A}_1^\top(\mathbf{A}_1\mathbb{I}_{(p+1)}\widehat{\mathbf{K}}_E^{-1}\mathbb{I}_{(p+1)}^\top\mathbf{A}_1^\top)^{-1}\mathbf{A}_1\widehat{\boldsymbol{\phi}}. \tag{2.45}$$

Para testar a hipótese de interesse

$$H_{02} : \alpha_2 = \alpha_3 = \cdots = \alpha_p = 0, \tag{2.46}$$

isto é, para testar se os instrumentos medem sem vício a característica de interesse x, vamos escrever H_{02} como

$$H_{02} : \mathbf{C}\boldsymbol{\theta}_* = \mathbf{0}_q, \tag{2.47}$$

em que $\mathbf{C} = [\mathbf{I}_q\mathbf{0}]$ é uma matriz $(p-1) \times (2p-1)$ de posto $q = p-1$. Assim, a estatística de Wald para testar H_{02}, denotada por W_{02}, é dada por

$$\begin{aligned} W_{02} &= n(\mathbf{C}\widehat{\boldsymbol{\theta}}_*)^\top[\mathbf{C}\widehat{\boldsymbol{\Omega}}_*\mathbf{C}^\top]^{-1}(\mathbf{C}\widehat{\boldsymbol{\theta}}_*) \\ &= n\widehat{\boldsymbol{\alpha}}^\top(\mathbb{I}_{(p)}\widehat{\mathbf{K}}_L^{-1}\mathbb{I}_{(p)}^\top)^{-1}\widehat{\boldsymbol{\alpha}}, \end{aligned} \tag{2.48}$$

que converge em distribuição para uma variável aleatória $\chi^2_{(p-1)}$ sob H_{02}. Portanto, H_{02} é rejeitado ao nível α se $W_{02} > \chi^2_{1-\alpha}$, onde $\chi^2_{1-\alpha}$ denota o percentil $100(1-\alpha)\%$ da distribuição qui-quadrado com $p-1$ graus de liberdade.

Finalmente, para testar se os instrumentos são igualmente precisos, a hipótese é

$$H_{03} : \phi_1 = \phi_2 = \cdots = \phi_p \quad ou \quad \mathbf{C}\boldsymbol{\theta}_* = \mathbf{0}_q, \tag{2.49}$$

em que $\mathbf{C} = [\mathbf{0}_q|\mathbf{A}_1]$ é uma matriz $(p-1)\times(2p-1)$ de posto $q = p-1$, com $\mathbf{A}_1$ como em (2.43). A estatística de Wald para testar H_{03}, denotada por W_{03} é dada por

$$\begin{aligned} W_{03} &= n(\mathbf{C}\widehat{\boldsymbol{\theta}}_*)^\top[\mathbf{C}\widehat{\boldsymbol{\Omega}}_*\mathbf{C}^\top]^{-1}(\mathbf{C}\widehat{\boldsymbol{\theta}}_*) \\ &= n\widehat{\boldsymbol{\phi}}^\top\mathbf{A}_1^\top(\mathbf{A}_1\mathbb{I}_{(p+1)}\widehat{\mathbf{K}}_E^{-1}\mathbb{I}_{(p+1)}^\top\mathbf{A}_1^\top)^{-1}\mathbf{A}_1\widehat{\boldsymbol{\phi}}, \end{aligned} \tag{2.50}$$

que converge em distribuição para uma variável aleatória $\chi^2_{(p-1)}$ sob H_{03}. Portanto, H_{03} é rejeitada ao nível α se $W_{03} > \chi^2_{1-\alpha}$, onde $\chi^2_{1-\alpha}$ denota o percentil $100(1-\alpha)\%$ da distribuição qui-quadrado com $p-1$ graus de liberdade.

É importante observar que das relações em (2.45), (2.48) e (2.50), segue que

$$W_{01} = W_{02} + W_{03}, \tag{2.51}$$

onde W_{02} e W_{03} são as estatísticas de Wald utilizadas para testar H_{02} e H_{03}, respectivamente.

2.6.2 Teste de Escore

Rao (1948) introduz a estatística de Escore dada por

$$E = \frac{1}{n}\left[\frac{\partial\ell(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\right]^\top[\mathbf{K}_F^{-1}(\tilde{\boldsymbol{\theta}})]\left[\frac{\partial\ell(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\right], \tag{2.52}$$

em que $\ell(\boldsymbol{\theta})$ é a função de log-verossimilhança definida em (2.11), $\tilde{\boldsymbol{\theta}}$ é o EMV sob a hipótese nula e $\mathbf{K}_F(\tilde{\boldsymbol{\theta}})$ é a matriz de informação de Fisher esperada definida no Teorema 2.2 avaliada no EMV sob a hipótese nula. Esta estatística é atrativa porque somente requer calcular os EMV sob hipótese nula, e é assintoticamente equivalente às estatísticas de Wald e da Razão de Verossimilhanças sob a hipótese nula.

A estatística de Escore tem sido utilizada no trabalho desenvolvido por Bedrick (2001), que apresenta um estudo de inferência no modelo de Grubbs supondo normalidade das observações. Sob normalidade, os EMV sob H_{01} e H_{02} têm forma fechada, assim a estatística de Escore também possui uma expressão fechada (veja Lachos-Dávila, 2002).

Dado que no modelo t-elíptico o EMV não possui forma explícita sob as hipóteses nulas, este pode ser obtida via um procedimento interativo (algoritmo EM). A implementação pode ser facilmente manipulada por meio de "softwares"como por exemplo Matlab. O algoritmo para obter o EMV sob as hipóteses nulas H_{01}, H_{02} e H_{03} são dados na Seção 2.5.

2.6.3 Teste da Razão de Verossimilhanças

A estatística da Razão de verossimilhaças é definida como sendo

$$Q = 2[\ell(\hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\theta}})], \tag{2.53}$$

em que $\ell(\boldsymbol{\theta})$ é a função de log-verossimilhança definida em (2.11), $\hat{\boldsymbol{\theta}}$ é o EMV do modelo irrestrito, e $\tilde{\boldsymbol{\theta}}$, é o EMV sob a hipótese nula que tem a forma $\mathbf{A}\boldsymbol{\theta}_* = \mathbf{q}_0$, em que $\boldsymbol{\theta}_*$ é como em (2.39) e $\mathbf{A}$ uma matriz de ordem $r \times (2p+1)$, com posto r.

Neste caso requeremos os EMV de $\boldsymbol{\theta}$ sob o modelo irrestrito e sob a hipótese nula , cujo método de implementação para obter $\hat{\boldsymbol{\theta}}$ e $\tilde{\boldsymbol{\theta}}$ são apresentados nas Seções 2.4 e 2.5, respectivamente.

Embora não seja possível obter uma expressão fechada da estatística Q, sua implementação é computacionalmente simples, dado que os EMV do modelo irrestrito e sob as hipóteses nulas H_{01}, H_{02} e H_{03} são de fácil implementação.

2.7 Estudo Assintótico dos Estimadores de Momentos

Considere a média amostral $\bar{\mathbf{Y}}$ e a matriz de covariância amostral $\mathbf{S}_n$ da sequência de observações $\mathbf{Y}_1, \cdots, \mathbf{Y}_n$ definidos em (2.10), isto é

$$\bar{\mathbf{Y}}_n = \frac{1}{n}\sum_{j=1}^{n}\mathbf{Y}_j \text{ e } \mathbf{S}_n = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{Y}_j - \bar{\mathbf{Y}})(\mathbf{Y}_j - \bar{\mathbf{Y}})^\top. \tag{2.54}$$

Sob o modelo estrutural normal, Grubbs (1948), considera os seguintes estimadores tipo momentos para o vetor de parâmetros $\boldsymbol{\theta}$ dados por $\tilde{\boldsymbol{\theta}} = (\tilde{\mu_x}, \tilde{\boldsymbol{\alpha}}^\top, \tilde{\phi_x}, \tilde{\boldsymbol{\phi}}^\top)^\top = (\tilde{\mu_x}, \tilde{\alpha_2}, \cdots, \tilde{\alpha_p}, \tilde{\phi_x}, \tilde{\phi_1}, \cdots, \tilde{\phi_p})$, em que

$$\begin{aligned}
\tilde{\mu_x} &= \frac{1}{n}\sum_{j=1}^{n}\mathbf{Y}_{1j}, \\
\tilde{\alpha_i} &= \frac{1}{n}\sum_{j=1}^{n}\mathbf{Y}_{ij} - \tilde{\mu_x},\ i = 2, \cdots, p \\
\tilde{\phi_x} &= \frac{2}{p(p-1)}\sum_{i=1}^{p}\sum_{j>i}^{n} S_{ij}, \\
\tilde{\phi_i} &= S_{ii} - \phi_x,
\end{aligned} \tag{2.55}$$

em que $i = 1, \cdots, p$ e S_{ij}, $i, j = 1, \cdots, p$, são elementos da matriz de covariância amostral $\mathbf{S}_n$.

Note que as estimativas $\tilde{\phi}$ e $\tilde{\phi_x}$ têm o inconveniente de poder assumir valores negativos.

Lachos-Dávila (2002) faz um estudo do comportamento assintótico de $\tilde{\theta}$ supondo normalidade das observações. Nesta seção, generalizaremos este estudo supondo distribuição elíptica para observações.

Para tal estudo, primeiro apresentaremos alguns resultados do comportamento assintótico de $\bar{\mathbf{Y}}_n$ e $\mathbf{S}_n$. As seguintes notações serão necessárias para uma explicação mais detalhada desta parte do trabalho. Sejam $\mathbf{A} = (a_{ij})$ e $\mathbf{B} = (b_{ij})$ matrizes de ordem $r \times s$

e $p \times q$, respectivamente. Então

1. $\mathbf{e}_i(j)$ denota o vetor de dimensão $i \times 1$ com um na j-ésima posição e zeros nos outros casos;

2. Vec($\mathbf{A}$) denota o vetor de dimensão $rs \times 1$ dado por $\text{Vec}(\mathbf{A}) = (\mathbf{A}_{1.}^\top, \cdots, \mathbf{A}_{s.}^\top)^\top$, em que $\mathbf{A}_{i.}$ denota a i-ésima coluna de $\mathbf{A}$;

3. $\mathbf{A} \otimes \mathbf{B}$ é uma matriz de ordem $rp \times sq$, que denota o produto de Kronecker de $\mathbf{A}$ e $\mathbf{B}$, isto é, $\mathbf{A} \otimes \mathbf{B} = (a_{ij}\mathbf{B})$;

4. Existe uma matriz $\mathbf{K}_{pq}$, chamada de matriz de permutação (veja Henderson e Searle, 1979), tal que

$$\text{Vec}(\mathbf{A}^\top) = \mathbf{K}_{pq}\text{Vec}(\mathbf{A}). \tag{2.56}$$

Quando $p = q$, escrevemos $\mathbf{K}_p$ em lugar de $\mathbf{K}_{pp}$. Além disso, se $\mathbf{A} = \mathbf{A}^\top$ então Vech($\mathbf{A}$) contém os $p(p+1)/2$ elementos distintos de $\mathbf{A}$ e, Vec($\mathbf{A}$) e $Vech(\mathbf{A}^\top)$ são são uma tranformação linear um do outro. Assim existe uma matriz $\mathbf{H}$ de ordem $p(p+1)/2 \times p^2$ tal que

$$\text{Vech}(\mathbf{A}^\top) = \mathbf{H}\text{Vec}(\mathbf{A}). \tag{2.57}$$

Para obter mais detalhes sobre o lema a seguir veja Gasco (1997).

Lema 2.1. *Suponha que $\boldsymbol{\varepsilon} \sim El_p(\mathbf{0}, \boldsymbol{\Sigma}; \phi)$ tem momentos de quarta ordem finitos. Então $E[\boldsymbol{\varepsilon}] = \mathbf{0}$, $Var[\boldsymbol{\varepsilon}] = \boldsymbol{\Psi} = 2\phi'(0)\boldsymbol{\Sigma}$ e*

$$Var[Vec(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top)] = (1+\kappa)(I_{p^2} + \mathbf{K}_p)(\boldsymbol{\Psi} \otimes \boldsymbol{\Psi}) + \kappa Vec(\boldsymbol{\Psi})Vec^\top(\boldsymbol{\Psi}),$$

em que $\kappa = \phi''(0)/(\phi'(0))^2 - 1$, conhecido como parâmetro de curtose, $\mathbf{K}_p$ como em (2.56) e I_{p^2} a matriz identidade de ordem $p^2 \times p^2$.

Observe que, se $\boldsymbol{\varepsilon} \sim t_p(\mathbf{0}, \boldsymbol{\Sigma}; \nu)$, então $\kappa = \frac{\nu-2}{\nu-4} - 1$, $\nu > 4$. No modelo normal $\kappa = 0$, que se obtem fazendo $\nu \to \infty$.

Agora, consideremos a sequência $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ definidos em (2.10) e $\boldsymbol{\varepsilon}_j = \mathbf{Y}_j - \boldsymbol{\mu}_j, j = 1, \cdots, n$. No seguinte resultado apresentamos o comportamento assintótico de $\bar{\mathbf{Y}}_n$ e $\mathbf{S}_n$.

Teorema 2.3. *Sejam $\bar{\boldsymbol{Y}}_n$ e $\boldsymbol{S}_n$ a média e matriz de covariância amostral dos vetores $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ definidos em (2.10). Então*
i) $\bar{\mathbf{Y}}_n \xrightarrow{q.c} \boldsymbol{a} + \boldsymbol{1}_p\, \mu_x = \boldsymbol{\mu}$ *e* $\mathbf{S}_n \xrightarrow{q.c} \boldsymbol{\Psi}$.

ii) Se $Var[Vec(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top)] = \boldsymbol{\Delta}$ é finito, isto é $\boldsymbol{\varepsilon}$ tem quarto momento finito, então $\sqrt{n}(\bar{\mathbf{Y}}_n - \mathbf{a} - \boldsymbol{1}_p\, \mu_x) \xrightarrow{d} N_p(\boldsymbol{0}, \boldsymbol{\Psi})$ *e* $\sqrt{n}\, Vec(\mathbf{S}_n - \boldsymbol{\Psi}) \xrightarrow{d} N_{p^2}(\mathbf{0}, \boldsymbol{\Delta})$,
em que $\boldsymbol{\Delta} = Var(\,Vec(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top)) = (1+\kappa)(I_{p^2} + \mathbf{K}_p)(\boldsymbol{\Psi} \otimes \boldsymbol{\Psi}) + \kappa Vec(\boldsymbol{\Psi}) Vec^\top(\boldsymbol{\Psi})$.

iii) $\bar{\mathbf{Y}}_n$ *e* $\boldsymbol{S}_n$ *são assintoticamente independentes.*

A prova deste Teorema pode ser encontrada, por exemplo em Vilca-Labra et al (1998).

Como $\mathbf{S}_n$ é uma matriz simétrica, então o Vec($\mathbf{S}_n$) contém elementos que são iguais, de modo que a distribuição assintótica de $\sqrt{n}$Vec($\mathbf{S}_n$) é singular. Contudo, se consideramos somente os $p(p+1)/2$ elementos diferentes de $\mathbf{S}_n$, isto é, os elementos do vetor $Vech(\mathbf{S}_n)$ é não-singular, como veremos no seguinte resultado.

Corolário 1. *Sob a condição do Teorema 2.3,*
$\sqrt{n}(\bar{\mathbf{Y}}_n - \mathbf{a} - \boldsymbol{1}_p\, \mu_x) \xrightarrow{d} N_p(\boldsymbol{0}, \boldsymbol{\Psi})$ *e* $\sqrt{n}\, Vech(\mathbf{S}_n - \boldsymbol{\Psi}) \xrightarrow{d} N_{p(p+1)/2}(\mathbf{0}, \boldsymbol{\nabla})$,
onde $\boldsymbol{\nabla} = (1+\kappa)\mathbf{H}(\boldsymbol{I}_{p^2} + \mathbf{K}_p)(\boldsymbol{\Psi} \otimes \boldsymbol{\Psi})\mathbf{H}^\top + \kappa\, Vech(\boldsymbol{\Psi})\, Vech(\boldsymbol{\Psi})^\top$.

O corolário acima segue como consequência imediata do Teorema 2.3, do Lema 2.1 e das relações em (2.56) e (2.57).

Note que o estimador $\widetilde{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$, apresentado em (2.55) depende da média amostral $\bar{\mathbf{Y}}_n$ e da matriz de covariância amostral $\mathbf{S}_n$, isto é, $\widetilde{\boldsymbol{\theta}} = F(\bar{\mathbf{Y}}_n, \mathrm{Vech}(\mathbf{S}_n))$, onde $\mathrm{Vech}(\mathbf{S}_n) = (S_{11}, S_{21}, \ldots, S_{p\,1}, S_{22}, S_{32}, \ldots, S_{p\,2}, \ldots, S_{p-1\,p-1}, S_{p-1\,p}, S_{pp})^\top$ e $\bar{\mathbf{Y}} = (\bar{Y}_1, \ldots, \bar{Y}_p)^\top$.

A seguir, apresentamos o comportamento assintótico do estimador $\widetilde{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$ para $p \geq 2$.

Teorema 2.4. *Seja o modelo definido em (2.9)-(2.10). Então $\widetilde{\boldsymbol{\theta}}$ é um estimador consistente de $\boldsymbol{\theta}$ e*

$$\sqrt{n}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N_{2p+1}(\mathbf{0}, \boldsymbol{\Lambda}),$$

onde $\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{A}^\top \boldsymbol{\Psi} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}^\top \boldsymbol{\nabla} \boldsymbol{C} \end{pmatrix}$, $\boldsymbol{A} = \boldsymbol{I}_p - \boldsymbol{B}$ *e* $\boldsymbol{C} = \begin{pmatrix} \boldsymbol{c}_1 \\ \vdots \\ \boldsymbol{c}_p \end{pmatrix}_{\frac{p(p+1)}{2} \times (p+1)}$, *com* $\boldsymbol{B} =$

$\begin{pmatrix} \boldsymbol{0} & -\boldsymbol{1}_q^\top \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$ *matriz* $p \times p$, $\boldsymbol{\nabla}$ *como no Corolário 2.1,*

$$\boldsymbol{c}_i = \begin{pmatrix} 0 & \boldsymbol{e}_p(i)^\top \\ \dfrac{2}{p(p-1)}\boldsymbol{1}_{p-i} & -\dfrac{2}{p(p-1)}\boldsymbol{1}_{p-i}\boldsymbol{1}_p^\top \end{pmatrix}$$

matriz $(p-i+1) \times (p+1)$ *para* $i = 1, \ldots, q$, $q = p-1$ *e* $\boldsymbol{c}_p = \boldsymbol{e}_{p+1}^\top(p+1)$*;* $\mathbf{e}_i(j)$ *foi definido no início da seção.*

Demonstração. Seja $(\bar{\mathbf{Y}}_n^\top, \text{Vech}^\top(\mathbf{S}_n))^\top = \Upsilon$ um vetor de dimensão $(p + p(p+1)/2) \times 1$. Como $\widetilde{\boldsymbol{\theta}} = F(\bar{\mathbf{Y}}_n, \text{Vech}(\mathbf{S}_n))$ é uma função diferenciável de $\bar{\mathbf{Y}}_n$ e $\text{Vech}(\mathbf{S}_n)$, então

$$\widetilde{\boldsymbol{\theta}} = F(\bar{\mathbf{Y}}_n, \text{Vech}(\mathbf{S}_n)) \xrightarrow{q.c.} F(\boldsymbol{\mu}\,, \text{Vech}(\boldsymbol{\Sigma})) = \boldsymbol{\theta}.$$

Por outro lado, do Corolário 2.1 e Teorema 2.3, temos que

$$\sqrt{n}\begin{pmatrix} \bar{\mathbf{Y}}_n - \boldsymbol{\mu} \\ \text{Vech}(\mathbf{S}_n) - \text{Vech}(\boldsymbol{\Sigma}) \end{pmatrix} \xrightarrow{d} N_{p+p(p+1)/2}(\mathbf{0}, \boldsymbol{\nabla}_*), \tag{2.58}$$

onde $\boldsymbol{\nabla}_* = \text{Diag}(\boldsymbol{\Psi}, (1+\kappa)\mathbf{H}(\mathbf{I}_{p^2} + \mathbf{K}_p)(\boldsymbol{\Psi} \otimes \boldsymbol{\Psi})\mathbf{H}^\top + \kappa Vech(\boldsymbol{\Psi})Vech(\boldsymbol{\Psi})^\top)$. Aplicando o método delta, temos que

$$\sqrt{n}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \sqrt{n}(F(\bar{\mathbf{Y}}_n, \text{Vech}(\mathbf{S}_n)) - F(\boldsymbol{\mu}\,, \text{Vech}(\boldsymbol{\Psi}))) \xrightarrow{d} N_{2p+1}(0, \mathbf{G}'\boldsymbol{\nabla}_*\mathbf{G})\,,$$

onde $\mathbf{G} = \dfrac{\partial F_i(\boldsymbol{\Upsilon})}{\partial \Upsilon_j}$, $i = 1, \ldots, 2p+1$, $j = 1, \ldots, p + p(p+1)/2$ tem a forma

$$\mathbf{G} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix},$$

onde $\mathbf{A} = \dfrac{\partial F_i(\boldsymbol{\Upsilon})}{\partial \Upsilon_j}$ $i, j = 1, \ldots, p$, $\mathbf{C} = \dfrac{\partial F_i(\boldsymbol{\Upsilon})}{\partial \Upsilon_j}$ $i = p+1, \ldots, 2p+1$, $j = p+1, \ldots, p + p(p+1)/2$. Após as manipulações algébricas obtém-se $\mathbf{A}$ e $\mathbf{C}$ como no enunciado do teorema, para mais detalhes sobre a obtenção das matrizes $\mathbf{K}_p$ e $\mathbf{H}$; veja por exemplo, Henderson e Searle (1979). □

Observação : Note que para $p = 3$, as matrizes $\mathbf{A}$, $\mathbf{C}$, $\mathbf{H}$ e $\mathbf{K}_p$, no Teorema acima, são respectivamente dadas por

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/3 & -1/3 & -1/3 & -1/3 \\ 1/3 & -1/3 & -1/3 & -1/3 \\ 0 & 0 & 1 & 0 \\ 1/3 & -1/3 & -1/3 & -1/3 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{K}_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

identicamente, para $p = 4$, as matrizes $\mathbf{A}$, $\mathbf{C}$, $\mathbf{H}$ e $\mathbf{K}_p$ são respectivamente

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$
\mathbf{C} = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 \\
1/6 & -1/6 & -1/6 & -1/6 & -1/6 \\
1/6 & -1/6 & -1/6 & -1/6 & -1/6 \\
1/6 & -1/6 & -1/6 & -1/6 & -1/6 \\
0 & 0 & 1 & 0 & 0 \\
1/6 & -1/6 & -1/6 & -1/6 & -1/6 \\
1/6 & -1/6 & -1/6 & -1/6 & -1/6 \\
0 & 0 & 0 & 1 & 0 \\
1/6 & -1/6 & -1/6 & -1/6 & -1/6 \\
0 & 0 & 0 & 0 & 1
\end{pmatrix},
$$

$$
\mathbf{H} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.5 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix},
$$

$$
\mathbf{K}_4 = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix} .
$$

3 Modelo de Grubbs com ponto de mudança

3.1 Introdução

Na literatura, estimação e testes de hipóteses em um modelo Gaussiano com ponto de mudança na média têm sido considerado por vários autores, como por exemplo, Chernoff e Zacks (1964), Gardner (1969), Hawkins (1992), Sen e Srivastava (1975) e Worsley (1979). Chen e Gupta (1997) estudam o problema da mudança na variância assumindo que as médias são iguais e conhecidas. Andrews et al. (1996) obtém uma classe de testes ótimos para o modelo de regressão linear normal com ponto de mudança supondo variância conhecida.

Nesta seção procederemos de maneira análoga ao trabalho de Alfaro(2000), que faz uma extensão do artigo de Chang e Huang (1997), onde se considera inferência para modelos lineares com erros nas variáveis que apresentam pontos de mudança quando as observações seguem a distribuição normal. Além disso, supondo que as observações seguem uma distribuição elíptica, apresentamos um estudo do comportamento assintótico dos estimadores de momentos conforme veremos a seguir.

O modelo considerado em Chang e Huang (1997) é o modelo clássico de regressão linear com erros nas variáveis (veja Fuller, 1987) que pode ser expresso como

$$Y_i = \alpha + \beta x_i + \varepsilon_i \quad e \quad X_i = x_i + \delta_i, \quad i = 1, \cdots, n, \tag{3.1}$$

em que α e β são constantes desconhecidas. As quantidades x_i não são observáveis e observa-se somente os pares de dados (X_i, Y_i), $i = 1, \cdots, n$. Supõe-se ainda que os erros de medida $(\varepsilon_i, \delta_i)$ são independentes e identicamente distribuídos com segundos momentos finitos.

No modelo de regressão estrutural com erros nas variáveis, usualmente assume-se que

$$x_i \sim N(\mu_x, \phi_x), \quad \varepsilon_i \sim N(0, \phi_\varepsilon), \quad \delta_i \sim N(0, \phi_\delta)$$

e independentes.

Entretanto, em muitas situações práticas, a suposição de que as quantidades desconhecidas x_j têm a mesma distribuição não é satisfeita. Por exemplo, no caso da medição das alturas das árvores, a fertilidade do solo, árvores com idades diferentes, ou outros fatores podem levar a grupos de árvores com alturas médias diferentes (veja Alfaro, 2000). Nestes casos é importante considerar um modelo que contenha tal informação, por exemplo o modelo com pontos de mudança.

Para contornar estes tipos de situações, pode ser utilizada a metodologia proposta por Chang e Huang (1997), que estendem o modelo dado em (3.1) considerando que

$$x_i \sim N(\mu_1, \phi_x), i = 1, \cdots, k \tag{3.2}$$

e

$$x_i \sim N(\mu_2, \phi_x), i = k+1, \cdots, n. \tag{3.3}$$

Seguindo a metodologia desenvolvida por Alfaro (2000) e Chang e Huang (1997), vamos apresentar um estudo do problema de ponto de mudança no contexto do modelo de Grubbs.

Considere o modelo de Grubbs definido por

$$Y_{ij} = \alpha_i + x_j + \varepsilon_{ij}, \quad i = 1, \cdots, p, \tag{3.4}$$

com $\alpha_1 = 0$. Neste contexto, introduzimos mudanças no parâmetro μ_x como segue: seja $\{x_j, \varepsilon_{1j}, \cdots, \varepsilon_{pj}\}_{j=1}^n$ uma seqüência de vetores aleatórios independentes que satisfazem

$$x_j \sim N(\mu_1, \phi_x), j = 1, \cdots, k, \tag{3.5}$$

$$x_j \sim N(\mu_2, \phi_x), j = k+1, \cdots, n \tag{3.6}$$

e

$$\varepsilon_{ij} \sim N(0, \phi_i), i = 1, \cdots, p \text{ e } j = 1, \cdots, n, \tag{3.7}$$

em que $k = [n\lambda]$ é o maior inteiro em $n\lambda$ e $\lambda \in (0, 1)$, $\mu_1, \mu_2, \phi_x, \phi_i$, $i = 1, \cdots, p$ são todos parâmetros desconhecidos.

O modelo (3.6) considera um ponto de mudança na média da distribuição da quantidade não observada, de modo que as primeiras k observações têm média μ_1 e as restantes $(n - k)$ observações têm média μ_2.

3.2 Teste para detectar a presença de ponto de mudança

O problema da detecção de pontos de mudança na distribuição da característica não observada x pode ser expresso através das seguintes hípóteses:

$$H_0 : \mu_1 = \mu_2 = \mu_x \text{ vs } H_a : \mu_1 \neq \mu_2$$

em que μ_x é uma constante desconhecida.

Sob o modelo definido em (3.4), considerando dois instrumentos ($p = 2$) e as suposições (3.6), a distribuição de $\mathbf{Y}_j = (Y_{1j}, Y_{2j})^\top$ pode ser escrita como

$$\mathbf{Y}_j \sim N_2(\mathbf{a} + \mathbf{1}_2\mu_1, \mathbf{\Sigma}), j = 1, \cdots, k \tag{3.8}$$

e

$$\mathbf{Y}_j \sim N_2(\mathbf{a} + \mathbf{1}_2\mu_2, \mathbf{\Sigma}), j = k + 1, \cdots, n, \tag{3.9}$$

em que $\mathbf{a} = (0, \alpha_2)^\top$ e

$$\mathbf{\Sigma} = \begin{pmatrix} \phi_x + \phi_1 & \phi_x \\ \phi_x & \phi_x + \phi_2 \end{pmatrix}.$$

Para detectar se há mudança no vetor de médias da distribuição dos vetores $\mathbf{Y}_j$,

$j = 1, \cdots, n$, vamos considerar a estatística do teste de razão de verossimilhança

$$\frac{\sup_{H_0} L(\boldsymbol{\theta})}{\sup_{H_a} L(\boldsymbol{\theta})}.$$

Os detalhes algébricos poder ser encontrados no Apêndice C.

Sob $H_0 : \mu_1 = \mu_2 = \mu_x$, o vetor $\mathbf{Y}_j = (Y_{1j}, Y_{2j})^\top \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $j = 1, \cdots, n$, em que $\boldsymbol{\mu} = \mathbf{a} + \mathbf{1}_2\mu_x$.

Assim a função de verossimilhança, sob H_0, pode ser escrita como

$$L(\boldsymbol{\theta}) = \frac{1}{(2\pi)^n |\boldsymbol{\Sigma}|^{n/2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{y}_j - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_j - \boldsymbol{\mu})\right\}, \tag{3.10}$$

em que

$$\boldsymbol{\theta} = (\mu_x, \alpha_2, \phi_x, \boldsymbol{\phi}^\top)^\top.$$

Considerando como uma reparametrização deste modelo para a distribuição normal bivariada com vetor de médias $\boldsymbol{\mu}_Y = (\mu_{Y1}, \mu_{Y2})^\top$ e matriz de covariâncias $\boldsymbol{\Sigma}$, a função de verossimilhança sob H_0 no modelo reparametrizado é a mesma que no modelo original.

Os estimadores de máxima verossimilhança de $\boldsymbol{\mu}_Y$ e $\boldsymbol{\Sigma}$ sob estas suposições são dados por

$$\widehat{\boldsymbol{\mu}}_Y = \overline{\mathbf{Y}} = \begin{bmatrix} \overline{Y}_1(0,n) \\ \overline{Y}_2(0,n) \end{bmatrix} \quad e \quad \widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{Y}_j - \overline{\mathbf{Y}})(\mathbf{Y}_j - \overline{\mathbf{Y}})^\top, \tag{3.11}$$

em que

$$\overline{Y}_i(a,b) = \frac{1}{b-a}\sum_{j=a+1}^{b} Y_{ij}, \quad i = 1, 2. \tag{3.12}$$

Sob $H_a : \mu_1 \neq \mu_2$ o vetor $\mathbf{Y}_j = (Y_{1j}, Y_{2j})^\top \sim N_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$, $j = 1, \cdots, k$ e $\mathbf{Y}_j = (Y_{ij}, Y_{2j})^\top \sim N_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, $j = k+1, \cdots, n$, de modo que

$$\boldsymbol{\mu}_1 = \mathbf{a} + \mathbf{1}_2\mu_1 \text{ e } \boldsymbol{\mu}_2 = \mathbf{a} + \mathbf{1}_2\mu_2.$$

Neste caso a função de verossimilhança é dada por

$$L(\boldsymbol{\theta}) = \frac{|\boldsymbol{\Sigma}|^{-n/2}}{(2\pi)^n}\; e^{-\frac{1}{2}\left[\sum_{j=1}^{k}(\mathbf{y}_j-\boldsymbol{\mu}_1)^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}_j-\boldsymbol{\mu}_1)+\sum_{j=k+1}^{n}(\mathbf{y}_j-\boldsymbol{\mu}_2)^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}_j-\boldsymbol{\mu}_2)\right]}, \tag{3.13}$$

em que

$$\boldsymbol{\theta} = (\mu_1, \mu_2, \alpha_2, \phi_x, \boldsymbol{\phi}^\top)^\top.$$

Os estimadores de máxima verossimilhança de $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ e $\boldsymbol{\Sigma}_Y$ são dados por

$$\widehat{\boldsymbol{\mu}}_1 = \overline{\mathbf{Y}}_1 = \begin{bmatrix} \overline{Y}_1(0,k) \\ \overline{Y}_2(0,k) \end{bmatrix}, \quad \widehat{\boldsymbol{\mu}}_2 = \overline{\mathbf{Y}}_2 = \begin{bmatrix} \overline{Y}_1(k,n) \\ \overline{Y}_2(k,n) \end{bmatrix} \tag{3.14}$$

e

$$\widehat{\boldsymbol{\Sigma}}_Y = \frac{B_1 + B_2}{n} = \frac{\sum_{j=1}^{k}(\mathbf{Y}_j - \overline{\mathbf{Y}}_1)(\mathbf{Y}_j - \overline{\mathbf{Y}}_1)^\top + \sum_{j=k+1}^{n}(\mathbf{Y}_j - \overline{\mathbf{Y}}_2)(\mathbf{Y}_j - \overline{\mathbf{Y}}_2)^\top}{n} \tag{3.15}$$

em que $\overline{\mathbf{Y}}_i(a,b)$, $i = 1,2$ é definida em (3.12).

O teste de razão de verossimilhança não é afetado pela reparametrização sob H_0 e o modelo é um modelo normal bivariado clássico com ponto de mudança sob H_a.

Depois de algumas manipulações algébricas pode-se mostrar que a estatística da razão de verossimilhança é dada por

$$\mathbf{W}_\lambda = \frac{1}{n-2}\left\{\sum_{j=1}^{k}\left(\mathbf{Y}_j - \overline{\mathbf{Y}}_1\right)\left(\mathbf{Y}_j - \overline{\mathbf{Y}}_1\right)^\top + \sum_{j=k+1}^{n}\left(\mathbf{Y}_j - \overline{\mathbf{Y}}_2\right)\left(\mathbf{Y}_j - \overline{\mathbf{Y}}_2\right)^\top\right\} \tag{3.16}$$

O vetor das diferenças entre as médias das observações antes e depois do ponto de mudança é dado por

$$\mathbf{Z}_\lambda = \left(\frac{k(n-k)}{n}\right)^{1/2}\left[\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2\right] \tag{3.17}$$

e a estatística de Hotelling (T^2) para testar diferença de médias na posição k é dada por

$$T_\lambda^2 = \mathbf{Z}_\lambda^\top \mathbf{W}_\lambda^{-1} \mathbf{Z}_\lambda. \tag{3.18}$$

O teste da razão de verossimilhanças para λ desconhecido pode ser baseado no valor máximo da estatística de Hotelling, ou seja,

$$T^2_{\hat{\lambda}} = \sup_{0<\lambda<1} \mathbf{Z}_\lambda^\top \mathbf{W}_\lambda^{-1} \mathbf{Z}_\lambda,$$

com $\mathbf{W}_\lambda$ e $\mathbf{Z}_\lambda$ calculados em (3.16) e (3.17), respectivamente. Valores grandes de $T^2_{\hat{\lambda}}$ indicam rejeição de H_0, ou seja, a presença de pontos de mudança na distribuição de x_j, $j = 1, \cdots, n$.

Caso a hipótese nula (H_0) seja rejeitada, os parâmetros podem ser estimados a partir do valor de λ obtidos através da maximização de $\mathbf{Z}_\lambda^\top \mathbf{W}_\lambda^{-1} \mathbf{Z}_\lambda$.

Observação: Para $p \geqslant 3$, pode ser usado o teste assintótico da Razão de Verossimilhanças, em que os EMV do modelo sob H_0 é o mesmo dado na Subseção 2.4.3 tomando $\nu \to \infty$. Sob o modelo sem restrição os EMV também podem ser obtidos via algoritmo EM conforme apresentaremos na próxima seção.

3.3 Estimação de parâmetros se a mudança é indicada

Se o teste indicou presença de mudança ou se é admitida a priori a existência de mudança no vetor de médias $\boldsymbol{\mu}$, então o parâmetro associado ao ponto de mudança λ pode ser estimado por

$$\widehat{\lambda} = j/n, \quad \text{onde } \ j \ \text{ maximiza } \ \mathbf{Z}_{j/n}^\top \mathbf{W}_{j/n}^{-1} \mathbf{Z}_{j/n}. \tag{3.19}$$

As estimativas dos outros parâmetros do modelo são calculados a seguir. O verdadeiro valor de λ será denotado por λ_0 e $k_0 = [n\lambda_0]$. Assim, quando λ_0 é conhecido os estimadores de momentos são dados por

$$\widehat{\mu}_1(\lambda_0) = \overline{Y}_1(0, k_0), \quad \widehat{\mu}_2(\lambda_0) = \overline{Y}_1(k_0, n), \tag{3.20}$$

$$\widehat{\alpha}_2(\lambda_0) = \frac{1}{2}[\{\overline{Y}_2(0,k_0)+\overline{Y}_2(k_0,n)\}-\{\overline{Y}_1(0,k_0)+\overline{Y}_1(k_0,n)\}], \quad (3.21)$$

$$\widehat{\phi}_x(\lambda_0) = \frac{1}{n}\{S_{12}(0,k_0)+S_{12}(k_0,n)\}\vee 0, \quad (3.22)$$

$$\widehat{\phi}_i(\lambda_0) = \left[\frac{1}{n}\{S_{ii}(0,k_0)+S_{ii}(k_0,n)\}-\widehat{\phi}_x(\lambda_0)\right]\vee 0, \quad (3.23)$$
$$i = 1, 2,$$

em que

$$S_{il}(a,b) = \sum_{j=a+1}^{b}\{Y_{ij}-\overline{Y}_i(a,b)\}\{Y_{lj}-\overline{Y}_l(a,b)\},$$

$$S_{ii}(a,b) = \sum_{j=a+1}^{b}\{Y_{ij}-\overline{Y}_i(a,b)\}^2$$

e $s \vee t = \max(s,t)$.

Para o caso geral, p instrumentos, os EMV do modelo sem restrição são obtidos através do algoritmo EM da seguinte forma:

Etapa E: Calculamos:

$$x_{1j}^{(m)} = \mu_x + \frac{\phi_x^{(m)}}{c^{(m)}}\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi}^{(m)})(\mathbf{Y}_j-\boldsymbol{\mu}_1^{(m)}),$$

$$x_{1j}^{2(m)} = \frac{\phi_x^{(m)}}{c^{(m)}}\frac{\nu+||\mathbf{T}_j^{(m)}||^2}{\nu+p-2}+(x_{1j}^{(m)})^2,\ j=1,\cdots,k_0$$

e

$$x_{2j}^{(m)} = \mu_x + \frac{\phi_x^{(m)}}{c^{(m)}}\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi}^{(m)})(\mathbf{Y}_j-\boldsymbol{\mu}_2^{(m)}),$$

$$x_{2j}^{2(m)} = \frac{\phi_x^{(m)}}{c^{(m)}}\frac{\nu+||\mathbf{T}_j^{(m)}||^2}{\nu+p-2}+(x_{2j}^{(m)})^2,\ j=k_0+1,\cdots,n,$$

com $\boldsymbol{\mu}_1^{(m)} = \mu_1^{(m)}\mathbf{1}_p+\mathbf{a}^{(m)}$, $\boldsymbol{\mu}_2^{(m)} = \mu_2^{(m)}\mathbf{1}_p+\mathbf{a}^{(m)}$, $c^{(m)} = 1+\phi_x^{(m)}\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi}^{(m)})\mathbf{1}_p$ e $\mathbf{T}_j^{(m)} = (\phi_x^{(m)}\mathbf{1}_p\mathbf{1}_p^\top+D(\phi^{(m)}))^{-1/2}(\mathbf{Y}_j-\boldsymbol{\mu}_l^{(m)}),\ l=1,2.$

Etapa M: Obtemos os EMV da etapa $m+1$ através das expressões:

$$
\begin{aligned}
\mu_1^{(m+1)} &= \frac{1}{k_0}\sum_{j=1}^{k_0} x_{1j}^{(m)}, \\
\mu_2^{(m+1)} &= \frac{1}{n-k_0}\sum_{j=k_0+1}^{n} x_{2j}^{(m)}, \\
\alpha_i^{(m+1)} &= \frac{1}{n}\sum_{j=1}^{n} Y_{ij} - \left(\mu_1^{(m+1)} + \mu_2^{(m+1)}\right),\ i = 2, \cdots, p, \\
\phi_x^{(m+1)} &= \frac{1}{n}\left(\sum_{j=1}^{k_0}(x_{1j}^{(m)} - \mu_1^{(m+1)})^2 + \sum_{j=k_0+1}^{n}(x_{2j}^{(m)} - \mu_2^{(m+1)})^2\right), \\
\phi_1^{(m+1)} &= \frac{1}{n}\left(\sum_{j=1}^{k_0}(x_{1j}^{(m)} - y_{1j})^2 + \sum_{j=k_0+1}^{n}(x_{2j}^{(m)} - y_{1j})^2\right), \\
\phi_i^{(m+1)} &= \frac{1}{n}\left(\sum_{j=1}^{k_0}(y_{ij} - \alpha_i^{(m+1)} - x_{1j}^{(m)})^2 + \sum_{j=k_0+1}^{n}(y_{ij} - \alpha_i^{(m+1)} - x_{2j}^{(m)})^2\right),\quad i = 2, \cdots, p.
\end{aligned}
$$

3.4 Distribuição assintótica dos estimadores de $\boldsymbol{\alpha}$

Nesta seção, faremos um estudo assintótico dos estimadores de momentos de $\boldsymbol{\alpha}$, considerando que as observações seguem uma distribuição elíptica. Faremos isto substituindo as suposições dadas em (3.8)-(3.9) por

$$
\mathbf{Y}_j \sim El_p(\mathbf{a} + \mathbf{1}_p\mu_1, \boldsymbol{\Sigma}; f), j = 1\cdots, k, \tag{3.24}
$$

e

$$
\mathbf{Y}_j \sim El_p(\mathbf{a} + \mathbf{1}_p\mu_2, \boldsymbol{\Sigma}; f), j = k+1\cdots, n, \tag{3.25}
$$

com $Var(\mathbf{Y}_j) = c_f\boldsymbol{\Sigma}$ e c_f uma constante que depende de f como em (A.8). Para efeito desta seção, vamos considerar que as observações possuem segundo momento finito.

Definimos a classe dos estimadores de $\boldsymbol{\alpha} = (\alpha_2, \cdots, \alpha_p)^\top$ como: $W = \{\hat{\alpha}_i(\lambda) : \hat{\alpha}_i(\lambda) = \frac{1}{2}(\overline{Y}_i(0,k) + \overline{Y}_i(k,n) - \overline{Y}_1(0,k) - \overline{Y}_1(k,n)), i = 2, \cdots, p; k = [n\lambda]; \lambda \in (0,1)\}$.

Teorema 3.1. *Para quaisquer $\lambda \in (0,1)$ e $\hat{\boldsymbol{\alpha}}(\lambda) = (\hat{\alpha}_2(\lambda), \cdots, \hat{\alpha}_p(\lambda))^\top$, com $\hat{\alpha}_i \in W$, temos*

a) $\hat{\boldsymbol{\alpha}}(\lambda) \longrightarrow \boldsymbol{\alpha}$, *quase certamente, quando $n \longrightarrow \infty$;*

b) $\sqrt{n}(\hat{\boldsymbol{\alpha}}(\lambda) - \boldsymbol{\alpha}) \xrightarrow{d} N_{p-1}(0, \boldsymbol{\Omega}(\lambda))$,

em que $\boldsymbol{\Omega}(\lambda) = c_f \frac{1}{4\lambda(1-\lambda)} \left(\phi_1 \mathbf{1}_{p-1}\mathbf{1}_{p-1}^\top + D(\boldsymbol{\phi}^*)\right)$, *com* $\boldsymbol{\phi}^* = (\phi_2, \cdots, \phi_p)$.

Prova de a): Consideremos os casos:

$$\begin{cases} 1. \quad \lambda \leqslant \lambda_0; \\ 2. \quad \lambda > \lambda_0, \end{cases}$$

com λ_0 sendo o verdadeiro valor do ponto de mudança.

Caso 1. $\lambda \leqslant \lambda_0$: Sejam

$$\underbrace{\mathbf{Y}_1, \cdots, \mathbf{Y}_{[n\lambda]}}_{G1}, \underbrace{\mathbf{Y}_{[n\lambda]+1}, \cdots, \mathbf{Y}_{[n\lambda_0]}, \mathbf{Y}_{[n\lambda_0]+1}, \cdots, \mathbf{Y}_n}_{G2}.$$

Em $G1$ temos que $E(\mathbf{Y}_j) = \mathbf{a} + \mathbf{1}_p\mu_1$. Portanto, pela Lei Forte de Kolmogorov (veja por exemplo James, 1996),

$$\overline{Y}_i(0,[n\lambda]) \xrightarrow{q.c.} \alpha_i + \mu_1,\ i = 1, \cdots, p.$$

No grupo $G2$, temos que os Y_j's são independentes mas não são identicamente distribuídos. Entretanto, para $i = 1, \cdots, p$, segue que

$$\lim_{n\to\infty} \sum_{j=[n\lambda]+1}^{n} \frac{Var(Y_{ij})}{j^2} = c_f(\phi_x + \phi_i) \lim_{n\to\infty} \sum_{j=[n\lambda]+1}^{n} \frac{1}{j^2} < \infty.$$

Além disso,

$$\begin{aligned} E(\overline{Y}_i([n\lambda], n)) &= \frac{([n\lambda_0] - [n\lambda])(\alpha_i + \mu_1) + (n - [n\lambda_0])(\alpha_i + \mu_2)}{n - [n\lambda]} \\ &\longrightarrow \frac{\lambda_0 - \lambda}{1 - \lambda}(\alpha_i + \mu_1) + \frac{1 - \lambda_0}{1 - \lambda}(\alpha_i + \mu_2), \end{aligned}$$

quando $n \longrightarrow \infty$, pois

$$\lim_{n\to\infty} \frac{[n\lambda_0]-[n\lambda]}{n-[n\lambda]} = \frac{\lambda_0-\lambda}{1-\lambda}$$

e

$$\lim_{n\to\infty} \frac{[n-n\lambda_0]}{n-[n\lambda]} = \frac{1-\lambda_0}{1-\lambda}.$$

Portanto, pela primeira Lei Forte de Kolmogorov (veja por exemplo James, 1996),

$$\overline{Y}_i([n\lambda],n) \xrightarrow{q.c.} \frac{\lambda_0-\lambda}{1-\lambda}(\alpha_i+\mu_1) + \frac{1-\lambda_0}{1-\lambda}(\alpha_i+\mu_2),\ i=1,\cdots,p.$$

Segue, dos resultados acima, que

$$\begin{aligned} \hat{\alpha}_i(\lambda) \quad &\xrightarrow{q.c.} \quad \frac{1}{2}\left[(\alpha_i+\mu_1)-\mu_1+\frac{\lambda_0-\lambda}{1-\lambda}(\alpha_i+\mu_1)+\frac{1-\lambda_0}{1-\lambda}(\alpha_i+\mu_2)\right. \\ &- \quad \left.\left(\frac{\lambda_0-\lambda}{1-\lambda}\mu_1+\frac{1-\lambda_0}{1-\lambda}\mu_2\right)\right] = \alpha_i,\ i=2,\cdots,p. \end{aligned}$$

A prova do *Caso 2* ($\lambda > \lambda_0$) segue de maneira análoga.

Prova de b): Sejam

$$\overline{Y}_i^*(0,[n\lambda]) = \frac{1}{[n\lambda]}\sum_{j=1}^{[n\lambda]}(Y_{ij}-E(Y_{ij})) = \overline{Y}_i(0,[n\lambda]) - E(\overline{Y}_i(0,[n\lambda]))$$

e

$$\overline{Y}_i^*([n\lambda],n) = \frac{1}{n-[n\lambda]}\sum_{j=[n\lambda]+1}^{n}(Y_{ij}-E(Y_{ij})) = \overline{Y}_i([n\lambda],n) - E(\overline{Y}_i([n\lambda],n)),$$

$i=1,\cdots,p$. Note que

$$\sqrt{n}\begin{pmatrix} \overline{Y}_1^*(0,[n\lambda]) \\ \overline{Y}_1^*([n\lambda],n) \\ \vdots \\ \overline{Y}_p^*(0,[n\lambda]) \\ \overline{Y}_p^*([n\lambda],n) \end{pmatrix} = A_{(n,\lambda)} \underbrace{\begin{pmatrix} \sqrt{[n\lambda]}\left\{\overline{Y}_1(0,[n\lambda]) - E[\overline{Y}_1(0,[n\lambda])]\right\} \\ \sqrt{[n-n\lambda]}\left\{\overline{Y}_1([n\lambda],n) - E[\overline{Y}_1([n\lambda],n)]\right\} \\ \vdots \\ \sqrt{[n\lambda]}\left\{\overline{Y}_p(0,[n\lambda]) - E[\overline{Y}_p(0,[n\lambda])]\right\} \\ \sqrt{[n-n\lambda]}\left\{\overline{Y}_p([n\lambda],n) - E[\overline{Y}_p([n\lambda],n)]\right\} \end{pmatrix}}_{\varphi\left(\overline{Y}_1(0,[n\lambda]),\overline{Y}_1([n\lambda],n),\cdots,\overline{Y}_p(0,[n\lambda]),\overline{Y}_p([n\lambda],n)\right)}, \qquad (3.26)$$

em que

$$A_{(n,\lambda)} = \begin{pmatrix} \frac{\sqrt{n}}{\sqrt{[n\lambda]}} & 0 & \cdots & 0 & 0 \\ 0 & \frac{\sqrt{n}}{\sqrt{n-[n\lambda]}} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{\sqrt{n}}{\sqrt{[n\lambda]}} & 0 \\ 0 & 0 & \cdots & 0 & \frac{\sqrt{n}}{\sqrt{n-[n\lambda]}} \end{pmatrix}.$$

Como

$$\lim_{n\to\infty} \frac{[n\lambda]}{n} = \lim_{n\to\infty} \frac{n\lambda}{n} = \lambda$$

e

$$\lim_{n\to\infty} \frac{n-[n\lambda]}{n} = \lim_{n\to\infty} \left(1 - \frac{n\lambda}{n}\right) = 1-\lambda,$$

temos que

$$A_{(n,\lambda)} \longrightarrow \begin{pmatrix} \frac{1}{\sqrt{\lambda}} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{\sqrt{1-\lambda}} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\lambda}} & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{\sqrt{1-\lambda}} \end{pmatrix} = A_\lambda.$$

Ainda, pelo Teorema Central do Limite (supondo segundo momento finito),

$$\varphi\left(\overline{Y}_1(0,[n\lambda]), \overline{Y}_1([n\lambda],n), \cdots, \overline{Y}_p(0,[n\lambda]), \overline{Y}_p([n\lambda],n)\right) \xrightarrow{d} N_{2p}(0,\boldsymbol{\Psi}),$$

em que

$$\boldsymbol{\Psi} = c_f \begin{pmatrix} \phi_x+\phi_1 & 0 & \phi_x & 0 & \cdots & \phi_x & 0 \\ 0 & \phi_x+\phi_1 & 0 & \phi_x & \cdots & 0 & \phi_x \\ \phi_x & 0 & \phi_x+\phi_2 & 0 & \cdots & \phi_x & 0 \\ 0 & \phi_x & 0 & \phi_x+\phi_2 & \cdots & 0 & \phi_x \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi_x & 0 & \phi_x & 0 & \cdots & \phi_x+\phi_p & 0 \\ 0 & \phi_x & 0 & \phi_x & \cdots & 0 & \phi_x+\phi_p \end{pmatrix}. \quad (3.27)$$

Logo, pelo Teorema de Slutsky, segue que

$$\sqrt{n}\begin{pmatrix}\overline{Y}_1^*(0,[n\lambda])\\ \overline{Y}_1^*([n\lambda],n)\\ \vdots\\ \overline{Y}_p^*(0,[n\lambda])\\ \overline{Y}_p^*([n\lambda],n)\end{pmatrix} \xrightarrow{d} N_{2p}(0,\boldsymbol{\Sigma}^*),$$

em que $\boldsymbol{\Sigma}^* = A_\lambda \boldsymbol{\Psi} A_\lambda$.

Agora, vamos estudar a distribuição assintótica de $\hat{\boldsymbol{\alpha}}(\lambda)$, considerando os seguintes casos

1. $\lambda = \lambda_0$
2. $\lambda < \lambda_0$
3. $\lambda > \lambda_0$

Caso 1. $\lambda = \lambda_0$: Sejam

$$\overline{\mathbf{Z}} = \left(\overline{Y}_1(0,[n\lambda]),\overline{Y}_1([n\lambda],n),\cdots,\overline{Y}_p(0,[n\lambda]),\overline{Y}_p([n\lambda],n)\right)^\top$$

e

$$\boldsymbol{\theta} = (\mu_1,\mu_2,\alpha_2+\mu_1,\alpha_2+\mu_2,\cdots,\alpha_p+\mu_1,\alpha_p+\mu_2)^\top,$$

então $E(\overline{\mathbf{Z}}) = \boldsymbol{\theta}$ e

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}(\lambda)-\boldsymbol{\alpha}) = \sqrt{n}\begin{pmatrix}\frac{1}{2}(\overline{Y}_2(0,[n\lambda])+\overline{Y}_2([n\lambda],n)-\overline{Y}_1(0,[n\lambda])-\overline{Y}_1([n\lambda],n))-\alpha_2\\ \frac{1}{2}(\overline{Y}_3(0,[n\lambda])+\overline{Y}_3([n\lambda],n)-\overline{Y}_1(0,[n\lambda])-\overline{Y}_1([n\lambda],n))-\alpha_3\\ \vdots\\ \frac{1}{2}(\overline{Y}_p(0,[n\lambda])+\overline{Y}_p([n\lambda],n)-\overline{Y}_1(0,[n\lambda])-\overline{Y}_1([n\lambda],n))-\alpha_p\end{pmatrix}.$$

Se

$$G(\overline{\mathbf{Z}}) = \hat{\boldsymbol{\alpha}}(\lambda) = (\hat{\alpha}_2(\lambda),\cdots,\hat{\alpha}_p(\lambda))^\top = (g_1(\overline{\mathbf{Z}}),\cdots,g_{p-1}(\overline{\mathbf{Z}}))^\top,$$

temos que

$$\begin{aligned}
g_1(\boldsymbol{\theta}) &= \frac{1}{2}(\alpha_2 + \mu_1 + \alpha_2 + \mu_2 - \mu_1 + \mu_2) = \alpha_2, \\
g_2(\boldsymbol{\theta}) &= \frac{1}{2}(\alpha_3 + \mu_1 + \alpha_3 + \mu_2 - \mu_1 + \mu_2) = \alpha_3, \\
&\vdots \\
g_{p-1}(\boldsymbol{\theta}) &= \frac{1}{2}(\alpha_p + \mu_1 + \alpha_p + \mu_2 - \mu_1 + \mu_2) = \alpha_p,
\end{aligned}$$

logo $G(\boldsymbol{\theta}) = \boldsymbol{\alpha}$.

Segue que

$$\mathbf{D} = \left.\frac{\partial G^\top}{\partial \mathbf{Z}}\right|_{\mathbf{Z}=\boldsymbol{\theta}} = \frac{1}{2}\begin{pmatrix} -1 & -1 & -1 & \cdots & -1 \\ -1 & -1 & -1 & \cdots & -1 \\ 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}_{(2p\times(p-1))}.$$

Logo, pelo método Delta,

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}(\lambda) - \boldsymbol{\alpha}) = \sqrt{n}(G(\overline{\mathbf{Z}}) - G(\boldsymbol{\theta})) \xrightarrow{d} N_{p-1}(0, \boldsymbol{\Omega}(\lambda)),$$

em que

$$\boldsymbol{\Omega}(\lambda) = \mathbf{D}^\top \boldsymbol{\Sigma}^* \mathbf{D} = c_f \frac{1}{4\lambda(1-\lambda)} \left(\phi_1 \mathbf{1}_{p-1}\mathbf{1}_{p-1}^\top + D(\boldsymbol{\phi}^*)\right), \tag{3.28}$$

com $\boldsymbol{\phi}^* = (\phi_2, \cdots, \phi_p)^\top$.

Caso 2. $\lambda < \lambda_0$: Neste caso temos que

$$E\begin{pmatrix} \overline{Y}_1(0,[n\lambda]) \\ \overline{Y}_1([n\lambda],n) \\ \overline{Y}_2(0,[n\lambda]) \\ \overline{Y}_2([n\lambda],n) \\ \vdots \\ \overline{Y}_p(0,[n\lambda]) \\ \overline{Y}_p([n\lambda],n) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \frac{1}{a_\lambda}(m_\lambda\mu_1 + n_\lambda\mu_2) \\ \alpha_2 + \mu_1 \\ \frac{1}{a_\lambda}[m_\lambda(\alpha_2+\mu_1) + n_\lambda(\alpha_2+\mu_2)] \\ \vdots \\ \alpha_p + \mu_1 \\ \frac{1}{a_\lambda}[m_\lambda(\alpha_p+\mu_1) + n_\lambda(\alpha_p+\mu_2)] \end{pmatrix},$$

em que $a_\lambda = n - [n\lambda], m_\lambda = [n\lambda_0] - [n\lambda]$ e $n_\lambda = n - [n\lambda_0]$. Ainda

$$\begin{aligned} E(\hat{\alpha}_i) &= \frac{1}{2}E[(\overline{Y}_i(0,[n\lambda]) + \overline{Y}_i([n\lambda],n) - \overline{Y}_1(0,[n\lambda]) - \overline{Y}_1([n\lambda],n))] \\ &= \frac{1}{2}\left[\alpha_i + \mu_1 + \frac{1}{a_\lambda}[m_\lambda(\alpha_i+\mu_1) + n_\lambda(\alpha_i+\mu_2)] - [\mu_1 + \frac{1}{a_\lambda}(m_\lambda\mu_1 + n_\lambda\mu_2)]\right] \\ &= \alpha_i, \ i = 2, \cdots, p. \end{aligned}$$

Prosseguindo de maneira análoga ao *Caso 1*, teremos pelo método Delta que

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}(\lambda) - \boldsymbol{\alpha}) \xrightarrow{d} N_{p-1}(0, \boldsymbol{\Omega}(\lambda)),$$

com $\boldsymbol{\Omega}(\lambda)$ como em (3.28).

Caso 3. $\lambda > \lambda_0$: Finalmente, temos neste caso que

$$E\begin{pmatrix} \overline{Y}_1(0,[n\lambda]) \\ \overline{Y}_1([n\lambda],n) \\ \overline{Y}_2(0,[n\lambda]) \\ \overline{Y}_2([n\lambda],n) \\ \vdots \\ \overline{Y}_p(0,[n\lambda]) \\ \overline{Y}_p([n\lambda],n) \end{pmatrix} = \begin{pmatrix} \frac{1}{[n\lambda]}([n\lambda_0]\mu_1 + r_\lambda\mu_2) \\ \mu_2 \\ \frac{1}{[n\lambda]}[[n\lambda_0](\alpha_2+\mu_1) + r_\lambda(\alpha_2+\mu_2)] \\ \alpha_2 + \mu_2 \\ \vdots \\ \frac{1}{[n\lambda]}[[n\lambda_0](\alpha_p+\mu_1) + r_\lambda(\alpha_p+\mu_2)] \\ \alpha_p + \mu_2 \end{pmatrix},$$

em que $r_\lambda = [n\lambda] - [n\lambda_0]$. Além disso

$$
\begin{aligned}
E(\hat{\alpha}_i) &= \frac{1}{2}E[(\overline{Y}_i(0,[n\lambda]) + \overline{Y}_i([n\lambda],n) - \overline{Y}_1(0,[n\lambda]) - \overline{Y}_1([n\lambda],n))] \\
&= \frac{1}{2}\left[\frac{1}{[n\lambda]}[[n\lambda_0](\alpha_i+\mu_1) + r_\lambda(\alpha_i+\mu_2)] + \alpha_p + \mu_2 - \frac{1}{[n\lambda]}([n\lambda_0]\mu_1 + r_\lambda\mu_2) - \mu_2\right] \\
&= \alpha_i, \ i = 2,\cdots,p.
\end{aligned}
$$

Novamente, teremos pelo método Delta que

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}(\lambda) - \boldsymbol{\alpha}) \xrightarrow{d} N_{p-1}(0, \boldsymbol{\Omega}(\lambda)),$$

com $\boldsymbol{\Omega}(\lambda)$ como em (3.28).

De modo que, dos resultados acima, temos a prova do resultado em **b)**.

Devemos notar que cada estimador em W é consistente para $\alpha_i, i = 2,\cdots,p$, mesmo se a partição dos dados em dois grupos por λ estiver incorreta.

4 Aplicações e Simulação

4.1 Introdução

Neste capítulo serão avaliados numericamente os desempenhos dos testes propostos no Capítulo 2. Gostaríamos de identificar, dentre os testes de Wald, Escore e Razão de Verossimilhanças, aquele que apresenta melhor comportamento em termos de tamanho e poder do teste.

Para tal objetivo, vamos considerar que um teste tem bom desempenho se apresentar baixas probabilidades dos erros tipo I e II, correspondentes à rejeição de que os instrumentos são precisos e não-viciados (H_{01}) quando eles realmente são e não rejeição de que os instrumentos são imprecisos e viciados quando de fato eles são. Os testes que apresentam, sob a hipótese nula, taxas de rejeição superior ao nível de significância nominal são chamados *testes anticonservativos*. O oposto disto são os *testes conservativos*. À primeira vista, pode parecer uma característica desejável o fato de um teste ser conservativo, uma vez que nessa situação o erro tipo I ocorre com baixa probabilidade. No entanto, em geral testes conservativos apresentam um baixo poder. Assim, busca-se identificar testes que tenham tamanhos empíricos razoavelmente próximos (ou menores) ao nível de significância nominal e, ao mesmo tempo, tenham um alto poder.

Também apresentamos um pequeno estudo de simulação para avaliar a capacidade do teste de Razão de Verossimilhanças, proposto no Capítulo 3, de detectar a presença de pontos de mudança, assim como a qualidade dos estimadores sob a presença deste.

Na seção que segue, apresentamos ilustrações numéricas dos resultados do Capítulo 2, envolvendo conjuntos de dados reais.

4.2 Ilustração Numérica

Nesta seção, dois exemplos serão apresentados com a intenção de ilustrar os resultados obtidos no Capítulo 2.

4.2.1 Exemplo 1

Neste exemplo, apresentaremos um estudo da estimação por máxima verossimilhança via algoritmo EM, conforme exposto na Subseção 2.4.3. Para isso, analisaremos o conjunto de dados apresentado em Jaech (1985) e reproduzidos na Tabela 4.1. Os dados são relativos a um experimento no qual a densidade de 43 pastilhas de combustível de urânio sintetizado para uso em reatores nucleares foram medidas por seis instrumentos, através de dois métodos, da seguinte forma:

Instrumento 1: Método geométrico, operador 1
Instrumento 2: Método geométrico, operador 2
Instrumento 3: Método geométrico, operador 3
Instrumento 4: Método de imersão, operador 4
Instrumento 5: Método de imersão, operador 5
Instrumento 6: Método de imersão, operador 6.

O método gemométrico consiste em pesar a pastilha e encontrar seu volume e, assim, obter a densidade. No método de imersão, verifica-se a mudança do peso da pastilha quando pesadas no ar e em um determinado líquido.

Logo, y_{ij} representa a densidade fornecida pelo instrumento i e a pastilha j, com $i = 1, \ldots, 6$, $j = 1, \ldots, 43$. Note que, neste caso, o vetor de parâmetros a ser estimado é dado por:

$$\boldsymbol{\theta} = (\mu_x, \alpha_2, \ldots, \alpha_6, \phi_x, \phi_1, \ldots, \phi_6)^\top \tag{4.1}$$

Lachos-Dávila (2002) mostra que os dados podem ser considerados normais e, portanto, estima os parâmetros do modelo considerando o modelo estrutural normal de

Grubbs. Com o intuito de avaliar os resultados da Subseção 2.4.3, mostraremos que as estimativas dos parâmetros do modelo coincidem com o caso normal quando os graus de liberdade da t-Student tendem a infinito. De fato, note na Tabela 4.2 como as estimativas dos parâmetros vão se aproximando daquelas encontradas em Lachos-Dávila (2002) à medida que os graus de liberdade assumem valores maiores.

Tabela 4.1: Medidas de densidade de pastilhas de urânio fornecidas por seis instrumentos

Obs.	Instrumento 1	Instrumento 2	Instrumento 3	Instrumento 4	Instrumento 5	Instrumento 6
1	4.30	4.06	4.14	4.68	5.22	4.75
2	4.51	4.35	4.53	5.15	4.45	4.79
3	4.42	4.52	4.57	4.78	4.79	4.72
4	4.90	4.67	4.75	5.55	5.10	4.75
5	4.60	4.70	4.84	5.08	4.82	5.05
6	4.29	4.30	4.21	4.80	4.47	4.93
7	4.30	4.47	4.01	4.41	4.27	4.66
8	4.69	4.59	4.47	5.32	4.95	4.75
9	4.02	4.05	4.09	4.12	4.48	4.40
10	4.61	4.78	4.60	4.41	4.75	5.18
11	4.35	4.25	4.40	4.37	4.06	4.39
12	4.46	4.55	4.70	4.73	5.14	4.75
13	4.08	4.12	3.85	4.60	4.22	4.33
14	3.95	3.95	4.13	4.62	4.49	3.94
15	4.21	4.27	3.94	4.34	4.29	4.62
16	4.62	4.47	4.50	5.09	5.44	4.70
17	4.81	4.81	4.98	5.26	4.86	5.22
18	4.47	4.34	4.33	5.05	4.62	4.38
19	4.38	4.24	4.47	4.60	4.27	4.20
20	4.43	4.37	4.57	5.23	5.04	4.51
21	4.74	4.52	4.58	5.24	5.36	4.77
22	4.35	4.37	4.45	4.42	4.66	4.54
23	4.11	4.23	4.07	4.46	4.48	4.26
24	3.97	3.98	4.19	4.07	4.69	4.29
25	4.40	4.36	4.33	4.55	4.73	4.62
26	4.24	4.30	4.36	4.84	4.56	4.97
27	4.51	4.52	4.57	4.70	4.20	4.69
28	4.49	4.33	4.61	4.85	4.26	4.67
29	4.42	4.14	4.22	4.35	4.39	4.46
30	4.49	4.59	4.69	5.06	5.50	4.75
31	4.14	4.32	4.45	4.02	4.30	4.58
32	4.38	4.25	4.44	4.86	4.81	4.63
33	4.45	4.60	4.28	4.40	4.17	4.81
34	4.39	4.38	4.64	4.86	4.98	4.54
35	4.56	4.35	4.83	4.53	5.36	4.46
36	4.38	4.40	4.60	4.96	4.92	4.56
37	4.44	4.34	4.66	4.60	5.59	4.66
38	4.21	4.24	4.07	4.21	4.39	4.40
39	4.51	4.52	4.54	4.79	4.85	4.77
40	4.23	4.12	4.35	4.62	4.45	4.15
41	4.33	4.47	4.68	5.05	4.29	4.92
42	4.27	4.21	4.45	4.74	5.24	4.41
43	4.67	4.52	4.60	4.42	5.06	4.73

Outra questão que é conveniente avaliar quando aplicamos métodos iterativos de es-

Tabela 4.2: Parâmetros estimados para diversos graus de liberdade

ν	μ_x	α_2	α_3	α_4	α_5	α_6	ϕ_x	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6
1	4.4036	-0.022	0.0572	0.3222	0.3426	0.2135	0.0374	0.0097	0.0088	0.0253	0.0814	0.1383	0.0279
2	4.4013	-0.0244	0.055	0.3252	0.339	0.2123	0.0331	0.0083	0.0074	0.0223	0.0713	0.1235	0.0252
3	4.3999	-0.0256	0.053	0.3263	0.3362	0.2123	0.0324	0.0078	0.007	0.0217	0.0691	0.1208	0.025
4	4.399	-0.0263	0.0513	0.3266	0.3341	0.2128	0.0324	0.0076	0.0068	0.0217	0.0686	0.1204	0.0252
5	4.3984	-0.0267	0.05	0.3267	0.3326	0.2133	0.0325	0.0075	0.0068	0.0218	0.0686	0.1207	0.0256
10	4.3970	-0.0273	0.046	0.3256	0.3284	0.2159	0.0334	0.0073	0.0068	0.0226	0.0698	0.1233	0.027
50	4.3967	-0.0272	0.0405	0.3212	0.3241	0.2209	0.0353	0.007	0.0073	0.0242	0.0736	0.129	0.0297
10^2	4.3969	-0.0271	0.0396	0.3201	0.3236	0.2218	0.0357	0.0069	0.0074	0.0246	0.0744	0.1301	0.0302
10^3	4.3972	-0.027	0.0387	0.319	0.3231	0.2227	0.036	0.0069	0.0076	0.0249	0.0752	0.1311	0.0308
10^4	4.3972	-0.027	0.0386	0.3189	0.323	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1312	0.0309
10^5	4.3972	-0.0270	0.0386	0.3188	0.3230	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1313	0.0309
10^6	4.3972	-0.027	0.0386	0.3188	0.323	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1313	0.0309
Lachos	4.3972	-0.027	0.0386	0.3188	0.323	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1313	0.0309

timação é a velocidade de convergência do algoritmo. Para tanto, fixando o número de iterações, consideramos diversos graus de liberdade para o modelo estrutural de Grubbs t-Student.

A Tabela 4.3 mostra algumas iterações do algoritmo EM considerando $\nu = 1$, ou seja, supomos que os dados seguem a distribuição de Cauchy, que é um caso particular da t-Student.

Tabela 4.3: Convergência do algoritmo EM considerando 1 grau de liberdade

Iter	μ_x	α_2	α_3	α_4	α_5	α_6	ϕ_x	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6
1	4.4049	-0.0255	0.0421	0.3108	0.3352	0.2095	0.0369	0.0103	0.0013	0.0309	0.0905	0.1414	0.0234
11	4.4052	-0.0225	0.0566	0.3203	0.3427	0.2122	0.0345	0.0091	0.0076	0.0234	0.0762	0.1283	0.0250
21	4.4037	-0.0219	0.0572	0.3219	0.3427	0.2135	0.0367	0.0095	0.0086	0.0248	0.0799	0.1357	0.0273
31	4.4036	-0.0220	0.0572	0.3221	0.3426	0.2135	0.0373	0.0097	0.0088	0.0251	0.0811	0.1377	0.0278
41	4.4036	-0.0220	0.0572	0.3222	0.3426	0.2135	0.0374	0.0097	0.0088	0.0252	0.0813	0.1382	0.0279
51	4.4036	-0.0220	0.0572	0.3222	0.3426	0.2135	0.0374	0.0097	0.0088	0.0252	0.0814	0.1383	0.0279
61	4.4036	-0.0220	0.0572	0.3222	0.3426	0.2135	0.0374	0.0097	0.0088	0.0253	0.0814	0.1383	0.0279
71	4.4036	-0.0220	0.0572	0.3222	0.3426	0.2135	0.0374	0.0097	0.0088	0.0253	0.0814	0.1383	0.0279
81	4.4036	-0.0220	0.0572	0.3222	0.3426	0.2135	0.0374	0.0097	0.0088	0.0253	0.0814	0.1383	0.0279
91	4.4036	-0.0220	0.0572	0.3222	0.3426	0.2135	0.0374	0.0097	0.0088	0.0253	0.0814	0.1383	0.0279
101	4.4036	-0.0220	0.0572	0.3222	0.3426	0.2135	0.0374	0.0097	0.0088	0.0253	0.0814	0.1383	0.0279
111	4.4036	-0.0220	0.0572	0.3222	0.3426	0.2135	0.0374	0.0097	0.0088	0.0253	0.0814	0.1383	0.0279

Observe que a convergência das estimativas de $\boldsymbol{\theta}$ é atingida rapidamente, aproximadamente na iteração 50. Nas Tabelas 4.4, 4.5, 4.6, 4.7 e 4.8 são apresentadas as iterações para, respectivamente, $\nu = 4$, $\nu = 10$, $\nu = 100$, $\nu = 10000$ e $\nu = 57779$, sendo que este último valor foi escolhido pois a partir dele todas as estimativas coincidem com as estimativas do modelo normal até, pelo menos, a quarta casa decimal. Observe ainda que em todos estes casos, a convergência de $\boldsymbol{\theta}$ é obtida sempre em torno da iteração 50.

Tabela 4.4: Convergência do algoritmo EM considerando 4 graus de liberdade

Iter	μ_x	α_2	α_3	α_4	α_5	α_6	ϕ_x	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6
1	4.4006	-0.0261	0.0416	0.3175	0.3293	0.2122	0.0359	0.0101	0.0012	0.0299	0.0858	0.1399	0.0236
11	4.3999	-0.0266	0.0515	0.3264	0.3348	0.2117	0.0319	0.0079	0.0060	0.0216	0.0688	0.1197	0.0239
21	4.3991	-0.0263	0.0513	0.3266	0.3341	0.2127	0.0323	0.0076	0.0068	0.0217	0.0685	0.1204	0.0252
31	4.3990	-0.0263	0.0513	0.3266	0.3341	0.2128	0.0324	0.0076	0.0068	0.0217	0.0686	0.1204	0.0252
41	4.3990	-0.0263	0.0513	0.3266	0.3341	0.2128	0.0324	0.0076	0.0068	0.0217	0.0686	0.1204	0.0252
51	4.3990	-0.0263	0.0513	0.3266	0.3341	0.2128	0.0324	0.0076	0.0068	0.0217	0.0686	0.1204	0.0252
61	4.3990	-0.0263	0.0513	0.3266	0.3341	0.2128	0.0324	0.0076	0.0068	0.0217	0.0686	0.1204	0.0252
71	4.3990	-0.0263	0.0513	0.3266	0.3341	0.2128	0.0324	0.0076	0.0068	0.0217	0.0686	0.1204	0.0252
81	4.3990	-0.0263	0.0513	0.3266	0.3341	0.2128	0.0324	0.0076	0.0068	0.0217	0.0686	0.1204	0.0252
91	4.3990	-0.0263	0.0513	0.3266	0.3341	0.2128	0.0324	0.0076	0.0068	0.0217	0.0686	0.1204	0.0252
101	4.3990	-0.0263	0.0513	0.3266	0.3341	0.2128	0.0324	0.0076	0.0068	0.0217	0.0686	0.1204	0.0252
111	4.3990	-0.0263	0.0513	0.3266	0.3341	0.2128	0.0324	0.0076	0.0068	0.0217	0.0686	0.1204	0.0252

Tabela 4.5: Convergência do algoritmo EM considerando 10 graus de liberdade

Iter	μ_x	α_2	α_3	α_4	α_5	α_6	ϕ_x	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6
1	4.3991	-0.0264	0.0402	0.3189	0.3253	0.2156	0.0361	0.0102	0.0012	0.0298	0.0849	0.1411	0.0244
11	4.3976	-0.0276	0.0465	0.3259	0.3291	0.2148	0.0332	0.0082	0.0054	0.0231	0.0716	0.1248	0.0252
21	4.3971	-0.0274	0.0460	0.3256	0.3285	0.2158	0.0334	0.0073	0.0067	0.0226	0.0699	0.1234	0.0269
31	4.3970	-0.0273	0.0460	0.3256	0.3284	0.2159	0.0334	0.0073	0.0068	0.0226	0.0698	0.1233	0.0270
41	4.3970	-0.0273	0.0460	0.3256	0.3284	0.2159	0.0334	0.0073	0.0068	0.0226	0.0698	0.1233	0.0270
51	4.3970	-0.0273	0.0460	0.3256	0.3284	0.2159	0.0334	0.0073	0.0068	0.0226	0.0698	0.1233	0.0270
61	4.3970	-0.0273	0.0460	0.3256	0.3284	0.2159	0.0334	0.0073	0.0068	0.0226	0.0698	0.1233	0.0270
71	4.3970	-0.0273	0.0460	0.3256	0.3284	0.2159	0.0334	0.0073	0.0068	0.0226	0.0698	0.1233	0.0270
81	4.3970	-0.0273	0.0460	0.3256	0.3284	0.2159	0.0334	0.0073	0.0068	0.0226	0.0698	0.1233	0.0270
91	4.3970	-0.0273	0.0460	0.3256	0.3284	0.2159	0.0334	0.0073	0.0068	0.0226	0.0698	0.1233	0.0270
101	4.3970	-0.0273	0.0460	0.3256	0.3284	0.2159	0.0334	0.0073	0.0068	0.0226	0.0698	0.1233	0.0270
111	4.3970	-0.0273	0.0460	0.3256	0.3284	0.2159	0.0334	0.0073	0.0068	0.0226	0.0698	0.1233	0.0270

Tabela 4.6: Convergência do algoritmo EM considerando 100 graus de liberdade

Iter	μ_x	α_2	α_3	α_4	α_5	α_6	ϕ_x	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6
1	4.3975	-0.0269	0.0388	0.3189	0.3229	0.2215	0.0375	0.0108	0.0011	0.0306	0.0867	0.1470	0.0264
11	4.3970	-0.0271	0.0398	0.3204	0.3237	0.2215	0.0356	0.0092	0.0044	0.0267	0.0795	0.1367	0.0271
21	4.3969	-0.0271	0.0396	0.3201	0.3236	0.2218	0.0356	0.0072	0.0071	0.0247	0.0748	0.1306	0.0299
31	4.3969	-0.0271	0.0396	0.3201	0.3236	0.2218	0.0357	0.0069	0.0074	0.0246	0.0744	0.1301	0.0302
41	4.3969	-0.0271	0.0396	0.3201	0.3236	0.2218	0.0357	0.0069	0.0074	0.0246	0.0744	0.1301	0.0302
51	4.3969	-0.0271	0.0396	0.3201	0.3236	0.2218	0.0357	0.0069	0.0074	0.0246	0.0744	0.1301	0.0302
61	4.3969	-0.0271	0.0396	0.3201	0.3236	0.2218	0.0357	0.0069	0.0074	0.0246	0.0744	0.1301	0.0302
71	4.3969	-0.0271	0.0396	0.3201	0.3236	0.2218	0.0357	0.0069	0.0074	0.0246	0.0744	0.1301	0.0302
81	4.3969	-0.0271	0.0396	0.3201	0.3236	0.2218	0.0357	0.0069	0.0074	0.0246	0.0744	0.1301	0.0302
91	4.3969	-0.0271	0.0396	0.3201	0.3236	0.2218	0.0357	0.0069	0.0074	0.0246	0.0744	0.1301	0.0302
101	4.3969	-0.0271	0.0396	0.3201	0.3236	0.2218	0.0357	0.0069	0.0074	0.0246	0.0744	0.1301	0.0302
111	4.3969	-0.0271	0.0396	0.3201	0.3236	0.2218	0.0357	0.0069	0.0074	0.0246	0.0744	0.1301	0.0302

É também interessante notar que, à medida que ν vai se aproximando de infinito, as estimativas de μ_x e dos vícios aditivos $\alpha_2, \ldots, \alpha_6$ vão convergindo cada vez mais rápido em relação às estimativas de ϕ_x e $\phi_1, \ldots, \phi_6$. Quando $\nu = 57779$, a convergência de μ_x e dos vícios aditivos ocorre na primeira iteração.

Tabela 4.7: Convergência do algoritmo EM considerando 10000 graus de liberdade

Iter	μ_x	α_2	α_3	α_4	α_5	α_6	ϕ_x	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6
1	4.3972	-0.0270	0.0386	0.3188	0.3230	0.2228	0.0380	0.0109	0.0011	0.0308	0.0874	0.1487	0.0269
11	4.3972	-0.0270	0.0386	0.3189	0.3230	0.2228	0.0361	0.0095	0.0042	0.0275	0.0812	0.1394	0.0274
21	4.3972	-0.0270	0.0386	0.3189	0.3230	0.2228	0.0360	0.0072	0.0073	0.0251	0.0759	0.1319	0.0304
31	4.3972	-0.0270	0.0386	0.3189	0.3230	0.2228	0.0361	0.0069	0.0076	0.0249	0.0754	0.1313	0.0308
41	4.3972	-0.0270	0.0386	0.3189	0.3230	0.2228	0.0361	0.0069	0.0076	0.0249	0.0753	0.1312	0.0309
51	4.3972	-0.0270	0.0386	0.3189	0.3230	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1312	0.0309
61	4.3972	-0.0270	0.0386	0.3189	0.3230	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1312	0.0309
71	4.3972	-0.0270	0.0386	0.3189	0.3230	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1312	0.0309
81	4.3972	-0.0270	0.0386	0.3189	0.3230	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1312	0.0309
91	4.3972	-0.0270	0.0386	0.3189	0.3230	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1312	0.0309
101	4.3972	-0.0270	0.0386	0.3189	0.3230	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1312	0.0309
111	4.3972	-0.0270	0.0386	0.3189	0.3230	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1312	0.0309

Tabela 4.8: Convergência do algoritmo EM considerando 57779 graus de liberdade

Iter	μ_x	α_2	α_3	α_4	α_5	α_6	ϕ_x	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6
1	4.3972	-0.0270	0.0386	0.3188	0.3230	0.2228	0.0380	0.0109	0.0011	0.0308	0.0874	0.1487	0.0269
11	4.3972	-0.0270	0.0386	0.3188	0.3230	0.2228	0.0361	0.0095	0.0042	0.0275	0.0813	0.1394	0.0274
21	4.3972	-0.0270	0.0386	0.3188	0.3230	0.2228	0.0360	0.0072	0.0073	0.0251	0.0759	0.1319	0.0304
31	4.3972	-0.0270	0.0386	0.3188	0.3230	0.2228	0.0361	0.0069	0.0076	0.0249	0.0754	0.1313	0.0308
41	4.3972	-0.0270	0.0386	0.3188	0.3230	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1313	0.0309
51	4.3972	-0.0270	0.0386	0.3188	0.3230	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1313	0.0309
61	4.3972	-0.0270	0.0386	0.3188	0.3230	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1313	0.0309
71	4.3972	-0.0270	0.0386	0.3188	0.3230	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1313	0.0309
81	4.3972	-0.0270	0.0386	0.3188	0.3230	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1313	0.0309
91	4.3972	-0.0270	0.0386	0.3188	0.3230	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1313	0.0309
101	4.3972	-0.0270	0.0386	0.3188	0.3230	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1313	0.0309
111	4.3972	-0.0270	0.0386	0.3188	0.3230	0.2228	0.0361	0.0068	0.0076	0.0249	0.0753	0.1313	0.0309

4.2.2 Exemplo 2

Consideremos os seguintes dados apresentados em Grubbs (1948), que representa os tempos de queima de fusíveis ("fuse-burning time"), em segundos, que são reportados por três observadores A, B e C. O tempo é medido através de três relógios elétricos que disparam automaticamente no início do processo, sendo cada um dos relógios acompanhado por um observador. Estes três observadores acompanham o processo simultaneamente e interrompem o relógio quando percebem que o fusível queima.

As observações y_{ij}, com $i = 1, 2, 3$ e $j = 1, \cdots, 29$ são apresentados na Tabela 4.9. Como considerado em Lachos-Dávila (2002), vamos supor que o observador B é o mais experiente, isto é, os dados referentes ao observador B serão considerados como os dados provenientes do instrumento de referência. Sendo assim e sem perda de generalidade consideremos $\alpha_1 = 0$.

Para ilustrar os resultados teóricos do capítulo anterior, vamos supor um modelo de

Tabela 4.9: Medidas de tempo de queima de fusíveis, em segundos

Observação	Observador A	Observador B	Observador C
1	10.10	10.07	10.07
2	9.98	9.90	9.90
3	9.89	9.85	9.86
4	9.79	9.71	9.70
5	9.67	9.65	9.65
6	9.89	9.83	9.83
7	9.82	9.75	9.79
8	9.59	9.56	9.59
9	9.76	9.68	9.72
10	9.93	9.89	9.92
11	9.62	9.61	9.64
12	10.24	10.23	10.24
13	9.84	9.83	9.86
14	9.62	9.58	9.63
15	9.60	9.60	9.65
16	9.74	9.73	9.74
17	10.32	10.32	10.34
18	9.86	9.86	9.86
19	9.65	9.64	9.65
20	9.50	9.49	9.50
21	9.56	9.56	9.55
22	9.54	9.53	9.54
23	9.89	9.89	9.88
24	9.53	9.52	9.51
25	9.52	9.52	9.53
26	9.44	9.43	9.45
27	9.67	9.67	9.67
28	9.77	9.76	9.78
29	9.86	9.84	9.86

Grubbs t-elíptico com $p = 3$ e $n = 29$. Este modelo tem uma vantagem em relação ao modelo normal que é dada pela sua generalidade, isto é, podemos absorver uma ampla classe de distribuições escolhendo apropriadamente os graus de liberdade, permitindo o ajuste da curtose da distribuição. Contudo, esta vantagem tem um custo que é a incorporação de um parâmetro adicional (ν), ou seja, temos mais um parâmetro para ser estimado.

Este parâmetro pode ser fixado a priori; Lange *et al.* (1989) e Berkane *et al.* (1994) recomendam $\nu = 4$, ou estimado desde os dados. Ver, também, Taylor (1992).

Galea-Rojas (1996) apresenta uma equação que deve ser resolvida numericamente dentro da etapa M do algoritmo EM para a obtenção do ν, caso ele não seja fixado a priori. Entretanto isto faz com que a convergência do algoritmo seja mais lenta.

Alternativamente, como sugerido por Lange *et al.* (1989), uma estimativa de ν pode ser obtida calculando o logaritmo da função de verossimilhança observada, $\ell(\boldsymbol{\theta})$, sobre uma tabela de valores de ν, e escolher o valor que maximiza $\ell(\boldsymbol{\theta})$.

Arellano (1995) apresenta um estimador consistente do tipo momentos quando $\nu > 4$. Este estimador é dado por

$$\hat{\nu}_M = 4\frac{\xi_n - 2}{\xi_n - 4},$$

em que $\xi_n = \frac{1}{n}\sum_{j=1}^{n}[(\mathbf{Y}_j - \bar{\mathbf{Y}})^\top \mathbf{S}_n^{-1}(\mathbf{Y}_j - \bar{\mathbf{Y}})]^2$, $\bar{\mathbf{Y}}$ e $\mathbf{S}_n$ são o vetor de médias e matriz de covariâncias amostrais, respectivamente.

Neste trabalho seguiremos Galea-Rojas (1996) utilizando a aproximação de Wilson-Hilferty (Johnson e Kotz, 1970), isto é,

$$z_j = \frac{\left(1 - \frac{2}{9\nu}\right) F_j^{1/3} - \left(1 - \frac{2}{9p}\right)}{\left[\left(\frac{2}{9\nu}\right) F_j^{2/3} - \left(\frac{2}{9p}\right)\right]^{1/2}}, \tag{4.2}$$

$j = 1, \cdots, n$, tem aproximadamente uma distribuição normal padrão se F_j tem uma distribuição F com p e ν graus de liberdade. De fato, sabemos que (ver Apêndice B) se $d_j^2 = (\mathbf{Y}_j - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y}_j - \boldsymbol{\mu})$ e Y_j como em (2.13), então d_j^2/p tem distribuição F com p e ν graus de liberdade com $j = 1, \cdots, n$. Ademais, $F_j = \hat{d}_j^2/p$ tem assintoticamente a mesma distribuição F, em que $\hat{d}_j^2 = d_j(\hat{\boldsymbol{\theta}}, \hat{\nu}), j = 1, \cdots, n$. Assim, um teste de normalidade encontrado em vários pacotes estatísticos, das distâncias transformadas $z_j, j = 1, \cdots, n$, pode ser usado para avaliar o ajuste da distribuição t multivariada. Contudo, exemplificaremos, também, alguns dos outros métodos acima com a finalidade de compará-los.

A Tabela 4.10 mostra as estimativas de máxima verossimilhança dos parâmetros do modelo de Grubbs t-Student, para diferentes graus de liberdade ($\nu = 1, 4, 4.7464, 5, 10, 50, 100$ e ∞, sendo este último equivalente ao caso normal).

As estimativas da Tabela 4.10 são obtidos via algoritmo EM e os estimadores do tipo momentos apresentados na Seção 2.8 são utilizados como valores iniciais.

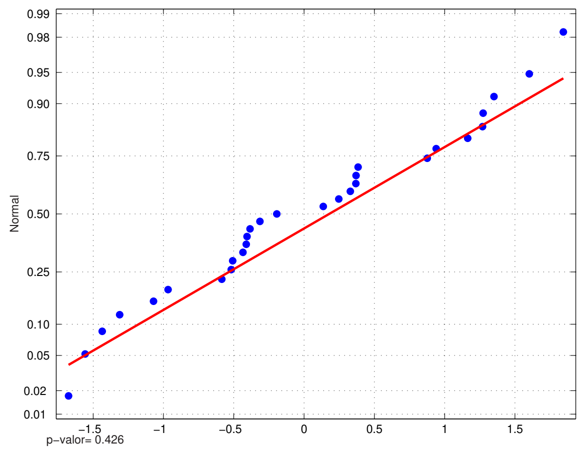
Prosseguindo a análise utilizando a aproximação de Wilson-Hilferty, temos conforme a Figura 4.2.2, que não existem evidências para a rejeição da hipótese de normalidade em todos os casos, ou seja, esta análise justificaria a escolha do modelo normal. Entretanto, note que quando se considera o modelo t-Student com poucos graus de liberdade, as evidências para se rejeitar a hipótese de normalidade são muito menores. Isso sugere, portanto, que o modelo que melhor se ajusta a estes dados é o modelo de Grubbs t-Student com 3 graus de liberdade conforme mostra a Tabela 4.11. No entanto, a literatura recomenda considerar, sempre que possível, distribuições que tenham $\nu > 4$, isto é,

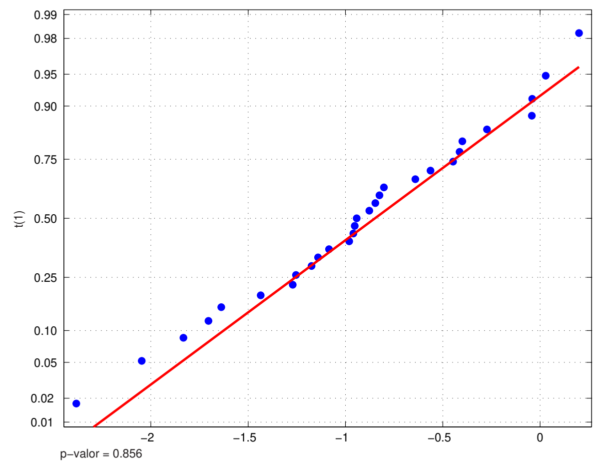
Tabela 4.10: Estimativas de máxima verossimilhança dos parâmetros do modelo de Grubbs

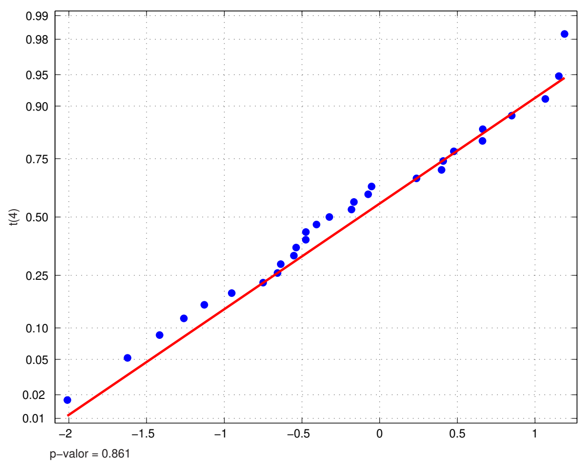
Parâmetro	Normal	t(1)	t(4)	t(4,7464)	t(5)	t(10)	t(50)	t(100)
μ_x	9,7414	9,7019	9,7148	9,7173	9,7180	9,7271	9,7380	9,7397
α_2	0,0238	0,0187	0,0196	0,0200	0,0201	0,0216	0,0233	0,0235
α_3	0,0141	0,0138	0,0137	0,0137	0,0137	0,0139	0,0141	0,0141
ϕ_x	0,0434	0,0514	0,0355	0,0357	0,0358	0,0379	0,0419	0,0426
ϕ_1	0,0001	0,0007	0,0002	0,0002	0,0002	0,0001	0,0001	0,0001
ϕ_2	0,0006	0,0009	0,0004	0,0004	0,0004	0,0005	0,0006	0,0006
ϕ_3	0,0002	0,0007	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002

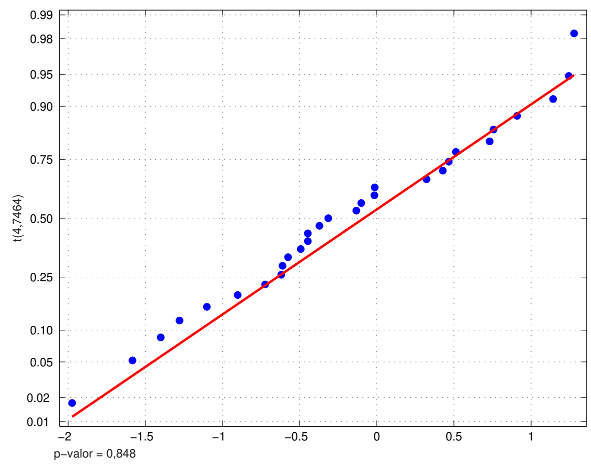
distribuições com o quarto momento finito, o que não ocorre com a t-Student com 3 graus de liberdade. Além disso, supondo que $\nu > 4$, temos que a estimativa do tipo momentos proposto por Arellano (1994) é $\hat{\nu}_M = 4,7464$. Portanto, utilizaremos esta estimativa em nossa análise.

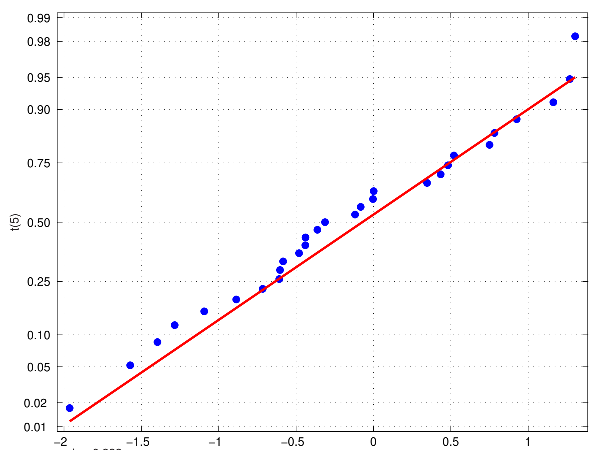
Gostaríamos de ressaltar aqui que não se encontra na literatura uma regra clássica para se estimar os graus de liberdade, de modo que os métodos apresentados acima são apenas sugestões.

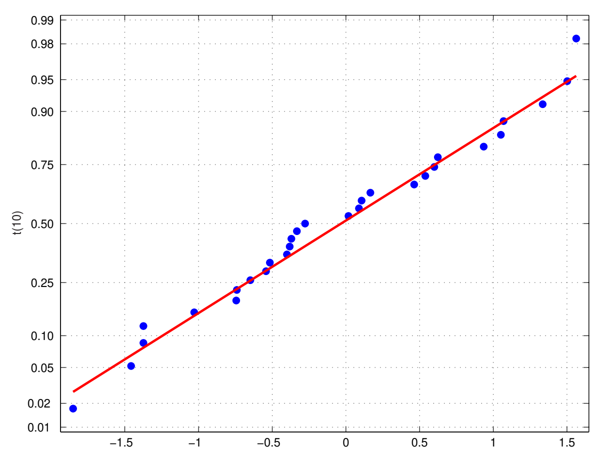

Normal
p−valor= 0.426


t(1)
p−valor = 0.856


t(4)
p−valor = 0.861


t(4,7464)
p−valor = 0,848


t(5)
p−valor=0,838


t(10)
p−valor = 0,509

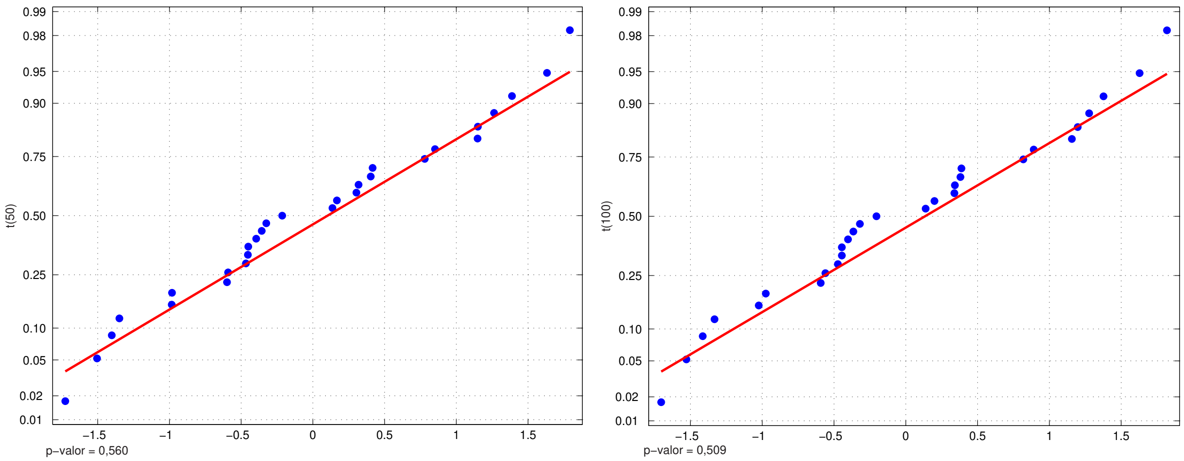


Figura 4.1: Teste de Normalidade para a Aproximação de Wilson-Hilferty

Tabela 4.11: Teste de Normalidade para a Aproximação de Wilson-Hilferty

Graus de Liberdade	1	2	3	4	4,7464
p-valor	0,856	0,821	0,865	0,861	0,848
Graus de Liberdade	6	7	8	9	10
p-valor	0,808	0,781	0,764	0,756	0,747

Antes de prosseguir com nosso exemplo, ilustraremos o seguinte fato: ao utilizarmos a sugestão de Lange et al. (1989), ou seja , fazer uma tabela com diferentes graus de liberdade e valores da log-verossimilhança para seus respectivos EMV, podemos notar que o método não consegue detectar as evidências a favor do modelo t, pois os valores de $\ell(\boldsymbol{\theta})$ aumentam conforme vai aumentando os graus de liberdade, indicando assim, que o modelo normal é o mais adequado. Veja Tabela 4.12.

Contudo, para mostrar a sensibilidade do algoritmo sob observações extremas, as três medições correspondentes ao fuzível número 1 foram substituídas pelo valor atípico 20. A Tabela 4.13 mostra agora que uma estimativa dos graus de liberdade seria $\hat{\nu} = 3$.

Podemos notar ainda, através da Tabela 4.14, que as estimativas para o ϕ_x são bem diferentes, isto é, no modelo normal esta estimativa é muito maior comparada a da Tabela 4.10, ao contrário do modelo t-Student com 3 graus de liberdade.

Considerando novamente o conjunto de dados da Tabela 4.1, continuaremos nossa

Tabela 4.12: Graus de liberdade × log-verossimilhança

ν	1	2	3	4	5
$\ell(\hat{\boldsymbol{\theta}})$	121.8933	136.2572	140.4599	142.2504	143.1885
ν	7	10	15	30	100
$\ell(\hat{\boldsymbol{\theta}})$	144.1121	144.6903	145.0712	145.3984	145.5992
ν	1000	10000	50000	100000	1000000
$\ell(\hat{\boldsymbol{\theta}})$	145.6621	145.6668	145.6672	145.6673	145.6673

Tabela 4.13: Graus de liberdade × log-verossimilhança para dados pertubados

ν	1	2	3	4	5	6	7
$\ell(\hat{\boldsymbol{\theta}})$	109,7052	121,2791	122,8023	122,0690	120,6213	111,7147	83,8389

Tabela 4.14: Estimativas de máxima verossimilhança para os dados pertubados

Parâmetro	Normal	t(3)
μ_x	10,0837	9,7012
α_2	0,0228	0,0184
α_3	0,0141	0,0140
ϕ_x	3,5469	0,0432
ϕ_1	0,0001	0,0002
ϕ_2	0,0006	0,0004
ϕ_3	0,0002	0,0002

análise assumindo o modelo t-Student com 4,7464 graus de liberdade e EMV dados pela Tabela 4.10.

Segue, do Teorema 2.2, que a matriz de informação esperada de Fisher é dada por

$$\mathbf{K}_F(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \hat{\mathbf{K}}_L & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{K}}_E \end{pmatrix},$$

em que

$$\hat{\mathbf{K}}_L = 10^3 \times \begin{pmatrix} 0,0222 & 0,0038 & 0,0084 \\ 0,0038 & 1,4725 & -0,6673 \\ 0,0084 & -0,6673 & 2,4173 \end{pmatrix}$$

e

$$\hat{\mathbf{K}}_E = 10^6 \times \begin{pmatrix} 0,0003 & -0,0045 & -0,0026 & -0,0043 \\ -0,0045 & 3,5714 & 0,0987 & 1,4072 \\ -0,0026 & 0,0987 & 1,1880 & -0,0090 \\ -0,0043 & 1,4072 & -0,0090 & 3,2016 \end{pmatrix}.$$

Para calcular as estatísticas de Escore e Razão de Verossimilhanças será necessário calcular os EMV sob as hipóteses nulas H_{01}, H_{02} e H_{03}. Assim, dos resultados da Seção 2.5, segue que: sob H_{01}, as estimativas das componentes de $\boldsymbol{\theta} = (\mu_x, \phi_x, \phi)^\top$ são

$$\tilde{\mu_x} = 9,7224,\ \tilde{\phi_x} = 0,0361 \text{ e } \tilde{\phi} = 0,0004.$$

Sob H_{02}, as estimativas das componentes de $\boldsymbol{\theta} = (\mu_x, \phi_x, \phi_1, \phi_2, \phi_3)^\top$ são

$$\tilde{\mu_x} = 9,7243,\ \tilde{\phi_x} = 0,0358,\ \tilde{\phi_1} = 0,0004,\ \tilde{\phi_2} = 0,0005 \text{ e } \tilde{\phi_3} = 0,0002.$$

Finalmente, sob H_{03}, as estimativas das componentes de $\boldsymbol{\theta} = (\mu_x, \alpha_2, \alpha_3, \phi_x, \phi)^\top$ são

$$\tilde{\mu_x} = 9,7146,\ \tilde{\alpha_2} = 0,0193,\ \tilde{\alpha_3} = 0,0138,\ \tilde{\phi_x} = 0,0360 \text{ e } \tilde{\phi} = 0,0003.$$

Usando estes resultados e as estatísticas de Wald (W), Escore (E) e Razão de Verossimilhanças (Q) dadas na Seção 2.6, podemos testar hipóteses de interesse H_{01}, H_{02} e H_{03} definidas no Capítulo 1. Assim, de acordo com a Tabela 4.15 e: 4 graus de liberdade para testar H_{01}; 2 graus de liberdade para H_{02}; 2 graus de liberdade para testar H_{03}, as hipóteses H_{01} e H_{02} são rejeitadas a um nível de significância de 5% sob o modelo t-Student. Entretanto, para o mesmo nível de significância, não há evidências para se rejeitar H_{03}. Portanto, podemos concluir que os instrumentos (relógio-observador) são viciados, isto é, medem o tempo de queima dos fuzíveis com vício, contudo não podemos concluir que eles não são igualmente precisos.

Se considerarmos o modelo normal, isto é, $\nu \to \infty$, então teremos que, para todas as estatísticas, todas as hipóteses nulas serão rejeitadas com 5% de significância. Os valores das estatísticas de Wald, Escore e Razão de Verossimilhança para este modelo também estão na Tabela 4.15.

Tabela 4.15: Estatísticas de Wald, Escore e Razão de Verossimilhanças para os modelos normal e t(4,7464)

	Normal			t(4,7464)		
	W	E	Q	W	E	Q
H_{01}	44,7483	25,5374	32,2895	21,6473	21,9626	25,4330
H_{02}	38,2356	16,4022	24,3051	19,6295	20,7828	22,2688
H_{03}	6,5127	8,6791	10,0375	2,0178	3,7608	4,3981

Finalmente, na Tabela 4.16, apresentamos as estimativas das confiabilidades definidas em (1.10), para os modelos normal e t-Student, com 4,7464 graus de liberdade.

Tabela 4.16: Estimativas das confiabilidades

	Normal	t(4.7464)
$\hat{\rho}_1$	0,9986	0,9952
$\hat{\rho}_2$	0,9860	0,9876
$\hat{\rho}_3$	0,9946	0,9943

Observe que os valores da Tabela 4.16, indicam que a proporção da variação observada "explicada"por instrumentos são muito próximas para os dois modelos, contudo o modelo normal rejeita a hipótese de igualdade das variâncias dos instrumentos que, de acordo com (1.11), é equivalente a igualdade das confiabilidades.

4.3 Estudo de Simulação dos Testes de Wald, Escore e Razão de Verossimilhanças

Nesta seção apresentamos um pequeno estudo de simulação, no qual o poder e o tamanho dos três testes (Wald, Escore e Razão de Verossimilhanças) para a hipótese H_{01} de igualdade de variâncias e vícios são considerados. Neste estudo utilizamos diferentes graus de liberdade e dois números de instrumentos diferentes: $p = 3$ e $p = 5$.

Para cada estimação Monte Carlo do poder dos testes foram geradas 1000 amostras de acordo com o modelo (2.12)-(2.13), com tamanhos $n = 25, 50, 100$ e 200, considerando H_{a1}: $\alpha_2 = 0, \alpha_3 = 0.5, \phi_1 = \phi_2 = 1$ e $\phi_3 = 1.5$ para $p = 3$ e $H_{a1} : \alpha_2 = \alpha_3 = \alpha_4 = 0, \alpha_5 = 0.5, \phi_1 = \cdots = \phi_4 = 1$ e $\phi_5 = 1.5$ para $p = 5$ como hipóteses alternativas. Nas Tabelas 4.17-4.26 apresentamos o poder dos testes estimado, em porcentagen, correspondente ao nível nominal de 5%. Em todos os casos a característica de interesse $x_j, j = 1, \cdots, n$, foi gerada conforme a distribuição dos erros, com $\phi_x = 0.01, 0.25, 1$ e 2.

Sob o critério de poder do teste, podemos notar nas Tabelas 4.17-4.22 que para valores pequenos de ν e ϕ_x, entre os valores testados, o teste de Escore apresenta taxas de rejeição maiores que os testes de Wald e Razão de Verossimilhanças. Para graus de liberdade pequenos e $\phi_x = 1$ ou $\phi_x = 2$, as estatísticas de Escore e Razão de Verossimilhanças apresentam valores mais altos para os poderes estimados, sendo que a ordem delas se altera para diferentes tamanhos de amostra.

Tabela 4.17: Poder estimado dos três testes para $p = 3$ e hipótese alternativa $H_{a1} : \alpha_2 = 0, \alpha_3 = 0.5, \phi_1 = \phi_2 = 1, \phi_3 = 1.5$ e $\nu = 4$

$\nu = 4$ e $p = 3$												
x_j	$t(10, 0.01; 4)$			$t(10, 0.25; 4)$			$t(10, 1; 4)$			$t(10, 2; 4)$		
n	W	E	Q	W	E	Q	W	E	Q	W	E	Q
25	17.6	52.5	24.8	9.1	32.0	21.2	4.1	16.0	17.6	3.8	10.5	14.0
50	25.9	92.6	43.1	15.4	61.2	38.2	11.8	30.3	32.6	8.1	23.7	26.5
100	48.7	98.9	68.9	39.4	73.5	66.2	28.5	61.6	56.5	21.7	51.4	52.7
200	86.9	100	91.9	80.8	81.9	92.6	67.1	92.7	87.4	56.7	85.1	83.3

Tabela 4.18: Poder estimado dos três testes para $p = 5$ e hipótese alternativa $H_{a1} : \alpha_2 = \alpha_3 = \alpha_4 = 0, \alpha_5 = 0.5, \phi_1 = \cdots = \phi_4 = 1, \phi_5 = 1.5$ e $\nu = 4$

$\nu = 4$ e $p = 5$												
x_j	$t(10, 0.01; 4)$			$t(10, 0.25; 4)$			$t(10, 1; 4)$			$t(10, 2; 4)$		
n	W	E	Q	W	E	Q	W	E	Q	W	E	Q
25	22.8	30.9	19.4	13.4	23.8	21.0	7.0	14.7	16.9	6.5	16.6	18.5
50	29.7	59.5	37.1	18.3	46.8	35.7	16.3	35.3	35.8	14.6	32.0	31.9
100	48.5	92.1	62.5	43.0	76.2	65.8	37.5	63.8	63.1	35.7	64.3	63.7
200	87.3	100	93.0	85.1	87.3	93.2	81.0	94.0	93.3	81.1	94.1	93.2

Tabela 4.19: Poder estimado dos três testes para $p = 3$ e hipótese alternativa $H_{a1} : \alpha_2 = 0, \alpha_3 = 0.5, \phi_1 = \phi_2 = 1, \phi_3 = 1.5$ e $\nu = 6$

$\nu = 6$ e $p = 3$												
x_j	$t(10, 0.01; 6)$			$t(10, 0.25; 6)$			$t(10, 1; 6)$			$t(10, 2; 6)$		
n	W	E	Q	W	E	Q	W	E	Q	W	E	Q
25	23.0	46.9	29.8	15.7	31.6	26.7	11.3	19.4	22.7	9.5	15.2	19.4
50	39.7	80.4	49.3	32.2	62.3	46.6	23.8	39.4	42.1	18.2	32.0	35.3
100	71.6	99.4	77.9	64.0	92.4	77.5	49.8	67.8	69.1	46.7	67.7	69.7
200	97.5	99.8	93.2	94.1	99.3	97.6	89.6	96.9	95.9	87.0	95.6	95.9

Tabela 4.20: Poder estimado dos três testes para $p = 5$ e hipótese alternativa $H_{a1} : \alpha_2 = \alpha_3 = \alpha_4 = 0, \alpha_5 = 0.5, \phi_1 = \cdots = \phi_4 = 1, \phi_5 = 1.5$ e $\nu = 6$

$\nu = 6$ e $p = 5$												
x_j	$t(10, 0.01; 6)$			$t(10, 0.25; 6)$			$t(10, 1; 6)$			$t(10, 2; 6)$		
n	W	E	Q	W	E	Q	W	E	Q	W	E	Q
25	27.5	32.4	24.1	17.4	28.5	25.8	13.7	23.4	24.8	11.5	19.7	23.0
50	37.5	59.8	40.7	28.9	51.1	45.0	22.8	40.2	39.8	24.9	40.5	42.0
100	66.3	91.6	74.9	62.9	85.8	77.4	55.7	75.6	74.0	55.8	74.1	74.3
200	96.0	99.9	97.7	92.6	98.4	98.4	93.3	97.8	97.4	92.7	97.7	97.5

Nas Tabelas 4.23-4.26, notamos que o teste de Escore quase sempre apresenta um poder maior que os demais testes. Deste modo, podemos concluir sob o critério de poder

Tabela 4.21: Poder estimado dos três testes para $p = 3$ e hipótese alternativa $H_{a1} : \alpha_2 = 0, \alpha_3 = 0.5, \phi_1 = \phi_2 = 1, \phi_3 = 1.5$ e $\nu = 10$

$\nu = 10$ e $p = 3$												
x_j	$t(10, 0.01; 10)$			$t(10, 0.25; 10)$			$t(10, 1; 10)$			$t(10, 2; 10)$		
n	W	E	Q	W	E	Q	W	E	Q	W	E	Q
25	43.0	49.5	36.2	27.8	34.7	32.4	17.4	21.4	26.4	14.5	19.6	24.8
50	67.2	80.8	54.4	45.1	59.5	53.7	36.4	45.5	48.0	33.7	41.9	46.2
100	88.5	98.5	67.4	78.8	92.0	83.7	69.4	77.1	79.9	67.3	76.9	78.2
200	99.3	99.8	76.3	98.2	99.9	98.9	96.5	98.1	98.1	96.3	98.2	98.6

Tabela 4.22: Poder estimado dos três testes para $p = 5$ e hipótese alternativa $H_{a1} : \alpha_2 = \alpha_3 = \alpha_4 = 0, \alpha_5 = 0.5, \phi_1 = \cdots = \phi_4 = 1, \phi_5 = 1.5$ e $\nu = 10$

$\nu = 10$ e $p = 5$												
x_j	$t(10, 0.01; 10)$			$t(10, 0.25; 10)$			$t(10, 1; 10)$			$t(10, 2; 10)$		
n	W	E	Q	W	E	Q	W	E	Q	W	E	Q
25	34.4	35.7	29.6	22.8	30.2	29.5	19.0	25.1	27.4	18.4	23.4	24.9
50	47.9	65.5	51.6	36.2	56.8	52.3	33.0	50.6	48.9	35.4	49.8	48.8
100	80.8	94.9	87.6	72.7	87.6	82.5	73.6	85.3	84.1	68.8	82.7	81.9
200	98.7	99.7	99.4	98.9	99.8	99.7	97.1	99.4	99.0	97.9	99.3	99.2

do teste, que a estatística de Escore é recomendável para modelos com graus de liberdade não tão pequenos. Nenhuma diferença é notada nos resultados quando variamos o número de instrumentos.

Ainda das Tabelas 4.23-4.26, podemos observar que para grandes valores de ν e valores moderados de n as estatísticas satisfazem a desigualdade

$$E \geqslant Q \geqslant W.$$

Sob normalidade, resultados semelhantes são encontrados em Lachos-Dávila (2002).

É interessante notar ainda que, em todos os casos, o poder estimado diminui quando se aumenta ϕ_x.

No estudo do tamanho dos testes, 1000 amostras independentes foram geradas de acordo com o modelo (2.12)-(2.13), para os casos $p = 3$ e $p = 5$ considerando, respecti-

Tabela 4.23: Poder estimado dos três testes para $p = 3$ e hipótese alternativa $H_{a1} : \alpha_2 = 0, \alpha_3 = 0.5, \phi_1 = \phi_2 = 1, \phi_3 = 1.5$ e $\nu = 100$

$\nu = 100$ e $p = 3$												
x_j	$t(10, 0.01; 100)$			$t(10, 0.25; 100)$			$t(10, 1; 100)$			$t(10, 2; 100)$		
n	W	E	Q	W	E	Q	W	E	Q	W	E	Q
25	54.2	69.9	70.4	34.6	42.9	45.3	34.4	31.4	36.8	32.2	29.1	33.4
50	71.5	85.7	84.7	60.7	67.0	66.5	60.9	64.4	64.3	56.9	62.7	62.2
100	93.2	97.9	97.1	91.8	93.3	93.0	89.5	91.9	91.9	87.3	89.7	89.6
200	99.9	100	100	99.9	99.9	99.9	99.7	99.7	99.7	99.6	99.8	99.7

Tabela 4.24: Poder estimado dos três testes para $p = 5$ e hipótese alternativa $H_{a1} : \alpha_2 = \alpha_3 = \alpha_4 = 0, \alpha_5 = 0.5, \phi_1 = \cdots = \phi_4 = 1, \phi_5 = 1.5$ e $\nu = 100$

$\nu = 100$ e $p = 5$												
x_j	$t(10, 0.01; 100)$			$t(10, 0.25; 100)$			$t(10, 1; 100)$			$t(10, 2; 100)$		
n	W	E	Q	W	E	Q	W	E	Q	W	E	Q
25	34.6	46.3	46.0	31.3	37.2	36.7	34.2	36.7	37.2	27.4	32.1	31.3
50	56.9	71.9	69.1	57.4	69.3	65.5	56.3	66.4	63.2	54.8	66.9	62.2
100	91.1	95.2	94.4	89.7	94.4	92.8	88.4	94.1	92.3	89.1	94.4	93.0
200	99.9	100	100	99.9	100	100	99.9	99.9	99.9	99.9	99.9	99.9

Tabela 4.25: Poder estimado dos três testes para $p = 3$ e hipótese alternativa $H_{a1} : \alpha_2 = 0, \alpha_3 = 0.5, \phi_1 = \phi_2 = 1, \phi_3 = 1.5$ e $\nu \to \infty$

Modelo Normal e $p = 3$												
x_j	$N(10, 0.01)$			$N(10, 0.25)$			$N(10, 1)$			$N(10, 2)$		
n	W	E	Q	W	E	Q	W	E	Q	W	E	Q
25	37.2	70.0	71.3	34.0	43.7	46.5	34.3	32.0	34.8	33.8	31.5	33.8
50	65.5	86.1	85.5	63.2	68.6	68.4	61.1	64.0	64.3	61.5	62.6	63.4
100	91.9	96.1	96.0	90.7	92.8	92.3	91.1	92.3	92.4	89.1	91.2	91.0
200	99.9	100	100	99.9	99.9	99.9	99.7	99.8	99.7	99.8	99.8	99.8

vamente, $\alpha_2 = \alpha_3 = 0$, $\phi_1 = \phi_2 = \phi_3 = 1$ e $\alpha_2 = \cdots = \alpha_5 = 0$, $\phi_1 = \cdots = \phi_5 = 1$. Tamanhos de amostras $n = 25, 35, 50$ e 100 foram considerados em ambos os casos.

Tabela 4.26: Poder estimado dos três testes para $p = 5$ e hipótese alternativa $H_{a1} : \alpha_2 = \alpha_3 = \alpha_4 = 0, \alpha_5 = 0.5, \phi_1 = \cdots = \phi_4 = 1, \phi_5 = 1.5$ e $\nu \to \infty$

Modelo Normal e $p = 5$												
x_j	$N(10, 0.01)$			$N(10, 0.25)$			$N(10, 1)$			$N(10, 2)$		
n	W	E	Q	W	E	Q	W	E	Q	W	E	Q
25	32.4	70.0	69.0	34.2	37.9	37.7	32.0	36.7	35.1	35.7	39.1	38.4
50	61.1	85.7	82.7	57.0	68.2	65.2	55.5	67.5	63.6	58.5	69.8	65.9
100	92.8	98.3	97.8	91.1	95.5	94.5	89.2	93.9	91.8	90.2	94.4	93.4
200	100	100	100	99.9	100	99.9	99.7	99.8	99.8	99.9	99.9	99.9

Nas Tabelas 4.27-4.29 apresentamos os tamanhos estimados, em porcentagens, correspondente ao nível nominal de 5%. Em todos os casos a característica de interesse $x_j, j = 1, \cdots, n$, foi gerada conforme a distribuição dos erros, com $\phi_x = 0.25$ e 1.

Para $\nu = 10$, podemos ver na Tabela 4.27, que a estatística de Razão de Verossimilhanças apresenta um tamanho mais próximo do nível nominal. Para graus de liberdade menores que dez, nossos estudos revelaram que a estatística da Razão de Verossimilhanças apresenta um bom comportamento em termos do tamanho do teste. Contudo, as estatísticas de Wald e Escore apresentaram uma certa instabilidade, a qual motivará futuros trabalhos.

Tabela 4.27: Tamanho dos testes de Wald (W), Escore (E) e Razão de Verossimilhanças (Q) com nível nominal de 5% e 10 graus de liberdade

$\nu = 10$							
	Tamanho da amostra	$p = 3$			$p = 5$		
x_j	n	W	E	Q	W	E	Q
$t(10, 0.25; 10)$	25	9.0	8.8	9.6	7.7	4.1	5.9
	35	7.2	7.2	7.4	5.8	5.4	5.9
	50	5.4	8.0	6.9	5.0	6.0	6.3
	100	4.5	9.2	6.2	5.5	8.0	5.7
$t(10, 1.0; 10)$	25	3.1	4.8	6.7	5.6	4.5	7.6
	35	2.9	3.8	5.8	4.0	5.1	6.2
	50	1.9	4.0	5.2	3.1	3.3	4.6
	100	2.7	4.3	6.2	3.8	3.2	4.6

Nas Tabelas 4.28-4.29 referindo-se respectivamente a $\nu = 100$ e $\nu \to \infty$, temos que em ambos os casos os testes de Wald e Razão de Verossimilhanças são anticonservativos, o que leva a concluir que o teste de Escore é melhor no contexto de minimizar o erro tipo I.

Tabela 4.28: Tamanho dos testes de Wald (W), Escore (E) e Razão de Verossimilhanças (Q) com nível nominal de 5% e 100 graus de liberdade

$\nu = 100$							
	Tamanho da amostra	$p = 3$			$p = 5$		
x_j	n	W	E	Q	W	E	Q
$t(10, 0.25; 100)$	25	8.2	10.5	12.8	8.1	5.0	7.4
	35	6.1	7.5	8.4	7.0	5.2	5.8
	50	7.0	7.3	7.9	7.8	4.9	7.1
	100	5.5	4.8	5.4	5.8	3.9	4.8
$t(10, 1.0; 100)$	25	8.1	4.3	7.7	11.6	5.3	8.3
	35	5.9	3.2	6.0	7.2	4.3	5.8
	50	5.1	3.9	5.3	6.4	4.0	5.6
	100	5.6	4.6	5.2	5.9	5.2	5.2

Tabela 4.29: Tamanho dos testes de Wald (W), Escore (E) e Razão de Verossimilhanças (Q) com nível nominal de 5% e $\nu \to \infty$ (Modelo Normal)

Modelo Normal							
	Tamanho da amostra	$p = 3$			$p = 5$		
x_j	n	W	E	Q	W	E	Q
$N(10, 0.25)$	25	5.1	9.6	11.1	8.5	6.4	8.1
	35	6.6	8.0	9.2	8.7	5.4	6.9
	50	6.4	6.2	7.2	6.3	5.1	6.3
	100	4.2	3.4	3.9	6.4	5.6	6.3
$N(10, 1.0)$	25	7.1	4.0	6.5	9.4	4.8	6.3
	35	7.2	5.6	6.8	7.0	3.9	4.9
	50	6.2	4.0	5.8	7.6	4.7	5.5
	100	6.0	4.0	5.3	6.2	4.5	5.1

Em todos os casos não parece haver mudança significativa dos testes com o aumento do número de instrumentos.

As Figuras 4.2-4.3 apresentam uma comparação das frequências acumuladas estimadas dos três testes estatísticos, sob H_{01} e $\nu = 10$, com as frequências acumuladas teóricas das correspondentes distribuições limites das estatíticas de Wald, Escore e Razão de Verossimilhança, isto é $\mathcal{X}_4^2$. Observe que mesmo para amostra de tamanhos pequenos as distribuições empíricas e teóricas tem formas bem parecidas.

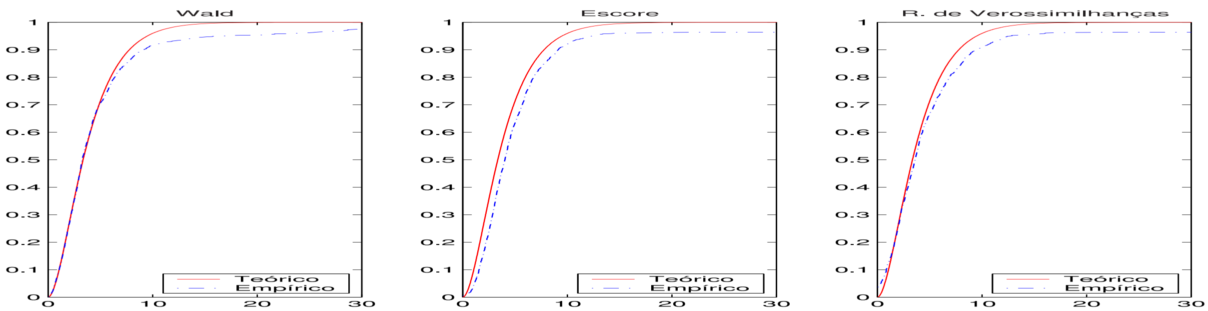


Figura 4.2: Distribuição empírica e teórica ($\mathcal{X}_4^2$) das estatisticas de Wald, Escore e Razão de Verossimilhanças para $n = 25$

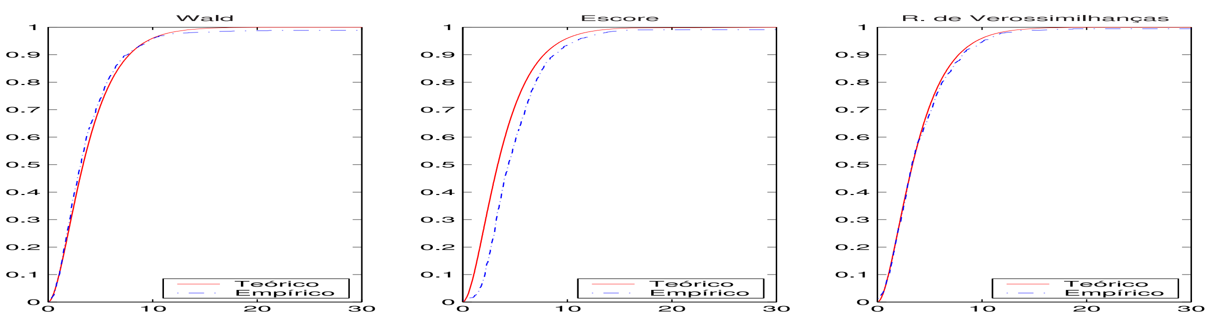


Figura 4.3: Distribuição empírica e teórica ($\mathcal{X}_4^2$) das estatisticas de Wald, Escore e Razão de Verossimilhanças para $n = 50$

4.4 Estudo de Simulação do Modelo com Ponto de Mudança

Nesta seção apresentamos como exemplos um estudo de simulação do modelo de Grubbs com ponto de mudança definido na Seção 3.1 considerando dois instrumentos de medição, isto é, $p = 2$.

Exemplo 4.1. *Um estudo de simulação considerando* $\mu_1 \neq \mu_2$

Neste exemplo foram geradas 1000 replicações de observações de diferentes tamanhos de amostra. Todas as variáveis envolvidas no modelo foram geradas a partir do seguinte experimento:

1. para cada um dos tamanhos de amostra, $n = 25, 50, 100$, fixar os valores de ponto de mudança $\lambda_0 = 0.1, 0.5, 0.8$, obtendo diferentes valores de $k_0 = [n\lambda_0]$;
2. gerar valores $x_1, \cdots, x_{k_0}$ de $x \sim N(\mu_1, \phi_x)$;
3. gerar valores $x_{k_0+1}, \cdots, x_n$ de $x \sim N(\mu_2, \phi_x)$;
4. gerar os erros de medição para cada k_0 considerado em (1), ou seja, gerar três conjuntos de valores $e_{i1}, \cdots, e_{ik_0}, e_{i(k_0+1)}, \cdots, e_{in}$ de $e_i \sim N(0, \phi_i)$, $i = 1, 2$;
5. calcular os dados observados a partir de:

$$\begin{pmatrix} Y_{1j} \\ Y_{2j} \end{pmatrix} = \begin{pmatrix} x_j + e_{1j} \\ \alpha_2 + x_j + e_{2j} \end{pmatrix}, j = 1, \cdots, k_0$$

e

$$\begin{pmatrix} Y_{1j} \\ Y_{2j} \end{pmatrix} = \begin{pmatrix} x_j + e_{1j} \\ \alpha_2 + x_j + e_{2j} \end{pmatrix}, j = k_0 + 1, \cdots, n.$$

Os parâmetros usados para gerar as observações foram os seguintes:

$$\mu_1 = 15, \mu_2 = 25, \alpha_2 = 2, \phi_1 = \phi_2 = 5$$

e dois valores para ϕ_x : 4 e 9.

O objetivo principal é estimar os parâmetros do modelo e calcular a partir destas estimativas o desvio padrão empírico e o erro quadrático médio empírico.

Temos:

i. $\hat{\theta}_{tl}$ a estimativa do componente t do vetor de parâmetros da amostra l, calculada segundo as expressões da Subseção 3.3;

ii. $EQM(\hat{\theta}_t)$ é o erro quadrático médio empírico da estimativa de $\hat{\theta}_t$ calculado como

$$EQM(\hat{\theta}_t) = \frac{1}{m}\sum_{l=1}^{m}(\hat{\theta}_{tl} - \theta_t)^2,$$

em que m é o número de amostras geradas.

Uma vez obtida a matriz de dados observados em cada amostra, calculamos a estatística do teste T^2, conforme expressão dada em (3.18), para todos os possíveis valores de k com o objetivo de calcular o $max_{\lambda}T^2_{\lambda}$. Este valor foi muito alto em todas as réplicas, portanto em todas as amostras rejeitamos H_0. A tabela contendo os valores críticos para este teste pode ser encontrada em Wang e Wang (1994).

Para que os valores dos desvios padrões (DP) e dos erros quadráticos médios empíricos das estimativas sejam comparáveis, obtivemos a raiz quadrada dos erros quadráticos médios (RQME).

Das Tabelas 4.30-4.35, podemos concluir que o estimador $\hat{\lambda}$ do ponto de mudança é muito próximo do verdadeiro valor em todos os casos. Não há mudança significativa das estimativas quando se muda os pontos de mudança, entretanto, há um aumento nos desvios padrões e nos erros quadráticos médios com um aumneto em ϕ_x. Perceba também, que os desvios padrões e os erros quadráticos médios diminuem com o aumento do tamanho das amostras.

Tabela 4.30: Média, desvio padrão e raiz do erro quadrático médio empírico das estimativas considerando $n = 25$ e $\phi_x = 4$

$n = 25, \phi_x = 4$									
	$\lambda = 0.1$			$\lambda = 0.5$			$\lambda = 0.8$		
	Média	DP	REQM	Média	DP	REQM	Média	DP	REQM
$\hat{\lambda}$	0.1264	0.0739	0.0785	0.5182	0.0153	0.0237	0.7621	0.1140	0.1201
$\hat{\mu}_1$	14.9927	1.9288	1.9278	14.9629	0.8442	0.8446	14.9462	0.8945	0.8957
$\hat{\mu}_2$	25.0079	0.6796	0.6793	25.0200	0.8973	0.8970	24.4364	2.1321	2.2043
$\hat{\alpha}_2$	1.9999	0.9857	0.9852	2.0087	0.6699	0.6696	1.9598	0.8710	0.8715
$\hat{\phi}_x$	3.7224	1.8948	1.9141	3.7991	2.0007	2.0098	4.4363	3.3656	3.3921
$\hat{\phi}_1$	4.5499	2.0755	2.1227	4.5940	2.1342	2.1714	4.4434	2.0975	2.1691
$\hat{\phi}_2$	4.6669	2.0658	2.0915	4.5456	2.0992	2.1468	4.6223	2.2284	2.2591

Tabela 4.31: Média, desvio padrão e raiz do erro quadrático médio empírico das estimativas considerando $n = 50$ e $\phi_x = 4$

$n = 50, \phi_x = 4$									
	$\lambda = 0.1$			$\lambda = 0.5$			$\lambda = 0.8$		
	Média	DP	REQM	Média	DP	REQM	Média	DP	REQM
$\hat{\lambda}$	0.1003	0.0071	0.0071	0.4998	0.0057	0.0057	0.7987	0.0064	0.0065
$\hat{\mu}_1$	15.0131	1.3614	1.3607	14.9764	0.5957	0.5959	15.0030	0.4729	0.4727
$\hat{\mu}_2$	25.0297	0.4426	0.4434	25.0046	0.6106	0.6103	25.0118	0.9799	0.9795
$\hat{\alpha}_2$	1.9760	0.7561	0.7561	1.9909	0.4608	0.4606	2.0325	0.5724	0.5730
$\hat{\phi}_x$	3.8370	1.4024	1.4111	3.7735	1.3914	1.4090	3.8245	1.4143	1.4245
$\hat{\phi}_1$	4.7508	1.5241	1.5436	4.7603	1.5147	1.5328	4.7488	1.5294	1.5492
$\hat{\phi}_2$	4.7468	1.4944	1.5149	4.8349	1.4976	1.5060	4.8670	1.4795	1.4848

Exemplo 4.2. *Um estudo de simulação considerando* $\mu_1 \simeq \mu_2$

Agora, geramos observações para o caso em que as médias são muito parecidas, de onde concluímos que o teste não detecta mudança.

Deste modo, consideramos $\lambda = 0.5, \mu_1 = 15$, $\mu_2 = 16$ e $\phi_x = 4$ mantendo os valores restantes iguais aos do exemplo anterior. Tomamos $n = 40$, de modo que $k = [n\lambda]$ e

Tabela 4.32: Média, desvio padrão e raiz do erro quadrático médio empírico das estimativas considerando $n = 100$ e $\phi_x = 4$

$n = 100, \phi_x = 4$									
	$\lambda = 0.1$			$\lambda = 0.5$			$\lambda = 0.8$		
	Média	DP	REQM	Média	DP	REQM	Média	DP	REQM
$\hat{\lambda}$	0.1002	0.0033	0.0033	0.5000	0.0026	0.0026	0.7995	0.0033	0.0034
$\hat{\mu}_1$	14.9859	0.9644	0.9640	14.9965	0.4120	0.4118	14.9943	0.3381	0.3380
$\hat{\mu}_2$	25.0118	0.3179	0.3180	24.9740	0.4383	0.4388	24.9800	0.7009	0.7008
$\hat{\alpha}_2$	2.0139	0.5347	0.5347	2.0113	0.3060	0.3060	1.9915	0.4119	0.4118
$\hat{\phi}_x$	3.9551	0.9370	0.9376	3.9090	0.9691	0.9729	3.8858	0.9744	0.9806
$\hat{\phi}_1$	4.8455	1.0470	1.0578	4.9065	1.0717	1.0753	4.9124	1.0713	1.0743
$\hat{\phi}_2$	4.9112	1.0874	1.0905	4.8779	1.0294	1.0361	4.8687	1.0377	1.0454

Tabela 4.33: Média, desvio padrão e raiz do erro quadrático médio empírico das estimativas considerando $n = 25$ e $\phi_x = 9$

$n = 25, \phi_x = 9$									
	$\lambda = 0.1$			$\lambda = 0.5$			$\lambda = 0.8$		
	Média	DP	REQM	Média	DP	REQM	Média	DP	REQM
$\hat{\lambda}$	0.1664	0.1819	0.1936	0.5130	0.0381	0.0402	0.7147	0.1820	0.2009
$\hat{\mu}_1$	15.1659	3.0923	3.0952	14.9274	1.1163	1.1181	14.8222	1.3802	1.3910
$\hat{\mu}_2$	24.9065	1.2029	1.2059	24.9018	1.1647	1.1682	23.8533	2.9835	3.1948
$\hat{\alpha}_2$	1.9615	1.0560	1.0562	2.0049	0.6464	0.6461	1.9769	0.9921	0.9919
$\hat{\phi}_x$	8.3242	3.4595	3.5232	8.4250	3.5537	3.5982	9.9754	5.1061	5.1959
$\hat{\phi}_1$	4.6334	2.4674	2.4932	4.6885	2.5010	2.5191	4.3988	2.7218	2.7861
$\hat{\phi}_2$	4.4381	2.4368	2.4995	4.4056	2.4879	2.5567	4.6267	2.6745	2.6991

$n - k = 20$, gerando assim os dados apresentados na Tabela 4.36:

Com os dados da Tabela 4.36 calculamos a estatística de Hotelling para todos os possíveis grupos com a finalidade de encontrar o máximo desta estatística. Os resultados são apresentados na Tabela 4.37.

Tabela 4.34: Média, desvio padrão e raiz do erro quadrático médio empírico das estimativas considerando $n = 50$ e $\phi_x = 9$

$n = 50, \phi_x = 9$									
	$\lambda = 0.1$			$\lambda = 0.5$			$\lambda = 0.8$		
	Média	DP	REQM	Média	DP	REQM	Média	DP	REQM
$\hat{\lambda}$	0.1060	0.0639	0.0642	0.4995	0.0108	0.0108	0.7951	0.0166	0.0173
$\hat{\mu}_1$	15.0682	1.8819	1.8822	14.9756	0.7413	0.7413	14.9534	0.5953	0.5968
$\hat{\mu}_2$	24.9793	0.7113	0.7113	25.0158	0.7258	0.7256	24.9179	1.1766	1.1788
$\hat{\alpha}_2$	1.9387	0.7518	0.7539	1.9851	0.4457	0.4457	2.0253	0.5649	0.5652
$\hat{\phi}_x$	8.5821	2.3706	2.4060	8.5674	2.2398	2.2801	8.4945	2.2510	2.3060
$\hat{\phi}_1$	4.7123	1.7705	1.7929	4.6883	1.7157	1.7430	4.7385	1.7230	1.7419
$\hat{\phi}_2$	4.7702	1.8630	1.8762	4.8433	1.8006	1.8065	4.7755	1.7318	1.7454

Tabela 4.35: Média, desvio padrão e raiz do erro quadrático médio empírico das estimativas considerando $n = 100$ e $\phi_x = 9$

$n = 100, \phi_x = 9$									
	$\lambda = 0.1$			$\lambda = 0.5$			$\lambda = 0.8$		
	Média	DP	REQM	Média	DP	REQM	Média	DP	REQM
$\hat{\lambda}$	0.1001	0.0060	0.0060	0.4996	0.0053	0.0053	0.7992	0.0056	0.0057
$\hat{\mu}_1$	14.9382	1.2016	1.2025	14.9995	0.5351	0.5348	14.9802	0.4146	0.4149
$\hat{\mu}_2$	24.9770	0.4113	0.4117	24.9942	0.5238	0.5236	25.0326	0.8674	0.8675
$\hat{\alpha}_2$	2.0309	0.5404	0.5410	1.9987	0.3136	0.3134	1.9842	0.3783	0.3784
$\hat{\phi}_x$	8.7692	1.5973	1.6131	8.7260	1.6570	1.6787	8.7081	1.6015	1.6270
$\hat{\phi}_1$	4.9062	1.2307	1.2337	4.9058	1.2942	1.2970	4.9279	1.3105	1.3118
$\hat{\phi}_2$	4.9240	1.2560	1.2577	4.9092	1.3063	1.3088	4.8773	1.3069	1.3120

Neste caso, $max_\lambda T^2_\lambda = 8.3872$ com $\hat{\lambda} = 0,6$, ou seja, $k = 24$. O valor da tabela dos valores críticos para este teste (veja Wang e Wang, 1994) é de 11,58 ($n = 40, p = 2$). Portanto, não se rejeita H_0.

Resultados similares aos obtidos nesta seção são encontrados em Alfaro (2000).

Tabela 4.36: Variáveis respostas Y_{ij}

grupo 1		grupo 2	
I_1	I_2	I_1	I_2
14.9238	19.7614	14.5896	14.8789
17.5910	17.5185	17.8829	16.8135
11.5592	16.4169	17.8092	20.5060
15.7707	15.3260	18.7433	16.1336
15.8643	18.9384	14.5119	20.1622
15.2081	9.7973	18.0883	21.4916
16.1743	17.0864	12.6101	18.8115
8.7665	15.0394	13.3343	18.9291
13.4711	13.1035	14.6912	16.9427
12.9388	13.4311	13.7980	18.8308
13.1629	14.4278	13.4464	21.2127
16.9084	19.1595	16.0546	16.4164
13.2855	13.3132	12.9080	16.6699
20.4463	21.7924	15.6199	20.9410
16.5529	14.7362	16.7750	17.6287
13.0156	13.0263	18.1323	29.0430
11.5221	18.1131	17.5232	20.7361
15.7606	22.4714	8.8580	13.4835
16.6382	18.4934	17.9505	20.5830
13.1441	16.6807	17.7670	20.5648

Tabela 4.37: Valores de T^2

	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
T^2	0.7280	0.1315	0.0531	0.1381	1.4631	1.9099	0.9807	2.0465	3.3005
	$k=11$	$k=12$	$k=13$	$k=14$	$k=15$	$k=16$	$k=17$	$k=18$	$k=19$
T^2	4.3830	3.3980	5.0942	3.7205	5.6768	7.7293	5.8297	3.2719	3.1542
	$k=20$	$k=21$	$k=22$	$k=23$	$k=24$	$k=25$	$k=26$	$k=27$	$k=28$
T^2	3.0488	4.1174	5.4088	4.9809	8.3872	6.3813	6.4671	4.4288	3.1490
	$k=29$	$k=30$	$k=31$	$k=32$	$k=33$	$k=34$	$k=35$	$k=36$	$k=37$
T^2	3.4036	2.5439	1.1497	1.9296	1.4947	1.1500	2.1268	3.3053	5.8458

5 *Considerações Finais*

5.1 Conclusões

O modelo de Grubbs tem sido amplamente utilizado no problema de comparação de instrumentos ou métodos de medição. Na literatura, vários autores exploram este modelo, contudo, quase sempre é assumida a hipótese de normalidade das observações. Essa suposição comprometeria uma análise estatística caso os dados fossem provenientes de distribuições com caudas mais pesadas ou mais leves que a cauda da distribuição normal. Isto sugere a necessidade de considerar uma análise estatística sobre uma nova classe de distribuições. Neste trabalho, consideramos um estudo de inferência no modelo t-Student, apresentando algumas extensões para os modelos Elípticos. Sob estes modelos, obtivemos alguns resultados originais como: as matrizes de informação de Fisher observadas e esperadas para os modelos elípticos dependente e independente, com especial atenção nas distribuições normal e t-Student; estimadores de máxima verossimilhança, via algoritmo EM, para o modelo t-Student com e sem restrição; propriedades assintóticas dos estimadores obtidos pelo método dos momentos e as estatísticas de Wald, Escore e Razão de Verossimilhanças para testar as hipóteses H_{01}, H_{02} e H_{03}.

Seguindo o trabalho de Alfaro (2000), propusemos o modelo de Grubbs com ponto de mudança; apresentamos o teste de Razão de Verossimilhanças para detectar a existência e posição da mudança; sob a existência da mudança, apresentamos estimadores do tipo momentos para os parâmetros do modelo e avaliamos estes estimadores através de um pequeno estudo de simulação, a partir do qual, podemos concluir que os estimadores produzem estimativas bem próximas dos verdadeiros parâmetros do modelo.

Finalmente, com respeito ao estudo de simulação desenvolvido na Seção 4.3, recomendamos o teste de Escore para os modelos com valores grandes dos graus de liberdade e o teste de Razão de Verossimilhanças para os modelos com graus de liberdade pequenos.

5.2 Pesquisa Futura

Neste trabalho, foi considerado o modelo de Grubbs estrutural elíptico. Considerando alguns resultados de Arellano *et al.* (1996), pode-se fazer um estudo assintótico, sob o modelo de Grubbs funcional e/ou ultraestrutural, considerando os estimadores obtidos pelo método dos momentos que são função da média amostral $\overline{\mathbf{Y}}$, e da matriz de covariâncias amostrais $\mathbf{S}_n$ definidas na Seção 2.7.

Para efeito de aplicação, temos considerado a distribuição t-Student, em que as estimativas dos parâmetros é obtido via algoritmo EM, em particular no modelo normal. Em geral, sob qualquer distribuição elíptica, a estimativa dos parâmetros é mais complexa. Uma forma de obtê-las é utilizar o algoritmo de Newton-Raphson e/ou Escore, usando a matriz de informação observada quando a matriz de informação esperada não é de fácil obtenção, por exemplo, sob a distribuição Exponencial-Potência.

Sob o modelo com ponto de mudança, gostaríamos de completar nosso estudo do comportamento assintótico dos parâmetros, obtendo as propriedades de $\hat{\phi}_i, i = 1, \cdots, p$, pois no modelo de Grubbs a qualidade dos intrumentos é avaliada em termos dos α_i's e ϕ_i's, de forma que o estimador do primeiro já foi estudado.

APÊNDICE A – Distribuições Elípticas

A.1 Definições e exemplos

Neste apêndice apresentamos definições e principais propriedades das distribuições elípticas juntamente com alguns exemplos ilustrativos. A classe de distribuições elípticas tem sido estudada por vários autores. Veja por exemplo, Kelker (1970), Muirhead (1980, 1982), Fang *et al.* (1990), Fang e Zhang (1990), Fang e Anderson (1990) e Arellano (1994).

As distribuições elípticas podem ser definidas de várias formas. Nós começamos definindo as distribuições esféricas e então estendemos a definição para as distribuições elípticas.

Definição 1. *O vetor aleatório* $\mathbf{X}(n \times 1)$ *é dito ter uma distribuição esférica (n-variada) se para cada matriz* $\mathbf{\Gamma}(n \times n)$ *ortogonal, temos que*

$$\mathbf{X} \stackrel{d}{=} \mathbf{\Gamma X}, \tag{A.1}$$

onde $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$ *indica que* $\mathbf{X}$ *e* $\mathbf{Y}$ *têm mesma distribuição.*

De (A.1) temos que um vetor aleatório $\mathbf{X}(n\times 1)$ tem distribuição esférica se, e somente se, a função característica de $\mathbf{X}$, $\psi_{\mathbf{X}}(\mathbf{t}) = E(e^{i\mathbf{t}^\top \mathbf{X}}), \mathbf{t} \in \mathbb{R}^n$ é da forma

$$\psi_{\mathbf{X}}(\mathbf{t}) = \phi(\mathbf{t}^\top \mathbf{t}), \tag{A.2}$$

para alguma função ϕ, com $\phi(u) \in \mathbb{R}, u \geq 0$. Segue também de (A.1) que se $\mathbf{X}$ tem densidade, ela é da forma

$$p_{\mathbf{X}}(\mathbf{x}) = f(\mathbf{x}^\top \mathbf{x}), \mathbf{x} \in \mathbb{R}^n, \tag{A.3}$$

para alguma função f, com $f(u) \geq 0, u \geq 0$.

Exemplo A.1. Seja $\mathbf{X} \sim N_n(\mathbf{0}, \mathbf{I}_n)$, a distribuição normal n-variada padrão. Neste caso, $\mathbf{X}$ tem função característica

$$\psi_{\mathbf{X}}(\mathbf{t}) = \phi(\mathbf{t}^\top \mathbf{t}) = e^{-\frac{1}{2}\mathbf{t}^\top \mathbf{t}}, \mathbf{t} \in \mathbb{R}^n, \tag{A.4}$$

e função de densidade

$$p_{\mathbf{X}}(\mathbf{t}) = f(\mathbf{x}^\top \mathbf{x}) = (2\pi)^{-\frac{1}{2}n} e^{-\frac{1}{2}\mathbf{x}^\top \mathbf{x}}, \mathbf{x} \in \mathbb{R}^n. \tag{A.5}$$

A definição das distribuições elípticas é dada seguindo Cambanis et al. (1981). Para motivar a definição, suponha que $\mathbf{Z}(m \times 1)$ é um vetor aleatório tendo distribuição esférica m-variada com função característica $\phi(\mathbf{s}^\top \mathbf{s}), \mathbf{s} \in \mathbb{R}^m$. Considere o vetor aleatório $\mathbf{X}(n \times 1)$ definido pela transformação linear

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z},$$

onde $\mathbf{A}$ é uma matriz $n \times n$ e $\boldsymbol{\mu}(n \times 1) \in \mathbb{R}^n$. Então, a função característica de $\mathbf{X}$ é dada por

$$\phi_{\mathbf{X}}(\mathbf{t}) = E(e^{i\mathbf{t}^\top(\boldsymbol{\mu} + \mathbf{A}\mathbf{Z})}) = e^{i\mathbf{t}^\top \boldsymbol{\mu}} \phi(\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}), \mathbf{t} \in \mathbb{R}^n,$$

onde $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^\top$. Com este resultado em mente, a seguinte definição da distribuição elíptica surge naturalmente.

Definição 2. *O vetor aleatório $\mathbf{X}(n \times 1)$ é dito ter distribuição elíptica (n-variada) com vetor de locação $\boldsymbol{\mu}(n \times 1) \in \mathbb{R}^n$ e matriz de dispersão $\boldsymbol{\Sigma}$ $(n \times n)$ definida não-negativa, se a função característica de $\mathbf{X}$ é da forma*

$$\psi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}^\top \boldsymbol{\mu}} \phi(\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}), \mathbf{t} \in \mathbb{R}^n, \tag{A.6}$$

para alguma função ϕ $(\phi(u) \in \mathbb{R}, u > 0)$. Em tal caso dizemos que $\mathbf{X}$ tem distribuição elíptica com parâmetros $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ e escrevemos

$$\mathbf{X} \sim El_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \phi).$$

Se a distribuição de $\mathbf{X}$ tem densidade, (com respeito à medida de Lebesgue em $\mathbb{R}^n$) então $\boldsymbol{\Sigma} > 0$ (definida positiva) (Arnold, 1981, p.41) e sua densidade é da forma

$$p_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} f((\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})), \mathbf{x} \in \mathbb{R}^n, \tag{A.7}$$

para alguma função $f, (f(u) \geq 0, u \geq 0)$. Neste caso costuma-se substituir ϕ por f na representação paramétrica dada (A.7). Note que, se em (A.7) $\boldsymbol{\mu} = 0$ e $\boldsymbol{\Sigma} = I_n$, isto é,

$$\mathbf{X} \sim El_n(\mathbf{0}, \mathbf{I}_n; \phi),$$

então (A.6) e (A.7) se reduzem a (A.2) e (A.3), respectivamente, e a distribuição de $\mathbf{X}$ é esférica.

Exemplo A.2. Seja $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, a distribuição normal n-variada com função característica

$$\psi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}^\top \boldsymbol{\mu}} e^{\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}}, \mathbf{t} \in \mathbb{R}^n.$$

Se $\boldsymbol{\Sigma} > 0$, então $\mathbf{X}$ tem densidade

$$p_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \mathbf{x} \in \mathbb{R}^n.$$

Exemplo A.3. Seja $\mathbf{X} \sim t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$, a distribuição t-Student n-variada com ν graus de liberdade e parâmetros de locação $\boldsymbol{\mu}$ e dispersão $\boldsymbol{\Sigma}$. Se $\boldsymbol{\Sigma} > 0$, então $\mathbf{X}$ tem densidade dada por

$$p_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \frac{\Gamma[\frac{1}{2}(\nu + n)] \nu^{\frac{1}{2}\nu}}{\Gamma(\frac{1}{2}\nu) \pi^{\frac{1}{2}n}} [\nu + (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})]^{-\frac{1}{2}(\nu+n)}, \mathbf{x} \in \mathbb{R}^n.$$

A função característica desta distribuição é derivada em Sutradhar (1986).

A.2 Propriedades

As propriedades dadas a seguir podem ser derivadas supondo que $\mathbf{X} \sim El_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ com $r(\boldsymbol{\Sigma}) = k \leq n$, excluindo o caso degenerado com $\boldsymbol{\Sigma} = 0$. Contudo, alguns resultados serão estabelecidos assumindo que $\mathbf{X}$ tem densidade.

Propriedade 1. *Se existem*

$$E(\mathbf{X}) = \boldsymbol{\mu} \quad e \quad Var(\mathbf{X}) = c_f \boldsymbol{\Sigma}, \tag{A.8}$$

onde c_f é uma constante que depende de f. Em termos da função característica esta constante é $c_f = -2\phi'(0)$.

Kelker (1970) nota que $u^{\frac{1}{2}(p-k)-1}f(u)$ é integrável, então os k-ésimos momentos de $\mathbf{X}$ existem. Também a variável aleatória $\mathbf{U} = (\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})$ tem densidade dada por

$$f_{\mathbf{U}}(u) = \{\pi^{p/2}/\Gamma(p/2)\}u^{\frac{p}{2}-1}f(u), \quad u \geq 0; \tag{A.9}$$

e em termos de $\mathbf{U}$, $c_f = E(\mathbf{U})/p$.

Propriedade 2. *Seja* $\mathbf{a} \in \mathbb{R}^q$, $q \leq p$, $\mathbf{A}$ *uma matriz* $q \times p$ *de posto* q *e* $\mathbf{Y} = \mathbf{a} + \mathbf{AX}$. *Então* $\mathbf{Y} \sim El_q(\mathbf{a} + \mathbf{A}\boldsymbol{\mu}, \mathbf{A\Sigma A}^\top)$.

Esta propriedade segue diretamente da definição 2.

Agora particionamos $\mathbf{X}$, $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ como:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \quad , \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad , \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

onde $\mathbf{X}_1$ e $\boldsymbol{\mu}_1$ são $q \times 1$, $q \leq p$ e $\boldsymbol{\Sigma}_{11}$ é $q \times q$.

Usamos novamente a definição 2 (ou a propriedade 2), temos:

Propriedade 3. *Se* $\mathbf{X}_1 \sim El_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, f)$, *então todas as distribuições marginais são também elípticas.*

Propriedade 4. *A distribuição condicional de* $\mathbf{X}_1$ *dado* $\mathbf{X}_2$ *é elíptica* q*-variada com média e variância (desde que existam)dadas por,*

$$\begin{aligned} E(\mathbf{X}_1|\mathbf{X}_2) &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2) \;\; e \\ Var(\mathbf{X}_1|\mathbf{X}_2) &= h(\mathbf{X}_2)(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}), \end{aligned}$$

para alguma função h *(ver por exemplo, Fang et al., 1990).*

Propriedade 5. *Seja* $\mathbf{X} \sim El_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *onde* $\boldsymbol{\Sigma}$ *é diagonal. Se os componentes de* $\mathbf{X}$ *são independentes, então* $\mathbf{X}$ *tem distribuição normal.*

Ver por exemplo, Muirhead (1982).

APÊNDICE B – Distribuição t-Student

Neste apêndice apresentamos um resumo das principais propriedades da distribuição t-Student. Os resultados que apresentamos a seguir podem ser encontrados em Arellano-Valle e Bolfarine (1995).

Definição 3. *Um vetor aleatório* $\mathbf{X} : n \times 1$ *é dito ter distribuição t-Student n-variada com vetor de locação* $\boldsymbol{\mu} : n \times 1$*, matriz de dispersão* $\boldsymbol{\Sigma} : n \times n$ *não negativa e* ν *graus de liberdade se,*

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + V^{\frac{1}{2}}\mathbf{Z}, \tag{B.1}$$

onde $\mathbf{Z} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$*,* $\text{posto}(\boldsymbol{\Sigma}) = k$ *e* $V \sim \nu/\chi^2_\nu$ *são independentes e escrevemos por*

$$\mathbf{X} \sim t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu) \tag{B.2}$$

Se $\text{posto}(\boldsymbol{\Sigma}) = k = n$, isto é, $\boldsymbol{\Sigma} > 0$, então $\mathbf{X}$ tem densidade dada por

$$p_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = k(n,\nu)\nu^{\nu/2}|\boldsymbol{\Sigma}|^{-1/2}[\nu + (\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})]^{-(\nu+n)/2}, \tag{B.3}$$

em que

$$k(n,\nu) = \frac{\Gamma(\frac{1}{2}(\nu+n))}{\pi^{\frac{1}{2}n}\Gamma(\frac{1}{2}\nu)}.$$

Em particular, se $\boldsymbol{\Sigma} = \mathbf{I}_n$, temos que $\mathbf{X}$ tem densidade dada por $f(\mathbf{x}^\top\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$, onde

$$f(u) = k(n,\nu)\nu^{\nu/2}(\nu+u)^{-(\nu+n)/2} \tag{B.4}$$

com $k(n,\nu)$ como em (B.3).

Se posto$(\mathbf{\Sigma}) = k < n$, então a distribuição de $(\mathbf{X}|V = \nu)$ é $N_n(\mathbf{0}, \nu\mathbf{\Sigma})$ singular. Conseqüentemente, a distribuição condicional de $\mathbf{X}$ é também singular. Este caso pode ser interpretado como uma versão singular da distribuição t-Student e em tal caso é tratado similarmente à distribuição elíptica no caso singular (veja Cambanis et al., 1981).

Uma extensão da distribuição $t_n(\boldsymbol{\mu}, \mathbf{\Sigma}; \nu)$, com $\mathbf{\Sigma} > 0$, pode ser obtida assumindo que em (B.1) a variável aleatória V tem distribuição $\lambda/\chi_\nu \equiv GI(\lambda/2, \nu/2)$. Neste caso, $\mathbf{X}$ tem densidade dada por

$$p_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{\Sigma}, \lambda, \nu) = k(n,\nu)\lambda^{\nu/2}|\mathbf{\Sigma}|^{-1/2}\{\lambda + (\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\}^{-(\nu+n)/2}, \tag{B.5}$$

$\mathbf{x} \in \mathbb{R}^n$, onde $k(n,\nu)$ é como em (B.3). Nesta situação, escrevemos $\mathbf{X} \sim t_n(\boldsymbol{\mu}, \mathbf{\Sigma}; \lambda, \nu)$ e diremos que $\mathbf{X}$ tem distribuição t-Student generalizada n-variada. Em particular, quando $\lambda = \nu$, temos a distribuião t-Student usual. Claramente, se $\mathbf{X}$ tem densidade dada por (B.5), então $\mathbf{X} \sim El_n(\boldsymbol{\mu}, \mathbf{\Sigma})$. Assim, podemos pensar também em uma versão singular da distribuição t-Student generalizada, definida similarmente como em (B.1), com $V \sim \lambda/\chi_\nu \equiv GI(\lambda/2, \nu/2)$ e $\mathbf{Z} \sim N_n(\mathbf{0}, \mathbf{\Sigma})$, posto$(\mathbf{\Sigma}) = k < n$, independentes.

Seja $\mathbf{X} \overset{d}{=} \boldsymbol{\mu} + V^{1/2}\mathbf{Z} \sim t_n(\boldsymbol{\mu}, \mathbf{\Sigma}; \lambda, \nu)$, com $V \sim \lambda/\chi^2_\nu$ e $\mathbf{Z} \sim N_n(\mathbf{0}, \mathbf{\Sigma})$ independentes. Então valem as seguintes propriedades:

i) $E[\mathbf{X}] = \boldsymbol{\mu}$, $\nu > 1$, $\text{Cov}[\mathbf{X}] = \frac{\lambda}{\nu-2}\mathbf{\Sigma}$, $\nu > 2$;

ii) se $\boldsymbol{\eta} \in \mathbb{R}^n$ e $\mathbf{A}$ é uma matriz $m \times n$ (não estocástica), então

$$\boldsymbol{\eta} + \mathbf{A}\mathbf{X} \sim t_n(\boldsymbol{\eta} + \mathbf{A}\mu, \mathbf{A}\mathbf{\Sigma}\mathbf{A}^\top; \lambda, \nu).$$

Em particular, se posto$(\mathbf{\Sigma}) = k = n$,

$$\mathbf{\Sigma}^{-1/2}(\mathbf{X} - \mu) \sim t_n(\mathbf{0}, \mathbf{I}_n; \lambda, \nu).$$

iii) $\frac{d^2}{k} \sim F_{k,\nu}$, em que $d^2 = (\mathbf{X} - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$.

iv) Partindo $\mathbf{X}$, $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ como

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad e \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

onde $\mathbf{X}_1$ e $\boldsymbol{\mu}_1$ são vetores $n \times 1$ e $\boldsymbol{\Sigma}_{11}$ é uma matriz $n \times n$, então

$$(\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2) \sim t_n(\boldsymbol{\mu}_1(\mathbf{x}_2), \boldsymbol{\Sigma}_{11.2}; \lambda_{q(\mathbf{x}_2)}, \nu_m),$$

em que $\boldsymbol{\mu}_1(\mathbf{x}_2) = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$, $\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ e $q(\mathbf{x}_2) = (\mathbf{x}_2 - \boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$ e os parâmetros $\lambda_{q(\mathbf{x}_2)}$ e ν_m são dados por

$$\begin{aligned} \lambda_{q(\mathbf{x}_2)} &= \lambda + q(\mathbf{x}_2), \\ \nu_m &= \nu + n - m, 1 \leq m \leq n. \end{aligned}$$

Além disso,

$$\begin{aligned} E[\mathbf{X}_1|\mathbf{X}_2] &= \boldsymbol{\mu}_1(\mathbf{X}_2), \nu_m > 1, \\ Cov[\mathbf{X}_1|\mathbf{X}_2] &= \frac{\lambda_q(\mathbf{x}_2)}{\nu_m - 2}\boldsymbol{\Sigma}_{11.2} \\ &= \frac{\lambda + (\mathbf{X}_2 - \boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2)}{\nu + n - m - 2}(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}), \nu_m > 2. \end{aligned}$$

A afirmação seguinte é importante para obter a matriz de informação de Fisher esperada sob o modelo t-Student.

Lema B.1. *Seja* $\mathbf{X} \sim t_n(\mathbf{0}, \mathbf{I}_n; \lambda, \nu)$. *Então*

$$E[||\mathbf{X}||^{2k}(\lambda + ||\mathbf{X}||^2)^{-m}] = \frac{B\left[\frac{1}{2}(p+2k), \frac{1}{2}(\nu + 2m - 2k)\right]}{\lambda^{(2m-2k)/2} B(\frac{1}{2}p, \frac{1}{2}\nu)},$$

$\nu + 2m - 2k > 0$, *onde* $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$ *é a função beta.*

APÊNDICE C – Teste da Razão de Verossimilhança do Modelo de Grubbs com Ponto de Mudança

C.1 Teste da razão de verossimilhanças

Consideremos a seguir a obtenção do teste da razão de verossimilhança para comparar as médias de 2 populações normais bivariadas. Para isso, necessitaremos de um resultado que enunciaremos a seguir:

Resultado 1. *Se* $\mathbf{B}$ *é uma matriz* $p \times p$ *simétrica e positiva definida e* b *uma constante positiva, então,*

$$\frac{1}{|\mathbf{\Sigma}|^b} exp\left\{-\frac{1}{2} tr\left[\mathbf{\Sigma}^{-1}\mathbf{B}\right]\right\} \leq \frac{1}{\mathbf{B}^b}(2b)^{pb} exp\{-bp\},$$

para toda $\mathbf{\Sigma}$, *matriz definida positiva.*

Temos que a igualdade ocorre para $\mathbf{\Sigma} = \frac{\mathbf{B}}{2b}$, *isto é,* $\mathbf{\Sigma} = \frac{\mathbf{B}}{2b}$ *produz o valor máximo de*

$$\frac{1}{|\mathbf{\Sigma}|^b} exp\left\{-\frac{1}{2} tr\left[\mathbf{\Sigma}^{-1}\mathbf{B}\right]\right\}.$$

Sob $H_0 : \mu_1 = \mu_2 = \mu_x$, o vetor $\mathbf{Y}_j = (Y_{1j}, Y_{2j})^\top \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, onde

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_x \\ \alpha_2 + \mu_x \end{bmatrix} \quad e \quad \boldsymbol{\Sigma} = \begin{pmatrix} \phi_x + \phi_1 & \phi_x \\ \phi_x & \phi_x + \phi_2 \end{pmatrix}, \quad j = 1, \cdots, n.$$

Podemos escrever a função de verossimilhança (3.10) como

$$\begin{aligned}
L(\boldsymbol{\theta}) &= \frac{1}{(2\pi)^n |\boldsymbol{\Sigma}|^{\frac{n}{2}}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n} \mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_j - \boldsymbol{\mu})(\mathbf{Y}_j - \boldsymbol{\mu})^\top\right]\right\} \\
&= \frac{1}{(2\pi)^n |\boldsymbol{\Sigma}|^{n/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}(\mathbf{Y}_j - \boldsymbol{\mu})(\mathbf{Y}_j - \boldsymbol{\mu})^\top\right]\right\} \\
&= \frac{1}{(2\pi)^n |\boldsymbol{\Sigma}|^{n/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}(\mathbf{Y}_j - \overline{\mathbf{Y}} + \overline{\mathbf{Y}} - \boldsymbol{\mu})(\mathbf{Y}_j - \overline{\mathbf{Y}} + \overline{\mathbf{Y}} - \boldsymbol{\mu})^\top\right]\right\} \\
&= \frac{1}{(2\pi)^n |\boldsymbol{\Sigma}|^{n/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}(\mathbf{Y}_j - \overline{\mathbf{Y}})(\mathbf{Y}_j - \overline{\mathbf{Y}})^\top\right.\right. \\
&+ \left.\left. n(\overline{Y} - \boldsymbol{\mu})(\overline{\mathbf{Y}} - \boldsymbol{\mu})^\top\right]\right\} \\
&= \frac{1}{(2\pi)^n |\boldsymbol{\Sigma}|^{\frac{n}{2}}} \exp\left\{-\frac{1}{2}\left\{\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{Y}_j - \overline{\mathbf{Y}})(\mathbf{Y}_j - \overline{\mathbf{Y}})^\top\right)\right]\right.\right. \\
&+ \left.\left. n\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\left((\overline{\mathbf{Y}} - \boldsymbol{\mu})(\overline{\mathbf{Y}} - \boldsymbol{\mu})^\top\right)\right]\right\}\right\}.
\end{aligned}$$

Seja

$$\mathbf{B} = \sum_{j=1}^{n}(\mathbf{Y}_j - \overline{\mathbf{Y}})(\mathbf{Y}_j - \overline{\mathbf{Y}})^\top,$$

temos que

$$\begin{aligned}
L_n(\boldsymbol{\theta}) &= \frac{1}{(2\pi)^n |\boldsymbol{\Sigma}|^{n/2}} \exp\left\{-\frac{1}{2}\left\{\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\mathbf{B}\right] + n\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\left((\overline{\mathbf{Y}} - \boldsymbol{\mu})(\overline{\mathbf{Y}} - \boldsymbol{\mu})^\top\right)\right]\right\}\right\} \\
&= \frac{1}{(2\pi)^n |\boldsymbol{\Sigma}|^{n/2}} \exp\left\{-\frac{1}{2}\left\{\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\mathbf{B}\right] + n(\overline{\mathbf{Y}} - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{Y}} - \boldsymbol{\mu})\right\}\right\}. \qquad \text{(C.1)}
\end{aligned}$$

Sob H_0 os EMV de $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ são respectivamente,

$$\widehat{\boldsymbol{\mu}} = \overline{\mathbf{Y}} = \begin{bmatrix} \overline{Y}_1(0,n) \\ \overline{Y}_2(0,n) \end{bmatrix} \quad e \quad \widehat{\boldsymbol{\Sigma}} = \frac{\mathbf{B}}{n} = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{Y}_j - \overline{\mathbf{Y}})(\mathbf{Y}_j - \overline{\mathbf{Y}})^\top. \tag{C.2}$$

onde $\overline{Y}_i(a,b)$, $i = 1,2$ é definida em (3.12).

Para verificar estes resultados acima, considere a expressão (C.1) da função de verossimilhança com $\boldsymbol{\Sigma}^{-1}$ positiva definida, i.e., $(\overline{\mathbf{Y}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\overline{\mathbf{Y}} - \boldsymbol{\mu}) > 0$ para $\overline{\mathbf{Y}} \neq \boldsymbol{\mu}$, então, temos que $L(\boldsymbol{\theta})$ é maximizado com relação à $\boldsymbol{\mu}$ no ponto $\boldsymbol{\mu} = \overline{\mathbf{Y}}$. Portanto, o EMV de $\boldsymbol{\mu}$ é $\widehat{\boldsymbol{\mu}} = \overline{\mathbf{Y}}$.

Agora devemos maximizar $L(\boldsymbol{\theta})$ com relação a $\boldsymbol{\Sigma}$. Fazemos isso usando o Resultado 1.

Então

$$L(\boldsymbol{\theta}) = \frac{1}{(2\pi)^n|\boldsymbol{\Sigma}|^{n/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\mathbf{B}\right]\right\} \leq \frac{1}{(2\pi)^n|\mathbf{B}|^{n/2}}\left(2\frac{n}{2}\right)^n \exp\left\{-n\right\},$$

onde a igualdade ocorre para $\boldsymbol{\Sigma} = \frac{\mathbf{B}}{n}$. Portanto $\widehat{\boldsymbol{\Sigma}} = \frac{\mathbf{B}}{n}$ é o EMV de $\boldsymbol{\Sigma}$. Ou seja, os EMV de $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ são as expressões dadas em (C.2).

O valor máximo da função de verossimilhança sob H_0 pode ser escrito como

$$\frac{1}{(2\pi)^n|\boldsymbol{\Sigma}|^{n/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\mathbf{B}\right] + 0\right\} = \frac{1}{(2\pi)^n|\boldsymbol{\Sigma}|^{n/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[n\mathbf{B}^{-1}\mathbf{B}\right]\right\}.$$

Sendo $\mathbf{B}^{-1}\mathbf{B} = \mathbf{I}_2$, temos que

$$\sup_{H_0} L_n(\boldsymbol{\theta}) = \frac{e^{-n}}{(2\pi)^n}|\widehat{\boldsymbol{\Sigma}}|^{-n/2} = K|\widehat{\boldsymbol{\Sigma}}|^{-n/2}.$$

Sob $H_a : \mu_1 \neq \mu_2$, temos as observações

$$\underbrace{\begin{pmatrix} Y_{11} \\ Y_{21} \end{pmatrix}, \cdots, \begin{pmatrix} Y_{1k} \\ Y_{2k} \end{pmatrix}}_{\text{antes do ponto de mudança}}, \underbrace{\begin{pmatrix} Y_{1(k+1)} \\ Y_{2(k+1)} \end{pmatrix}, \cdots, \begin{pmatrix} Y_{1n} \\ Y_{2n} \end{pmatrix}}_{\text{após a mudança}}.$$

Então, o vetor

$$
\begin{aligned}
\mathbf{Y}_j &= (Y_{1j}, Y_{2j})^\top \sim N_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}), j = 1, \cdots, k, \\
&\text{e} \\
\mathbf{Y}_j &= (Y_{1j}, Y_{2j})^\top \sim N_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}), j = k+1, \cdots, n,
\end{aligned}
$$

de modo que

$$
\boldsymbol{\mu}_1 = \begin{bmatrix} \mu_1 \\ \alpha_2 + \mu_1 \end{bmatrix} \quad e \quad \boldsymbol{\mu}_2 = \begin{bmatrix} \mu_2 \\ \alpha_2 + \mu_2 \end{bmatrix}.
$$

Sendo

$$
\overline{\mathbf{Y}}_1 = \begin{bmatrix} \overline{Y}_1(0,k) \\ \overline{Y}_2(0,k) \end{bmatrix} \quad e \quad \overline{\mathbf{Y}}_2 = \begin{bmatrix} \overline{Y}_1(k,n) \\ \overline{Y}_2(k,n) \end{bmatrix}, \tag{C.3}
$$

onde $\overline{Y}_i(a,b)$, $i = 1, 2$ é definida em (3.12), temos

$$
\mathbf{B}_1 = \sum_{j=1}^{k} (\mathbf{Y}_j - \overline{\mathbf{Y}}_1)(\mathbf{Y}_j - \overline{\mathbf{Y}}_1)^\top = \sum_{j=1}^{k} \begin{pmatrix} Y_{1j} - \overline{Y}_1(0,k) \\ Y_{2j} - \overline{Y}_2(0,k) \end{pmatrix} \begin{pmatrix} Y_{1j} - \overline{Y}_1(0,k) \\ Y_{2j} - \overline{Y}_2(0,k) \end{pmatrix}^\top \tag{C.4}
$$

e

$$
\mathbf{B}_2 = \sum_{j=k+1}^{n} (\mathbf{Y}_j - \overline{\mathbf{Y}}_2)(\mathbf{Y}_j - \overline{\mathbf{Y}}_2)^\top = \sum_{j=k+1}^{n} \begin{pmatrix} Y_{1j} - \overline{Y}_1(k,n) \\ Y_{2j} - \overline{Y}_2(k,n) \end{pmatrix} \begin{pmatrix} Y_{1j} - \overline{Y}_1(k,n) \\ Y_{2j} - \overline{Y}_2(k,n) \end{pmatrix}^\top
$$

Novamente, temos que a função de verossimilhança dada no Capítulo 2, expressão (3.13) pode ser reescrita como

$$
L_n(\boldsymbol{\theta}) = \frac{1}{(2\pi)^n |\boldsymbol{\Sigma}|^{n/2}} \exp\left\{ -\frac{1}{2} \left\{ \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\mathbf{B}_1\right] + \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\mathbf{B}_2\right] + \right.\right.
$$

$$
\left.\left. + k(\overline{\mathbf{Y}}_1 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}^{-1} (\overline{\mathbf{Y}}_1 - \boldsymbol{\mu}_1) + (n-k)(\overline{\mathbf{Y}}_2 - \boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}^{-1} (\overline{\mathbf{Y}}_2 - \boldsymbol{\mu}_2) \right\} \right\}.
$$

Esta função é maximizada em relação a $\boldsymbol{\mu}_1$ e $\boldsymbol{\mu}_2$, para $\boldsymbol{\mu}_1 = \overline{\mathbf{Y}}_1$ e $\boldsymbol{\mu}_2 = \overline{\mathbf{Y}}_2$, já que $(\overline{\mathbf{Y}}_1 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}^{-1} (\overline{\mathbf{Y}}_1 - \boldsymbol{\mu}_1) \geq 0$, $(\overline{\mathbf{Y}}_2 - \boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}^{-1} (\overline{\mathbf{Y}}_2 - \boldsymbol{\mu}_2) \geq 0$ se as igualdades ocorrem quando $\boldsymbol{\mu}_1 = \overline{\mathbf{Y}}_1$ e $\boldsymbol{\mu}_2 = \overline{\mathbf{Y}}_2$ respectivamente. Portanto $\widehat{\boldsymbol{\mu}}_1 = \overline{\mathbf{Y}}_1$ e $\widehat{\boldsymbol{\mu}}_2 = \overline{\mathbf{Y}}_2$ são os estimadores de máxima verossimilhança de $\boldsymbol{\mu}_1$ e $\boldsymbol{\mu}_2$ respectivamente.

Para obter o EMV de $\boldsymbol{\Sigma}$ basta maximizar $L_n(\boldsymbol{\theta})$ em relação a $\boldsymbol{\Sigma}$, isto é, obter o ponto máximo da função

$$\begin{aligned} &\frac{1}{(2\pi)^n|\boldsymbol{\Sigma}|^{\frac{n}{2}}}\exp\left\{-\frac{1}{2}\left\{\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\mathbf{B}_1\right]+\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\mathbf{B}_2\right]\right\}\right\} \\ &= \frac{1}{(2\pi)^n|\boldsymbol{\Sigma}|^{\frac{n}{2}}}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}(\mathbf{B}_1+\mathbf{B}_2)\right]\right\}. \end{aligned}$$

Usando o Resultado 1 no Apêndice A, temos que

$$\begin{aligned} \frac{1}{(2\pi)^n|\boldsymbol{\Sigma}|^{\frac{n}{2}}}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}(\mathbf{B}_1+\mathbf{B}_2)\right]\right\} &\leq \frac{\left(2\frac{n}{2}\right)^n e^{-n}}{(2\pi)^n|\mathbf{B}_1+\mathbf{B}_2|^{\frac{n}{2}}} \\ &= \frac{n^n e^{-n}}{(2\pi)^n|\mathbf{B}_1+\mathbf{B}_2|^{\frac{n}{2}}} \end{aligned}$$

e ocorre a igualdade quando $\boldsymbol{\Sigma} = \frac{1}{n}(\mathbf{B}_1+\mathbf{B}_2)$. Então,

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}(\mathbf{B}_1+\mathbf{B}_2) = \frac{1}{n}\left[\sum_{j=1}^{k}(\mathbf{Y}_j-\overline{\mathbf{Y}}_1)(\mathbf{Y}_j-\overline{\mathbf{Y}}_1)^\top + \sum_{j=k+1}^{n}(\mathbf{Y}_j-\overline{\mathbf{Y}}_2)(\mathbf{Y}_j-\overline{\mathbf{Y}}_2)^\top\right]$$

é o EMV de $\boldsymbol{\Sigma}$.

Logo, os estimadores de máxima verossimilhança de $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ e $\boldsymbol{\Sigma}$ são dados por

$$\widehat{\boldsymbol{\mu}}_1 = \overline{\mathbf{Y}}_1 = \begin{bmatrix} \overline{Y}_1(0,k) \\ \overline{Y}_2(0,k) \end{bmatrix} \quad , \quad \widehat{\boldsymbol{\mu}}_2 = \overline{\mathbf{Y}}_2 = \begin{bmatrix} \overline{Y}_2(k,n) \\ \overline{Y}_2(k,n) \end{bmatrix}$$

e

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}(\mathbf{B}_1+\mathbf{B}_2) = \frac{\sum_{j=1}^{k}(\mathbf{Y}_j-\overline{\mathbf{Y}}_1)(\mathbf{Y}_j-\overline{\mathbf{Y}}_1)^\top + \sum_{j=k+1}^{n}(\mathbf{Y}_j-\overline{\mathbf{Y}}_2)(\mathbf{Y}_j-\overline{\mathbf{Y}}_2)^\top}{n},$$

de modo que

$$\sup_{H_1} L_n = L_n(\widehat{\boldsymbol{\theta}}) = \frac{n^n e^{-n}}{(2\pi)^n|\mathbf{B}_1+\mathbf{B}_2|^{\frac{n}{2}}} = K\frac{1}{|\mathbf{B}_1+\mathbf{B}_2|^{-\frac{n}{2}}}.$$

Finalmente para λ fixado, a razão da verossimilhança é:

$$\frac{\sup_{H_0} L_n}{\sup_{H_1} L_n} = K\frac{|\mathbf{B}|^{-\frac{n}{2}}}{|\mathbf{B}_1 + \mathbf{B}_2|^{-\frac{n}{2}}}$$

Assim, rejeita-se $H_0 \Longleftrightarrow \left(\frac{|\mathbf{B}|}{|\mathbf{B}_1+\mathbf{B}_2|}\right)^{-\frac{n}{2}} < c \Longleftrightarrow \frac{|\mathbf{B}_1+\mathbf{B}_2|}{|\mathbf{B}|} < c.$

$$\Longleftrightarrow \frac{|\sum_{j=1}^{k}(\mathbf{Y}_j - \overline{\mathbf{Y}}_1)(\mathbf{Y}_j - \overline{\mathbf{Y}}_1)^\top + \sum_{j=k+1}^{n}(\mathbf{Y}_j - \overline{\mathbf{Y}}_2)(\mathbf{Y}_j - \overline{\mathbf{Y}}_2)^\top|}{|\sum_{j=1}^{n}(\mathbf{Y}_j - \overline{\mathbf{Y}})(\mathbf{Y}_j - \overline{\mathbf{Y}})^\top|} < c.$$

Agora vamos mostrar que este teste é equivalente ao teste T^2 de Hotelling. Note que:

$$|\sum_{j=1}^{n}(\mathbf{Y}_j - \overline{\mathbf{Y}})(\mathbf{Y}_j - \overline{\mathbf{Y}})^\top| = |\sum_{j=1}^{k}(\mathbf{Y}_j - \overline{\mathbf{Y}})(\mathbf{Y}_j - \overline{\mathbf{Y}})^\top + \sum_{j=k+1}^{n}(\mathbf{Y}_j - \overline{\mathbf{Y}})(\mathbf{Y}_j - \overline{\mathbf{Y}})^\top|$$

$$\begin{aligned}
= \;& |\sum_{j=1}^{k}(\mathbf{Y}_j - \overline{\mathbf{Y}}_1 + \overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}})(\mathbf{Y}_j - \overline{\mathbf{Y}}_1 + \overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}})^\top \\
+ \;& \sum_{j=k+1}^{n}(\mathbf{Y}_j - \overline{\mathbf{Y}}_2 + \overline{\mathbf{Y}}_2 - \overline{\mathbf{Y}})(\mathbf{Y}_j - \overline{\mathbf{Y}}_2 + \overline{\mathbf{Y}}_2 - \overline{\mathbf{Y}})^\top| \\
= \;& |\sum_{j=1}^{k}(\mathbf{Y}_j - \overline{\mathbf{Y}}_1)(\mathbf{Y}_j - \overline{\mathbf{Y}}_1)^\top + \sum_{j=k+1}^{n}(\mathbf{Y}_j - \overline{\mathbf{Y}}_2)(\mathbf{Y}_j - \overline{\mathbf{Y}}_2)^\top + k(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}})(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}})^\top \\
+ \;& (n-k)(\overline{\mathbf{Y}}_2 - \overline{\mathbf{Y}})(\overline{\mathbf{Y}}_2 - \overline{\mathbf{Y}})^\top| \\
= \;& |\mathbf{B}_1 + \mathbf{B}_2 + k(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}})(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}})^\top + (n-k)(\overline{\mathbf{Y}}_2 - \overline{\mathbf{Y}})(\overline{\mathbf{Y}}_2 - \overline{\mathbf{Y}})^\top|.(**)
\end{aligned}$$

Além disso

$$\begin{aligned}
\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}} &= \overline{\mathbf{Y}}_1 - \frac{k\overline{\mathbf{Y}}_1 + (n-k)\overline{\mathbf{Y}}_2}{n} = \frac{(n-k)}{n}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2) \\
&\text{e} \\
\overline{\mathbf{Y}}_2 - \overline{\mathbf{Y}} &= \overline{\mathbf{Y}}_2 - \frac{k\overline{\mathbf{Y}}_1 + (n-k)\overline{\mathbf{Y}}_2}{n} = -\frac{k}{n}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2),
\end{aligned}$$

de modo que

$$\begin{aligned}
(**) &= |\mathbf{B}_1 + \mathbf{B}_2 + \frac{k(n-k)^2}{n^2}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)^\top \\
&+ \frac{(n-k)k^2}{n^2}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)^\top| \\
&= |\mathbf{B}_1 + \mathbf{B}_2 + \frac{k(n-k)}{n^2}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)^\top((n-k)+k)| \\
&= |\mathbf{B}_1 + \mathbf{B}_2 + \frac{k(n-k)}{n}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)^\top|.
\end{aligned}$$

Portanto,

$$|\sum_{j=1}^{n}(\mathbf{Y}_j - \overline{\mathbf{Y}})(\mathbf{Y}_j - \overline{\mathbf{Y}})^\top| = |\mathbf{B}_1 + \mathbf{B}_2 + \frac{k(n-k)}{n}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)^\top|,$$

isto é,

$$|\mathbf{B}| = |\mathbf{B}_1 + \mathbf{B}_2 + \frac{k(n-k)}{n}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)^\top|.$$

Então,

$$|\mathbf{B}| = |\mathbf{B}_1 + \mathbf{B}_2|.\left(1 + \frac{k(n-k)}{n}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)^\top(\mathbf{B}_1 + \mathbf{B}_2)^{-1}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)\right).$$

Logo

$$\begin{aligned}
\frac{|\mathbf{B}|}{|\mathbf{B}_1 + \mathbf{B}_2|} &= 1 + \frac{k(n-k)}{n}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)^\top(\mathbf{B}_1 + \mathbf{B}_2)^{-1}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2) \\
&= 1 + \frac{\sqrt{\frac{k(n-k)}{n}}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)^\top\left(\frac{\mathbf{B}_1 + \mathbf{B}_2}{n-2}\right)^{-1}\sqrt{\frac{k(n-k)}{n}}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)}{n-2}.
\end{aligned}$$

Fazendo

$$\mathbf{Z}_\lambda = \sqrt{\frac{k(n-k)}{n}}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)$$

e

$$\mathbf{W}_\lambda = \frac{\mathbf{B}_1 + \mathbf{B}_2}{n-2} = \frac{\sum_{j=1}^{k}(\mathbf{Y}_j - \overline{\mathbf{Y}}_1)(\mathbf{Y}_j - \overline{\mathbf{Y}}_1)^\top + \sum_{j=k+1}^{n}(\mathbf{Y}_j - \overline{\mathbf{Y}}_2)(\mathbf{Y}_j - \overline{\mathbf{Y}}_2)^\top}{n-2},$$

podemos escrever

$$\frac{|\mathbf{B}|}{|\mathbf{B}_1 + \mathbf{B}_2|} = 1 + \frac{\mathbf{Z}_\lambda^\top \mathbf{W}_\lambda^{-1} \mathbf{Z}_\lambda}{n-2}.$$

Em resumo, rejeitamos H_0 quando

$$\frac{|\mathbf{B}_1 + \mathbf{B}_2|}{|\mathbf{B}|} < c \Longleftrightarrow \frac{|\mathbf{B}|}{|\mathbf{B}_1 + \mathbf{B}_2|} > c \Longleftrightarrow 1 + \frac{\mathbf{Z}_\lambda^\top \mathbf{W}_\lambda^{-1} \mathbf{Z}_\lambda}{n-2} > c \Longleftrightarrow \mathbf{Z}_\lambda^\top \mathbf{W}_\lambda^{-1} \mathbf{Z}_\lambda > c',$$

que é equivalente a rejeitar $H_0 \Longleftrightarrow T^2 > c$, onde $T^2 = \mathbf{Z}_\lambda^\top \mathbf{W}_\lambda^{-1} \mathbf{Z}_\lambda$ e c é o valor crítico segundo o nível de significância α.

Referências Bibliográficas

ALFARO, K.H.G. (2000). Modelo de calibração comparativa em grupos. Tese de doutorado, IME - USP.

ANDERSON, T.W.E.; FANG, K.T.E. e HSU, H. (1986). Maximum-likelihood estimates and likelihood-ratio criteria for multivariate elliptical contoured distributions. *The Canandian Journal of Statistics*, **44**, 55-59.

ANDREWS, D.W.K.; LEE, I. e PLOBERGER, W. (1996). Optimal changepoint tests for normal linear regression. *Journal of Econometrics*, 70, 9-38.

ARELLANO, R.B. (1994). Distribuições elípticas: propriedades, inferência e aplicações a modelos de regressão. Tese de doutorado, IME - USP.

ARELLANO, R. e BOLFARINE , H. (1995). A note on the simple structural regression model. *Annals of the Institute of Statistical Mathematics*, **48**, 111-125.

ARELLANO, R.B.; BOLFARINE, H. e VILCA- LABRA, F.E. (1996). Ultrastructural elliptical models. *The Canadian Journal of Statistics*, **24**(2), 207-216.

ARNOLD, S.F. (1981). *The Theory of Linear Models and Multivariate Analysis.* New York: Wiley.

BARNETT, V.D. (1969). Simultaneous pairwise linear structural relationships. *Biometrics*, **25**, 129-142.

BARNETT, V.D. e LEWIS, T. (1994). *Outliers in Statistical Data*. 3rd Edition. John Wiley

BEDRICK, E.J. (2001). An efficient scores test for comparing several measuring devices. *Journal of Quality Technology*, **33**, **1**, 96-103.

BELSLEY, D.A.; KUH, E. e WELSCH, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Scores of Collinearity*, John Wiley. New York.

BERKANE, M.; KANO, Y. e BENTLER, P.M.(1994). Pseudo maximun likelihood estimation in elliptical theory: Effects of misspecification. *Computational Statistics* & *Data*

Analysis, **18**, 255-267.

BOLFARINE, H. e ARELLANO-VALLE, R.B.(1994). Robust modeling in measurement error models using the Student-t distributions. *Brazilian Journal of Probability and Statistics*, **8**, 67-84.

BRINDLEY, D.A. e BRADLEY, R.A. (1985). Some new results on Grubb's estimator. *Journal of the American Statistical Association*, **80**, 711-714.

CAMBANIS, S.; HUANG, S e SIMONS, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, **11**, 368-385.

CARTER, R. (1981): Restricted maximum likelihood estimation of bias and reliability in the comparison of several measuring methods . *Biometrics*, **37**, 733-741.

CHANG, Y. e HUANG, W. (1997). A inference for the linear errors-in-variables with changepoint models. *Journal of the American Statistical Association*, **92**, 171-178.

CHATTERJEE, S. e HADI, A.S. (1988). *Sensivity Analysis in Linear Regression*, John Wiley. New York.

CHEN, J. e GUPTA, A.K. (1997). Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical Association*, **92**, 739-747.

CHENG, C. e VAN NESS, J. (1990). Bounded influence errors in variables regression. *Contemporary Mathematics*, **112**, 227-241.

CHERNOFF, H. e ZACKS, S. (1964). Estimating the current mean of a normal distribution wich is subject to Changes in time. *Annals of Mathematical Statistics*, **35**, 999-1018.

CHIB, S.; TIWARI, R.C. e JAMMALAMADAKA, S.R. (1988). Bayes prediction in regressions with elliptical errors. *Journal of Econometrics*, **38**, 349-360.

CHRISTENSEN, R. e BLACKWOOD, L.G (1993). Tests for accuracy of multiple measuring devices. *Technometrics* **35**, 411-420.

COOK, R.D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society*, B, **4**8, 133-169.

COOK, R.D. and WEISBERG, S. (1982). *Residuals and Influence in Regression*, London: Chapman and Hall.

DEMPSTER, A.; LAIRD, N.; e RUBIN, D. B. (1977). Maximum likeihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal*

Statistical Society, B, **39**, 1-38.

EVANS, G.B.A. e SAVIN, N.E. (1982). Conflict among the criteria revisited; the W, LR and LM tests. *Econometrica*, **50**, 737-748.

FANG, K.T. e ANDERSON, T.W. (1990). *Statistical Inference in Elliptical Contoured and Related Distributions*, New York: Allerton Press Inc..

FANG, K.T.; KOTZ, S. e Ng, K.W. (1990). *Symmetric multivariate and related distributions*. London: Chapman and Hall.

FANG,K.T. e ZHANG, Y.T. (1990). *Generalized Multivariate Analysis*, Springer-Verlag. London.

FULLER, W.A. (1987). *Measurement Error Models*, Wiley, New York.

GALEA-ROJAS, M.J. (1996).Calibração comparativa estrutural e funcional. Tese de doutorado, IME - USP.

GALEA-ROJAS, M.J.; BOLFARINE, H.; e Vilca, F.E.V. (2002). Local influence in comparative calibration under elliptical t-distribution. Relatório de pesquisa, IMECC, Unicamp. RP13/02.

GARDNER, L.A. (1969). On detecting change in the mean of normal variates. *Annals of Mathematical Statistics*, **40**, 116-126.

GASCO, C.L. (1997). Previsão e preditivismo em modelos lineares com e sem erros nas variáveis. Tese de Doutorado, IME-USP.

GHOSH, M. e KUMAR SINHA, B. (1980). On the robustness of least squares procedures in regression models. *Journal of Multivariate Analysis*, **10**, 332-342.

GRUBBS, F.E. (1948). On estimating precision of measuring instruments and product variability. *Journal of the American Statistical Association*, **43**, 243-264.

GRUBBS, F.E. (1973). Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments. *Tecnometrics*, **15**, 53-66.

HAMPEL, F.R., RONCHETTI, E.M., ROUSSEEUW, P.J. e STAHEL, W.A. (1986). *Robust Statistics*, John Wiley. New York.

HAWKINS, D.M. (1980). *Identification of Outliers*, Chapman and Hall. London.

HAWKINS, D.M. (1992). Detecting shifts in functions of multivariate location and covariance parameters. *Journal of Statistical Planning and Inference*, **33**, 233-244.

HENDERSON, H.V. e SEARLE, S.R. (1979). Vec and Vech operators for matrices, with some uses in jacobians and multivariates statiscs. *The Canadian Journal of Statistics 7*, **1**, 65-81.

HUBER, P.Y. (1981). *Robust Statistics.*, New York: Wiley.

JAECH, J.L. (1985). Statistical Analysis of Measurement Errors. *Exxon Monographs* . John Wiley, New York.

JAMES, B.R. (1996). *Probabilidade: Um Curso em Nível Intermediário*, 2ed. Rio de Janeiro: Instituto de Matemática Pura e Aplicada.

JOHNSON, N. e KOTZ, S. (1970). *Distributions in statistics: continuous univariate distributions*, 2ed. New York. Houghton Mifflin.

KANO, Y.; BERKANE, M. e BENTLER, P.M. (1993). Statistical inference based on pseudo maximum likelihood in elliptical population. *Journal of the American Statistical Association*, **84**, 881-896.

KELKER, D. (1970). Distribution theory of spherical distributions and a location scale parameter generalization. *Sankhga A*, **32**, 419-430.

KELLY, G. (1984).The influence function in the errors in variables problem. *The Annals of Statistics*, **12**, 87-100.

LACHOS-DÁVILA, V.H. (2002). Inferência e diagnóstico em modelos de Grubbs . Dissertação de mestrado, IMECC - UNICAMP.

LANGE, K.L., LITTLE, R.J. e TAYLOR, J. (1989). Robust statistical modelling using the t-distribution. *Journal of the American Statistical Association*, **84**, 881-896.

LITTLE, R.J. (1988). Robust estimation of the mean and covariance matrix from data with missing values. *Applied Statistics*, **37**, 23-38.

MCLACHLAN, G.J. e KRISHNAN, T. (1993). *The EM Algorithm and Extensions*. New York: Wiley Series in Probability and Statistics.

MUIRHEAD, R. (1980). The effects of elliptical distributions on some standard procedures involving correlation coefficients. *Multivariate Statistical Analysis*, North-Holland, 143-159.

MUIRHEAD, R. (1982). Aspects of Multivariate Statistical Theory. John Wiley. New York.

PEÑA, D. e YOHAI, V. (1995). The detection of influential subsets in linear regression by using an influence matrix. *Journal of Royal Statistical Society*, B, **57**, 145-156.

RAO, C.R. (1948). Large sample tests of statistical hypotesis concerning several parameters with aplications to problems of estimation. *Proceeding of the Cambridge Philosophical Society*, **44**, 50-57.

ROUSSEEUW, P.J. e LEROY, A.M. (1987). *Robust Regression and Outlier Detection*, John Wiley. New York.

RUBIN, D.B. (1983). Iteratively reweighted least squares. *Encyclopedia of the Statistical Sciences*, **4**, 272-275.

SEN, P.K. e SINGER, J.M. (1993). *Large Sample Methods in Statistics: An Introduction With Applications*, Chapman and Hall. New York.

SEN, P.K. e SRIVASTAVA, M.S. (1975). On tests for detecting change in mean. *The Annals of Statistics,* **3**, 98-108.

SHYR, J.Y. e GLESER, L.J. (1986). Inference about comparative precision in linear structural relationships. *Journal of the Statistical Planning and Inference*, **14**, 339-358.

STAUDTE, R.G. e SHEATHER, S.J. (1990). *Robust Estimation and Testing*, John Wiley. New York.

SUTRADHAR, B.C. (1986). On the characteristic function of multivariate student t distribution. *Journal of Multivariate Annalysis*, **46**, 1-12.

SUTRADHAR, B.C. (1993). Score tests for the covariance matrix of the elliptical *t*-distribution. *Journal of Multivariate Analysis*, **46**, 1-12.

TAYLOR, J. (1992). Properties of modelling the error distribution with an extra shape parameter. *Computational Statistics* & *Data Analysis*, **13**, 33-46.

THEOBALD, C.M. e MALLISON, J.R. (1978). Comparative calibration, linear structural relationship and congeneric measurements. *Biometrics*, **34**, 39-45.

ULLAH, A. e ZINDE-WALSH, V. (1984). On the robustness of LM, LR, and W test in regression models. *Econometrica*, **52(4)**, 1055-1066.

VILCA- LABRA, F.; ARELLANO, R. e BOLFARINE, H.; (1998). Functional Elliptical Models:*Journal of Multivariate Analysis* **64**, 36-57.

VILCA-LABRA, F.; LACHOS, V. e BOLFARINE, H. (2002). On testing statistics for comparing several measuring devices. Relatório de pesquisa, IMECC, Unicamp. RP51/02.

WANG, J.L. e WANG, J. (1994). The test and confidence interval for a change-point in mean vector of multivariate normal distribution. *Multivariate Analysis and its Applications*, **92**, 171-178.

WELLMAN, J.M. e GUNST, R.F. (1991). Influence diagnostics for linear measurement error models. *Biometrika*, **78**, 373-380.

WORSLEY, K.J. (1979). On the likelihood ratio test for a shift in location of normal populations, *Journal of the American Statistical Association*, **74**, 365-367.

ZAMAR, R.H. (1989). Robust estimation in the errors-in-variables model. *Biometrika*, **76**, 149-160.

ZELLNER, A. (1976). Bayesian and non-bayesian analysis of the regression model with multivariate student t-error term. *Journal of the American Statistics Association*, **71**, 400-405.