

Markus Chagas Stein

*Testes da Razão de Verossimilhanças em Modelos
Lineares Mistos*

Belo Horizonte - MG

Dezembro / 2008

Markus Chagas Stein

*Testes da Razão de Verossimilhanças em Modelos
Lineares Mistos*

Dissertação apresentada como requisito parcial
para obtenção do grau de Mestre em Estatística
pela Universidade Federal de Minas Gerais.

Orientador: Prof. Dr. Michel Ferreira da Silva

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA
INSTITUTO DE CIÊNCIAS EXATAS
UNIVERSIDADE FEDERAL DE MINAS GERAIS

Belo Horizonte – MG

Dezembro / 2008

*Aos meus pais, Luiz Henrique e Denise,
a minha futura esposa, Camille, e a minha filha, Ana,
dedico este trabalho.*

Agradecimentos

Primeiramente agradeço a força divina que sempre guiou meus passos e me fez seguir perseverante até a conclusão deste trabalho.

Ao professor Michel Ferreira da Silva pela orientação, amizade e por todo o esforço que demandou para ver o desfecho deste trabalho. Sou grato também pela sua contribuição em minha formação acadêmica e pessoal.

Aos demais professores do Departamento de Estatística da UFMG que contribuíram para minha formação acadêmica. Em especial aos professores Luiz Duczmal, Gregório Atuncar, Lourdes Montenegro e Enrico Colosimo pela oportunidade que tive de trabalhar com cada um.

Aos amigos que me acolheram quando cheguei em Belo Horizonte, Fábio, Carlito, Luciano, Max e Ronaldo, pela convivência “quase” sempre harmoniosa e pelos bons momentos que compartilhamos. E aos demais colegas do Mestrado e do Doutorado pela amizade compartilhada neste dois anos.

Aos meus conterrâneos e grandes amigos, Rodrigo e Márcia, pelas experiências que juntos passamos, pela amizade e pelo convívio muito divertido.

Aos meus pais, que contribuíram de forma incondicional para minha formação pessoal e por nunca terem medido esforços para que eu pudesse continuar nesta caminhada. Agradeço também pela incomensurável ajuda que deram na criação de minha filha neste longo período que estive ausente.

A minha namorada, Camille, que se mostrou uma heróica companheira, agradeço pelo apoio moral, pela paciência e sobre tudo pelo papel que teve de desempenhar durante estes dois anos suprindo minha falta na criação de nossa filha.

A FAPEMIG, pelo apoio financeiro.

Aos participantes da banca examinadora pelas sugestões.

Resumo

Freqüentemente modelos lineares mistos são utilizados para análise de medidas repetidas. Em particular, dados longitudinais podem ser ajustados por tais modelos. Dependendo da quantidade de covariáveis, da estrutura de covariâncias considerada e do tamanho amostral, inferência baseada na função de verossimilhança sobre parte do vetor paramétrico pode ser bastante influenciada pelos demais parâmetros desconhecidos como, por exemplo, o teste da razão de verossimilhanças. Neste trabalho consideramos três aperfeiçoamentos deste teste para inferência sobre os efeitos fixos em modelos lineares mistos: correção de Bartlett, um teste baseado no método *bootstrap* e um teste da razão de verossimilhanças perfiladas modificadas. Aqui utilizamos uma aproximação, baseada em Severini (1998), para a modificação da função de verossimilhança perfilada proposta por Barndorff-Nielsen (1983). Simulações de Monte Carlo foram realizadas e, em todos os casos simulados, as taxas de rejeição observadas sob a hipótese nula dos três testes alternativos considerados foram mais próximas dos níveis nominais do que as taxas relativas ao teste usual.

Abstract

Mixed linear models are often used for the analysis of repeated measures. In particular, longitudinal data can be adjusted by such models. Depending on the amount of covariates, the structure of covariance considered and the sample size, inference based on the likelihood function on part of the parametric vector can be strongly influenced by other unknown parameters, such as the likelihood ratio test. In this work we consider three improvements of this test for fixed effects inference in mixed linear models: Bartlett correction, a test based on the bootstrap method and an adjusted profile likelihood ratio test. Here we use an approximation to the modified profile likelihood proposed by Barndorff-Nielsen (1983), the approximation having been obtained using the results by Severini (1998). Monte Carlo simulations were performed, and in all cases simulated, the three alternative tests displayed smaller size distortions than the usual likelihood ratio test.

Sumário

Lista de Figuras		p. vii
Lista de Tabelas		p. ix
1 Introdução		p. 1
1.1 Introdução		p. 1
1.2 Organização da Dissertação		p. 2
1.3 Recursos Computacionais		p. 3
2 Modelo Linear Misto Geral		p. 5
2.1 Introdução		p. 5
2.2 Modelo Linear Misto Geral		p. 6
2.3 Exemplo		p. 7
3 Refinamentos do Teste da Razão de Verossimilhanças		p. 9
3.1 Introdução		p. 9
3.2 Função de Verossimilhança Perfilada		p. 10
3.3 Verossimilhança Perfilada Modificada Proposta por Barndorff-Nielsen		p. 11
3.4 Aproximação para a Função de Verossimilhança Proposta por Barndorff-Nielsen		p. 12
3.5 Exemplos		p. 12
3.6 Teste da Razão de Verossimilhanças Baseado no Método <i>Bootstrap</i>		p. 18
3.7 Correção de Bartlett para a Estatística da Razão de Verossimilhanças		p. 19
4 Ajuste de Barndorff-Nielsen no Modelo Linear de Efeitos Mistos Geral		p. 21
4.1 Introdução		p. 21

4.2	Modelo Misto Linear Geral	p. 21
4.3	Exemplos	p. 23
5	Resultados Numéricos	p. 28
5.1	Introdução	p. 28
5.2	Modelo com Intercepto Aleatório	p. 29
5.3	Modelo com Intercepto e Inclinação Aleatórios	p. 38
6	Conclusões	p. 46
	Apêndice A - Cálculos para os Ajustes	p. 48
	Derivadas do Modelo Misto Linear Geral	p. 48
	Apêndice B - Gráficos de Discrepâncias Relativas de Quantis	p. 49
	Modelo com Intercepto Aleatório	p. 49
	Modelo com Intercepto e Inclinação Aleatórios	p. 53
	Apêndice C - Programa Computacional	p. 58
	Referências Bibliográficas	p. 78

Lista de Figuras

- 5.1 Gráfico das discrepâncias relativas de quantis para $N = 5$ e $\rho = 0,8$ no modelo com intercepto aleatório ($H_0 : \beta_3 = 0$). p. 33
- 5.2 Gráfico das discrepâncias relativas de quantis para $N = 10$ e $\rho = 0,8$ no modelo com intercepto aleatório ($H_0 : \beta_3 = 0$). p. 34
- 5.3 Gráfico das discrepâncias relativas de quantis para o teste de $H_0 : (\beta_2, \beta_3) = (0, 0)$ contra $H_1 : (\beta_2, \beta_3) \neq (0, 0)$, com $N = 5$, $\rho = 0,95$ e $\tau^2 = 1,05$ no modelo com intercepto aleatório. p. 37
- 5.4 Gráfico das discrepâncias relativas de quantis para $N = 15$, $\alpha_2 = 0$ e $\alpha_3 = 0,5$ no modelo com intercepto e inclinação aleatórios ($H_0 : \beta_3 = 0$). p. 42
- 5.5 Gráfico das discrepâncias relativas de quantis para $N = 20$, $\alpha_2 = 0$ e $\alpha_3 = 0,5$ no modelo com intercepto e inclinação aleatórios ($H_0 : \beta_3 = 0$). p. 43
- 5.6 Gráfico das discrepâncias relativas de quantis para o teste de $H_0 : (\beta_2, \beta_3) = (0, 0)$ contra $H_1 : (\beta_2, \beta_3) \neq (0, 0)$, com $N = 15$, $\alpha_2 = 0$ e $\alpha_3 = 0,5$ no modelo com intercepto e inclinação aleatórios. p. 44
- 6.1 Gráfico das discrepâncias relativas de quantis para $N = 5$ e $\rho = 0,9$ ($H_0 : \beta_3 = 0$), no modelo com intercepto aleatório. p. 49
- 6.2 Gráfico das discrepâncias relativas de quantis para $N = 5$ e $\rho = 0,95$ ($H_0 : \beta_3 = 0$), no modelo com intercepto aleatório. p. 50
- 6.3 Gráfico das discrepâncias relativas de quantis para $N = 10$ e $\rho = 0,9$ ($H_0 : \beta_3 = 0$), no modelo com intercepto aleatório. p. 50
- 6.4 Gráfico das discrepâncias relativas de quantis para $N = 10$ e $\rho = 0,95$ ($H_0 : \beta_3 = 0$), no modelo com intercepto aleatório. p. 51
- 6.5 Gráfico das discrepâncias relativas de quantis para $N = 5$, $\rho = 0,95$ e $\tau^2 = 10$ ($H_0 : \beta_3 = 0$), no modelo com intercepto aleatório. p. 51
- 6.6 Gráfico das discrepâncias relativas de quantis para $N = 10$, $\rho = 0,5$ ($H_0 : \beta_3 = 0$), no modelo com intercepto aleatório. p. 52

- 6.7 Gráfico das discrepâncias relativas de quantis para $N = 15$, $\alpha_2 = 0$ e $\alpha_3 = 1$ ($H_0 : \beta_3 = 0$), no modelo com intercepto e inclinação aleatórios. p. 53
- 6.8 Gráfico das discrepâncias relativas de quantis para $N = 15$, $\alpha_2 = 0,35$ e $\alpha_3 = 0,5$ ($H_0 : \beta_3 = 0$), no modelo com intercepto e inclinação aleatórios. p. 53
- 6.9 Gráfico das discrepâncias relativas de quantis para $N = 15$, $\alpha_2 = 0,35$ e $\alpha_3 = 1$ ($H_0 : \beta_3 = 0$), no modelo com intercepto e inclinação aleatórios. p. 54
- 6.10 Gráfico das discrepâncias relativas de quantis para $N = 20$, $\alpha_2 = 0$ e $\alpha_3 = 1$ ($H_0 : \beta_3 = 0$), no modelo com intercepto e inclinação aleatórios. p. 54
- 6.11 Gráfico das discrepâncias relativas de quantis para $N = 20$, $\alpha_2 = 0,35$ e $\alpha_3 = 0,5$ ($H_0 : \beta_3 = 0$), no modelo com intercepto e inclinação aleatórios. p. 55
- 6.12 Gráfico das discrepâncias relativas de quantis para $N = 20$, $\alpha_2 = 0,35$ e $\alpha_3 = 1$ ($H_0 : \beta_3 = 0$), no modelo com intercepto e inclinação aleatórios. p. 55
- 6.13 Gráfico das discrepâncias relativas de quantis para $N = 15$, $\alpha_2 = 0$ e $\alpha_3 = 1$ ($H_0 : \beta_3 = 0$ e $\beta_2 = 0$), no modelo com intercepto e inclinação aleatórios. p. 56
- 6.14 Gráfico das discrepâncias relativas de quantis para $N = 15$, $\alpha_2 = 0,35$ e $\alpha_3 = 0,5$ ($H_0 : \beta_3 = 0$ e $\beta_2 = 0$), no modelo com intercepto e inclinação aleatórios. p. 56
- 6.15 Gráfico das discrepâncias relativas de quantis para $N = 15$, $\alpha_2 = 0,35$ e $\alpha_3 = 1$ ($H_0 : \beta_3 = 0$ e $\beta_2 = 0$), no modelo com intercepto e inclinação aleatórios. p. 57

Lista de Tabelas

5.1	Estatísticas descritivas das estimativas para β_3 no modelo com intercepto aleatório.	p. 30
5.2	Taxas de rejeição (%) da hipótese nula ($H_0 : \beta_3 = 0$) calculadas a partir de 3000 réplicas de Monte Carlo e 200 amostras <i>bootstrap</i> no modelo com o intercepto aleatório.	p. 31
5.3	Médias e variâncias amostrais de RV , RV_{ABN} e RV^* e seus respectivos valores assintóticos, no modelo com o intercepto aleatório.	p. 32
5.4	Taxas de rejeição (%) da hipótese nula ($H_0 : \beta_3 = 0$) calculadas a partir de 3000 réplicas de Monte Carlo e 100, 200 e 500 amostras <i>bootstrap</i> , para $N = 5$ e $\rho = 0,95$ no modelo com o intercepto aleatório.	p. 35
5.5	Taxas de rejeição (%) da hipótese nula ($H_0 : \beta_3 = 0$) calculadas a partir de 3000 réplicas de Monte Carlo e 200 amostras <i>bootstrap</i> , utilizando $\tau^2 = 10$, $N = 5$ e $\rho = 0,95$ no modelo com o intercepto aleatório.	p. 36
5.6	Taxas de rejeição (%) da hipótese nula ($H_0 : \beta_3 = 0$) calculadas a partir de 3000 réplicas de Monte Carlo e 200 reamostras <i>bootstrap</i> , utilizando $\rho = 0,5$ e $N = 10$ no modelo com o intercepto aleatório.	p. 36
5.7	Taxas de rejeição (%) da hipótese nula ($H_0 : (\beta_2, \beta_3) = (0,0)$) calculadas a partir de 3000 réplicas de Monte Carlo e 200 reamostras <i>bootstrap</i> , utilizando $\tau^2 = 1,05$, $\rho = 0,95$ e $N = 5$ no modelo com o intercepto aleatório.	p. 37
5.8	Estatísticas descritivas das estimativas para β_3 no modelo com intercepto e inclinação aleatórios.	p. 39
5.9	Taxas de rejeição (%) da hipótese nula ($H_0 : \beta_3 = 0$) calculadas a partir de 3000 réplicas de Monte Carlo no modelo com intercepto e inclinação aleatórios. . . .	p. 40
5.10	Médias e variâncias amostrais de RV , RV_{ABN} e seus respectivos valores assintóticos, no modelo com intercepto e inclinação aleatórios.	p. 41
5.11	Taxas de rejeição (%) da hipótese nula ($H_0 : (\beta_2, \beta_3) = (0,0)$) calculadas a partir de 3000 réplicas de Monte Carlo, para $N = 15$ no modelo com intercepto e inclinação aleatórios.	p. 43

5.12 Médias e variâncias amostrais de RV , RV_{ABN} e seus respectivos valores assintóticos, para o teste de $H_0 : (\beta_2, \beta_3) = (0, 0)$ contra $H_1 : (\beta_2, \beta_3) \neq (0, 0)$ considerando $N = 15$ p. 45

1 Introdução

1.1 Introdução

Quando construímos um modelo estatístico para explicar relações entre variáveis, geralmente temos interesse em inferir sobre apenas parte do vetor paramétrico para a compreensão do fenômeno em estudo. Neste caso podemos definir um vetor de parâmetros de interesse (μ) e um vetor de parâmetros de incômodo (ν) tal que $\theta = (\mu, \nu)$ é o vetor de parâmetros do modelo.

Dependendo da dimensão de ν , relativamente ao tamanho amostral, as inferências para μ podem ser bastante afetadas. Quando o tamanho amostral diminui, esta influência aumenta. Por exemplo, o teste da razão de verossimilhanças de $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$ tem estatística dada por $RV = 2 \log [L(\hat{\mu}, \hat{\nu}) / L(\mu_0, \hat{\nu}_{\mu_0})]$, onde $\hat{\mu}$ e $\hat{\nu}$ são as estimativas de máxima verossimilhança de μ e ν , respectivamente, e $\hat{\nu}_{\mu_0}$ é a estimativa de máxima verossimilhança de ν sob H_0 . Sob esta hipótese, a distribuição assintótica dessa estatística de teste é qui-quadrado com grau de liberdade igual a dimensão de μ . Se o tamanho amostral for pequeno, a aproximação da distribuição amostral de RV pela qui-quadrado pode não ser satisfatória, acarretando um teste com taxas de rejeição bastante distorcidas. Note que essa estatística de teste é baseada na verossimilhança perfilada $L_p(\mu) = L(\mu, \hat{\nu}_\mu)$, onde $\hat{\nu}_\mu$ é a estimativa de máxima verossimilhança de ν para μ fixado, pois $L(\hat{\mu}, \hat{\nu}) = L_p(\hat{\mu})$ e $L(\mu_0, \hat{\nu}_{\mu_0}) = L_p(\mu_0)$.

$L_p(\mu)$ não é obtida diretamente de uma distribuição conjunta, fazendo com que não tenha algumas propriedades interessantes e inerentes às funções de verossimilhanças genuínas. Por exemplo, a esperança da função score correspondente não será necessariamente igual a zero e a informação poderá ser viesada (ou seja, a igualdade da informação poderá não ser observada). Portanto, ajustes na função de verossimilhança perfilada são necessários e foram propostos a fim de se reduzir os problemas provenientes do fato desta função não ser uma verossimilhança genuína. Com tais ajustes, pretendemos melhorar a taxa de convergência da distribuição amostral de RV para a distribuição qui-quadrado. Assim, no caso de pequenas amostras, teremos taxas de rejeição sob H_0 mais próximas dos níveis nominais.

Zucker, Lieberman e Manor (2000) apresentaram resultados numéricos relativos a testes da razão de verossimilhanças sobre um parâmetro dos efeitos fixos do modelo linear misto baseados na verossimilhança perfilada modificada proposta por Cox e Reid (1987), considerando tamanhos

amostrais relativamente pequenos. Sob a hipótese nula, tais testes foram sempre conservadores e, em algumas situações, apresentaram taxas de rejeição bem menores do que os níveis nominais. Nestes casos, o teste usual apresentou taxas de rejeição bem maiores do que os níveis nominais. Diferentemente daquele, este teste foi sempre liberal. Os autores também consideraram correção de Bartlett(1937) para o teste da razão de verossimilhanças usual e para o teste baseado no ajuste de Cox e Reid (1987). Os fatores de correção foram obtidos analiticamente e as taxas de rejeição sob a hipótese nula de ambos os testes ficaram próximas dos níveis nominais.

Neste trabalho, optamos pelo uso da verossimilhança perfilada modificada proposta por Barndorff-Nielsen (1983), mais precisamente, uma aproximação para esta função baseada em Severini (1998). Considerando situações semelhantes àquelas apresentadas em Zucker, Lieberman e Manor (2000), o teste baseado nesse ajuste apresentou taxas de rejeição sob a hipótese nula próximas dos níveis nominais, sem a necessidade da aplicação da correção de Bartlett. Esta é a principal contribuição deste trabalho. Também consideramos um teste baseado no método *bootstrap* (Efron, 1979) e usamos o esquema de reamostragem desse método para obter numericamente o fator de correção de Bartlett para a estatística de teste usual, diferentemente de Zucker, Lieberman e Manor (2000). Esses testes também apresentaram taxas de rejeição próximas aos níveis nominais e representam outra contribuição desta dissertação.

Avaliamos o desempenho desses testes para um modelo de regressão que possui apenas um efeito aleatório, aqui denominado de modelo com intercepto aleatório. Neste caso, as observações para o mesmo indivíduo são igualmente correlacionadas (Diggle et al., 1994). Esta situação não foi considerada por Zucker, Lieberman e Manor (2000). Também apresentamos resultados numéricos para outro modelo que, além do intercepto aleatório, incorpora o tempo da observação como um efeito aleatório. Assim a distância no tempo entre duas observações influencia a correlação entre elas. Este foi o caso considerado por Zucker, Lieberman e Manor (2000), aqui denominado de modelo com intercepto e inclinação aleatórios. Os ajustes encontrados neste trabalho referem-se a classe geral de modelos lineares mistos, podendo ser aplicados em diversos contextos. Porém temos interesse em estudos longitudinais pois, em muitas situações, são estudos caros e/ou envolvem um pequeno número de indivíduos observados, surgindo assim a necessidade de melhorarmos os testes de hipóteses baseados em resultados assintóticos.

1.2 Organização da Dissertação

A presente dissertação está dividida em seis capítulos e três apêndices. No Capítulo 2, são apresentados conceitos relacionados aos modelos lineares mistos, tais como a formulação do modelo e aspectos relacionados à inferência sobre os parâmetros deste modelo. Além disso, discutimos algumas dificuldades computacionais que podem ser encontradas ao maximizar numericamente a função de verossimilhança e que observamos ao longo deste estudo.

No terceiro capítulo, consideramos refinamentos para o teste da razão de verossimilhanças usual. Apresentamos o conceito de verossimilhança perfilada, a modificação proposta por Barndorff-Nielsen (1983), bem como uma aproximação para esta proposta sugerida por Severini (1998). Apresentamos também exemplos de obtenção dessas funções no contexto do modelo normal. Inicialmente, supomos observações independentes e identicamente distribuídas e, posteriormente, supomos observações apenas independentes. Outros conceitos, como correção de Bartlett (1937) e teste *bootstrap* (Efron, 1979), utilizados como refinamentos do teste da razão de verossimilhanças usual, também são mostrados.

No Capítulo 4, para o modelo linear misto geral, derivamos a função de verossimilhança perfilada e a modificação proposta por Barndorff-Nielsen (1983). Em seguida, calculamos essas funções de verossimilhança em dois particulares modelos lineares mistos, que serão considerados na avaliação numérica do desempenho dos testes.

No quinto capítulo, apresentamos resultados numéricos relativos a simulações de Monte Carlo. Comparamos o desempenho das funções de verossimilhança usual e verossimilhança perfilada ajustada, tanto no caso de estimação pontual (método da máxima verossimilhança) quanto em testes de hipóteses (teste da razão de verossimilhanças). Em se tratando de testes de hipóteses, também comparamos o teste usual corrigido via Bartlett e o teste *bootstrap*. Por fim, no Capítulo 6, apresentamos as conclusões obtidas com este trabalho e sugerimos tópicos de pesquisa que podem vir a ser desenvolvidos a partir desta dissertação.

No Apêndice A, mostramos algumas derivadas utilizadas nos termos da aproximação para a modificação proposta por Barndorff-Nielsen (1983). No Apêndice B, apresentamos vários gráficos de discrepâncias relativas entre os quantis amostrais e os quantis assintóticos de estatísticas da razão de verossimilhanças. No Apêndice C, encontra-se o código do programa implementado na linguagem de programação `0x`, que foi utilizado para obtenção dos resultados numéricos.

1.3 Recursos Computacionais

Os resultados numéricos apresentados neste trabalho foram obtidos utilizando-se a versão 5.10 da linguagem de programação `0x` para o sistema operacional Windows. Esta foi desenvolvida por Jurgen A. Doornik com base na linguagem de programação `C`. Versões desta linguagem, para uso acadêmico, podem ser encontradas em <http://www.doornik.com>. Uma documentação introdutória está em Doornik e Ooms (2007), maiores detalhes podem ser vistos em Doornik (2007).

Os resultados gráficos e o tratamento dos dados foram realizados na linguagem `R`, versão 2.6.2. Esta linguagem foi desenvolvida por Ross Ihaka e Robert Gentleman (1996), sendo um ambiente que proporciona facilidades para manipulação de dados, obtenção de gráficos, bem como análises estatísticas em geral. Pode ser obtida gratuitamente no endereço <http://www.r-project.org>.

Para a digitação desta dissertação foi utilizado o programa $T_E X$. Este programa foi desenvolvido por Donald Knuth, no final dos anos 70, e é um programa que pode ser utilizado na edição de textos com excelente apresentação gráfica. Uma implementação do $T_E X$ para o sistema operacional Windows, o $MiKTeX$, encontra-se disponível no endereço <http://miktex.org>. Informações sobre este programa podem ser obtidas em Knuth (1986).

2 *Modelo Linear Misto Geral*

2.1 Introdução

Em determinadas situações, podemos estar interessados em analisar um conjunto de dados obtidos a partir de múltiplas mensurações sobre a mesma unidade experimental. Esta classe de estudos é denominada análise de medidas repetidas (Diggle et al., 1994).

Em particular, análises de dados longitudinais têm essa característica, pois, por exemplo, podemos estar interessados em avaliar o comportamento de determinada variável ao longo do tempo, bem como possíveis relações com covariáveis. Assim esperamos que as medidas repetidas ao longo do tempo para a mesma unidade experimental sejam correlacionadas. Tal correlação deve ser modelada através de uma estrutura de covariâncias apropriada para que a inferência sobre a média da variável resposta seja válida.

Uma maneira de se incluir num modelo a correlação entre as medidas da mesma unidade experimental é através dos modelos lineares mistos, apresentados, por exemplo, em Harville (1977). No caso de dados longitudinais, Laird e Ware (1982) utilizam o modelo em dois estágios. Por este modelo, os vetores de observações para as diferentes unidades experimentais seguem a mesma família de distribuição (família normal multivariada) e determinados efeitos variam através das unidades experimentais de acordo com alguma distribuição de probabilidades (distribuição normal). Verbeke e Molenberghs (2000) denominam este modelo de modelo linear de efeitos mistos geral e também o utilizam no contexto de dados longitudinais, apresentando resultados de inferência e ajuste do modelo.

Dentre algumas referências para análise de modelos lineares mistos estão Littell et al. (2006) e Pinheiro e Bates (2000). Outras referências que apresentam a teoria dos modelos lineares mistos e ainda fazem a generalização destes modelos para o chamado modelo linear misto generalizado (família exponencial) são McCulloch e Searle (2001), Demidenko (2004) e Jiang (2007).

Neste capítulo apresentamos a formulação do modelo linear misto geral, seguido de um exemplo. Também comentamos sobre alguns aspectos para realização de inferências neste modelo.

2.2 Modelo Linear Misto Geral

Seja a variável aleatória y_{ij} definida como a resposta da i -ésima unidade experimental medida no instante t_{ij} , para todo $i = 1, \dots, N$ e $j = 1, \dots, n_i$. O vetor aleatório $\mathbf{Y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^\top$, de dimensão $n_i \times 1$, representa todas as observações da unidade i . Define-se o modelo linear misto geral como

$$\mathbf{Y}_i = X_i\beta + Z_i\mathbf{b}_i + \epsilon_i, \quad (2.1)$$

onde X_i ($n_i \times p$) é a matriz de delineamento dos efeitos fixos (covariáveis de efeitos fixos), Z_i ($n_i \times q$) é a matriz de delineamento dos efeitos aleatórios (covariáveis de efeitos aleatórios), β ($p \times 1$) é o vetor desconhecido de coeficientes associados aos efeitos fixos, \mathbf{b}_i ($q \times 1$) é o vetor referente aos efeitos aleatórios, ϵ_i ($n_i \times 1$) é o vetor de erros aleatórios e $\mathbf{b}_1, \dots, \mathbf{b}_N, \epsilon_1, \dots, \epsilon_N$ são independentes. Assumimos que $\mathbf{b}_i \sim Normal_q(\mathbf{0}, D)$ e $\epsilon_i \sim Normal_{n_i}(\mathbf{0}, V_i), \forall i$, onde D e V_i são matrizes positivas-definidas contendo parâmetros desconhecidos. V_i é uma matriz diagonal.

Condicional aos efeitos b_i , o vetor \mathbf{Y}_i segue uma distribuição *Normal* de média $X_i\beta + Z_i\mathbf{b}_i$ e matriz de covariâncias V_i . Marginalmente, os efeitos fixos influenciam na estrutura de média e os efeitos aleatórios influenciam na estrutura de covariâncias do vetor \mathbf{Y}_i . Então temos que $\mathbf{Y}_i \sim Normal_{n_i}(X_i\beta, \Sigma_i)$, onde $\Sigma_i = Z_i D Z_i^\top + V_i, \forall i$.

Neste trabalho, adotamos a abordagem marginal, assim, para $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_N)^\top$, temos o modelo

$$\mathbf{Y} = X\beta + Z\mathbf{b} + \epsilon, \quad (2.2)$$

onde \mathbf{Y} é um vetor aleatório de tamanho $T \times 1$, $T = \sum_{i=1}^N n_i$, com distribuição *Normal* $_T(X\beta, \Sigma)$. Aqui $\mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_N)^\top$, $\epsilon = (\epsilon_1, \dots, \epsilon_N)^\top$ e $X = (X_1, \dots, X_N)^\top$ são obtidos empilhando-se os vetores \mathbf{b}_i e ϵ_i e as matrizes X_i , respectivamente. $\Sigma = Z(I_N \otimes D)Z^\top + V$, onde Z e V são matrizes bloco-diagonais com blocos iguais a (Z_1, \dots, Z_N) e (V_1, \dots, V_N) , respectivamente, ou, simplesmente, Σ é uma matriz bloco-diagonal com blocos $\Sigma_i, \forall i$. I_N denota a matriz identidade de ordem N e \otimes denota o produto de Kronecker entre matrizes.

Aqui o vetor de parâmetros presentes na matriz Σ é denotado por $\alpha = (\alpha_1, \dots, \alpha_m)$. Ou seja, $\Sigma = \Sigma(\alpha)$ e $\Sigma_i = \Sigma_i(\alpha), \forall i$. Assim o vetor de parâmetros do modelo marginal (2.2) é $\theta = (\beta, \alpha)$ e o espaço paramétrico é dado por $\Theta = \Theta_\beta \times \Theta_\alpha$, com $\Theta_\beta \subseteq R^p$ e Θ_α o conjunto de todos os valores de α tais que Σ seja positiva-definida.

Baseado no modelo definido em (2.2), o logaritmo da função de verossimilhança de θ é dada por

$$\ell(\theta) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (\mathbf{y} - X\beta)^\top \Sigma^{-1} (\mathbf{y} - X\beta), \quad (2.3)$$

onde $|\Sigma|$ denota o determinante da matriz Σ .

Para a estimação pontual de θ , geralmente são usados os estimadores de máxima verossimilhança (*EMV*) ou os estimadores de máxima verossimilhança restrita (*EMVR*). Estes últimos são obtidos a partir da maximização da seguinte função:

$$\ell_{EMVR}(\theta) = -\frac{1}{2} \log \left| \sum_{i=1}^N X_i^\top \Sigma_i^{-1}(\alpha) X_i \right| + \ell(\theta). \quad (2.4)$$

O estimador de máxima verossimilhança restrita de Σ é não viesado, diferentemente do estimador de máxima verossimilhança.

Na literatura referente a modelos lineares mistos encontram-se alguns métodos de maximização numérica, como o de Newton-Raphson, para encontrar os *EMV* e *EMVR*. Embora os modelos lineares mistos sejam flexíveis para acomodar a estrutura de covariâncias dos dados, os ajustes desses modelos envolvem grandes dificuldades computacionais na estimação dos parâmetros. Lindstrom e Bates (1988) desenvolveram uma implementação eficiente baseada no algoritmo de Newton-Raphson. Além de reduzir o tempo computacional, os autores mostraram que uma das vantagens sobre os métodos existentes até então consiste em assegurar, a cada iteração, que a matriz de covariâncias do modelo seja positiva-definida.

Um aspecto que também dificulta a estimação de parâmetros de um modelo linear misto são os dados incompletos. Laird e Ware (1982) fazem uma discussão acerca de obter estimativas de máxima verossimilhança na ausência de algumas observações e apresentam como solução uma implementação do algoritmo EM. Outros problemas são relatados em Verbeke e Molenberghs (2000).

Bates e DebRoy (2003) apresentam um método de estimação dos parâmetros do efeito fixo denominado de mínimos quadrados penalizados. Este método é capaz de produzir estimadores com formas fechadas, diminuindo assim o custo computacional envolvido na obtenção de estimativas.

2.3 Exemplo

Ilustrações do uso de modelos mistos em dados longitudinais podem ser encontrados em Verbeke e Molenberghs (2000). Dentre essas, uma refere-se a um experimento realizado no Departamento de Ortodontia da Universidade Católica de Leuven na Bélgica. Iremos transcrever parte do experimento.

Atualmente, em Ciências Médicas, está ocorrendo um aumento no interesse do uso terapêutico de hormônios. No entanto, essas terapias drásticas exigem conhecimento detalhado dos seus efeitos sobre os diferentes aspectos do crescimento. No estudo citado, o principal objetivo foi investigar o efeito da inibição da produção de testosterona em ratos machos no seu crescimento

craniofacial. Um total de 50 ratos machos foram aleatorizados em três grupos: grupo controle; grupo que recebeu uma dose baixa da droga Decapeptyl, que é um inibidor da produção de testosterona em ratos; e grupo que recebeu uma dose alta da droga Decapeptyl.

A variável resposta que Verbeke e Molemberghs (2000) consideraram (apenas uma variável do estudo original) corresponde a uma medição que pode ser usada para caracterizar a altura do crânio.

O tratamento começou quando os ratos completaram 45 dias de idade e a primeira medição foi realizada com a idade de 50 dias. As medições foram realizadas a cada 10 dias e a última medição foi feita quando os ratos completaram 110 dias.

O principal interesse é estimar o efeito das mudanças de crescimento ao longo do tempo e testar se estas mudanças são dependentes do tratamento.

Nem todos os ratos têm medições até à idade de 110 dias. Isto se deve ao fato de que muitos ratos não sobreviveram à anestesia e, portanto, saíram antes do final do estudo. Enquanto que 50 ratos foram aleatorizados no início do experimento, somente 22 deles sobreviveram às primeiras 6 medições.

Um modelo sugerido para a resposta de cada rato é dado por

$$Y_{ij} = \beta_0 + \beta_1 L_i t_{ij} + \beta_2 H_i t_{ij} + \beta_3 C_i t_{ij} + b_{i0} + b_{i1} t_{ij} + \epsilon_{ij},$$

para todo $j = 1, \dots, n_i$, onde t_{ij} é o tempo da observação, L_i , H_i e C_i são variáveis indicadoras que assumem o valor um se o rato pertence ao grupo que recebeu dose baixa, dose alta ou é do grupo controle, respectivamente, e zero caso contrário. E os erros, ϵ_{ij} , são independentes e identicamente distribuídos segundo uma distribuição normal com média zero e variância σ^2 .

A covariância entre duas observações do mesmo indivíduo é dada por,

$$\begin{aligned} Cov(Y_{ir}, Y_{is}) &= \begin{pmatrix} 1 & t_{ir} \end{pmatrix} D \begin{pmatrix} 1 \\ t_{is} \end{pmatrix} \\ &= d_{22} t_{ir} t_{is} + d_{12} (t_{ir} + t_{is}) + d_{11}, \end{aligned}$$

onde d_{11} , d_{12} , d_{22} são elementos da matriz D , matriz de covariâncias dos efeitos aleatórios. Aqui, para efeitos de interpretação do estudo, necessitaríamos estimar os parâmetros dos efeitos fixos e testar algumas hipóteses.

3 Refinamentos do Teste da Razão de Verossimilhanças

3.1 Introdução

O nosso objetivo nesta dissertação é realizar e comparar o desempenho de testes da razão de verossimilhanças sobre parte do vetor de parâmetros dos efeitos fixos em modelos lineares mistos. Esses são nossos parâmetros de interesse. Denotando tais parâmetros por μ , vimos, no capítulo introdutório deste trabalho, que a estatística da razão de verossimilhanças para testar $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$ é dada por

$$RV = 2 \log [L(\hat{\mu}, \hat{\nu}) / L(\mu_0, \hat{\nu}_{\mu_0})] = 2 \log [L_p(\hat{\mu}) / L_p(\mu_0)], \quad (3.1)$$

onde $\hat{\mu}$ e $\hat{\nu}$ são os *EMV* de μ e ν , respectivamente, $\hat{\nu}_{\mu_0}$ é o *EMV* de ν sob H_0 e $L_p(\mu) = L(\mu, \hat{\nu}_\mu)$ é a função de verossimilhança perfilada. Sob a hipótese nula, a distribuição assintótica da estatística RV é qui-quadrado com grau de liberdade igual a dimensão de μ . Vale ressaltar que se forem utilizados os *EMVR* ao invés dos *EMV*, esse resultado assintótico não é válido. Caso haja interesse em testar os parâmetros da estrutura de covariâncias, ver Crainiceanu e Ruppert (2004).

Quando modelos são ajustados pelo método da máxima verossimilhança restrita (*MVR*), as comparações dos modelos através dos testes da razão de verossimilhanças restritas devem ser limitadas aos modelos que podem ser encaixados de acordo com os parâmetros da matriz de covariâncias. Dois modelos que têm diferentes matrizes X , não podem ser comparados com base em *MVR* (Littel et al., 2006).

Um método baseado no teste da razão de verossimilhanças sobre os parâmetros dos efeitos fixos utilizando *EMVR* foi apresentado por Welham e Thompson (1997). Eles também obtiveram a correção de Bartlett (1937) para o teste.

Para pequenas amostras, em geral, o desempenho do teste da razão de verossimilhanças usual não é satisfatório. Neste capítulo, apresentamos três métodos de aperfeiçoamento desse teste: correção de Bartlett, *bootstrap* e verossimilhança perfilada modificada.

Na Seção 3.2 definimos a função de verossimilhança perfilada e destacamos suas deficiências.

Na Seção 3.3 apresentamos uma modificação da função de verossimilhança perfilada proposta por Barndorff-Nielsen (1983) e na Seção 3.4 mostramos uma aproximação para esta modificação baseada em covariâncias de derivadas do logaritmo da função de verossimilhança. A Seção 3.5 contém exemplos de obtenção destes ajustes no modelo normal: (i) no caso de observações independentes e identicamente distribuídas e (ii) no caso do modelo de regressão normal linear. Nas Seções 3.6 e 3.7 apresentamos uma breve introdução sobre testes baseados no método *bootstrap* e sobre correções de testes de hipóteses via Bartlett, respectivamente.

3.2 Função de Verossimilhança Perfilada

Seja $\theta = (\mu, \nu)$ o vetor paramétrico do modelo cuja função de verossimilhança é dada por $L(\theta)$, onde μ é um vetor de dimensão $m \times 1$ e ν é um vetor de dimensão $t \times 1$. Seja $\Theta \subseteq \mathbb{R}^{m+t}$ o espaço paramétrico. Define-se a função de verossimilhança perfilada como

$$L_p(\mu) = L(\mu, \hat{\nu}_\mu),$$

em que $\hat{\nu}_\mu$ é a estimativa de máxima verossimilhança de ν para μ fixado, obtida como raiz da equação $\ell_\nu(\theta) = \partial \ell(\theta) / \partial \nu = 0$, e $\ell(\theta)$ representa o logaritmo da função $L(\theta)$. Note que $L_p(\mu)$ não é uma função de verossimilhança genuína. Portanto, a esperança da função escore perfilada pode ser diferente de zero e a igualdade da informação pode não valer. Ou seja, as igualdades

$$\mathbb{E}[u_p(\mu)] = 0 \quad \text{e} \quad \mathbb{E}\left[u_p(\mu) u_p^\top(\mu)\right] + \mathbb{E}\left[\frac{\partial u_p(\mu)}{\partial \mu^\top}\right] = 0,$$

onde $u_p(\mu) = \partial \log L_p(\mu) / \partial \mu$ é a função escore perfilada, não são necessariamente obedecidas. Mas $L_p(\mu)$ apresenta duas propriedades interessantes: (i) a estimativa de máxima verossimilhança de μ obtida dessa função ($\hat{\mu}_p$) é igual a respectiva estimativa obtida de $L(\mu, \nu)$, (ii) a estatística da razão de verossimilhanças usual é baseada nessa função.

Dependendo da dimensão de ν , relativamente ao tamanho amostral, as inferências para μ podem ser bastante afetadas. Por exemplo, a aproximação da distribuição amostral de RV pela qui-quadrado pode não ser satisfatória, acarretando um teste com taxas de rejeição bastante distorcidas. Portanto, no caso de inferência que envolva a presença de parâmetros de perturbação, é necessário realizar ajustes sobre a verossimilhança perfilada. Ferreira da Silva (2005) faz uma revisão de vários ajustes propostos na literatura. O objetivo dessas modificações consiste em incorporar à verossimilhança perfilada as boas propriedades inerentes a uma verossimilhança genuína.

3.3 Verossimilhança Perfilada Modificada Proposta por Barndorff-Nielsen

Uma proposta de modificação da verossimilhança perfilada bastante consagrada foi sugerida por Barndorff-Nielsen (1983). O ajuste proposto produz uma verossimilhança que é uma aproximação para a verossimilhança condicional ou marginal de μ , se alguma dessas existir. Esta proposta se baseia na aproximação para a função densidade condicional do estimador de máxima verossimilhança do vetor paramétrico do modelo dada uma estatística ancilar, denominada fórmula p^* (Barndorff-Nielsen, 1980). Sua expressão é dada por

$$L_{BN}(\mu) = \left| \frac{\partial \hat{\nu}_\mu}{\partial \hat{\nu}} \right|^{-1} |j_{\nu\nu}(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu}, a)|^{-1/2} L_p(\mu), \quad (3.2)$$

onde $j_{\nu\nu}(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu}, a)$ é a matriz de informação observada relativa ao vetor de parâmetros de incômodo ν , $-\partial^2 \log [L(\mu, \nu; \hat{\mu}, \hat{\nu}, a)] / \partial \nu^2$, avaliada em $(\mu, \hat{\nu}_\mu)$; a é uma estatística ancilar para θ ; e $\partial \hat{\nu}_\mu / \partial \hat{\nu}$ é uma derivada com respeito ao espaço amostral ($\hat{\nu}_\mu$ é uma função de $\hat{\nu}$). $L_{BN}(\mu)$ possui erro relativo de ordem $O(n^{-3/2})$ quando comparada com a verossimilhança marginal ou condicional exatas, caso esta existam. Para realizarmos inferências, é comum trabalharmos com o logaritmo da expressão (3.2), dado por

$$\ell_{BN}(\mu) = \ell_p(\mu) - \frac{1}{2} \log |j_{\nu\nu}(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu}, a)| - \log \left| \frac{\partial \hat{\nu}_\mu}{\partial \hat{\nu}} \right|, \quad (3.3)$$

onde $\ell_p(\mu) = \log L_p(\mu)$. A estimativa de máxima verossimilhança perfilada modificada proposta por Barndorff-Nielsen (1983) é o ponto que maximiza $\ell_{BN}(\mu)$, denotada aqui por $\hat{\mu}_{BN}$. A função escore correspondente a esta verossimilhança é dada por $u_{BN}(\mu) = \partial \ell_{BN}(\mu) / \partial \mu$ e

$$\mathbb{E}[u_{BN}(\mu)] = O(n^{-1}) \quad \text{e} \quad \mathbb{E}\left[u_{BN}(\mu) u_{BN}^\top(\mu)\right] + \mathbb{E}\left[\frac{\partial u_{BN}(\mu)}{\partial \mu^\top}\right] = O(n^{-1}).$$

Como $\hat{\nu}_\mu$ satisfaz a igualdade

$$\frac{\partial \ell(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu}, a)}{\partial \nu} = 0,$$

podemos derivar esta equação com relação a $\hat{\nu}$ e obtermos

$$\frac{\partial^2 \ell(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu}, a)}{\partial \nu^2} \frac{\partial \hat{\nu}_\mu}{\partial \hat{\nu}} + \frac{\partial}{\partial \hat{\nu}} \left(\frac{\partial \ell(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu}, a)}{\partial \nu} \right) = 0.$$

Logo,

$$\frac{\partial \hat{\nu}_\mu}{\partial \hat{\nu}} = j_{\nu\nu}(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu}, a)^{-1} \ell_{\nu; \hat{\nu}}(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu}, a), \quad (3.4)$$

onde $\ell_{\nu; \hat{\nu}}(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu}, a) = \partial [\partial \ell(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu}, a) / \partial \nu] / \partial \hat{\nu}$ é uma derivada com relação ao espaço amostral. Usando (3.4), podemos apresentar (3.3) como

$$\ell_{BN}(\mu) = \ell_p(\mu) + \frac{1}{2} \log |j_{\nu\nu}(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu}_\mu, a)| - \log |\ell_{\nu; \hat{\nu}}(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu}, a)|. \quad (3.5)$$

Nos casos em que os estimadores de máxima verossimilhança dos parâmetros são estatísticas suficientes minimais, não se faz necessária a estatística ancilar a .

A estatística da razão de verossimilhanças perfilada modificada proposta por Barndorff-Nielsen para se testar $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$ é dada por

$$RV_{BN} = 2 \log [L_{BN}(\hat{\mu}_{BN}) / L_{BN}(\mu_0)]. \quad (3.6)$$

3.4 Aproximação para a Função de Verossimilhança Proposta por Barndorff-Nielsen

Para utilizarmos a função de verossimilhança perfilada modificada encontrada em (3.5), temos que conhecer a forma analítica do estimador de máxima verossimilhança do vetor de incômodo, $\hat{\nu}$. Além disso, precisamos especificar a estatística ancilar a , e o termo $\ell_{\nu; \hat{\nu}}(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu}, a)$ requer o conhecimento de ambas as estatísticas. Devido à dificuldade de cálculo deste termo, foram propostas algumas aproximações para o mesmo; para uma revisão, ver Severini (2000).

A aproximação que utilizaremos é a baseada nas covariâncias de derivadas do logaritmo da função de verossimilhança (Severini, 1998). Neste caso, mostra-se que $\ell_{\nu; \hat{\nu}}(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu}, a)$ pode ser aproximada por $I_\nu(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu})$, onde, genericamente,

$$I_\nu(\mu, \nu; \mu_0, \nu_0) = \mathbb{E}_{(\mu_0, \nu_0)} [\ell_\nu(\mu, \nu) \ell_\nu(\mu_0, \nu_0)^T].$$

Dessa forma, podemos aproximar $\ell_{BN}(\mu)$ por

$$\ell_{ABN}(\mu) = \ell_p(\mu) + \frac{1}{2} \log |j_{\nu\nu}(\mu, \hat{\nu}_\mu)| - \log |I_\nu(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu})|. \quad (3.7)$$

O estimador de máxima verossimilhança, neste caso, será denotado por $\hat{\mu}_{ABN}$ e a função escore, por $u_{ABN}(\mu) = \partial \ell_{ABN}(\mu) / \partial \mu$. Esta aproximação possui um erro relativo de ordem $O(n^{-1/2})$.

A estatística da razão de verossimilhanças perfilada modificada baseada em Severini (1998) para se testar $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$ é dada por

$$RV_{ABN} = 2 \log [L_{ABN}(\hat{\mu}_{ABN}) / L_{ABN}(\mu_0)]. \quad (3.8)$$

3.5 Exemplos

Nesta seção apresentaremos duas ilustrações da obtenção das funções descritas anteriormente no contexto do modelo normal. Inicialmente, consideraremos observações independentes e identicamente distribuídas. Posteriormente, teremos apenas observações independentes (modelo de regressão). Para cada situação, obteremos: (i) $\ell_p(\mu)$, $\ell_{BN}(\mu)$ e $\ell_{ABN}(\mu)$, cujas expressões serão constituídas somente por parcelas que dependem dos parâmetros; (ii) estimadores de máxima

verossimilhança dos parâmetros de interesse.

Observações Independentes e Identicamente Distribuídas

Seja y variável aleatória cuja distribuição de probabilidade é dada pelo modelo normal com média μ e variância σ^2 ($-\infty < \mu < +\infty$; $\sigma^2 > 0$). Sua função densidade é dada por

$$p(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-1}{2\sigma^2} (y - \mu)^2 \right\}, \quad -\infty < y < +\infty.$$

Considere (y_1, \dots, y_T) a realização de uma amostra aleatória da variável y . O logaritmo da função de verossimilhança para (μ, σ^2) é

$$\ell(\mu, \sigma^2) = -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^T (y_i - \mu)^2.$$

As estimativas de máxima verossimilhança dos parâmetros são dadas por

$$\hat{\mu} = \bar{y} \quad \text{e} \quad \hat{\sigma}^2 = \frac{1}{T} \sum_{i=1}^T (y_i - \hat{\mu})^2.$$

Supondo que temos interesse no parâmetro μ , o parâmetro de incômodo será σ^2 . Passaremos agora a encontrar a função de verossimilhança perfilada e perfilada modificada proposta por Barndorff-Nielsen (1983). Para μ fixado, temos que a estimativa de máxima verossimilhança para σ^2 é dada por

$$\hat{\sigma}_\mu^2 = \frac{1}{T} \sum_{i=1}^T (y_i - \mu)^2.$$

Substituindo σ^2 por $\hat{\sigma}_\mu^2$ na função de verossimilhança, temos que o logaritmo da função de verossimilhança perfilada (desprezando termos que não dependam dos parâmetros) é

$$\ell_p(\mu) = \ell(\mu, \hat{\sigma}_\mu^2) = -\frac{T}{2} \log \left[\sum_{i=1}^T (y_i - \mu)^2 \right].$$

A função escore perfilada é dada por

$$u_p(\mu) = T \frac{\sum_{i=1}^T (y_i - \mu)}{\sum_{i=1}^T (y_i - \mu)^2}.$$

Igualando-se a função escore a zero, temos que a estimativa de máxima verossimilhança perfilada para μ é a média amostral, como era de se esperar.

Tendo a expressão da função de verossimilhança perfilada, vamos obter os termos do ajuste proposto por Barndorff-Nielsen (1983). A informação observada relativa a σ^2 e avaliada em

$(\mu, \hat{\sigma}_\mu^2)$ é dada por

$$j_{\sigma^2\sigma^2}(\mu, \hat{\sigma}_\mu^2) = \frac{T^3}{2 \left[\sum_{i=1}^T (y_i - \mu)^2 \right]^2}.$$

O outro termo que compõe o ajuste não depende do parâmetro de interesse, então não aparecerá na função de verossimilhança ajustada. Com isso, a expressão da função de verossimilhança perfilada modificada proposta por Barndorff-Nielsen (1983) é

$$\ell_{BN}(\mu) = \left(1 - \frac{T}{2}\right) \log \left[\sum_{i=1}^T (y_i - \mu)^2 \right],$$

cuja função escore é

$$u_{BN}(\mu) = (T - 2) \frac{\sum_{i=1}^T (y_i - \mu)}{\sum_{i=1}^T (y_i - \mu)^2}.$$

Igualando-se a função escore a zero temos que a estimativa de máxima verossimilhança perfilada modificada também é a média amostral. Ou seja, neste exemplo,

$$\hat{\mu}_{BN} = \hat{\mu}_p = \bar{y},$$

o que significa que os estimadores pontuais para o parâmetro de interesse são idênticos.

Como podemos perceber,

$$\ell_{BN}(\mu) = \frac{2 - T}{2} \log \left[\sum_{i=1}^T (y_i - \mu)^2 \right] = \frac{T - 2}{T} \ell_p(\mu).$$

Portanto, neste caso, as estatísticas da razão de verossimilhanças baseadas em $\ell_p(\mu)$ (RV) e $\ell_{BN}(\mu)$ (RV_{BN}) tem a seguinte relação

$$RV_{BN} = \frac{T - 2}{T} RV.$$

Dessa forma, com probabilidade um, percebe-se que $RV_{BN} < RV$. Assim, utilizando a estatística RV_{BN} , pode-se observar menores taxas de rejeição da hipótese nula do que utilizando RV .

Para encontrarmos a aproximação baseada nas covariâncias de derivadas do logaritmo da função de verossimilhança precisamos calcular o termo $I_{\sigma^2}(\mu, \hat{\sigma}_\mu^2; \hat{\mu}, \hat{\sigma}^2)$. Assim, temos que

$$\ell_{ABN}(\mu) = -\frac{T}{2} \log \left[\sum_{i=1}^T (y_i - \mu)^2 \right] + \frac{1}{2} \log \left[\frac{1}{\left(\sum_{i=1}^T (y_i - \mu)^2 \right)^2} \right] - \log \left[\frac{1}{\left(\sum_{i=1}^T (y_i - \mu)^2 \right)^2} \right] = \ell_{BN}(\mu).$$

Ou seja, a aproximação é exata. Com isso, as inferências baseadas em $\ell_{ABN}(\mu)$ herdam as mesmas propriedades das baseadas em $\ell_{BN}(\mu)$.

Modelo de Regressão Normal Linear Simples

No exemplo anterior obtivemos a função de verossimilhança e suas modificações para o caso de observações independentes e identicamente distribuídas (i.i.d) de uma variável aleatória com distribuição normal. Agora especificamos uma estrutura de regressão para explicar o comportamento de uma variável aleatória y , quebrando assim a suposição de observações identicamente distribuídas. Veremos como obter a função de verossimilhança perfilada e perfilada modificada para este caso.

Seja y_i a i -ésima observação de uma amostra aleatória da variável y . Definimos que o i -ésimo valor observado é dado por

$$y_i = \alpha + \beta x_i + \epsilon_i,$$

onde α e β são coeficientes desconhecidos, x_i é a covariável medida na i -ésima observação e ϵ_i é um erro aleatório, para todo $i = 1, \dots, T$. Supomos ainda que ϵ_i e ϵ_k são independentes para todo $i \neq k$ e ϵ_i segue uma distribuição normal de média zero e variância σ^2 , para todo $i = 1, \dots, T$.

A função de densidade para y_i é dada por

$$p(y_i; \alpha, \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-1}{2\sigma^2} (y_i - \alpha - x_i\beta)^2 \right\}, -\infty < y_i < +\infty,$$

para todo $i = 1, \dots, T$.

O vetor de dimensão $T \times 1$, $\mathbf{Y} = (y_1, \dots, y_T)^\top$, representa uma amostra aleatória da variável y . Então temos que

$$\mathbf{Y} = \mathbf{1}\alpha + \mathbf{X}\beta + \boldsymbol{\epsilon},$$

onde $\mathbf{1}$ de tamanho $(T \times 1)$ possui o número um em todas as posições, X e $\boldsymbol{\epsilon}$, ambos de tamanho $(T \times 1)$, são obtidos empilhando-se x_i e ϵ_i , respectivamente. Ou, na forma matricial,

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \alpha + \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{pmatrix} \beta + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix}.$$

Assim \mathbf{Y} segue uma distribuição normal T -variada de média $\alpha + X\beta$ e variância $\mathbb{I}\sigma^2$, onde \mathbb{I} é a matriz identidade de tamanho $T \times T$. O logaritmo da função de verossimilhança pode ser escrito como

$$\ell(\alpha, \beta, \sigma^2) = -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta).$$

As estimativas de máxima verossimilhança para os parâmetros do modelo são

$$\begin{aligned}\hat{\alpha} &= (\mathbf{1}^\top \mathbf{1})^{-1} \mathbf{1}^\top (\mathbf{Y} - \mathbf{X}\hat{\beta}), \\ \hat{\beta} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\mathbf{Y} - \mathbf{1}\hat{\alpha}) \\ &\text{e} \\ \hat{\sigma}^2 &= \frac{1}{T} (\mathbf{Y} - \mathbf{1}\hat{\alpha} - \mathbf{X}\hat{\beta})^\top (\mathbf{Y} - \mathbf{1}\hat{\alpha} - \mathbf{X}\hat{\beta}).\end{aligned}$$

Vamos supor que nosso interesse seja produzir inferências para o vetor $\mu = (\alpha, \beta)$. Neste caso, o parâmetro de incômodo será $\nu = \sigma^2$. Fixando μ , a estimativa do parâmetro de incômodo ($\hat{\nu}_\mu = \hat{\sigma}_\mu^2$) é dada por

$$\hat{\sigma}_\mu^2 = \frac{1}{T} (\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta).$$

Substituindo ν por $\hat{\nu}_\mu$ na função de verossimilhança usual, encontramos o logaritmo da função de verossimilhança perfilada

$$\ell_p(\mu) = \ell_p(\alpha, \beta) = -\frac{T}{2} \log \left[(\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta) \right].$$

O vetor escore perfilado $u_p(\mu) = (u_p(\alpha), u_p(\beta))$ é dado por

$$\begin{aligned}u_p(\alpha) &= T \left[(\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta) \right]^{-1} \mathbf{1}^\top (\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta), \\ &\text{e} \\ u_p(\beta) &= T \left[(\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta) \right]^{-1} \mathbf{X}^\top (\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta).\end{aligned}$$

Igualando-se cada elemento do vetor escore a zero e resolvendo o sistema de equações, obtemos $\hat{\alpha}_p = \hat{\alpha}$ e $\hat{\beta}_p = \hat{\beta}$, como era de se esperar.

O primeiro termo do ajuste da verossimilhança perfilada proposta por Barndorff-Nielsen (1983) é dado por

$$j_{\sigma^2 \sigma^2}(\mu, \sigma^2) = \frac{1}{(\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta)}.$$

O outro termo que compõe o ajuste, neste caso, não dependerá dos parâmetros de interesse. Então, a expressão do logaritmo da função de verossimilhança perfilada modificada é dada por

$$\ell_{BN}(\alpha, \beta) = \left(1 - \frac{T}{2}\right) \log \left[(\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta) \right]$$

e a respectiva função escore $u_{BN}(\alpha, \beta) = (u_{BN}(\alpha), u_{BN}(\beta))$, por

$$u_{BN}(\alpha) = 2 \left(1 - \frac{T}{2}\right) \left[(\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta) \right]^{-1} \mathbf{1}^\top (\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta)$$

e

$$u_{BN}(\beta) = 2 \left(1 - \frac{T}{2}\right) \left[(\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta) \right]^{-1} \mathbf{X}^\top (\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta).$$

Igualando a função escore a zero e resolvendo o sistema, obtemos $\hat{\alpha}_{BN} = \hat{\alpha}_p = \hat{\alpha}$ e $\hat{\beta}_{BN} = \hat{\beta}_p = \hat{\beta}$.

As mesmas relações entre as estatísticas da razão de verossimilhanças encontradas no exemplo anterior valem para este exemplo. Assim se desejamos testar $H_0 : (\alpha, \beta) = (0, 0)$ contra $H_1 : (\alpha, \beta) \neq (0, 0)$ utilizando a estatística RV_{BN} , esperamos observar menores taxas de rejeição da hipótese nula do que utilizando RV . Aqui também $\ell_{ABN}(\alpha, \beta) = \ell_{BN}(\alpha, \beta)$.

Se estivermos interessados apenas em produzir inferências para o parâmetro β , o vetor de parâmetros de incômodo será $\nu = (\alpha, \sigma^2)$.

Para facilitar os cálculos, podemos reescrever o logaritmo da função de verossimilhança da seguinte forma

$$\begin{aligned} \ell(\alpha, \beta, \sigma^2) &= -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{1}\alpha - \mathbf{X}\beta) \\ &= -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left\{ (\mathbf{Y} - \mathbf{1}\hat{\alpha} - \mathbf{X}\hat{\beta})^\top (\mathbf{Y} - \mathbf{1}\hat{\alpha} - \mathbf{X}\hat{\beta}) + T(\hat{\alpha} - \alpha)^2 + (\hat{\beta} - \beta)^\top X^\top X (\hat{\beta} - \beta) \right\} \\ &= -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left[T\hat{\sigma}^2 + T(\hat{\alpha} - \alpha)^2 + (\hat{\beta} - \beta)^\top X^\top X (\hat{\beta} - \beta) \right]. \end{aligned}$$

Fixando β , a estimativa de máxima verossimilhança do vetor de parâmetros de incômodo $\hat{\nu}_\beta = (\hat{\alpha}_\beta, \hat{\sigma}_\beta^2)$ é dada por

$$\hat{\sigma}_\beta^2 = \frac{T\hat{\sigma}^2 + (\hat{\beta} - \beta)^\top X^\top X (\hat{\beta} - \beta)}{T} \quad \text{e} \quad \hat{\alpha}_\beta = \hat{\alpha}.$$

Substituindo ν por $\hat{\nu}_\beta$ na função de verossimilhança usual, encontramos o logaritmo da função de verossimilhança perfilada

$$\ell_p(\beta) = -\frac{T}{2} \log \left[T\hat{\sigma}^2 + (\hat{\beta} - \beta)^\top X^\top X (\hat{\beta} - \beta) \right].$$

A função escore perfilada é dada por

$$u_p(\beta) = \frac{T \left[\mathbf{1}^\top X^\top X (\hat{\beta} - \beta) \right]}{\log \left[T\hat{\sigma}^2 + (\hat{\beta} - \beta)^\top X^\top X (\hat{\beta} - \beta) \right]}.$$

Igualando a função escore perfilada a zero, obtemos $\hat{\beta}_p = \hat{\beta}$.

O primeiro termo do ajuste da verossimilhança perfilada proposto por Barndorff-Nielsen (1983)

é dado por

$$j_{\nu\nu}(\beta, \hat{\nu}_\beta) = \begin{pmatrix} T/(2\hat{\sigma}_\beta^2) & 0 \\ 0 & T/\hat{\sigma}_\beta^2 \end{pmatrix}.$$

O segundo termo que compõe o ajuste é dado por

$$\ell_{\nu;\hat{\nu}}(\beta, \hat{\nu}_\beta; \hat{\beta}, \hat{\nu}) = \begin{pmatrix} T/[2(\hat{\sigma}_\beta^2)^2] & 0 \\ 0 & T/\hat{\sigma}_\beta^2 \end{pmatrix}.$$

O logaritmo de função de verossimilhança perfilada modificada é

$$\ell_{BN}(\beta) = \left(2 - \frac{T}{2}\right) \log \left[T\hat{\sigma}^2 + (\hat{\beta} - \beta)^\top X^\top X (\hat{\beta} - \beta) \right]$$

e a função escore, por

$$u_{BN}(\beta) = -2 \left(2 - \frac{T}{2}\right) \frac{[\mathbf{1}^\top X^\top X (\hat{\beta} - \beta)]}{\log \left[T\hat{\sigma}^2 + (\hat{\beta} - \beta)^\top X^\top X (\hat{\beta} - \beta) \right]}.$$

Igualando a função escore a zero, obtemos $\hat{\beta}_{BN} = \hat{\beta}_p = \hat{\beta}$.

Se desejamos testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$ utilizando a estatística RV_{BN} , também esperamos observar menores taxas de rejeição da hipótese nula do que utilizando RV .

A aproximação para Barndorff-Nielsen (1983), neste caso, também é exata. Ou seja, $\ell_{ABN}(\beta) = \ell_{BN}(\beta)$.

3.6 Teste da Razão de Verossimilhanças Baseado no Método *Bootstrap*

Nas situações práticas, em geral, temos acesso a apenas uma amostra $\mathbf{x} = (x_1, \dots, x_T)$. Efron (1979) desenvolveu uma maneira de se obter K coleções $\mathbf{x}^\dagger = (x_1^\dagger, \dots, x_T^\dagger)$, que representam valores numéricos assumidos pelo vetor aleatório $\mathbf{X} = (X_1, \dots, X_T)$. Para isto, ele explorou a capacidade dos computadores de realizar tarefas repetitivas e propôs que gerássemos as K coleções, basicamente, de duas formas: (i) a partir de uma estimativa paramétrica da função de distribuição associada ao modelo considerado (*bootstrap* paramétrico), (ii) a partir da função de distribuição empírica construída da amostra \mathbf{x} (*bootstrap* não-paramétrico).

Nesta dissertação utilizaremos apenas o chamado *bootstrap* paramétrico, pois no nosso caso assumimos um modelo paramétrico conhecido para os dados observados. O *bootstrap* paramétrico é caracterizado pela suposição de uma função de distribuição F pertencente a uma família paramétrica F_τ e propõe a geração de vários conjuntos de medidas \mathbf{x}^\dagger a partir da estimativa

paramétrica $F_{\hat{\tau}}$. A estimativa $\hat{\tau}$ é obtida da amostra observada \mathbf{x} .

Em inferência estatística, diversas vezes nos deparamos com o problema de estimação de alguma grandeza θ da distribuição F . Para isto, usamos um estimador $\hat{\theta}$. As K amostras *bootstrap* proporcionam K réplicas de $\hat{\theta}$ utilizadas para a construção de uma função de distribuição empírica, e utilizamos esta distribuição como uma aproximação para a distribuição de $\hat{\theta}$. Esta é a idéia do método *bootstrap*.

No nosso caso, queremos encontrar uma boa aproximação para a distribuição de RV sob H_0 . Baseado na função de distribuição empírica de RV obtida a partir das K amostras *bootstrap* geradas sob H_0 , podemos utilizar um quantil para definir a região de rejeição do teste da razão de verossimilhanças ou calcularmos um p-valor como $p_{boot} = \#\{RV_t^\dagger \geq RV\}/K$, onde RV_t^\dagger denota a estatística da razão de verossimilhanças obtida da t -ésima amostra *bootstrap* e RV , obtida da amostra inicial (\mathbf{x}). $\#$ denota a cardinalidade de um conjunto. Assim p_{boot} é a proporção de valores de RV obtidos das amostras *bootstrap* maiores ou iguais ao valor de RV obtido da amostra inicial.

3.7 Correção de Bartlett para a Estatística da Razão de Verossimilhanças

Utilizando a notação introduzida na Seção 3.1, nosso interesse consiste em testar $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$. A estatística da razão de verossimilhanças (RV) sob a hipótese nula e em problemas regulares tem, assintoticamente, uma distribuição qui-quadrado com q graus de liberdade (χ_q^2), onde q é o número de restrições impostas por H_0 .

Para amostras de tamanhos pequenos, a aproximação da distribuição de RV pela χ_q^2 pode ser insatisfatória, produzindo um teste com taxas de rejeição bastante distorcidas. Uma solução para melhorar esta aproximação foi proposta por Bartlett (1937), onde o autor sugere a obtenção de um fator de correção c tal que $RV^* = RV/c$ é a estatística da razão de verossimilhanças modificada via Bartlett.

Mais precisamente, Bartlett (1937) propôs a seguinte estatística modificada:

$$\widetilde{RV} = q \frac{RV}{E(RV)},$$

onde $E(RV)$ é calculada sob $(\mu_0, \hat{\nu}_{\mu_0})$. Entretanto, em geral, não é possível determinar este valor esperado explicitamente e, assim, baseado na expansão

$$E(RV) = q [1 + B(\mu, \nu) + O(n^{-2})],$$

onde $B(\mu, \nu) = O(n^{-1})$, costuma-se usar a estatística ajustada definida por

$$RV^* = \frac{RV}{1 + B(\mu_0, \hat{\nu}_{\mu_0})}. \quad (3.9)$$

Lawley (1956) apresenta uma fórmula para $B(\mu, \nu)$ em termos dos valores esperados das quatro primeiras derivadas de $\ell(\mu, \nu)$ e de derivadas desses valores esperados. Considerando algumas suposições sobre a ordem estocástica de derivadas de $\ell(\mu, \nu)$, ele mostra que os cumulantes de RV^* são iguais aos cumulantes de uma variável qui-quadrado com q graus de liberdade até termos de ordem $O(n^{-2})$.

Aqui o fator de correção $B(\mu_0, \hat{\nu}_{\mu_0})$ será calculado por simulação, usando os valores da estatística de teste baseados nas amostras bootstrap. Note que podemos fazer

$$\widehat{E}(RV) = \frac{1}{K} \sum_{t=1}^K RV_t^\dagger = q[1 + \hat{B}(\mu_0, \hat{\nu}_{\mu_0})].$$

Então,

$$\hat{B}(\mu_0, \hat{\nu}_{\mu_0}) = \frac{1}{q} \left(\frac{1}{K} \sum_{t=1}^K RV_t^\dagger \right) - 1$$

e, por conseguinte, a estatística da razão de verossimilhanças ajustada via Bartlett será aproximada por

$$RV^* \approx \frac{RV}{1 + \hat{B}(\mu_0, \hat{\nu}_{\mu_0})}. \quad (3.10)$$

Observe que o ajuste sobre a estatística da razão de verossimilhanças proporcionado pelo uso de uma verossimilhança perfilada modificada é preventivo (ajustamos a função de verossimilhança antes de construir a estatística), enquanto o enfoque de Bartlett (1937) é corretivo (construímos a estatística e depois a corrigimos).

4 *Ajuste de Barndorff-Nielsen no Modelo Linear de Efeitos Mistos Geral*

4.1 Introdução

Conforme já argumentamos, dependendo do tamanho amostral, a maximização da função de verossimilhança perfilada pode produzir estimativas bastante viesadas e o teste da razão de verossimilhanças pode apresentar taxas de rejeição sob a hipótese nula bastante distorcidas. Neste capítulo derivamos a aproximação, baseada em Severini (1998), para o ajuste de Barndorff-Nielsen sobre a verossimilhança perfilada (Seção 3.4) no contexto do modelo linear misto geral. Na Seção 4.2 apresentamos a modificação da verossimilhança perfilada no modelo linear misto geral. Consideramos como vetor de parâmetros de interesse uma parte dos parâmetros dos efeitos fixos. E na Seção 4.3, consideramos dois particulares modelos lineares mistos que serão utilizados para avaliar o desempenho dos métodos de aperfeiçoamento do teste da razão de verossimilhanças discutidos nesta dissertação através de simulações de Monte Carlo.

4.2 Modelo Misto Linear Geral

O modelo misto linear geral é dado por (ver Seção 2.2)

$$\mathbf{Y} = X\beta + Z\mathbf{b} + \epsilon, \quad (4.1)$$

onde \mathbf{Y} é um vetor aleatório de tamanho $T \times 1$, $T = \sum_{i=1}^N n_i$, com distribuição $Normal_T(X\beta, \Sigma)$. $\Sigma = \Sigma(\alpha)$.

Suponha que nosso interesse seja inferir sobre os q ($q \leq p$) últimos parâmetros do vetor β , situação esta que é muito comum na prática. Neste caso o vetor de interesse é $\mu = (\beta_{p-q+1}, \beta_{p-q+2}, \dots, \beta_p)^\top$ e o vetor de incômodo é dado por $\nu = (\beta_1, \dots, \beta_{p-q}, \alpha)^\top = (\tilde{\beta}, \alpha)^\top$. Definimos também \tilde{X}_q como a matriz contendo as q últimas colunas de X e \tilde{X} a matriz com as demais $(p-q)$ colunas. Temos então que $X\beta = \tilde{X}_q\mu + \tilde{X}\tilde{\beta}$. Podemos reescrever o logaritmo da função de verossimilhança (2.3) em termos de $\theta = (\mu, \tilde{\beta}, \alpha)$ como

$$\ell(\theta) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma(\alpha)| - \frac{1}{2} \left(\mathbf{y} - \tilde{X}_q\mu - \tilde{X}\tilde{\beta} \right)^\top \Sigma^{-1}(\alpha) \left(\mathbf{y} - \tilde{X}_q\mu - \tilde{X}\tilde{\beta} \right). \quad (4.2)$$

O estimador $\hat{\nu}_\mu = (\hat{\beta}_\mu, \hat{\alpha}_\mu)$ não tem forma fechada. Conseqüentemente, a função de verossimilhança perfilada não possui forma tratável analiticamente. Assim obtemos valores de $\ell_p(\mu)$ maximizando numericamente $\ell(\mu, \tilde{\beta}, \alpha)$ sujeita às restrições $\partial\ell/\partial\tilde{\beta} = 0$ e $\partial\ell/\partial\alpha = 0$.

O primeiro termo do ajuste de Barndorff-Nielsen (1983) (ver 3.5) é a informação observada relativa a ν avaliada em $(\mu, \hat{\nu}_\mu)$. Aqui essa informação observada é dada pela matriz

$$j_{\nu\nu}(\mu, \nu) = - \begin{pmatrix} \frac{\partial^2 \ell(\mu, \nu)}{\partial \alpha_1^2} & \frac{\partial^2 \ell(\mu, \nu)}{\partial \alpha_1 \partial \alpha_2} & \dots & \frac{\partial^2 \ell(\mu, \nu)}{\partial \alpha_1 \partial \alpha_m} & \frac{\partial^2 \ell(\mu, \nu)}{\partial \alpha_1 \partial \tilde{\beta}^\top} \\ \frac{\partial^2 \ell(\mu, \nu)}{\partial \alpha_2 \partial \alpha_1} & \frac{\partial^2 \ell(\mu, \nu)}{\partial \alpha_2^2} & \dots & \frac{\partial^2 \ell(\mu, \nu)}{\partial \alpha_2 \partial \alpha_m} & \frac{\partial^2 \ell(\mu, \nu)}{\partial \alpha_2 \partial \tilde{\beta}^\top} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial^2 \ell(\mu, \nu)}{\partial \alpha_m \partial \alpha_1} & \frac{\partial^2 \ell(\mu, \nu)}{\partial \alpha_m \partial \alpha_2} & \dots & \frac{\partial^2 \ell(\mu, \nu)}{\partial \alpha_m^2} & \frac{\partial^2 \ell(\mu, \nu)}{\partial \alpha_m \partial \tilde{\beta}^\top} \\ \hline \frac{\partial^2 \ell(\mu, \nu)}{\partial \tilde{\beta} \partial \alpha_1} & \frac{\partial^2 \ell(\mu, \nu)}{\partial \tilde{\beta} \partial \alpha_2} & \dots & \frac{\partial^2 \ell(\mu, \nu)}{\partial \tilde{\beta} \partial \alpha_m} & \frac{\partial^2 \ell(\mu, \nu)}{\partial \tilde{\beta} \partial \tilde{\beta}^\top} \end{pmatrix},$$

que, calculada e avaliada em $(\mu, \hat{\nu}_\mu)$, tem elementos $(i, j, k = 1, \dots, m)$

$$\begin{aligned} \frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial \alpha_i \partial \alpha_j} &= \frac{1}{2} \text{tr} \left(\Sigma^{-1}(\hat{\alpha}_\mu) \dot{\Sigma}_i(\hat{\alpha}_\mu) \Sigma^{-1}(\hat{\alpha}_\mu) \dot{\Sigma}_j(\hat{\alpha}_\mu) \right) \\ &- \frac{1}{2} \left(\mathbf{Y} - \tilde{\mathbf{X}}_q \mu - \tilde{\mathbf{X}} \hat{\beta}_\mu \right)^\top \Sigma^{-1}(\hat{\alpha}_\mu) \dot{\Sigma}_i(\hat{\alpha}_\mu) \Sigma^{-1}(\hat{\alpha}_\mu) \dot{\Sigma}_j(\hat{\alpha}_\mu) \Sigma^{-1}(\hat{\alpha}_\mu) \left(\mathbf{Y} - \tilde{\mathbf{X}}_q \mu - \tilde{\mathbf{X}} \hat{\beta}_\mu \right), \end{aligned}$$

$$\frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial \tilde{\beta} \partial \alpha_i} = \tilde{\mathbf{X}}^\top \Sigma^{-1}(\hat{\alpha}_\mu) \dot{\Sigma}_i(\hat{\alpha}_\mu) \Sigma^{-1}(\hat{\alpha}_\mu) \left(\mathbf{Y} - \tilde{\mathbf{X}}_q \mu - \tilde{\mathbf{X}} \hat{\beta}_\mu \right) = \left(\frac{\partial^2 \ell(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu})}{\partial \alpha_i \partial \tilde{\beta}^\top} \right)^\top,$$

$$\frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial \tilde{\beta} \partial \tilde{\beta}^\top} = \tilde{\mathbf{X}}^\top \Sigma^{-1}(\hat{\alpha}_\mu) \tilde{\mathbf{X}},$$

onde $\dot{\Sigma}_k(\alpha) = \partial \Sigma(\alpha) / \partial \alpha_k$ (ver Apêndice A).

Como não é possível derivar o segundo termo do ajuste de Barndorff-Nielsen (1983), usamos uma aproximação proposta por Severini (1998) e baseada na matriz

$$I_\nu(\mu, \nu; \mu_0, \nu_0) = E_{(\mu_0, \nu_0)} \begin{pmatrix} \frac{\partial \ell(\mu, \nu)}{\partial \alpha_1} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \alpha_1} & \frac{\partial \ell(\mu, \nu)}{\partial \alpha_1} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \alpha_2} & \dots & \frac{\partial \ell(\mu, \nu)}{\partial \alpha_1} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \alpha_m} & \frac{\partial \ell(\mu, \nu)}{\partial \alpha_1} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \tilde{\beta}^\top} \\ \frac{\partial \ell(\mu, \nu)}{\partial \alpha_2} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \alpha_1} & \frac{\partial \ell(\mu, \nu)}{\partial \alpha_2} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \alpha_2} & \dots & \frac{\partial \ell(\mu, \nu)}{\partial \alpha_2} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \alpha_m} & \frac{\partial \ell(\mu, \nu)}{\partial \alpha_2} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \tilde{\beta}^\top} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial \ell(\mu, \nu)}{\partial \alpha_m} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \alpha_1} & \frac{\partial \ell(\mu, \nu)}{\partial \alpha_m} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \alpha_2} & \dots & \frac{\partial \ell(\mu, \nu)}{\partial \alpha_m} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \alpha_m} & \frac{\partial \ell(\mu, \nu)}{\partial \alpha_m} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \tilde{\beta}^\top} \\ \hline \frac{\partial \ell(\mu, \nu)}{\partial \tilde{\beta}} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \alpha_1} & \frac{\partial \ell(\mu, \nu)}{\partial \tilde{\beta}} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \alpha_2} & \dots & \frac{\partial \ell(\mu, \nu)}{\partial \tilde{\beta}} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \alpha_m} & \frac{\partial \ell(\mu, \nu)}{\partial \tilde{\beta}} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \tilde{\beta}^\top} \end{pmatrix}.$$

Calculando o valor esperado dessa matriz e, posteriormente, avaliando (μ, ν) em $(\mu, \hat{\nu}_\mu)$ e (μ_0, ν_0)

em $(\hat{\mu}, \hat{\nu})$ temos os elementos $(i, j = 1, \dots, m)$

$$\begin{aligned} \left\{ E_{(\mu_0, \nu_0)} \left[\frac{\partial \ell(\mu, \nu)}{\partial \alpha_i} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \alpha_j} \right] \right\} \Bigg|_{(\mu, \nu) = (\mu, \hat{\nu}_\mu), (\mu_0, \nu_0) = (\hat{\mu}, \hat{\nu})} &= \frac{1}{2} \text{tr} \left[\Sigma^{-1}(\hat{\alpha}_\mu) \dot{\Sigma}_i(\hat{\alpha}_\mu) \Sigma^{-1}(\hat{\alpha}_\mu) \dot{\Sigma}_j(\hat{\alpha}) \right], \\ \left\{ E_{(\mu_0, \nu_0)} \left[\frac{\partial \ell(\mu, \nu)}{\partial \tilde{\beta}} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \alpha_i} \right] \right\} \Bigg|_{(\mu, \nu) = (\mu, \hat{\nu}_\mu), (\mu_0, \nu_0) = (\hat{\mu}, \hat{\nu})} &= \mathbf{0}_{(p-q)} \quad (\text{vetor nulo } (p-q) \times 1), \\ \left\{ E_{(\mu_0, \nu_0)} \left[\frac{\partial \ell(\mu, \nu)}{\partial \tilde{\beta}} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \tilde{\beta}^\top} \right] \right\} \Bigg|_{(\mu, \nu) = (\mu, \hat{\nu}_\mu), (\mu_0, \nu_0) = (\hat{\mu}, \hat{\nu})} &= \tilde{\mathbf{X}}^\top \Sigma^{-1}(\hat{\alpha}_\mu) \tilde{\mathbf{X}}, \\ \left\{ E_{(\mu_0, \nu_0)} \left[\frac{\partial \ell(\mu, \nu)}{\partial \alpha_i} \frac{\partial \ell(\mu_0, \nu_0)}{\partial \tilde{\beta}^\top} \right] \right\} \Bigg|_{(\mu, \nu) = (\mu, \hat{\nu}_\mu), (\mu_0, \nu_0) = (\hat{\mu}, \hat{\nu})} &= \left[\tilde{\mathbf{X}}^\top \Sigma^{-1}(\hat{\alpha}_\mu) \dot{\Sigma}_i(\hat{\alpha}_\mu) \Sigma^{-1}(\hat{\alpha}_\mu) \right. \\ &\quad \left. \times \left(\tilde{\mathbf{X}}_q \hat{\mu} + \tilde{\mathbf{X}} \hat{\beta} - \tilde{\mathbf{X}}_q \mu - \tilde{\mathbf{X}} \hat{\beta}_\mu \right) \right]^\top. \end{aligned}$$

Portanto, a aproximação para o ajuste de Barndorff-Nielsen $\ell_{ABN}(\mu)$ é igual a

$$\ell_{ABN}(\mu) = \ell_{ABN}(\mu, \tilde{\beta}, \alpha) = \ell(\mu, \tilde{\beta}, \alpha) + \frac{1}{2} \log \left| j_{\nu\nu}(\mu, \tilde{\beta}, \alpha) \right| - \log \left| I_\nu(\mu, \tilde{\beta}, \alpha; \hat{\mu}, \hat{\beta}, \hat{\alpha}) \right|$$

sujeita às restrições $\partial \ell(\mu, \tilde{\beta}, \alpha) / \partial \tilde{\beta} = 0$ e $\partial \ell(\mu, \tilde{\beta}, \alpha) / \partial \alpha = 0$. Maximizando $\ell_{ABN}(\mu)$ sob essas restrições, obtemos a estimativa de máxima verossimilhança perfilada ajustada $\hat{\mu}_{ABN}$. Se desejarmos encontrar o valor da estatística da razão de verossimilhanças perfiladas modificadas

$$RV_{ABN} = 2 \{ \ell_{ABN}(\hat{\mu}_{ABN}) - \ell_{ABN}(\mu_0) \}$$

para testar $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$, precisamos maximizar $\ell_{ABN}(\mu_0)$ sujeita às restrições $\partial \ell(\mu_0, \tilde{\beta}, \alpha) / \partial \tilde{\beta} = 0$ e $\partial \ell(\mu_0, \tilde{\beta}, \alpha) / \partial \alpha = 0$. Sob a hipótese nula, assintoticamente, esta estatística tem distribuição qui-quadrado com q graus de liberdade.

4.3 Exemplos

Nesta seção desenvolvemos os ajustes para dois particulares modelos lineares mistos que serão utilizados nas avaliações numéricas do próximo capítulo. O primeiro caso denominamos de modelo com intercepto aleatório e o segundo, além do intercepto, possui a inclinação da reta através do tempo também aleatória. Esses modelos foram utilizados por Laird e Ware (1982) em aplicações com dados reais.

Modelo com Intercepto Aleatório

Definimos o modelo com intercepto aleatório como

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 x_{2i} + \beta_3 x_{3i} + b_{0i} + \epsilon_{ij}, \quad (4.3)$$

onde t_{ij} é o tempo da observação j do indivíduo i , $b_{0i} \sim Normal(0, \sigma_b^2)$ e $\epsilon_{ij} \sim Normal(\mathbf{0}, \sigma_\epsilon^2)$. Em notação matricial, as observações para o indivíduo i são dadas por

$$\mathbf{Y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{pmatrix} = \begin{pmatrix} 1 & t_{i1} & x_{2i} & x_{3i} \\ 1 & t_{i2} & x_{2i} & x_{3i} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_{in_i} & x_{2i} & x_{3i} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} b_{0i} + \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{in_i} \end{pmatrix}.$$

Assim, $\mathbf{Y}_i \sim Normal_{n_i}(\mathbf{X}_i\beta, V_{0i})$, onde a matriz V_{0i} pode ser dada por

$$V_{0i} = \tau^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}_{n_i \times n_i},$$

com $\tau^2 = \sigma_b^2 + \sigma_\epsilon^2$ e $\rho = \sigma_b^2/\tau^2$.

Dessa forma, assumimos que as observações para o mesmo indivíduo são igualmente correlacionadas, conhecido como modelo de correlação uniforme (Diggle et al., 1994). Logo $\mathbf{Y} \sim Normal_T(\mathbf{X}\beta, \Sigma)$, onde $T = \sum_{i=1}^N n_i$, Σ é uma matriz bloco-diagonal com blocos V_{0i} ($i = 1, \dots, N$) e $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^\top$; ver Seção 2.2. O logaritmo da função de verosimilhança associado a este modelo é dado por

$$\ell(\beta, \tau^2, \rho; \mathbf{Y}) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\tau^2) - \frac{1}{2} \log |P| - \frac{1}{2\tau^2} (\mathbf{Y} - \mathbf{X}\beta)^\top P^{-1} (\mathbf{Y} - \mathbf{X}\beta),$$

onde P é matriz bloco-diagonal com blocos $[(1 - \rho)I_{n_i} + \rho J_{n_i}]$ e J_q é uma matriz quadrada $q \times q$ com todos os elementos iguais a um. Para este exemplo, temos que $\theta = (\beta, \tau^2, \rho)$. A função escore $u(\theta) = \partial\ell(\theta)/\partial\theta$ tem componentes:

$$\frac{\partial\ell}{\partial\beta} = \mathbf{X}^\top \Sigma^{-1} (\mathbf{Y} - \mathbf{X}\beta),$$

$$\frac{\partial\ell}{\partial\tau^2} = -\frac{T}{2\tau^2} + \frac{1}{2(\tau^2)^2} (\mathbf{Y} - \mathbf{X}\beta)^\top P^{-1} (\mathbf{Y} - \mathbf{X}\beta)$$

e

$$\frac{\partial\ell}{\partial\rho} = -\frac{1}{2} \text{tr}(P^{-1}\mathbb{P}) + \frac{1}{2\tau^2} (\mathbf{Y} - \mathbf{X}\beta)^\top P^{-1} \mathbb{P} P^{-1} (\mathbf{Y} - \mathbf{X}\beta),$$

em que \mathbb{P} representa $\partial P/\partial\rho$. Igualando essas três derivadas a zero e resolvendo o sistema de

equações, obtemos os estimadores de máxima verossimilhança (EMV) dos componentes de θ . Os estimadores dos parâmetros β e τ^2 têm formas fechadas, já o estimador de ρ não possui:

$$\hat{\beta} = \left(\mathbf{X}^\top \hat{\Sigma}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^\top \hat{\Sigma}^{-1} \mathbf{Y},$$

$$\hat{\tau}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\beta})^\top \hat{P}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\beta})}{T}.$$

Como esses estimadores dependem do estimador de ρ , utilizamos métodos de maximização numérica para obter as estimativas de máxima verossimilhança dos componentes de θ .

No próximo capítulo, as comparações de desempenho dos testes da razão de verossimilhanças são feitas considerando como interesse tanto o parâmetro $\mu = \beta_3$ quanto o vetor $\mu = (\beta_3, \beta_2)$ e, para ambos os casos, $\nu = (\tilde{\beta}, \tau^2, \rho)$ denota o vetor de parâmetros de incômodo. Devido ao EMV de ρ não possuir forma fechada, a função de verossimilhança perfilada não será tratável analiticamente. Assim, para obtermos valores de $\ell_p(\mu)$, maximizamos numericamente $\ell(\mu, \nu; \mathbf{Y})$ restrita a $\partial\ell(\theta)/\partial\nu = 0$.

A matriz de informação observada relativa ao vetor de parâmetros de incômodo ν avaliada em $(\mu, \hat{\nu}_\mu)$ é composta pelos seguintes elementos:

$$\begin{aligned} \frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial(\tau^2)^2} &= \frac{T}{2(\hat{\tau}_\mu^2)^2} - \frac{1}{(\hat{\tau}_\mu^2)^3} \left(\mathbf{Y} - \tilde{\mathbf{X}}_q \mu - \tilde{\mathbf{X}} \hat{\beta}_\mu \right)^\top \hat{P}_\mu^{-1} \left(\mathbf{Y} - \tilde{\mathbf{X}}_q \mu - \tilde{\mathbf{X}} \hat{\beta}_\mu \right), \\ \frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial\tau^2 \partial\rho} &= -\frac{1}{2(\hat{\tau}_\mu^2)^2} \left(\mathbf{Y} - \tilde{\mathbf{X}}_q \mu - \tilde{\mathbf{X}} \hat{\beta}_\mu \right)^\top \hat{P}_\mu^{-1} \hat{\mathbb{P}}_\mu \hat{P}_\mu^{-1} \left(\mathbf{Y} - \tilde{\mathbf{X}}_q \mu - \tilde{\mathbf{X}} \hat{\beta}_\mu \right) \\ &= \frac{\partial \ell(\mu, \hat{\nu}_\mu)}{\partial\rho \partial\tau^2}, \\ \frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial\tilde{\beta} \partial\tau^2} &= -\frac{1}{(\hat{\tau}_\mu^2)^2} \tilde{\mathbf{X}}^\top \hat{P}_\mu^{-1} \left(\mathbf{Y} - \tilde{\mathbf{X}}_q \mu - \tilde{\mathbf{X}} \hat{\beta}_\mu \right) \\ &= \left(\frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial\tau^2 \partial\tilde{\beta}^\top} \right)^\top, \\ \frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial\rho^2} &= \frac{1}{2} \text{tr} \left(\hat{P}_\mu^{-1} \hat{\mathbb{P}}_\mu \hat{P}_\mu^{-1} \hat{\mathbb{P}}_\mu \right) \\ &\quad - \frac{1}{2\hat{\tau}_\mu^2} \left(\mathbf{Y} - \tilde{\mathbf{X}}_q \mu - \tilde{\mathbf{X}} \hat{\beta}_\mu \right)^\top \hat{P}_\mu^{-1} \hat{\mathbb{P}}_\mu \hat{P}_\mu^{-1} \hat{\mathbb{P}}_\mu \hat{P}_\mu^{-1} \left(\mathbf{Y} - \tilde{\mathbf{X}}_q \mu - \tilde{\mathbf{X}} \hat{\beta}_\mu \right), \\ \frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial\tilde{\beta} \partial\rho} &= -\frac{1}{2\hat{\tau}_\mu^2} \tilde{\mathbf{X}}^\top \hat{P}_\mu^{-1} \hat{\mathbb{P}}_\mu \hat{P}_\mu^{-1} \hat{\mathbb{P}}_\mu \hat{P}_\mu^{-1} \left(\mathbf{Y} - \tilde{\mathbf{X}}_q \mu - \tilde{\mathbf{X}} \hat{\beta}_\mu \right) \\ &= \left(\frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial\rho \partial\tilde{\beta}^\top} \right)^\top, \\ \frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial\tilde{\beta} \partial\tilde{\beta}^\top} &= -\frac{1}{2\hat{\tau}_\mu^2} \tilde{\mathbf{X}}^\top \hat{P}_\mu^{-1} \tilde{\mathbf{X}}, \end{aligned}$$

que são dispostos na forma

$$j_{\nu\nu}(\mu, \hat{\nu}_\mu) = - \left(\begin{array}{cc|c} \frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial(\tau^2)^2} & \frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial\tau^2 \partial\rho} & \frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial\tau^2 \partial\tilde{\beta}^\top} \\ \frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial\rho \partial\tau^2} & \frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial\rho^2} & \frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial\rho \partial\tilde{\beta}^\top} \\ \hline \frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial\tilde{\beta} \partial\tau^2} & \frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial\tilde{\beta} \partial\rho} & \frac{\partial^2 \ell(\mu, \hat{\nu}_\mu)}{\partial\tilde{\beta} \partial\tilde{\beta}^\top} \end{array} \right).$$

O subscripto μ nas matrizes $\hat{\mathbb{P}}_\mu$, \hat{P}_μ e $\hat{\Sigma}_\mu$ indica que as matrizes \mathbb{P} , P e Σ estão sendo avaliadas em $(\mu, \hat{\nu}_\mu)$, respectivamente. Quando avaliadas em $(\hat{\mu}, \hat{\nu})$, as matrizes \mathbb{P} , P e Σ são denotadas por $\hat{\mathbb{P}}$, \hat{P} e $\hat{\Sigma}$.

Como mencionamos anteriormente, tamanhos amostrais demasiadamente pequenos podem induzir a testes de hipóteses baseados na razão de verossimilhanças usual (ou verossimilhanças perfiladas) insatisfatórios. Como alternativa utilizaremos a aproximação para a modificação proposta por Barndorff-Nielsen (1983), que requer o cálculo da matriz $I_\nu(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu})$, constituída dos seguintes termos:

$$\begin{aligned} \left\{ E_{(\mu_0, \nu_0)} \left[\frac{\partial \ell(\mu, \nu)}{\partial\tau^2} \frac{\partial \ell(\mu_0, \nu_0)}{\partial\tau^2} \right] \right\} \Big|_{(\mu, \nu) = (\mu, \hat{\nu}_\mu), (\mu_0, \nu_0) = (\hat{\mu}, \hat{\nu})} &= \frac{1}{2\hat{\tau}_\mu^2} \text{tr} \left(\hat{\Sigma}_\mu^{-1} \hat{P} \right), \\ \left\{ E_{(\mu_0, \nu_0)} \left[\frac{\partial \ell(\mu, \nu)}{\partial\tau^2} \frac{\partial \ell(\mu_0, \nu_0)}{\partial\rho} \right] \right\} \Big|_{(\mu, \nu) = (\mu, \hat{\nu}_\mu), (\mu_0, \nu_0) = (\hat{\mu}, \hat{\nu})} &= \frac{\hat{\tau}^2}{2\hat{\tau}_\mu^2} \text{tr} \left(\hat{\Sigma}_\mu^{-1} \hat{\mathbb{P}} \right), \\ \left\{ E_{(\mu_0, \nu_0)} \left[\frac{\partial \ell(\mu, \nu)}{\partial\tau^2} \frac{\partial \ell(\mu_0, \nu_0)}{\partial\tilde{\beta}^\top} \right] \right\} \Big|_{(\mu, \nu) = (\mu, \hat{\nu}_\mu), (\mu_0, \nu_0) = (\hat{\mu}, \hat{\nu})} &= \frac{1}{\hat{\tau}_\mu^2} \tilde{\mathbf{X}}^\top \hat{P}_\mu^{-1} \left(\tilde{\mathbf{X}}_q \hat{\mu} + \tilde{\mathbf{X}} \hat{\beta} - \tilde{\mathbf{X}}_q \mu - \tilde{\mathbf{X}} \hat{\beta}_\mu \right), \\ \left\{ E_{(\mu_0, \nu_0)} \left[\frac{\partial \ell(\mu, \nu)}{\partial\rho} \frac{\partial \ell(\mu_0, \nu_0)}{\partial\tau^2} \right] \right\} \Big|_{(\mu, \nu) = (\mu, \hat{\nu}_\mu), (\mu_0, \nu_0) = (\hat{\mu}, \hat{\nu})} &= \frac{1}{2\hat{\tau}_\mu^2} \text{tr} \left(\hat{P}_\mu^{-1} \hat{\mathbb{P}}_\mu \hat{P}_\mu^{-1} \hat{P} \right), \\ \left\{ E_{(\mu_0, \nu_0)} \left[\frac{\partial \ell(\mu, \nu)}{\partial\rho} \frac{\partial \ell(\mu_0, \nu_0)}{\partial\rho} \right] \right\} \Big|_{(\mu, \nu) = (\mu, \hat{\nu}_\mu), (\mu_0, \nu_0) = (\hat{\mu}, \hat{\nu})} &= \frac{\hat{\tau}^2}{2\hat{\tau}_\mu^2} \text{tr} \left(\hat{P}_\mu^{-1} \hat{\mathbb{P}}_\mu \hat{P}_\mu^{-1} \hat{\mathbb{P}} \right), \\ \left\{ E_{(\mu_0, \nu_0)} \left[\frac{\partial \ell(\mu, \nu)}{\partial\rho} \frac{\partial \ell(\mu_0, \nu_0)}{\partial\tilde{\beta}^\top} \right] \right\} \Big|_{(\mu, \nu) = (\mu, \hat{\nu}_\mu), (\mu_0, \nu_0) = (\hat{\mu}, \hat{\nu})} &= \frac{1}{\hat{\tau}_\mu^2} \tilde{\mathbf{X}}^\top \hat{P}_\mu^{-1} \hat{\mathbb{P}}_\mu \hat{P}_\mu^{-1} \left(\tilde{\mathbf{X}}_q \hat{\mu} + \tilde{\mathbf{X}} \hat{\beta} - \tilde{\mathbf{X}}_q \mu - \tilde{\mathbf{X}} \hat{\beta}_\mu \right), \\ \left\{ E_{(\mu_0, \nu_0)} \left[\frac{\partial \ell(\mu, \nu)}{\partial\tilde{\beta}} \frac{\partial \ell(\mu_0, \nu_0)}{\partial\tau^2} \right] \right\} \Big|_{(\mu, \nu) = (\mu, \hat{\nu}_\mu), (\mu_0, \nu_0) = (\hat{\mu}, \hat{\nu})} &= \mathbf{0}, \\ \left\{ E_{(\mu_0, \nu_0)} \left[\frac{\partial \ell(\mu, \nu)}{\partial\tilde{\beta}} \frac{\partial \ell(\mu_0, \nu_0)}{\partial\rho} \right] \right\} \Big|_{(\mu, \nu) = (\mu, \hat{\nu}_\mu), (\mu_0, \nu_0) = (\hat{\mu}, \hat{\nu})} &= \mathbf{0}, \\ \left\{ E_{(\mu_0, \nu_0)} \left[\frac{\partial \ell(\mu, \nu)}{\partial\tilde{\beta}} \frac{\partial \ell(\mu_0, \nu_0)}{\partial\tilde{\beta}^\top} \right] \right\} \Big|_{(\mu, \nu) = (\mu, \hat{\nu}_\mu), (\mu_0, \nu_0) = (\hat{\mu}, \hat{\nu})} &= \tilde{\mathbf{X}}^\top \hat{\Sigma}_\mu^{-1} \tilde{\mathbf{X}}, \end{aligned}$$

que são derivados da matriz

$$I_\nu(\mu, \nu; \mu_0, \nu_0) = E_{(\mu_0, \nu_0)} \left(\begin{array}{cc|cc} \frac{\partial \ell(\mu, \nu)}{\partial \tau^2} & \frac{\partial \ell(\mu_0, \nu_0)}{\partial \tau^2} & \frac{\partial \ell(\mu, \nu)}{\partial \tau^2} & \frac{\partial \ell(\mu_0, \nu_0)}{\partial \rho} \\ \frac{\partial \ell(\mu, \hat{\nu}_\mu)}{\partial \rho} & \frac{\partial \ell(\mu_0, \nu_0)}{\partial \tau^2} & \frac{\partial \ell(\mu, \hat{\nu}_\mu)}{\partial \rho} & \frac{\partial \ell(\mu_0, \nu_0)}{\partial \rho} \\ \hline \frac{\partial \ell(\mu, \hat{\nu}_\mu)}{\partial \hat{\beta}} & \frac{\partial \ell(\mu_0, \nu_0)}{\partial \tau^2} & \frac{\partial \ell(\mu, \hat{\nu}_\mu)}{\partial \hat{\beta}} & \frac{\partial \ell(\mu_0, \nu_0)}{\partial \rho} \\ \frac{\partial \ell(\mu, \hat{\nu}_\mu)}{\partial \hat{\beta}^\top} & \frac{\partial \ell(\mu_0, \nu_0)}{\partial \tau^2} & \frac{\partial \ell(\mu, \hat{\nu}_\mu)}{\partial \hat{\beta}^\top} & \frac{\partial \ell(\mu_0, \nu_0)}{\partial \rho} \end{array} \right).$$

A aproximação para a função de verossimilhança proposta por Barndorf-Nielsen tem a seguinte expressão:

$$\ell_{ABN}(\mu) = \ell_p(\mu) + \frac{1}{2} \log |j_{\nu\nu}(\mu, \hat{\nu}_\mu)| - \log |I_\nu(\mu, \hat{\nu}_\mu; \hat{\mu}, \hat{\nu})|.$$

Como, no nosso caso, a expressão acima é função das quantidades desconhecidas $\hat{\rho}$ e $\hat{\rho}_\mu$, obteremos os valores de $\ell_{ABN}(\mu)$ necessários à realização dos testes da razão de verossimilhanças maximizando numericamente a função

$$\ell_{ABN}(\mu) = \ell(\mu, \tilde{\beta}, \tau^2, \rho) + \frac{1}{2} \log |j_{\nu\nu}(\mu, \tilde{\beta}, \tau^2, \rho)| - \log |I_\nu(\mu, \tilde{\beta}, \tau^2, \rho; \hat{\mu}, \hat{\tilde{\beta}}, \hat{\tau}^2, \hat{\rho})|$$

sujeita às restrições $\partial \ell(\mu, \tilde{\beta}, \tau^2, \rho) / \partial \tilde{\beta} = 0$, $\partial \ell(\mu, \tilde{\beta}, \tau^2, \rho) / \partial \tau^2 = 0$ e $\partial \ell(\mu, \tilde{\beta}, \tau^2, \rho) / \partial \rho = 0$.

Modelo com Intercepto e Inclinação Aleatórios

O segundo modelo que utilizaremos na avaliação numérica possui um efeito aleatório a mais que o do exemplo anterior, efeito este que está associado ao tempo em que ocorreu a observação. Com isso não assumimos mais um modelo de correlação uniforme e sim a distância entre pares de observações para o mesmo indivíduo influencia na correlação entre ambas. O modelo é dado por

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 x_{2i} + \beta_3 x_{3i} + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij}. \quad (4.4)$$

Neste caso, como não utilizamos alguma transformação de variáveis, os cálculos são os mesmos apresentados na Seção 4.3. Como aqui temos dois efeitos aleatórios, assumimos para cada indivíduo que o vetor $\mathbf{b}_i = (b_{0i}, b_{1i})$ segue uma distribuição $Normal_2(\mathbf{0}, G)$, onde pode-se assumir uma estrutura bastante geral para a matriz G ,

$$G = \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{pmatrix}.$$

Supomos que os erros aleatórios, ϵ_{ij} , seguem uma distribuição $Normal(0, \alpha_4)$ e são independentes para todo $i = 1, \dots, N$ e $j = 1, \dots, n_i$.

5 *Resultados Numéricos*

5.1 Introdução

Neste capítulo avaliamos numericamente (simulações de Monte Carlo) a qualidade das inferências para alguns parâmetros de interesse nos dois modelos mistos apresentados na Seção 4.3. Essa comparação é feita tanto no caso de estimação pontual quanto em testes de hipóteses. Na Seção 5.2 faremos esta avaliação para o modelo com apenas o intercepto aleatório e na Seção 5.3 utilizaremos o modelo com intercepto e inclinação aleatórios. Consideramos inferências sobre um e dois parâmetros dos efeitos fixos.

No que se refere à estimação pontual, avaliamos os comportamentos dos estimadores obtidos pela maximização da função de verossimilhança usual ($\hat{\mu}$) e pela maximização da aproximação para a função de verossimilhança perfilada ajustada proposta por Barndorff-Nielsen (1983) ($\hat{\mu}_{ABN}$). Em cada caso, calculamos a média das estimativas, a variância, o erro quadrático médio (EQM), a assimetria e o excesso de curtose.

Nos testes de hipóteses, comparamos os desempenhos dos testes baseados na estatística da razão de verossimilhanças usual (RV , dado por (3.1)), na estatística da razão de verossimilhanças perfiladas modificadas (RV_{ABN} , ver equação (3.8)), na estatística da razão de verossimilhanças usual corrigida via Bartlett (RV^* , equação (3.10)) e o teste da razão de verossimilhanças usual baseado no método *bootstrap* (RV_{boot} , ver Seção 3.6). Basicamente, avaliamos as taxas de rejeição dos testes sob a hipótese nula e também comparamos os quantis amostrais das estatísticas RV , RV_{ABN} e RV^* com seus valores assintóticos. Essas comparações nos permitem avaliar a proximidade da distribuição nula das estatísticas de teste com relação a distribuição assintótica qui-quadrado. Gráficos de discrepâncias relativas de quantis são apresentados nas seções 5.2, 5.3 e no Apêndice B.

Em todos os casos simulados, utilizamos 3.000 amostras de Monte Carlo. Zucker, Lieberman e Manor (2000) afirmam que, a partir de 3.000 amostras, seus resultados já foram muito próximos dos obtidos com 10.000 amostras. A geração das amostras foi realizada da seguinte maneira:

1. definimos o número de observações para cada indivíduo;
2. geramos as matrizes X e Z ;

3. construímos então a matriz $\Sigma(\alpha)$, para α especificado;
4. com isso a amostra é dada por

$$\mathbf{Y}^* = \mathbf{X}\beta + \mathbf{N}^* \mathbf{A},$$

onde β é um vetor de parâmetros especificados, \mathbf{N}^* é um vetor gerado da distribuição $Normal_T(\mathbf{0}, I_T)$ e A é a matriz triangular superior obtida da decomposição de Cholesky $\Sigma = A^\top A$. Dessa forma, $\mathbb{E}(Y^*) = \mathbb{E}(X\beta + N^*A) = X\beta$ e $Var(Y^*) = Var(X\beta + N^*A) = A^\top Var(N^*)A = A^\top A = \Sigma$.

As simulações foram realizadas utilizando a linguagem de programação `0x`. A maximização da função de verossimilhança original foi feita utilizando o algoritmo quasi-Newton BFGS. Já na maximização do ajuste da função de verossimilhança perfilada modificada, como necessitamos maximizar uma função sujeita a determinadas restrições, foi utilizado o algoritmo SQP (*Sequential Quadratic Programming*). Ambos os métodos estão implementados na linguagem `0x`.

5.2 Modelo com Intercepto Aleatório

Para avaliar o desempenho do modelo com intercepto aleatório, consideramos $N = 5$ e $N = 10$ (indivíduos ou unidades experimentais). A quantidade de observações por indivíduo foi escolhida de tal forma que $n_i \in \{2, 3, 4, 5, 6\}$, para todo $i = 1, 2, \dots, N$. Consideramos três valores para o coeficiente de correlação ρ : 0,95; 0,9 e 0,8. Os valores verdadeiros dos demais parâmetros são: $\tau^2 = 1,05$; $(\beta_0, \beta_1, \beta_2, \beta_3) = (0; 0,2; 0; 0)$. Para cada indivíduo, obtemos os valores dos tempos (dados longitudinais), t_{ij} , de acordo com a distribuição uniforme padrão. E as variáveis x_{2i} e x_{3i} são variáveis dicotômicas. β_3 é o parâmetro de interesse.

Nos melhoramentos dos testes baseados num esquema de reamostragem (correção de Bartlett e teste *bootstrap*), devido ao custo computacional, optamos por gerar apenas 200 amostras *bootstrap*.

Com relação a estimação pontual, Tabela 5.1, o estimador de máxima verossimilhança usual ($\hat{\beta}_3$) e o de máxima verossimilhança perfilada modificada ($\hat{\beta}_{3_{ABN}}$) apresentaram desempenhos muito semelhantes. Ambos apresentaram estimativas muito próximas do verdadeiro valor de β_3 , zero. Por exemplo, para $N = 5$ e $\rho = 0,8$, observa-se que as médias das estimativas, $\hat{\beta}_3 = 0,0069$ e $\hat{\beta}_{3_{ABN}} = 0,0064$, diferem apenas na quarta casa decimal. As estimativas das variâncias dos estimadores $\hat{\beta}_{3_{ABN}}$ e $\hat{\beta}_3$ também não foram tão discrepantes, acarretando estimativas dos erros quadráticos médios (EQM) muito próximas. As estimativas dos coeficientes de assimetria e dos excessos de curtose desses estimadores de máxima verossimilhança foram bem próximos dos seus valores assintóticos, 0 e 0, respectivamente. Lembre que, assintoticamente, tais estimadores são normalmente distribuídos. Para os demais valores de ρ e de N , os comportamentos dos estimadores foram muito semelhantes ao caso de $N = 5$ e $\rho = 0,8$.

Tabela 5.1: Estatísticas descritivas das estimativas para β_3 no modelo com intercepto aleatório.

N	ρ	Estimador	Medidas				
			Média	Variância	EQM	Assimetria	Curtose
5	0,8	$\hat{\beta}_3$	0,0069	0,0814	0,0814	-0,0543	0,0550
		$\hat{\beta}_{3_{ABN}}$	0,0064	0,0794	0,0795	-0,0709	0,0396
	0,9	$\hat{\beta}_3$	-0,0050	0,0424	0,0424	-0,0980	-0,0182
		$\hat{\beta}_{3_{ABN}}$	-0,0050	0,0418	0,0418	-0,0973	-0,0076
	0,95	$\hat{\beta}_3$	-0,0043	0,0209	0,0209	-0,1071	0,0281
		$\hat{\beta}_{3_{ABN}}$	-0,0041	0,0207	0,0208	-0,1024	0,0228
10	0,8	$\hat{\beta}_3$	0,0018	0,0386	0,0386	-0,0677	-0,0285
		$\hat{\beta}_{3_{ABN}}$	0,0015	0,0385	0,0385	-0,0712	-0,0333
	0,9	$\hat{\beta}_3$	-0,0012	0,0194	0,0194	-0,0981	0,0527
		$\hat{\beta}_{3_{ABN}}$	-0,0012	0,0194	0,0194	-0,0980	0,0509
	0,95	$\hat{\beta}_3$	-0,0031	0,0095	0,0095	-0,0679	0,0977
		$\hat{\beta}_{3_{ABN}}$	-0,0031	0,0095	0,0095	-0,0677	0,1021

Ao testar $H_0 : \beta_3 = 0$ contra $H_1 : \beta_3 \neq 0$, os refinamentos do teste da razão de verossimilhanças apresentam resultados mais satisfatórios do que o teste original. Na Tabela 5.2, em geral, os testes baseados nas estatísticas da razão de verossimilhanças usuais tenderam a ser mais liberais quando comparados às propostas de melhoramento.

Para $N = 5$, $\rho = 0,8$ e nível nominal de 10%, o teste baseado na razão de verossimilhanças usual (RV) apresentou taxa de rejeição sob H_0 56,7% superior a esse nível: $(15,67\% - 10\%)/10\%$. O teste baseado na verossimilhança perfilada ajustada (RV_{ABN}) apresentou taxa de rejeição igual a 9,77%, ou seja, 2,3% inferior ao nível nominal. A taxa correspondente a correção do tipo Bartlett ($RV^* = 9,87\%$) foi 1,3% menor, enquanto o teste baseado no método *bootstrap* ($RV_{boot} = 10,30\%$) teve taxa 3% maior.

À medida que diminuimos o nível nominal, o teste baseado na estatística da razão de verossimilhanças usual tendeu a ser menos satisfatório. Ou seja, ainda para $N = 5$ e $\rho = 0,8$, mas agora para 5% de nível nominal, o teste baseado na RV apresentou taxa de rejeição 84,6% superior e, para o nível de 3%, obteve taxa 101,0% superior. Este comportamento não se verifica com tanta intensidade nos testes refinados. Estes resultados mostram o bom desempenho dos melhoramentos dos testes para um tamanho amostral relativamente pequeno. As taxas de rejeição sob H_0 ficaram próximas aos níveis nominais.

Ainda para $N = 5$, a comparação dos testes nos outros dois valores de ρ é semelhante a

Tabela 5.2: Taxas de rejeição (%) da hipótese nula ($H_0 : \beta_3 = 0$) calculadas a partir de 3000 réplicas de Monte Carlo e 200 amostras *bootstrap* no modelo com o intercepto aleatório.

N	ρ	Nível Nominal											
		10%				5%				3%			
		<i>RV</i>	<i>RV_{ABN}</i>	<i>RV*</i>	<i>RV_{boot}</i>	<i>RV</i>	<i>RV_{ABN}</i>	<i>RV*</i>	<i>RV_{boot}</i>	<i>RV</i>	<i>RV_{ABN}</i>	<i>RV*</i>	<i>RV_{boot}</i>
	0,8	15,67	9,77	9,87	10,30	9,23	4,67	4,73	5,10	6,03	3,07	3,13	3,23
5	0,9	15,60	9,17	9,87	10,37	8,63	5,07	5,33	5,57	5,93	3,20	3,57	3,93
	0,95	15,50	9,90	10,50	10,77	9,23	4,73	5,17	5,70	6,07	2,77	3,20	3,70
10	0,8	13,17	10,30	10,17	10,80	7,20	5,53	5,63	5,97	4,93	3,20	3,60	4,07
	0,9	12,77	9,80	10,10	10,33	7,00	5,07	5,20	5,53	4,47	2,90	3,20	3,63
	0,95	12,60	9,80	9,90	10,17	6,93	5,03	5,27	5,50	4,40	3,10	3,13	3,57

discutida para $\rho = 0,8$. A piora no desempenho do teste baseado na estatística *RV* com a diminuição do nível nominal se deve a uma aproximação insatisfatória da distribuição amostral da estatística de teste para sua distribuição assintótica na cauda.

Aumentando o tamanho amostral, agora para $N = 10$, todas as taxas de rejeição dos testes baseados na estatística da razão de verossimilhanças usuais tornaram-se mais próximas dos níveis nominais, como era de se esperar. Com $\rho = 0,8$, o teste que utiliza a estatística *RV* apresenta superioridade nas taxas de 31,7%, 44,0% e 64,3% em relação aos níveis 10%, 5% e 3% respectivamente. Percentuais que representam uma melhor aproximação se comparados aos resultados para $N = 5$. Nem todos os refinamentos dos testes apresentaram taxas mais próximas dos níveis nominais, ocorrendo em alguns casos um pequeno afastamento das taxas com relação aos resultados para $N = 5$. Mas como as taxas para $N = 5$ já são bastante satisfatórias, pode ser razoável pensar que para $N = 10$ alguma amostra Monte Carlo produza uma rejeição ou aceitação a mais provocando um pequeno distanciamento entre as taxas e seus correspondentes níveis nominais.

Comparando as propostas de refinamento do teste da razão de verossimilhanças, na maioria dos casos simulados, o teste baseado na razão de verossimilhanças perfilada modificada obteve taxas de rejeição mais próximas dos níveis nominais. Isso nos dá indícios de uma boa aproximação da distribuição amostral de *RV_{ABN}*, sob H_0 , para a distribuição assintótica χ_1^2 . Devemos salientar que, devido ao excessivo tempo computacional, optamos por utilizar 200 amostras *bootstrap* para realizar o teste *bootstrap* bem como a correção do tipo Bartlett, o que pode ser um número pequeno.

Outra questão relevante é que, em apenas um caso ($N = 5, \rho = 0,8$ e nível 5%), o teste *bootstrap* apresentou melhor desempenho comparado ao baseado na estatística da razão de verossimilhanças usual corrigida via Bartlett. Como observamos apenas um caso, pode-se suspeitar

que seja necessário um menor número de amostras *bootstrap* para se ter taxas mais próximas dos níveis nominais com o teste baseado em RV^* do que com o próprio teste *bootstrap*.

Tabela 5.3: Médias e variâncias amostrais de RV , RV_{ABN} e RV^* e seus respectivos valores assintóticos, no modelo com o intercepto aleatório.

N	ρ	Momento	χ_1^2	RV	RV_{ABN}	RV^*
5	0,8	Média	1	1,3471	0,9757	0,9930
		Variância	2	3,7559	1,9468	2,0205
	0,9	Média	1	1,3542	0,9902	1,0240
		Variância	2	3,6225	1,9557	2,1158
	0,95	Média	1	1,3284	0,9804	1,0096
		Variância	2	3,5259	1,8690	2,0551
10	0,8	Média	1	1,1811	1,0123	1,0299
		Variância	2	2,8113	2,1991	2,2265
	0,9	Média	1	1,1789	1,0340	1,0315
		Variância	2	2,8444	2,7156	2,2216
	0,95	Média	1	1,1353	1,0276	0,9911
		Variância	2	2,6629	2,9670	2,0389

Uma maneira de avaliarmos a aproximação da distribuição amostral das estatísticas dos testes, sob a hipótese nula, para sua respectiva distribuição assintótica consiste em comparar os momentos destas distribuições. Na Tabela 5.3 apresentamos as médias e variâncias amostrais das estatísticas baseadas na razão de verossimilhanças usual, da razão de verossimilhanças perfiladas modificadas e das estatísticas usuais corrigidas via Bartlett. Neste caso, não podemos comparar o teste *bootstrap* pois este não fornece uma estatística de teste que possibilite a avaliação de sua distribuição amostral. Os momentos amostrais observados, média e variância, da estatística RV estão mais afastados dos momentos assintóticos comparativamente aos momentos amostrais dos melhoramentos (RV_{ABN} e RV^*). Ou seja, a aproximação da distribuição amostral da estatística RV , resumida pelos dois primeiros momentos, não parece tão satisfatória quanto a aproximação para as demais propostas de estatísticas de teste. Esta é uma provável explicação desempenho insatisfatório dos testes baseados na estatística RV .

A estatística RV , para $N = 5$ e $\rho = 0,8$, obteve média 34,7% superior a média da distribuição assintótica, enquanto que a estatística RV_{ABN} apresenta média 2,4% inferior ao momento assintótico e a RV^* , 0,7% inferior. Já a variância da estatística RV apresenta-se 87,8% maior do que seu valor assintótico, ao passo que a estatística RV_{ABN} obteve variância 2,7% inferior e a RV^* apenas 1,0% superior. Para os demais valores de ρ , as médias têm comportamentos semelhantes.

Para $N = 10$ indivíduos, as médias e variâncias amostrais de RV se aproximaram dos seus respectivos valores assintóticos. Os refinamentos mantiveram suas médias próximas da média da distribuição assintótica, porém suas variâncias aumentaram.

Outra maneira de se avaliar a aproximação da distribuição amostral das estatísticas de teste para sua distribuição assintótica é através das discrepâncias relativas de quantis (Figura 5.1). A discrepância, para testar k parâmetros, é definida como

$$\frac{ET(q) - \chi_k^2(q)}{\chi_k^2(q)}$$

onde $ET(q)$ é o quantil amostral de ordem q do conjunto de valores da Estatística de Teste avaliada e $\chi_k^2(q)$ é o respectivo quantil da distribuição qui-quadrado com k graus de liberdade. Note que, quanto mais próxima de zero for a discrepância, melhor é a aproximação do quantil amostral para seu respectivo valor assintótico sob a hipótese nula.

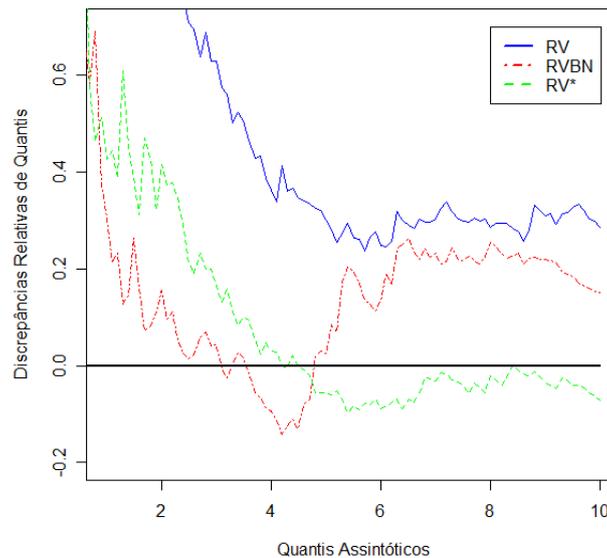


Figura 5.1: Gráfico das discrepâncias relativas de quantis para $N = 5$ e $\rho = 0,8$ no modelo com intercepto aleatório ($H_0 : \beta_3 = 0$).

A Figura 5.1 apresenta o gráfico das discrepâncias relativas de quantis *versus* seus correspondentes quantis assintóticos. Nela percebe-se que a curva de discrepâncias produzida pelo teste usual foi maior do que a curva das propostas de melhoramento, para todos os quantis avaliados. O teste baseado na RV_{ABN} obteve, para quantis próximos de zero, uma melhor aproximação em relação as outras duas propostas de teste, mas aproximadamente a partir do quantil 5 começa a apresentar um maior afastamento em relação ao teste que utiliza a estatística RV^* . Percebe-se também para quantis maiores que 5, aproximadamente, que o teste baseado em RV^* apresenta

a melhor aproximação.

Para $N = 10$ (Figura 5.2), a curva de discrepâncias do teste usual tendeu a se aproximar mais da ordenada zero mas em relação a curva para $N = 5$. Porém, com exceção dos quantis próximos de zero, a curva continua apresentando as maiores discrepâncias se comparada às demais curvas. As discrepâncias dos melhoramentos apresentam comportamento similar até quantis próximos de 5, a partir deste quantil o teste baseado na RV^* mostra menores discrepâncias. Os gráficos relativos às demais situações foram semelhantes e estão no Apêndice B dessa dissertação.

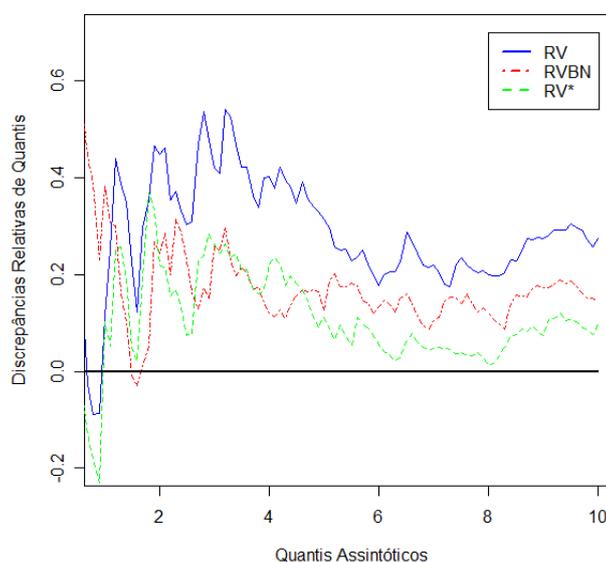


Figura 5.2: Gráfico das discrepâncias relativas de quantis para $N = 10$ e $\rho = 0,8$ no modelo com intercepto aleatório ($H_0 : \beta_3 = 0$).

Anteriormente comentamos que o desempenho das propostas baseadas no teste *bootstrap* e na correção de Bartlett poderiam não ser tão satisfatórios devido ao número de amostras *bootstrap* que utilizamos. Como forma de avaliar esse efeito nos desempenhos destes testes, realizamos dois conjuntos de simulações com 100 e 500 amostras *bootstrap*, para $\rho = 0,95$ e $N = 5$. Na Tabela 5.4, apresentamos as taxas de rejeição sob $H_0 : \beta_3 = 0$ dos testes que utilizam *bootstrap* e comparamos às taxas obtidas com 200 amostras *bootstrap*.

Nos testes ao nível de significância de 10%, observamos claramente, com o aumento da quantidade de amostras *bootstrap*, que os testes apresentam taxas de rejeição sob a hipótese nula mais próximas desse nível. Para os níveis 5% e 3% as taxas do teste baseado em RV^* com 500 reamostras *bootstrap* obtiveram um pequeno afastamento dos respectivos níveis nominais em relação ao caso com 200 reamostras, já o teste *bootstrap* apresentou melhora em todos os casos de aumento do número de reamostras. Além disso, com exceção de um caso, o teste corrigido

via Bartlett conseguiu as taxas mais próximas dos níveis nominais.

A diferença entre as taxas de rejeição dos dois testes diminuiu com o aumento do número de reamostras. Para 100 amostras bootstrap, ao nível 10%, a taxa do teste baseado em RV_{boot} foi 3,70% superior a do baseado em RV^* , para 200 reamostras essa diferença diminuiu para 2,57% e com 500 reamostras a diferença foi 2,28%. Considerando o nível de significância de 5% as taxas do teste *bootstrap* foram 17,23% e 10,25% superiores para 100 e 200 respectivamente, e com 500 reamostras a taxa do teste *bootstrap* foi 1,27% inferior. Realizando o teste ao nível de 3% observam-se as maiores diminuições nas diferenças: com 100 reamostras a diferença foi de 28,39%; para 200 reamostras a diferença passou para 15,63%; e com 500 reamostras o teste *bootstrap* teve taxa mais próxima do nível nominal que o teste corrigido via Bartlett e a diferença foi de 3,66%. Este resultado nos dá indícios de que utilizando números de reamostras relativamente grandes, as taxas para quantis mais próximos de zero podem ter melhor aproximação para os níveis assintóticos no teste *bootstrap* do que no teste usual corrigido via Bartlett.

Um aspecto importante a se destacar é que os testes corrigidos via Bartlett e *bootstrap*, para os níveis nominais 10% e 5%, não apresentaram melhora nas taxas de rejeição em relação as taxas obtidas por RV_{ABN} (Tabela 5.2). Já considerando o nível nominal de 3% o teste baseado em RV^* tem melhor desempenho que o RV_{ABN} para 100 e 200 réplicas e o teste *bootstrap* apresenta melhoria apenas quando observa-se 500 amostras *bootstrap*.

Tabela 5.4: Taxas de rejeição (%) da hipótese nula ($H_0 : \beta_3 = 0$) calculadas a partir de 3000 réplicas de Monte Carlo e 100, 200 e 500 amostras *bootstrap*, para $N = 5$ e $\rho = 0,95$ no modelo com o intercepto aleatório.

Quantidade de Reamostras	Nível Nominal					
	10%		5%		3%	
	RV^*	RV_{boot}	RV^*	RV_{boot}	RV^*	RV_{boot}
100	10,80	11,20	5,63	6,60	3,17	4,07
200	10,50	10,77	5,17	5,70	3,20	3,70
500	10,07	10,30	4,73	4,67	2,73	2,83

Simulamos uma situação complementar para um valor de $\tau^2 = 10$, aumentando assim a variabilidade dos dados observados. A Tabela 5.5 apresenta as taxas de rejeição dos testes sob a hipótese nula $H_0 : \beta_3 = 0$ para $N = 5$ e $\rho = 0,95$. Com exceção do teste ao nível de significância de 5% baseados em RV^* e em RV_{boot} , todas as taxas apresentam uma suave melhora. Ou seja, com o aumento da variabilidade dos dados observados e considerando uma correlação positiva demasiadamente alta entre as observações do mesmo indivíduo, em geral, observamos taxas de rejeição sob a hipótese nula um pouco mais próximas dos níveis nominais assintóticos do que as

taxas observadas na Tabela 5.2.

Tabela 5.5: Taxas de rejeição (%) da hipótese nula ($H_0 : \beta_3 = 0$) calculadas a partir de 3000 réplicas de Monte Carlo e 200 amostras *bootstrap*, utilizando $\tau^2 = 10$, $N = 5$ e $\rho = 0,95$ no modelo com o intercepto aleatório.

Nível Nominal	RV	RV_{ABN}	RV^*	RV_{boot}
10%	14,90	9,53	10,27	10,40
5%	8,87	4,80	5,27	5,77
3%	5,97	3,03	3,20	3,50

Em todas as simulações apresentadas até aqui, consideramos uma alta correlação entre as observações do mesmo indivíduo. Na tabela 5.6 são mostrados resultados para os mesmos parâmetros definidos no início desta seção, porém mudando a correlação intra-indivíduo para $\rho = 0,5$ e consideramos apenas $N = 10$. As taxas de rejeição para o teste da razão de verossimilhanças usuais ficaram mais próximas dos níveis assintóticos se comparadas às taxas obtidas para $N = 10$ na Tabela 5.2. Para as propostas de melhoramento esta relação não está muito evidente.

Ainda na Tabela 5.6, observa-se que as taxas de rejeição utilizando o teste *bootstrap* são mais próximas dos níveis nominais, em todos os três valores avaliados, do que as taxas obtidas no teste baseado na estatística RV_{ABN} .

Tabela 5.6: Taxas de rejeição (%) da hipótese nula ($H_0 : \beta_3 = 0$) calculadas a partir de 3000 réplicas de Monte Carlo e 200 reamostras *bootstrap*, utilizando $\rho = 0,5$ e $N = 10$ no modelo com o intercepto aleatório.

Nível Nominal	RV	RV_{ABN}	RV^*	RV_{boot}
10%	12,37	9,43	9,33	9,50
5%	6,90	4,33	5,03	5,13
3%	3,93	2,50	2,47	2,83

Suponha agora que estejamos interessados em testar $H_0 : (\beta_2, \beta_3) = (0, 0)$ contra $H_1 : (\beta_2, \beta_3) \neq (0, 0)$. Para $N = 5$, $\rho = 0,95$ e $\tau^2 = 1,05$, sob H_0 , as quatro propostas de testes avaliadas nesta dissertação apresentam taxas de rejeição mais afastadas dos níveis nominais, se comparadas às taxas obtidas no teste apenas de β_3 (Tabela 5.7). Os maiores afastamentos percentuais ocorreram com o teste baseado na estatística RV .

Na Tabela 5.7 percebe-se que ao nível nominal 10% a taxa de rejeição mais próxima foi do teste RV_{ABN} , para o nível 5% os testes baseados em RV_{ABN} e RV^* obtiveram a mesma taxa cujo valor é o mais próximo do nível nominal. Já para o teste ao nível 3%, das propostas de

melhoramento dos testes, a taxa de rejeição do baseado em RV_{ABN} foi a mais afastada.

Tabela 5.7: Taxas de rejeição (%) da hipótese nula ($H_0 : (\beta_2, \beta_3) = (0, 0)$) calculadas a partir de 3000 réplicas de Monte Carlo e 200 reamostras *bootstrap*, utilizando $\tau^2 = 1,05$, $\rho = 0,95$ e $N = 5$ no modelo com o intercepto aleatório.

Nível Nominal	RV	RV_{ABN}	RV^*	RV_{boot}
10%	16,77	10,27	10,50	10,90
5%	9,70	5,57	5,57	5,87
3%	6,67	3,67	3,43	3,57

Na Figura 5.3, apresentamos as discrepâncias relativas de quantis no caso de testarmos $H_0 : (\beta_2, \beta_3) = (0, 0)$ contra $H_1 : (\beta_2, \beta_3) \neq (0, 0)$. Em grande parte dos quantis avaliados o teste da razão de verossimilhanças usuais apresentou as maiores discrepâncias relativas de quantis, enquanto que o teste da razão de verossimilhanças usual corrigido via Bartlett foi o de menores discrepâncias.

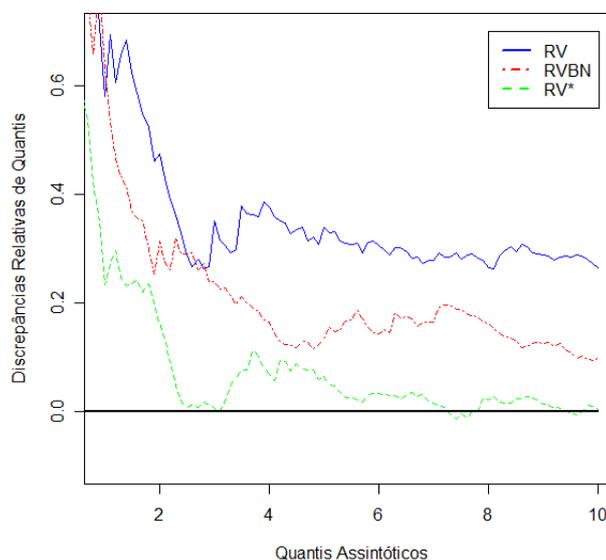


Figura 5.3: Gráfico das discrepâncias relativas de quantis para o teste de $H_0 : (\beta_2, \beta_3) = (0, 0)$ contra $H_1 : (\beta_2, \beta_3) \neq (0, 0)$, com $N = 5$, $\rho = 0,95$ e $\tau^2 = 1,05$ no modelo com intercepto aleatório.

Os resultados para o modelo misto com apenas um efeito aleatório, o intercepto, favorecem, de certo modo, o teste da razão de verossimilhanças perfiladas modificadas baseado em Severini (1998). Em Zucker, Lieberman e Manor (2000), para um modelo com o intercepto

e inclinação aleatórios, as taxas de rejeição, sob a hipótese nula, obtidas pelo teste da razão de verossimilhanças perfiladas ajustadas proposta por Cox e Reid (1987) apresentam-se mais próximas dos níveis nominais do que as baseadas no teste usual, mas ainda sim essas taxas são demasiadamente afastadas dos correspondentes níveis nominais. Então, além de modificar a verossimilhança perfilada, os autores utilizam correção do tipo Bartlett na estatística do teste da razão de verossimilhanças perfiladas modificadas a fim de tornar as taxas de rejeição mais próximas dos níveis nominais. E ainda assim algumas taxas estiveram mais afastadas dos níveis do que o próprio teste usual corrigido via Bartlett. Já com o ajuste na função de verossimilhança que utilizamos, baseado em Severini (1998), os resultados mostram que em muitas situações o teste baseado apenas na razão de verossimilhanças perfiladas modificadas apresenta taxas bem próximas dos respectivos níveis nominais, sem a necessidade da aplicação da correção de Bartlett. Veremos na próxima seção se os resultados obtidos para o modelo com o intercepto aleatório seguirão o mesmo comportamento se adicionarmos um efeito aleatório associado ao tempo das observações.

Devido a utilização de métodos de maximização numérica para encontrarmos as estimativas de máxima verossimilhança para este modelo, em cada simulação ocorreram alguns erros devido aos algoritmos de maximização não encontrarem o máximo global da função. Quando se trabalha no contexto de pequenas amostras, estamos sujeitos ainda mais a tais erros. O maior percentual de erros que ocorreram para as simulações deste modelo foi de 0,67%, o que representa um percentual relativamente pequeno.

5.3 Modelo com Intercepto e Inclinação Aleatórios

Nesta seção apresentamos alguns resultados para o segundo modelo apresentado na Seção 4.3. As simulações para este modelo foram realizadas considerando os números de indivíduos $N = 15$ e $N = 20$, e quantidade de observações por indivíduo $n_i \in \{2, 3, 4, 5, 6, 7, 8, 9\}$, para todo $i = 1, 2, \dots, N$. Os valores dos parâmetros utilizados são: $\alpha_1 = 1$; $\alpha_2 = 0$ e $0,35$; $\alpha_3 = 0,5$ e 1 ; $\alpha_4 = 0,05$; e $(\beta_0, \beta_1, \beta_2, \beta_3) = (0; 0, 2; 0; 0)$. O tempo das observações dos indivíduos, t_{ij} , foram obtidos de acordo com a distribuição Uniforme[0;100], e as variáveis x_{2ij} e x_{3ij} são variáveis dicotômicas.

As propostas de melhoramento dos testes que utilizam *bootstrap* não serão comparadas para este modelo. O tempo computacional para realizar as simulações foi bastante elevado, por isso decidimos comparar somente o teste usual com o baseado na razão de verossimilhanças perfiladas modificadas. Assim como vimos na seção anterior, veremos aqui que o teste baseado na estatística RV_{ABN} também é bastante satisfatório comparado ao usual.

Começaremos avaliando o caso onde o parâmetro de interesse é β_3 . As médias das estimativas pontuais para β_3 (Tabela 5.8) derivadas da maximização da verossimilhança usual e

da verossimilhança perfilada modificada são muito próximas, tanto para $N = 15$ quanto para $N = 20$.

Tabela 5.8: Estatísticas descritivas das estimativas para β_3 no modelo com intercepto e inclinação aleatórios.

N	Medidas							
	α_2	α_3	<i>Estimador</i>	Média	Variância	EQM	Assimetria	Curtose
15	0	0,5	$\hat{\beta}_3$	-0,0043	0,1171	0,1171	0,0452	0,0246
			$\hat{\beta}_{3_{ABN}}$	-0,0051	0,1151	0,1152	0,0454	-0,0109
	0	1,0	$\hat{\beta}_3$	-0,0041	0,1171	0,1171	0,0429	0,0259
			$\hat{\beta}_{3_{ABN}}$	-0,0050	0,1152	0,1152	0,0436	-0,0096
	0,35	0,5	$\hat{\beta}_3$	-0,0047	0,1141	0,1141	0,0423	0,0171
			$\hat{\beta}_{3_{ABN}}$	-0,0050	0,1117	0,1117	0,0460	-0,0160
0,35	1,0	$\hat{\beta}_3$	-0,0040	0,1159	0,1159	0,0433	0,0157	
		$\hat{\beta}_{3_{ABN}}$	-0,0049	0,1138	0,1139	0,0473	-0,0111	
20	0	0,5	$\hat{\beta}_3$	-0,0055	0,1067	0,1067	-0,0940	0,0795
			$\hat{\beta}_{3_{ABN}}$	-0,0041	0,1055	0,1055	-0,0913	0,0616
	0	1,0	$\hat{\beta}_3$	-0,0056	0,1066	0,1066	-0,0952	0,0769
			$\hat{\beta}_{3_{ABN}}$	-0,0042	0,1054	0,1055	-0,0921	0,0597
	0,35	0,5	$\hat{\beta}_3$	-0,0065	0,1028	0,1028	-0,1053	0,0839
			$\hat{\beta}_{3_{ABN}}$	-0,0048	0,1014	0,1014	-0,0961	0,0626
0,35	1,0	$\hat{\beta}_3$	-0,0063	0,1049	0,1050	-0,1032	0,0796	
		$\hat{\beta}_{3_{ABN}}$	-0,0046	0,1036	0,1036	-0,0938	0,0589	

Do mesmo modo, as variabilidades das estimativas foram muito próximas. Em todos os casos, as estimativas de máxima verossimilhança perfilada modificada apresentaram menor variabilidade, mas esta diferença também é observada apenas na terceira casa decimal. À medida que aumentamos o número de indivíduos observados, a variabilidade das estimativas obtidas por ambos os métodos diminui. Sendo as médias e variâncias muito semelhantes, os EQMs dos métodos também é muito similar. E os coeficientes de assimetria e excesso de curtose são próximos de zero, e não diferem muito entre os métodos.

No caso de testarmos $H_0 : \beta_3 = 0$ contra $H_1 : \beta_3 \neq 0$ veremos que o desempenho do teste da razão de verossimilhanças perfiladas modificadas foi mais satisfatório do que o teste da razão de verossimilhanças usual, em todos os casos analisados. Na Tabela 5.9, são apresentadas as taxas

Tabela 5.9: Taxas de rejeição (%) da hipótese nula ($H_0 : \beta_3 = 0$) calculadas a partir de 3000 réplicas de Monte Carlo no modelo com intercepto e inclinação aleatórios.

N	α_2	α_3	Nível Nominal					
			10%		5%		3%	
			RV	RV_{ABN}	RV	RV_{ABN}	RV	RV_{ABN}
15	0	0,5	13,00	10,50	7,40	5,27	4,83	3,30
	0	1,0	12,97	10,43	7,40	5,27	4,87	3,33
	0,35	0,5	13,10	10,77	7,67	5,53	4,93	3,30
	0,35	1,0	13,10	10,53	7,47	5,37	4,83	3,30
20	0	0,5	11,13	9,20	6,00	4,40	3,53	2,70
	0	1,0	10,93	9,10	6,00	4,37	3,50	2,70
	0,35	0,5	11,30	9,37	6,33	4,57	3,70	2,83
	0,35	1,0	11,03	9,20	6,10	4,40	3,73	2,90

de rejeição, sob a hipótese nula, de ambas as propostas de teste.

As taxas de rejeição obtidas pelo teste baseado na estatística RV não apresentaram-se tão afastadas dos níveis nominais quanto as obtidas no modelo com apenas o intercepto aleatório, mas ainda este teste mostrou-se bastante liberal. Analisando o caso de $N = 15$, $\alpha_2 = 0$ e $\alpha_3 = 0,5$, o teste baseado na razão de verossimilhanças usual apresentou taxa de rejeição 30,0% superior ao nível nominal 10%, já o teste baseado na aproximação proposta por Severini (1998) apresentou taxa de rejeição 5,0% superior ao mesmo nível. O teste de hipóteses utilizando a estatística RV ao nível de significância 5% obteve taxa de rejeição 48,0% superior, enquanto que o baseado na RV_{ABN} obteve taxa 5,4% superior a esse nível. E ao nível 3%, o teste usual mostrou taxa 61,0% superior, ao passo que o teste baseado em RV_{ABN} mostrou taxa 10% superior. Percebe-se claramente a deterioração de ambos os testes conforme consideramos níveis nominais menores. Nas demais situações simuladas para $N = 15$, as taxas têm comportamentos muito semelhantes aos observados para $\alpha_2 = 0$ e $\alpha_3 = 0,5$.

Com $N = 20$ indivíduos, todas as taxas de rejeição do teste da razão de verossimilhanças usual ficaram mais próximas dos níveis nominais, comparadas as do caso onde $N = 15$. Considerando $N = 20$, $\alpha_2 = 0$ e $\alpha_3 = 0,5$, os percentuais de afastamento das taxas são 11,3%, 20,0% e 17,7% para os níveis nominais de 10%, 5% e 3%, respectivamente. Para o teste da razão de verossimilhanças perfiladas modificadas se observa-se os afastamentos de 8,0%, 12,0% e 10%, para os mesmos níveis nominais, respectivamente. As demais situações para $N = 20$ também seguem o mesmo comportamento observado para $\alpha_2 = 0$ e $\alpha_3 = 0,5$. Isto quer dizer que mesmo observando taxas de rejeição do teste baseado em RV mais próximas dos níveis nominais, ainda

o teste baseado em RV_{ABN} apresenta melhor aproximação. Note que este teste foi conservador, rejeitando a hipótese nula aquém dos níveis nominais.

Tabela 5.10: Médias e variâncias amostrais de RV , RV_{ABN} e seus respectivos valores assintóticos, no modelo com intercepto e inclinação aleatórios.

N	α_2	α_3	Momento	χ_1^2	RV	RV_{ABN}	
15	0	0,5	Média	1	1,1942	1,0443	
			Variância	2	2,9237	2,2747	
	0	1,0	Média	1	1,1936	1,0404	
			Variância	2	2,9228	2,1930	
	15	0,35	0,5	Média	1	1,2119	1,0429
				Variância	2	3,0087	2,2024
0,35		1,0	Média	1	1,2037	1,0428	
			Variância	2	2,9691	2,2014	
20	0	0,5	Média	1	1,0850	0,9756	
			Variância	2	2,5108	2,0092	
	0	1,0	Média	1	1,0838	0,9749	
			Variância	2	2,5028	2,0043	
	20	0,35	0,5	Média	1	1,0926	0,9733
				Variância	2	2,5527	2,0037
0,35		1,0	Média	1	1,0896	0,9734	
			Variância	2	2,2191	1,9989	

Na Tabela 5.10, apresentamos as médias e variâncias amostrais das estatísticas da razão de verossimilhanças usual e da razão de verossimilhanças perfiladas modificadas. Como observamos um melhor desempenho do teste melhorado, é de se esperar uma melhor aproximação da distribuição amostral de RV_{ABN} , sob a hipótese nula, para a distribuição qui-quadrado com um grau de liberdade. Consequentemente, os momentos da distribuição amostral de RV_{ABN} seriam mais próximos dos assintóticos.

Comparando as estatísticas da razão de verossimilhanças, percebe-se que a baseada na verossimilhança perfilada modificada apresenta seus dois primeiros momentos mais próximos dos momentos assintóticos em todos os casos avaliados. Em ambas propostas, as médias e variâncias, para $N = 15$ indivíduos, estão mais afastadas dos seus respectivos valores assintóticos do que para $N = 20$, ou seja, com o aumento do número de indivíduos, temos melhores aproximações.

Os gráficos das discrepâncias relativas de quantis indicam o melhor desempenho do teste

da razão de verossimilhanças perfiladas modificadas. Na Figura 5.4, é possível comparar as curvas de discrepâncias relativas as duas propostas para $N = 15$, $\alpha_2 = 0$ e $\alpha_3 = 0,5$. A curva de discrepâncias da estatística usual apresenta-se mais próxima da ordenada zero até aproximadamente o quantil dois e, após esse valor a curva da estatística melhorada está mais próxima da ordenada zero.

O bom desempenho do teste melhorado, para o modelo com intercepto e inclinação aleatórios, também pode ser explicado pela melhor aproximação dos quantis amostrais para os quantis assintóticos, indicando novamente uma melhor aproximação da distribuição amostral de RV_{ABN} para a distribuição assintótica sob a hipótese nula se comparada à distribuição amostral de RV .

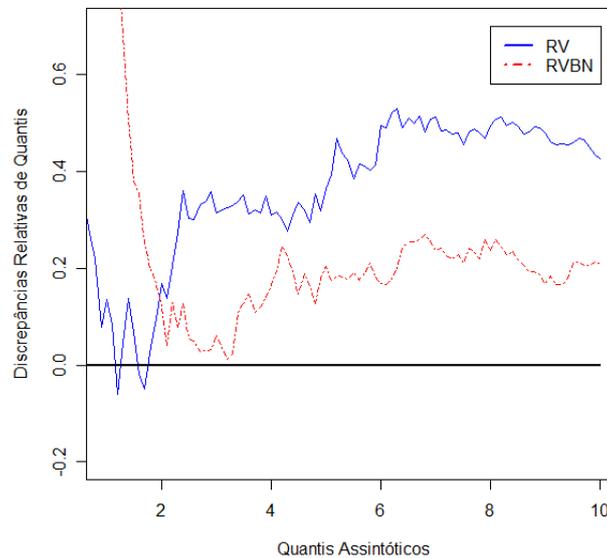


Figura 5.4: Gráfico das discrepâncias relativas de quantis para $N = 15$, $\alpha_2 = 0$ e $\alpha_3 = 0,5$ no modelo com intercepto e inclinação aleatórios ($H_0 : \beta_3 = 0$).

A Figura 5.5 mostra as discrepâncias para $N = 20$, $\alpha_2 = 0$ e $\alpha = 0,5$. A curva de discrepâncias da estatística de teste usual fica mais próxima da ordenada zero do que a do teste melhorado apenas numa pequena região em torno do quantil dois e do quantil oito. Nos demais casos a estatística melhorado apresenta aproximação mais satisfatória.

No modelo com o intercepto e inclinação aleatórios também simulamos o caso de testar $H_0 : (\beta_2, \beta_3) = (0, 0)$ contra $H_1 : (\beta_2, \beta_3) \neq (0, 0)$. A tabela 5.11 apresenta as taxas de rejeição considerando apenas $N = 15$.

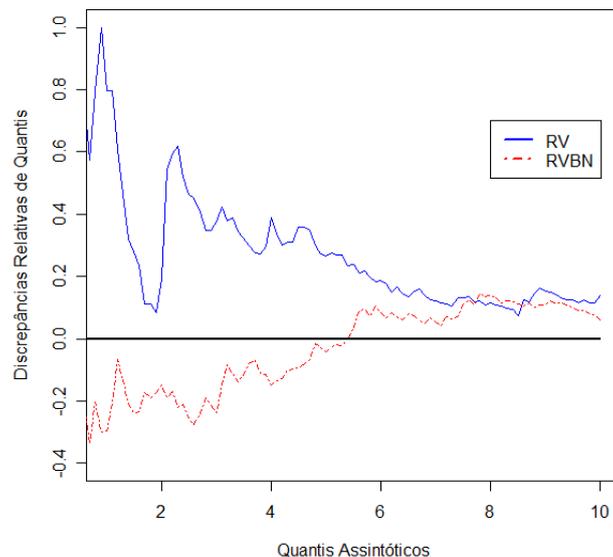


Figura 5.5: Gráfico das discrepâncias relativas de quantis para $N = 20$, $\alpha_2 = 0$ e $\alpha = 0,5$ no modelo com intercepto e inclinação aleatórios ($H_0 : \beta_3 = 0$).

Tabela 5.11: Taxas de rejeição (%) da hipótese nula ($H_0 : (\beta_2, \beta_3) = (0, 0)$) calculadas a partir de 3000 réplicas de Monte Carlo, para $N = 15$ no modelo com intercepto e inclinação aleatórios.

N	α_2	α_3	Nível Nominal					
			10%		5%		3%	
			RV	RV_{ABN}	RV	RV_{ABN}	RV	RV_{ABN}
	0	0,5	13,43	10,43	7,07	5,57	4,63	3,57
15	0	1,0	13,47	10,40	7,03	5,57	4,67	3,57
	0,35	0,5	13,63	10,53	7,27	5,70	4,80	3,53
	0,35	1,0	13,73	10,67	7,33	5,73	4,73	3,53

O teste usual, ao nível 10%, teve um pequeno afastamento nas taxas de rejeição se comparado aos resultados da Tabela 5.9, enquanto que o teste melhorado manteve taxas semelhantes. Nos outros níveis nominais, em geral, o teste usual apresentou uma melhora e o teste da razão de verossimilhanças perfiladas modificadas mostrou taxas com pequenos afastamentos. Mas, em todos os casos, ainda a proposta de melhoramento do teste apresenta-se mais satisfatória do que o teste da razão de verossimilhanças usual.

Na Figura 5.6, apresentamos as discrepâncias relativas de quantis no caso de testarmos $H_0 : (\beta_2, \beta_3) = (0, 0)$ contra $H_1 : (\beta_2, \beta_3) \neq (0, 0)$. Percebe-se que, para quantis próximos de zero, há uma maior variabilidade nas discrepâncias e as propostas se alternam quanto a melhor aproximação. Mas, para quantis mais afastados de zero, vemos a melhor aproximação da

estatística da razão de verossimilhanças perfiladas modificadas.

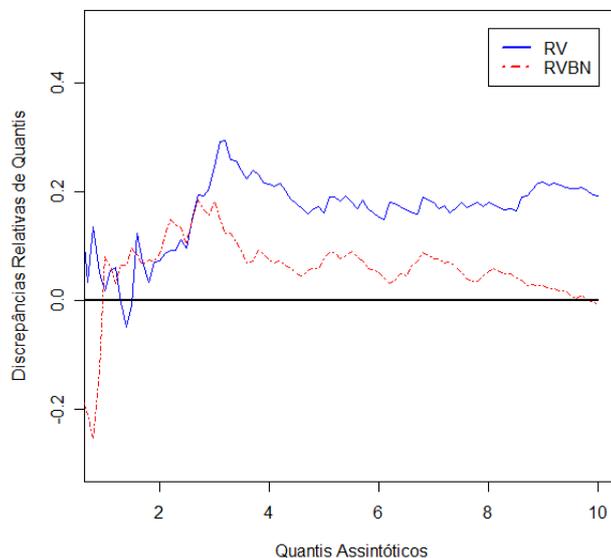


Figura 5.6: Gráfico das discrepâncias relativas de quantis para o teste de $H_0 : (\beta_2, \beta_3) = (0, 0)$ contra $H_1 : (\beta_2, \beta_3) \neq (0, 0)$, com $N = 15$, $\alpha_2 = 0$ e $\alpha_3 = 0,5$ no modelo com intercepto e inclinação aleatórios.

Por fim analisamos os dois primeiros momentos das estatísticas de teste comparadas (Tabela 5.12), na situação de testarmos β_2 e β_3 , considerando $N = 15$ indivíduos. Ao contrário do caso de testar apenas β_3 , aqui as estimativas dos dois primeiros momentos da distribuição amostral da estatística RV_{ABN} são mais próximos dos seus respectivos momentos assintóticos.

Da mesma forma que discutimos ao final da Seção 5.2, aqui os resultados favorecem o teste da razão de verossimilhanças perfiladas modificadas baseado em Severini (1998). Comentamos que Zucker, Lieberman e Manor (2000), com o mesmo modelo desta seção, obtiveram as taxas de rejeição, sob a hipótese nula, para o teste da razão de verossimilhanças perfiladas ajustadas baseado em Cox e Reid (1987) mais próximas dos níveis nominais do que o teste usual, mas ainda sim tais taxas foram demasiadamente afastadas dos correspondentes níveis nominais. Eles apenas fazem o teste para um parâmetro, β_3 .

Apesar de os resultados dos testes baseados na estatística RV_{ABN} serem satisfatórios, podemos compará-lo com o desempenho do teste *bootstrap* e com o teste da razão de verossimilhanças usuais corrigido via Bartlett. Acreditamos que neste caso o teste corrigido via Bartlett e o teste da razão de verossimilhanças perfiladas modificadas se alternariam quanto ao desempenho mais satisfatório.

Tabela 5.12: Médias e variâncias amostrais de RV , RV_{ABN} e seus respectivos valores assintóticos, para o teste de $H_0 : (\beta_2, \beta_3) = (0, 0)$ contra $H_1 : (\beta_2, \beta_3) \neq (0, 0)$ considerando $N = 15$.

N	α_2	α_3	Momento	χ_1^2	RV	RV_{ABN}
	0	0,5	Média	2	2,2893	2,0884
			Variância	4	5,4136	4,4946
	0	1,0	Média	2	2,2896	2,0887
			Variância	4	5,4148	4,4963
15	0,35	0,5	Média	2	2,3120	2,0991
			Variância	4	5,5199	4,7570
	0,35	1,0	Média	2	2,3036	2,0930
			Variância	4	5,4760	4,5142

Também ocorreram problemas numéricos para este modelo, representando, no máximo 0,7% das amostras de Monte Carlo geradas. Assim como na situação anterior, esse percentual é relativamente pequeno.

6 Conclusões

Ao longo deste trabalho apresentamos três propostas de melhoramentos do teste da razão de verossimilhanças cno contexto do modelo linear misto geral, como forma de melhorarmos o desempenho do teste da razão de verossimilhanças em pequenas amostras. Comentamos algumas vezes a utilização do modelo linear misto no contexto de dados longitudinais, mas é importante ficar claro que os ajustes encontrados neste trabalho referem-se a classe geral de modelos lineares mistos, podendo ser aplicados em outros contextos. Fazemos um apelo maior aos estudos longitudinais, pois em muitas situações são estudos caros e/ou envolvem um pequeno número de indivíduos observados.

Nos resultados apresentados, em geral, vimos que a aproximação do ajuste de Barndorff-Nielsen (1983) proposta por Severini (1998) foi bastante satisfatória quanto aos testes de hipóteses. Vale ressaltar que este desempenho satisfatório representa um grande ganho se comparado aos resultados obtidos por Zucker, Lieberman e Manor (2000). Pois os autores avaliaram o desempenho de outra proposta de melhoramento de teste, baseado na razão de verossimilhanças perfiladas modificadas proposta por Cox e Reid (1987), e não obtiveram resultados tão satisfatórios, tendo que aplicar a correção de Bartlett na estatística baseada no ajuste.

No modelo com o intercepto aleatório além da comparação do teste usual com o baseado no ajuste da verossimilhança perfilada, pudemos comparar o desempenho do teste *bootstrap* e do teste da razão de verossimilhanças usuais corrigido via Bartlett. Os testes com melhores desempenhos foram aqueles baseado na função de verossimilhança perfilada modificada e na correção de Bartlett. Uma vantagem de se utilizar o teste baseado no ajuste da verossimilhança perfilada se deve ao fato de não haver expressão analítica para a estatística da razão de verossimilhanças usual e portanto não é possível encontrar analiticamente o termo de ajuste da correção de Bartlett. A alternativa que utilizamos foi de encontrar o termo que faz parte do ajuste da correção, a esperança da estatística da razão de verossimilhanças, através do método *bootstrap*. E isso faz com que o tempo computacional para a análise dos dados seja multiplicado pelo número de amostras *bootstrap* que se utiliza.

Ainda no modelo com o intercepto aleatório, em apenas um caso, a taxa de rejeição do teste *bootstrap* esteve mais próxima do nível nominal se comparada a taxa do teste usual corrigido via

Bartlett. Este caso foi utilizando-se 500 reamostras. Isto nos dá indícios que para o teste *bootstrap* ter taxas de rejeição próximas as do teste corrigido via Bartlett seja necessário uma quantidade maior de reamostras. Ou seja, que para construir a distribuição amostral da estatística da razão de verossimilhanças usual seja necessária uma quantidade maior de reamostras do que a quantidade utilizada para se encontrar o termo da correção.

Um aspecto de grande relevância é que os cálculos envolvidos na obtenção da aproximação proposta por Severini (1998), no caso em que os dados seguem um modelo misto linear geral, são relativamente simples. O maior problema computacional é a necessidade de utilizarmos métodos de maximização numérica, o que pode ocasionar em alguns erros do algoritmo escolhido.

Como sugestões de pesquisas futuras, citamos: avaliar outras propostas de ajuste da função de verossimilhança perfilada, realizar testes sobre os parâmetros da estrutura de covariâncias e utilizar alguma transformação nas variáveis afim de se obter expressão analítica para os estimadores e assim reduzir o tempo computacional.

Apêndice A - Cálculos para os Ajustes

Derivadas do Modelo Misto Linear Geral

As primeiras derivadas do logaritmo da função de verossimilhança dada por (2.2) são

$$\frac{\partial \ell(\theta; \mathbf{Y})}{\partial \alpha_i} = -\frac{1}{2} \text{tr} \left(\Sigma^{-1} \dot{\Sigma}_i \right) + \frac{1}{2} (\mathbf{Y} - \mathbf{X}\beta)^\top \Sigma^{-1} \dot{\Sigma}_i \Sigma^{-1} (\mathbf{Y} - \mathbf{X}\beta),$$

$$\frac{\partial \ell(\theta; \mathbf{Y})}{\partial \beta} = \mathbf{X}^\top \Sigma^{-1} (\mathbf{Y} - \mathbf{X}\beta).$$

As segundas derivadas do logaritmo da função de verossimilhança dada por (2.2) são

$$\frac{\partial^2 \ell(\theta; \mathbf{Y})}{\partial \alpha_i \partial \alpha_j} = \frac{1}{2} \text{tr} \left(\Sigma^{-1} \dot{\Sigma}_i \Sigma^{-1} \dot{\Sigma}_j \right) - \frac{1}{2} (\mathbf{Y} - \mathbf{X}\beta)^\top \Sigma^{-1} \dot{\Sigma}_i \Sigma^{-1} \dot{\Sigma}_j \Sigma^{-1} (\mathbf{Y} - \mathbf{X}\beta),$$

$$\frac{\partial^2 \ell(\theta; \mathbf{Y})}{\partial \alpha_i \partial \beta} = \mathbf{X}^\top \Sigma^{-1} \dot{\Sigma}_i \Sigma^{-1} (\mathbf{Y} - \mathbf{X}\beta) = \frac{\partial^2 \ell(\theta; \mathbf{Y})}{\partial \beta \partial \alpha_i},$$

$$\frac{\partial^2 \ell(\theta; \mathbf{Y})}{\partial \beta^2} = \mathbf{X}^\top \Sigma^{-1} \mathbf{X}.$$

Apêndice B - Gráficos de Discrepâncias Relativas de Quantis

Modelo com Intercepto Aleatório

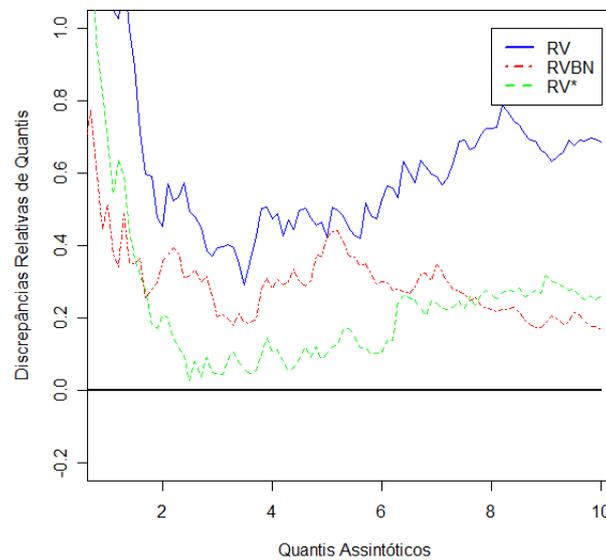


Figura 6.1: Gráfico das discrepâncias relativas de quantis para $N = 5$ e $\rho = 0,9$ ($H_0 : \beta_3 = 0$), no modelo com intercepto aleatório.

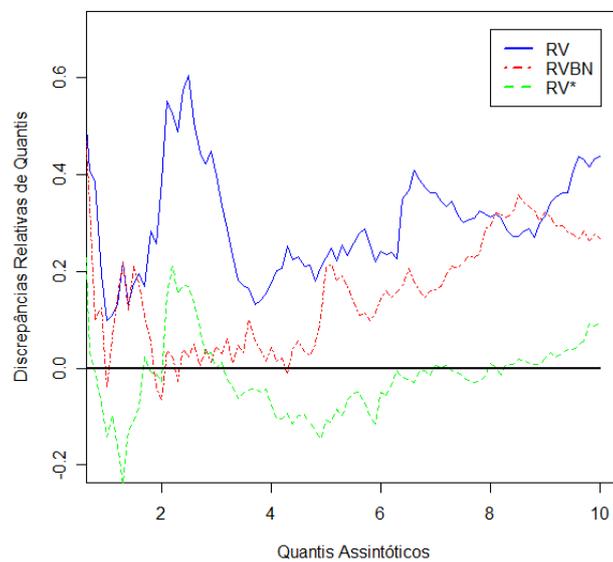


Figura 6.2: Gráfico das discrepâncias relativas de quantis para $N = 5$ e $\rho = 0,95$ ($H_0 : \beta_3 = 0$), no modelo com intercepto aleatório.

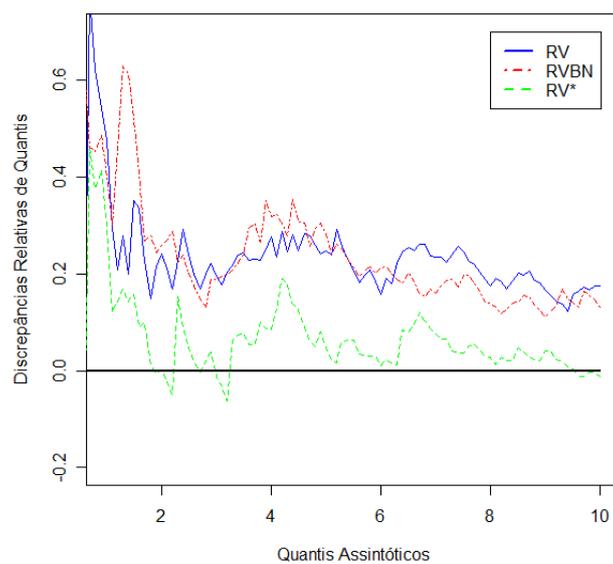


Figura 6.3: Gráfico das discrepâncias relativas de quantis para $N = 10$ e $\rho = 0,9$ ($H_0 : \beta_3 = 0$), no modelo com intercepto aleatório.

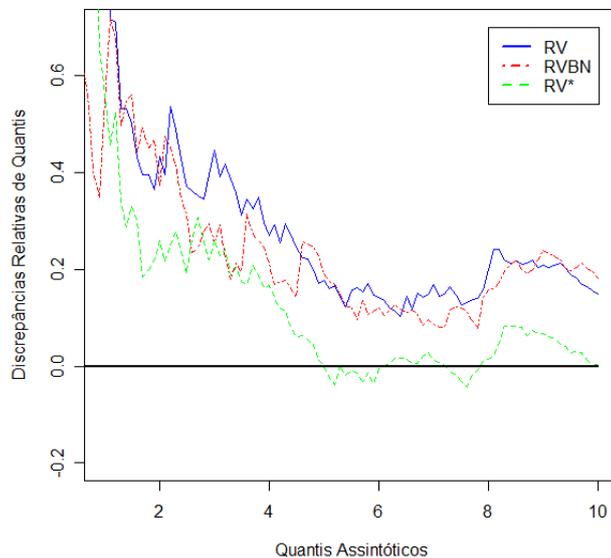


Figura 6.4: Gráfico das discrepâncias relativas de quantis para $N = 10$ e $\rho = 0,95$ ($H_0 : \beta_3 = 0$), no modelo com intercepto aleatório.

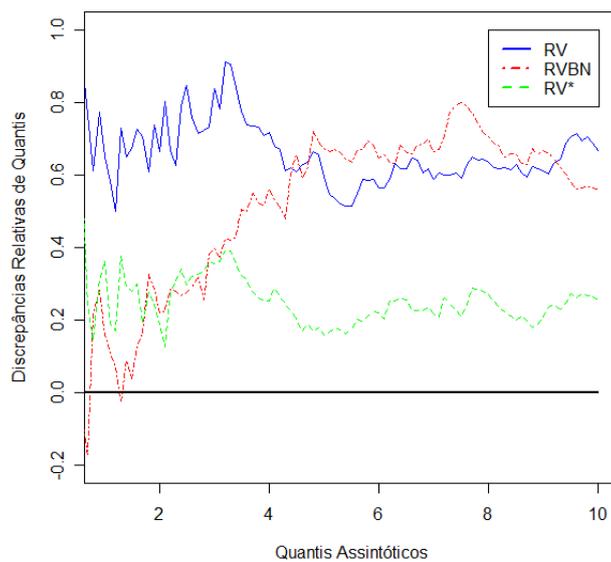


Figura 6.5: Gráfico das discrepâncias relativas de quantis para $N = 5$, $\rho = 0,95$ e $\tau^2 = 10$ ($H_0 : \beta_3 = 0$), no modelo com intercepto aleatório.

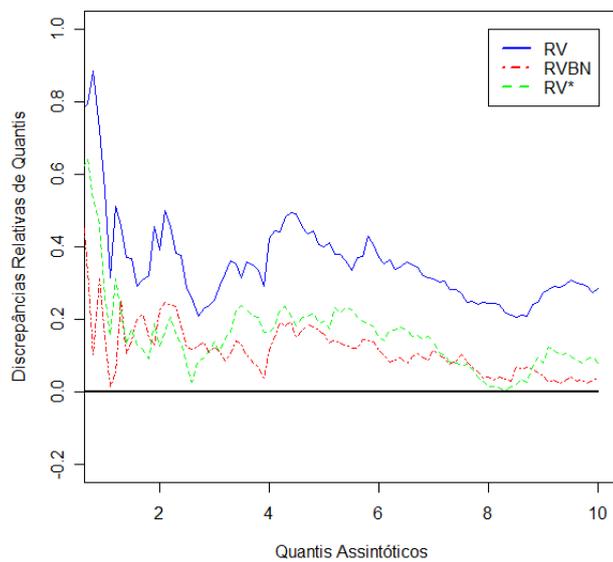


Figura 6.6: Gráfico das discrepâncias relativas de quantis para $N = 10$, $\rho = 0,5$ ($H_0 : \beta_3 = 0$), no modelo com intercepto aleatório.

Modelo com Intercepto e Inclinação Aleatórios

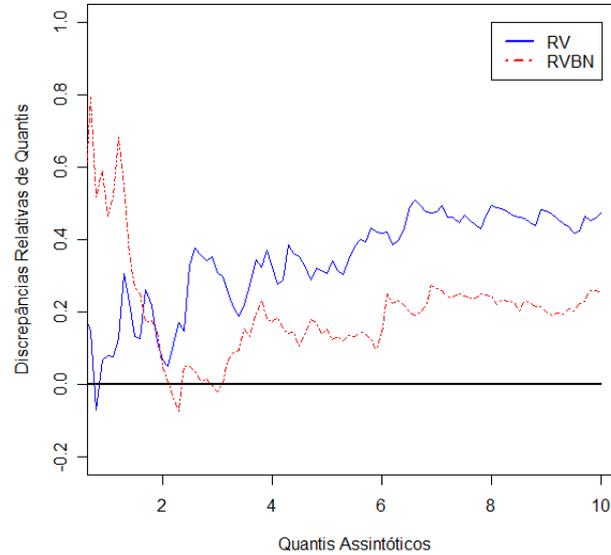


Figura 6.7: Gráfico das discrepâncias relativas de quantis para $N = 15$, $\alpha_2 = 0$ e $\alpha_3 = 1$ ($H_0 : \beta_3 = 0$), no modelo com intercepto e inclinação aleatórios.

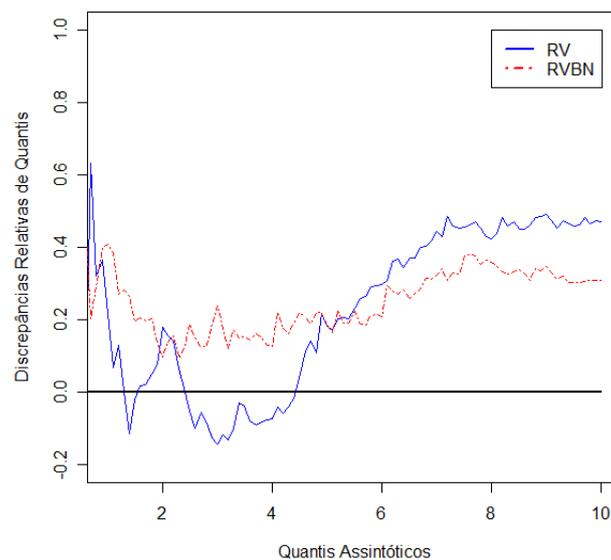


Figura 6.8: Gráfico das discrepâncias relativas de quantis para $N = 15$, $\alpha_2 = 0,35$ e $\alpha_3 = 0,5$ ($H_0 : \beta_3 = 0$), no modelo com intercepto e inclinação aleatórios.

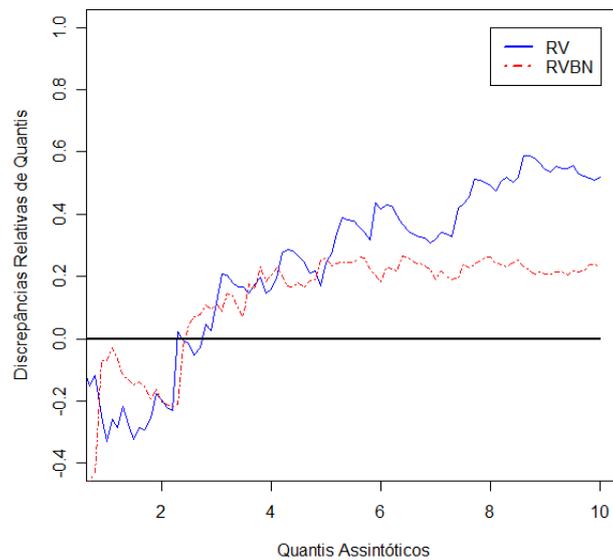


Figura 6.9: Gráfico das discrepâncias relativas de quantis para $N = 15$, $\alpha_2 = 0,35$ e $\alpha_3 = 1$ ($H_0 : \beta_3 = 0$), no modelo com intercepto e inclinação aleatórios.

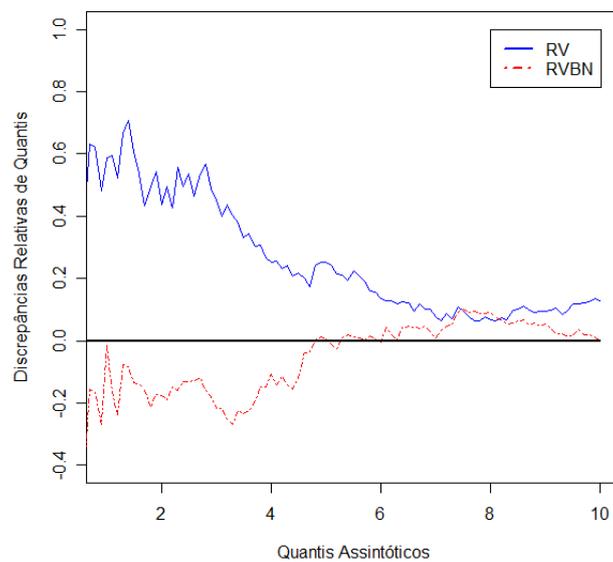


Figura 6.10: Gráfico das discrepâncias relativas de quantis para $N = 20$, $\alpha_2 = 0$ e $\alpha_3 = 1$ ($H_0 : \beta_3 = 0$), no modelo com intercepto e inclinação aleatórios.

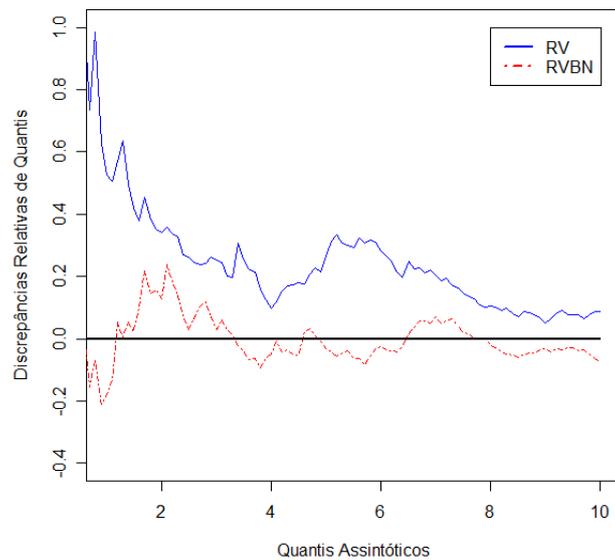


Figura 6.11: Gráfico das discrepâncias relativas de quantis para $N = 20$, $\alpha_2 = 0,35$ e $\alpha_3 = 0,5$ ($H_0 : \beta_3 = 0$), no modelo com intercepto e inclinação aleatórios.

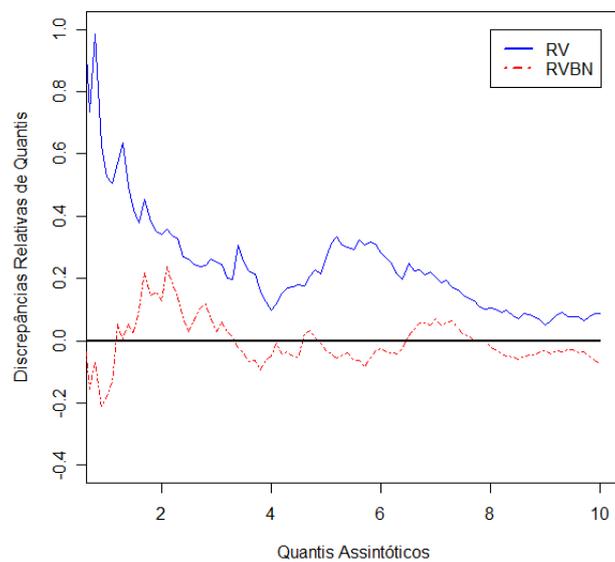


Figura 6.12: Gráfico das discrepâncias relativas de quantis para $N = 20$, $\alpha_2 = 0,35$ e $\alpha_3 = 1$ ($H_0 : \beta_3 = 0$), no modelo com intercepto e inclinação aleatórios.

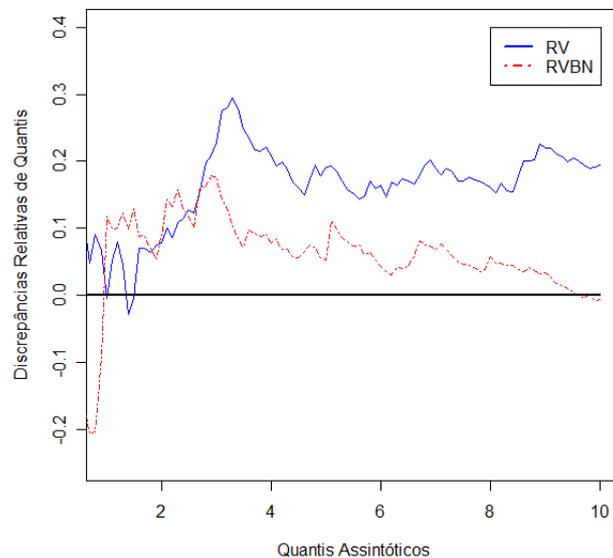


Figura 6.13: Gráfico das discrepâncias relativas de quantis para $N = 15$, $\alpha_2 = 0$ e $\alpha_3 = 1$ ($H_0 : \beta_3 = 0$ e $\beta_2 = 0$), no modelo com intercepto e inclinação aleatórios.

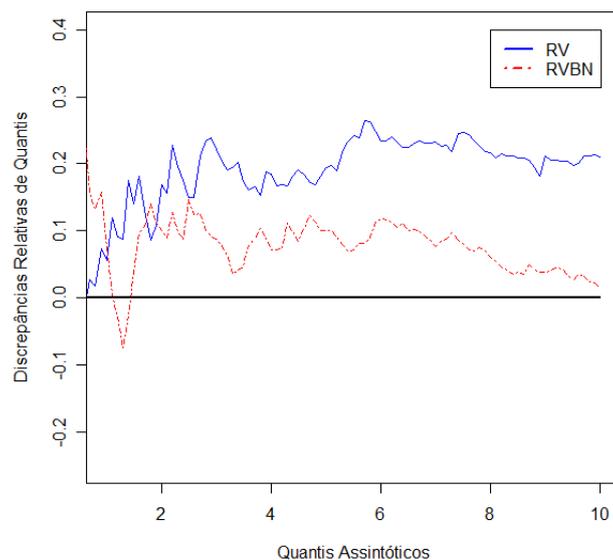


Figura 6.14: Gráfico das discrepâncias relativas de quantis para $N = 15$, $\alpha_2 = 0,35$ e $\alpha_3 = 0,5$ ($H_0 : \beta_3 = 0$ e $\beta_2 = 0$), no modelo com intercepto e inclinação aleatórios.

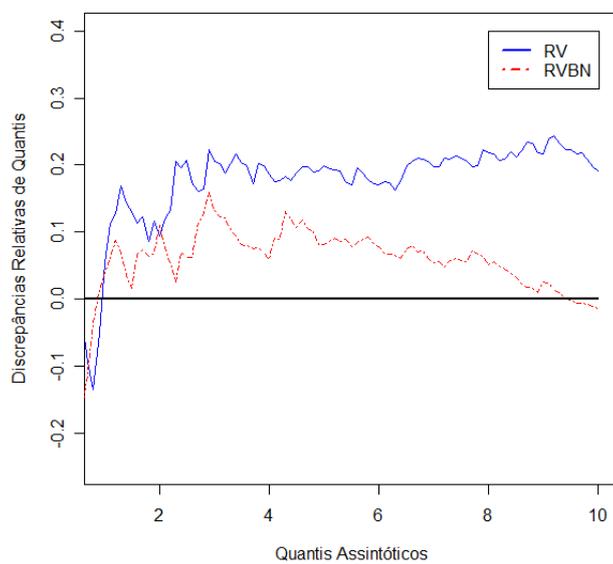


Figura 6.15: Gráfico das discrepâncias relativas de quantis para $N = 15$, $\alpha_2 = 0,35$ e $\alpha_3 = 1$ ($H_0 : \beta_3 = 0$ e $\beta_2 = 0$), no modelo com intercepto e inclinação aleatórios.

Apêndice C - Programa Computacional

```

/*#####
Programa: Exemplo3.ox
Versão: 3-7
Modelo misto ->  $Y_i = \beta_0 + \beta_1 t_i + \beta_2 * x_{2i} + \beta_3 * x_{3i} + b_{0i} + b_{1i} * t_i + e_i$ 

Autor: Markus Chagas Stein

Data: 18 de setembro de 2008
#####*/

#include <oxstd.h>
#include <oxfloat.h>
#include <maximize.h>
#include <oxprob.h>
#pragma link("maximize.oxo")
#import <maximize>
#import <maxsqp>

//#####DECLARACOES DAS VARIAVEIS GLOBAIS

static decl g_vY;           //vetor de todas as observacoes
static decl g_mX;           //matriz de delineamento do EFEITOS FIXOS
static decl g_mZ;           //matriz de delineamento do EFEITOS ALEATORIOS
static decl g_vn;           //vetor do numero de observacoes para cada individuo
static decl g_vw;           //vetor de parâmetros da matriz de covariâncias
static decl g_vbeta;        //vetor de parâmetros beta
static decl g_mSigma;       //matriz de covariância de todas as observações
static decl g_vEMV;         //EMV do vetor de parâmetros
static decl g_vEMVO;        //EMV do vetor de parâmetros sob H0
static decl g_vEMVB;        //EMV do vetor de parâmetros BOOTSTRAP
static decl g_inpar;        //Número de parametros do modelo
static decl g_inpartest;    //Número de parametros testados
static decl g_vParNTes;     // vetor indicador de parametros nao-testados
static decl g_inparNAOtest; //Número de parametros NAO testados
static decl g_iCovarv;      // numero de covariaveis no modelo
static decl g_iLogLik;      // Valor da funcao de log-verossimilhança
//decl file1()  ##

//#####Função para criar V0 - Matriz de correlações das observações do individuo i
Zi(const i){
    decl v0;
    v0 = zeros(g_vn[i],2);
    if(i == 0){
        v0 = g_mX[0:g_vn[0]-1][0:1];
    }
    else{

```

```

        v0 = g_mX[sumr(g_vn[0:i-1]) : sumr(g_vn[0:i-1])+g_vn[i]-1][0:1];
    }
    return(v0);
} //FIM Zi

#####Função para criar P - Matriz bloco-diagonal de correlações de todas as observações
Z(const n){
    decl v0,p,k;
    v0 = zeros(n[0],2);
    v0 = Zi(0);
    p = v0;
    for(k=1;k<columns(n);k++){
        v0 = zeros(n[k],2);
        v0 = Zi(k);
        p = diagcat(p,v0);
    }
    return(p);
} //FIM Z

#####Função para criar a derivada de G
dGi(const i){
    decl v0,G;
    v0 = zeros(2,2);

    if(i == 1) v0 = <1,0;0,0>;
    if(i == 2) v0 = <0,1;1,0>;
    if(i == 3) v0 = <0,0;0,1>;

    return(v0);
} //FIM dGi

#####Função para criar G - Matriz de covariâncias dos efeitos aleatorios
G(const w){
    decl v0, G;

    G = w[0] * dGi(1) + w[1] * dGi(2) + w[2] * dGi(3);

    return(G);
} //FIM G

#####Função para criar Sigma - Matriz de covariâncias de Y
Sigma(const w){
    decl v0, G;
    v0 = zeros(sumr(g_vn),sumr(g_vn));

    G = w[0] * dGi(1) + w[1] * dGi(2) + w[2] * dGi(3);

    v0 = g_mZ * (unit(columns(g_vn),columns(g_vn))*G) * g_mZ' + w[3] * unit(sumr(g_vn),sumr(g_vn));

    return(v0);
} //FIM Sigma

#####Função para criar a derivada de Sigma
dSigma(const i){

```

```

decl v0,p,k;
v0 = zeros(sumr(g_vn),sumr(g_vn));

if(i == 4){
    v0 = unit(sumr(g_vn),sumr(g_vn));
}
else{
    v0 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dGi(i)) * g_mZ';
}
return(v0);
} //FIM Sigma

//####Funcao de log-Verossimilhanca da dist. Normal(BETA, SIGMA) dividida pelo tamanho da amostra

//***** IRRESTRITA *****
/*vP[0:3] é o vetor w = (w1, w2, w3, w4)
vP[4:7] é o vetor beta*/

fLoglik(const vP, const adFunc, const avScore, const amHess){
    decl vw,result,SIGMA,dG1,dG2,dG3,dSigma1,dSigma2,dSigma3,dSigma4, isign,vai;

    //println("antes",vP);

    dG1 = <1,0;0,0>;
    dG2 = <0,1;1,0>;
    dG3 = <0,0;0,1>;

    dSigma4 = unit(sumr(g_vn),sumr(g_vn));
    dSigma1 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG1) * g_mZ';
    dSigma2 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG2) * g_mZ';
    dSigma3 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG3) * g_mZ';

    vw = zeros(columns(g_vw),1);
    result = zeros(g_inpar,1);

    vw[0] = exp(vP[0]);
    vw[1] = vP[1];
    vw[2] = exp(vP[2]);
    vw[3] = exp(vP[3]);

    SIGMA = g_mZ * (unit(columns(g_vn),columns(g_vn))*(vw[0:1]' | vw[1:2]')) * g_mZ'
        + vw[3] * unit(sumr(g_vn),sumr(g_vn));

    vai = choleski(SIGMA);

    adFunc[0] = - (1/2) * log(2 * M_PI)
        - (1/(2*sumr(g_vn))) * logdet(vai*vai', &isign)
        - (1/(2*sumr(g_vn))) * (g_vY - g_mX * vP[4:g_inpar-1])'
        * invertgen(vai*vai') * (g_vY - g_mX * vP[4:g_inpar-1]);

    if (isnan(adFunc[0]) || isdotinf(adFunc[0]) || isign <= 0 || isign == 2)

```

```

return 0;

if(avScore){
  (avScore[0])[0] = ( - (1/(2*sumr(g_vn))) * trace(invertgen(SIGMA) * dSigma1)
    + (1/(2*sumr(g_vn))) * (g_vY - g_mX * vP[4:g_inpar-1]))'
    * invertgen(SIGMA) * dSigma1 * invertgen(SIGMA)
    * (g_vY - g_mX * vP[4:g_inpar-1]) * exp(vP[0]);

  (avScore[0])[1] = - (1/(2*sumr(g_vn))) * trace(invertgen(SIGMA) * dSigma2)
    + (1/(2*sumr(g_vn))) * (g_vY - g_mX * vP[4:g_inpar-1]))'
    * invertgen(SIGMA) * dSigma2 * invertgen(SIGMA)
    * (g_vY - g_mX * vP[4:g_inpar-1]);

  (avScore[0])[2] = ( - (1/(2*sumr(g_vn))) * trace(invertgen(SIGMA) * dSigma3)
    + (1/(2*sumr(g_vn))) * (g_vY - g_mX * vP[4:g_inpar-1]))'
    * invertgen(SIGMA) * dSigma3 * invertgen(SIGMA)
    * (g_vY - g_mX * vP[4:g_inpar-1]) * exp(vP[2]);

  (avScore[0])[3] = ( - (1/(2*sumr(g_vn))) * trace(invertgen(SIGMA) * dSigma4)
    + (1/(2*sumr(g_vn))) * (g_vY - g_mX * vP[4:g_inpar-1]))'
    * invertgen(SIGMA) * dSigma4 * invertgen(SIGMA)
    * (g_vY - g_mX * vP[4:g_inpar-1]) * exp(vP[3]);

  (avScore[0])[4:g_inpar-1] = (1/sumr(g_vn)) * g_mX' * invertgen(SIGMA)
    * (g_vY - g_mX * vP[4:g_inpar-1]);
}

result[0:3] = vw;
result[4:g_inpar-1] = vP[4:g_inpar-1];

g_vEMV = result;

return 1;
} //FIM fLoglik

//***** SOB A HIPOTESE NULA *****
/*vP[0] é o parâmetro tal2
  vP[1] é o parâmetro ro*/

fLoglik_Re(const vP, const adFunc, const avScore, const amHess){

  decl vPar_Re, vai;
  vPar_Re = vP[4:g_inparNA0test-1][0] | zeros(g_inpartest, 1);

  decl vw,result,SIGMA,dG1,dG2,dG3,dSigma1,dSigma2,dSigma3,dSigma4, isign;

  dG1 = <1,0;0,0>;
  dG2 = <0,1;1,0>;
  dG3 = <0,0;0,1>;

  dSigma4 = unit(sumr(g_vn),sumr(g_vn));

```

```

dSigma1 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG1) * g_mZ';
dSigma2 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG2) * g_mZ';
dSigma3 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG3) * g_mZ';

vw = zeros(columns(g_vw),1);
result = zeros(g_inpar,1);

vw[0] = exp(vP[0]);
vw[1] = vP[1];
vw[2] = exp(vP[2]);
vw[3] = exp(vP[3]);

SIGMA = g_mZ * (unit(columns(g_vn),columns(g_vn))*(vw[0:1]' | vw[1:2]')) * g_mZ'
        + vw[3] * unit(sumr(g_vn),sumr(g_vn));

vai = choleski(SIGMA);

adFunc[0] = - (1/2) * log(2 * M_PI)
            - (1/(2*sumr(g_vn))) * logdet(vai*vai', &isign)
            - (1/(2*sumr(g_vn))) * (g_vY - g_mX * vPar_Re)'
            * invertgen(vai*vai') * (g_vY - g_mX * vPar_Re);

if (isnan(adFunc[0]) || isdotinf(adFunc[0]) || isign <= 0 || isign == 2)
    return 0;

if(avScore){
    (avScore[0])[0] = ( - (1/(2*sumr(g_vn))) * trace(invertgen(SIGMA) * dSigma1)
                    + (1/(2*sumr(g_vn))) * (g_vY - g_mX * vPar_Re)'
                    * invertgen(SIGMA) * dSigma1 * invertgen(SIGMA)
                    * (g_vY - g_mX * vPar_Re) ) * exp(vP[0]);

    (avScore[0])[1] = - (1/(2*sumr(g_vn))) * trace(invertgen(SIGMA) * dSigma2)
                    + (1/(2*sumr(g_vn))) * (g_vY - g_mX * vPar_Re)'
                    * invertgen(SIGMA) * dSigma2 * invertgen(SIGMA)
                    * (g_vY - g_mX * vPar_Re);

    (avScore[0])[2] = ( - (1/(2*sumr(g_vn))) * trace(invertgen(SIGMA) * dSigma3)
                    + (1/(2*sumr(g_vn))) * (g_vY - g_mX * vPar_Re)'
                    * invertgen(SIGMA) * dSigma3 * invertgen(SIGMA)
                    * (g_vY - g_mX * vPar_Re) ) * exp(vP[2]);

    (avScore[0])[3] = ( - (1/(2*sumr(g_vn))) * trace(invertgen(SIGMA) * dSigma4)
                    + (1/(2*sumr(g_vn))) * (g_vY - g_mX * vPar_Re)'
                    * invertgen(SIGMA) * dSigma4 * invertgen(SIGMA)
                    * (g_vY - g_mX * vPar_Re) ) * exp(vP[3]);

    (avScore[0])[4:g_inparNA0test-1] = (1/sumr(g_vn)) * g_mX' * invertgen(SIGMA)
                                        * (g_vY - g_mX * vPar_Re);
}

result[0:3] = vw;
result[4:g_inpar-1] = vPar_Re;

```

```

    g_vEMV0 = result;

    return 1;
} //FIM fLoglik

#####Definicao das restricoes das derivadas da log-verossimilhanca iguais a 0

//***** IRRESTRITA *****
//      -> (avF[0])[0] é a restricao para derivada de tal2
//      -> (avF[0])[1] é a restricao para derivada de ro

fRest(const avF, const vP){
    decl SIGMA,dG1,dG2,dG3,dSigma1,dSigma2,dSigma3,dSigma4;
    decl c = new matrix[g_inparNA0test][1];

    dG1 = <1,0;0,0>;
    dG2 = <0,1;1,0>;
    dG3 = <0,0;0,1>;

    dSigma4 = unit(sumr(g_vn),sumr(g_vn));
    dSigma1 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG1) * g_mZ';
    dSigma2 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG2) * g_mZ';
    dSigma3 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG3) * g_mZ';

    SIGMA = g_mZ * (unit(columns(g_vn),columns(g_vn))*(vP[0:1]' | vP[1:2]')) * g_mZ'
            + vP[3] * unit(sumr(g_vn),sumr(g_vn));

    c[0] = - (1/2) * trace(invertgen(SIGMA) * dSigma1) + (1/2) * (g_vY - g_mX * vP[4:g_inpar-1])'
            * invertgen(SIGMA) * dSigma1 * invertgen(SIGMA) * (g_vY - g_mX * vP[4:g_inpar-1]);

    c[1] = - (1/2) * trace(invertgen(SIGMA) * dSigma2) + (1/2) * (g_vY - g_mX * vP[4:g_inpar-1])'
            * invertgen(SIGMA) * dSigma2 * invertgen(SIGMA) * (g_vY - g_mX * vP[4:g_inpar-1]);

    c[2] = - (1/2) * trace(invertgen(SIGMA) * dSigma3) + (1/2) * (g_vY - g_mX * vP[4:g_inpar-1])'
            * invertgen(SIGMA) * dSigma3 * invertgen(SIGMA) * (g_vY - g_mX * vP[4:g_inpar-1]);

    c[3] = - (1/2) * trace(invertgen(SIGMA) * dSigma4) + (1/2) * (g_vY - g_mX * vP[4:g_inpar-1])'
            * invertgen(SIGMA) * dSigma4 * invertgen(SIGMA) * (g_vY - g_mX * vP[4:g_inpar-1]);

    c[4:g_inparNA0test-1]=selectifr((g_mX' * invertgen(SIGMA) * (g_vY - g_mX * vP[4:g_inpar-1])),
            g_vParNTes[4:g_inpar-1]');

    avF[0] = c;
    return 1;
} //FIM fRest

//***** SOB A HIPOTESE NULA *****
fRest_Re(const avF, const vP){
    decl SIGMA,dG1,dG2,dG3,dSigma1,dSigma2,dSigma3,dSigma4;
    decl c = new matrix[g_inparNA0test][1];

```

```

dG1 = <1,0;0,0>;
dG2 = <0,1;1,0>;
dG3 = <0,0;0,1>;

dSigma4 = unit(sumr(g_vn),sumr(g_vn));
dSigma1 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG1) * g_mZ';
dSigma2 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG2) * g_mZ';
dSigma3 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG3) * g_mZ';

SIGMA = g_mZ * (unit(columns(g_vn),columns(g_vn))*(vP[0:1]' | vP[1:2]')) * g_mZ'
      + vP[3] * unit(sumr(g_vn),sumr(g_vn));

decl vPar_Re;
vPar_Re = vP[4:g_inparNA0test-1][0] | zeros(g_inpartest, 1);

c[0] = - (1/2) * trace(invertgen(SIGMA) * dSigma1) + (1/2) * (g_vY - g_mX * vPar_Re)'
      * invertgen(SIGMA) * dSigma1 * invertgen(SIGMA) * (g_vY - g_mX * vPar_Re);

c[1] = - (1/2) * trace(invertgen(SIGMA) * dSigma2) + (1/2) * (g_vY - g_mX * vPar_Re)'
      * invertgen(SIGMA) * dSigma2 * invertgen(SIGMA) * (g_vY - g_mX * vPar_Re);

c[2] = - (1/2) * trace(invertgen(SIGMA) * dSigma3) + (1/2) * (g_vY - g_mX * vPar_Re)'
      * invertgen(SIGMA) * dSigma3 * invertgen(SIGMA) * (g_vY - g_mX * vPar_Re);

c[3] = - (1/2) * trace(invertgen(SIGMA) * dSigma4) + (1/2) * (g_vY - g_mX * vPar_Re)'
      * invertgen(SIGMA) * dSigma4 * invertgen(SIGMA) * (g_vY - g_mX * vPar_Re);

c[4:g_inparNA0test-1]=selectifr((g_mX' * invertgen(SIGMA) * (g_vY - g_mX * vPar_Re)),
      g_vParNTes[4:g_inpar-1]');

avF[0] = c;
return 1;
} //FIM fRest_Re

/*/####Funcao de log-Verossimilhanca PERFILADA ajustada baseada em covariancias de derivadas da log-verossimilhanca
da dist. Normal(BETA0, SIGMA) dividida pelo tamanho da amostra*/

//***** IRRESTRITA *****
/*vP[0] é o parâmetro tal2
vP[1] é o parâmetro ro
vP[2] é o parâmetro beta0*/

fLoglik_BN_COVDER(const vP, const adFunc, const avScore, const amHess){
decl SIGMA,SIGMAEMV,invSigma,dG1,dG2,dG3,dSigma1,dSigma2,dSigma3,dSigma4,ismgn,ismgn2,ismgn3,vai;

dG1 = <1,0;0,0>;
dG2 = <0,1;1,0>;
dG3 = <0,0;0,1>;

dSigma4 = unit(sumr(g_vn),sumr(g_vn));
dSigma1 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG1) * g_mZ';

```

```

dSigma2 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG2) * g_mZ';
dSigma3 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG3) * g_mZ';

SIGMA = g_mZ * (unit(columns(g_vn),columns(g_vn))*(vP[0:1]' | vP[1:2]')) * g_mZ'
        + vP[3] * unit(sumr(g_vn),sumr(g_vn));

vai = choleski(SIGMA);

invSigma = zeros(sumr(g_vn),sumr(g_vn));

invSigma = invertgen(vai*vai');

//INFORMACAO OBSERVADA
decl j = new matrix[g_inparNA0test][g_inparNA0test];

j[0][0] = (-1/2) * trace(invSigma * dSigma1 * invSigma * dSigma1)
        + (g_vY - g_mX * vP[4:g_inpar-1])' * invSigma * dSigma1 * invSigma
        * dSigma1 * invSigma * (g_vY - g_mX * vP[4:g_inpar-1]);
j[1][1] = (-1/2) * trace(invSigma * dSigma2 * invSigma * dSigma2)
        + (g_vY - g_mX * vP[4:g_inpar-1])' * invSigma * dSigma2 * invSigma
        * dSigma2 * invSigma * (g_vY - g_mX * vP[4:g_inpar-1]);
j[2][2] = (-1/2) * trace(invSigma * dSigma3 * invSigma * dSigma3)
        + (g_vY - g_mX * vP[4:g_inpar-1])' * invSigma * dSigma3 * invSigma
        * dSigma3 * invSigma * (g_vY - g_mX * vP[4:g_inpar-1]);
j[3][3] = (-1/2) * trace(invSigma * dSigma4 * invSigma * dSigma4)
        + (g_vY - g_mX * vP[4:g_inpar-1])' * invSigma * dSigma4 * invSigma
        * dSigma4 * invSigma * (g_vY - g_mX * vP[4:g_inpar-1]);

j[0][1] = j[1][0] = (-1/2) * trace(invSigma * dSigma1 * invSigma * dSigma2)
        + (g_vY - g_mX * vP[4:g_inpar-1])' * invSigma * dSigma1 * invSigma
        * dSigma2 * invSigma * (g_vY - g_mX * vP[4:g_inpar-1]);
j[0][2] = j[2][0] = (-1/2) * trace(invSigma * dSigma1 * invSigma * dSigma3)
        + (g_vY - g_mX * vP[4:g_inpar-1])' * invSigma * dSigma1 * invSigma
        * dSigma3 * invSigma * (g_vY - g_mX * vP[4:g_inpar-1]);
j[0][3] = j[3][0] = (-1/2) * trace(invSigma * dSigma1 * invSigma * dSigma4)
        + (g_vY - g_mX * vP[4:g_inpar-1])' * invSigma * dSigma1 * invSigma
        * dSigma4 * invSigma * (g_vY - g_mX * vP[4:g_inpar-1]);

j[1][2] = j[2][1] = (-1/2) * trace(invSigma * dSigma2 * invSigma * dSigma3)
        + (g_vY - g_mX * vP[4:g_inpar-1])' * invSigma * dSigma2 * invSigma
        * dSigma3 * invSigma * (g_vY - g_mX * vP[4:g_inpar-1]);
j[1][3] = j[3][1] = (-1/2) * trace(invSigma * dSigma2 * invSigma * dSigma4)
        + (g_vY - g_mX * vP[4:g_inpar-1])' * invSigma * dSigma2 * invSigma
        * dSigma4 * invSigma * (g_vY - g_mX * vP[4:g_inpar-1]);

j[2][3] = j[3][2] = (-1/2) * trace(invSigma * dSigma3 * invSigma * dSigma4)
        + (g_vY - g_mX * vP[4:g_inpar-1])' * invSigma * dSigma3 * invSigma
        * dSigma4 * invSigma * (g_vY - g_mX * vP[4:g_inpar-1]);

j[4:g_inparNA0test-1][0] = g_mX' * invSigma * dSigma1 * invSigma * (g_vY - g_mX * vP[4:g_inpar-1]);
j[0][4:g_inparNA0test-1] = j[4:g_inparNA0test-1][0]';

```

```

j[4:g_inparNAOtest-1][1] = g_mX' * invSigma * dSigma2 * invSigma * (g_vY - g_mX * vP[4:g_inpar-1]);
j[1][4:g_inparNAOtest-1] = j[4:g_inparNAOtest-1][1]';

j[4:g_inparNAOtest-1][2] = g_mX' * invSigma * dSigma3 * invSigma * (g_vY - g_mX * vP[4:g_inpar-1]);
j[2][4:g_inparNAOtest-1] = j[4:g_inparNAOtest-1][2]';

j[4:g_inparNAOtest-1][3] = g_mX' * invSigma * dSigma4 * invSigma * (g_vY - g_mX * vP[4:g_inpar-1]);
j[3][4:g_inparNAOtest-1] = j[4:g_inparNAOtest-1][3]';

j[4:g_inparNAOtest-1][4:g_inparNAOtest-1] = g_mX' * invSigma * g_mX;

//COVARIANCIAS EMPIRICAS
decl i = new matrix[g_inparNAOtest][g_inparNAOtest];

i[0][0] = (1/2) * trace(invSigma * dSigma1 * invSigma * dSigma1);
i[1][1] = (1/2) * trace(invSigma * dSigma2 * invSigma * dSigma2);
i[2][2] = (1/2) * trace(invSigma * dSigma3 * invSigma * dSigma3);
i[3][3] = (1/2) * trace(invSigma * dSigma4 * invSigma * dSigma4);

i[0][1] = i[1][0] = (1/2) * trace(invSigma * dSigma1 * invSigma * dSigma2);
i[0][2] = i[2][0] = (1/2) * trace(invSigma * dSigma1 * invSigma * dSigma3);
i[0][3] = i[3][0] = (1/2) * trace(invSigma * dSigma1 * invSigma * dSigma4);

i[1][2] = i[2][1] = (1/2) * trace(invSigma * dSigma2 * invSigma * dSigma3);
i[1][3] = i[3][1] = (1/2) * trace(invSigma * dSigma2 * invSigma * dSigma4);

i[2][3] = i[3][2] = (1/2) * trace(invSigma * dSigma3 * invSigma * dSigma4);

i[0][4:g_inparNAOtest-1] = (g_mX * g_vEMV[4:g_inpar-1] - g_mX * vP[4:g_inpar-1])' * invSigma
    * dSigma1 * invSigma * g_mX;
i[1][4:g_inparNAOtest-1] = (g_mX * g_vEMV[4:g_inpar-1] - g_mX * vP[4:g_inpar-1])' * invSigma
    * dSigma2 * invSigma * g_mX;
i[2][4:g_inparNAOtest-1] = (g_mX * g_vEMV[4:g_inpar-1] - g_mX * vP[4:g_inpar-1])' * invSigma
    * dSigma3 * invSigma * g_mX;
i[3][4:g_inparNAOtest-1] = (g_mX * g_vEMV[4:g_inpar-1] - g_mX * vP[4:g_inpar-1])' * invSigma
    * dSigma4 * invSigma * g_mX;

i[4:g_inparNAOtest-1][4:g_inparNAOtest-1] = g_mX' * invSigma * g_mX;

adFunc[0] = - (1/2) * log(2 * M_PI) - (1/(2*sumr(g_vn))) * logdet(vai*vai', &isign)
    - (1/(2*sumr(g_vn))) * (g_vY - g_mX * vP[4:g_inpar-1])' * invertgen(vai*vai')
    * (g_vY - g_mX * vP[4:g_inpar-1]) + (1/(2 * sumr(g_vn))) * logdet(j,&isign2)
    - (1/sumr(g_vn)) * logdet(i,&isign3);

if (isnan(adFunc[0]) || isdotinf(adFunc[0]) || isign <= 0 || isign == 2
    || isign2 <= 0 || isign2 == 2 || isign3 <= 0 || isign3 == 2)
    return 0;

return 1;
} //FIM fLoglik_BN_COVDER

//***** SOB A HIPOTESE NULA *****

```

```

/*vP[0] é o parâmetro tal2
   vP[1] é o parâmetro ro*/

fLoglik_BN_COVDER_Re(const vP, const adFunc, const avScore, const amHess){
  decl SIGMA,SIGMAEMV,invSigma,dG1,dG2,dG3,dSigma1,dSigma2,dSigma3,dSigma4,ismgn,ismgn2,ismgn3,vai;

  dG1 = <1,0;0,0>;
  dG2 = <0,1;1,0>;
  dG3 = <0,0;0,1>;

  dSigma4 = unit(sumr(g_vn),sumr(g_vn));
  dSigma1 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG1) * g_mZ';
  dSigma2 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG2) * g_mZ';
  dSigma3 = g_mZ * (unit(columns(g_vn),columns(g_vn))*dG3) * g_mZ';

  SIGMA = g_mZ * (unit(columns(g_vn),columns(g_vn))*(vP[0:1]' | vP[1:2]')) * g_mZ'
    + vP[3] * unit(sumr(g_vn),sumr(g_vn));

  vai = choleski(SIGMA);

  invSigma = zeros(sumr(g_vn),sumr(g_vn));

  invSigma = invertgen(vai*vai');

  decl vPar_Re;
  vPar_Re = vP[4:g_inparNA0test-1][0] | zeros(g_inpartest, 1);

  //INFORMACAO OBSERVADA
  decl j = new matrix[g_inparNA0test][g_inparNA0test];

  j[0][0] = (-1/2) * trace(invSigma * dSigma1 * invSigma * dSigma1)
    + (g_vY - g_mX * vPar_Re)' * invSigma * dSigma1 * invSigma
    * dSigma1 * invSigma * (g_vY - g_mX * vPar_Re);
  j[1][1] = (-1/2) * trace(invSigma * dSigma2 * invSigma * dSigma2)
    + (g_vY - g_mX * vPar_Re)' * invSigma * dSigma2 * invSigma * dSigma2
    * invSigma * (g_vY - g_mX * vPar_Re);
  j[2][2] = (-1/2) * trace(invSigma * dSigma3 * invSigma * dSigma3)
    + (g_vY - g_mX * vPar_Re)' * invSigma * dSigma3 * invSigma * dSigma3
    * invSigma * (g_vY - g_mX * vPar_Re);
  j[3][3] = (-1/2) * trace(invSigma * dSigma4 * invSigma * dSigma4)
    + (g_vY - g_mX * vPar_Re)' * invSigma * dSigma4 * invSigma * dSigma4
    * invSigma * (g_vY - g_mX * vPar_Re);

  j[0][1] = j[1][0] = (-1/2) * trace(invSigma * dSigma1 * invSigma * dSigma2)
    + (g_vY - g_mX * vPar_Re)' * invSigma * dSigma1 * invSigma * dSigma2
    * invSigma * (g_vY - g_mX * vPar_Re);
  j[0][2] = j[2][0] = (-1/2) * trace(invSigma * dSigma1 * invSigma * dSigma3)
    + (g_vY - g_mX * vPar_Re)' * invSigma * dSigma1 * invSigma * dSigma3
    * invSigma * (g_vY - g_mX * vPar_Re);
  j[0][3] = j[3][0] = (-1/2) * trace(invSigma * dSigma1 * invSigma * dSigma4)
    + (g_vY - g_mX * vPar_Re)' * invSigma * dSigma1 * invSigma * dSigma4
    * invSigma * (g_vY - g_mX * vPar_Re);

```

```

j[1][2] = j[2][1] = (-1/2) * trace(invSigma * dSigma2 * invSigma * dSigma3)
      + (g_vY - g_mX * vPar_Re)' * invSigma * dSigma2 * invSigma * dSigma3
      * invSigma * (g_vY - g_mX * vPar_Re);
j[1][3] = j[3][1] = (-1/2) * trace(invSigma * dSigma2 * invSigma * dSigma4)
      + (g_vY - g_mX * vPar_Re)' * invSigma * dSigma2 * invSigma * dSigma4
      * invSigma * (g_vY - g_mX * vPar_Re);

j[2][3] = j[3][2] = (-1/2) * trace(invSigma * dSigma3 * invSigma * dSigma4)
      + (g_vY - g_mX * vPar_Re)' * invSigma * dSigma3 * invSigma * dSigma4
      * invSigma * (g_vY - g_mX * vPar_Re);

j[4:g_inparNAOtest-1][0] = g_mX' * invSigma * dSigma1 * invSigma * (g_vY - g_mX * vPar_Re);
j[0][4:g_inparNAOtest-1] = j[4:g_inparNAOtest-1][0]';

j[4:g_inparNAOtest-1][1] = g_mX' * invSigma * dSigma2 * invSigma * (g_vY - g_mX * vPar_Re);
j[1][4:g_inparNAOtest-1] = j[4:g_inparNAOtest-1][1]';

j[4:g_inparNAOtest-1][2] = g_mX' * invSigma * dSigma3 * invSigma * (g_vY - g_mX * vPar_Re);
j[2][4:g_inparNAOtest-1] = j[4:g_inparNAOtest-1][2]';

j[4:g_inparNAOtest-1][3] = g_mX' * invSigma * dSigma4 * invSigma * (g_vY - g_mX * vPar_Re);
j[3][4:g_inparNAOtest-1] = j[4:g_inparNAOtest-1][3]';

j[4:g_inparNAOtest-1][4:g_inparNAOtest-1] = g_mX' * invSigma * g_mX;

//COVARIANCIAS EMPIRICAS
decl i = new matrix[g_inparNAOtest][g_inparNAOtest];

i[0][0] = (1/2) * trace(invSigma * dSigma1 * invSigma * dSigma1);
i[1][1] = (1/2) * trace(invSigma * dSigma2 * invSigma * dSigma2);
i[2][2] = (1/2) * trace(invSigma * dSigma3 * invSigma * dSigma3);
i[3][3] = (1/2) * trace(invSigma * dSigma4 * invSigma * dSigma4);

i[0][1] = i[1][0] = (1/2) * trace(invSigma * dSigma1 * invSigma * dSigma2);
i[0][2] = i[2][0] = (1/2) * trace(invSigma * dSigma1 * invSigma * dSigma3);
i[0][3] = i[3][0] = (1/2) * trace(invSigma * dSigma1 * invSigma * dSigma4);

i[1][2] = i[2][1] = (1/2) * trace(invSigma * dSigma2 * invSigma * dSigma3);
i[1][3] = i[3][1] = (1/2) * trace(invSigma * dSigma2 * invSigma * dSigma4);

i[2][3] = i[3][2] = (1/2) * trace(invSigma * dSigma3 * invSigma * dSigma4);

i[0][4:g_inparNAOtest-1] = (g_mX * g_vEMV[4:g_inpar-1] - g_mX * vPar_Re)' * invSigma
      * dSigma1 * invSigma * g_mX;
i[1][4:g_inparNAOtest-1] = (g_mX * g_vEMV[4:g_inpar-1] - g_mX * vPar_Re)' * invSigma
      * dSigma2 * invSigma * g_mX;
i[2][4:g_inparNAOtest-1] = (g_mX * g_vEMV[4:g_inpar-1] - g_mX * vPar_Re)' * invSigma
      * dSigma3 * invSigma * g_mX;
i[3][4:g_inparNAOtest-1] = (g_mX * g_vEMV[4:g_inpar-1] - g_mX * vPar_Re)' * invSigma
      * dSigma4 * invSigma * g_mX;

i[4:g_inparNAOtest-1][4:g_inparNAOtest-1] = g_mX' * invSigma * g_mX;

```

```

adFunc[0] = - (1/2) * log(2 * M_PI) - (1/(2*sumr(g_vn))) * logdet(vai*vai', &isign)
           - (1/(2*sumr(g_vn))) * (g_vY - g_mX * vPar_Re)' * invertgen(vai*vai')
           * (g_vY - g_mX * vPar_Re) + (1/(2 * sumr(g_vn))) * logdet(j,&isign2)
           - (1/sumr(g_vn)) * logdet(i,&isign3);

if (isnan(adFunc[0]) || isdotinf(adFunc[0]) || isign <= 0 || isign == 2
    || isign2 <= 0 || isign2 == 2 || isign3 <= 0 || isign3 == 2)
    return 0;

return 1;
} //FIM fLoglik_BN_COVDER_Re

##### PROGRAMA PRINCIPAL
main(){

    oxwarning(0);

    ##### DECLARACAO DAS VARIAVEIS
    decl
        Y, N, n, nobs, nsim, teta, beta0, beta0p, beta0h0, sigma2, d, v, ro, tal2, vemv,
        vh0, soman, i, j, aux0, aux1, aux2, naux, cont, cont2, rvp, rvbn, txcob1, txcob2,
        qq, ww, pp, derivro, deriv, beta0EMV, tal2EMV, dfunc0, dfunc1, vero, vP, a, inv,
        tal2_p, k, u, p, npar, nB, jB, ERV, tij, x3, x2;

    ##### INICIALIZACAO DAS VARIAVEIS
    N = 15;                               //Numero de individuos
    nsim = 3000;                           //Numero de replicas MC
    g_vbeta = <0,0.2,0,0>;                 //Vetor de parametros referentes aos efeitos fixos
    g_vw = <1,0.35,1.0,0.05>;              //Vetor de parametros referentes aos efeitos aleatorios
    g_iCovarv = columns(g_vbeta);           //Numero de covariaveis no modelo
    g_vParNTes = <1, 1, 1, 1, 1, 1, 0, 0>; /* Vetor indicador dos parametros a serem testados
                                           (deve ter dimensao igual a g_iCovarv):
                                           0 = parametro testado,
                                           1 = parametro nao-testado */
    g_inpar = columns(g_vParNTes);          //Numero de parametros do modelo
    g_inparNAOtest = sumr(g_vParNTes);      //Numero de parametros NAO testados
    g_inpartest = g_inpar - g_inparNAOtest; //Numero de parametros testados
    nB = 0;                                 //Numero de amostras BOOTSTRAP

    // Geracao dos numeros de obs por individuo e matriz de especificacao

    //DECLARACAO DE ARQUIVOS
    decl ESTmc = fopen("ESTmc.txt","w");    //Estimativas dos parametros de interesse
    decl RVmc = fopen("RVmc.txt","w");      //Estatisticas da razao de verossimilhanças

    g_vn = reshape(loadmat("vNtat3.txt",1),1,N)[0:N-1]; //Vetor do n de observacoes por individuo
    println(g_vn');

    tij = ranu(sumr(g_vn),1); //Tempos das Observacoes
    g_mX = reshape(loadmat("mX.txt",1),sumr(g_vn[0:N-1]),g_iCovarv); //Matriz de covariaveis X

```

```

g_mZ = Z(g_vn);           //Matriz de especificacao efeitos aleatorios Z

##### ESCOLHA DO GERADOR DE NUMEROS ALEATORIOS
ranseed("GM");

#####DECLARACAO DOS CONTADORES
// ***** IRRESTRITO *****
decl cf_max_0, cmax, cf, cf_max_P, cf_max_BN_COVDER;
cf_max_0 = cmax = cf = cf_max_P = cf_max_BN_COVDER = 0;

// ***** SOB A HIPOTESE NULA *****
decl cf_max_0_re, cf_max_P_re, cf_max_BN_COVDER_re;
cf_max_0_re = cf_max_P_re = cf_max_BN_COVDER_re = 0;

// ***** TIPOS DE ERROS NA CONVERGENCIA DA MAXIMIZACAO *****
decl erromaxite, erroeps, errofunc, erronuncamax;
erromaxite = erroeps = errofunc = erronuncamax = 0;

//***** CONTADORES DAS ESTATISTICAS DA RV COM SINAL NEGATIVO *****
decl csin, csin_RV, csin_RV_P, csin_RV_BN_COVDER;
csin = csin_RV = csin_RV_P = csin_RV_BN_COVDER = 0;

//***** CONTADORES DE REJEICAO DA HIPOTES NULA *****
decl cRV, cRV_P, cRV_BN_COVDER;
cRV = cRV_P = cRV_BN_COVDER = zeros(9, 1);

##### DECLARACAO DAS MATRIZES DE RESULTADOS
//***** ESTIMADORES PARA BETA *****
decl mRESTo, mRESTp, mRESTbn_covder;
mRESTo = zeros(nsim,g_inpar);
mRESTp = zeros(nsim,g_inpar);
mRESTbn_covder = zeros(nsim,g_inpar);

//***** VALORES DAS ESTATISTICAS DA RAZAO DE VEROSSIMILHANCAS *****
decl mRV;
mRV = zeros(nsim,6);

// Niveis de significancia do teste

##### VARIAVEIS UTILIZADAS NOS TESTES DE HIPOTHESES
decl vsig, vquantil;
vsig = <0.15; 0.10; 0.05; 0.03; 0.01; 0.005; 0.001; 0.0005; 0.0001>;

//***** Quantis da distribuicao qui-quadrado *****
vquantil = quanchi(1 - vsig, g_inpartest);

//***** MATRIZES DE RESULTADOS *****
decl mres2, mres3, mres4;

```

```

mres2 = zeros(9, 6); // teste de hipotese
mres3 = zeros(2, 6); // momentos da est. da RV
mres4 = zeros(9, 6); // quantis empiricos

#####VARIABLES DE CONTROLE PARA MAXIMIZACAO
MaxControl(100000, 0); // (num. max. iter. , impr. iter.)

#####DECLARACAO DO VETOR DE PARAMETROS P/ MAXIMIZACAO e ESPACO PARAMETRICO
// ***** IRRESTRITO *****
decl vtheta, vtheta_p, vtheta_bn_covder, vLiminf, vLimsup;
vLiminf = vLimsup = vtheta = vtheta_p = vtheta_bn_covder = zeros(g_inpar, 1);

// ***** SOB A HIPOTESE NULA *****
decl vtheta_re, vtheta_p_re, vtheta_bn_covder_re, vLiminf_re, vLimsup_re;
vLiminf_re = vLimsup_re = vtheta_re = vtheta_p_re = vtheta_bn_covder_re = zeros(g_inparNA0test, 1);

#####DECLARACAO DAS MATRIZES DE RESTRICOES, VALORES USADOS NAS MAXIMIZACOES
// ***** IRRESTRITO *****
decl vo, dfunc, //Verossimilhanca original
vp, dfunc_p, mrest_p, //Verossimilhanca perfilada
v_bn_covder, dfunc_bn_covder, mrest_bn_covder; //Verossimilhanca perfilada ajustada BN_COV

dfunc = dfunc_p = dfunc_bn_covder = 0;
mrest_p = mrest_bn_covder = zeros(g_inpar, 1);

// ***** SOB A HIPOTESE NULA *****
decl vo_re, dfunc_re, //Verossimilhanca original
vp_re, dfunc_p_re, mrest_p_re, //Verossimilhanca perfilada
v_bn_covder_re, dfunc_bn_covder_re, mrest_bn_covder_re; //Verossimilhanca perfilada ajustada BN_COV

dfunc_re = dfunc_p_re = dfunc_bn_covder_re = 0;
mrest_p_re = mrest_bn_covder_re = zeros(g_inpar - g_inpartest, 1);

##### CONTROLE DA EXECUCAO DO PROGRAMA
decl cprint;
cprint = floor(nsim / 10);

#####INICIALIZACAO DO TIMER
decl extime;
extime = timer();

##### INÍCIO DA SIMULACAO
for(j=0;j<nsim + cf;j++){

#####GERACAO DA AMOSTRA

```

```

g_vY = g_mX * g_vbeta' + choleski(Sigma(g_vw)) * rann(sumr(g_vn),1);

#####

//## MAXIMIZACOES DAS VEROSSIMILHANCAS

//***** Verossimilhanca Original *****

//***** IRRESTRITA *****
// Valores iniciais dos parametros

vtheta[0] = log(g_vw[0]);
vtheta[1] = g_vw[1];
vtheta[2:3] = log(g_vw[2:3]);
vtheta[4:g_inpar-1] = g_vbeta+0.2;

vtheta = vtheta * 1.2;

vLiminf = <0;-.Inf;0;0;-.Inf;-.Inf;-.Inf;-.Inf>; //Limite Inferior do espaço paramétrico
vLimsup = <+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf>; //Limite Superior

// Variaveis de controle para maximizacoes
MaxControlEps(0.0001, 0.0001); // (eps1, eps2)

// Maximizacao da log-verossimilhanca
vo = MaxBFGS(fLoglik, &vtheta, &dfunc, 0, 0);

// Checa a convergencia
if(vo != MAX_CONV && vo != MAX_WEAK_CONV){
    ++ cf_max_0;
    ++ cmax;
    ++ cf;

    println(g_vEMV);
    println("\n erro original: ", MaxConvergenceMsg(vo));
}

if(vo == MAX_MAXIT) ++ erromaxite;
if(vo == MAX_LINE_FAIL) ++ erroeps;
if(vo == MAX_FUNC_FAIL) ++ errofunc;
if(vo == MAX_NOCONV) ++ erronuncamax;

// Sai do loop MC quando nao ha convergencia
if(vo != MAX_CONV && vo != MAX_WEAK_CONV) continue;

vtheta = g_vEMV; //Definicao do EMV do vetor de parametros

//***** SOB A HIPOTESE NULA *****

// Valores iniciais dos parametros sob a hipotese nula
vtheta_re[0] = log(g_vEMV[0]);

```

```

vtheta_re[1] = g_vEMV[1];
vtheta_re[2:3] = log(g_vEMV[2:3]);
vtheta_re[4:g_inparNA0test-1] = g_vEMV[4:g_inparNA0test-1];

vLiminf_re = <0;-.Inf;0;0;-.Inf;-.Inf>; //Limite Inferior do espaço paramétrico restrito
vLimsup_re = <+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf>; //Limite Superior

// Variaveis de controle para maximizacoes
MaxControlEps(0.0001, 0.0001); // (eps1, eps2)

// Maximizacao da log-verossimilhanca sob a hipotese nula
vo_re = MaxBFGS(fLoglik_Re, &vtheta_re, &dfunc_re, 0, 0);

// Checa a convergencia
if(vo_re != MAX_CONV && vo_re != MAX_WEAK_CONV){
    ++ cf_max_0_re;
    ++ cmax;
    ++ cf;

    println(g_vEMV0);
    println("\n erro original - SOB H0: ", MaxConvergenceMsg(vo_re));
}

if(vo_re == MAX_MAXIT) ++ erromaxite;
if(vo_re == MAX_LINE_FAIL) ++ erroeps;
if(vo_re == MAX_FUNC_FAIL) ++ errofunc;
if(vo_re == MAX_NOCONV) ++ erronuncamax;

// Sai do loop MC quando nao ha convergencia
if(vo_re != MAX_CONV && vo_re != MAX_WEAK_CONV) continue;

//***** Verossimilhanca Perfilada Ajustada BN_COVDER *****

//***** IRRESTRITA *****
// Valores iniciais dos parametros

vtheta_bn_covder = vtheta;

// Variaveis de controle para maximizacoes
MaxControlEps(0.00001, 0.00001); // (eps1, eps2)

// Maximizacao conjunta da log-verossimilhanca
v_bn_covder = MaxSQP(fLoglik_BN_COVDER, &vtheta_bn_covder, &dfunc_bn_covder, 0, 1, 0,
    fRest(&mrest_bn_covder, vtheta_bn_covder), vLiminf, vLimsup);

// Checa a convergencia
if(v_bn_covder != MAX_CONV && v_bn_covder != MAX_WEAK_CONV){
    ++ cf_max_BN_COVDER;
    ++ cmax;
    ++ cf;

    println(vtheta_bn_covder);

```

```

        println("\n erro BN_COVDER: ", MaxConvergenceMsg(v_bn_covder));
    }

    if(v_bn_covder == MAX_MAXIT) ++ erromaxite;
    if(v_bn_covder == MAX_LINE_FAIL) ++ erroeps;
    if(v_bn_covder == MAX_FUNC_FAIL) ++ errofunc;
    if(v_bn_covder == MAX_NOCONV) ++ erronuncamax;

    // Sai do loop quando nao ha convergencia
    if(v_bn_covder != MAX_CONV && v_bn_covder != MAX_WEAK_CONV) continue;

    //***** SOB A HIPOTESE NULA *****
    //println("SOB H0");
    // Valores iniciais dos parametros

    vtheta_bn_covder_re = vtheta_bn_covder[0:g_inparNA0test-1];

    // Variaveis de controle para maximizacoes
    MaxControlEps(0.00001, 0.00001);      // (eps1, eps2)

    // Maximizacao conjunta da log-verossimilhanca
    v_bn_covder_re = MaxSQP(fLoglik_BN_COVDER_Re, &vtheta_bn_covder_re, &dfunc_bn_covder_re, 0, 1, 0,
        fRest_Re(&mrest_bn_covder_re, vtheta_bn_covder_re), vLiminf_re, vLimsup_re);

    // Checa a convergencia
    if(v_bn_covder_re != MAX_CONV && v_bn_covder_re != MAX_WEAK_CONV){
        ++ cf_max_BN_COVDER_re;
        ++ cmax;
        ++ cf;

        println(vtheta_bn_covder);
        println("\n erro BN_COVDER_re: ", MaxConvergenceMsg(v_bn_covder_re));
    }

    if(v_bn_covder_re == MAX_MAXIT) ++ erromaxite;
    if(v_bn_covder_re == MAX_LINE_FAIL) ++ erroeps;
    if(v_bn_covder_re == MAX_FUNC_FAIL) ++ errofunc;
    if(v_bn_covder_re == MAX_NOCONV) ++ erronuncamax;

    // Sai do loop quando nao ha convergencia
    if(v_bn_covder_re != MAX_CONV && v_bn_covder_re != MAX_WEAK_CONV) continue;

    //***** ESTADISTICAS DA RAZAO DE VEROSSIMILHANCAS *****
    //***** RV ORIGINAL*****
    mRV[j - cf][0] = 2 * sumr(g_vn) * (dfunc - dfunc_re);
    if(mRV[j-cf][0] < 0){
        ++ csin_RV;
        ++ csin;
    }

```

```

        ++ cf;
        continue;
    }

    //***** RV Perfilada          Modificada BN_COVDER *****
    mRV[j - cf][1] = 2 * sumr(g_vn) * (dfunc_bn_covder - dfunc_bn_covder_re);
    if(mRV[j - cf][1] < 0){
        ++ csin_RV_BN_COVDER;
        ++ csin;
        ++ cf;
        continue;
    }

    //### NIVEL DE SIGNIFICANCIA: 100 * vsig[cq][0] % //
    for(i = 0; i < 9; ++i){
        if(mRV[j-cf][0] >= vquantil[i][0]) ++cRV[i][0];           // original
        if(mRV[j-cf][1] >= vquantil[i][0]) ++cRV_BN_COVDER[i][0]; // perfilada modificada BN_COVDER
        if(mRV[j-cf][2] >= vquantil[i][0]) ++cRVB[i][0];         // perfilada modificada BN_COVDER
        if(mRV[j-cf][3] >= vquantil[i][0]) ++cRV_BN_COVDERB[i][0]; // perfilada modificada BN_COVDER
    }

    //*****

    //###Estimadores
    //
    //
    //
    //
    println("Original: ",vtheta);
    println("Perfilada: ",vtheta_p);
    println("Perfilada COVDER: ",vtheta_bn_covder);
    println("Perfilada COVEMP: ",vtheta_bn_covemp);

    //### Resultados da estimacao
    mRESTo[j - cf][0:g_inpar-1] = vtheta';
    mRESTp[j - cf][0:g_inpar-1] = vtheta_p';
    mRESTbn_covder[j - cf][0:g_inpar-1] = vtheta_bn_covder';

    //###imprime o estagio do progresso
    if (imod(j - cf + 1, cprint) == 0){
        print(j - cf + 1, " replicas de MC realizadas: ",
            ((j - cf + 1)/nsim) * 100, " por cento\n");
    }

} //#####FIM SIMULACAO
fprintf(ESTmc,mRESTo[][g_inparNAOtest:g_inpar-1] ~ mRESTbn_covder[][g_inparNAOtest:g_inpar-1]);
fprintf(RVmc,mRV);

fclose(ESTmc);
fclose(RVmc);

##### RESULTADOS
##### TAMANHO DO TESTE #####

```

```

        for(i = 0; i < 9; ++i){
            mres2[i][0] = 100 * (cRV[i][0] / nsim);           // Original
            mres2[i][1] = 100 * (cRV_BN_COVDER[i][0] / nsim); // Perfilada modificada BN_COVDER
            mres2[i][2] = 100 * (cRVB[i][0] / nsim);         // Original
            mres2[i][3] = 100 * (cRV_BN_COVDERB[i][0] / nsim); // Perfilada modificada BN_COVDER
        }

println("\n Media MC dos EMVs para beta:");
println(meanc(mRESTo));
println(mRESTo);
println("\n Media MC dos EMVPs modificada BN_COVDER para beta:");
println(meanc(mRESTbn_covder));
println(mRESTbn_covder);
println("Media MC das estatisticas da RV:");
println(meanc(mRV));
println(mRV);
println("\n Dados observados:");
println( "Numero de individuos: ", N);
println( "Numero maximo de observacoes por individuo: ", nobs);
println( "Fator de desbalanceamento: ", p);
println("\n Verdadeiros Parametros:");
println( "Beta: ", g_vbeta);
println( "W: ", g_vw);
println("\n Parametros Testados:", g_vParNTes);

print( "\n\t ** NUM. REPLICAS TOTAL **: ", nsim + cf );
print( "\n\t NUM. REPLICAS MC DESCARTADAS: ", cf );
print( "\n\t\t Problemas de maximizacao: ", cmax );
print( "\n\t\t\t Veros. Original: ", cf_max_0 );
print( "\n\t\t\t Veros. Original - SOB H0: ", cf_max_0_re );
print( "\n\t\t\t Veros. Perfilada: ", cf_max_P );
print( "\n\t\t\t Veros. Perfilada - SOB H0: ", cf_max_P_re );
print( "\n\t\t\t Veros. Perfilada Ajustada BN_COVDER: ", cf_max_BN_COVDER );
print( "\n\t\t\t Veros. Perfilada Ajustada BN_COVDER - SOB H0: ", cf_max_BN_COVDER_re );
print( "\n\t\t Sinal negativo da estatistica da RV: ", csin );
print( "\n\t\t\t Veros. Original: ", csin_RV );
print( "\n\t\t\t Veros. Perfilada: ", csin_RV_P );
print( "\n\t\t\t Veros. Perfilada Ajustada BN_COVDER: ", csin_RV_BN_COVDER );
print( "\n DATA: ", date() );
println( "\n HORA: ", time() );

print( "\n\t **** ERROS **** ");
print( "\n\t ERRO: Maximo de Iteracoes Excedeu - ", erromaxite);
print( "\n\t ERRO: Eps nao foi atingido - ", erroeps);
print( "\n\t ERRO: Funcao avaliada falhou - ", errofunc);
print( "\n\t ERRO: Provavelmente nao ha maximo - ", erronuncamax);

//##### TEMPO DE EXECUCAO

```

```
print( "\n\n\t\t TEMPO DE EXECUCAO: ", timespan(extime) );
print( "\n" );

##### Resultados do teste de hipotese

print( "\n\nResultados do teste de hipotese: ");
print( "\n\n", "%8.3f", "%c", {"0", "P", "BN_COVDER", "BN_COVEMP"}, "%r",
      {"15%", "10%", "5%", "3%", "1%", "0.5%", "0.1%", "0.05%", "0.01%"}, mres2 );

}
```

Referências Bibliográficas

- Barndorff-Nielsen, O. E. (1980). Conditionality resolutions. *Biometrika*, **67**, 293-310.
- Barndorff-Nielsen, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343-365.
- Bartlett, M. S. (1937). Properties of sufficiency and statistical test. *Proceeding of the Royal Society A*, **160**, 268-282.
- Bates, D. M. e DebRoy, S. (2004). Linear mixed models and penalized least squares. *Journal of Multivariate Analysis*, **91**, 1-17.
- Cox, D. R. e Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society B*, **49**, 1-39.
- Crainiceanu, C. M. e Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society B*, **66**, 165-185.
- Diggle, P. J., Liang, K. Y. e Zeger, S. L. (1994). *Analysis of Longitudinal Data*. New York: Oxford, 253p.
- Demidenko, E. (2004). *Mixed Models: Theory and Applications*. New York: John Wiley e Sons. Inc.
- Doornik, J.A. (2007). *Ox: An Object-oriented Matrix Programming Language*. London: Timberlake Consultants Press.
- Doornik, J.A. e Ooms, M. (2007). *Introduction to Ox: An Object-Oriented Matrix Language*. London: Timberlake Consultants Press.
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, **7**, n. 1, 1-26.
- Efron, B. e Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman e Hall.
- Ferrari, S. L. P., Cribari-Neto, F. e Ferreira da Silva, M. (2005). Adjusted profile likelihood for two-parameter exponential family models. *Communications in Statistics. Theory and Methods*, **34**, n. 2, 33-53.
- Ferreira da Silva, M. (2005). Estimação e teste de hipótese baseados em verossimilhanças perfiladas. *Tese de Doutorado - Instituto de Matemática e Estatística - Universidade de São Paulo*
- Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72**, 320-340.
- Ihaka, R. e Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of Computational Graphics and Statistics*, **5**, 299-314.

- Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*. New York: Springer.
- Kenward, M. G. e Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, **53**, 983-997.
- Knuth, D. (1986). *The T_EXbook*. New York: Addison-Wesley.
- Laird, N. M. e Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*, **38**, 963-974.
- Lindstrom, M. e Bates, D. (1988). Newton Raphson and EM algorithms for linear mixed-effects modelos for repeated-measures data. *Journal of the American Statistical Association*, **83**, 1014-1022.
- Littel, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D. e Schadenderger, O. (2006). *SAS for Mixed Models*. 2nd ed. Cary, NC: SAS Institute Inc.
- Manor, O. e Zucker, D. M. (2004). Small sample inference for the fixed effects in the mixed linear model. *Computational Statistica and Data Analysis*, **46**, 801-817.
- McCulloch, C. E. e Searle, R. S. (2001). *Generalized, linear and mixed models*. New York: John Wiley e Sons, Inc.
- Pinheiro, J. C. e Bates, D. M. (2000). *Mixed -Effects Models in S and S-PLUS*. New York: Springer.
- Severini, T. A. (1998). An approximation to the modified profile likelihood function. *Biometrika*, **85**, 403-411.
- Severini, T. A. (1999). An empirical adjustment to the likelihood ratio statistic. *Biometrika*, **86**, 235-247.
- Severini, T. A. (2000). *Likelihood Methods in Statistics*. Oxford: Oxford University Press.
- Verbeke, G. e Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer, 569p.
- Welham, S. J. e Thompson, R. (1997). Likelihood ratio tests for fixed model terms using residual maximum likelihood. *Journal of the Royal Statistical Society B*, **59**, 701-714.
- Zucker, D. M., Lieberman, O. e Manor, O. (2000). Improved small sample inference in the mixed linear model: Bartlett correction and adjusted likelihood. *Journal of the Royal Statistical Society B*, **62**, 827-838.