

JHENNIFER DOS SANTOS NASCIMENTO

**COMPARAÇÃO DOS AJUSTES DE MODELOS COM ERRO NORMAL E
*SKEW-NORMAL***

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

Orientador: Paulo César Emiliano

Coorientadores: Altemir da Silva Braga

Edcarlos Miranda de Souza

**VIÇOSA - MINAS GERAIS
2020**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

N244c
2020 Nascimento, Jhennifer dos Santos, 1996-
Comparação dos ajustes de modelos com erro normal e
skew-normal / Jhennifer dos Santos Nascimento. – Viçosa, MG,
2020.
73 f. : il. (algumas color.) ; 29 cm.

Inclui apêndices.

Orientador: Paulo César Emiliano.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f.64-66.

1. Variáveis aleatórias. 2. Assimetria. 3. Delineamento de
experimento. 4. Monte Carlo, Método de. I. Universidade
Federal de Viçosa. Departamento de Estatística. Programa de
Pós-Graduação em Estatística Aplicada e Biometria. II. Título.

CDD 22. ed. 519.2

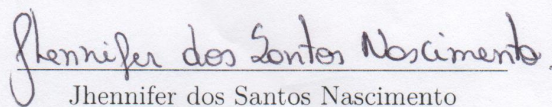
JHENNIFER DOS SANTOS NASCIMENTO

COMPARAÇÃO DOS AJUSTES DE MODELOS COM ERRO NORMAL E
SKEW-NORMAL

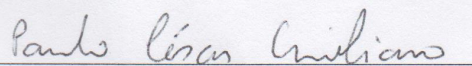
Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

APROVADA: 16 de outubro de 2020.

Assentimento:



Jhennifer dos Santos Nascimento
Autora



Paulo César Emiliano
Orientador

Dedico à minha grande família, que sem o suporte, apoio e cuidado, eu não teria conseguido.

AGRADECIMENTOS

Agradeço primeiramente a Deus, que é tão perfeito em sua maneira de agir e ensinar. Deus tem agido minha vida de forma que me emociona, quem sabe de onde venho e como fui criada sabe que minha vida poderia ser completamente oposta ao que é, e embora eu acredite na aleatoriedade da vida e que é necessário agir e perseverar para alcançar um objetivo, também acredito que Deus nos guia em seus caminhos, basta ter fé. Deus é perfeito em tudo que faz e não há motivo pra eu estar onde estou se não porque Deus quis assim, cada pessoa em meu caminho, cada amigo, cada professor, me trouxe a esta cidade, a este programa e a este departamento que é diferenciado, e sou grata a Ele por tudo.

Agradeço imensamente a minha família, que mesmo não compreendendo minhas escolhas me deram apoio incondicional e demonstraram fazer mais que o possível para que meus sonhos se realizassem. À minha mãe, Glacya, ao pai que me criou Ângelo, às minhas tias Ildelania e Ilnara, à minha avó querida Josete, a todos os meus seis irmãos, ao amor da minha vida Lucas Matheus, à minha sogra Valdenira e sua mãe Marina e aos muitos membros da família não citados aqui, meu sincero obrigada pelo cuidado, carinho, paciência e por todos os sacrifícios.

Aos amigos e companheiros do programa, agradeço pela forma que me acolheram, por toda a ajuda e diversão que proporcionaram mesmo em meio a tanto estresse e ansiedade, acredito que juntos somos mais fortes, sem vocês eu não teria conseguido permanecer tanto tempo longe das pessoas que amo e também não teria conseguido passar pelas temidas disciplinas. Foram vários fins de semana estudando e rindo juntos, jamais esquecerei desses momentos, sou muitíssimo grata pelo companheirismo e já sinto saudade dessa fase.

Aos professores da graduação e pós-graduação, obrigada por serem fonte de inspiração, por todos os ensinamentos e auxílio, vocês me fazem querer ser uma professora tão boa quanto vocês. À Paulo Emiliano, meu orientador, sou grata por ter me recepcionado e se mostrado um professor de excelência, exemplar e admiravelmente correto. Aos meus coorientadores, Altemir Braga, sou grata pela instrução, pelo paciente direcionamento em todas as atividades e pelas muitíssimas horas em reuniões, e à Edcarlos Miranda agradeço pela confiança e impulsionamento, suas aulas fizeram despertar meu interesse por estatística, se cheguei até aqui, você foi graças ao professor que você é. Agradeço ao professor Djair Frade pela participação e colaboração e a toda equipe que tornou este trabalho possível.

Agradeço a Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Estatística Aplicada e Biometria pela oportunidade, e a CAPES agradeço pela concessão da bolsa de estudo, o presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

“Nossa maior fraqueza está em desistir. O caminho mais certo de vencer é tentar mais uma vez.”

Thomas Edison

“Seja você quem for, seja qual for a posição social que você tenha na vida, a mais alta ou a mais baixa, tenha sempre como meta muita força, muita determinação e sempre faça tudo com muito amor e com muita fé em Deus, que um dia você chega lá. De alguma maneira você chega lá.”

Ayrton Senna

RESUMO

NASCIMENTO, Jhennifer dos Santos, M.Sc., Universidade Federal de Viçosa, outubro de 2020. **Comparação dos ajustes de modelos com erro normal e *skew-normal*.** Orientador: Paulo César Emiliano. Coorientadores: Altemir da Silva Braga e Edcarlos Miranda de Souza.

Na modelagem de fenômenos aleatórios, em que uma variável aleatória quantitativa é estudada em função de uma variável aleatória qualitativa, é muito comum ser utilizada a análise de variâncias (ANOVA). Porém, para essa análise ser validada é necessário que quatro pressuposições sejam verificadas com testes estatísticos, dentre elas, a normalidade do componente aleatório do modelo. Entretanto, na prática, alguns resultados de ensaios experimentais violam uma ou mais pressuposições da ANOVA tornando-se necessária uma outra metodologia para analisar os dados. Com isso, o objetivo deste trabalho é, em vez de utilizar a metodologia anteriormente mencionada, fazer o ajuste de um delineamento inteiramente casualizado (DIC), atribuindo a distribuição *skew-normal* para o componente aleatório do modelo, utilizando o método da máxima verossimilhança. Neste estudo, foram utilizados dois conjuntos de dados não normais com indícios de dispersão assimetria que foram imprescindíveis para fazer comparação, através dos critérios de comparação AIC e BIC, entre a distribuição *skew-normal* e o seu submodelo normal. Para o primeiro conjunto de dados, referente a temperatura, foi ajustado um modelo de intercepto, e o segundo conjunto de dados refere-se a um experimento realizado segundo um DIC, que avaliou o percentual de gordura de idosos que praticaram diferentes intervenções de atividade física. Neste trabalho, conduziu-se todos os testes de comparação de contrastes de duas médias e intervalos de confiança para as distribuições normal e *skew-normal*. Além dos critérios de informação de Akaike e Bayesiano, para avaliar o ajuste dos modelos aos dados, foi realizada análise de resíduos quantílicos, e para finalizar, foram realizados estudos de simulação para avaliar inferência e as propriedades assintóticas dos estimadores da distribuição *skew-normal* para os modelos ajustados. Conclui-se, assim, que esses estimadores ganham eficiência com o aumento do tamanho das amostras e fornecem melhores estimativas dos dados estudados em relação a distribuição normal, fato que foi comprovado no estudo dos envelopes simulados utilizando resíduos quantílicos e no estudo de simulação.

Palavras-chave: *Assimetria. Delineamentos experimentais. Monte Carlo. Verossimilhança.*

ABSTRACT

NASCIMENTO, Jhennifer dos Santos, M.Sc., Universidade Federal de Viçosa, October, 2020. **Comparison of model fitting with normal and *skew*-normal errors.** Advisor: Paulo César Emiliano. Co-advisors: Altemir da Silva Braga and Edcarlos Miranda de Souza.

In the modeling of random phenomena, where a quantitative random variable is studied in function of a qualitative random variable, it's very common to use analysis of variances (ANOVA). However, for this analysis to be validated, it's necessary that four assumptions be verified with statistical tests, among them, the normality of the random component of the model. However, in practice, some results of experimental tests violate one or more assumptions of ANOVA making it necessary another methodology to analyze the data. Therefore, the objective of this work is, instead of using the previously mentioned methodology, to adjust a completely randomized design, assigning the *skew*-normal distribution to the random component of the model, using the maximum likelihood method. In this study, two non-normal data sets with evidence of asymmetric dispersion were used, which were essential to make a comparison, using the AIC and BIC comparison criteria, between the *skew*-normal distribution and its normal sub-model. For the first data set, referring to temperature, an intercept model was adjusted, and the second data set refers to an experiment carried out according to a completely randomized design, which evaluated the percentage of fat in the elderly who practiced different interventions of physical activity. In this work, all tests of comparison of contrasts of two means and confidence intervals for the normal and *skew*-normal distributions were conducted. In addition to the Akaike and Bayesian information criteria, to assess the fit of the models to the data, quantitative residue analysis was performed, and finally, simulation studies were carried out to assess inference and the asymptotic properties of the *skew*-normal distribution estimators for the adjusted models. It is concluded, therefore, that these estimators gain efficiency with the increase of the sample size and provide better estimates of the studied data in relation to the normal distribution, a fact that was proven in the study of the simulated envelopes using quantile residues and in the simulation study.

Keywords: *Asymmetry. Experimental designs. Monte Carlo. Likelihood.*

LISTA DE FIGURAS

Figura 1: Curva do coeficiente de assimetria da distribuição $SN(\lambda)$, em função de λ	25
Figura 2: Comportamento da distribuição <i>skew</i> -normal de acordo com a variação do parâmetro de assimetria λ	25
Figura 3: Saída da função <i>summary.lm</i> do ajuste de um DIC, com três tratamentos e duas repetições, feito pela função <i>aov</i> no <i>software</i> R.	35
Figura 4: Histograma com a frequência absoluta e <i>boxplot</i> para a variável temperatura. .	48
Figura 5: Histograma para a variável temperatura com as densidades das distribuições normal e <i>skew</i> -normal (SN).	50
Figura 6: Envelopes simulados dos quantis normais versus os resíduos quantílicos para o modelo de intercepto com distribuição normal e <i>skew</i> -normal, respectivamente. .	51
Figura 7: Funções de densidade da distribuição SN nos valores verdadeiros dos parâmetros e nas estimativas médias para $n=30$; $n=100$; $n=365$	53
Figura 8: <i>Boxplot</i> para os tratamentos da variável percentual de gordura.	54
Figura 9: Envelopes simulados dos quantis normais versus os resíduos quantílicos para o modelo inteiramente casualizado com distribuição normal e <i>skew</i> -normal, respectivamente.	57
Figura 10: Densidades da distribuição <i>skew</i> -normal com $\lambda = -2,5$, nos valores verdadeiros dos parâmetros e nas estimativas médias para $n=20$; $n=68$; $n=120$	60
Figura 11: Densidades da distribuição <i>skew</i> -normal com $\lambda = -4,43$, nos valores verdadeiros dos parâmetros e nas estimativas médias para $n=20$; $n=68$; $n=120$	61
Figura 12: Densidades da distribuição <i>skew</i> -normal com $\lambda = 4$, nos valores verdadeiros dos parâmetros e nas estimativas médias para $n=20$; $n=68$; $n=120$	62

LISTA DE TABELAS

Tabela 1: Teste t e intervalos de confiança para os parâmetros utilizando-se as distribuições normal e <i>skew</i> -normal para a variável temperatura.	49
Tabela 2: Critérios de informação para as distribuições normal e <i>skew</i> -normal para a variável temperatura.	49
Tabela 3: Estimativa dos parâmetros, variância e erro quadrático médio, baseados em 1000 simulações da distribuição <i>skew</i> -normal para o modelo de intercepto.	52
Tabela 4: Teste t e intervalos de confiança para os parâmetros das distribuições normal e <i>skew</i> -normal em um DIC.	55
Tabela 5: Critérios de informação para os modelos normal e <i>skew</i> -normal e experimento em DIC.	56
Tabela 6: Estimativa dos parâmetros, variância e erro quadrático médio, baseados em 1000 simulações da distribuição <i>skew</i> -normal para o modelo em DIC.	59

SUMÁRIO

1	Introdução	11
2	Referencial teórico	13
2.1	Delineamentos experimentais	13
2.1.1	Delineamento inteiramente casualizado - DIC	13
2.1.2	Delineamento em blocos casualizados - DBC	14
2.1.3	Delineamento em quadrado latino - DQL	14
2.2	Modelo de intercepto	15
2.3	Distribuição normal	15
2.3.1	Função geradora de momentos	16
2.4	Distribuição <i>skew</i> -normal	18
2.4.1	Função geradora de momentos	20
2.5	Método da máxima verossimilhança	25
2.6	Critérios de seleção de modelos	27
2.7	Resíduos quantílicos	28
3	Estimação dos parâmetros dos modelos	29
3.1	EMV para $Y \sim N(\mu, \sigma^2)$	29
3.2	EMV para $Y \sim N(m + \tau_i, \sigma^2)$	31
3.3	EMV para $Y \sim SN(\mu, \sigma, \lambda)$	37
3.4	EMV para $Y \sim SN(m + \tau_i, \sigma, \lambda)$	41
3.5	EMV para $Y \sim SN(m + \tau_i + \beta_j, \sigma, \lambda)$	43
3.6	EMV para $Y \sim SN(m + \alpha_i + \tau_j + \beta_k, \sigma, \lambda)$	45
4	Resultados e discussão	47
4.1	Dados de temperatura	47
4.1.1	Ajuste	48
4.1.2	Análise dos resíduos	50
4.1.3	Estudo de simulação	51
4.2	Dados de percentual de gordura	53
4.2.1	Ajuste	55
4.2.2	Análise dos resíduos	56
4.2.3	Estudo de simulação	57
5	Conclusões	63

1 Introdução

A seleção de modelos é uma parte importante de toda pesquisa em modelagem estatística e envolve a procura de um modelo que seja o mais simples possível e que descreva bem o processo gerador dos valores observados que surgem em diversas áreas do conhecimento como agricultura, demografia, ecologia, economia, engenharia, geologia, medicina, ciência política, sociologia e zootecnia, entre outras (CORDEIRO e DEMÉTRIO, 2013).

Ao realizar um experimento e comparar efeitos de tratamentos, em geral, é utilizado o modelo linear de Gauss Markov normal, no qual há a pressuposição de normalidade para os resíduos e, como consequência, os dados também precisam ser normais para que sejam atendidas as exigências do modelo e as conclusões sejam válidas. Inúmeros trabalhos, em diversas áreas do conhecimento, apresentam comparações de efeitos de tratamentos utilizando análise de variâncias (ANOVA).

Pinto et al. (2015) realizaram um experimento de produção de três tipos de cerveja artesanal, em que cada uma possuía um percentual diferente de acerola e abacaxi e testaram via ANOVA, sob o delineamento inteiramente casualizado (DIC), se houve diferença significativa entre as propriedades das novas cervejas. Verificaram, também, se tais propriedades estavam de acordo com o padrão definido pela base de formulações típicas de cervejas do tipo *Pale Ale*, que são cervejas de alta fermentação. Para este ensaio, os autores relataram que os resíduos foram considerados normais pelo testes de normalidade e os testes de comparação múltipla indicaram que houve diferença significativa entre oito das nove características essenciais testadas entre as três cervejas produzidas.

Freitas et al. (2018), por sua vez, avaliaram a produtividade de três tipos de minicepas em três diferentes volumes de tubetes, em coletas sucessivas. Neste artigo, foram apresentados resultados referentes a três experimentos, cada um referente a uma espécie de minicepas, sendo que, quando os dados não atenderam às pressuposições da anova eles foram transformados. Utilizando a análise de variâncias, sob um DIC, concluíram que houve diferença significativa entre os tratamentos para todos os três experimentos.

Entretanto, nem todos os experimentos produzem dados que atendem às pressuposições da ANOVA, que são aditividade do modelo, homogeneidade, independência e normalidade dos erros. Essa última, sob a teoria de modelos lineares, implica na normalidade dos dados que, por sua vez, não é atendida quando a variável aleatória que descreve um fenômeno segue qualquer distribuição de probabilidade que não seja a normal, por razões intrínsecas ao fenômeno (PINO, 2014).

Muitas pesquisas, nas mais diversas áreas do conhecimento, produzem dados assimétricos e, nesses casos, pode-se atribuir outra distribuição para os erros, uma que caracterize melhor os dados em questão, e em vez de proceder com uma ANOVA, fazer a transformação dos dados em busca da normalidade ou fazer análise não paramétrica, podemos utilizar o método da máxima verossimilhança para estimação dos parâmetros e tirar conclusões acerca da diferença entre os efeitos de tratamentos com um teste de médias.

Segundo Cordeiro e Demétrio (2013), não há razão para se restringir à distribuição normal para o componente aleatório do modelo e à suposição de homogeneidade de variâncias. Com os recursos computacionais que estão a cada dia mais evoluídos, ensaios experimentais que não se

restringem a essas pressuposições já estão sendo feitos e há trabalhos que utilizam especificamente a distribuição *skew*-normal para o erro aleatório do modelo de um DIC.

Guedes et al. (2014) realizaram um experimento para verificar se o crescimento de plantas de *Stevia rebaudiana* (Bert) Bertoni em três tipos de bandejas de isopor com diferentes volumes de células influenciavam na altura dos mesmos. Os autores ajustaram modelos uni e multivariados de regressão a estes dados, supondo as distribuições normal e *skew*-normal para os resíduos sob o DIC. Observando o *boxplot* apresentado no artigo, pode-se perceber que os tratamentos apresentaram assimetria. Os autores apontam que a suposição de normalidade ao ajustar modelos de regressão com erros normais não foi satisfeita, entretanto não mencionam se algum teste foi utilizado. O modelo que melhor ajustou, tanto pela regressão linear quanto pela regressão linear multivariada foi aquele que considerou a *skew*-normal como distribuição para os erros.

Said, Ning e Tian (2018), propuseram um procedimento para detectar alterações em um modelo de regressão linear com erros *skew*-normais baseados em um Critério de Informação Modificado (MIC). Os autores realizaram estudos de simulações para avaliar o desempenho do procedimento proposto e a consistência dos estimadores locais de mudança e, concluíram que a suposição de erro *skew*-normal foi bem sucedida ao ser testada em um banco de dados da Nasa.

Muitos autores baseados no trabalho de Azzalini (1985), consideraram a distribuição *skew*-normal e a aplicaram em diferentes áreas como economia, finanças, oceanografia, engenharia e ciências biomédicas, entre outras, (FERREIRA, BOLFARINE e LABRA, 2018). Resultados complementares a respeito desta distribuição e suas aplicações podem ser encontrados em Azzalini (2005) e Lachos, Bolfarine e Montenegro (2008).

E com isso, o objetivo deste trabalho não é, de forma alguma, se opor a distribuição normal, pois esta é suficientemente apropriada para resolver a maioria dos problemas de experimentação quando os experimentos são planejados e executados conforme algumas literaturas clássicas instruem, por exemplo, Banzatto e Kronka (2006) ou Montgomery (2017). Pretendemos propor uma alternativa para análise de dados quando a distribuição normal não é adequada e, ainda, não se deseja fazer transformação dos dados e nem análise não paramétrica. Neste caso, podemos fazer ajustes de modelos utilizando o método da máxima verossimilhança a dados que não apresentam normalidade e verificar se a distribuição *skew*-normal tem melhor desempenho quando comparada a normal.

Fizemos o estudo dessa metodologia utilizando dois bancos de dados reais não normais que apresentaram indícios de dispersão e assimetria, em que a cada um foi ajustado um modelo com erro *skew*-normal e feita a comparação do ajuste com seus respectivos modelos com erro normal utilizando os critérios de informação de Akaike (AIC), Akaike corrigido (AICc), Bayesiano (BIC) e envelopes simulados dos resíduos quantílicos. Para finalizar, foi realizado um estudo de simulação para avaliar inferência e as propriedades assintóticas dos estimadores de máxima verossimilhança da distribuição *skew*-normal para os modelos ajustados, pois esta apresentou melhor ajuste quando comparada a distribuição normal.

2 Referencial teórico

Nesta seção é apresentado o embasamento teórico que norteia este trabalho. Na Subseção 2.1 trataremos de delineamentos experimentais, na Subseção 2.2 do modelo de média, nas Subseções 2.3 e 2.4 teremos, respectivamente, as distribuições normal e *skew*-normal, na Subseção 2.5 o método da máxima verossimilhança, na Subseção 2.6 os métodos para seleção de modelos e na Subseção 2.7 apresentamos brevemente os resíduos quantílicos.

2.1 Delineamentos experimentais

A experimentação é a área da estatística utilizada, tanto para desenvolver, quanto para aprimorar processos de fabricação, além de ter muitas outras áreas de aplicações, tais como, marketing, operações de serviço e operações gerais de negócios (MONTGOMERY, 2017).

Segundo Banzatto e Kronka (2006), delineamento ou esquema experimental é a estratégia que a experimentação utiliza para que os tratamentos sejam designados às suas unidades experimentais. São exemplos de delineamentos experimentais, o delineamento inteiramente casualizado, o delineamento em blocos casualizados e o delineamento em quadrado latino conhecidos como DIC, DBC e DQL, respectivamente, como veremos nas Seções 2.1.1 a 2.1.3.

2.1.1 Delineamento inteiramente casualizado - DIC

O DIC é o delineamento mais simples entre todos os delineamentos experimentais, sendo utilizado somente em experimentos que possuem homogeneidade das condições experimentais entre as parcelas, também conhecidas como unidades experimentais que são os elementos que recebem os tratamentos.

Esse delineamento experimental utiliza apenas dois dos três princípios básicos da experimentação, repetição e casualização, e os tratamentos são designados às unidades experimentais ou parcelas de forma completamente aleatória, ou seja, por meio de algum tipo de sorteio e sem que haja nenhuma restrição na casualização. A casualização pode ser feita com números iguais ou diferentes de repetições por tratamento, sendo o mais comum utilizar este delineamento com mesmo número de repetições (BANZATTO e KRONKA, 2006). O modelo do delineamento inteiramente casualizado é dado pela seguinte expressão, que possui erro aleatório normal que são independentes e identicamente distribuídos (iid).

$$y_{ij} = m + \tau_i + \epsilon_{ij}, \quad \text{com } \epsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma^2) \quad \text{e } \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \end{cases}$$

em que,

y_{ij} é o valor observado na parcela que recebeu o tratamento i na repetição j ;

m é uma constante comum a todas as unidades experimentais;

τ_i é o efeito do tratamento i ;

ϵ_{ij} é o efeito dos fatores não controlados na parcela que recebeu o tratamento i na repetição j ;

I é o número de tratamentos do experimento;

J é o número de repetições dos tratamentos e $IJ = n$ é o número total de parcelas.

Desse modo, sob estas condições, pode-se demonstrar que Y_{ij} é uma variável aleatória com distribuição $N(m + \tau_i, \sigma^2)$.

2.1.2 Delineamento em blocos casualizados - DBC

Quando em um experimento há a presença de alguma perturbação na homogeneidade das unidades experimentais, e este fator perturbador pode ser controlado utilizando-se o controle local, que é um dos princípios básicos da experimentação, o delineamento em blocos casualizados é utilizado para eliminar sistematicamente o efeito deste fator nas comparações estatísticas entre tratamentos.

Segundo Montgomery (2017), este delineamento reduz o erro residual do experimento, devido a remoção da variabilidade inerente ao fator perturbador que é conhecido e controlável. Além do controle local, este delineamento experimental utiliza os princípios da casualização e repetição e tem modelo dado pela seguinte expressão, com erro aleatório independente e identicamente distribuídos com distribuição normal.

$$y_{ij} = m + \tau_i + \beta_j + \epsilon_{ij}, \quad \epsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma^2) \quad \text{e} \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \end{cases}$$

em que,

y_{ij} é o valor observado na parcela que recebeu o tratamento i no bloco j ;

m é uma constante comum a todas as unidades experimentais;

τ_i é o efeito do tratamento i ;

β_j é o efeito do j -ésimo bloco;

ϵ_{ij} é o efeito dos fatores não controlados na parcela que recebeu o tratamento i no bloco j ;

I é o número de tratamentos do experimento;

J é o número de blocos.

Desse modo, essas condições implicam que Y_{ij} é uma variável aleatória com distribuição $N(m + \tau_i + \beta_j, \sigma^2)$.

2.1.3 Delineamento em quadrado latino - DQL

Este delineamento utiliza o princípio do controle local duas vezes e com isto retira do erro experimental os efeitos de duas fontes de variação conhecidas e controláveis, segundo Montgomery (2017), isto equivale a utilizar o esquema em blocos casualizados sistematicamente em duas direções.

O DQL com I tratamentos, corresponde a um “quadrado” com I linhas e I colunas, contendo I^2 unidades experimentais, em que as linhas e colunas representam duas restrições à casualização.

O modelo estatístico do delineamento em quadrado latino é dado por

$$y_{ijk} = m + \alpha_i + \tau_j + \beta_k + \epsilon_{ijk} \quad \text{com} \quad \epsilon_{ijk} \stackrel{\text{iid}}{\sim} N(0, \sigma^2) \quad \text{e} \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, I \\ k = 1, 2, \dots, I \end{cases}$$

em que,

y_{ijk} é o valor observado na i -ésima linha e k -ésima coluna do j -ésimo tratamento;

m é uma constante comum a todas as unidades experimentais;

α_i é o efeito da i -ésima linha, primeiro controle local;

β_k é o efeito da k -ésima coluna, segundo controle local;

τ_j é o efeito do j -ésimo tratamento;

ϵ_{ijk} é o efeito dos fatores não controlados.

2.2 Modelo de intercepto

Um modelo ainda mais simples que os modelos dos delineamentos apresentados na Seção 2.1 é o *cell means model*, também conhecido como modelo de intercepto ou modelo de média. Diferentemente dos modelos anteriores, este tem todos os efeitos que causam variação concentrados em um único parâmetro, m , ou seja, não há distinção destes fatores. O modelo de intercepto é dado por

$$y_{ij} = m_i + \epsilon_{ij}, \quad \text{com} \quad \epsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma^2) \quad (1)$$

em que, y_{ij} é a i -ésima observação, da j -ésima repetição do evento de interesse; m_i é a i -ésima média em estudo e ϵ_{ij} é o erro associado a observação y_{ij} . Este modelo tem a propriedade de que o número de parâmetros é exatamente igual ao número de células preenchidas, dando origem à natureza de classificação completa das equações normais (SEARLE, 1997).

2.3 Distribuição normal

A distribuição normal, também denominada distribuição gaussiana, é amplamente conhecida e utilizada na estatística. Caracterizada por dois parâmetros, μ e σ^2 , esta é uma distribuição simétrica em torno da média μ e tem gráfico em forma de sino. Mood, Graybill e Boes (1974) definem a distribuição normal conforme a Definição 1.

Definição 1. Uma variável aleatória Y é dita ser normalmente distribuída se sua função densidade de probabilidade for dada por

$$f_Y(y) = f_Y(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \left(\frac{y - \mu}{\sigma} \right)^2 \right] \quad (2)$$

em que, os parâmetros μ e σ^2 são tais que $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}_+^*$ e $y \in \mathbb{R}$.

Qualquer variável aleatória cuja função densidade de probabilidade for dada por (2), é dita ter distribuição normal. Fazendo a transformação $z = \frac{y - \mu}{\sigma}$, então teremos a distribuição conhecida

como normal padrão, $Z \sim N(0,1)$, com função densidade de probabilidade dada por

$$f_Z(z) = f_Z(z|0,1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right).$$

2.3.1 Função geradora de momentos

Segundo Casella e Berger (2010), a função geradora de momentos é uma função que caracteriza uma distribuição e, quando existe, pode-se encontrar todos os momentos de uma variável aleatória utilizando-a.

Seja $Y \sim N(\mu, \sigma^2)$, sua função geradora de momentos é dada por:

$$\begin{aligned} M_Y(t) &= E[e^{Yt}] \\ &= \int_{-\infty}^{+\infty} e^{yt} \cdot f_Y(y) dy = \int_{-\infty}^{+\infty} e^{yt} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \frac{(y-\mu)^2}{\sigma^2}\right] dy \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[yt - \frac{(y-\mu)^2}{2\sigma^2} \right] dy \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{2\sigma^2 yt - y^2 + 2y\mu - \mu^2}{2\sigma^2} \right) dy \\ &= \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \cdot \frac{y^2 - 2\sigma^2 yt - 2y\mu}{\sigma^2}\right) dy \\ &= \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \cdot \frac{y^2 - 2y(\sigma^2 t + \mu)}{\sigma^2}\right] dy, \end{aligned}$$

completando quadrado, tem-se

$$\begin{aligned} M_Y(t) &= \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \cdot \exp\left[\frac{(\sigma^2 t + \mu)^2}{2\sigma^2}\right] \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \cdot \frac{y^2 - 2y(\sigma^2 t + \mu) + (\sigma^2 t + \mu)^2}{\sigma^2}\right] dy \\ &= \exp\left[\frac{-\mu^2 + (\sigma^2 t + \mu)^2}{2\sigma^2}\right] \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \cdot \frac{[y - (\sigma^2 t + \mu)]^2}{\sigma^2}\right] dy \\ &= \exp\left[\frac{-\mu^2 + (\sigma^2 t + \mu)^2}{2\sigma^2}\right] \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \cdot \frac{(y - \mu^*)^2}{\sigma^2}\right] dy, \quad \text{com } \mu^* = \sigma^2 t + \mu, \end{aligned}$$

mas

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \cdot \frac{(y - \mu^*)^2}{\sigma^2}\right] dy = 1$$

e, portanto

$$\begin{aligned}
 M_Y(t) &= \exp\left[\frac{-\mu^2 + (\sigma^2 t + \mu)^2}{2\sigma^2}\right] \\
 &= \exp\left[\frac{-\mu^2 + (\sigma^2 t)^2 + 2\sigma^2 t\mu + \mu^2}{2\sigma^2}\right] \\
 &= \exp\left[\frac{(\sigma^2 t)^2 + 2\sigma^2 t\mu}{2\sigma^2}\right] \\
 &= \exp\left[\frac{2\sigma^2\left(\frac{\sigma^2 t^2}{2} + t\mu\right)}{2\sigma^2}\right] \\
 &= \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right).
 \end{aligned}$$

Assim, temos que

$$\frac{d}{dt}[M_Y(t)] = \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right) \cdot (\mu + t\sigma^2),$$

portanto

$$\begin{aligned}
 E[Y] &= \left. \frac{d}{dt}[M_Y(t)] \right|_{t=0} \\
 &= \mu.
 \end{aligned} \tag{3}$$

Precisamos de $E[Y^2]$ para encontrar $\text{var}(Y)$. Daí, vem que

$$\frac{d^2}{dt^2}[M_Y(t)] = \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right) \cdot (\mu + t\sigma^2)^2 + \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right) \cdot \sigma^2,$$

assim

$$\begin{aligned}
 E[Y^2] &= \left. \frac{d^2}{dt^2}[M_Y(t)] \right|_{t=0} \\
 &= \mu^2 + \sigma^2.
 \end{aligned} \tag{4}$$

Utilizando (3) e (4) teremos a variância de Y , dada por

$$\begin{aligned}
 \text{var}(Y) &= E[Y^2] - [E[Y]]^2 \\
 &= \mu^2 + \sigma^2 - \mu^2 \\
 &= \sigma^2.
 \end{aligned}$$

O coeficiente de assimetria de Y , segundo Mood, Graybill e Boes (1974) é dado por

$$\gamma_1 = \frac{E[(Y - \mu_Y)^3]}{(\sqrt{\sigma_Y^2})^3}. \quad (5)$$

Temos que

$$\begin{aligned} \frac{d^3}{dt^3}[M_Y(t)] &= \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right) \cdot (\mu + t\sigma^2)^3 + \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right) \cdot 2(\mu + t\sigma^2) \cdot \sigma^2 \\ &+ \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right) \cdot (\mu + t\sigma^2) \cdot \sigma^2, \end{aligned}$$

e o valor esperado de Y^3 é dado por

$$\begin{aligned} E[Y^3] &= \left. \frac{d^3}{dt^3}[M_Y(t)] \right|_{t=0} \\ &= \mu^3 + 2\mu\sigma^2 + \mu\sigma^2 \\ &= \mu^3 + 3\mu\sigma^2. \end{aligned} \quad (6)$$

Daí, usando (3), (4) e (6), teremos que

$$\begin{aligned} E[(Y - \mu_Y)^3] &= E[Y^3 - 3Y^2\mu_Y + 3Y\mu_Y^2 - \mu_Y^3], \text{ mas } \mu_Y = \mu \\ &= E[Y^3] - 3\mu E[Y^2] + 3\mu^2 E[Y] - \mu^3 \\ &= \mu^3 + 3\mu\sigma^2 - 3\mu(\mu^2 + \sigma^2) + 3\mu^3 - \mu^3 \\ &= 3\mu\sigma^2 - 3\mu^3 - 3\mu\sigma^2 + 3\mu^3 \\ &= 0, \end{aligned}$$

portanto para a distribuição normal, o coeficiente de assimetria é

$$\gamma_1 = \frac{0}{(\sqrt{\sigma^2})^3} = 0$$

e como esperávamos, o coeficiente de assimetria da distribuição normal é zero, pois esta é uma distribuição simétrica.

2.4 Distribuição *skew-normal*

Azzalini (1985) propôs uma nova classe de distribuições de probabilidades que contém a distribuição normal como caso particular. Esta classe de distribuição depende de um parâmetro de forma λ , sendo este um parâmetro que interfere na assimetria da distribuição.

Definição 2. Se uma variável aleatória Z tem função densidade de probabilidade dada por

$$f_Z(z) = f_Z(z|\lambda) = 2\phi(z)\Phi(\lambda z), \quad z \in \mathbb{R} \quad (7)$$

em que, $\phi(\cdot)$ e $\Phi(\cdot)$ são as funções densidade de probabilidade e de distribuição acumulada da distribuição normal padrão, respectivamente, então dizemos que Z é uma variável aleatória com distribuição *skew-normal* e parâmetro de assimetria λ , denotada por $Z \sim SN(0, 1, \lambda)$ ou $Z \sim SN(\lambda)$.

Segundo Azzalini (1985), a função de distribuição acumulada de (7) é dada por

$$F(z|\lambda) = \Phi(z) - 2T(z, \lambda)$$

em que, $T(h, a)$ é a função T de Owen, estudada por Owen (1956), dada por

$$T(h, a) = \frac{1}{2\pi} \int_0^a \frac{e^{-\frac{1}{2}h^2(1+z^2)}}{1+z^2} dz.$$

Se $\lambda = 0$, então a densidade dada em (7) corresponderá a função densidade de probabilidade da distribuição normal padrão. Note que diferentemente de outros casos, a distribuição normal é um caso particular da *skew-normal*, e não um caso limite, como ocorre com a distribuição de \bar{X}_n , a média amostral de X_1, \dots, X_n , uma amostra aleatória de X , e as condições do teorema central do limite forem satisfeitas.

Para inserir parâmetros de locação e escala na função densidade de probabilidade de Z , será utilizado o Teorema 1.

Teorema 1 (Casella e Berger, 2010; p.108). Se $f(\cdot)$ é uma função densidade de probabilidade com $\mu, \sigma \in \mathbb{R}$, e $\sigma > 0$, então Y é uma variável aleatória com função densidade de probabilidade

$$f_Y(y|\mu, \sigma) = \frac{1}{\sigma} \cdot f\left(\frac{y-\mu}{\sigma}\right)$$

se, e somente se, existe uma variável aleatória Z com função densidade de probabilidade $f(z)$ e $Y = \mu + \sigma Z$.

Sendo assim, se Z é uma variável aleatória com distribuição $SN(0, 1, \lambda)$ e $Y = \mu + \sigma Z$, então a densidade de Y será dada por

$$f(y|\mu, \sigma, \lambda) = \frac{2}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) \Phi\left(\lambda \frac{y-\mu}{\sigma}\right), \quad (8)$$

e com função distribuição acumulada dada por

$$F(y|\mu, \sigma, \lambda) = \Phi\left(\frac{y-\mu}{\sigma}\right) - 2T\left(\frac{y-\mu}{\sigma}, \lambda\right). \quad (9)$$

Dessa forma, se $Z \sim SN(0, 1, \lambda)$, a variável aleatória $Y = \mu + \sigma Z$, com função densidade de probabilidade dada por (8), é definida como tendo distribuição *skew-normal*, com parâmetro de locação

μ , de escala σ e de assimetria λ , e será denotada por $Y \sim SN(\mu, \sigma, \lambda)$.

2.4.1 Função geradora de momentos

Seja Z uma variável aleatória com distribuição *skew-normal*, denotado por $Z \sim SN(0, 1, \lambda)$, ou simplesmente $Z \sim SN(\lambda)$, sua função geradora de momentos é dada por:

$$\begin{aligned} M_Z(t) &= E[e^{tZ}] \\ &= \int_{-\infty}^{+\infty} e^{tz} \cdot f_Z(z) dz = \int_{-\infty}^{+\infty} e^{tz} \cdot 2\phi(z)\Phi(\lambda z) dz \\ &= 2 \int_{-\infty}^{+\infty} e^{tz} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \cdot \Phi(\lambda z) dz = 2 \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(-2tz+z^2)} \cdot \Phi(\lambda z) dz, \end{aligned}$$

completando quadrado teremos

$$\begin{aligned} M_Z(t) &= 2 \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(z-t)^2 + \frac{t^2}{2}} \cdot \Phi(\lambda z) dz \\ &= 2e^{\frac{t^2}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(z-t)^2} \cdot \Phi(\lambda z) dz \end{aligned}$$

fazendo $x = z - t$, então $dx = dz$ e, portanto

$$\begin{aligned} M_Z(t) &= 2e^{\frac{t^2}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} \cdot \Phi[\lambda(x+t)] dx \\ &= 2e^{\frac{t^2}{2}} \int_{-\infty}^{+\infty} \phi(x) \cdot \Phi(\lambda x + \lambda t) dx \\ &= 2e^{\frac{t^2}{2}} \cdot E[\Phi(\lambda X + \lambda t)] \end{aligned} \tag{10}$$

e, para calcular a esperança em (10), será utilizado o Lema 1, que é apresentado sem demonstração.

Lema 1 (Azzalini e Capitanio, 2014; p.26). Se $X \sim N(0, 1)$, então

$$E[\Phi(hX + k)] = \Phi\left(\frac{k}{\sqrt{1+h^2}}\right)$$

com $h, k \in \mathbb{R}$.

Como $X \sim N(0,1)$, fazendo-se $h = \lambda$ e $k = \lambda t$ obteremos o valor esperado, da seguinte forma

$$\begin{aligned} E[\Phi(\lambda X + \lambda t)] &= \Phi\left(\frac{\lambda t}{\sqrt{1+\lambda^2}}\right) \\ &= \Phi\left(\frac{\lambda}{\sqrt{1+\lambda^2}} \cdot t\right). \end{aligned}$$

Retomando a (10), obtemos a função geradora de momentos de $Z \sim SN(0,1,\lambda)$

$$\begin{aligned} M_Z(t) &= 2e^{\frac{t^2}{2}} \cdot E[\Phi(\lambda X + \lambda t)] \\ &= 2e^{\frac{t^2}{2}} \cdot \Phi\left(\frac{\lambda}{\sqrt{1+\lambda^2}} \cdot t\right). \end{aligned}$$

Fazendo $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$, a primeira derivada de $M_Z(t)$ em relação a t é dada por

$$\begin{aligned} \frac{d}{dt}[M_Z(t)] &= 2 \cdot \frac{2t}{2} \cdot e^{\frac{t^2}{2}} \cdot \Phi(\delta t) + 2e^{\frac{t^2}{2}} \cdot \delta \cdot \phi(\delta t) \\ &= 2te^{\frac{t^2}{2}} \cdot \Phi(\delta t) + 2e^{\frac{t^2}{2}} \cdot \delta \cdot \phi(\delta t) \\ &= 2e^{\frac{t^2}{2}} \cdot [t \cdot \Phi(\delta t) + \delta \cdot \phi(\delta t)]. \end{aligned}$$

Para encontrar o primeiro momento, que corresponde a $E[Z] = \mu_Z$, basta avaliar a derivada acima em $t = 0$, como segue

$$\begin{aligned} E[Z] &= \left. \frac{d}{dt}[M_Z(t)] \right|_{t=0} \\ &= 2e^{\frac{0^2}{2}} \cdot [0 \cdot \Phi(\delta \cdot 0) + \delta \cdot \phi(\delta \cdot 0)] \\ &= 2 \cdot \delta \cdot \phi(0) \\ &= 2 \cdot \delta \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{0^2}{2}} \\ &= \delta \cdot \frac{2}{\sqrt{2\pi}} = \delta \sqrt{\frac{2}{\pi}}. \end{aligned} \tag{11}$$

Para encontrar $\text{var}[Z] = \sigma_Z^2$, precisamos de $E[Z^2]$, o qual é obtido pela segunda derivada de $M_Z(t)$ avaliada em $t = 0$, assim

$$\begin{aligned} \frac{d^2}{dt^2}[M_Z(t)] &= \left(2e^{\frac{t^2}{2}} + 2t^2 e^{\frac{t^2}{2}}\right) \cdot \Phi(\delta t) + 2te^{\frac{t^2}{2}} \cdot \phi(\delta t) \cdot \delta \\ &\quad + \delta \cdot 2te^{\frac{t^2}{2}} \cdot \phi(\delta t) + 2e^{\frac{t^2}{2}} \cdot \delta^2 \cdot \phi'(\delta t) \\ &= 2e^{\frac{t^2}{2}} (1+t^2) \Phi(\delta t) + 4\delta te^{\frac{t^2}{2}} \phi(\delta t) + 2\delta^2 e^{\frac{t^2}{2}} \phi'(\delta t) \\ &= 2e^{\frac{t^2}{2}} \left[(1+t^2) \Phi(\delta t) + 2\delta t \phi(\delta t) + \delta^2 \phi'(\delta t) \right], \end{aligned}$$

e avaliando em $t = 0$, tem-se que

$$\begin{aligned}
 E[Z^2] &= \left. \frac{d^2}{dt^2} [M_Z(t)] \right|_{t=0} \\
 &= 2 \cdot \Phi(0) \\
 &= 2 \cdot \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\
 &= 2 \cdot \frac{1}{2} \\
 &= 1
 \end{aligned} \tag{12}$$

e, portanto, a variância de Z é dada por

$$\begin{aligned}
 \text{var}[Z] &= E[Z^2] - [E(Z)]^2 \\
 &= 1 - \left(\delta \cdot \sqrt{\frac{2}{\pi}} \right)^2 \\
 &= 1 - \delta^2 \cdot \frac{2}{\pi}.
 \end{aligned} \tag{13}$$

Temos interesse, também, em encontrar o coeficiente de assimetria de Z , que é dado pela Equação (5). Para encontrar o terceiro momento de Z centrado na média, que corresponde ao numerador de (5), precisaremos de $E[Z^3]$. Sendo assim, será calculado a derivada terceira de $M_Z(t)$ que em seguida será avaliada em $t = 0$, conforme adiante.

$$\begin{aligned}
 \frac{d^3}{dt^3} (M_Z(t)) &= \left(2te^{\frac{t^2}{2}} + 4te^{\frac{t^2}{2}} + 2t^3e^{\frac{t^2}{2}} \right) \cdot \Phi(\delta t) + \left(2e^{\frac{t^2}{2}} + 2t^2e^{\frac{t^2}{2}} \right) \cdot \phi(\delta t) \cdot \delta \\
 &+ \left(2e^{\frac{t^2}{2}} + 2t^2e^{\frac{t^2}{2}} \right) \cdot \phi(\delta t) \cdot \delta + 2te^{\frac{t^2}{2}} \cdot \phi'(\delta t) \cdot \delta^2 \\
 &+ \left(2e^{\frac{t^2}{2}} + 2t^2e^{\frac{t^2}{2}} \right) \cdot \phi(\delta t) \cdot \delta + 2te^{\frac{t^2}{2}} \cdot \phi'(\delta t) \cdot \delta^2 \\
 &+ 2te^{\frac{t^2}{2}} \cdot \phi'(\delta t) \cdot \delta^2 + 2e^{\frac{t^2}{2}} \cdot \phi''(\delta t) \cdot \delta^3,
 \end{aligned} \tag{14}$$

avaliando (14) em $t = 0$, teremos que

$$\left. \frac{d^3}{dt^3} (M_Z(t)) \right|_{t=0} = 3 \cdot 2 \cdot \phi(0) \cdot \delta + 2 \cdot \delta^3 \cdot \phi''(\delta t) \Big|_{t=0},$$

mas

$$\left. \phi''(\delta t) \right|_{t=0} = \frac{1}{\sqrt{2\pi}} e^{-\frac{0^2}{2}} = \frac{1}{\sqrt{2\pi}},$$

e

$$\frac{d^2}{dt^2}(\phi(0)) = -\frac{1}{\sqrt{2\pi}}e^{-\frac{0^2}{2}} + \frac{0^2}{\sqrt{2\pi}}e^{-\frac{0^2}{2}} = -\frac{1}{\sqrt{2\pi}},$$

portanto,

$$\begin{aligned} E[Z^3] &= 3 \cdot \frac{2}{\sqrt{2\pi}} \cdot \delta - \frac{2}{\sqrt{2\pi}} \cdot \delta^3 \\ &= 3 \cdot \sqrt{\frac{2}{\pi}} \cdot \delta - \sqrt{\frac{2}{\pi}} \cdot \delta^3. \end{aligned} \quad (15)$$

Daí, usando os resultados (11), (12) e (15), temos que

$$\begin{aligned} E[(Z - \mu_Z)^3] &= E[Z^3 - 3Z^2\mu_Z + 3Z\mu_Z^2 - \mu_Z^3] \\ &= E[Z^3] - 3\mu_Z E[Z^2] + 3\mu_Z^2 E[Z] - \mu_Z^3 \\ &= E[Z^3] - 3\mu_Z + 2\mu_Z^3 \\ &= 3\delta \cdot \sqrt{\frac{2}{\pi}} - \delta^3 \sqrt{\frac{2}{\pi}} - 3\delta \cdot \sqrt{\frac{2}{\pi}} + 2\delta^3 \left(\sqrt{\frac{2}{\pi}}\right)^3 \\ &= 2\delta^3 \left(\sqrt{\frac{2}{\pi}}\right)^3 - \delta^3 \sqrt{\frac{2}{\pi}} \\ &= \delta^3 \sqrt{\frac{2}{\pi}} \left[2 \left(\sqrt{\frac{2}{\pi}}\right)^2 - 1 \right] \\ &= \delta^3 \sqrt{\frac{2}{\pi}} \left[\frac{4}{\pi} - 1 \right] \\ &= \frac{4 - \pi}{\pi} \cdot \delta^3 \sqrt{\frac{2}{\pi}}. \end{aligned} \quad (16)$$

Utilizando os resultados (13) e (16) em (5), o coeficiente de assimetria de $Z \sim SN(0, 1, \lambda)$ é dado por

$$\begin{aligned} \gamma_1(Z) &= \frac{\frac{4-\pi}{\pi} \cdot \delta^3 \sqrt{\frac{2}{\pi}}}{\left(\sqrt{1 - \delta^2 \cdot \frac{2}{\pi}}\right)^3} \\ &= \frac{4 - \pi}{\pi} \cdot \frac{\delta^3 \sqrt{\frac{2}{\pi}}}{\left(1 - \delta^2 \cdot \frac{2}{\pi}\right)^{\frac{3}{2}}}, \end{aligned}$$

resultado este condizente com Pewsey (2000).

Após algumas manipulações matemáticas em (16), obtém-se o resultado encontrado em

Azzalini e Capitanio (2014). Assim,

$$\begin{aligned}
 E[(Z - \mu_Z)^3] &= \frac{4 - \pi}{\pi} \cdot \delta^3 \sqrt{\frac{2}{\pi}} \cdot \frac{2}{2} \\
 &= \frac{4 - \pi}{2} \cdot \delta^3 \cdot \frac{2}{\pi} \cdot \sqrt{\frac{2}{\pi}} \\
 &= \frac{4 - \pi}{2} \cdot \delta^3 \cdot \left(\sqrt{\frac{2}{\pi}}\right)^2 \cdot \sqrt{\frac{2}{\pi}} \\
 &= \frac{4 - \pi}{2} \cdot \delta^3 \cdot \left(\sqrt{\frac{2}{\pi}}\right)^3 \\
 &= \frac{4 - \pi}{2} \cdot \left(\delta \sqrt{\frac{2}{\pi}}\right)^3
 \end{aligned}$$

logo,

$$\gamma_1(Z) = \frac{4 - \pi}{2} \cdot \frac{\left(\delta \sqrt{\frac{2}{\pi}}\right)^3}{\left(1 - \delta^2 \cdot \frac{2}{\pi}\right)^{\frac{3}{2}}}, \quad \text{em que } \delta = \frac{\lambda}{\sqrt{1 + \lambda^2}}.$$

Tendo em vista que $\gamma_1(Z)$ depende somente de δ , que por sua vez depende somente de λ , é relativamente simples estudar o comportamento do coeficiente de assimetria da distribuição *skew-normal*. Na Figura 1, observa-se que o coeficiente de assimetria está no intervalo $(-1, 1)$, para qualquer valor de $\lambda \in \mathbb{R}$, pois há uma assíntota horizontal em $y = 1$ e outra em $y = -1$. Isto significa que não importa quão grande, ou pequeno, o parâmetro λ seja, a assimetria será próxima de 1 ou -1 , o que tem influência no comportamento da densidade da distribuição *skew-normal*.

Ao estudar as densidades desta distribuição para certos valores empiricamente relevantes do parâmetro de assimetria λ , a fim de analisar o comportamento visual da curva ao variar este parâmetro, $\lambda \in \{-50, -20, -10, -5, -2, 0, 2, 5, 10, 20, 50\}$, é observado que o comportamento é alterado de modo expressivo quando λ está no intervalo $(-10, 10)$. Para valores além destes, não existem evidências, ao menos visuais, de extensas mudanças na forma da densidade, o que se deve ao coeficiente de assimetria desta distribuição, pois este, pertence ao intervalo $(-1, 1)$. Devido a isto, é observado na Figura 2, gráfico (a), que quando $\lambda = 20$ e $\lambda = 50$ as curvas estão mais próximas da densidade para $\lambda = 10$ do que as curvas com assimetria iguais a 0, 2 e 5. O mesmo ocorre para valores negativos de λ na Figura 2, gráfico (b). Segundo Pewsey (2000), para $\lambda = 2$, a densidade é moderadamente assimétrica, enquanto que para $\lambda = 20$ a densidade difere pouco da distribuição *folded-normal* padrão limite.

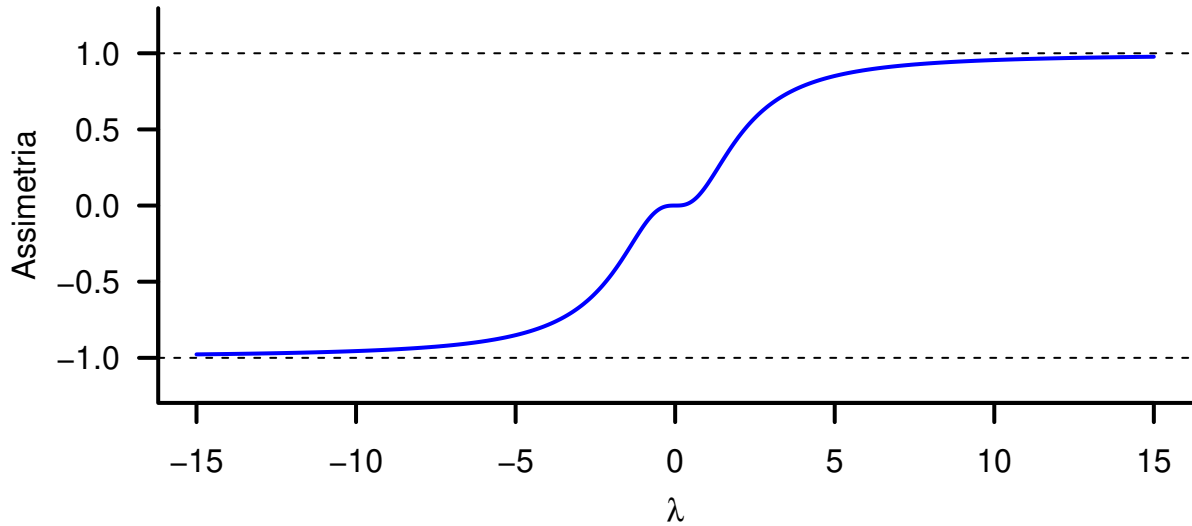


Figura 1: Curva do coeficiente de assimetria da distribuição $SN(\lambda)$, em função de λ .

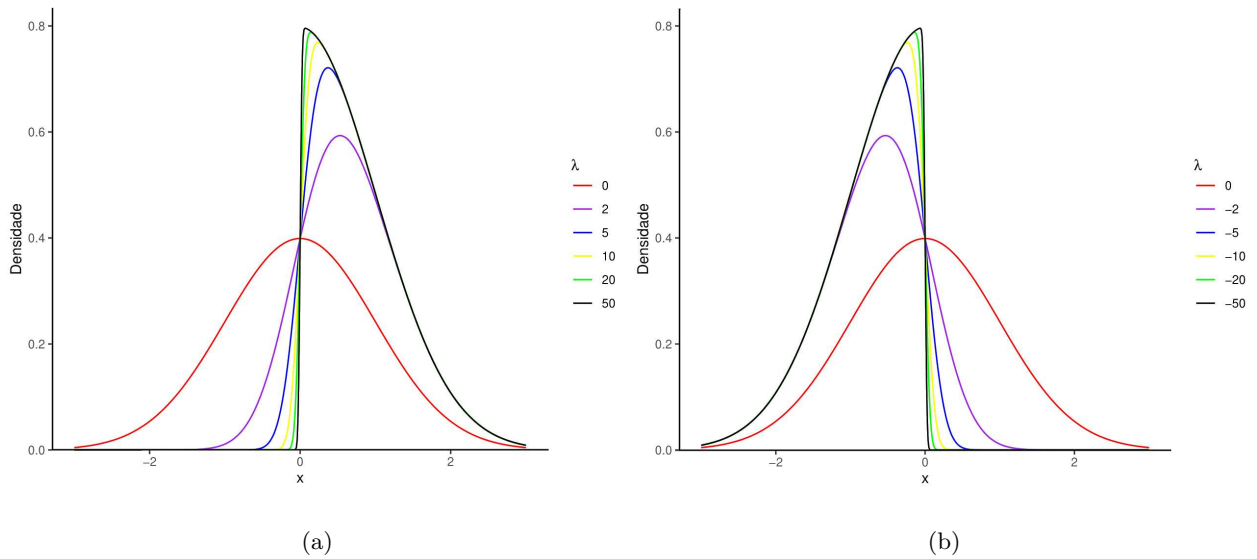


Figura 2: Comportamento da distribuição *skew*-normal de acordo com a variação do parâmetro de assimetria λ .

2.5 Método da máxima verossimilhança

O método da máxima verossimilhança foi proposto por Carl Friedrich Gauss em 1821 e teve suas propriedades estudadas no ano seguinte por Walter Lowrie Fisher. Os estimadores encontrados a partir deste método são, em geral, estimadores consistentes e assintoticamente eficientes, em que

a estimativa obtida é o valor do parâmetro para o qual a amostra observada é mais provável, (CASELLA E BERGER, 2010).

Mood; Graybill e Boes (1974), definem a função de verossimilhança e estimador de máxima verossimilhança como nas Definições 3 e 4 que se seguem.

Definição 3. A função de verossimilhança de n variáveis aleatórias X_1, \dots, X_n é definida como a densidade conjunta das n variáveis aleatórias, $f_{X_1, \dots, X_n}(x_1, \dots, x_n | \theta)$, considerada como função do parâmetro θ . Se X_1, \dots, X_n é uma amostra aleatória da densidade $f(x | \theta)$, então sua função de verossimilhança é

$$f(x_1 | \theta) f(x_2 | \theta) \cdots f(x_n | \theta) = \prod_{i=1}^n f(x_i | \theta),$$

que será denotada por $L(\theta | x_1, \dots, x_n)$.

Definição 4. Seja $L(\theta) = L(\theta | x_1, \dots, x_n)$, a função de verossimilhança das variáveis aleatórias X_1, \dots, X_n . Se $\hat{\theta}$, é o valor de θ em $\bar{\Theta}$, que maximiza a função de verossimilhança $L(\theta)$, então $\hat{\Theta}$ é o estimador de máxima verossimilhança de θ e $\hat{\theta}$ é a estimativa de máxima verossimilhança de θ para a amostra x_1, \dots, x_n , em que,

$\bar{\Theta}$ é o espaço paramétrico associado a θ ;

$\hat{\theta} = \hat{\vartheta}(x_1, \dots, x_n)$ é uma função das observações x_1, \dots, x_n ;

$\hat{\Theta} = \hat{\vartheta}(X_1, \dots, X_n)$ é função das variáveis aleatórias X_1, \dots, X_n , sendo portanto uma estatística.

Segundo Mood; Graybill e Boes (1974), sob certas condições de regularidade, o estimador de máxima verossimilhança (EMV), é a solução da seguinte equação

$$\frac{d}{d\theta} (\log L(\theta)) = \frac{d\ell(\theta)}{d\theta} = 0. \quad (17)$$

A solução da equação (17) é o valor que maximiza a função de verossimilhança. Entretanto, deseja-se que o estimador de máxima verossimilhança seja um máximo global da função de verossimilhança. A primeira derivada de $\ell(\theta)$ igualada a zero é uma condição necessária para obtermos um ponto de máximo, mas não é suficiente para que tenha-se um máximo global, sendo assim, considerando o caso uniparamétrico precisamos ainda verificar se a segunda derivada de $\ell(\theta)$ aplicada em $\theta = \hat{\theta}$ é negativa. Para os casos multiparamétricos precisa-se verificar se a matriz hessiana, que é a matriz de derivadas parciais de segunda ordem, é negativa definida, (CECON et al., 2012).

Segundo Azzalini e Capitanio (2014), sob as condições de regularidade para estimadores eficientes e estatísticas suficientes, em grandes amostras, os estimadores de máxima verossimilhança do vetor de parâmetros $\boldsymbol{\theta}$, em seu espaço paramétrico, é um estimador eficiente e $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ tem distribuição assintótica normal multivariada $N(\mathbf{0}, K(\boldsymbol{\theta})^{-1})$, em que, $K(\boldsymbol{\theta})$ é a matriz de informação esperada. Segundo Eloy et al. (2019), a matriz de covariâncias assintótica $K(\boldsymbol{\theta})^{-1}$ de $\boldsymbol{\theta}$ pode ser aproximada pela inversa da matriz de informação observada $-I(\boldsymbol{\theta})$, de ordem $n \times n$, ou seja, pode-se inverter a matriz de informação observada da função dos parâmetros para obter uma aproximação da matriz de covariâncias. Assim, as estimativas de máxima verossimilhança podem fornecer re-

giões de confiança utilizando a normalidade assintótica. As inferências assintóticas para o vetor de parâmetros $\boldsymbol{\theta}$ podem ser realizadas utilizando a aproximação normal $N(\mathbf{0}, -I(\boldsymbol{\theta})^{-1})$ para $\hat{\boldsymbol{\theta}}$, e os erros-padrão para as estimativas de máxima verossimilhança podem ser obtidas pela raiz quadrada dos elementos da diagonal principal da inversa da matriz de informação observada e podem ser formuladas hipóteses e serem testadas.

Os elementos da matriz de informação observada para um modelo de regressão de um DIC, $(I+2) \times (I+2)$, são dadas por:

$$I(\boldsymbol{\theta}) = \begin{bmatrix} I_{\tau_1\tau_1} & I_{\tau_1\tau_2} & \cdots & I_{\tau_1\tau_I} & I_{\tau_1\sigma} & I_{\tau_1\lambda} \\ \cdot & I_{\tau_2\tau_2} & \cdots & I_{\tau_2\tau_I} & I_{\tau_2\sigma} & I_{\tau_2\lambda} \\ \cdot & \cdot & \ddots & \vdots & \vdots & \vdots \\ \cdot & \cdot & \cdots & I_{\tau_I\tau_I} & I_{\tau_I\sigma} & I_{\tau_I\lambda} \\ \cdot & \cdot & \cdots & \cdot & I_{\sigma\sigma} & I_{\sigma\lambda} \\ \cdot & \cdot & \cdots & \cdot & \cdot & I_{\lambda\lambda} \end{bmatrix}$$

em que, o número de tratamentos $i = 1, \dots, I$ é obtido numericamente. A distribuição normal assintótica $N_{I+2}(\mathbf{0}, -I(\boldsymbol{\theta})^{-1})$ pode ser utilizada para construir regiões aproximadas de confiança para o vetor de parâmetros $\boldsymbol{\theta}$. Neste caso, os intervalos de confiança assintóticos $100(1 - \frac{\alpha}{2})\%$ para cada componente do vetor de parâmetros θ_r é dado por:

$$IC(\theta_r) = \left(\hat{\theta}_r - z_{\alpha/2} \sqrt{-\hat{I}^{r,r}}, \hat{\theta}_r + z_{\alpha/2} \sqrt{-\hat{I}^{r,r}} \right),$$

em que $-\hat{I}^{r,r}$ é o r -ésimo elemento da diagonal de $-I(\hat{\boldsymbol{\theta}})^{-1}$ de $\hat{\boldsymbol{\theta}}$, para $r = 1, \dots, I+2$ e $z_{\alpha/2}$ é o quantil, $1 - \alpha/2$, da distribuição normal padrão.

2.6 Critérios de seleção de modelos

Existem vários métodos para comparar modelos, alguns são baseados na teoria da informação, que é uma teoria usada para quantificar, ou medir, o valor esperado de informação. Segundo Fabozzi et al. (2014), a medida de Kullback e Leibler minimiza a perda de informações para um bom modelo.

O critério de informação Akaike (AIC) proposto por Akaike (1974), é uma medida relativa de qualidade do ajuste obtida por um modelo estatístico, (EMILIANO, VIVANCO e MENEZES, 2014). Este critério para a seleção de modelos foi derivado por Akaike, que estabeleceu uma relação entre a medida de Kullback-Leibler e o método de estimação de máxima verossimilhança e é dado por

$$AIC = -2 \log L(\hat{\boldsymbol{\theta}}) + 2k$$

em que,

$\boldsymbol{\theta}$ é o conjunto (vetor) de parâmetros do modelo;

$L(\hat{\boldsymbol{\theta}})$ é a verossimilhança do modelo candidato quando avaliado no estimador de máxima verossimi-

lança de θ ;

k é o número de parâmetros estimados no modelo candidato.

O critério de informação Akaike de segunda ordem (AICc), também baseado na medida de Kullback-Leibler, deve ser usado em vez do AIC para amostras pequenas, quando $\frac{n}{p} < 40$ de acordo com Burnham e Anderson (2004). O AICc é dado por

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$

em que,

n é o tamanho da amostra,

k é o número de parâmetros estimados no modelo.

Quatro anos após Akaike propor o AIC, Schwarz (1978) propôs uma correção para este critério dando origem ao Critério de Informação Bayesino (BIC), que por sua vez é dado por

$$BIC = -2\log L(\hat{\theta}) + k\log(n),$$

em que, os termos acima são os mesmos apresentados na descrição do AIC.

O modelo que apresenta o menor de seus respectivos valores, é considerado o melhor modelo, dentre os comparados. De acordo com Fabozzi et al. (2014), todos os critérios determinam quão bem os dados suportam cada modelo, levando em consideração a qualidade do ajuste e o número de parâmetros no modelo. A diferença entre eles é que o BIC penaliza um pouco mais pelo número de parâmetros do modelo candidato. Emiliano, Vivanco e Menezes (2014) compararam estes critérios utilizando simulações de Monte Carlo e mostram que há casos em que cada critério se mostra superior, e desta forma aqui serão utilizados os três. Caso ocorra discordância entre eles, utilizaremos aqueles indicados pela maioria.

2.7 Resíduos quantílicos

Os resíduos quantílicos aleatórios, foram apresentados por Dunn e Smyth (1996). Segundo eles, os resíduos quantílicos, rq , são facilmente calculados e permitem que análises residuais sejam realizadas em muitas situações comuns nas quais as definições habituais de resíduos falham. Os autores definem resíduo quantílico por:

$$rq = \Phi^{-1}\{F(y_i | \mu_i, \boldsymbol{\theta})\}$$

em que, $\Phi(\cdot)$ é a função de distribuição acumulada da normal padrão; $F(y_i | \mu_i, \boldsymbol{\theta})$ é a função de distribuição acumulada de $f(y_i | \boldsymbol{\theta})$, que por sua vez é a distribuição de y_1, \dots, y_n observações independentes; $\mu_i = E(y_i)$ e $\boldsymbol{\theta}$ é o vetor de parâmetros comum a todos os y_i .

Assim, se tomarmos y_1, \dots, y_n observações independentes da distribuição *skew-normal*, que

tem função distribuição acumulada dada por (9), cada y_i tem função quantílica estimada dada por

$$\widehat{r}q_{SN} = \Phi^{-1} \left\{ \Phi \left(\frac{y_i - \widehat{\mu}}{\widehat{\sigma}} \right) - 2T \left(\frac{y_i - \widehat{\mu}}{\widehat{\sigma}}, \widehat{\lambda} \right) \right\}.$$

Para a análise dos resíduos, serão construídos os envelopes simulados de confiança, sugerido por Atkinson (1987). Para interpretação do gráfico *half*-normal de probabilidades, Eloy et. al (2019), recomendam os seguintes passos:

1. Ajustar o modelo proposto e calcular os resíduos $\widehat{r}q_{ij}$;
2. Simular n amostras da variável resposta utilizando o modelo ajustado;
3. Ajustar o modelo proposto para cada uma das amostras simuladas e calcular os seus respectivos resíduos simulados $\widehat{r}q_{ij}$, $i = 1, \dots, I$ e $j = 1, \dots, J$;
4. Colocar cada grupo de IJ resíduos em ordem crescente, obtendo $\widehat{r}q_{(ij)n}$;
5. Para cada ij , calcular a média, o mínimo e o máximo dos resíduos $\widehat{r}q_{(ij)n}$;
6. Incluir as médias, os mínimos e os máximos conjuntamente com os valores $\widehat{r}q_{ij}$ contra os quantis esperados da distribuição normal padrão.

O modelo *skew*-normal estará bem ajustado, se os resíduos $\widehat{r}q_{ij}$ apresentarem um comportamento aleatório em torno de zero.

3 Estimação dos parâmetros dos modelos

Os estimadores encontrados a partir do método da máxima verossimilhança (EMV) são, em geral, estimadores consistentes e assintoticamente eficientes (Casella e Berger, 2010). Nesta seção, os parâmetros das distribuições normal e *skew*-normal serão estimados utilizando este método, abrangendo os delineamentos inteiramente casualizado, blocos casualizados e quadrado latino. Iniciaremos com a distribuição normal e estenderemos para a distribuição *skew*-normal que, por sua vez, tem como caso particular a distribuição normal. Aqui, daremos sugestões matemáticas de como fazer ajuste de modelos para os delineamentos já mencionados com erros distribuídos tanto normalmente quanto pela distribuição *skew*-normal.

3.1 EMV para $Y \sim N(\mu, \sigma^2)$

Seja y_1, y_2, \dots, y_n uma amostra aleatória, de tamanho n , de $Y \sim N(\mu, \sigma^2)$. Então, o estimador de máxima verossimilhança do vetor de parâmetros $\boldsymbol{\theta} = (\mu, \sigma^2)^T$ é obtido utilizando a função

de verossimilhança dada por:

$$\begin{aligned}
 L(\boldsymbol{\theta}) &= L(\mu, \sigma^2 | y_1, \dots, y_n) \\
 &= \prod_{i=1}^n f(y_i) \\
 &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - \mu)^2}{2\sigma^2}\right\} \\
 &= (2\pi\sigma^2)^{-\frac{n}{2}} \cdot \exp\left\{\sum_{i=1}^n -\frac{(y_i - \mu)^2}{2\sigma^2}\right\}.
 \end{aligned}$$

A função log-verossimilhança é expressa por:

$$\begin{aligned}
 \ell(\boldsymbol{\theta}) &= \ell(\mu, \sigma^2 | y_1, \dots, y_n) \\
 &= \ln\left\{(2\pi\sigma^2)^{-\frac{n}{2}} \cdot \exp\left[\sum_{i=1}^n -\frac{(y_i - \mu)^2}{2\sigma^2}\right]\right\} \\
 &= -\frac{n}{2} \ln(2\pi\sigma^2) - \sum_{i=1}^n \frac{(y_i - \mu)^2}{2\sigma^2}.
 \end{aligned}$$

As derivadas parciais de $\ell(\boldsymbol{\theta})$, em relação a cada parâmetro do vetor $\boldsymbol{\theta}$, são as componentes função escore $U(\boldsymbol{\theta})$ e são dadas por:

$$\begin{aligned}
 U_{\mu}(\boldsymbol{\theta}) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu} \\
 &= -\frac{2 \sum_{i=1}^n (y_i - \mu)}{2\sigma^2} \cdot (-1) \\
 &= \frac{\sum_{i=1}^n (y_i - \mu)}{\sigma^2} \\
 &= \frac{1}{\sigma^2} \left[\sum_{i=1}^n y_i - n\mu \right]
 \end{aligned}$$

e

$$\begin{aligned}
 U_{\sigma^2}(\boldsymbol{\theta}) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma^2} \\
 &= -\frac{n}{2} \cdot \frac{1}{2\pi\sigma^2} \cdot 2\pi + \frac{\sum_{i=1}^n (y_i - \mu)^2}{2(\sigma^2)^2} \\
 &= -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (y_i - \mu)^2}{2(\sigma^2)^2}.
 \end{aligned}$$

Igualando $U_\mu(\boldsymbol{\theta})$ e $U_{\sigma^2}(\boldsymbol{\theta})$ a zero, encontra-se as expressões matemáticas para os candidatos a estimadores de máxima verossimilhança, de μ e σ^2 , dados respectivamente, por:

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

Para que $\hat{\mu}$ e $\hat{\sigma}^2$ sejam os estimadores de máxima verossimilhança de μ e σ^2 é necessário verificar duas condições: a primeira é se a derivada de segunda ordem em relação a cada um dos parâmetros é menor do que zero e, a segunda é calcular o determinante da matriz hessiana (H), formada pelas derivadas parciais em relação aos parâmetros, e averiguar se o determinante avaliado no estimador é maior que zero, o que garante que os estimadores dos parâmetros são máximos globais. Tal fato ocorre aqui e pode ser visto, por exemplo, em Cecon et al. (2012).

3.2 EMV para $Y \sim N(m + \tau_i, \sigma^2)$

Nesta subseção é apresentada a função de verossimilhança para um delineamento em DIC, cujo modelo é descrito da seguinte forma:

$$y_{ij} = m + \tau_i + \epsilon_{ij}, \quad \text{com} \quad \epsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma^2) \quad \text{e} \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \end{cases}$$

em que, y_{ij} é o valor observado na parcela que recebeu o tratamento i na repetição j ; m é uma constante comum a todas as unidades experimentais; τ_i é o efeito do tratamento i na j -ésima repetição; ϵ_{ij} é o efeito dos fatores não controlados na parcela; I é o número de tratamentos do experimento; J é o número de repetições dos tratamentos e $IJ = n$ é o número de parcelas. Tem-se que a variável aleatória Y_{ij} tem distribuição normal, ou seja, $Y_{ij} \sim N(m + \tau_i, \sigma^2)$. Assim, a função de verossimilhança para uma amostra aleatória y_1, \dots, y_n de tamanho n é dada por:

$$\begin{aligned} L(\boldsymbol{\theta}) &= L(m, \tau_1, \dots, \tau_I, \sigma^2 | y_1, \dots, y_n) \\ &= \prod_{i=1}^I \prod_{j=1}^J f(y_{ij}) \\ &= (2\pi\sigma^2)^{-\frac{IJ}{2}} \cdot \exp \left\{ - \sum_{i=1}^I \sum_{j=1}^J \frac{[y_{ij} - (m + \tau_i)]^2}{2\sigma^2} \right\}, \end{aligned}$$

consequentemente, a função suporte é dada por:

$$\begin{aligned}
\ell(\boldsymbol{\theta}) &= \ell(m, \tau_1, \dots, \tau_I, \sigma^2 | y_1, \dots, y_n) \\
&= \ln[L(m, \tau_1, \dots, \tau_I, \sigma^2 | \mathbf{y})] \\
&= \ln \left\{ (2\pi\sigma^2)^{-\frac{IJ}{2}} \cdot \exp \left[- \sum_{i=1}^I \sum_{j=1}^J \frac{[y_{ij} - (m + \tau_i)]^2}{2\sigma^2} \right] \right\} \\
&= -\frac{IJ}{2} \ln(2\pi\sigma^2) - \frac{\sum_{i=1}^I \sum_{j=1}^J [y_{ij} - (m + \tau_i)]^2}{2\sigma^2}.
\end{aligned}$$

As componentes de $U(\boldsymbol{\theta})$, em relação aos parâmetros, m , τ_i e σ^2 , são expressas por (18) a (20) que se seguem

$$\begin{aligned}
U_m(\boldsymbol{\theta}) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial m} \\
&= \frac{2 \sum_{i=1}^I \sum_{j=1}^J [y_{ij} - (m + \tau_i)]}{2\sigma^2} \cdot (-1) \\
&= \frac{\sum_{i=1}^I \sum_{j=1}^J [y_{ij} - (m + \tau_i)]}{\sigma^2},
\end{aligned} \tag{18}$$

$$\begin{aligned}
U_{\tau_i}(\boldsymbol{\theta}) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \tau_i} \\
&= \frac{2 \sum_{j=1}^J [y_{ij} - (m + \tau_i)]}{2\sigma^2} \cdot (-1) \\
&= \frac{\sum_{j=1}^J [y_{ij} - (m + \tau_i)]}{\sigma^2},
\end{aligned} \tag{19}$$

$$\begin{aligned}
U_{\sigma^2}(\boldsymbol{\theta}) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma^2} \\
&= -\frac{IJ}{2} \cdot \frac{1}{2\pi\sigma^2} \cdot 2\pi + \frac{\sum_{i=1}^I \sum_{j=1}^J [y_{ij} - (m + \tau_i)]^2}{2(\sigma^2)^2} \\
&= -\frac{IJ}{2\sigma^2} + \frac{\sum_{i=1}^I \sum_{j=1}^J [y_{ij} - (m + \tau_i)]^2}{2(\sigma^2)^2}
\end{aligned} \tag{20}$$

primeiro tratamento. Fato evidenciado no Exemplo 1.

Exemplo 1. Considerando um ensaio experimental, em DIC, com três tratamentos ($i = 1, 2, 3$) e duas repetições ($j = 1, 2$), de acordo com o modelo linear de Gauss Markov,

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

sendo sua forma matricial dada por

$$\begin{bmatrix} 5 \\ 4 \\ 9 \\ 10 \\ 21 \\ 22 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{m} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \\ \hat{\tau}_3 \end{bmatrix} + \begin{bmatrix} \hat{\epsilon}_{11} \\ \hat{\epsilon}_{12} \\ \hat{\epsilon}_{21} \\ \hat{\epsilon}_{22} \\ \hat{\epsilon}_{31} \\ \hat{\epsilon}_{32} \end{bmatrix}.$$

Sabe-se, neste modelo, que

$$E[\mathbf{y}] = X\boldsymbol{\beta}, \text{ pois } \epsilon_{ij} \sim N(0, \sigma^2),$$

sendo assim, utilizando o resultado obtido em (21), tem-se:

$$\begin{cases} 5 = \hat{m} + \hat{\tau}_1 \\ 4 = \hat{m} + \hat{\tau}_1 \end{cases} \Rightarrow 9 = 2\hat{m} + 2\hat{\tau}_1$$

$$\begin{cases} 9 = \hat{m} + \hat{\tau}_2 \\ 10 = \hat{m} + \hat{\tau}_2 \end{cases} \Rightarrow 19 = 2\hat{m} + 2\hat{\tau}_2$$

$$\begin{cases} 21 = \hat{m} + \hat{\tau}_3 \\ 22 = \hat{m} + \hat{\tau}_3 \end{cases} \Rightarrow 43 = 2\hat{m} + 2\hat{\tau}_3.$$

Daí, quando $\tau_1 = 0$, o vetor estimado $\hat{\boldsymbol{\beta}}$ é dado por:

$$\hat{\boldsymbol{\beta}} = [\hat{m} \quad \hat{\tau}_1 \quad \hat{\tau}_2 \quad \hat{\tau}_3]^T = [4,5 \quad 0 \quad 5 \quad 17]^T$$

que é exatamente o que encontra-se com o estimador de máxima verossimilhança de τ_i ao utilizar a

restrição $\tau_1 = 0$, como deduzido pela Equação (21). De fato,

$$\begin{aligned}\hat{m} &= \frac{\sum_{j=1}^J y_{1j}}{J} = \frac{5+4}{2} = 4,5 = \bar{y}_{1j} \\ \hat{\tau}_2 &= \frac{\sum_{j=1}^J y_{2j} - \sum_{j=1}^J y_{1j}}{J} = \frac{(9+10)}{2} - \frac{(5+4)}{2} = 9,5 - 4,5 = 5 = \bar{y}_{2j} - \bar{y}_{1j} \\ \hat{\tau}_3 &= \frac{\sum_{j=1}^J y_{3j} - \sum_{j=1}^J y_{1j}}{J} = \frac{(21+22)}{2} - \frac{(5+4)}{2} = 21,5 - 4,5 = 17 = \bar{y}_{3j} - \bar{y}_{1j},\end{aligned}$$

que corresponde exatamente a saída da função `summary.lm` ao fazer o ajuste de um DIC utilizando a função `aov` no *software* R (2020), como na Figura 3.

```
> FT <- relevel(tra,1)
> mod1 <- aov(prod ~ FT, data = dados)
> summary.lm(mod1)
Coefficients:
(Intercept)  4.5000  0.5000  9.000 0.002896 **
FTB          5.0000  0.7071  7.071 0.005816 **
FTC         17.0000  0.7071  24.042 0.000158 ***
---
> dados
tra r prod
1  A 1  5
2  A 2  4
3  B 1  9
4  B 2 10
5  C 1 21
6  C 2 22
> Intercept <- (5+4)/2
> Intercept
[1] 4.5
> FTB <- (9+10)/2 - Intercept
> FTB
[1] 5
> FTC <- (21+22)/2 - Intercept
> FTC
[1] 17
```

Figura 3: Saída da função `summary.lm` do ajuste de um DIC, com três tratamentos e duas repetições, feito pela função `aov` no *software* R.

Ainda na Figura 3, pode-se notar que o *intercept* corresponde a média do tratamento A. Isso ocorre porque a função `relevel` está fazendo a restrição em relação ao primeiro tratamento, ou seja $\tau_1 = 0$, todavia pode-se fazer a restrição em relação a qualquer tratamento, ou seja qualquer $\tau_i = 0$, que as médias dos demais tratamentos serão comparadas com a média deste.

Para avaliar as propriedades desejáveis, tais como viés e consistência, de $\hat{\tau}_i$, será calculado o caso geral, para qualquer i , do seu valor esperado e erro quadrático médio. Sendo assim, considerando o estimador de máxima verossimilhança de τ_i , para um $i' > i''$, com $\tau_i'' = 0$, teremos que

$$\hat{\tau}_i = \bar{y}_{i'j} - \bar{y}_{i''j}, \quad \text{com } i' > i'',$$

aplicando a esperança, vem que

$$\begin{aligned}
E(\hat{\tau}_i) &= E(\bar{y}_{i'j} - \bar{y}_{i''j}), \quad \text{com } i' > i'' \\
&= E(\bar{y}_{i'j}) - E(\bar{y}_{i''j}) \\
&= E\left(\frac{\sum_{j=1}^J y_{i'j}}{J}\right) - E\left(\frac{\sum_{j=1}^J y_{i''j}}{J}\right) \\
&= \frac{1}{J} \left[\sum_{j=1}^J E(y_{i'j}) - \sum_{j=1}^J E(y_{i''j}) \right], \quad \text{mas } y_{ij} \sim N(m + \tau_i; \sigma^2) \\
&= \frac{1}{J} [J(m + \tau_{i'}) - J(m + \tau_{i''})] \\
&= (m + \tau_{i'}) - (m + \tau_{i''}) \\
&= \mu_{i'} - \mu_{i''} \\
&= \tau_i.
\end{aligned}$$

em que, $\mu_{i'} = m + \tau_{i'}$ e $\mu_{i''} = m + \tau_{i''}$ são, respectivamente, as médias dos tratamentos i' e i'' , e $\tau_i = \mu_{i'} - \mu_{i''}$, é essa diferença entre a média dos tratamentos i' e i'' devido à restrição imposta. Portanto, $\hat{\tau}_i$ é um estimador não viesado de τ_i e, além disso, seu erro quadrático médio é igual a variância do estimador de τ_i . Desse modo,

$$\begin{aligned}
\text{var}(\hat{\tau}_i) &= \text{var}(\bar{y}_{i'j} - \bar{y}_{i''j}) \\
&= \text{var}\left(\frac{\sum_{j=1}^J y_{i'j}}{J}\right) + \text{var}\left(\frac{\sum_{j=1}^J y_{i''j}}{J}\right) \\
&= \frac{\sum_{j=1}^J \text{var}[y_{i''j}] + \sum_{j=1}^J \text{var}[y_{i'j}]}{J^2}, \quad \text{mas } y_{ij} \sim N(m + \tau_i; \sigma^2) \\
&= \frac{J\sigma^2 + J\sigma^2}{J^2} \\
&= \frac{2\sigma^2}{J}, \tag{22}
\end{aligned}$$

e como $\lim_{J \rightarrow \infty} \text{var}(\hat{\tau}_i) = 0$, $\hat{\tau}_i$ é um estimador consistente e não viesado.

Utilizando o estimador da variância, $\hat{\sigma}^2$, na expressão (22) tem-se o estimador da variância do estimador de τ_i , $\widehat{\text{var}}(\hat{\tau}_i)$, que será utilizado com o intuito de construir intervalos de confiança e testar hipóteses. O teste de hipóteses para τ_i é dado por:

$$\begin{cases} H_0 & : \tau_i = 0 \\ H_1 & : \tau_i \neq 0 \end{cases}$$

Segundo Cecon et al. (2012), a estatística T_c , do teste t de Student sob H_0 verdadeira é dada por

$$T_c = \frac{\hat{\tau}_i - \tau_i}{\sqrt{\widehat{\text{var}}(\hat{\tau}_i)}} = \frac{\bar{y}_{i'j} - \bar{y}_{i''j}}{\sqrt{\frac{2\hat{\sigma}^2}{J}}}$$

em que, $T_c \sim t_{n-1}$, $n-1$ são os graus de liberdade desta distribuição e $n = IJ$. Neste teste, rejeita-se a hipótese H_0 , se $|t_c| > |t_{tab}|$ ou se o *valor-p* $= 2P[T_c > |t| | H_0]$ for menor ou igual ao nível de significância α , em que t_c significa t calculado e t_{tab} , tabelado. Caso contrário, não se rejeita a hipótese H_0 .

Muitas vezes o pesquisador tem interesse em estimar a diferença entre as médias populacionais dos tratamentos. Isto pode ser feito utilizando o seguinte intervalo de confiança para τ_i :

$$IC_{1-\frac{\alpha}{2}}(\tau_i) = \left(\bar{y}_{i'j} - \bar{y}_{i''j} \pm t_{(\frac{\alpha}{2}, n-1)} \sqrt{\frac{2\hat{\sigma}^2}{J}} \right)$$

em que $t_{(\frac{\alpha}{2}, \nu)}$ é o quantil da distribuição t de Student com $\nu = n-1$ e ao nível de confiança $100(1 - \frac{\alpha}{2})\%$.

Podemos utilizar o intervalo de confiança anteriormente obtido a fim de testar as hipóteses

$$\begin{cases} H_0 & : \tau_i = 0 \\ H_1 & : \tau_i \neq 0 \end{cases}$$

e rejeitaremos H_0 se o intervalo de confiança não contiver o zero.

A distribuição normal é uma das poucas densidades, em que é possível encontrar os estimadores de máxima verossimilhança analiticamente. Para a distribuição *skew-normal*, apresentada a seguir, encontramos somente o vetor de funções escore e as estimativas dos parâmetros podem ser encontrados somente através de métodos numéricos.

3.3 EMV para $Y \sim SN(\mu, \sigma, \lambda)$

Considere uma amostra aleatória de $Y \sim SN(\mu, \sigma, \lambda)$, de tamanho $n = 1$. A função de verossimilhança de Y é dada por:

$$\begin{aligned} L(\boldsymbol{\theta}) &= L(\mu, \sigma, \lambda | y) \\ &= \frac{2}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right) \Phi\left(\lambda \frac{y - \mu}{\sigma}\right), \quad \text{com } \boldsymbol{\theta} = (\mu, \sigma, \lambda), \end{aligned}$$

e função suporte dada por:

$$\begin{aligned}
\ell(\boldsymbol{\theta}) &= \ln L(\boldsymbol{\theta}|y) \\
&= \ln \left[\frac{2}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) \Phi\left(\lambda \frac{y-\mu}{\sigma}\right) \right] \\
&= \ln 2 - \ln \sigma + \ln \phi\left(\frac{y-\mu}{\sigma}\right) + \ln \Phi\left(\lambda \frac{y-\mu}{\sigma}\right) \\
&= \ln 2 - \ln \sigma + \ln \phi(z) + \ln \Phi(\lambda z), \quad \text{com } z = \frac{y-\mu}{\sigma} \\
&= \ln 2 - \ln \sigma + \ln \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \right) + \ln \Phi(\lambda z) \\
&= \ln 2 - \ln \sigma + \ln \left(\frac{1}{\sqrt{2\pi}} \right) - \frac{z^2}{2} + \ln \Phi(\lambda z), \quad \text{com } k = \ln 2 + \ln \left(\frac{1}{\sqrt{2\pi}} \right) \\
&= k - \ln \sigma - \frac{z^2}{2} + \ln \Phi(\lambda z).
\end{aligned}$$

As derivadas parciais em relação a cada parâmetro são as componentes do vetor escore $U(\boldsymbol{\theta})$ e são obtidas a seguir.

Derivando-se em relação a μ teremos

$$\begin{aligned}
U_{\mu}(\boldsymbol{\theta}) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu} = -\frac{2z}{2} \cdot \left(-\frac{1}{\sigma}\right) + \left(-\frac{1}{\sigma}\right) \cdot \lambda \cdot \phi(\lambda z) \cdot \frac{1}{\Phi(\lambda z)} \\
&= \frac{z}{\sigma} - \frac{\lambda}{\sigma} \cdot \frac{\phi(\lambda z)}{\Phi(\lambda z)},
\end{aligned}$$

em relação a σ :

$$\begin{aligned}
U_{\sigma}(\boldsymbol{\theta}) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma} = -\frac{1}{\sigma} - \frac{2z}{2} \cdot (y-\mu) \cdot \left(-\frac{1}{\sigma^2}\right) + (y-\mu) \cdot \left(-\frac{1}{\sigma^2}\right) \cdot \lambda \cdot \frac{\phi(\lambda z)}{\Phi(\lambda z)} \\
&= -\frac{1}{\sigma} + \frac{z}{\sigma} \cdot \frac{y-\mu}{\sigma} - \frac{1}{\sigma} \cdot \frac{y-\mu}{\sigma} \cdot \lambda \cdot \frac{\phi(\lambda z)}{\Phi(\lambda z)} \\
&= -\frac{1}{\sigma} + \frac{z^2}{\sigma} - \frac{\lambda z}{\sigma} \cdot \frac{\phi(\lambda z)}{\Phi(\lambda z)},
\end{aligned}$$

finalmente, em relação a λ :

$$U_{\lambda}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \lambda} = z \cdot \frac{\phi(\lambda z)}{\Phi(\lambda z)}.$$

Estendendo para uma amostra aleatória de tamanho n , considere y_1, \dots, y_n amostrados de $Y \sim SN(\mu, \sigma, \lambda)$. Por definição, sua função de verossimilhança é:

$$L(\boldsymbol{\theta}|y) = \prod_{i=1}^n f(y_i)$$

com função suporte dada por:

$$\begin{aligned}
\ell(\boldsymbol{\theta}) &= \ln \prod_{i=1}^n f(y_i) = \sum_{i=1}^n \ln f(y_i) \\
&= \sum_{i=1}^n \ln \left[\frac{2}{\sigma} \phi \left(\frac{y_i - \mu}{\sigma} \right) \Phi \left(\lambda \frac{y_i - \mu}{\sigma} \right) \right] \\
&= n \cdot \ln 2 - n \cdot \ln \sigma + \sum_{i=1}^n \ln \phi \left(\frac{y_i - \mu}{\sigma} \right) + \sum_{i=1}^n \ln \Phi \left(\lambda \frac{y_i - \mu}{\sigma} \right) \\
&= n \cdot \ln 2 - n \cdot \ln \sigma - n \cdot \ln(\sqrt{2\pi}) - \sum_{i=1}^n \frac{(y_i - \mu)^2}{2\sigma^2} + \sum_{i=1}^n \ln \Phi \left(\lambda \frac{y_i - \mu}{\sigma} \right) \\
&= k - n \cdot \ln \sigma - \sum_{i=1}^n \frac{(y_i - \mu)^2}{2\sigma^2} + \sum_{i=1}^n \ln \Phi \left(\lambda \frac{y_i - \mu}{\sigma} \right) \\
&= k - n \cdot \ln \sigma - \sum_{i=1}^n \frac{z_i^2}{2} + \sum_{i=1}^n \ln \Phi(\lambda z_i), \tag{23}
\end{aligned}$$

com $z_i = \frac{y_i - \mu}{\sigma}$ e $k = n \left[\ln 2 - \ln(\sqrt{2\pi}) \right]$.

As componentes do vetor escore, $U(\boldsymbol{\theta})$, são expressas por:

$$\begin{aligned}
U_\mu(\boldsymbol{\theta}) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu} \\
&= -\frac{2 \sum_{i=1}^n z_i}{2} \cdot \left(-\frac{1}{\sigma} \right) + \sum_{i=1}^n \frac{1}{\Phi(\lambda z_i)} \cdot \phi(\lambda z_i) \cdot \frac{\lambda}{\sigma} \cdot (-1) \\
&= \frac{\sum_{i=1}^n z_i}{\sigma} - \frac{\lambda}{\sigma} \cdot \sum_{i=1}^n \frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)} \tag{24}
\end{aligned}$$

$$\begin{aligned}
U_\sigma(\boldsymbol{\theta}) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma} \\
&= -\frac{n}{\sigma} - \frac{2 \sum_{i=1}^n z_i \cdot (y_i - \mu) \cdot \frac{(-1)}{\sigma^2}}{2} + \sum_{i=1}^n \frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)} \cdot \frac{\lambda(y_i - \mu)}{\sigma^2} \cdot (-1) \\
&= -\frac{n}{\sigma} + \frac{\sum_{i=1}^n z_i^2}{\sigma} - \frac{\lambda}{\sigma} \sum_{i=1}^n z_i \cdot \frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)} \tag{25}
\end{aligned}$$

$$\begin{aligned}
U_\lambda(\boldsymbol{\theta}) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \lambda} \\
&= \sum_{i=1}^n z_i \cdot \frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)}.
\end{aligned}$$

Fazendo-se $U_\mu(\boldsymbol{\theta})$, $U_\sigma(\boldsymbol{\theta})$ e $U_\lambda(\boldsymbol{\theta})$ iguais a zero, teremos o seguinte sistema

$$\left\{ \begin{array}{l} \frac{\sum_{i=1}^n z_i}{\hat{\sigma}} - \frac{\hat{\lambda}}{\hat{\sigma}} \cdot \sum_{i=1}^n \frac{\phi(\hat{\lambda}z_i)}{\Phi(\hat{\lambda}z_i)} = 0 \\ -\frac{n}{\hat{\sigma}} + \frac{\sum_{i=1}^n z_i^2}{\hat{\sigma}} - \frac{\hat{\lambda}}{\hat{\sigma}} \sum_{i=1}^n z_i \frac{\phi(\hat{\lambda}z_i)}{\Phi(\hat{\lambda}z_i)} = 0 \\ \sum_{i=1}^n z_i \cdot \frac{\phi(\hat{\lambda}z_i)}{\Phi(\hat{\lambda}z_i)} = 0 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} \sum_{i=1}^n z_i - \hat{\lambda} \cdot \sum_{i=1}^n \frac{\phi(\hat{\lambda}z_i)}{\Phi(\hat{\lambda}z_i)} = 0 \\ \sum_{i=1}^n z_i^2 - \hat{\lambda} \sum_{i=1}^n z_i \frac{\phi(\hat{\lambda}z_i)}{\Phi(\hat{\lambda}z_i)} = n \\ \sum_{i=1}^n z_i \cdot \frac{\phi(\hat{\lambda}z_i)}{\Phi(\hat{\lambda}z_i)} = 0 \end{array} \right.$$

Observe que essas equações são não lineares nos parâmetros devido a não linearidade de $\frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)}$ nas expressões e, conseqüentemente, os estimadores de máxima verossimilhança só poderão ser obtidos por métodos numéricos.

A função *optim* do *software* R é utilizada para minimizar ou maximizar funções segundo seis métodos de otimizações, segundo Nelder e Mead (1965), o método “Nelder-Mead”, que é o *default* da função *optim*, é um método que depende da comparação de valores de função nos vértices $(4n - 1)$ de um simplex geral, seguido pela substituição do vértice pelo valor mais alto por outro ponto. Segundo Nocedal e Wright (2006), o método “BFGS” é o algoritmo quase Newton mais popular. O método “CG”, gradientes conjugados, é um método iterativo baseado em Fletcher e Reeves (1964), e segundo Hestenes e Stiefel (1952), usado para resolver sistemas de equações lineares. O método “L-BFGS-B”, segundo Byrd et al. (1996), é descrito como um algoritmo para resolver grandes problemas de otimização não linear com limites simples, é baseado no método de projeção em gradiente e pode ser aplicado quando a matriz Hessiana não for prática para calcular, pois o algoritmo não requer segundas derivadas ou conhecimento da estrutura da função objetivo. O método “SANN” é uma das variações dos algoritmos de anulação simulada dada em Bélisle (1992), que por sua vez, pertence à classe de métodos estocásticos de otimização global e que usa apenas valores de função, o que o torna relativamente lento. E o método de “Brent” é apenas para problemas unidimensionais.

Neste estudo, será utilizada a função *optim* do *software* R para maximizar a função log-verossimilhança dada em (23). Ademais, as estimativas dos parâmetros e seus, respectivos, erros padrão serão obtidos pelos métodos “L-BFGS-B” ou “Nelder-Mead”. As funções e scripts deste estudo são encontrados no Apêndice A. Decorre assim, que para estimar os parâmetros da distribuição $SN(\mu, \sigma, \lambda)$, primeiramente, serão estimados os parâmetros μ e σ^2 da distribuição $N(\mu, \sigma^2)$, visto que tal distribuição é um submodelo da *skew-normal*. Dessa forma, supondo $\lambda = 0$, nas Equações (24) e (25) têm-se os valores iniciais a serem utilizados como chutes iniciais no método numérico iterativo dados por:

$$\sum_{i=1}^n z_i = 0 \Rightarrow \sum_{i=1}^n \frac{(y_i - \hat{\mu})}{\hat{\sigma}} = 0 \Rightarrow \hat{\mu} = \frac{\sum_{i=1}^n y_i}{n}$$

e também,

$$\sum_{i=1}^n z_i^2 = n \Rightarrow \sum_{i=1}^n \frac{(y_i - \hat{\mu})^2}{\hat{\sigma}^2} = n \Rightarrow \hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{\mu})^2}{n}.$$

Para estimar o parâmetro λ , será utilizado o método de tentativa e erro, iniciando com valores iniciais baixos, próximos ao número um, no caso do parâmetro de assimetria ser pequeno a ponto de a distribuição *skew-normal* estimar igual a distribuição normal, prosseguir com valores maiores que um a fim de encontrar um estimativa para $\lambda \leq 10$, e em seguida com valores para $\lambda \geq -10$, até encontrar uma estimativa que esteja no intervalo $(-10, 10)$ conforme visto na Seção 2.4.1.

3.4 EMV para $Y \sim SN(m + \tau_i, \sigma, \lambda)$

Nesta subseção é apresentada a função de verossimilhança para um delineamento em DIC considerando ϵ_{ij} com distribuição *skew-normal*. A densidade e a demonstração de que $Y_{ij} \sim SN(m + \tau_i, \sigma, \lambda)$ quando $\epsilon_{ij} \sim SN(0, 1, \lambda)$, se encontra no Apêndice B.

O modelo de um DIC com a adequação, na qual foi acrescentado o parâmetro σ de escala, é expresso da seguinte forma:

$$y_{ij} = m + \tau_i + \sigma \epsilon_{ij}, \text{ com } \epsilon_{ij} \sim SN(0, 1, \lambda).$$

A distribuição *skew-normal* com parâmetros de posição, escala e forma, respectivamente, $m + \tau_i$, σ , λ tem sua função densidade de probabilidade expressa por:

$$f(y_{ij}) = \frac{2}{\sigma} \cdot \phi\left(\frac{y_{ij} - (m + \tau_i)}{\sigma}\right) \cdot \Phi\left(\lambda \frac{y_{ij} - (m + \tau_i)}{\sigma}\right)$$

ou seja,

$$Y_{ij} \sim SN(m + \tau_i, \sigma, \lambda)$$

com esperança e variância dadas por

$$\begin{aligned} E[Y_{ij}] &= E[m + \tau_i + \sigma \epsilon_{ij}] & \text{var}[Y_{ij}] &= \text{var}[m + \tau_i + \sigma \epsilon_{ij}] \\ &= m + \tau_i + \sigma E[\epsilon_{ij}] & &= \sigma^2 \text{var}[\epsilon_{ij}] \\ &= m + \tau_i + \frac{\sigma \lambda}{\sqrt{\lambda^2 + 1}} \cdot \sqrt{\frac{2}{\pi}} & &= \sigma^2 - \frac{\sigma^2 \lambda^2}{\lambda^2 + 1} \cdot \frac{2}{\pi}, \end{aligned}$$

o que nos permitirá determinar a função de verossimilhança e os estimadores de máxima verossimilhança de $\theta = (m, \tau_1, \dots, \tau_I, \sigma, \lambda)$.

Seja y_1, \dots, y_n uma amostra aleatória de tamanho n oriunda de $Y_{ij} \sim SN(m + \tau_i, \sigma, \lambda)$.

Sua função de verossimilhança é dada por

$$L(\boldsymbol{\theta}|y_{ij}) = \prod_{i=1}^I \prod_{j=1}^J f(y_{ij}),$$

e a função suporte é:

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= \ln \left[\prod_{i=1}^I \prod_{j=1}^J f(y_{ij}) \right] = \sum_{i=1}^I \sum_{j=1}^J \ln f(y_{ij}) \\ &= \sum_{i=1}^I \sum_{j=1}^J \ln \left[\frac{2}{\sigma} \phi \left(\frac{y_{ij} - (m + \tau_i)}{\sigma} \right) \Phi \left(\lambda \frac{y_{ij} - (m + \tau_i)}{\sigma} \right) \right] \\ &= IJ \cdot \ln 2 - IJ \cdot \ln \sigma + \sum_{i=1}^I \sum_{j=1}^J \ln \phi \left(\frac{y_{ij} - (m + \tau_i)}{\sigma} \right) + \sum_{i=1}^I \sum_{j=1}^J \ln \Phi \left(\lambda \frac{y_{ij} - (m + \tau_i)}{\sigma} \right) \\ &= IJ \cdot \ln 2 - IJ \cdot \ln \sigma - IJ \cdot \ln (\sqrt{2\pi}) - \sum_{i=1}^I \sum_{j=1}^J \frac{[y_{ij} - (m + \tau_i)]^2}{2\sigma^2} + \sum_{i=1}^I \sum_{j=1}^J \ln \Phi \left(\lambda \frac{y_{ij} - (m + \tau_i)}{\sigma} \right) \\ &= k - IJ \cdot \ln \sigma - \sum_{i=1}^I \sum_{j=1}^J \frac{[y_{ij} - (m + \tau_i)]^2}{2\sigma^2} + \sum_{i=1}^I \sum_{j=1}^J \ln \Phi \left(\lambda \frac{y_{ij} - (m + \tau_i)}{\sigma} \right) \\ &= k - IJ \cdot \ln \sigma - \sum_{i=1}^I \sum_{j=1}^J \frac{z_{ij}^2}{2} + \sum_{i=1}^I \sum_{j=1}^J \ln \Phi(\lambda z_{ij}), \end{aligned} \quad (26)$$

com $z_{ij} = \frac{y_{ij} - (m + \tau_i)}{\sigma}$ e $k = IJ \left[\ln 2 - \ln (\sqrt{2\pi}) \right]$.

As componentes do vetor escore, $U(\boldsymbol{\theta})$, que são as derivadas parciais da função suporte em relação a cada parâmetro, são expressas por:

$$\begin{aligned} U_m(\boldsymbol{\theta}) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial m} \\ &= -\frac{2 \sum_{i=1}^I \sum_{j=1}^J z_{ij}}{2} \cdot \left(-\frac{1}{\sigma} \right) + \sum_{i=1}^I \sum_{j=1}^J \frac{1}{\Phi(\lambda z_{ij})} \cdot \phi(\lambda z_{ij}) \cdot \frac{\lambda}{\sigma} \cdot (-1) \\ &= \frac{\sum_{i=1}^I \sum_{j=1}^J z_{ij}}{\sigma} - \frac{\lambda}{\sigma} \cdot \sum_{i=1}^I \sum_{j=1}^J \frac{\phi(\lambda z_{ij})}{\Phi(\lambda z_{ij})}, \end{aligned}$$

em relação a τ_i :

$$\begin{aligned} U_{\tau_i}(\boldsymbol{\theta}) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \tau_i} \\ &= -\frac{2 \sum_{j=1}^J z_{ij}}{2} \cdot \left(-\frac{1}{\sigma}\right) + \sum_{j=1}^J \frac{1}{\Phi(\lambda z_{ij})} \cdot \phi(\lambda z_{ij}) \cdot \frac{\lambda}{\sigma} \cdot (-1) \\ &= \frac{\sum_{j=1}^J z_{ij}}{\sigma} - \frac{\lambda}{\sigma} \cdot \sum_{j=1}^J \frac{\phi(\lambda z_{ij})}{\Phi(\lambda z_{ij})}, \end{aligned}$$

em relação a σ :

$$\begin{aligned} U_{\sigma}(\boldsymbol{\theta}) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma} \\ &= -\frac{IJ}{\sigma} - \frac{2 \sum_{i=1}^I \sum_{j=1}^J \left(z_{ij} \cdot [y_{ij} - (m + \tau_i)] \cdot \frac{(-1)}{\sigma^2} \right)}{2} + \sum_{i=1}^I \sum_{j=1}^J \frac{\phi(\lambda z_{ij})}{\Phi(\lambda z_{ij})} \cdot \lambda \frac{(y_{ij} - (m + \tau_i))}{\sigma^2} \cdot (-1) \\ &= -\frac{IJ}{\sigma} + \frac{\sum_{i=1}^I \sum_{j=1}^J z_{ij}^2}{\sigma} - \frac{\lambda}{\sigma} \sum_{i=1}^I \sum_{j=1}^J z_{ij} \cdot \frac{\phi(\lambda z_{ij})}{\Phi(\lambda z_{ij})} \end{aligned}$$

e finalmente, em relação a λ , tem-se:

$$\begin{aligned} U_{\lambda}(\boldsymbol{\theta}) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \lambda} \\ &= \sum_{i=1}^I \sum_{j=1}^J z_{ij} \cdot \frac{\phi(\lambda z_{ij})}{\Phi(\lambda z_{ij})}. \end{aligned}$$

Igualando as funções $U_m(\boldsymbol{\theta})$, $U_{\tau_i}(\boldsymbol{\theta})$, $U_{\sigma}(\boldsymbol{\theta})$, $U_{\lambda}(\boldsymbol{\theta})$ a zero e, resolvendo-as, simultaneamente, obtêm-se as estimativas de máxima verossimilhança dos parâmetros utilizando métodos numéricos. Conforme Eloy et al. (2019), utilizaremos a função *optim* do software R para maximizar a função (26) e obter as estimativas dos parâmetros e seus, respectivos erros padrão serão obtidos utilizando os métodos “L-BFGS-B” ou “Nelder–Mead”.

Para os delineamentos em blocos casualizados (DBC) e em quadrado latino (DQL), a função de verossimilhança e derivadas parciais em relação a cada parâmetro serão análogas àquelas obtidas para o DIC.

3.5 EMV para $Y \sim SN(m + \tau_i + \beta_j, \sigma, \lambda)$

Para o DBC, sob a suposição de que o erro tem distribuição $SN(0, 1, \lambda)$ e com a adequação do parâmetro σ de escala, teremos o seguinte modelo

$$y_{ij} = m + \tau_i + \beta_j + \sigma\epsilon_{ij}, \quad \epsilon_{ij} \sim SN(\lambda) \text{ e } \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \end{cases}$$

em que, y_{ij} é o valor observado na parcela que recebeu o tratamento i no bloco j ; m é uma constante comum a todas as unidades experimentais; τ_i é o efeito do tratamento i aplicado no bloco j ; β_j é o efeito do bloco; ϵ_{ij} é o efeito dos fatores não controlados; I é o número de tratamentos do experimento; J é o número de blocos e assim

$$Y_{ij} \sim SN(m + \tau_i + \beta_j, \sigma, \lambda)$$

com esperança e variância dadas por

$$\begin{aligned} E[Y_{ij}] &= E[m + \tau_i + \beta_j + \sigma\epsilon_{ij}] & \text{var}[Y_{ij}] &= \text{var}[m + \tau_i + \beta_j + \sigma\epsilon_{ij}] \\ &= m + \tau_i + \beta_j + \sigma E[\epsilon_{ij}] & &= \sigma^2 \text{var}[\epsilon_{ij}] \\ &= m + \tau_i + \beta_j + \frac{\sigma\lambda}{\sqrt{\lambda^2 + 1}} \cdot \sqrt{\frac{2}{\pi}} & &= \sigma^2 - \frac{\sigma^2\lambda^2}{\lambda^2 + 1} \cdot \frac{2}{\pi}. \end{aligned}$$

Daí,

$$f_{Y_{ij}}(y_{ij}) = \frac{2}{\sigma} \cdot \phi\left(\frac{y_{ij} - (m + \tau_i + \beta_j)}{\sigma}\right) \cdot \Phi\left(\lambda \frac{y_{ij} - (m + \tau_i + \beta_j)}{\sigma}\right)$$

e a função suporte de $\boldsymbol{\theta} = (m, \tau_1, \dots, \tau_I, \beta_1, \dots, \beta_J, \sigma, \lambda)^T$ é dada por:

$$\ell(\boldsymbol{\theta}) = k - IJ \cdot \ln \sigma - \sum_{i=1}^I \sum_{j=1}^J \frac{z_{ij}^2}{2} + \sum_{i=1}^I \sum_{j=1}^J \ln \Phi(\lambda z_{ij}), \quad (27)$$

com $z_{ij} = \frac{y_{ij} - (m + \tau_i + \beta_j)}{\sigma}$ e $k = IJ [\ln 2 - \ln(\sqrt{2\pi})]$.

As componentes do vetor escore $U(\boldsymbol{\theta})$ são dadas por:

$$U_m(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial m} = \frac{\sum_{i=1}^I \sum_{j=1}^J z_{ij}}{\sigma} - \frac{\lambda}{\sigma} \cdot \sum_{i=1}^I \sum_{j=1}^J \frac{\phi(\lambda z_{ij})}{\Phi(\lambda z_{ij})},$$

em relação a τ_i :

$$U_{\tau_i}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \tau_i} = \frac{\sum_{j=1}^J z_{ij}}{\sigma} - \frac{\lambda}{\sigma} \cdot \sum_{j=1}^J \frac{\phi(\lambda z_{ij})}{\Phi(\lambda z_{ij})},$$

em relação a β_j :

$$U_{\beta_j}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \frac{\sum_{i=1}^I z_{ij}}{\sigma} - \frac{\lambda}{\sigma} \cdot \sum_{i=1}^I \frac{\phi(\lambda z_{ij})}{\Phi(\lambda z_{ij})},$$

em relação a σ :

$$U_{\sigma}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma} = -\frac{IJ}{\sigma} + \frac{\sum_{i=1}^I \sum_{j=1}^J z_{ij}^2}{\sigma} - \frac{\lambda}{\sigma} \sum_{i=1}^I \sum_{j=1}^J z_{ij} \cdot \frac{\phi(\lambda z_{ij})}{\Phi(\lambda z_{ij})},$$

e finalmente, em relação a λ , tem-se:

$$U_{\lambda}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \lambda} = \sum_{i=1}^I \sum_{j=1}^J z_{ij} \cdot \frac{\phi(\lambda z_{ij})}{\Phi(\lambda z_{ij})}.$$

Igualando as funções $U_m(\boldsymbol{\theta})$, $U_{\tau_i}(\boldsymbol{\theta})$, $U_{\beta_j}(\boldsymbol{\theta})$, $U_{\sigma}(\boldsymbol{\theta})$, $U_{\lambda}(\boldsymbol{\theta})$ a zero e, resolvendo-as, simultaneamente, obtêm-se as estimativas de máxima verossimilhança dos parâmetros que podem ser encontradas utilizando a função *optim* com os métodos numéricos “L-BFGS-B” ou “Nelder–Mead”, conforme Eloy et al. (2019).

3.6 EMV para $Y \sim SN(m + \alpha_i + \tau_j + \beta_k, \sigma, \lambda)$

Ao considerarmos o DQL, também sob a suposição de que o erro tem distribuição $SN(0, 1, \lambda)$ e com a adequação do parâmetro σ de escala, teremos o seguinte modelo

$$y_{ijk} = m + \alpha_i + \tau_j + \beta_k + \sigma \epsilon_{ijk}, \quad \epsilon_{ijk} \sim SN(\lambda) \text{ e } \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, K \end{cases}$$

em que, y_{ijk} é o valor observado na i -ésima linha e k -ésima coluna do j -ésimo tratamento; m é uma constante comum a todas as unidades experimentais; α_i é o efeito da i -ésima linha; β_k é o efeito da k -ésima coluna; τ_j é o efeito do j -ésimo tratamento; ϵ_{ijk} é o efeito dos fatores não controlados; I é o número de linhas do experimento; K é o número de colunas do experimento; J é o número de tratamentos. Assim

$$Y_{ijk} \sim SN(m + \alpha_i + \tau_j + \beta_k, \sigma, \lambda)$$

com esperança e variância dadas por

$$\begin{aligned}
E[Y_{ijk}] &= E[m + \alpha_i + \tau_j + \beta_k + \sigma \epsilon_{ijk}] & \text{var}[Y_{ijk}] &= \text{var}[m + \alpha_i + \tau_j + \beta_k + \sigma \epsilon_{ijk}] \\
&= m + \alpha_i + \tau_j + \beta_k + \sigma E[\epsilon_{ijk}] & &= \sigma^2 \text{var}[\epsilon_{ijk}] \\
&= m + \alpha_i + \tau_j + \beta_k + \frac{\sigma \lambda}{\sqrt{\lambda^2 + 1}} \cdot \sqrt{\frac{2}{\pi}} & &= \sigma^2 - \frac{\sigma^2 \lambda^2}{\lambda^2 + 1} \cdot \frac{2}{\pi}.
\end{aligned}$$

Assim, a densidade de Y_{ijk} será dada por

$$f_{Y_{ijk}}(y_{ijk}) = \frac{2}{\sigma} \cdot \phi\left(\frac{y_{ij} - (m + \alpha_i + \tau_j + \beta_k)}{\sigma}\right) \cdot \Phi\left(\lambda \frac{y_{ij} - (m + \alpha_i + \tau_j + \beta_k)}{\sigma}\right)$$

a função suporte de $\boldsymbol{\theta} = (m, \alpha_1, \dots, \alpha_I, \tau_1, \dots, \tau_J, \beta_1, \dots, \beta_K, \lambda)^T$ é dada por

$$\ell(\boldsymbol{\theta}) = k - IJK \cdot \ln \sigma - \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \frac{z_{ijk}^2}{2} + \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \ln \Phi(\lambda z_{ijk}), \quad (28)$$

com $z_{ijk} = \frac{y_{ij} - (m + \alpha_i + \tau_j + \beta_k)}{\sigma}$ e $k = IJK \left[\ln 2 - \ln(\sqrt{2\pi}) \right]$.

As componentes do vetor escore $U(\boldsymbol{\theta})$ são dadas por:

$$U_m(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial m} = \frac{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K z_{ijk}}{\sigma} - \frac{\lambda}{\sigma} \cdot \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \frac{\phi(\lambda z_{ijk})}{\Phi(\lambda z_{ijk})},$$

$$U_{\alpha_i}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha_i} = \frac{\sum_{j=1}^J \sum_{k=1}^K z_{ijk}}{\sigma} - \frac{\lambda}{\sigma} \cdot \sum_{j=1}^J \sum_{k=1}^K \frac{\phi(\lambda z_{ijk})}{\Phi(\lambda z_{ijk})},$$

$$U_{\beta_k}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_k} = \frac{\sum_{i=1}^I \sum_{j=1}^J z_{ijk}}{\sigma} - \frac{\lambda}{\sigma} \cdot \sum_{i=1}^I \sum_{j=1}^J \frac{\phi(\lambda z_{ijk})}{\Phi(\lambda z_{ijk})},$$

$$U_{\tau_j}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \tau_j} = \frac{\sum_{i=1}^I \sum_{k=1}^K z_{ijk}}{\sigma} - \frac{\lambda}{\sigma} \cdot \sum_{i=1}^I \sum_{k=1}^K \frac{\phi(\lambda z_{ijk})}{\Phi(\lambda z_{ijk})},$$

$$U_{\sigma}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma} = -\frac{IJK}{\sigma} + \frac{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K z_{ijk}^2}{\sigma} - \frac{\lambda}{\sigma} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K z_{ijk} \cdot \frac{\phi(\lambda z_{ijk})}{\Phi(\lambda z_{ijk})},$$

$$U_{\sigma}(\lambda) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \lambda} = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K z_{ijk} \cdot \frac{\phi(\lambda z_{ijk})}{\Phi(\lambda z_{ijk})}.$$

Igualando-se as funções $U_m(\boldsymbol{\theta})$, $U_{\alpha_i}(\boldsymbol{\theta})$, $U_{\beta_k}(\boldsymbol{\theta})$, $U_{\tau_j}(\boldsymbol{\theta})$, $U_{\sigma}(\boldsymbol{\theta})$, $U_{\lambda}(\boldsymbol{\theta})$ a zero e, resolvendo-as, simultaneamente, obtêm-se as estimativas de máxima verossimilhança dos parâmetros que podem ser encontradas utilizando a função *optim* com os métodos numéricos “L-BFGS-B” ou “Nelder–Mead”, conforme Eloy et al. (2019).

4 Resultados e discussão

Nesta seção é apresentada a análise de dois bancos de dados reais, a Subseção 4.1 refere-se ao banco de dados de temperatura da cidade de Piracicaba, São Paulo, que por sua vez corresponde a um modelo de intercepto, ou seja, um modelo de média e a Subseção 4.2 trata de um experimento realizado sob o delineamento inteiramente casualizado, em que avaliou-se o percentual de gordura de idosos quando submetidos a diferentes intervenções de atividade física. Em ambos, vamos comparar o ajuste das distribuições normal e *skew-normal*, construir os gráficos de envelopes simulados para análise de resíduos quantílicos, conforme o passo a passo presente na Subseção 2.7, e comparar o percentual de pontos fora dos envelopes considerando as duas distribuições, e ainda, através do estudo de simulação de Monte Carlo, vamos avaliar a inferência dos estimadores de máxima verossimilhança tanto para o modelo de intercepto quanto para o modelo de um DIC, com respectivos erros dados pela distribuição *skew-normal*.

4.1 Dados de temperatura

Este primeiro conjunto de dados refere-se à variável temperatura (°C) obtida das leituras diárias do período de 1º de janeiro a 31 de dezembro de 2011, na cidade de Piracicaba, constituindo-se de 365 valores observados. Os dados foram fornecidos pelo Departamento de Engenharia de Biosistemas da Escola Superior de Agricultura Luiz de Queiroz (ESALQ), associado à Universidade de São Paulo (USP). As temperaturas foram medidas na Estação Meteorológica de Piracicaba, localizada a latitude 22° 42’30”S, longitude 47° 38’30”W e altitude de 546 metros.

Nesse ano, a temperatura média em Piracicaba foi de 22,33 °C, com desvio padrão de 2,95 °C e temperaturas mínima e máxima, respectivamente, 14,68 °C e 27,25 °C. A mediana da temperatura foi de 22,90 °C, próxima a temperatura média, e além disso, este conjunto de dados apresentou coeficientes de assimetria e curtose iguais a, respectivamente, $-0,50$ e $-0,73$.

Conforme a Figura 4, gráfico (a), o comportamento observado no histograma, mostra que aparentemente os dados apresentam pequena assimetria a esquerda, e embora este gráfico aponte a forma da distribuição, ele não informa sobre *outliers*. Todavia, o *boxplot* dos dados na Figura 4, gráfico (b), aponta que além de os dados possuírem média e mediana diferentes, a mediana é superior à média, o que evidencia visualmente a assimetria a esquerda, em concordância com a análise descritiva em que o coeficiente de assimetria é $-0,5$.

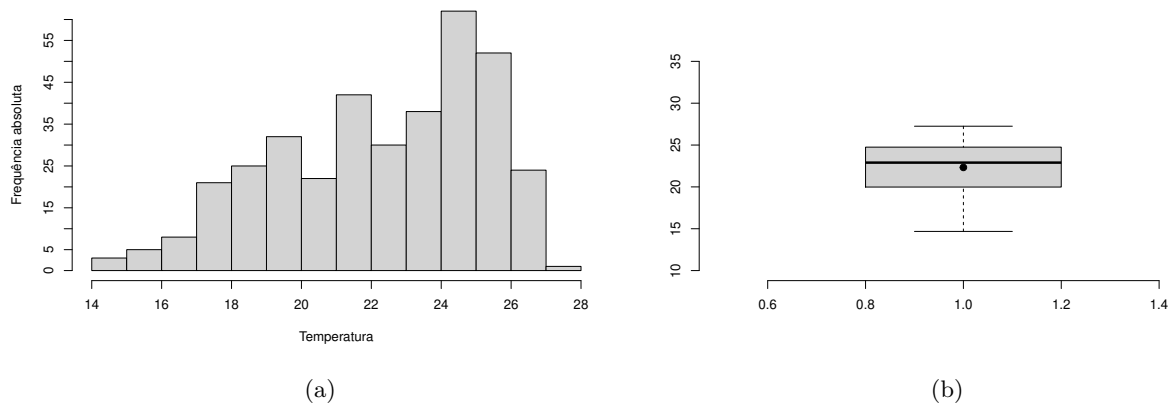


Figura 4: Histograma com a frequência absoluta e *boxplot* para a variável temperatura.

4.1.1 Ajuste

As estimativas dos parâmetros para modelo de intercepto, considerando as distribuições normal e *skew*-normal para o erro, foram estimadas através da função *optim* no *software* R. Tais estimativas se encontram na terceira coluna da Tabela 1, em que é observado que na distribuição *skew*-normal os parâmetros estimados possuem valores maiores que aqueles estimados pela distribuição normal. Neste trabalho, as demonstrações para encontrar os estimadores de máxima verossimilhança do modelo de intercepto não foram feitas, mas embora representados por m tanto para a distribuição normal quanto para a *skew*-normal, esses parâmetros não representam o mesmo, eles possuem exatamente o mesmo estimador de máxima verossimilhança somente quando $\lambda = 0$ para o modelo *skew*-normal.

Seguidos das estimativas dos parâmetros, na mesma tabela, encontra-se o erro padrão, denotado por EP, e os testes t e intervalos de confiança para cada um dos parâmetros do modelo, em que LI e LS correspondem aos limites inferior e superior dos intervalos de confiança, respectivamente. A hipótese de nulidade tanto para o teste t quanto para os intervalos de confiança, que supõem que os parâmetros são iguais a zero, foi rejeitada para todas as estimativas dos parâmetros, para as duas distribuições. Além disso, podemos observar que a amplitude dos intervalos de confiança para as estimativas obtidas pela distribuição *skew*-normal são menores quando comparados com os da normal.

Tabela 1: Teste t e intervalos de confiança para os parâmetros utilizando-se as distribuições normal e *skew-normal* para a variável temperatura.

Distribuição	θ	$\hat{\theta}$	EP	t_{cal}	valor- p	LI	LS
Normal	m	22,33	0,15	144,77	0,0000	22,02	22,63
	σ^2	2,95	0,11	27,02	0,0000	2,73	3,16
<i>Skew-normal</i>	m	26,36	0,13	197,86	0,0000	26,09	26,62
	σ	4,99	0,21	23,35	0,0000	4,57	5,41
	λ	-9,70	2,44	-3,97	0,0008	-14,51	-4,90

Para avaliar o ajuste das distribuições, vamos usar os critérios de informação de Akaike, Akaike corrigido e critério de informação bayesiano dados na Tabela 2. Sabendo que o modelo que apresenta menor valor de AIC, AICc e BIC, é considerado o melhor modelo dentre os comparados, podemos observar que, como esperado, a distribuição *skew-normal* teve melhor ajuste quando comparada a distribuição normal.

Tabela 2: Critérios de informação para as distribuições normal e *skew-normal* para a variável temperatura.

Distribuição	AIC	AICc	BIC
Normal	1828,64	1828,75	1836,44
<i>Skew-normal</i>	1764,64	1764,81	1776,34

Percebe-se que a distribuição *skew-normal* teve melhor ajuste para os dados de temperatura, pois conseguiu captar melhor as características intrínsecas aos dados do que a distribuição normal, que por sua vez não possui o parâmetro λ de assimetria. Na Figura 5 observa-se o *plot* das densidades das distribuições normal e *skew-normal* com as respectivas estimativas de máxima verossimilhança dos parâmetros de cada distribuição, e nela é possível visualizar o ajuste de cada distribuição e notar que a distribuição *skew-normal*, de fato, tem melhor ajuste, conforme visto com os critérios de seleção de modelos.

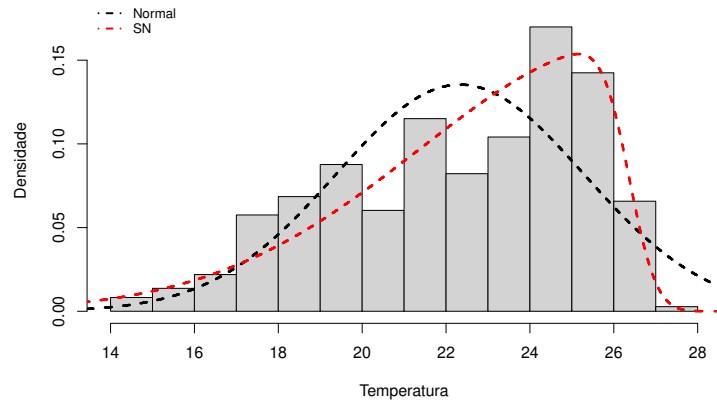


Figura 5: Histograma para a variável temperatura com as densidades das distribuições normal e *skew-normal* (SN).

4.1.2 Análise dos resíduos

O resíduo quantílico estimado para o modelo de intercepto, considerando a distribuição *skew-normal*, é dado por

$$\widehat{r}q_{SN} = \Phi^{-1} \left\{ \Phi \left(\frac{y - \widehat{m}}{\widehat{\sigma}} \right) - 2T \left(\frac{y - \widehat{m}}{\widehat{\sigma}}, \widehat{\lambda} \right) \right\},$$

em que, $\Phi^{-1}(\cdot)$ é a função quantílica da distribuição normal padrão composta pela função de distribuição acumulada da *skew-normal*. Vale ressaltar que se $\lambda = 0$ então teremos o resíduo quantílico da distribuição normal.

Observando os envelopes simulados na Figura 6, constata-se que em 2000 simulações de Monte Carlo, a distribuição *skew-normal* apresentou somente 4,11% dos pontos fora dos envelopes enquanto a distribuição normal apresentou pouco mais 63%, confirmando o já visto na análise de dados e no ajuste através dos critérios de informação de Akaike, Akaike corrigido e Bayesiano, que a distribuição *skew-normal* apresentou de fato melhor ajuste aos dados de temperatura quando comparada a distribuição normal, pois esta capta melhor a assimetria presente neste conjunto de dados, devido ao parâmetro de assimetria λ que esta possui a mais que a distribuição normal.

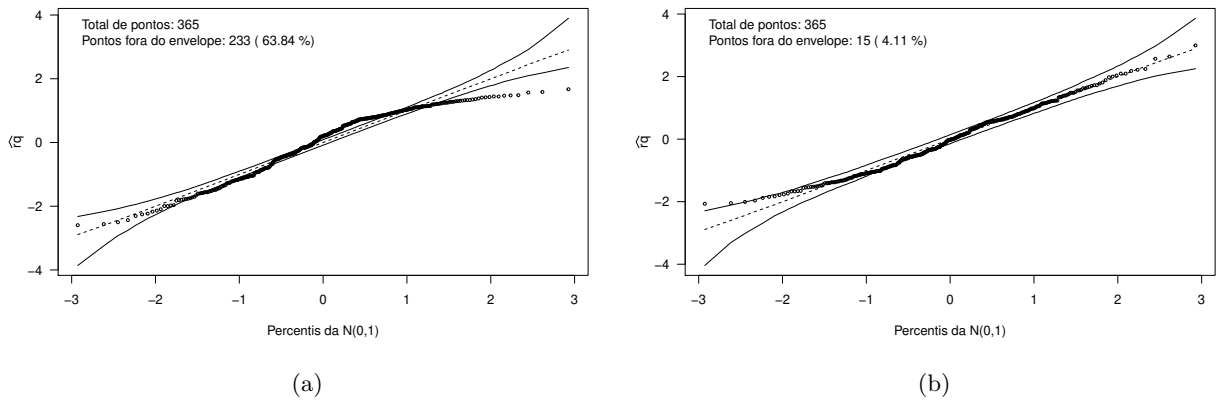


Figura 6: Envelopes simulados dos quantis normais versus os resíduos quantílicos para o modelo de intercepto com distribuição normal e *skew*-normal, respectivamente.

4.1.3 Estudo de simulação

Como na Subseção 4.1.1 a distribuição *skew*-normal apresentou melhor ajuste aos dados, será feito um estudo de simulação para avaliar o comportamento das estimativas dos parâmetros do modelo de intercepto somente com erro *skew*-normal. Os parâmetros dos seguintes cenários considerados tiveram configurações tomadas a partir das estimativas encontradas no ajuste, Subseção 4.1.1, em que também foi variado o tamanho amostral para avaliar se as estimativas dos parâmetros são boas para amostras menores que 365, que é o tamanho correspondente ao nosso conjunto de dados, e além disso, foi variado o parâmetro de assimetria, λ , para averiguar se a distribuição *skew*-normal continua estimando bem para assimetrias menores do que a encontrada na Subseção 4.1.1. Para avaliação das estimativas dos parâmetros, utilizaremos a variância, erro quadrático médio associados a cada estimativa, além de ilustrações gráficas.

- **Cenário 1:** $n = (30, 100, 365)$, $m = 26$, $\sigma = 4$, $\lambda = -2$;
- **Cenário 2:** $n = (30, 100, 365)$, $m = 26$, $\sigma = 4$, $\lambda = -5$;
- **Cenário 3:** $n = (30, 100, 365)$, $m = 26$, $\sigma = 4$, $\lambda = -9$.

Em cada cenário foram feitas 1000 simulações Monte Carlo, e na Tabela 3 encontra-se as estimativas de máxima verossimilhança de cada parâmetro de acordo com a variação no tamanho das amostras simuladas. Nela pode-se observar que os parâmetros m e σ da distribuição *skew*-normal têm estimativas próximas aos parâmetros tomados como verdadeiros mesmo para $n = 30$, que é a amostra de menor tamanho. A maior variância observada neste cenário é 1,88 para m , e tem pequena magnitude quando comparada a estimativa de seu parâmetro.

Ao aumentar o tamanho da amostra para 100, os estimadores de máxima verossimilhança de m e σ ficam mais próximos dos parâmetros verdadeiros em todos os cenários, com exceção

do cenário 1. Ao observar as amostras de tamanho 365, que corresponde ao tamanho da amostra dos dados reais, as estimativas dos parâmetros diferem do verdadeiro por no máximo 0.15, e com exceção do parâmetro m , nos cenários 1 e 2, o erro quadrático médio apresentou diminuição para todos.

O parâmetro de assimetria, λ , foi o parâmetro que apresentou as piores estimativas em todos os cenários. Entretanto, conforme aumenta-se o tamanho da amostra, suas estimativas ficam cada vez mais próximas dos parâmetros verdadeiros com diminuição em suas variâncias e erros quadrático médios.

Tabela 3: Estimativa dos parâmetros, variância e erro quadrático médio, baseados em 1000 simulações da distribuição *skew-normal* para o modelo de intercepto.

Cenário 1									
	$n = 30$			$n = 100$			$n = 365$		
θ	Estim.	Var.	EQM	Estim.	Var.	EQM	Estim.	Var.	EQM
$m = 26$	25,41	1,88	2,22	24,85	2,10	3,43	25,93	0,20	0,20
$\sigma = 4$	3,69	0,73	0,82	3,49	0,31	0,57	3,96	0,10	0,10
$\lambda = -2$	-2,32	4,65	4,75	-1,43	2,04	2,36	-2,00	0,26	0,26
Cenário 2									
	$n = 30$			$n = 100$			$n = 365$		
θ	Estim.	Var.	EQM	Estim.	Var.	EQM	Estim.	Var.	EQM
$m = 26$	25,53	0,78	1,00	25,92	0,11	0,12	25,85	0,47	0,49
$\sigma = 4$	3,65	0,54	0,67	3,92	0,14	0,14	3,92	0,14	0,14
$\lambda = -5$	-4,56	6,72	6,92	-5,30	3,83	3,92	-5,05	2,54	2,54
Cenário 3									
	$n = 30$			$n = 100$			$n = 365$		
θ	Estim.	Var.	EQM	Estim.	Var.	EQM	Estim.	Var.	EQM
$m = 26$	25,54	0,45	0,65	25,84	0,09	0,12	25,99	0,01	0,01
$\sigma = 4$	3,55	0,38	0,58	3,80	0,11	0,15	4,02	0,02	0,02
$\lambda = -9$	-5,97	7,25	16,44	-7,20	4,18	7,42	-9,03	1,40	1,40

A Figura 7, confirma as informações que constam na Tabela 3. Nela, conforme aumenta-se os tamanhos das amostras, em geral, mais os parâmetros estimados aproximam-se dos parâmetros verdadeiros. E na figura, quanto maior o tamanho amostral, mais as curvas das densidades plotadas com seus respectivos parâmetros estimados, se aproximam da curva verdadeira, que é a de cor preta. A curva de cor azul é a correspondente as amostras de tamanho 365, em ambos os cenários, ela é visualmente a densidade mais próxima da densidade com os parâmetros verdadeiros.

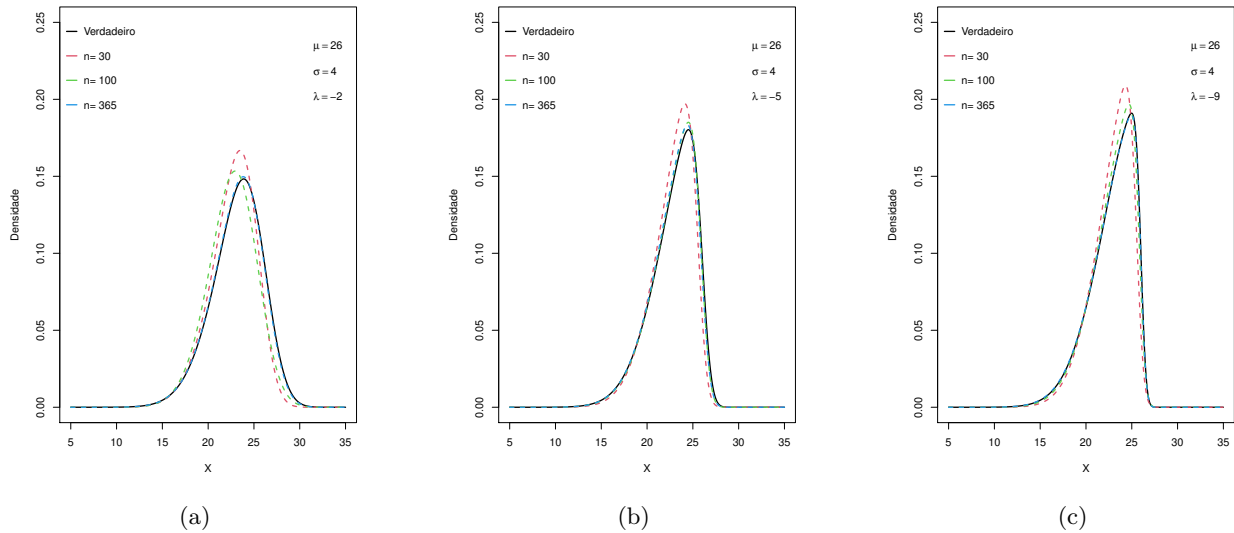


Figura 7: Funções de densidade da distribuição SN nos valores verdadeiros dos parâmetros e nas estimativas médias para $n=30$; $n=100$; $n=365$.

Este estudo de simulação evidencia que a distribuição *skew*-normal apresentou estimativas para os parâmetros do modelo de intercepto próximas aos parâmetros verdadeiros, que foram tomados a partir dos dados de temperatura vistos na Seção 4.1, e com isso é uma opção de distribuição viável para estudos semelhantes a este.

4.2 Dados de percentual de gordura

Este segundo banco de dados, foi gentilmente cedido por Aristéia Nunes Sampaio, pós-graduanda do programa Vigilância em Saúde da Universidade Federal do Acre, e refere-se a um experimento realizado sob o delineamento inteiramente casualizado, em que foram analisadas mudanças ocorridas nos perfis antropométrico, lipídico, glicêmico e de condicionamento físico de idosos participantes de dois programas de atenção à saúde do Município de Rio Branco, Acre. Os idosos selecionados para este estudo são integrantes dos programas Estratégia de Saúde da Família (ESF) e programa de extensão Idoso Ativo na UFAC (IA UFAC). A população de estudo foi composta por 60 idosos de ambos os sexos, com idades entre 60 e 80 anos, divididos em quatro grupos experimentais desbalanceados.

O primeiro grupo, correspondente ao tratamento A, com 17 idosos, é o grupo de treinamento em academias ao ar livre do programa ESF, o tratamento consistiu em atividades nos aparelhos durante 30 minutos seguidos de atividades recreativas e alongamento. O tratamento B, com 11 idosos, corresponde ao grupo de treinamento de exercícios aeróbicos e localizados, também do programa ESF, este grupo realizou 20 minutos de exercícios aeróbicos seguidos de 30 minutos de exercícios localizados para membros superiores e inferiores e mais 10 minutos de alongamentos. O tratamento C, com 17 idosos, é o grupo de treinamento em exercícios resistidos do programa

IA UFAC, que por sua vez executou um programa de exercícios resistidos realizados na sala de musculação da Universidade Federal do Acre, composto por 60 minutos de exercícios para membros superiores e inferiores com movimentos mono e biarticular nos equipamentos de musculação. E o tratamento D, com 15 idosos, corresponde ao grupo controle, composto por não praticantes de atividades físicas pertencentes ao programa ESF.

Os grupos experimentais foram submetidos a 16 semanas de treinamento, sendo avaliados antes e após a intervenção dos exercícios, sendo que o grupo controle realizou as avaliações, mas não praticou exercícios físicos durante o período. Este experimento coletou dados de treze variáveis respostas, e neste trabalho será avaliado somente o percentual de gordura que equivale a diferença entre o percentual de gordura corpórea avaliada antes e após os exercícios físicos. Como será avaliada uma diferença, então é interessante que sejam observados valores negativos, se o objetivo de diminuir o percentual de gordura com atividades físicas for alcançado. O intercepto do modelo do delineamento em DIC corresponde a média do tratamento D, ou seja, a comparação entre tratamentos será feita com o grupo controle.

Fazendo a análise descritiva dos dados observou-se que as porcentagens mínima e máxima do percentual de gordura dos idosos são $-6,80\%$, $2,20\%$, respectivamente, com uma média de $-1,57\%$, e desvio padrão de $1,74\%$, apontando que, em média, o percentual de gordura dos idosos diminuiu após as dezesseis semanas de treino. Os coeficientes de assimetria e curtose foram $-0,81$ e $1,57$, respectivamente.

Assim como em Guedes et al. (2014), os tratamentos apresentam indícios de assimetria. O que é possível perceber visualmente na Figura 8, em que *boxplots* além de apresentarem assimetria visível dentro dos quatro tratamentos, há também a presença possíveis *outliers*, influenciando na variância de cada um. Para testar a homogeneidade de variâncias, procedeu-se o teste de Bartlett, a 5% de significância, que teve valor $-p = 0,1215$ evidenciando que as variâncias podem ser consideradas homogêneas.

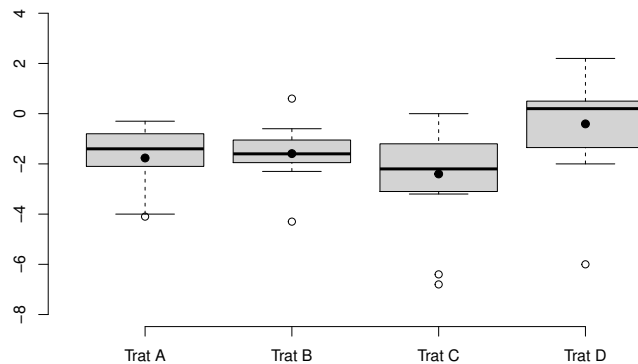


Figura 8: *Boxplot* para os tratamentos da variável percentual de gordura.

4.2.1 Ajuste

As estimativas de máxima verossimilhança dos parâmetros de cada modelo foram obtidos através da função *optim*, e os chutes iniciais para esta função são dados a partir das estimativas de mínimos quadrados ordinários obtidos no ajuste pela função *aov*, no *software* R. Lembrando que a restrição que está sendo utilizada nesse estudo é $\tau_4 = 0$, o que implica que o intercepto do modelo corresponde a média do tratamento D, e então a comparação entre as médias dos tratamentos será feita com a média do grupo controle.

As estimativas de máxima verossimilhança dos parâmetros dos modelos do DIC com erro normal e *skew-normal*, seguidos do erro padrão (EP), teste *t* e intervalos de confiança, em que os limites inferior e superior são dados por LI e LS, constam na Tabela 4. Nela é observado que diferentemente de quando se considera a distribuição *skew-normal*, o teste *t* para a distribuição normal evidencia que o efeito de τ_2 não difere estatisticamente de zero, o que mostra que a distribuição normal não detectou diferença entre os tratamentos B e D, pois τ_2 expressa a diferença entre as médias destes tratamentos, ou seja, de acordo com a distribuição normal, o percentual de gordura do grupo de treinamento de exercícios aeróbicos e localizados não difere do grupo não praticante de atividade física.

Já a distribuição *skew-normal* constatou que o percentual de gordura dos idosos dos três tipos de treinos testados diferiu do percentual de gordura do grupo controle, e como as estimativas para os efeitos de τ_1 , τ_2 e τ_3 são todas negativas, então o percentual de gordura de ambos os tratamentos diminuiu quando comparado ao percentual dos idosos pertencentes ao grupo controle. Quanto aos intervalos de confiança, com exceção do intervalo para a variância, a distribuição *skew-normal* apresentou menor amplitude intervalar para as estimativas dos demais parâmetros, quando comparados com os intervalos de confiança gerados a partir da distribuição normal.

Tabela 4: Teste *t* e intervalos de confiança para os parâmetros das distribuições normal e *skew-normal* em um DIC.

Distribuição	θ	$\hat{\theta}$	EP	Hipóteses	t_{cal}	valor- <i>p</i>	LI	LS
Normal	m	-0,41	0,40	$H_0 : m = 0$	-1,01	0,3159	-1,21	0,40
	τ_1	-1,36	0,55	$H_0 : \mu_A - \mu_D = 0$	-2,46	0,0168	-2,46	-0,25
	τ_2	-1,18	0,62	$H_0 : \mu_B - \mu_D = 0$	-1,91	0,0605	-2,42	0,05
	τ_3	-1,99	0,55	$H_0 : \mu_C - \mu_D = 0$	-3,61	0,0006	-3,10	-0,89
	σ^2	1,56	0,14	$H_0 : \sigma^2 = 0$	10,95	0,0000	1,27	1,84
<i>Skew-normal</i>	m	1,61	0,41	$H_0 : m = 0$	3,91	0,0002	0,78	2,43
	τ_1	-1,82	0,47	$H_0 : \mu_A - \mu_D = 0$	-3,86	0,0003	-2,76	-0,88
	τ_2	-1,48	0,50	$H_0 : \mu_B - \mu_D = 0$	-2,95	0,0045	-2,49	-0,48
	τ_3	-2,03	0,43	$H_0 : \mu_C - \mu_D = 0$	-4,78	0,0000	-2,88	-1,18
	σ	2,40	0,29	$H_0 : \sigma^2 = 0$	8,20	0,0000	1,81	2,98
	λ	-4,43	2,21	$H_0 : \lambda = 0$	-2,01	0,0491	-8,84	-0,02

O ajuste das distribuições normal e *skew*-normal, realizados pelo método da máxima verossimilhança, serão avaliados através dos critérios AIC, AICc e BIC, que constam na Tabela 5.

Tabela 5: Critérios de informação para os modelos normal e *skew*-normal e experimento em DIC.

Distribuição	AIC	AICc	BIC
Normal	233,53	235,68	244,00
<i>Skew</i> -normal	222,55	225,38	235,12

Os valores dos critérios de seleção, AIC, AICc e BIC, indicam que o modelo com distribuição *skew*-normal para o erro obteve melhor ajuste para os dados de percentual de gordura corporal dos idosos do que o modelo com distribuição normal, em concordância com Guedes et al. (2014) e Said, Ning e Tian, (2018) que também ajustaram um DIC com erro *skew*-normal. Além disso, as estimativas dos efeitos de tratamentos através do modelo com distribuição *skew*-normal detectou que os três programas de treinamento diferiram do grupo não praticante de atividade física.

4.2.2 Análise dos resíduos

O resíduo quantílico estimado do delineamento inteiramente casualizado, com distribuição *skew*-normal, é dado por

$$\widehat{r}q_{SN} = \Phi^{-1} \left\{ \Phi \left(\frac{y - (\widehat{m} + \widehat{\tau}_i)}{\widehat{\sigma}} \right) - 2T \left(\frac{y - (\widehat{m} + \widehat{\tau}_i)}{\widehat{\sigma}}, \widehat{\lambda} \right) \right\},$$

em que, $\Phi^{-1}(\cdot)$ é a função quantílica da distribuição normal padrão composta pela função de distribuição acumulada de $Y \sim SN(m + \tau_i, \sigma, \lambda)$. Considerando $\lambda = 0$ tem-se o modelo do delineamento em DIC com distribuição normal e conseqüentemente o resíduo quantílico desta.

Na Figura 9, é expresso o quão bem a distribuição *skew*-normal se ajustou aos dados de percentual de gordura dos idosos. Em 2000 simulações de Monte Carlo nenhum ponto ficou fora dos envelopes simulados, enquanto a distribuição normal teve um pouco mais de 40% dos pontos do lado de fora.

A distribuição *skew*-normal captou as características intrínsecas aos dados de percentual de gordura, e de fato obteve melhor ajuste quando comparada a distribuição normal, como já visto na Seção 4.2 e confirmando seu resultado.

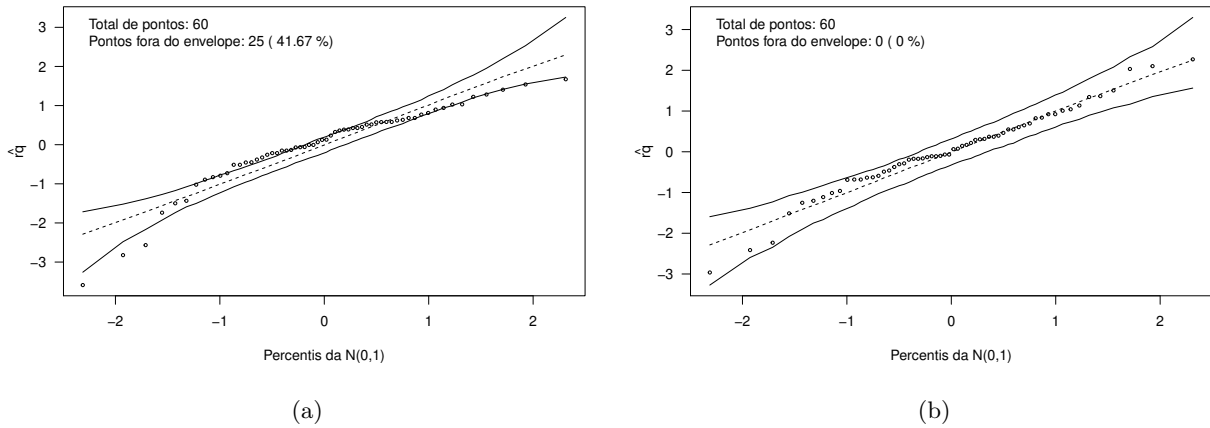


Figura 9: Envelopes simulados dos quantis normais versus os resíduos quantílicos para o modelo inteiramente casualizado com distribuição normal e *skew*-normal, respectivamente.

4.2.3 Estudo de simulação

Na análise de dados do percentual de gordura, com os critérios de comparação de modelos, foi concluído que a distribuição *skew*-normal apresentou melhor ajuste aos dados. Nesta subseção apresentaremos um estudo de simulação para identificar se as estimativas dos parâmetros que esta distribuição gera são de fato boas estimativas.

Assim como no estudo de simulação baseado nos dados de temperatura, serão considerados três cenários com as configurações a partir das estimativas dos parâmetros obtidas na Subseção 4.2, em que também foram variados o tamanho amostral e o parâmetro de assimetria, a fim de avaliarmos se a distribuição *skew*-normal continuará fazendo estimativas próximas dos parâmetros que foram simulados para um número de repetições maior e menor e assimetria mais baixa e mais alta do que as apresentadas na Subseção 4.2.

Avaliaremos o erro quadrático médio e variância associados a cada estimativa, além de ilustrações gráficas para percepção visual da proximidade entre as curvas estimadas e a verdadeira.

- **Cenário 1:** $n = (20, 68, 120)$, $m = 1,61$, $\tau_1 = -1,82$, $\tau_2 = -1,48$, $\tau_3 = -2,03$, $\sigma = 2,40$, $\lambda = -2,50$;
- **Cenário 2:** $n = (20, 68, 120)$, $m = 1,61$, $\tau_1 = -1,82$, $\tau_2 = -1,48$, $\tau_3 = -2,03$, $\sigma = 2,40$, $\lambda = -4,43$;
- **Cenário 3:** $n = (20, 68, 120)$, $m = 1,61$, $\tau_1 = -1,82$, $\tau_2 = -1,48$, $\tau_3 = -2,03$, $\sigma = 2,40$, $\lambda = 4,00$.

Em cada cenário foram feitos 1000 simulações de Monte Carlo e calculada média, variância e erro quadrático médio da cada estimativa para avaliar o quão bem a distribuição *skew*-normal está

estimando os parâmetros do modelo. Na Tabela 6, constam estas estimativas e nela é observado que para 5 repetições por tratamento, que corresponde a uma amostra de tamanho 20, as estimativas para m , σ , e λ não foram boas em nenhum dos cenários. O maior interesse dos pesquisadores em um experimento é nas estimativas dos τ_i , que indicam a existência de diferença entre os tratamentos. Mesmo para um número de repetições baixo, as estimativas destes parâmetros foram próximas aos verdadeiros valores, com baixa variância e baixo erro quadrático médio.

Ao aumentar o número de repetições para 17, que é aproximadamente o número de repetições do nosso conjunto de dados reais, correspondendo a uma amostra de tamanho 68, houve melhora nas estimativas de todos os parâmetros, principalmente nas estimativas de m , σ , e λ , que apresentaram menor variância e erro quadrático médio do que para $j = 5$. Para 30 repetições, a estimativa somente do parâmetro λ foi mais distante do parâmetro verdadeiro do que para 17 repetições, as estimativas dos demais parâmetros diferiram de seus respectivos parâmetros verdadeiros por no máximo 0.05.

Tabela 6: Estimativa dos parâmetros, variância e erro quadrático médio, baseados em 1000 simulações da distribuição *skew-normal* para o modelo em DIC.

Cenário 1									
θ	$j = 5$			$j = 17$			$j = 30$		
	Estim,	Var,	EQM	Estim,	Var	EQM	Estim,	Var,	EQM
$m = 1,61$	0,28	0,86	2,63	1,42	0,39	0,42	1,57	0,18	0,18
$\tau_1 = -1,82$	-1,89	0,95	0,96	-1,82	0,30	0,30	-1,82	0,16	0,16
$\tau_2 = -1,48$	-1,45	0,96	0,96	-1,49	0,28	0,28	-1,49	0,15	0,15
$\tau_3 = -2,03$	-2,06	1,08	1,08	-2,06	0,26	0,26	-2,03	0,15	0,15
$\sigma = 2,40$	1,51	0,16	0,95	2,26	0,14	0,16	2,36	0,08	0,08
$\lambda = -2,50$	-0,58	0,79	4,49	-2,77	2,69	2,76	-2,94	1,94	2,14

Cenário 2									
θ	$j = 5$			$j = 17$			$j = 30$		
	Estim,	Var,	EQM	Estim,	Var	EQM	Estim,	Var,	EQM
$m = 1,61$	-0,04	0,71	3,44	1,45	0,17	0,20	1,56	0,07	0,08
$\tau_1 = -1,82$	-1,76	0,90	0,91	-1,82	0,19	0,19	-1,82	0,10	0,10
$\tau_2 = -1,48$	-1,46	0,97	0,97	-1,49	0,20	0,20	-1,47	0,10	0,10
$\tau_3 = -2,03$	-1,94	0,96	0,97	-2,04	0,19	0,19	-2,02	0,11	0,11
$\sigma = 2,40$	1,30	0,12	1,34	2,24	0,09	0,11	2,35	0,05	0,05
$\lambda = -4,43$	-0,49	0,84	16,36	-4,42	3,58	3,58	-5,09	3,58	4,02

Cenário 3									
θ	$j = 5$			$j = 17$			$j = 30$		
	Estim,	Var,	EQM	Estim,	Var	EQM	Estim,	Var,	EQM
$m = 1,61$	3,11	0,74	3,00	1,78	0,22	0,24	1,66	0,09	0,09
$\tau_1 = -1,82$	-1,79	0,89	0,89	-1,83	0,22	0,22	-1,81	0,10	0,10
$\tau_2 = -1,48$	-1,50	0,91	0,91	-1,51	0,21	0,21	-1,49	0,11	0,11
$\tau_3 = -2,03$	-1,97	0,92	0,92	-2,03	0,20	0,20	-2,03	0,11	0,11
$\sigma = 2,40$	1,33	0,13	1,27	2,24	0,10	0,13	2,34	0,05	0,05
$\lambda = 4$	0,55	0,85	12,75	4,10	3,60	3,61	4,57	3,28	3,60

Nas Figuras 10, 11 e 12 estão plotadas as curvas das densidades com as respectivas estimativas dos parâmetros variando o número de repetições, para os quatro tratamentos. Como esperado, ao aumentar o número de repetições a densidade aproxima-se da densidade com valores verdadeiros, a de cor preta.

Sendo assim, este estudo evidenciou que a distribuição *skew-normal* apresentou estimativas para os parâmetros do modelo do delineamento inteiramente casualizado próximas aos parâmetros verdadeiros, principalmente para os parâmetros que expressam diferença entre tratamentos. E portanto, a distribuição *skew-normal* se mostrou viável para ajuste a dados que possuem *outliers* e assimetria dentro dos tratamentos, como os dados da Seção 4.2.

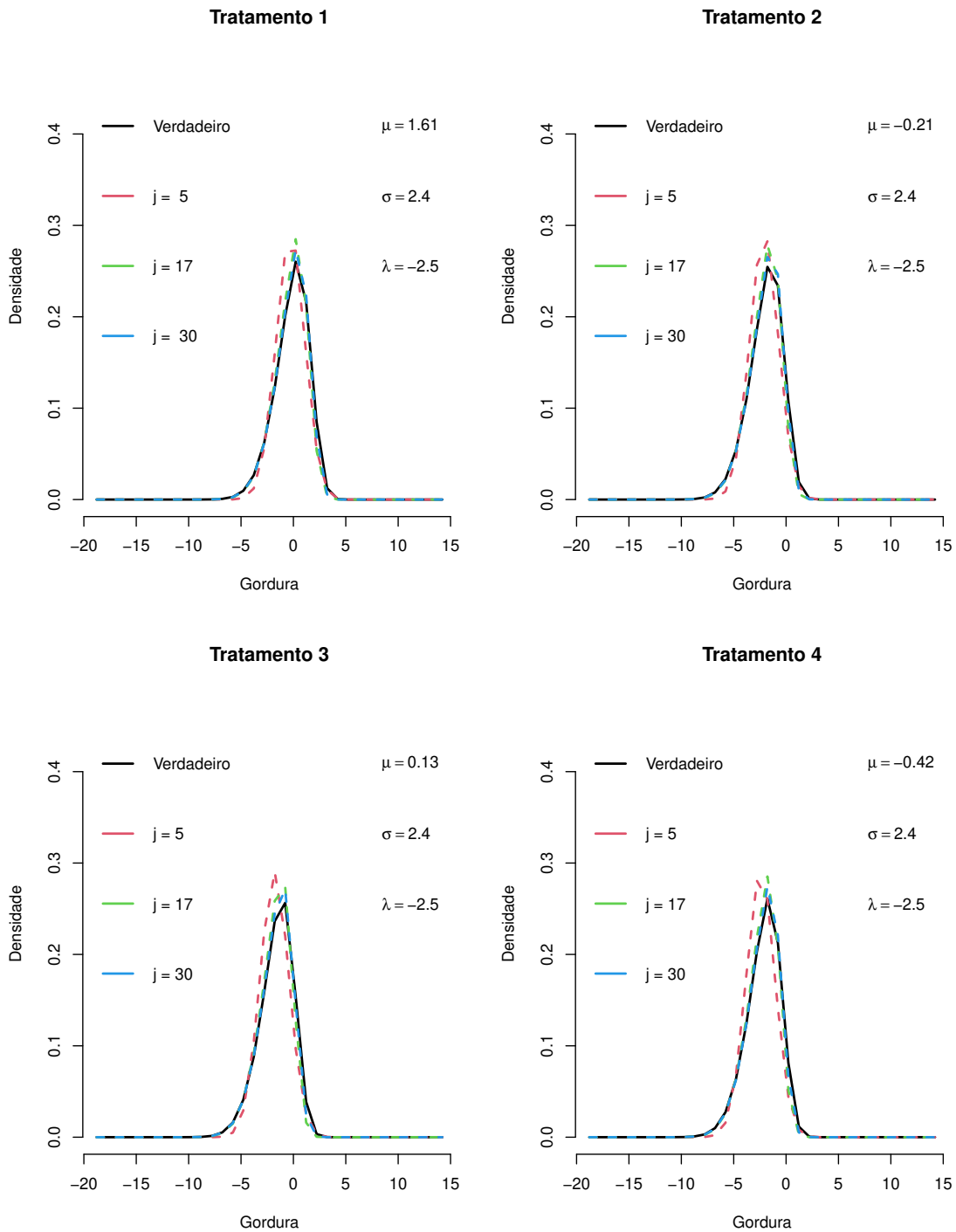


Figura 10: Densidades da distribuição *skew*-normal com $\lambda = -2,5$, nos valores verdadeiros dos parâmetros e nas estimativas médias para $n=20$; $n=68$; $n=120$.

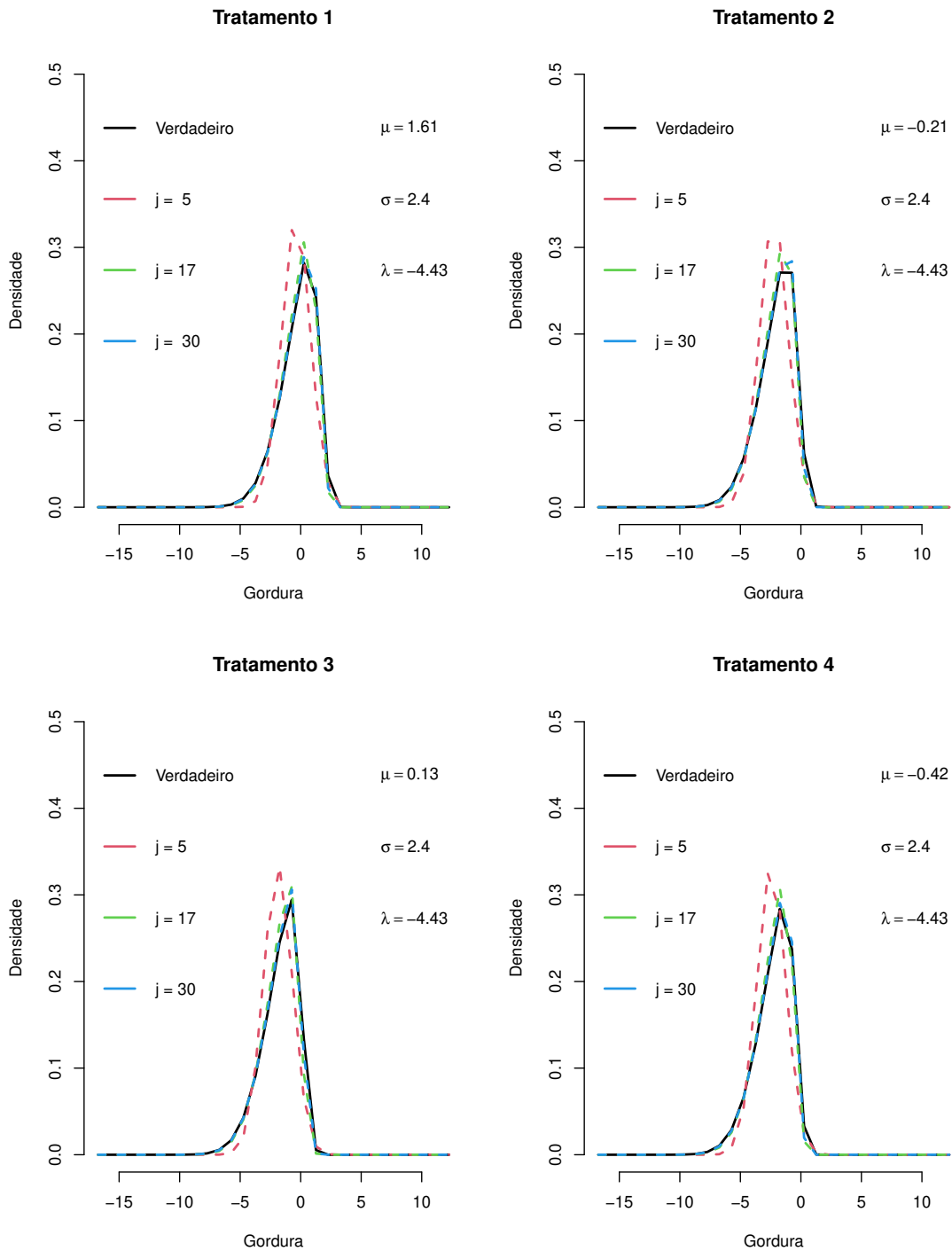


Figura 11: Densidades da distribuição *skew-normal* com $\lambda = -4,43$, nos valores verdadeiros dos parâmetros e nas estimativas médias para $n=20$; $n=68$; $n=120$.

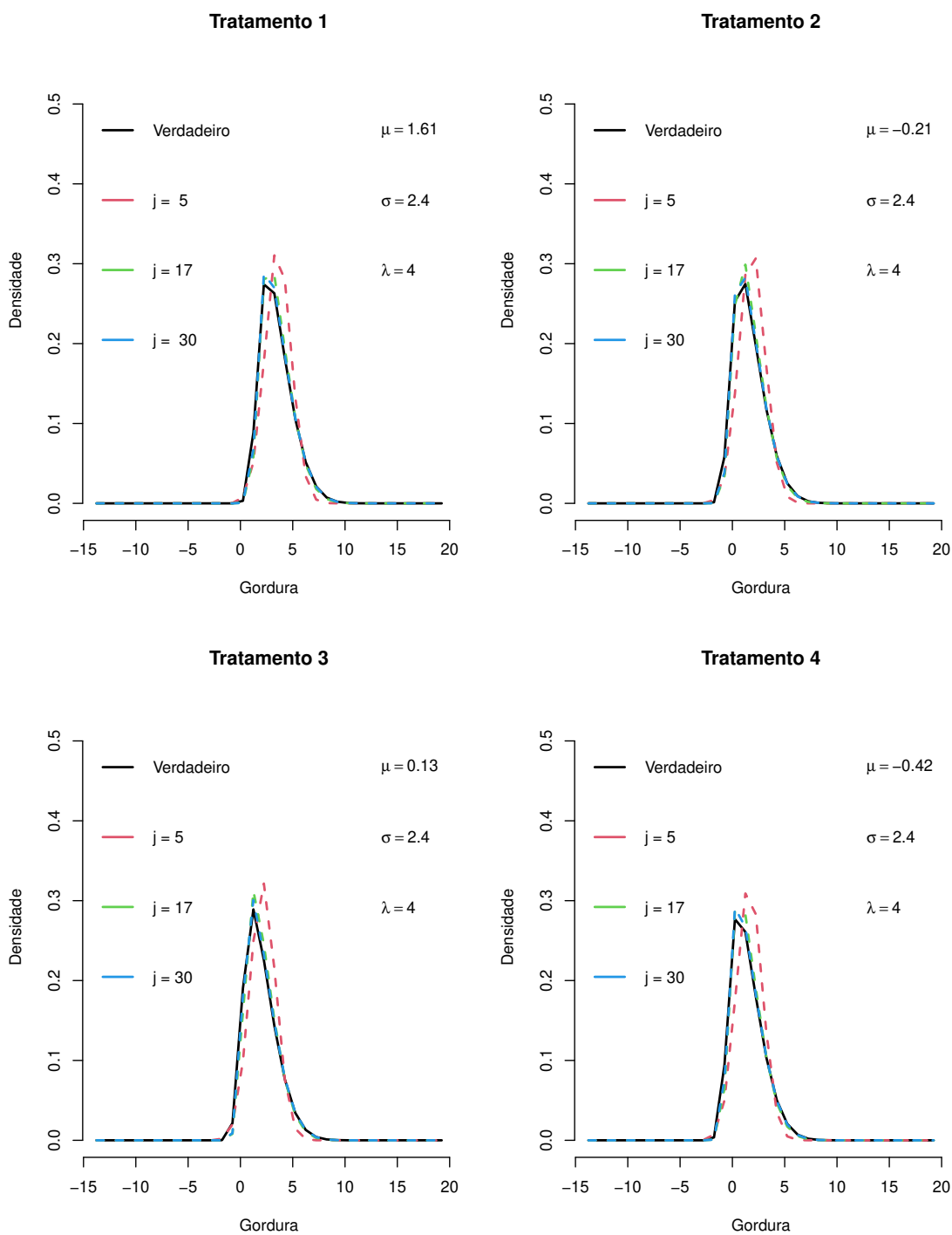


Figura 12: Densidades da distribuição *skew-normal* com $\lambda = 4$, nos valores verdadeiros dos parâmetros e nas estimativas médias para $n=20$; $n=68$; $n=120$.

5 Conclusões

Neste trabalho, nós analisamos dois conjuntos de dados a fim de comparar o ajuste de dois modelos a cada um dos conjuntos de dados, sendo que os modelos foram um com erro distribuído normalmente e outro com erro *skew-normal*. A primeira variável aleatória mensurou a temperatura média de Piracicaba no ano de 2011, e a segunda avaliou o percentual de gordura de 60 idosos que participaram de 4 programas de treinamento corporal. Além disso, procedeu-se à análise de resíduos quantílicos e estudo de simulação para estes conjuntos de dados.

Na análise dos dados, segundo os critérios de comparação de modelos, AIC, AICc e BIC, a distribuição *skew-normal* apresentou melhor ajuste tanto para os dados de temperatura quanto para os dados de gordura, captando melhor as características intrínsecas aos dados quando comparada a distribuição normal, que por sua vez não possui o parâmetro λ de assimetria.

Na análise de resíduos quantílicos, nós confirmamos os resultados obtidos no ajuste. A distribuição *skew-normal* apresentou melhor ajuste quando comparada a distribuição normal, para os dois conjuntos de dados. Para os dados de temperatura, somente 5,21% dos pontos da distribuição *skew-normal* ficaram fora dos envelopes simulados contra 63% dos pontos para a distribuição normal. Considerando-se os dados de percentual de gordura nenhum ponto ficou fora dos envelopes simulados quando considerada a distribuição *skew-normal* contra 40% dos pontos para a distribuição normal.

Segundo o estudo de simulação, que foi feito somente para avaliar as estimativas dos parâmetros dos modelos com erro *skew-normal*, tanto os parâmetros do modelo de intercepto quanto os parâmetros do modelo do delineamento inteiramente casualizado obtiveram estimativas próximas aos parâmetros verdadeiros, diferindo dos parâmetros verdadeiros, respectivamente, por no máximo 0,15 para a amostra de tamanho 365, e 0,05 para 17 repetições.

Desta maneira, nós concluímos para os bancos de dados utilizados nesse trabalho, que dentre as distribuições normal e *skew-normal*, a distribuição *skew-normal* se ajustou melhor, e segundo o estudo de simulação essa distribuição teve estimadores de máxima verossimilhança precisos, bem próximos dos verdadeiros, ou seja pode ser considerada como opção quando não é viável a utilização da distribuição normal e não se deseja fazer transformação dos dados ou análise não paramétrica de dados assimétricos. Além disso, pode-se inferir que se os dados tiverem indício de assimetria, provavelmente, a distribuição normal não se ajustará bem e que o melhor ajuste será de uma outra distribuição que pode ou não ser a distribuição *skew-normal*.

Referências

- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, Boston, v. 19, n. 6, p. 716-723, 1974.
- ATKINSON, A. C. *Plots, transformations and regression: an introduction to graphical methods of diagnostics regression analysis*, 1. ed. Oxford: Clarendon press, 1987.
- AZZALINI, A. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, v. 12, n. 2, p. 171-178, 1985.
- AZZALINI, A. The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, v. 32, p. 159-188, 2005.
- AZZALINI, A.; CAPITANIO, A. *The skew-normal and related families*, 1. ed. New York: Cambridge University Press, 2014.
- BANZATTO, D. A.; KRONKA, S. N. *Experimentação agrícola*, 4. ed. Jaboticabal: FUNEP, 2006.
- BÉLISLE, A. Convergence theorems for a class of simulated annealing algorithms on \mathbb{R}^d . *Journal of Applied Probability*, v. 29, n. 4, p. 885-895, 1992.
- BOLFARINE, H.; SANDOVAL, M. C. *Introdução à inferência estatística*, 2. ed. Rio de Janeiro: SBM, 2010.
- BURNHAM, K. P.; ANDERSON, D. R. Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods and Research*, v. 33, n. 2, p. 261-304, 2004.
- BYRD, R. H.; LU, P.; NOCEDAL, J.; ZHU, C. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing Journal of Computational and Graphical Statistics*, v. 16, p. 1190-1208, 1995.
- CASELLA, G.; BERGER, R. L. *Inferência estatística*, 2. ed. São Paulo: Cengage Learning, 2010.
- CECON, P. R.; ANDESON, R. S.; NASCIMENTO, M.; FERREIRA, A. *Métodos estatísticos*, 1. ed. Viçosa: Editora UFV, 2012.
- CORDEIRO, G. M.; DEMÉTRIO, C. G. B. *Modelos lineares generalizados e extensões*, 1. ed. Recife, PE: [s.n.], 2013.
- DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals *Journal of Computational and Graphical Statistics*, v. 5, n. 3, p. 236-244, 1996.

- ELOY, R. A. O.; BRAGA, A. S.; FRADE, D. D. R.; PEIXOTO, A. P. B.; SILVA I. D. Distribuição assimétrica t-student tipo 3: uma aplicação a delineamentos inteiramente casualizados. *South American Journal of Basic Education, Technical and Technological*, v. 6, n. 2, p. 55-77, 2019.
- EMILIANO, P. C.; VIVANCO, M. J. F.; MENEZES, F. S. Information criteria: How do they behave in different models? *Computational Statistics and Data Analysis*, v. 69, p. 141-153, 2014.
- FABOZZI, F. J.; FOCARDI, S. M.; RACHEV, S. T.; ARSHANAPALLI, B. G.; HOCHSTOTTER, M. *The basics of financial econometrics: Tools, concepts and asset management applications*, 1. ed. Hoboken, New Jersey: [s.n.], 2014.
- FERREIRA, C. S.; BOLFARINE, H.; LABRA, F. E. V. Diagnostics analysis for skew normal regression models. *Brazilian Journal of Probability and Statistics*, v. 32, n. 3, p. 525-544, 2018.
- FLETCHER, R.; REEVES, C. M. Function minimization by conjugate gradients. *The Computer Journal*, v. 7, n. 2, p. 149-154, 1964.
- FREITAS, T. A. S.; SOUZA, S. S. M.; SANTOSA, L. B.; MENDONÇA, A. V. R. Produtividade de minicepas de três espécies florestais em diferentes tamanhos de tubetes. *Brazilian Journal of Forestry Research*, v. 38, p. 1-11, 2018.
- GUEDES, T. A.; ROSSI, R. M.; MARTINS, A. B. T.; JANEIRO, V.; CARNEIRO, J. W. P. Applying regression models with skew-normal errors to the height of bedding plants of *Stevia rebaudiana* (Bert) Bertoni. *Acta Scientiarum*, v. 36, n. 3, p. 463-468, 2014.
- HESTENES, M. R.; STIEFEL, E. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, v. 49, n. 6, p. 409-436, 1952.
- LACHOS, V. H.; BOLFARINE, H.; MONTENEGRO, L. C. Inference and assessment of local influence in skew-normal null intercept measurement error models. *Journal of Statistical Computation and Simulation*, v. 78, n. 3, p. 395-419, 2008.
- MONTGOMERY, D. C. *Design and analysis of experiments*, 9. ed. New Jersey: John Wiley & Sons, 2017.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, C. *Introduction to the theory of Statistics*, 3. ed. New York: McGraw Hill, 1974.
- NELDER, J. A.; MEAD, R. A simplex method for function minimization. *The Computer Journal*, v. 7, n. 4, p. 308-313, 1965.
- NOCEDAL J.; WRIGHT, S. *Numerical Optimization*, 2. ed. New York: Springer, 2006.
- OWEN, D. B. Tables for computing bivariate normal probabilities. *Annals of Mathematical Statistics*, v. 27, n. 4, p. 1075-1090, 1956.

- PAES, A. T. O que fazer quando a distribuição não é normal? *Einstein: Educação Continuada em Saúde*, v. 2, n. 1, p. 3-4, 2009.
- PAULA, G. A. *Modelos de regressão com apoio computacional*, 1. ed. São Paulo: [s.n.], 2010.
- PEWSEY, A. Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, v. 27, n. 7, p. 859-870, 2000.
- PINO, F. A. A questão da não normalidade: uma revisão. *Revista de Economia Agrícola*, v. 61, n. 2, p. 17-33, 2014.
- PINTO, L. I. F.; ZAMBELLI, R. A.; SANTOS JÚNIOR, E. C.; PONTES, D. F. Desenvolvimento de cerveja artesanal com acerola (*Malpighia emarginata* DC) e Abacaxi (*Ananas comosus* L. Merrill). *Revista Verde de Agroecologia e Desenvolvimento Sustentável*, v. 10, n. 4, p. 67-71, 2015.
- R DEVELOPMENT CORE TEAM. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, 2020. Disponível em: <<http://www.r-project.org>>.
- SAID, K. K.; NING, W.; TIAN, Y. Detecting changes in linear regression models with skew normal errors. *Random Operators and Stochastic Equations*, v. 26, p. 1-10, 2018.
- SCHWARZ, G. Estimating the dimensional of a model. *Annals of Statistics*, v. 6, n. 2, p. 461-464, 1978.
- SEARLE, S. R. *Linear models*, 1. ed. New York: ISBN, 1997.

Apêndice A

Script análise de dados

```

\hspace{-1.6cm} dadostemperatura <- c(24.75, ..., 25.20)
dadosdic <- data.frame(tr,y)

tr y
A -1.2
A -2.1
A -1.1
...
D -0.5
D -6.0

# Funções para, respectivamente, critérios de informação, intervalo de confiança e
funções de verossimilhança das distribuições normal e skew-normal

criterios = function(value,d,n)\{
l = 2*value
AIC = l + 2*d
...
return(resul)\}
IC = function(parametros,hessiana,n){
  Var = solve(hessiana)
  EP = diag(sqrt(Var))
  tvalue = parametros/EP
  $ \cdots $
return(resul)\}
Logver = function(param,y,x){
  nbeta <- ncol(x)
  beta <- param[1:nbeta]
  ...
return(soma)\} \vspace{0.5cm}\
Logver = function(param,y,x){
  nbeta <- ncol(x)
  beta <- param[1:nbeta]
  sigma <- param[nbeta+1]
  lambda <- param[nbeta+2]

```

```

...
return(soma)\}

# Foi realizado o ajuste utilizando o pacote ExpDes.pt e utilizado a função
summary.lm para os chutes iniciais, criado um vetor para estes chutes que
foi utilizado na função \textit{optim} e por último chamadas as funções.

dic(tr, y, quali = TRUE, mcomp = "sk", sigT = 0.05, sigF = 0.05)
tra <- relevel(tr,4);tra
is.factor(tra)
M1 <- aov( y ~ tra, data = dados)
summary.lm(M1)

vaini <- $c(x_1,x_2, \cdots, x_n)$
valorinicial = optim(par = vaini, Logver, y = y, x=x, method = "Nelder-Mead",
hessian = TRUE)

n = length(y)
estima = valorinicial
dados11 = criterios(estima\$value)
parametros = estima\par
hessiana = estima\hessian
IC = round(IC(parametros,hessiana,n))

\subsection*{Script análise de resíduos}

\hspace{-1.6cm} require(sn)
require(fGarch)
require(hnp)
require(gamlss)
require(ExpDes.pt)

attach(dados)
y <- GORD
names(dados)

Verossimilhanca sn
logverosn <- function(param,y,x){

```

```

    nbeta <- ncol(x)
...
    soma = - sum(lv)
    return(soma)}

fit.sn <- function(data,tratamento){
    y1 = data
...
object <- list(parametros = fit\$par, residuos = t, matrix = x)
    return(object)}

d.fun <- function(obj) obj\$residuos

s.fun <- function(n,obj) {
    beta <- obj\$parametros[1:(length(obj\$parametros)-2)]
...
rsn(y, xi = mu, omega = sigma, alpha = lambda)}

my.data <- data.frame(y,trat)

f.fun <- function(y.) fit.sn(y.,tra)

# Ajuste do modelo utilizando dados reais
fit.Sn <- fit.sn(data = y,tratamento = tra)

\subsection*{Script estudo de simulação}

\hspace{-1.62cm} Geradora <- function(Namostra = NULL, Trep = NULL,
Ntrat = NULL, par, t = NULL,\\ print.leg=FALSE, auto.save=FALSE,
nomeimagem="padrao",metodo="Nelder-Mead", tamanho.leg=NULL,legenda.x=NULL,
legenda.y=NULL){
options(warn=-1)
carregar <- function(x){
    for(i in x){
        if(!require(i, character.only = TRUE)){
            install.packages(i, dependencies = TRUE)
            require(i, character.only = TRUE )
        }}
}
suppressPackageStartupMessages({

```

```

carregar(c("sn","fGarch","gamlss","tools","stargazer","tables"))
...
compa <- c(beta,sigma,lambda)
EQM <- function(x) var(x) + (mean(x) - compa)2
resultado[[j]] <- data.frame
("Estimativa" = c(round(colMeans(parametro),2))
"Variancia" = c(round(colVars(parametro),2)),
"EQM" = c(round(diag(apply(parametro, 2, EQM)),2)),
"verdadeiro" = c(beta,sigma,lambda),
row.names=c(paste("beta", 1:length(beta), sep = "-"),
"$sigma$","$lambda$"))}
...
for(h in 1:caso){
invisible(readline(prompt="Copie o codigo abaixo e Press [enter] to continue"))
stargazer(resultado[[h]],summary = FALSE)
save <- booktabs()
latex(tabular( (Parametro) ~ Format(digits = 2)*
(Namostra1 + Namostra2 + Namostra3)*(mean + var + EQM), data=Teste))
return(resultado)}

A <- Geradora(Namostra = 1000, Trep = c(5,17,30), Ntrat = 4,
par = c(1.61,-1.82,-1.48,-2.03,2.40,4), metodo = "BFGS",
legenda.x = "Gordura", legenda.y="Densidade" ,tamanho.leg=1).

```

Apêndice B

Considerando o delineamento inteiramente casualizado, para determinarmos sua função de verossimilhança e estimadores de máxima verossimilhança precisamos primeiramente verificar se Y_{ij} tem de fato distribuição *skew-normal*.

Seja $Z \sim SN(0, 1, \lambda)$ e

$$Y = \mu + \sigma Z, \quad \text{em que } \mu \in \mathbb{R}, \sigma \in \mathbb{R}_+^* \quad (29)$$

queremos determinar a distribuição de Y , e para isto utilizaremos o Teorema 2.

Teorema 2 (Casella e Berger, 2010; p.47). Seja X com função densidade de probabilidade $f_X(x)$ e $y = g(x)$, em que, $g(x)$ é uma função monótona. Considere os espaços amostrais $\mathcal{X} = \{x; f_X(x) > 0\}$ e $\mathcal{Y} = \{y; y = g(x), \text{ para algum } x \in \mathcal{X}\}$. Suponha que $f_X(x)$ seja contínua em \mathcal{X} e que $g^{-1}(y)$ tenha derivada contínua em \mathcal{Y} . Então, a função densidade de probabilidade de Y é dada por:

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} (g^{-1}(y)) \right|, & y \in \mathcal{Y} \\ 0, & \text{caso contrário} \end{cases}.$$

Considerando a variável aleatória $Y = \mu + \sigma Z$, temos que

$$y = g(z) = \mu + \sigma z \quad \Rightarrow \quad g^{-1}(y) = \frac{y - \mu}{\sigma},$$

assim

$$\begin{aligned} f_Y(y) &= f_Z(g^{-1}(y)) \cdot \left| \frac{d}{dy} (g^{-1}(y)) \right| \\ &= f_Z\left(\frac{y - \mu}{\sigma}\right) \cdot \left| \frac{d}{dy} \left(\frac{y - \mu}{\sigma}\right) \right| \\ &= f_Z\left(\frac{y - \mu}{\sigma}\right) \cdot \frac{1}{\sigma}. \end{aligned}$$

Como $Z \sim SN(0, 1, \lambda)$, temos que

$$f_Z(z|\lambda) = 2 \cdot \phi(z) \cdot \Phi(\lambda z)$$

assim

$$f_Y(y) = f_Y(y|\mu, \sigma, \lambda) = \frac{2}{\sigma} \cdot \phi\left(\frac{y - \mu}{\sigma}\right) \cdot \Phi\left(\lambda \frac{y - \mu}{\sigma}\right),$$

a qual é a função densidade de probabilidade de $Y \sim SN(\mu, \sigma, \lambda)$.

Utilizando o resultado anterior para determinar a distribuição de Y_{ij} para o DIC com $\epsilon \sim SN(0, 1, \lambda)$. Assim, considerando o delineamento inteiramente casualizado com erro *skew-normal*,

teremos a seguinte variável aleatória

$$Y_{ij} = m + \tau_i + \epsilon_{ij}, \quad \text{em que } \epsilon_{ij} \sim SN(0, 1, \lambda).$$

Por analogia, podemos considerar a função monótona

$$\begin{aligned} y_{ij} &= g(\epsilon_{ij}) = m + \tau_i + \epsilon_{ij} \\ \Rightarrow g^{-1}(y_{ij}) &= y_{ij} - (m + \tau_i) \end{aligned}$$

assim,

$$\begin{aligned} f_{Y_{ij}}(y_{ij}) &= f_{\epsilon_{ij}}(g^{-1}(y_{ij})) \cdot \left| \frac{d}{dy_{ij}}(g^{-1}(y_{ij})) \right| \\ &= f_{\epsilon_{ij}}(y_{ij} - (m + \tau_i)) \cdot \left| \frac{d}{dy_{ij}}(y_{ij} - (m + \tau_i)) \right| \\ &= f_{\epsilon_{ij}}(y_{ij} - (m + \tau_i)) \end{aligned}$$

como $\epsilon_{ij} \sim SN(0, 1, \lambda)$, então

$$f_{\epsilon_{ij}}(z|\lambda) = 2 \cdot \phi(z) \cdot \Phi(\lambda z)$$

e portanto,

$$f_{Y_{ij}}(y_{ij}) = 2 \cdot \phi(y_{ij} - (m + \tau_i)) \cdot \Phi[\lambda(y_{ij} - (m + \tau_i))]$$

é a função densidade de probabilidade da variável aleatória Y_{ij} , dada por (29) se tomarmos

$$\mu = m + \tau_i; \quad \sigma = 1; \quad \text{e } Z = \epsilon_{ij}.$$

Assim,

$$f_{Y_{ij}}(y_{ij}) = \frac{2}{1} \cdot \phi\left(\frac{y_{ij} - (m + \tau_i)}{1}\right) \cdot \Phi\left(\lambda \frac{y_{ij} - (m + \tau_i)}{1}\right)$$

e portanto

$$Y_{ij} \sim SN(m + \tau_i, 1, \lambda).$$

Consideraremos agora, o modelo para o DIC com $\epsilon \sim SN(0, 1, \lambda)$ e com introdução do parâmetro de escala. Se fizermos a seguinte adequação, na qual é acrescentado o parâmetro σ de escala, teremos

$$Y_{ij} = m + \tau_i + \sigma \epsilon_{ij}, \quad \text{com } \epsilon_{ij} \sim SN(0, 1, \lambda),$$

por analogia ao demonstrado anteriormente,

$$f_{Y_{ij}}(y_{ij}) = \frac{2}{\sigma} \cdot \phi\left(\frac{y_{ij} - (m + \tau_i)}{\sigma}\right) \cdot \Phi\left(\lambda \frac{y_{ij} - (m + \tau_i)}{\sigma}\right)$$

ou seja,

$$Y_{ij} \sim SN(m + \tau_i, \sigma, \lambda).$$

como queríamos mostrar.