UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
DEPARTAMENTO DE MEDICINA SOCIAL

ROBERTO MOLINA DE SOUZA

**Modelagem estatística em estudos de
bioequivalência sob o enfoque Bayesiano**

Ribeirão Preto
2015

ROBERTO MOLINA DE SOUZA

Modelagem estatística em estudos de bioequivalência sob o enfoque Bayesiano

Tese apresentada ao Departamento de Medicina Social da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do título de Doutor em Ciências Médicas.

Área de Concentração: Saúde na Comunidade.

Orientador: Prof. Dr. Jorge Alberto Achcar.

Ribeirão Preto
2015

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

FICHA CATALOGRÁFICA

Souza, Roberto Molina

Modelagem estatística em estudos de bioequivalência sob o enfoque Bayesiano 130p.:il.; 29,7cm

Tese (Doutorado) apresentada à Faculdade de Medicina de Ribeirão Preto/USP - Área de concentração: Saúde na Comunidade.

Orientador: Achcar, Jorge Alberto

1. Bioequivalência. 2. Inferência Bayesiana. 3. Medicamento genérico 4. Distribuições assimétricas 5. Funções cópulas

FOLHA DE APROVAÇÃO

Roberto Molina de Souza

Modelagem estatística em estudos de bioequivalência sob o enfoque Bayesiano

Tese apresentada ao Departamento de Medicina Social da Faculdade
de Medicina de Ribeirão Preto da Universidade de São Paulo para
obtenção do título de Doutor em Ciências Médicas.

Área de Concentração: Saúde na Comunidade

Aprovado em: ____/____/____

Banca Examinadora

Prof(a). Dr(a).: _____

Instituição: _____ Assinatura: _____

Prof(a). Dr(a).: _____

Instituição: _____ Assinatura: _____

Prof(a). Dr(a).: _____

Instituição: _____ Assinatura: _____

Prof(a). Dr(a).: _____

Instituição: _____ Assinatura: _____

Prof(a). Dr(a).: _____

Instituição: _____ Assinatura: _____

*“Creio que as maiores inovações do século XXI
estarão na intersecção da biologia com a tecnologia”. Steve Jobs*

— —

Agradecimentos

A Deus pelas inúmeras bençãos nesta caminhada.

A minha esposa, Flávia Leite, que sempre esteve do meu lado em todos os momentos.

Aos meus pais Gustavo e Odacir e aos meus irmãos Dênis (e esposa Silvana) e Elisabeth (e marido Bruno) pelo apoio nas minhas decisões.

Pelo carinho da minha avó Dolores e meu Tio Domingos (in memoriam).

Ao meu orientador, Prof. Dr. Jorge Alberto Achcar, pela sua disposição e capacidade de entender e respeitar minhas limitações e explorar minhas qualidades.

A Prof. Dra. Sílvia Regina Costa Lopes, ao Prof. Dr. Antonio Carlos Pedroso de Lima, ao Prof. Dr. Eduardo Barbosa Coelho e ao Prof. Dr. Altacílio Aparecido Nunes por aceitarem o convite de compor a banca examinadora e por sua valiosa contribuição.

Ao coordenador deste programa de pós-graduação, prof. Dr. Edson Zangiacomi Martinez, pela sua contribuição neste trabalho e demais parcerias.

Ao prof. Dr. Josmar Mazucheli que esteve sempre disposto a ajudar-me desde a graduação, tendo grande influência na minha escolha pela área acadêmica.

Aos amigos Emílio, Josmar e Davi, companheiros de longa data, pelas longas conversas e momentos de descontração.

A Mônica Elisabet Knack (secretária do departamento de Medicina Social), Paula Maria Pereira e Sérgio Carlos do Nascimento (secretários da Pós Graduação) pela ajuda prestada frente as burocracias exigidas neste período.

Ao Prof. Dr. Heleno Bolfarine, Prof. Dr. Edson Zangiacomi Martinez e Prof. Dra. Vera Lúcia Lanchote que formaram a banca do exame de qualificação.

Aos membros do LCCA-Laboratório de Computação Científica Avançada da Universidade de São Paulo pelo auxílio no uso de computadores de alto desempenho.

Ao diretor da UTFPR, Campus Cornélio Procópio, Prof. Me. Devanil Antônio Francisco, ao gerente de pesquisa, Prof. Dr. Luciano Tadeu Esteves Pansanato e ao gerente de ensino, Prof. Me. Edson Luis Bassetto que sempre apoiaram minha pós graduação.

Aos amigos e colegas de trabalho da UTFPR (parte 1: Londrina) Emílio, Rafael, Vinícius, Gláucia, Fabiano, pelos momentos de descontração.

Aos amigos e colegas de trabalho da UTFPR (parte 2: Cornélio Procópio) André, Cristiane, Evandro, Devanil, Sandra, Armando, Maria Lúcia, Elisabete e Isaac (in memoriam) também pelos momentos de descontração.

Aos amigos e colegas de trabalho da UTFPR (parte 3: Demais localidades) Milton, Simone (e marido Paulo) e Joselene mais uma vez pelos momentos de descontração.

Aos professores da UPtime Londrina (escola de Inglês) Loui, Ana e Julia e ao proprietário Abner que ofereceram um curso em alto nível que me encorajou na participação de um evento internacional.

Agradeço também a **FA** (Fundação Araucária) do Paraná, pelo fomento concedido nos 6 meses finais do Doutorado em forma de bolsa (CP13/2014).

RESUMO

SOUZA, R. M. “**Modelagem estatística em estudos de bioequivalência sob o enfoque Bayesiano**”. Tese (Doutorado) Faculdade de Medicina de Ribeirão Preto - USP, Ribeirão Preto - SP - Brasil, 2015

O interesse pelos estudos de bioequivalência iniciou-se na década de 60, sendo o *FDA* (EUA) a primeira agência reguladora a se interessar por esta questão. No Brasil, uma lei de 1999 regulamentou o medicamento genérico no país, sendo este um importante meio de acesso aos medicamentos pela população e fazendo parte da política de medicamentos do SUS. No Brasil, a ANVISA é responsável por inspecionar os centros de bioequivalência bem como dar as diretrizes para estes. Um modelo paramétrico padrão para a etapa estatística é disponibilizado para a decisão de bioequivalência média e espera-se que este ajuste-se aos dados obtidos nos estudos de bioequivalência, o que nem sempre acontece. Neste sentido, é proposto nesta tese o uso de modelos paramétricos mais abrangentes baseados em outras distribuições de probabilidade para a decisão de bioequivalência média e que possam modelar a assimetria dos dados, dispensando o uso da transformação logarítmica para os parâmetros farmacocinéticos o que afeta a amplitude dos limites de bioequivalência. Propõe-se também o uso de modelos bivariados para a tomada conjunta da decisão de bioequivalência média, quando são analisados simultaneamente dois parâmetros farmacocinéticos. Foram utilizados métodos Bayesianos para a estimação dos parâmetros dado a flexibilidade deste enfoque quando combinado ao uso dos métodos *MCMC* facilitados a partir do uso de softwares livres. Nesta tese é apresentado um estudo do poder empírico dos testes de hipóteses para os modelos univariados propostos bem como são introduzidos quatro exemplos, sendo os três primeiros voltados à aplicação da decisão de bioequivalência média e o quarto para a aplicação da bioequivalência populacional e individual. Nos três primeiros exemplos foram observados ganhos em termos de ajuste dos novos modelos propostos aos dados com relação ao modelo padrão segundo os critérios de seleção de modelos utilizados. O exemplo quatro apresenta uma análise padrão de bioequivalência populacional e individual com o código computacional para a obtenção dos resultados disponível no apêndice A bem como outros códigos para os demais exemplos. Embora a padronização das análises estatísticas para os estudos de bioequivalência seja importante, não deve-se esperar que um modelo padrão ajuste-se a qualquer conjunto de dados originados destes tipo de estudos. Neste sentido, são apresentadas alternativas que podem auxiliar o pesquisador na tomada de uma decisão em termos de bioequivalência média mais segura.

palavras-chaves: bioequivalência, inferência Bayesiana, medicamento genérico, distribuições assimétricas, funções cópulas

ABSTRACT

SOUZA, R. M. “**Statistical modeling in bioequivalence studies under Bayesian approach**”. Thesis (Doctorate) Faculdade de Medicina de Ribeirão Preto - USP, Ribeirão Preto - SP - Brasil, 2015

The interest in bioequivalence studies began in the early 1960s especially in the United States of America where the FDA was the first regulatory agency to show interest upon this issue. In Brazil, this interest started in 1999 the year when a law regulated the generic drugs in the country. The ANVISA is the Brazilian regulatory agency responsible for inspecting the bioequivalence centers and giving guidelines for this issue. In general, a standard parametric model for the statistical step is indicated for the average bioequivalence decision and this model is expected to be fitted by the data obtained in the bioequivalence studies. In some cases, this model would not be appropriate. In this way, this thesis proposes the use of more comprehensive parametric models based on other probability distributions for the average bioequivalence decision and that can model asymmetrical data, a common situation in bioequivalence studies. In addition, there is no need of a logarithmic transformation for the pharmacokinetic parameters which could affect the range of the bioequivalence limits. We also propose the use of parametrical bivariate models for the joint decision of the average bioequivalence decision, since these measures are usually analyzed simultaneously with two pharmacokinetic parameters. We use Bayesian methods to estimate the parameters, given the great flexibility of this approach when combined with the use of MCMC methods using free available softwares. This thesis also presents a study of the empirical power of hypothesis testing for the proposed univariate models and four examples are introduced. In the examples one, two and three we apply the average bioequivalence decision and in the fourth example we consider for the implementation of population and individual bioequivalence. In the examples one, two and three were observed gains in the fitting of the proposed new models for the data where some existing approaches were used in the selection criteria for the proposed models. Example four provides a standard analysis of population and individual bioequivalence with the computer code for obtaining the results available in the Appendix A, as well as other codes. Although the standardization of statistical analysis for bioequivalence studies is important, a standard model is not expected to be well fitted to any data set originated by such studies. In this way, we present alternatives that can help researchers in making a decision in terms of average bioequivalence with more security.

keywords: bioequivalence, Bayesian inference, generic drugs, asymmetric distributions, copula functions

Lista de Abreviaturas ou Siglas

<i>ANOVA</i>	<i>Analysis Of Variance</i>
ANVISA	Agência Nacional de Vigilância Sanitária
ASC_{0-t}	Área sob a curva de concentração plasmática do tempo 0 ao último tempo (t) observado
<i>BUGS</i>	<i>Bayesian inference Using Gibbs Sampling</i>
C_{max}	Concentração máxima
CV	Coeficiente de Variação
<i>DIC</i>	<i>Deviance Information Criterion</i>
<i>FDA</i>	<i>Food and Drug Administration</i>
FGM	Farlie-Gumbel-Morgenstern
GGE	Gama Generalizada Estendida
IC_r	Intervalo de Credibilidade
<i>HPB</i>	<i>Health Protection Branch</i>
<i>HPD</i>	<i>Highest Posterior Density</i>
<i>JAGS</i>	<i>Just Another Gibbs Samples</i>
ln	logarítmo neperiano
<i>LPML</i>	<i>Logarithm of the Pseudo Marginal Likelihood</i>
<i>MCMC</i>	<i>Markov Chain Monte Carlo</i>
MO	Marshall-Olkin
R	Referência
<i>SN</i>	<i>Skew-Normal</i>
<i>ST</i>	<i>Skew-t</i>
T	Teste
T_{max}	Tempo em que ocorre a concentração máxima
<i>TN</i>	<i>Tilted Normal</i>
WMW	Wilcoxon-Mann-Whitney

Lista de Figuras

1	Representação gráfica de alguns parâmetros farmacocinéticos.	18
2	Gráficos quantil-quantil para os resíduos do modelo padrão de alguns conjuntos de dados (<i>i</i>).	28
3	Gráficos quantil-quantil para os resíduos do modelo padrão de alguns conjuntos de dados (<i>ii</i>).	29
4	Gráficos para a densidade da distribuição GGE assumindo diferentes valores para os parâmetros desta distribuição.	32
5	Gráficos para a densidade da distribuição <i>TN</i> assumindo diferentes valores para os parâmetros desta distribuição.	36
6	Gráficos para a densidade da distribuição <i>SN</i> assumindo diferentes valores para os parâmetros desta distribuição.	38
7	Gráficos para a densidade da distribuição <i>ST</i> assumindo diferentes valores para os parâmetros desta distribuição.	40
8	Região dos limites de bioequivalência para uma decisão simultânea de dois parâmetros farmacocinéticos.	45
9	Histogramas para os parâmetros farmacocinéticos segundo as formulações - Exemplo 1.	60
10	Gráfico de dispersão segundo C_{max} e ASC_{0-t} para o modelo GGE - Exemplo 1.	60
11	Exemplo de duas cadeias que convergem com sucesso.	63
12	Gráficos quantil-quantil para os parâmetros farmacocinéticos segundo os modelos univariados ajustados introduzidos na seção 3.1 - Exemplo 1. .	64
13	Histogramas para os parâmetros farmacocinéticos com as curvas teóricas das funções densidade de probabilidade ajustadas - Exemplo 1.	65
14	Gráficos de dispersão das médias a posteriori segundo cada observação para o modelo GGE - Exemplo 1.	67
15	Valores de <i>DIC</i> e <i>LPML</i> variando o parâmetro v da distribuição multivariada <i>ST</i> - Exemplo 1.	68
16	Região <i>HPD</i> para a decisão conjunta de bioequivalência média para C_{max} e ASC_{0-t} do modelo GGE sob a função cópula de Clayton - Exemplo 1.	69

17	Gráficos de dispersão das médias a posteriori segundo cada observação para o modelo GGE sob a função cópula de Clayton - Exemplo 1. . . .	70
18	Histogramas para os parâmetros farmacocinéticos segundo as formulações - Exemplo 2.	71
19	Gráfico de dispersão segundo C_{max} e ASC_{0-t} para o modelo GGE - Exemplo 2.	71
20	Gráficos quantil-quantil para os parâmetros farmacocinéticos segundo os modelos univariados ajustados introduzidos na seção 3.1 - Exemplo 2. .	73
21	Histogramas para os parâmetros farmacocinéticos com as curvas teóricas das funções densidade de probabilidade ajustadas - Exemplo 2.	73
22	Gráficos de dispersão das médias a posteriori segundo cada observação para o modelo GGE - Exemplo 2.	75
23	Região <i>HPD</i> para a decisão conjunta de bioequivalência média para C_{max} e ASC_{0-t} do modelo GGE sob a função cópula Gaussiana - Exemplo 2.	76
24	Gráficos de dispersão das médias a posteriori segundo cada observação para o modelo GGE sob a função cópula Gaussiana - Exemplo 2. . . .	77
25	Histogramas para os parâmetros farmacocinéticos segundo as formulações - Exemplo 3.	78
26	Gráfico de dispersão segundo C_{max} e ASC_{0-t} - Exemplo 3.	78
27	Gráficos quantil-quantil para os parâmetros farmacocinéticos segundo os modelos univariados ajustados introduzidos na seção 3.1 - Exemplo 3. .	79
28	Histogramas para os parâmetros farmacocinéticos com as curvas teóricas das funções densidade de probabilidade ajustadas - Exemplo 3.	79
29	Gráficos de dispersão das médias a posteriori segundo cada observação para o modelo GGE - Exemplo 3.	81
30	Região <i>HPD</i> para a decisão conjunta de bioequivalência média para C_{max} e ASC_{0-t} do modelo GGE sob a função cópula Gaussiana - Exemplo 3.	82
31	Gráficos de dispersão das médias a posteriori segundo cada observação para o modelo GGE sob a função cópula Gaussiana - Exemplo 3. . . .	83
32	Gráfico quantil-quantil para os resíduos do modelo ajustado - Exemplo 4.	85

Lista de Tabelas

1	Esquema da disposição dos dados para um modelo cruzado 2×2 (2 períodos, 2 formulações e 2 sequências).	24
2	Resumo da análise de resíduos para alguns conjuntos de dados.	27
3	Esquema da disposição dos dados para um modelo cruzado com 4 períodos, 2 tratamentos, 2 sequências e 2 réplicas para cada tratamento. .	50
4	Estudo do poder empírico do teste para os modelos propostos.	57
5	Medidas resumos para os dados do exemplo 1.	59
6	Valores de <i>DIC</i> e <i>LPML</i> para os modelos propostos na seção 3.1 para cada parâmetro farmacocinético - Exemplo 1.	65
7	Intervalos <i>HPD</i> para a decisão de bioequivalência média - Exemplo 1. .	66
8	Valores de <i>DIC</i> e <i>LPML</i> e intervalos <i>HPD</i> para os parâmetros de correlação dos modelos propostos na seção 3.2 - Exemplo 1.	68
9	Medidas resumos para os dados do exemplo 2.	70
10	Valores de <i>DIC</i> e <i>LPML</i> para os modelos propostos na seção 3.1 para cada parâmetro farmacocinético - Exemplo 2.	74
11	Intervalos <i>HPD</i> para a decisão de bioequivalência média - Exemplo 2. .	74
12	Valores de <i>DIC</i> e <i>LPML</i> e intervalos <i>HPD</i> para os parâmetros de correlação dos modelos propostos na seção 3.2 - Exemplo 2.	76
13	Medidas resumos para os dados do exemplo 3.	78
14	Valores de <i>DIC</i> e <i>LPML</i> para os modelos propostos na seção 3.1 para cada parâmetro farmacocinético - Exemplo 3.	80
15	Intervalos <i>HPD</i> para a decisão de bioequivalência média - Exemplo 3. .	80
16	Valores de <i>DIC</i> e <i>LPML</i> e intervalos <i>HPD</i> para os parâmetros de correlação dos modelos propostos na seção 3.2 - Exemplo 3.	82
17	Medidas resumos para os dados do exemplo 4.	83
18	Intervalos <i>HPD</i> para as decisões de bioequivalência individual, populacional e média - Exemplo 4.	85
19	Médias a posteriori (e desvio padrão) - Exemplo 1 - Univariado.	114
20	Médias a posteriori (e desvio padrão) - Exemplo 1 - Bivariado.	115

21	Médias a posteriori (e desvio padrão) - Exemplo 2 - Univariado.	116
22	Médias a posteriori (e desvio padrão) - Exemplo 2 - Bivariado.	117
23	Médias a posteriori (e desvio padrão) - Exemplo 3 - Univariado.	118
24	Médias a posteriori (e desvio padrão) - Exemplo 3 - Bivariado.	119

Sumário

1	INTRODUÇÃO	17
2	MOTIVAÇÃO E OBJETIVOS	23
3	METODOLOGIA	31
3.1	Bioequivalência média - Enfoque univariado	31
3.1.1	Distribuição Gama Generalizada Estendida (GGE)	31
3.1.2	Distribuição <i>Tilted Normal</i> (<i>TN</i>)	35
3.1.3	Distribuição <i>Skew-Normal</i> (<i>SN</i>)	37
3.1.4	Distribuição <i>Skew-t</i> (<i>ST</i>)	39
3.2	Bioequivalência média - Enfoque multivariado	41
3.2.1	Distribuição GGE sob funções cópulas	41
3.2.2	Distribuição <i>TN</i> sob funções cópulas	45
3.2.3	Distribuição <i>SN</i> multivariada	46
3.2.4	Distribuição <i>ST</i> multivariada	48
3.3	Bioequivalência populacional e individual	49
3.4	Seleção de modelos	52
3.4.1	Gráfico quantil-quantil e teste de aderência	52
3.4.2	Ajuste da função densidade de probabilidade a um histograma .	53
3.4.3	<i>Deviance Information Criterion</i> (<i>DIC</i>)	53
3.4.4	<i>Conditional Predictive Ordinate</i> (<i>CPO</i>)	53
3.4.5	<i>Logarithm of the Pseudo Marginal Likelihood</i> (<i>LPML</i>)	54
3.5	Poder do teste	54
4	APLICAÇÕES	59
4.1	Exemplo 1: Ondansetrona	59
4.1.1	Conjunto de dados	59
4.1.2	Elicitação das distribuições a priori	61

4.1.3	Resultados	64
4.2	Exemplo 2: Risperidona	69
4.2.1	Conjunto de dados	69
4.2.2	Elicitação das distribuições a priori	69
4.2.3	Resultados	72
4.3	Exemplo 3: Tinidazol	77
4.3.1	Conjunto de dados	77
4.3.2	Resultados	77
4.4	Bioequivalência individual e populacional: <i>FDA</i>	83
4.4.1	Conjunto de dados	83
4.4.2	Elicitação das distribuições a priori	84
4.4.3	Resultados	84
5	CONCLUSÕES E PERSPECTIVAS FUTURAS	87
	REFERÊNCIAS BIBLIOGRÁFICAS	91
	APÊNDICES	99
A	Códigos BUGS e JAGS para os exemplos	99
B	Estimativas dos parâmetros	113
C	Artigos submetidos para publicação	121
D	Capítulo de livro em evento	129

1 INTRODUÇÃO

A partir de um levantamento bibliográfico realizado por BARRET et al. (2000), o interesse em buscar evidências para a diferença entre biodisponibilidades para duas formulações iniciou-se por volta de 1960. Nesta direção, GRIZZLE (1965) publicou um trabalho usando um modelo de Análise de Variância (*ANOVA*) para comparar biodisponibilidades utilizando alguns parâmetros farmacocinéticos.

Na década de 70 o *FDA* (*Food and Drug Administration*), dos Estados Unidos da América, começou a questionar estes tipos de estudos (BARRET et al., 2000). Como o objetivo destes estudos estava voltado para a fabricação de medicamentos com menor custo (METZLER, 1974) e que apresentassem o mesmo efeito de medicamentos com segurança e eficácia comprovada, o interesse no uso de testes estatísticos para as comparações estava na não rejeição da igualdade de alguns parâmetros farmacocinéticos segundo as formulações em estudo. Neste sentido, WESTLAKE (1972) propôs um método de decisão a partir de intervalos de confiança para a diferença (ou razão) de médias como alternativa aos métodos usuais de testes de hipóteses. Neste método, o autor esperava que os intervalos de confiança construídos para a diferença média dos parâmetros farmacocinéticos segundo as formulações em estudo estivessem totalmente contidos em intervalos clinicamente estabelecidos para que fosse evidenciada a biodisponibilidade entre estas.

METZLER (1974) levantou várias questões biológicas no contexto de biodisponibilidade que precederam os métodos estatísticos utilizados para as comparações destas, por exemplo, o termo biodisponibilidade de duas formulações que tornou-se popularmente a contração do termo biodisponibilidade relativa sendo o acesso a esta relativo à comparação de alguns parâmetros farmacocinéticos. Segundo este autor, a contribuição da Estatística neste processo se dá em três etapas: (i) escolha, estimação e comparação dos parâmetros farmacocinéticos relativos à biodisponibilidade; (ii) formulação de um modelo estatístico e o planejamento do experimento e (iii) determinação do tamanho amostral.

Em 1977 o *FDA* publicou um documento regulatório referente a procedimentos para testes de bioequivalência e biodisponibilidade para seres humanos. Na década de 80, alguns autores apresentaram estudos sobre o poder dos testes estabelecidos para bioequivalência na literatura considerando a transformação logarítmica como também outras possibilidades de construção do teste de hipóteses (MANDALLAZ; MAU, 1981; ANDERSON; HAUCK, 1983).

SELWYN et al. (1981) apresentaram um artigo em que introduziram o uso de métodos Bayesianos em estudos de bioequivalência, explorando a distribuição a posteriori da diferença de médias na decisão de bioequivalência respeitando os limites clinicamente

estabelecidos em um planejamento cruzado com dois períodos e duas sequências na administração das formulações. No mesmo sentido, SELWYN e HALL (1984) publicaram um outro artigo em que consideraram mais períodos e sequências em um delineamento experimental.

Ainda sob o ponto de vista da utilização de métodos Bayesianos em estudos de bioequivalência, RACINE et al. (1986) publicaram um trabalho em que consideraram vários métodos Bayesianos aplicados à dados da indústria farmacêutica da época, dentre estes, algumas aplicações em bioequivalência. No ano seguinte, RACINE-POON et al. (1987) publicaram um trabalho em que realizaram um estudo de bioequivalência em dois estágios explorando, mais uma vez a abordagem Bayesiana. Os resultados obtidos no primeiro estágio foram utilizados como informação a priori para o segundo estágio.

Em 1984 o *FDA* autorizou a comercialização de medicamentos genéricos com o objetivo de gerar competitividade na fabricação de medicamentos e, conseqüentemente, a redução de preço dos mesmos. Em 1986, o *FDA* ouviu diversos profissionais das áreas envolvidas neste processo para avaliar os procedimentos que a própria agência havia colocado em vigor em 1977, sendo as ideias apresentadas e discutidas em um encontro sobre bioequivalência no ano de 1988 (BARRET et al., 2000).

Em 1992 o *FDA* editou um guia para regular os estudos de bioequivalência nos Estados Unidos da América. Neste documento, a decisão de bioequivalência foi baseada no método de intervalos de confiança proposto por WESTLAKE (1972) construído para os parâmetros farmacocinéticos concentração máxima (C_{max}) e área sob a curva de concentração plasmática do início da administração do fármaco até um tempo previamente determinado (ASC_{0-t}). A Figura 1 traz uma representação da curva tempo/concentração com estes parâmetros farmacocinéticos para uma situação qualquer.

Figura 1: Representação gráfica de alguns parâmetros farmacocinéticos.

Segundo WESTLAKE (1972), para cada parâmetro farmacocinético constrói-se um intervalo com 90% de confiança para a diferença de médias (ou razão) segundo as formulações em estudo e decide-se por bioequivalência se este intervalo de confiança estiver totalmente contido em um intervalo de $\pm 0,20$ da média da formulação referência (ou entre 0,80 a 1,20 para a razão de médias). Este procedimento é chamado de bioequivalência média. Neste mesmo ano, o HPB (*Health Protection Branch*) do Canadá também publicou sua regulamentação para estudos de bioequivalência (BARRET et al., 2000).

ANDERSON e HAUCK (1990) propuseram um método alternativo ao de bioequivalência média, chamado pelos autores de bioequivalência individual e, nesta mesma direção, HAUCK e ANDERSON (1992) publicaram um outro trabalho em que concluíram que o uso deste procedimento tornaria a decisão de bioequivalência mais segura. Isto se deve ao fato deste tipo de estudo necessitar de um planejamento mais complexo, com mais períodos e que considera fontes de variação adicionais ao modelo de bioequivalência média. A avaliação de bioequivalência individual está relacionada à comutabilidade (*switchability*) das formulações (GHOSH; NTZOUFRAS, 2005), ou seja, a troca de um medicamento referência por uma cópia genérica deste para um mesmo indivíduo cuja concentração dos ingredientes ativos da formulação referência seja considerada estável, eficaz e segura para este indivíduo (CHOW; LIU, 2009). Para que este tipo de avaliação de bioequivalência seja alcançado, no planejamento cruzado do estudo, um mesmo indivíduo deve receber ao menos duas vezes cada formulação, enquanto no procedimento de bioequivalência média basta que o indivíduo receba uma vez cada formulação em estudo.

Um caso particular da avaliação de bioequivalência individual é a avaliação de bioequivalência populacional que está associada à capacidade de prescrição (*prescribability*) das formulações (GHOSH; NTZOUFRAS, 2005). De acordo com CHOW e LIU (2009) a capacidade de prescrição refere-se a escolha do médico entre um medicamento referência e alguma cópia genérica deste medicamento na prescrição de um medicamento apropriado para seus novos pacientes. Tanto o procedimento de avaliação de bioequivalência individual quanto o populacional foram considerados pelo FDA em uma publicação de 1999 (FDA, 1999) sugerindo o uso de técnicas de reamostragem *Bootstrap* (EFRON, 1979) para a estimação e construção de intervalos de confiança para os parâmetros associados às decisões de bioequivalência individual e populacional.

Baseados nesta publicação do FDA de 1999, SHAO et al. (2000) publicaram um artigo em que questionaram o uso da técnica de reamostragem para a construção dos intervalos de confiança sob o ponto de vista de bioequivalência individual, alegando que em alguns casos o poder do teste é pequeno. Estes autores propuseram o uso de estimadores de máxima verossimilhança. GHOSH e NTZOUFRAS (2005) introduziram

o uso da inferência Bayesiana no contexto dos modelos de bioequivalência individual e populacional.

O FDA (2001) publicou a última versão de seu guia para estudos de bioequivalência vigente até hoje, complementando o mesmo em 2003 (FDA, 2003).

Avançando nas pesquisas relacionadas aos métodos estatísticos para a decisão de bioequivalência média, HSU et al. (1995) publicaram um trabalho em que consideraram métodos multivariados para a análise dos dados, partindo do princípio que um mesmo indivíduo fornece um conjunto de parâmetros farmacocinéticos, e estes poderiam estar correlacionados. Neste sentido, os autores propuseram um modelo baseado na distribuição normal bivariada considerando como componentes do vetor de respostas os parâmetros farmacocinéticos ASC_{0-t} e C_{max} .

No Brasil, a regulamentação dos medicamentos genéricos ocorreu com a publicação da Lei nº 9787 de 1999, que dispõe sobre a vigilância sanitária e estabelece o medicamento genérico, dispõe sobre a utilização de nomes genéricos em produtos farmacêuticos e administra temas relacionados a essa área. Antes desta lei já havia a comercialização dos medicamentos similares no Brasil que, embora apresentassem o mesmo princípio ativo do medicamento de referência, não requeriam testes de bioequivalência, o que não garante evidências de intercambialidade com os medicamentos de referência (ANVISA, 2002b).

Ainda no Brasil, o medicamento genérico começou a fazer parte da política nacional de medicamentos, sendo direito do cidadão receber a prescrição medicamentosa com o nome genérico do medicamento receitado (SALAZAR et al., 2003). Alguns autores publicaram trabalhos sobre o impacto social dos medicamentos genéricos (DIAS; ROMANO-LIEBER, 2006; QUENTAL et al., 2008) bem como sobre o impacto econômico destes no país (VIEIRA; ZUCCHI, 2006; NISHIJIMA, 2003).

A regulamentação dos estudos de bioequivalência no Brasil ficou sob responsabilidade da Agência Nacional de Vigilância Sanitária (ANVISA) que aproveitou o momento para conscientizar a população sobre o uso responsável de medicamentos (ANVISA, 2002b), visto que a introdução dos medicamentos genéricos no mercado nacional tornou-os mais acessíveis à população devido ao menor preço praticado.

Após iniciar as inspeções nos centros que realizavam estudos de bioequivalência, os profissionais da ANVISA perceberam a necessidade de estabelecer alguns padrões para a realização dos estudos de bioequivalência no Brasil. Em 2002 foi publicado o Manual de Boas Práticas em Biodisponibilidade e Bioequivalência (ANVISA, 2002a).

Com os protocolos definidos em muitos países do mundo a partir de suas agências reguladoras, mas com algumas incertezas dos pesquisadores sobre este processo dinâmico, as contribuições metodológicas para os estudos de bioequivalência continuaram

buscando disponibilizar técnicas para as mais diversas situações encontradas. Assim, uma série de artigos com o uso de inferência Bayesiana foram publicados: GHOSH e KHATTREE (2003), GHOSH e ROSNER (2007), GHOSH e GÖNEN (2008), SOUZA et al. (2009). Neste sentido, recentemente o *FDA* publicou um guia para o uso da inferência Bayesiana em ensaios clínicos (FDA, 2010) e BERRY et al. (2011) um livro também explorando este tema.

Além dos guias e artigos considerados como referência em bioequivalência, destaca-se um livro publicado por CHOW e LIU (2009) que teve sua primeira edição publicada em 1992. Outra referência importante, que teve sua primeira edição publicada recentemente, é o livro de HAUSCHKE et al. (2007).

Em resumo, para avaliar a intercambialidade de duas ou mais formulações é recomendado o uso de ensaios de bioequivalência que tem como objetivo apresentar evidências de similaridade na biodisponibilidade destes (HYSLOP et al., 2000). De acordo com o guia publicado pelo FDA (2003) para provas de bioequivalência, biodisponibilidade é definida como a velocidade e a extensão de absorção de um princípio ativo em uma forma de dosagem, a partir de sua curva tempo/concentração na circulação sistêmica ou sua excreção na urina. Um medicamento genérico pode ser comercializado quando apresenta-se bioequivalente ao medicamento de referência.

Como forma de quantificar e, conseqüentemente, comparar a biodisponibilidade de duas ou mais formulações em um estudo de bioequivalência, as mesmas são administradas aos indivíduos, geralmente sob um planejamento cruzado, e a concentração do fármaco absorvida por cada um é medida em vários instantes de tempo pré-determinados após esta administração. A partir da curva tempo/concentração fornecida para cada indivíduo e para cada formulação recebida, a quantificação deste processo é realizada a partir de alguns parâmetros farmacocinéticos.

Os parâmetros farmacocinéticos usualmente considerados em um teste de bioequivalência são: concentração máxima (C_{max}) que indica a maior concentração sanguínea da formulação obtida após administração e área sob a curva de concentração plasmática do início da administração do fármaco até um tempo previamente determinado (ASC_{0-t}). De acordo com a especificidade de alguma formulação, ainda podem ser relevantes no estudo o tempo em que ocorre a concentração máxima (T_{max}) e área sob a curva de concentração plasmática do início da administração do fármaco extrapolada para o infinito ($ASC_{0-\infty}$). O método numérico utilizado para obtenção da ASC_{0-t} é o método dos trapézios e a extrapolação utilizada para $ASC_{0-\infty}$ é baseada na inclinação de uma reta de regressão linear para as últimas concentrações tomadas nos últimos tempos de cada curva de concentração plasmática. Estes dois métodos são apresentados em detalhes por CHOW e LIU (2009).

Para a escolha dos parâmetros farmacocinéticos relevantes em um estudo de bioequivalência leva-se em conta que a concentração máxima está diretamente relacionada com a fração absorvida do fármaco, o tempo em que esta ocorre está associada à velocidade de absorção e a área sob a curva de concentração do fármaco absorvido representa a fração que entra na circulação e é independente da velocidade de absorção.

Embora o mesmo indivíduo apresente mais de um parâmetro farmacocinético, é usual que as análises dos parâmetros farmacocinéticos sejam conduzidas de forma independente, não levando em conta uma possível dependência existente entre estas medidas (BRAMLAGE; GOLDIS, 2008; ISMAIL et al., 2004; KHANDAVE et al., 2010). Para esta situação, modelos multivariados que contemplam a análise simultânea dos parâmetros farmacocinéticos (BERGER; HSU, 1996; SOUZA et al., 2009) são propostos, porém os mesmos estão limitados aos pressupostos de normalidade dos resíduos ou dados.

Neste contexto, esta tese está organizada como segue: na próxima seção será introduzido o modelo padrão para análise de dados de bioequivalência bem como algumas análises de dados de bioequivalência utilizando este modelo que motivam a busca por melhores modelos paramétricos para análise de bioequivalência média. Também serão introduzidos nesta seção os objetivos gerais e específicos desta tese. Na seção 3, toda a metodologia proposta nesta tese será introduzida bem como o estudo de poder empírico dos testes introduzidos sob o ponto de vista univariado. Na seção 4 serão considerados 3 exemplos de estudos de bioequivalência com aplicação dos métodos propostos e um exemplo adicional com a aplicação da metodologia de bioequivalência populacional e individual. Na última seção serão apresentadas as conclusões e perspectivas deste trabalho. No Apêndice A estão disponibilizados os códigos computacionais utilizados para a aplicação da metodologia proposta.

2 MOTIVAÇÃO E OBJETIVOS

Em geral, o procedimento de bioequivalência média (WESTLAKE, 1972) é adotado para a decisão de intercambialidade entre formulações por agências reguladoras como ANVISA e FDA. Este procedimento consiste em construir um intervalo com $(1 - \alpha) \times 100\%$ de confiança para a diferença entre os parâmetros farmacocinéticos segundo duas formulações em estudo em que α é um nível de significância fixado previamente e definido pela agência reguladora de cada país ou região. Daí espera-se que estes intervalos estejam totalmente contidos em limites previamente especificados, denominados limites de bioequivalência (LB), também estabelecidos pelas agências reguladoras. Segundo o FDA (2001), o uso adequado de ferramentas estatísticas voltadas à análise dos ensaios de bioequivalência é essencial para a garantia da segurança e eficácia da formulação.

Um modelo geral para aplicação em estudos cruzados de bioequivalência, em que um mesmo indivíduo recebe ao menos uma dose de cada formulação, apresentando em ANVISA (2002a), é dado por:

$$y_{ijk} = \mu + S_{ik} + P_j + F_{(j,k)} + R_{(j-1,k)} + e_{ijk} \quad (1)$$

em que y_{ijk} denota a resposta de um parâmetro farmacocinético na escala original ou logarítmica para o i -ésimo indivíduo, aleatorizado na k -ésima sequência e recebendo alguma formulação no j -ésimo período. Logo, μ é a média geral do parâmetro farmacocinético em estudo; S_{ik} é o efeito aleatório do indivíduo i ($i = 1, 2, \dots, n_k$) na sequência k ($k = 1, 2, \dots, g$) para g sequências; P_j é o efeito fixo do período j ($j = 1, 2, \dots, p$) para p períodos tal que $\sum_{j=1}^p P_j = 0$; $F_{(j,k)}$ é o efeito fixo da formulação administrada na sequência k e no período j , tal que $\sum_{j=1}^p \sum_{k=1}^g F_{(j,k)} = 0$; $R_{(j-1,k)}$ é o efeito residual fixo de primeira ordem da formulação administrada na sequência k e período $(j - 1)$, tal que $R_{(0,k)} = 0$ e $\sum_{j=1}^p \sum_{k=1}^g R_{(j-1,k)} = 0$; e_{ijk} é o erro aleatório associado ao modelo. Para completar a especificação deste modelo, assume-se que S_{ik} e e_{ijk} são independentes, identicamente e normalmente distribuídos com média zero e variâncias σ_s^2 e σ_e^2 , respectivamente.

Assumindo uma situação usual de um planejamento cruzado 2×2 , considerando $k = 1, 2$ sequências, $j = 1, 2$ períodos e duas formulações (R =Referência e T =Teste) sob o modelo (1), tem-se um esquema da disposição destes dados na Tabela 1.

Utilizando covariáveis, o modelo (1) sob um planejamento cruzado 2×2 torna-se mais fácil de ser implementado na maioria dos softwares estatísticos. Esse modelo é dado por:

$$y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3i} + \omega_i + e_{ij} \quad (2)$$

Tabela 1: Esquema da disposição dos dados para um modelo cruzado 2×2 (2 períodos, 2 formulações e 2 sequências).

Indivíduo*	Período (Formulação)	Sequência	ASC _{0-t} ou C _{max}
i	j	k	y_{ijk}
1	1 (R)	RT	y_{111}
	2 (T)		y_{121}
2	1 (R)		y_{211}
	2 (T)		y_{221}
$\vdots$	$\vdots$		$\vdots$
n_1	1 (R)		$y_{n_1 11}$
	2 (T)		$y_{n_1 21}$
<hr/>			
1	1 (T)	TR	y_{112}
	2 (R)		y_{122}
2	1 (T)		y_{212}
	2 (R)		y_{222}
$\vdots$	$\vdots$		$\vdots$
n_2	1 (T)		$y_{n_2 12}$
	2 (R)		$y_{n_2 22}$

*Os indivíduos da sequência *TR* são diferentes dos indivíduos da sequência *RT*.

para $i = 1, 2, \dots, n$ indivíduos e $j = 1, 2$ formulações. No modelo (2) y_{ij} representa um parâmetro farmacocinético na escala original ou logarítmica dos dados e β_0 é o intercepto deste modelo; $X_{1_{ij}}$ é uma variável independente representando a formulação recebida pelo indivíduo i em que $X_{1_{i1}} = 0$ (formulação *R*) e $X_{1_{i2}} = 1$ (formulação *T*); $X_{2_{ij}}$ é uma variável independente representando o período em que o indivíduo i recebe a formulação j ($X_{2_{ij}} = 0$ para o primeiro período e $X_{2_{ij}} = 1$ para o segundo período); X_{3_i} é uma variável independente representando a sequência em que o indivíduo recebeu as formulações ($X_{3_i} = 0$ para a sequência *RT* e $X_{3_i} = 1$ para a sequência *TR*). Consequentemente, β_1 é o efeito fixo de formulação, β_2 é o efeito fixo de período e β_3 o efeito fixo de sequência; ω_i é o efeito aleatório (uma variável latente) responsável por capturar a possível dependência entre os períodos. Neste caso, assume-se que ω_i tem uma distribuição normal com média 0 e variância σ_ω^2 ; e_{ij} é o erro associado ao modelo que tem como pressuposto ser normal, independente e com variância constante σ_e^2 . Assume-se que e_{ij} e ω_i são independentes.

Quando um planejamento cruzado é utilizado existem duas questões importantes envolvidas: o tempo de descanso dos indivíduos de um período para o outro (*washout*) com o objetivo de evitar o efeito residual (*carryover*) no próximo período. O tempo de descanso geralmente é calculado com base no parâmetro farmacocinético $t_{1/2}$ que estima o tempo em que 50% da concentração do fármaco é eliminada do organismo de um indivíduo. Este tempo de descanso geralmente é sugerido pelas agências regulatórias

de cada país. A não evidência de efeitos de período ($\alpha = 0,05$) e sequência ($\alpha = 0,10$) no modelo (2) são fortes indícios da não existência de efeitos residuais segundo ANVISA (2002a).

Em (2) tem-se um modelo de efeitos mistos (efeitos aleatórios e fixos) discutido em detalhes por MCLEAN et al. (1991). Nos modelos de regressão, em geral, é comum o pesquisador estar interessado apenas em verificar o possível efeito das variáveis independentes na sua resposta, tomando sua decisão a partir de um p-valor. No contexto de bioequivalência as hipóteses triviais $H_0 : \beta_1 = 0$ contra $H_1 : \beta_1 \neq 0$ não são suficientes para evidenciar que um parâmetro farmacocinético tenha o mesmo comportamento para diferentes formulações pois uma amostra pequena ou com alta variabilidade no parâmetro farmacocinético, por exemplo, levaria a não rejeição de H_0 . Foi neste sentido que WESTLAKE (1972) propôs o uso dos intervalos de confiança na decisão de bioequivalência, ou seja, um intervalo de confiança para β_1 deve estar totalmente contido dentro de limites de bioequivalência previamente especificados. Considerando a regra de $\pm 20\%$ da média do medicamento referência adotada pela ANVISA (2002a) na construção dos limites de bioequivalência, tem-se:

$$-0,20\mu_R < \mu_T - \mu_R < 0,20\mu_R \quad (3)$$

em que μ_R denota a média de um parâmetro farmacocinético para a formulação R e μ_T denota a média do mesmo parâmetro farmacocinético para a formulação T . Logo, existem evidências de igualdade entre as formulações para o parâmetro farmacocinético em questão se um intervalo de confiança $(1 - \alpha)$ para $\mu_T - \mu_R$ estiver totalmente incluído nos limites de bioequivalência $(\pm 0,20\mu_R)$.

Por outro lado, como estes intervalos geralmente são assintóticos, ou seja, quanto maior o tamanho da amostra menor será o intervalo de confiança, o cálculo do tamanho amostral também é padronizado pelas agências reguladoras. Além disso, existem as questões éticas que envolvem este tipo de estudo por necessitar da internação dos voluntários para coletas periódicas de amostras para a construção das curvas tempo / concentração de cada indivíduo em cada período, conforme exemplificada na Figura 1.

Ainda no modelo (2), note que se y_{ij} representar os dados na escala original, β_1 será diretamente a diferença de médias entre as duas formulações em estudos porém, se a transformação logarítmica for considerada, β_1 será a diferença de médias na escala logarítmica e os limites de bioequivalência devem ser adequados, ou seja, dividindo-se (3) por μ_R :

$$-0,20 < \frac{\mu_T}{\mu_R} - 1 < 0,20 \quad (4)$$

ou

$$0,80 < \frac{\mu_T}{\mu_R} < 1,20 \quad (5)$$

Aplicando o logaritmo em (5) têm-se os limites de bioequivalência para os dados na escala logarítmica:

$$-0,2231 < \ln(\mu_T) - \ln(\mu_R) < 0,1823 \quad (6)$$

Observe que os limites obtidos em (6) não são simétricos e a decisão do teste de hipótese seria diferente se fosse utilizado $\ln(\mu_T) - \ln(\mu_R)$ ou $\ln(\mu_R) - \ln(\mu_T)$. Para resolver este problema agências reguladoras como *FDA* e *ANVISA* propuseram o uso de intervalos simétricos, ou seja:

$$0,80 < \frac{\mu_T}{\mu_R} < 1,25 \quad (7)$$

que resulta em:

$$-0,2231 < \ln(\mu_T) - \ln(\mu_R) < 0,2231 \quad (8)$$

Assim, a amplitude inicial do intervalo igual a 0,40 é aumentada para 0,45 (12,5% de acréscimo) com o uso da transformação logarítmica para os dados. É importante notar que esta é a única transformação admitida para os estudos de bioequivalência, visto que qualquer transformação que seja realizada nos dados tem impacto direto nos limites de bioequivalência. Este fato dá uma característica muito própria aos modelos propostos para estudos de bioequivalência, diferenciando-os das análises de regressão em que o pesquisador tenta diversas transformações nos dados a fim de alcançar resíduos normais interpretando os resultados somente em termos dos efeitos das variáveis independentes e utilizando apenas os *p*-valores para suas conclusões.

Caso os pressupostos do modelo (2) sejam atendidos sem o uso da transformação logarítmica a mesma seria desnecessária porém, caso isto não ocorra, esta seria uma opção desde que a transformação resolva efetivamente o problema dos pressupostos.

No sentido de explorar o modelo padrão de bioequivalência dado em (2), a partir de alguns conjuntos de dados reais disponíveis na literatura, realizou-se a análise dos parâmetros farmacocinéticos ASC_{0-t} e C_{max} tanto na escala original quanto logarítmica. A partir da análise de resíduos a Tabela 2 e as Figuras 2 e 3 com gráficos quantil-quantil (ver seção 3.4.1) foram construídas.

Na Tabela 2 além das referências e formulações dos conjuntos de dados são apresentadas análises de resíduos baseadas no teste de SHAPIRO e WILK (1965) fixando um nível de 0,05 de significância. Observe nesta tabela que, para as formulações propranolol, ondansetrone e risperidona na escala original dos dados, em que não evidenciou-se normalidade para os resíduos, a transformação logarítmica foi efetiva somente para o parâmetro ASC_{0-t} da formulação propranolol (Figuras 2 (a) e (b)). Para a formulação

Tabela 2: Resumo da análise de resíduos para alguns conjuntos de dados.

Autores	Formulação	Resíduos normais?	
		C_{max}	ASC_{0-t}
ISMAIL et al. (2004)	Propranolol	Sim	Após transformação
KOONO (2005)	Tinidazol	Sim	Sim
ARMANDO (2006)	Ondansetrona	Sim	Não
BRIOSCHI (2006)	Cloridrato de ciclobenzaprina	Sim	Sim
SAMPAIO (2008)		Sim	Sim
SILVA (2009)	Metronidazol	Sim	Sim
BELOTTO (2010)	Risperidona	Não	Não

ondansetrona (parâmetro farmacocinético ASC_{0-t} - Figuras 2 (k) e (l)) e risperidona (parâmetros farmacocinéticos ASC_{0-t} e C_{max} - Figuras 3 (i)-(l)) mesmo usando a transformação logarítmica não observou-se evidência de normalidade para os resíduos.

Mesmo com a tabela 2 construída com base no teste de SHAPIRO e WILK (1965), evidenciando dois conjuntos de dados em que os pressupostos de normalidade não foram atendidos para algum parâmetro farmacocinético, observando as Figuras 2 e 3, o pesquisador poderia não estar confortável na validação dos pressupostos de normalidade para outras situações, o que poderia motivar a busca por um melhor modelo. Neste sentido, dentre alguns modelos alternativos na literatura, GHOSH e ROSNER (2007) propuseram um modelo semiparamétrico com mistura de distribuições normais estendendo a ideia para o caso multivariado (GHOSH; GÖNEN, 2008). O uso destes modelos é bastante complexo por se tratar de um modelo de mistura de distribuições.

Paralelamente ao modelo padrão de bioequivalência, o Manual de Boas Práticas em Bioequivalência/Biodisponibilidade (ANVISA, 2002a) sugere o uso do teste não paramétrico de Wilcoxon-Mann-Whitney (WMW) caso os pressupostos do modelo padrão não sejam atendidos. Estudos apresentados por SCHUIRMANN (1987) mostram que este teste não paramétrico tem poder menor que o teste paramétrico padrão.

A partir destes desafios, o objetivo geral desta tese é introduzir novos modelos baseados em distribuições assimétricas que sejam adequados a um número maior de situações encontradas em análise de dados de bioequivalência. Serão apresentados também os modelos de bioequivalência individual e populacional em uma análise sob o enfoque Bayesiano.

Para atingir o objetivo geral, os objetivos específicos são:

- Introduzir o uso das distribuições gama generalizada estendida e *tilted normal* para a decisão de bioequivalência média sob um modelo cruzado tanto no contexto univariado quanto bivariado. A construção destes modelos bivariados será

Figura 2: Gráficos quantil-quantil para os resíduos do modelo padrão de alguns conjuntos de dados (*i*).

(a) Lorazepam: C_{max}

(b) $\ln(C_{max})$

(c) ASC_{0-t}

(d) $\ln(ASC_{0-t})$

(e) Metronizadol: C_{max}

(f) $\ln(C_{max})$

(g) ASC_{0-t}

(h) $\ln(ASC_{0-t})$

(i) Risperidona: C_{max}

(j) $\ln(C_{max})$

(k) ASC_{0-t}

(l) $\ln(ASC_{0-t})$

Figura 3: Gráficos quantil-quantil para os resíduos do modelo padrão de alguns conjuntos de dados (*ii*).

realizada com o uso de funções cópulas;

- Introduzir o uso da distribuição *skew-normal* e *skew-t* para a decisão de bioequivalência média sob um modelo cruzado tanto no contexto univariado quando no contexto multivariado;
- Apresentar um estudo do poder empírico dos testes introduzidos nesta tese sob o enfoque univariado a fim de comparar o poder destes testes com o poder dos testes do modelo padrão na escala original e logarítmica e o teste não paramétrico de WMW;
- Apresentar algumas aplicações baseadas em dados reais de estudos de bioequivalência utilizando todos os modelos introduzidos nesta tese e comparar a adequabilidade destes modelos aos dados entre os modelos propostos e com relação ao modelo padrão utilizando alguns critérios de seleção de modelos introduzidos na seção 3.4;
- Apresentar os enfoques de bioequivalência individual e populacional bem como uma aplicação utilizando dados reais disponibilizados pelo *FDA* de um estudo cruzado com 4 períodos 2 formulações e 2 sequências sob os enfoques de bioequivalência individual, populacional e média;
- Disponibilizar os códigos computacionais utilizados nesta tese de forma que os mesmos possam ser úteis na aplicação dos modelos propostos em outros conjuntos de dados de bioequivalência ou problemas similares por outros pesquisadores.

3 METODOLOGIA

3.1 Bioequivalência média - Enfoque univariado

3.1.1 Distribuição Gama Generalizada Estendida (GGE)

Como primeira alternativa, considera-se a distribuição assimétrica gama generalizada estendida introduzida por PRENTICE (1974) obtida a partir de uma reparametrização da distribuição gama generalizada introduzida por STACY (1962). Esta distribuição é interessante na construção de um modelo para dados provenientes de um estudo de bioequivalência em que os parâmetros farmacocinéticos $C_{\max}$ e ASC_{0-t} , geralmente em sua escala original, apresentem assimetria.

Assim, seja Y uma variável aleatória assumindo distribuição gama generalizada estendida denotada por $GGE(\mu, \sigma, q)$ com função densidade de probabilidade:

$$f(y; \mu, \sigma, q) = \begin{cases} |q| [\sigma \Gamma(q^{-2})]^{-1} \exp[wq^{-2} - \exp(w)] & \text{para } q \neq 0, \\ \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left[-\frac{1}{2\sigma^2}(y - \mu)^2\right] & \text{para } q = 0 \end{cases} \quad (9)$$

em que $y \geq 0$, $-\infty < \mu, q < \infty$ e $\sigma \geq 0$. $\Gamma(x) = \int_0^\infty \exp(-t) t^{x-1} dt$ é a função gama e $w = (y - \mu) \sigma^{-1} q + \Psi(q^{-2})$ em que $\Psi(q^{-2})$ denota a função digamma de q^{-2} ; μ é o parâmetro de locação; q é o parâmetro de assimetria: se $q < 0$ a densidade tem assimetria positiva e se $q > 0$ assimetria negativa e σ é o parâmetro de escala desta distribuição. A construção da função densidade de probabilidade dada em (9), bem como a obtenção de ω a partir da reparametrização da distribuição gama generalizada introduzida por STACY (1962) é apresentada em detalhes por PRENTICE (1974) e discutida em FAREWELL e PRENTICE (1977) e COX et al. (2007). A esperança desta distribuição ($E(Y) = \int_0^\infty y f(y) dy$) pode ser calculado utilizando algum método numérico de integração.

Na função densidade de probabilidade apresentada em (9), observe que para $q = 0$ tem-se a distribuição normal (μ, σ^2) , que resultaria no modelo padrão introduzido na seção 2.

Considerando que Y_{ij} represente um parâmetro farmacocinético na escala original ($C_{\max}$ ou ASC_{0-t}) obtido de um estudo cruzado 2×2 , assume-se:

$$Y_{ij} \sim GGE(\mu_{ij}, \sigma, q) \quad (10)$$

em que y_{ij} é o valor da variável resposta proveniente do indivíduo i , recebendo a formulação j seguindo uma distribuição gama generalizada estendida com densidade

Figura 4: Gráficos para a densidade da distribuição GGE assumindo diferentes valores para os parâmetros desta distribuição.

(9). Baseado no modelo (2), tem-se:

$$\mu_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3i} + \omega_i \quad (11)$$

em que β_0 é o intercepto deste modelo; β_1 é o efeito fixo de formulação; β_2 é o efeito fixo de período e β_3 é o efeito fixo de sequência; X_{3i} , X_{2ij} e X_{1ij} estão definidos na seção 2; ω_i é um efeito aleatório introduzido para capturar a dependência entre as medidas repetidas. Assume-se que este efeito tenha distribuição normal com média 0 e variância σ_ω^2 .

A função de verossimilhança para $\boldsymbol{\vartheta} = (\beta_l, \sigma, q, \omega_i, \sigma_\omega^2)$, $l = 0, 1, 2, 3$ e $i = 1, 2, \dots, n$ para $q \neq 0$ é dada por

$$f(\mathbf{y} | \boldsymbol{\vartheta}) = \prod_{i=1}^n \prod_{j=1}^2 |q| [\sigma \Gamma(q^{-2})]^{-1} \exp[w_{ij} q^{-2} - \exp(w_{ij})] \quad (12)$$

em que $w_{ij} = (y_{ij} - \mu_{ij}) \sigma^{-1} q + \Psi(q^{-2})$ e μ_{ij} é definido em (11). Para que w_{ij} possa ser calculado, a função $\Psi(q^{-2})$ pode ser aproximada numericamente por $\Psi(k) \approx -\gamma - (\frac{1}{k}) + k \sum_{l=1}^{\infty} [l(k+l)]^{-1}$ em que γ é a constante de Euler ($\gamma \approx 0,5772$) (PRENTICE, 1974).

Sob o enfoque Bayesiano é necessário assumir as distribuições a priori para $\boldsymbol{\vartheta}$. Neste sentido, devido a presença do efeito aleatório no modelo, será utilizada uma análise

Bayesiana hierárquica (GELMAN, 2006). No primeiro estágio, serão assumidas:

$$\begin{aligned}\beta_l &\sim N(a_l; s_l^2) \\ \sigma^{-2} &\sim G(c, d) \\ q &\sim N(0; 1) I(-\infty, 0) \text{ ou } q \sim N(0; 1) I(0, \infty)\end{aligned}\tag{13}$$

para $l = 0, 1, 2, 3$ e $j = 1, 2$; $N(a; s^2)$ denota uma distribuição normal com média a e variância s^2 , $G(c; d)$ denota uma distribuição gama com média c/d e variância c/d^2 , $N(0; 1) I(-\infty, 0)$ denota uma distribuição normal truncada assumindo valores negativos e $N(0; 1) I(0, \infty)$ denota uma distribuição normal truncada assumindo valores positivos. Observe que a escolha desta distribuição a priori deve ser realizada a partir de alguma informação sobre a assimetria dos dados (um histograma, por exemplo).

No segundo estágio, assume-se $\omega_i \sim N(0, \sigma_\omega^2)$, $i = 1, \dots, n$ e $\sigma_\omega^2 \sim IG(u, v)$ em que $IG(u, v)$ denota uma distribuição gama inversa com média $v/(u-1)$ e variância $v^2/(u-1)^2(u-2)$. Assume-se que as distribuições a priori são independentes. Observe aqui a necessidade do uso de uma análise Bayesiana hierárquica (mais de um estágio), pois σ_ω^2 também é um parâmetro do modelo, logo necessita estar condicionado a uma distribuição de probabilidade como os demais parâmetros.

As quantidades $a_1, s_1, \dots, u, v$ são chamadas de hiperparâmetros. Nos hiperparâmetros o pesquisador pode introduzir alguma informação a priori. Caso exista esta informação prévia, a análise é conduzida assumindo distribuições a priori informativas. Caso não exista a necessidade de inclusão de alguma informação prévia, estes hiperparâmetros podem ser escolhidos de forma que a distribuição a priori torne-se aproximadamente não informativa, fazendo com que a estimação de $\boldsymbol{\theta}$ seja dominada pela função de verossimilhança, o que leva a resultados próximos aos obtidos pelos estimadores de máxima verossimilhança. Observe que, em ambos os casos, os hiperparâmetros são conhecidos.

O uso do algoritmo *MCMC* (*Markov chain Monte Carlo*) que vem popularizando e facilitando o uso da inferência Bayesiana nas últimas décadas requer que sejam conhecidas as distribuições condicionais (distribuição a posteriori para cada parâmetro). Logo, deve-se multiplicar a função de verossimilhança (12) com as distribuições a priori dada em (13) e com as distribuições a priori do segundo estágio. O resultado deste produto de distribuições é a distribuição a posteriori conjunta e, a partir desta, obtêm-se as distribuições marginais para cada parâmetro que é a distribuição condicional, ou distribuição a posteriori de cada parâmetro.

Se as distribuições condicionais tiverem função densidade de probabilidade conhecidas, bastaria gerar amostrar e obter as medidas resumo de interesse como média, desvio padrão e intervalos de credibilidade das estimativas. Este procedimento é conhecido

como amostrador de Gibbs (GELFAND; SMITH, 1990). As distribuições condicionais do modelo padrão de bioequivalência introduzido em (1) são apresentados em Souza et al. (2007).

Para modelos mais complexos como o introduzido nesta seção e os demais modelos que serão introduzidos nesta tese, as distribuições condicionais apresentam função densidade de probabilidade desconhecida, o que impede a geração de valores diretamente das distribuições condicionais, necessitando assim que a geração das amostras seja realizada a partir de possíveis núcleos destas distribuições (com função densidade de probabilidade conhecida) e submetidas a critérios de aceitação / rejeição como os propostos no algoritmo Metropolis Hastings (CHIB; GREENBERG, 1995).

Implementar estes amostradores para cada modelo de interesse é uma tarefa complexa e, como uma tendência na Estatística, o desenvolvimento de softwares para facilitar a estimação dos parâmetros tanto do ponto de vista frequentista, por reamostragem (Bootstrap) ou Bayesiano, vem tornando esta tarefa mais simples e possibilitando o desenvolvimento de diversos modelos, permitindo assim que o Estatístico fique menos preso a pressupostos de modelos pré estabelecidos e com liberdade para buscar outros modelos que melhor se ajuste aos seus conjuntos de dados.

O software mais popular em inferência Bayesiana surgiu de um projeto chamado *BUGS* (*Bayesian inference Using Gibbs Sampling*) de 1989 resultando no software livre WinBUGS (versão para o sistema operacional Windows) e OpenBUGS (LUNN et al., 2000) para os demais sistemas operacionais. Outras opções de softwares para o uso da inferência Bayesiana que surgiram ao longo dos anos são algumas bibliotecas do software R (R Core Team, 2014), o programa *JAGS* (*Just another Gibbs sampler*) (PLUMMER, 2003), o procedimento *MCMC* no software SAS (SAS Institute Inc., 2011), entre outros.

Nestes softwares, geralmente o usuário precisa informar a distribuição dos dados, modelo utilizado, distribuições a priori com seus respectivos hiperparâmetros, entender como o software trata a entrada dos dados para fazê-la de forma correta e analisar a convergência das cadeias geradas para que os resultados sejam validados, ou seja, o uso destes softwares não exige o conhecimento prévio de inferência Bayesiana do usuário tal como qualquer outro software desenvolvido para facilitar o uso da Estatística. Para quem não está familiarizado com a inferência Bayesiana é importante que o mesmo venha compreendê-la, visto que estes softwares são do tipo “caixa-preta”, o que pode levar um usuário desatento ao erro.

Sendo assim, a aplicação do modelo proposto nesta seção bem como os demais modelos que serão apresentados terão seus resultados obtidos a partir do software *OpenBUGS* ou *JAGS* e os códigos utilizados serão disponibilizados no Apêndice A.

Assim, para cada parâmetro de interesse é gerada uma cadeia de valores (ou simplesmente amostras geradas) que permitirá o cálculo das estimativas (médias dessas amostras), variabilidade do estimador (desvio-padrão destas amostras) e intervalos de credibilidade (IC_r) ou intervalos *HPD* (*Highest Posterior Density*) para um nível de significância α fixo.

Um IC_r $(1 - \alpha) \times 100\%$ representa um intervalo em que o limite inferior é o percentil $\alpha/2 \times 100$ e o limite superior é o percentil $(1 - \alpha/2) \times 100$ dos valores da amostra gerada. Um intervalo *HPD* é o menor intervalo tal que $(1 - \alpha) \times 100\%$ dos valores da amostra gerada estejam contidos neste intervalo. Se a distribuição a posteriori de um parâmetro for simétrica, o *HPD* e o IC_r serão bem próximos.

Para concluir a introdução do modelo desta seção, observe que os efeitos das covariáveis são avaliados observando os intervalos contruídos para cada parâmetro do modelo de regressão dado em (11). Se o valor 0 estiver contido no intervalo associado à covariável, a mesma não apresenta evidências que seja significativa para a resposta de interesse. Para a decisão de bioequivalência é necessário a construção de um intervalo para a diferença entre os parâmetros farmacocinéticos segundo as formulações. Neste sentido, β_1 do modelo de regressão (11) fornece esta informação diretamente bastando obter um intervalo *HPD* ou IC_r para este parâmetro.

3.1.2 Distribuição *Tilted Normal* (*TN*)

Uma distribuição assimétrica recentemente introduzida na literatura é proveniente da classe de distribuições de Marshall Olkin (MO) dada por MARSHALL e OLKIN (1997), denominada distribuição *Tilted Normal* (*TN*) introduzida por MAITI e DEY (2012). Assim, seja um parâmetro farmacocinético representado por $\mathbf{y}$, a função densidade de probabilidade da distribuição *TN* é dada por:

$$f(\mathbf{y} \mid \mu, \sigma^2, \lambda) = \frac{\lambda \sigma^{-1} \phi\left(\frac{y-\mu}{\sigma}\right)}{\left\{1 - (1 - \lambda) \left[1 - \Phi\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^2} \quad (14)$$

para $-\infty < y, \mu < \infty$ e $\lambda, \sigma^2 > 0$; em que ϕ é a densidade da distribuição normal padrão e Φ é a função de probabilidade acumulada da distribuição normal padrão. A esperança desta distribuição ($E(Y) = \int_0^\infty y f(y) dy$) pode ser calculado utilizando algum método numérico de integração. Na Figura 5 tem-se alguns gráficos da densidade desta distribuição assumindo diferentes valores para o parâmetro λ da distribuição *TN*.

Observe na Figura 5 e equação (14) que para $\lambda = 1$ tem-se uma distribuição normal com média μ e variância σ^2 . Observe também que esta distribuição comporta tanto assimetria positiva ($\lambda > 1$) quanto negativa ($\lambda < 1$). Esta distribuição é denotada por $TN(\mu, \sigma^2, \lambda)$ em que μ é o parâmetro de localização, σ^2 é o parâmetro de escala e λ é o

Figura 5: Gráficos para a densidade da distribuição TN assumindo diferentes valores para os parâmetros desta distribuição.

parâmetro de assimetria.

Considerando $j = 1, 2$ formulações e n indivíduos ($i = 1, \dots, n$) assume-se

$$Y_{ij} \sim TN(\mu_{ij}, \sigma^2, \lambda) \quad (15)$$

em que y_{ij} representa a resposta de um parâmetro farmacocinético referente ao indivíduo i recebendo a formulação j seguindo distribuição TN com densidade (14). Assim,

$$\mu_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3i} + \omega_i \quad (16)$$

em que β_0 é o intercepto deste modelo; β_1 é o efeito fixo de formulação; β_2 é o efeito fixo de período e β_3 é o efeito fixo de sequência; X_{3i} , X_{2ij} e X_{1ij} estão definidos na seção 2; ω_i é um efeito aleatório introduzido para capturar a dependência entre as medidas repetidas. Assume-se que este efeito tenha distribuição normal com média 0 e variância σ_ω^2 .

A função de verossimilhança para $\boldsymbol{\vartheta} = (\lambda, \beta_l, \sigma^2, \omega_i, \sigma_\omega^2)$, para $l = 0, 1, 2, 3$ e $i = 1, \dots, n$ é dada por:

$$p(\mathbf{y} \mid \boldsymbol{\vartheta}) = \prod_{i=1}^n \prod_{j=1}^2 \frac{\lambda \sigma^{-1} \phi\left(\frac{y_{ij} - \mu_{ij}}{\sigma}\right)}{\left\{1 - (1 - \lambda) \left[1 - \Phi\left(\frac{y_{ij} - \mu_{ij}}{\sigma}\right)\right]\right\}^2} \quad (17)$$

em que μ_{ij} é dada por (16).

Para o primeiro estágio da análise Bayesiana, assumindo a função de verossimilhança definida em (17), considera-se as seguintes distribuições a priori:

$$\begin{aligned}\beta_l &\sim N(a_l; s_l^2) \\ \sigma^{-2} &\sim G(c; d) \\ \lambda &\sim G(e; f)\end{aligned}\tag{18}$$

Para o segundo estágio, assume-se $\omega_i \sim N(0, \sigma_\omega^2)$, $i = 1, \dots, n$ e $\sigma_\omega^2 \sim IG(u, v)$. Assume-se que todos os hiperparâmetros sejam conhecidos e as distribuições a priori independentes.

Para a decisão de bioequivalência, basta construir os intervalos *HPD* ou *IC_r* para diferença (ou razão) do parâmetro β_1 do modelo introduzido em (16).

3.1.3 Distribuição *Skew-Normal* (*SN*)

Assumindo a classe de distribuições *Skew-Normal* (*SN*) introduzida por AZZALINI (1985) para um parâmetro farmacocinético representado por $\mathbf{y}$, a função densidade de probabilidade para a distribuição *SN* é dada por:

$$f(\mathbf{y} \mid \mu, \sigma^2, \delta) = \frac{2}{\sqrt{\sigma^2 + \delta^2}} \phi\left(\frac{y - \mu}{\sqrt{\sigma^2 + \delta^2}}\right) \Phi\left(\frac{\delta}{\sigma} \frac{y - \mu}{\sqrt{\sigma^2 + \delta^2}}\right)\tag{19}$$

para $-\infty < y, \mu, \delta < \infty$ e $\sigma^2 > 0$; em que ϕ é a densidade da distribuição normal padrão e Φ é a função distribuição de probabilidade acumulada da distribuição normal padrão. Na Figura 6 tem-se alguns gráficos da densidade da distribuição assumindo diferentes valores de δ para a distribuição *SN*.

Observe na Figura 6 e equação (19) que para $\delta = 0$ tem-se uma distribuição normal com média μ e variância σ^2 . Observe também que esta distribuição comporta tanto assimetria positiva ($\delta > 0$) quanto negativa ($\delta < 0$). Esta distribuição é denotada por *SN* (μ, σ^2, δ) em que μ é o parâmetro de localização, σ^2 é o parâmetro de escala e δ é o parâmetro de assimetria. A média desta distribuição é dada por $E(Y) = \mu + \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \delta$ e a variância é dada por $Var(Y) = \sigma^2 + \left(1 - \frac{2}{\pi}\right) \delta^2$.

Para esta distribuição, existem duas estratégias para a construção da análise Bayesiana:

i) A primeira estratégia, considerando o modelo usual cruzado 2×2 de bioequivalência, em que y_{ij} representa um parâmetro farmacocinético do indivíduo i ($i = 1, \dots, n$) recebendo a formulação j ($j = 1, 2$) assumindo distribuição *SN*, a função de verossimilhança para o vetor $\boldsymbol{\vartheta} = (\delta, \beta_l, \sigma^2, \omega_i, \sigma_\omega^2)$, para $l = 0, 1, 2, 3$ e $i = 1, \dots, n$ é dada

Figura 6: Gráficos para a densidade da distribuição SN assumindo diferentes valores para os parâmetros desta distribuição.

por:

$$p(\mathbf{y} | \boldsymbol{\vartheta}) = \prod_{i=1}^n \prod_{j=1}^2 \frac{2}{\sqrt{\sigma^2 + \delta^2}} \phi\left(\frac{y_{ij} - \mu_{ij}}{\sqrt{\sigma^2 + \delta^2}}\right) \Phi\left(\frac{\delta}{\sigma} \frac{y_{ij} - \mu_{ij}}{\sqrt{\sigma^2 + \delta^2}}\right) \quad (20)$$

em que

$$\mu_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3i} + \omega_i \quad (21)$$

onde β_0 é o intercepto deste modelo; β_1 é o efeito fixo de formulação; β_2 é o efeito fixo de período e β_3 é o efeito fixo de sequência. As covariáveis X_{3i} , X_{2ij} e X_{1ij} estão definidas na seção 2; ω_i é um efeito aleatório introduzido para capturar a dependência entre as medidas repetidas. Assume-se que este efeito tenha distribuição normal com média 0 e variância σ_ω^2 .

Para o primeiro estágio da análise Bayesiana, assumindo a função de verossimilhança definida em (20), considera-se as seguintes distribuições a priori:

$$\begin{aligned} \beta_l &\sim N(a_l; s_l^2) \\ \delta &\sim N(a; s^2) \\ \sigma^{-2} &\sim G(c; d) \end{aligned} \quad (22)$$

para $l = 0, 1, 2, 3$ and $i = 1, 2, \dots, n$.

Para o segundo estágio, assume-se $\omega_i \sim N(0, \sigma_\omega^2)$, $i = 1, \dots, n$ e $\sigma_\omega^2 \sim IG(u, v)$. Assume-se que todos os hiperparâmetros sejam conhecidos e independentes.

ii) Na segunda estratégia, considerando o mesmo planejamento apresentado em

(i), assume-se o modelo regressão introduzido por SAHU et al. (2003):

$$y_{ji} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3i} + \omega_i + \delta z_i + e_i \quad (23)$$

em que β_0 é o intercepto deste modelo; β_1 é o efeito fixo de formulação; β_2 é o efeito fixo de período e β_3 é o efeito fixo de sequência; X_{3i} , X_{2ij} e X_{1ij} estão definidos na seção 2; ω_i é um efeito aleatório introduzido para capturar a dependência entre as medidas repetidas. Assume-se que este efeito tenha distribuição normal com média 0 e variância σ_ω^2 . O parâmetro δ é o parâmetro de assimetria do parâmetro farmacocinético em estudo; z_i é uma variável latente associada a δ e e_i é o erro deste modelo de regressão (ver SAHU et al. (2003)).

Esta estratégia de implementação para a distribuição SN tem um grande atrativo computacional sob o procedimento Bayesiano, pois permite que seja usada como base a distribuição normal, como segue:

$$\mathbf{Y}_i \mid \mathbf{z}_i, \boldsymbol{\beta}, X_i, \sigma^2, \delta \sim N(X_i' \boldsymbol{\beta} + \delta \mathbf{z}_i, \sigma^2) \quad (24)$$

Para uma análise Bayesiana hierárquica assume-se as mesmas distribuições a priori introduzidas em (22) além de $z_i \sim N(0; 1) I(z > 0)$ em que $N(0; 1) I(z > 0)$ denota uma distribuição normal truncada assumindo somente valores positivos (SAHU et al., 2003).

Em ambas as estratégias, a construção do intervalo HPD ou IC_r da diferença entre formulações para o parâmetro farmacocinético pode ser realizada diretamente em β_1 .

3.1.4 Distribuição *Skew-t* (ST)

Uma extensão da distribuição SN apresentada na seção 3.1.3 é a distribuição *Skew-t* (ST) baseada na distribuição t de Student (AZZALINI; CAPITANIO, 2003). Para um parâmetro farmacocinético representado por $\mathbf{y}$, a função densidade de probabilidade para a distribuição ST é dada por:

$$f(\mathbf{y} \mid \mu, \sigma^2, \delta, v) = \frac{2}{\sqrt{\sigma^2 + \delta^2}} \frac{\Gamma\left[\frac{(v+1)}{2}\right]}{\Gamma\left(\frac{v}{2}\right) (v\pi)^{\frac{1}{2}}} \left[1 + \frac{(y - \mu)^2}{v(\sigma^2 + \delta^2)}\right]^{-\frac{v+1}{2}} \times \quad (25)$$

$$T_{v+1} \left\{ \left[\frac{v + q(y)}{v + 1} \right]^{-\frac{1}{2}} \frac{\delta}{\sigma} \frac{(y - \mu)}{\sqrt{\sigma^2 + \delta^2}} \right\}$$

para $-\infty < y, \mu, \delta < \infty$ e $\sigma^2 > 0$; em que $q(y) = \frac{(y - \mu)^2}{\sigma^2 + \delta^2}$ e T_{v+1} é a função de probabilidade acumulada da distribuição t de Student padrão com $v + 1$ graus de

liberdade. Na Figura 7 tem-se alguns gráficos da densidade da distribuição assumindo diferentes valores dos parâmetros para a distribuição ST .

Figura 7: Gráficos para a densidade da distribuição ST assumindo diferentes valores para os parâmetros desta distribuição.

Observe na figura (7) que para valores pequenos de v a distribuição apresenta caudas mais pesadas. A medida que os valores de v aumentam, a distribuição ST converge para a distribuição SN . Esta distribuição é denotada por $ST(\mu, \sigma^2, \delta, v)$ em que μ é o parâmetro de locação, σ^2 é o parâmetro de escala, δ é o parâmetro de assimetria e v é o número de graus de liberdade da distribuição. A média desta distribuição é dada por $E(Y) = \mu + \left(\frac{v}{\pi}\right)^{\frac{1}{2}} \frac{\Gamma\left[\frac{(v-1)}{2}\right]}{\Gamma\left(\frac{v}{2}\right)} \delta$ e a variância é dada por $Var(Y) = (\sigma^2 + \delta^2) \frac{v}{v-2} - \frac{v}{\pi} \left\{ \frac{\Gamma\left[\frac{(v-1)}{2}\right]}{\Gamma\left(\frac{v}{2}\right)} \right\}^2 \delta^2$.

Por esta distribuição ser um pouco mais complexa que a distribuição SN , sugere-se a aplicação da mesma a partir da estratégia similar a apresentada na seção 3.1.3 (ii). Também introduzida por SAHU et al. (2003) tem-se:

$$y_{ji} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3i} + \omega_i + \delta z_i + e_i \quad (26)$$

em que β_0 é o intercepto deste modelo; β_1 é o efeito fixo de formulação; β_2 é o efeito fixo de período e β_3 é o efeito fixo de sequência; X_{3i} , X_{2ij} e X_{1ij} estão definidos na seção 2; ω_i é um efeito aleatório introduzido para capturar a dependência entre as medidas repetidas. Assume-se que este efeito tenha distribuição normal com média 0 e variância σ_ω^2 . O parâmetro δ é o parâmetro de assimetria do parâmetro farmacocinético em estudo, z_i é uma variável latente associada a δ e e_i é o erro deste modelo de regressão (ver SAHU et al. (2003)).

Usando a distribuição normal como base, tem-se:

$$\mathbf{Y}_i \mid \mathbf{z}_i, \boldsymbol{\beta}, X_i, \sigma^2, \delta, \zeta_i \sim N \left(X_i' \boldsymbol{\beta} + \delta \mathbf{z}_i, \frac{\sigma^2}{\zeta_i} \right) \quad (27)$$

em que ζ_i é o peso relativo a cada indivíduo no modelo de regressão ($i = 1, \dots, n$) assumindo uma distribuição gama,

$$\zeta_i \mid \mathbf{z}_i, \boldsymbol{\beta}, X_i, \sigma^2, \delta, v \sim G(v/2, v/2) \quad (28)$$

com média 1 e variância $2/v$. Observe que quando $v \rightarrow \infty$, a variância de ζ_i será muito pequena, levando a pesos próximos de 1, convergindo para uma distribuição normal. Usando as expressões (27) e (28) sob o enfoque Bayesiano, ζ_i pode ser gerado diretamente de uma distribuição gama. Este tipo de implementação específico da distribuição t de Student utilizando a distribuição normal como base é chamado de distribuição t hierárquica e o procedimento a regressão realizado pertence a classe das regressões robustas.

Para uma análise Bayesiana desse modelo, pode-se assumir o mesmo conjunto de distribuições a priori assumidas para a distribuição SN . O parâmetro v apresenta dificuldades para ser estimado, assim sugere-se fixar diferentes valores de v e escolher aquele que apresente melhores critérios de ajuste, como aqueles que minimizem o valor de DIC (ver sessão 3.4), por exemplo.

3.2 Bioequivalência média - Enfoque multivariado

3.2.1 Distribuição GGE sob funções cópulas

Função cópula é associada a uma distribuição de probabilidade multivariada em que a distribuição de probabilidade marginal de cada variável é uniforme. As funções cópulas são utilizadas para descrever a dependência entre variáveis aleatórias (NELSEN, 2006). A função cópula para m variáveis aleatórias é definida por

$$F(y_1, y_2, \dots, y_m) = C[F_1(y_1), F_2(y_2), \dots, F_m(y_m); \theta] \quad (29)$$

em que $F(y_i) = \int_{-\infty}^u f(y_i) dy_i$ denota a função de distribuição acumulada da k -ésima variável aleatória ($k = 1, 2, \dots, m$) e θ é o parâmetro de cópula que quantifica uma possível dependência entre estas variáveis (TRIVEDI; ZIMMER, 2005). Pelo teorema de SKLAR (1973), se $F(y_1, y_2, \dots, y_m)$ é uma função distribuição com distribuições marginais $F_1(y_1), F_2(y_2), \dots, F_m(y_m)$ então existe uma função cópula C para todo $y_1, y_2, \dots, y_m$ definida entre 0 e 1.

A seguir, serão descritas três funções cópulas escolhidas estrategicamente. A função

cópula FGM é usada para modelar correlações fracas tanto positiva quanto negativa; a função cópula Gaussiana é definida para modelar qualquer grau de dependência em ambas as direções e a função cópula de Clayton é usada para modelar correlações positivas com qualquer grau de dependência.

Função cópula Farlie-Gumbel-Morgenstern (FGM): A função cópula FGM foi proposta por MORGENSTERN (1956) e posteriormente estendida por FARLIE (1960) e GUMBEL (1960). Em adição, propriedades desta função cópula foram estudadas por AMBLARD e GIRARD (2002, 2005, 2009). Considerando duas variáveis aleatórias ($C_{\max}$ e ASC_{0-t} , por exemplo), tem-se:

$$C[F_1(y_1), F_2(y_2); \theta] = F_1(y_1) F_2(y_2) \{1 + \theta [1 - F_1(y_1)] [1 - F_1(y_2)]\} \quad (30)$$

em que $-1 \leq \theta \leq 1$; Y_1 e Y_2 são as duas variáveis aleatórias com parâmetro de dependência definido por θ .

A função densidade de probabilidade conjunta para as variáveis aleatórias Y_1 e Y_2 derivada de (30) é dada por

$$f(y_1, y_2; \theta) = f_1(y_1) f_2(y_2) \{1 + \theta [1 - 2F_1(y_1)] [1 - 2F_2(y_2)]\} \quad (31)$$

Se $\theta = 0$ tem-se independência entre as variáveis aleatórias Y_1 e Y_2 . Se $\theta \neq 0$ tem-se algumas relações interessantes para os coeficientes de correlação não paramétrico de Kendall (τ) e Spearman (ρ) dados, respectivamente, por (TRIVEDI; ZIMMER, 2005):

$$\tau(\theta) = \frac{2}{9}\theta \text{ e } \rho(\theta) = \frac{1}{3}\theta \quad (32)$$

Para valores extremos de θ , os valores de correlação máximos absolutos para $\tau(\theta)$ e $\rho(\theta)$ são 0,22 e 0,33 (correlação fraca), respectivamente.

Função cópula Gaussiana: A função cópula Gaussiana (LEE, 1983), como a função cópula FGM, pode ser usada para modelar dependência em ambas as direções (correlações positivas ou negativas). Para duas variáveis aleatórias, a função densidade de probabilidade bivariada é dada por:

$$f(y_1, y_2; \theta) = f_1(y_1) f_2(y_2) \frac{1}{\sqrt{(1 - \theta^2)}} \exp \left\{ \left[-\frac{1}{2(1 - \theta^2)} (u^2 + z^2 - 2\theta uz) \right] + \frac{1}{2} (u^2 + z^2) \right\} \quad (33)$$

em que $u = \Phi^{-1}[F_1(y_1)]$; $z = \Phi^{-1}[F_2(y_2)]$ e $-1 < \theta < 1$.

Da mesma forma que na função FGM, podem ser obtidas as relações do parâmetro de dependência θ com os coeficientes não paramétricos de correlação Kendall (τ) e

Spearman (ρ). Estas relações são dadas por (TRIVEDI; ZIMMER, 2005):

$$\tau(\theta) = \frac{2}{\pi} \arcsin(\theta) \text{ e } \rho(\theta) = \frac{6}{\pi} \arcsin\left(\frac{\theta}{2}\right) \quad (34)$$

Função cópula de Clayton: Diferentemente das funções cópulas FGM e Gaussiana, a função cópula de Clayton (CLAYTON, 1978) é apropriada somente para modelar dependências positivas. Para duas variáveis aleatórias, sua função densidade de probabilidade bivariada é dada por:

$$f(y_1, y_2; \theta) = f_1(y_1) f_2(y_2) (1 + \theta) [F_1(y_1) F_2(y_2)]^{-\theta-1} [F_1(y_1) + F_2(y_2) - 1]^{-2-\frac{1}{\theta}} \quad (35)$$

em que $\theta > 0$. O coeficiente de correlação não paramétrico de Kendall (τ), relacionado com o parâmetro de dependência θ é dado por, (TRIVEDI; ZIMMER, 2005):

$$\tau(\theta) = \frac{\theta}{\theta + 2} \quad (36)$$

A relação com o coeficiente de correlação não paramétrico de Spearman (ρ) não é dada em uma forma explícita para a cópula de Clayton.

Função de verossimilhança e distribuições a priori para a distribuição gama generalizada estendida bivariada sob diferentes funções cópulas: Assumindo agora k parâmetros farmacocinéticos ($k = 1, 2$), j formulações e i indivíduos $i = 1, \dots, n$ a função de verossimilhança para $\boldsymbol{\vartheta} = (\sigma_k, q_k, \beta_{kl}, \omega_{ki}, \tau_1^2, \tau_2^2, \rho_\omega, \theta)$, para $l = 0, 1, 2, 3$, é dada por:

$$p(\mathbf{y} | \boldsymbol{\vartheta}) = \prod_{k=1}^2 \prod_{j=1}^2 \prod_{i=1}^n |q_k| [\sigma_k \Gamma(q_k^{-2})]^{-1} \exp[w_{kji} q_k^{-2} - \exp(w_{kji})] \times C_{12}(\theta) \quad (37)$$

em que

$$\mu_{kji} = \beta_{k0} + \beta_{k1} X_{1ji} + \beta_{k2} X_{2ji} + \beta_{k3} X_{3i} + \omega_{ki} \quad (38)$$

e $C_{12}(\theta)$ é parte da função densidade da cópula assumida. Logo:

- Para a cópula FGM:

$$C_{12}(\theta) = \{1 + \theta [1 - 2F_1(y_1)] [1 - 2F_2(y_2)]\}$$

- Para a cópula Gaussiana:

$$C_{12}(\theta) = \frac{1}{\sqrt{(1 - \theta^2)}} \exp \left\{ \left[-\frac{1}{2(1 - \theta^2)} (u^2 + z^2 - 2\theta uz) \right] + \frac{1}{2} (u^2 + z^2) \right\}$$

- Para a cópula de Clayton:

$$C_{12}(\theta) = (1 + \theta) [F_1(y_1) F_2(y_2)]^{-\theta-1} [F_1(y_1) + F_2(y_2) - 1]^{-2-\frac{1}{\theta}}$$

Para o primeiro estágio da análise Bayesiana, assumindo a função de verossimilhança definida em (37), considera-se as seguintes distribuições a priori:

$$\begin{aligned}\beta_{kl} &\sim N(a_{kl}; s_{kl}^2) \\ \sigma_k^{-2} &\sim G(c_k, d_k) \\ q_k &\sim N(0; 1) I(-\infty, 0) \text{ ou } q_k \sim N(0; 1) I(0, \infty)\end{aligned}\tag{39}$$

para $j = 1, 2$; $k = 1, 2$ e $l = 0, 1, 2, 3$. Também assume-se as distribuições a priori para θ dadas por:

$$\begin{aligned}\theta &\sim U(-1; 1) \text{ para funções cópulas FGM e Gaussiana} \\ \theta &\sim G(g; h) \text{ para a função cópula de Clayton}\end{aligned}\tag{40}$$

em que $U(-1; 1)$ denota uma distribuição uniforme no intervalo $(-1; 1)$.

Para o segundo estágio da análise Bayesiana hierárquica, assume-se que ω_{ki} tenha uma distribuição normal bivariada com vetor de médias $\mathbf{0}$ e matriz de covariâncias $\mathbf{S} = \begin{pmatrix} \tau_1^2 & \rho_\omega \tau_1 \tau_2 \\ \rho_\omega \tau_1 \tau_2 & \tau_2^2 \end{pmatrix}$. Assim, assume-se as seguintes distribuições a priori:

$$\begin{aligned}\tau_1^{-2} &\sim G(t_{11}; t_{12}) \\ \tau_2^{-2} &\sim G(t_{21}; t_{22}) \\ \rho_\omega &\sim U(-1; 1)\end{aligned}\tag{41}$$

Assume-se também que todos os hiperparâmetros sejam conhecidos e que as distribuições a priori sejam independentes.

Para a decisão de bioequivalência no contexto bivariado pode-se construir uma região $(1 - \alpha)\%$ *HPD* para $\beta_1 = (\beta_{11}, \beta_{21})$. Para a construção da região *HPD*, sugere-se a estimação da densidades kernel bivariada (DUONG; HAZELTONA, 2003) (função *kde2d* da biblioteca *MASS* do software R) utilizando as amostras geradas para a diferença entre os parâmetros farmacocinéticos. Para que se possa concluir que as formulações em estudos são bioequivalentes de forma conjunta, espera-se que a região $(1 - \alpha)\%$ *HPD* construída esteja totalmente incluída na área verde da Figura 8.

Figura 8: Região dos limites de bioequivalência para uma decisão simultânea de dois parâmetros farmacocinéticos.

3.2.2 Distribuição TN sob funções cópulas

Para k parâmetros farmacocinéticos ($k = 1, 2$), j formulações e i indivíduos $i = 1, \dots, n$ a função de verossimilhança para $\boldsymbol{\vartheta} = (\lambda_k, \sigma_k, \beta_{kl}, \omega_{ki}, \tau_1^2, \tau_2^2, \rho_\omega, \theta)$, para $l = 0, 1, 2, 3$, é dada por:

$$p(\mathbf{y} | \boldsymbol{\vartheta}) = \prod_{k=1}^2 \prod_{j=1}^2 \prod_{i=1}^n \frac{\lambda_k \sigma_k^{-1} \phi\left(\frac{y_{kji} - \mu_{kji}}{\sigma_k}\right)}{\left\{1 - (1 - \lambda_k) \left[1 - \Phi\left(\frac{y_{kji} - \mu_{kji}}{\sigma_k}\right)\right]\right\}^2} \times C_{12}(\theta) \quad (42)$$

em que

$$\mu_{kji} = \beta_{k0} + \beta_{k1}X_{1ji} + \beta_{k2}X_{2ji} + \beta_{k3}X_{3i} + \omega_{ki} \quad (43)$$

e $C_{12}(\theta)$ é relacionada com a função densidade de alguma das três funções cópulas apresentadas na seção 3.2.1.

Para o primeiro estágio da análise Bayesiana, assumindo a função de verossimilhança definida em (42), considera-se as seguintes distribuições a priori:

$$\begin{aligned} \beta_{kl} &\sim N(a_{kl}; s_{kl}^2) \\ \sigma_k^{-2} &\sim G(c_k; d_k) \\ \lambda_k &\sim G(e_k; f_k) \end{aligned} \quad (44)$$

para $j = 1, 2$; $k = 1, 2$ e $l = 0, 1, 2, 3$. Para θ , assume-se as mesmas distribuições a priori dadas em (40) conforme a função cópula assumida.

Para o segundo estágio da análise Bayesiana hierárquica, assume-se que ω_{ki} tenha

distribuição normal bivariada com vetor de médias $\mathbf{0}$ e matriz de covariâncias $\mathbf{S} = \begin{pmatrix} \tau_1^2 & \rho_\omega \tau_1 \tau_2 \\ \rho_\omega \tau_1 \tau_2 & \tau_2^2 \end{pmatrix}$. Assim, assume-se as seguintes distribuições a priori:

$$\begin{aligned} \tau_1^{-2} &\sim G(t_{11}; t_{12}) \\ \tau_2^{-2} &\sim G(t_{21}; t_{22}) \\ \rho_\omega &\sim U(-1; 1) \end{aligned} \quad (45)$$

Assume-se também que todos os hiperparâmetros sejam conhecidos e que as distribuições a priori sejam independentes.

Para a decisão de bioequivalência neste contexto, também constrói-se uma região $(1 - \alpha)\%$ HPD para $\beta_1 = (\beta_{11}, \beta_{21})$. Para que se possa concluir que as formulações em estudos são bioequivalentes de forma conjunta, espera-se que a região $(1 - \alpha)\%$ HPD construída esteja totalmente incluída na área verde da Figura 8.

3.2.3 Distribuição *SN* multivariada

Para m variáveis aleatórias (parâmetros farmacocinéticos, por exemplo) a distribuição normal multivariada (AZZALINI; VALLE, 1996) é dada por:

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, D) = 2^m |\Sigma + D^2|^{-\frac{1}{2}} \phi_m \left[(\Sigma + D^2)^{-\frac{1}{2}} (\mathbf{y} - \boldsymbol{\mu}) \right] P(\mathbf{V} > 0) \quad (46)$$

em que

$$\mathbf{V} \sim N_m \left[D (\Sigma + D^2)^{-1} (\mathbf{y} - \boldsymbol{\mu}), I - D (\Sigma + D^2)^{-1} D \right]$$

e ϕ_m é a densidade de uma distribuição normal com m dimensões, média 0 e matriz de covariâncias identidade. A distribuição será denotada por $SN(\boldsymbol{\mu}, \Sigma, D)$ em que $\boldsymbol{\mu}$ é um vetor de dimensão m , Σ é uma matriz positiva definida $m \times m$ e D é uma matriz diagonal com elementos $\delta_1, \delta_2, \dots, \delta_k$. A média desta distribuição é dada por $E(\mathbf{Y}) = \boldsymbol{\mu} + \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \boldsymbol{\delta}$ e a matriz de covariâncias $cov(\mathbf{Y}) = \Sigma + \left(1 - \frac{2}{\pi}\right) D^2$.

Considerando ainda m parâmetros farmacocinéticos em um estudo cruzado com $i = 1, \dots, n$ indivíduos, duas formulações e dois períodos, a função de verossimilhança gerada a partir de (46) seria extremamente complexa para a implementação computacional. Neste caso a estratégia introduzida por SAHU et al. (2003) torna-se fundamental para que este modelo seja aplicado. Assim, assume-se o seguinte modelo de regressão:

$$y_{kji} = \beta_{k0} + \beta_{k1} X_{1ji} + \beta_{k2} X_{2ji} + \beta_{k3} X_{3ji} + \omega_{ki} + \delta_k z_{ki} + e_{ki} \quad (47)$$

em que $k = 1, 2, \dots, m$; β_{k0} é o intercepto do modelo para o parâmetro farmacocinético

k ; β_{k1} é o efeito fixo de formulação para o parâmetro farmacocinético k ; β_{k2} é o efeito fixo de período parâmetro farmacocinético k e β_{k3} é o efeito fixo de sequência do parâmetro farmacocinético k ; X_{3i} , X_{2ij} e X_{1ij} estão definidos na seção 2 e ω_{ki} é um efeito aleatório introduzido para capturar a dependência entre as medidas repetidas do parâmetro farmacocinético k . Assume-se que este efeito tenha distribuição normal multivariada com vetor de médias $\mathbf{0}$ e matriz de covariâncias Σ_ω ; $\delta_1, \delta_2, \dots, \delta_m$ são os parâmetros de assimetria para cada parâmetro farmacocinético, z_{ki} é uma variável latente associada a δ_k e e_{ki} é o erro do modelo de regressão (ver SAHU et al. (2003)).

Assumindo a distribuição normal multivariada como base da implementação deste modelo, segue que:

$$\mathbf{Y}_i \mid \mathbf{z}_i, \boldsymbol{\beta}, X_i, \Sigma, D \sim N_m(X_i' \boldsymbol{\beta} + D \mathbf{z}_i, \Sigma) \quad (48)$$

Para o primeiro estágio da análise Bayesiana, assumindo o modelo definido em (48), considera-se as seguintes distribuições a priori:

$$\begin{aligned} \beta_{kl} &\sim N(a_{kl}; s_{kl}^2) \\ \delta_k &\sim N(a_{k\delta}; s_{k\delta}^2) \\ z_{ki} &\sim N(0; 1) I(z > 0) \\ \Sigma^{-1} &\sim W(\mathbf{G}, g) \end{aligned} \quad (49)$$

para $k = 1, 2, \dots, m$ e $l = 0, 1, 2, 3$ em que $N(0; 1) I(z > 0)$ denota uma distribuição normal padrão truncada assumindo valores positivos e $\Sigma^{-1} \sim W(\mathbf{G}; g)$ denota uma distribuição de Wishart com média $g\mathbf{H}$ e variância $g(h_{ij}^2 + h_{ii}h_{jj})$ em que $\mathbf{H} = \mathbf{G}^{-1}$.

Para o segundo estágio da análise Bayesiana hierárquica, assume-se que $\boldsymbol{\omega}_k$ tenha distribuição normal bivariada com vetor de médias $\mathbf{0}$ e matriz de covariâncias Ω em que $\Omega^{-1} \sim W(\mathbf{G}_\omega, g_\omega)$.

Para a decisão de bioequivalência considerando $m = 2$ (apenas dois parâmetros farmacocinéticos) pode-se construir a região HPD para $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{21})$ sugerida na seção 3.2.1 e verificar se a mesma está totalmente contida na área verde da Figura 8. Para $m > 2$, pode-se construir todos os $\binom{m}{2}$ pares e decidir a partir das respectivas regiões HPD dois a dois.

Observe que a proposta da distribuição SN multivariada apresenta mais alternativas às distribuições GGE e TN bivariadas construídas a partir das funções cópulas pois comporta mais de dois parâmetros farmacocinéticos enquanto as distribuições GGE e TN introduzidas nas seções 3.2.1 e 3.2.2, respectivamente, comportam apenas o caso bivariado. A construção de uma distribuição multivariada sob funções cópula torna-se muito complexa a medida em que o número de variáveis aleatórias dependentes vai aumentando (CZADO, 2010).

3.2.4 Distribuição ST multivariada

Ainda mais robusta do que a distribuição SN multivariada é a distribuição ST multivariada (AZZALINI; CAPITANIO, 2003). Para m variáveis aleatórias, a função densidade de probabilidade desta distribuição é dada por

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, D, v) = 2^m t_{m,v}(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma + D^2) P(\mathbf{V} > \mathbf{0}) \quad (50)$$

em que

$$t_{m,v}(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma + D^2) = \frac{\Gamma\left(\frac{v+m}{2}\right)}{\Gamma\left(\frac{v}{2}\right) (v\pi)^{\frac{m}{2}} |\Sigma + D^2|^{\frac{1}{2}}} \left[1 + \frac{1}{v} (\mathbf{y} - \boldsymbol{\mu})^T (\Sigma + D^2)^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right]^{-\frac{(v+m)}{2}}$$

e $\mathbf{V}$ denota uma distribuição t multivariada $t_{m,v+m}$ (ver SAHU et al. (2003)).

Esta distribuição é denotada por $ST(\boldsymbol{\mu}, \Sigma, D, v)$ em que $\boldsymbol{\mu}$ é um vetor com m dimensões, Σ em uma matriz $m \times m$ definida positiva, D é uma matriz diagonal com elementos $\delta_1, \delta_2, \dots, \delta_m$ responsável por capturar a assimetria de cada parâmetro farmacocinético e v denota os graus de liberdade ($v > 2$). A média desta distribuição é dada por $E(\mathbf{Y}) = \boldsymbol{\mu} + (v/\pi)^{\frac{1}{2}} \frac{\Gamma[(v-1)/2]}{\Gamma[v/2]} \boldsymbol{\delta}$ e a covariância é dada por $cov(\mathbf{Y}) = (\Sigma + D^2) \frac{v}{v-2} - \frac{v}{\pi} \left[\frac{\Gamma(\frac{v-1}{2})}{\Gamma(\frac{v}{2})} \right]^2 D^2$.

Considerando m parâmetros farmacocinéticos em um estudo cruzado com $i = 1, \dots, n$ indivíduos, duas formulações e dois períodos e adotando a estratégia introduzida por SAHU et al. (2003) tem-se o seguinte modelo de regressão:

$$y_{kji} = \beta_{k0} + \beta_{k1} X_{1ji} + \beta_{k2} X_{2ji} + \beta_{k3} X_{3ji} + \omega_{ki} + \delta_k z_{ki} + e_{ki} \quad (51)$$

em que $k = 1, 2, \dots, m$; β_{k0} é o intercepto do modelo para o parâmetro farmacocinético k ; β_{k1} é o efeito fixo de formulação para o parâmetro farmacocinético k ; β_{k2} é o efeito fixo de período parâmetro farmacocinético k e β_{k3} é o efeito fixo de sequência do parâmetro farmacocinético k . As covariáveis X_{3i} , X_{2ij} e X_{1ij} estão definidas na seção 2; ω_{ki} é um efeito aleatório introduzido para capturar a dependência entre as medidas repetidas do parâmetro farmacocinético k . Assume-se que este efeito tenha distribuição normal multivariada com vetor de médias $\mathbf{0}$ e matriz de covariâncias Σ_ω . $\delta_1, \delta_2, \dots, \delta_m$ são os parâmetros de assimetria para cada parâmetro farmacocinético, z_{ki} é uma variável latente associada a δ_k e e_{ki} é o erro do modelo de regressão (ver SAHU et al. (2003)).

Assumindo a distribuição normal multivariada como base da implementação deste

modelo, segue que:

$$\mathbf{Y}_i \mid \mathbf{z}_i, \boldsymbol{\beta}, X_i, \Sigma, D, \zeta_i \sim N_m \left(X_i' \boldsymbol{\beta} + D \mathbf{z}_i, \frac{\Sigma}{\zeta_i} \right) \quad (52)$$

em que ζ_i é o peso relativo a cada indivíduo no modelo de regressão ($i = 1, \dots, n$) assumindo uma distribuição gama,

$$\zeta_i \mid \mathbf{z}_i, \boldsymbol{\beta}, X_i, \Sigma, D, v \sim G(v/2, v/2) \quad (53)$$

com média 1 e variância $2/v$.

Pode-se assumir as mesmas distribuições a priori introduzidas na seção 3.2.3 para a distribuição *SN* multivariada. Quanto aos graus de liberdade, o mesmo problema discutido na distribuição *ST* univariada é encontrado aqui. Assim, a mesma estratégia introduzida na seção 3.1.4 (basear-se em algum critério de seleção de modelos) pode ser aplicada para viabilizar a utilização deste modelo. Os procedimentos para a decisão de bioequivalência são os mesmos introduzidos na seção 3.2.3.

3.3 Bioequivalência populacional e individual

Com o objetivo de apresentar tanto a decisão de bioequivalência populacional quanto a de bioequivalência individual nesta seção, será adotado um único modelo cruzado assumindo a não existência de efeitos residuais introduzido por CHINCHILLI e ESINHART (1996):

$$y_{ijkl} = \mu_k + \gamma_{ikl} + \delta_{ijk} + e_{ijkl} \quad (54)$$

em que y_{ijkl} representa o valor observado de um parâmetro farmacocinético na escala logarítmica; $i = 1, \dots, s$ indica a sequência das formulações recebidos, $j = 1, \dots, n_i$ indica o indivíduo na sequência i em que o mesmo foi aleatorizado, k indica o tratamento recebido pelo indivíduo e $j = 1, \dots, p_{ik}$ indica o número da réplica, sendo que p_{ik} representa o número de réplicas para a k -ésima formulação na i -ésima sequência. Se $l = 1$, este modelo reduz-se ao modelo restrito usual para avaliação de bioequivalência média; μ_k é a média populacional para a k -ésima formulação, ω_{ikl} representa o efeito fixo para a réplica l da formulação k na sequência i . Este efeito representa a interação entre os efeitos de período e sequência. Logo, é necessário impor que $\sum_{i=1}^s \sum_{l=1}^{p_{ik}} \omega_{ikl} = 0$ (CHINCHILLI; ESINHART, 1996; HYSLOP et al., 2000); δ_{ijk} é o efeito aleatório para o indivíduo j , na sequência i e formulação k ; e_{ijkl} é o erro aleatório para o indivíduo j na sequência i , réplica l e tratamento k .

Um caso particular do modelo apresentado em (54) é o do planejamento cruzado com 4 períodos, 2 formulações sendo que a formulação referência será denotada por R e a formulação em teste por T , 2 sequências e 2 réplicas para cada tratamento.

Um esquema de como os dados são dispostos neste tipo de planejamento é introduzido na Tabela 3. As especificações deste planejamento serão utilizadas para o modelo apresentado em (54) no restante desta seção.

Tabela 3: Esquema da disposição dos dados para um modelo cruzado com 4 períodos, 2 tratamentos, 2 sequências e 2 réplicas para cada tratamento.

Sequência i	Indivíduo* j	Período (Formulação) k	Réplica l	Logarítmo de ASC_{0-t} ou C_{max}
$TRTR$	1	1 (T)	1	y_{11T1}
		2 (R)	1	y_{11R1}
		3 (T)	2	y_{11T2}
		4 (R)	2	y_{11R2}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	n_1	1 (T)	1	y_{1n_1T1}
		2 (R)	1	y_{1n_1R1}
		3 (T)	2	y_{1n_1T2}
		4 (R)	2	y_{1n_1R2}
$RTRT$	1	1 (R)	1	y_{21R1}
		2 (T)	1	y_{21T1}
		3 (R)	2	y_{21R2}
		4 (T)	2	y_{21T2}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	n_2	1 (R)	1	y_{2n_2R1}
		2 (T)	1	y_{2n_2T1}
		3 (R)	2	y_{2n_2R2}
		4 (T)	2	y_{2n_2T2}

*Os indivíduos da sequência $TRTR$ são diferentes dos indivíduos da sequência $RTRT$.

Para que a variância entre os indivíduos seja capturada pelo modelo segundo as formulações recebidas, define-se:

$$\boldsymbol{\delta}_{ij} = (\delta_{ijR}, \delta_{ijT})^T \sim N_2(\mathbf{0}, \boldsymbol{\Omega}) \text{ em que } \boldsymbol{\Omega} = \begin{bmatrix} \sigma_{ER}^2 & \rho\sigma_{ER}\sigma_{ET} \\ & \sigma_{ET}^2 \end{bmatrix} \quad (55)$$

em que σ_{ER}^2 e σ_{ET}^2 representam a variância entre os indivíduos para a formulação referência e teste, respectivamente, e ρ captura a correlação entre as respostas do mesmo indivíduo correspondente às duas formulações.

Finalmente, $e_{ijkl} \sim N(0, \sigma_{D_k}^2)$, assumidos independentes em que $\sigma_{D_k}^2$ é a variância dos indivíduos segundo a formulação k ($k = R, T$). Além disso, δ_{ijk} e e_{ijkl} são assumidos independentes.

Tomando como referência os procedimentos de bioequivalência populacional e individual adotados pelo FDA (1999), a capacidade de prescrição (*prescribability*) de uma

formulação pode ser medida por:

$$\frac{E(y_T - y_R)^2 - E(y_R - y'_R)^2}{\max\left[\frac{1}{2}E(y_R - y'_R)^2, \sigma_{T_0}^2\right]} \quad (56)$$

em que y_R e y'_R são as respostas de dois indivíduos selecionadas aleatoriamente que recebem a formulação referência e y_T um outro indivíduo selecionado independentemente que recebe a formulação em teste. Sob o modelo (54), o parâmetro de interesse para a decisão de bioequivalência populacional é dado por:

$$\Theta_{BP} = \frac{(\mu_T - \mu_R)^2 + \sigma_{FT}^2 - \sigma_{FR}^2}{\max[\sigma_{FR}^2, \sigma_{F_0}^2]} \quad (57)$$

em que $\sigma_{FT}^2 = \sigma_{DT}^2 + \sigma_{ET}^2$ é a variância total da formulação em teste, $\sigma_{FR}^2 = \sigma_{DR}^2 + \sigma_{ER}^2$ é variância total da formulação referência e $\sigma_{F_0}^2$ é uma constante fixada no valor 0,04 pelo FDA (1999). Assim, decide-se por bioequivalência populacional se $\Theta_{BP} < 1,7448$ sendo que este valor também é fixado pelo FDA.

Sob (56), do ponto de vista da decisão de bioequivalência individual (comutabilidade ou *switchability*), tem-se:

$$\Theta_{BI} = \frac{(\mu_T - \mu_R)^2 + \sigma_I^2 + \sigma_{DT}^2 - \sigma_{DR}^2}{\max[\sigma_{DR}^2, \sigma_{D_0}^2]} \quad (58)$$

em que $\sigma_I^2 = (\sigma_{ET} - \sigma_{ER})^2 + 2(1 - \rho)\sigma_{ET}\sigma_{ER}$ é a variância da interação entre sujeito e formulação e $\sigma_{D_0}^2 = 0,04$ fixado pelo FDA (1999). Logo, decide-se por bioequivalência individual se $\Theta_{BI} < 2,4948$ em que este valor também é fixado pelo FDA.

Para a construção de uma análise Bayesiana hierárquica (GELMAN, 2006), assume-se que $y_{ijkl} \sim N(m_{ijkl}, \sigma_{D_k}^2)$ em que

$$m_{ijkl} = \mu_k + \gamma_{ikl} + \delta_{ijk} \quad (59)$$

e $f(\delta_{ij} | \Omega) \sim N_2(\mathbf{0}, \Omega)$; δ_{ij} e Ω foram definidos em (55). Utilizando softwares como JAGS (PLUMMER, 2003) ou BUGS (LUNN et al., 2000) tem-se grande simplificação na obtenção dos estimadores para o modelo proposto. Visto que a distribuição dos dados já foi definida, é necessário definir as distribuições a priori para o modelo. Neste sentido, a escolha será orientada por GHOSH e NTZOUFRAS (2005). Para o primeiro

estágio,

$$\begin{aligned}
\mu_k &\sim N(\mu_{0k}; \sigma_{0k}^2) \\
\sigma_{Dk}^{-2} &\sim G(a_{Dk}, b_{Dk}) \\
\delta_{ij} &\sim N_2(\mathbf{0}; \mathbf{\Omega}) \\
\gamma_{ikl} &\sim N(0, \sigma_k^2) \text{ para } (i, l) \neq 1
\end{aligned} \tag{60}$$

em que $\gamma_{1k1} = -\sum_{l=2}^{p_{1k}} \gamma_{1kl} - \sum_{i=2}^s \sum_{l=2}^{p_{1k}} \gamma_{ikl}$ para $k = R, T$.

Para o segundo estágio, assume-se

$$\begin{aligned}
\mu_{0k} &\sim N(\mu_{00}; \sigma_{00}^2) \\
\sigma_{0k}^{-2} &\sim G(a, b) \\
\sigma_k^{-2} &\sim G(a_k, b_k) \\
\sigma_{ER}^{-2} &\sim G(a_{ER}, b_{ER}) \\
\sigma_{ET}^{-2} &\sim G(a_{ET}, b_{ET}) \\
\rho &= 2U - 1 \text{ com } U \sim \text{Beta}(a_\rho, b_\rho)
\end{aligned} \tag{61}$$

em que $\text{Beta}(a, b)$ denota um distribuição beta com média $\frac{a}{a+b}$ e variância $\frac{ab}{(a+b)^2(a+b+1)}$. Assume-se ainda $\sigma_{00}^{-2} \sim G(a_0, b_0)$, todos os hiperparâmetros conhecidos e independência entre as distribuições a priori.

3.4 Seleção de modelos

3.4.1 Gráfico quantil-quantil e teste de aderência

Um gráfico muito utilizado para verificar o ajuste de uma distribuição a partir dos parâmetros estimados à um conjunto de dados é o gráfico quantil-quantil (WILK; GNANADESIKAN, 1968). A ideia deste método é construir um gráfico de dispersão em que um eixo represente os valores observados e o outro eixo representa o quantil calculado a partir da distribuição teórica proposta correspondente a probabilidade acumulada empírica para cada observação. Espera-se que estes valores acompanhem uma reta que passa pela origem e tem inclinação 1. Para construir um gráfico quantil-quantil pode-se seguir os passos abaixo:

1.) Ordenar o vetor de observações e calcular a probabilidade acumulada empírica para cada observação, ou seja, $P_i = \frac{k_i - 0,5}{n}$ em que $k_1 = 1$ para o menor valor, $k_2 = 2$ para o segundo menor valor e assim sucessivamente; $i = 1, 2, \dots, n$ em que n é o comprimento do vetor;

2.) Para cada valor de P_i calcular o quantil correspondente, q_i da distribuição

teórica em estudo, ou seja, $q_i = \int_{-\infty}^{q_i} f(y) dy = P_i$.

Além do gráfico pode-se aproveitar os quantis calculados da distribuição teórica e tomá-los como valores esperados, calculando a estatística do teste chi-quadrado de Pearson para avaliar a aderência dos dados a uma distribuição dado por:

$$Q^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (62)$$

em que n é o número de observações; O_i o i -ésimo valor observado e E_i o i -ésimo valor esperado; $Q^2 \sim \chi_{n-1}^2$ (Q^2 tem aproximadamente distribuição chi-quadrado com $n - 1$ graus de liberdade). Para um nível de significância α fixo, rejeita-se a hipótese nula (ou seja, existem evidências que a distribuição não adere aos dados), se $Q^2 > \chi_{\alpha, n-1}^2$.

3.4.2 Ajuste da função densidade de probabilidade a um histograma

Outra alternativa gráfica menos criteriosa que o gráfico quantil-quantil (que é calculado ponto a ponto) é obter a curva teórica suavizada da função densidade de probabilidade a partir dos parâmetros estimados e desenhá-la sobre um histograma com eixo das ordenadas representando a densidade das observações. Este critério pode ser útil para comparar em um mesmo gráfico mais de uma distribuição proposta.

3.4.3 Deviance Information Criterion (DIC)

O *Deviance Information Criterion* (*DIC*) (SPIEGELHALTER et al., 2002) é um importante critério Bayesiano usado para a seleção de modelos. *DIC* é similar ao *Akaike information criterion* (AKAIKE, 1974). O *DIC* é dado por:

$$DIC = \hat{D} + 2p_D \quad (63)$$

em que $\hat{D}$ é a *deviance* (desvio) estimado a priori para os parâmetros de interesse e p_D é o número efetivo de parâmetros do modelo, dado por $p_D = \bar{D} - \hat{D}$, em que $\bar{D}$ é a média a posteriori da *deviance*. Quando dois valores de *DIC* são comparados, o menor valor de *DIC* implica no melhor modelo.

3.4.4 Conditional Predictive Ordinate (CPO)

Para a discriminação do melhor modelo, também pode-se considerar a *Conditional Predictive Ordinate* (*CPO*) (GELFAND, 1996) para cada indivíduo. Com $\mathbf{y}_{(i)} =$

$(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n)$ (n indivíduos), a CPO é dada por:

$$CPO_i = p(y_i | \mathbf{y}_{(i)}) = \int p(y_i | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{y}_{(i)}) d\boldsymbol{\theta} \quad (64)$$

em que $p(\boldsymbol{\theta} | \mathbf{y}_{(i)})$ é a densidade a posteriori para o vetor de parâmetros $\boldsymbol{\theta}$ dado $\mathbf{y}_{(i)}$. Também pode-se obter as estimativas de Monte Carlo para CPO_i baseado nas amostras $MCMC$ geradas. Quando dois valores de CPO_i (para cada sujeito) são comparados, maiores valores de CPO_i (em média) indicam um melhor modelo.

3.4.5 *Logarithm of the Pseudo Marginal Likelihood (LPML)*

Como os valores CPO são calculados para cada indivíduo, estes poderiam ser resumidos em um único valor por modelo (como o DIC , por exemplo). Assim o logaritmo da verossimilhança pseudo marginal (*Logarithm of the Pseudo Marginal Likelihood - LPML*) também pode ser calculado. Define-se c_i para a estimativa de Monte Carlo de (64); então, $LPML = \sum \log(c_i)$. Quando dois valores de $LPML$ são comparados, o modelo com maior valor de $LPML$ indica um melhor ajuste.

3.5 Poder do teste

Sejam as seguintes hipóteses para um teste de bioequivalência:

$$\begin{aligned} H_0 &: \mu_R - \mu_T \leq LB_I \text{ ou } \mu_R - \mu_T \geq LB_S \\ H_1 &: LB_I < \mu_R - \mu_T < LB_S \end{aligned} \quad (65)$$

em que LB_I e LB_S denotam os limites inferiores e superiores de bioequivalência, H_0 é a hipótese nula (as formulações não são bioequivalentes) e H_1 a hipótese alternativa (as formulações são bioequivalentes). Em um modelo estatístico para o uso em bioequivalência, o poder de um teste é definido como a probabilidade deste teste não rejeitar a bioequivalência entre duas formulações dado $\theta = \mu_R - \mu_T$ (SCHUIRMANN, 1987; PHILLIPS, 1990).

Para estudar o poder dos testes de forma empírica referente aos modelos univariados introduzidos na seção 3.1 realizou-se o seguinte procedimento:

1.) Foram geradas 100 amostras de tamanho $n = 24$ que representassem dados provenientes de estudos de bioequivalência para 2 diferentes valores do coeficiente de variação total: 15% e 25%; 3 valores diferentes de θ : 0, 5% e 10% em que 0 significa que as médias μ_R e μ_T são iguais, 5% que $\mu_T = 1,05\mu_R$ e para 10% tem-se $\mu_T = 1,10\mu_R$; 3 situações de assimetria: positiva, negativa e nula (simétrica). Logo, foram geradas amostras para cada uma das 18 combinações possíveis.

2.) A distribuição base para a geração destas amostras foi a distribuição gama generalizada estendida disponível no software R (função *rgengamma*) visto que a geração de dados a partir desta distribuição permite a obtenção de coeficientes de assimetria mais acentuados quando comparados a distribuição *skew-normal*, por exemplo.

3.) A escolha de $n = 24$ deve-se fato da Resolução 478 (ANVISA, 2002c) sugerir um número mínimo de 24 voluntários para estudos cruzados 2×2 de bioequivalência média.

Para os modelos introduzidos na seção 3.1 foram assumidos os seguintes hiperparâmetros para as distribuições a priori:

(a) Distribuição GGE e TN : Para as distribuições a priori em (13) e (18) foram escolhidos valores para os hiperparâmetros que levam à distribuições aproximadamente não informativas. Desta forma, as estimativas obtidas são dominadas apenas pela informação dos dados (como na inferência clássica) sem influência das distribuições a priori. Logo, assume-se: $a_l = 0$; $s_l^2 = 1000^2$; $c = d = u = v = 0,0001$ e $e = f = 1$ para $l = 0, 1$, assumindo modelos restritos que consideram somente o efeito de formulação. Para o parâmetro $\mathbf{q}$ da distribuição GGE assume-se $q \sim N(0; 1)I(-\infty, 0)$ ou $q \sim N(0; 1)I(0, \infty)$ de acordo com a direção da assimetria do vetor de dados gerados. Para uma convergência mais rápida das cadeias foi escolhido como chute inicial para β_0 a média do vetor de dados.

(b) Distribuições SN e ST : Para as distribuições a priori em (22), assume-se: $a_l = 0$; $s_l^2 = 1000^2$; $a = 0$ e $s^2 = 10^2$ para $l = 0, 1$; $c = d = u = v = 0,0001$. Para uma convergência mais rápida das cadeias foi escolhido como chute inicial para β_0 a média do vetor de dados.

A escolha dos graus de liberdade para a distribuição de probabilidade ST foi baseada em LANGE et al. (1989) e que sugerem o uso de $v = 4$ graus de liberdade como uma boa escolha em muitas aplicações. GHOSH e NTZOUFRAS (2005) também chegaram as mesmas conclusões para um exemplo que usam a distribuição t de *Student* para um estudo de bioequivalência.

Para todos os modelos foram simuladas 100000 amostras de Gibbs para cada parâmetro usando o software JAGS a partir do software R (biblioteca R2jags). Do total de amostras geradas foram descartadas 50000 amostras ("*burn-in-sample*") para evitar o possível efeito dos valores iniciais até as cadeias tornarem-se estáveis. Das 50000 amostras restantes para cada parâmetro foram tomados valores de 50 em 50 a fim de se obter amostras não correlacionadas. Por necessitar de muito tempo de máquina, estes resultados foram obtidos a partir do LCCA-Laboratório de Computação Científica Avançada da Universidade de São Paulo usando um computador de grande desempenho.

Na Tabela 4 são apresentados os resultados do estudo do poder empírico descrito acima para o modelo padrão na escala original e logarítmica, o teste não paramétrico de Wilcoxon-Mann-Whitney (WMW) sugerido pela ANVISA (2002a) como alternativa ao modelo padrão, bem como os 4 modelos propostos na seção 3.1. Também nesta tabela são apresentadas taxas de aderência dos modelos aos dados: para o modelo padrão foi considerado o teste de SHAPIRO e WILK (1965) enquanto para os modelos propostos foi utilizado o teste chi-quadrado de aderência apresentado na seção 3.4.1.

Observe na Tabela 4 que para um coeficiente de variação total de aproximadamente 15%, quando não existe assimetria, o modelo GGE apresenta poder muito parecido com o modelo padrão na escala original. O modelo padrão na escala logarítmica apresenta maior poder pois tem o intervalo para a decisão de bioequivalência inflacionado conforme apresentado na seção 2. Os outros 3 modelos propostos tem poder menor. Quando são considerados dados com assimetria positiva forte, observa-se que o poder do modelo padrão diminui um pouco para $\theta = 10\%$ enquanto aumenta na escala logarítmica. Observe que a taxa de ajuste dos resíduos deste modelo na escala original é pequena, aumentando quando considerada a transformação logarítmica. O modelo GGE tem poder parecido com os dados na escala logarítmica e grande capacidade de ajuste. Os demais modelos tem poder ligeiramente menor que o modelo GGE e boa aderência aos dados.

Ainda para o coeficiente de variação de 15%, para assimetria negativa forte, o poder do modelo padrão diminui e, como esperado, a escala logarítmica é ineficaz nestes casos tornando a aderência dos resíduos ao modelo ainda pior. Logo, o modelo padrão (simétrico) não tem capacidade de capturar a assimetria negativa e a escala logarítmica não corrige esta assimetria. Por outro lado, o modelo GGE recupera o poder do teste, modelando esta assimetria e apresentando um poder do teste próximo ao das amostras simétrica bem como alto índice de aderência aos dados. Os demais modelos tem poder do teste menor.

Para um coeficiente de variação total de aproximadamente 25%, para os casos em que não se tem assimetria, os modelos propostos tem poder próximo. Quando considerada uma assimetria moderada positiva, da mesma forma que para o coeficiente de variação de 15%, o modelo GGE tem poder parecido com o modelo padrão na escala logarítmica (transformação esta que tende a corrigir esta assimetria). Os demais modelos propostos tem poder menor. Para assimetria negativa moderada, o modelo GGE recupera parte do poder do teste, sendo o que traz resultados mais próximos a situação em que não foi levada em conta assimetria. Os demais modelos tem poder menor, inclusive o teste não paramétrico.

Tabela 4: Estudo do poder empírico do teste para os modelos propostos.

CV	Assimetria	θ	Modelos - Poder (Taxa de ajuste)						ST
			Padrão	Padrão (log)	WMW	GGE	TN	SN	
15%	Nao	10%	0,80(0,84)	0,90(0,80)	0,72	0,81(0,85)	0,72(1,00)	0,76(1,00)	0,77(1,00)
		5%	0,93(0,84)	1,00(0,80)	0,93	0,95(0,88)	0,91(1,00)	0,91(1,00)	0,92(1,00)
		0%	1,00(0,84)	1,00(0,80)	1,00	1,00(0,94)	1,00(1,00)	1,00(1,00)	1,00(1,00)
	Positiva forte	10%	0,76(0,21)	0,95(0,42)	0,81	0,92(0,99)	0,84(1,00)	0,74(0,98)	0,80(0,97)
		5%	0,94(0,21)	1,00(0,43)	0,94	0,99(1,00)	0,97(0,96)	0,93(0,98)	0,96(0,97)
		0%	1,00(0,21)	1,00(0,47)	0,99	1,00(1,00)	1,00(0,99)	0,99(0,98)	1,00(0,96)
	Negativa forte	10%	0,62(0,21)	0,61(0,07)	0,62	0,80(1,00)	0,60(0,98)	0,58(0,93)	0,64(1,00)
		5%	0,84(0,21)	0,79(0,10)	0,86	0,99(1,00)	0,87(0,97)	0,83(0,93)	0,89(1,00)
		0%	0,96(0,21)	0,78(0,10)	0,96	1,00(0,96)	0,99(0,98)	0,96(0,91)	0,98(0,99)
	Nao	10%	0,43(0,85)	0,51(0,68)	0,41	0,46(0,76)	0,40(0,95)	0,41(1,00)	0,43(0,56)
		5%	0,61(0,90)	0,68(0,58)	0,56	0,61(0,83)	0,58(0,99)	0,57(1,00)	0,61(0,57)
		0%	0,74(0,90)	0,73(0,57)	0,65	0,70(0,82)	0,69(0,99)	0,69(1,00)	0,70(0,66)
25%	Positiva moderada	10%	0,55(0,51)	0,74(0,86)	0,53	0,62(0,99)	0,55(0,87)	0,50(0,91)	0,58(0,35)
		5%	0,69(0,51)	0,85(0,86)	0,73	0,84(0,99)	0,76(0,85)	0,67(0,91)	0,73(0,31)
		0%	0,74(0,51)	0,93(0,86)	0,81	0,89(0,99)	0,82(0,83)	0,71(0,89)	0,81(0,21)
	Negativa moderada	10%	0,24(0,41)	0,13(0,02)	0,26	0,33(0,95)	0,19(0,72)	0,21(0,83)	0,27(0,83)
		5%	0,35(0,43)	0,11(0,01)	0,35	0,48(0,93)	0,33(0,70)	0,28(0,89)	0,32(0,86)
		0%	0,38(0,46)	0,06(0,02)	0,39	0,56(0,92)	0,32(0,71)	0,32(0,82)	0,37(0,87)

sk: coeficiente de assimetria (*skewness*)

4 APLICAÇÕES

Para a aplicação da metodologia proposta foram escolhidos três conjuntos de dados: O conjunto de dados introduzido por ARMANDO (2006) (ondansetrona), que teve o pressuposto de normalidade do modelo padrão rejeitado para ASC_{0-t} e apresenta assimetria positiva; o conjunto de dados introduzido por BELOTTO (2010) (risperidona) que também teve o pressuposto de normalidade rejeitado tanto para ASC_{0-t} e C_{max} e apresenta assimetria negativa; o conjunto de dados introduzido por (KOONO, 2005) (tinidazol) que apresenta o melhor ajuste (Figuras 2 (e) e (g)) sob o modelo padrão.

4.1 Exemplo 1: Ondansetrona

4.1.1 Conjunto de dados

Introduzido por ARMANDO (2006), estes dados representam um estudo de bio-equivalência cruzado 2×2 com 23 indivíduos. O objetivo principal deste estudo é comparar a biodisponibilidade relativa de 8 mg de ondansetrona, sendo uma formulação em comprimidos convencionais e a outra em comprimidos de desintegração oral produzidas por diferentes laboratórios. A formulação referência (R) foi fabricada por Zofran[®] e a formulação teste (T) por Vonau[®] flash.

Um resumo do conjunto de dados segundo os parâmetros farmacocinéticos C_{max} e ASC_{0-t} considerados neste exemplo é apresentado na Tabela 5. Na Figura 9 foram construídos histogramas para os parâmetros farmacocinéticos e observa-se comportamento assimétrico tanto para C_{max} quanto para ASC_{0-t} . Na Figura 10 foram construídos os gráficos de dispersão para os parâmetros farmacocinéticos e calculadas as estimativas dos coeficientes de correlação não-paramétricos de Kendall ($\hat{\tau}$) e Spearman ($\hat{\rho}$). Observa-se um grau de correlação forte entre os dois parâmetros farmacocinéticos.

Tabela 5: Medidas resumos para os dados do exemplo 1.

Parâmetros farmacocinéticos	Formulação	Média	Desvio Padrão	Coeficiente de Variação	Coeficiente de Assimetria
C_{max}	R	31,87	11,50	0,36	0,64
	T	30,42	11,72	0,39	
ASC_{0-t}	R	227,66	102,47	0,45	1,08
	T	223,68	117,49	0,53

Figura 9: Histogramas para os parâmetros farmacocinéticos segundo as formulações - Exemplo 1.

Figura 10: Gráfico de dispersão segundo C_{max} e ASC_{0-t} para o modelo GGE - Exemplo 1.

4.1.2 Elicitação das distribuições a priori

(a) **Distribuição GGE e TN sob funções cópulas:** Para as distribuições a priori em (39), (40), (41), (44) e (45) foram escolhidos valores para os hiperparâmetros que levam à distribuições aproximadamente não informativas. Desta forma, as estimativas obtidas são dominadas apenas pela informação dos dados (como na inferência clássica) sem influência das distribuições a priori. Logo, assume-se: $a_{kl} = 0$; $s_{kl}^2 = 1000^2$; $c_k = d_k = 0,0001$ e $e_k = f_k = 1$ para $j = 1, 2$; $k = 1, 2$ e $l = 0, 1, 2, 3$; $g = h = 0,1$ e $t_{k1} = t_{k2} = u = v = 0,01$ para $k = 1, 2$. Para o parâmetro $\mathbf{q}$ da distribuição GGE assume-se $q_k \sim N(0; 1) I(-\infty, 0)$.

(b) **Distribuições SN e ST multivariadas:** Para as distribuições a priori em (49), assume-se: $a_{kl} = 0$; $s_{kl}^2 = 10000^2$; $a_{k\delta} = 0$ e $s_{k\delta}^2 = 10^2$ para $k = 1, 2$ e $l = 0, 1, 2, 3$; $\mathbf{G} = (100/g) R^{-1}$ em que $g = 3$ e R é uma matriz diagonal em que cada entrada da diagonal é o quadrado da amplitude do vetor correspondente a este componente no conjunto de dados (ver SAHU et al. (2003)). Também assume-se $G_\omega = \begin{pmatrix} 10^{-3} & 0 \\ 0 & 10^{-3} \end{pmatrix}$ e $g_\omega = 2$ (ver GHOSH e GÖNEN (2008)).

Para o caso univariado utilizou-se as mesmas distribuições a priori introduzidas na seção 3.5. Para todos os casos utilizou-se a média do vetor dos dados (C_{max} ou ASC_{0-t}) como chute inicial para o parâmetro β_0 .

Para todos os modelos foram simuladas 150000 amostras de Gibbs para cada parâmetro usando o software JAGS a partir do software R (biblioteca R2jags) para as distribuições GGE e TN sob funções cópulas e o software *OpenBUGS* a partir do software R (biblioteca R2OpenBUGS) para as distribuições SN e ST multivariada. Do total de amostras geradas foram descartadas 50000 amostras (“burn-in-sample”) para evitar o possível efeito dos valores iniciais até as cadeias tornarem-se estáveis. Das 100000 amostras restantes para cada parâmetro foram tomados valores de 50 em 50 a fim de se obter amostras não correlacionadas. Foram consideradas no mínimo 2 cadeias de Markov para cada parâmetro com o objetivo de validar as amostras simuladas.

Optou-se pelo software *OpenBUGS* quando a distribuição assumida para os dados já estava disponível diretamente no software (no caso, distribuição normal univariada e multivariada). Nos casos em quem a distribuição não estava disponível no grupo de distribuições previamente implementadas, necessitando assim introduzir a função de verossimilhança da mesma, optou-se pelo software *JAGS* devido ao melhor desempenho computacional deste em termos de tempo gasto para obtenção dos resultados.

Para as amostras simuladas, a convergência foi verificada utilizando-se gráficos temporais, histogramas e gráficos de autocorrelação. Para os resultados da próxima sessão, observou-se grande indicativo de convergência dos algoritmos utilizados para

todos os parâmetros. Isto é fundamental para a credibilidade dos resultados.

Os códigos utilizados para este e os demais exemplos, necessitando apenas da adequação dos hiperparâmetros conforme o exemplo considerado são apresentados no Apêndice A. Um exemplo de duas cadeias que convergem com sucesso é apresentado na Figura 11. As médias a posteriori e seus respectivos desvios padrão da amostra gerada para cada parâmetro dos exemplos 1, 2 e 3 são apresentados no Apêndice B.

Figura 11: Exemplo de duas cadeias que convergem com sucesso.

4.1.3 Resultados

Enfoque Univariado: Considerando os parâmetros farmacocinéticos C_{max} e ASC_{0-t} na escala original serão apresentadas as análises conduzidas a partir dos modelos introduzidos na seção 3.1 em que cada parâmetro farmacocinético é analisado separadamente.

A escolha da melhor distribuição que ajuste os dados do exemplo e consequentemente a escolha do melhor modelo para os dados pode ser realizada combinando os 4 critérios de seleção propostos na seção 3.4. Neste sentido, a Figura 12 apresenta os gráficos quantil-quantil, a Figura 13 os histogramas com as curvas ajustadas e na Tabela 6 os valores de DIC e $LPML$.

Figura 12: Gráficos quantil-quantil para os parâmetros farmacocinéticos segundo os modelos univariados ajustados introduzidos na seção 3.1 - Exemplo 1.

Na Tabela 6 observa-se que o modelo baseado na distribuição de probabilidade GGE minimiza os valores de DIC e maximiza os valores de $LPML$ para os dois parâmetros farmacocinéticos sendo este um indicador numérico relevante para rejeitar

Figura 13: Histogramas para os parâmetros farmacocinéticos com as curvas teóricas das funções densidade de probabilidade ajustadas - Exemplo 1.

Tabela 6: Valores de DIC e $LPML$ para os modelos propostos na seção 3.1 para cada parâmetro farmacocinético - Exemplo 1.

Modelos	Parâmetros Farmacocinéticos			
	C_{max}		$ASC_{(0-t)}$	
	DIC	$LPML$	DIC	$LPML$
GGE	279, 6	-131, 3	449, 4	-216, 3
TN	283, 4	-133, 4	459, 1	-221, 9
ST ($v = 4$)	300, 9	-131, 5	479, 3	-218, 7
SN	301, 6	-134, 2	484, 3	-223, 4
Padrão	303, 8	-134, 3	484, 4	-223, 7

os demais modelos em favor deste. Na Figura 12 (a), observa-se claramente o melhor ajuste da distribuição de probabilidade GGE aos dados. O melhor ajuste do parâmetro farmacocinético $ASC_{(0-t)}$ pode ser observado na Figura 13 (b), sendo o modelo GGE aquele que melhor acompanha a forma do histograma quando comparado as demais distribuições.

Na Tabela 7 são apresentados os intervalos *HPD* para a diferença de médias dos parâmetros farmacocinéticos segundo cada formulação. Estes intervalos são fundamentais na decisão de bioequivalência média, ou seja, espera-se que os mesmos estejam totalmente contidos nos limites de bioequivalência. Neste sentido, tomando a média da formulação referência apresentada na Tabela 5 para C_{max} , multiplica-se a mesma por $\pm 20\%$, ou seja, $31,87 \times \pm 0,20 = (-6,37; 6,37)$, que são os limites de bioequivalência para este parâmetro farmacocinético. Para ASC_{0-t} , tem-se $227,66 \times \pm 0,20 = (-45,53; 45,53)$ que são os limites de bioequivalência para este parâmetro farmacocinético. Observa-se que os intervalos *HPD* para a melhor distribuição dentre as propostas (GGE) está totalmente contido dentro dos limites de bioequivalência, o que levaria a não rejeição de bioequivalência média entre os fármacos, o que de acordo com a ANVISA ou *FDA* faz dos mesmos intercambiáveis.

Tabela 7: Intervalos *HPD* para a decisão de bioequivalência média - Exemplo 1.

Parâmetro farmacocinético	Modelo	média (d.p.)	Intervalo HPD (90%)	Amplitude
C_{max}	GGE	-1,39 (1,52)	-3,72; 1,29	5,01
	Padrão	-1,31 (1,51)	-3,75; 1,09	4,84
ASC_{0-t}	GGE	-8,82 (9,74)	-23,72; 7,07	30,79
	Padrão	-2,75 (10,63)	-20,13; 13,78	33,91

Ainda na Tabela 7 foram apresentados os intervalos *HPD* considerando o modelo padrão dado em (2) para a decisão de bioequivalência média. Observe que a decisão de bioequivalência média seria a mesma, porém os intervalos apresentam-se um pouco diferentes, o que poderia gerar diferentes decisões em uma situação de proximidade de algum limite do intervalo a algum extremo dos limites de bioequivalência. Logo esta discussão de ajustes a partir de diversos critérios deve ser imprescindível.

Para finalizar o enfoque univariado deste exemplo, na Figura 14 são apresentados gráficos de dispersão para cada parâmetro farmacocinético em que confronta-se os valores observados com as médias a posteriori calculadas no modelo proposto para cada indivíduo. Observa-se um bom ajuste das médias a posteriori com os dados observados.

Enfoque Bivariado: Neste enfoque os parâmetros farmacocinéticos são analisados conjuntamente a partir das distribuições propostas na seção 3.2. Neste sentido, na

Figura 14: Gráficos de dispersão das médias a posteriori segundo cada observação para o modelo GGE - Exemplo 1.

Tabela 8 são apresentados os valores de DIC e $LPML$ bem como os intervalos HPD para os parâmetros que capturam a correlação dos modelos propostos. Como a Figura 10 apresenta evidências de correlação forte entre os parâmetros farmacocinéticos deste exemplo, a cópula FGM (que captura correlação fraca) foi descartada neste exemplo.

Na Tabela 8, observa-se que todos os modelos capturam a correlação entre os dois vetores de parâmetros farmacocinéticos, evidenciando assim a importância de levar em conta, do ponto de vista estatístico, esta relação. Entre as distribuições, concordando com a escolha feita no caso univariado, a distribuição GGE apresenta os maiores valores de $LPML$. Como estes são bem próximos, opta-se pela escolha do modelo GGE sob a função cópula de Clayton que apresenta menor valor de DIC entre todos os modelos.

Na Figura 15 tem-se a variação dos valores de DIC e $LPML$ conforme os graus de liberdade variam na distribuição de probabilidade ST , o que justifica a escolha de $v = 4$ conforme discutido por LANGE et al. (1989) e GHOSH e NTZOUFRAS (2005).

Para a decisão de bioequivalência simultânea entre os parâmetros farmacocinéticos constrói-se uma região HPD para os mesmos. Neste sentido, espera-se que esta região esteja contida no retângulo construído com os limites de bioequivalência de $\pm 20\%$ conforme calculado no enfoque univariado deste exemplo. Logo, conforme a decisão univariada um a um, a decisão bivariada apresentada na Figura 16 também resulta em bioequivalência média entre as formulações. Também não foram observadas evidências de efeito de período e sequência em nenhum dos dois enfoques desta aplicação.

Para finalizar este exemplo, na Figura 17 são apresentados gráficos de dispersão

Tabela 8: Valores de DIC e $LPML$ e intervalos HPD para os parâmetros de correlação dos modelos propostos na seção 3.2 - Exemplo 1.

Modelo	Função Cópula	Parâmetro de Correlação HPD 90%	DIC	LPML
GGE	Gaussiana	$\rho(\hat{\theta}) = (0, 80; 0, 90)$ $\tau(\hat{\theta}) = (0, 60; 0, 72)$	718, 9	-311, 5
	Clayton	$\tau(\hat{\theta}) = (0, 54; 0, 67)$	653, 6	-313, 0
TN	Gaussiana	$\rho(\hat{\theta}) = (0, 77; 0, 88)$ $\tau(\hat{\theta}) = (0, 58; 0, 71)$	733, 4	-322, 2
	Clayton	$\tau(\hat{\theta}) = (0, 54; 0, 68)$	670, 7	-321, 9
SN Bivariada	—	$\hat{\rho} = (0, 54; 0, 89)$	738, 6	-340, 3
ST ($v = 4$) Bivariada	—	$\hat{\rho} = (0, 36; 0, 87)$	711, 9	-324, 1

Figura 15: Valores de DIC e $LPML$ variando o parâmetro v da distribuição multivariada ST - Exemplo 1.

Figura 16: Região *HPD* para a decisão conjunta de bioequivalência média para C_{max} e ASC_{0-t} do modelo GGE sob a função cópula de Clayton - Exemplo 1.

para cada parâmetro farmacocinético em que confronta-se os valores observados com as médias a posteriori calculadas no modelo proposto para cada indivíduo. Observa-se um bom ajuste das médias a posteriori com os dados observados.

4.2 Exemplo 2: Risperidona

4.2.1 Conjunto de dados

Introduzido por BELOTTO (2010), estes dados representam um estudo de bioequivalência cruzado 2×2 com 23 indivíduos. O objetivo principal deste estudo é avaliar a bioequivalência e intercambialidade entre dois antipsicóticos (um referência e outro equivalente farmacêutico) contendo 2 mg de risperidona.

Um resumo do conjunto de dados segundo os parâmetros farmacocinéticos C_{max} e ASC_{0-t} considerados neste exemplo é apresentado na Tabela 9. Na Figura 18 foram construídos histogramas para os parâmetros farmacocinéticos. Na Figura 18 observa-se comportamento assimétrico tanto para C_{max} quanto para ASC_{0-t} . Na Figura 19 foram construídos os gráficos de dispersão para os parâmetros farmacocinéticos e calculadas as estimativas coeficientes de correlação não paramétrico de Kendall ($\hat{\tau}$) e Spearman ($\hat{\rho}$). Observa-se um grau de correlação forte entre os dois parâmetros farmacocinéticos.

4.2.2 Elicitação das distribuições a priori

(a) Distribuição GGE e *TN* sob funções cópulas: Para as distribuições a priori em (39), (40), (41), (44) e (45) escolheu-se valores para os hiperparâmetros que

Figura 17: Gráficos de dispersão das médias a posteriori segundo cada observação para o modelo GGE sob a função cópula de Clayton - Exemplo 1.

Tabela 9: Medidas resumos para os dados do exemplo 2.

Parâmetros farmacocinéticos	Formulação	Média	Desvio Padrão	Coefficiente de Variação	Coefficiente de Assimetria
C_{max}	R	16,02	2,32	0,14	-0,72
	T	12,65	2,68	0,21	
ASC_{0-t}	R	341,45	47,76	0,14	-0,79
	T	279,63	66,10	0,24

Figura 18: Histogramas para os parâmetros farmacocinéticos segundo as formulações - Exemplo 2.

Figura 19: Gráfico de dispersão segundo C_{max} e ASC_{0-t} para o modelo GGE - Exemplo 2.

levam à distribuições a priori aproximadamente não informativas. Logo, assume-se: $a_{kl} = 0$; $s_{kl}^2 = 1000^2$; $c_k = d_k = 0,0001$ e $e_k = f_k = 1$ para $j = 1, 2$; $k = 1, 2$ e $l = 0, 1, 2, 3$; $g = h = 0,1$ e $t_{k1} = t_{k2} = u = v = 0,01$ para $k = 1, 2$. Para o parâmetro $\mathbf{q}$ da distribuição GGE assume-se $q_k \sim N(0; 1) I(0, \infty)$.

(b) Distribuições SN e ST multivariadas: Para as distribuições a priori em (49), assume-se: $a_{kl} = 0$; $s_{kl}^2 = 10000^2$; $a_{k\delta} = 0$ e $s_{k\delta}^2 = 10^2$ para $k = 1, 2$ e $l = 0, 1, 2, 3$; $\mathbf{G} = (100/g) R^{-1}$ em que $g = 3$ e R é uma matriz diagonal em que cada entrada da diagonal é o quadrado da amplitude do vetor correspondente a este componente no conjunto de dados (ver SAHU et al. (2003)). Também assume-se $G_\omega = \begin{pmatrix} 10^{-3} & 0 \\ 0 & 10^{-3} \end{pmatrix}$ e $g_\omega = 2$ (ver GHOSH e GÖNEN (2008)).

Para o caso univariado utilizou-se as mesmas distribuições a priori introduzidas na seção 3.5. Para todos os casos utilizou-se a média do vetor dos dados (C_{max} ou ASC_{0-t}) como chute inicial para o parâmetro β_0 .

Para todos os modelos foram simuladas 150000 amostras de Gibbs para cada parâmetro usando o software JAGS a partir do software R (biblioteca R2jags) para as distribuições GGE e TN sob funções cópulas e o software BUGS a partir do software R (biblioteca R2OpenBUGS) para as distribuições SN e ST multivariada. Do total de amostras geradas foram descartadas 50000 amostras (*“burn-in-sample”*) para evitar o possível efeito dos valores iniciais até as cadeias tornarem-se estáveis. Das 100000 amostras restantes para cada parâmetro foram tomados valores de 50 em 50 a fim de se obter amostras não correlacionadas. Foram consideradas no mínimo 2 cadeias de Markov para cada parâmetro com o objetivo de validar as amostras simuladas.

Para as amostras simuladas, a convergência foi verificada utilizando-se gráficos temporais, histogramas e gráficos de autocorrelação. Para os resultados da próxima sessão, observou-se grande indicativo de convergência dos algoritmos utilizados para todos os parâmetros.

4.2.3 Resultados

Enfoque Univariado: Considerando os parâmetros farmacocinéticos C_{max} e ASC_{0-t} na escala original serão apresentadas as análises conduzidas a partir dos modelos introduzidos na seção 3.1.

Na Tabela 10 observa-se que o modelo baseado na distribuição de probabilidade GGE minimiza os valores de DIC e maximiza os valores de $LPML$ para os dois parâmetros farmacocinéticos sendo este um indicador numérico relevante para rejeitar os demais modelos em favor deste. Na Figura 20, observa-se o melhor ajuste da distribuição de probabilidade GGE aos dados, principalmente na extremidades, o que pode ser

Figura 20: Gráficos quantil-quantil para os parâmetros farmacocinéticos segundo os modelos univariados ajustados introduzidos na seção 3.1 - Exemplo 2.

Figura 21: Histogramas para os parâmetros farmacocinéticos com as curvas teóricas das funções densidade de probabilidade ajustadas - Exemplo 2.

Tabela 10: Valores de DIC e $LPML$ para os modelos propostos na seção 3.1 para cada parâmetro farmacocinético - Exemplo 2.

Modelos	Parâmetros Farmacocinéticos			
	C_{max}		$ASC_{(0-t)}$	
	DIC	$LPML$	DIC	$LPML$
GGE	110, 2	-41, 9	368, 1	-175, 7
TN	120, 0	-49, 0	390, 7	-187, 9
ST ($v = 4$)	114, 8	-45, 9	409, 8	-184, 1
SN	120, 0	-49, 5	413, 3	-189, 0
Padrão	134, 5	-49, 8	413, 3	-189, 2

visto na Figura 21.

Na Tabela 11 são apresentados os intervalos HPD para a diferença de médias dos parâmetros farmacocinéticos segundo cada formulação. Tomando a média da formulação referência apresentada na tabela 9 para C_{max} , multiplica-se a mesma por $\pm 20\%$, ou seja, $16,02 \times \pm 0,20 = (-3,20; 3,20)$, que são os limites de bioequivalência para este parâmetro farmacocinético. Para ASC_{0-t} , tem-se $341,45 \times \pm 0,20 = (-68,29; 68,29)$ que são os limites de bioequivalência para este parâmetro farmacocinético. Observa-se que o intervalo HPD do parâmetro C_{max} para a distribuição GGE não está totalmente contido dentro dos limites de bioequivalência, o que levaria a rejeição de bioequivalência média entre os fármacos. Observe também que tanto para C_{max} quanto para ASC_{0-t} os efeitos de formulação são rejeitados (o valor 0 não está incluído nos intervalos HPD), porém rejeitar ou não o efeito de formulação não implica em rejeitar ou não bioequivalência entre duas formulações conforme discutido na seção 2.

Tabela 11: Intervalos HPD para a decisão de bioequivalência média - Exemplo 2.

Parâmetro farmacoc.	Modelo	média (d.p.)	Intervalo HPD (90%)	Amplitude
C_{max}	GGE	-3,34 (0,22)	-3,72; -2,98	0,74
	Padrão	-3,37 (0,25)	-3,81; -3,00	0,81
ASC_{0-t}	GGE	-54,68 (4,92)	-63,02; -47,24	15,78
	Padrão	-61,82 (6,08)	-72,30; -52,99	19,31

Para finalizar o enfoque univariado deste exemplo, na Figura 22 são apresentados gráficos de dispersão para cada parâmetro farmacocinético em que confronta-se os valores observados com as médias a posteriori calculadas no modelo proposto para cada indivíduo. Observa-se um bom ajuste das médias a posteriori com os dados observados.

Enfoque Bivariado: Na Tabela 12 são apresentados os valores de DIC e $LPML$ bem como os intervalos HPD para os parâmetros que capturam a correlação dos mo-

Figura 22: Gráficos de dispersão das médias a posteriori segundo cada observação para o modelo GGE - Exemplo 2.

delos propostos na seção 3.2. Como a Figura 19 apresenta evidências de correlação forte entre os parâmetros farmacocinéticos deste exemplo, a cópula FGM (que captura correlação fraca) foi descartada neste exemplo.

Na tabela 12, observe que todos os modelos capturam a correlação entre os dois vetores de parâmetros farmacocinéticos. Entre as distribuições, concordando com a escolha feita no caso univariado, a distribuição GGE sob a função cópula Gaussiana maximiza os valores de $LPML$ e minimiza os valores de DIC .

Para a decisão de bioequivalência simultânea entre os parâmetros farmacocinéticos construiu-se uma região HPD para os mesmos segundo o modelo GGE sob a cópula Gaussiana. Neste sentido, espera-se que esta região esteja contida no retângulo construído com os limites de bioequivalência de $\pm 20\%$ conforme calculado no enfoque univariado deste exemplo. Logo, conforme a decisão univariada um a um, a decisão bivariada apresentada na Figura 23 também leva a rejeição da bioequivalência média entre as formulações pois a região HPD não está totalmente contida na área dos limites de bioequivalência. Não foram observadas evidências de efeito de período e sequência em nenhum dos dois enfoques desta aplicação.

Para finalizar este exemplo, na figura 24 são apresentados gráficos de dispersão para cada parâmetro farmacocinético em que confronta-se os valores observados com as médias a posteriori calculadas no modelo proposto para cada indivíduo. Observa-se um bom ajuste das médias a posteriori com os dados observados.

Tabela 12: Valores de DIC e $LPML$ e intervalos HPD para os parâmetros de correlação dos modelos propostos na seção 3.2 - Exemplo 2.

Modelo	Função Cópula	Parâmetro de Correlação HPD 90%	DIC	LPML
GGE	Gaussiana	$\rho(\hat{\theta}) = (0, 88; 0, 94)$ $\tau(\hat{\theta}) = (0, 69; 0, 79)$	399, 4	-175, 3
	Clayton	$\tau(\hat{\theta}) = (0, 71; 0, 80)$	403, 4	-179, 5
TN	Gaussiana	$\rho(\hat{\theta}) = (0, 89; 0, 94)$ $\tau(\hat{\theta}) = (0, 71; 0, 79)$	486, 3	-188, 5
	Clayton	$\tau(\hat{\theta}) = (0, 70; 0, 80)$	415, 6	-194, 3
SN Bivariada	—	$\hat{\rho} = (0, 45; 0, 85)$	502, 8	-225, 0
ST ($v = 4$) Bivariada	—	$\hat{\rho} = (0, 26; 0, 81)$	478, 9	-202, 9

Figura 23: Região HPD para a decisão conjunta de bioequivalência média para C_{max} e ASC_{0-t} do modelo GGE sob a função cópula Gaussiana - Exemplo 2.

Figura 24: Gráficos de dispersão das médias a posteriori segundo cada observação para o modelo GGE sob a função cópula Gaussiana - Exemplo 2.

4.3 Exemplo 3: Tinidazol

4.3.1 Conjunto de dados

Introduzido por KOONO (2005) estes dados representam um estudo de bioequivalência cruzado 2×2 com 24 indivíduos. O objetivo principal deste estudo é verificar a bioequivalência de duas marcas comerciais de comprimidos (produto teste: Amplium[®] - FARMASA; produto referência: Pletil[®] - Pharmacia do Brasil Ltda) contendo 500 mg de tinidazol em voluntários sadios.

Um resumo do conjunto de dados segundo os parâmetros farmacocinéticos C_{max} e ASC_{0-t} considerados neste exemplo é apresentado na Tabela 13. Na Figura 25 foram construídos histogramas para os parâmetros farmacocinéticos. Na Figura 25 observa-se uma leve assimetria tanto para C_{max} quanto para ASC_{0-t} . Na Figura 26 foram construídos os gráficos de dispersão para os parâmetros farmacocinéticos e calculadas as estimativas coeficientes de correlação não paramétrico de Kendall ($\hat{\tau}$) e Spearman ($\hat{\rho}$). Observa-se um grau de correlação moderada entre os dois parâmetros farmacocinéticos.

4.3.2 Resultados

Enfoque Univariado: Assumindo as mesmas distribuições a priori introduzidas no exemplo 2, considerando os parâmetros farmacocinéticos C_{max} e ASC_{0-t} na escala original serão apresentadas as análises conduzidas a partir dos modelos introduzidos na seção 3.1.

Tabela 13: Medidas resumos para os dados do exemplo 3.

Parâmetros farmacocinéticos	Formulação	Média	Desvio Padrão	Coeficiente de Variação	Coeficiente de Assimetria
C_{max}	R	11,34	1,83	0,16	-0,50
	T	11,11	1,69	0,15	
ASC_{0-t}	R	201,92	27,44	0,14	-0,41
	T	198,15	27,33	0,14

Figura 25: Histogramas para os parâmetros farmacocinéticos segundo as formulações - Exemplo 3.

Figura 26: Gráfico de dispersão segundo C_{max} e ASC_{0-t} - Exemplo 3.

Figura 27: Gráficos quantil-quantil para os parâmetros farmacocinéticos segundo os modelos univariados ajustados introduzidos na seção 3.1 - Exemplo 3.

Figura 28: Histogramas para os parâmetros farmacocinéticos com as curvas teóricas das funções densidade de probabilidade ajustadas - Exemplo 3.

Tabela 14: Valores de DIC e $LPML$ para os modelos propostos na seção 3.1 para cada parâmetro farmacocinético - Exemplo 3.

Modelos	Parâmetros Farmacocinéticos			
	C_{max}		$ASC_{(0-t)}$	
	DIC	$LPML$	DIC	$LPML$
GGE	143,7	-59,4	388,2	-184,4
TN	146,3	-62,3	388,2	-186,3
ST ($v = 4$)	143,6	-60,3	410,7	-184,6
SN	147,5	-63,1	410,1	-187,1
Padrão	160,0	-62,9	410,0	-187,0

Nas Figuras 27 e 28 observa-se muita proximidade entre o ajuste dos modelos visto que os mesmos apresentam pouca assimetria, sendo modelado de forma parecida pelas 5 distribuições consideradas. Os indicadores dados na tabela 14 apresentam pequena vantagem no ajuste do modelo GGE.

Na Tabela 15 são apresentados os intervalos HPD para a diferença de médias dos parâmetros farmacocinéticos segundo cada formulação. Tomando a média da formulação referência apresentada na tabela 13 para C_{max} , multiplica-se a mesma por $\pm 20\%$, ou seja, $11,34 \times \pm 0,20 = (-2,27; 2,27)$, que são os limites de bioequivalência para este parâmetro farmacocinético. Para ASC_{0-t} , tem-se $201,92 \times \pm 0,20 = (-40,38; 40,38)$ que são os limites de bioequivalência para este parâmetro farmacocinético. Observa-se que o intervalo HPD dos parâmetros farmacocinéticos estão totalmente contidos nos limites de bioequivalência, o que levaria a não rejeição de bioequivalência média entre os fármacos. Além disso, da mesma forma que os critérios de ajustes foram muito parecidos entre os modelos, os intervalos também são bastante similares.

Tabela 15: Intervalos HPD para a decisão de bioequivalência média - Exemplo 3.

Parâmetro farmacoc.	Modelo	média (d.p.)	Intervalo HPD (90%)	Amplitude
C_{max}	GGE	-0,32 (0,30)	-0,80; 0,18	0,98
	Padrão	-0,24 (0,31)	-0,73; 0,28	1,01
ASC_{0-t}	GGE	-4,02 (3,71)	-10,47; 2,68	13,15
	Padrão	-4,08 (4,08)	-10,46; 2,92	13,38

Para finalizar o enfoque univariado deste exemplo, na Figura 29 são apresentados gráficos de dispersão para cada parâmetro farmacocinético em que confronta-se os valores observados com as médias a posteriori calculadas no modelo proposto para cada indivíduo. Observa-se um bom ajuste das médias a posteriori com os dados observados.

Enfoque Bivariado: Na Tabela 16 são apresentados os valores de DIC e $LPML$

Figura 29: Gráficos de dispersão das médias a posteriori segundo cada observação para o modelo GGE - Exemplo 3.

bem como os intervalos *HPD* para os parâmetros que capturam a correlação dos modelos propostos na seção 3.2. Como a Figura 19 apresenta evidências de correlação moderada entre os parâmetros farmacocinéticos deste exemplo, a cópula FGM (que captura correlação fraca) foi descartada neste exemplo.

Na tabela 16 observa-se que todos os modelos capturam a correlação entre os dois vetores de parâmetros farmacocinéticos, porém os modelos *SN* e *ST* tem o valor 0 incluído em seus intervalos enquanto os modelos GGE e *TN* não tem. Entre as distribuições, concordando com a escolha feita no caso univariado, a distribuição GGE sob a função cópula Gaussiana maximiza os valores de *LPML* e minimiza os valores de *DIC*.

Para a decisão de bioequivalência simultânea entre os parâmetros farmacocinéticos construiu-se uma região *HPD* para os mesmos segundo o modelo GGE sob a cópula Gaussiana. Neste sentido, espera-se que esta região esteja contida no retângulo construído com os limites de bioequivalência de $\pm 20\%$ conforme calculado no enfoque univariado deste exemplo. Logo, conforme a decisão univariada um a um, a decisão bivariada apresentada na Figura 30 também leva a não rejeição da bioequivalência média entre as formulações pois a região *HPD* está totalmente contida na área dos limites de bioequivalência. Não foram observadas evidências de efeito de período e sequência em nenhum dos dois enfoques desta aplicação.

Para finalizar este exemplo, na figura 31 são apresentados gráficos de dispersão para cada parâmetro farmacocinético em que confronta-se os valores observados com

Tabela 16: Valores de DIC e $LPML$ e intervalos HPD para os parâmetros de correlação dos modelos propostos na seção 3.2 - Exemplo 3.

Modelo	Função Cópula	Parâmetro de Correlação HPD 90%	DIC	LPML
GGE	Gaussiana	$\rho(\hat{\theta}) = (0, 53; 0, 75)$ $\tau(\hat{\theta}) = (0, 37; 0, 55)$	482, 3	-225, 9
	Clayton	$\tau(\hat{\theta}) = (0, 27; 0, 47)$	489, 8	-229, 2
TN	Gaussiana	$\rho(\hat{\theta}) = (0, 52; 0, 76)$ $\tau(\hat{\theta}) = (0, 36; 0, 56)$	523, 1	-231, 5
	Clayton	$\tau(\hat{\theta}) = (0, 26; 0, 46)$	498, 2	-235, 3
SN Bivariada	—	$\hat{\rho} = (-0, 04; 0, 58)$	541, 6	-247, 0
ST ($v = 4$) Bivariada	—	$\hat{\rho} = (-0, 01; 0, 80)$	540, 1	-234, 5

Figura 30: Região HPD para a decisão conjunta de bioequivalência média para C_{max} e ASC_{0-t} do modelo GGE sob a função cópula Gaussiana - Exemplo 3.

as médias a posteriori calculadas no modelo proposto para cada indivíduo. Observa-se um bom ajuste das médias a posteriori com os dados observados.

Figura 31: Gráficos de dispersão das médias a posteriori segundo cada observação para o modelo GGE sob a função cópula Gaussiana - Exemplo 3.

4.4 Bioequivalência individual e populacional: *FDA*

4.4.1 Conjunto de dados

Para aplicação da metodologia apresentada na seção 3.3 será considerado o conjunto de dados disponibilizado pelo *FDA* em <http://www.fda.gov/Drugs/ScienceResearch/ucm301277.htm> (referência 14d) sobre formulações inibidoras da MAO (monoamina oxidase). Estes dados representam um estudo de bioequivalência cruzado 4×2 (4 períodos e 2 sequências) com 38 indivíduos. O objetivo principal é comparar a bioequivalência individual e populacional entre as duas formulações *R* e *T*. Um resumo do conjunto de dados na escala logarítmica segundo os parâmetros farmacocinéticos C_{max} e ASC_{0-t} considerados neste exemplo é apresentado na Tabela 17.

Tabela 17: Medidas resumos para os dados do exemplo 4.

Parâmetros farmacocinéticos	Formulação	Média	Desvio Padrão	Coefficiente de Variação
$\ln(C_{max})$	T	1,29	0,23	0,18
	R	1,33	0,23	0,18
$\ln(ASC_{0-t})$	T	4,63	0,30	0,06
	R	4,64	0,30	0,06

4.4.2 Elicitação das distribuições a priori

Para as distribuições a priori em (60) e (61) escolheu-se valores para os hiperparâmetros que levem à distribuições aproximadamente não informativas. Logo, assume-se: $a_{Dk} = b_{Dk} = a = b = a_k = b_k = a_{ER} = b_{ER} = a_{ET} = b_{ET} = a_{00} = b_{00} = 0,0001$; $a_\rho = b_\rho = 1$ e $\mu_{00} = 0$.

Foram simuladas 150000 amostras de Gibbs para cada parâmetro usando o software JAGS a partir do software R (biblioteca R2jags). Do total de amostras geradas foram descartadas 50000 amostras (*“burn-in-sample”*) para evitar o possível efeito dos valores iniciais até as cadeias tornarem-se estáveis. Das 100000 amostras restantes para cada parâmetro foram tomados valores de 50 em 50 a fim de se obter amostras não correlacionadas. Foram consideradas no mínimo 2 cadeias de Markov para cada parâmetro com o objetivo de validar as amostras simuladas.

Para as amostras simuladas, a convergência foi verificada utilizando-se gráficos temporais, histogramas e gráficos de autocorrelação. Para os resultados da próxima sessão, observou-se grande indicativo de convergência dos algoritmos utilizados para todos os parâmetros.

4.4.3 Resultados

Como o objetivo a respeito da decisão de bioequivalência populacional e individual é apresentar apenas o procedimento, disponibilizando o código para a obtenção das estimativas do mesmo sob o enfoque Bayesiano (Apêndice A), não haverá comparação de modelos, assumindo-se então o modelo introduzido em (54). Assim, na figura 32 são apresentados gráficos quantil-quantil para verificar o ajuste da distribuição normal assumida para os resíduos. Observa-se um ajuste razoável dos resíduos à distribuição de probabilidade proposta para os dois parâmetros farmacocinéticos.

Na tabela 18 são apresentados os intervalos *HPD* para os 3 tipos de decisão de bioequivalência. Conforme apresentado na seção 3.3, os limites superiores dos intervalos para $\hat{\Theta}_{BI}$ e $\hat{\Theta}_{BP}$ não devem ser maiores do que 2,4948 e 1,7448 para que não sejam rejeitadas as bioequivalências individual e populacional, respectivamente. Neste sentido, observa-se que tanto para o parâmetro farmacocinético $C_{\max}$ quanto para $ASC_{(0-t)}$ pode-se concluir em bioequivalência individual e populacional entre as formulações. Como os dados foram transformados para a escala logarítmica os limites da bioequivalência para a decisão de bioequivalência média são $-0,2231; 0,2231$ (ver equação 8). Assim, também conclui-se em bioequivalência média pois os intervalos *HPD* para os dois parâmetros farmacocinéticos estão totalmente contidos nestes limites.

Figura 32: Gráfico quantil-quantil para os resíduos do modelo ajustado - Exemplo 4.

Tabela 18: Intervalos *HPD* para as decisões de bioequivalência individual, populacional e média - Exemplo 4.

Parâmetro Farmacocinético	Tipo de bioequivalência	Média (d.p.)	Intervalo <i>HPD</i> (90%)
C_{max}	Individual: $\hat{\Theta}_{BI}$	0,96 (0,23)	0,72; 1,30
	Populacional: $\hat{\Theta}_{BP}$	0,05 (0,24)	-0,25; 0,42
	Média: $\hat{\mu}_R - \hat{\mu}_T$	0,04 (0,05)	-0,03; 0,11
$ASC_{(0-t)}$	Individual: $\hat{\Theta}_{BI}$	1,43 (0,43)	1,05; 1,96
	Populacional: $\hat{\Theta}_{BP}$	0,06 (0,21)	-0,23; 0,42
	Média: $\hat{\mu}_R - \hat{\mu}_T$	0,01 (0,06)	-0,08; 0,11

5 CONCLUSÕES E PERSPECTIVAS FUTURAS

Sabe-se que os testes de bioequivalência são regulamentados por importantes agências, geralmente vinculadas ao governo de cada país, e a etapa estatística para a análise dos dados geralmente é padronizada. Isto é importante para o direcionamento das análises, mas em algumas situações, o modelo paramétrico padrão pode não adequar-se aos dados ou até mesmo podem existir melhores modelos que ajustem os dados, o que causa a inquietude dos pesquisadores. Nos exemplos 1 e 2 apresentados nesta tese, observa-se que melhores ajustes podem ser alcançados quando novos modelos são propostos levando em conta algumas especificidades dos conjuntos de dados como a assimetria, por exemplo.

Logo, acredita-se que é importante procurar melhores modelos, uma vez que tal escolha se justifique e que seja guiada pelo bom senso e ética dos profissionais envolvidos. A literatura também apresenta outros estudos com diferentes metodologias para a análise Estatística dos estudos de bioequivalência (por exemplo, GHOSH e KHATTREE (2003), GHOSH e ROSNER (2007), GHOSH e GÖNEN (2008)). Outros modelos e outras distribuições de probabilidade poderiam ser escolhidos para os conjuntos de dados introduzidos nos exemplos deste trabalho levando a bons resultados também, mas acredita-se que a metodologia introduzida nesta tese seja uma boa alternativa no contexto de estudos de bioequivalência. Destaca-se também que para os modelos propostos não houve necessidade de transformações dos dados como as vezes é usado em estudos de bioequivalência para obter um bom ajuste do modelo proposto para o conjunto de dados. A vantagem de analisar os dados em sua escala original está em poder construir os limites de bioequivalência dentro da regra de $\pm 20\%$.

Do ponto de vista das estimativas obtidas utilizando a inferência Bayesiana, métodos tradicionais de estimação (como os estimadores de máxima verossimilhança, por exemplo) também poderiam ser utilizados, embora para os exemplos apresentados, o procedimento iterativo necessário para o uso destes métodos dependeria de valores iniciais cuidadosamente escolhidos, fazendo com que esta tarefa torne-se extremamente complexa em cada caso específico, o que torna o processo pouco atraente. Os resultados apresentados nos exemplos propostos utilizando a inferência Bayesianos foram obtidos assumindo a maioria dos valores iniciais aleatórios bem como distribuições a priori aproximadamente não informativas.

Como qualquer software Estatístico, o uso dos softwares livres para inferência Bayesiana requer um conhecimento prévio do pesquisador sobre a metodologia utilizada para a obtenção das estimativas. O FDA (2010) disponibilizou um guia para o uso de métodos Bayesianos em ensaios clínicos e grande parte das publicações recentes na área de bioequivalência tem a metodologia Bayesiana como estratégia na obtenção das estima-

tivas de interesse. BERRY et al. (2011) publicaram um livro com uma extensa revisão bibliográfica recente e direcionamentos neste sentido.

Nos exemplos 1 e 2 introduzidos neste texto, observou-se que o uso da distribuição GGE destacou-se no ajuste dos modelos para os dados destes exemplos tanto sob o ponto de vista univariado quanto bivariado segundo os critérios de seleção de modelos considerados na seção 3.4 com relação aos demais modelos. A distribuição GGE apresentou-se mais sensível à assimetria dos dados dos exemplos apresentados. Observou-se também a importância da utilização de estrutura de dependência, uma vez que todos os coeficientes de correlação apresentaram evidências de significância (os respectivos *HPD* (90%) não contêm o valor 0). Logo, como a decisão de bioequivalência média geralmente é tomada considerando os parâmetros farmacocinéticos C_{max} e ASC_{0-t} , a decisão simultânea a partir de uma região pode ser mais adequada visto que os modelos propostos tiveram grande capacidade de capturar a correlação entre os parâmetros farmacocinéticos.

Para o exemplo 3, em que não se tem problemas relacionados aos pressupostos de distribuição normal inerentes ao modelo padrão, o ajuste dos modelos propostos, mesmo o baseado na distribuição GGE que tem maior sensibilidade para capturar assimetria, traz ajustes e resultados muito parecidos com o modelo padrão, o que torna estes modelos muito interessantes pois além de serem alternativas para situações não modeladas pelo modelo padrão, ainda tem ajustes e resultados similares quando adotado o modelo padrão.

Quanto aos conceitos de bioequivalência populacional e individual, estes ainda tem pouca aplicação prática, visto que o procedimento de bioequivalência média tem sido aceito pela maioria das agências que regulam os estudos de bioequivalência pelo mundo. Além disso, o planejamento dos estudos de bioequivalência individual e populacional são bastante complexos, necessitando de uma disponibilidade de tempo grande dos voluntários, pois são estudos realizados com no mínimo 4 períodos. Neste sentido, grande parte dos estudos referentes a melhoria dos modelos para a decisão de bioequivalência tem sido direcionados à bioequivalência média.

Além das aplicações em estudos de bioequivalência abordadas neste trabalho, os modelos propostos poderiam ser utilizados na análise de dados em que a variável resposta de interesse é contínua e apresenta assimetria, bastando algumas adaptações nos modelos de regressão. Geralmente o pesquisador recorre a diversos tipos de transformações nos dados, o que traz bastante prejuízo em termos da interpretação dos resultados.

Para uma continuidade desta pesquisa, pode-se propor outras distribuições de probabilidade além das consideradas neste texto, construindo assim novos modelos para a

decisão de bioequivalência média; introduzir misturas de distribuições para os modelos propostos neste trabalho; estender os modelos propostos neste trabalho da decisão de bioequivalência média para a decisão de bioequivalência populacional e individual.

Finalmente, nas situações em que o pesquisador não está seguro quanto aos pressupostos que devem ser assumidos pelo modelo padrão usado para a decisão de bioequivalência média, principalmente nos conjuntos de dados com bastante assimetria como os introduzidos nos exemplos 1 e 2, sugere-se que sejam utilizados os modelos introduzidos neste trabalho no sentido de encontrar um ajuste que possa ser considerado satisfatório e a decisão seja tomada de forma segura no que diz respeito a etapa Estatística levando em conta que os modelos propostos nesta tese tem o modelo padrão como caso particular.

Referências

- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716–723, 1974.
- AMBLARD, C.; GIRARD, S. Symmetry and dependence properties within a semiparametric family of bivariate copulas. **Journal of Nonparametric Statistics**, v. 14, n. 6, p. 715–727, 2002.
- AMBLARD, C.; GIRARD, S. Estimation procedures for a semiparametric family of bivariate copulas. **Journal of Computational and Graphical Statistics**, v. 14, n. 2, p. 363–377, 2005.
- AMBLARD, C.; GIRARD, S. A new extension of bivariate FGM copulas. **Metrika**, v. 70, n. 1, p. 1–17, 2009.
- ANDERSON, S.; HAUCK, W. W. A new procedure for testing equivalence in comparative bioavailability and other clinical trials. **Communications in Statistics - Theory and Methods**, v. 12, n. 23, p. 2663–2692, 1983.
- ANDERSON, S.; HAUCK, W. W. Consideration of individual bioequivalence. **Journal of Pharmacokinetics and Pharmacodynamics**, v. 18, n. 3, p. 259–273, 1990.
- ANVISA. **Manual de boas práticas em biodisponibilidade bioequivalência**. 2002. Gerência-Geral de Inspeção e Controle de Medicamentos e Produtos (GGIMP), Brasília, 2002. Disponível em: <<http://www.anvisa.gov.br/medicamentos/bioequivalencia/publicacoes/index.asp>>. Acesso em: 2014-11-12.
- ANVISA. **Medicamentos genéricos: Oriente-se**. 2002. Gerência-Geral de Regulação Econômica e Monitoramento de Mercado (GGREM), Brasília, 2002. Disponível em: <http://portal.anvisa.gov.br/wps/wcm/connect/a09ae180483c0adca421af0d8b4275ce/genericos_cartilha.pdf?MOD=AJPERES>. Acesso em: 2014-11-12.
- ANVISA. **Resolução número 478, de 19 de março de 2002: Guia para Provas de Bioequivalência de Medicamentos Genéricos**. 2002. Direção da Diretoria Colegiada, Brasília, 2002. Disponível em: <http://www.anvisa.gov.br/hotsite/genericos/legis/resolucoes/2002/478_02re.htm>. Acesso em: 2014-11-12.
- ARMANDO, Y. P. **Avaliação da bioequivalência entre comprimido convencional e comprimido de desintegração oral contendo 8 mg de**

ondansetrona. 2006. Dissertação (Mestrado em Produção e Controle Farmacêuticos) - Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, 2008. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/9/9139/tde-08102009-190008/>>. Acesso em: 2014-11-14.

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.; CAPITANIO, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 65, n. 2, p. 367–389, 2003.

AZZALINI, A.; VALLE, A. D. The multivariate skew-normal distribution. **Biometrika**, v. 83, n. 4, p. 715–726, 1996.

BARRET, J. S. et al. PhRMA perspective on population and individual bioequivalence. **The Journal of Clinical Pharmacology**, v. 40, n. 6, p. 561–570, 2000.

BELOTTO, K. C. R. **Comparação da bioequivalência de duas formulações de risperidona.** 2010. Dissertação (Mestrado em Psiquiatria) - Faculdade de Medicina, Universidade de São Paulo, São Paulo, 2010. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/5/5142/tde-08062010-152653/>>. Acesso em: 2014-11-14.

BERGER, R. L.; HSU, J. C. Bioequivalence trials, intersection-union tests and equivalence confidence sets. **Statistical Science**, v. 11, n. 4, p. pp. 283–302, 1996.

BERRY, S. M. et al. **Bayesian adaptive methods for clinical trials.** 1. ed. Boca Raton: Chapman & Hall, 2011. 311 p.

BRAMLAGE, P.; GOLDIS, A. Bioequivalence study of three ibuprofen formulations after single dose administration in healthy volunteers. **BMC Pharmacology**, v. 8, n. 18, p. 1–9, 2008.

BRIOSCHI, T. M. D. L. S. **Avaliação da bioequivalência de comprimidos contendo 10 mg de cloridrato de ciclobenzaprina.** 2006. Dissertação (Mestrado em Produção e Controle Farmacêuticos) - Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, 2006. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/9/9139/tde-13122006-091912/>>. Acesso em: 2014-11-14.

- CHIB, S.; GREENBERG, E. Understanding the Metropolis-Hastings algorithm. **The American Statistician**, v. 49, n. 4, p. 327–335, 1995.
- CHINCHILLI, V. M.; ESINHART, J. D. Design and analysis of intra-subject variability in cross-over experiments. **Statistics in Medicine**, v. 15, n. 15, p. 1619–1634, 1996.
- CHOW, S.; LIU, J. **Design and analysis of bioavailability and bioequivalence studies**. 3. ed. Boca Raton: Chapman and Hall/CRC, 2009. 758 p.
- CLAYTON, D. G. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. **Biometrika**, v. 65, n. 1, p. 141–151, 1978.
- COX, C. et al. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. **Statistics in Medicine**, v. 26, n. 23, p. 4352–4374, 2007.
- CZADO, C. Pair-copula constructions of multivariate copulas. In: JAWORSKI, P. et al. (Ed.). **Copula Theory and Its Applications**. Berlin Heidelberg: Springer, 2010, (Lecture Notes in Statistics, v. 198). p. 93–109.
- DIAS, C. A. R. C.; ROMANO-LIEBER, N. S. Processo da implantação da política de medicamentos genéricos no Brasil. **Cadernos de Saúde Pública**, v. 22, p. 1661 – 1669, 2006.
- DUONG, T.; HAZELTONA, M. Plug-in bandwidth matrices for bivariate kernel density estimation. **Journal of Nonparametric Statistics**, v. 1, n. 15, p. 17–30, 2003.
- EFRON, B. Bootstrap methods: Another look at the Jackknife. **The Annals of Statistics**, v. 7, n. 1, p. 1–26, 1979.
- FAREWELL, V. T.; PRENTICE, R. L. A study of distributional shape in life testing. **Technometrics**, v. 19, n. 1, p. 69–75, 1977.
- FARLIE, D. J. G. The performance of some correlation coefficients for a general bivariate distribution. **Biometrika**, v. 47, n. 3-4, p. 307–323, 1960.
- FDA. **Guidance for industry: Average, population, and individual approaches to establishing bioequivalence**. 1999. Center for Drug Evaluation and Research (CDER), Rockville, 1999. Disponível em: <<http://www.fda.gov/OHRMS/DOCKETS/98fr/3657gd1.pdf>>. Acesso em: 2014-11-12.

FDA. Guidance for industry: Statistical approaches to establishing bioequivalence. 2001. Center for Drug Evaluation and Research (CDER), Rockville, 2001. Disponível em: <<http://www.fda.gov/downloads/.../Guidances/ucm070244.pdf>>. Acesso em: 2014-11-12.

FDA. Guidance for industry: Bioavailability and bioequivalence studies for orally administered drug products - General considerations. 2003. Center for Drug Evaluation and Research (CDER), Rockville, 2003. Disponível em: <<http://www.fda.gov/downloads/Drugs/.../Guidances/ucm070124.pdf>>. Acesso em: 2014-11-12.

FDA. Guidance for the use of Bayesian statistics in medical device clinical trials. 2010. Center for Drug Evaluation and Research (CDER), Rockville, 2010. Disponível em: <<http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm071072.htm>>. Acesso em: 2014-11-12.

GELFAND, A. E. Model determination using sampling-based methods. In: GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. J. (Ed.). **Markov Chain Monte Carlo in practice**. London, UK: Chapman & Hall, 1996. cap. 9, p. 145–161.

GELFAND, A. E.; SMITH, A. F. M. Sampling-based approaches to calculating marginal distributions. **Journal of the American Statistical Association**, v. 85, n. 410, p. 398–409, 1990.

GELMAN, A. Prior distributions for variance parameters in hierarchical models. **Bayesian Analysis**, v. 1, n. 3, p. 1–19, 2006.

GHOSH, P.; GÖNEN, M. Bayesian modeling of multivariate average bioequivalence. **Statistics in Medicine**, v. 27, n. 13, p. 2402–19, 2008.

GHOSH, P.; KHATTREE, R. Bayesian approach to average bioequivalence using Bayes' factor. **Journal of Biopharmaceutical Statistics**, v. 13, n. 4, p. 719–734, 2003.

GHOSH, P.; NTZOUFRAS, I. **Testing population and individual bioequivalence: A hierarchical Bayesian approach**. Athens University of Economics and Business, 2005. 31 p.

GHOSH, P.; ROSNER, G. L. A semi-parametric Bayesian approach to average bioequivalence. **Statistics in Medicine**, v. 26, n. 6, p. 1224–36, 2007.

- GRIZZLE, J. E. The two-period change-over design and its use in clinical trials. **Biometrics**, v. 21, n. 2, p. 467–480, 1965.
- GUMBEL, E. J. Bivariate exponential distributions. **Journal of American Statistical Association**, v. 55, n. 292, p. 307–323, 1960.
- HAUCK, W. W.; ANDERSON, S. Types of bioequivalence and related statistical considerations. **International Journal of Clinical Pharmacology & Toxicology**, v. 30, n. 5, p. 181–187, 1992.
- HAUSCHKE, D.; STEINIJANS, V.; PIGEOT, I. **Bioequivalence studies in drug development**. 1. ed. Hoboken: John Wiley & Sons Ltd, 2007. 312 p.
- HSU, H. C.; LU, H. L.; CHAN, K. K. A novel multivariate approach for estimating the degree of similarity in bioavailability between two pharmaceutical products. **Journal of Pharmaceutical Sciences**, v. 84, n. 6, p. 768–772, 1995.
- HYSLOP, T.; HSUAN, F.; HOLDER, D. J. A small sample confidence interval approach to assess individual bioequivalence. **Statistics in Medicine**, v. 19, n. 20, p. 2885–2897, 2000.
- ISMAIL, Z.; WAHAB, M.; RAHMAN, A. Bioequivalence study of propranolol tablets. **European Journal of General Medicine**, v. 1, n. 4, p. 42–47, 2004.
- KHANDAVE, S. S. et al. Comparative bioequivalence studies of tramadol hydrochloride sustained-release 200 mg tablets. **Drug Design, Development and Therapy**, v. 25, n. 4, p. 367–74, 2010.
- KOONO, E. E. M. **Avaliação de bioequivalência de comprimidos contendo 500mg de tinidazol**. 2005. Dissertação (Mestrado em Produção e Controle Farmacêuticos) - Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, 2005. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/9/9139/tde-18092007-111223/>>. Acesso em: 2014-11-12.
- LANGE, K. L.; LITTLE, R. J. A.; TAYLOR, J. M. G. Robust statistical modeling using t distribution. **Journal of the American Statistical Association**, v. 84, n. 408, p. 881–896, 1989.
- LEE, L.-F. Generalized econometric models with selectivity. **Econometrica**, v. 51, n. 2, p. 507–512, 1983.
- LUNN, D. J. et al. WinBUGS: A Bayesian modelling framework: Concepts, structure, and extensibility. **Statistics and Computing**, v. 10, n. 4, p. 325–337, 2000.

- MAITI, S. S.; DEY, M. Tilted normal distribution and its survival properties. **Journal of Data Science**, v. 10, p. 225–240, 2012.
- MANDALLAZ, D.; MAU, J. Comparison of different methods for decision-making in bioequivalence assessment. **Biometrics**, v. 37, n. 2, p. 213–222, 1981.
- MARSHALL, A. W.; OLKIN, I. A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. **Biometrika**, v. 84, n. 3, p. 641–652, 1997.
- MCLEAN, R. A.; SANDERS, W. L.; STROUP, W. W. A unified approach to mixed linear models. **American Statistical Association**, v. 45, n. 1, p. 54–64, 1991.
- METZLER, C. M. Bioavailability—a problem in equivalence. **Biometrics**, v. 30, n. 2, p. 309–317, 1974.
- MORGENSTERN, D. Einfache beispiele zweidimensionaler verteilungen. **Mitteilungsblatt für Mathematische Statistik**, v. 8, p. 234–235, 1956.
- NELSEN, R. **An introduction to copulas**. Segunda edição. New York: Springer Science+Business Media, Inc., 2006. 269 p.
- NISHIJIMA, M. **Análise econômica dos medicamentos genéricos no Brasil**. 2003. Tese (Doutorado em Teoria Econômica) - Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2003. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/12/12138/tde-12122005-101205/>>. Acesso em: 2014-11-12.
- PHILLIPS, K. F. Power of the two one-sided tests procedure in bioequivalence. **Journal of Pharmacokinetics and Biopharmaceutics**, v. 18, n. 2, p. 137–144, 1990.
- PLUMMER, M. **JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling**. 2003.
- PRENTICE, R. L. A log gamma model and its maximum likelihood estimation. **Biometrika**, v. 61, p. 539–544, 1974.
- QUENTAL, C. et al. Medicamentos genéricos no Brasil: impactos das políticas públicas sobre a indústria nacional. **Ciência & Saúde Coletiva**, v. 13, p. 619 – 628, 2008.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2014.

RACINE, A. et al. Bayesian Methods in Practice: Experiences in the Pharmaceutical Industry. **Applied Statistics**, v. 35, n. 2, p. 93–150, 1986.

RACINE-POON, A. et al. A two-stage procedure for bioequivalence studies. **Biometrics**, v. 43, n. 4, p. 847–856, 1987.

SAHU, S. K.; DEY, D. K.; BRANCO, M. D. A new class of multivariate skew distributions with applications to bayesian regression models. **Canadian Journal of Statistics**, v. 31, n. 2, p. 129–150, 2003.

SALAZAR, A. et al. **O SUS pode ser seu melhor plano de saúde**. Brasília, 2003. Acessado em: 29/10/2014. Disponível em: <http://conselho.saude.gov.br/biblioteca/livros/sus_plano_saude.pdf>.

SAMPAIO, M. R. d. M. **Avaliação da bioequivalência de formulações contendo lorazepam através de método bioanalítico utilizando a cromatografia líquida acoplada ao sistema de detecção por espectrometria de massa**. 2008. Dissertação (Mestrado em Produção e Controle Farmacêuticos) - Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, 2008. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/9/9139/tde-16062008-222314/>>. Acesso em: 2014-11-12.

SAS Institute Inc. **SAS/STAT Software, Version 9.3**. Cary, NC, 2011.

SCHUIRMANN, D. J. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. **Journal of Pharmacokinetics and Biopharmaceutics**, v. 15, n. 6, p. 657–680, 1987.

SELWYN, M. R.; DEMPSTER, A. P.; HALL, N. R. A Bayesian approach to bioequivalence for the 2 x 2 changeover design. **Biometrics**, v. 37, n. 1, p. 11–21, 1981.

SELWYN, M. R.; HALL, N. R. On Bayesian methods for bioequivalence. **Biometrics**, v. 40, n. 4, p. 1103–1108, 1984.

SHAO, J.; CHOW, S.-C.; WANG, B. The bootstrap procedure in individual bioequivalence. **Statistics in Medicine**, v. 19, n. 20, p. 2741–2754, 2000.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, v. 52, n. 3-4, p. 591–611, 1965.

SILVA, M. d. F. **Determinação da bioequivalência do metronidazol a partir de comprimidos revestidos**. 2009. Dissertação (Mestrado

- em Produção e Controle Farmacêuticos) - Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, 2009. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/9/9139/tde-16112009-181138/>>. Acesso em: 2014-11-12.
- SKLAR, A. Random variables, distribution functions, and copulas. **Kybernetika**, v. 9, n. 6, p. 449–460, 1973.
- SOUZA, R. M.; ACHCAR, J. A.; MARTINEZ, E. Z. Use of Bayesian methods for multivariate bioequivalence measures. **Journal of Biopharmaceutical Statistics**, v. 19, n. 1, p. 42–66, 2009.
- SOUZA, R. M. et al. Bioequivalência: Uso de métodos Bayesianos. **Revista Brasileira de Biometria**, v. 25, n. 4, p. 115–134, 2007.
- SPIEGELHALTER, D. J. et al. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 64, n. 4, p. 583–639, 2002.
- STACY, E. W. A generalization of the gamma distribution. **The Annals of Mathematical Statistics**, v. 33, n. 3, p. 1187–1192, 1962.
- TRIVEDI, P. K.; ZIMMER, D. M. Copula modeling: An introduction for practitioners. **Foundations and Trends in Econometrics**, v. 1, n. 1, p. 1–111, 2005.
- VIEIRA, F.; ZUCCHI, P. Diferenças de preços entre medicamentos genéricos e de referência no Brasil. **Revista de Saúde Pública**, v. 40, n. 3, p. 444–449, 2006.
- WESTLAKE, W. J. Use of confidence intervals in analysis of comparative bioavailability trials. **Journal of Pharmaceutical Sciences**, v. 61, n. 8, p. 1340–1341, 1972.
- WILK, M. B.; GNANADESIKAN, R. Probability plotting methods for the analysis of data. **Biometrika**, v. 55, n. 1, p. 1–17, 1968.

APÊNDICE

A Códigos BUGS e JAGS para os exemplos

Listagem 1: Código JAGS - GGE Univariada

```
data 1
{ 2
  for(i in 1:n) 3
  { 4
    for(j in 1:2) 5
    { 6
      zeros[i,j]<- 0 7
    } 8
  } 9
} 10
11

model 12
{ 13
  for(i in 1:n) 14
  { 15
    for(j in 1:2) 16
    { 17
      phi[i,j]<- -log.L[i,j] 18
      zeros[i,j] ~ dpois(phi[i,j]) 19
    } 20
    mu[i,j]<- b0 + b1*X1[i,j] + b2*X2[i,j] + b3*X3[i,j] + omega[i] 21
    log.L[i,j]<- log(abs(q)) -log(sigma) -loggam(k) + w[i,j]*k - 22
      exp(w[i,j]) 23
    w[i,j]<- (y[i,j]-mu[i,j])*(1/sigma)*q + dig.k 24
  } 25
} 26
27
for(i in 1:n) 28
{ 29
  omega[i] ~ dnorm(0,tau.omega) 30
} 31
32
b0 ~ dnorm(0.0, 1.0E-6) 33
b1 ~ dnorm(0.0, 1.0E-6) 34
b2 ~ dnorm(0.0, 1.0E-6) 35
b3 ~ dnorm(0.0, 1.0E-6) 36
37
```

<code>tau ~ dgamma(0.001, 0.001)</code>	38
<code>sigma <- 1 / sqrt(tau)</code>	39
	40
	41
<code>tau.omega ~ dgamma(0.001, 0.001)</code>	42
<code>sigma.omega <- 1/sqrt(tau.omega)</code>	43
	44
<code>q ~ dnorm(0,1)I(,0) "OU" q ~ dnorm(0,1)I(0,)</code>	45
	46
<code>k<- pow((1/q),2)</code>	47
	48
<code>dig.k<- -0.5772 - (1/k) + k*sum(serie[])</code>	49
	50
<code>for(l in 1:1000)</code>	51
<code>{</code>	52
<code>serie[l]<- 1/(1*(k+1))</code>	53
<code>}</code>	54
<code>}</code>	55

Listagem 2: Código JAGS - TN Univariada

<code>data</code>	1
<code>{</code>	2
<code>for(i in 1:n)</code>	3
<code>{</code>	4
<code>for(j in 1:2)</code>	5
<code>{</code>	6
<code>zeros[i,j]<- 0</code>	7
<code>}</code>	8
<code>}</code>	9
<code>}</code>	10
	11
<code>model</code>	12
<code>{</code>	13
<code>for(i in 1:n)</code>	14
<code>{</code>	15
<code>for(j in 1:2)</code>	16
<code>{</code>	17
<code>phi[i,j]<- -log.L[i,j]</code>	18
<code>zeros[i,j] ~ dpois(phi[i,j])</code>	19
<code>}</code>	20
<code>mu[i,j]<- b0 + b1*X1[i,j] + b2*X2[i,j] + b3*X3[i,j] + omega[i]</code>	21

```

log.L[i,j]<- log(beta) - log(sqrt(2*3.14159*pow(sigma,2))) - 22
0.5*pow(((y[i,j]-mu[i,j])/sigma),2) - 23
2*log(1-(1-beta)*(1-phi((y[i,j]-mu[i,j])/sigma))) 24
} 25
} 26
} 27
} 28

for(i in 1:n) 29
{ 30
  omega[i] ~ dnorm(0,tau.omega) 31
} 32
} 33

b0 ~ dnorm(0.0, 1.0E-6) 34
b1 ~ dnorm(0.0, 1.0E-6) 35
b2 ~ dnorm(0.0, 1.0E-6) 36
b3 ~ dnorm(0.0, 1.0E-6) 37
} 38

tau ~ dgamma(0.001, 0.001) 39
sigma <- 1 / sqrt(tau) 40
} 41

tau.omega ~ dgamma(0.001, 0.001) 42
sigma.omega <- 1/sqrt(tau.omega) 43
} 44

beta ~ dgamma(1,1) 45

```

Listagem 3: Código BUGS - SN Univariada (Estratégia ii)

```

model 1
{ 2
  for(i in 1:n) 3
  { 4
    for(j in 1:2) 5
    { 6
      y[i,j] ~ dnorm(mu[i,j],tau) 7
      mu[i,j]<- b0 + b1*X1[i,j] + b2*X2[i,j] + b3*X3[i,j] + 8
        delta*z[i] + omega[i] 9
    } 10
  } 11
} 12

for(i in 1:n) 13
{ 14
  omega[i] ~ dnorm(0,tau.omega) 15
}

```

z[i] ~ dnorm(0,1)T(0,)	16
}	17
	18
b0 ~ dnorm(0.0, 1.0E-6)	19
b1 ~ dnorm(0.0, 1.0E-6)	20
b2 ~ dnorm(0.0, 1.0E-6)	21
b3 ~ dnorm(0.0, 1.0E-6)	22
	23
tau ~ dgamma(0.001, 0.001)	24
sigma <- 1 / sqrt(tau)	25
	26
tau.omega ~ dgamma(0.001, 0.001)	27
sigma.omega <- 1/sqrt(tau.omega)	28
	29
delta ~ dnorm(0,0.1)	30

Listagem 4: Código BUGS - ST Univariada (Estratégia ii)

model	1
{	2
for(i in 1:n)	3
{	4
for(j in 1:2)	5
{	6
y[i,j] ~ dnorm(mu[i,j], tau2[i])	7
mu[i,j] <- b0 + b1*X1[i,j] + b2*X2[i,j] + b3*X3[i,j] +	8
delta*z[i] + omega[i]	9
}	10
}	11
	12
for(i in 1:n)	13
{	14
zeta[i] ~ dgamma(v/2, v/2)	15
tau2[i] <- tau1*zeta[i]	16
	17
omega[i] ~ dnorm(0, tau.omega)	18
z[i] ~ dnorm(0,1)T(0,)	19
}	20
	21
b0 ~ dnorm(0.0, 1.0E-6)	22
b1 ~ dnorm(0.0, 1.0E-6)	23
b2 ~ dnorm(0.0, 1.0E-6)	24

```

b3 ~ dnorm(0.0 , 1.0E-6) 25
tau ~ dgamma(0.001 , 0.001) 26
sigma <- 1 / sqrt(tau) 27
tau.omega ~ dgamma(0.001 , 0.001) 28
sigma.omega <- 1/sqrt(tau.omega) 29
delta ~ dnorm(0,0.1) 30
v<- 4 31

```

Listagem 5: Código JAGS - GGE Bivariada - Cópula Gaussiana

```

data 1
{ 2
  for(i in 1:n) 3
  { 4
    for(j in 1:2) 5
    { 6
      zeros[i,j]<- 0 7
    } 8
  } 9
} 10

model 11
{ 12
  for(i in 1:n) 13
  { 14
    for(j in 1:2) 15
    { 16
      phi[i,j]<- -log.L[i,j] 17
      zeros[i,j] ~ dpois(phi[i,j]) 18
    } 19
  } 20

  log.L[i,j]<- log(abs(q1)) -log(sigma1) -loggam(k1) + 21
               w1[i,j]*k1 - exp(w1[i,j]) + log(abs(q2)) 22
               -log(sigma2) -loggam(k2) + w2[i,j]*k2 23
               -exp(w2[i,j]) + l.C12[i] 24
    } 25

  w1[i,j]<- (y1[i,j]-mu1[i,j])*(1/sigma1)*q1 + dig.k1 26
  w2[i,j]<- (y2[i,j]-mu2[i,j])*(1/sigma2)*q2 + dig.k2 27
    } 28

  mul[i,j]<- b01 + b11*X1[i,j] + b21*X2[i,j] + b31*X3[i,j] + 29

```



```

                                omega[i,1]                                30
mu2[i,j]<- b02 + b12*X1[i,j] + b22*X2[i,j] + b32*X3[i,j] +          31
                                omega[i,2]                                32
    }                                                                    33
l.C12[i]<- -0.5*log(1-pow(theta,2)) - (1/(2*(1-pow(theta,2)))) * 34
                                (pow(u[i],2)+pow(z[i],2)-2*theta*u[i]*z[i]) + 35
                                0.5*(pow(u[i],2)+pow(z[i],2))          36
omega[i,1:2]~dmnorm(mu0[, ],tau.omega[, ])                            37
    }                                                                    38
                                                                    39
k1<- pow((1/q1),2)                                                    40
dig.k1<- -0.5772 - (1/k1) + k1*sum(serie1 [])                        41
                                                                    42
for(l in 1:1000)                                                       43
{                                                                        44
    serie1[l]<- 1/(l*(k1+l))                                            45
}                                                                        46
                                                                    47
k2<- pow((1/q2),2)                                                    48
dig.k2<- -0.5772 - (1/k2) + k2*sum(serie2 [])                        49
                                                                    50
for(l in 1:1000)                                                       51
{                                                                        52
    serie2[l]<- 1/(l*(k2+l))                                            53
}                                                                        54
                                                                    55
tau.omega[1:2,1:2]<-inverse(sigma.omega[, ])                          56
                                                                    57
sigma.omega[1,1]<-1/(sigma.a3)                                          58
sigma.a3 ~ dgamma(0.01,0.01)                                          59
sigma.omega[2,2]<-1/(sigma.b3)                                          60
sigma.b3 ~ dgamma(0.01,0.01)                                          61
                                                                    62
sigma.omega[1,2]<-rho*sqrt(1/(sigma.a3)*1/(sigma.b3))                 63
sigma.omega[2,1]<-sigma.omega[1,2]                                     64
                                                                    65
rho ~ dunif(-1,1)                                                      66
                                                                    67
tau1 ~ dgamma(0.0001,0.0001)                                          68
sigma1<- 1/sqrt(tau1)                                                  69
                                                                    70

```

tau2 ~ dgamma(0.0001,0.0001)	71
sigma2<- 1/sqrt(tau2)	72
	73
q1 ~ dnorm(0,1)I(0,) "OU" q1 ~ dnorm(0,1)I(,0)	74
q2 ~ dnorm(0,1)I(0,) "OU" q2 ~ dnorm(0,1)I(,0)	75
	76
b01 ~ dnorm(0,0.0001)	77
b02 ~ dnorm(0,0.0001)	78
b11 ~ dnorm(0,0.0001)	79
b12 ~ dnorm(0,0.0001)	80
b21 ~ dnorm(0,0.0001)	81
b22 ~ dnorm(0,0.0001)	82
b31 ~ dnorm(0,0.0001)	83
b32 ~ dnorm(0,0.0001)	84
	85
theta ~ dunif(-1,1)	86
	87
tau.theta<- (2/3.14159) * arcsin(theta)	88
rho.theta<- (6/3.14159) * arcsin(theta/2)	89
}	90

Listagem 6: Código JAGS - TN Bivariada - Cópula de Clayton

data	1
{	2
for(i in 1:n)	3
{	4
for(j in 1:2)	5
{	6
zeros[i,j]<- 0	7
}	8
}	9
}	10
	11
model	12
{	13
for(i in 1:n)	14
{	15
for(j in 1:2)	16
{	17
phi[i,j]<- -log.L[i,j]	18
zeros[i,j] ~ dpois(phi[i,j])	19

```

log.L[i,j]<- log(lambda1) - log(sqrt(2*3.14159*pow(sigma1,2)))
-0.5*pow(((y1[i,j]-mu1[i,j])/sigma1),2) -
2*log(1-(1-lambda1)*(1-phi((y1[i,j]-mu1[i,j])/sigma1))) +
log(lambda2) - log(sqrt(2*3.14159*pow(sigma2,2))) -
0.5*pow(((y2[i,j]-mu2[i,j])/sigma2),2) -
2*log(1-(1-lambda2)*(1-phi((y2[i,j]-mu2[i,j])/sigma2))) +
1.C12[i]

mul[i,j]<- b01 + b11*X1[i,j] + b21*X2[i,j] + b31*X3[i,j] +
omega[i,1]
mu2[i,j]<- b02 + b12*X1[i,j] + b22*X2[i,j] + b32*X3[i,j] +
omega[i,2]
}

1.C12[i]<- log(1+theta) - (theta+1)*log(u1[i]*u2[i]) -
(2+(1/theta))*log(pow(u1[i],-theta) + pow(u2[i],-theta) - 1)
omega[i,1:2]~dmnorm(mu0[, ],tau.omega[, ])
}

tau.omega[1:2,1:2]<-inverse(sigma.omega[, ])

sigma.omega[1,1]<-1/(sigma.a3)
sigma.a3 ~ dgamma(0.01,0.01)
sigma.omega[2,2]<-1/(sigma.b3)
sigma.b3 ~ dgamma(0.01,0.01)

sigma.omega[1,2]<-rho*sqrt(1/(sigma.a3)*1/(sigma.b3))
sigma.omega[2,1]<-sigma.omega[1,2]

rho ~ dunif(-1,1)

tau1 ~ dgamma(0.0001,0.0001)
sigma1<- 1/sqrt(tau1)

tau2 ~ dgamma(0.0001,0.0001)
sigma2<- 1/sqrt(tau2)

lambda1 ~ dgamma(1,1)
lambda2 ~ dgamma(1,1)

```

b01 ~ dnorm(0,0.0001)	61
b02 ~ dnorm(0,0.0001)	62
b11 ~ dnorm(0,0.0001)	63
b12 ~ dnorm(0,0.0001)	64
b21 ~ dnorm(0,0.0001)	65
b22 ~ dnorm(0,0.0001)	66
b31 ~ dnorm(0,0.0001)	67
b32 ~ dnorm(0,0.0001)	68
	69
theta ~ dgamma(1,1)	70
	71
tau.theta<- theta/(theta + 2)	72
}	73

Listagem 7: Código BUGS - SN Bivariada (Estratégia ii)

model	1
{	2
for(i in 1:n)	3
{	4
for(j in 1:2)	5
{	6
y[i,j,1:2] ~ dmnorm(mu[i,j,],tau[,])	7
	8
mu[i,j,1]<- b0[1] + b1[1]*X1[i,j] + b2[1]*X2[i,j] +	9
b3[1]*X3[i,j] + omega[i,1] + delta[1]*z[i,1]	10
mu[i,j,2]<- b0[2] + b1[2]*X1[i,j] + b2[2]*X2[i,j] +	11
b3[2]*X3[i,j] + omega[i,2] + delta[2]*z[i,2]	12
}	13
	14
omega[i,1:2]~dmnorm(mu0[,],tau.omega[,])	15
	16
z[i,1] ~ dnorm(0,1)T(0,)	17
z[i,2] ~ dnorm(0,1)T(0,)	18
}	19
	20
tau[1:2,1:2] ~ dwish(RT1[,],3)	21
sigma[1:2,1:2]<- inverse(tau[1:2,1:2])	22
rho<- sigma[1,2]/sqrt(sigma[1,1]*sigma[2,2])	23
	24
rho3<- sigma.omega[1,2]/sqrt(sigma.omega[1,1]*sigma.omega[2,2])	25
tau.omega[1:2,1:2] ~ dwish(RT3[,],2)	26

```

sigma.omega[1:2,1:2] <- inverse(tau.omega[1:2,1:2])      27

                                                                28
RT1[1,1] <- 0.001                                         29
RT1[2,2] <- 0.001                                         30
RT1[1,2] <- 0                                             31
RT1[2,1] <- 0                                             32
                                                                33
RT3[1,1] <- 0.01                                          34
RT3[2,2] <- 0.01                                          35
RT3[1,2] <- 0                                             36
RT3[2,1] <- 0                                             37
                                                                38
delta[1] ~ dnorm(0,0.1)                                   39
delta[2] ~ dnorm(0,0.1)                                   40
                                                                41
b0[1] ~ dnorm(0,0.0001)                                   42
b1[1] ~ dnorm(0,0.0001)                                   43
b2[1] ~ dnorm(0,0.0001)                                   44
b3[1] ~ dnorm(0,0.0001)                                   45
                                                                46
b0[2] ~ dnorm(0,0.0001)                                   47
b1[2] ~ dnorm(0,0.0001)                                   48
b2[2] ~ dnorm(0,0.0001)                                   49
b3[2] ~ dnorm(0,0.0001)                                   50
}                                                         51

```

Listagem 8: Código BUGS - ST Bivariada (Estratégia ii)

```

model                                                    1
{                                                        2
  for(i in 1:n)                                         3
  {                                                    4
    for(j in 1:2)                                       5
    {                                                  6
      y[i,j,1:2] ~ dmnorm(mu[i,j,],tau1d[i,,])        7
                                                                8
      mu[i,j,1] <- b0[1] + b1[1]*X1[i,j] + b2[1]*X2[i,j] +
        b3[1]*X3[i,j] + omega[i,1] + delta[1]*z[i,1]    9
      mu[i,j,2] <- b0[2] + b1[2]*X1[i,j] + b2[2]*X2[i,j] +
        b3[2]*X3[i,j] + omega[i,2] + delta[2]*z[i,2]   10
    }                                                  11
  }                                                    12
}                                                        13

```

```

zeta[i] ~ dgamma(pv,pv) 15

16

tauld[i,1,1]<- tau[1,1]*zeta[i] 17
tauld[i,1,2]<- tau[1,2]*zeta[i] 18
tauld[i,2,1]<- tau[2,1]*zeta[i] 19
tauld[i,2,2]<- tau[2,2]*zeta[i] 20
21

omega[i,1:2] ~ dmnorm(mu0[,],tau.omega[,]) 22
23

sigma1d[i,1:2,1:2]<- inverse(tauld[i,,]) 24
rho[i]<- sigma1d[i,1,2]/sqrt(sigma1d[i,1,1]*sigma1d[i,2,2]) 25
26

z[i,1] ~ dnorm(0,1)T(0,) 27
z[i,2] ~ dnorm(0,1)T(0,) 28
} 29
30

tau[1:2,1:2] ~ dwish(RT1[,],3) 31
sigma[1:2,1:2]<- inverse(tau[1:2,1:2]) 32
rho1<- sigma[1,2]/sqrt(sigma[1,1]*sigma[2,2]) 33
34

tau.omega[1:2,1:2] ~ dwish(RT3[,],2) 35
sigma.omega[1:2,1:2]<- inverse(tau.omega[1:2,1:2]) 36
rho3<- sigma3[1,2]/sqrt(sigma3[1,1]*sigma3[2,2]) 37
38

RT1[1,1]<- 0.001 39
RT1[2,2]<- 0.001 40
RT1[1,2]<- 0 41
RT1[2,1]<- 0 42
43

RT3[1,1]<- 0.01 44
RT3[2,2]<- 0.01 45
RT3[1,2]<- 0 46
RT3[2,1]<- 0 47
48

pv<- v/2 49
v <- 4 50
51

delta[1]~dnorm(0,0.1) 52
delta[2]~dnorm(0,0.1) 53
54

b0[1]~dnorm(0,0.0001) 55

```

b1[1] ~ dnorm(0,0.0001)	56
b2[1] ~ dnorm(0,0.0001)	57
b3[1] ~ dnorm(0,0.0001)	58
	59
b0[2] ~ dnorm(0,0.0001)	60
b1[2] ~ dnorm(0,0.0001)	61
b2[2] ~ dnorm(0,0.0001)	62
b3[2] ~ dnorm(0,0.0001)	63
}	64

Listagem 9: Código JAGS - Bioequivalência individual e populacional

model	1
{	2
for(i in 1:2)	3
{	4
for(j in 1:20)	5
{	6
for(k in 1:2)	7
{	8
for(l in 1:2)	9
{	10
y[i,j,k,l] ~ dnorm(mu[i,j,k,l],tau.W[k])	11
mu[i,j,k,l] <- m[k] + gamma[i,k,l] + delta[i,j,k]	12
}	13
}	14
}	15
}	16
	17
for(i in 1:2)	18
{	19
for(j in 1:20)	20
{	21
delta[i,j,1:2] ~ dmnorm(mu0[,],tau4[,])	22
}	23
}	24
	25
tau4[1:2,1:2] <- inverse(sigma4[,])	26
	27
sigma4[1,1] <- 1/(sigma.a4)	28
sigma.a4 ~ dgamma(0.0001,0.0001)	29
sigma4[2,2] <- 1/(sigma.b4)	30

```

sigma.b4 ~ dgamma(0.0001,0.0001) 31

sigma4[1,2] <- rho4*sqrt(sigma4[1,1]*sigma4[2,2]) 32
sigma4[2,1] <- sigma4[1,2] 33
34
35
U ~ dbeta(1,1) 36
37
rho4 <- (2*U) - 1 38
39
tau00 ~ dgamma(0.0001,0.0001) 40
41
sigma2.TT<- pow(sigma.W[1],2) + sigma4[1,1] 42
sigma2.TR<- pow(sigma.W[2],2) + sigma4[2,2] 43
44
delta.PBE<- (pow(m[1]-m[2],2) + sigma2.TT - sigma2.TR)/ 45
max(sigma2.TR,0.04) 46
47
delta.IBE<- (pow(m[1]-m[2],2) + sigma2D + pow(sigma.W[1],2) + 48
pow(sigma.W[2],2))/max(pow(sigma.W[2],2),0.04) 49
50
delta.ABE<- mean(mu[2]) - mean(mu[1]) 51
52
sigma2D<- pow(sqrt(sigma4[1,1]) - sqrt(sigma4[2,2]),2) + 53
2*(1-rho4)*sqrt(sigma4[1,1]*sigma4[2,2]) 54
55
for(k in 1:2) 56
{ 57
tau.W[k] ~ dgamma(0.0001,0.0001) 58
sigma.W[k]<- 1/sqrt(tau.W[k]) 59
60
m[k] ~ dnorm(mu00[k],tau0[k]) 61
mu00[k] ~ dnorm(0,tau00) 62
tau0[k] ~ dgamma(0.0001,0.0001) 63
sigma0[k] <- 1/sqrt(tau0[k]) 64
65
gamma[1,k,1] ~ dnorm(0,tau1[k]) 66
67
gamma[1,k,2] ~ dnorm(0,tau1[k]) 68
gamma[2,k,1] ~ dnorm(0,tau1[k]) 69
gamma[2,k,2] ~ dnorm(0,tau1[k]) 70
71

```


tau1[k] ~ dgamma(0.0001,0.0001)	72
sigma1[k] <- 1/sqrt(tau1[k])	73
}	74
}	75

B Estimativas dos parâmetros

Tabela 19: Médias a posteriori (e desvio padrão) - Exemplo 1 - Univariado.

Exemplo	GGE		TN		SN		ST	
	$C_{\max}$	ASC_{0-t}	$C_{\max}$	ASC_{0-t}	$C_{\max}$	ASC_{0-t}	$C_{\max}$	ASC_{0-t}
1	36,08 (4,01)	253,06 (33,65)	38,57 (6,03)	277,11 (50,61)	36,05 (4,37)	248,75 (34,02)	36,05 (4,46)	249,97 (33,69)
$\hat{\beta}_0$	-1,37 (1,53)	-8,55 (9,79)	-1,51 (1,44)	-5,17 (11,10)	-1,48 (1,47)	-3,66 (11,31)	-1,45 (1,60)	-6,64 (9,57)
$\hat{\beta}_1$	0,65 (1,47)	6,25 (10,06)	0,80 (1,52)	10,27 (10,13)	1,05 (1,49)	12,52 (10,28)	0,87 (1,61)	10,07 (9,77)
$\hat{\beta}_2$	-7,90 (4,64)	-49,64 (44,36)	-8,51 (4,63)	-45,08 (40,45)	-9,08 (4,38)	-55,22 (43,87)	-8,96 (4,49)	-52,60 (46,41)
$\hat{\beta}_3$	-0,79 (0,54)	-1,12 (0,57)	-	-	-	-	-	-
$\hat{q}$	-	-	0,89 (0,91)	0,68 (0,85)	-	-	-	-
$\hat{\lambda}$	-	-	-	-	0,18 (3,37)	0,07 (3,21)	0,22 (3,33)	0,11 (3,20)
$\hat{\delta}$	4,36 (1,07)	25,65 (7,25)	5,37 (1,05)	38,64 (7,71)	5,11 (0,88)	35,98 (5,97)	4,63 (0,91)	30,61 (5,43)
$\hat{\sigma}$	10,12 (1,94)	102,93 (17,23)	10,06 (1,95)	104,21 (17,86)	9,68 (2,00)	105,49 (18,13)	9,71 (2,06)	105,05 (17,75)
$\hat{\sigma}_\omega$								

Tabela 20: Médias a posteriori (e desvio padrão) - Exemplo 1 - Bivariado.

Exemplo	GGE (Clayton)		TN (Clayton)		SN		ST	
	$k = 1$	$k = 2$	$k = 1$	$k = 2$	$k = 1$	$k = 2$	$k = 1$	$k = 2$
1	35, 80 (3, 85)	252, 30 (30, 86)	37, 41 (5, 29)	282, 76 (48, 27)	32, 81 (4, 10)	251, 49 (32, 48)	33, 85 (4, 58)	254, 31 (31, 13)
$\hat{\beta}_{k0}$	-0, 96 (1, 34)	-9, 35 (9, 29)	-1, 29 (1, 43)	-5, 98 (9, 86)	-1, 43 (1, 51)	-3, 32 (10, 45)	-1, 68 (1, 34)	-9, 85 (9, 52)
$\hat{\beta}_{k1}$	0, 56 (1, 41)	4, 99 (8, 97)	0, 88 (1, 49)	10, 24 (10, 27)	0, 97 (1, 49)	12, 82 (10, 65)	0, 26 (1, 38)	7, 40 (9, 24)
$\hat{\beta}_{k2}$	-7, 80 (4, 24)	-47, 70 (40, 96)	-7, 90 (4, 35)	-46, 60 (43, 18)	-9, 01 (4, 02)	-56, 96 (41, 74)	-9, 27 (4, 44)	-55, 52 (45, 98)
$\hat{\beta}_{k3}$	-0, 82 (0, 56)	-1, 28 (0, 57)	-	-	-	-	-	-
$\hat{q}_k$	-	-	0, 95 (0, 97)	0, 50 (0, 65)	-	-	-	-
$\hat{\lambda}_k$	-	-	-	-	-	-	-	-
$\hat{\delta}_k$	-	-	-	-	4, 17 (3, 78)	-0, 04 (3, 16)	3, 27 (4, 25)	-0, 06 (3, 15)
$\hat{\sigma}_{kk}$	4, 11 (0, 90)	23, 85 (6, 71)	5, 04 (0, 88)	37, 69 (7, 10)	4, 90 (0, 80)	34, 10 (5, 30)	3, 70 (0, 80)	25, 50 (5, 20)
$\hat{\theta}$	3, 14 (0, 51)	3, 24 (0, 56)	3, 24 (0, 56)	0, 70 (0, 12)	0, 70 (0, 12)	0, 61 (0, 17)	0, 61 (0, 17)	0, 61 (0, 17)
$\hat{\sigma}_{\omega_{kk}}^2$	9, 65 (1, 79)	98, 41 (16, 98)	9, 83 (1, 71)	100, 58 (16, 87)	8, 80 (1, 70)	104, 0 (17, 30)	8, 70 (1, 80)	100, 70 (17, 40)
ρ_{ω}	0, 95 (0, 05)	0, 95 (0, 05)	0, 95 (0, 05)	0, 95 (0, 05)	0, 99 (0, 01)	0, 99 (0, 01)	0, 99 (0, 01)	0, 99 (0, 01)

Tabela 21: Médias a posteriori (e desvio padrão) - Exemplo 2 - Univariado.

Exemplo	GGE			TN			SN			ST		
	$C_{\max}$	ASC_{0-t}	$C_{\max}$	ASC_{0-t}	$C_{\max}$	ASC_{0-t}	$C_{\max}$	ASC_{0-t}	$C_{\max}$	ASC_{0-t}	$C_{\max}$	ASC_{0-t}
2	16, 45 (0, 78)	348, 43 (15, 44)	16, 59 (1, 02)	352, 81 (23, 84)	16, 67 (1, 29)	351, 70 (18, 11)	17, 06 (1, 29)	350, 73 (18, 12)				
$\hat{\beta}_0$	-3, 34 (0, 22)	-54, 68 (4, 92)	-3, 36 (0, 26)	-61, 16 (23, 84)	-3, 38 (0, 27)	-61, 76 (6, 33)	-3, 33 (0, 25)	-59, 34 (5, 98)				
$\hat{\beta}_1$	-0, 14 (0, 24)	0, 93 (4, 64)	0, 03 (0, 26)	3, 53 (5, 92)	0, 05 (0, 26)	3, 55 (6, 59)	-0, 02 (0, 25)	3, 35 (5, 99)				
$\hat{\beta}_2$	-0, 86 (1, 06)	-21, 83 (21, 65)	-0, 93 (1, 10)	-24, 78 (24, 18)	-0, 90 (1, 11)	-24, 66 (24, 31)	-0, 79 (1, 01)	-24, 55 (25, 37)				
$\hat{\beta}_3$	1, 27 (0, 63)	1, 65 (0, 61)	-	-	-	-	-	-				
$\hat{q}$	-	-	1, 30 (1, 16)	1, 57 (1, 37)	-	-	-	-				
$\hat{\lambda}$	-	-	-	-	-0, 27 (1, 32)	-0, 01 (3, 13)	-0, 74 (1, 40)	-0, 04 (3, 22)				
$\hat{\delta}$	0, 59 (0, 18)	10, 94 (3, 68)	0, 88 (0, 16)	20, 81 (3, 73)	0, 85 (0, 14)	20, 27 (3, 37)	0, 75 (0, 15)	17, 53 (3, 45)				
$\hat{\sigma}$	2, 33 (0, 43)	49, 80 (8, 90)	2, 46 (0, 44)	55, 77 (10, 16)	2, 22 (0, 65)	55, 72 (9, 58)	2, 02 (0, 76)	55, 72 (10, 17)				
$\hat{\sigma}_w$												

Tabela 22: Médias a posteriori (e desvio padrão) - Exemplo 2 - Bivariado.

Exemplo	GGE (Gaussiana)		TN (Gaussiana)		SN		ST	
	$k = 1$	$k = 2$	$k = 1$	$k = 2$	$k = 1$	$k = 2$	$k = 1$	$k = 2$
2	16, 17 (1, 23)	346, 95 (18, 00)	16, 82 (1, 11)	354, 47 (23, 07)	15, 84 (1, 18)	351, 91 (18, 45)	15, 88 (0, 93)	349, 90 (17, 16)
$\hat{\beta}_{k0}$	-3, 24 (0, 24)	-57, 29 (5, 10)	-3, 39 (0, 25)	-61, 05 (5, 80)	-3, 39 (0, 25)	-61, 94 (5, 69)	-3, 28 (0, 24)	-56, 16 (4, 51)
$\hat{\beta}_{k1}$	-0, 09 (0, 24)	2, 03 (4, 56)	0, 03 (0, 24)	3, 59 (5, 76)	0, 05 (0, 28)	3, 43 (5, 94)	-0, 05 (0, 25)	1, 56 (4, 41)
$\hat{\beta}_{k2}$	-0, 81 (1, 00)	-20, 00 (20, 86)	-0, 84 (1, 08)	-22, 74 (24, 41)	-0, 94 (1, 14)	-24, 57 (25, 93)	-0, 86 (0, 99)	-21, 50 (23, 43)
$\hat{\beta}_{k3}$	0, 97 (0, 61)	1, 43 (0, 66)	-	-	-	-	-	-
$\hat{q}_k$	-	-	0, 90 (1, 08)	1, 27 (1, 27)	-	-	-	-
$\hat{\lambda}_k$	-	-	-	-	0, 83 (1, 07)	-0, 12 (3, 18)	0, 82 (0, 83)	0, 03 (3, 38)
$\hat{\delta}_k$	0, 66 (0, 20)	12, 21 (4, 30)	0, 86 (0, 15)	20, 05 (3, 40)	0, 90 (0, 10)	19, 40 (3, 20)	0, 70 (0, 10)	12, 70 (3, 00)
$\hat{\sigma}_{kk}$	0, 92 (0, 02)	0, 93 (0, 02)	-	-	0, 64 (0, 13)	0, 53 (0, 18)	-	-
$\hat{\theta}$	2, 26 (0, 38)	49, 34 (8, 52)	2, 38 (0, 40)	53, 68 (8, 72)	2, 20 (0, 40)	54, 20 (9, 50)	2, 10 (0, 40)	53, 30 (9, 40)
$\hat{\sigma}_{\omega_{kk}}^2$	0, 92 (0, 06)	0, 96 (0, 04)	0, 96 (0, 04)	0, 99 (0, 01)	0, 99 (0, 01)	0, 99 (0, 01)	0, 99 (0, 01)	0, 99 (0, 01)
ρ_{ω}	-	-	-	-	-	-	-	-

Tabela 23: Médias a posteriori (e desvio padrão) - Exemplo 3 - Univariado.

Exemplo	GGE			TN			SN			ST		
	$C_{\max}$	ASC_{0-t}	$C_{\max}$	ASC_{0-t}	$C_{\max}$	ASC_{0-t}	$C_{\max}$	ASC_{0-t}	$C_{\max}$	ASC_{0-t}	$C_{\max}$	ASC_{0-t}
3	10, 93 (0, 54)	195, 78 (11, 03)	11, 19 (0, 90)	204, 62 (13, 19)	11, 46 (1, 02)	199, 90 (8, 33)	11, 37 (1, 02)	199, 34 (8, 42)	11, 37 (1, 02)	199, 34 (8, 42)	11, 37 (1, 02)	199, 34 (8, 42)
$\hat{\beta}_0$	-0, 32 (0, 30)	-4, 02 (3, 71)	-0, 25 (0, 31)	-4, 00 (3, 89)	-0, 22 (0, 29)	-3, 84 (4, 08)	-0, 30 (0, 31)	-3, 72 (3, 96)	-0, 30 (0, 31)	-3, 72 (3, 96)	-0, 30 (0, 31)	-3, 72 (3, 96)
$\hat{\beta}_1$	-0, 01 (0, 28)	1, 76 (3, 96)	0, 04 (0, 30)	2, 35 (3, 86)	0, 04 (0, 31)	1, 98 (4, 04)	0, 03 (0, 30)	2, 30 (4, 00)	0, 03 (0, 30)	2, 30 (4, 00)	0, 03 (0, 30)	2, 30 (4, 00)
$\hat{\beta}_2$	0, 81 (0, 64)	5, 70 (10, 82)	0, 84 (0, 67)	2, 53 (10, 95)	0, 85 (0, 66)	2, 70 (11, 46)	0, 87 (0, 65)	3, 00 (11, 10)	0, 87 (0, 65)	3, 00 (11, 10)	0, 87 (0, 65)	3, 00 (11, 10)
$\hat{\beta}_3$	0, 91 (0, 59)	0, 71 (0, 46)	-	-	-	-	-	-	-	-	-	-
$\hat{q}$	-	-	1, 15 (1, 07)	0, 92 (0, 86)	-	-	-	-	-	-	-	-
$\hat{\lambda}$	-	-	-	-	-0, 67 (1, 14)	0, 04 (0, 98)	-0, 51 (1, 12)	0, 04 (3, 18)	-0, 51 (1, 12)	0, 04 (3, 18)	-0, 51 (1, 12)	0, 04 (3, 18)
$\hat{\delta}$	0, 85 (0, 23)	11, 95 (2, 65)	1, 06 (0, 19)	13, 94 (2, 49)	1, 03 (0, 18)	13, 52 (2, 20)	0, 94 (0, 17)	12, 25 (2, 14)	0, 94 (0, 17)	12, 25 (2, 14)	0, 94 (0, 17)	12, 25 (2, 14)
$\hat{\sigma}$	1, 42 (0, 33)	25, 28 (4, 66)	1, 48 (0, 30)	25, 34 (4, 61)	1, 09 (0, 50)	25, 23 (4, 56)	1, 10 (0, 49)	25, 25 (4, 77)	1, 10 (0, 49)	25, 25 (4, 77)	1, 10 (0, 49)	25, 25 (4, 77)
$\hat{\sigma}_\omega$												

Tabela 24: Médias a posteriori (e desvio padrão) - Exemplo 3 - Bivariado.

Exemplo	GGE (Gaussiana)		TN (Gaussiana)		SN		ST	
	$k = 1$	$k = 2$	$k = 1$	$k = 2$	$k = 1$	$k = 2$	$k = 1$	$k = 2$
3	10, 94 (0, 72)	197, 51 (10, 70)	10, 99 (0, 75)	202, 12 (11, 94)	10, 88 (0, 97)	199, 90 (8, 21)	10, 91 (0, 84)	200, 15 (7, 72)
$\hat{\beta}_{k0}$	-0, 29 (0, 27)	-4, 25 (3, 71)	-0, 22 (0, 29)	-3, 77 (3, 97)	-0, 23 (0, 28)	-3, 79 (3, 85)	-0, 35 (0, 31)	-4, 23 (4, 54)
$\hat{\beta}_{k1}$	-0, 10 (0, 28)	1, 63 (3, 73)	0, 05 (0, 30)	2, 36 (3, 96)	0, 05 (0, 29)	2, 18 (3, 83)	-0, 04 (0, 31)	2, 59 (4, 45)
$\hat{\beta}_{k2}$	0, 74 (0, 67)	3, 86 (11, 20)	0, 84 (0, 61)	2, 44 (10, 19)	0, 80 (0, 71)	2, 26 (11, 07)	0, 87 (0, 64)	2, 72 (10, 30)
$\hat{\beta}_{k3}$	0, 76 (0, 54)	0, 70 (0, 51)	—	—	—	—	—	—
$\hat{q}_k$	—	—	1, 35 (1, 15)	1, 17 (1, 03)	—	—	—	—
$\hat{\lambda}_k$	—	—	—	—	0, 08 (1, 06)	-0, 09 (3, 28)	0, 18 (0, 95)	-0, 02 (1, 00)
$\hat{\delta}_k$	0, 85 (0, 20)	11, 63 (2, 61)	1, 01 (0, 16)	13, 69 (2, 24)	1, 00 (0, 10)	13, 10 (2, 10)	0, 90 (0, 20)	14, 30 (5, 60)
$\hat{\sigma}_{kk}$	0, 67 (0, 07)	0, 67 (0, 07)	0, 67 (0, 07)	0, 67 (0, 07)	0, 25 (0, 19)	0, 25 (0, 19)	0, 37 (0, 25)	0, 37 (0, 25)
$\hat{\theta}$	1, 46 (0, 27)	25, 18 (4, 35)	1, 45 (0, 27)	24, 60 (4, 31)	1, 30 (0, 30)	24, 30 (4, 50)	1, 00 (0, 50)	16, 30 (11, 80)
$\hat{\sigma}_{\omega_{kk}}^2$	0, 88 (0, 10)	0, 88 (0, 10)	0, 87 (0, 11)	0, 87 (0, 11)	0, 98 (0, 03)	0, 98 (0, 03)	0, 80 (0, 45)	0, 80 (0, 45)
ρ_ω	—	—	—	—	—	—	—	—

C Artigos submetidos para publicação

Use of asymmetric distributions in average bioequivalence

Journal:	<i>Statistics in Medicine</i>
Manuscript ID:	SIM-14-0571.R1
Wiley - Manuscript type:	Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Souza, Roberto; Federal Technological University of Paraná, Coordination of Mathematics; University of São Paulo, Department of Social Medicine Achcar, Jorge; University of São Paulo, Department of Social Medicine Martinez, Edson; University of São Paulo, Department of Social Medicine Mazucheli, Josmar; State University of Maringa, Department of Statistics
Keywords:	Multivariate skew-t distribution, Extended generalized gamma distribution, Copula functions, Average bioequivalence, Bayesian inference

Use of asymmetric distributions in average bioequivalence

Roberto Molina de Souza^{*1,2}, Jorge Alberto Achcar¹, Edson Zangiacomi Martinez¹, and Josmar Mazucheli³

¹*University of São Paulo, Department of Social Medicine*

²*Federal Technological University of Paraná, Department of Mathematics*

³*State University of Maringá, Department of Statistics*

Abstract

Generic drugs are marketed by many countries in the world. Most of these countries approve the marketing of a generic drug when there is bioequivalence evidence between this generic drug with a reference one. Usually, the responsibility for the application of bioequivalence tests is of the pharmaceutical industry under the supervision of an official regulatory office. This procedure has its conclusions based on a statistical data analysis. Despite many regulatory agencies use standard statistical analysis based on preestablished guidelines, we could have situations where the data set clearly can not be fitted by the standard models specified in these guidelines. We propose two classes of univariate and bivariate asymmetrical models based on an extended generalized gamma distribution formulated using existing copula functions and a multivariate skew-t distribution under a Bayesian approach to analyze bioequivalence data. We also present under a classical inference approach, a study of the empirical power for hypotheses tests for univariate models showing advantages for the use of an extended generalized gamma distribution. Three bioequivalence data sets are analysed under the new approach and compared to other existing proposed by the official guidelines. To verify that the proposed asymmetrical distributions are usually better fitted by the data when compared to standard models proposed by the regulatory agencies, we have used some existing Bayesian discrimination methods, as the DIC and quantile-quantile plots. We conclude that, in general, the use of extended

^{*}Electronic address: rmolinasouza@utfpr.edu.br; Corresponding author

A new approach to evaluate average bioequivalence in the presence of asymmetric data

Journal:	<i>Biometrical Journal</i>
Manuscript ID:	Draft
Wiley - Manuscript type:	Research Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Souza, Roberto; Federal Technological University of Paraná, Academic Department of Mathematics; University of São Paulo, Social Medicine Achcar, Jorge; University of São Paulo, Social Medicine Mazucheli, Josmar; State University of Maringá, Department of Statistics
Keywords:	AVERAGE BIOEQUIVALENCE, BAYESIAN APPROACH, MARSHALL OLKIN DISTRIBUTION, SKEW NORMAL DISTRIBUTION, TILTED NORMAL DISTRIBUTION

SCHOLARONE™
Manuscripts

A new approach to evaluate average bioequivalence in the presence of asymmetric data

Roberto M. de Souza* ^{1,2}, Jorge A. Achcar ², and Josmar Mazucheli ³

¹ Avenida Alberto Carazzai, 1640, Cornélio Procopio, PR, Brazil, 86300-000 (*Academic Department of Mathematics*)

² Avenida Bandeirantes, 3900, Ribeirão Preto, SP, Brazil, 14049-900 (*Department of Social Medicine*)

³ Avenida Colombo, 5790, Maringá, PR, Brazil, 87020-900 (*Department of Statistics*)

Received zzz, revised zzz, accepted zzz

A common approach to evaluate a generic drug interchangeability to a reference drug is the use of a bioequivalence test between both formulations. Despite many regulatory agencies use standard statistical analysis based on preestablished guidelines, we could have situations where the data set clearly can not be fitted by the standard models specified in these guidelines. In this way, we propose two classes of asymmetrical models based on Marshall-Olkin extended normal distribution, called as tilted normal, and on a skew normal distribution under a Bayesian approach to analyze bioequivalence data. To apply the proposed methodology, we made a study of empirical power of hypothesis testing and two bioequivalence data sets in two different asymmetric forms are analyzed under the new approach and compared to other existing models given by the official guidelines. We observe that the power of hypothesis testing for the tilted normal model is more consistent because the tilted normal model is the model which loses less power in presence of asymmetric data due to the great flexibility of fit of this model to bioequivalence measures. In addition, the standard model is a particular case of the proposed models. In these examples, we have used some existing Bayesian discrimination methods, as the DIC Criterion to verify that the proposed models are better fitted by the data when compared to the standard models assumed by the regulatory agencies. These results could be a good alternative in situations that the standard model for bioequivalence study does not work.

Key words: AVERAGE BIOEQUIVALENCE; BAYESIAN APPROACH; MARSHALL OLKIN DISTRIBUTION; SKEW NORMAL DISTRIBUTION; TILTED NORMAL DISTRIBUTION

Supporting Information for this article is available from the author or on the WWW under <http://dx.doi.org/10.1022/bimj.XXXXXXX> (please delete if not applicable)

1 Introduction

To evaluate the exchangeability of two or more drug formulations it is recommended the use of bioequivalence trials which has as main goal to get evidences of similarity among the pharmacokinetic parameters of these formulations (Barrett et al., 2000; Hyslop et al., 2000; Rani and Pargal, 2004)

Following a FDA published guideline (FDA, 2003), bioequivalence is defined as “absence of a significant difference in the rate and extent to which the active ingredient in pharmaceutical equivalents or pharmaceutical alternatives become available at the site of drug action when administered at the same molar dose under similar conditions in an appropriately designed study”. Bioavailability is defined by the FDA as the speed and extend of the absorption of an active principle in the form of a dosage, measured

*Corresponding author: e-mail: rmolinasouza@utfpr.edu.br, Phone: +55-43-3520-3905, Fax: +55-43-3520-3905

Robust Regression for Average Bioequivalence: A Bayesian approach assuming a multivariate Student's t-distribution

Journal:	<i>Journal of Applied Statistics</i>
Manuscript ID:	Draft
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Souza, Roberto; Federal Technological University of Paraná, Coordination of Mathematics; University of São Paulo, Social Medicine Achcar, Jorge; University of São Paulo, Mazucheli, Josmar; State University of Maringa, Martinez, Edson; Universidade de São Paulo, Medicina Social
Keywords:	robust regression, multivariate t-Student distribution, average bioequivalence, Bayesian Inference, MCMC

SCHOLARONE™
Manuscripts

To appear in the *Journal of Applied Statistics*
Vol. 00, No. 00, Month 20XX, 1–10

Robust Regression for Average Bioequivalence: A Bayesian approach assuming a multivariate Student's t-distribution

(Received 00 Month 20XX; accepted 00 Month 20XX)

Many countries have approved the marketing of a generic drug when there are bioequivalence evidences between this generic drug and a reference drug. This procedure has its conclusion based on a statistical analysis. Despite many regulatory agencies use standardized statistical analysis based on preestablished guidelines, usually we could have different data structures that could invalidate the assumptions of the statistical model assumed by the regulatory agency in the data analysis. It is important to explore alternative models in the analysis of bioequivalent data. We propose a model based on a multivariate Student's t-distribution under a Bayesian approach to analyze bioequivalent data. Two applications and some simulation studies show that this new proposed methodology could be a good alternative for the statistical phase of a bioequivalence study. We have used the Deviance Information Criterion to verify that the new model based on a multivariate Student's t-distribution is better fitted by the data, when compared to the existing standard assumption of normality for the data. In the simulation studies, we observed that more accurate inferences are obtained using the new modeling approach when compared to the obtained ranges of confidence intervals obtained using the standard approach of the regulatory agencies.

Keywords: robust regression; multivariate t-Student distribution; average bioequivalence; Bayesian Inference; MCMC

1. Introduction

In the last decades there has been a great interest of pharmaceutical industries, or the health regulatory agencies and researchers in the procedures related to bioequivalence studies [1]. These studies are used, generally, in the approval of generic drugs (or generics).

To assess the relative biodisponibility of two or more formulations [15], the plasma concentration data are used to assess key pharmacokinetic parameters such as the area under the curve (AUC_{0-t}) limited to the last time (t) and peak concentration (C_{max}). It is also possible to consider the time where the maximum concentration (t_{max}) occurs.

The data concerning the plasma concentration are usually obtained from curves, as shown in Figure 1, that relate the time and the concentration of the formulation of interest absorbed by an individual that received the formulation in study. Generally these individuals are hospitalized and after receiving a dose of the formulation in study it is obtained blood samples in pre-determined times and, from theses samples, it is established the ordered pairs (time;concentration) that enable us in the construction of the plasma concentration curves for each individual and the achievement of the pharmacokinetic parameters.

The maximum concentration and the time it occurs are the parameters obtained directly from the concentration curve constructed for each individual, as exemplified in Figure 1 and these, are respectively associated to the absorbed fraction and the absorption speed of the formulation. The AUC_{0-t} represents the fraction of formulation that enters the bloodstream and it is independent from the absorption speed. This fraction is usually estimated by trapezoidal method [7]; however some authors propose some other forms (numeric or theoretical) to obtain this measure [19, 31].

D Capítulo de livro em evento

DE SOUZA, ROBERTO MOLINA; Achcar, Jorge Alberto; Zangiacomi Martinez, Edson; Mazucheli, Josmar. Use of a Generalized Multivariate Gamma Distribution Based on Copula Functions in the Average Bioequivalence. In: Ansgar Steland; Ewaryst Rafajlowicz; Krzysztof Szajowski. (Org.). Springer Proceedings in Mathematics Statistics. 1ed. Cham: Springer International Publishing, 2015, v. 122, p. 343-350.

Chapter 38

Use of a Generalized Multivariate Gamma Distribution Based on Copula Functions in the Average Bioequivalence

Roberto Molina de Souza, Jorge Alberto Achcar,
Edson Zangiacomi Martinez, and Josmar Mazucheli

Abstract Bioequivalence studies have been generally used to compare a test formulation with a reference, in order to validate the interchangeability between them. Some pharmacokinetic (PK) parameters are compared in this type of study, typically using a model which assumes independence among PK parameters, the same variance for the different formulations, logarithmic transformation for the data and normal distribution for the residuals. We propose an alternative model based on a generalized gamma distribution, which permits the presence of positive asymmetry for the data and possible differences in the variances for the different formulations which could have more flexibility in this case. For the multivariate structure, we use a Gaussian copula function to capture the possible dependence between the PK parameters. We use Bayesian inference methods to obtain the results of interest. We also introduce a real data example from where we observe a good fit of the proposed model for the dataset. From this study, we conclude that the proposed model could be a good alternative in some applications where the distribution of the bioequivalence data presents a positive asymmetric distribution.

R.M. de Souza (✉) · J.A. Achcar · E. Zangiacomi Martinez
Faculty of Medicine of Ribeirão Preto, Department of Social Medicine, University of São Paulo,
3900, Bandeirantes Avenue, Ribeirão Preto, SP 14040-900, Brazil
e-mail: rmolinasouza@utfpr.edu.br

J.A. Achcar
e-mail: achcar@fmrp.usp.br

E. Zangiacomi Martinez
e-mail: edson@fmrp.usp.br

R.M. de Souza
Department of Mathematics, Federal Technological University of Paraná, 1640, Alberto Carazzai
Avenue, Cornélio Procópio, PR 86300-000, Brazil

J. Mazucheli
Department of Statistics, State University of Maringá, 5790, Colombo Avenue, Maringá, PR
87020-900, Brazil
e-mail: jmazucheli@uem.br