



**ELAYNE PENHA VEIGA**

**DESEMPENHO DA MEDIDA L NA SELEÇÃO DE  
MODELOS NORMAIS**

**LAVRAS - MG  
2017**

**ELAYNE PENHA VEIGA**

**DESEMPENHO DA MEDIDA L NA SELEÇÃO DE MODELOS NORMAIS**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Orientador

Dr. Mario Javier Ferrua Vivanco

Coorientador

Dr. Fortunato Silva de Menezes

Coorientador

Dr. Júlio Sílvio de Sousa Bueno Filho

**LAVRAS - MG**

**2017**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca  
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Veiga, Elayne Penha.

Desempenho da Medida L na seleção de modelos normais /  
Elayne Penha Veiga. – 2016.  
103 p. : il.

Orientador (a): Mario Javier Ferrua Vivanco.

Coorientador(a): Fortunato Silva de Menezes, Julio Sílvio de Sousa  
Bueno Filho. Tese – Universidade Federal de Lavras, 2016.  
Bibliografia.

1. Inferência preditiva. 2. Comparação de modelos. 3. Medida L.  
I. Vivanco, Mario Javier Ferrua. II. Menezes, Fortunato Silva de. III.  
Bueno Filho, Julio Sílvio de Sousa. IV. Título.

**ELAYNE PENHA VEIGA**

**DESEMPENHO DA MEDIDA L NA SELEÇÃO DE MODELOS NORMAIS**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 15 de setembro de 2016.

Dr. Mario Javier Ferrua Vivanco	UFLA
Dr. Fortunato Silva de Menezes	UFLA
Dr. Júlio Sílvio de Sousa Bueno Filho	UFLA
Dra. Thelma Sáfadi	UFLA
Dr. Renato Ribeiro de Lima	UFLA
Dr. Clécio da Silva Ferreira	UFJF
Dra. Carla Regina Guimarães Brighenti	UFSJ

Dr. Mario Javier Ferrua Vivanco  
Orientador  
Dr. Fortunato Silva de Menezes  
Dr. Júlio Sílvio de Sousa Bueno Filho  
Coorientadores

**LAVRAS - MG**

**2017**

*Aos meus pais, Vicente Delly Veiga Junior e Angela Maria Penha Veiga; à minha  
irmã, Renata Penha Veiga.  
**dedico.***

## AGRADECIMENTOS

Antes de qualquer coisa, agradeço a Deus, e à Maria, pela vida.

Aos meus familiares, por torcerem por mim sempre.

À Angela, minha mãe, minha rainha, que me ensinou a ter um pensamento lógico.

Ao Vicente, meu pai, meu Super Herói, por me ensinar que se deve amar o que faz.

À Renata, minha irmã, por ter paciência comigo durante o período da escrita da tese e por cuidar tão bem dos nossos cães.

Ao Allan, meu amor, por estar comigo, ao meu lado SEMPRE, e pelos gráficos no R.

À Kellen Ferreira, à Isabel de Sousa Amorim, ao Paulo Henrique Sales Guimarães, meus amigos, pela amizade inestimável e apoio. Aos ETERNOS, por serem eternos!

Ao professor Daniel Furtado Ferreira e ao Paulo César Emiliano, pela ajuda.

À Universidade Federal de Lavras (UFLA). Ao Departamento de Ciências Exatas (DEX), em especial aos professores que me acompanharam e me ajudaram tanto, e às funcionárias pela atenção e carinho.

Ao professor Mário Javier Ferrua Vivanco, meu orientador, por me mostrar que eu posso.

Ao professor Fortunato Silva de Menezes. Não tenho palavras pra agradecer o apoio.

Ao professor Júlio Silvio de S. B. Filho por me mostrar que "é muito simples".

Ao Departamento de Administração e Economia (DAE) e à UFLA Júnior Consultoria Administrativa. Ao Centro Universitário de Lavras, em especial aos meus professores do curso de Matemática.

Aos meus amigos das faculdades e meus amigos da Empresa Júnior e da comissão de formatura UFLA-Julho/2009 que se tornaram parte da minha família.

Obrigada!

“A vida, príncipe Leon, pode muito bem ser comparada a estes jogos. Na imensa multidão aqui reunida alguns vieram à procura de lucros, outros foram trazidos pelas esperanças e ambições da fama e da glória. Mas entre eles existem uns poucos que vieram para observar e entender tudo o que se passa aqui. Com a vida acontece a mesma coisa. Alguns são influenciados pela busca da riqueza, enquanto outros são dominados pela febre do poder e da dominação. Mas os melhores entre os homens se dedicam à descoberta do significado e do propósito da vida. Eles tentam descobrir os segredos da natureza. Este tipo de homem eu chamo de filósofo, pois embora nenhuma homem seja completamente sábio em todos os assuntos, ele pode amar a sabedoria como a chave para os segredos da natureza.”

(Pitágoras)

## RESUMO

Os modelos estatísticos tentam explicar fenômenos naturais ou experimentais. É comum que mais de um modelo seja formulado para um mesmo fenômeno e, dessa forma, faz-se necessária a escolha entre eles daquele que melhor o descreve. Existem muitos critérios na literatura para a comparação de modelos, como os critérios de informação de Akaike (AIC), Akaike corrigido (AICc), de informação Bayesiano (BIC), entre outros, que tentam minimizar a perda de informação envolvida no processo de modelagem. Esses critérios tem resultados assintóticos. A Medida L é uma medida para comparação de modelos que se preocupa com a predição de valores advindos de um mesmo experimento ou de experimentos semelhantes usando conceitos como densidade preditiva em sua definição e, dessa forma, comparando o que é predito ao que é observado para fazer a escolha entre modelos. No presente trabalho foram calculadas as taxas de verdadeiro positivo (TP), falso positivo (FP), falso negativo (FN) e verdadeiro negativo (TN) para a Medida L, bem como a sensibilidade em relação aos diferentes tamanhos amostrais para modelos normais. Quando foi considerada uma distribuição preditiva bastante próxima à verdadeira distribuição preditiva, os resultados das taxas TP e TN se mostraram baixos assim como os resultados para a sensibilidade. Nas outras configurações consideradas para o estudo, com distribuições preditivas diferentes e distantes da distribuição preditiva verdadeira, os resultados das taxas TP e TN foram altos assim como os resultados para sensibilidade. De maneira geral, a Medida L é mais eficaz que os critérios AIC, AICc e BIC para amostras menores que 60 quando se compara modelos normais.

**Palavras-chave:** Inferência preditiva. Comparação de modelos. Medida L.



## ABSTRACT

Statistical models attempt to explain phenomena, natural or experimental. It is common to formulate more than one model to the same phenomenon and thus it is necessary to choose that one the best describes it. There are many criteria in the literature for comparison of models such as the Akaike information criterion (AIC), corrected Akaike criterion (AIC), Bayesian information criterion (BIC), among others, that try to minimize the loss of information in the modeling process. These criteria have asymptotic results. The L-measure is a measure for comparison of models concerned with the prediction values arising from the same or similar experiments using concepts such as predictive density in its definition, and thus, by comparing what is predicted to what is observed to make choice between models. In this work were calculated the rate of true positives (TP), false positives (FP), false negatives (FN) and true negatives (TN) for L-measure, as well as sensitivity to different sample sizes, smaller than 60. When considered predictive distributions quite close to the true predictive distribution, the results of the rates of TP and TN were low as well as the results for sensitivity. In other configurations considered for the study, with different predictive distributions from true predictive distribution, the results of the rates of TP and TN were high as well as the results for sensitivity. In general, the L-measure presented best performance than the AIC criteria, AIC and BIC for samples smaller than 60.

**Keywords:** Predictive inference. Model comparison. L-measure.

## Lista de Figuras

1	Representação gráfica para a densidade preditiva <i>a posteriori</i> de $Z$ dado $Y$ . . . . .	29
2	Sensibilidade <i>versus</i> tamanho amostral: AIC, AICc, BIC. Médias iguais, variâncias iguais. Grupo 1. . . . .	46
3	Sensibilidade <i>versus</i> tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias diferentes. Grupo 1. . . . .	47
4	Sensibilidade <i>versus</i> tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias iguais. Grupo 1. . . . .	48
5	Sensibilidade <i>versus</i> tamanho amostral: AIC, AICc, BIC. Médias iguais, variâncias diferentes. Grupo 1. . . . .	48
6	Taxas de TP, FP, TN e FN considerando médias iguais e variâncias iguais. . . . .	49
7	Taxas de TP, FP, TN e FN considerando médias iguais e variâncias diferentes. . . . .	50
8	Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias iguais. . . . .	51
9	Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias diferentes. . . . .	51
10	Sensibilidade <i>versus</i> tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias diferentes. Grupo 2. . . . .	52
11	Sensibilidade <i>versus</i> tamanho amostral: AIC, AICc, BIC. Médias iguais, variâncias diferentes. Grupo 2. . . . .	53
12	Taxas de TP, FP, TN e FN considerando médias iguais e variâncias iguais. Grupo 2. . . . .	54
13	Taxas de TP, FP, TN e FN considerando médias iguais e variâncias diferentes. Grupo 2. . . . .	54
14	Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias iguais. Grupo 2. . . . .	55
15	Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias diferentes. Grupo 2. . . . .	56

16	Sensibilidade <i>versus</i> tamanho amostral. Caso 1 - Situação 1 . . . . .	60
17	Sensibilidade <i>versus</i> tamanho amostral. Caso 1 - Situação 2 . . . . .	60
18	Sensibilidade <i>versus</i> tamanho amostral. Caso 1 - Situação 3 . . . . .	61
19	Sensibilidade <i>versus</i> tamanho amostral. Caso 2 - Situação 1 . . . . .	62
20	Sensibilidade <i>versus</i> tamanho amostral. Caso 2 - Situação 2 . . . . .	62
21	Sensibilidade <i>versus</i> tamanho amostral. Caso 2 - Situação 3 . . . . .	63
22	Taxas de TP, FP, FN e TN. Caso 1 . . . . .	65
23	Taxas de TP, FP, FN e TN. Caso 2 . . . . .	65
24	Taxas de TP, FP, FN e TN. Caso 3 . . . . .	66
25	Sensibilidade <i>versus</i> tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias iguais. Grupo 1. . . . .	78
26	Sensibilidade <i>versus</i> tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias diferentes. Grupo 1. . . . .	78
27	Taxas de TP, FP, TN e FN considerando médias iguais e variâncias iguais. Grupo 1. . . . .	80
28	Taxas de TP, FP, TN e FN considerando médias iguais e variâncias diferentes. Grupo 1. . . . .	80
29	Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias iguais. Grupo 1. . . . .	81
30	Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias diferentes. Grupo 1. . . . .	81
31	Sensibilidade <i>versus</i> tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias diferentes. Grupo 2. . . . .	82
32	Taxas de TP, FP, TN e FN considerando médias iguais e variâncias iguais. Grupo 2. . . . .	84
33	Taxas de TP, FP, TN e FN considerando médias iguais e variâncias diferentes. Grupo 2. . . . .	84
34	Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias iguais. Grupo 2. . . . .	85
35	Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias diferentes. Grupo 2. . . . .	85

36	Sensibilidade <i>versus</i> tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias iguais. Grupo 1. . . . .	87
37	Sensibilidade <i>versus</i> tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias diferentes. Grupo 1. . . . .	87
38	Taxas de TP, FP, TN e FN considerando médias iguais e variâncias iguais. Grupo 1. . . . .	89
39	Taxas de TP, FP, TN e FN considerando médias iguais e variâncias diferentes. Grupo 1. . . . .	89
40	Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias iguais. Grupo 1. . . . .	90
41	Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias diferentes. Grupo 1. . . . .	90
42	Sensibilidade <i>versus</i> tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias diferentes. Grupo 2. . . . .	91
43	Taxas de TP, FP, TN e FN considerando médias iguais e variâncias iguais. Grupo 2. . . . .	93
44	Taxas de TP, FP, TN e FN considerando médias iguais e variâncias diferentes. Grupo 2. . . . .	93
45	Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias iguais. Grupo 2. . . . .	94
46	Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias diferentes. Grupo 2. . . . .	94
47	Sensibilidade <i>versus</i> tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias iguais. Grupo 1. . . . .	96
48	Sensibilidade <i>versus</i> tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias diferentes. Grupo 1. . . . .	96
49	Taxas de TP, FP, TN e FN considerando médias iguais e variâncias iguais. Grupo 1. . . . .	98
50	Taxas de TP, FP, TN e FN considerando médias iguais e variâncias diferentes. Grupo 1. . . . .	98
51	Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias iguais. Grupo 1. . . . .	99

52	Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias diferentes. Grupo 1. . . . .	99
53	Sensibilidade <i>versus</i> tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias diferentes. Grupo 2. . . . .	100
54	Taxas de TP, FP, TN e FN considerando médias iguais e variâncias iguais. Grupo 2. . . . .	102
55	Taxas de TP, FP, TN e FN considerando médias iguais e variâncias diferentes. Grupo 2. . . . .	102
56	Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias iguais. Grupo 2. . . . .	103
57	Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias diferentes. Grupo 2. . . . .	103

## Lista de Tabelas

1	Sensibilidade e especificidade . . . . .	32
2	Caso 1 - Grupo 1 . . . . .	43
3	Caso 1 - Grupo 2 . . . . .	44
4	Caso 2 - Grupo 1 . . . . .	44
5	Caso 2 - Grupo 2 . . . . .	44
6	Caso 3 - Grupo 1 . . . . .	44
7	Caso 3 - Grupo 2 . . . . .	45
8	Caso 4 - Grupo 1 . . . . .	45
9	Caso 4 - Grupo 2 . . . . .	45
10	Situações possíveis na seleção de modelos . . . . .	58
11	Caso 1 . . . . .	58
12	Caso 2 . . . . .	59
13	Caso 3 . . . . .	59
14	Valores das taxas TP, FP, FN e TN para AIC, AICc e BIC para o Caso 1, Grupo 1. . . . .	75
15	Valores das taxas TP, FP, FN e TN para AIC, AICc e BIC para o Caso 1, Grupo 2. . . . .	76
16	Valores das taxas de TP, FP, FN e TN para AIC, AICc e BIC para o Caso 2, Grupo 1. . . . .	79
17	Valores das taxas de TP, FP, FN e TN para AIC, AICc e BIC para o Caso 2, Grupo 2. . . . .	83
18	Valores das taxas de TP, FP, FN e TN para AIC, AICc e BIC para o Caso 3, Grupo 1. . . . .	88
19	Valores das taxas de TP, FP, FN e TN para AIC, AICc e BIC para o Caso 3, Grupo 2. . . . .	92
20	Valores das taxas de TP, FP, FN e TN para AIC, AICc e BIC para o Caso 4, Grupo 1. . . . .	97
21	Valores das taxas de TP, FP, FN e TN para AIC, AICc e BIC para o Caso 4, Grupo 2. . . . .	101

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>16</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>18</b>
<b>2.1</b>	<b>Critérios para comparação de modelos</b>	<b>18</b>
<b>2.2</b>	<b>Considerações sobre os critérios AIC, AICc e BIC</b>	<b>19</b>
<b>2.2.1</b>	<b>Resultados do trabalho de Emiliano, Vivanco e Menezes (2014) para modelos normais</b>	<b>20</b>
<b>2.3</b>	<b>Conceitos bayesianos</b>	<b>22</b>
<b>2.3.1</b>	<b>Teorema de Bayes</b>	<b>22</b>
<b>2.3.2</b>	<b>Distribuição <i>a priori</i> e distribuição <i>a posteriori</i></b>	<b>23</b>
<b>2.4</b>	<b>Inferência preditiva bayesiana</b>	<b>25</b>
<b>2.5</b>	<b>Densidade preditiva</b>	<b>27</b>
<b>2.6</b>	<b>Sensibilidade e especificidade</b>	<b>30</b>
<b>2.6.1</b>	<b>Taxas de verdadeiro positivo (TP) , verdadeiro negativo (TN), falso positivo (FP) e falso negativo (FN)</b>	<b>30</b>
<b>3</b>	<b>MEDIDA L</b>	<b>34</b>
<b>3.1</b>	<b>Interpretações para a Medida L</b>	<b>35</b>
<b>3.1.1</b>	<b>Como uma função perda quadrática</b>	<b>35</b>
<b>3.1.2</b>	<b>Como um erro quadrático médio</b>	<b>38</b>
<b>3.1.3</b>	<b>Como uma distância euclidiana</b>	<b>38</b>
<b>3.2</b>	<b>Medida L como medida de adequação para modelos lineares</b>	<b>39</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>43</b>
<b>4.1</b>	<b>Resultados para valores de sensibilidade para AIC, AICc e BIC para modelos normais com amostras menores que 60</b>	<b>43</b>
<b>4.1.1</b>	<b>Resultados para o Caso 1</b>	<b>46</b>
<b>4.1.1.1</b>	<b>Grupo 1</b>	<b>46</b>

4.1.1.2	Grupo 2 . . . . .	52
4.2	Simulação com dados normais para Medida L . . . . .	57
4.3	Resultados para Medida L: sensibilidade <i>versus</i> tamanhos amostrais . . . . .	59
4.3.1	Caso 1 . . . . .	59
4.3.2	Caso 2 . . . . .	61
4.3.3	Caso 3 . . . . .	63
4.3.4	Resultados das simulações para as taxas de TP (Sensibilidade), FP, FN e TN (Especificidade) . . . . .	63
5	CONCLUSÕES . . . . .	67
	REFERÊNCIAS . . . . .	68
	APÊNDICE A . . . . .	71
	APÊNDICE B . . . . .	75



## 1 INTRODUÇÃO

Na natureza, fenômenos e experimentos estudados pelos cientistas e pesquisadores não podem ser completamente conhecidos, isso ocorre devido à complexidade desses fenômenos e aos erros associados ao processo de coleta e análise de dados. A observação de um evento, de forma controlada (ou não), pode gerar padrões que ao serem modelados, auxiliam em estudos futuros ou previsões de interesse. Esses padrões são representados por variáveis e os modelos podem ser matemáticos ou estatísticos. Os modelos matemáticos (ou determinísticos) tentam explicar fenômenos objetivando representar todas ou quase todas as variáveis envolvidas. Os modelos estatísticos incluem variáveis envolvidas no processo que não são conhecidas ou observadas, mas podem ser representadas e compõem o erro da modelagem.

A modelagem estatística, como ferramenta do estudo estatístico de experimentos, auxilia no melhoramento de processos e produtos. Porém, mais de um modelo pode ser obtido para o mesmo conjunto de observações e, nesse sentido, é importante que o modelo seja válido, ou seja, mais próximo à realidade observada. Estes modelos serão comparados com o objetivo da escolha do “melhor” modelo ou “modelo ótimo” que será usado para fazer previsões/inferências.

Existem muitas medidas para a comparação de modelos e entre elas tem-se o teste de razão de verossimilhanças, os critérios de informação de Akaike (Akaike, 1974) e bayesiano (Schwarz, 1978), e o critério de validação cruzada, introduzido por Mosier (1951). Os critérios de informação de Akaike (AIC), bayesiano (BIC) e de Akaike corrigido (AICc) são bastante conhecidos e utilizados já que estão implementados em muitos softwares estatísticos. A maioria desses critérios já têm estudos sobre suas características mais importantes, como comportamento em grandes amostras, taxas de falso positivo, falso negativo e verdadeiro positivo (Emiliano, 2013). Apesar de não tão recente, mas pouco utilizada, também podemos citar a Medida L introduzida por Ibrahim e Laud (1994). Diante disso, faz-se necessário o estudo apropriado desses aspectos da Medida L.

Nesse sentido, pretende-se estudar o desempenho da Medida L na comparação de modelos normais a fim de fazer considerações sobre sua utilidade e

diferenças de abordagem em relação aos critérios AIC, BIC e AICc.

Mais especificamente, pretende-se avaliar o desempenho da Medida L na comparação de modelos lineares normais a partir das taxas de Verdadeiro Positivo, Verdadeiro Negativo, Falso Positivo e Falso Negativo e do cálculo da sensibilidade em relação a diferentes tamanhos amostrais para a classe de modelos lineares normais, mais especificamente modelos de regressão, e compará-la com os resultados dos critérios AIC, AICc e BIC para o mesmo grupo de modelos apresentados por Emiliano, Vivanco e Menezes (2014).

O trabalho está dividido em uma revisão bibliográfica geral na seção 2 e os conceitos para a Medida L na seção 3. A seção 4 apresenta os resultados da simulação para o estudo de desempenho; em 4.1 para AIC, AICc e BIC e em 4.2 são apresentadas a metodologia das simulações para a Medida L e seus resultados. A seção 5 trata das conclusões obtidas. Por fim, o Anexo A apresenta as distribuições de probabilidade utilizadas neste trabalho e o Anexo B apresenta outros resultados de desempenho dos critérios AIC, AICc e BIC para amostras menores que 60.

## 2 REFERENCIAL TEÓRICO

### 2.1 Critérios para comparação de modelos

De maneira bastante comum na experimentação, ao estudar um fenômeno, os pesquisadores objetivam criar um modelo que o explique de maneira satisfatória, ou seja, objetivam criar um modelo que o descreva bem e que permita tirar conclusões sobre experimentos futuros de mesma natureza. Existem os modelos matemáticos (ou determinísticos) e os modelos estatísticos. Os modelos matemáticos objetivam representar todas ou quase todas as variáveis envolvidas no fenômeno. Os modelos estatísticos permitem a inclusão de variáveis que não são observadas (ou conhecidas) por meio do erro da modelagem, representado por uma distribuição de probabilidade.

Ao modelarmos um fenômeno, mais de um modelo pode ser especificado para a sua representação. Dessa forma, existe a necessidade de que esses modelos sejam comparados entre si para a seleção daquele mais adequado.

Busemeyer e Wang (2000) comentam que, em geral, os cientistas concordam quanto aos objetivos da ciência no que concerne à construção de teorias gerais que são baseadas em princípios científicos confiáveis que expliquem os fenômenos de maneira rigorosa e útil para possíveis predições. Quando dois ou mais modelos são comparados, o objetivo é determinar qual “teoria geral” é “melhor”. Esse objetivo tem fundamento na hipótese assumida comumente de que existe um modelo correto ou pelo menos um modelo que seja útil para fazer inferências a partir dos dados. Ainda de acordo com os autores, cientistas geralmente distinguem os modelos entre dois tipos, quais sejam, os modelos super simplificados e os modelos extremamente complexos. Ambos os tipos tem vantagens matemáticas ou propriedades estatísticas e desvantagens como um número muito pequeno ou muito grande de parâmetros.

Veiga e Vivanco (2012) citam a modelagem estatística como uma das principais ferramentas do estudo estatístico de experimentos ou fenômenos já que auxiliam no melhoramento de processos e produtos. Comentam ainda sobre a importância da validade de um modelo no sentido de ser mais próximo à realidade

observada, já que este, sendo ótimo, em concordância com os autores Busemeyer e Wang (2000), será usado para fazer previsões ou inferências.

A seleção de modelos é uma atividade que tem se tornado cada vez mais importante nas ciências, principalmente com os avanços computacionais que permitem ajustar modelos incrivelmente complexos. Os métodos ou critérios existentes atualmente para comparação de modelos tendem, muitas vezes, a escolher entre dois tipos de modelos: os super simples ou super complexos. Assim, diante da importância da modelagem tornam-se necessários estudos para a comparação ou calibração desses critérios.

Inúmeros critérios propostos na literatura, como o Critério de Informação de Akaike (Akaike, 1973) e o Critério de Informação Bayesiano (Schwarz, 1978) têm calibrações propostas para modelos bastante complexos. Trabalhos que estudam seus comportamentos em grandes amostras, como Emiliano et al. (2009) e estudo das taxas de Falso Positivo, Verdadeiro Positivo e Falso Negativo como Emiliano, Vivanco e Menezes (2014) são encontrados na literatura. Outros critérios bastante conhecidos da literatura são o Teste de Qui-Quadrado baseado na estatística da razão de verossimilhanças (MOOD; GRAYBILLI, 1963) e o Critério de Validação Cruzada apresentado por Mosier (1951).

Menos conhecidos, ou menos utilizados, são os Critérios de Informação de Akaike Corrigido, proposto por Bozdogan (1987) e a Medida L proposta por Ibrahim e Laud (1994). Este último, ainda menos conhecido, é menos utilizado entre outros motivos, por não estar implementado em pacotes estatísticos.

## **2.2 Considerações sobre os critérios AIC, AICc e BIC**

Em Emiliano, Vivanco e Menezes (2014) são apresentados alguns resultados pertinentes ao comportamento assintótico dos critérios para comparação de modelos amplamente utilizados na experimentação e literatura acadêmica atual, a saber: o critério de Akaike (AIC) (Akaike, 1974), o critério de Akaike Corrigido (AICc) (Sugiura, 1978) e o critério bayesiano (BIC) (Schwarz, 1978). A literatura detalhada sobre a formação desses critérios é apresentada em Emiliano (2009) e Emiliano (2013).

**Definição 1.** O Critério de Informação de Akaike, AIC, é definido por:

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

em que  $\log L(\hat{\theta})$  é a função suporte maximizada e  $p$  é o número de parâmetros do modelo.

**Definição 2.** O Critério de Informação de Akaike Corrigido, AICc, é uma variante de segunda ordem do AIC. Sua utilização é indicada quando  $\frac{n}{p} \geq 40$ , e é definido por:

$$AICc = -2 \log L(\hat{\theta}) + 2p \left( \frac{n}{n-p-1} \right)$$

**Definição 3.** Seja  $F(x_n|\hat{\theta})$  um modelo estatístico estimado através do método da máxima verossimilhança. Então o Critério de Informação Bayesiano, BIC, é dado por:

$$BIC = -2 \log[f(x_n)|\hat{\theta}] + p \log(n)$$

em que  $f(x_n|\hat{\theta})$  é o modelo selecionado para o cálculo,  $p$  é o número de parâmetros a serem estimados e  $n$  é o número de observações da amostra.

Em Emiliano, Vivanco e Menezes (2014), as performances desses critérios foram avaliadas por meio de simulações Monte Carlo para diferentes tamanhos amostrais. Os autores apresentam os resultados da taxa de verdadeiro positivo *versus* tamanhos amostrais diferentes e os histogramas das taxas de verdadeiro positivo, falso positivo e falso negativo para cada um dos critérios em diferentes situações e diferentes classes de modelos, a saber: modelos normais, séries temporais e modelos de crescimento.

### 2.2.1 Resultados do trabalho de Emiliano, Vivanco e Menezes (2014) para modelos normais

Nesta seção é apresentado um resumo dos resultados obtidos por Emiliano, Vivanco e Menezes (2014) para Modelos Normais. Foram simulados dois diferentes conjuntos de dados normais em cinco situações, a saber:

- Médias e variâncias iguais

- Médias iguais,  $\mu_1 = \mu_2 = 10$ , e variâncias diferentes porém próximas,  $\sigma_1^2 = 1$  e  $\sigma_2^2 = 0,81$ .
- Médias iguais,  $\mu_1 = \mu_2 = 10$ , e variâncias diferentes porém bem distintas,  $\sigma_1^2 = 1$  e  $\sigma_2^2 = 0,25$ .
- Médias diferentes,  $\mu_1 = 10$  e  $\mu_2 = 10,5$ , porém bem próximas e variâncias diferentes porém próximas,  $\sigma_1^2 = 1$  e  $\sigma_2^2 = 0,81$ .
- Médias diferentes,  $\mu_1 = 10$  e  $\mu_2 = 10,5$ , porém bem próximas e variâncias diferentes porém bem distintas,  $\sigma_1^2 = 1$  e  $\sigma_2^2 = 0,25$ .

Para cada um dos casos acima, ajustou-se aos dados simulados diferentes distribuições normais, a saber: médias iguais e variâncias iguais (MIVI), médias iguais e variâncias diferentes (MIVD), médias diferentes e variâncias iguais (MDVI) e, médias diferentes e variâncias diferentes (MDVD). Para cada configuração ajustada calculou-se cada um dos critérios. Repetiu-se o procedimento 100 vezes e contou-se quantas vezes cada um dos critérios selecionou a distribuição que foi ajustada corretamente aos dados.

Quando foram simulados dados de modelos normais com médias diferentes porém próximas,  $\mu_1 = 10$  e  $\mu_2 = 10,5$ , e variâncias diferentes, bem distintas,  $\sigma_1^2 = 1$  e  $\sigma_2^2 = 0,25$ , o critério bayesiano, BIC, mostrou-se superior ao AIC e AICc já que apresentou taxas de Verdadeiro Positivo superiores às taxas para estas medidas.

Quando foram simulados modelos normais com médias diferentes, porém próximas,  $\mu_1 = 10$  e  $\mu_2 = 10,5$ , e variâncias diferentes, porém próximas,  $\sigma_1^2 = 1$  e  $\sigma_2^2 = 0,81$ , AIC e AICc tiveram um desempenho similar: as taxas de Verdadeiro Positivo mantiveram-se estáveis desde o início para casos de variâncias iguais e quando eram diferentes, as taxas se estabilizaram somente com tamanhos amostrais superiores a 1200.

O critério bayesiano, BIC, apresentou desempenho superior, pois as taxas de Verdadeiro Positivo foram superiores às outras medidas quando as variâncias eram iguais. Quando as variâncias eram diferentes, BIC não teve um bom desempenho, selecionando os modelos como oriundos de populações com variâncias diferentes somente com amostras superiores a 1200.

Portanto, pode-se concluir que nenhum dos critérios estudados por Emiliano, Vivanco e Menezes (2014) foi eficaz ao selecionar os modelos com de fato variâncias distintas quando estas, na verdade, eram distintas porém muito próximas. Vale ressaltar que mesmo assim, para amostras superiores a 1200, o critério bayesiano é mais eficaz que o critério de Akaike e Akaike corrigido.

### 2.3 Conceitos bayesianos

Antes da fundamentação teórica para a Medida L, alguns conceitos se fazem necessários, como os conceitos relacionados à teoria bayesiana. A Estatística bayesiana está baseada na ideia da probabilidade condicional, traduzida pelo Teorema de Bayes e na ideia de probabilidade como grau de credibilidade. O conhecimento prévio sobre o evento de interesse pode ser traduzido por meio de distribuições de probabilidade chamadas de distribuições *a priori*. O Teorema de Bayes atualiza o conhecimento prévio ao calcular a distribuição *a posteriori*.

#### 2.3.1 Teorema de Bayes

O Teorema de Bayes é o teorema postulado por Thomas Bayes sobre probabilidades condicionadas, conhecido como Regra de Bayes. Pode ser também interpretado como um processo de aprendizagem por meio dos dados, já que há a atualização do conhecimento prévio, traduzido pela distribuição *a priori*,  $\pi(\theta)$ , por meio da distribuição *a posteriori*  $\pi(\theta|x)$ .

De acordo com Veiga (2012) e Bernardo e Smith (2000), o teorema quando exposto por Bayes em 1773 considerava os dados com distribuição Binomial e os parâmetros tinham distribuição Uniforme. Porém, desde então, o teorema tem sido generalizado para outras distribuições de dados e prioris e ainda de acordo com os mesmos autores foi Laplace que generalizou o teorema para o caso discreto.

Considerando o parâmetro desconhecido  $\theta$  e a observação de  $n$  dados  $X = x_1, x_2, \dots, x_n$ , a função de densidade dos dados é representada por  $f(x|\theta)$  e a função de verossimilhança por  $L(\theta; x) = \prod_{i=1}^n f(x_i|\theta)$ .

O teorema de Bayes pode ser dado por:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}, \quad (1)$$

em que a integral  $\int f(x|\theta)\pi(\theta)d\theta$  é chamada de densidade marginal ou preditiva *a priori* de  $X$ .

Já que  $k^{-1} = \int f(x|\theta)\pi(\theta)d\theta$  não depende de  $\theta$  e funciona como uma constante normalizadora, a equação (1) pode ser reescrita a menos da constante normalizadora:

$$\pi(\theta|x) \propto L(\theta; x)\pi(\theta). \quad (2)$$

### 2.3.2 Distribuição *a priori* e distribuição *a posteriori*

Na estatística clássica o parâmetro,  $\theta$  ou o vetor de parâmetros  $\theta$  são desconhecidos e fixos. Do ponto de vista Bayesiano, é um escalar ou vetor aleatório. Nesse sentido, Paulino, Turkman e Murteira (2003) comentam que esse parâmetro ou vetor aleatório é incerto e que esta incerteza deve ser quantificada em termos de probabilidade (VEIGA; VIVANCO, 2012).

A distribuição *a priori*,  $\pi(\theta)$ , pode ser entendida como um conhecimento prévio sobre o parâmetro antes da observação dos dados e, de acordo com Box e Tiao (1973), tem um papel importante na análise bayesiana.

Gelman et al. (2003) consideram duas interpretações para as distribuições *a priori*: a interpretação *populacional* e a interpretação do *estado de conhecimento*. A primeira considera que a distribuição *a priori* representa uma população de possíveis parâmetros, da qual  $\theta$  se origina e a segunda considera o princípio de que temos que expressar nosso conhecimento (ou incerteza) sobre o parâmetro como se este fosse uma realização aleatória da distribuição *a priori*.

As distribuições *a priori* podem se caracterizar em objetivas e subjetivas e entre elas ainda se subdividir em informativas, não informativas, próprias e impróprias. Para a escolha de prioris, algumas questões devem ser consideradas, como por exemplo a conveniência matemática e a importância da influência da priori na análise.

Veiga e Vivanco (2012) comentam que distribuições *a priori* objetivas são



aquelas em que se tenta traduzir muito pouca informação disponível sobre o parâmetro antes da observação dos dados e que seja a crença comum da maioria das pessoas. Para esse tipo de priori, como as *prioris* em Políticas Públicas - que tentam refletir a opinião de um grande grupo de pessoas - o Princípio da Razão Insuficiente de Laplace sugere que na ausência de qualquer razão ao contrário, todos os valores do parâmetro desconhecido deveriam ser igualmente prováveis a priori. Jeffreys (1961), citado por PRESS (2003) e por Gelman et al. (2003), seguindo o mesmo princípio, concluem que quando o parâmetro desconhecido encontra-se em um intervalo finito, a distribuição Uniforme pode ser usada como distribuição a priori objetiva.

As *prioris* subjetivas tentam traduzir um conhecimento prévio - antes da observação dos dados atuais - advindo de experiências semelhantes observadas anteriormente, ou seja, um conhecimento prévio incorporado através de observações de experimentos semelhantes. É importante comentar que diferentes pesquisadores podem especificar diferentes distribuições *a priori* para um mesmo parâmetro e que, geralmente, essa tarefa torna-se complicada já que nem sempre é fácil traduzir um conhecimento em termos de distribuições de probabilidade antes de se observar os dados.

Entre as subjetivas, podemos destacar as distribuições *a priori* informativas. As distribuições *a priori* informativas são aquelas que de fato introduzem informação sobre o parâmetro desconhecido no sentido de não serem vagas, ou seja, não refletem a falta de informação probabilística do pesquisador quanto aos valores que esse pode assumir. Entre as *prioris* informativas, destacam-se as distribuições que pertencem às *famílias conjugadas naturais*. As distribuições *a priori* e *a posteriori conjugadas* caracterizam-se por pertencer à mesma classe de distribuições.

Gelman et al. (2003) definem distribuição *a priori* conjugada e comentam que tem uma vantagem prática, além da conveniência computacional, de serem interpretáveis como os dados adicionais.

**Definição 4.** Se  $\mathfrak{S}$  é uma classe de distribuições de amostragem  $\pi(x|\theta)$ , e  $\wp$  é a classe de distribuições a priori para  $\theta$ , então a classe  $\wp$  é conjugada à  $\mathfrak{S}$  se

$$\pi(\theta|x) \in \wp \forall \pi(\bullet|\theta) \in \mathfrak{S} \text{ e } \pi(\bullet) \in \wp$$

Para dados com distribuição Normal ( $X \sim N$ ), podemos citar a família conjugada Normal-Gamma para os parâmetros  $((\beta, \tau) \sim \text{Normal} - \text{Gamma})$  como será apresentado na seção (3.2).

O conhecimento prévio a respeito do(s) parâmetro(s), traduzido por meio da distribuição *a priori*, pode ser atualizado após a observação dos dados por meio da distribuição *a posteriori*.

A distribuição *a posteriori* é a atualização do conhecimento prévio sobre o parâmetro ( $\pi(\theta)$ ) feita por meio do Teorema de Bayes (equações 1 e 2), representada por  $\pi(\theta|x)$ .

## 2.4 Inferência preditiva bayesiana

Predizer novas observações de experimentos científicos tem sido o principal objetivo por séculos.

A inferência preditiva é o conjunto de metodologias para se fazer inferência. Está preocupada com os problemas preditivos, que são aqueles em que as quantidades desconhecidas de interesse são variáveis aleatórias futuras advindas do mesmo experimento em questão ou de experimentos semelhantes. Ao formularmos modelos estatísticos para descrever determinado acontecimento e/ou experimento também é interessante prever observações futuras desse fenômeno, a fim de entendê-lo e tomar decisões diante dessas previsões.

Broemeling (1985) comenta que a análise preditiva é uma metodologia com o objetivo de prever novas observações. Portanto, a abordagem preditiva é centrada na predição de variáveis aleatórias futuras e não de parâmetros. Comenta ainda que dentre as várias maneiras para prever novas observações, a abordagem bayesiana é “natural” já que a predição de novas observações está baseada na distribuição condicional do “futuro” dado o “passado”.

Trabalhos como Martini e Spezzaferrri (1984), Ibrahim e Laud (1994) e Chen e Ibrahim (2000) usam a abordagem preditiva. Martini e Spezzaferrri (1984) propõem um critério para comparação de modelos utilizando a função utilidade e a abordagem preditiva. Seu critério escolherá aquele modelo em que a função utilidade esperada é máxima, ou seja, por exemplo, ao comparar dois modelos será

escolhido aquele que apresentar maior função utilidade. Ibrahim e Laud (1994), considerando a análise de experimentos um problema de seleção de modelos, apresentam a Medida L, que será apresentada na seção 3, utilizando a abordagem preditiva bayesiana e apresenta uma breve comparação com os critérios AIC, BIC e Cp de Mallow. Chen e Ibrahim (2000) consideram a abordagem preditiva para propor uma classe de distribuições *a priori* informativas para modelos de séries temporais e também utilizam a Medida L (seção 3) para avaliar as predições.

Press (2003) comenta que a filosofia preditivista é centrada no desenvolvimento de modelos e hipóteses matemáticas e estatísticas, baseados nos dados observados, que podem ser usados para prever novos resultados sobre novas condições.

Ainda de acordo com o autor, cientistas testam suas teorias usando uma formulação matemática da teoria chamada *modelo* e predizem os valores das futuras observações a partir dele. Ao fazer essa predição, a diferença entre os valores preditos e observados formam o “erro de predição” que sob boas condições para formulação do modelo, deve ser pequeno. Isto é, erros de predição pequenos indicariam um bom ajuste do modelo aos dados observados.

“A qualidade de qualquer teoria científica é medida por quão bem a teoria prediz observações futuras” (PRESS, 2003, p. 234). Se a teoria prediz mal, não será uma teoria aceitável, mesmo que se ajuste bem aos dados observados. Existe ainda a teoria que prediz bem mas que não se ajusta bem aos dados observados. A segunda situação é preferível em relação à primeira, porém a situação ideal é uma teoria que se ajusta bem aos dados observados e prediz bem (PRESS, 2003).

De acordo com Ghosh, Delampady e Samanta (2006) um princípio geralmente aceito pelos bayesianos é que alguma validação deve ser feita sempre quando possível e por validação se entende habilidade preditiva. Neste sentido, Gelman et al. (2003) comentam que para fazer inferências a respeito de uma observação desconhecida, deve-se seguir uma lógica similar à da inferência Bayesiana no sentido de verificar o ajuste do modelo aos dados e sua plausibilidade para o propósito para o qual foi construído. Uma das maneiras de fazer essa verificação é usando o modelo para fazer predições de dados futuros, coletá-los numa situação posterior e comparar o que foi coletado com o que foi predito. Ainda de acordo

com o autor, falhas na especificação do modelo levam a problemas práticos como criar falsas inferências.

Uma justificativa para essa forma de validação do modelo é que ao gerar-mos dados por meio do modelo, estes deveriam estar tão próximos quanto possíveis daqueles que foram realmente observados. Mais especificamente, os dados devem parecer plausíveis sob a *distribuição preditiva a posteriori*,  $p(z|x)$ , que é a função de densidade de probabilidade relacionada às observações futuras ( $Z$ ), condicionadas aos dados observados ( $X$ ). O conceito dessa distribuição será dado na seção 2.5.

A distribuição preditiva *a posteriori* a ser apresentada na seção 2.5, é usada para o cálculo da Medida L que será descrita na seção 3.

## 2.5 Densidade preditiva

Para o desenvolvimento da teoria e estudos de simulação das características que são objetivo desse trabalho, faz-se necessária a definição de densidade preditiva.

A distribuição preditiva é a função de densidade de probabilidades usada no preditivismo para prever novas observações ou observações futuras. Existem dois tipos: a distribuição preditiva *a priori* e a distribuição preditiva *a posteriori*. A preditiva *a priori*, comumente chamada de distribuição marginal é a constante de normalização no teorema de Bayes dada pelo denominador de (1). A preditiva *a posteriori* é a distribuição de observações futuras condicionada aos dados já observados e considera a distribuição *a posteriori* dos parâmetros para o seu cálculo.

Before the data  $x$  are considered, the distribution of the unknown but observable  $x$  is  $f(x) = \int \pi(x, \theta) d\theta = \int \pi(\theta) \pi(x|\theta) d\theta$ . This is often called the marginal distribution of  $x$ , but a more informative name is the *prior predictive distribution*: prior because it is not conditional on a previous observation of the process, and predictive because it is the distribution for a quantity that is observable. (GELMAN et al., 2003, p. 8)

Após os dados serem observados, podemos ainda estar interessados em prever novas observações provenientes do mesmo processo, ou seja, prever  $\tilde{X}$  após observarmos  $X$ . A distribuição de  $\tilde{X}$  será chamada *distribuição preditiva a posteriori*:

$$f(\tilde{x}|x) = \int f(\tilde{x},\theta|x)d\theta = \int f(\tilde{x}|\theta,x)\pi(\theta|x)d\theta = \int f(\tilde{x}|\theta)\pi(\theta|x)d\theta,$$

em que  $\pi(\theta|x)$  é a distribuição *a posteriori*.

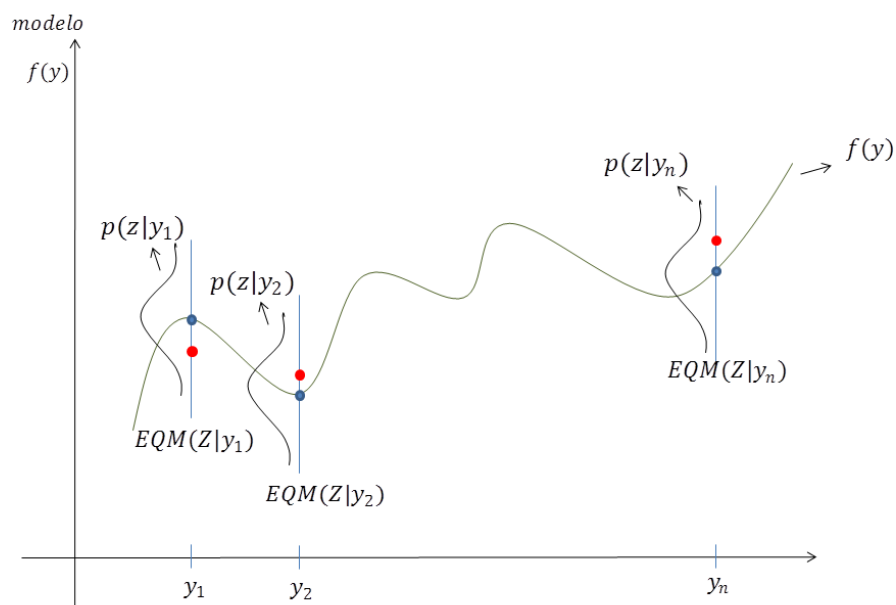
Berger (1985) introduz a metodologia preditiva considerando um exemplo simples de predição de uma variável aleatória  $Z$  com densidade  $f(z|\theta)$  tendo disponíveis dados  $Y$  com densidade  $f(y|\theta)$ . Em ambas as situações,  $\theta$  é desconhecido. As densidades de  $Z$  e  $Y$  são iguais, já que  $Z$  provém de um experimento replicado de  $Y$ . Pode-se definir a densidade pela distribuição conjunta de  $Z$  e  $\theta$  dado  $Y$ , que integrada em relação a  $\theta$  resultará na distribuição de  $Z$  dado  $Y$ , então chamada de densidade preditiva. Assim, tem-se que a densidade preditiva é dada por:

$$p(z|y) = \int f(z|\theta)\pi(\theta|y)d\theta = \int f(z|\theta)\pi(\theta)f(y|\theta)d\theta, \quad (3)$$

em que  $p(z|y)$  é chamada densidade preditiva *a posteriori* da variável  $z$  e  $\pi(\theta|y)$  é a *posteriori* de  $\theta$ . Graficamente, pode-se representar como na Figura 1. A figura em questão representa um experimento qualquer em que os pontos vermelhos são os valores observados e a curva é o modelo ajustado aos dados. Os pontos azuis representam as médias dos pontos pertencentes às curvas de densidade  $p(z|y_i)$ . Essas densidades são as densidades preditivas, ou seja, representam o conjunto de pontos que poderiam ser observados caso o experimento fosse replicado. Como será visto nas seções 3.1.2 e 3.1.3, a distância entre os pontos vermelhos e azuis pode ser interpretada, respectivamente, como um erro quadrático médio entre os pontos observados (vermelhos) e os pontos preditos (azuis) e como uma distância euclidiana respectivamente.

A densidade apresentada em (3) é a densidade utilizada por Ibrahim e Laud

Figura 1 – Representação gráfica para a densidade preditiva *a posteriori* de  $Z$  dado  $Y$ .



Fonte: Do autor (2016)

(1994) para definição da Medida L, como será apresentado na seção 3.

De acordo com Broemeling (1985), a densidade preditiva bayesiana pode ser usada para uma variedade de problemas e não somente para prever valores futuros da variável aleatória como, por exemplo, para determinar os hiperparâmetros da densidade *a priori* para os parâmetros de modelos lineares.

## 2.6 Sensibilidade e especificidade

Sensibilidade e especificidade quantificam o desempenho de um teste diagnóstico. São taxas inicialmente introduzidas pela área médica e são também conhecidas, respectivamente, como taxa de verdadeiro positivo e de verdadeiro negativo. Estão relacionadas aos falsos positivos e falsos negativos que serão conceituados na seção a seguir.

### 2.6.1 Taxas de verdadeiro positivo (TP) , verdadeiro negativo (TN), falso positivo (FP) e falso negativo (FN)

As taxas de verdadeiro positivo, ou *true positive* (TP) e verdadeiro negativo, ou *true negative* (TN) são medidas estatísticas para classificação de testes.

Mais especificamente, a sensibilidade e a especificidade estão bem definidas na área médica. Como comenta Emiliano, Vivanco e Menezes (2014), um teste de diagnóstico é uma das experiências mais presentes no dia a dia da prática médica. O teste diagnóstico objetiva verificar a presença ou ausência de uma doença em um paciente. Altman e Bland (1994) citam um teste de diagnóstico simples, que utiliza raio x ou biópsia, cujo objetivo é classificar pacientes em dois grupos de acordo com a presença ou ausência de um sintoma ou um sinal. Comentam ainda que sensibilidade e especificidade são uma forma de quantificar o desempenho do teste diagnóstico.

Swets (1988) comenta que os sistemas de diagnósticos são usados para diagnosticar pessoas com doenças, mal funcionamentos em centrais nucleares, falhas em produtos manufaturados, entre outros.

Historicamente, esses sistemas diagnósticos tiveram seu início na Segunda Guerra Mundial com o desenvolvimento da curva *ROC* (*Receiver Operators Characteristic*), que tinham o objetivo de descrever de forma quantitativa a habilidade dos operadores dos radares (*receiver operators*) em distinguir um sinal de um ruído.

Martinez, Louzada-Neto e Pereira (2003) explicitam o objetivo da curva *ROC*, qual seja, descrever de forma quantitativa o desempenho de um teste di-

agnóstico. Ainda de acordo com os autores, essa metodologia foi difundida na área biomédica nos anos 70, cujo objetivo principal era auxiliar na classificação de indivíduos entre doente e não doentes.

Sensibilidade e especificidade são independentes, já que não são calculadas sobre os mesmos indivíduos e ainda, não sofrem o efeito da prevalência da doença, por exemplo. Na Tabela 1, adaptada de Emiliano (2013) e Martinez, Louzada-Neto e Pereira (2003), têm-se as expressões matemáticas para sensibilidade e especificidade.

A taxa de verdadeiro positivo, também chamada de sensibilidade, mede a proporção de positivos que são corretamente identificados como positivos. A taxa de verdadeiro negativo, também chamada de especificidade, mede a proporção de negativos que são corretamente identificados como negativos.

No contexto de testes de hipóteses são estabelecidas as hipóteses, a saber: a hipótese nula ( $H_0$ ) e a hipótese alternativa ( $H_1$ ) e procede-se o teste de modo a rejeitar ou não rejeitar a hipótese nula. De acordo com Casella e Berger (2010), uma hipótese estatística é uma afirmação sobre um parâmetro populacional.

Ao rejeitar ou não rejeitar a hipótese nula, se está sujeito a cometer dois tipos de erro: o erro tipo I e o erro tipo II. O erro tipo I comete-se ao rejeitar a hipótese nula, sendo ela verdadeira e o erro tipo II comete-se ao não rejeitar a hipótese nula, sendo ela falsa.

O erro tipo I está associado ao falso negativo, ou seja, relaciona-se à probabilidade de caracterizar negativo algo que seja positivo. O erro tipo II está associado ao falso positivo, ou seja, relaciona-se à probabilidade de caracterizar positivo algo que seja negativo.

Para a relação entre essas medidas de desempenho e a seleção de modelos, considera-se primeiramente, um elemento (ou um conjunto de elementos) que desejamos classificar e que podem assumir valores positivos e negativos, respectivamente,  $p$  e  $n$ . Ao classificar esse elemento (ou conjuntos de elementos), 4 situações podem ocorrer:

- Situação 1: se o elemento é  $p$  e é classificado como  $p$  pelo classificador, temos um verdadeiro positivo, *true positive*.
- Situação 2: se o elemento é  $p$  e é classificado como  $n$  pelo classificador,



Tabela 1 – Sensibilidade e especificidade

Resultado do teste	Realidade do paciente	
	Positivo (doente)	Negativo (não doente)
Positivo	Verdadeiro Positivo (TP)	Falso Positivo (FP) = Erro tipo II
Negativo	Falso Negativo (FN) = Erro tipo I	Verdadeiro Negativo (TN)
Desempenho	Sensibilidade = $TP/(TP+FN)$	Especificidade = $TN/(TN+FP)$

temos um falso negativo, *false negative*.

- Situação 3: se o elemento é  $n$  e é classificado como  $n$  pelo classificador, temos um verdadeiro negativo, *true negative*.
- Situação 4: se o elemento é  $n$  e é classificado como  $p$  pelo classificador, temos um falso positivo, *false positive*.

Emiliano (2013) resume a relação entre sensibilidade e especificidade e a seleção de modelos. O autor considera a situação de selecionar, entre vários, um modelo como gerador dos dados  $Y = (y_1, \dots, y_n)$ .

Considera-se o modelo C e uma suposição: os dados se originam deste modelo C. Suponha-se ainda que a suposição será testada. Ao realizar o teste, haverá quatro situações a considerar:

- Se classificarmos os dados como oriundos do modelo C e de fato eles se originam de C, temos um verdadeiro positivo, TP.
- Se classificarmos os dados como não oriundos do modelo C e de fato eles se originam de C, temos um falso negativo, FN.
- Se classificarmos os dados como não oriundos do modelo C e de fato eles não se originam de C, temos um verdadeiro negativo, TN.
- Se classificarmos os dados como oriundos do modelo C e de fato ele não se originam de C, temos um falso positivo, FP.

Em fenômenos ou experimentos reais, o grande problema no cálculo dessas taxas é que, na maioria das vezes, não é possível saber a realidade, ou seja, dificilmente o verdadeiro modelo do qual os dados realmente se originam será conhecido.

Pode-se considerar ainda as relações em termos de testes de hipóteses:

- A *sensibilidade* está relacionada à taxa de *verdadeiro positivo* que dá o *nível de confiança do teste* ( $1 - \alpha$ ).
- A *especificidade* está relacionada à taxa de *verdadeiro negativo* que dá o *poder do teste* ( $1 - \beta$ ).
- A taxa de *falso positivo* dá a probabilidade de cometermos o *erro tipo II* ( $\beta$ ).
- A taxa de *falso negativo* dá a probabilidade de cometermos o *erro tipo I* ou *nível de significância do teste* ( $\alpha$ ).

### 3 MEDIDA L

A Medida L baseia-se no preditivismo, cuja preocupação é a predição de valores advindos de um mesmo experimento ou de experimentos semelhantes. O cálculo dessa medida utiliza o conceito de densidade preditiva em sua definição. Portanto, compara o que é predito ao que é observado para fazer a escolha entre modelos, justificada pela abordagem inferencial preditiva bayesiana apresentada na seção 2.4. Pode ser interpretada como uma função perda quadrática, como uma distância euclidiana entre pontos ou como o erro quadrático médio.

De acordo com Ibrahim e Laud (1994), considerando  $Y = (Y_1, \dots, Y_n)$  como  $n$  variáveis aleatórias com função de densidade  $f(y|\theta)$ ,  $Z = (Z_1, \dots, Z_n)$  como  $n$  observações futuras advindas do mesmo experimento gerador de  $Y$  e com mesma densidade de  $Y$ , ou seja,  $f(Z|\theta)$ , e  $f(Z|y)$  a densidade preditiva a posteriori de  $Z$  dado  $Y$  a forma geral da Medida L pode ser apresentada como:

$$L_{IL}^2 = E[(Z - y)'(Z - y)], \quad (4)$$

em que  $L_{IL}$  é a notação dada pelos autores para apresentar a forma geral. A partir desse ponto, as letras subscritas serão suprimidas para melhor entendimento.

Ibrahim e Laud (1994), Laud e Ibrahim (1995) e Ibrahim, Chen e Sinha (2001) comentam que a medida pode ser escrita como a soma de dois termos: um envolvendo a variância preditiva e outro um viés envolvendo a diferença quadrada entre a média preditiva e os dados observados, por meio de simples manipulação algébrica de (4).

A medida L é definida a partir das observações  $y_{i's}$  que podem ser totalmente observadas, como  $Y = (y_1, \dots, y_n)$  com densidade conjunta amostral dada por  $f(y|\theta)$ , em que  $\theta$  é um vetor de parâmetros e ainda considerando valores futuros de um experimento replicado imaginário,  $Z = (z_1, \dots, z_n)$  com mesma densidade amostral de  $Y$ . As médias e variâncias preditivas são calculadas a partir da densidade preditiva, dada na equação (3). Em termos:

$$L^2 = \sum_{i=1}^n [(E(Z_i) - y_i)^2 + var(Z_i)]. \quad (5)$$

Ibrahim, Chen e Sinha (2001) ainda apresentam a forma geral da medida dada por:

$$L^2 = E[(z - y)'(z - y)] + k(y - b)'(y - b), \quad (6)$$

em que a esperança também é tomada com respeito à distribuição preditiva de  $z|y$ ;  $b = (b_1, \dots, b_n)$  é um vetor de posição arbitrário e  $k$  é um escalar não negativo que pondera a discrepância entre os valores futuros em relação aos observados. Se  $\mu_i = E(Z_i|y)$ ,  $b$  é o valor que minimiza  $L(y, b, k)$ , qual seja:  $\hat{b} = (1 - \nu)\mu_i + \nu y_i$  em que  $\nu = k/(k + 1)$  ( $\nu$  é o “peso” do viés da variância). Assim, a Medida L pode ser reescrita por:

$$L^2 = \sum_i^n Var(z_i | y) + \nu \sum_i^n (\mu_i - y_i)^2. \quad (7)$$

### 3.1 Interpretações para a Medida L

A Medida L pode ser interpretada de formas diferentes dependendo da abordagem, a saber: função perda quadrática; distância Euclidiana e erro quadrático médio (VEIGA; VIVANCO, 2012).

Interpretando a Medida L como uma função perda quadrática, distância euclidiana ou erro quadrático médio, pode-se observar de uma forma mais simples que ela pode servir para comparar modelos. Isto é, quanto menor o valor da Medida L calculada para determinado modelo, mais plausível será.

#### 3.1.1 Como uma função perda quadrática

Em Estatística, a tomada de decisão envolve tomar decisões em um ambiente de incerteza e essa incerteza pode ser trabalhada probabilisticamente. Alguns tipos de decisão amplamente discutidos são, por exemplo, problemas na área agropecuária como decidir entre qual fertilizante utilizar, o lançamento ou não de um novo produto farmacêutico, ou ainda o tipo de tratamento de uma doença.

A área da estatística que se preocupa com a estrutura do processo de to-

mada de decisão é a Teoria da Decisão.

Segundo Berger (1985), a Teoria da Decisão é um conjunto de métodos para a tomada de decisão em ambientes de incerteza que permitem resultados confiáveis.

Na Teoria da Decisão Bayesiana, o parâmetro desconhecido  $\theta$  representa o estado da natureza, afeta o processo de decisão e pertence a  $\Theta$  que é o conjunto de todos os possíveis estados da natureza. Berger (1985) comenta que ao executar um experimento para obter informações a respeito de  $\theta$ , as observações são distribuídas de acordo com alguma distribuição de probabilidade com este parâmetro como desconhecido. Dessa forma,  $\Theta$  refere-se ao espaço paramétrico.

Decisões são ações e podem ser representadas por  $a$ , e o conjunto de todas as ações que podem ser tomadas podem ser chamadas de  $A$ . Se uma particular ação  $a_1$  é tomada e  $\theta_1$  é o estado da natureza, então a perda  $L(\theta_1, a_1)$  ocorrerá (BERGER, 1985).

As ações são tomadas pelo pesquisador com base nas informações sobre o verdadeiro estado da natureza e podem ser avaliadas numericamente. Na teoria da decisão, os números que quantificam as consequências das ações são chamados de *utilidades*.

Um problema de decisão pode ser resolvido utilizando uma função relacionada a essas utilidades, mais especificamente, a função utilidade. De acordo com Berger (1985), cada par  $(\theta, a)$  relacionado com a ação tomada em função do verdadeiro estado da natureza, tem uma consequência definida como “recompensa”  $r$ . Essa recompensa tem uma utilidade  $U(r)$ . Esta função pode ser representada por  $U(\theta, a)$ .

A partir da obtenção da função utilidade, a função perda pode ser obtida:

$$L(\theta, a) = -U(\theta, a),$$

sendo  $\theta$  o parâmetro desconhecido e  $a$  a ação a ser tomada,  $\theta \in \Theta$  e  $a \in A$ .

De acordo com Veiga (2012) ao tomar decisões em ambientes de incerteza, a função perda não será conhecida, então pode-se considerar a perda esperada. A definição de Perda Esperada Bayesiana é dada por:

**Definição 5.** Se  $\pi(\theta|x)$  é a distribuição de probabilidade de  $\theta$  no momento da

decisão, a Perda Esperada Bayesiana de uma ação  $a$  é

$$\rho[\pi(\theta|x), a] = E_{\pi(\theta|x)} L(\theta, a) = \int L(\theta, a) dF_{\pi(\theta|x)}(\theta) = \int L(\theta, a) \pi(\theta|x) d\theta,$$

em que a integral é calculada no espaço paramétrico  $\Theta$  e  $\pi(\theta|x)$  representa a distribuição a posteriori do parâmetro. A Perda Esperada Bayesiana é também definida como Função Risco. Na tomada de decisão ou na avaliação de uma regra de decisão, importantes funções perda padrões são utilizadas como, por exemplo, a função perda erro quadrático, definida por:

$$L(\theta, a) = (\theta - a)^2. \quad (8)$$

Veiga (2012) cita as três mais importantes razões citadas por Berger (1985) para o uso de (8) para avaliar regras de decisão:

1. É utilizada em problemas de inferência quando um estimador não viciado para o parâmetro é utilizado, já que a função risco representa a variância desse estimador. Essa variância será a esperança com relação ao parâmetro da função perda considerada, ou seja,

$$R[\theta, \delta(X)] = E_{\theta}[L(\theta, \delta(X))] = E_{\theta}[\theta - \delta(X)]^2,$$

em que  $R[\theta, \delta(X)]$  é a função risco de uma regra de decisão  $\delta(X)$  e a esperança é tomada com relação ao parâmetro  $\theta$ .

**Definição 6.** Uma regra de decisão  $\delta$  é uma função definida em  $\Omega$  que assume valores em  $A$ , ou seja:  $\delta : \Omega \rightarrow A$ .

2. Existe uma relação entre a função perda erro-quadrático e a teoria clássica de mínimos quadrados, já que se a função perda quadrática é empregada, a perda esperada reduz-se ao Erro Quadrático Médio.
3. Para a maioria dos problemas de análise de decisão, o uso da função perda erro quadrático torna os cálculos relativamente fáceis e simples. Uma generalização da função perda erro-quadrático é a função perda quadrática, que é uma extensão natural para situações multivariadas e é dada por:  $L(\theta, a) =$

$(\theta - a)'Q(\theta - a)$ , em que  $\theta = (\theta_1, \dots, \theta_p)'$  é o vetor de parâmetros a ser estimado por  $a = (a_1, \dots, a_p)'$  e  $Q$  é uma matriz positiva definida  $p \times p$ . Se  $Q$  é diagonal, então a função pode ser reescrita como  $L(\theta, a) = \sum_{i=1}^p q_i(\theta_i - a_i)^2$ .

Veiga (2012) concluiu que pela Teoria da Decisão, a medida L é a função perda quadrática. Neste sentido, ao comparar modelos, o objetivo é diminuir esta perda ao se escolher um modelo em detrimento de outro.

### 3.1.2 Como um erro quadrático médio

O erro quadrático médio é comumente utilizado para avaliar o desempenho de um estimador (BOLFARINE; SANDOVAL, 2000). Sua definição é apresentada pelos mesmos autores.

**Definição 7.** *O erro quadrático médio (EQM) de um estimador  $\hat{\theta}$  do parâmetro  $\theta$  é dado por:*

$$EQM[\hat{\theta}] = E[(\hat{\theta} - \theta)^2]$$

De acordo com Veiga (2012) o desenvolvimento algébrico da função perda quadrática resultará no cálculo do erro quadrático médio.

Já que bons modelos terão pequenos valores para a Medida L, o cálculo deste erro quadrático médio está condicionado a cada ponto observado  $y_i$ , ou seja, observados os valores iniciais  $Y$  e predizendo  $Z$ , a variabilidade entre estes dois valores deve ser bem pequena para que o modelo tenha a menor Medida L.

### 3.1.3 Como uma distância euclidiana

A distância euclidiana representa a distância entre dois pontos no espaço métrico.

A distância euclidiana entre dois pontos,  $P = (p_1, \dots, p_n)$  e  $Q = (q_1, \dots, q_n)$ , num espaço euclidiano n-dimensional é definida por:

$$DE_n = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (9)$$

Para pontos unidimensionais entre dois pontos  $P = (p_x)$  e  $Q = (q_x)$ , a distância dada por (9), pode ser reescrita:

$$DE_1 = \sqrt{(p_1 - q_1)^2}.$$

Ibrahim e Laud (1994) e Laud e Ibrahim (1995) citados por Ibrahim, Chen e Sinha (2001) definem a Medida L como a esperança do quadrado da distância euclidiana entre  $Y$  e  $Z$  como apresentado na equação (5).

### 3.2 Medida L como medida de adequação para modelos lineares

O cálculo da Medida L para modelos lineares dependerá da escolha das *prioris*. Laud e Ibrahim (1995) comentam a respeito da dificuldade de se selecionar uma *priori* informativa para um determinado modelo já que, geralmente, não é possível interpretar cada componente da matriz dos parâmetros para cada possível modelo. Alguns resultados podem ser utilizados para modelos lineares, como os apresentados por Box e Tiao (1973).

Considere, a partir de Ibrahim e Laud (1994), Laud e Ibrahim (1995), Hoeting e Ibrahim (1998) e Veiga e Vivanco (2012), o modelo usual normal de regressão linear dado por:

$$Y = X\beta + \epsilon, \quad (10)$$

em que  $Y$  é um vetor de respostas  $n$ -dimensional,  $X$  é a matriz de preditores observados,  $\beta$  é o vetor de coeficientes de regressão de dimensão  $(k + 1)$  e  $\epsilon$  é um vetor de erros aleatórios  $n$ -dimensional. O modelo (10) pode ser detalhado e apresentado por:



$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

Usualmente,  $\epsilon$  tem distribuição normal multivariada com média  $\mathbf{0}$  e matriz de precisão  $\tau I$ , em que  $\tau$  é um escalar positivo e  $I$  é uma matriz identidade com dimensão  $(n \times n)$ . Em termos:

$$\epsilon \sim N_n(\mathbf{0}, \tau I). \quad (11)$$

Considere ainda o problema de seleção de variáveis, o que permite que o modelo (10) seja reescrito da forma:

$$Y = X_m \beta^{(m)} + \epsilon, \quad (12)$$

em que  $X_m$  é a matriz de incidência sobre o modelo  $m$ , de posto completo e dimensão  $(n \times k_m)$  e  $\beta^{(m)}$  o vetor de coeficientes correspondente. O subíndice  $m$  refere-se a cada modelo.

Diante do modelo de regressão (12), as prioris são dadas em Box e Tiao (1973) e Broemeling (1985) e apresentadas por Laud e Ibrahim (1995):

- $\beta^{(m)} | \tau \sim N_n(\mu^{(m)}, \tau T)$  em que:

$$\mu^{(m)} = (X_m' X_m)^{-1} X_m' \eta_0$$

$$T = c(X_m' X_m) \text{ e } c \geq 0;$$

- $\tau \sim \text{Gama}(\delta_0/2, \gamma_0/2)$ .

O que resulta em uma priori conjugada Normal-Gama para  $(\beta, \tau)$ :

$$(\beta, \tau) \sim NG(\mu^{(m)}, \tau T_m, \delta_0/2, \gamma_0/2)$$

Usando a verossimilhança Normal para os dados ( $Y$ ), a distribuição *a posteriori* será também uma Normal-Gama. De acordo com Broemeling (1985) a distribuição *a priori* Normal-Gama é um membro de uma classe de distribuições conjugadas e as famílias conjugadas têm a vantagem de que o pesquisador tem a escala para julgar a quantidade de informação trazida pela amostra, além daquela dada pela priori.

Então, considerando a *priori* conjugada Normal-Gama para  $(\beta, \tau)$  e a *posteriori*, também Normal-Gama para  $(\beta, \tau|Y)$ , a distribuição preditiva para  $Z$  será *t*:

$$Z \sim t_n(n + \delta_0, \eta_m, s_m^2(I + (1 - \gamma)P_m)) \quad (13)$$

em que

- $\gamma = c/(1 + c)$
- $P_m = X_m(X_m'X_m)^{-1}X_m'$
- $\eta_m = P_m(\gamma\eta_0 + (1 - \gamma)y)$
- $s_m^2 = (n + \delta_0)^{-1}(q_m + \gamma p_m + \gamma_0)$
- $q_m = y'(I - P_m)y$
- $p_m = (y - \eta_0)'P_m(y - \eta_0)$

Diante dessa densidade preditiva, o cálculo da medida L para modelos lineares pode ser feita e sua forma é apresentada em Laud e Ibrahim (1995):

$$L_m = \{(1 + \lambda_m)q_m + \gamma(\gamma + \lambda_m)p_m + \lambda_m\gamma_0\}^{1/2}. \quad (14)$$

em que:

- $\lambda_m = \frac{n+(1-\gamma)k}{n+\delta_0-2}$ ;
- $p_m$ : representa a penalidade para uma priori ruim;
- $q_m$ : representa a soma de quadrados do erro para o modelo selecionado.

$L_m$  é a notação referente à situação apresentada acima. O subíndice  $m$  refere-se ao modelo especificado.

## 4 RESULTADOS E DISCUSSÃO

Nessa seção são apresentados os resultados para as simulações de modelos normais e cálculos de sensibilidade (taxa de TP) em relação à diferentes tamanhos amostrais e taxas de FP, TN e FN.

Na seção 4.1 são apresentados os resultados para AIC, AICc e BIC e na seção 4.2 para a Medida L, que é o objetivo desse trabalho.

### 4.1 Resultados para valores de sensibilidade para AIC, AICc e BIC para modelos normais com amostras menores que 60

Para efeito de comparação, foram simulados dados seguindo a mesma metodologia de Emiliano, Vivanco e Menezes (2014) apresentada na seção 2.2.1 porém com parâmetros que foram utilizados na simulação para estudo da medida L (Tabelas 11, 12 e 13 apresentadas na seção 4.2), com tamanhos amostrais entre 5 e 60 e 1000 repetições. Estes resultados não contemplam discussão já que não compreendem o objetivo deste trabalho.

Os dados simulados têm distribuição Normal. São considerados 4 casos (casos 1, 2, 3 e 4), cada um com 2 grupos; o grupo 1 é aquele que quando as variâncias são diferentes, são próximas e o grupo 2 é aquele que quando as variâncias são diferentes, são distantes.

Para melhor visualização da simulação dos dados, são apresentadas as tabelas a seguir:

Tabela 2 – Caso 1 - Grupo 1

$\mu_1$	$\mu_2$	$\sigma_1^2$	$\sigma_2^2$
10	10	1	1
10	10	1	0,81
10	12	1	1
10	12	1	0,81

Tabela 3 – Caso 1 - Grupo 2

$\mu_1$	$\mu_2$	$\sigma_1^2$	$\sigma_2^2$
10	10	1	1
10	10	1	2
10	12	1	1
10	12	1	2

Tabela 4 – Caso 2 - Grupo 1

$\mu_1$	$\mu_2$	$\sigma_1^2$	$\sigma_2^2$
10	10	1	1
10	10	1	0,81
10	25	1	1
10	25	1	0,81

Tabela 5 – Caso 2 - Grupo 2

$\mu_1$	$\mu_2$	$\sigma_1^2$	$\sigma_2^2$
10	10	1	1
10	10	1	2
10	25	1	1
10	25	1	2

Tabela 6 – Caso 3 - Grupo 1

$\mu_1$	$\mu_2$	$\sigma_1^2$	$\sigma_2^2$
10	10	1	1
10	10	1	0,81
10	40	1	1
10	40	1	0,81

Tabela 7 – Caso 3 - Grupo 2

$\mu_1$	$\mu_2$	$\sigma_1^2$	$\sigma_2^2$
10	10	1	1
10	10	1	2
10	40	1	1
10	40	1	2

Tabela 8 – Caso 4 - Grupo 1

$\mu_1$	$\mu_2$	$\sigma_1^2$	$\sigma_2^2$
10	10	1	1
10	10	1	0,81
10	100	1	1
10	100	1	0,81

Tabela 9 – Caso 4 - Grupo 2

$\mu_1$	$\mu_2$	$\sigma_1^2$	$\sigma_2^2$
10	10	1	1
10	10	1	2
10	100	1	1
10	100	1	2

A variância igual a 0,81 foi considerada como na metodologia apresentada por Emiliano, Vivanco e Menezes (2014) já que esse valor não foi contemplado na metodologia apresentada nesse trabalho para a simulação da Medida L.

Ao considerar os 4 casos acima, três configurações, a saber: médias iguais e variâncias iguais,  $\mu_1 = \mu_2 = 10$  e  $\sigma_1^2 = \sigma_2^2 = 1$ ; médias iguais e variâncias diferentes (grupo 1),  $\mu_1 = \mu_2 = 10$  e  $\sigma_1^2 = 1$  e  $\sigma_2^2 = 0.81$  e médias iguais e variâncias diferentes (grupo 2),  $\mu_1 = \mu_2 = 10$  e  $\sigma_1^2 = 1$  e  $\sigma_2^2 = 2$  pertencem a todos os casos. Portanto, os gráficos com os resultados dessas configurações para sensibilidade *versus* tamanho amostral são apresentados na seção 4.1.1 que apresenta os resultados para o caso 1 e omitidos para os outros casos.

### 4.1.1 Resultados para o Caso 1

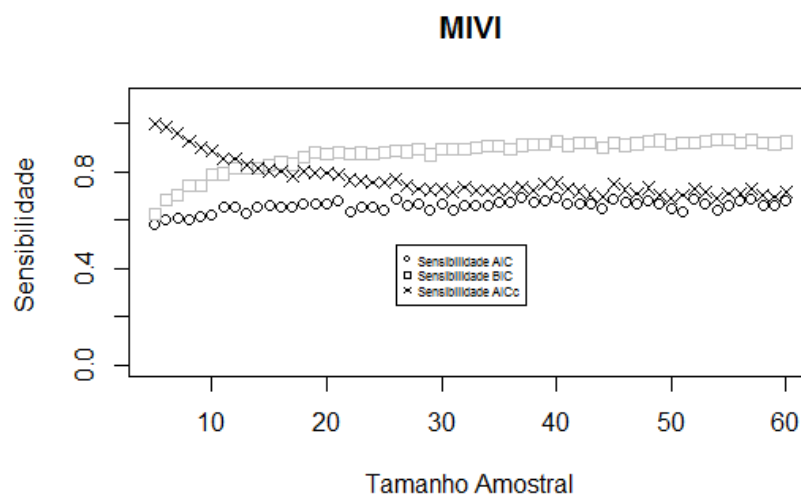
Nessa seção são apresentados os resultados considerando a metodologia de Emiliano, Vivanco e Menezes (2014), porém com parâmetros iguais aos do caso 1 apresentados nas Tabelas 2 e 3.

#### 4.1.1.1 Grupo 1

O grupo 1 é aquele em que quando as variâncias são diferentes, elas são próximas.

Quando foram simulados dados com médias iguais e variâncias iguais, os valores para sensibilidade de BIC se estabilizaram a partir do tamanho amostral igual a 20 e AICc e AIC a partir do tamanho amostral igual a 30. BIC apresentou melhores resultados com os valores em torno de 87%, seguido do AICc com média de 77% e AIC com média de 66% (Figura 2).

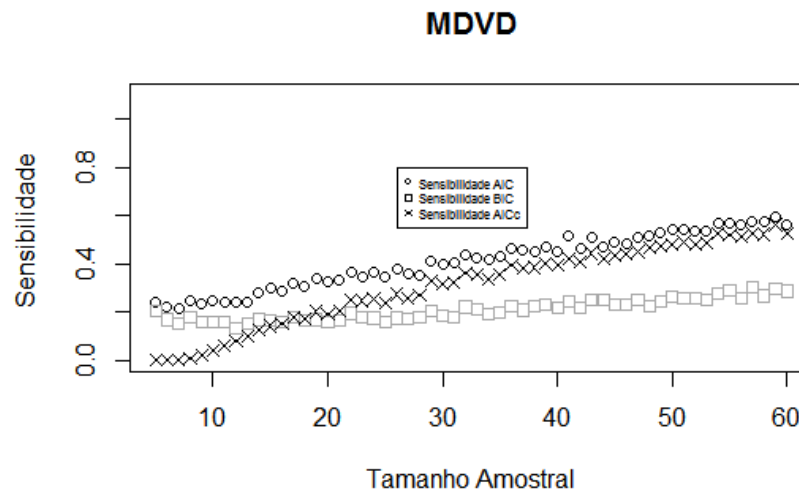
Figura 2 – Sensibilidade *versus* tamanho amostral: AIC, AICc, BIC. Médias iguais, variâncias iguais. Grupo 1.



Quando foram simulados dados com médias diferentes e variâncias di-

ferentes (Figura 3), AIC e AICc mostraram desempenho melhor que BIC. AIC e AICc apresentaram valores crescentes para sensibilidade, porém menores que 60% e BIC apresentou média de 21%. Ou seja, em 60% das vezes ou menos, AIC foi capaz de identificar os dados com médias e variâncias diferentes como realmente com médias e variâncias diferentes e em 21% das vezes BIC foi capaz de fazê-lo.

Figura 3 – Sensibilidade *versus* tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias diferentes. Grupo 1.



Quando foram simulados dados com médias diferentes e variâncias iguais, conforme Figura 4, BIC apresentou melhores resultados, seguido do AICc e AIC. A partir do tamanho amostral igual a 20, BIC apresentou média de sensibilidade de 91%, AICc apresentou média de 84% e AIC de 81%.

Ao serem simulados dados com médias iguais e variâncias diferentes, Figura 5, AICc e AIC apresentaram maiores valores de sensibilidade que BIC a partir do tamanho amostral igual a 18 porém não maiores que 58%. BIC apresentou média de 23%.



Figura 4 – Sensibilidade *versus* tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias iguais. Grupo 1.

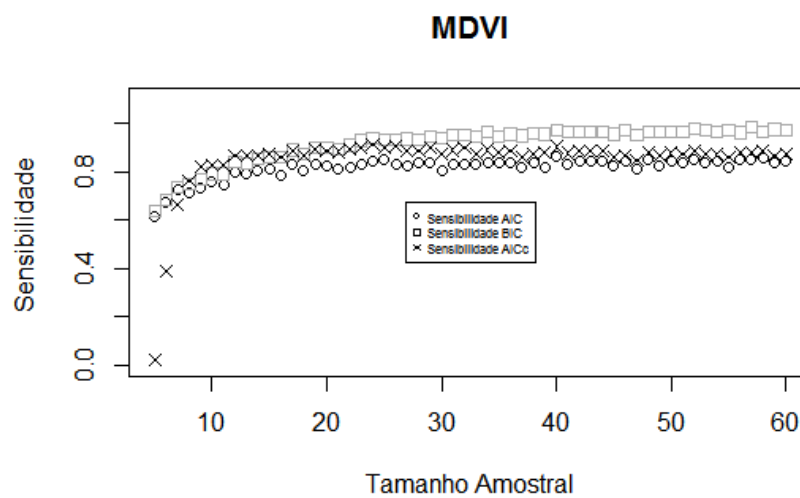
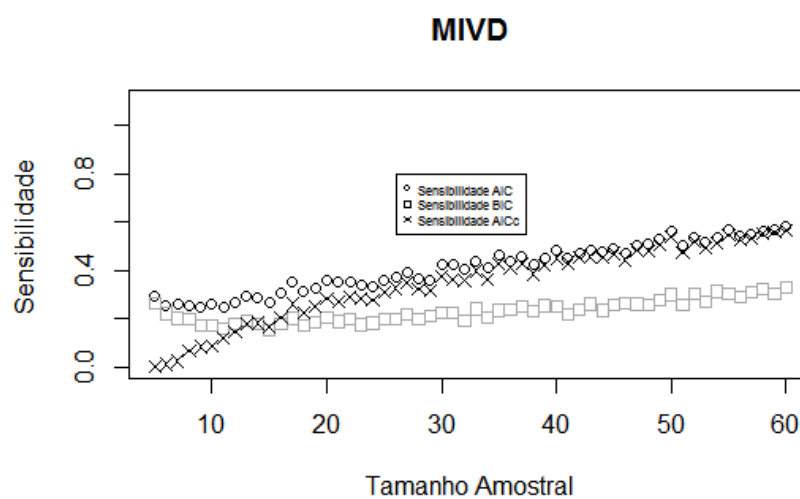


Figura 5 – Sensibilidade *versus* tamanho amostral: AIC, AICc, BIC. Médias iguais, variâncias diferentes. Grupo 1.



Em relação às taxas, quando foram simulados dados com médias iguais e variâncias iguais, AIC apresentou poder do teste (taxa de TN) maior que AICc e BIC respectivamente; porém também apresentou maior nível de significância do teste, ou taxa de FN. Apesar do BIC apresentar maior sensibilidade (taxa de TP), também apresentou maior probabilidade para o erro tipo II (taxa de FP) ( Figura 6).

Na Figura 7, quando foram simulados dados com médias iguais e variâncias diferentes, os resultados das taxas para os critérios foi bastante semelhante. BIC apresentou maior poder do teste (taxa de TN) e AIC apresentou maior sensibilidade (taxa de TP).

Figura 6 – Taxas de TP, FP, TN e FN considerando médias iguais e variâncias iguais.

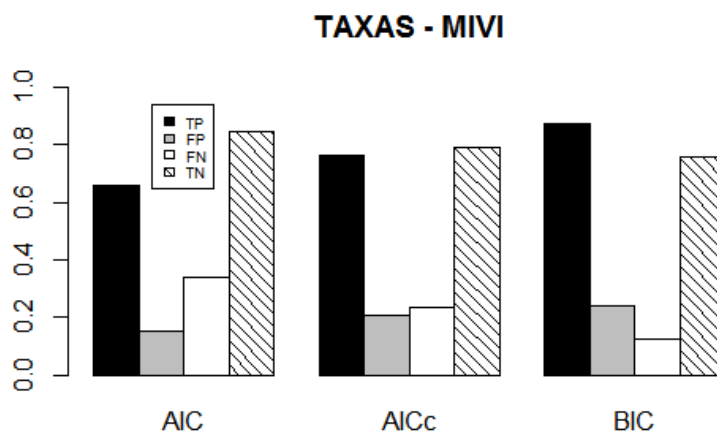
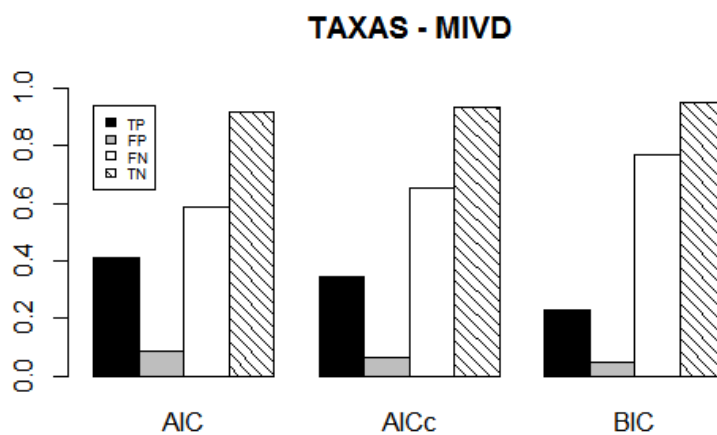


Figura 7 – Taxas de TP, FP, TN e FN considerando médias iguais e variâncias diferentes.



Ao simularmos dados com médias diferentes e variâncias iguais, ao observar a Figura 8, BIC apresentou maior sensibilidade e probabilidade do erro tipo II, taxas de TP e FP respectivamente, e AIC apresentou maior nível de significância, taxa de FN, porém os resultados de todas as taxas foram bastante semelhantes.

Quando foram simulados dados com médias diferentes e variâncias diferentes (Figura 9), BIC apresentou maior nível de significância e poder do teste (taxas de FN e TN respectivamente) que AICc e AIC. AIC apresentou maior sensibilidade (taxa de TP) e maior probabilidade do erro tipo II (taxa de FP).

A tabela (Tabela 14) com os resultados numéricos das taxas das Figuras 6, 7, 8 e 9 está no Apêndice B.

Figura 8 – Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias iguais.

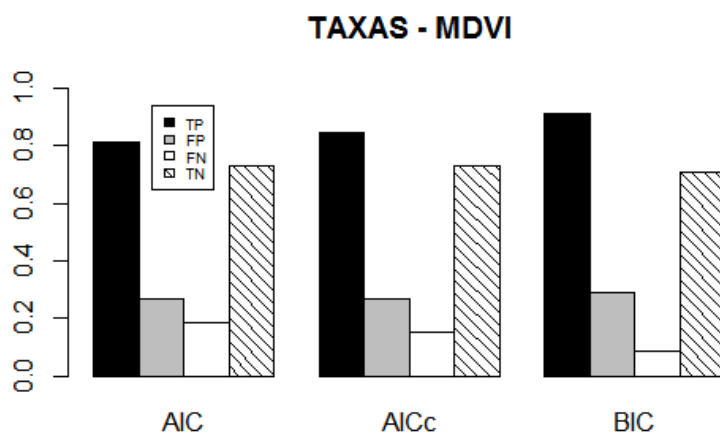
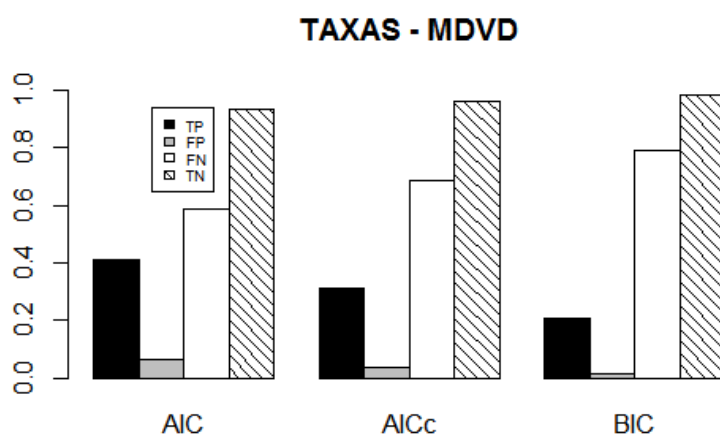


Figura 9 – Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias diferentes.



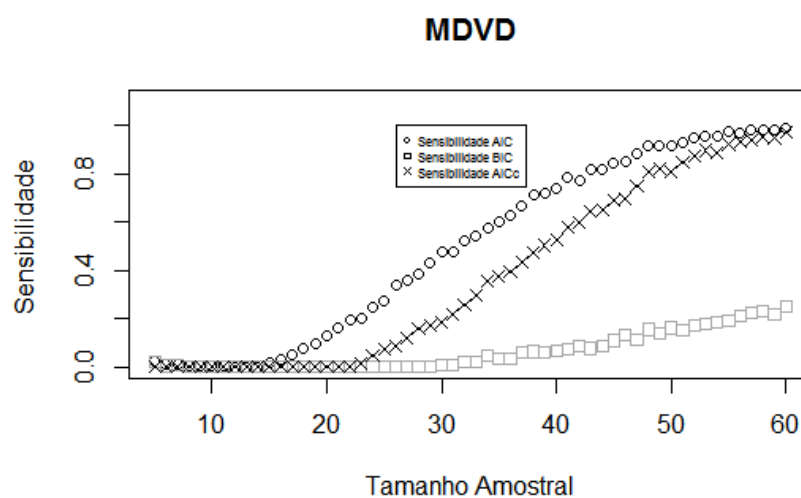
#### 4.1.1.2 Grupo 2

O grupo 2 é aquele em que quando as variâncias são diferentes, elas são distantes.

É importante notar que neste grupo, existem duas configurações para a simulação dos dados que já foram consideradas no grupo 1, a saber: médias iguais e variâncias iguais,  $\mu_1 = \mu_2 = 10$  e  $\sigma_1^2 = \sigma_2^2 = 1$ , e médias diferentes e variâncias iguais,  $\mu_1 = 10$ ;  $\mu_2 = 12$  e  $\sigma_1^2 = \sigma_2^2 = 1$ , portanto, os gráficos de sensibilidade *versus* tamanho amostral não são apresentados novamente, já que foram apresentados anteriormente nas Figuras 2 e 4 na seção anterior.

Quando foram simulados dados com médias diferentes e variâncias diferentes, como observado na Figura 10, AIC apresenta melhores resultados que AICc e BIC. AIC e AICc tem sensibilidades maiores que 80% somente a partir do tamanho amostral igual a 40 e BIC não tem valores maiores que 25%.

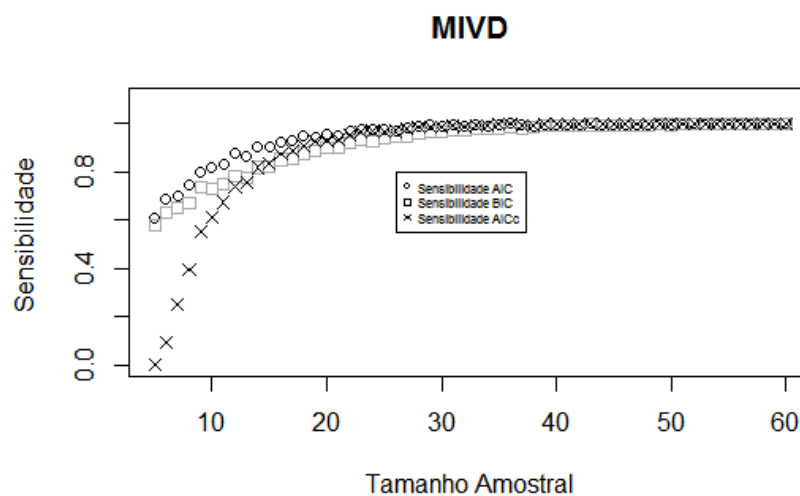
Figura 10 – Sensibilidade *versus* tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias diferentes. Grupo 2.



Quando os dados simulados tinham médias iguais e variâncias diferentes, Figura 11, os valores de sensibilidade se estabilizaram a partir do tamanho amos-

tral igual a 30. A partir deste tamanho amostral, para todos os critérios, os valores de sensibilidade foram iguais a 100%.

Figura 11 – Sensibilidade *versus* tamanho amostral: AIC, AICc, BIC. Médias iguais, variâncias diferentes. Grupo 2.



Em relação às taxas, quando simulados dados com médias iguais e variâncias iguais, Figura 12, o poder do teste, ou taxa de TN e significância do teste, ou taxa de TN, do AIC foi maior que para BIC e AICc.

Quando simulados dados com médias iguais e variâncias diferentes, Figura 13, AIC apresentou sensibilidade (taxa de TP) e poder do teste (taxa de TN) maior que BIC e AICc. BIC apresentou maior probabilidade do erro tipo II (taxa de FP) que AICc e AIC.

Figura 12 – Taxas de TP, FP, TN e FN considerando médias iguais e variâncias iguais. Grupo 2.

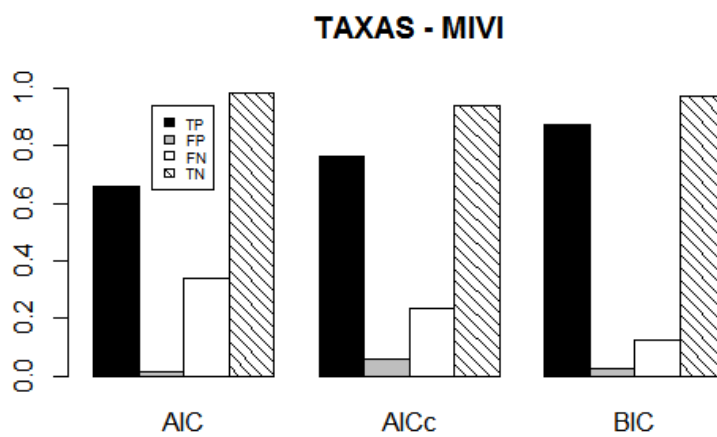
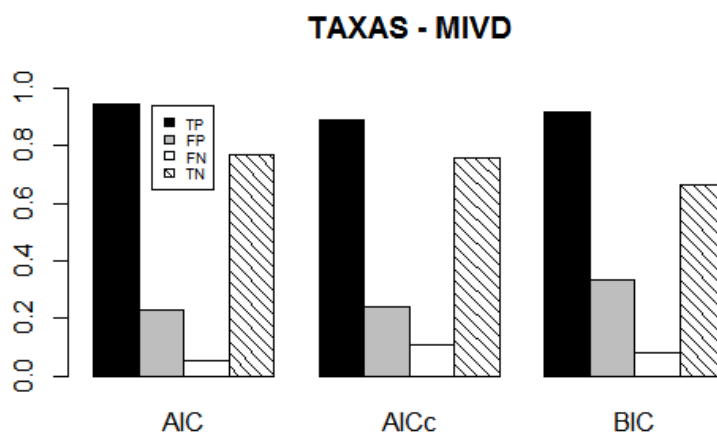
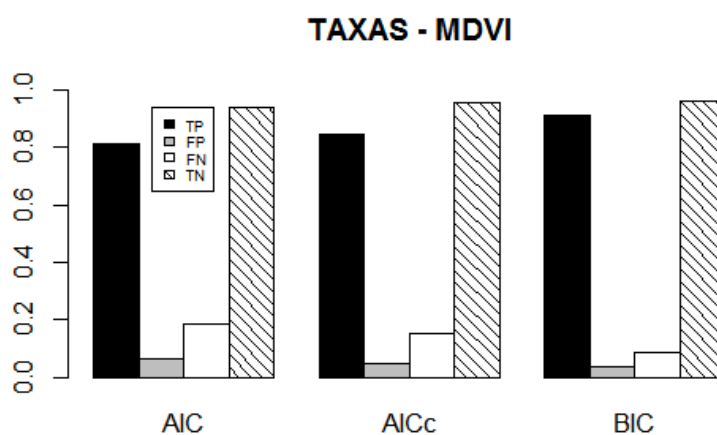


Figura 13 – Taxas de TP, FP, TN e FN considerando médias iguais e variâncias diferentes. Grupo 2.



Quando os dados simulados tinham médias diferentes e variâncias iguais, Figura 14, os resultados dos poderes do teste e sensibilidade (taxas de TN e TP) para os três critérios foram bastante semelhantes, assim como aconteceu para o grupo 1 (Figura 8). AIC obteve maior nível de significância (taxa de FN) que AICc e BIC e por fim, a probabilidade do erro tipo II para todos os critérios foi bastante semelhante (7% para AIC, 7% para AICc e 6% para BIC).

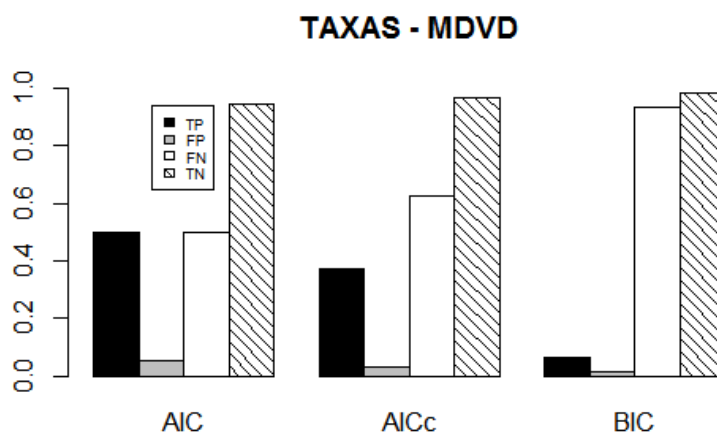
Figura 14 – Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias iguais. Grupo 2.



Quando foram simulados dados com médias diferentes e variâncias diferentes, Figura 15, AIC apresentou maior valor de sensibilidade (taxa de TP), seguido do AICc e BIC. BIC apresentou maior nível de significância (taxa de FN) que AICc e AIC. Os poderes do teste (taxa de TN) foram semelhantes, porém BIC apresentou maior valor.



Figura 15 – Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias diferentes. Grupo 2.



A tabela, Tabela 15, com os resultados numéricos das taxas das Figuras 12, 13, 14 e 15 esta no Apêndice B.

Os resultados referentes aos casos 2, 3 e 4 também estão no Apêndice B e seguiram a mesma metodologia apresentada nesta seção, ou seja, cada caso apresenta dois grupos, um com variâncias diferentes porém próximas e um com variâncias diferentes porém distantes como apresentado anteriormente nas Tabelas 4, 5, 6, 7, 8 e 9.

Os resultados foram bastante semelhantes aos apresentados nessa seção, por isso optou-se por colocá-los no apêndice.

## 4.2 Simulação com dados normais para Medida L

Nesta seção são feitos os estudos da sensibilidade *versus* tamanho amostral e os cálculos das outras taxas, a saber: FP, FN e TN.

Para as simulações, considera-se um modelo de regressão linear múltipla,

$$Y = X\beta + \epsilon$$

sendo  $Y$  um vetor de respostas de tamanho  $n$ ,  $\beta$  um vetor de coeficientes da regressão de tamanho  $k + 1$  e  $\epsilon$  um vetor de erros de tamanho  $n$ . Ainda,  $\epsilon \sim N_n(\mathbf{0}, \tau I)$  (matriz de precisão  $\tau I$ ).

Considera-se ainda que o conjunto de dados  $Y = (y_1, y_2, \dots, y_n)$  são distribuídos normalmente, ou seja,

$$Y \sim N_n(X(X'X)^{-1}X'\eta_0; \tau I)$$

em que  $X$  é uma matriz de preditores de posto completo formada por uma coluna de 1's e uma coluna de valores normais com média igual a 0 e variância igual a 1. A distribuição dos parâmetros  $\beta|\tau$  é Normal com média  $(X'X)^{-1}X'\eta_0$  e variância  $\tau I$ . O parâmetro de dispersão  $\tau$  tem distribuição Gamma com parâmetro  $\delta_0/2$  e  $\gamma_0/2$ , ou seja,  $\tau \sim \text{Gamma}(\delta_0/2, \gamma_0/2)$  e os seus valores para a simulação foram:  $\delta_0 = 25$  e  $\gamma_0 = 125$  para garantir  $\delta_0/\gamma_0 = 0,2$ .  $\eta_0$  é um vetor fixo, de valores futuros observados, independente do modelo sob consideração.

Ao se considerar distribuição Normal para os dados, de acordo com a seção 3.2 tem-se uma densidade preditiva, para os dados  $Z$ ,  $t$  multivariada para uma replicação do experimento ou ainda a continuação do mesmo:  $Z \sim t_n(n + \delta_0, \eta_m, s^2[I + (1 - \gamma)P_m])$ .

Para cada tamanho amostral a simulação foi repetida 1000 vezes. Foram consideradas ainda amostras de tamanho 5, 6, 7, ..., 60 e o vetor  $\eta_0$  distribuído normalmente com média 10 e variância igual a 1, ou seja,  $\eta_0 \sim N_n(\mu_{\eta_0} = 10, \sigma_{\eta_0}^2 = 1)$ .

Vale ressaltar que a escolha de  $\eta_0$  vai alterar as configurações das distribuições preditivas a posteriori que serão usadas para os cálculos das taxas em questão,

já que na distribuição preditiva dada em (13),  $p_m$  depende de  $\eta_0$ .

Na dinâmica de simulação deve-se considerar a Tabela 10.

Tabela 10 – Situações possíveis na seleção de modelos

Suposição	Teste de classificação	Realidade	Situação
Os dados são oriundos de C	Os dados são oriundos do modelo C	Os dados, realmente, são oriundos do modelo C	Verdadeiro Positivo
Os dados são oriundos de C	Os dados não são oriundos do modelo C	Os dados, realmente, são oriundos do modelo C	Falso Negativo
Os dados são oriundos de C	Os dados são oriundos do modelo C	Os dados, realmente, são não oriundos do modelo C	Falso Positivo
Os dados são oriundos de C	Os dados não são oriundos do modelo C	Os dados, realmente, são não oriundos do modelo C	Verdadeiro Negativo

Para o cálculo da taxas, considera-se três casos, cada um com os dados seguindo uma distribuição preditiva diferente e para cada um dos casos, uma situação diferente, apresentados nas Tabelas 11, 12, 13 a seguir.

Todos os casos compreendem o cálculo da Medida L entre a distribuição dos dados  $Y$  e as diferentes distribuições preditivas consideradas em cada situação.

Tabela 11 – Caso 1

Situação	Distribuição Original dos dados			Distribuição preditiva "verdadeira"			Distribuição preditiva próxima			Distribuição preditiva distante		
	$\mu_{\eta_0}$	$\sigma_{\eta_0}^2$	c	$\mu_{\eta_0}$	$\sigma_{\eta_0}^2$	c	$\mu_{\eta_0}$	$\sigma_{\eta_0}^2$	c	$\mu_{\eta_0}$	$\sigma_{\eta_0}^2$	c
1	10	1	1	10	1	1	12	1	5	100	2	5
2	12	1	5	12	1	5	10	1	1	100	2	5
3	100	2	5	100	2	5	12	1	5	10	1	1

Tabela 12 – Caso 2

Situação	Distribuição Original dos dados			Distribuição preditiva "verdadeira"			Distribuição preditiva próxima			Distribuição preditiva distante		
	$\mu_{\eta_0}$	$\sigma_{\eta_0}^2$	c	$\mu_{\eta_0}$	$\sigma_{\eta_0}^2$	c	$\mu_{\eta_0}$	$\sigma_{\eta_0}^2$	c	$\mu_{\eta_0}$	$\sigma_{\eta_0}^2$	c
1	10	1	1	10	1	1	25	1	5	100	2	5
2	25	1	5	25	1	5	10	1	1	100	2	5
3	100	2	5	100	2	5	25	1	5	10	1	1

Tabela 13 – Caso 3

Situação	Distribuição Original dos dados			Distribuição preditiva "verdadeira"			Distribuição preditiva próxima			Distribuição preditiva distante		
	$\mu_{\eta_0}$	$\sigma_{\eta_0}^2$	c	$\mu_{\eta_0}$	$\sigma_{\eta_0}^2$	c	$\mu_{\eta_0}$	$\sigma_{\eta_0}^2$	c	$\mu_{\eta_0}$	$\sigma_{\eta_0}^2$	c
1	10	1	1	10	1	1	40	1	5	100	2	5
2	40	1	5	40	1	5	10	1	1	100	2	5
3	100	2	5	100	2	5	40	1	5	10	1	1

### 4.3 Resultados para Medida L: sensibilidade *versus* tamanhos amostrais

#### 4.3.1 Caso 1

Na avaliação da situação 1, quando a distribuição preditiva “verdadeira”  $t$  tinha vetor  $\eta_0 \sim N_n(\mu_{\eta_0} = 10, \sigma_{\eta_0}^2 = 1)$ , os valores das sensibilidades para os diferentes tamanhos amostrais apresentam um comportamento estável em torno de 70%, com valores entre 63,6% e 74,1%, como apresentado na Figura 16.

Na situação 2, conforme a Figura 17, pode-se observar que a sensibilidade apresenta um comportamento decrescente. Seu valor mínimo é de 44,1% e o valor máximo é de 60,4%.

Na situação 3, conforme a Figura 18, observa-se que a sensibilidade mostrou um comportamento estável com valores iguais a 100%.

Pode-se observar que os valores de sensibilidade e nível de confiança do teste, para as situações 1 e 2, não são considerados na literatura como bons, já que o mais usual é considerar um valor de 95%, ou seja,  $1 - \alpha = 0,95$ . Já para a

situação 3, os valores das sensibilidades, para os diferentes tamanhos amostrais, são considerados bons.

Figura 16 – Sensibilidade *versus* tamanho amostral. Caso 1 - Situação 1

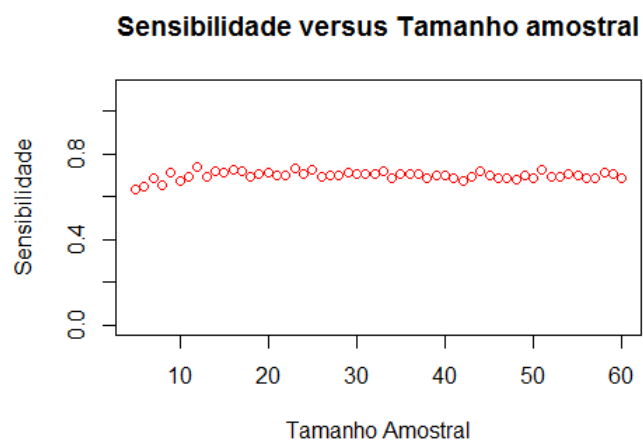


Figura 17 – Sensibilidade *versus* tamanho amostral. Caso 1 - Situação 2

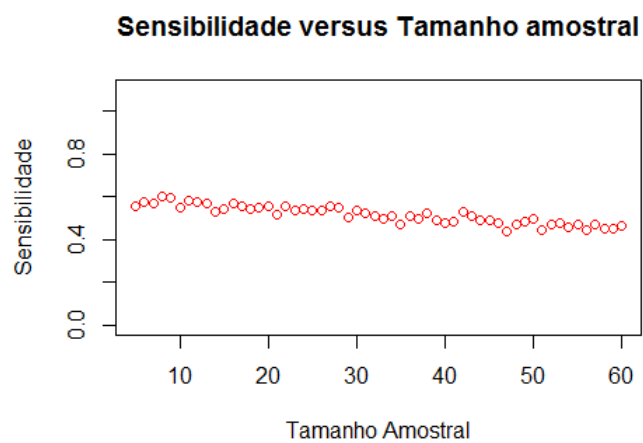
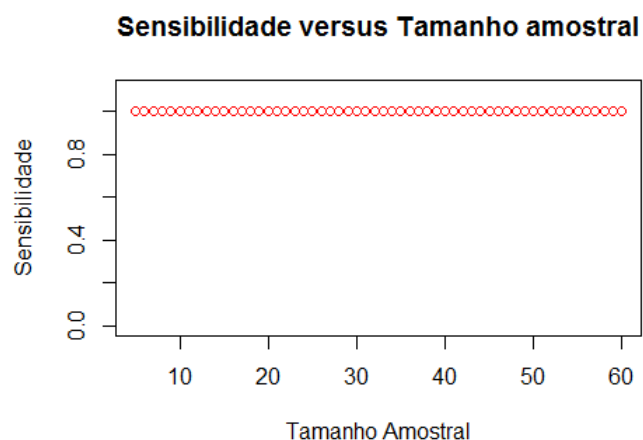


Figura 18 – Sensibilidade *versus* tamanho amostral. Caso 1 - Situação 3

#### 4.3.2 Caso 2

Na avaliação da situação 1, as sensibilidades para os diferentes tamanhos amostrais se mantiveram estáveis em torno de 99% e apresentou menor valor de 99% e maior valor de 100%, como apresentado na Figura 19.

Para a situação 2, os valores para as sensibilidades mantiveram-se estáveis em torno de 99,8%, apresentando mínimo de 94,6% e máximo de 99,9% (Figura 20); e para a situação 3, as taxas se mantiveram estáveis em 100% (Figura 21).

No caso 2, para todas as situações apresentadas, os valores das sensibilidades para diferentes tamanhos amostrais mostraram-se bons, já que, ao se considerar o valor de 95% como referência para o nível de confiança, 100% mostra-se excelente.

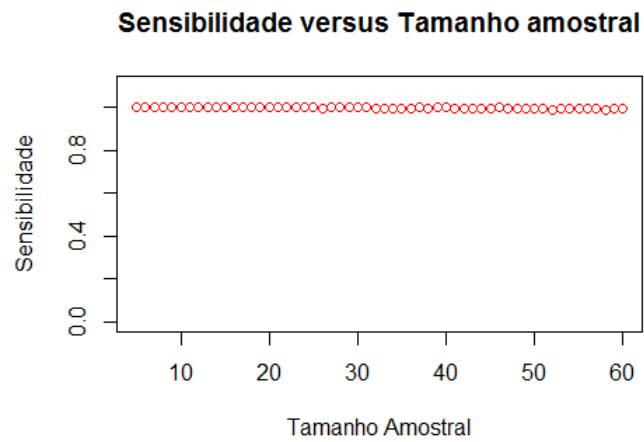
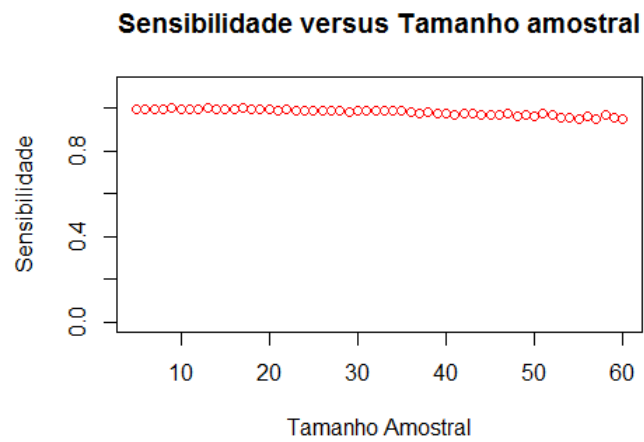
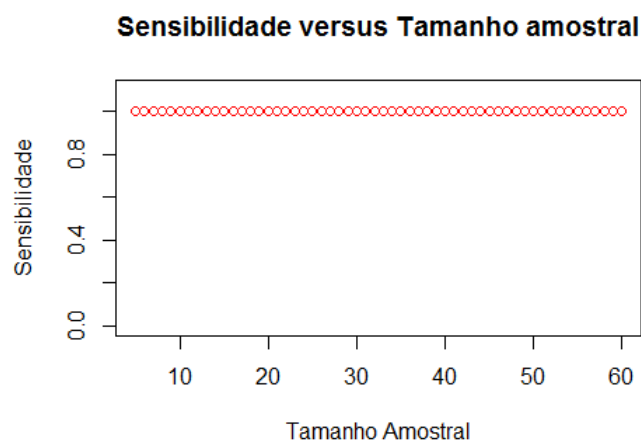
Figura 19 – Sensibilidade *versus* tamanho amostral. Caso 2 - Situação 1Figura 20 – Sensibilidade *versus* tamanho amostral. Caso 2 - Situação 2

Figura 21 – Sensibilidade *versus* tamanho amostral. Caso 2 - Situação 3

#### 4.3.3 Caso 3

No caso 3, assim como para o caso 2, os resultados dos valores das sensibilidades apresentados para os diferentes tamanhos amostrais foram excelentes.

#### 4.3.4 Resultados das simulações para as taxas de TP (Sensibilidade), FP, FN e TN (Especificidade)

As Figuras 22, 23 e 24 mostram as taxas de verdadeiro positivo (TP), falso positivo (FP) (probabilidade do erro tipo II), falso negativo (FN) (significância do teste) e verdadeiro negativo (TN) (poder do teste ou especificidade) para as três situações consideradas.

Analisando a Figura 22, observa-se os resultados das taxas para o caso 1. As taxas de verdadeiro positivo foram de 70%, 53% e 100% respectivamente para as situações 1, 2 e 3 e suas variâncias foram 4,31%, 6,72% e 0%. Esses são os valores médios dos valores das sensibilidades apresentadas na seção anterior (Seção 4.3). Suas variâncias podem ser utilizadas para o cálculo dos coeficientes de variabilidade, que são, respectivamente iguais a 2,96%, 4,8% e 0%, o que mostra



uma baixa variabilidade nos valores de sensibilidade para os diferentes tamanhos amostrais pesquisados. As taxas de falso positivo, que dão indicativo da probabilidade do erro tipo II ( $\beta$ ) foram, respectivamente, de 24%, 14% e 0%. Valores altos de  $\beta$  não são indicados, portanto, as taxas para as situações 1 e 2 podem não ser adequadas.

Ainda para o caso 1 (Figura 22), os valores de falso negativo, que indicam a probabilidade de cometermos o erro tipo I, são complementares às sensibilidades, portanto podem ser interpretadas de maneira análoga. E os valores das especificidades são complementares ao valores das taxas de falso positivo e também podem ser interpretados de maneira análoga.

Analisando a Figura 23, referente ao caso 2, pode-se observar que para todas as situações as médias das taxas de verdadeiro positivo (ou seja, das sensibilidades, como apresentadas na seção anterior) foram maiores que 98% e as taxas de verdadeiro negativo maiores que 99%. As taxas de falso positivo (valores para  $\beta$ ) foram menores que 1%. As variâncias da taxa de verdadeiro positivo foram para a situação 1, 2 e 3, respectivamente, 0,38%, 1,97% e 0%. As taxas de verdadeiro negativo que dão os valores para especificidade, ou seja, os poder do teste, apresentaram-se elevadas, portanto considerados boas. E os valores para as taxas de falso negativo ( $\alpha$ ) também foram considerados bons.

Na Figura 24, observa-se que para o caso 3, em todas as situações apresentadas, as taxas de verdadeiro positivo e as taxas de verdadeiro negativo foram iguais a 100%. As variâncias das taxas de verdadeiro positivo foram iguais a 0%, portanto, seus coeficientes de variabilidade também foram iguais a 0%. E os valores para as taxas de falso positivo ( $\beta$ ) e falso negativo ( $\alpha$ ) foram iguais a 0%. Da mesma forma como nas interpretações para os casos anteriores, o caso 3 mostrou-se o de melhor avaliação em termos de valores para as taxas avaliadas.

Figura 22 – Taxas de TP, FP, FN e TN. Caso 1

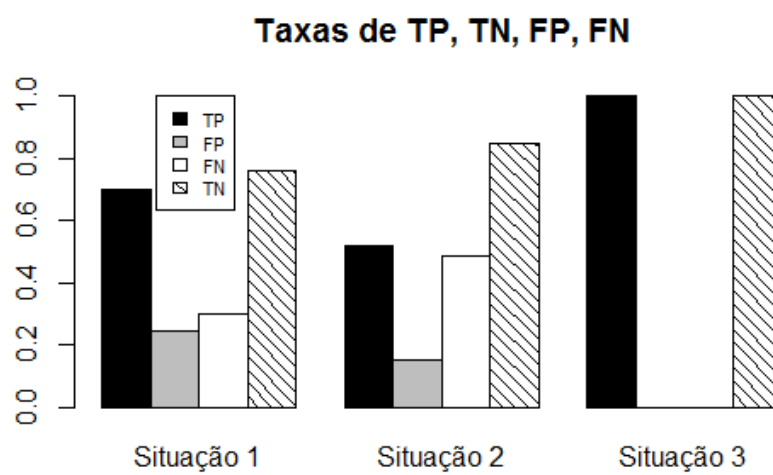


Figura 23 – Taxas de TP, FP, FN e TN. Caso 2

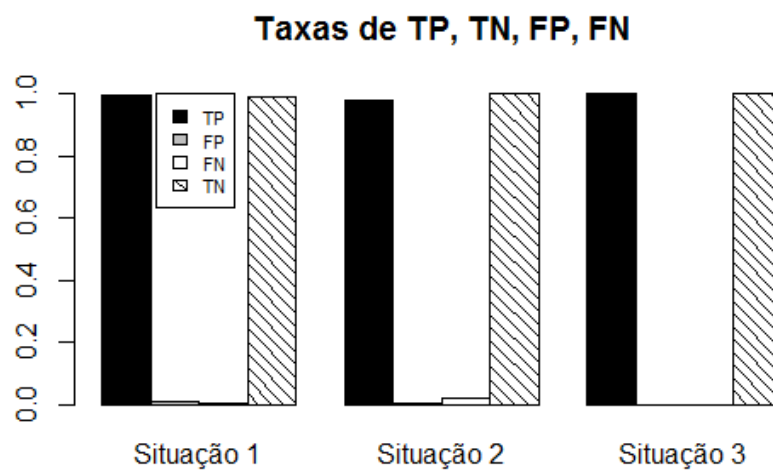
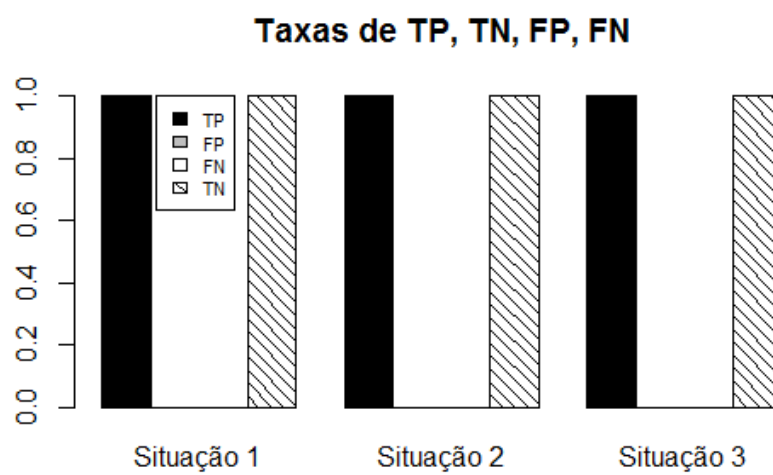


Figura 24 – Taxas de TP, FP, FN e TN. Caso 3



## 5 CONCLUSÕES

Os critérios de Akaike, Akaike corrigido e Bayesiano penalizam a verossimilhança e a Medida L é baseada no preditivismo. Apesar de ter que considerar a diferença de abordagem dos critérios, verifica-se que a Medida L apresentou valores de sensibilidade mais altos que os valores de sensibilidade para os critérios AIC, AICc e BIC na seleção de modelos lineares com amostras menores que 60.

Considerando a Medida L, critério estudado nesse trabalho, podemos observar que, dos diferentes casos apresentados, quando a distribuição preditiva era muito próxima à distribuição preditiva realmente associada aos dados simulados, como apresentado no caso 1, o poder e a significância do teste eram reduzidos.

Quando a comparação envolvia uma distribuição preditiva relativamente próxima, como apresentado no caso 2, o poder e a significância apresentaram bons resultados. Assim como quando a comparação envolvia uma distribuição preditiva “distante” da distribuição preditiva associada aos dados simulados (caso 3), quando estes apresentaram resultados excelentes.

Ao contrário dos resultados apresentados para os critérios AIC, AICc e BIC, a Medida L mostrou-se mais eficaz para amostras pequenas, apresentando taxas estáveis e elevadas para os casos 2 e 3.

## REFERÊNCIAS

ALTMAN, D. G.; BLAND, J. MARTIN. Statistics Notes: Diagnostics tests 1: sensitivity and specificity. **British Medical Journal**, v. 308, p. 1552, 1994.

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, Boston, v.9, n.6, p. 716-723, Dec. 1974.

BERGER, J. O. **Statistical Decision Theory and Bayesian Analysis**. 2a ed. New York: Springer-Verlag, 1985. 617p.

BERNARDO, J. M.; SMITH, A. F. M. **Bayesian Theory**. Chichester: John Wiley & Sons, Ltd, 2000. 586 p.

BOLFARINE, H.; SANDOVAL, M. C. **Introdução à inferência estatística**. São Paulo: Sociedade Brasileira de Matemática, 2000. 125 p.

BOX, G. E. P.; TIAO, G. C. **Bayesian Inference in statistical analysis**. Wiley Classics Library, 1992. 588 p.

BOZDOGAN H. Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. **Psychometrika** , n. 52, p. 345-370, 1987.

BROEMELING, L. D. **Bayesian Analysis of Linear Models**. STATISTICS: textbooks and monographs, v. 60, 1985. 454 p.

BUSEMEYER, J. R.; WANG, Y-M. Model Comparisons and Model Selections Based on Generalization Criterion Methodology. **Journal of Mathematical Psychology**, v.44, p.171-189, 2000.

CASELA, G.; BERGER, R. L. **Inferência Estatística**. 2a ed. Cengage Learning, 2010. 588 p.

CHEN, M-H.; IBRAHIM, J. G. Bayesian Predictive Inference for Time Series Count Data. **Biometrics**, v. 56, p. 678-685, 2000.

EMILIANO, P. C. **Critérios de Informação: como eles se comportam em diferentes modelos?**. 2013. 190 f. Tese (Doutorado), Universidade Federal de Lavras,

Lavras, 2013.

EMILIANO, P. C.; VIVANCO, M. J. F.; MENEZES, F. S. M.; AVELAR, F. G. Foundations and comparison of information criteria: Akaike and Bayesian. **Revista Brasileira de Biometria**, São Paulo, v.27, n.3, p.394-411, 2009.

EMILIANO, P. C.; VIVANCO, M. J. F.; MENEZES, F. S.de. Information Criteria: How do they behave in different models?. **Computational Statistics and Data Analysis**, v. 69, P. 141-153, 2014.

GELMAN, A. et al. **Bayesian data analysis**. Boca Raton, FL, USA: Chapman & Hall/CRC, 2003.

GHOSH, J. K.; DELAMPADY, M.; SAMANTA, T. An Introduction to Bayesian Analysis: Theory and Methods. 2006.

HOETING, J. A.; IBRAHIM, J. G. Bayesian Predictive Simultaneous Variable and Transformation Selection in Linear Model. **Journal of Computational Statistics and Data Analysis**, v. 28, p. 87-103, 1998.

IBRAHIM, J. G.; CHEN, M-H.; SINHA, D. Criterion-based for Bayesian Model Assessment. **Statistica Sinica**, v.11, p.419-443, 2001.

IBRAHIM, J. G.; LAUD, P. W. A Predictive Approach to the Analysis of Designed Experiments. **Journal of the American Statistical Association**, v.89, n.425, p.309-319, 1994.

LAUD, P. W.; IBRAHIM, J. G. Predictive Model Selection. **Journal of the Royal Statistical Society Series B**, v.57, n.1, p. 247-262, 1995.

MARTINEZ, E. Z.; LOUZADA-NETO, F.; PEREIRA, B. de B. A curva ROC para testes diagnósticos. **Cadernos Saúde Coletiva**, v.7, p 7-31, 2003.

MARTINI, A. S.; SPEZZAFERRI, F. A predictive model selection criterion. **Journal of the Royal Statistical Society. Series B (Methodological)**, p. 296-303, 1984.

MOSIER, C. I. The need and means of cross validation. I. Problems and designs of cross-validation. **Educational and Psychological Measurement**, 1951.

PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. **Estatística Bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2003. 446 p.

PRESS, S. J. **Subjective and objective Bayesian statistics: principals, models and applications**. 2. ed. New Jersey: Wiley-Interscience, 2003. 591 p.

R DEVELOPMENT CORE TEAM. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2009.

SUGIURA, N. Further analysis of the data by Akaike's information criterion and the finite corrections. **Communications in Statistics - Theory and Methods**, Ontario, v. 7, n.1, p.13-26, 1978.

SCHWARZ, G. Estimating the dimensional of a model. **Annals of Statistics**, Hayward, v.6, n.2, p.461-464, Mar. 1978.

SWETS, J. A. Measuring the Accuracy of Diagnostic Systems. **Science, New Series**, v. 240, p. 1285-1293, 1988.

VEIGA, E. P. **A Medida L como critério de comparação de modelos: uma revisão de literatura**. 2012. 64 f. Dissertação (Mestrado), Universidade Federal de Lavras, Lavras, 2012.

VEIGA, E. P.; VIVANCO, M. J. F. A Medida L como critério de comparação de modelos: uma revisão de literatura. **Revista Brasileira de Biometria**, São Paulo, v. 30, n. 3, p. 343-352, 2012.

## APÊNDICE A

### Distribuições Univariadas

#### Distribuição Normal

Uma variável aleatória  $X$  tem distribuição Normal com parâmetros  $\mu$  e  $\sigma^2$  se a função de densidade é dada por

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(X - \mu)^2\right), \mu \in \Re, X \in \Re, \sigma^2 > 0.$$

Notação:  $X \sim N(X|\mu, \sigma^2)$ . A média e a variância são dadas por  $E[X] = \mu$ ,  $Var[X] = \sigma^2$ .

Se  $\tau^{-1} = \sigma^2$ , diz-se que  $\tau$  representa a precisão da distribuição e a função de densidade pode ser reescrita como

$$f(x|\mu, \tau) = \frac{\tau^{\frac{1}{2}}}{\sqrt{2\pi}} \exp\left(-\frac{\tau}{2}(x - \mu)^2\right), \tau > \Re, x \in \Re.$$

#### Distribuição Gama

Uma variável aleatória  $X$  tem distribuição Gama com parâmetros  $\alpha$  e  $\beta$  se a função de densidade é dada por

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), x > 0, \alpha > 0, \beta > 0.$$

Notação:  $X \sim Gama(x|\alpha, \beta)$ . A média e a variância são dadas por  $E[X] = \frac{\alpha}{\beta}$  e  $Var[X] = \frac{\alpha}{\beta^2}$ .

#### Distribuição Gama-Inversa

Uma variável aleatória  $X$  tem distribuição Gama-inversa com parâmetros  $\alpha$  e  $\beta$  se a função de densidade é dada por

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} \exp\left(-\frac{\beta}{x}\right), x > 0, \alpha > 0, \beta > 0.$$



Notação:  $X \sim GI(x|\alpha, \beta)$ . A média e a variância são dadas por  $E[X] = \frac{\beta}{\alpha-1}$ ,  $\alpha > 0$  e  $Var[X] = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$ ,  $\alpha > 2$ .

Observação: Se  $Y \sim Gama(\alpha, \beta) \Rightarrow X = Y^{-1} \sim GI(\alpha, \beta)$ .

### Distribuição $t$ de Student

Uma variável aleatória  $X$  tem distribuição  $t$  de Student com parâmetros  $\mu$ ,  $\lambda$  e  $\alpha$  ( $\mu \in \Re$ ,  $\lambda > 0$ ,  $\alpha > 0$ ) se a sua densidade é

$$f(x|\mu, \lambda, \alpha) = c \left[ 1 + \frac{\lambda}{\alpha} (X - \mu)^2 \right]^{-(\alpha+1)/2}, x \in \Re,$$

em que

$$c = \frac{\Gamma((\alpha + \frac{1}{2}))}{\Gamma(\frac{\alpha}{2})\Gamma(\frac{1}{2})} \left( \frac{\lambda}{\alpha} \right)^{\frac{1}{2}}$$

A distribuição é simétrica em  $x = \mu$ , e tem uma única moda  $M[X] = \mu$ . A média e a variância são dadas por

$$E[X] = \mu \text{ se } \alpha > 1,$$

$$V[X] = \frac{1}{\lambda} \frac{\alpha}{(\alpha-2)} \text{ se } \alpha > 2.$$

O parâmetro  $\alpha$  é chamado *graus de liberdade* da distribuição.

A distribuição é gerada (Dickey, 1968) pela mistura

$$f(x|\mu, \lambda, \alpha) = \int_0^{\infty} N(x|\mu, \lambda y) Ga\left(y \left| \frac{\alpha}{2}, \frac{\alpha}{2} \right.\right) dy,$$

e inclui a distribuição Normal como um caso especial, quando  $N(x|\mu, \lambda y) = \lim_{\alpha \rightarrow \infty} f(x|\mu, \lambda, \alpha)$ .

Se  $y = \lambda^{\frac{1}{2}}(x - \mu)$ , em que  $x$  tem função de densidade  $f(x|\mu, \lambda, \alpha)$ , então  $y$  tem densidade  $t$   $f(y|0, 1, \alpha)$ . Se  $\alpha = 1$ ,  $x$  é dito ter distribuição de Cauchy com densidade  $Ca(x|\mu, \lambda)$ .

Se  $x$  tem distribuição normal padrão,  $y$  tem distribuição  $\chi^{2\nu}$  e  $x$  e  $y$  são mutuamente independentes, então

$$z = \frac{x}{\left(\frac{y}{\nu}\right)^{\frac{1}{2}}}$$

tem distribuição  $t$  de Student  $f(z|0, 1, \nu)$ .

## Distribuições Multivariadas

### Distribuição $t$ de Student Multivariada

Um vetor aleatório contínuo  $\mathbf{x} = (x_1, \dots, x_k)'$  tem distribuição  $t$  de Student multivariada de dimensão  $k$ , com parâmetros  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)'$ ,  $\boldsymbol{\lambda}$  e  $\alpha$  ( $\boldsymbol{\mu} \in \Re^k$ ,  $\boldsymbol{\lambda}$  é uma matriz simétrica, positiva definida ( $k \times k$ ) e  $\alpha > 0$ ) se sua densidade  $f_k(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\lambda}, \alpha)$  é

$$f_k(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\lambda}, \alpha) = c \left[ 1 + \frac{1}{\alpha} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\lambda} (\mathbf{x} - \boldsymbol{\mu}) \right]^{-\frac{(\alpha+k)}{2}}$$

$$\mathbf{x} \in \Re^k \text{ e } c = \frac{\Gamma(\frac{\alpha+k}{2})}{\Gamma(\frac{\alpha}{2})(\alpha\pi)^{\frac{k}{2}}} |\boldsymbol{\lambda}|^{\frac{1}{2}}.$$

Se  $k = 1$ , então  $\boldsymbol{\lambda}$  é um escalar, então  $f_k(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\lambda}, \alpha)$  se reduz a distribuição  $t$  de Student univariada.

A média e a variância são dadas por  $E[\mathbf{x}] = \boldsymbol{\mu}$  e  $V[\mathbf{x}] = \boldsymbol{\lambda}^{-1} \left( \frac{\alpha}{\alpha-2} \right)$ . O parâmetro  $\boldsymbol{\lambda}$  é frequentemente referido como *matriz de precisão* da distribuição.

### Distribuição Normal Multivariada

Um vetor aleatório contínuo  $\mathbf{x} = (x_1, \dots, x_k)'$  tem uma distribuição Normal Multivariada de dimensão  $k$ , com parâmetros  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)'$ ,  $\boldsymbol{\lambda}$  em que  $\boldsymbol{\mu} \in \Re^k$ ,  $\boldsymbol{\lambda}$  é uma matriz simétrica, positiva definida ( $k \times k$ ) se sua densidade  $f_k(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\lambda})$  é

$$f_k(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\lambda}) = \frac{|\boldsymbol{\lambda}|^{\frac{1}{2}}}{(2\pi)^{\frac{k}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\lambda} (\mathbf{x} - \boldsymbol{\mu}) \right\}.$$

Em geral,  $E[x_i] = \mu_i$  e  $\Sigma = \boldsymbol{\lambda}^{-1}$  de elementos  $\sigma_{ij}$ , sendo  $Var(x_i) = \sigma_{ii}$  e  $Cov(x_i, x_j) = \sigma_{ij}$  tal que  $Var(\mathbf{x}) = \boldsymbol{\lambda}^{-1}$ .

O parâmetro  $\boldsymbol{\mu}$  é denominado vetor de médias e  $\boldsymbol{\lambda}$  é a matriz de precisão (inversa da matriz de covariância).

### Distribuição Normal-Gama

Um vetor aleatório contínuo  $(\mathbf{x}, \mathbf{y})$  tem distribuição Normal-Gama com parâmetros  $\mu$ ,  $\sigma^2$ ,  $\alpha$  e  $\beta$  se a função de densidade é dada por

$$f(\mathbf{x}, \mathbf{y}|\mu, \sigma^2, \alpha, \beta) = f(x|\mu, \sigma^2) f(y|\alpha, \beta), \mu \in \Re, \sigma^2 > 0, \alpha > 0, \beta > 0, x \in \Re, y > 0,$$

em que  $f(x|\mu, \sigma^2)$  e  $f(y|\alpha, \beta)$  já foram definidas acima como distribuição Normal e distribuição Gama respectivamente.

Notação:  $(X, Y) \sim NG(x, y|\mu, \sigma^2, \alpha, \beta)$ .

## APÊNDICE B

### Tabelas com os resultados das taxas de TP, FP, FN e TN para AIC, AICc e BIC para o Caso 1, Grupos 1 e 2

Nessa seção são apresentadas as tabelas referentes aos resultados das taxas de TP, FP, FN e TN para os critérios AIC, AICc e BIC nas simulações do caso 1, grupos 1 (Tabela 14) e 2 (Tabela 15).

As Figuras foram mostradas na Seção 4.1; Figuras 6, 7, 8, 9 para o grupo 1 e Figuras 12, 13, 14, 15 para o grupo 2.

Tabela 14 – Valores das taxas TP, FP, FN e TN para AIC, AICc e BIC para o Caso 1, Grupo 1.

Grupo 1		MIVI	MIVD	MDVI	MDVD
AIC	TP	0,66	0,41	0,82	0,41
	FP	0,15	0,08	0,27	0,06
	FN	0,34	0,59	0,18	0,59
	TN	0,85	0,92	0,73	0,94
AICc	TP	0,77	0,35	0,85	0,31
	FP	0,21	0,06	0,27	0,03
	FN	0,23	0,65	0,15	0,69
	TN	0,79	0,94	0,73	0,97
BIC	TP	0,87	0,23	0,91	0,21
	FP	0,24	0,05	0,29	0,01
	FN	0,13	0,77	0,09	0,79
	TN	0,76	0,95	0,71	0,99

Tabela 15 – Valores das taxas TP, FP, FN e TN para AIC, AICc e BIC para o Caso 1, Grupo 2.

Grupo 2		MIVI	MIVD	MDVI	MDVD
AIC	TP	0,66	0,95	0,82	0,50
	FP	0,01	0,23	0,06	0,05
	FN	0,34	0,05	0,18	0,50
	TN	0,99	0,77	0,94	0,95
AICc	TP	0,77	0,89	0,85	0,37
	FP	0,06	0,24	0,05	0,03
	FN	0,23	0,11	0,15	0,63
	TN	0,94	0,76	0,95	0,97
BIC	TP	0,87	0,92	0,91	0,07
	FP	0,03	0,33	0,04	0,01
	FN	0,13	0,08	0,09	0,93
	TN	0,97	0,67	0,96	0,99

**Resultados para Sensibilidade *versus* tamanho amostral e taxas de TP, FP, FN e TN de AIC, AICc, BIC para amostras menores que 60: casos 2, 3 e 4.**

Nessa seção são apresentados os gráficos para Sensibilidade *versus* tamanho amostral para os critérios AIC, AICc, BIC para amostras menores que 60, 1000 repetições conforme metodologia apresentada na seção 4.1 para os casos 2 e 3 e 4.

A metodologia para as simulações foi baseada no trabalho de Emiliano, Vivanco e Menezes (2014) e segue o seguinte procedimento: foram simulados dados com diferentes configurações, e para cada configuração foram ajustados modelos com *mivi*, *mdvi*, *mivd* e *mdvd*. E por fim, foram calculados os valores dos critérios para cada situação.

Para exemplificar a simulação:

1. Simula-se os dados com *mivi*
2. Ajusta-se o modelo *mivi* aos dados e calcula AIC, AICc, BIC
3. Ajusta-se o modelo *mdvi* aos dados e calcula AIC, AICc, BIC
4. Ajusta-se o modelo *mivd* aos dados e calcula AIC, AICc, BIC
5. Ajusta-se o modelo *mdvd* aos dados e calcula AIC, AICc, BIC

6. Procede-se com o cálculo das taxas TP, FP, TN e FN.

Para a apresentação dos resultados das taxas TP, FP, FN e TN, optou-se pela apresentação somente dos gráficos e das tabelas correspondentes.

## Caso 2

Nessa seção são apresentados os resultados referentes ao caso 2, grupos 1 e 2. É importante salientar que existem três configurações para a simulação dos dados que já foram consideradas no caso 1, a saber: médias iguais e variâncias iguais ( $\mu_1 = \mu_2 = 10$  e  $\sigma_1^2 = \sigma_2^2 = 1$ ) e médias iguais e variâncias diferentes próximas e distantes ( $\mu_1 = \mu_2 = 10$ ;  $\sigma_1^2 = 1$ ;  $\sigma_2^2 = 0.81$  e  $\mu_1 = \mu_2 = 10$ ;  $\sigma_1^2 = 1$ ;  $\sigma_2^2 = 2$ ), portanto, os gráficos de sensibilidade *versus* tamanho amostral não serão apresentados novamente, já que foram apresentadas anteriormente nas Figuras 2 e 5 e 11, na seção 4.1.

## Grupo 1

O grupo 1 é aquele em que quando as variâncias são diferentes, elas são próximas.

### Sensibilidade *versus* tamanho amostral

Quando foram simulados dados com médias diferentes e variâncias iguais, Figura 25, BIC apresentou melhores resultados a partir do tamanho amostral igual a 20, com média de 99%. AIC apresentou média de 83% e AICc apresentou comportamento decrescente, estabilizando a partir do tamanho amostral igual a 30 com média de 94%.

Quando foram simulados dados com médias diferentes e variâncias diferentes todos os critérios apresentaram valores de sensibilidade menores que 60%. AIC apresentou melhores resultados que AICc e BIC. BIC apresentou um comportamento estável em torno de 21% (Figura 26).

Figura 25 – Sensibilidade *versus* tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias iguais. Grupo 1.

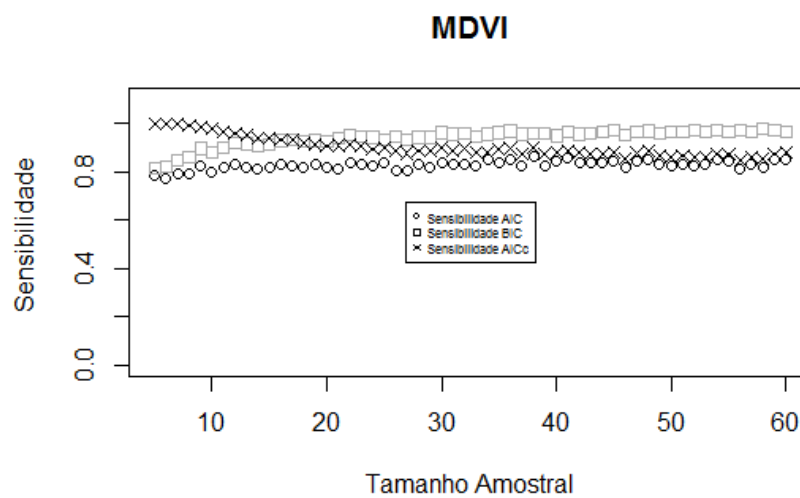
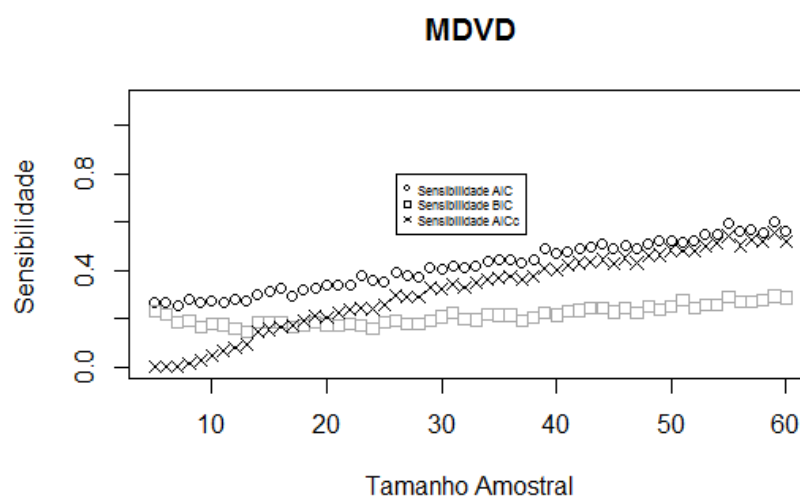


Figura 26 – Sensibilidade *versus* tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias diferentes. Grupo 1.



### Taxas de TP, FP, FN e TN

A Tabela 16 apresenta os valores para as taxas apresentadas nas Figuras 27, 28, 29 e 30 que referem-se às simulações de dados do caso 2, grupo 1, a saber: dados com médias iguais e variâncias iguais (MIVI), médias iguais e variâncias diferentes (MIVD), médias diferentes e variâncias iguais (MDVI) e médias diferentes e variâncias diferentes (MDVD) respectivamente.

Tabela 16 – Valores das taxas de TP, FP, FN e TN para AIC, AICc e BIC para o Caso 2, Grupo 1.

Caso 1		MIVI	MIVD	MDVI	MDVD
AIC	TP	0,66	0,41	0,83	0,42
	FP	0,15	0,07	0,27	0,07
	FN	0,34	0,59	0,17	0,58
	TN	0,85	0,93	0,73	0,93
AICc	TP	0,77	0,35	0,90	0,31
	FP	0,19	0,05	0,28	0,04
	FN	0,23	0,65	0,10	0,69
	TN	0,81	0,95	0,72	0,96
BIC	TP	0,87	0,23	0,94	0,21
	FP	0,24	0,03	0,29	0,02
	FN	0,13	0,77	0,06	0,79
	TN	0,76	0,97	0,71	0,98



Figura 27 – Taxas de TP, FP, TN e FN considerando médias iguais e variâncias iguais. Grupo 1.

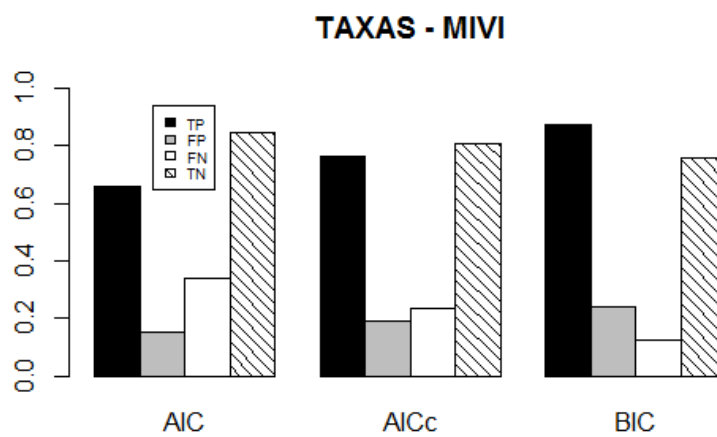


Figura 28 – Taxas de TP, FP, TN e FN considerando médias iguais e variâncias diferentes. Grupo 1.

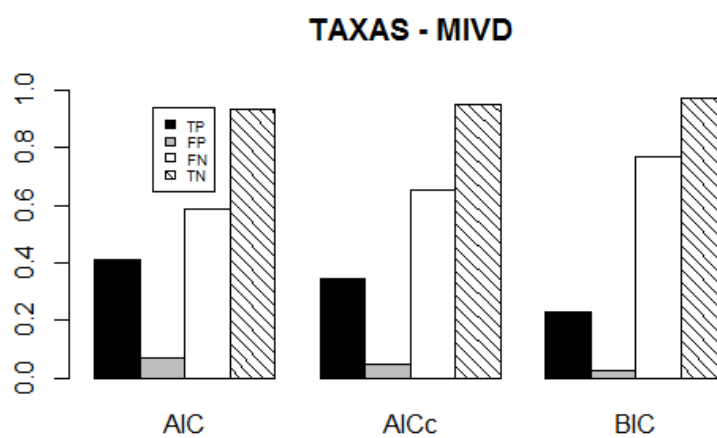


Figura 29 – Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias iguais. Grupo 1.

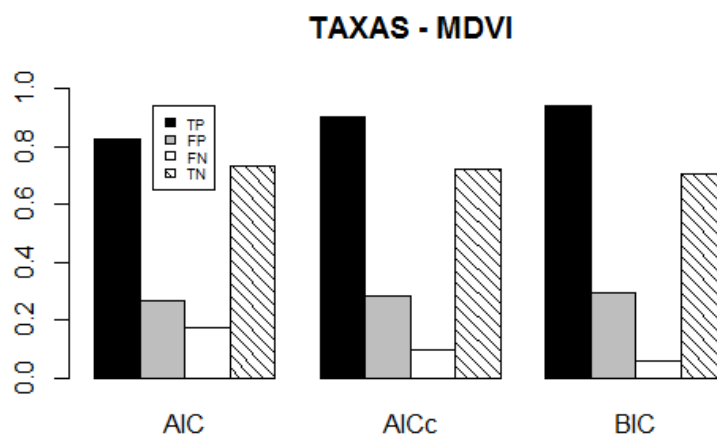
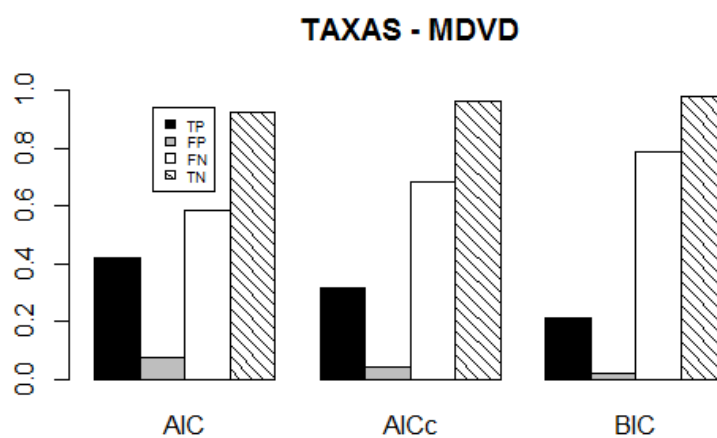


Figura 30 – Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias diferentes. Grupo 1.



## Grupo 2

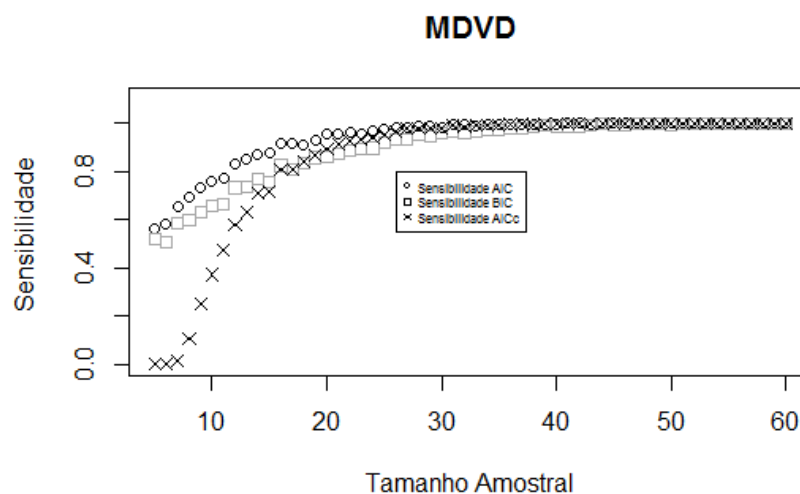
O grupo 2 é aquele em que quando as variâncias são diferentes, elas são distantes.

É importante notar que neste grupo, existe uma configuração para a simulação dos dados que já foi considerada no grupo 1, a saber: médias diferentes e variâncias iguais (MDVI), portanto, o gráfico de sensibilidade *versus* tamanho amostral não será apresentado novamente, já que foi apresentado anteriormente na Figura 25.

### Sensibilidade *versus* tamanho amostral

Quando os dados simulados tinham médias diferentes e variâncias diferentes, os valores de sensibilidade para todos os critérios estabilizou-se a partir do tamanho amostral igual a 38, em torno de 99%. Até este tamanho amostral, AIC apresentou maiores valores, seguido do BIC e AICc (Figura 31).

Figura 31 – Sensibilidade *versus* tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias diferentes. Grupo 2.



### Taxas de TP, FP, FN e TN

A Tabela 17 apresenta os valores para as taxas apresentadas nas Figuras 32, 33, 34 e 35 que referem-se às simulações de dados com médias iguais e variâncias iguais (MIVI), médias iguais e variâncias diferentes (MIVD), médias diferentes e variâncias iguais (MDVI) e médias diferentes e variâncias diferentes (MDVD), respectivamente.

Tabela 17 – Valores das taxas de TP, FP, FN e TN para AIC, AICc e BIC para o Caso 2, Grupo 2.

Caso 2		MIVI	MIVD	MDVI	MDVD
AIC	TP	0,66	0,95	0,83	0,94
	FP	0,01	0,07	0,07	0,06
	FN	0,34	0,05	0,17	0,06
	TN	0,99	0,93	0,93	0,94
AICc	TP	0,77	0,89	0,90	0,85
	FP	0,04	0,05	0,08	0,03
	FN	0,23	0,11	0,10	0,15
	TN	0,96	0,95	0,92	0,97
BIC	TP	0,87	0,92	0,94	0,90
	FP	0,02	0,03	0,05	0,02
	FN	0,13	0,08	0,06	0,10
	TN	0,98	0,97	0,95	0,98

Figura 32 – Taxas de TP, FP, TN e FN considerando médias iguais e variâncias iguais. Grupo 2.

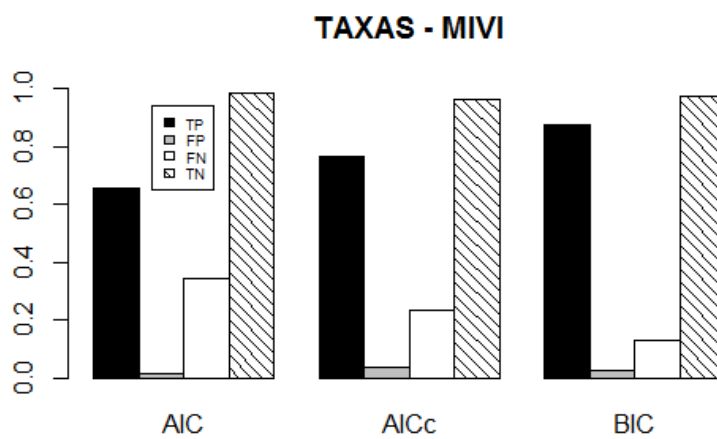


Figura 33 – Taxas de TP, FP, TN e FN considerando médias iguais e variâncias diferentes. Grupo 2.

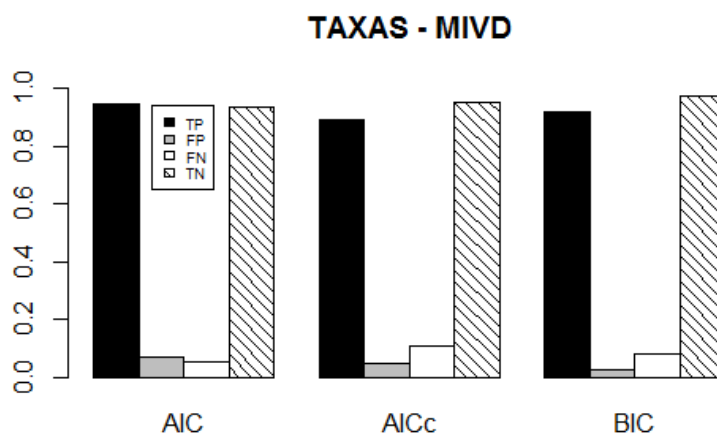


Figura 34 – Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias iguais. Grupo 2.

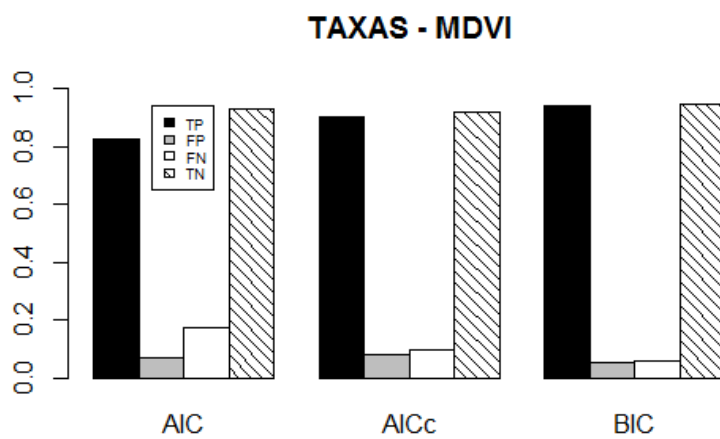
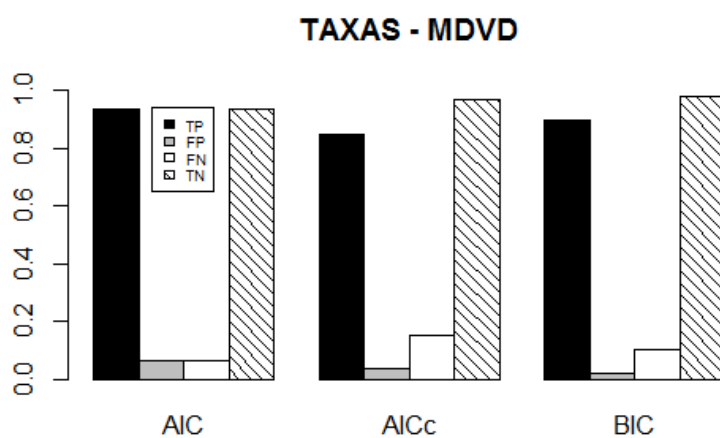


Figura 35 – Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias diferentes. Grupo 2.



### Caso 3

Nessa seção são apresentados os resultados referentes ao caso 3, grupos 1 e 2. É importante salientar que existem três configurações para a simulação dos dados que já foram consideradas no caso 1, a saber: médias iguais e variâncias iguais ( $\mu_1 = \mu_2 = 10$  e  $\sigma_1^2 = \sigma_2^2 = 1$ ) e médias iguais e variâncias diferentes próximas e distantes ( $\mu_1 = \mu_2 = 10$ ;  $\sigma_1^2 = 1$ ;  $\sigma_2^2 = 0.81$  e  $\mu_1 = \mu_2 = 10$ ;  $\sigma_1^2 = 1$ ;  $\sigma_2^2 = 2$ ), portanto, os gráficos de sensibilidade *versus* tamanho amostral não serão apresentados novamente, já que foram apresentadas anteriormente nas Figuras 2 e 5 e 11, na seção 4.1.

#### Grupo 1

O grupo 1 é aquele em que quando as variâncias são diferentes, elas são próximas.

#### Sensibilidade *versus* tamanho amostral

Quando os dados simulados tinham médias diferentes e variâncias iguais, Figura 36, a partir do tamanho amostral igual a 14, BIC apresentou maiores valores de sensibilidade que AICc e AIC. A partir deste tamanho amostral, a sensibilidade do BIC variou em torno de 98%, do AICc em torno de 91% e do AIC estabilizou-se em torno de 85%.

Quando os dados simulados tinham médias diferentes e variâncias diferentes, Figura 37, AIC apresentou maiores valores de sensibilidade que AICc e BIC. O valor máximo de sensibilidade apresentado pelo AIC foi de 60%. Os valores para a sensibilidade do BIC estabilizaram-se em torno de 21%; AIC e AICc apresentaram valores crescentes, porém até o valor máximo de 60%.

Figura 36 – Sensibilidade *versus* tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias iguais. Grupo 1.

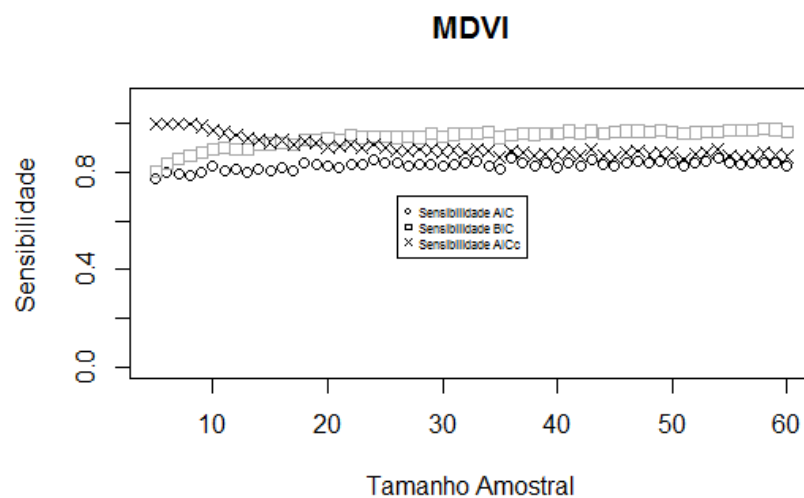
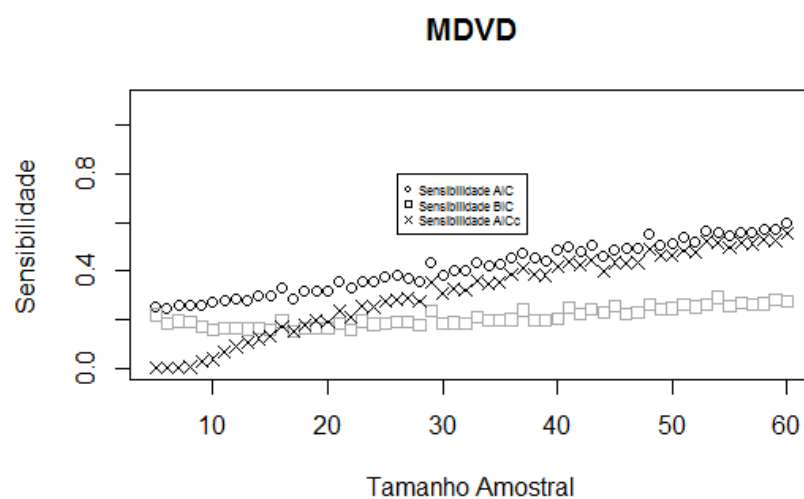


Figura 37 – Sensibilidade *versus* tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias diferentes. Grupo 1.





### Taxas de TP, FP, FN e TN

A Tabela 18 apresenta os valores para as taxas apresentadas nas Figuras 38, 39, 40 e 41 que referem-se às simulações de dados com médias iguais e variâncias iguais (MIVI), médias iguais e variâncias diferentes (MIVD), médias diferentes e variâncias iguais (MDVI) e médias diferentes e variâncias diferentes (MDVD), respectivamente.

Tabela 18 – Valores das taxas de TP, FP, FN e TN para AIC, AICc e BIC para o Caso 3, Grupo 1.

Caso 1		MIVI	MIVD	MDVI	MDVD
AIC	TP	0,66	0,41	0,83	0,42
	FP	0,15	0,07	0,27	0,07
	FN	0,34	0,59	0,17	0,58
	TN	0,85	0,93	0,73	0,93
AICc	TP	0,77	0,35	0,90	0,31
	FP	0,19	0,05	0,28	0,04
	FN	0,23	0,65	0,10	0,69
	TN	0,81	0,95	0,72	0,96
BIC	TP	0,87	0,23	0,94	0,21
	FP	0,24	0,03	0,29	0,02
	FN	0,13	0,77	0,06	0,79
	TN	0,76	0,97	0,71	0,98

Figura 38 – Taxas de TP, FP, TN e FN considerando médias iguais e variâncias iguais. Grupo 1.

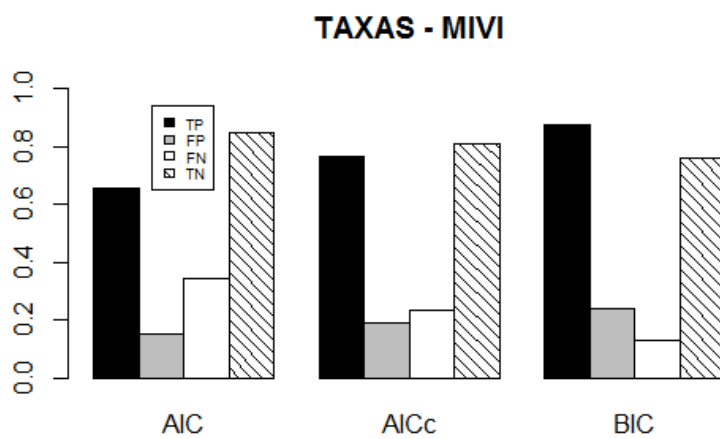


Figura 39 – Taxas de TP, FP, TN e FN considerando médias iguais e variâncias diferentes. Grupo 1.

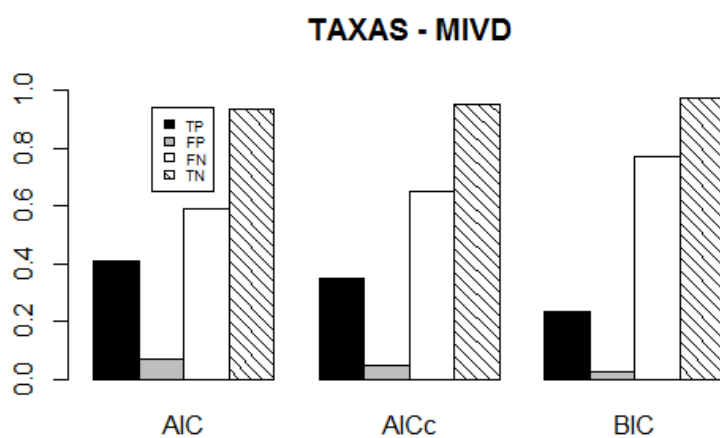


Figura 40 – Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias iguais. Grupo 1.

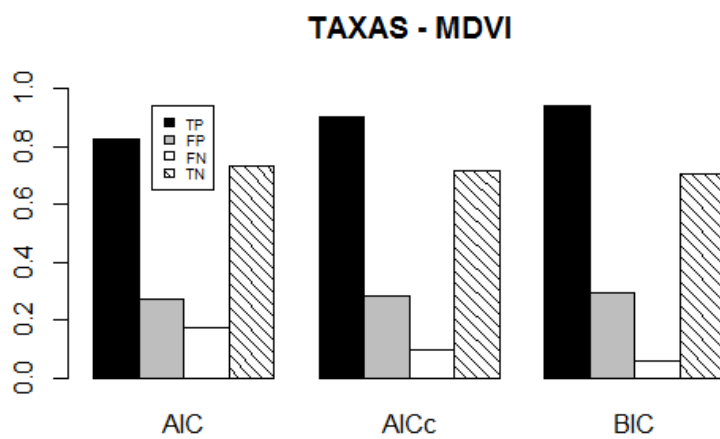
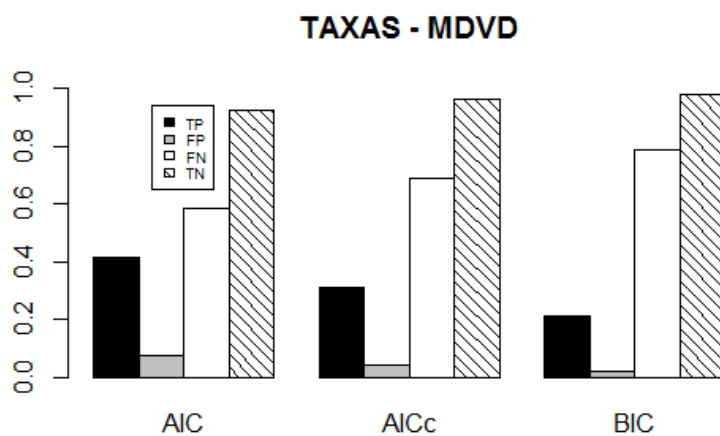


Figura 41 – Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias diferentes. Grupo 1.



## Grupo 2

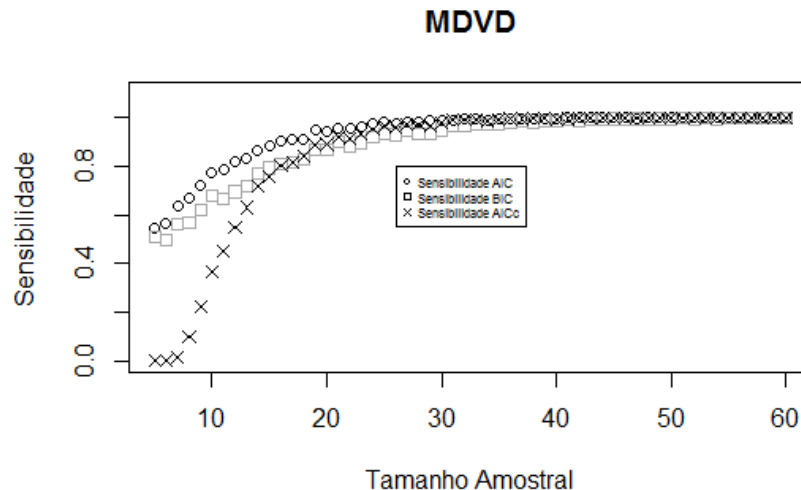
O grupo 2 é aquele em que quando as variâncias são diferentes, elas são distantes.

### Sensibilidade *versus* tamanho amostral

É importante notar que neste grupo, existe uma configuração para a simulação dos dados que já foi considerada no grupo 1, a saber: médias diferentes e variâncias iguais (MDVI), portanto, o gráfico de sensibilidade *versus* tamanho amostral não será apresentado novamente, já que foi apresentado anteriormente na Figura 36.

Quando os dados simulados tinham médias diferentes e variâncias diferentes, Figura 42, o valor da sensibilidade para os três critérios estabilizou-se a partir do tamanho amostral igual a 32 e em 100%.

Figura 42 – Sensibilidade *versus* tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias diferentes. Grupo 2.



### Taxas de TP, FP, FN e TN

A Tabela 19 apresenta os valores para as taxas apresentadas nas Figuras 43, 44, 45 e 46 que referem-se às simulações de dados com médias iguais e

variâncias iguais (MIVI), médias iguais e variâncias diferentes (MIVD), médias diferentes e variâncias iguais (MDVI) e médias diferentes e variâncias diferentes (MDVD), respectivamente.

Tabela 19 – Valores das taxas de TP, FP, FN e TN para AIC, AICc e BIC para o Caso 3, Grupo 2.

Caso 2		MIVI	MIVD	MDVI	MDVD
AIC	TP	0,66	0,95	0,83	0,94
	FP	0,01	0,07	0,07	0,06
	FN	0,34	0,05	0,17	0,06
	TN	0,99	0,93	0,93	0,94
AICc	TP	0,77	0,89	0,90	0,85
	FP	0,04	0,05	0,08	0,03
	FN	0,23	0,11	0,10	0,15
	TN	0,96	0,95	0,92	0,97
BIC	TP	0,87	0,92	0,94	0,90
	FP	0,02	0,03	0,05	0,02
	FN	0,13	0,08	0,06	0,10
	TN	0,98	0,97	0,95	0,98

Figura 43 – Taxas de TP, FP, TN e FN considerando médias iguais e variâncias iguais. Grupo 2.

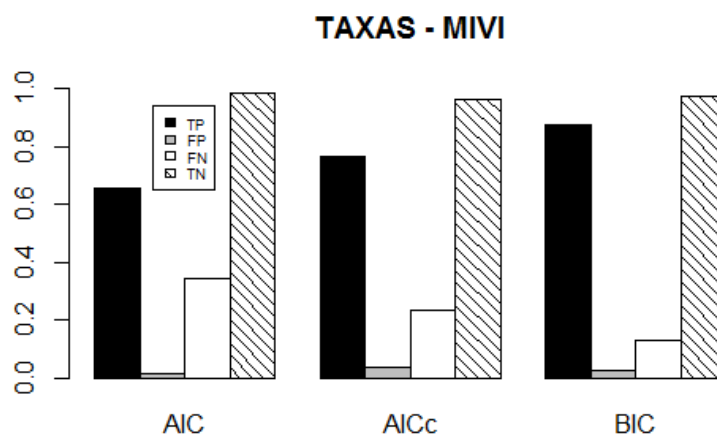


Figura 44 – Taxas de TP, FP, TN e FN considerando médias iguais e variâncias diferentes. Grupo 2.

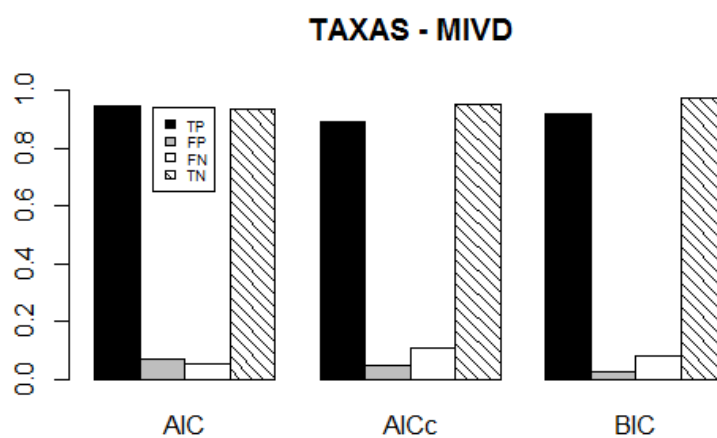


Figura 45 – Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias iguais. Grupo 2.

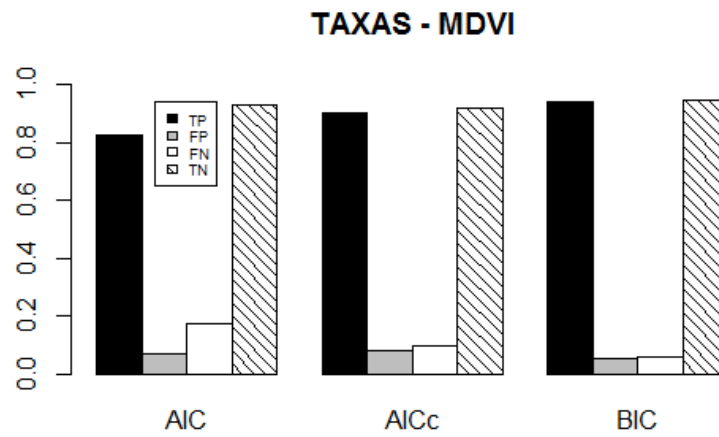
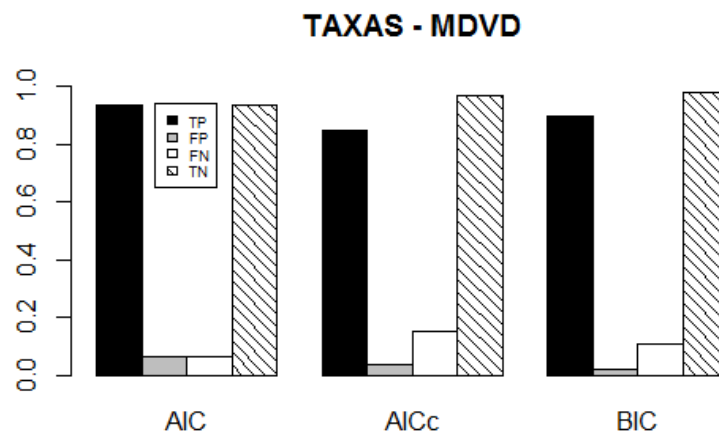


Figura 46 – Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias diferentes. Grupo 2.



## Caso 4

Nessa seção são apresentados os resultados referentes ao caso 4, grupos 1 e 2. É importante salientar que existem três configurações para a simulação dos dados que já foram consideradas no caso 1, a saber: médias iguais e variâncias iguais ( $\mu_1 = \mu_2 = 10$  e  $\sigma_1^2 = \sigma_2^2 = 1$ ) e médias iguais e variâncias diferentes próximas e distantes ( $\mu_1 = \mu_2 = 10$ ;  $\sigma_1^2 = 1$ ;  $\sigma_2^2 = 0.81$  e  $\mu_1 = \mu_2 = 10$ ;  $\sigma_1^2 = 1$ ;  $\sigma_2^2 = 2$ ), portanto, os gráficos de sensibilidade *versus* tamanho amostral não serão apresentados novamente, já que foram apresentadas anteriormente nas Figuras 2 e 5 e 11, na seção 4.1.

### Grupo 1

O grupo 1 é aquele em que quando as variâncias são diferentes, elas são próximas.

#### Sensibilidade *versus* tamanho amostral

Quando os dados simulados tinham médias diferentes e variâncias iguais, Figura 47, BIC apresentou maiores valores de sensibilidade que AICc e AIC a partir do tamanho amostral igual a 16. Os valores de sensibilidade para o BIC estabilizou-se em torno de 98%, para AIC estabilizou-se em torno de 85% e para AICc em torno de 90%.

Quando os dados simulados tinham médias diferentes e variâncias diferentes, Figura 48, AIC e AICc apresentaram valores de sensibilidade crescentes porém menores que 58%. BIC apresentou resultados de sensibilidade estáveis em torno de 21%.



Figura 47 – Sensibilidade *versus* tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias iguais. Grupo 1.

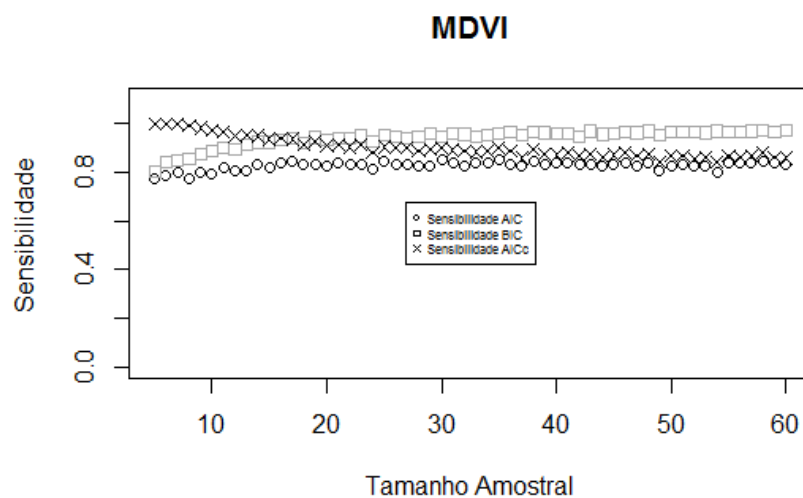
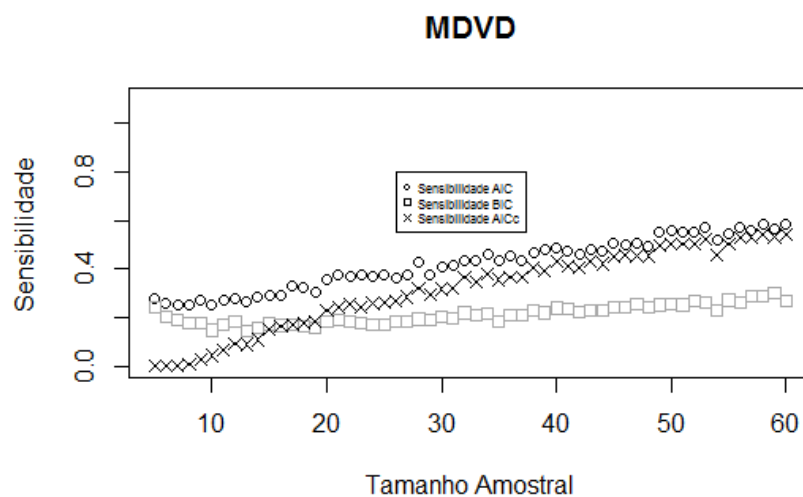


Figura 48 – Sensibilidade *versus* tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias diferentes. Grupo 1.



### Taxas de TP, FP, FN e TN

A Tabela 20 apresenta os valores para as taxas apresentadas nas Figuras 49, 50, 51 e 52 que referem-se às simulações de dados com médias iguais e variâncias iguais (MIVI), médias iguais e variâncias diferentes (MIVD), médias diferentes e variâncias iguais (MDVI) e médias diferentes e variâncias diferentes (MDVD), respectivamente.

Tabela 20 – Valores das taxas de TP, FP, FN e TN para AIC, AICc e BIC para o Caso 4, Grupo 1.

Grupo 1		MIVI	MIVD	MDVI	MDVD
AIC	TP	0,66	0,41	0,83	0,42
	FP	0,15	0,07	0,27	0,07
	FN	0,34	0,59	0,17	0,58
	TN	0,85	0,93	0,73	0,93
AICc	TP	0,77	0,35	0,90	0,32
	FP	0,19	0,05	0,28	0,04
	FN	0,23	0,65	0,10	0,68
	TN	0,81	0,95	0,72	0,96
BIC	TP	0,87	0,23	0,94	0,21
	FP	0,24	0,03	0,29	0,02
	FN	0,13	0,77	0,06	0,79
	TN	0,76	0,97	0,71	0,98

Figura 49 – Taxas de TP, FP, TN e FN considerando médias iguais e variâncias iguais. Grupo 1.

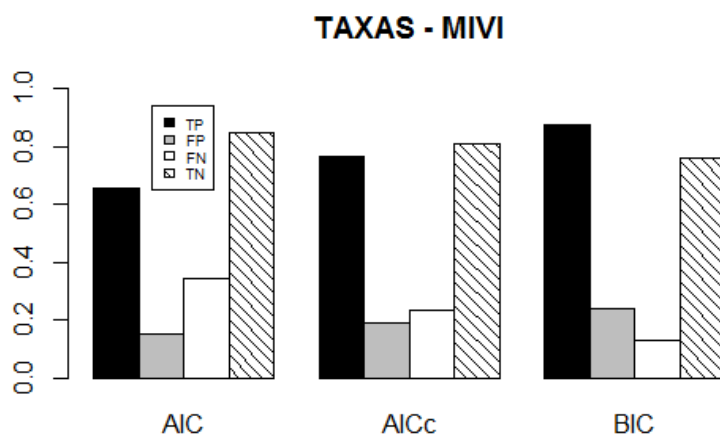


Figura 50 – Taxas de TP, FP, TN e FN considerando médias iguais e variâncias diferentes. Grupo 1.

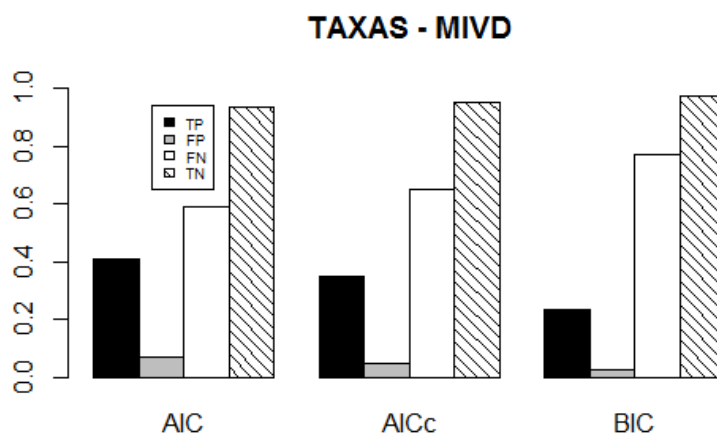


Figura 51 – Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias iguais. Grupo 1.

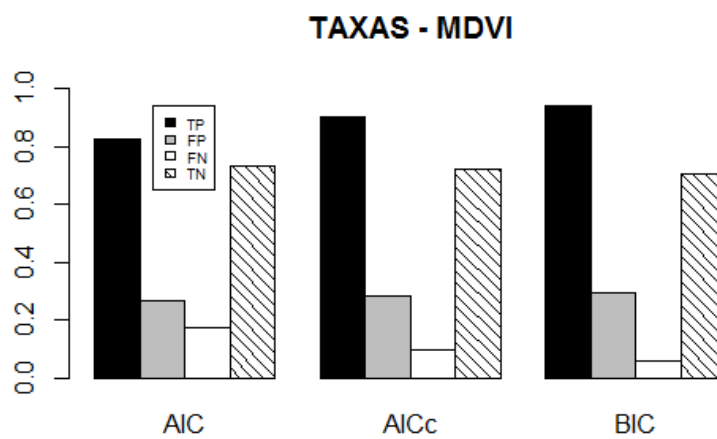
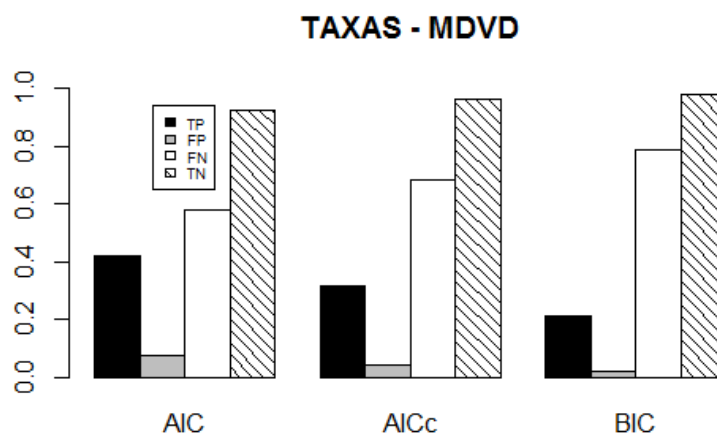


Figura 52 – Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias diferentes. Grupo 1.



## Grupo 2

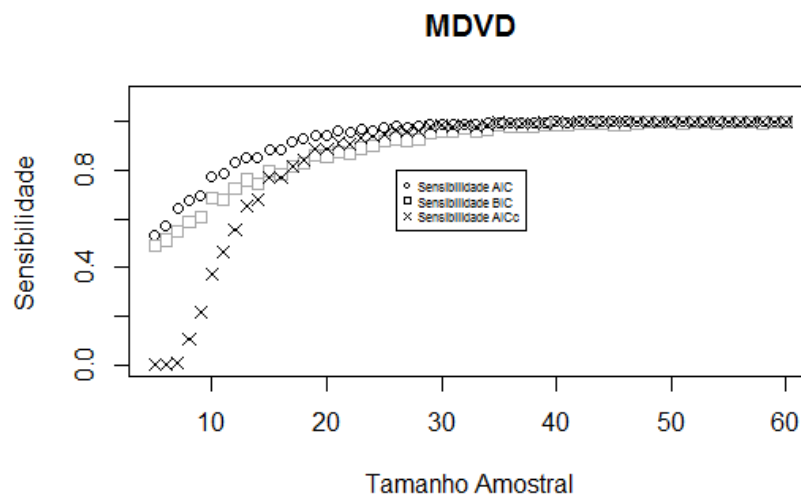
O grupo 2 é aquele em que quando as variâncias são diferentes, elas são distantes.

### Sensibilidade *versus* tamanho amostral

É importante notar que neste grupo, existe uma configuração para a simulação dos dados que já foi considerada no grupo 1, a saber: médias diferentes e variâncias iguais (MDVI), portanto, o gráfico de sensibilidade *versus* tamanho amostral não será apresentado novamente, já que foi apresentado anteriormente na Figura 47.

Quando os dados simulados tinham médias diferentes e variâncias diferentes, Figura 53, os valores de sensibilidade para os três critérios estabilizaram-se a partir do tamanho amostral igual a 36 em 100%.

Figura 53 – Sensibilidade *versus* tamanho amostral: AIC, AICc, BIC. Médias diferentes, variâncias diferentes. Grupo 2.



### Taxas de TP, FP, FN e TN

A Tabela 21 apresenta os valores para as taxas apresentadas nas Figuras 54, 55, 56 e 57 que referem-se às simulações de dados com médias iguais e variâncias iguais (MIVI), médias iguais e variâncias diferentes (MIVD), médias diferentes e variâncias iguais (MDVI) e médias diferentes e variâncias diferentes (MDVD), respectivamente.

Tabela 21 – Valores das taxas de TP, FP, FN e TN para AIC, AICc e BIC para o Caso 4, Grupo 2.

Grupo 2		MIVI	MIVD	MDVI	MDVD
AIC	TP	0,66	0,95	0,83	0,94
	FP	0,01	0,07	0,07	0,06
	FN	0,34	0,05	0,17	0,06
	TN	0,99	0,93	0,93	0,94
AICc	TP	0,77	0,89	0,90	0,85
	FP	0,04	0,05	0,08	0,03
	FN	0,23	0,11	0,10	0,15
	TN	0,96	0,95	0,92	0,97
BIC	TP	0,87	0,92	0,94	0,89
	FP	0,02	0,03	0,05	0,02
	FN	0,13	0,08	0,06	0,11
	TN	0,98	0,97	0,95	0,98

Figura 54 – Taxas de TP, FP, TN e FN considerando médias iguais e variâncias iguais. Grupo 2.

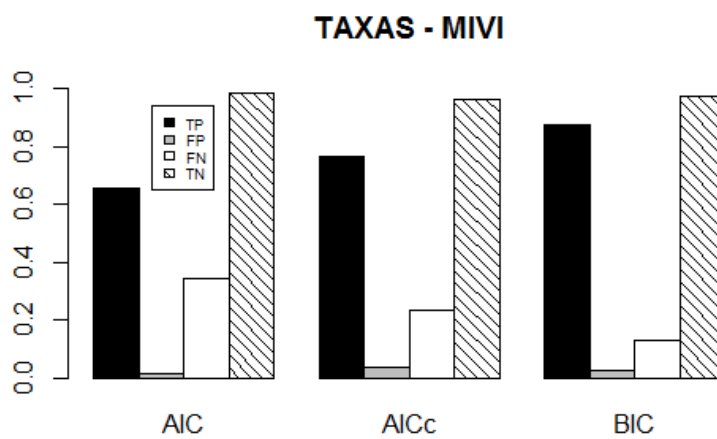


Figura 55 – Taxas de TP, FP, TN e FN considerando médias iguais e variâncias diferentes. Grupo 2.

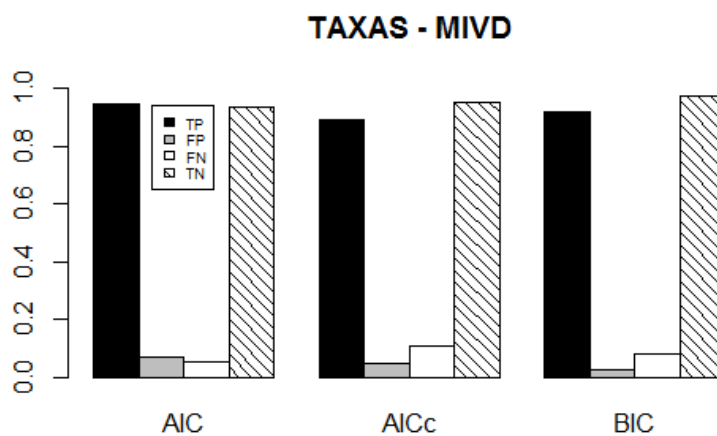


Figura 56 – Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias iguais. Grupo 2.

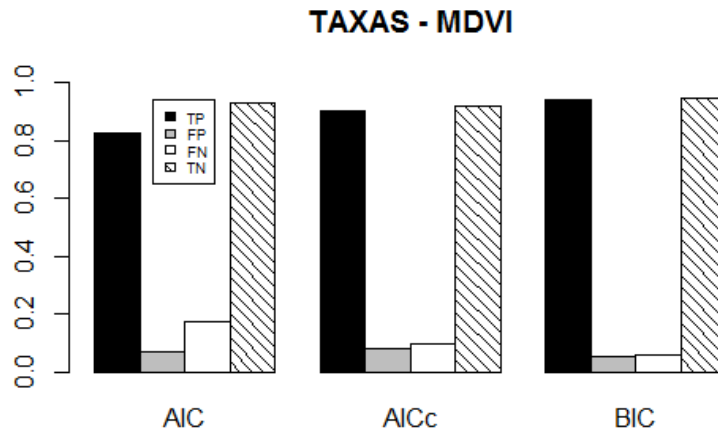


Figura 57 – Taxas de TP, FP, TN e FN considerando médias diferentes e variâncias diferentes. Grupo 2.

