

Rodrigo Citton Padilha dos Reis

**Técnicas Estatísticas para Avaliação de Novos Marcadores
de Risco: Aplicações Envolvendo o Modelo de Cox**

Belo Horizonte, março de 2010

Rodrigo Citton Padilha dos Reis

Técnicas Estatísticas para Avaliação de Novos Marcadores de Risco: Aplicações Envolvendo o Modelo de Cox

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

Orientador: Prof. Dr. Enrico Antônio Colosimo
Co-Orientadora: Prof.^a Dr.^a Maria do Carmo Pereira Nunes
Colaborador: Prof. Dr. Antônio Luiz Pinho Ribeiro

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

Belo Horizonte, março de 2010

*Para Yuri Petri.
Um amigo que se foi
e nos deixa muita saudade.*

Agradecimentos

Ao professor Enrico, meu orientador, que apostou nesta dissertação, pelos ensinamentos que me passou e pela amizade. Também sou grato pelo incentivo e oportunidade que me proporcionou desde a minha chegada no curso de mestrado.

À professora Carminha, minha co-orientadora, por todo incentivo, pelos conhecimentos transmitidos de medicina, pelos conselhos e acima de tudo pela oportunidade de desenvolver com ela esta dissertação.

Ao Programa de Pós-Graduação em Ciências da Saúde: Infectologia e Medicina Tropical. Em especial ao professor Manoel Otávio, pela oportunidade que me deu de trabalhar em conjunto com o programa, ao Tom, colaborador desta dissertação e ao colega e amigo Felipe, pelo apoio e parceria nas tarefas desenvolvidas no grupo.

Aos meus colegas e amigos da Pós-Graduação em Estatística. Em especial à Márcia, ao Gustavo, ao Max e ao Fábio, grandes amigos que me ajudaram a suportar a distância de casa. Aprendi muito com todos neste período.

Ao meu grande amigo Markus, desde os tempos de graduação, um exemplo de persistência e dedicação. Agradeço por todo apoio no momento que cheguei a Belo Horizonte, pelo companheirismo, e pelas vibrações positivas enviadas nestas últimas etapas do curso.

À minha família, meus pais e irmãos, maiores responsáveis e incentivadores desta empreitada. Agradeço por todo amor, carinho, torcida, apoio e por sempre cuidarem de mim. A minha coragem sempre foi, e ainda é inspirada em vocês.

À Laura, minha namorada, por ter compreendido a importância da minha partida. Agradeço também pelas atentas leituras das primeiras versões da dissertação e dicas a respeito das normas. Seu amor, seu companheirismo e sua paciência foram fundamentais para que eu conseguisse realizar esta dissertação.

Aos professores e funcionários do Departamento de Estatística da UFMG pela oportunidade e por terem contribuído na minha formação, aos professores da banca pelas valiosas sugestões e à FAPEMIG pelo apoio financeiro.

À cidade de Belo Horizonte, e ao povo mineiro, que me recebeu muito bem aqui.

Resumo

Novos marcadores de predição de risco vêm sendo propostos na literatura. Assim, novas formulações do risco são investigadas na intenção de incorporar estes marcadores. É importante avaliar a utilidade de um novo marcador para a predição do risco. A curva ROC tem mostrado limitações para a avaliação de tais modelos. Apresentamos as seguintes técnicas estatísticas recentes para avaliação de marcadores e modelos de predição de risco: a tabela de estratificação de risco, a curva de previsibilidade, a melhora da reclassificação líquida (NRI) e a melhora da discriminação integrada (IDI). Simulações de Monte Carlo foram utilizadas para avaliar o desempenho dos testes relacionados ao NRI e IDI. Sob a hipótese nula o teste do NRI apresentou taxas de rejeição mais próximas ao nível nominal do que o teste do IDI. Sob a hipótese alternativa, o teste do IDI apresentou mais poder em relação ao teste do NRI e é menos sensível à variação do tamanho de amostra. Um objetivo importante desta dissertação foi avaliar a inclusão da razão E/E' no desempenho do modelo de predição de risco de morte precoce em pacientes cardiopatas com doença de Chagas. Todas as técnicas, com exceção do IDI, apontaram que a razão E/E' não acrescenta valor preditivo ao modelo.

Abstract

New risk prediction markers have been proposed in the literature. Thus, new formulations of risk are investigated with the intention of incorporating these markers. It is important to evaluate the usefulness of a new marker for risk prediction. The ROC curve has shown limitations in the evaluation of such models. We present in this dissertation the following recent statistical techniques for markers evaluation and models of risk prediction: the risk stratification table, the predictiveness curve, the net reclassification improvement (NRI) and the integrated discrimination improvement (IDI). Monte Carlo simulations were used in order to evaluate the performance of tests related to NRI and IDI. Under the null hypothesis, the NRI test showed rejection rates closer to the nominal level than the IDI one. Under the alternative hypothesis, the IDI test had more power in relation to the NRI test and is less sensitive to the sample size variation. An important aim of this work was to assess the importance of the E/E' ratio in the performance of the model to predict risk of early death in patients with cardiac Chagas disease. All techniques, except for the IDI, pointed out that the E/E' ratio does not add predictive value to the model.

Sumário

Lista de Abreviaturas	vi
Lista de Símbolos	vii
Lista de Figuras	viii
Lista de Tabelas	xv
1 Introdução	1
1.1 Considerações Preliminares	1
1.2 Motivação e Importância do Trabalho	2
1.3 Objetivos	3
1.4 Organização do Trabalho	4
2 Conceitos Básicos e Notação	5
2.1 Introdução	5
2.2 Tempos de Falha e Censura	5
2.3 Funções de Sobrevivência e de Taxa de Falha	6
2.4 Modelo de Cox	7
2.4.1 Método da Máxima Verossimilhança Parcial	8
2.4.2 Estimação das Funções de Base	9
2.5 Considerações Finais sobre a Notação	9
3 Técnicas para Avaliação de Marcadores de Risco	11
3.1 Introdução	11

3.2	Curva ROC	11
3.3	Tabela de Estratificação de Risco	13
3.4	Curva de Previsibilidade	15
3.5	NRI e IDI	17
3.5.1	Melhora da Reclassificação Líquida (NRI)	17
3.5.2	Melhora da Discriminação Integrada (IDI)	18
3.6	Gráfico de Dispersão dos Riscos	20
4	Estudo de Simulação	21
4.1	Introdução	21
4.2	Descrição da Simulação	21
4.3	Descrição dos Cenários Simulados	22
4.4	Resultados	27
4.4.1	Comentários Finais	31
5	Análise de um Estudo de Cardiomiopatia Chagásica	33
5.1	Introdução	33
5.2	Descrição dos Dados e Ajuste do Modelo	33
5.3	Comparação dos Modelos	34
5.4	Análise de Sensibilidade do NRI	37
6	Conclusões	39
6.1	Considerações Finais	39
A	Gráficos do Estudo de Simulação	41
	Referências Bibliográficas	61

Lista de Abreviaturas

AUC	Área sob a curva (<i>Area under the curve</i>).
FPF	Fração de falsos positivos (<i>False positive fraction</i>).
IDI	Melhora da discriminação integrada (<i>Integrated discrimination improvement</i>).
IFPF	Integral da fração de falsos positivos.
ITPF	Integral da fração de verdadeiros positivos.
HR	Razão de riscos (<i>Hazard ratio</i>).
NRI	Melhora da reclassificação líquida (<i>Net reclassification improvement</i>).
ROC	Característica operadora receptora (<i>Receiver operating characteristic</i>).
TPF	Fração de verdadeiros positivos (<i>True positive fraction</i>).

Lista de Símbolos

p_L	Limiar de baixo risco.
p_H	Limiar de alto risco.
π	Risco total.

Lista de Figuras

3.1	Exemplo de uma curva ROC.	12
3.2	Exemplo de uma curva de previsibilidade.	16
3.3	Exemplo do gráfico de dispersão dos riscos estimados.	20
4.1	Gráfico de dispersão entre riscos para o Cenário 1; $n = 5.000$	23
4.2	Gráfico de dispersão entre riscos para o Cenário 2; $n = 5.000$	24
4.3	Gráfico de dispersão entre riscos para o Cenário 3; $n = 5.000$	24
4.4	Gráfico de dispersão entre riscos para o Cenário 4; $n = 5.000$	24
4.5	Gráfico de dispersão entre riscos para o Cenário 5; $n = 5.000$	25
4.6	Gráfico de dispersão entre riscos para o Cenário 6; $n = 5.000$	25
4.7	Gráfico de dispersão entre riscos para o Cenário 7; $n = 5.000$	26
4.8	Gráfico de dispersão entre riscos para o Cenário 8; $n = 5.000$	26
4.9	Gráfico de dispersão entre riscos para o Cenário 9; $n = 5.000$	26
4.10	Gráfico de dispersão entre riscos para o Cenário 10; $n = 5.000$	27
4.11	Histograma de 3.000 réplicas das estatísticas de teste para o Cenário 1 : (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.	28
5.1	Gráfico de dispersão entre riscos para o estudo de cardiomiopatia chagásica.	35
5.2	Curva ROC dos riscos estimados dos modelos sem e com E/E' para o estudo de cardiomiopatia chagásica.	35
5.3	Curva de previsibilidade dos riscos estimados dos modelos sem e com E/E' para o estudo de cardiomiopatia chagásica.	37

- A.1 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 1**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 41
- A.2 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 1**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 41
- A.3 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 1**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 42
- A.4 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 1**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 42
- A.5 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 1**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 42
- A.6 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 2**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$ 43
- A.7 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 2**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$ 43
- A.8 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 2**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$ 43
- A.9 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 2**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 44
- A.10 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 2**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 44
- A.11 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 2**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 44
- A.12 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 3**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$ 45

- A.13 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 3**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 45
- A.14 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 3**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 45
- A.15 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 3**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 46
- A.16 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 3**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 46
- A.17 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 3**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 46
- A.18 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 4**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$ 47
- A.19 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 4**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 47
- A.20 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 4**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 47
- A.21 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 4**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 48
- A.22 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 4**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 48
- A.23 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 4**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 48
- A.24 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 5**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 49

- A.25 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 5**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 49
- A.26 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 5**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 49
- A.27 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 5**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 50
- A.28 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 5**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 50
- A.29 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 5**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 50
- A.30 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 6**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 51
- A.31 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 6**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 51
- A.32 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 6**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 51
- A.33 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 6**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 52
- A.34 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 6**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 52
- A.35 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 6**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 52

- A.36 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 7**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 53
- A.37 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 7**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 53
- A.38 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 7**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 53
- A.39 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 7**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 54
- A.40 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 7**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 54
- A.41 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 7**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 54
- A.42 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 8**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 55
- A.43 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 8**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 55
- A.44 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 8**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 55
- A.45 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 8**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 56
- A.46 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 8**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 56

- A.47 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 8**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 56
- A.48 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 9**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 57
- A.49 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 9**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 57
- A.50 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 9**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 57
- A.51 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 9**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 58
- A.52 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 9**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 58
- A.53 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 9**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 58
- A.54 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 10**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 59
- A.55 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 10**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 59
- A.56 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 10**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 59
- A.57 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 10**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 60

- A.58 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 10**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 60
- A.59 Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 10**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão. 60

Lista de Tabelas

3.1	Exemplo da tabela de estratificação: Risco de morte em até um ano pelos modelos sem e com E/E'.	14
4.1	Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o Cenário 1	28
4.2	Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o Cenário 2	29
4.3	Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o Cenário 3	29
4.4	Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o Cenário 4	29
4.5	Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o Cenário 5	30
4.6	Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o Cenário 6	30
4.7	Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o Cenário 7	30
4.8	Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o Cenário 8	31
4.9	Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o Cenário 9	31
4.10	Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o Cenário 10	31
5.1	Resultados do modelo de Cox para os modelos sem e com E/E' do estudo de cardiomiopatia chagásica.	34

5.2	Tabela de estratificação: Risco de morte em um até um ano pelos modelos sem e com E/E'.	36
5.3	Resultados da análise de sensibilidade para o NRI.	38

Capítulo 1

Introdução

1.1 Considerações Preliminares

Um marcador de predição de risco é qualquer medida obtida de um paciente que é usada para calcular a probabilidade de se ter uma doença ou ocorrer algum evento adverso (1,2), geralmente num período de tempo estipulado (por exemplo, o aparecimento de um câncer em até cinco anos, morte em dez anos, entre outros). Chamaremos esta probabilidade de “risco”. Este marcador pode ser uma medida quantitativa ou qualitativa, tal como: a idade do indivíduo, o sexo, níveis de colesterol, pressão sistólica, entre outros. Modelos estatísticos, como a regressão logística e de Cox, têm sido utilizados para combinar as informações de marcadores e assim predizer o risco de um evento ocorrer para cada indivíduo. Neste contexto de prevenção, estes modelos têm o papel de classificar os indivíduos em categorias de risco clinicamente relevantes. Assim, decisões médicas são tomadas em relação a tratamentos ou intervenções preventivas de acordo com a classificação de cada paciente. Por exemplo, pacientes com cardiomiopatia chagásica classificados como alto risco de morte precoce (morte em até um ano após o diagnóstico) devem receber o tratamento diferenciado, ou serem encaminhados para o transplante. Já os pacientes classificados como baixo risco de morte precoce devem ser monitorados, porém sem a necessidade de transplante.

A contínua descoberta de novos marcadores pelos cientistas apresenta oportunidades e desafios para estatísticos e clínicos para avaliar esses marcadores e desenvolver novas formulações de risco que os incorporem (3). A questão principal é avaliar a utilidade de um novo marcador para a predição do risco. Necessariamente um marcador deve ser estatisticamente associado a resposta. Porém, significância estatística não implica em significância clínica ou melhora no desempenho do modelo (3). Desta forma, medidas de ajustamento do modelo, tais como a estatística da razão de verossimilhanças, o critério de informação de Bayes (BIC), o critério de informação de Akaike (AIC), entre outros, não devem ser consideradas como um “ponto de chegada” na avaliação de modelos de predição de risco.

O desempenho de um modelo de predição de risco pode ser avaliado através do quanto o risco calculado do modelo pode refletir a fração de indivíduos da população com eventos (calibração

do modelo); as proporções em que a população é estratificada em categorias de risco clinicamente significativas (capacidade de estratificação do modelo); e o grau com que indivíduos com eventos são designados para categorias de alto risco e aqueles sem eventos são designados para categorias de baixo risco (acurácia da classificação do modelo) (1). A utilidade de um marcador para a predição de risco deve ser avaliada através da melhora (ou piora) destas características do modelo.

Modelos de predição de risco são frequentemente avaliados através da análise da curva ROC (*receiver operating characteristic*) e sua principal estatística resumo, a área sob a curva (AUC) (4). A AUC é uma estatística que pode ser interpretada como a probabilidade do risco de um indivíduo com evento ser maior que o risco de um indivíduo sem evento. Com o acréscimo de um novo marcador no modelo de predição de risco, pesquisadores avaliam a melhora na predição do risco através do aumento da AUC. Porém, estudos mostraram que para um aumento significativo na AUC, é necessário uma associação muito grande do marcador com a resposta (5,6). Raramente tais marcadores atingem os níveis de associação mencionados nestes estudos, demonstrando uma limitação da análise através da curva ROC. Apesar do seu importante papel na avaliação de testes diagnósticos, a análise da curva ROC como método de avaliação de modelos de predição de risco vem sendo fortemente criticada (3,6–8), e uma das principais críticas está no fato de que a AUC não tem uma interpretação clínica relevante.

Em resposta a estas críticas, técnicas alternativas para avaliação de modelos de predição de risco foram propostas. Cook, Buring e Ridker (6) apresentaram a “tabela de reclassificação” para avaliar o valor adicional de um novo marcador de risco quando combinado a um conjunto de marcadores de risco, e posteriormente Cook (7) formalizou esta proposta. Este método estratifica o risco em categorias de relevância clínica e compara o modelo sem o novo marcador com o modelo com o novo marcador através de uma tabela de dupla entrada, observando o percentual de indivíduos reclassificados. Janes, Pepe e Gu (1) exploraram a ideia da tabela de reclassificação, incrementando-a, e assim extraindo informações mais relevantes. Esses autores passaram a chamar esta técnica de “tabela de estratificação de risco”. Pepe *et al.* (8,9) também propuseram uma nova abordagem para avaliar a utilidade de um novo marcador, a “curva de previsibilidade”. Esta técnica não depende da especificação de categorias de risco. Recentemente, Pencina *et al.* (3) apresentaram duas novas medidas resumo para avaliação de um novo marcador de risco. Uma medida é baseada na reclassificação dos indivíduos em categorias de risco clinicamente relevantes, a “melhora da reclassificação líquida” (NRI), e a outra é uma proposta de generalização da primeira, em um contexto que categorias de risco não podem ser (ou não necessitam ser) estipuladas, a “melhora da discriminação integrada” (IDI). Os autores de (3) ainda propuseram testes assintóticos para verificar a significância estatística destas medidas.

1.2 Motivação e Importância do Trabalho

A doença de Chagas, causada pelo parasita protozoário *Trypanosoma cruzi*, infecta cerca de 18 milhões de pessoas na América Latina (10). A cardiomiopatia crônica é a mais importante

e severa manifestação da doença de Chagas humana, eventualmente afetando de 20 a 30% dos indivíduos na fase crônica da doença e, neste grupo, causando falha cardíaca, arritmias, paradas cardíacas, tromboembolismo, derrame e morte súbita (10,11). Além disso, o tratamento da doença de Chagas é custoso e frequentemente afeta os pacientes nos seus anos mais produtivos de trabalho. A identificação de fatores que possam prever o alto risco de eventos cardíacos deve resultar em uma maior gestão médica intensiva ou até um transplante de coração, levando à melhora na sobrevivência do paciente (12).

O ecocardiograma é um exame não invasivo, que permite ao médico obter informações a respeito do coração do paciente. Com o avanço das tecnologias (em particular as tecnologias de imagem), novos marcadores vêm sendo propostos por pesquisadores da área médica (cardiologia). Este trabalho foi motivado por estudos do grupo de pesquisa Ecocardiografia nas Doenças Tropicais do Programa de Pós-Graduação de Infectologia e Medicina Tropical da Universidade Federal de Minas Gerais (UFMG). Este grupo vem desenvolvendo pesquisas para incorporar novos marcadores, desenvolvidos a partir do exame do ecocardiograma, aos modelos de predição de risco de morte precoce em pacientes com doença de Chagas (12,13), e portanto, a avaliação adequada destes modelos deve ser feita. Em particular, avaliamos se a inclusão de um destes marcadores, a razão E/E' , melhora o desempenho do modelo de predição de risco de morte precoce de um grupo de pacientes com cardiomiopatia chagásica, diagnosticados e acompanhados pelo Hospital das Clínicas da UFMG. A razão E/E' é um marcador acurado das pressões de enchimento (14), com valor prognóstico estabelecido em várias patologias cardíacas, mas não determinado na cardiopatia chagásica. Esta avaliação foi realizada utilizando as técnicas citadas na Seção 1.1. Como modelo de predição de risco, daremos enfoque ao modelo de Cox pela natureza de nossos dados, porém salientamos desde já que outros modelos podem ser utilizados para a predição do risco.

No âmbito da tomada de decisões médicas, a predição do risco (probabilidade de ocorrência de um evento de interesse) vem ocupando um importante papel e, conseqüentemente, sua análise adequada é uma área muito ativa de investigação estatística atualmente (1–3,6–9). Tendo em vista que as propostas citadas anteriormente são recentes, e que há certa discordância quando utilizadas para um mesmo banco de dados (7,15), vemos a necessidade de aprofundar o estudo destas técnicas.

1.3 Objetivos

Esta dissertação tem como objetivos:

1. Apresentar as técnicas de avaliação de modelos de predição de risco, em particular as técnicas que avaliam a utilidade de um novo marcador de risco quando combinado com um conjunto prévio de marcadores de risco (a análise da curva ROC, a tabela de estratificação de risco, a curva de previsibilidade, o NRI e o IDI);
2. Realizar simulações para comparar os testes de hipóteses relacionados ao NRI e IDI;

3. Realizar a análise dos dados referentes aos pacientes com cardiomiopatia chagásica do Hospital das Clínicas da UFMG e avaliar se a inclusão da razão E/E' melhora o desempenho do modelo de predição de risco de morte precoce destes pacientes;
4. Realizar uma análise de sensibilidade para o NRI com os dados da análise.

1.4 Organização do Trabalho

A presente dissertação está dividida em seis capítulos e um apêndice. No Capítulo 2 apresentamos alguns conceitos básicos da análise de sobrevivência para estabelecer uma terminologia e notação. Também consideramos os métodos de estimação para os parâmetros e funções relacionadas ao modelo de Cox. No Capítulo 3 expomos algumas técnicas para avaliação de marcadores e modelos de predição de risco. Consideramos a análise da curva ROC, a tabela de estratificação de risco, a curva de previsibilidade, a melhora da reclassificação líquida (NRI), a melhora da discriminação integrada (IDI) e o gráfico de dispersão dos riscos. No Capítulo 4 apresentamos os resultados do estudo de simulação que fizemos para avaliação dos testes de hipóteses relacionados ao NRI e IDI. Comparamos as taxas de rejeição dos testes relacionados a estas duas estatísticas para diferentes cenários. No Capítulo 5 fazemos a análise de um banco de dados de pacientes com cardiomiopatia chagásica. Avaliamos se o novo marcador, a razão E/E' , acrescenta valor preditivo ao modelo de predição de risco de morte. Por fim, no Capítulo 6 apresentamos as conclusões obtidas com este trabalho e sugerimos tópicos de pesquisa que podem vir a ser desenvolvidos a partir desta dissertação.

No Apêndice A apresentamos alguns gráficos que ilustram os cenários simulados no Capítulo 4, e histogramas dos resultados das simulações.

Capítulo 2

Conceitos Básicos e Notação

2.1 Introdução

O objetivo principal das técnicas que abordamos nesta dissertação é avaliar se a inclusão de um novo marcador melhora o desempenho do modelo de predição de risco. Dado um vetor de covariáveis \mathbf{x} observadas para um indivíduo, o risco é definido como a probabilidade condicional a este vetor de covariáveis de ocorrer o evento de interesse no período de tempo especificado:

$$risco(\mathbf{x}) = P(\text{“evento ocorre até o tempo } t\text{”} \mid \mathbf{x}). \quad (2.1)$$

É necessário um método para estimarmos esta probabilidade. Veremos mais adiante que em um contexto de análise de sobrevivência esta probabilidade é uma função do tempo de falha, e que o risco é função da função de sobrevivência. Neste capítulo apresentamos uma revisão de conceitos básicos da análise de sobrevivência para estabelecer uma terminologia e notação. Daremos enfoque ao modelo de Cox, porém ressaltamos que outros modelos podem ser empregados para a estimação do risco, inclusive modelos que não levam em conta o tempo até a ocorrência do evento (modelos com resposta binária, por exemplo, o modelo de regressão logística). Na Seção 2.2 introduzimos os conceitos de tempo de falha e censura. Na Seção 2.3 são abordadas as funções relacionadas ao tempo de falha. Na Seção 2.4 apresentamos o modelo de Cox, o método de estimação dos parâmetros do modelo e a estimação de funções do modelo. Por fim, na Seção 2.5 fazemos algumas considerações finais a respeito da notação.

2.2 Tempos de Falha e Censura

Na análise de sobrevivência, o interesse centra-se em um grupo ou grupos de indivíduos para os quais existe um evento definido, frequentemente chamado de falha, que ocorre após um período de tempo chamado de tempo de falha (16). Exemplos de tempos de falha na área médica podem ser o tempo até a morte de pacientes, o tempo até a ocorrência de um infarto, ou um acidente vascular cerebral, entre outros. A análise de tempos de falha também se estende a outras áreas como engenharia, economia, física e outras.

Em alguns casos, estamos apenas preocupados com o comportamento destes tempos de falha para algum grupo de indivíduos. Embora existam métodos bem conhecidos para estimar as distribuições dos tempos de falha, geralmente estamos interessados em examinar a relação entre a sobrevivência e um ou mais preditores, usualmente denominados covariáveis.

Uma característica especial dos estudos de sobrevivência é a presença de censuras. Para alguns indivíduos pode não ser possível observar o evento de interesse durante o período de observação, ou seja, estes dados são censurados. Para estes indivíduos a única informação que temos é que o tempo de falha deve ser maior do que o tempo de observação.

Três tipos de censura são encontradas na literatura (17):

- Censura do tipo I: ocorre em estudos onde o período de observação é pré-estabelecido;
- Censura do tipo II: ocorre em estudos que se encerram após um determinado número de falhas ocorrer;
- Censura aleatória: ocorre quando o indivíduo é retirado no decorrer do estudo sem ter ocorrido a falha, ou no caso de ocorrer a falha por um motivo diferente do estudado.

A censura aleatória pode ser representada considerando uma variável aleatória T (tempo de falha) e C (tempo de censura), uma segunda variável aleatória independente de T . Então a observação consiste de $t = \min(T, C)$, junto com a variável indicadora $\delta = 1$ se $T \leq C$ (falha), $\delta = 0$ se $T > C$ (censura). Nesta dissertação abordamos apenas o caso de dados com censura aleatória, que é o caso que ocorre na área médica.

Na análise de sobrevivência a variável resposta observada é então representada pelo par (t_i, δ_i) sendo t_i o tempo de falha ou de censura e δ_i a variável indicadora de falha ou censura (17). Aqui, i é o índice para observação, e temos que $i = 1, \dots, n$, em que n é o número total de observações.

2.3 Funções de Sobrevivência e de Taxa de Falha

Seja T um tempo de falha. Consideramos T como uma variável aleatória não-negativa com função de distribuição acumulada $F(t) = P(T \leq t)$ e densidade $f(t) = dF(t)/dt$. A função de sobrevivência $S(t)$ é o complementar da função de distribuição, $S(t) = P(T > t) = 1 - F(t)$. Outra representação dos tempos de sobrevivência é a função de taxa de falha $\lambda(t)$, que representa a taxa de falha instantânea no tempo t condicional à sobrevivência até o tempo t (17):

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}. \quad (2.2)$$

A função de taxa de falha também é conhecida como função de risco, porém para que não haja confusão entre $\lambda(t)$ e a Expressão (2.1), continuaremos a chamá-la de função de taxa de falha. A

função de taxa de falha acumulada é definida como $\Lambda(t) = \int_0^t \lambda(u)du$. Uma relação importante entre a função de sobrevivência e a função de taxa de falha acumulada é:

$$S(t) = \exp\{-\Lambda(t)\}.$$

2.4 Modelo de Cox

O modelo de regressão de Cox (18) examina a relação entre a função de taxa de falha $\lambda(t)$ e um vetor de p covariáveis, $\mathbf{x} = (x_1, \dots, x_p)'$. A função de taxa de falha é modelada como:

$$\lambda_i(t) = \lambda_0(t)g(\mathbf{x}'_i\boldsymbol{\beta}). \quad (2.3)$$

Aqui, $\lambda_0(t)$ é a função de taxa de falha para um indivíduo quando $\mathbf{x} = \mathbf{0}$, também conhecida como função de taxa de falha de base. A função g é não-negativa e requer $g(0) = 1$. O mais comum, e que utilizamos nesta dissertação, é especificar g como:

$$g(\mathbf{x}'\boldsymbol{\beta}) = \exp\{\mathbf{x}'\boldsymbol{\beta}\},$$

em que $\boldsymbol{\beta}$ é o vetor de parâmetros a ser estimado. Desta forma o modelo de Cox é especificado como:

$$\begin{aligned} \lambda_i(t) &= \lambda_0(t) \exp\{\mathbf{x}'_i\boldsymbol{\beta}\} \\ &= \lambda_0(t) \exp\{\beta_1 x_{i1} + \dots + \beta_p x_{ip}\}. \end{aligned} \quad (2.4)$$

Este modelo é *semiparamétrico*, pois enquanto a função de taxa de falha de base pode assumir qualquer forma, as covariáveis entram no modelo linearmente. O modelo de Cox também é denominado modelo de riscos proporcionais, pois a razão das taxas de falha de dois indivíduos diferentes é constante no tempo (17). Ou seja, para dois indivíduos i e i' que diferem em seus valores de \mathbf{x} , com os correspondentes preditores lineares

$$\mathbf{x}'_i\boldsymbol{\beta} = \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

e

$$\mathbf{x}'_{i'}\boldsymbol{\beta} = \beta_1 x_{i'1} + \dots + \beta_p x_{i'p},$$

a razão das funções de taxa de falha para os indivíduos i e i' é

$$\frac{\lambda_i(t)}{\lambda_{i'}(t)} = \frac{\lambda_0(t) \exp\{\mathbf{x}'_i\boldsymbol{\beta}\}}{\lambda_0(t) \exp\{\mathbf{x}'_{i'}\boldsymbol{\beta}\}} = \exp\{\mathbf{x}'_i\boldsymbol{\beta} - \mathbf{x}'_{i'}\boldsymbol{\beta}\},$$

que não depende do tempo. Esta propriedade deve ser utilizada para a interpretação dos coeficientes. Tomando-se a mesma razão das taxas de falha destes indivíduos, porém agora, i e i' têm

os mesmos valores para as covariáveis com exceção da l -ésima, teremos:

$$\frac{\lambda_i(t)}{\lambda_{i'}(t)} = \exp\{\beta_l(x_{il} - x_{i'l})\},$$

que pode ser interpretado como a razão de riscos (*hazard ratio*, HR).

2.4.1 Método da Máxima Verossimilhança Parcial

Consideremos agora inferência a respeito do vetor β quando a função de taxa de falha de base $\lambda_0(t)$ é completamente desconhecida. Cox em seu artigo original (18) sugere uma verossimilhança condicional para a estimação dos parâmetros do modelo, e posteriormente reconhece que esta verossimilhança é uma verossimilhança parcial (19). O método da máxima verossimilhança parcial assume que em uma amostra de n indivíduos, k falhas ocorram em tempos distintos $t_1 < \dots < t_k$, e contínuos. Denotemos $R(t_i)$ o conjunto dos índices dos indivíduos sob risco em t_i , ou seja, o conjunto dos índices de indivíduos que não falharam ou foram censurados até o tempo t_i . Assim, a probabilidade do i -ésimo indivíduo falhar no tempo t_i condicional a história de falhas e censuras até o tempo t_i é:

$$\frac{\lambda_i(t|\mathbf{x}_i)}{\sum_{j \in R(t_i)} \lambda_j(t|\mathbf{x}_j)} = \frac{\exp\{\mathbf{x}'_i \beta\}}{\sum_{j \in R(t_i)} \exp\{\mathbf{x}'_j \beta\}}. \quad (2.5)$$

Seja \mathbf{x}_j o valor de \mathbf{x} para o j -ésimo indivíduo e \mathbf{x}_i o valor de \mathbf{x} para o indivíduo cujo tempo de falha é t_i . Temos então a seguinte função de verossimilhança parcial

$$L(\beta) = \prod_{i=1}^k \frac{\exp\{\mathbf{x}'_i \beta\}}{\sum_{j \in R(t_i)} \exp\{\mathbf{x}'_j \beta\}}. \quad (2.6)$$

Os valores de β que maximizam a função de verossimilhança (2.6) são os estimadores de máxima verossimilhança parcial dos parâmetros do modelo. Quando ocorrem empates entre os tempos de falha, temos problemas na estimação de β , e aproximações para a função de verossimilhança parcial devem ser feitas. A aproximação proposta por Breslow (20) é a que utilizamos nesta dissertação. Seja \mathbf{s}_i o vetor formado pela soma das correspondentes p covariáveis para os indivíduos que falham no mesmo tempo t_i ($i = 1, \dots, k$) e d_i o número de falhas neste mesmo tempo, então temos a seguinte função de verossimilhança parcial (17):

$$L(\beta) = \prod_{i=1}^k \frac{\exp\{\mathbf{s}'_i \beta\}}{\left[\sum_{j \in R(t_i)} \exp\{\mathbf{s}'_j \beta\} \right]^{d_i}}. \quad (2.7)$$

A Expressão (2.7) se reduz a (2.6) quando não houver empates (17). Os estimadores de máxima verossimilhança parcial possuem as propriedades assintóticas, como consistência e normalidade,

necessárias para construção de intervalos de confiança e testes de hipóteses sobre os coeficientes do modelo (17).

2.4.2 Estimação das Funções de Base

O modelo de Cox possui o termo não especificado $\lambda_0(t)$, que é a função de taxa de falha de base. Associada a esta função ainda temos a função de taxa de falha acumulada de base $\Lambda_0(t) = \int_0^t \lambda_0(u)du$, e a correspondente função de sobrevivência de base $S_0(t) = \exp\{-\Lambda_0(t)\}$.

Na maior parte dos estudos de sobrevivência a atenção está voltada para estimação dos parâmetros do modelo. Porém, quando estamos interessados em concluir nossas análises em termos da função de sobrevivência de indivíduos, a estimação das funções de base se tornam essenciais. Temos interesse de estimar a função de sobrevivência de um indivíduo i em um tempo especificado t . Temos que:

$$\begin{aligned} S_i(t) &= \exp\left\{-\int_0^t \lambda_0(u) \exp\{\mathbf{x}'_i \boldsymbol{\beta}\} du\right\} \\ &= \exp\left\{-\exp\{\mathbf{x}'_i \boldsymbol{\beta}\} \log(S_0(t))\right\} \\ &= [S_0(t)]^{\exp\{\mathbf{x}'_i \boldsymbol{\beta}\}}. \end{aligned} \tag{2.8}$$

Cox e Oakes (16) apresentam uma maneira não-paramétrica de estimar as funções de base do modelo de Cox. Sejam d_j o número de falhas em t_j e $\hat{\boldsymbol{\beta}}$ os estimadores de máxima verossimilhança parcial de $\boldsymbol{\beta}$, temos:

$$\hat{\Lambda}_0(t) = \sum_{j:t_j < t} \frac{d_j}{\sum_{l \in R(t_j)} \exp\{\mathbf{x}'_l \hat{\boldsymbol{\beta}}\}}. \tag{2.9}$$

A função de sobrevivência de base pode ser estimada, a partir de (2.9), por $\hat{S}_0(t) = \exp\{-\hat{\Lambda}_0(t)\}$. Combinando este estimador com os estimadores dos parâmetros do modelo, temos que para o i -ésimo indivíduo do estudo a sua função de sobrevivência é estimada por:

$$\hat{S}_i(t) = [\hat{S}_0(t)]^{\exp\{\mathbf{x}'_i \hat{\boldsymbol{\beta}}\}}. \tag{2.10}$$

2.5 Considerações Finais sobre a Notação

Começamos este capítulo definindo o risco como uma probabilidade de ocorrer um evento de interesse em um período de tempo especificado. Agora, atualizamos nossa definição da Expressão (2.1) pela seguinte expressão:

$$\begin{aligned} \text{risco}(\mathbf{x}_i) &= P(i\text{-ésimo indivíduo falhar até } t | \mathbf{x}_i) \\ &= 1 - S_i(t), \end{aligned} \tag{2.11}$$

em que t é um valor fixo ou pré-especificado.

Quando o modelo de Cox é utilizado, de (2.11) temos:

$$risco(\mathbf{x}_i) = 1 - [S_0(t)]^{\exp\{\mathbf{x}'_i\boldsymbol{\beta}\}}, \quad (2.12)$$

e portando,

$$r\hat{isco}(\mathbf{x}_i) = 1 - [\hat{S}_0(t)]^{\exp\{\mathbf{x}'_i\hat{\boldsymbol{\beta}}\}}. \quad (2.13)$$

A Expressão (2.13) é o estimador que utilizamos para o risco nesta dissertação em nossas análises. Para simplificar a notação utilizamos $\hat{r}_i = r\hat{isco}(\mathbf{x}_i)$, e entende-se por \hat{r}_i o risco estimado para o i -ésimo indivíduo no tempo t . Mais uma vez, lembramos que quando outros modelos estatísticos são utilizados para a predição do risco, outras formulações do risco são derivadas a partir do modelo escolhido.

Quando avaliamos o risco predito em uma amostra, em algumas situações é de interesse estipular limiares que estratificam o risco em categorias. Nesta dissertação usamos dois limiares de risco: p_L , o limiar de baixo risco e p_H , o limiar de alto risco. Assim temos três categorias de risco:

- categoria de baixo risco, se $0 \leq \hat{r} < p_L$;
- categoria intermediária, se $p_L \leq \hat{r} < p_H$;
- categoria de alto risco, se $p_H \leq \hat{r} \leq 1$.

As definições de p_L e p_H são inteiramente clínicas, ou seja, são obtidas a partir de interesses e experiência da área médica.

Ainda em relação a amostra, para alguns indivíduos o evento de interesse terá ocorrido. Por exemplo, na amostra de pacientes com cardiomiopatia chagásica, alguns dos indivíduos morreram precocemente (o evento de interesse é a morte em até um ano após o diagnóstico). Estes indivíduos chamaremos de eventos (E), e os indivíduos para os quais o evento não ocorreu chamaremos de não-eventos (NE). Assim, $\hat{r}_{i,E}$ é o risco estimado para um indivíduo da população de eventos, e $\hat{r}_{i,NE}$ é o risco estimado para um indivíduo da população de não-eventos.

Capítulo 3

Técnicas para Avaliação de Marcadores de Risco

3.1 Introdução

Neste capítulo consideramos algumas das técnicas mais recentes para a avaliação de marcadores e modelos de predição de risco. A maioria destas técnicas está preocupada em avaliar se a inclusão de um novo marcador melhora o desempenho de um modelo já estudado. Para fins de notação chamaremos de “*modelo antigo*” o modelo sem o novo marcador, e “*modelo novo*” o modelo com os mesmos preditores do modelo antigo, porém com o acréscimo do novo marcador. Primeiramente na Seção 3.2 apresentamos a curva ROC, uma interpretação para a área sob a curva e ressaltamos os principais tópicos do debate em torno do método. Na Seção 3.3 mostramos a tabela de estratificação de risco e salientamos as vantagens e desvantagens da técnica. Na Seção 3.4 apresentamos o método gráfico da curva de previsibilidade. Na Seção 3.5 abordamos o NRI e o IDI, duas medidas resumo para avaliação de um novo marcador de risco, e os testes de hipóteses relacionados a estas medidas. E por fim, na Seção 3.6 apresentamos mais um método gráfico para a comparação de riscos estimados de dois modelos.

3.2 Curva ROC

A curva ROC (*receiver operating characteristic*) é uma ferramenta estatística muito conhecida e utilizada na avaliação de testes diagnósticos e também para caracterizar o potencial prognóstico de um marcador (ou modelo) (4). Utilizando um limiar c , se $\hat{r} \geq c$ a classificação é positiva (se risco é maior que um limiar c , então o indivíduo é classificado como evento), e $\hat{r} < c$ a classificação é negativa (se o risco é menor que c , então o indivíduo é classificado como não-evento). Seja D a variável indicadora de evento ($D = 1$, se evento; $D = 0$, caso contrário). Denotamos por fração de verdadeiros positivos e fração de falsos positivos segundo o limiar c como $\text{TPF}(c)$ e $\text{FPF}(c)$, respectivamente, em que:

$$\text{TPF}(c) = P(\hat{r} > c | D = 1), \quad (3.1)$$

$$\text{FPF}(c) = P(\hat{r} > c | D = 0). \quad (3.2)$$

Quando nenhum valor de c é indicado o *spectrum* completo de TPF's e FPF's pode ser caracterizado utilizando uma curva ROC que plota a fração de verdadeiros positivos *versus* a fração de falsos positivos para todo $c \in (-\infty, \infty)$ (21). Também podemos escrever a curva ROC como

$$\text{ROC}(\cdot) = \{(u, \text{ROC}(u)), \quad u \in (0, 1)\}, \quad (3.3)$$

em que $u = \text{FPF}(c)$, e $\text{ROC}(u) = \text{TPF}(c)$ (22).

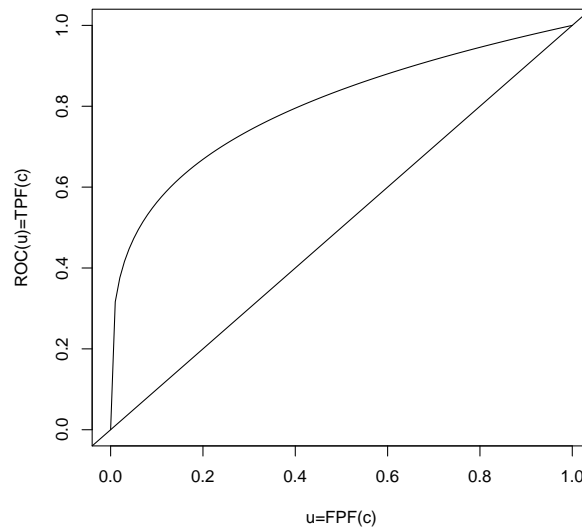


Figura 3.1: Exemplo de uma curva ROC.

A curva ROC é uma função monotonamente crescente no quadrante positivo, como pode ser visto na Figura 3.1. Um marcador (ou de uma maneira mais geral, um modelo) não-informativo é aquele em que o \hat{r} não tem relação com o status do evento, ou seja, a distribuição do risco é a mesma nos eventos e nos não-eventos, e portanto, para qualquer limiar c temos $\text{TPF}(c) = \text{FPF}(c)$. A reta $\text{TPF}(c) = \text{FPF}(c)$ na Figura 3.1 representa a curva ROC para um marcador inútil para predição do risco. Já um marcador perfeito é aquele que o risco associado a ele separa completamente os indivíduos em eventos e não-eventos. Para algum limiar c temos que $\text{TPF}(c) = 1$ e $\text{FPF}(c) = 0$. A curva ROC do risco de tal marcador é plotada ao longo das bordas esquerda e superior do quadrante positivo (22).

A medida resumo mais utilizada na análise da curva ROC é a área sob a curva (AUC) (22), definida por

$$\text{AUC} = \int_0^1 \text{ROC}(u) du. \quad (3.4)$$

A AUC é a probabilidade que o risco estimado para um indivíduo selecionado aleatoriamente da população de eventos seja maior que o risco estimado para um indivíduo selecionado aleatoriamente

da população de não-eventos (21), isto é,

$$\text{AUC} = P(\hat{r}_E > \hat{r}_{NE}).$$

A demonstração deste resultado pode ser vista em (22). Quanto maior a AUC melhor é o marcador para a predição do risco.

A avaliação de um novo marcador de risco geralmente é feita comparando as curvas ROC para os riscos estimados dos dois modelos (*modelo antigo* e *modelo novo*). Mais especificamente, é calculada a diferença entre as áreas das duas curvas. Se a diferença é significativa, e positiva, então consideramos que o novo marcador acrescenta valor preditivo ao modelo. Detalhes da estimação da AUC podem ser vistos em (22) ou em (4), e um método para a comparação de duas AUC's quando utilizados os mesmos dados pode ser visto em (23).

Existe um crescente reconhecimento que a curva ROC possui sérias limitações na avaliação de marcadores de predição de risco (1, 3, 7, 24). Pepe *et al.* (5), através de simulações numéricas, demonstraram a relação entre associação, através do *odds ratio*, e classificação, descrita pela AUC. Eles concluíram que a significância estatística na associação por si só não caracteriza a capacidade discriminatória de um marcador, e que um aumento significativo na AUC requer uma associação com magnitude raramente vista em estudos epidemiológicos. O trabalho de Cook, Buring e Ridker (6), mostrou que não houve mudança significativa na AUC quando foram acrescentados ao modelo marcadores com importância preditiva já reconhecida.

Em particular, uma característica chave da curva ROC é que ela não envolve a escala de medida original dos marcadores, ou seja, os limiares de risco não podem ser visualizados. Se é de interesse comparar o desempenho de dois diferentes modelos utilizando o limiar “ $p_L = 0,20$ ”, por exemplo, isto não pode ser feito utilizando as curvas ROC, pois não é possível localizar na curva o ponto que corresponde a este limiar (8). Para finalizar, a interpretação da AUC como a probabilidade de corretamente ordenar os riscos de dois indivíduos aleatoriamente sorteados, um da população de eventos e outro da população de não-eventos, parece ter pouca relevância clínica, pois os indivíduos nunca são apresentados aos pares na prática médica (25).

3.3 Tabela de Estratificação de Risco

A tabela de estratificação de risco foi primeiramente sugerida por Cook, Buring e Ridker (6,7) para avaliar o valor adicional de um novo marcador, ou o benefício de se adicionar um novo marcador, a um conjunto de marcadores de risco já estabelecido. Nesta tabela de estratificação, os riscos são estimados dos modelos sem e com o novo marcador, os indivíduos em estudo são estratificados em categorias de risco clinicamente relevantes, e então estas duas classificações são cruzadas. Janes, Pepe e Gu (1) examinaram a abordagem da estratificação de risco em detalhes, e concluíram que a informação que pode ser extraída da tabela de estratificação de risco é um enorme avanço sobre

medidas geralmente utilizadas para avaliar o desempenho do modelo de predição de risco. Através de um exemplo, mostramos como informações a respeito da calibração do modelo, capacidade de estratificação e acurácia da classificação podem ser obtidos de uma tabela de estratificação de risco e utilizadas para avaliar o valor de um novo marcador para decisões médicas.

Na Tabela 3.1 apresentamos um exemplo da tabela de estratificação de risco. Os dados deste exemplo referem-se a um trabalho desenvolvido no decorrer desta dissertação (15). Estes dados também se referem ao risco de morte precoce (morte em até um ano após o diagnóstico) em pacientes com cardiomiopatia chagásica (no Capítulo 5 fazemos a análise com os dados atualizados e considerando as demais técnicas de avaliação de marcadores de risco). O *modelo antigo* dispunha de quatro variáveis (volume do átrio esquerdo, fração de ejeção, classificação NYHA e o índice de Tei do ventrículo direito), e estamos interessados em avaliar se a inclusão do marcador E/E' (a razão da onda E do fluxo mitral com a onda E' do *Doppler* tecidual) melhora o desempenho do modelo. Os riscos foram estimados a partir do modelo de Cox. Os autores de (15) definiram os limiares de baixo e alto risco, respectivamente, $p_L = 0,02$ e $p_H = 0,10$.

Tabela 3.1: Exemplo da tabela de estratificação: Risco de morte em até um ano pelos modelos sem e com E/E'.

Risco 1 ano modelo sem E/E'		Risco 1 ano modelo com E/E'			Total
		0-2%	2-10%	$\geq 10\%$	
0-2%	N	48	2	0	50
	Eventos	0	0	0	0
	Não-eventos	48	2	0	50
	Prop. Eventos	0,00	0,00	-	0,00
2-10%	N	3	36	5	44
	Eventos	0	2	0	2
	Não-eventos	3	34	5	42
	Prop. Eventos	0,00	5,56	0,00	4,55
$\geq 10\%$	N	0	8	35	43
	Eventos	0	0	22	22
	Não-eventos	0	8	13	21
	Prop. Eventos	-	0,00	62,86	51,16
Total	N	51	46	40	137
	Eventos	0	2	22	24
	Não-eventos	51	44	18	113
	Prop. Eventos	0,00	4,35	55,00	17,52

Na Tabela 3.1, para cada classe de risco, temos o número de indivíduos alocados na classe (N), o número de eventos na classe, o número de não-eventos na classe, e a proporção de eventos na classe (Prop. Eventos). Janes, Pepe e Gu (1) perceberam que era possível acessar informações a respeito da calibração do modelo, capacidade de estratificação e acurácia da classificação através da tabela de estratificação.

A calibração descreve o quão concordantes são os riscos preditos e as respostas observadas. Na tabela de estratificação de risco o modelo está bem calibrado quando a taxa de eventos observados para cada categoria encontra-se dentro da faixa de valores de cada categoria. Podemos verificar isto através das marginais da Tabela 3.1, comparando a proporção de eventos da classe com a faixa de valores da classe. Por exemplo, na classe “0 a 2%” de risco, os dois modelos apresentam 0% de eventos, e portanto dentro dos valores da faixa; na classe “2 a 10%”, o modelo sem E/E’ apresenta 4,55% de eventos, este valor pertence a faixa de valores da classe, assim como a proporção de eventos no modelo com E/E’ (4,35%). Se para todas as classes, as proporções de classe estão dentro dos valores de cada categoria, então o modelo é considerado bem calibrado. Em nosso exemplo os dois modelos tem um desempenho parecido quanto a calibração e, segundo o critério utilizado na tabela de estratificação, estão bem calibrados.

A capacidade de estratificação de um modelo de predição de risco pode ser descrita pela proporção de indivíduos alocados nas categorias de risco. Um melhor modelo deve colocar mais indivíduos nas classe extremas (baixo e alto risco), onde as decisões médicas são mais claras. Através da tabela de estratificação de risco, a capacidade de estratificação é avaliada através do percentual de indivíduos que estão nas classes intermediárias. Quanto menor for este percentual, melhor é a capacidade de estratificação do modelo. Mais uma vez voltamos nossa atenção para as marginais da Tabela 3.1, onde fazemos a razão entre os indivíduos da classe intermediária e o total de indivíduos do estudo. Em nosso exemplo, aproximadamente 32% ($44 \div 137 \times 100$) dos indivíduos são alocados na classe “2 a 10%” no modelo sem E/E’, e 34% ($46 \div 137 \times 100$) são alocados nesta classe no modelo com E/E’, ou seja, o acréscimo do novo marcador ligeiramente piora a capacidade de estratificação do modelo de predição de risco.

Utilizando um limiar de alto risco, medidas de acurácia da classificação, como a fração de verdadeiros positivos e a fração de falsos positivos, $TPF(p_H)$ e $FPF(p_H)$, podem ser extraídas das margens da tabela de estratificação de risco. Estas medidas são estimadas através da razão entre eventos na classe de alto risco e o total de eventos (o $TPF(p_H)$), e a razão de não-eventos na classe de alto risco e o total de não-eventos (o $FPF(p_H)$). Da Tabela 3.1, $TPF(p_H) = 91,67\%$ ($22 \div 24 \times 100$) para os dois modelos e $FPF(p_H) = 18,58\%$ ($21 \div 113 \times 100$) para o modelo sem E/E’ e $FPF(p_H) = 15,93\%$ ($18 \div 113 \times 100$) para o modelo com E/E’.

Janes, Pepe e Gu (1) ainda alertam para o cuidado de compararmos modelos encaixados de predição de risco para que a técnica da tabela de estratificação de risco tenha sentido de ser utilizada. A utilização da tabela de estratificação com modelos que não são encaixados apresenta informação somente a respeito da correlação entre os riscos dos dois modelos. Uma possível desvantagem desta técnica é a escolha dos limiares de risco.

3.4 Curva de Previsibilidade

Pepe *et al.* (8,9) propuseram um novo gráfico, a curva de previsibilidade para descrever a capacidade preditiva de um marcador ou modelo de predição de risco. Esta curva apresenta a

distribuição do risco através de quantis de risco, $R(\nu)$ versus ν (24), em que

$$P(\text{risco}(\mathbf{x}) \leq R(\nu)) = \nu.$$

O inverso da curva de previsibilidade é simplesmente a função de distribuição acumulada do $\text{risco}(\mathbf{x})$

$$R^{-1}(p) = P(\text{risco}(\mathbf{x}) \leq p) = F_{\text{risco}}(p),$$

e correspondentemente

$$R(\nu) = F_{\text{risco}}^{-1}(\nu).$$

Na Figura 3.2 apresentamos a curva de previsibilidade para um marcador fictício. Este gráfico facilita a comparação de marcadores e modelos de predição de risco, pois utiliza a mesma escala ν . A reta horizontal no gráfico representa o risco total, ou a prevalência ($\pi = P(\text{evento ocorrer})$) e serve de referência, como um marcador que atribui o mesmo risco para todos indivíduos.

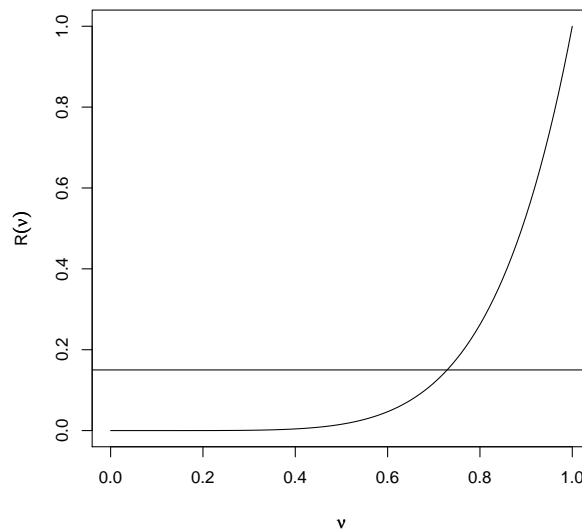


Figura 3.2: Exemplo de uma curva de previsibilidade.

Para a construção da curva da Figura 3.2, primeiramente estimamos os riscos para cada indivíduo a partir do modelo e os ordenamos do menor para o maior. Estimamos os quantis do risco e plotamos quantis *versus* os riscos ordenados. Da Figura 3.2 vemos que em 0,8 no eixo x o risco é aproximadamente 0,26. Isto nos indica que para 80% dos indivíduos o risco estimado para o evento de interesse com base no marcador é menor que 0,26, e para 20% dos indivíduos o risco é maior que 0,26. Outra maneira de utilizar o gráfico é começando a partir do valor do risco no eixo y (8). Por exemplo, com o risco $\hat{r} = 0,20$, vemos que ν é aproximadamente 0,77, ou seja, estimamos que 23%

dos indivíduos têm riscos estimados acima de 0,20. Com o risco $\hat{r} = 0,10$, ν é aproximadamente 0,68, e portanto, 68% dos indivíduos têm riscos estimados abaixo de 0,10. Se 0,10 e 0,20 fossem os respectivos limiares de baixo e alto risco, teríamos que apenas 9% dos indivíduos são alocados na classe intermediária.

3.5 NRI e IDI

Pencina *et al.* (3) propuseram duas novas estatísticas para quantificar o aumento no desempenho do modelo quando um novo marcador é adicionado a um conjunto de preditores já especificado para a predição de um evento de interesse. Estas estatísticas são conhecidas como o NRI e o IDI, e testes de hipóteses foram propostos para avaliar a significância estatística destas (3).

3.5.1 Melhora da Reclassificação Líquida (NRI)

Consideremos o mesmo contexto apresentado na Seção 3.3 em que temos limiares de risco definidos para a formação de categorias de risco. Para Pencina *et al.* (3), indivíduos da população de eventos, que se movem de uma classe de risco para outra mais alta (por exemplo, indivíduos classificados na categoria “0 a 2%” segundo o risco estimado pelo *modelo antigo* e se movem para a categoria “2 a 10%” com seu risco estimado pelo *modelo novo*) implicam em melhora da classificação, e indivíduos eventos que se movem de uma classe de risco para outra mais baixa (por exemplo, indivíduos que estavam na categoria “> 10%” pelo *modelo antigo* e passam para a categoria “2 a 10%”) indicam piora na classificação. A interpretação é inversa para indivíduos da população de não-eventos. Assim, na população de não-eventos, indivíduos que se movem para baixo, implicam melhora na classificação, e indivíduos que se movem para cima, a piora na classificação. Desta maneira, a melhora na reclassificação pode ser quantificada como a soma da “proporção de indivíduos movendo para cima” menos a “proporção de indivíduos movendo para baixo” na população de eventos, e a “proporção de indivíduos movendo para baixo” menos a “proporção de indivíduos movendo para cima” na população de não-eventos. Pencina *et al.* (3) chamaram esta soma de NRI (*net reclassification improvement*). Utilizando a mesma notação da Seção 3.2 da variável indicadora de evento D , e “cima” e “baixo” para denotar mudanças de uma ou mais categorias de risco nas direções ascendente e descendente, respectivamente, o NRI é definido como

$$\text{NRI} = [P(\text{cima}|D = 1) - P(\text{baixo}|D = 1)] - [P(\text{cima}|D = 0) - P(\text{baixo}|D = 0)]. \quad (3.5)$$

Para estimar o NRI, Pencina *et al.* (3) utilizaram proporções amostrais para estimar as prob-

abilidades que compõem o NRI:

$$\hat{P}(\text{cima}|D = 1) = \hat{p}_{\text{cima},E} = \frac{\#E \text{ movendo para cima}}{n_E}, \quad (3.6)$$

$$\hat{P}(\text{baixo}|D = 1) = \hat{p}_{\text{baixo},E} = \frac{\#E \text{ movendo para baixo}}{n_E}, \quad (3.7)$$

$$\hat{P}(\text{cima}|D = 0) = \hat{p}_{\text{cima},NE} = \frac{\#NE \text{ movendo para cima}}{n_{NE}}, \quad (3.8)$$

$$\hat{P}(\text{baixo}|D = 0) = \hat{p}_{\text{baixo},NE} = \frac{\#NE \text{ movendo para baixo}}{n_{NE}}, \quad (3.9)$$

em que n_E e n_{NE} são o número de indivíduos eventos e número de indivíduos não-eventos, respectivamente. O NRI é estimado por

$$\hat{\text{NRI}} = (\hat{p}_{\text{cima},E} - \hat{p}_{\text{baixo},E}) - (\hat{p}_{\text{cima},NE} - \hat{p}_{\text{baixo},NE}). \quad (3.10)$$

Note, que com esta definição, o NRI é uma estatística resumo que pode ser diretamente calculada da tabela de estratificação de risco. Voltemos ao exemplo da Tabela 3.1. Nenhum indivíduo dos eventos mudou de categoria. Entre os não-eventos, 2 indivíduos mudaram da classe mais baixa para a intermediária e 5 mudaram da intermediária para a mais alta (para estes 7 indivíduos a reclassificação foi “negativa”). Ainda entre os não-eventos, 3 indivíduos mudaram da classe intermediária para a mais baixa e 8 mudaram da mais alta para a intermediária. Desta forma, temos para o exemplo da Seção 3.3 que o NRI é:

$$\begin{aligned} \hat{\text{NRI}} &= \left(\frac{0 - 0}{24} \right) - \left(\frac{2 + 5 - 3 - 8}{113} \right) \\ &= \frac{4}{113} = 0,0354. \end{aligned}$$

Assumindo independência entre os indivíduos das populações de eventos e não-eventos, Pencina *et al.* (3) obtiveram a seguinte estatística de teste para a hipótese nula $\text{NRI} = 0$ ($H_0 : \text{NRI} = 0$):

$$z_{\text{NRI}} = \frac{\hat{\text{NRI}}}{\sqrt{\frac{\hat{p}_{\text{cima},E} + \hat{p}_{\text{baixo},E}}{n_E} + \frac{\hat{p}_{\text{cima},NE} + \hat{p}_{\text{baixo},NE}}{n_{NE}}}}. \quad (3.11)$$

A estatística z_{NRI} tem distribuição assintótica normal padrão, sob H_0 .

3.5.2 Melhora da Discriminação Integrada (IDI)

A alta dependência na escolha das classes de risco foi reconhecida e apontada como uma possível desvantagem do NRI (3). Para superar esta limitação, Pencina *et al.* (3) propuseram o IDI (*integrated discrimination improvement*). Imagine que o risco estimado para cada indivíduo é a sua própria categoria de risco, então qualquer mudança no risco estimado proporciona um movimento

para baixo ou para cima. Pencina *et al.* (3) mostraram que isto equivale a quantificar melhoras conjuntas nas curvas de sensibilidade (fração de verdadeiros positivos) e especificidade (um menos a fração de falsos positivos). Utilizando a mesma notação da Seção 3.2, temos que as frações de verdadeiros e falsos positivos integradas, ITPF e IFPF, respectivamente, são

$$\text{ITPF} = \int_0^1 P(\hat{r} > c | D = 1) dc, \quad (3.12)$$

$$\text{IFPF} = \int_0^1 P(\hat{r} > c | D = 0) dc. \quad (3.13)$$

Assim, a melhora da discriminação integrada (IDI) entre o “*modelo novo*” e o “*modelo antigo*” é definida como:

$$\text{IDI} = (\text{ITPF}_{\text{novo}} - \text{ITPF}_{\text{antigo}}) - (\text{IFPF}_{\text{novo}} - \text{IFPF}_{\text{antigo}}). \quad (3.14)$$

Observe que as Expressões (3.12) e (3.13) são as esperanças do risco nas populações de eventos e não-eventos, respectivamente

$$\text{ITPF} = E(\hat{r} | D = 1), \quad (3.15)$$

$$\text{IFPF} = E(\hat{r} | D = 0), \quad (3.16)$$

e estas podem ser estimadas pelas respectivas médias amostrais

$$\widehat{\text{ITPF}} = \bar{r}_E = \frac{\sum_i \hat{r}_i}{n_E}, \quad \forall i \text{ evento}, \quad (3.17)$$

$$\widehat{\text{IFPF}} = \bar{r}_{NE} = \frac{\sum_i \hat{r}_i}{n_{NE}}, \quad \forall i \text{ não-evento}, \quad (3.18)$$

assim o IDI é estimado por

$$\widehat{\text{IDI}} = (\bar{r}_{\text{novo},E} - \bar{r}_{\text{antigo},E}) - (\bar{r}_{\text{novo},NE} - \bar{r}_{\text{antigo},NE}). \quad (3.19)$$

Como os eventos atuais não dependem do modelo (“*antigo*” ou “*novo*”), o desvio padrão de $(\bar{r}_{\text{novo},E} - \bar{r}_{\text{antigo},E})$ pode ser calculado como o erro padrão das diferenças pareadas dos riscos dos dois modelos para indivíduos da população de eventos, $\hat{e}p_E$ (3). Da mesma forma o correspondente estimador do desvio padrão de $(\bar{r}_{\text{novo},NE} - \bar{r}_{\text{antigo},NE})$ é $\hat{e}p_{NE}$. Assumindo independência entre eventos e não-eventos, Pencina *et al.* (3) propuseram a seguinte estatística de teste para hipótese nula $\text{IDI} = 0$ ($H_0 : \text{IDI} = 0$):

$$z_{\text{IDI}} = \frac{\widehat{\text{IDI}}}{\sqrt{(\hat{e}p_E)^2 + (\hat{e}p_{NE})^2}}. \quad (3.20)$$

A estatística z_{IDI} tem distribuição assintótica normal padrão, sob H_0 .

3.6 Gráfico de Dispersão dos Riscos

Um método gráfico que não foi formalmente sugerido na literatura, mas que é bastante útil para compararmos os riscos de dois modelos é o gráfico de dispersão dos riscos estimados. Este método pode ser visto no artigo de Gu e Pepe (2). Nesta dissertação utilizamos este gráfico em nossas análises e para descrever alguns cenários que desenvolvemos em nossas simulações.

Nesta técnica, fizemos o gráfico dos riscos estimados a partir do *modelo antigo versus* os riscos estimados do *modelo novo*, separadamente para eventos e não-eventos. Os pontos da Figura 3.3 representam os riscos estimados de uma população fictícia. A reta $y = x$ representa a igualdade dos riscos, ou seja, se os pontos estão dispostos nesta reta, então os riscos dos dois modelos se equivalem. Se os pontos estão acima desta reta nos eventos, então o *modelo novo* melhora a predição do risco nos eventos (quanto maior o risco para eventos, mais fácil e acuradamente são tomadas decisões médicas). Se os pontos estão abaixo desta reta nos eventos, então o *modelo novo* piora a predição do risco nos eventos. O raciocínio é exatamente o contrário para os não-eventos. As retas tracejadas na horizontal e vertical são os limiares de baixo e alto risco, e servem para visualizarmos o movimento de uma classe de risco para outra.

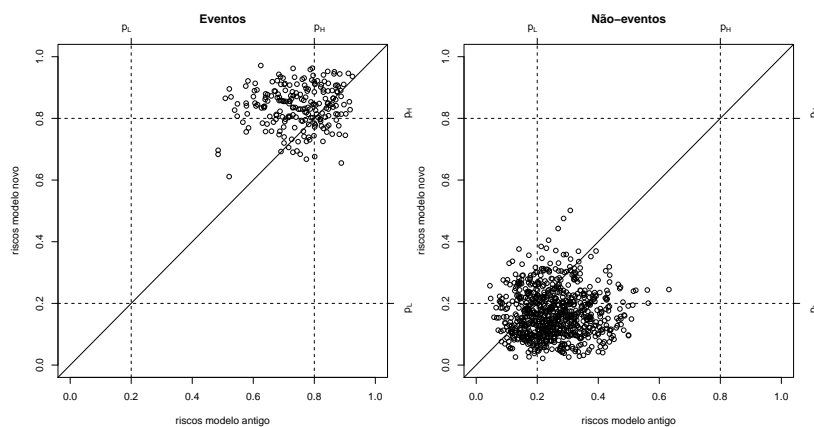


Figura 3.3: Exemplo do gráfico de dispersão dos riscos estimados.

Capítulo 4

Estudo de Simulação

4.1 Introdução

Neste capítulo avaliamos através de simulações Monte Carlo a qualidade das inferências para as medidas da melhora da reclassificação líquida (NRI) e da melhora da discriminação integrada (IDI) introduzidos por Pencina *et al.* (3) e descritos na Seção 3.5. Fizemos comparações em relação aos testes de hipóteses propostos para verificar a significância do NRI e IDI. As comparações foram feitas utilizando dois modelos para a predição do risco de um determinado evento. O primeiro modelo com uma única covariável (X_1), denominado de *modelo antigo*, e o segundo modelo com as covariáveis X_1 e X_2 , o *modelo novo*. Desta forma o NRI e o IDI avaliam se o novo marcador (a covariável X_2) melhora o desempenho do modelo. Escolhemos o modelo exponencial para dados de sobrevivência para o ajuste dos modelos, pois este é um caso particular do modelo de Cox.

As simulações foram realizadas utilizando a linguagem de programação R 2.9.0 (26). Em todos os casos simulados, utilizamos 3.000 réplicas de Monte Carlo. Na Seção 4.2 descrevemos os passos da simulação. Na Seção 4.3 descrevemos os cenários que simulamos. Por fim, na Seção 4.4 apresentamos e discutimos os resultados encontrados.

4.2 Descrição da Simulação

Realizamos as simulações da seguinte maneira:

1. Geramos n observações de X_1 e X_2 uma única vez, tal que $\mathbf{X} = (X_1, X_2)'$ e

$$\mathbf{X} \sim N_2(\boldsymbol{\mu}; \boldsymbol{\Sigma}), \text{ em que } \boldsymbol{\mu} = (\mu_{X_1}, \mu_{X_2})' \text{ e } \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{X_1}^2 & \rho_{X_1, X_2}(\sigma_{X_1}\sigma_{X_2}) \\ \rho_{X_1, X_2}(\sigma_{X_1}\sigma_{X_2}) & \sigma_{X_2}^2 \end{bmatrix};$$

2. Para cada amostra de Monte Carlo geramos os tempos de falha T_i , tal que

$$T_i = \exp\{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}\} \epsilon_i, \text{ em que:}$$

- (a) β_0 , β_1 e β_2 são fixos e pertencem ao conjunto dos números reais;
- (b) os ϵ_i 's são *iid* com distribuição exponencial de média unitária, $i = 1, \dots, n$;

3. Falhas: se $T_i < t$, então T_i é tempo de falha; t é fixo tal que $t = -\log(1 - \pi) \exp\{\beta_0 + \beta_1 \mu_{X_1} + \beta_2 \mu_{X_2}\}$, e $\pi = P(T_i < t)$ é o risco total (prevalência); assim espera-se que em média $100\pi\%$ das observações sejam tempos de falha em cada amostra de Monte Carlo;
4. Modelos: ajustamos o modelo exponencial para dados de sobrevivência sem e com a covariável X_2 em cada amostra de Monte Carlo;
5. Encontrando o risco para t fixo: computamos os riscos para cada observação para os modelos sem e com a covariável X_2 , tal que:
 - (a) $\hat{r}_{i,antigo} = 1 - \exp\left\{-\frac{t}{\exp\{\hat{\beta}_0 + \hat{\beta}_1 x_{1i}\}}\right\}$;
 - (b) $\hat{r}_{i,novo} = 1 - \exp\left\{-\frac{t}{\exp\{\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}\}}\right\}$;
6. Categorias de risco: os riscos foram categorizados a partir de dois limiares de risco p_L e p_H ; desta maneira temos três categorias de risco:
 - (a) baixo risco, se $0 \leq \hat{r}_i < p_L$;
 - (b) médio risco, se $p_L \leq \hat{r}_i < p_H$;
 - (c) alto risco, se $p_H \leq \hat{r}_i \leq 1$;
7. Estimamos o NRI e IDI conforme as Expressões (3.10) e (3.19) para cada amostra de Monte Carlo; calculamos as estatísticas de teste z_{NRI} e z_{IDI} para cada amostra Monte Carlo.

4.3 Descrição dos Cenários Simulados

O primeiro objetivo de simularmos os métodos do NRI e do IDI foi de verificar se as taxas de rejeição das hipóteses nulas ($H_0 : NRI = 0$ e $H_0 : IDI = 0$) estavam próximas dos níveis nominais de significância, e comparar o desempenho dos testes. Este objetivo demanda um exercício de pensarmos o que significa a hipótese nula para cada um dos testes. Pepe, Feng e Gu (25) demonstraram que testar a hipótese $H_0 : IDI = 0$ é equivalente a testar $H_0 : \beta_2 = 0$, em que β_2 é o coeficiente associado a covariável X_2 no modelo de regressão. Porém, não encontramos na literatura uma hipótese equivalente para $H_0 : NRI = 0$. É claro que se $\beta_2 = 0$, então $NRI = 0$, pois $\hat{r}_{antigo} = \hat{r}_{novo}$ e desta forma não há reclassificação. Mas, qualquer reclassificação não implica em melhora na reclassificação, como argumentaram Pencina *et al.* (3). Desta forma, pensamos em outras situações que podem implicar em $NRI = 0$, como por exemplo, um cenário que existe reclassificação, mas as reclassificações positivas anulam as reclassificações negativas.

Com este desafio, criamos dez cenários para estudarmos o comportamento dos testes. O **Cenário 1** corresponde a situação de simularmos sob H_0 . Os demais cenários são descritos a seguir e foram gerados sob a hipótese alternativa ($H_1 : NRI \neq 0$ e $H_1 : IDI \neq 0$). Para ilustrar os diferentes cenários, mostramos os gráficos de dispersão dos riscos de uma amostra de 5.000 observações gerados conforme descrito na Seção 4.2. Como vimos anteriormente o processo de

geração dos dados envolve uma quantidade razoável de parâmetros a serem especificados. Para gerar os diferentes cenários, utilizamos diferentes valores para estes parâmetros, mas alguns deixamos fixos para todos os cenários, tais como: o risco total ($\pi = 0,2$), os limiares de baixo e alto risco ($p_L = 0,02$ e $p_H = 0,10$), o intercepto ($\beta_0 = 0$), o vetor de médias da distribuição conjunta das covariáveis ($\mu = \mathbf{0}$) e o nível nominal de 5%. Para cada cenário consideramos os tamanhos de amostra $n = \{100, 300, 500, 1.000, 2.000, 5.000\}$.

Cenário 1: Este cenário é correspondente as hipóteses nula $H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$; geramos este cenário com os parâmetros $\beta_1 = 1$, $\beta_2 = 0$, $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 1$ e $\rho_{X_1, X_2} = 0$; o comportamento dos riscos pode ser observado pela Figura 4.1.

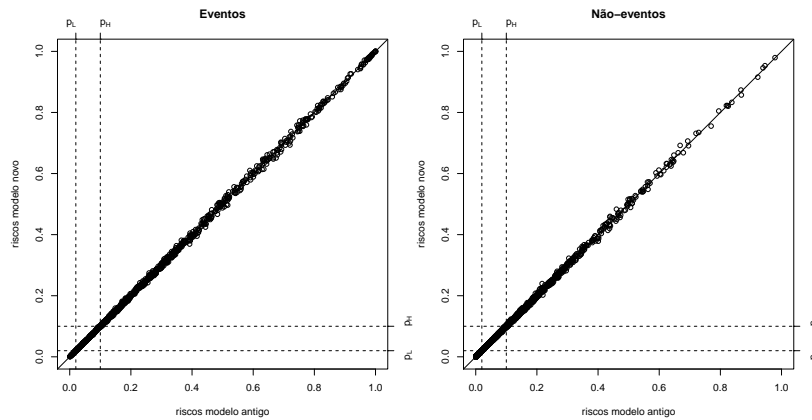


Figura 4.1: Gráfico de dispersão entre riscos para o Cenário 1; $n = 5.000$.

Cenário 2: Com este cenário podemos avaliar o poder do teste, pois simulamos sob as hipóteses alternativas $H_1 : \text{NRI} \neq 0$ e $H_1 : \text{IDI} \neq 0$; demos a mesma importância para X_1 e X_2 ; geramos este cenário com os parâmetros $\beta_1 = \beta_2 = 1$, $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 1$ e $\rho_{X_1, X_2} = 0$; vemos claramente pela Figura 4.2 que com o acréscimo de X_2 (β_2 significativo) há o aumento do risco para os indivíduos eventos.

Cenário 3: Neste cenário introduzimos correlação entre X_1 e X_2 ; geramos este cenário com os parâmetros $\beta_1 = \beta_2 = 1$, $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 1$ e $\rho_{X_1, X_2} = 0,5$; note pela Figura 4.3 que este cenário é bem semelhante ao **Cenário 2**.

Cenário 4: Neste cenário aumentamos a correlação entre X_1 e X_2 ; geramos este cenário com os parâmetros $\beta_1 = \beta_2 = 1$, $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 1$ e $\rho_{X_1, X_2} = 0,7$; nota-se pela Figura 4.4 que quanto maior a correlação, menos X_2 acrescenta no desempenho do modelo.

Cenário 5: Neste cenário aumentamos a correlação entre as covariáveis; geramos este cenário com os parâmetros $\beta_1 = \beta_2 = 1$, $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 1$ e $\rho_{X_1, X_2} = 0,85$; vemos pela Figura 4.5 que quanto maior a correlação, mais os pontos do gráfico se aproximam da reta de referência.

Cenário 6: Neste colocamos forte correlação entre X_1 e X_2 ; geramos este cenário com os

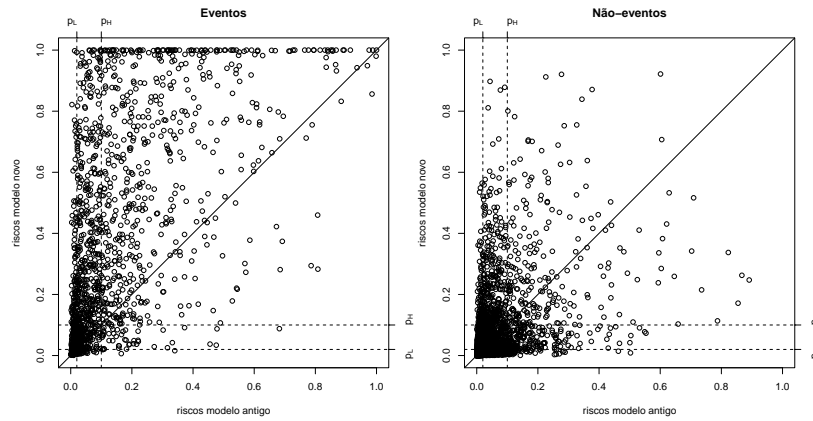


Figura 4.2: Gráfico de dispersão entre riscos para o Cenário 2; $n = 5.000$.

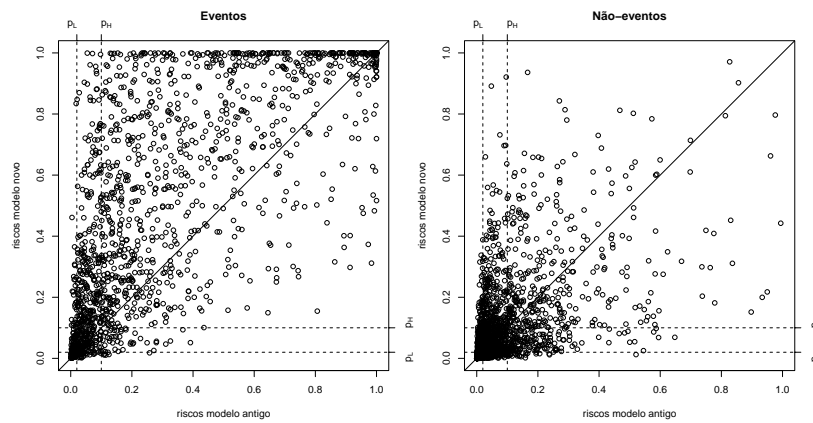


Figura 4.3: Gráfico de dispersão entre riscos para o Cenário 3; $n = 5.000$.

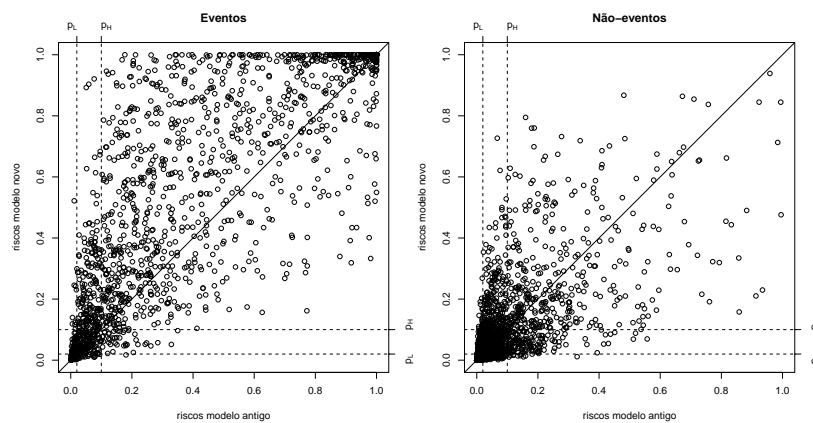


Figura 4.4: Gráfico de dispersão entre riscos para o Cenário 4; $n = 5.000$.

parâmetros $\beta_1 = \beta_2 = 1$, $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 1$ e $\rho_{X_1, X_2} = 0,95$; na Figura 4.6 podemos ver que os pontos do gráfico estão mais concentrados em torno da reta de referência.

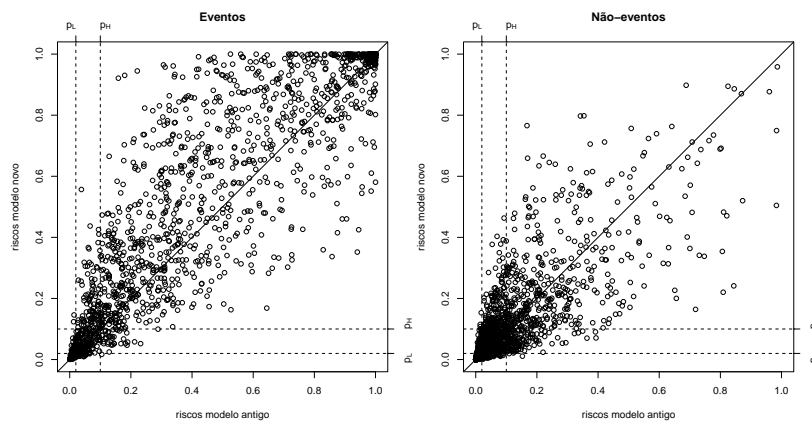


Figura 4.5: Gráfico de dispersão entre riscos para o Cenário 5; $n = 5.000$.

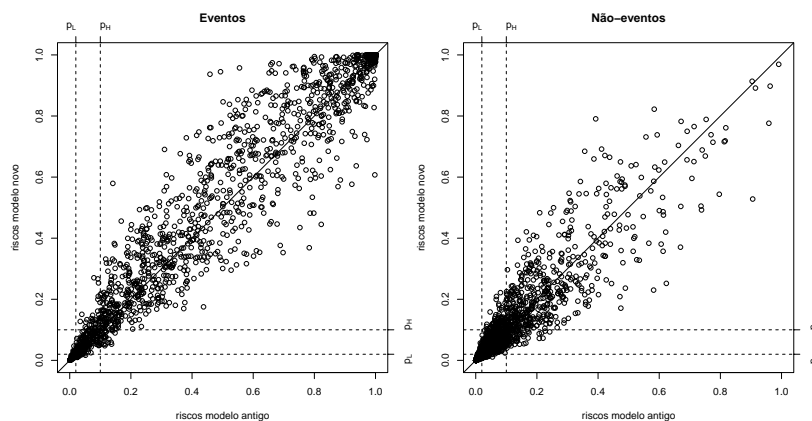


Figura 4.6: Gráfico de dispersão entre riscos para o Cenário 6; $n = 5.000$.

Cenário 7: Neste cenário mantivemos a correlação entre as covariáveis, porém alteramos as variâncias; geramos este cenário com os parâmetros $\beta_1 = \beta_2 = 1$, $\sigma_{X_1}^2 = 0,7$, $\sigma_{X_2}^2 = 0,3$ e $\rho_{X_1, X_2} = 0,85$; nota-se pela Figura 4.7 que isto provoca uma ligeira alteração em relação ao **Cenário 6**.

Cenário 8: Neste cenário alteramos os pesos de X_1 e X_2 ; geramos este cenário com os parâmetros $\beta_1 = 0,9$, $\beta_2 = 0,5$, $\sigma_{X_1}^2 = 0,7$, $\sigma_{X_2}^2 = 0,3$ e $\rho_{X_1, X_2} = 0,85$; note pela Figura 4.8 que nos aproximamos bastante da hipótese nula (quando os pontos estão todos sobre a reta de referência).

Cenário 9: Neste cenário diminuimos ainda mais a variância de X_2 ; geramos este cenário com os parâmetros $\beta_1 = 0,9$, $\beta_2 = 0,5$, $\sigma_{X_1}^2 = 0,7$, $\sigma_{X_2}^2 = 0,1$ e $\rho_{X_1, X_2} = 0,85$; podemos ver pela Figura 4.9 que cada vez mais nos aproximamos da ideia de que há reclassificação, porém a reclassificação positiva anula a reclassificação negativa ($H_0 : NRI = 0$).

Cenário 10: Neste cenário aumentamos a variância de X_1 ; geramos este cenário com os

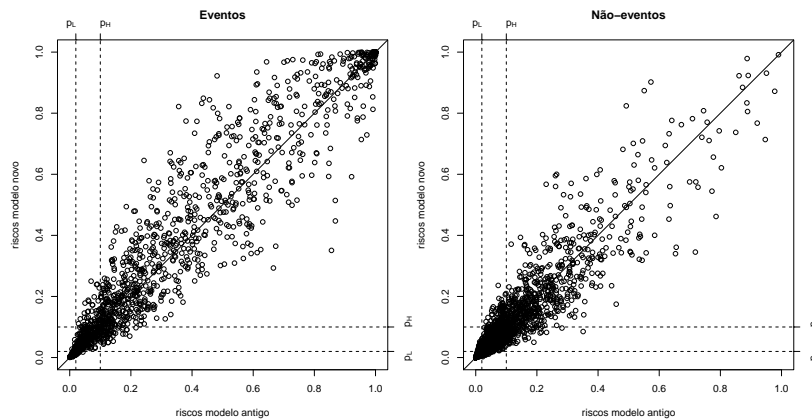


Figura 4.7: Gráfico de dispersão entre riscos para o Cenário 7; $n = 5.000$.

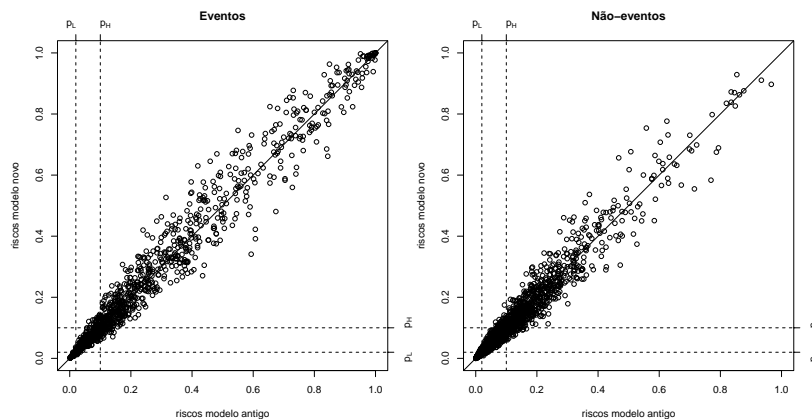


Figura 4.8: Gráfico de dispersão entre riscos para o Cenário 8; $n = 5.000$.

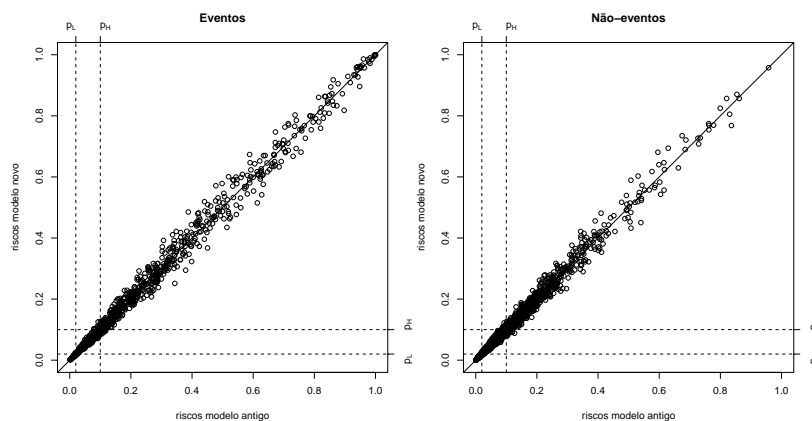


Figura 4.9: Gráfico de dispersão entre riscos para o Cenário 9; $n = 5.000$.

parâmetros $\beta_1 = 0,9$, $\beta_2 = 0,5$, $\sigma_{X_1}^2 = 0,9$, $\sigma_{X_2}^2 = 0,1$ e $\rho_{X_1, X_2} = 0,85$; nota-se pela Figura 4.10 que isto provoca pouca alteração em relação ao **Cenário 9**.

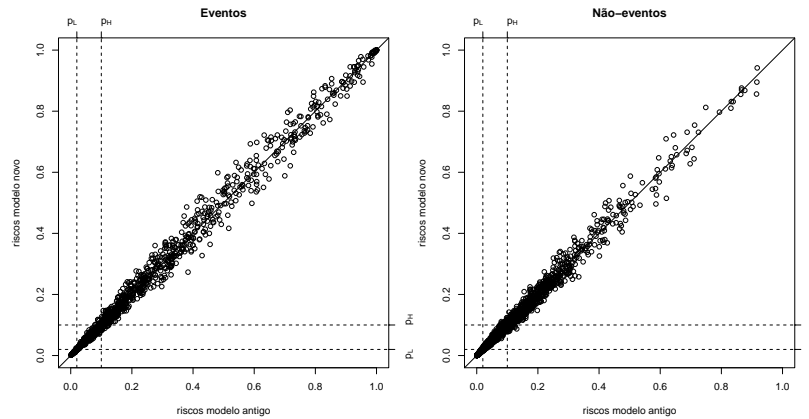


Figura 4.10: Gráfico de dispersão entre riscos para o Cenário 10; $n = 5.000$.

4.4 Resultados

Simulamos os cenários descritos na Seção 4.3, considerando diferentes tamanhos de amostra e o nível nominal de 5%. Para os **Cenários 1 e 5** ao **10** tivemos problemas com algumas das amostras geradas, relacionado ao cálculo da estatística de teste z_{NRI} , que por vezes apresentou uma indeterminação do tipo “zero sobre zero”. Isto ocorreu quando os riscos (dos modelos *antigo* e *novo*) foram iguais, ou tão parecidos que não houve reclassificação. Para estes casos descartamos a amostra. O número máximo de amostras descartadas (Amostras Mal Sucedidas) foi 257 no **Cenário 1** para tamanho de amostra 100. Consideramos este número baixo, em relação ao número de replicações realizadas (3.000), a ponto de não interferir em nossos resultados. Nas tabelas a seguir apresentamos as taxas de rejeição das hipóteses nula ($H_0 : NRI = 0$ e $H_0 : IDI = 0$).

Na Tabela 4.1 apresentamos os resultados para o **Cenário 1**, em que simulamos sob H_0 . Nota-se que as taxas de rejeição da hipótese nula $H_0 : NRI = 0$ ficaram bem próximas do nível nominal de 5%, mas sempre abaixo deste. Já as taxas de rejeição da hipótese nula $H_0 : IDI = 0$ ficaram sempre acima do nível nominal de 5%. Também percebemos que, com exceção dos tamanhos de amostra 100 e 300, as taxas de rejeição da hipótese nula $H_0 : NRI = 0$ ficaram mais próximas do nível nominal em relação às taxas de rejeição da hipótese nula $H_0 : IDI = 0$.

Na Figura 4.11 apresentamos os histogramas das 3.000 réplicas das estatísticas de teste z_{NRI} e z_{IDI} para o **Cenário 1** com tamanho de amostra 5.000. Lembramos que neste cenário, simulamos sob H_0 , e portanto a distribuição assintótica, tanto de z_{NRI} quanto de z_{IDI} , é uma normal padrão. Podemos perceber que a distribuição exata da estatística de teste z_{NRI} tem uma aproximação melhor para a distribuição assintótica de referência do que a distribuição da estatística de teste z_{IDI} . A distribuição de z_{IDI} está aparentemente deslocada para a direita e com uma variabilidade menor. Conforme diminuimos o tamanho de amostra, a aproximação das distribuições das estatísticas de teste pela distribuição assintótica piora, mas o padrão se mantém semelhante com o observado para $n = 5.000$ (veja os histogramas no Apêndice A, Figuras A.1- A.5).

Tabela 4.1: Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o **Cenário 1**.

n	Amostras		
	Mal Sucedidas	NRI	IDI
5.000	38	4,63	5,64
2.000	57	4,04	11,45
1.000	85	4,08	7,72
500	164	4,37	6,73
300	209	3,33	5,05
100	257	3,86	6,05

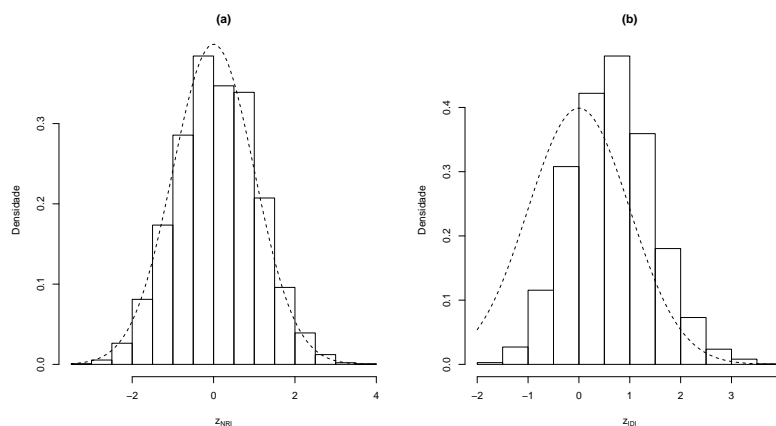


Figura 4.11: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 1**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

Nas Tabelas 4.2 e 4.3 apresentamos os resultados das simulações para os cenários **2** e **3**. Estes dois cenários são os casos que mais nos afastamos da geração dos dados sob H_0 . Não sabemos os verdadeiros valores do NRI e IDI, porém é bastante claro que a covariável X_2 melhora o desempenho do modelo para predição do risco (pelo menos para os eventos). As taxas de rejeição demonstraram que os dois testes são bastante poderosos para detectar a importância de X_2 para a predição do risco, nestes cenários. Para tamanhos de amostra menores ou igual a 500, o teste relacionado ao IDI se mostrou mais poderoso em relação ao teste do NRI.

Nos cenários **4** e **5** simulamos a situação que X_2 acrescenta no desempenho do modelo de predição de risco, porém de forma moderada. Apresentamos os resultados das simulações nas Tabelas 4.4 e 4.5. Assim como para os cenários **2** e **3**, o teste relacionado ao IDI se mostrou muito poderoso para detectar o acréscimo proporcionado por X_2 na predição do risco, e levou vantagem em relação ao teste do NRI para todos os tamanhos de amostra. Conforme o tamanho de amostra diminui, o teste relacionado ao NRI perde poder.

Nos cenários **6** e **7**, simulamos a situação na qual a melhora que X_2 proporciona na predição

Tabela 4.2: Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o **Cenário 2**.

n	NRI	IDI
5.000	100,00	100,00
2.000	100,00	100,00
1.000	100,00	100,00
500	99,40	100,00
300	84,20	99,97
100	57,73	99,10

Tabela 4.3: Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o **Cenário 3**.

n	NRI	IDI
5.000	100,00	100,00
2.000	100,00	100,00
1.000	98,63	100,00
500	84,17	100,00
300	41,57	100,00
100	17,43	70,40

Tabela 4.4: Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o **Cenário 4**.

n	NRI	IDI
5.000	100,00	100,00
2.000	98,00	100,00
1.000	72,47	100,00
500	52,83	100,00
300	16,43	99,63
100	7,33	42,57

do risco é muito pequena. Os resultados das simulações para estes cenários são apresentados nas Tabelas 4.6 e 4.7. Mais uma vez o teste do IDI se mostrou bastante poderoso, superando as taxas de rejeição do teste do NRI, exceto quando o tamanho de amostra foi 100. O NRI mostrou um poder muito baixo nestes dois cenários.

Nos cenários de **8** a **10** os riscos gerados se comportam de tal forma que, aparentemente, estamos simulando sob H_0 . Desta forma esperavamos que as taxas de rejeição das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) fossem bem próximas ao nível nominal de 5% de significância. Nas Tabelas 4.8, 4.9 e 4.10 apresentamos os resultados das simulações para estes cenários. O teste

Tabela 4.5: Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o **Cenário 5**.

n	Amostras		
	Mal Sucedidas	NRI	IDI
5.000	0	99,17	100,00
2.000	0	72,97	100,00
1.000	0	18,90	100,00
500	0	18,60	99,83
300	0	4,83	88,53
100	2	3,67	14,47

Tabela 4.6: Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o **Cenário 6**.

n	Amostras		
	Mal Sucedidas	NRI	IDI
5.000	0	48,37	100,00
2.000	0	28,47	99,97
1.000	0	3,23	98,90
500	0	4,63	48,53
300	0	3,83	47,07
100	141	2,76	1,71

Tabela 4.7: Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o **Cenário 7**.

n	Amostras		
	Mal Sucedidas	NRI	IDI
5.000	0	39,57	100,00
2.000	0	24,47	99,96
1.000	0	22,37	86,93
500	0	6,20	55,03
300	0	9,13	23,63
100	34	9,13	2,00

relacionado ao IDI apresentou taxas de rejeição muito maiores que as taxas de rejeição do teste relacionado ao NRI. O teste do NRI apresentou taxas relativamente próximas ao nível nominal, porém o tamanho de amostra parece influenciar nos resultados. Quanto maiores os tamanhos de amostra, mais se distanciam as taxas de rejeição do nível nominal.

Tabela 4.8: Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o **Cenário 8**.

n	Amostras		
	Mal Sucedidas	NRI	IDI
5.000	0	13,13	97,73
2.000	0	13,03	67,07
1.000	1	6,44	56,89
500	12	4,58	20,85
300	43	6,39	19,78
100	193	5,56	10,44

Tabela 4.9: Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o **Cenário 9**.

n	Amostras		
	Mal Sucedidas	NRI	IDI
5.000	0	8,67	62,17
2.000	2	5,34	35,32
1.000	17	6,74	26,12
500	64	4,09	12,33
300	117	4,40	11,76
100	219	5,93	6,33

Tabela 4.10: Taxas de rejeição (%) das hipóteses nula ($H_0 : \text{NRI} = 0$ e $H_0 : \text{IDI} = 0$) com base em 3.000 réplicas de Monte Carlo para o **Cenário 10**.

n	Amostras		
	Mal Sucedidas	NRI	IDI
5.000	0	7,73	61,15
2.000	3	5,67	28,89
1.000	14	6,40	24,55
500	49	4,13	8,74
300	113	4,36	10,08
100	221	5,15	6,91

4.4.1 Comentários Finais

Os resultados que encontramos nos mostram que o teste relacionado ao NRI é mais conservador. Quando simulamos sob H_0 , o teste do NRI mostrou taxas de rejeição mais próximas ao nível nominal de 5%. Quando não temos condições de afirmar que estamos sob H_0 , mas X_2 claramente não acrescenta ao modelo de predição de risco (cenários **8 a 10**) o teste relacionado ao NRI também

teve taxas de rejeição mais próximas do nível nominal. Porém, quando simulamos sob a hipótese alternativa (cenários **2** a **7**), o teste do NRI perdeu poder conforme variamos os cenários. Nos casos que X_2 claramente adiciona valor preditivo ao modelo, o teste relacionado ao NRI mostrou alto poder, mas este decresce rapidamente conforme o tamanho da amostra diminui. Para os cenários que X_2 acrescenta no desempenho do modelo de forma moderada para fraca, o teste do NRI apresentou baixo poder para praticamente todos os tamanhos de amostra.

O teste relacionado ao IDI mostrou leve desvantagem em relação ao NRI, quando simulamos sob H_0 . Neste caso, para todos os tamanhos de amostra o teste do IDI apresentou taxas de rejeição acima do nível nominal, demonstrando ser um teste menos conservador, no sentido que rejeita mais do que o valor nominal. Para os cenários que consideramos estar simulando muito próximos da situação sob H_0 , o teste do IDI rejeitou mais do que o esperado. Assim, qualquer mudança no comportamento do risco estimado, quando incorporado um novo marcador ao modelo, parece não passar despercebida pelo teste do IDI. Quando simulamos sob a hipótese alternativa, o teste do IDI apresentou os melhores resultados do que o teste do NRI, mostrando ser muito poderoso. Com exceção do tamanho de amostra 100, o teste relacionado ao IDI apresentou taxas de rejeição para os cenários de **2** a **7** sempre acima de 40,0%.

Capítulo 5

Análise de um Estudo de Cardiomiopatia Chagásica

5.1 Introdução

Este trabalho foi motivado por estudos do grupo de pesquisa Ecocardiografia nas Doenças Tropicais do Programa de Pós-Graduação de Infectologia e Medicina Tropical da Universidade Federal de Minas Gerais (UFMG). Este grupo vem desenvolvendo pesquisas para incorporar novos marcadores, desenvolvidos a partir do exame do ecocardiograma, aos modelos de predição de risco de morte precoce em pacientes com doença de Chagas (12, 13). Neste capítulo fazemos a análise do banco de dados referente aos pacientes com cardiomiopatia chagásica. O objetivo da análise foi verificar se a inclusão do marcador E/E' no modelo de predição de risco, melhora o desempenho do modelo. Consideramos o risco de morte precoce, ou seja, morte até um ano após o diagnóstico (em particular, 13 meses equivale a um ano em nossa análise). Implementamos as rotinas da análise através da linguagem de programação R 2.9.0 (26). Na Seção 5.2 fazemos uma breve descrição dos dados e apresentamos os resultados para o ajuste dos modelos. Na Seção 5.3 fazemos a comparação dos modelos utilizando as técnicas apresentadas no Capítulo 3. Por fim, na Seção 5.4 fazemos uma análise de sensibilidade para a melhora da reclassificação líquida (NRI).

5.2 Descrição dos Dados e Ajuste do Modelo

Os dados deste estudo se referem a 224 pacientes com cardiomiopatia chagásica, diagnosticados no período de 1999 a 2008 e acompanhados até o mês de outubro de 2009, no Hospital das Clínicas da Universidade Federal de Minas Gerais. Após a exclusão de indivíduos com valores faltantes, 158 pacientes permaneceram na análise. Os pacientes foram acompanhados por pelo menos 13 meses, ou até a ocorrência da morte. O tempo mediano de acompanhamento foi 37,59 meses. Dos 158 pacientes 24 morreram antes de 13 meses.

Utilizamos o modelo de Cox para posteriormente estimarmos os riscos de morte precoce para cada paciente. Dois modelos foram ajustados, sem e com a covariável E/E' (“*modelo antigo*” e “*modelo novo*”). As covariáveis comuns aos dois modelos foram o volume do átrio esquerdo (volae), a fração de ejeção do ventrículo esquerdo (fe), a classe funcional segundo classificação de NYHA

(*New York Heart Association*) e o índice de Tei do ventrículo direito (vdtei). A classe funcional foi dicotomizada (classe I e II, NYHA = 0; classe III e IV, NYHA = 1). Os coeficientes estimados dos modelos, assim como as razões de risco (HR), e níveis de significância para as covariáveis são apresentados na Tabela 5.1. A razão E/E' foi altamente significativa (HR = 0,88, valor $p < 0,001$), e o AIC diminuiu de 173,99 para 167,42 em favor do modelo que inclui E/E'.

Tabela 5.1: Resultados do modelo de Cox para os modelos sem e com E/E' do estudo de cardiomiopatia chagásica.

Modelo sem E/E'				
Covariável	Coeficiente- β	Erro Padrão	Valor p	HR
volae	0,0255	0,0104	0,0147	1,0258
fe	-0,0535	0,0263	0,0421	0,9479
NYHA	1,5358	0,5829	0,0084	4,6449
vdtei	1,2862	0,5102	0,0117	3,6190
Modelo com E/E'				
Covariável	Coeficiente- β	Erro Padrão	Valor p	HR
volae	0,0325	0,0105	0,0020	1,0331
fe	-0,0687	0,0269	0,0106	0,9337
NYHA	1,5509	0,6165	0,0119	4,7156
vdtei	1,7347	0,5002	0,0005	5,6671
E/E'	-0,1254	0,0378	0,0009	0,8821

5.3 Comparação dos Modelos

A partir dos resultados da Tabela 5.1, e das estimativas para a função de sobrevivência de base no tempo 13 ($\hat{S}_0(13) = 0,9578$ para o modelo sem E/E', e $\hat{S}_0(13) = 0,8502$ para o modelo com E/E'), estimamos os riscos de morte precoce para cada indivíduo, utilizando a Expressão (2.13). Foram utilizados os limiares de risco $p_L = 0,02$ e $p_H = 0,10$ para estipular as classes de baixo, médio e alto risco. Na Figura 5.1 apresentamos o gráfico de dispersão dos riscos estimados pelos dois modelos. Por este gráfico, é possível ver que a inclusão da razão E/E' ao modelo fez o risco aumentar na maioria dos eventos (pontos que estão acima da reta de referência). Apesar disto não houve reclassificação entre os eventos. Entre os não-eventos, aparentemente não houve mudança no comportamento dos riscos, com a inclusão da razão E/E'.

Na Figura 5.2 apresentamos as curvas ROC para os dois modelos. Por esta figura, é visualmente nítido que não há mudança na curva ROC com a inclusão da covariável E/E' no modelo. As áreas sob a curva (AUC) foram estimadas em 0,928 para o modelo sem E/E' e 0,942 para o modelo com E/E'. O aumento da AUC não foi significativo (valor $p = 0,731$) utilizando o método proposto por (23) para a comparação das AUC's.

A Tabela 5.2 apresenta os resultados da estratificação dos riscos dos modelos sem e com E/E'.

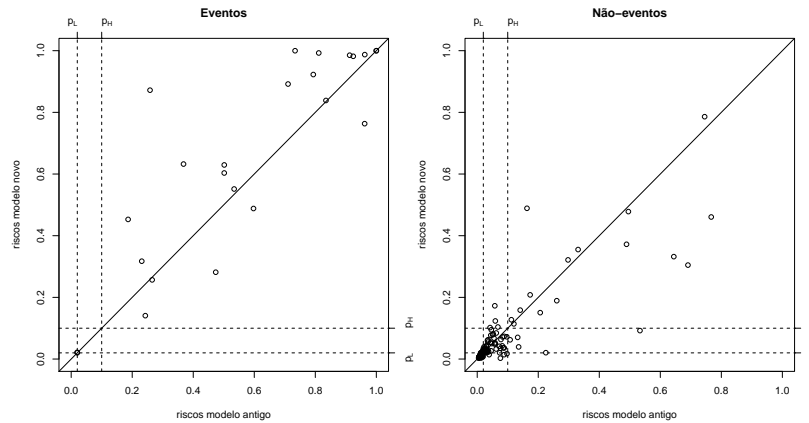


Figura 5.1: Gráfico de dispersão entre riscos para o estudo de cardiomiopatia chagásica.

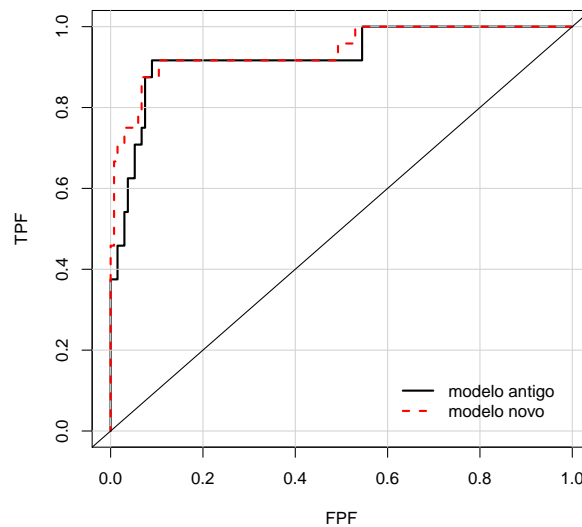


Figura 5.2: Curva ROC dos riscos estimados dos modelos sem e com E/E' para o estudo de cardiomiopatia chagásica.

Como dissemos na Seção 3.3, através da tabela de estratificação de risco podemos acessar informações a respeito da calibração dos modelos, capacidade de estratificação e acurácia da classificação. Observando as proporções de eventos nas marginais da Tabela 5.2, percebemos que o modelo sem a razão E/E' não está bem calibrado, pois 1,85% dos indivíduos da classe “2 a 10%” são eventos. Neste aspecto o modelo com a razão E/E' leva pequena vantagem em relação ao modelo que não a inclui. Quanto a capacidade de estratificação, o desempenho dos dois modelos foi muito semelhante. O modelo sem E/E' designa aproximadamente 34% dos indivíduos à classe intermediária (“2 a 10%”), enquanto o modelo com E/E' designa 35% dos indivíduos à classe intermediária. A acurácia da classificação também foi praticamente igual para os dois modelos. Da

Tabela 5.2 estimamos $TPF(p_H) = 92\%$ para os dois modelos e $FPF(p_H) = 15\%$ para o modelo sem E/E' e $FPF(p_H) = 14\%$ para o modelo com E/E'.

Tabela 5.2: Tabela de estratificação: Risco de morte em um até um ano pelos modelos sem e com E/E'.

		Risco 1 ano modelo sem E/E'			Risco 1 ano modelo com E/E'		
		0-2%	2-10%	$\geq 10\%$	Total		
0-2%	N	57	5	0	62		
	Eventos	0	1	0	1		
	Não-eventos	57	4	0	61		
	Prop. Eventos	0,00	20,00	-	1,61		
2-10%	N	4	46	4	54		
	Eventos	0	1	0	1		
	Não-eventos	4	45	4	53		
	Prop. Eventos	0,00	2,17	0,00	1,85		
$\geq 10\%$	N	0	5	37	42		
	Eventos	0	0	22	22		
	Não-eventos	0	5	15	20		
	Prop. Eventos	-	0,00	59,46	52,38		
Total	N	61	56	41	158		
	Eventos	0	2	22	24		
	Não-eventos	61	54	19	134		
	Prop. Eventos	0,00	3,57	53,66	15,19		

A semelhança no desempenho dos modelos também pode ser vista pela Figura 5.3, onde apresentamos a curva de previsibilidade referente aos dois modelos. As curvas estão muito próximas. Para o limiar de baixo risco $p_L = 0,02$, nos dois modelos, aproximadamente 39% dos indivíduos têm riscos estimados abaixo de p_L . Notamos que ao destacar os limiares de baixo e alto risco, obtemos a mesma informação da Tabela 5.2 em relação ao poder de estratificação dos modelos.

Por fim, testamos a inclusão da razão E/E' utilizando as estatísticas NRI e IDI. O NRI foi estimado em 0,0491 e não foi significativo (valor $p = 0,3429$), nos apontando que a razão E/E' não acrescenta valor preditivo ao modelo. Entretanto, o IDI foi estimado em 0,0873 e foi significativo (valor $p = 0,0145$). Observe que os resultados dos dois testes são inconsistentes. Esta inconsistência já foi observada no Capítulo 4. Os resultados das simulações com tamanhos de amostra parecidos com o tamanho de amostra do banco de dados (100 e 300), em situações que o comportamento dos riscos estimados foi parecido com o que encontramos na análise, o teste do NRI se comportou melhor do que o teste do IDI. Além disso, todas as outras técnicas apontam que a razão E/E' não acrescenta no desempenho do modelo de predição de risco de morte precoce em pacientes com cardiomiopatia chagásica.

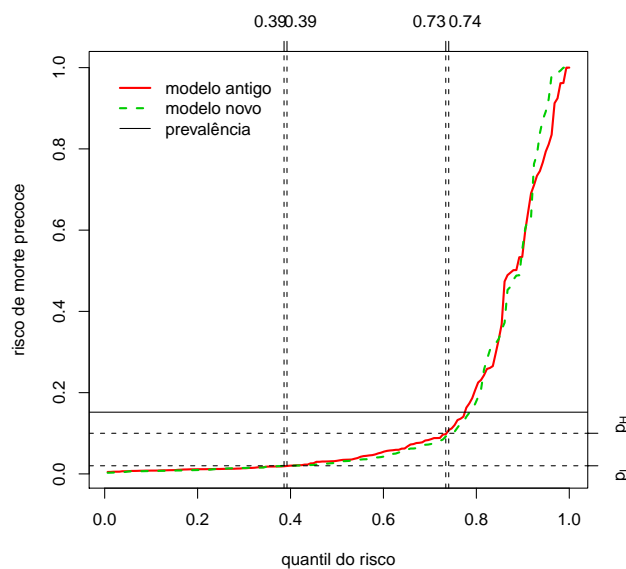


Figura 5.3: Curva de previsibilidade dos riscos estimados dos modelos sem e com E/E' para o estudo de cardiomiopatia chagásica.

5.4 Análise de Sensibilidade do NRI

Como já mencionamos, a melhora da reclassificação líquida (NRI) é dependente dos limiares de risco. A escolha desses limiares é feita com base na experiência clínica. Para avaliar quanto o resultado que encontramos na Seção 5.3 para o NRI pode ter sido influenciado pela definição das classes de risco, fizemos uma análise de sensibilidade variando os limiares p_L e p_H . Apresentamos os resultados desta análise na Tabela 5.3. Notamos que para algumas combinações de p_L e p_H , os testes forneceram resultados significativos (valor $p < 0,05$). Em especial, quando utilizamos o ponto 0,6, tanto como limiar de baixo como de alto risco, os resultados dos testes foram significativos ou marginalmente significativos. Para as demais combinações, os resultados não foram significativos, reforçando a conclusão de que a razão E/E' não melhora o desempenho do modelo de predição de risco de morte precoce em pacientes com cardiomiopatia chagásica. Devemos ter o devido cuidado ao analisar esta tabela, pois realizamos 36 testes.

Tabela 5.3: Resultados da análise de sensibilidade para o NRI.

p_L	p_H	NRI	Valor p
0,1	0,2	0,0075	0,9083
	0,3	0,0759	0,3828
	0,4	0,1057	0,2273
	0,5	0,0715	0,3509
	0,6	0,1965	0,0243
	0,7	0,0566	0,2373
	0,8	0,1325	0,1669
	0,9	0,0908	0,2927
0,2	0,3	0,0491	0,6060
	0,4	0,0790	0,4098
	0,5	0,0790	0,4098
	0,6	0,2040	0,0509
	0,7	0,0641	0,3919
	0,8	0,1399	0,2105
	0,9	0,0983	0,3433
0,3	0,4	0,1399	0,1406
	0,5	0,0983	0,3433
	0,6	0,2233	0,0457
	0,7	0,0833	0,3250
	0,8	0,1592	0,1793
	0,9	0,1592	0,1793
0,4	0,5	0,0715	0,4513
	0,6	0,1965	0,0581
	0,7	0,1132	0,1846
	0,8	0,1891	0,1122
	0,9	0,1891	0,1122
0,5	0,6	0,1549	0,1008
	0,7	0,0715	0,3318
	0,8	0,1549	0,1639
	0,9	0,1549	0,1639
0,6	0,7	0,1891	0,0250
	0,8	0,2724	0,0216
	0,9	0,2724	0,0216
0,7	0,8	0,1325	0,1564
	0,9	0,1325	0,1564
0,8	0,9	0,1667	0,1025

Capítulo 6

Conclusões

6.1 Considerações Finais

A estimação do risco de um determinado evento é de grande importância para as decisões médicas. Quanto melhor o desempenho do modelo para predição do risco, maior a qualidade das ações médicas tomadas. Com o avanço das tecnologias, novos marcadores têm surgido, e desta forma, novas formulações para a estimação do risco são investigadas. Nesta dissertação apresentamos algumas das mais recentes técnicas estatísticas para avaliação de marcadores e modelos de predição de risco. Utilizamos o modelo de Cox como modelo de predição de risco, mas salientamos que outros modelos podem ser utilizados para a estimação do risco, e a avaliação do desempenho destes modelos poderia ser feita da mesma forma como ilustramos aqui.

A análise da curva ROC tem mostrado sérias limitações na avaliação de modelos de predição de risco e não deve ser a única ferramenta utilizada para tal objetivo. Quando limiares de risco podem ser especificados, sugerimos a utilização da tabela de estratificação de risco (1), pois esta técnica consegue descrever as características essenciais de um modelo de predição de risco. A curva de previsibilidade (8,9) é uma boa alternativa, quando não existem limiares de risco pré-estabelecidos. Este gráfico permite a comparação de dois ou mais modelos pois utiliza uma escala em comum, os quantis do risco.

Através de simulações Monte Carlo, fizemos comparações em relação aos testes de hipóteses propostos para verificar a significância das estatísticas melhora da reclassificação líquida (NRI) e melhora da discriminação integrada (IDI) (3). Os resultados que encontramos mostram que sob a hipótese nula, o teste do NRI apresentou taxas de rejeição mais próximas ao nível nominal do que o teste do IDI. Sob a hipótese alternativa, o teste do IDI apresentou mais poder em relação ao teste do NRI e é menos sensível à variação do tamanho de amostra.

Utilizamos as técnicas apresentadas nesta dissertação para avaliar a inclusão da razão E/E' no desempenho do modelo de predição de risco de morte precoce em pacientes cardiopatas com doença de Chagas. Todas as técnicas, com exceção do IDI, apontaram que a razão E/E' não acrescenta valor preditivo ao modelo.

Como sugestões de pesquisas futuras, sugerimos a avaliação do desempenho dos testes relacionados ao NRI e IDI, estimando a variância das estatísticas NRI e IDI via método de *bootstrap*. Essencialmente, o IDI pode ter um ganho com este procedimento pois a aproximação da estatística de teste é muito lenta para a normal padrão.

Apêndice A

Gráficos do Estudo de Simulação

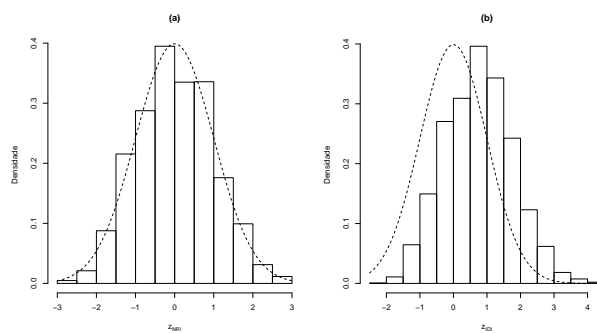


Figura A.1: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 1**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

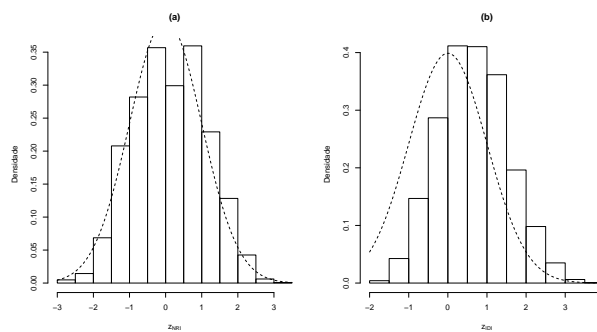


Figura A.2: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 1**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

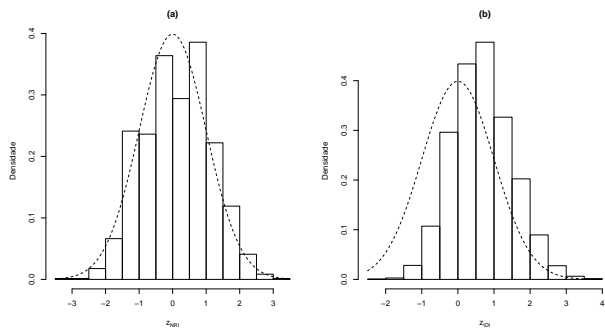


Figura A.3: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 1**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

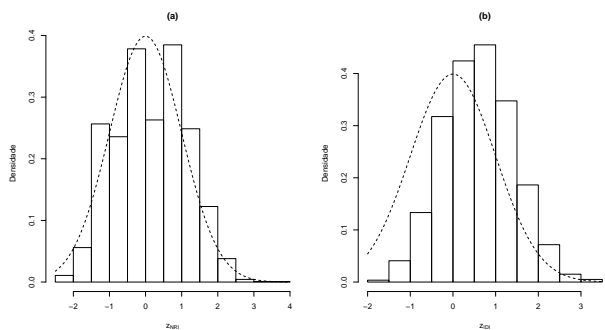


Figura A.4: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 1**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

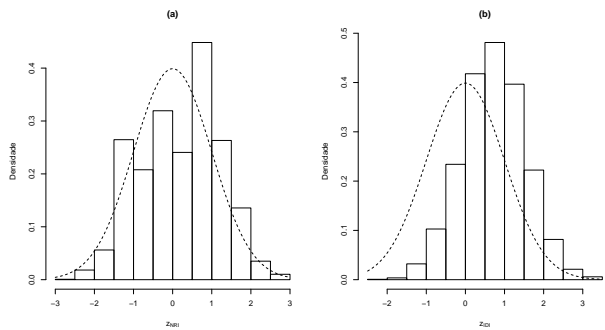


Figura A.5: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 1**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

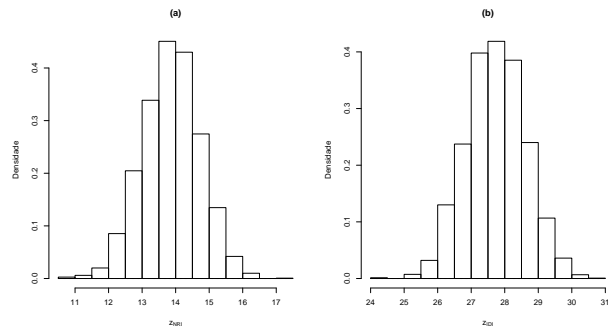


Figura A.6: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 2**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$.

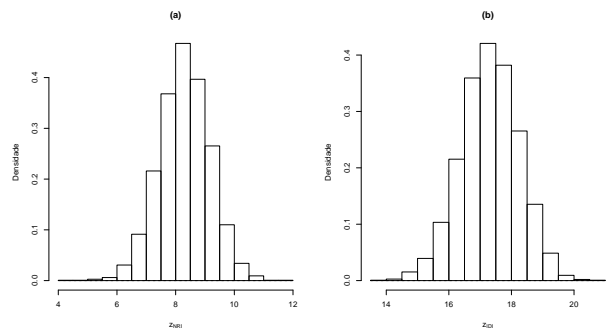


Figura A.7: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 2**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$.

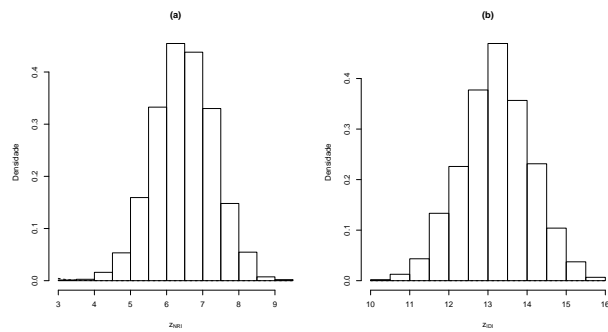


Figura A.8: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 2**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$.

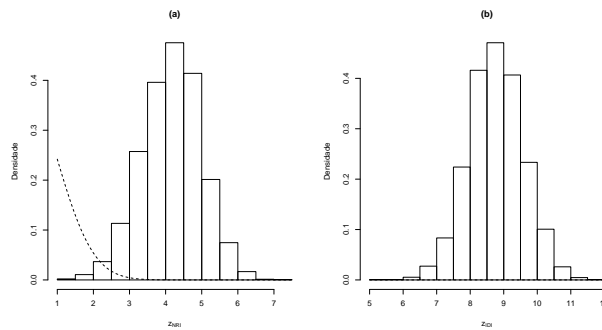


Figura A.9: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 2**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

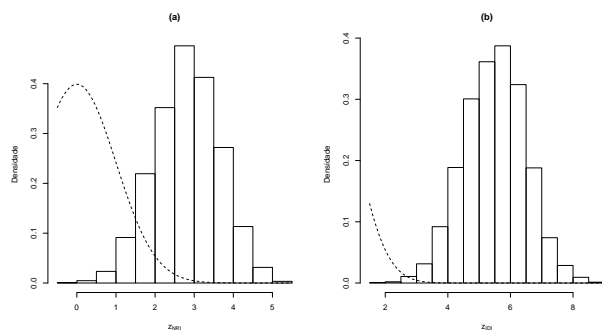


Figura A.10: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 2**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

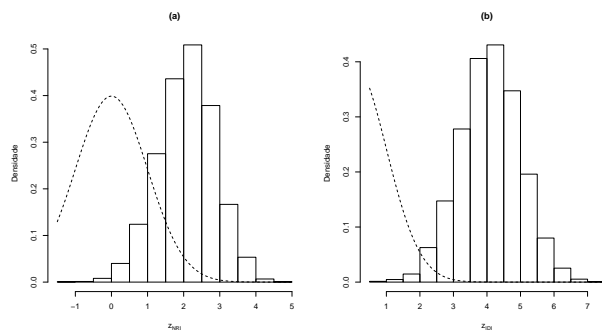


Figura A.11: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 2**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

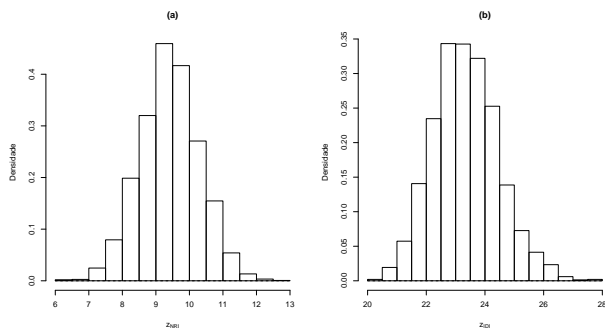


Figura A.12: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 3**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$.

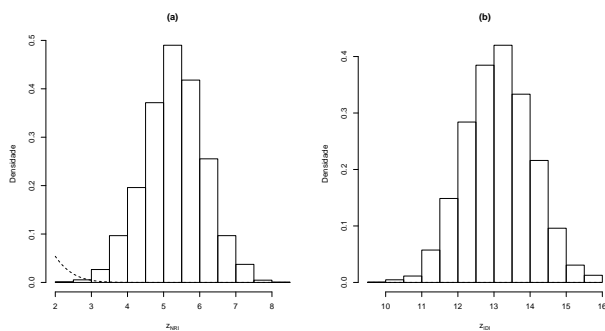


Figura A.13: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 3**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

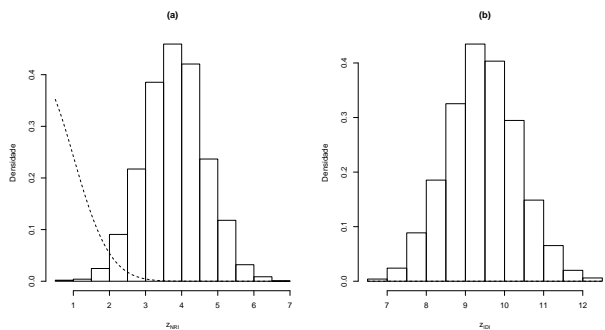


Figura A.14: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 3**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

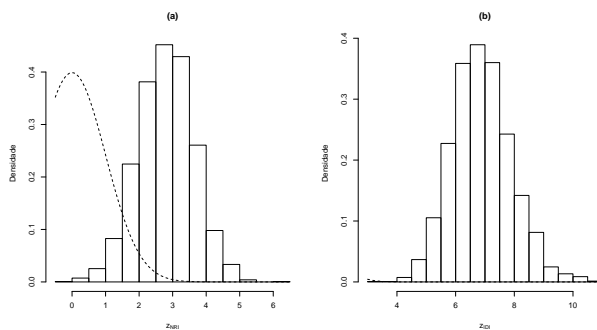


Figura A.15: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 3**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

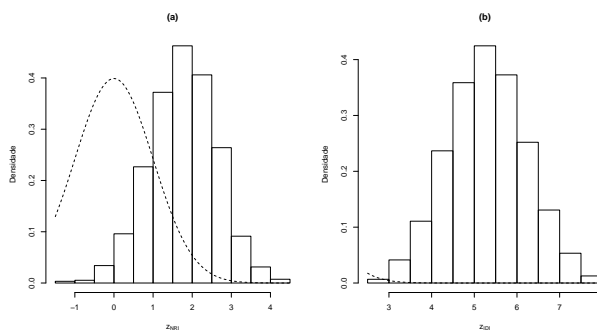


Figura A.16: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 3**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

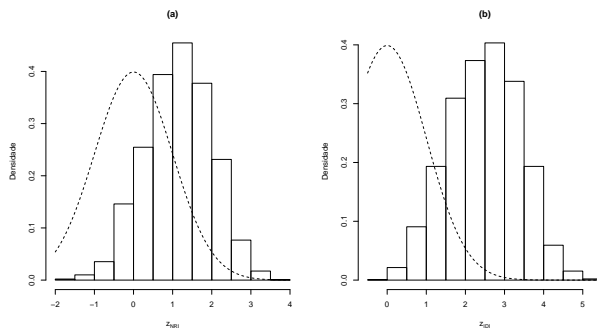


Figura A.17: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 3**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

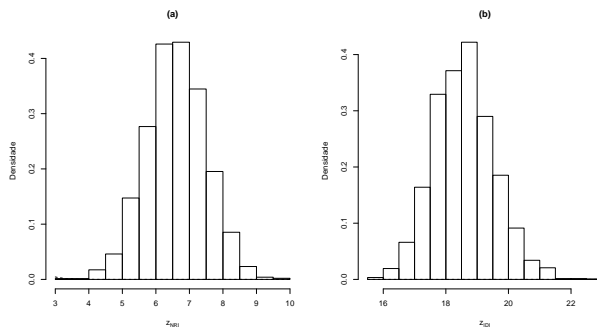


Figura A.18: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 4**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$.

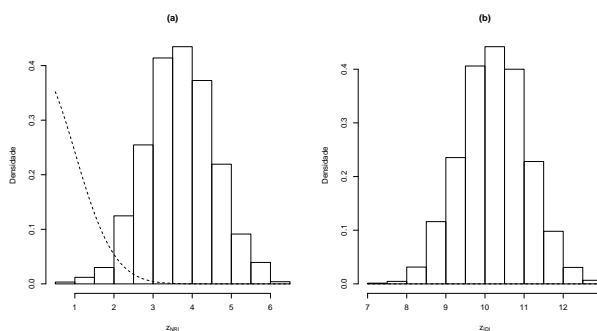


Figura A.19: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 4**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

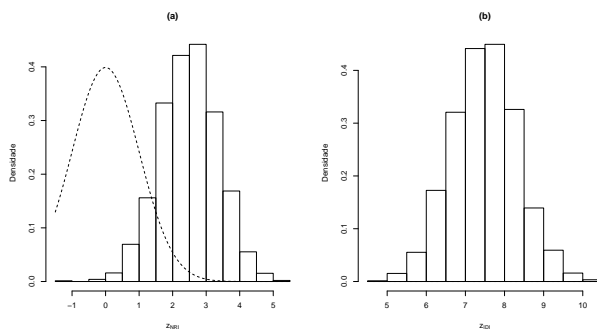


Figura A.20: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 4**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

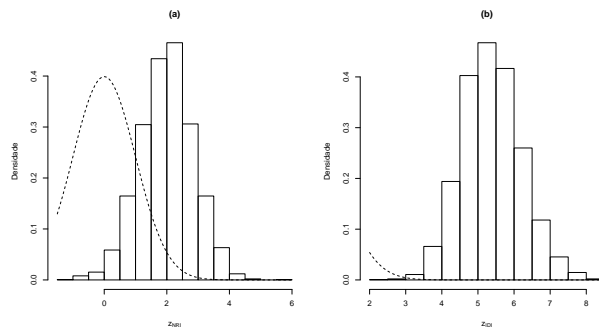


Figura A.21: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 4**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

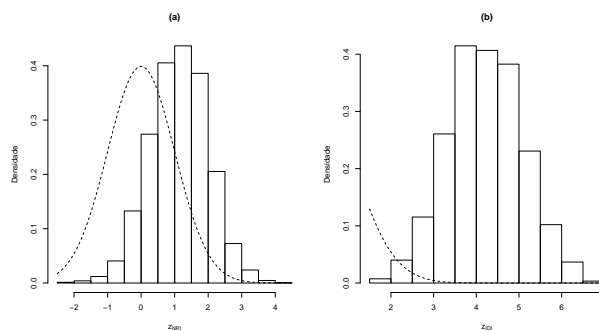


Figura A.22: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 4**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

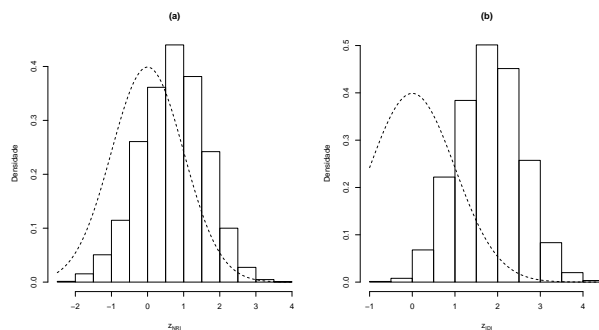


Figura A.23: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 4**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

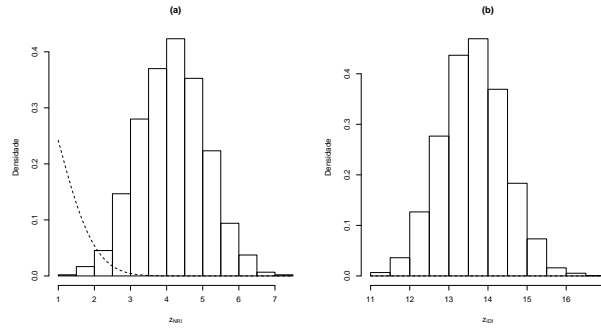


Figura A.24: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 5**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$; a linha racejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

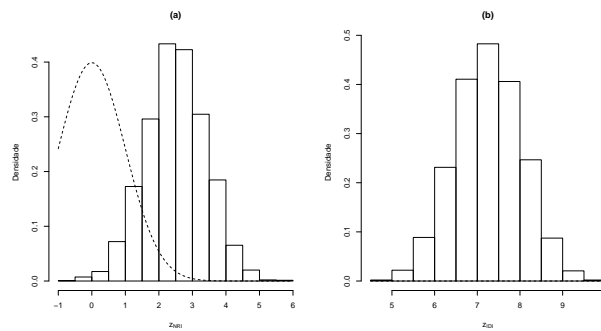


Figura A.25: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 5**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

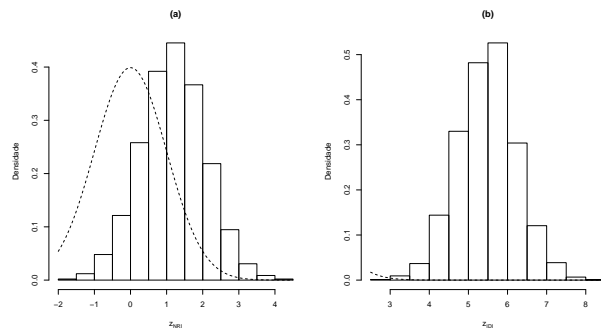


Figura A.26: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 5**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

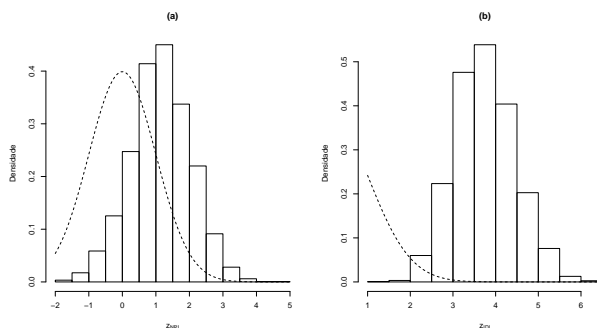


Figura A.27: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 5**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

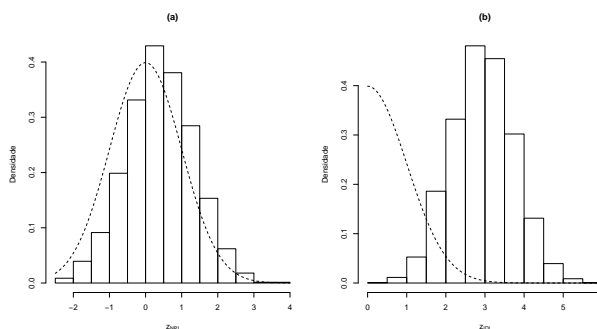


Figura A.28: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 5**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

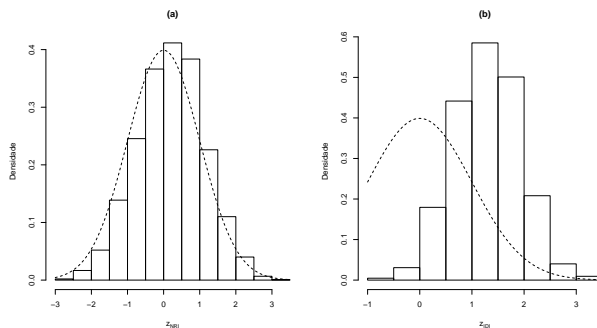


Figura A.29: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 5**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

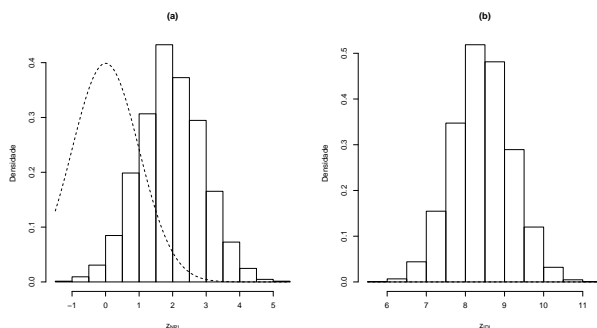


Figura A.30: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 6**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

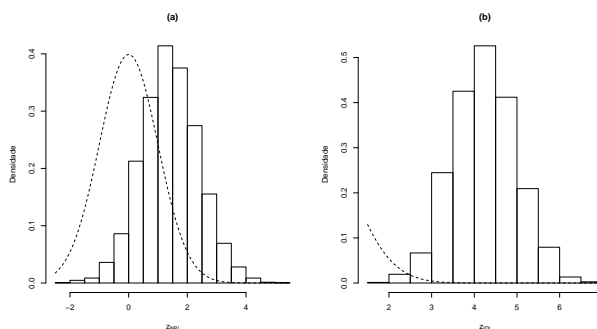


Figura A.31: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 6**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

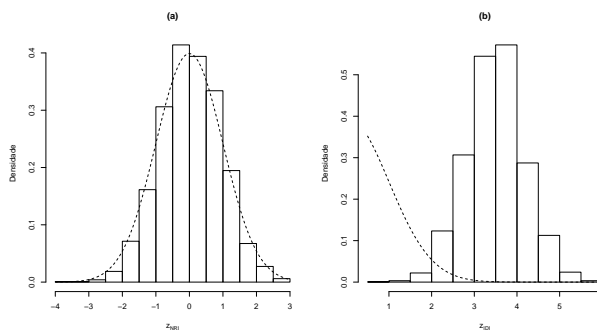


Figura A.32: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 6**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

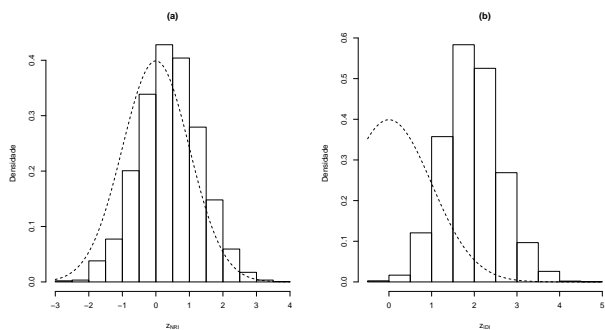


Figura A.33: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 6**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

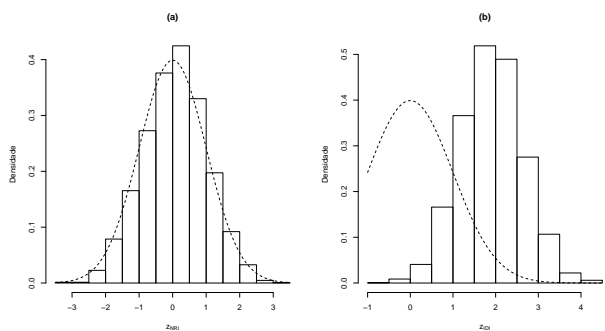


Figura A.34: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 6**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

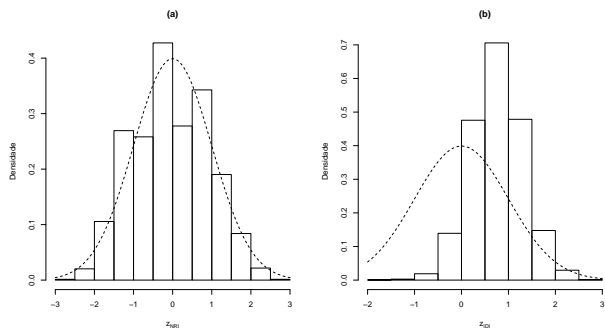


Figura A.35: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 6**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

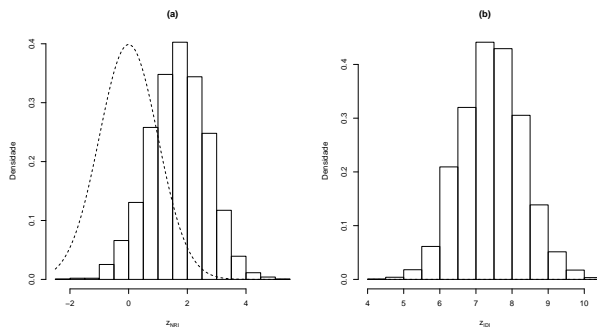


Figura A.36: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 7**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

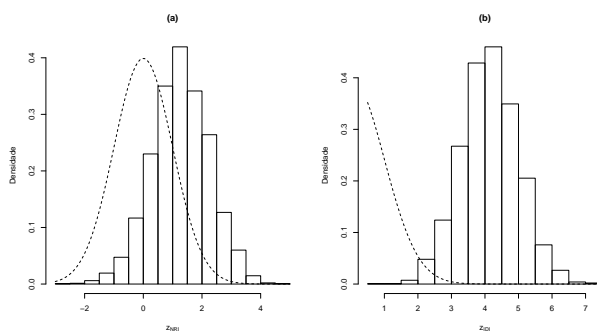


Figura A.37: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 7**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

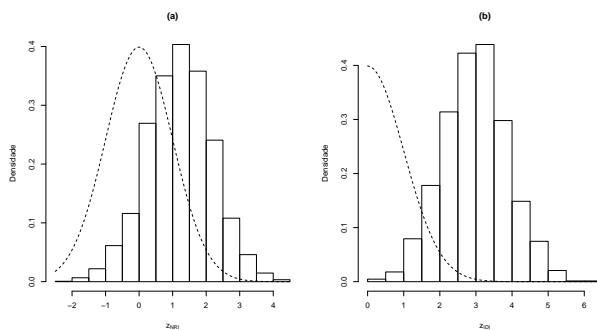


Figura A.38: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 7**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

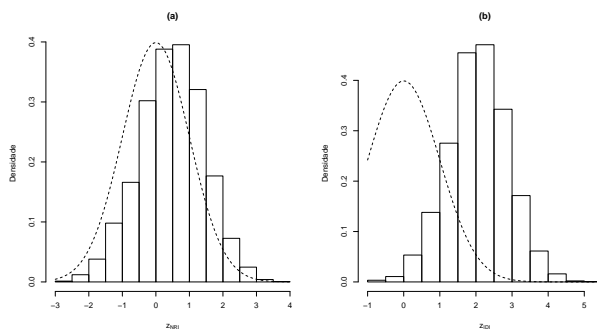


Figura A.39: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 7**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

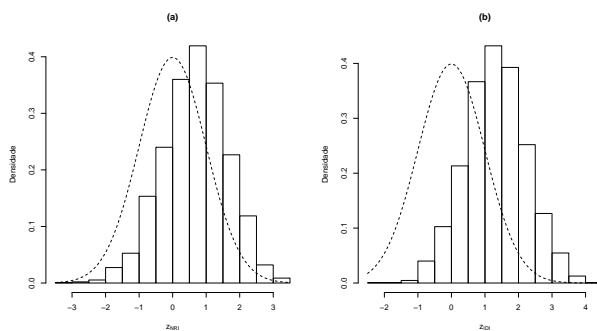


Figura A.40: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 7**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

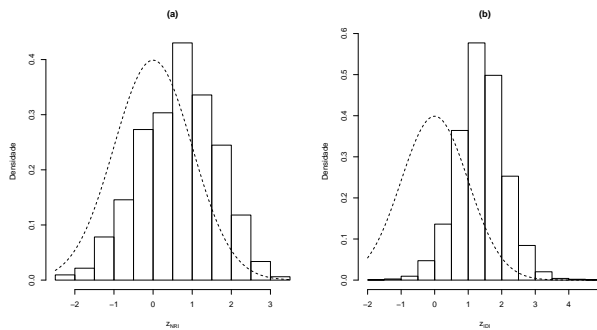


Figura A.41: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 7**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

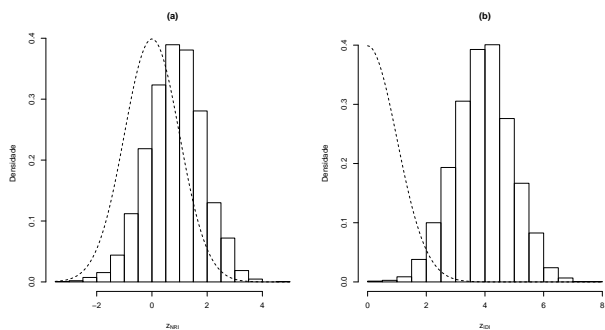


Figura A.42: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 8**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

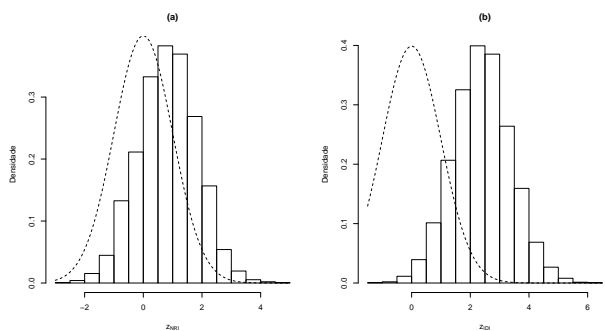


Figura A.43: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 8**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

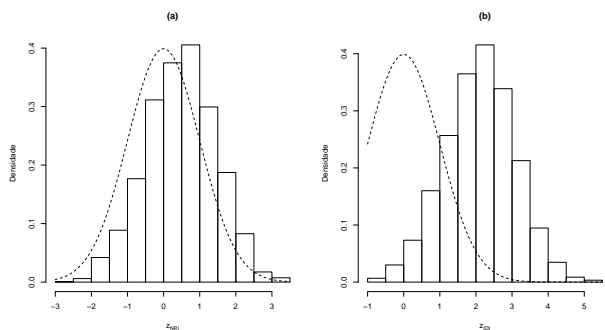


Figura A.44: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 8**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

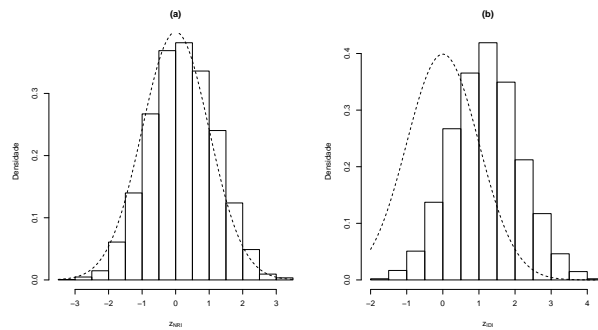


Figura A.45: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 8**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

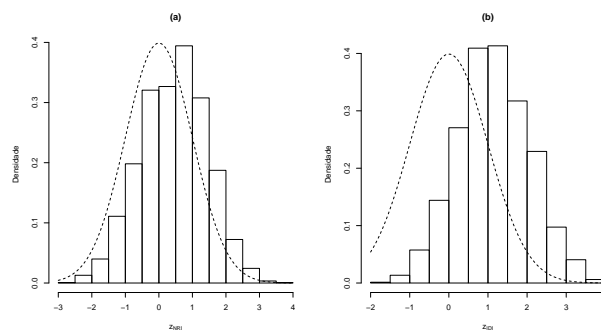


Figura A.46: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 8**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

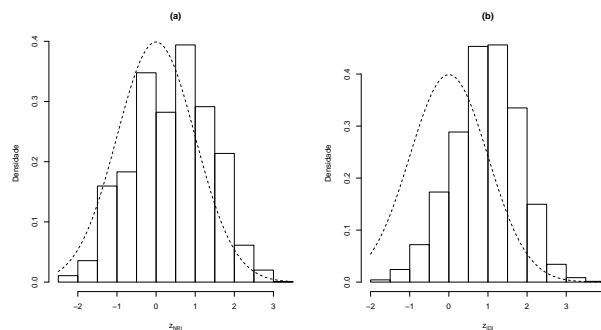


Figura A.47: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 8**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

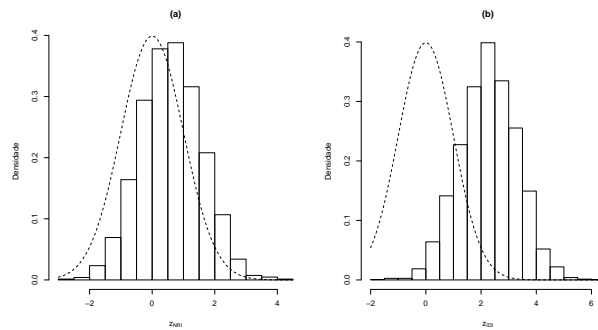


Figura A.48: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 9**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

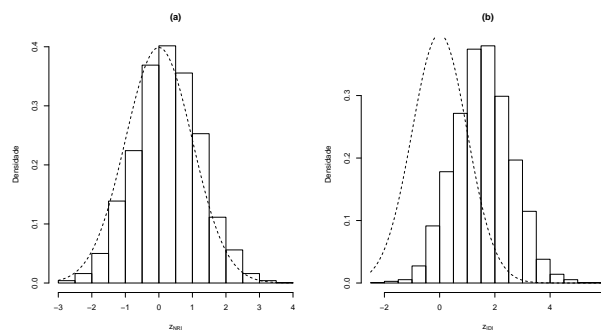


Figura A.49: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 9**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

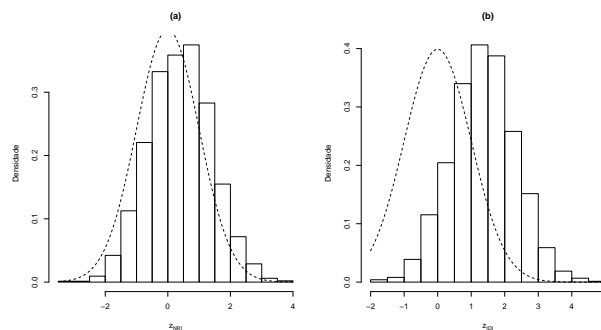


Figura A.50: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 9**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

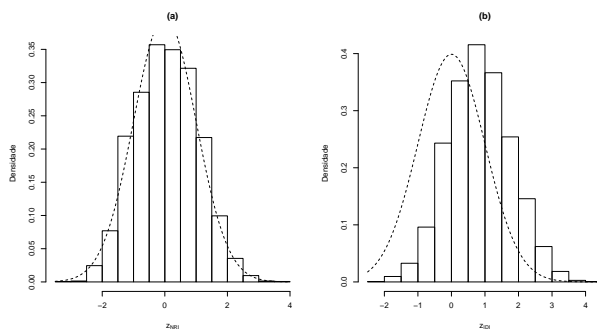


Figura A.51: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 9**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

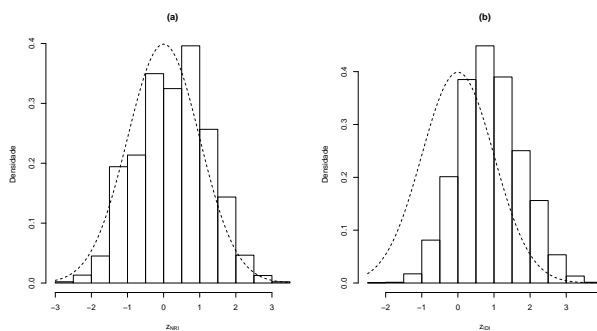


Figura A.52: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 9**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

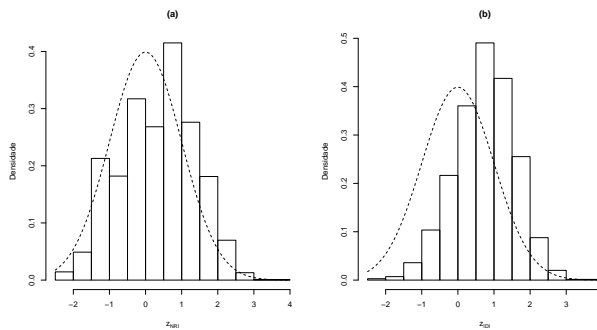


Figura A.53: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 9**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

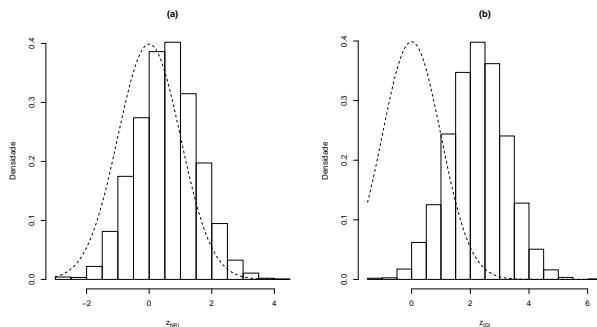


Figura A.54: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 10**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 5.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

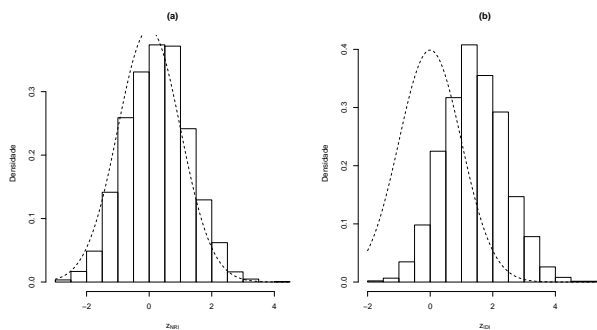


Figura A.55: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 10**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 2.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

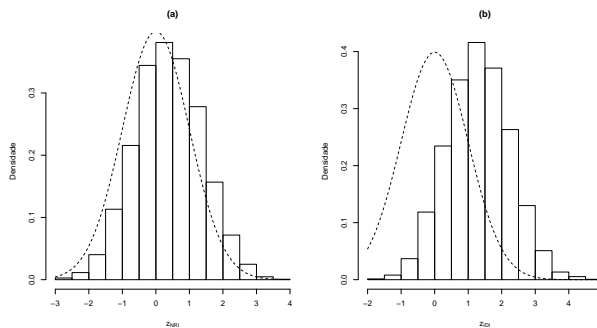


Figura A.56: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 10**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 1.000$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

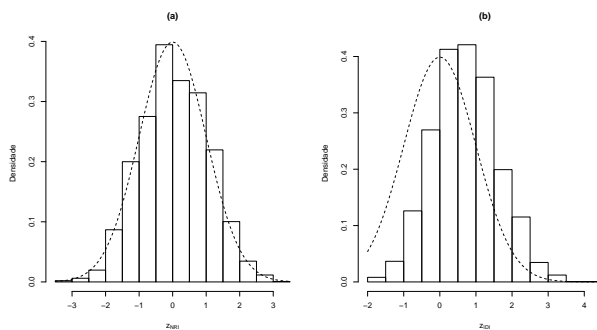


Figura A.57: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 10**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 500$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

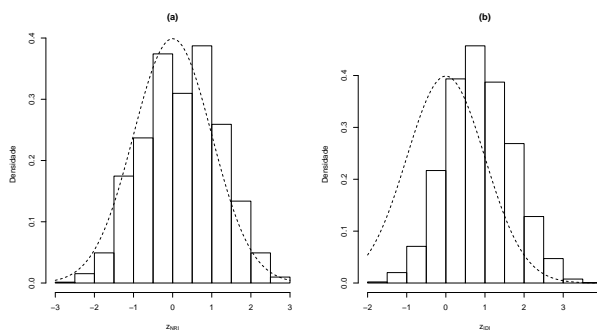


Figura A.58: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 10**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 300$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

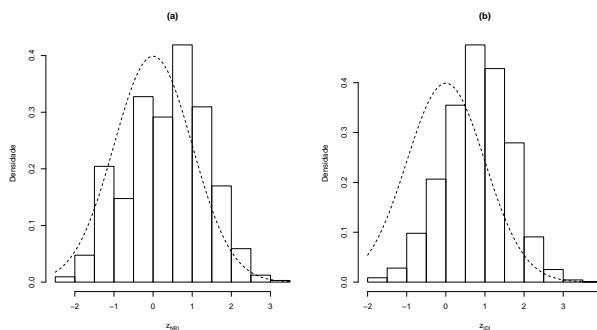


Figura A.59: Histograma de 3.000 réplicas das estatísticas de teste para o **Cenário 10**: (a) z_{NRI} ; (b) z_{IDI} ; $n = 100$; a linha tracejada representa a distribuição de referência sob H_0 , a distribuição normal padrão.

Referências Bibliográficas

- 1 JANES, H.; PEPE, M. S.; GU, W. Accessing the value of risk predictions by using risk stratification tables. *Annals of Internal Medicine*, v. 149, p. 751–760, 2008. 1, 2, 3, 13, 14, 15, 39
- 2 GU, W.; PEPE, M. S. Estimating the capacity for improvement in risk prediction with a marker. *Biostatistics*, v. 10, p. 172–186, 2008. 1, 3, 20
- 3 PENCINA, M. J.; D’AGOSTINO Sr., R. B.; D’AGOSTINO Jr., R. B.; VASAN, R. S. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine*, v. 27, p. 157–172, 2008. 1, 2, 3, 13, 17, 18, 19, 21, 22, 39
- 4 HANLEY, J. A.; MCNEIL, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, v. 143, p. 29–36, 1982. 2, 11, 13
- 5 PEPE, M. S.; JANES, H.; LONGTON, G.; LEISENRING, W.; NEWCOMB, P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology*, v. 159, p. 882–889, 2004. 2, 13
- 6 COOK, N. R.; BURING, J. E.; RIDKER, P. M. The effect of including C-reactive protein in cardiovascular risk prediction models for women. *Annals of Internal Medicine*, v. 145, p. 21–29, 2006. 2, 3, 13
- 7 COOK, N. R. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*, v. 115, p. 928–935, 2007. 2, 3, 13
- 8 PEPE, M. S.; FENG, Z.; HUANG, Y.; LONGTON, G.; PRENTICE, R.; THOMPSON, I. M.; ZHENG, Y. Integrating the predictiveness of a marker with its performance as a classifier. *American Journal of Epidemiology*, v. 167, p. 362–368, 2007. 2, 3, 13, 15, 16, 39
- 9 HUANG, Y.; PEPE, M. S.; FENG, Z. Evaluating the predictiveness of a continuous marker. *Biometrics*, v. 63, p. 1181–1188, 2007. 2, 3, 15, 39
- 10 ANONYMOUS. Chagas’ disease - an endemic that can no longer be ignored. *Lancet*, v. 368, p. 619, 2006. Editorial. 2, 3
- 11 WORLD HEALTH ORGANIZATION. *Control of Chagas disease*. Geneva, 2002. 3
- 12 NUNES, M. C. P.; ROCHA, M. O. C.; RIBEIRO, A. L. P.; COLOSIMO, E. A.; REZENDE, R. A.; CARMO, G. A. A.; BARBOSA, M. M. Rightventricular dysfunction is an independent predictor of survival in patients with dilated chronic Chagas’ cardiomyopathy. *International Journal of Cardiology*, v. 127, p. 372–379, 2007. 3, 33

- 13 RIBEIRO, A. L. P.; CAVALCANTI, P. S.; LOMBARDI, F.; NUNES, M. C. P.; BARROS, M. V. L.; ROCHA, M. O. C. Prognostic value of signal-averaged electrocardiogram in Chagas disease. *Journal of Cardiovascular Electrophysiology*, v. 19, p. 502–509, 2008. 3, 33
- 14 OMMEN, S. R.; NISHIMURA, R. A.; APPLETON, C. P.; MILLER, F. A.; OH, J. K.; REDFIELD, M. M.; TAJIK, A. J. Clinical utility of doppler echocardiography and tissue doppler imaging in the estimation of left ventricular filling pressures: a comparative simultaneous doppler catheterization study. *Circulation*, v. 102, p. 1788–1794, 2000. 3
- 15 REIS, R. C. P.; COLOSIMO, E. A.; NUNES, M. C. P.; RIBEIRO, A. L. P. (Poster), *Estudo de Técnicas para Avaliação de Novos Marcadores de Risco - Análise de um Estudo de Cardiopatia Chagásica*. São Carlos, 2009. 54^a Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria e 13^o Simpósio de Estatística Aplicada à Experimentação Agrônômica. 3, 14
- 16 COX, D. R.; OAKES, D. *Analysis of Survival Data*. London: Chapman and Hall, 1984. (Monographs on Statistics and Applied Probability). 5, 9
- 17 COLOSIMO, E. A.; GIOLO, S. R. *Análise de Sobrevida Aplicada*. São Paulo: Edgard Blücher, 2006. (ABE - Projeto Fisher). 6, 7, 8, 9
- 18 COX, D. R. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, B*, v. 34, p. 187–220, 1972. 7, 8
- 19 COX, D. R. Partial likelihood. *Biometrika*, v. 66, p. 269–276, 1975. 8
- 20 BRESLOW, N. Covariance analysis of censored survival data. *Biometrics*, v. 30, p. 89–99, 1974. 8
- 21 HEAGERTY, P. J.; ZHENG, Y. Survival model predictive accuracy and ROC curves. *Biometrics*, v. 61, p. 92–105, 2005. 12, 13
- 22 PEPE, M. S. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford: Oxford University Press, 2003. (Oxford Statistical Science Series). 12, 13
- 23 HANLEY, J. A.; MCNEIL, B. J. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, v. 148, p. 839–843, 1983. 13, 34
- 24 GU, W.; PEPE, M. S. *Measures to Summarize and Compare the Predictive Capacity Markers*. February 2009. Working Paper 342. Disponível em: <www.bepress.com/uwbiostat/paper342>. 13, 16
- 25 PEPE, M. S.; FENG, Z.; GU, W. Comments on ‘Evaluating the added predictive ability of a new marker: from the area under the ROC curve to reclassification and beyond’. *Statistics in Medicine*, v. 27, p. 153–181, 2008. 13, 22
- 26 R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2009. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>. 21, 33