



**Universidade do Minho**  
Escola de Engenharia

Joana Margarida Rodrigues Barros de Araújo

**Avaliação do Desempenho de Indicadores com  
Base na Metodologia ROC [Receiver Operating  
Characteristic]**

Dissertação de Mestrado

**Mestrado em Bioinformática**

Trabalho efectuado sob a orientação de

**Professora Doutora Ana Cristina Braga**

Março de 2015

# Resumo

O objetivo deste estudo é avaliar e comparar o desempenho de dois indicadores de previsão de risco de mortalidade neonatal para recém-nascidos de muito baixo peso (<1500g), o CRIB (Clinical Risk Index for Babies) e o SNAPPE-II (Score for Neonatal Acute Physiology-Perinatal Extension II), com recurso à metodologia ROC (Receiver Operating Characteristic).

A execução prática deste estudo foi suportada com auxílio a programas estatísticos próprios para a análise da metodologia ROC, como o SPSS, ROCNPA, Comp2ROC, ROCR e caTools.

Os dados que contemplam o presente estudo foram recolhidos pelas unidades de cuidados intensivos neonatais do território português entre 2010 e 2012, e enviados para o Registo Nacional de Recém-Nascidos de Muito Baixo Peso (RNMBP), que é a entidade responsável pelo armazenamento desta informação.

Será aferida também a comparação e avaliação de variáveis de elevada expressão na previsão da mortalidade, que compõem os indicadores de mortalidade em estudo, sendo elas, o Peso à Nascimento e a Idade Gestacional.

A amostra em estudo é composta por 789 recém-nascidos de muito baixo peso, dos quais 51,3% são do género masculino. Em média os recém-nascidos em questão apresentam um peso médio ao nascimento de 1214 g  $\pm$ 343,1 e 29,8  $\pm$ 2,5 semanas de gestação e, dos integrantes na amostra 11,3% foram declarados óbitos hospitalares.

A exatidão dos indicadores de mortalidade e das variáveis foi obtida através do cálculo da AUC, área abaixo da curva ROC, que para o CRIB foi de 0,876  $\pm$ 0,025, para o SNAPPE-II de 0,867  $\pm$ 0,026, seguindo-se das variáveis idade gestacional e o peso ao nascimento com 0,785  $\pm$ 0,032 e 0,782  $\pm$ 0,028, respetivamente.

Com base nos resultados obtidos durante a elaboração do presente estudo, o CRIB provou ser melhor em prever a mortalidade para recém-nascidos de muito baixo peso, e tem a seu favor um menor número de variáveis comparativamente ao SNAPPE-II.

**Palavras-chave:** Recém-nascidos de muito baixo peso, CRIB, SNAPPE-II, metodologia ROC



# Abstract

The aim of this study is to evaluate and compare the performance of two scores of neonatal mortality risk prediction in very low birth weight infants (<1500g), CRIB (Clinical Risk Index for Babies) and the SNAPPE-II (Score for Neonatal Acute Physiology Perinatal Extension-II), using the ROC (Receiver Operating Characteristic) methodology.

The execution of this study was supported with statistical programs for the analysis of ROC methodology, such as SPSS, ROCNPA, Comp2ROC, ROCR and caTools.

The dataset used in present work was collected by the neonatal intensive care units of the Portuguese territory between 2010 and 2012, and sent to the Registo Nacional de Recém-Nascidos de Muito Baixo Peso (RNMBP), organization responsible by the storage of this information.

It will be also assess the comparison and evaluation of high expression variables in the mortality prediction, which are part of the risk mortality scores in question, that are, the Birth Weight and Gestational Age.

The dataset consists in 789 very low birth weight infants, of which 51.3% are males. The newborns in study have an average birth weight of 1214 g  $\pm$ 343.1g and an average gestational age of 29.8  $\pm$ 2.5 weeks, and 11.3% of selected newborns were declared hospital deaths.

The accuracy of the risk mortality scores and variables was obtained by calculating the area under the ROC curve (AUC), for the CRIB was 0.876  $\pm$ 0.025 and 0.867 SNAPPE-II  $\pm$ 0.026 following gestational age and birth weight with 0.785  $\pm$ 0.032 and 0.782  $\pm$ 0.028, respectively.

Based on the results obtained during the elaboration of this study, CRIB proved be better to predict mortality in very low birth weight infants, and has a minor number of variables compared to SNAPPE-II.

**keywords:** Very low birth weight infants, CRIB, SNAPPE-II, ROC methodology



# Agradecimentos

Os meus sinceros agradecimentos e reconhecimentos a todos os que me acompanharam e auxiliaram na execução da minha dissertação, bem como nestes dois anos de mestrado.

Em primeiro lugar, à minha orientadora **Professora Doutora Ana Cristina Braga**, pela forma como me orientou, pela disponibilidade ao longo destes meses de trabalho, pela partilha de conhecimentos, pelo esclarecimento de dúvidas e pelas sugestões na melhoria do meu trabalho. Obrigada pelo apoio, motivação, compreensão e amizade.

À minha família, em especial à minha **mãe, pai e irmã**, pelo carinho, motivação e compreensão ao longo desta trajetória.

Ao Grupo do Registo Nacional do Muito Baixo Peso pela cedência dos dados que possibilitaram a realização do presente trabalho, em especial à **Dr.<sup>a</sup> Alexandra Almeida** e à **Dr.<sup>a</sup> Gabriela Mimoso**.

Ao **Departamento de Produção e Sistemas**, pelas instalações disponibilizadas para a execução do meu trabalho. E a todas as pessoas com que tive a oportunidade de cruzar-me e privar durante estes meses.

Aos meus amigos, pela amizade, encorajamento e ajuda que me prestaram.



# **Dedicatória**

Aos meus pais e irmã,  
pelo apoio incondicional





# Conteúdo

<b>Resumo</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>Agradecimentos</b>	<b>v</b>
<b>Dedicatória</b>	<b>vii</b>
<b>Lista de Tabelas</b>	<b>xiii</b>
<b>Lista de Figuras</b>	<b>xv</b>
<b>Siglas e Acrónimos</b>	<b>xvii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	1
1.2 Objetivos . . . . .	2
1.3 Estrutura da Dissertação . . . . .	3
<b>2 Metodologia da Curva ROC</b>	<b>5</b>
2.1 Desempenho de Classificadores . . . . .	6
2.2 Espaço ROC . . . . .	9
2.2.1 Modelo Estatístico . . . . .	9
2.2.2 Curva ROC . . . . .	11
2.2.3 Área abaixo da curva . . . . .	13

2.2.4	Comparação de Desempenho de Amostras Emparelhadas . . . . .	18
2.3	Índices Clínicos de Mortalidade Neonatal:	
	CRIB e SNAPPE-II . . . . .	20
2.3.1	CRIB: Clinical Risk Index for Babies . . . . .	21
2.3.2	SNAPPE-II: Score for Neonatal Acute Physiology-Perinatal Extension Version II . . . . .	22
2.4	Programas Estatísticos para Análise ROC . . . . .	25
2.4.1	IBM® SPSS® Statistics . . . . .	25
2.4.2	R . . . . .	25
2.4.3	ROCNP . . . . .	26
<b>3</b>	<b>Aplicação da Metodologia ROC ao Caso em Estudo</b>	<b>27</b>
3.1	Caracterização da Amostra em Estudo . . . . .	27
<b>4</b>	<b>Comparação de Duas Amostras Emparelhadas Utilizando Curvas ROC</b>	<b>37</b>
4.1	Comparação dos indicadores de mortalidade: CRIB e SNAPPE-II . . . . .	38
4.1.1	Comparação dos indicadores de mortalidade através do <b>SPSS</b> . . . . .	38
4.1.2	Comparação dos índices de mortalidade através do <b>ROCR</b> . . . . .	40
4.1.3	Comparação dos índices de mortalidade através do <b>Comp2ROC</b> . . . . .	40
4.1.4	Comparação dos índices de mortalidade através do <b>caTools</b> . . . . .	42
4.1.5	Comparação dos índices de mortalidade através do <b>ROCNP</b> . . . . .	42
4.2	Comparação do desempenho das variáveis Peso ao Nascimento e Idade Gestacional . . . . .	45
4.3	Comparação do desempenho dos indicadores de mortalidade com a variável peso através do ROCNP 46	46
<b>5</b>	<b>Discussão</b>	<b>49</b>
<b>6</b>	<b>Conclusões e Trabalho Futuro</b>	<b>53</b>
	<b>Bibliografia</b>	<b>55</b>
	<b>Apêndices</b>	<b>59</b>
<b>A</b>	<b>Tabelas Relativas aos Índices</b>	<b>61</b>

<b>B</b>	<b>Tabela de Coeficientes de Correlação entre Áreas</b>	<b>63</b>
<b>C</b>	<b>Resultado de saída do pacote Comp2ROC</b>	<b>65</b>
<b>D</b>	<b>Teste de Comparação de Áreas</b>	<b>67</b>



## Lista de Tabelas

2.1	Classificação dos resultados de testes por estado de doença . . . . .	7
2.2	Pontuações do índice CRIB, adaptado de [30] . . . . .	22
2.3	Pontuações do índice SNAPPE-II, adaptado de [14] . . . . .	24
3.1	Características da amostra em estudo . . . . .	29
3.2	Características dos indivíduos sobreviventes e falecidos por PN e IG . . . . .	30
3.3	Estatísticas sumárias do Peso ao Nascimento . . . . .	31
3.4	Distribuição da mortalidade por IG . . . . .	32
3.5	Estatísticas da pontuação dos índices em estudo por estado, óbito e sobrevivo . . . . .	32
3.6	Valores da AUC, Erro Padrão e Intervalos de Confiança (95%) para o CRIB, SNPAPPE-II, IG e PN . . . . .	33
3.7	Distribuição dos recém-nascidos quanto aos Graus do CRIB e a respetiva mortalidade . . . . .	34
3.8	Distribuição dos recém-nascidos pelos cinco aos graus do SNAPPE-II e a respetiva mortalidade . . . . .	35
4.1	Correlação entre os índices para o estado de sobrevivência . . . . .	39
4.2	Correlação entre os índices para o estado de mortalidade . . . . .	39
4.3	Valores das áreas . . . . .	44
4.4	Testes de comparação múltipla para variáveis correlacionadas, valores de $p$ . . . . .	47
5.1	Resumo dos valores obtidos para o índice CRIB . . . . .	50
5.2	Resumo dos valores obtidos para o índice SNAPPE-II . . . . .	50
5.3	Resumo dos valores obtidos para as variáveis IG e PN . . . . .	51
5.4	Valores da AUC, para o CRIB, SNAPPE-II, PN e IG em vários estudos . . . . .	51

A.1	Componentes da escala APGAR, adaptado de [31] . . . . .	61
A.2	Valores limite para o peso consoante a idade gestacional, adaptado de [27] . . . . .	62
B.1	Tabela de Coeficientes de Correlação entre áreas, adaptado de [18] . . . . .	64

## Lista de Figuras

2.1	Exemplo de um modelo representativo das metodologia ROC, adaptado de [6] [22]. . . . .	10
2.2	Ponto no espaço ROC . . . . .	12
2.3	Curva ROC . . . . .	13
2.4	Representação de curvas ROC com 3 graus de discriminação diferentes . . . . .	14
2.5	Cálculo da AUC pela regra do trapézio . . . . .	16
3.1	Distribuição dos recém-nascidos de muito baixo peso por NUTS II . . . . .	28
3.2	Distribuição do número de neonatos por distritos . . . . .	29
3.3	Histograma representativo da distribuição da variável PN . . . . .	30
3.4	Peso ao nascimento por sobrevivido e falecido . . . . .	31
3.5	Curvas ROC empírica, do CRIB e dos Grupos do CRIB . . . . .	34
3.6	Curvas ROC, do SNAPPE-II e dos Grupos do SNAPPE-II . . . . .	36
4.1	Curvas ROC dos índices CRIB e SNAPPE-II . . . . .	38
4.2	Curva empírica através do <b>ROCR</b> . . . . .	40
4.3	Diferença de áreas entre as curvas e respetivos intervalos de confiança através do <b>Comp2ROC</b> . . . . .	41
4.4	Curvas empíricas através do <b>Comp2ROC</b> . . . . .	41
4.5	Curva Empírica através do <b>caTools</b> . . . . .	43
4.6	Curva Empírica através do <b>ROCNP</b> A . . . . .	43
4.7	Gráfico da distribuição dos indicadores com recurso ao <b>ROCNP</b> A . . . . .	44
4.8	Curvas ROC para o peso à nascença e a idade gestacional . . . . .	45
4.9	Curvas ROC para os indicadores CRIB e SNAPPE-II e variável peso . . . . .	46





## Siglas e Acrónimos

<b>AUC</b>	Area Under Curve
<b>CRIB</b>	Clinical Risk Index for Babies
<b>IG</b>	Idade Gestacional
<b>FN</b>	Falsos Negativos
<b>FP</b>	Falsos Positivos
<b>FFP</b>	Fração de Falsos Positivos
<b>FVP</b>	Fração de Verdadeiros Positivos
<b>FFN</b>	Fração de Falsos Negativos
<b>FFP</b>	Fração de Falsos Positivos
<b>NUTS II</b>	Nomenclatura das Unidades Territoriais para fins Estatísticos
<b>RNMBP</b>	Registo Nacional de Muito Baixo Peso
<b>ROC</b>	Receiver Operating Characteristics
<b>PN</b>	Peso ao Nascimento
<b>PTG</b>	Pequeno para o Tempo de Gestação
<b>SE</b>	Erro Padrão
<b>VN</b>	Verdadeiros Negativos
<b>VP</b>	Verdadeiros Positivos



# Capítulo 1

## Introdução

### 1.1 Motivação

A análise ROC (Receiver Operating Characteristic) baseia-se na teoria de deteção de sinal e foi originalmente desenvolvida durante a segunda guerra mundial [4] [12] [32]. Um operador de radar tinha de decidir se um ponto no ecrã representava um inimigo, um aliado ou simplesmente ruído [22]. Posteriormente a metodologia ROC foi alargada a outras áreas científicas, na medicina este conceito foi introduzido por Lusted [1968], finais da década de 60 [13] [32]. Neste campo esta metodologia é aplicada em larga escala como, em áreas clínicas que dependam de triagem, testes de diagnóstico, epidemiologia, radiologia e bioinformática [34]. É uma ferramenta comumente usada em decisões médicas devido à sua capacidade discriminativa. [4]

O desempenho de testes de diagnóstico, tem a capacidade de classificar um indivíduo em duas classes (ou categorias), ter presença ou ausência de condições anormais [12] [13] [17] [34]. Este conceito pode ser estendido para avaliação de diagnóstico de modelos de regressão binária [4].

A representação gráfica bidimensional da curva ROC corresponde aos pares  $x$  e  $y$ , que por sua vez correspondem a 1-especificidade e à sensibilidade, respetivamente [4] [21]. Esta técnica serve para visualizar, organizar e seleccionar classificadores com base no seu desempenho. Devido a importantes propriedades estatísticas é a ferramenta mais utilizada na metodologia ROC [12] [13]. Deste modo, surge um dos indicadores mais utilizados na metodologia apresentada, a área abaixo da curva ROC, ou simplesmente AUC (*Area Under Curve*) [12].

Hanley e McNeil definem a AUC de um classificador como a representação da probabilidade de um classificador

ter uma maior propensão para ao acaso, escolher um exemplo positivo, corretamente classificado, do que um exemplo negativo [17]. Assim, valores de AUC próximos de 1 indicam que o modelo de classificação tem um elevado poder discriminativo [13].

Dada a versatilidade da metodologia ROC para analisar e comparar desempenho de testes de diagnóstico, durante a elaboração deste estudo, esta metodologia será aplicada na comparação de dois modelos de previsão do risco de mortalidade neonatal, CRIB (Critical Risk Index for Babies) e SNAPPE-II (Score for Neonatal Acute Physiology-Perinatal Extension II).

O risco de mortalidade em recém-nascidos prematuros que são admitidos numa unidade de cuidados intensivos neonatais varia consoante o tempo de gestação e o peso à nascença. Um recém-nascido é considerado um recém-nascido de muito baixo peso se tiver menos de 32 semanas de gestação ou menos de 1500g de peso à nascença [23] [26]. O índice CRIB foi criado para prever a mortalidade para bebés com um peso à nascença inferior a 1500g, com base na idade gestacional, presença ou ausência de malformações congénitas, o máximo excesso de base (BE) e a fração apropriada do mínimo e o máximo de oxigénio inspirado durante as primeiras 12 horas de vida,  $FiO_2$ , [15] [23] [11]. O índice SNAPPE-II tem a capacidade de prever a mortalidade de um recém-nascido pelo, peso à nascença inferior ou igual a 750g, pequeno para o tempo de gestação, APGAR inferior a 7 aos 5 minutos, valor da diurese, valor mínimo do pH sérico, convulsões múltiplas, valor da razão  $PO_2/FiO_2$ , menor temperatura registada e pressão arterial média, estes registos são contabilizados durante as primeiras 12h depois da admissão [11] [15].

Os índices anteriormente apresentados podem ainda ser indicadores de outros fatores, como, a evolução de práticas médicas e o desempenho das unidades neonatais [20] [23] [30].

Em Portugal, os recém-nascidos de muito baixo peso representam 1% dos nados-vivos, correspondendo a cerca de 1000 por ano [24].

O crescente uso da metodologia ROC em várias áreas da medicina, o poder no apoio da decisão médica e a escassez de estudos com recém-nascidos de muito baixo peso aliados a esta metodologia e aos indicadores, CRIB e SNAPPE-II, no território português, foram as grandes motivações para o desenvolvimento do presente estudo.

## 1.2 Objetivos

Os objetivos do trabalho a desenvolver passam pelo levantamento das medidas de avaliação de desempenho para comparação de diagnósticos com base nas curvas ROC. Para a realização deste estudo será necessário analisar

os indicadores de desempenho associados à curva ROC. Com aplicação de programas específicos de análise para esta metodologia. E deste modo, estudar através da metodologia ROC dos indicadores de gravidade clínica, CRIB e SNAPPE-II, numa população representativa de recém-nascidos de muito baixo peso registados entre 2010 e 2012 em Portugal.

## 1.3 Estrutura da Dissertação

No **Capítulo 1** da dissertação é apresentada uma introdução ao tema com a respetiva motivação, os objetivos assim como a estrutura da mesma.

No **Capítulo 2** apresenta-se o estado da arte, passando por uma revisão bibliográfica sobre a metodologia ROC, a sua utilização e principais índices de desempenho. Será efetuada também uma breve descrição sobre uma aplicação utilizando indicadores de gravidade clínica para recém-nascidos de muito baixo peso, CRIB e SNAPPE-II.

Será também descrita a metodologia que se encontra desenvolvida no **IBM® SPSS® Statistics** e no R, em especial nos pacotes **ROCR** e no **caTools** para indicadores de avaliação de desempenho da curva ROC. Serão também descritos 2 programas utilizados para comparação de duas ou mais curvas ROC; o **ROCNP**A e o **Comp2ROC**.

No **Capítulo 3**, é analisada em pormenor a amostra em estudo, tanto a nível geográfico como estatístico. O conjunto de dados que compõem a amostra em estudo é referente a uma porção aleatória de 25% da população total de recém-nascidos de muito baixo peso que nasceram em Portugal entre 2010 e 2012. Neste capítulo também são analisados e avaliados os indicadores de risco de mortalidade, CRIB e SNAPPE-II, através da utilização da metodologia ROC com auxílio do programa **IBM® SPSS® Statistics**.

No **Capítulo 4** é realizada uma análise e interpretação do desempenho dos índices de mortalidade, CRIB e SNAPPE-II, pelos diferentes programas, **IBM® SPSS® Statistics**, **ROCNP**A e **R** (pacotes: **caTools**, **Comp2ROC**, **ROCR**). São abordadas comparações de duas amostras relacionadas, entre os indicadores de mortalidade e as variáveis Peso ao Nascimento e Idade Gestacional.

No **Capítulo 5** é apresentada a discussão, onde estão apresentados os resultados mais relevantes obtidos durante a realização do presente estudo. Também são realizadas e expostas comparações com resultados obtidos por outros autores em estudos idênticos.

No **Capítulo 6** dar-se-á lugar à conclusão, onde estão expostas as principais conclusões alcançadas com o desenvolvimento do presente trabalho, seguindo-se de propostas para trabalho futuro.

## Capítulo 2

### Metodologia da Curva ROC

Numa experiência de detecção de sinal podem ser observados dois tipos de eventos, um evento de sinal e um evento de ruído mas é essencial discriminar estes dois eventos. Sabe-se que, os eventos de sinal tendem a produzir impressões mais fortes que os eventos de ruído. Porém a intensidade de impressão varia consoante o tipo de experiência. Sendo a intensidade de impressão apenas uma linha de apoio primária para a discriminação de eventos [1]. Considerem-se os eventos de sinal e ruído como eventos binários, caracterizados como sinais sim/não. Desta forma o "sim" corresponde à presença de sinal e o "não" à sua ausência, ruído [1].

A teoria da detecção de sinal levou ao desenvolvimento da metodologia ROC, fundamentada pela teoria de decisão estatística [32]. Na causa do seu desenvolvimento estiveram problemas relacionados com os radares durante a segunda guerra mundial. Em meados da década de 60 a análise ROC foi utilizada em grande escala e com sucesso em várias áreas de estudo, com ênfase na área médica [4] [12] [32]. Na mesma década, Lusted enumera as vantagens da análise ROC no apoio às decisões médicas [22].

Devido à capacidade discriminatória intrínseca à metodologia ROC, esta é ainda usada no apoio a decisões médicas com o intuito de prever e classificar problemas de foro médico. Os testes de diagnóstico, e tendo em conta os diversos tipos, numa vertente médica são a primeira fase de controlo, e são cruciais para permitirem geralmente um tratamento adequado. Porém quando há uma tomada de decisão sobre a conduta do tratamento há sempre uma relação de compromisso, designada por Metz por uma relação de custo/benefício [22]. Pois as decisões tomadas podem estar corretas ou incorretas [17] [22].



## 2.1 Desempenho de Classificadores

A exatidão dos testes diagnósticos é inerente à natureza das variáveis dos classificadores, que podem ser ordinais ou contínuas. As variáveis contínuas podem tomar qualquer valor num intervalo de números finitos ou infinitos, como por exemplo, peso e altura dos pacientes, enquanto que nas variáveis ordinais os valores são medidos por intensidade, como por exemplo, ausência de dor, dor leve, dor moderada ou dor severa. Sendo o resultado de um teste de diagnóstico uma variável binária, com resposta "sim"/ "não" para um determinado estímulo [19] [25].

Para testes binários considerem-se as seguintes analogias, onde a variável  $D$  corresponde a um estado de doença verdadeiro.

$$D = \begin{cases} 1 & \text{Doente} \\ 0 & \text{Não doente} \end{cases}$$

E a variável  $Y$  representa o resultado do teste de diagnóstico. Por convenção, um maior valor de  $Y$  é indicativo de uma maior propensão para a presença de doença.

$$Y = \begin{cases} 1 & \text{Positivo para doença} \\ 0 & \text{Negativo para doença} \end{cases}$$

Com base no resultado de um teste diagnóstico, a equipa médica pode classificar um indivíduo como saudável ou doente. Mas nem sempre o resultado de um teste diagnóstico é conclusivo, sendo possível a equipa médica cometer erros na classificação.

Numa abordagem estatística as possibilidades de classificação variam entre um verdadeiro positivo (VP), um verdadeiro negativo (VN), um falso positivo (FP) ou um falso negativo (FN), como se pode comprovar pela observação da Tabela 2.1.

Tendo em consideração as analogias expostas, um verdadeiro positivo ocorre quando um teste com resultado positivo classifica como doente um indivíduo doente. Um falso negativo ocorre quando um teste dá negativo na presença de doença. Do mesmo modo, um verdadeiro negativo ou um falso positivo ocorrem quando o resultado do teste dá negativo ou positivo, respetivamente, para a ausência de doença [12] [25].

Como referido anteriormente, os resultados dos testes podem induzir a um diagnóstico errado, cometendo-se dois tipos de erros. Em estatística estes erros têm a designação de, erro do tipo I e do tipo II, que correspondem respetivamente à fração de falsos positivos e à fração de falsos negativos.

Tabela 2.1: Classificação dos resultados de testes por estado de doença

	$D = 0$	$D = 1$
$Y = 0$	Verdadeiros Negativos <b>(VN)</b>	Falsos Negativos <b>(FN)</b>
$Y = 1$	Falsos Positivos <b>(FP)</b>	Verdadeiros Positivos <b>(VP)</b>

$$\text{Erro do Tipo I} = \text{Fração de Falsos Positivos (FFP)} = P[\text{classificar } Y = 1 | D = 0]$$

$$\text{Erro do Tipo II} = \text{Fração de Falsos Negativos (FFN)} = P[\text{classificar } Y = 0 | D = 1]$$

A partir da matriz de confusão apresentada na Tabela 2.1 várias métricas podem ser calculadas, destacando-se as seguintes métricas, a Fração de Verdadeiros Positivos e a Fração de Falsos Positivos que representam a taxa de acerto e o falso alarme, respetivamente [12] [21]. E o Total de Negativos que corresponde aos indivíduos não doentes, e o Total de Positivos que corresponde ao total de indivíduos com doença [4].

Assim, podem ser definidas as seguintes entidades:

$$\text{Total de Negativos (N)} = \text{Verdadeiros Negativos (VN)} + \text{Falsos Positivos (FP)}$$

$$\text{Total de Positivos (P)} = \text{Falsos Negativos (FN)} + \text{Verdadeiros Positivos (VP)}$$

$$\text{Fração de Falsos Positivos} = \text{FFP} = P[Y = 1 | D = 0] = \frac{P(Y = 1 \cap D = 0)}{P(D = 0)}$$

$$\text{Sensibilidade} = \text{FVP} = \frac{\text{Número de Verdadeiros Positivos (VP)}}{\text{Total de Positivos (P)}} = P[Y = 1 | D = 1]$$

$$\text{Fração de Verdadeiros Positivos} = \text{FVP} = P[Y = 1|D = 1] = \frac{P(Y = 1 \cap D = 1)}{P(D = 1)}$$

$$\text{FFP} = \frac{\text{Número de Falsos Positivos (FP)}}{\text{Total de Negativos (N)}}$$

Comum à matriz de confusão, os valores que se apresentam ao longo da diagonal principal representam as decisões feitas corretamente [12] [25].

Um teste ideal tem a Fração de Falsos Positivo (FFP) igual a zero e a Fração de Verdadeiros Positivos (FVP) igual a um. De outro modo, quando a FFP=FVP, o teste não é conclusivo sobre a presença de doença. Na área médica a sensibilidade (FVP) e a especificidade (1-FFP) são frequentemente usados para descrever o desempenho da classificação de um teste de diagnóstico [25].

Sendo que, a Sensibilidade, representa a capacidade de um teste para detetar um dado atributo e a Especificidade, representa a capacidade de um teste para excluir indivíduos isentos do atributo [4]. Então pode-se afirmar que ambas representam dois tipos de exatidão, a primeira para casos positivos e a segunda para os casos negativos [21]. Extrapolando as definições consideradas para uma aplicação médica, podem-se fazer as seguintes analogias, se o valor da sensibilidade for elevado, há uma maior probabilidade de indivíduos doentes serem classificados como tal. E por sua vez, se o valor da sensibilidade for baixo, indivíduos saudáveis podem ser classificados como doentes.

$$\text{Sensibilidade } (1 - \beta) = \frac{\text{Número de Verdadeiros Positivos (VP)}}{\text{Total de Positivos (P)}} = \text{FVP}$$

$$\text{Especificidade } (1 - \alpha) = \frac{\text{Número de Verdadeiros Negativos (VN)}}{\text{Total de Negativos (N)}} = 1 - \text{FFP}$$

É possível calcular a Precisão, também denominada Valor Preditivo Positivo (VPP), que corresponde à porção de pacientes verdadeiramente positivos entre os diagnosticados como positivos. Exprime, portanto, a taxa de acerto do classificador [12]. E a Exatidão que representa a taxa de acerto de todo o classificador, isto é, a razão entre a soma dos pacientes verdadeiramente positivos e verdadeiramente negativos e o número total de pacientes diagnosticados como positivos e negativos [12] [21].

$$\text{Precisão} = \frac{\text{Verdadeiros Positivos (VP)}}{\text{Verdadeiros Positivos (VP)} + \text{Falsos Positivos (FP)}}$$

$$\text{Exatidão} = \frac{\text{Verdadeiros Positivos (VP)} + \text{Verdadeiros Negativos (VN)}}{\text{Total de Positivos (P)} + \text{Total de Negativos (N)}}$$

## 2.2 Espaço ROC

### 2.2.1 Modelo Estatístico

A curva ROC pode ser modelada em termos estatísticos com base nas distribuições das variáveis de decisão. Uma das distribuições representa os indivíduos com uma determinada característica, que corresponde à hipótese alternativa, e a outra distribuição representa o grupo sem a característica em questão, que corresponde à hipótese nula [17].

O recurso a testes de hipóteses é muito utilizado para prever o resultado de um teste de diagnóstico, caracterizando-se então, um indivíduo saudável como  $H_0$  e um indivíduo doente como  $H_1$  [4] [25].

O conjunto de valores que podem gerar as curvas ROC, podem ser considerados como variáveis contínuas ou variáveis ordinais, isto é, variáveis que podem tomar qualquer valor num intervalo de números finitos ou infinitos, então pode considerar-se que um teste de diagnóstico produz um resultado binário mas dependendo de um valor de corte [19]. Denomina-se como Valor de Corte, um valor crítico de  $c$  que é representado pela linha vertical na Figura 2.1, separando as decisões positivas das decisões negativas [22].

De acordo com a Figura 2.1, cada decisão é baseada na variável quantitativa, denominada por  $X$ . Os valores que se apresentam à esquerda do valor de  $c$  representam a não ocorrência do evento em questão e os valores

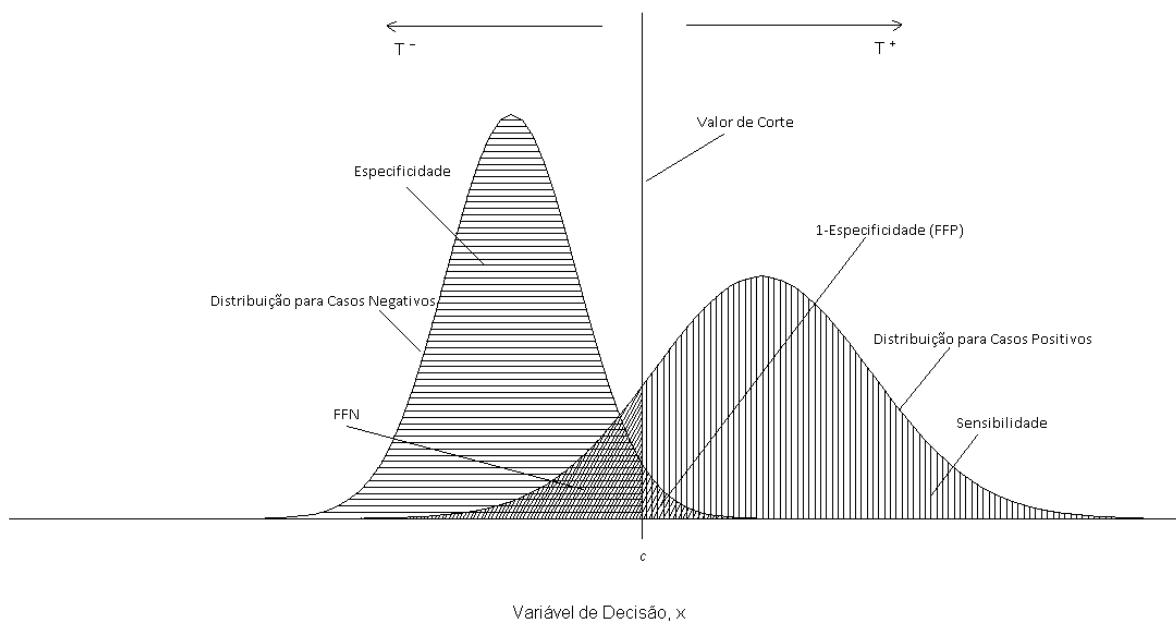


Figura 2.1: Exemplo de um modelo representativo das metodologia ROC, adaptado de [6] [22].

à direita, representam os valores da ocorrência do evento. Então, para valores inferiores ao valor de corte,  $c$ , o indivíduo é considerado saudável e para valores superiores a  $c$ , o indivíduo tem uma patologia. Elevados valores de  $c$ , representam maior probabilidade de ocorrerem casos positivos mas também correspondem a critérios muito restritos e produzem baixos valores tanto de VP como de FP [32].

A fixação do Valor de Corte pode trazer implicações médicas, em termos de custo/benefício, pois o melhor ponto de operação numa dada curva ROC é o ponto de operação que minimiza a média dos custos e por conseguinte, maximiza a média dos benefícios, em uma situação particular de diagnóstico. Deste modo, é possível mover o ponto de operação ao longo da variável de decisão usando um diferente Valor de Corte sendo desta forma possível alterar a média dos valor dos custos [21].

Com referência ao valor de corte  $c$  define-se um teste binário,  $Y$ , a partir de valores contínuos como:

$$\text{Positivo se } Y \geq c \quad (T^+)$$

$$\text{Negativo se } Y < c \quad (T^-)$$

Por conseguinte, a FVP e a FFP são descritas em função do valor de corte,  $c$ , como  $FVP(c)$  e  $FFP(c)$ ,

respetivamente. De salientar a relação de compromisso que existe, em que à medida que o valor de corte,  $c$ , aumenta o valor de  $FFP(c)$  diminuindo  $FVP(c)$ . Também associado aos testes de hipóteses estão os erros, erros de tipo I e II.

$$FVP(c) = P[Y \geq c | D = 1]$$

$$FFP(c) = P[Y \geq c | D = 0]$$

$$P(\text{Erro do Tipo I}) = \alpha = P[\text{rejeitar } H_0 | H_0] = P[T^+ | D = 0] = P[X > c | D = 0] = FFP$$

$$P(\text{Erro do Tipo II}) = \beta = P[\text{não rejeitar } H_0 | H_1] = P[T^- | D = 1] = P[X < c | D = 1] = FFN$$

A curva ROC é definida como um conjunto de valores de FFP e FVP, em função de diferentes valores de corte, num contexto discreto pode ser descrita por [32] :

$$ROC = \{(FFP(c), FVP(c)), c \in (-\infty, +\infty)\}$$

### 2.2.2 Curva ROC

A curva ROC é representada num espaço bidimensional, onde no eixo do  $x$  estão representados os valores da FFP e no eixo do  $y$  estão representados os valores da FVP [12].

O gráfico ROC retrata a relação entre os benefícios (VP) e os custos (FP), deste modo um classificador discreto produz um par (FFP, FVP) no espaço ROC.

Vários pontos são importantes no espaço ROC, entre eles, o ponto inferior esquerdo (0,0) que representa uma estratégia que nunca gera uma classificação positiva, isto é, não classifica como FP nem como VP, porém o ponto superior direito (1,1) só gera VP, e o ponto (0,1) representa uma classificação perfeita.

Classificadores discretos que se apresentam mais perto do ponto (0,0) são denominados conservativos, como o ponto B na Figura 2.2, estes pontos cometem poucos erros do tipo I. Já os classificadores que se encontram mais

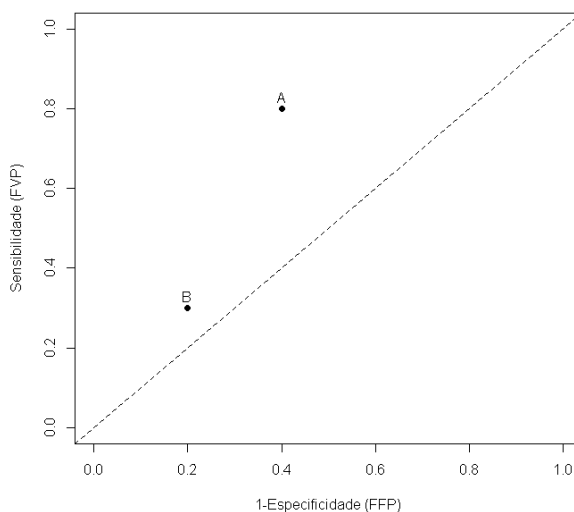


Figura 2.2: Ponto no espaço ROC

perto do ponto (1,1) são denominados por liberais como o ponto A na Figura 2.2, estes pontos cometem mais erros do tipo I [12].

De uma forma geral, um ponto no espaço ROC é tanto melhor quanto mais a noroeste se encontrar, isto é, um maior valor de FVP para um menor valor de FFP, que indica um maior poder discriminativo [12].

Pela observação da Figura 2.2 pode-se afirmar que o ponto B é mais conservativo que o ponto A, porém o ponto A tem um maior poder discriminativo que o ponto B.

Como foi referido anteriormente, a curva ROC representa todos os compromissos entre a sensibilidade e a especificidade. Através da Figura 2.3 é possível visualizar a típica curva ROC empírica [12] [32].

A representação gráfica de curvas ROC é uma técnica que serve para visualizar, organizar e seleccionar classificadores com base no seu desempenho. Um melhor desempenho está intimamente relacionado com uma maior capacidade discriminante.

Uma curva ROC que se aproxima mais do canto superior esquerdo, indicará uma capacidade discriminatória maior, porque há um maior valor de FVP do que o valor de FFP, e por conseguinte uma curva ROC mais afastada deste canto indica uma capacidade discriminatória menor.

Uma propriedade que reflete o compromisso entre a sensibilidade e a especificidade é o facto da curva ROC ser crescente. Neste caso, é possível estabelecer uma relação entre a ter e a não ter uma determinada característica.

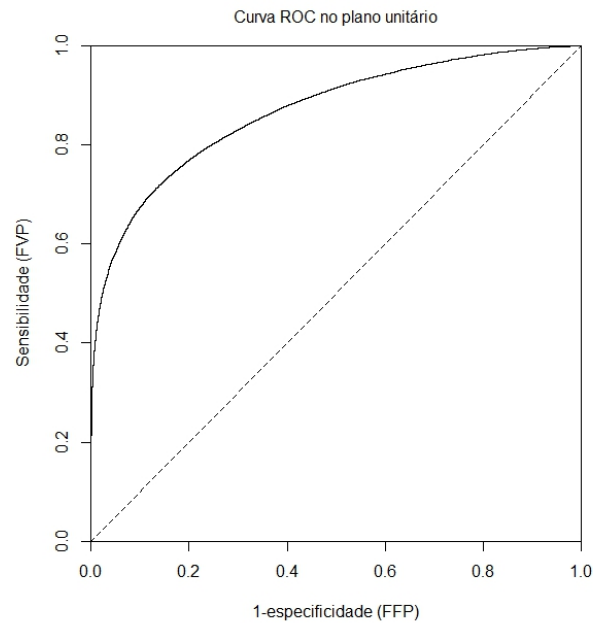


Figura 2.3: Curva ROC

Porém, se a curva gerada no plano bidimensional for diagonal não é possível estabelecer qualquer relação, [21], [13]. Na Figura 2.4 estão ilustradas três curvas que representam vários graus de discriminação, a Curva 1 apresenta um baixo poder discriminante, a Curva 2 apresenta um médio poder discriminante e a Curva 3, um elevado poder discriminante [32].

É possível, então, enumerar algumas propriedades da curva ROC, [13]:

- A curva ROC é uma função monótona crescente, e se para qualquer ponto de corte  $c$ , se tiver  $FFP(c) = FVP(c)$ , o modelo de classificação não é informativo;
- A curva ROC resulta da representação gráfica de todos os pares  $(FFP, FVP)$ ;
- A curva ROC deverá estar definida acima da diagonal principal, e desta forma, os valores da área abaixo da curva variam entre 0,5 e 1.



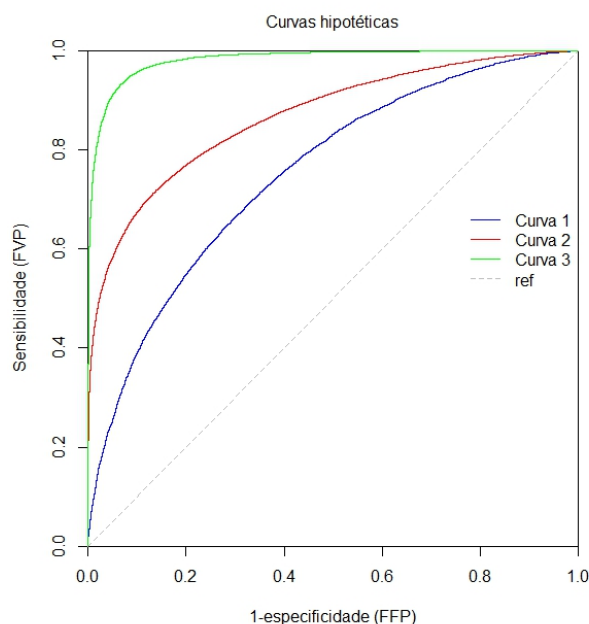


Figura 2.4: Representação de curvas ROC com 3 graus de discriminação diferentes

### 2.2.3 Área abaixo da curva

A área abaixo da curva está associada ao poder discriminante de um teste diagnóstico [17]. Segundo Gagliardi et al. [15] a discriminação num teste diagnóstico é a habilidade dos índices predizerem corretamente dois estados, como por exemplo, vida/morte ou saudável/doente e pode ser calculada com recurso à área abaixo da curva, AUC. Este método tem uma propriedade estatística importante, tal que a AUC de um classificador tem uma maior probabilidade de ao acaso escolher uma instância positiva do que uma instância negativa. De acordo com Green e Swets [1966], citado por Hanley e McNeil [17] a área abaixo da curva ROC corresponde à probabilidade de identificar corretamente dois estímulos, ter ou não ter uma determinada característica.

Segundo Bradley [3] a área a baixo da curva é uma demonstração bidimensional do desempenho de um classificador e fornece informação importante sobre os casos em estudo. E sendo esta uma porção do quadrado unitário, no espaço ROC, os valores variam entre 0,5 e 1. Quando o valor da AUC é igual a 0,5, isto é, a Fração de Verdadeiros e Falsos Positivos é igual ( $FFP = FVP$ ), o classificador não tem poder discriminante, mas à medida que o valor da AUC aumenta, isto é, o valor de  $FVP$  aumenta e o valor de  $FFP$  diminui, aumenta o poder discriminante de um classificador. Até ao valor máximo que este pode atingir com uma  $AUC = 1$ , havendo neste caso uma

discriminação perfeita, este fenómeno observa-se quando a curva está no lado superior esquerdo, tal que a Fração de Verdadeiros Positivos é um [32].

Os autores, Bradley [3], Braga [4] e Fawcett [12] referem os seguintes métodos estatístico mais utilizados para o cálculo da AUC de um classificador:

- Estatística de Wilcoxon-Mann-Whitney;
- Regra do Trapézio;
- Área Binormal.

Para valores contínuos, a área abaixo da curva pode ser calculada pela função densidade de probabilidade [1] [25].

$$AUC = \int_0^1 ROC(t) dt$$

$$AUC = P[Y_D > Y_{\bar{D}}]$$

Porém quando os dados são discretos é possível calcular a AUC pela estatística de Wilcoxon-Mann-Whitney, regra do trapézio ou pela área binormal [1] [3].

Considere-se uma amostra de dimensão  $n_A$  para os indivíduos classificados como anormais,  $A$ , e outra de dimensão  $n_N$  para indivíduos classificados como normais,  $N$ . Este teste estatístico consiste em fazer todas as  $n_A \cdot n_N$  comparações possíveis entre os valores  $x_A$  da amostra  $n_A$  e os valores  $x_N$  da amostra  $n_N$ , graduando cada comparação de acordo com a regra:

$$S(x_A, x_N) = \begin{cases} 1 & \text{se } x_A > x_N \\ \frac{1}{2} & \text{se } x_A = x_N \\ 0 & \text{se } x_A < x_N \end{cases}$$

E fazendo a média de todos os  $S'$ s para todas as  $n_A \cdot n_N$  comparações, vem:

$$W = \frac{1}{n_A \cdot n_N} \sum_1^{n_A} \sum_1^{n_N} S(x_A, x_N)$$

que é uma estatística que não depende dos valores de  $x$ , mas depende das graduações, designada como estatística de Wilcoxon-Mann-Whitney [17].

Como cada comparação é classificada por  $1, \frac{1}{2}$  ou 0, o valor médio de  $W$  estará entre 0 e 1, e reflete qual a proporção de  $x_A$ 's que são maiores que  $x_N$  [17].

Quando se varia o valor de corte, obtém-se um conjunto de pontos pertencentes à curva ROC, e desta forma consegue-se um método simples para calcular a área abaixo da curva ROC: a regra do trapézio [3]. A Figura 2.5 ilustra o processo de cálculo da AUC através da regra do trapézio, onde estão evidenciados os sub-intervalos,  $N$ .

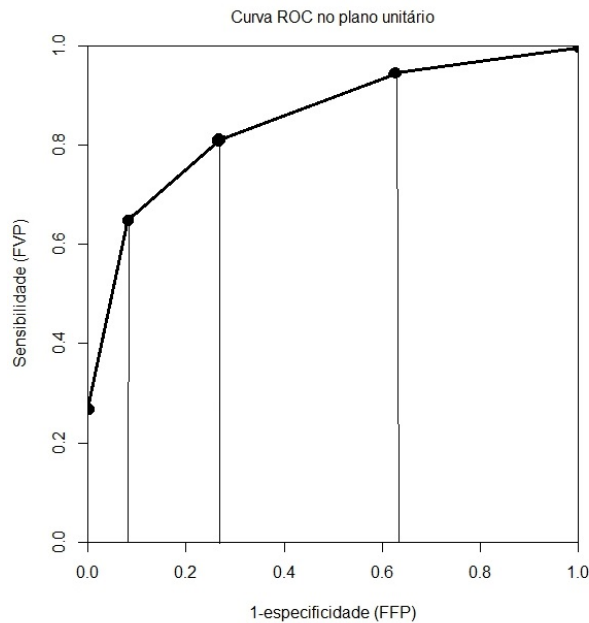


Figura 2.5: Cálculo da AUC pela regra do trapézio

Então:

$$AUC = \sum_{i=1}^N \left\{ (FVP_{i-1} \cdot \Delta FFP) + \frac{1}{2} [\Delta FVP \cdot \Delta FFP] \right\}$$

Onde:

$$\Delta FVP = FVP_i - FVP_{i-1}$$

$$\Delta FFP = FFP_i - FFP_{i-1}$$

O cálculo da área abaixo da curva para o modelo binormal é dado por:

$$AUC = \Phi \left( \frac{a}{\sqrt{1 + b^2}} \right)$$

Com,

$$a = \frac{\mu_1 - \mu_0}{\sigma_1}$$

$$b = \frac{\sigma_0}{\sigma_1}$$

Sendo,  $\mu_1$  e  $\sigma_1$ , a média e a variância para casos anormais e  $\mu_0$  e  $\sigma_0$  para casos normais.

O erro padrão, *Standard Error*, SE, é uma medida de extrema importância para o teste de significância de uma classificação. Convencionalmente existem três formas de calcular a variabilidade associada à AUC [3] [17]:

- A partir dos intervalos de confiança associados à máxima verossimilhança estimada da AUC ( $\hat{\theta}$ );
- A partir do erro padrão da estatística de Wilcoxon, SE(W);
- A partir da aproximação da estatística de Wilcoxon que assume que as distribuições subjacentes positivas e negativas do tipo exponenciais negativas.

O erro padrão,  $SE(W)$  é dado por:

$$SE(W) = \sqrt{\frac{\theta(1 - \theta) + (n_A - 1)(Q_1 - \theta^2) + (n_N - 1)(Q_2 - \theta^2)}{n_A n_N}}$$

Onde  $\theta$  representa o valor da estimativa da AUC, o  $n_N$  e  $n_A$  os número de exemplos negativos e positivos, respetivamente e,

$$Q_1 = \frac{\theta}{(2 - \theta)}$$

$$Q_2 = \frac{2\theta^2}{(1 + \theta)}$$

### 2.2.4 Comparação de Desempenho de Amostras Emparelhadas

Segundo Darlington [1973] citado por [1], as propriedades da AUC fazem desta uma útil propriedade na medida da diferença entre dois ou mais testes para a população em estudo. Pois é possível medir a exatidão, na medida em que cada curva representa a mesma população mas em dois testes diferentes [1].

De forma a comparar classificadores, múltiplos testes são aplicados ao objeto em estudo, sendo que os resultados dos testes são susceptíveis a serem correlacionados [25]. Quando duas ou mais curvas empíricas são construídas com base em testes de desempenho diferentes mas para o mesmo conjunto de dados, é necessário efetuar uma análise estatística entre as curvas construídas, de forma a saber qual dos testes tem um melhor desempenho [4] [18] [10].

Neste estudo, os dois indicadores de risco de mortalidade neonatal são aplicados ao mesmo conjunto de indivíduos. Consideram-se portanto amostras emparelhadas, a que a cada indivíduo corresponde um par de resultados.

É necessário efetuar uma comparação estatística das áreas entre as curvas [18]. Para esse fim, será abordado um método não paramétrico da análise das áreas abaixo da curva ROC empíricas correlacionadas.

#### Área abaixo da curva- Abordagem não paramétrica

A avaliação da diferença entre as duas áreas abaixo da curva ROC derivadas do mesmo conjunto de dados é possível, recorrendo ao cálculo da razão crítica,  $Z$ , definida como:

$$Z = \frac{A_1 - A_2}{\sqrt{SE_1^2 + SE_2^2 - 2rSE_1SE_2}} \quad (2.2.1)$$

Onde  $A_1$  e  $SE_1$  e  $A_2$  e  $SE_2$  são referentes à área e ao erro padrão estimado da curva ROC, referentes aos classificadores 1 e 2, respetivamente. O parâmetro  $r$  representa a correlação estimada entre  $A_1$  e  $A_2$ .

A quantidade  $Z$ , reporta às tabelas da distribuição normal padrão, que para valores de  $Z$  acima de um determinado valor evidencia-se, estatisticamente, que as verdadeiras áreas abaixo da curva ROC são diferentes. A importância da introdução do termo  $2rSE_1SE_2$ , no denominador da expressão é devido ao facto de os dados estarem relacionados, porque foram escolhidos da mesma amostra, e a ausência desse termo implicaria um denominador de maior valor e, conseqüentemente o valor de  $Z$ , mais pequeno o que, eventualmente reduziria a possibilidade de detetar diferenças significativas entre os dois classificadores [4] [18].

### Cálculo do Coeficiente de Correlação

O método sugerido por Hanley e McNeil [18] utiliza uma tabela para determinação do coeficiente de correlação,  $r$ , entre as áreas  $A_1$  e  $A_2$  e através de dois coeficientes de correlação intermédios. Determinam-se os coeficientes de correlação intermédia,  $r_N$  e  $r_A$ , a partir das classificações dadas para pacientes normais e anormais, respetivamente, para as duas modalidades. Cada um destes coeficientes pode ser calculado pelas formas tradicionais utilizando, quer o método de cálculo do produto dos momentos para a correlação de Pearson, quer o método de tau Kendall. Como as curvas ROC em medicina são normalmente obtidas numa escala ordinal, utiliza-se o tau de Kendall para calcular  $r_N$  e  $r_A$ .

O coeficiente de correlação médio,  $\frac{r_N + r_A}{2}$ , e a área média,  $\frac{A_1 + A_2}{2}$ , vão constituir os parâmetros de entrada na Tabela B.1 construída por Hanley e McNeil [18] a partir da qual se obtém o valor do coeficiente de correlação,  $r$ .

Os coeficientes de correlação de áreas podem também ser determinados através do método sugerido por DeLong et al. [10]. Este método utiliza uma aproximação não paramétrica ao cálculo da área abaixo de curvas ROC, para conjuntos de dados correlacionados, utilizando a teoria das estatísticas U-generalizadas para a estimação da matriz de covariâncias quando se comparam duas ou mais curvas ROC [4] [10].

Segundo este método, admitindo que se tem  $m$  indivíduos que apresentam na realidade a doença e  $n$  indivíduos que não têm a doença, a matriz de covariâncias estimada para o vetor de parâmetros (área abaixo da curva ROC)  $\hat{\theta} = (\hat{\theta}^1, \hat{\theta}^2, \dots, \hat{\theta}^k)$ , em que  $k$  representa o número de modalidades a comparar, é tal que:

$$S = \frac{1}{m}S_{10} + \frac{1}{n}S_{01}$$

Seja  $X_i, i = 1, 2, \dots, m$  e  $Y_j, j = 1, 2, \dots, n$  os valores das variáveis nos quais o teste de diagnóstico é baseado, e supondo que valores elevados da variável teste estão associados à presença de doença, as matrizes  $S_{10}$  e  $S_{01}$  com dimensões  $k \times k$  são definidas, respetivamente, para o  $(r, s)$  elemento, pelas expressões das seguintes equações:

$$s_{10}^{r,s} = \frac{1}{m-1} \sum_{i=1}^m [V_{10}^r(X_i) - \hat{\theta}^r][V_{10}^s(X_i) - \hat{\theta}^s]$$

$$s_{01}^{r,s} = \frac{1}{n-1} \sum_{j=1}^n [V_{01}^r(Y_j) - \hat{\theta}^r][V_{01}^s(Y_j) - \hat{\theta}^s]$$

$V_{10}^r$  e  $V_{01}^r$  representam as componentes em  $X$  e  $Y$ , respetivamente, para a  $r$ -ésima estatística  $\hat{\theta}^r$ , definidas por:

$$V_{10}^r = \frac{1}{n} \sum_{j=1}^n \psi(X_i^r, Y_j^r) \quad (i = 1, 2, \dots, m)$$

$$V_{01}^r = \frac{1}{m} \sum_{i=1}^m \psi(X_i^r, Y_j^r) \quad (j = 1, 2, \dots, n)$$

Com  $\psi(X, Y)$  definida através da expressão da equação:

$$\psi = \begin{cases} 1 & Y < X \\ \frac{1}{2} & Y = X \\ 1 & Y > X \end{cases}$$

De salientar que a média desta função  $\psi(X, Y)$ , conduz à estimativa da estatística de Wilcoxon-Mann-Whitney, que corresponde a um estimador da área abaixo da curva ROC ( $\hat{\theta}$ ) [10].

## 2.3 Índices Clínicos de Mortalidade Neonatal:

### CRIB e SNAPPE-II

O desenvolvimento de medidas de risco de mortalidade neonatal é essencial para agrupar os recém-nascidos por risco de mortalidade. Servindo como parâmetro para comparar as unidades de cuidados intensivos neonatais, avaliando as práticas médicas e a qualidade dos serviços intensivos de neonatologia [15] [20] [23]. Entre os vários índices que foram desenvolvidos com o intuito de prever o risco de mortalidade neonatal, neste estudo serão abordados o CRIB (Clinical Risk Index for Babies) e o SNAPPE-II (Score for Neonatal Acute Physiology-Perinatal Extension, Version II) [20] [27]. Estes são os índices mais utilizados mundialmente nas unidades de cuidados intensivos neonatais [23].

A sobrevivência de recém-nascidos de muito baixo peso depende do peso à nascença e do tempo de gestação, para tal, os bebés sujeitos a esta análise têm de nascer com menos 32 semanas de gestação e/ou menos 1500g de peso à nascença [16] [23] [26]. Mas também outros fatores perinatais e condições fisiológicas podem afetar individualmente os recém-nascidos, em particular complicações que possam surgir nas primeiras 12 horas de vida [15]. Em ambos os índices, CRIB e SNAPPE-II, a equipa médica tem um papel preponderante, pois muitos dos dados necessários são recolhidos durante as primeiras 12 horas depois da admissão [23] [26] [27].

#### 2.3.1 CRIB: Clinical Risk Index for Babies

O índice CRIB foi criado para prever a mortalidade para recém-nascidos com menos de 32 semanas de gestação. Este índice foi desenvolvido no Reino Unido e com dados pertencentes a recém-nascidos admitidos nas unidades de neonatologia entre 1988 e 1990. A amostra era constituída por 812 recém-nascidos de muito baixo peso, dos quais 25% acabariam por falecer [11].

Os autores com recurso à regressão logística conseguiram identificar as seis variáveis que melhor prediziam a mortalidade. Sendo elas, o peso à nascença, tempo de gestação, malformações congénitas, excesso de base máximo nas primeiras 12 horas e os valores mínimos e máximo de  $FiO_2$  (fração inspirada de oxigénio) nas primeiras 12 horas após a admissão [11] [30].

Cada variável do índice CRIB tem um valor numérico pré-determinado que varia conforme a sua gravidade como mostra a Tabela 2.2, a pontuação final deste índice é a soma dos pesos destas seis variáveis, que pode variar entre 0 e 23 pontos.



Tabela 2.2: Pontuações do índice CRIB, adaptado de [30]

Variável		Pontuação
Peso ao Nascimento (g)	> 1350	0
	851 a 1350	1
	701 a 850	4
	≤ 700	7
Idade Gestacional (semanas)	> 24	0
	≤ 24	1
Malformação Congénita	Ausente	0
	Sem risco de vida	1
	Com risco de vida	3
Excesso de Base máximo (mmol/l)	> -7,0	0
	-7,0 a -9,9	1
	-10,0 a -14,9	2
	≤ -15,0	3
Mínimo de FiO <sub>2</sub>	≤ 0,40	0
	0,41 a 0,60	2
	0,61 a 0,90	3
	0,91 a 1,00	4
Máximo de FiO <sub>2</sub>	≤ 0,40	0
	0,41 a 0,80	1
	0,81 a 0,90	3
	0,91 a 1,00	5

Vários autores, entre eles, Sarquis et al [30] sugerem agrupar os indivíduos pelo resultado final do índice, padronizando assim os indivíduos por gravidade clínica. Criando deste modo, 4 graus de classificação, Grau 1 para valores entre 0 a 5, Grau 2 de 6 a 10, Grau 3 de 11 a 15, e Grau 4 para pontuações superiores a 15 [30].

Uma das grandes vantagens deste índice é a facilidade de recolha dos dados [11]. Segundo Dorling et al. [11],

no estudo original, os autores mostraram que este índice era um bom discriminador, contestando com um valor de 0,90 para a área abaixo da curva.

### **2.3.2 SNAPPE-II: Score for Neonatal Acute Physiology-Perinatal Extension Version**

#### **II**

Muitos outros índices foram desenvolvidos, e alguns desenvolvidos a partir de adaptações e alterações a índices pré-existent sempre com a finalidade de melhorar a capacidade discriminativa do mesmo. Um desses casos é o índice SNAPPE-II [11]. Devido à dificuldade de recolha de dados para os índices, SNAP (Score for Neonatal Acute Physiology) e SNAP-PE (Score for Neonatal Acute Physiology-Perinatal Extension), os autores sentiram necessidade de desenvolver uma versão mais simples e prática. Então, com dados de recém-nascidos de 30 unidades de neonatologia Norte Americanas, que abrangiam 10 018 recém-nascidos conseguiram modelar novos índices.

As mudanças incluíram a redução do período de recolha de dados para 12 horas após a admissão e a redução do número de variáveis para seis, as que demonstraram ter uma forte associação estatística com a mortalidade. Sendo estas, a média da pressão arterial, menor temperatura, pior fração de pressão parcial de oxigénio ( $PO_2$ ) pela fração inspirada de oxigénio ( $PO_2/FiO_2$ ), valor mínimo do pH sérico, convulsões múltiplas e débito urinário, isto é, se a produção de urina é mais baixa que os valores de referência padronizados [11] [15].

Estas seis variáveis expostas no parágrafo anterior deram origem ao índice SNAP-II mas este carenciava de outros fatores perinatais, que poderiam ser úteis na previsão da mortalidade. Então os autores sentiram a necessidade de criar o SNAPPE-II, uma extensão do SNAP-II com a inclusão de mais 3 variáveis, o valor do APGAR aos 5 minutos, peso à nascença e o parâmetro, pequeno para o tempo de gestação (PTG) [11] [15].

A escala APGAR (Tabela A.1) é uma avaliação usada em todos os recém-nascidos, e avalia as condições do recém-nascido nos primeiros minutos de vida e determina a possibilidade de uma intervenção médica. Tal como em todas as escalas, para cada parâmetro está associada uma pontuação que varia conforme a gravidade, os sinais avaliados são, a frequência cardíaca, tipo de respiração, tónus muscular, irritabilidade reflexa e cor da pele [31]. O parâmetro pequeno para o tempo de gestação (PTG) é uma avaliação feita com suporte numa tabela de referência (Tabela A.2), em que para cada idade gestacional em semanas, está associada um valor mínimo do peso, em gramas, que corresponde ao 3º percentil [27].

A cada variável que compõe o índice SNAPPE-II está associada uma pontuação como mostra a Tabela 2.3 que

Tabela 2.3: Pontuações do índice SNAPPE-II, adaptado de [14]

Variável		Pontuação
Média da pressão arterial (mm Hg)	> 29	0
	20 a 29	9
	< 20	19
Temperatura Corporal (°C)	> 35,6	0
	35 a 35,6	8
	< 35	15
Peso à nascença (g)	> 999	0
	750 a 999	10
	< 750	17
PTG < 3º percentil	não	0
	sim	12
PO <sub>2</sub> /FIO <sub>2</sub>	> 2,49	0
	1 a 2,49	5
	0,3 a 0,99	16
	< 0,3	28
Valor Mínimo de pH sérico	> 7,19	0
	7,10 a 7,19	7
	< 7,10	16
Débito Urinário (ml/kg/h)	> 0,9	0
	0,1 a 0,9	5
	< 0,1	18
Convulsões Múltiplas	não	0
	sim	19
APGAR aos 5 minutos	> 7	0
	< 7	18

varia consoante a gravidade clínica da mesma, após a contabilização da totalidade das variáveis é obtida a pontuação final do índice em questão, que pode variar entre 0 e 162 pontos.

Vários autores, entre eles Fortes Filho et al [14] propõem no seu estudo agrupar os indivíduos pelo valor do resultado do índice, criando desta forma 5 graus de classificação. Grau 1 para pontuações entre 0 e 20, Grau 2 entre 21 a 40, Grau 3 entre 41 e 60, Grau 4 entre 61 a 80 e o Grau 5 para pontuações superiores a 81. Estas classificações expostas são adaptadas das sugeridas por Fortes Filho et al [14]. Uma das grandes vantagens deste índice, tal como o CRIB, é a facilidade de recolha dos dados [11].

## 2.4 Programas Estatísticos para Análise ROC

### 2.4.1 IBM® SPSS® Statistics

O IBM® SPSS® Statistics versão 22.0 foi a versão utilizada na elaboração deste trabalho. Mais conhecido por SPSS, que assim será designado ao longo deste estudo, é um programa estatístico comumente utilizado devido à sua robustez na análise estatística. Através deste, é possível obter a representação bidimensional da curva ROC no plano unitário, assim como os valores da AUC, sensibilidade, especificidade e respetivos intervalos de confiança. É possível calcular os coeficientes de correlação das áreas com recurso à metodologia tau de Kendall, apresentada por Hanley e McNeil [18]. Também calcula o intervalo de confiança e o teste de comparação com o valor da  $AUC=0,5$ .

### 2.4.2 R

O programa estatístico R, neste estudo utilizada a versão 3.1, disponibiliza vários pacotes para o cálculo da curva ROC. Entre eles, **caTools**, **ROCR** e **Comp2ROC**.

#### Pacote caTools

O pacote **caTools** contém enumeras funções, entre elas a função *co/AUC*, que permite o cálculo da AUC, pelo método não paramétrico, estatística de Wilcoxon-Mann-Whitney e a visualização das curvas ROC.

Apesar de não ser um pacote robusto e direcionado exclusivamente à metodologia ROC, a função *co/AUC* tem aspetos positivos, sendo de fácil implementação e de rápida execução. Apresentando porém, algumas limitações

relativas a informações estatísticas relevantes para a análise desta metodologia, como a impossibilidade do cálculo do erro padrão, do intervalo de confiança e dos coeficientes de correlação.

### **Pacote ROCR**

O pacote **ROCR** é uma ferramenta útil para a criação bidimensional da curva ROC. Vários métodos estatísticos e várias medidas de desempenho estão disponíveis neste pacote e serão usadas ao longo do estudo prático, entre elas, o índice da área abaixo da curva ROC que é determinado pelo método não paramétrico, estatística de Wilcoxon-Mann-Whitney. Porém este programa apresenta algumas restrições, pois não calcula o erro padrão, e como não compara as áreas de diferentes classificadores, não calcula o coeficiente de correlação.

### **Pacote Comp2ROC**

No estudo do desempenho de dois classificadores, recorrendo ao método da área abaixo da curva ROC, pode ocorrer que as curvas pertencentes a cada classificador se intersectem, originando alternância no poder discriminante.

As curvas podem intersectar-se em um ou mais pontos, sendo necessário abordar este caso de uma forma diferente daquela que é abordada quando as curvas ROC dos classificadores não se cruzam. Para tal efeito, recorre-se às ferramentas existentes no pacote **Comp2ROC**, elaborado por Braga et al[5].

O índice área abaixo da curva ROC é determinado por dois processos diferentes, a regra do trapézio e a aproximação não paramétrica à estatística de Wilcoxon-Mann-Whitney. Os valores dos erros padrão e o valor do coeficiente de correlação entre áreas,  $r$ , são estimados pelo sugerido por DeLong et al[10].

### **2.4.3 ROCNPA**

O programa **ROCNPA** (Receiver Operating Characteristic Non Parametric Analysis) desenvolvido em linguagem Java, foi concebido com o intuito de suprimir algumas lacunas encontradas nos programas até então existentes de análise da metodologia ROC [4]. Partindo desta motivação Braga [4] criou um programa que se tornasse versátil para a análise ROC, que se encontra disponível através do link: <http://pessoais.dps.uminho.pt/acb/englab/feedback.htm>. Tendo como objetivos principais:

- Determinar um ajuste para a curva ROC;

- Avaliar o desempenho do teste do teste diagnóstico através de um índice de determinação simples e livre de hipóteses distribucionais;
- Comparar mais que três diagnósticos, quer os dados sejam provenientes de amostras independentes, quer de amostras correlacionadas.

O índice área abaixo da curva ROC é determinado por três processos diferentes, a regra do trapézio, a aproximação não paramétrica à estatística de Wilcoxon-Mann-Whitney e a aproximação ao plano binormal através dos coeficientes estimados da reta de regressão. Os valores dos erros padrão são estimados pelo sugerido por Hanley e McNeil [18] [4].



## Capítulo 3

# Aplicação da Metodologia ROC ao Caso em Estudo

Numa primeira fase será caracterizada a amostra em estudo, tanto a nível geográfico como estatístico. Seguindo-se um estudo pormenorizado sobre os indicadores de mortalidade e algumas das variáveis que os compõem, sendo elas, o peso ao nascimento e a idade gestacional. Estas abordagens foram realizadas com o auxílio do programa **SPSS Statistics** versão 22.0.

### 3.1 Caracterização da Amostra em Estudo

O Registo Nacional do Recém-Nascido de Muito Baixo Peso (RNMBP) funciona em Portugal desde 1994, e tem como objetivo primordial, o registo da população sobrevivente de recém-nascidos de muito baixo peso para avaliação futura do neurodesenvolvimento e das sequelas. Integram neste registo todos os recém-nascidos de muito baixo peso, isto é com menos de 1500g e recém-nascidos com idade gestacional inferior a 32 semanas, independentemente do peso ao nascimento [9] [24].

Os dados que compõem o presente estudo foram facultados e recolhidos pelo RNMBP, referentes a um período entre 2010 e 2012, contemplando todo o território português.

De forma a obter uma amostra representativa para o desenvolvimento do presente estudo, selecionaram-se aleatoriamente 25% dos dados totais, obtendo-se uma amostra composta por 789 recém-nascidos.



Estes são provenientes de várias unidades hospitalares Portuguesas, a Figura 3.1 apresenta a distribuição dos recém-nascidos de muito baixo peso a nível regional, NUTS II. As regiões que integram a NUTS II são 7, das quais 5 no continente (Norte, Centro, Lisboa e Vale do Tejo, Alentejo e Algarve) e os territórios das regiões autónomas dos Açores e Madeira.

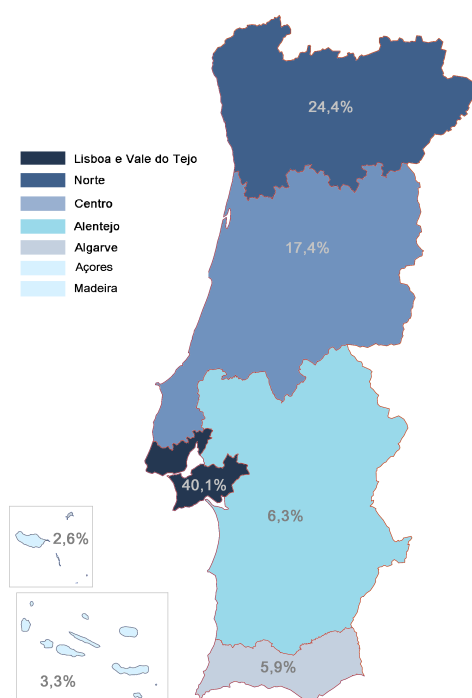


Figura 3.1: Distribuição dos recém-nascidos de muito baixo peso por NUTS II

Num enquadramento geográfico mais detalhado, é possível através da Figura 3.2 obter-se uma perceção da distribuição do número de recém-nascidos de muito baixo peso em função dos distritos onde estava sediada a unidade hospitalar. Sendo os distritos portugueses com maior densidade populacional os que apresentam a maior percentagem de neonatos.

Os recém-nascidos que integram este estudo contemplam algumas das características expostas na Tabela 3.1, com uma média e um desvio-padrão do peso ao nascimento de  $1214,3g \pm 343,1$  e uma média e um desvio-padrão de idade gestacional de  $29,8 \pm 2,5$  semanas. Os valores mínimo e máximo recaem para, as 23 e 38 semanas de idade gestacional e para 400 e 2170 gramas de peso à nascença. Do total da amostra, com 789 recém-nascidos, 405 são do género masculino que representam 51,3 % da amostra em estudo, e 384 são do género feminino. Tendo sido declarados 89 óbitos hospitalares, que representam 11,3% da amostra em estudo.

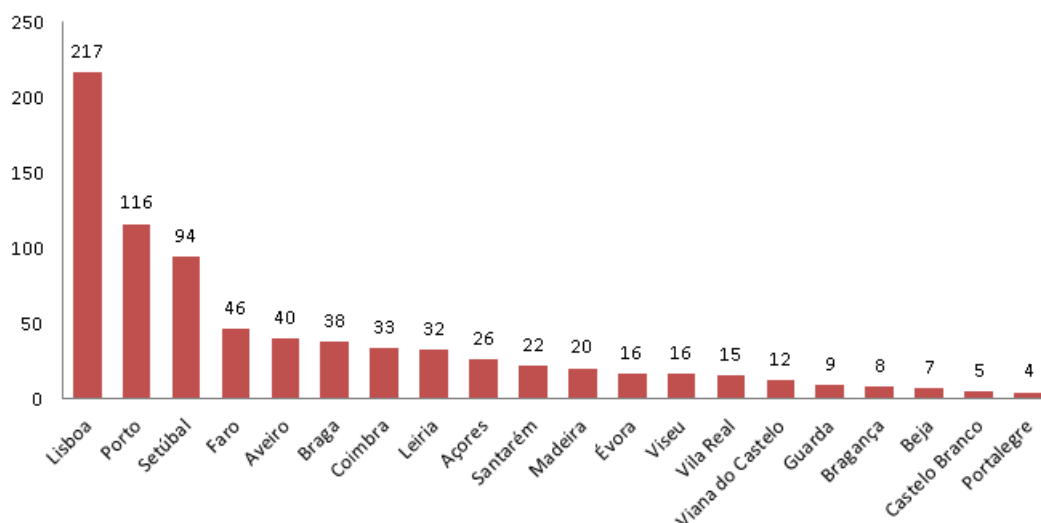


Figura 3.2: Distribuição do número de neonatos por distritos

Tabela 3.1: Características da amostra em estudo, média e desvio-padrão, respetivamente

Peso à nascença (g)	1214,3 (343,1)
Idade gestacional (semanas)	29,8 (2,5)

Na Tabela 3.2 é exposta uma análise estatística dos recém-nascidos sobreviventes e falecidos em função da idade (IG) e do peso ao nascimento (PN). A incidência de mortalidade do género masculino é de 11,6 %, sendo ligeiramente superior à do género feminino, com 10,9%, também verificado pelo estudo levado a cabo por Marshall et al. [20]. Para os recém-nascidos sobreviventes e falecidos a idade média das progenitoras é de 30,65 e 30,84 anos, respetivamente.

Em geral, os sobreviventes apresentam um maior peso à nascença e maior idade gestacional que os que não sobrevivem. É facto que uma idade gestacional tardia está intimamente relacionada com a maturação dos órgãos vitais.

Sendo o peso uma variável de grande expressão nos indicadores de mortalidade, a Figura 3.3 representa a distribuição do peso dos 789 recém-nascidos de muito baixo peso, e na Figura 3.4 está representada a variação dos pesos dos recém-nascidos no ato da admissão por estado, sobrevivente e falecido.

Na Tabela 3.3 estão expostas algumas estatísticas sumárias da variável peso ao nascimento. Para o caso em estudo, a mortalidade para recém-nascidos com mais de 500g é inferior a 50%, e para valores de peso superiores a

Tabela 3.2: Características dos indivíduos sobreviventes e falecidos por PN e IG

	Peso ao nascimento (g)		Idade gestacional (semanas)	
	Óbito	Sobrevivente	Óbito	Sobrevivente
Média	899,54	1254,36	27,29	30,13
Mínimo	410	400	23	24
Máximo	2130	2170	38	36
Desvio-Padrão	337,81	322,62	3,3	2,28

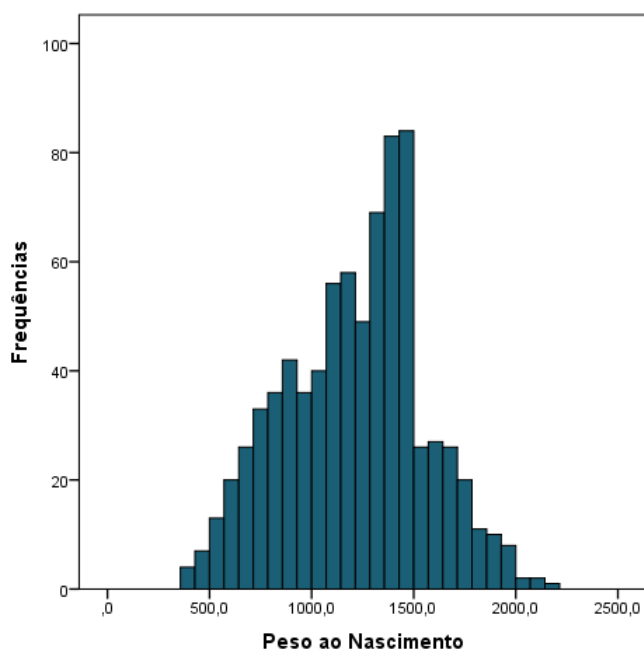


Figura 3.3: Histograma representativo da distribuição da variável PN

1000g a probabilidade de mortalidade é inferior a 10%.

Pela análise da Tabela 3.4 constata-se que a partir das 26 semanas a probabilidade de mortalidade é inferior a 50% porém com apenas 23 semanas a taxa de sobrevivência é nula. Os estudos desenvolvidos por Baumer et al. [2] e por Rocha et al. [29] apresentam resultados idênticos, com a mortalidade de recém-nascidos a partir das 26 semanas inferior a 50%.

Quanto maior é a idade gestacional maior é a probabilidade de sobrevivência dos recém-nascidos, que a cada

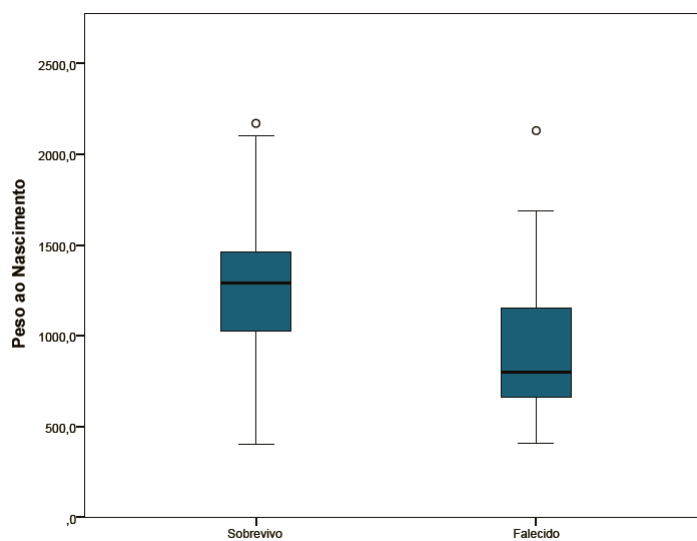


Figura 3.4: Peso ao nascimento por sobrevivido e falecido

Tabela 3.3: Estatísticas sumárias do Peso ao Nascimento

Peso ao nascimento (g)	
Média	1214,3
Desvio-Padrão	343,1
Mínimo	400
Máximo	2170
Mediana	1250
Percentil 10	743
Percentil 25	960
Percentil 75	1450
Percentil 90	1649

semana de gestação adicional em que um recém-nascido é admitido há uma redução de 10,96% do risco de mortalidade. Valor muito semelhante ao apresentado por Marshall et al. [20], que referindo uma taxa de 12%.

O índice CRIB tem uma média de pontuação de 2,47, com mediana de 1, e com a pontuação final a variar entre 0

Tabela 3.4: Distribuição da mortalidade por IG

IG (semanas)	23	24	25	26	27	28	29	30	31	32
Total	3	21	31	39	59	72	107	100	129	142
Sobreviventes	0	9	11	29	47	66	100	95	124	141
Falecidos	3	12	20	10	12	6	7	5	5	1
Mortalidade (%)	100	57,1	64,5	25,6	20,3	8,3	6,5	5	3,9	1,4

e 18. Por sua vez, o SNAPPE-II tem uma média de pontuação de 18,91, com mediana de 12, e variação de pontuação final entre 0 e 115. Na Tabela 3.5 estão expostas outras estatísticas complementares dos indicadores de mortalidade.

Tabela 3.5: Estatísticas da pontuação dos índices em estudo por estado, óbito e sobrevivente

	CRIB		SNAPPE-II	
	Óbito	Sobrevivente	Óbito	Sobrevivente
Média	7,81	1,80	52,61	15,02
Mediana	8	1	54	9
Mínimo	0	0	0	0
Máximo	18	13	116	101
Desvio-Padrão	4,86	2,44	28,71	17,54

Na Tabela 3.6, encontram-se os valores da AUC, Erro Padrão e o Intervalo de Confiança a 95% para os índices CRIB e SNAPPE-II, peso ao nascimento e idade gestacional, que foram calculados pelo **SPSS**. Observou-se que o maior valor da área abaixo da curva pertence ao índice CRIB, seguindo em ordem decrescente, pelo índice SNAPPE-II, idade gestacional e peso ao nascimento. Os indicadores de mortalidade, CRIB e SNAPPE-II, predizem melhor a mortalidade que qualquer uma das duas variáveis que os compõem, idade gestacional e peso. Sendo o índice CRIB o que demonstra a melhor capacidade de desempenho.

Os autores, Baumer et al. [2], Braga et al. [6], Fortes Filho et al. [14], Richardson et al. [28] e Sarquis et al. [30], nos estudos por eles elaborados propõem agrupar os indicadores de mortalidade, CRIB e SNAPPE-II, por graus

Tabela 3.6: Valores da AUC, Erro Padrão e Intervalos de Confiança (95%) para o CRIB, SNAPPE-II, IG e PN

Variável	N	AUC	Erro-Padrão	Intervalos de Confiança a 95%
Índice CRIB	788	0,876	0,025	0,828 - 0,925
Índice SNAPPE-II	600	0,867	0,026	0,816 - 0,918
Idade Gestacional		0,785	0,032	0,723 - 0,848
Peso ao Nascimento		0,782	0,028	0,727 - 0,836

de mortalidade clínica. Tendo como argumento principal a facilidade na previsão da mortalidade, desta forma, com o valor da pontuação final do índice conseguia-se saber em que grupo estaria inserido e consequentemente a respetiva probabilidade de mortalidade.

Partindo desta motivação agrupam-se os indivíduos pertencentes à amostra em estudo em 4 grupos, consoante a pontuação do índice CRIB. Desta forma os recém-nascidos são classificados em 4 Graus: Grau 1 para valores entre 0 a 5, Grau 2 de 6 a 10, Grau 3 de 11 a 15, e Grau 4 para pontuações superiores a 15.

Dos 789 recém-nascidos que constituem a amostra em estudo só 788 preencheram os critérios para esta análise estatística. Deste modo, os recém-nascidos foram divididos em 4 faixas crescentes de gravidade, obtendo-se a mortalidade correspondente a cada grau de gravidade como mostra a Tabela 3.7.

Tabela 3.7: Distribuição dos recém-nascidos quanto aos Graus do CRIB e a respetiva mortalidade

Classificação	N	%	Mortalidade	Mortalidade (%)
Grau 1	658	83,50	35	5,32%
Grau 2	100	12,69	29	29 %
Grau 3	23	2,92	18	78,26%
Grau 4	7	0,89	7	100%

Os resultados obtidos nos Graus do índice CRIB são análogos aos apresentados por Baumer et al. [2] e Sarquis et al. [30]. É notório que à medida que os Graus aumentam também aumenta o número de óbitos hospitalares.

Com recurso às funcionalidades da metodologia ROC presentes no **SPSS**, procedeu-se à avaliação das duas

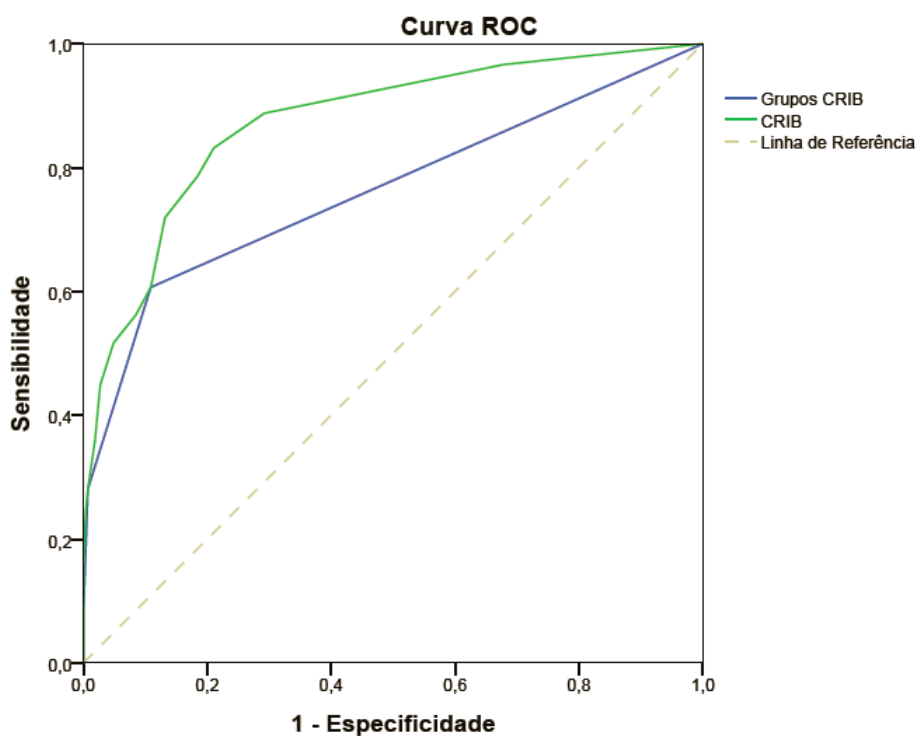


Figura 3.5: Curvas ROC empírica, do CRIB e dos Grupos do CRIB

curvas ROC, uma referente ao índice CRIB, e outra pertencente aos Grupos do índice CRIB, Figura 3.5.

Os valores obtidos para as AUC's, foram de 0,876 para o CRIB e 0,762 para os Grupos do CRIB. Este último valor revela que adotar esta estratégia não traz benefícios numa previsão correta da mortalidade. Por vezes o facto de um recém-nascido ter um valor do índice CRIB elevado não significa que este não irá sobreviver, o mesmo se aplica inversamente, um valor mais baixo não está intimamente relacionado com a sobrevivência.

A mesma abordagem é aplicada no índice SNAPPE-II, que do total dos recém-nascidos que integram o estudo somente 600 são elegíveis para esta análise. A cada variável que o compõe está associada uma pontuação que varia consoante a gravidade clínica da mesma, após a contabilização da totalidade das variáveis é obtida a pontuação final do índice em questão. Desta forma os recém-nascidos são classificados em 5 graus, Grau 1 para pontuações entre 0 e 20, Grau 2 entre 21 a 40, Grau 3 entre 41 e 60, Grau 4 entre 61 a 80 e o Grau 5 para pontuações superiores a 81, estas classificações expostas são adaptadas das sugeridas por [14].

A taxa de mortalidade por Graus no índice SNAPPE-II é muito elevada nos graus 4 e 5, que atingem mais de 60% e 70% em cada grupo respetivamente.

Tabela 3.8: Distribuição dos recém-nascidos pelos cinco graus do SNAPPE-II e a respetiva mortalidade

Classificação	N	%	Mortalidade	Mortalidade (%)
Grau 1	396	66	9	2,27 %
Grau 2	111	18,5	12	10,81 %
Grau 3	55	9,17	16	29,01 %
Grau 4	24	4	15	62,5%
Grau 5	14	2,33	10	71,43 %

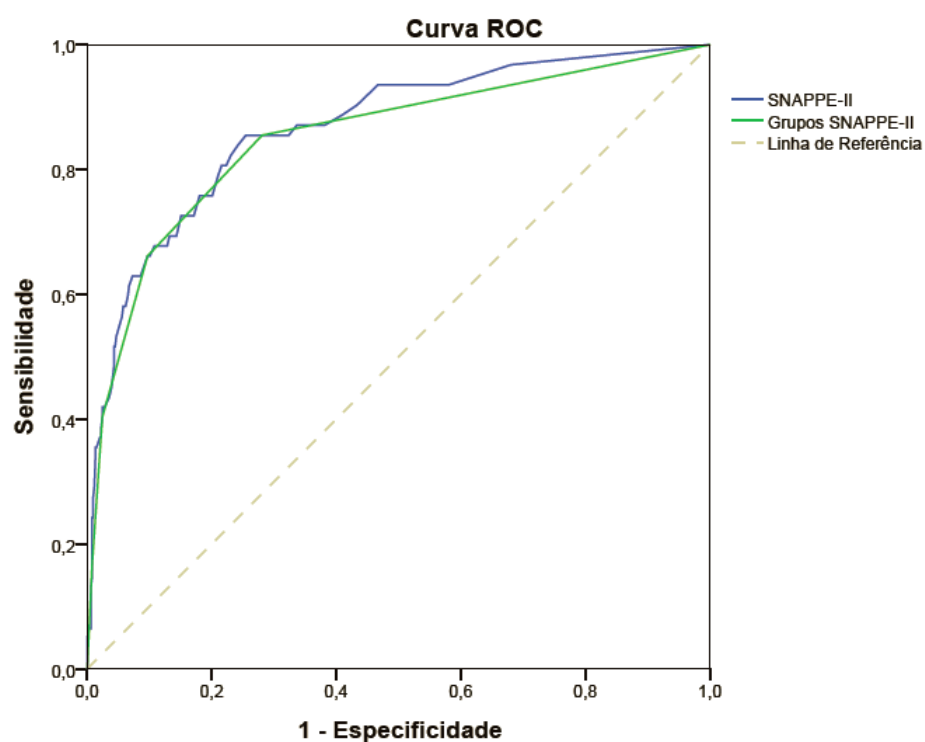


Figura 3.6: Curvas ROC, do SNAPPE-II e dos Grupos do SNAPPE-II

De forma a avaliar o desempenho do SNAPPE-II e dos Grupos do SNAPPE-II na previsão da mortalidade, recorreu-se à funcionalidade da metodologia ROC contida no programa **SPSS**. Visualmente apresentado na Figura 3.6, e com os valores de AUC, de 0,867 para o SNAPPE-II e para os Grupos do SNAPPE-II de 0,851.



Apesar das diferenças das áreas não ser tão grande como a diferença apresentada para o CRIB e os Grupos do CRIB, Figura 3.7, o índice SNAPPE-II prediz melhor a mortalidade que os Grupos do SNAPPE-II.

Pela análise das Figuras 3.7 e 3.8 e pelos valores da AUC obtidos é evidente que o desempenho dos indicadores de mortalidade agrupados é inferior aos alcançados pelos indicadores de mortalidade na sua forma original, sem serem agrupados. Motivo pelo qual, nas avaliações e comparações futuras serão utilizados os índices na sua forma não agrupada.

## Capítulo 4

# Comparação de Duas Amostras Emparelhadas Utilizando Curvas ROC

Neste capítulo, serão abordadas comparações de duas amostras relacionadas, entre os indicadores de mortalidade e entre as variáveis de maior expressão que os compõem, utilizando a metodologia ROC e as suas propriedades estatísticas. As análises serão realizadas com recurso aos pacotes **caTools**, **ROCR** e **Comp2ROC** pertencentes ao **R** e aos programas **SPSS** e **ROCNP**A.

Na amostra apresentada e caracterizada no capítulo anterior, nem todos os recém-nascidos integrantes contêm simultaneamente informação referente ao índice CRIB e SNAPPE-II, sendo uma condição estritamente necessária para trabalhar no pacote **Comp2ROC** e no programa **ROCNP**A. Por este motivo, selecionaram-se somente os indivíduos que preenchiam estes critérios, ter as variáveis CRIB e SNAPPE-II, que conseqüentemente reduziu a amostra para 599 recém-nascidos.

Desta forma, neste capítulo, efetuar-se-á um estudo pormenorizado de avaliação discriminativa com o intuito de estimar se os indicadores de mortalidade e as variáveis, peso ao nascimento e idade gestacional, têm uma boa capacidade em prever entre dois estados, sobrevivência ou não dos prematuros. Para tal comparar-se-ão curvas ROC e os respetivos valores de AUC.

## 4.1 Comparação dos indicadores de mortalidade: CRIB e SNAPPE-II

Ao longo desta secção serão apresentadas várias comparações dos indicadores de mortalidade CRIB e SNAPPE-II, com recurso ao **SPSS**, **ROCR**, **Comp2ROC**, **caTools** e **ROCNP**A, utilizando para o efeito um conjunto de dados em que os dois indicadores estão definidos, correspondendo a um total de 599. Com isto, pretende-se perceber qual destes pacotes e programas, é o mais versátil na análise da metodologia ROC.

### 4.1.1 Comparação dos indicadores de mortalidade através do SPSS

De forma a avaliar o desempenho dos dois índices, aplicou-se a metodologia ROC com auxílio do programa **SPSS**. Na Figura 4.1 é possível observar a comparação entre as duas curvas ROC, a curva a azul representa o desempenho do índice CRIB e por sua vez a curva verde representa o desempenho do índice SNAPPE-II. Obtiveram-se as áreas abaixo da curva ROC, referentes a cada índice, de forma a qualificar o desempenho destes. Apresentando para o CRIB um valor de AUC de 0,876, um desvio padrão de 0,025 e um intervalo de confiança de 0,828 a 0,925. E por sua vez, para o índice SNAPPE-II um valor de AUC de 0,867, um desvio padrão de 0,026 e um intervalo de confiança de 0,816 a 0,918.

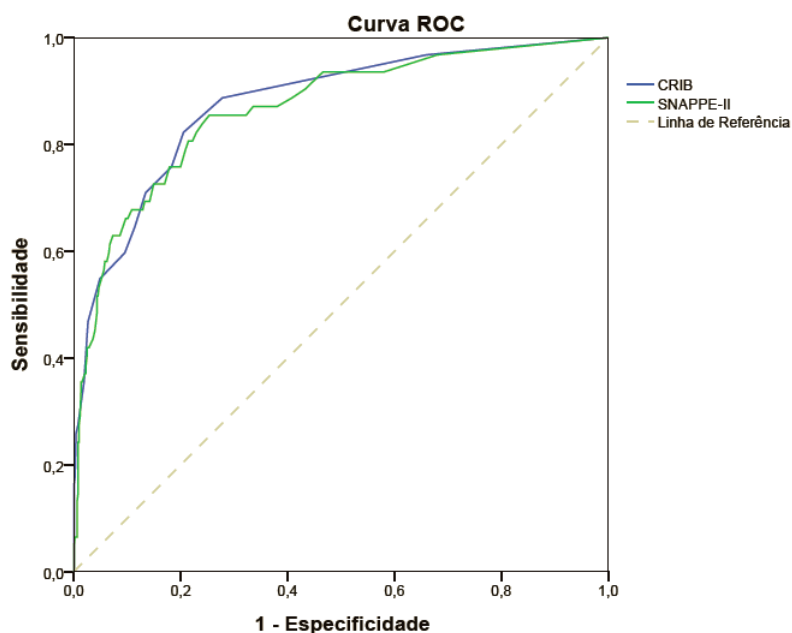


Figura 4.1: Curvas ROC dos índices CRIB e SNAPPE-II

As Tabelas 4.1 e 4.2 apresentam os valores dos coeficientes de correlação entre os indicadores de mortalidade para cada estado, sobrevivência e mortalidade, respetivamente. A medida usada foi o tau de kendall proposto por Hanley and McNeil [18].

Tabela 4.1: Correlação entre os índices para o estado de sobrevivência

		CRIB	SNAPPE-II
CRIB	coeficiente de correlação	1,00	0,507
	N	537	537
SNAPPE-II	coeficiente de correlação	0,507	1,00
	N	537	537

Tabela 4.2: Correlação entre os índices para o estado de mortalidade

		CRIB	SNAPPE-II
CRIB	coeficiente de correlação	1,00	0,675
	N	62	62
SNAPPE-II	coeficiente de correlação	0,675	1,00
	N	62	62

Com base nas informações expostas nas Tabelas 4.1 e 4.2, foi possível calcular o coeficiente de correlação médio e a área média, obtendo-se os valores 0,5895 e 0,8715, para as respetivas medidas. Com estes dados e com o auxílio da Tabela B.1 é possível determinar a correlação entre as áreas,  $r$ , que neste caso é aproximadamente 0,51. De modo a avaliar a diferença entre as duas áreas abaixo da curva ROC, recorre-se ao cálculo da razão crítica,  $Z$ , definida na equação 2.2.1, pela qual se obtém um valor de  $Z=0,356$ . A partir deste valor de  $Z$  calcula-se o valor  $p$  com recurso ao programa **R**, através do qual se obtém um valor de  $p$  de 0,7218406, pelos resultados pode-se afirmar que não existe diferença estatisticamente significativa entre os dois indicadores de mortalidade.

### 4.1.2 Comparação dos índices de mortalidade através do ROCR

A abordagem à metodologia ROC foi, aplicada no pacote **ROCR**, na linguagem **R**, obtendo-se a Figura 4.2, e também as medidas da área abaixo da curva para os dois índices, com um valor de 0,873 para o CRIB e 0,867 para o SNAPPE-II.

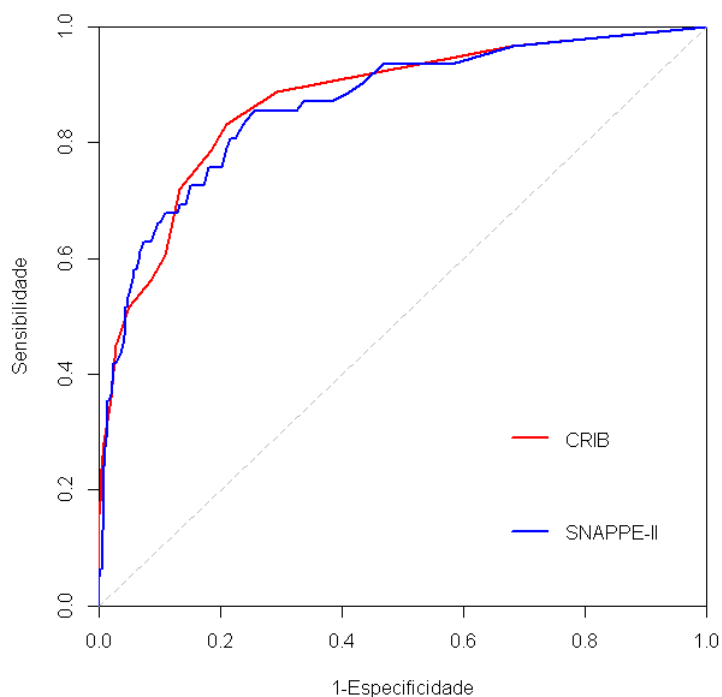


Figura 4.2: Curva empírica através do **ROCR**

Pela observação da Figura 4.1 são notórios pontos de interseção das duas curvas e momentos em que há alternância do desempenho dos indicadores de mortalidade. De forma a estudar este comportamento irá utilizar-se o pacote **Comp2ROC** na próxima secção.

### 4.1.3 Comparação dos índices de mortalidade através do Comp2ROC

Na utilização do pacote **Comp2ROC** é necessário uma atenção especial com o conjunto de dados a trabalhar, os valores da variável a comparar necessitam de estar por ordem crescente, de modo ao maior o valor dos índices corresponderem a uma maior probabilidade de mortalidade.

No que diz respeito à alternância de desempenho dos classificadores, o pacote **Comp2ROC**, fornece um estudo

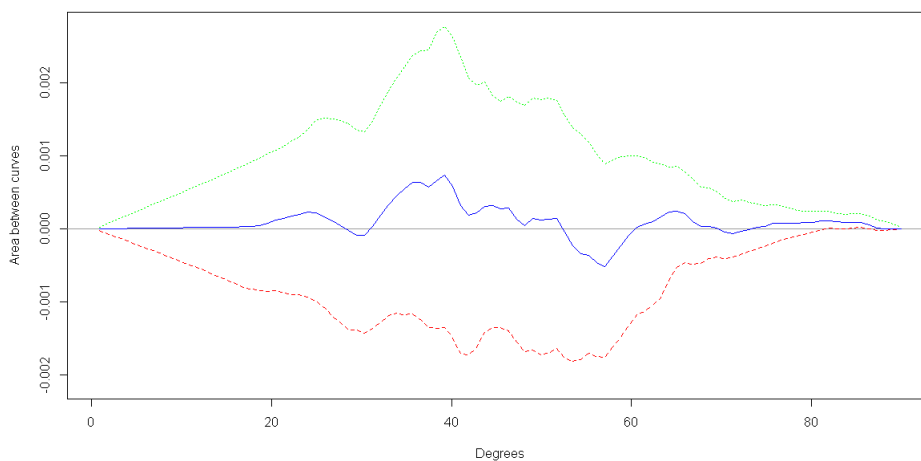


Figura 4.3: Diferença de áreas entre as curvas e respectivos intervalos de confiança através do **Comp2ROC**

estatístico sobre a análise comparativa das curvas ROC, pertencentes aos índices, bastante robusto. Através das funcionalidades inerentes a este pacote, consegue-se saber qual dos índices tem uma maior superação e o número de vezes em que se sucede esta alternância.

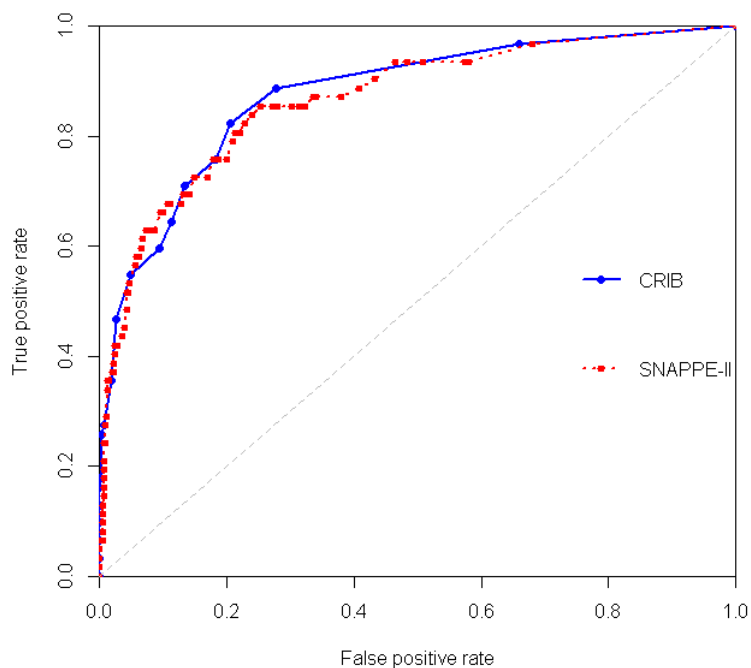


Figura 4.4: Curvas empíricas através do **Comp2ROC**

No caso em estudo, em 81,2% do espaço ROC, o CRIB tem melhor desempenho que o SNAPPE-II, identificando-

se seis pontos de interseção entre as curvas dos classificadores. A partir deste programa obtêm-se um gráfico que expressa a diferença de área entre as curvas e os respectivos intervalos de confiança, Figura 4.3, assim como os valores dessas medidas, de 0,00927 para a diferença de área entre as curvas e o respectivo intervalo de confiança, com limite inferior de -0,0609 e o limite superior de 0,0756.

As curvas ROC por este pacote desenhadas são visíveis na Figura 4.4. Numa análise isolada aos índices, obtém-se o valor da AUC, que para o índice CRIB é de 0,876 (0,025) , com um intervalo de confiança entre 0,83 a 0,92, como limite mínimo e máximo. E para o índice SNAPPE-II o valor da AUC é de 0,867 (0,026), com um intervalo de confiança de 0,81 a 0,91, como limite mínimo e máximo, respetivamente.

O programa **Comp2ROC** também calcula o coeficiente de correlação entre as áreas,  $r$ , que para o caso em estudo é de 0,733, pelo método proposto por DeLong et al. [10]. Como testes estatísticos sobre as diferenças entre as áreas, o programa usa a estatística de  $Z$ , proposta por Hanley e McNeil [18], pelo qual se obtém um valor de  $Z=0,497$ , e um valor de  $p$  associado ao teste de 0,6185, pelo qual se pode assumir que não existe diferença estatisticamente significativa entre os dois indicadores de mortalidade. No Apêndice C é possível observar o documento que contempla os dados citados anteriormente, obtidos através do Comp2ROC.

#### **4.1.4 Comparação dos índices de mortalidade através do caTools**

De forma abordar outras funções e pacotes disponíveis para o programa **R**, usou-se o pacote **caTools** com recurso à função **colAUC**, que permitiu o cálculo da AUC tanto do índice CRIB como do SNAPPE-II assim como a visualização gráfica das curvas ROC associadas. Obtiveram-se para os valores da AUC, 0,876 e 0,867 para o CRIB e o SNAPPE-II, respetivamente. Apesar desta função ser de fácil implementação, somente é possível o cálculo da área abaixo da curva e a visualização bidimensional do desempenho dos indicadores de mortalidade, 4.5. Porém apresenta algumas limitações como, no cálculo do desvio padrão, do intervalo de confiança e dos coeficientes de correlação.

#### **4.1.5 Comparação dos índices de mortalidade através do ROCNPA**

O programa **ROCNPA** é uma ferramenta bastante versátil no estudo da metodologia ROC. Os dados a trabalhar neste programa devem obedecer aos dois critérios, primeiramente selecionam-se os recém-nascidos que têm informação do CRIB e do SNAPPE-II simultaneamente e de seguida, ordenar os dados por ordem crescente da variável a comparar.

Através deste programa é possível obter as curvas ROC empíricas, como mostra a Figura 4.6, gráficos de distribui-

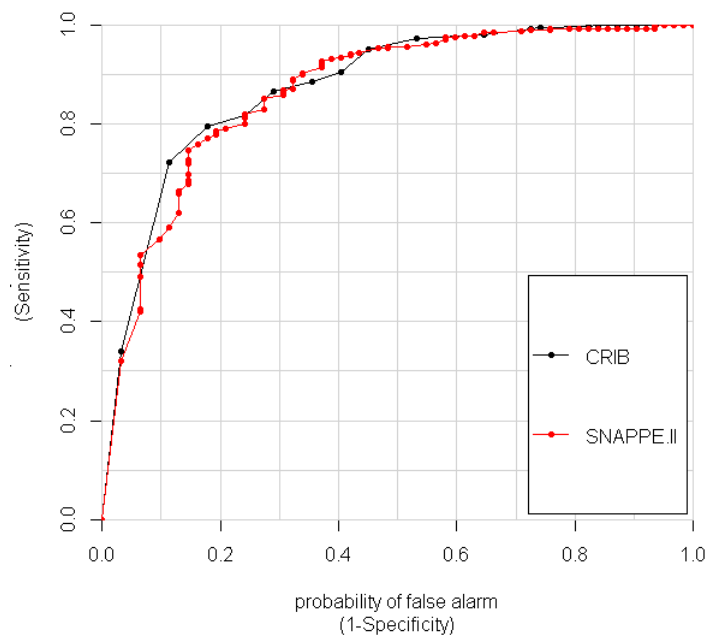


Figura 4.5: Curva Empirica através do **caTools**

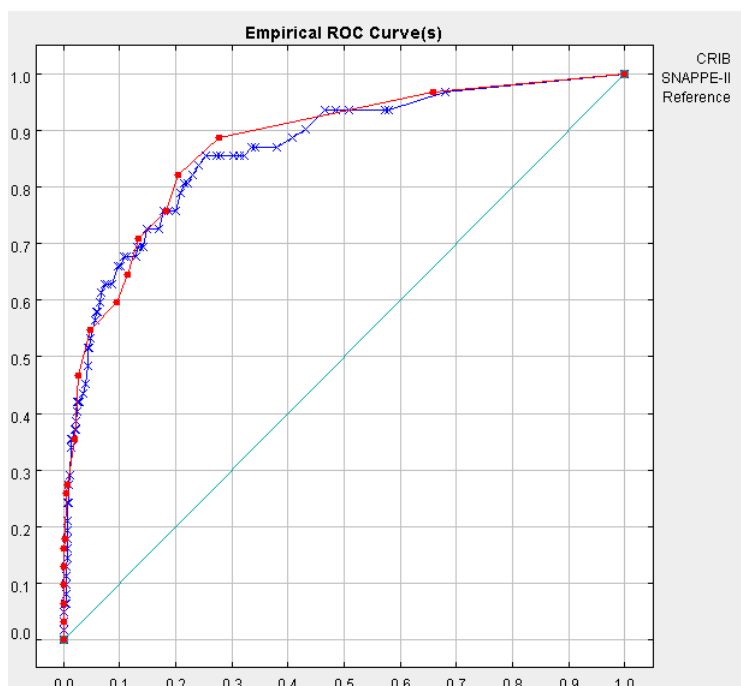


Figura 4.6: Curva Empirica através do **ROCNPA**



ção, representados pelas Figuras 4.7a e 4.7b referentes aos dois indicadores de mortalidade, mostram a frequência de sobrevividos e falecidos para cada pontuação final dos respectivos indicadores.

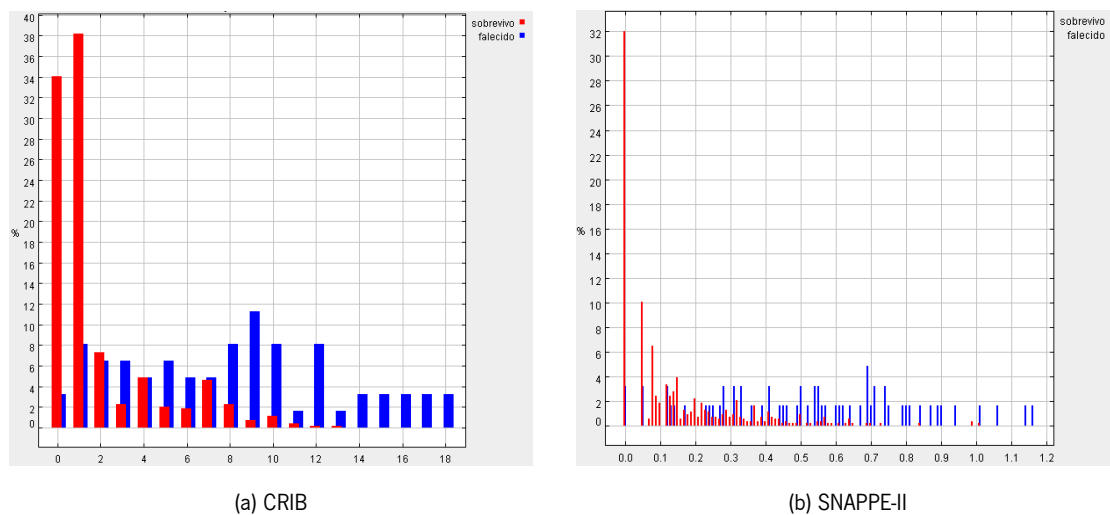


Figura 4.7: Gráfico da distribuição dos indicadores com recurso ao **ROCNP**A

### Correlação entre as AUC do CRIB e AUC do SNAPPE-II

Uma funcionalidade característica deste programa e que o distingue dos demais é, a possibilidade de fazer testes de comparação de áreas, e calcular as diferentes estimativas, para a AUC, apresentadas na Tabela 4.3. O valor do teste de comparação entre o CRIB e o SNAPPE-II é de 0,617, e o coeficiente de correlação entre estes dois indicadores de mortalidade, é de 0,733.

Tabela 4.3: Valores das áreas

	CRIB	SNAPPE-II
Wilcoxon	0,876	0,867
Regra do Trapézio	0,876	0,867
Binormal	0,846	0,860

## 4.2 Comparação do desempenho das variáveis Peso ao Nascimento e Idade Gestacional

Como referido anteriormente os indicadores de mortalidade em questão, CRIB e SNAPPE-II, são formados por um conjunto de variáveis fisiológicas que são determinadas no momento da admissão e durante as 12h seguintes. Mas de forma a tentar prever o mais antecipadamente possível o óbito do recém-nascido e de aliviar a motorização por parte da equipa médica, avaliaram-se duas variáveis isoladamente o peso à nascença (PN) e a idade gestacional (IG), também estudadas por outros autores devido à sua elevada capacidade discriminatória. Após a avaliação do desempenho destas com recurso à área abaixo da curva ROC no **SPSS**, obtiveram-se os valores de 0,785 (0,032) e 0,782 (0,028) para a idade gestacional e o peso ao nascimento, respetivamente. Na Figura 4.8 é possível observar as curvas ROC das variáveis peso à nascença e idade gestacional. A variável idade gestacional apresenta um intervalo de confiança a variar entre 0,723 e 0,848 para limite inferior e superior, respetivamente. E por sua vez, a variável peso à nascença apresenta um intervalo de confiança com um valor de 0,727 para o limite inferior e valor de 0,836 para o limite superior.

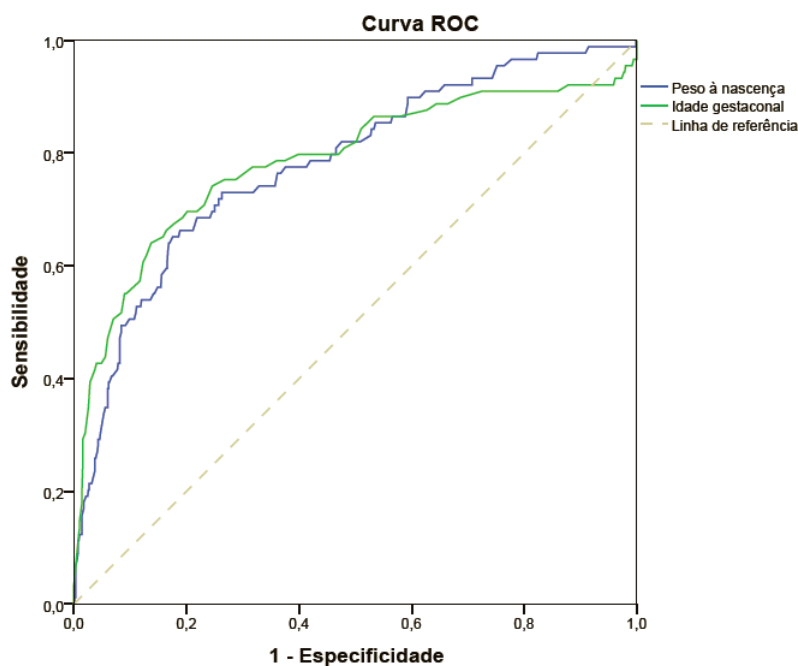


Figura 4.8: Curvas ROC para o peso à nascença e a idade gestacional

### 4.3 Comparação do desempenho dos indicadores de mortalidade com a variável peso através do ROCNPA

Com recurso ao programa **ROCNPA**, os índices, CRIB e SNAPPE-II e a variável peso, foram trabalhados simultaneamente de forma a fazer-se comparações múltiplas, a Figura 4.9 mostra as três curvas ROC num plano bidimensional. Os valores da AUC calculados foram de 0,876 (0,025), 0,867 (0,026) e 0,822 (0,091) para o CRIB, o SNAPPE-II e a variável peso. Para o índice CRIB e a variável peso o coeficiente de correlação é de -0,197 e para o índice SNAPPE-II e a variável peso é de -0,206.

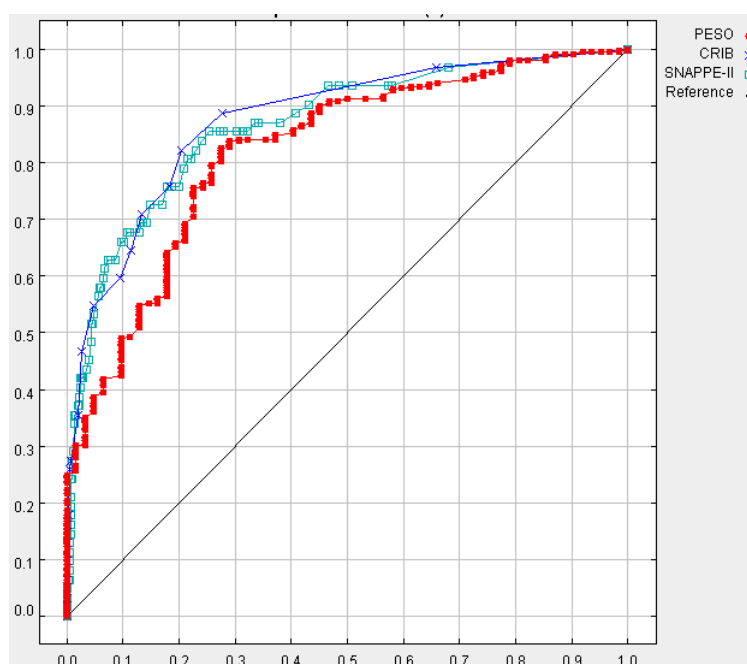


Figura 4.9: Curvas ROC para os indicadores CRIB e SNAPPE-II e variável peso

A correção de Bonferroni foi utilizada para efetuar as comparações múltiplas. Esta correção baseia-se na divisão do nível de significância,  $\alpha$ , pelo número de testes efetuados simultaneamente,  $p$ . Neste caso,  $\frac{\alpha}{p} = \frac{0,05}{3} = 0,017$ . O valor obtido é comparado com o valor de  $p$  obtido para cada um deles, Tabela 4.4. Neste caso, não são verificadas diferenças estatisticamente significativas.

No Apêndice D, é apresentado o formato de saída dos dados no programa **ROCNPA** e mais informações estatísticas, no qual integra também a Tabela 4.4.

#### 4.3. COMPARAÇÃO DO DESEMPENHO DOS INDICADORES DE MORTALIDADE COM AVARIÁVEL PESO ATRAVÉS DO RO

Tabela 4.4: Testes de comparação múltipla para variáveis correlacionadas, valores de  $p$

	CRIB	SNAPPE-II
PESO	0.58231931	0.645516252
CRIB	-	0.617075066



# Capítulo 5

## Discussão

Ao longo do presente estudo pretendeu-se avaliar o desempenho dos classificadores de mortalidade e a versatilidade das ferramentas adaptadas ao estudo da metodologia ROC.

Quando se fala em desempenho de classificadores, referimo-nos a uma medida quantitativa, ao valor da AUC. Neste estudo, os valores apresentados são obtidos pelo cálculo da estatística de Wilcoxon-Mann-Whitney, [18]. A esta medida está associado o valor SE, que representa a imprecisão associada à estimativa de um determinado parâmetro. Com o aumento da amostra, aumenta a precisão da estimativa e conseqüentemente o valor do SE diminui. As Tabelas 5.1 e 5.2 apresentam os valores da AUC e os SE, em cada programa utilizado, para o índice CRIB e SNAPPE-II, respetivamente.

Relativamente aos dois indicadores de mortalidade, o CRIB apresenta um desempenho superior, com a AUC de 0,876 ( 0,025) comparativamente ao SNAPPE-II, com a AUC de 0,867 (0,026).

Os resultados da área abaixo da curva para os indicadores de mortalidade no caso em estudo vão ao encontro aos sustentados por diversos autores, Tabela 5.4.

Os estudos sobre o índice SNAPPE-II são bastante reduzidos sobre o ponto de vista do seu desempenho através da metodologia ROC. Dos autores estudados somente Gagliardi et al. [15] efetuaram a comparação do desempenho destes dois indicadores de mortalidade. Afirmando que o índice CRIB tem um melhor poder discriminante face ao SNAPPE-II, no presente estudo faz-se a mesma observação crítica, porém com valores diferentes aos apresentados pelos mesmos.

Entre estes há uma notória diferença nos valores de AUC, o que traduz-se uma diferença de desempenho, que

poderá estar associada às variáveis que diferem entre os índices em questão. Segundo Zupancic et al. (2007) citado por Buhner [7], uma das causas possíveis para o poder discriminante do SNAPPE-II ser inferior ao CRIB, poderá estar relacionado com facto da variável, anomalias congénitas não estar incluída no conjunto de variáveis que compõem o indicador de mortalidade SNAPPE-II.

Tabela 5.1: Resumo dos valores obtidos para o índice CRIB

	SPSS	ROCR	caTools	Comp2ROC	ROCNP
AUC	0,876	0,873	0,876	0,876	0,876
SE(AUC)	0,025			0,025	0,025

Tabela 5.2: Resumo dos valores obtidos para o índice SNAPPE-II

	SPSS	ROCR	caTools	Comp2ROC	ROCNP
AUC	0,867	0,867	0,867	0,867	0,867
SE(AUC)	0,026			0,026	0,026

Como é perceptível pela análise das Tabelas 5.1 e 5.2, os valores de AUC obtidos pelos vários programas foram iguais, à exceção do pacote **ROCR** que para o índice CRIB deu um resultados ligeiramente mais baixo que os demais. Os pacotes **caTools** e **ROCR** apresentam algumas limitações, no que diz respeito ao cálculo do SE, foi impossível através dos mesmos, razão pela qual esta medida não aparece nas Tabelas 5.1 e 5.2.

Dos programas estudados o **ROCNP** revelou-se o mais versátil, para além de suprimir algumas lacunas existentes em alguns programas. Tem como características dominantes, a realização de cálculos de forma rápida e simples, a apresentação gráfica de duas ou mais curvas ROC empíricas, permite a comparação de mais de três curvas ROC quer sejam dados provenientes de amostras independentes ou correlacionadas, utilizando a estatística de teste  $Z$ , definida por Hanley e McNeil [18].

Pollack et al. [26], enaltecem o desempenho da variável PN, sublinhando o ótimo resultado alcançado, com um valor de AUC de 0,869, motivação que levou ao estudo desta variável da variável PN.

Tabela 5.3: Resumo dos valores obtidos para as variáveis IG e PN

	SPSS	ROCNP
AUC (IG)	0,785	
AUC (PN)	0,782	0,822

No presente estudo, as variáveis PN e IG, avaliadas isoladamente não têm um desempenho tão bom como os indicadores de mortalidade referidos, Tabela 5.3. Porém o valor da AUC alcançado através do **ROCNP** é francamente superior ao obtido pelo **SPSS**. Esta divergência pode estar relacionada com o formato dos dados, enquanto que no **ROCNP** trabalha-se com amostras emparelhadas, sendo uma condição necessária que todos os campos das variáveis tenham um valor válido, ao invés do que acontece com o SPSS, que analisa as variáveis isoladamente, não apresentado restrição se as amostras não forem emparelhadas.

Na Tabela 5.4, são apresentados os resultados obtidos pelos autores mencionados ao longo deste estudo.

Tabela 5.4: Valores da AUC, para o CRIB, SNAPPE-II, PN e IG em vários estudos

Estudo	N	CRIB	SNAPPE-II	Peso ao Nascimento	Idade Gestacional
Bumer et al. [2]	112	0,93	-	0,83	0,83
Braga et al. [6]	169	0,90	-	0,77	-
Buhrer et al. [7]	1358	0,82	-	0,74	0,71
Gagliardi et al. [15]	720	0,903	0,837	-	-
SNC e INN [16]	695	0,83	-	0,74	-
Marshall et al. [20]	1801	0,79	-	0,79	0,77
Parry et al. [23]	3027	0,79	-	0,77	0,80
Pollack et al. [26]	522	0,891	-	0,869	-
Richardson et al. [28]	1768	-	0,91	-	-
Sarquis et al. [30]	100	0,877	-	-	-





## Capítulo 6

### Conclusões e Trabalho Futuro

Ao longo do desenvolvimento do presente estudo, com base nos resultados obtidos é possível mencionar o CRIB como o indicador de risco de mortalidade neonatal com melhor desempenho, com uma AUC de  $0,876 \pm 0,025$  comparativamente ao SNAPPE-II, com uma AUC de  $0,867 \pm 0,026$ , no entanto não foram detetadas, entre estes, diferenças estatisticamente significativas, com um  $Z$  de 0,497 e um valor  $p$  de 0,6185. O valor da área abaixo da curva e o erro padrão permitem fundamentar o supracitado, para além deste indicador de mortalidade conter um menor número de parâmetros de entrada. No entanto, também foram estudadas variáveis com grande expressão na previsão da mortalidade de recém-nascidos de muito baixo peso que compõem os índices em estudo, como é o caso do peso à nascença e a idade gestacional.

A exatidão destas na previsão da mortalidade é menor que a apresentada pelos indicadores de mortalidade porém numa primeira abordagem a sua análise é crucial. Tendo em conta que são variáveis em que a recolha da informação é simples, e que são conhecidas ao longo do período de gestação tanto pela progenitora como pela equipa médica, tornam estas variáveis bons indicadores na toma da decisão da indução do parto.

Esta análise refletiva só foi possível com recurso a vários programas de análise da metodologia ROC, dos quais se realça o **ROCNP**, devido à versatilidade.

Numa perspetiva de trabalho futuro, seria oportuno fazer um estudo com os sobreviventes que integram este conjunto de dados, com o intuito de fazer um levantamento das morbilidades que possam estar-lhes associadas. E comparar a existência das mesmas, com a pontuação do índice a que foram admitidas nos serviços de cuidados intensivos de neonatologia.

Visto que nas unidades neonatais, os recém-nascidos de muito baixo peso são colocados num ambiente de elevada tecnologia que, apesar de necessário para a sua sobrevivência é de elevado risco para o cérebro, pulmões, intestino e órgãos dos sentidos [8] [9]. Esta condição está intimamente relacionadas com a morbidade, segundo Cunha et al. [8], sendo a prevalência das sequelas ao nível, da surdez, cegueira, problemas neurológicos, paralisia cerebral e atraso no desenvolvimento.

## Bibliografia

- [1] Bamber, D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology* (1974).
- [2] Baumer, J. H., Wright, D., and Mill, T. Illness severity measured by crib score: a product of changes in perinatal care? *Archives of Disease in Childhood* (1997).
- [3] Bradley, A. P. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition* (1998).
- [4] Braga, A. C. *Curvas ROC: Aspectos Funcionais e Aplicações*. PhD thesis, Universidade do Minho, 2000.
- [5] Braga, A. C. Compare two roc curves that intersect. *Repository CRAN* (2013).
- [6] Braga, A. C., and Oliveira, P. Diagnostic analysis based on roc curves: theory and applications in medicine. *International Journal of Health Care Quality Assurance* (2003).
- [7] Buhner, C., Metze, B., and Obladen, M. Crib, crib ii, birth weight or gestational age to assess mortality risk in very low birth weight infants? *Acta Paediatrica* (2008).
- [8] Cunha, M., Bettencourt, A., Almeida, A., Mimoso, G., Soares, P., Tomé, T., and do Registo Nacional de Muito Baixo Peso, G. recém nascido de extremo baixo peso. estado aos 2-3 anos. resultados do registo nacional de muito baixo peso de 2005 e 2006. *Acta Pediátrica Portuguesa - Sociedade Portuguesa de Pediatria* (2013).

- [9] Cunha, M., Cadete, A., Virella, D., and do Registo Nacional de Muito Baixo Peso, G. Acompanhamento dos recém-nascidos de muito baixo peso em Portugal. *Acta Pediátrica Portuguesa - Sociedade Portuguesa de Pediatria* (2010).
- [10] DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* (1988).
- [11] Dorling, J., Field, D. J., and Manktelow, B. Neonatal disease severity scoring systems. *Arch Dis Child Neonatal* (2005).
- [12] Fawcett, T. An introduction to roc analysis. *ELSEVIER* (2005).
- [13] Fortes, C. *Aplicação da Metodologia ROC na Análise de Dados em Microarrays*. PhD thesis, Universidade de Lisboa, 2012.
- [14] Fortes Filho, J., Dill, J., Ishizaki, A., Aguiar, W., Silveira, R., and Procianoy, R. Score for neonatal acute physiology and perinatal extension ii as a predictor of retinopathy of prematurity: Study in 304 very-low-birth-weight preterm infants. *Ophthalmologica* (2009).
- [15] Gagliardi, L., Cavazza, A., Brunelli, A., Battaglioli, M., Merazzi, D., Tandoi, F., Cella, D., Perotti, G. F., Pelti, M., Stucchi, I., Frisone, F., Avanzini, A., R, B., and study group, N. Assessing mortality risk in very low birthweight infants: a comparison of crib, crib-ii, and snappe-ii. *Arch Dis Child Fetal Neonatal Ed* (2004).
- [16] Group, S. N. C. C. S., and the International Neonatal Network. Crib (clinical risk index for babies), mortality, and impairment after neonatal intensive care. *Lancet* (1995).
- [17] Hanley, J., and McNeil, B. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology* (1982).
- [18] Hanley, J. A., and McNeil, B. J. The method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* (1983).
- [19] Krzanowski, W., and Hand, D. J. Roc curves for continuous data - monographs on statistics and applied probability; 111. *Taylor and Francis Group, LLC* (2009).

- [20] Marshall, G., Tapia, J., Ivonne, D., Grandi, C., Barros, C., Alegria, A., Standen, J., Panizza, R., Roldan, L., Musante, G., Bancari, A., Bambaren, E., Lacarruba, J., Hubner, M., fabres, J., Decaro, M., Mariani, G., Kurlat, I., and Gonzalez, A. A new score for predicting neonatal very low birth weight mortality risk in the neocosur south american network. *Journal of Perinatology* (2005).
- [21] Metz, C. Basic principles of roc analysis. *Seminars in Nuclear Medicine* (1978).
- [22] Metz, C. Roc methodology in radiologic imaging. *Investigative Radiology* (1986).
- [23] Parry, G., Tucker, J., and Tarnow-Mordi, W. Crib ii: an update of the clinical risk index for babies score. *The Lancet* (2003).
- [24] Peixoto, J., Guimarães, H., Machado, M., Martins, V., Mimoso, G., Neto, T., Virella, D., and do Registo Nacional do Recém-Nascido de Muito Baixo Peso, G. Nascer prematuro em portugal. estudo multicêntrico nacional 1996-2000. *Edições Bial* (2002).
- [25] Pepe, M. S. The statistical evaluation of medic tests for classification and prediction. *Oxford University Press* (2004).
- [26] Pollack, M., Koch, M., Bartel, D., Rapoport, I., Dhanireddy, R., El-mohandes, A., Harkavy, K., Subramanian, S., and of Columbia Neonatal Network, D. A comparison of neonatal mortality risk prediction models in very low birth weight infants. *Pediatrics* (2000).
- [27] Richardson, D. K., Corcoran, J. D., Escobar, G. J., and Lee, S. K. Snap-ii and snappe-ii: Simplified newborn illness severity and mortality risk scores. *The Journal of Pediatrics* (2001).
- [28] Richardson, D. K., Phibbs, C. S., Gray, J. E., McCormick, M. C., Workman-Daniels, K., and Goldmann, D. A. Birth weight and illness severity: Independent predictors of neonatal mortality. *Pediatrics* (1993).
- [29] Rocha, G., Rocha, A., Martins, A., Silva, G., Guedes, B., and Guimarães, H. Aspectos clínicos da prematuridade extrema parte i- morbidade e mortalidade neonatais. *Acta Pediátrica Portuguesa - Sociedade Portuguesa de Pediatria* (2002).
- [30] Sarquis, A. L., Miyaki, M., and Cat, M. Aplicação do escore crib para avaliar o risco de mortalidade neonatal. *Jornal de Pediatria* (2002).

- [31] Schnell, F. N. The apgar in perspective. *Canadian Family Physician* (1983).
- [32] Swets, J. A. Signal detection theory and roc analysis in psychology and diagnostics. *Scientific Psychology Series* (1996).
- [33] Tuszynski, J. Tools: moving window statistics, gif, base64, roc auc, etc. *Repository CRAN* (2014).
- [34] Zou, K., O'Malley, J., and Mauri, L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *American Heart Association* (2007).

# **Apêndices**





# Apêndice A

## Tabelas Relativas aos Índices

Tabela A.1: Componentes da escala APGAR, adaptado de [31]

	0	1	2
Frequência Cardíaca	Ausência	<100	>100
Respiração	Ausência	Irregular	Choro forte
Irritabilidade reflexa	Ausência	Careta	Choro
Tônus muscular	Flácido	Flexão mediana	Boa Flexão
Cor	Pálidez	Acrocianose	Rosado

Tabela A.2: Valores limite para o peso consoante a idade gestacional, adaptado de [27]

Idade Gestacional (semanas)	3° Percentil Peso à Nascimento (g)
22	320
23	380
24	430
25	500
26	580
27	670
28	740
29	820
30	920
31	1030
32	1140
33	1280
34	1420
35	1580
36	1750
37	1920
38	2120
39	2350
40	2520
41	2660
>41	2750

## **Apêndice B**

### **Tabela de Coeficientes de Correlação entre Áreas**

Tabela B.1: Tabela de Coeficientes de Correlação entre áreas, adaptado de [18]

Média das Correlações	Média das áreas											
	.70	.725	.750	.775	.800	.825	.850	.875	.900	.925	.950	.975
0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01
0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02
0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.03	0.02
0.08	0.07	0.07	0.07	0.07	0.07	0.06	0.06	0.06	0.06	0.05	0.04	0.03
0.10	0.09	0.09	0.09	0.09	0.08	0.08	0.08	0.07	0.07	0.06	0.06	0.04
0.12	0.11	0.11	0.11	0.10	0.10	0.10	0.09	0.09	0.08	0.08	0.07	0.05
0.14	0.13	0.12	0.12	0.12	0.12	0.11	0.11	0.11	0.10	0.09	0.08	0.06
0.16	0.14	0.14	0.14	0.14	0.13	0.13	0.13	0.12	0.11	0.11	0.09	0.07
0.18	0.16	0.16	0.16	0.16	0.15	0.15	0.14	0.14	0.13	0.12	0.11	0.09
0.20	0.18	0.18	0.18	0.17	0.17	0.17	0.16	0.15	0.15	0.14	0.12	0.10
0.22	0.20	0.20	0.19	0.19	0.19	0.18	0.18	0.17	0.16	0.15	0.14	0.11
0.24	0.22	0.22	0.21	0.21	0.21	0.20	0.19	0.19	0.18	0.17	0.15	0.12
0.26	0.24	0.23	0.23	0.23	0.22	0.22	0.21	0.20	0.19	0.18	0.16	0.13
0.28	0.26	0.25	0.25	0.25	0.24	0.24	0.23	0.22	0.21	0.20	0.18	0.15
0.30	0.27	0.27	0.27	0.26	0.26	0.25	0.25	0.24	0.23	0.21	0.19	0.16
0.32	0.29	0.29	0.29	0.28	0.28	0.27	0.26	0.26	0.24	0.23	0.21	0.18
0.34	0.31	0.31	0.31	0.30	0.30	0.29	0.28	0.27	0.26	0.25	0.23	0.19
0.36	0.33	0.33	0.32	0.32	0.31	0.31	0.30	0.29	0.28	0.26	0.24	0.21
0.38	0.35	0.35	0.34	0.34	0.33	0.33	0.32	0.31	0.30	0.28	0.26	0.22
0.40	0.37	0.37	0.36	0.36	0.35	0.35	0.34	0.33	0.32	0.30	0.28	0.24
0.42	0.39	0.39	0.38	0.38	0.37	0.36	0.36	0.35	0.33	0.32	0.29	0.25
0.44	0.41	0.40	0.40	0.40	0.39	0.38	0.38	0.37	0.35	0.34	0.31	0.27
0.46	0.43	0.42	0.42	0.42	0.41	0.40	0.39	0.38	0.37	0.35	0.33	0.29
0.48	0.45	0.44	0.44	0.43	0.43	0.42	0.41	0.40	0.39	0.37	0.35	0.30
0.50	0.47	0.46	0.46	0.45	0.45	0.44	0.43	0.42	0.41	0.39	0.37	0.32
0.52	0.49	0.48	0.48	0.47	0.47	0.46	0.45	0.44	0.43	0.41	0.39	0.34
0.54	0.51	0.50	0.50	0.49	0.49	0.48	0.47	0.46	0.45	0.43	0.41	0.36
0.56	0.53	0.52	0.52	0.51	0.51	0.50	0.49	0.48	0.47	0.45	0.43	0.38
0.58	0.55	0.54	0.54	0.53	0.53	0.52	0.51	0.50	0.49	0.47	0.45	0.40
0.60	0.57	0.56	0.56	0.55	0.55	0.54	0.53	0.52	0.51	0.49	0.47	0.42
0.62	0.59	0.58	0.58	0.57	0.57	0.56	0.55	0.54	0.53	0.51	0.49	0.45
0.64	0.61	0.60	0.60	0.59	0.59	0.58	0.58	0.57	0.55	0.54	0.51	0.47
0.66	0.63	0.62	0.62	0.62	0.61	0.60	0.60	0.59	0.57	0.56	0.53	0.49
0.68	0.65	0.64	0.64	0.64	0.63	0.62	0.62	0.61	0.60	0.58	0.56	0.51
0.70	0.67	0.66	0.66	0.66	0.65	0.65	0.64	0.63	0.62	0.60	0.58	0.54
0.72	0.69	0.69	0.68	0.68	0.67	0.67	0.66	0.65	0.64	0.63	0.60	0.56
0.74	0.71	0.71	0.70	0.70	0.69	0.69	0.68	0.67	0.66	0.65	0.63	0.59
0.76	0.73	0.73	0.72	0.72	0.72	0.71	0.71	0.70	0.69	0.67	0.65	0.61
0.78	0.75	0.75	0.75	0.74	0.74	0.73	0.73	0.72	0.71	0.70	0.68	0.64
0.80	0.77	0.77	0.77	0.76	0.76	0.76	0.75	0.74	0.73	0.72	0.70	0.67
0.82	0.79	0.79	0.79	0.79	0.78	0.78	0.77	0.77	0.76	0.75	0.73	0.70
0.84	0.82	0.81	0.81	0.81	0.81	0.80	0.80	0.79	0.78	0.77	0.76	0.73
0.86	0.84	0.84	0.83	0.83	0.83	0.82	0.82	0.81	0.81	0.80	0.78	0.75
0.88	0.86	0.86	0.86	0.85	0.85	0.85	0.84	0.84	0.83	0.82	0.81	0.79
0.90	0.88	0.88	0.88	0.88	0.87	0.87	0.87	0.86	0.86	0.85	0.84	0.82

## Apêndice C

### Resultado de saída do pacote Comp2ROC

ROC SAMPLING RESULTS

Number of sampling lines: 100

Proportion CRIB : 0.8118812

Proportion SNAPPE-II : 0.1584158

Proportion ties: 0.02970297

---

#### CRIB

---

Area: 0.8763441

Standard Error: 0.02450573

Area through Trapezoidal Method: 0.8763048

CI Upper bound (Percentil Method): 0.9194652

CI Lower bound (Percentil Method): 0.8261995

---

---

#### SNAPPE-II

---

Area: 0.8670782

Standard Error: 0.02624983

Area through Trapezoidal Method: 0.8670746

CI Upper bound (Percentil Method): 0.9135291

CI Lower bound (Percentil Method): 0.8147095

---

Correlation Coefficient between areas: 0.7332236

#### TEST OF DIFFERENCES

Z stats: 0.4979519

p-value: 0.6185179

Sum of Global Areas Differences (TS): 0.009265934

CI Upper bound (Percentil Method): 0.07562699

CI Lower bound (Percentil Method): -0.06089761

Number of Crossings: 6

## Apêndice D

### Teste de Comparação de Áreas

#### Multiple comparison tests for correlated variables

	CRIB	SNAPPE-II
PESO	0.58231931	0.645516252
CRIB	-	0.617075066

#### Coefficients of correlation

	PESO	CRIB	SNAPPE-II
PESO	1.0	-0.19739022116572974	-0.20639900085638796
CRIB	-	1.0	0.7332236465096464
SNAPPE-II	-	-	1.0

---

#### PESO

Wilcoxon Area: 0.8216345287439178

Standard error (SE): 0.09107408105946108



Area calculated with the trapezoidal rule: 0.821634528743918

Regression line:

- $a = 1.1730666658229638$
- $b = 0.7641354201230738$
- coefficient of determination (r-square) = 0.9261922740008052

Binormal area: 0.82381448

---

### **CRIB**

Wilcoxon Area: 0.8763440860215054

Standard error (SE): 0.0245057335825185

Area calculated with the trapezoidal rule: 0.8763440860215053

Regression line:

- $a = 1.3030926848843156$
- $b = 0.8026086187967427$
- coefficient of determination (r-square) = 0.8733032481682249

Binormal area: 0.846135756

---

### **SNAPPE-II**

Wilcoxon Area: 0.8670781522196191

Standard error (SE): 0.026249826319163252

Area calculated with the trapezoidal rule: 0.867078152219619

Regression line:

- $a = 1.4859761378270069$
- $b = 0.9418401702331904$

- coefficient of determination (r-square) = 0.9236722225158245

Binormal area: 0.859928875

