

UNIVERSIDADE ABERTA



UNIVERSIDADE
AbERTA
www.uab.pt

Fatores preditores de complicações após transplantes renais: Análise Estatística Multivariada

Rosa Ildebranda Pedrosa Bompastor Cerqueira

Mestrado em Bioestatística e Biometria, Universidade Aberta.

2016

UNIVERSIDADE ABERTA



UNIVERSIDADE
AbERTA
www.uab.pt

Fatores preditores de complicações após transplantes renais: Análise Estatística Multivariada

Rosa Ildebranda Pedrosa Bompastor Cerqueira

Mestrado em Bioestatística e Biometria, Universidade Aberta.

**Dissertação orientada pela Professora Doutora Catarina S Nunes,
Universidade Aberta**

2016

RESUMO

O transplante renal é uma oportunidade de vida para pacientes em estadios finais de doença renal. Muitos dos pacientes propostos para transplante renal, apresentam múltiplas comorbidades, o que torna a sua abordagem um desafio. Por outro lado, o transplante renal exige que um conjunto de condições sejam tomadas em consideração para a seleção do dador e do recetor, com implicações decorrentes desses critérios, na evolução do transplante.

O presente trabalho consiste na análise de fatores que possam prever a ocorrência de complicações após transplantes renais, quer sejam de dadores vivos, quer sejam de dadores cadáver.

Pretende-se compreender quais os fatores relacionados com a ocorrência de complicações tais como aumento do tempo de internamento, reinternamento, reintervenção, falhas hepáticas, infeções, complicações cardiovasculares, respiratórias, urorenais e endócrinas. Para tal foi utilizada uma base de dados dos 153 doentes transplantados entre janeiro 2012 e dezembro de 2013, num dos maiores centros de transplantes renais do país e na qual foram registadas informações pré e pós transplante relativas a características clínicas e demográficas de dadores e recetores, totalizando 58 variáveis. A morbilidade associada a esta doença leva à necessidade de identificação e melhoria das variáveis dependentes do paciente que influenciam os resultados, de modo a melhorar a curto prazo os resultados de sucesso.

Estudou-se a influência de variáveis como: tipo de dador, idade, sexo e ASA do doente, parentesco e diferença de idades entre dador e recetor, comorbidades, tempo e técnica de diálise, tempo de isquemia, técnica anestésica e tipo de analgesia pós operatória, nos resultados pós operatórios.

Foram utilizados métodos de análise multivariada, incluindo modelos de regressão linear e logística. Neste sentido e usando como variáveis resultado, o tempo de internamento, as complicações pós operatórias e as reintervenções e, recorrendo a técnicas de seleção de variáveis implementadas no SPSS, seleção passo a passo progressiva ou regressiva (Forward e Backward) foram construídos modelos que permitiram selecionar as variáveis que influenciam o sucesso dos transplantes renais. Com o objetivo de avaliar o desempenho dos modelos construídos recorreu-se à Curva ROC e respetiva área. Este trabalho permitiu identificar que a presença de hipertensão arterial, a condição física do doente (ASA) e o tipo de dador cadáver são fatores de risco na previsão de complicações pós operatórias; as complicações urorenais, complicações infecciosas, tempo em diálise, DAP e DCI são fatores de risco na previsão do tempo de internamento e que as complicações cardiovasculares, complicações endócrinas e complicações urorenais são fatores de risco na previsão de reintervenções.

Palavras chave

Transplante Renal, Fatores preditores, Dadores vivos e cadáver, Complicações pós operatórias, Estatística Multivariada

ABSTRACT

Kidney transplantation is an opportunity of life for patients with end-stage kidney disease. Many patients proposed for renal transplant, have multiple comorbidities, which makes this approach a challenge. Furthermore, renal transplantation requires a set of conditions to be taken into account for the selection of the donor and recipient. This selection criteria is important since it has implications in the transplant final results.

This work consists the factors analysis that may predict the occurrence of complications after kidney transplants from living or cadaver donors.

The aim is understand what factors are related to the occurrence of complications such as increased length of hospital stay, readmission, reoperation, liver failure, infections, cardiovascular, respiratory, urologic and endocrine. A database of 153 transplant patients between January 2012 and December 2013, of one of the largest of kidney transplants centers in the country, will be used in this work. The data of 58 clinical variables (pre and post transplant) related to the donors and recipients were obtained. The patient's or donor's identification depend factors that can influence the transplant outcomes are an important step to improve the short-temp success of this treatment.

We studied the influence of variables such as type of donor, patient age, age difference between donor and recipient, comorbidities, time on dialysis and technique, ischemia time, anesthetic technique and type of postoperative analgesia; on postoperative outcomes.

Multivariate analysis methods were used, including linear and logistic regression models. The time of hospital stay, the occurrence of post operative complications and re-interventions, are used as outcome measurements for the kidney transplant. Several techniques for variable selection were implemented in SPSS, in order to build significant prediction models. This work allows us to find (dependent of the outcome being considered) that the presence of hypertension, the receptor physical status (ASA) and the type of deceased donor; dialysis time, DAP and DCI are risk factors for the short term success of renal transplant. And that the cardiovascular, endocrine and urologic complications are independent risk factors in for post-transplant re-interventions.

Keywords

Renal Transplant, Predictors, living donors and cadaver, post operative complications, Multivariate Statistics

AGRADECIMENTOS

Início esta secção com o meu sincero agradecimento, pela compreensão, apoio e disponibilidade da minha orientadora Professora Doutora Catarina S. Nunes, quer seja no esclarecimento de dúvidas como no fornecimento de sugestões e orientações para a elaboração desta tese.

Agradeço à Universidade Aberta e a todos os docentes com quem tive o prazer de partilhar dúvidas e assimilar conhecimentos que tornaram possível a realização deste trabalho.

Agradeço à Dr^a Manuela Casal (Médica Anestesiologista do Centro Hospitalar do Porto), pela disponibilização da base de dados que tornou possível este trabalho.

E não seria justa se não referisse aqui, que embora a realização deste meu mestrado não tenha ocorrido num período muito favorável da minha vida, que por razões várias, pessoais e familiares, me impediram de trabalhar nele com a garra e dedicação que me é característica, tive a felicidade de poder contar com o apoio, a força e o carinho dos meus três filhos e que nem sempre soube retribuir e demonstrar o quanto me faziam bem. A vocês os três, António, Érica e Francisco, o meu muito obrigada pela vossa paciência e apoio.

Agradeço também aos meus pais, apesar de já falecidos, António Bompastor e Rosa Maria, pelos ensinamentos que sempre me transmitiram e que me incutiram a garra e o espírito de luta com que me proponho a atingir os meus objetivos.

De uma forma geral agradeço a todos os meus familiares que estiveram ao meu lado e acreditaram nas minhas capacidades, agradecendo em especial ao meu marido, Orlando pela paciência e amizade que disponibilizou sempre.

CONTEÚDO

RESUMO	iii
ABSTRACT	v
AGRADECIMENTOS	vii
CONTEÚDO	ix
ÍNDICE DE TABELAS	xi
ÍNDICE DE GRÁFICOS	xiii
ÍNDICE DE FIGURAS	xiv
Lista de abreviaturas	xv
INTRODUÇÃO	1
PARTE I	5
ENQUADRAMENTO TEÓRICO	5
1 – ENQUADRAMENTO E MOTIVAÇÃO	7
2 - TÉCNICAS DE INFERÊNCIA ESTATÍSTICA NÃO PARAMÉTRICA	9
3 – SOFTWARE UTILIZADO	12
4 - ANÁLISE DE REGRESSÃO	13
4.1. Correlação Múltipla e Coeficiente de Determinação	16
4.2. Testes de Significância	18
4.3. Pressupostos da Análise de Regressão	19
4.3.1. Linearidade	19
4.3.2. Independência dos resíduos	20
4.3.3. Homocedasticidade	21
4.3.4. Normalidade dos resíduos	22
4.4. Multicolinearidade	23
5 - REGRESSÃO LOGÍSTICA	24
5.1. Técnicas de Inferência Estatística aplicadas à regressão	25
5.1.1. Inferência no modelo de regressão	25
5.1.1.1. Teste de Wald	25
5.1.1.2. Teste da Razão de Verossimilhança	26
5.1.1.3. Intervalos de Confiança na Regressão Logística	27
5.1.1.4. Razão de possibilidades (odds ratio)	27
5.1.1.5. Teste de Ajustamento do modelo	29
5.1.1.6. Análise de resíduos	31

5.1.1.7. Classificação da previsão do modelo.....	34
6 - ANÁLISE DE CLUSTERS	37
6.1. Metodologia da análise de clusters.....	37
7 - ANÁLISE DISCRIMINANTE	39
PARTE II.....	43
MATERIAL E RESULTADOS	43
1 - PARTICIPANTES E VARIÁVEIS DO ESTUDO.....	45
2 - ANÁLISE DESCRITIVA	46
3 - CORRELAÇÕES	49
4 - TESTE χ -QUADRADO	53
5 - REGRESSÃO LINEAR MÚLTIPLA.....	56
6 - REGRESSÃO LOGÍSTICA.....	67
1 - Variável dependente: Complicações-pos-op30diasAlta.....	67
2 – Variável dependente: Complicações-pos-op30diasAlta	72
3 – Variável dependente Reintervensões.....	78
4 – Variável dependente: Cada uma das Complicações-pos-op30diasAlta	79
7 - ANÁLISE DE CLUSTERS	82
7.1 - Para o tempo de Internamento.....	82
7.1.1 - Análise Discriminante	82
7.2 - Para o tempo de Diálise.....	84
7.2.1 - Análise Discriminante	85
PARTE III.....	87
CONCLUSÕES	87
BIBLIOGRAFIA	93

ÍNDICE DE TABELAS

Tabela 1: Classificação do poder discriminativo do modelo de regressão	37
Tabela 2: Análise Descritiva dos dados numéricos	46
Tabela 3: Média de idades em função da existência ou não de comorbilidades	52
Tabela 4: Teste de Normalidade	52
Tabela 5: Comparação de Ranking.....	53
Tabela 6: Estatística do Teste U de Mann-Whitney.....	53
Tabela 7: Relação entre as Reintervenções e as Complicações Cardiovasculares	54
Tabela 8: Valor da estimativa da razão de possibilidades e respetivo intervalo de confiança (95%)	54
Tabela 9: Relação entre as Complicações pós operatórias e HTA	54
Tabela 10: Valor da estimativa da razão de possibilidades e respetivo intervalo de confiança (95%)	54
Tabela 11: Relação entre as Reinternamento 3º mês e as Complicações Respiratórias.....	55
Tabela 12: Valor da estimativa da razão de possibilidades e respetivo intervalo de confiança (95%)	55
Tabela 13: Análise descritiva das variáveis usadas na Regressão Linear Múltipla.....	56
Tabela 14: Sumário do modelo	57
Tabela 15: ANOVA do modelo inicial	57
Tabela 16: Significância das variáveis no modelo – Output do SPSS	57
Tabela 17: Processo de seleção das variáveis preditoras - Output do SPSS	58
Tabela 18: Sumário do modelo com as variáveis preditoras ASA e Tempo em diálise - Output do SPSS.....	58
Tabela 19:ANOVA do modelo com as variáveis preditoras - Output do SPSS	59
Tabela 20: Tabela de coeficientes - Output do SPSS.....	59
Tabela 21: Processo de seleção das variáveis preditoras - Output do SPSS	61
Tabela 22:Sumário do modelo com as variáveis preditoras Tempo em diálise, DAP e DCI - Output do SPSS.....	61
Tabela 23: ANOVA do modelo com as variáveis preditoras - Output do SPSS	62
Tabela 24: Tabela de coeficientes - Output do SPSS.....	62
Tabela 25: Processo de seleção das variáveis preditoras - Output do SPSS	64
Tabela 26: Sumário do modelo com as variáveis preditoras Urorenal e Infeciosas - Output do SPSS	64
Tabela 27:Anova do modelo com as variáveis preditoras - Output do SPSS	64
Tabela 28: Tabela dos coeficientes das variáveis preditoras - Output do SPSS.....	65
Tabela 29: Tabela de observações candidatas a outlier – Output do SPSS	68
Tabela 30: Omnibus Test – Output do SPSS.....	69
Tabela 31: Sumário do modelo – Output do SPSS	69
Tabela 32: Hosmer Test – Output do SPSS.....	69
Tabela 33: Tabela de classificação do modelo – Output do SPSS	69
Tabela 34: resumo dos coeficientes do modelo e a sua significância – Output do SPSS.....	70
Tabela 35: Área da Curva Roc e respetivo Intervalo de Confiança – Output do SPSS	72
Tabela 36: Área da Curva Roc e respetivo Intervalo de Confiança – Output do SPSS	72
Tabela 37: Omnibus Test – Output do SPSS.....	74

Tabela 38: Sumário do modelo – Output do SPSS.....	74
Tabela 39: Hosmer Test – Output do SPSS.....	74
Tabela 40: Tabela de classificação do modelo – Output do SPSS.....	75
Tabela 41: resumo dos coeficientes do modelo e a sua significância.....	75
Tabela 42: Área da Curva Roc e respetivo Intervalo de Confiança – Output do SPSS.....	77
Tabela 43: Área da Curva Roc e respetivo Intervalo de Confiança – Output do SPSS.....	78
Tabela 44: Hosmer Test – Output do SPSS.....	78
Tabela 45: Resumo dos coeficientes do modelo – Output do SPSS.....	78
Tabela 46: Área da Curva Roc e respetivo Intervalo de Confiança – Output do SPSS.....	79
Tabela 47: Área da Curva Roc e respetivo Intervalo de Confiança – Output do SPSS.....	80
Tabela 48: Área da Curva Roc e respetivo Intervalo de Confiança – Output do SPSS.....	81
Tabela 49: Média dos tempos de internamento e nº de casos por cluster.....	82
Tabela 50: ANOVA de comparação das médias dos cluster.....	82
Tabela 51: Teste de igualdade de médias de grupo.....	83
Tabela 52: Funções discriminantes.....	83
Tabela 53: Teste de significância das funções discriminantes.....	83
Tabela 54: Matriz de estrutura das funções.....	84
Tabela 55: Média dos tempos de internamento e nº de casos por cluster.....	84
Tabela 56: ANOVA de comparação das médias dos cluster.....	85
Tabela 57: Teste de igualdade de médias de grupo.....	85
Tabela 58: Função discriminante.....	85
Tabela 59: Teste de significância da função discriminante.....	85
Tabela 60: Matriz de estrutura das funções.....	86

ÍNDICE DE GRÁFICOS

Gráfico 1: Diagrama de extremos e quartis das idades em função da existência ou não de comorbilidades.....	51
Gráfico 2: Normalidade da idade dos pacientes sem e com comorbilidades.....	52
Gráfico 3: Importância dos preditores encontrados - Output do SPSS.....	60
Gráfico 4: Importância dos preditores - Output do SPSS.....	63
Gráfico 5: Importância dos preditores - Output do SPSS.....	66
Gráfico 6: Análise gráfica dos resíduos – Output do SPSS	67
Gráfico 7: Diagnóstico de observações influentes – Output do SPSS.....	68
Gráfico 8: Gráfico de probabilidades – Output do SPSS	70
Gráfico 9: Curva Roc do Modelo 1.1 – Output do SPSS	71
Gráfico 10: Curva Roc do Modelo 1.2 – Output do SPSS	72
Gráfico 11: Análise gráfica dos resíduos – Output do SPSS	73
Gráfico 12: Diagnóstico de observações influentes – Output do SPSS.....	73
Gráfico 13: Gráfico de probabilidades – Output do SPSS	76
Gráfico 14: Curva Roc do Modelo 2.1. – Output do SPSS	76
Gráfico 15: Curva Roc do Modelo 2.2. – Output do SPSS	77
Gráfico 16: Curva Roc do Modelo	79
Gráfico 17: Curva Roc do Modelo 4.1. – Output do SPSS	80
Gráfico 18: Curva Roc do Modelo 4.3. – Output do SPSS	81
Gráfico 19: Função discriminante 1 para o cluster 2	86
Gráfico 20: Função discriminante 1 para o cluster 1	86

ÍNDICE DE FIGURAS

Figura 1: Representação dos Processos (http://www.portalection.com.br/analise-de-regressao)	13
Figura 2: Ajuste das observações à linha de regressão (REGRESSÃO MÚLTIPLA: uma digressão sobre seus usos - http://www.ibci.com.br/Regressao_Multipla.pdf).....	17
Figura 3: Variância constante dos resíduos (REGRESSÃO MÚLTIPLA: uma digressão sobre seus usos - http://www.ibci.com.br/Regressao_Multipla.pdf)	21
Figura 4: Gráfico da Curva ROC, para uma dada capacidade de discriminação, com a variação do critério de decisão. Fonte: Braga(2000)	36
Figura 5: Acurácia do modelo - Output do SPSS.....	60
Figura 6: Acurácia do modelo - Output do SPSS.....	63
Figura 7: Acurácia do modelo - Output do SPSS.....	66

Lista de abreviaturas

ASA: American Society of Anesthesiologists
AVC: Acidente Vascular Cerebral
CHP: Centro Hospitalar do Porto
CI: Intervalo de Confiança
CNS: Sistema Nervoso Central
DCI: Doença Cardíaca Isquêmica
DAP: Doença Arterial Periférica
DM: Diabetes Mellitus
DRC: Doença Renal Crónica
HTA: Hipertensão Arterial
HTTP: Hipertiroidismo
PCA: Analgesia Controlada pelo Paciente
SPSS: Statistical Package for the Social Sciences
M.M.V: Método de Máxima Verosimilhança
OR: Odds Ratio
r: Coeficiente de correlação de Pearson
ROC: Receiver Operating Characteristic

INTRODUÇÃO

O número de transplantes renais, em doentes com doença renal crónica em estado avançado tem aumentado juntamente com a maturação das técnicas de transplante. Também os tempos de sobrevivência dos transplantados têm aumentado, graças à melhoria da taxa de sobrevivência durante o período perioperatório e à melhoria nas terapias imunossupressoras.

O transplante renal é assim o tratamento de eleição para pacientes em estadio final da doença renal crónica.

As características demográficas e clínicas dos pacientes transplantados, bem como características cirúrgicas, pós-cirúrgicas e do dador são potenciais fatores de risco para complicações a curto e longo prazo. Os três estudos já realizados sobre este tema e intitulados "Living-donor Kidney Transplantation: Predictive Factors and Impact on Post-transplant Outcome", "Deceased-donor Kidney Transplantation: Predictive Factors and Impact on Postoperative Outcome" e "Living-donor and Deceased-donor Renal Transplantation: Differences in Early Outcome-A Single-center Experience", da autoria respetivamente de Araujo AM, Santos F, Guimaraes J, Nunes CS, Casal M (2015), de Santos F, Guimaraes J, Araujo AM, Nunes CS, Casal M (2015) e de Guimaraes J, Araujo AM, Santos F, Nunes CS, Casal M (2015), focaram os seguintes objetivos, respetivamente:

- Analisar características demográficas e clínicas de dadores e recetores de rins, para prever resultados pós-transplante (complicações pós-transplante, tempo de hospitalização, reintervenções e readmissões hospitalares), depois de um transplante de dador vivo. Concluiu que a doação de rins vivos parece ser uma aceitável e justificável opção para pacientes em estágio final de doença renal crónica e que uma cuidada seleção do dador e do recetor surge como um pré-requisito para o sucesso do transplante.
- Avaliar os fatores de risco para complicações pós-transplante em recetores de rins de dadores cadáver. Permitiu concluir que o sucesso do transplante renal depende de fatores independentes do paciente, que incluem a capacidade técnica da equipa cirúrgica, a organização e otimização da equipa que realiza a gestão da colheita do enxerto e os meios técnicos disponíveis. Quando estas variáveis são controladas a necessidade de perceber os fatores dependentes do paciente submetido ao transplante torna-se crucial. Em suma, mostrou o impacto de alguns fatores de risco

clínico comuns à maioria dos pacientes em estágio final de doença renal crónica e que uma boa gestão destes fatores de risco contribuem para a melhoria a curto prazo no sucesso do transplante.

- Identificar diferenças nas complicações imediatamente a seguir ao pós-operatório, as reintervenções e readmissões hospitalares, entre recetores de dadores vivos e recetores de dadores cadáver concluindo que as variáveis pré-transplante do recetor, com diferenças significativas entre pacientes recetores de dadores cadáver e de dadores vivos foram identificadas como predictoras de resultados imediatos.

O objetivo deste estudo é avaliar os fatores preditores de complicações após transplantes renais, quer sejam recetores de dadores vivos quer sejam recetores de dadores cadáver.

Os dados foram recolhidos de 153 pacientes que receberam transplante de rins de dadores vivos e dadores cadáver, dos quais foram registadas 58 variáveis agrupadas nas seguintes categorias: **caraterísticas do recetor:** idade; sexo; ASA; Tempo em diálise; técnica dialítica; técnica anestésica; Etiologia DRC; Comorbilidades; Tipos de Comorbilidades: HTA, DCI, DAP, Tabagismo, DM, Hiperuricemia, Dislipidemia, AVC, HPT, Anemia; **caraterísticas cirurgicas:** Tempo de cirurgia; Tempo de isquemia; Enxerto safena; **caraterísticas pós-cirurgicas:** Analgesia pós-operatória; PCA, N°_pedidos_1^{as}24horas; N°_dias; BNE, N°_dias_Após; Escala_dor_R; Escala_dor_M; Tempo_internamento; Complicações_pós_oper_30_dias_ALTA; Tipos de complicações: Cardiovasculares, Respiratórias, CNS, Urorenal, Infeciosas, Hematológicas e Endócrinas; Reintervenções; Reinternamento_3^o_mês e consulta follow_up; **caraterísticas do dador no caso de dadores vivos:** Sexo_dador; Idade_dador; ASA_dador; Tipo de relação; Relação do 1^o grau; Técnica anestésica_dador; T_isquemia quente_rim_dador e diferença_idade_maior_10_anos. De acordo com as tabelas de correlações obtidas no SPSS e constantes do ANEXO 1, os potenciais fatores de risco considerados neste estudo são os seguintes:

- Fatores de risco no Tempo_Internamento: complicações cardiovasculares, Urorenais, Infeciosas, Hematológicas e Endócrinas, DAP, Tempo em diálise, Tempo de isquemia e ASA_dador.
- Fatores de risco na reintervenções: complicações cardiovasculares.
- Fatores de risco nas complicações_pós_oper_30_dias_ALTA: HTA, ASA_dador.

- Fatores de risco nas complicações Urorenais: Tempo em diálise e Tempo de isquemia

O presente estudo foi aprovado por Quality Management, Risk, Health and Safety do Centro Hospitalar do Porto (CHP) e foi conduzida em concordância com a declaração de Helsínquia de 1975 (tal com revista em 1983). Os dados foram obtidos de registos médicos eletrónicos dos pacientes e foram tratados de acordo com a lei portuguesa de confidencialidade e proteção de dados clínicos.

Os dados contêm todos os procedimentos efetuados a dadores vivos e recetores de rins transplantados de dadores vivos e dadores cadáver efetuados entre janeiro de 2012 e dezembro de 2013.

Foi realizada análise estatística com recurso SPSS. Modelos de regressão linear multivariada e regressão logística foram utilizados para identificar fatores de risco independentes para complicações pós operatórias, com a obtenção de p-values, Odds Ratio (OR) e intervalos de confiança (CI) a 95%. Análise de clusters e análise discriminante foram utilizadas com o objetivo de estabelecer uma diferenciação de grupos por tempos de internamento e tempo em diálise minimizando as probabilidades de erro na previsão de tempos de internamento, na presença de determinados fatores.

É oportuno salientar neste ponto que, a pesquisa bibliográfica efetuada aos recursos identificados na secção da Bibliografia com os números [14] ao [24], foram recursos pesquisados mas que não se revelaram de interesse para a realização deste trabalho.

Este trabalho está estruturado em 3 partes:

Parte I - Enquadramento teórico, com uma breve apresentação do software utilizado, a fundamentação da regressão linear multivariada e logística, finalizando com a parte da análise dos resíduos com recurso à curva ROC e com a análise de clusters e análise discriminante.

Parte II – Material e Resultados incluindo a análise exploratória dos dados, apresentação de resultados obtidos após análise de dados com recurso ao SPSS.

Parte III - Conclusões, onde são apresentadas as conclusões obtidas das análises realizadas e algumas considerações sobre o trabalho realizado.

PARTE I

ENQUADRAMENTO TEÓRICO

1 - ENQUADRAMENTO E MOTIVAÇÃO

Este estudo, tal como já referido incide sobre o transplante renal. De acordo com a Sociedade Portuguesa de Urologia, no seu artigo de informação sobre o transplante renal, publicado em janeiro de 2010, “O transplante renal é um método de substituição da função renal, estando indicada a sua realização quando existe insuficiência renal, ou seja quando os rins deixam de realizar de uma forma adequada as suas normais funções. As doenças que causam insuficiência renal são variadas, sendo as mais comuns a diabetes, a hipertensão arterial, malformações, infeções e inflamações dos rins. Os doentes com um só rim mas que funcione bem não necessitam de tratamentos para substituir a função renal. As opções de substituição da função renal são o transplante renal e a hemodiálise e suas variantes (como por exemplo a diálise peritoneal). O transplante de um rim permite que este desempenhe de uma forma satisfatória a função antes realizada pelos dois rins do doente. As grandes vantagens que o transplante renal apresenta em relação às técnicas dialíticas resultam de o rim transplantado ser um órgão vivo que possui todas as funções que lhe são próprias. Para além de, por meios naturais, regular o metabolismo de diversas substâncias (água, sódio, potássio, etc.) e de eliminar outras que são tóxicas, também produz hormonas que regulam funções muito importantes, designadamente, entre outras, as que promovem a formação dos glóbulos vermelhos do sangue e as que regulam regeneração dos ossos. Por isso, e por permitir que o doente se liberte de algumas das restrições impostas pela diálise, o transplante renal é a modalidade terapêutica que faculta uma melhor qualidade de vida e uma maior esperança de vida.

Só os doentes com insuficiência renal terminal necessitam ser transplantados. A proposta para avaliação da possibilidade de transplante é geralmente feita pelo médico assistente (nefrologista) do doente e o estudo realizado em centros hospitalares especializados, onde é realizado um estudo aprofundado dos candidatos.

As contra-indicações ao transplante podem ser de diversa ordem: infeções, doenças que aumentem muito o risco da cirurgia, antecedentes de cancro há pouco tempo, patologia cardiovascular, psiquiátrica ou pulmonar severa. A idade por si só, não é hoje em dia, um critério de exclusão absoluto, embora acima dos 70 anos os riscos do transplante aumentem bastante.

Os órgãos a ser transplantados - designados por enxertos - têm origem em dadores, que podem ser cadáveres (indivíduos em morte cerebral - que apesar de apresentarem uma

lesão irreversível das funções cerebrais apresentam os órgãos em bom estado) ou dadores vivos (pessoas saudáveis que voluntariamente abdicam de um rim para ser implantado no doente insuficiente renal).

Para ser colocado em lista de espera para transplante de dador cadáver é necessário que o doente esteja em programa regular de hemodiálise. Caso exista um dador vivo disponível, o doente pode eventualmente ser transplantado antes de entrar em diálise.

Existem critérios para a seleção do par dador-recetor que estão perfeitamente definidos e legislados. Fazem parte destes critérios (com ponderações variadas): tempo em diálise, idade (crianças e jovens têm prioridade), critérios de histocompatibilidade (semelhança entre os tecidos do dador e recetor), diferenças de idade entre dador e recetor. Os critérios de histocompatibilidade são fundamentais com vista a minimizar a possibilidade de rejeição e têm um peso decisivo na seleção.

Quando existe um dador são analisados todos os recetores possíveis e é selecionado aquele que reúna o maior número de critérios. Existem situações de exceção em que recetores são colocados em posição de urgência – por exemplo quando o transplante seja o único tratamento adequado e exista risco de vida.

Existem atualmente no nosso país 8 centros com atividade regular de transplantação renal:

- Hospitais da Universidade de Coimbra
- Hospital Geral Santo António - Porto
- Hospital de São João - Porto
- Hospital Santa Maria - Lisboa
- Hospital Curry Cabral - Lisboa
- Hospital Santa Cruz - Lisboa
- Hospital Cruz Vermelha - Lisboa
- Hospital Garcia da Orta - Almada

Só alguns destes centros realizam transplantes de dador vivo e transplante renal em idades pediátricas.

De acordo com um trabalho publicado pela Sociedade Portuguesa de Urologia (janeiro 2010), no Serviço de Urologia e Transplantação Renal dos Hospitais da Universidade de Coimbra - com cerca de 2000 transplantes realizados nos últimos 30 anos, realizaram-se durante o ano de 2009 - 177 transplantes renais, sendo um dos centros com maior atividade a nível europeu. A nossa legislação, Despacho nº 6537/2007, permite que cada doente

possa estar inscrito em duas unidades de transplantação. Dados fornecidos pela Sociedade Portuguesa de Transplantação de Órgãos e até finais de 2008 mostram que tinham sido realizados até aquela data 8136 transplantes renais em Portugal. No ano de 2008 foram realizados 525 transplantes (sendo 49 de dador vivo - 9,3%). O número de doentes que entram anualmente em diálise é significativamente superior ao número de doentes que são transplantados. Nem todos os doentes em diálise têm indicação para ser transplantados mas mesmo assim existe uma escassez de órgãos para as necessidades, havendo uma tendência para o aumento do número de doentes em lista de espera. O tempo médio de permanência em diálise dos doentes transplantados em 2008 foi de 6,65 anos. A percentagem de dadores vivos é, entre nós, muito reduzida (menos de 10%) ao contrário do que acontece noutros países onde as taxas se podem aproximar ou ultrapassar os 50%. As taxas de dadores cadáver são, no entanto, elevadas (26,7/milhão de habitantes) e situam-nos numa posição de relevo a nível mundial (terceira posição). O número de transplantes renais em 2008 colocou-nos em sexto lugar a nível mundial (48 / milhão de habitantes)” (http://www.apurologia.pt/publico/transplante_renal.htm)[29]. Todo este conjunto de informações vem dar ênfase aos estudos até agora realizados sobre este tema.

2 - TÉCNICAS DE INFERÊNCIA ESTATÍSTICA NÃO PARAMÉTRICA

As técnicas de Inferência Estatística Não Paramétrica surgiram como um processo para resolver problemas de difícil resolução no campo da Estatística Paramétrica.

Os métodos paramétricos, univariados e multivariados, conduzem na maior parte das vezes à obrigatoriedade de considerar pressupostos que nem sempre refletem o modelo estudado e que, quando não são verificados, implicam erros graves na análise e conclusões.

Segundo Cristina Ferreira, na sua tese de Mestrado (2013), “embora em 1710 se encontrem referências à utilização de métodos de estatística não paramétrica, estes surgem só bastante mais tarde, em 1942, com Wolfowitz, e só então se começa a verificar o seu desenvolvimento e um maior impacto, sendo hoje a estatística não paramétrica considerada como um ramo extremamente importante da Estatística.” (<http://www-history.mcs.st-andrews.ac.uk/history/Biographies/Wolfowitz.html>).

A estatística não paramétrica constitui um conjunto de ferramentas mais adequado em pesquisas onde não se conhece bem a distribuição da população e os seus parâmetros. Na

eventualidade de um desconhecimento da população, o uso dos testes não paramétricos são uma ferramenta de grande importância na análise de pesquisas.

A origem da designação destes testes como “Métodos ou Testes não Paramétricos” reside no facto de estes não terem a pretensão de testar ou estimar parâmetros de uma dada distribuição, mas sim estudar o ajustamento de certas funções aos dados, averiguar a independência ou comparar duas ou mais distribuições (OLIVEIRA, T., 2004).

Os testes não paramétricos são também conhecidos por “distribution-free tests”, MAROCO (2007), ou seja, testes adistribucionais, pois não necessitam do conhecimento prévio da distribuição da variável em estudo (distribuição normal), sendo por isso uma boa alternativa aos testes paramétricos quando falha a normalidade e a homogeneidade das variâncias entre os grupos, relativamente à variável em estudo.

As principais diferenças entre estes dois tipos de testes podem ser sintetizadas da seguinte forma:

Testes paramétricos:

- Exigem que as amostras sejam provenientes de uma distribuição normal, sendo especialmente importante quando as amostras são pequenas, ou seja, quando a dimensão é inferior a 30, tamanho mais usual na literatura.
- Nas amostras de dimensão superior a 30, a distribuição da média amostral aproxima-se bem da distribuição normal viabilizando a aplicação dos testes paramétricos.

Testes não paramétricos:

- É prescindível a assunção de pressupostos muito estritos sobre a distribuição da população donde foram extraídos os dados (como p. ex. a normalidade). São também indicados quando as amostras são pequenas.
- São alternativas para quando a amostra não se ajusta à distribuição normal ou quando, apesar da amostra ter uma dimensão superior a 30, se opta por conclusões mais conservadoras.
- São passíveis de serem usados em dados que não são obrigatoriamente numéricos, em escala intervalar ou de razão, por exemplo sobre dados ordinais.
- A desvantagem destes testes é que, verificando-se os pressupostos dos testes paramétricos, estes são menos potentes, isto é, com os testes não paramétricos não se encontram tantas diferenças entre os dados, quando essas diferenças realmente existem.

Neste estudo, os testes não paramétrico a utilizar são o teste do χ^2 e o teste U de Mann Whitney.

O teste do qui-quadrado (χ^2) é utilizado para comprovar se existem diferenças estatisticamente significativas entre duas distribuições medindo o grau de discrepância entre um conjunto de frequências observadas e um conjunto de frequências esperadas e o seu valor será zero quando não houver diferença entre os valores observados e os esperados.

Este teste quantifica a associação entre variáveis qualitativas, comparando as diferenças entre os valores observados e os esperados. Quanto maior for o valor da estatística χ^2 , maior será o grau de associação existente entre as duas variáveis.

No teste do χ^2 a estatística de teste é dada por,

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad (2.1)$$

pois sendo desconhecida a distribuição da amostra e para testar a hipótese do ajustamento do modelo aos dados, baseia-se na estimativa das probabilidades de sucesso de cada uma das n observações, sendo a estatística de teste a utilizar, obtida com um teste do χ^2 a uma tabela de contingência $2 \times g$. Também será utilizado este teste para testar a hipótese nula de que não existe relação entre duas variáveis da amostra em estudo. Neste caso o teste χ^2 de Independência verifica a existência de associação estatisticamente significativa entre duas variáveis categóricas, medidas em escala nominal ou ordinal. Não havendo associação, dizemos que as variáveis são independentes.

O teste U de Mann Whitney é apropriado para averiguar se são iguais as medianas μ_X e μ_Y de duas populações contínuas e independentes, X e Y, respetivamente. As duas amostras envolvidas não têm que ter a mesma dimensão. As hipóteses a testar são:

$$H_0: \mu_X = \mu_Y$$

$$H_1: \mu_X \neq \mu_Y$$

De acordo com o seguinte procedimento: juntam-se as duas amostras numa só (amostra combinada) identificando a origem (X ou Y) de cada elemento desta nova amostra. Ordena-se a amostra combinada e observam-se as posições (ranks) ocupadas pelos elementos X e pelos elementos Y. Se X e Y ficarem aleatoriamente distribuídos, então a decisão final será favorável a H_0 (pois, sob H_0 , a mediana da população subjacente à

amostra combinada será igual às medianas μ_X e μ_Y). Caso contrário, rejeita-se H_0 . O teste é desenvolvido com base na soma das posições (Sum of Ranks) ocupadas, na amostra ordenada, pelos elementos da amostra inicial de menor dimensão. No caso de haver uma ou mais observações iguais nas duas amostras iniciais (empates=Ties) é feita uma “correção” usando as posições médias das observações empatadas.

3 - SOFTWARE UTILIZADO

A análise estatística dos dados realizou-se com recurso ao software SPSS (Statistical Package for the Social Sciences) que teve origem em 1968, na Chicago University, tendo sido divulgado e com utilização global desde então. Este software foi desde 1994, representado em Portugal pela PSE- Produtos e Serviços de Estatística, Lda (www.pse.pt). O software SPSS é um sistema que permite organizar dados e executar análises estatísticas. Tem um ambiente gráfico muito apelativo com o qual, para a maioria das análises a efetuar, basta a seleção das respetivas opções em menus e caixas de diálogos (Laureano e Botelho, 2010).

A utilização do SPSS apresenta várias vantagens:

- Flexibilidade para variáveis de diferentes naturezas;
- Facilidade de utilização, sendo um programa de fácil manuseamento, com diversos níveis de complexidade, de acordo com as necessidades dos seus utilizadores;
- Participação em todo o processo analítico, desde o planeamento até à recolha de dados para a análise, possibilitando a elaboração de relatórios, quer pelo próprio programa, quer por uma articulação com um processador de texto (Laureano e Botelho, 2010).

4 - ANÁLISE DE REGRESSÃO

Em diversos problemas das áreas médica, biológica, industrial, química entre outras, é de grande interesse verificar se duas ou mais variáveis estão relacionadas de alguma forma. Para expressar esta relação é muito importante estabelecer um modelo matemático. Este tipo de modelagem é chamado de regressão, e ajuda a entender como determinadas variáveis influenciam outra variável, ou seja, verifica como o comportamento de uma(s) variável(is) pode mudar o comportamento de outra.

Esta relação pode ser analisada como um processo. Neste processo, os valores de X são chamados de Variáveis de Entrada ou Regressoras (inputs) e Y de Variável de Saída ou Resposta (output).

Processo

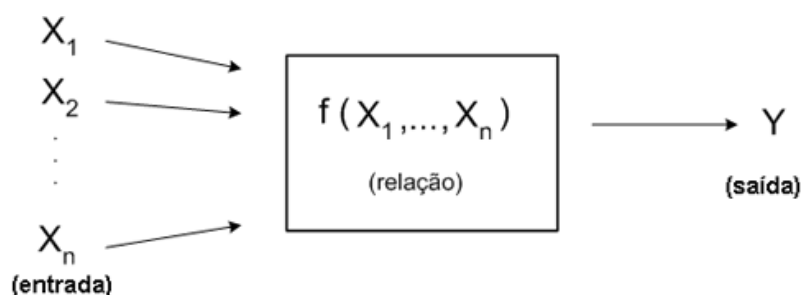


Figura 1: Representação dos Processos (<http://www.portalaction.com.br/analise-de-regressao>)

A Análise de Regressão possibilita encontrar uma relação razoável entre as variáveis de entrada e saída, por meio de relações empíricas. A recolha de dados permite conhecer a natureza da relação entre as variáveis e realizar estudos capazes de acomodar situações inesperadas, como por exemplo, variabilidade na matéria prima, condições ambientais, máquina e operadores, entre outros.

Para tal, muito contribuiu o desenvolvimento e rápido crescimento da tecnologia informática, onde *Hardware* significa mais velocidade e mais capacidade e *Software* significa mais possibilidades.

Assim, desenvolveu-se uma panóplia de técnicas de regressão (e.g., regressões robustas, regressão envolvendo respostas como as séries temporais, curvas de crescimento, regressão nas quais as variáveis preditoras e resposta são curvas, imagens, gráficos, regressão com dados omissos, regressão não-paramétrica, métodos Bayesianos para regressão, regressão onde as variáveis preditoras são medidas com erro, regressão com mais variáveis

preditoras do que observações, inferência com regressão). Porém, de todas as formas, na atualidade a técnica estatística da análise de regressão linear sobressai como uma das mais utilizadas e importantes na investigação e modelação da relação estatística entre duas ou mais variáveis submetidas a um processo. Na base desta manifesta importância está a facilidade com que uma equação é capaz de expressar a relação suscetível de existir entre as variáveis de interesse. Quando a equação de regressão for utilizada com a intenção de controlar, assume especial importância que a relação entre a variável de interesse e a(s) variável(eis) usadas para o controlo sejam do tipo causa-efeito (Landim, 2010). Não obstante, se a equação for usada apenas para predição, é dispensada a relação causa-efeito, ou seja, neste caso a causalidade não deve ser assumida, independentemente da bondade do ajuste (*goodness of fit*) estatístico do modelo (Neter et al., 1983, Freedman et al., 2004). Nesta situação, apenas importará que as relações estabelecidas, originalmente usadas na construção da equação, ainda sejam válidas (Freedman et al., 2004). A explicação oferecida pelo modelo da relação entre Y e X é parcial, dada a possibilidade de poderem existir outros fatores suscetíveis de influenciar Y , para além de afetarem X , bem como a possibilidade de parte da variabilidade observada poder estar associada com a variabilidade imputada a Y . Ou seja, segundo Neter et al. (1983), a análise de regressão processa as informações contidas num conjunto de dados de tal modo que seja gerado um modelo que represente o relacionamento existente entre as variáveis de interesse. Segundo Pestana & Gageiro (2005b), os mais importantes objetivos da análise de regressão situam-se ao nível da descrição (da relação entre variáveis), controlo (e.g., controlo estatístico da qualidade), predição (do valor da variável dependente) e estimação.

Os modelos de regressão podem ser classificados segundo o número de variáveis independentes. Na presença de apenas uma variável independente denomina-se “Modelo de Regressão Simples” e quando existe mais de uma variável independente denomina-se “Modelo de Regressão Múltipla”. Também é possível classificar o modelo de acordo com o tipo de função que o define, podendo ser representada por diferentes tipos de equações (linear, polinomial, etc.).

Existem vários métodos para construir uma equação de regressão, sendo o método dos mínimos quadrados o mais utilizado. Este método, tem algumas propriedades estatísticas que fizeram dele um dos mais poderosos e populares métodos na construção de uma equação de regressão e foi atribuído ao matemático alemão Carl Friedrich Gauss. A reta de regressão que se obtém por método passa pela média da amostra dos valores das variáveis

dependente e independente (X e Y), mostrando se existe alguma relação entre elas, assim como se são positivamente ou negativamente correlacionadas.

A Regressão Múltipla é um dos inúmeros modelos estatísticos explanatórios causais referentes ao tratamento de séries temporais de dados. A sua base estatística advém da Regressão Linear, que se restringe a duas variáveis e a apenas uma equação funcional do primeiro grau

$$Y = a + bX \quad (4.1)$$

de ajustamento.

A análise de Regressão Múltipla é pois uma metodologia estatística de previsão de valores e explicação do comportamento de uma ou mais variáveis de *resposta* (Dependentes) através de um conjunto de variáveis *explicativas* (Independentes) controlando o efeito de outras variáveis. Esta metodologia pode ser utilizada também para a avaliação dos efeitos das variáveis explicativas como preditoras das variáveis de resposta; isto é, serve para contribuir na obtenção de respostas a perguntas do tipo “Qual é o melhor estimador para ... ?”.

O formato geral da equação de Regressão Linear Múltipla é :

$$Y = a + b_1X_1 + b_2X_2 + \dots b_kX_k \quad (4.2)$$

- Y é a Variável Dependente;
- a corresponde a um coeficiente técnico fixo, (vulgar constante) um valor de base a partir do qual começa Y ;
- b_k corresponde aos coeficientes técnicos acoplados às Variáveis Independentes;
- X_k as Variáveis Independentes

A sua aplicação é especialmente importante pois permite que se estime o valor de uma variável com base num conjunto de outras variáveis. Quanto mais significativo for o peso de uma variável isolada, ou de um conjunto de variáveis explicativas, tanto mais se poderá afirmar que alguns fatores afetam mais o comportamento de uma variável de resposta especificamente procurada, do que outros.

O que distingue o modelo de regressão logística do da regressão linear é que a variável resposta (variável dependente) na regressão logística é usualmente binária (dicotómica). Esta diferença entre regressão logística e linear é refletida quer na escolha de um modelo

paramétrico, quer nas hipóteses a serem consideradas. Desde que esta diferença seja tida em conta, os métodos empregues na análise usando a regressão logística seguem os mesmos princípios usados na regressão linear. Então, as técnicas usadas na análise de regressão linear deverão motivar uma aproximação à regressão logística, que será abordada na secção seguinte.

A regressão logística permite o uso de um método de regressão para calcular ou prever a probabilidade de um evento específico. Desta forma, esta usa-se quando se tem uma variável dependente em escala nominal e uma variável independente nominal e/ou contínua, e serve para descrever a relação entre a variável dependente nominal e o conjunto de variáveis independentes através da função *logit* (Braga, 1994).

Quando a regressão logística é usada deve-se primeiramente achar o modelo que melhor se ajuste aos dados em análise, com o intuito de se obter um modelo moderado e biologicamente razoável, que permita descrever a relação entre a variável dependente e um conjunto de variáveis independentes (Braga, 1994).

4.1. Correlação Múltipla e Coeficiente de Determinação

Na Regressão Linear é bastante comum que a Variável Dependente (Y) se relacione com a Variável Independente (X), mas é incorreto afirmar que o valor da primeira depende *em causa e efeito* das alterações no valor da segunda. Neste caso a inter-relação entre as variáveis é demonstrada através da correlação. O coeficiente de correlação, r (4.3), é a medida de inter-relação entre a Variável Dependente e a Variável Independente. Ele pode variar de 0 (que indica ausência de correlação) a ± 1 (que indica correlação perfeita). Quando o coeficiente de correlação é maior que 0, as duas variáveis são positivamente correlacionadas, em contrapartida quando é menor que 0, as duas variáveis são negativamente correlacionadas. O sinal do coeficiente de correlação numa Regressão Linear é sempre o mesmo sinal do coeficiente de regressão, b .

O coeficiente de correlação, r , é calculado a partir da seguinte fórmula (onde n é o Número de Observações (pontos) necessários para o ajuste da linha de regressão):

$$r = \frac{n \sum XY - \sum X \sum Y}{\sqrt{(n \sum X^2 - (\sum X)^2)(n \sum Y^2 - (\sum Y)^2)}} \quad (4.3)$$

Dessa expressão, deduz-se o coeficiente de determinação r^2 (4.3), que nada mais é que o quociente da variação explicada sobre a variação total, ou seja:

$$r^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{(n \sum XY - \sum X \sum Y)^2}{[n \sum X^2 - (\sum X)^2][n \sum Y^2]} \quad (4.4)$$

Onde Y_i são os valores observados da Variável Dependente, \hat{Y}_i são os valores estimados da Variável Dependente e \bar{Y} a média das observações.

Assim, r^2 é a medida de quão bem as observações se ajustam ao longo da linha de regressão. Para ajuste dos pontos Y_i , teríamos graficamente :

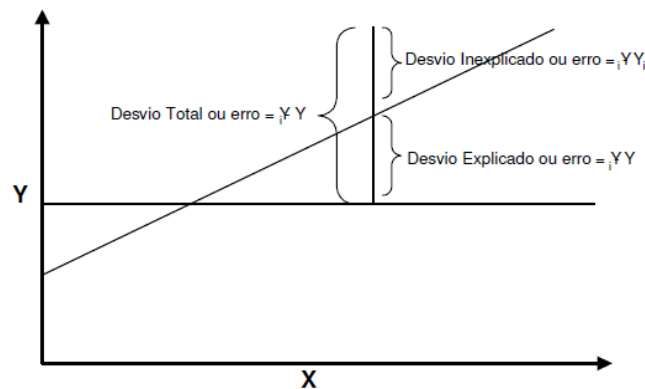


Figura 2: Ajuste das observações à linha de regressão (REGRESSÃO MÚLTIPLA: uma digressão sobre seus usos - http://www.ibci.com.br/Regressao_Multipla.pdf)

No caso de uma Regressão Múltipla o coeficiente de determinação deve calcular também o quociente entre a variação explicada e a variação total, para todas as variáveis independentes.

A Matriz de Correlação possui grande significado informativo para a Regressão Múltipla porque estabelece como os pares de variáveis Dependentes (Y) e Independentes (X_1, X_2, \dots, X_k) se correlacionam. Esta informação é utilizada na seleção das variáveis que deverão fazer parte da equação de regressão - ou seja, variáveis com correlação elevada e positiva deverão ser incluídas no modelo proposto, enquanto que as variáveis na condição inversa deverão ser descartadas.

A Correlação Múltipla e o Coeficiente de Determinação (r^2) também indicam como a relação expressa através da equação de regressão explica as variações da Variável Dependente (Y).

4.2. Testes de Significância

A significância estatística dos resultados obtidos na Análise de Regressão deve ser estabelecida antes do uso desses resultados numa previsão. A determinação dos coeficientes técnicos (b_1, b_2, \dots, b_k) é baseada simplesmente nas observações históricas. O propósito dos testes de significância estatística é determinar a confiança que pode ser depositada nos resultados da regressão e a sua aplicabilidade na população de valores possíveis. A existência de uma relação significativa entre a variável dependente e as variáveis independentes ou explicativas pode ser avaliada pelo seguinte teste de hipóteses:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0 \text{ (a relação entre as variáveis é não linear)}$$

$$H_1: \exists \beta_i \neq 0$$

Este teste é conhecido como teste do F total. A estatística do teste é a relação entre a variância decorrente da regressão linear múltipla e a variância dos resíduos:

$$F = \frac{QM\ Reg}{QM\ Res} \quad (4.5)$$

QM Reg -> Quadrados médios da Regressão, obtidos da ANOVA

QM Res -> Quadrados médios dos Resíduos, obtidos da ANOVA

A hipótese nula será aceite se,

$$F < F(\alpha, p, n - p - 1) \quad (4.6)$$

Onde

α é o nível de significância

p é o número de variáveis independentes

p e $n-p-1$ são os graus de liberdade da distribuição F de Snedecor

A contribuição de uma variável explicativa (independente) ao modelo de regressão múltipla, pode ser determinada pelo critério do teste do F parcial. Segundo este critério, avalia-se a contribuição de uma variável independente para a soma dos quadrados devido à regressão, após a inclusão no modelo das restantes variáveis independentes.

A verificação se a inclusão de uma variável X_k melhora significativamente o modelo de regressão é realizada por meio do seguinte teste de hipóteses:

H_0 : a variável X_k não melhora significativamente o modelo

H_1 : a variável X_k melhora significativamente o modelo

A estatística do teste é dada por:

$$F_p = \frac{SQReg(X_k)}{QMRes} \quad (4.7)$$

Onde

$SQReg(X_k)$ -> Soma dos Quadrados devidos à Regressão

$QMRes$ -> Soma dos Quadrados dos Resíduos

A hipótese nula deve ser rejeitada se a estatística F_p (4.7) for maior ou igual ao valor crítico da distribuição F de Snedecor, onde,

n é o tamanho da amostra

α é o nível de significância

p é o número de variáveis explicativas incluindo X_k

p e $n-p-1$ são os graus de liberdade da distribuição F de Snedecor

ou seja, rejeita-se H_0 se $F_p \geq F(\alpha, 1, n - p - 1)$.

4.3. Pressupostos da Análise de Regressão

Para obtenção dos resultados, a análise de regressão baseia-se em quatro pressupostos básicos:

4.3.1. Linearidade

Apesar de parecer um pressuposto restritivo, matematicamente toda função não-linear pode ser transformada numa função linear através de técnicas logarítmicas, polinomiais e de relações recíprocas. Não é objetivo deste texto discutir as formulações matemáticas de transformação, porém a sua existência é de fundamental importância uma vez que a análise de regressão não pode ser aplicada se a função não puder ser transformada para a forma linear (http://www.ibci.com.br/Regressao_Multipla.pdf).

4.3.2. Independência dos resíduos

A violação do pressuposto da independência dos resíduos implica na existência de forte correlação (autocorrelação) entre os residuais sucessivos. A falta de independência não afeta o valor dos parâmetros estimados, mas afeta diretamente as variâncias estimadas. A falta de independência dos resíduos implica em r^2 e estatística F elevados e, teste t reduzido se a autocorrelação é positiva e, todos os testes com resultados elevados se a autocorrelação for negativa.

A autocorrelação pode ser resultante de: especificação incorreta (a inclusão de um número não ótimo de variáveis), o que causa dependência entre as Variáveis Independentes; forma funcional incorreta (deficiência de modelagem); ou forte tendência entre as variáveis. A autocorrelação pode ser visualizada através do gráfico dos resíduos, no entanto, existe um teste estatístico, o de Durbin-Watson (teste D-W), (<http://www.portaaction.com.br/analise-de-regressao/33-diagnostico-de-independencia>) que pode ser utilizado para verificar a existência de autocorrelação nos resíduos de uma análise de regressão. Autocorrelação significa que observações adjacentes são correlacionadas. Quando essas observações são correlacionadas, a regressão de mínimos quadrados subestima o erro padrão dos coeficientes e os preditores podem parecer significantes quando na verdade não o são, isto é, diminui a precisão na estimação dos parâmetros e aumenta os erros de cálculo decorrentes de arredondamentos (Pestana & Gageiro, 2005b). Este teste é baseado na suposição de que os erros no modelo de regressão são gerados por um processo autoregressivo de primeira ordem, de acordo com

$$e_i = \rho e_{i-1} + a_i \quad (4.8)$$

Em que e_i é o termo do erro do modelo (resíduo) na i -ésima observação e $a_i \stackrel{iid}{\sim} N(0, \sigma_a^2)$ e $\rho (|\rho| < 1)$ é o parâmetro de autocorrelação. Testa-se as hipóteses de autocorreção por meio das hipóteses:

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

Sendo e_i o resíduo associado à i -ésima observação, temos que a estatística Durbin-Watson é definida como:

$$dw = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (4.9)$$

Onde o numerador é apenas a diferença ao quadrado entre resíduos em instantes sucessivos e o denominador é a soma do quadrado dos resíduos. O valor da estatística de Durbin-Watson é sempre um valor tal que $0 \leq dw \leq 4$. Podemos tomar a decisão comparando o valor de dw com os valores críticos dL e dU da Tabela de Durbin-Watson. Assim,

- se $0 \leq dw < dL$, existe evidência de autocorrelação positiva nos erros, então rejeitamos H_0 (há dependência);
- se $dL \leq dw \leq dU$ então o teste é inconclusivo;
- se $dU < dw < 4-dU$, então não rejeitamos H_0 (há independência);
- se $4-dU \leq dw \leq 4-dL$ então o teste é inconclusivo;
- se $4-dL < dw \leq 4$, existe evidência de autocorrelação negativa nos erros, então rejeitamos H_0 (dependência).

Quando $0 \leq dw < dL$ temos evidência de uma correlação positiva. Já quando $4-dL < dw \leq 4$, a correlação é negativa. No caso em que não rejeitamos H_0 , temos que não existe autocorrelação, ou seja, os resíduos são independentes. Portanto a autocorrelação entre os resíduos é indicada por valores de dw significativamente diferentes de 2. Podemos também tomar a decisão pelo p -valor.

4.3.3. Homocedasticidade

Se os resíduos não estão distribuídos ao longo da linha de regressão em torno de todo o intervalo de observações, o pressuposto da variância constante, ou homocedasticidade, é violado. O gráfico a seguir ilustra o significado da variância constante dos resíduos:

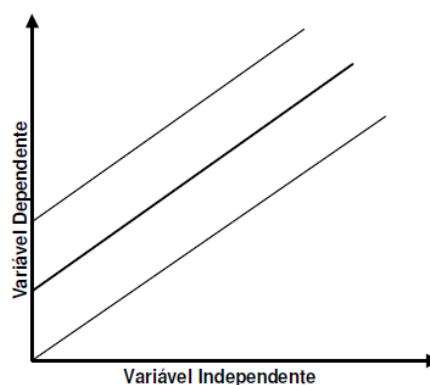


Figura 3: Variância constante dos resíduos (REGRESSÃO MÚLTIPLA: uma digressão sobre seus usos - http://www.ibci.com.br/Regressao_Multipla.pdf)

A ocorrência de variâncias não constantes nos resíduos é chamada de heterocedasticidade. A sua ocorrência pode estar condicionada a especificações incorretas no modelo de regressão, e a sua detecção é possível através do estudo residual dos erros.

O teste Durbin-Watson (4.9) pode indicar heterocedasticidade e a sua correção está vinculada à eliminação de algumas variáveis ou à transformação matemática do modelo, trazendo uniformidade dos erros percentuais ao longo da linha de regressão.

4.3.4. Normalidade dos resíduos

Esta hipótese também apresenta características pouco restritivas uma vez que os resíduos são resultantes de um sem número de fatores menos importantes no que tange a influência no comportamento da variável dependente (senão deveriam ser incluídos na equação de regressão, perdendo a sua característica residual). A normalidade dos resíduos é uma suposição essencial para que os resultados do ajuste do modelo de regressão linear sejam confiáveis. Podemos verificar essa suposição por meio do gráfico de Papel de Probabilidade e por meio de testes tais como Shapiro-Wilk, Anderson-Darling e Kolmogorov-Smirnov, (<http://www.portaaction.com.br/analise-de-regressao/31-diagnostico-de-normalidade>). Estatisticamente, se possuímos um número de observações superior a 30, a previsão de dados assume a “normalidade”. Isto porque a distribuição amostral dos estimadores pode ser aproximada à curva normal onde n possua amplitude suficiente, o que na maior parte ocorre quando n é igual a 30. O Teorema do Limite Central da estatística permite esta aproximação e torna possível o uso da curva normal na avaliação da dispersão dos dados, inclusive dos resíduos, da amostra em torno do parâmetro central (média). Assim ao calcularmos a sua média e variância, a extensão de possíveis erros pode ser avaliada, o que introduz um intervalo de confiança de 30 observações para a variância.

Quando o pressuposto da normalidade dos resíduos é questionado, não existem testes estatísticos específicos para sua avaliação. Todavia os resíduos podem ser representados graficamente com vista à detecção da sua distribuição próxima à normal e o seu intervalo de variação (o maior menos o menor valor) pode ser medido com vista à determinação de sua dispersão (se próxima a 6.0 é considerado dentro da distribuição normal).

4.4. Multicolinearidade

A multicolinearidade é um problema computacional que se desenvolve quando duas ou mais variáveis independentes possuem forte correlação. O resultado é uma Matriz de Correlação com variabilidade única (próximo de zero) tendo em vista o efeito da divisão de um número por uma variação absoluta extremamente pequena (próxima de zero). O resultado desta divisão é um número, com um número de casas decimais bastante elevado, o que torna a aproximação computacional totalmente ineficaz. A existência de multicolinearidade introduz erros grosseiros no resultado da regressão, produzindo sérios erros na previsão da variável dependente. Felizmente a multicolinearidade é de fácil detecção e correção. A ocorrência de multicolinearidade dá-se quando um ou mais dos testes t assume resultado(s) muito pequenos (não significativos) e os valores de r^2 ou da estatística F são muito grandes. Se este for o caso, a Matriz de Correlação deve ser avaliada e, possivelmente, uma das Variáveis Independentes com forte correlação ser eliminada (regra geral, um coeficiente de correlação superior ou próximo a 0,7 entre duas Variáveis Independentes indica problemas de multicolinearidade), (REGRESSÃO MÚLTIPLA: uma digressão sobre seus usos. Autores: Istvan Karoly Kasznar, PhD Professor Titular da FGV e Presidente da IBCI, Bento Mario Lages Gonçalves, MSc Consultor Senior da IBCI).

A multicolinearidade sendo um problema no ajuste do modelo pode causar impactos na estimativa dos parâmetros. Podemos diagnosticar Multicolinearidade por meio do VIF (Variance Inflation Factor). Assim, o VIF_j é dado por:

$$VIF_j = \frac{1}{1-r_j^2} \quad (4.10)$$

Onde r_j^2 é o coeficiente de determinação da regressão de X_j sobre as outras variáveis explicativas.

Verificamos que VIF_j mede o quanto a variância do coeficiente β_j é inflacionada por sua colinearidade. Geralmente, o VIF é indicativo de problemas de multicolinearidade se $VIF > 10$ (<http://www.portalaaction.com.br/analise-de-regressao/362-multicolinearidade>).

5 - REGRESSÃO LOGÍSTICA

Tal como na análise de regressão linear múltipla é possível estimar efeitos através de parâmetros de regressão. A regressão logística é uma técnica estatística que tem como objetivo produzir, a partir de um conjunto de observações, um modelo que permita a predição de valores tomados por uma variável categórica, frequentemente binária, a partir de uma série de variáveis explicativas contínuas e/ou binárias. Na análise de regressão logística múltipla, o modelo de regressão da probabilidade de se observar um dado evento é dado por:

$$P_{estimada} = \frac{e^L}{1+e^L} \text{ em que } L = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_k * X_k \quad (5.1),$$

ou de forma equivalente,

$$P_{estimada} = \frac{1}{1+e^{-L}} \text{ em que } L = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_k * X_k \quad (5.2)$$

Em que $P_{estimada}$ é a probabilidade estimada de ocorrer um dado evento, sendo $e \cong 2,718$, número de Nepper, o valor utilizado na função exponencial, e $\beta_1, \beta_2, \dots, \beta_k$, os k coeficientes de regressão estimados e que correspondem às k variáveis independentes, com β_0 a constante estimada do modelo. Os parâmetros (coeficientes) considerados são estimados pelo método de máxima verosimilhança, que consiste em determinar os valores dos parâmetros que maximizem a probabilidade de obter o conjunto de valores observados. Atendendo a Hosmer e Lemeshow 1989, o M.M.V. permite estimar os coeficientes de regressão, que maximizam a probabilidade de obter as realizações da variável dependente da amostra.

A função de verosimilhança expressa a probabilidade dos dados observados, como uma função dos parâmetros desconhecidos. Os estimadores de máxima verosimilhança destes parâmetros, são escolhidos de modo a ser aqueles que maximizam a função de verosimilhança, cujas equações são dadas por:

$$\begin{cases} \frac{\delta L}{\delta \beta_0} = 0 \\ \frac{\delta L}{\delta \beta_j} = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^n [y_i - \pi(x_i)] = 0 \\ \sum_{i=1}^n x_{ij} [y_i - \pi(x_i)] = 0 \end{cases} \quad j = 1, \dots, m \quad (5.3)$$

5.1. Técnicas de Inferência Estatística aplicadas à regressão

“A Inferência Estatística consiste em fazer afirmações probabilísticas sobre as características do modelo probabilístico, que se supõe representar uma população, a partir dos dados de uma amostra aleatória (probabilística) dessa mesma população”.

Fazer uma afirmação probabilística sobre uma característica qualquer é associar à declaração feita, uma probabilidade de que tal declaração esteja correta (e, portanto, a probabilidade complementar de que esteja errada). Quando se usa uma amostra da população haverá SEMPRE uma probabilidade de se cometer um erro (justamente por ser usada uma amostra) daí que a diferença entre os métodos estatísticos e os outros, reside no fato de que os métodos estatísticos permitem calcular essa probabilidade de erro. E para que isso seja possível a amostra da população precisa ser aleatória.

As afirmações probabilísticas sobre o modelo da população podem ser basicamente:

=> estimar quais são os possíveis valores dos parâmetros (Estimação de Parâmetros):

- qual é o valor da média de uma variável que segue uma distribuição normal?
- qual é o valor da proporção de um dos 2 resultados possíveis de uma variável que segue uma distribuição binomial.

=> testar hipóteses sobre as características do modelo: parâmetros, forma da distribuição de probabilidades, etc. (Testes de Hipóteses).

- o valor da média de uma variável que segue uma distribuição é maior do que um certo valor?
- o modelo probabilístico da população é uma distribuição normal?
- o valor da média de uma variável que segue uma distribuição normal numa população é diferente da mesma média noutra população?

5.1.1. Inferência no modelo de regressão

Assim como na Inferência num modelo de regressão logística simples, também na regressão logística múltipla, podemos testar a significância dos parâmetros pelo teste de Wald, score e Razão de Verossimilhança (TRV).

5.1.1.1. *Teste de Wald*

Na realização deste teste testam-se as seguintes hipóteses:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

Para testar esta hipótese, a estatística de Wald é realizada da seguinte forma:

$$W_i = \frac{\hat{\beta}_i}{\sqrt{\text{Var}(\hat{\beta}_i)}} \quad (5.4)$$

Se não rejeitarmos H_0 temos que a variável X_i não explica a variável resposta.

No entanto, o teste de *Wald*, falha quando se rejeitam coeficientes que são estatisticamente significativos (*Hauck e Donner, 1977*, citado por *Crichton (2001)*). Assim, aconselha-se que os coeficientes, identificados por este teste como sendo estatisticamente não significativos, sejam testados novamente pelo teste da razão de verossimilhança.

5.1.1.2. Teste da Razão de Verossimilhança

O teste da razão de verossimilhança é realizado da mesma maneira no modelo de regressão logística simples ou num modelo de regressão logística múltipla, visando a significância dos p coeficientes das variáveis independentes.

$$G^2 = -2 \ln \left[\frac{(\text{verossimilhança sem a variável})}{(\text{verossimilhança com a variável})} \right] \quad (5.5)$$

Ou ainda

$$G^2 = -2 \ln(L_s) + 2 \ln(L_c) \quad (5.6)$$

L_s = verossimilhança do modelo sem a covariável

L_c = verossimilhança do modelo com a covariável

Na regressão múltipla, interessa-nos saber se há pelo menos uma variável significativa para o modelo. Sob a hipótese nula, os p coeficientes são iguais a zero, e a estatística G^2 tem distribuição Qui-Quadrado com p graus de liberdade. Nesse caso L_c = verossimilhança do modelo com as p variáveis explicativas e L_s = verossimilhança do modelo apenas com o intercepto. Como a Verossimilhança L , é inferior a 1 é usual usar $LL = \ln(L)$ que é um

número negativo, pelo que se multiplica por -2 para torna-lo positivo, maior, e com distribuição amostral conhecida. O $-2LL$ é assim um indicador da mediocridade do ajustamento do modelo aos dados, pois quanto maior for $-2LL$ pior o ajustamento e por outro lado se $-2LL=0$, o ajustamento é perfeito. Assim o modelo completo é estatisticamente significativo apenas quando a adição de uma ou mais variáveis independentes ao modelo reduz significativamente o $-2LL$.

5.1.1.3. *Intervalos de Confiança na Regressão Logística*

A estimação intervalar consiste na determinação de um intervalo onde, com uma certa confiança (probabilidade), esteja o parâmetro β desconhecido, tendo-se em conta um seu estimador. Assim,

$$P(L1 < \beta < L2) = \alpha \quad (5.7)$$

significa que a probabilidade do intervalo aleatório $(L1, L2)$ conter o valor exato β é α .

O intervalo $(L1, L2)$ é designado por intervalo de confiança para o parâmetro β , com um nível de confiança α .

O intervalo de confiança para um parâmetro β_j é baseado no seu teste de Wald. O intervalo de confiança de $100(1 - \alpha)\%$ para o parâmetro β_j é:

$$IC(\beta_j, 1 - \alpha) = [\hat{\beta}_j - z_{1-\alpha/2} DP[(\hat{\beta}_j)]; \hat{\beta}_j + z_{1-\alpha/2} DP[(\hat{\beta}_j)]] \quad (5.8)$$

Vantagem: É possível determinar o erro máximo cometido na estimação, com uma certa confiança. No entanto devemos ter em conta as variações das estatísticas amostrais de amostra para amostra e que nunca podemos ter intervalos com 100% de confiança.

5.1.1.4. *Razão de possibilidades (odds ratio)*

Atualmente muitos investigadores optam por analisar a relação entre duas variáveis de escala nominal através do rácio de produtos cruzados – *razão de possibilidade*, pois tem uma interpretação mais fácil do que o teste de *Qui Quadrado* (Bessa, 2007).

De acordo com Bessa (2007), a razão de possibilidade é uma medida antiga tendo sido usada por *Snow* no seu trabalho de identificação do fator risco da propagação da cólera em

Londres (1853). Esta medida é também utilizada como medida de associação em estudos de “caso- controle” e em estudos transversais controlados.

A razão das possibilidades (odds ratio - OR) é definida como a probabilidade de que um evento ocorra dividido pela probabilidade de que ele não ocorra, ou seja é o quociente entre a quantidade de eventos e a quantidade livre de eventos, sendo as OR calculadas da seguinte forma:

$$\text{odds} = \frac{\text{"Probabilidade de um acontecimento ocorrer"}}{\text{"Probabilidade de um acontecimento não ocorrer"}} \quad (5.9)$$

Assim, a razão de possibilidade é uma forma de se comparar se a probabilidade de um determinado evento é a mesma para dois grupos (*Wagner e Callegari-Jacques, 1998; Rumel in Revista Saúde Pública, 1986*). Desta forma, conclui-se que:

- *razão de possibilidades* =1 implica que o evento é igualmente provável em ambos os grupos;
- *razão de possibilidades* >1 significa que o evento é mais provável no 1º grupo;
- *razão de possibilidades* <1 implica que o evento é menos provável no 1º grupo.

Pode concluir-se que o significado da razão de probabilidade é semelhante ao risco relativo obtido em estudos de *coorte*, e expressa a força de associação entre o evento e o grupo (Pestana e Gageiro, 2005).

O Risco Relativo (*RR*) é a medida da força da associação entre a exposição e o evento (resultado). Com o *RR* avalia-se a variabilidade amostral da *OR* através do cálculo do intervalo de confiança e a interpretação é a mesma do *RR*, ou seja, se o intervalo de confiança da *OR* contiver a unidade, significa ausência de associação). A *OR* pode ser estimada em estudo tipo *coorte* (população seguida da causa para o efeito) e em estudo tipo caso- controle (população seguida do efeito para causa). Já o *RR* só pode ser estimado em estudos tipo *coorte* e não em estudos tipo caso-controle, porque este tipo de estudo não permite estimativa da taxa populacional do evento nos grupos estudados.

Quando o *RR* ou a *OR* é igual a 1, significa que não há efeito com o tratamento, ou seja, o efeito do tratamento é o mesmo nos dois grupos. Se o *RR/OR* é < 1, significa que o risco do evento é menor no grupo tratado que no controle. A variabilidade amostral do estudo pode

ser avaliada através de testes de significância ou via intervalo de confiança. Aqui daremos preferência à segunda abordagem. Para um dado nível de significância ($p\text{-value} \leq 0,05$), o intervalo de confiança representa o intervalo onde deve estar o parâmetro, ou seja, o risco relativo verdadeiro. Se o valor 1, que se refere à nulidade da associação, não estiver contido no intervalo, temos uma confiança que na população de onde a nossa amostra foi extraída, o *RR* é diferente de 1, sendo, portanto significativo o achado da amostra. Para a realização dos cálculos do risco relativo e o intervalo de 95% de confiança, assim como o seu significado, consulte www.paulomargotto.com.br, em *Entendendo Bioestatística e Exercício da Medicina Baseado em Evidências*.

O intervalo de confiança mostra o limite dentro do qual existe a certeza do verdadeiro efeito do tratamento; estima a magnitude da associação e informa a variabilidade da estimativa, através dos limites inferior e superior.

5.1.1.5. Teste de Ajustamento do modelo

Para testar a significância do ajustamento do modelo torna-se necessário formular as seguintes hipóteses:

H_0 : O modelo ajusta-se aos dados

H_1 : O modelo não se ajusta aos dados

Como resultado do cruzamento dos preditores qualitativos, estando os dados agrupados em J células, a estatística de teste é o χ^2 de Pearson (Hosmer & Lemeshow, 2000)

$$\chi_P^2 = \sum_{j=1}^J \left(\frac{y_j - n_j \hat{\pi}_j}{\sqrt{n_j \hat{\pi}_j (1 - \hat{\pi}_j)}} \right)^2 = \sum_{j=1}^J \frac{(O_j - E_j)^2}{E_j} \quad (5.10)$$

Onde y_j é o número de sucessos na célula j , $\hat{\pi}_j$ é a probabilidade estimada da célula j , e n_j é o número de elementos da célula j ou, usando a notação tradicional do teste do χ^2 , O_j e E_j são o número de sucessos observados e o número de sucessos esperados na célula j , respetivamente. Alguns autores (ver e.g., McCullag & Nelder, 1989) usam a estatística Deviance

$$D = 2 \sum_{j=1}^J \left[y_j \ln \left(\frac{y_j}{n_j \hat{\pi}_j} \right) + (n_j - y_j) \ln \left(\frac{n_j - y_j}{n_j (1 - \hat{\pi}_j)} \right) \right] \quad (5.11)$$

Supondo H_0 verdadeira, e para amostras grandes, χ^2 e D apresentam distribuição assintótica χ^2 com $j-p-1$ graus de liberdade, rejeitando-se H_0 se o p -value for inferior ou igual ao nível de significância fixado. No entanto, se o modelo incluir variáveis quantitativas, caso em que o número de células é aproximadamente igual ao número de elementos na amostra ($J \approx n$) estas estatísticas não podem ser usadas com dados individuais. Neste cenário, a distribuição χ^2 da estatística de teste não é, geralmente, válida, pois uma das condições de aplicação do teste do χ^2 exige que 80% dos $E_j \geq 5$. Hosmer & Lemeshow (1982), propuseram uma outra estatística para testar a hipótese do ajustamento do modelo aos dados, baseada na estimativa das probabilidades de sucesso de cada uma das n observações. Esta estatística de teste é obtida com um teste do χ^2 a uma tabela de contingência $2 \times g$. Esta tabela é construída classificando as duas classes da variável dependente dicotômica por g grupos definidos pelos decís das probabilidades de “sucesso” (i. é, $[0;0,1[$, $[0,1;0,2[$, ..., $[0,9;1]$) estimadas pelo modelo para cada uma das observações. Geralmente g é igual a 10, mas se o modelo tiver variáveis independentes qualitativas g pode ser inferior ou igual a 10, em função do número de células distintas resultantes do cruzamento das classes das variáveis independentes. O número de sucessos da variável dependente em cada classe definida pelos decís (O_i) e os valores esperados nessas classes, usando a média das probabilidades estimadas para cada um dos grupos da variável dependente e cada uma das classes decílicas ($E_i = n_i \times \hat{\pi}_j$) são então usados no cálculo da estatística de teste:

$$\chi_{HL}^2 = \sum_{j=1}^J \frac{(O_j - E_j)^2}{E_j} \quad (5.12)$$

Para amostras de grande dimensão a estatística χ_{HL}^2 tem distribuição assintótica $\chi_{(g-2)}^2$. Finalmente rejeita-se H_0 se o p – value $\leq \alpha$. Para que o modelo completo se ajuste aos dados, é necessário que os valores observados sejam suficientemente próximos dos valores esperados, isto é, que não se rejeite H_0 . Se o ajustamento do modelo for bom, a maioria dos “1 – Sucesso” encontra-se nas classes decílicas altas, e a maioria dos “0 – Insucessos” encontra-se nas classes decílicas baixas, mas se o ajustamento não for bom, os “0” e “1” encontram-se distribuídos aleatoriamente pelas classes decílicas.

5.1.1.6. *Análise de resíduos*

Os diagnósticos da regressão logística, baseados na análise de resíduos, servem os mesmos propósitos dos diagnósticos descritos para a regressão linear. A análise de resíduos permite identificar a presença de outliers e casos influentes na estimação do modelo. As medidas usadas para a análise de resíduos são praticamente as mesmas da regressão linear (resíduos, Leverage, Distância de Cook e DfBetas). Contudo, na regressão logística é necessário considerar que as variâncias das observações não são constantes como acontece na regressão linear e as fórmulas de cálculo têm de ser adaptadas a esse facto (Pregibon(1981) e Hosmer & Lemshow(2000)).

Os resíduos ou erros são definidos como sendo a diferença entre o número de sucessos observados (y_j) e o número de sucessos estimados (\hat{y}_j) pelo modelo para a célula j , isto é:

$$e_j = y_j - \hat{y}_j = y_j - n_j \hat{\pi}_j^2 \quad (5.13)$$

Onde n_j é o número de observações da célula j e $\hat{\pi}_j$ é a probabilidade de sucesso estimada na célula j . Estes resíduos podem também ser calculados simplesmente como a diferença entre a probabilidade observada e a probabilidade estimada pelo modelo para o sujeito j , da seguinte forma (procedimento de cálculo usado pelo SPSS):

$$e_{pj} = p_j - \hat{\pi}_j \quad (5.14)$$

Estes resíduos designam-se por resíduos não estandardizados.

Para obter os resíduos estandardizados, também designados por resíduos de Pearson, é necessário dividir os resíduos não estandardizados pela estimativa do desvio padrão dos valores estimados:

$$e'_j = \frac{e_j}{\sqrt{n_j \hat{\pi}_j (1 - \hat{\pi}_j)}} \quad (5.15)$$

Estes resíduos apresentam média 0 e desvio-padrão aproximadamente igual a 1. Para amostras de grande dimensão $e'_j \simeq N(0,1)$. Assim, com recurso às tabelas da distribuição Normal, 95% dos valores $|e'_j|$ devem ser inferiores a 1,96 ($\alpha = 0,05$ ($z_{0,975} = 1,96$)), e qualquer observação com $|e'_j|$ superior a 1,96 ≈ 2 pode ser classificada como um outlier para $\alpha = 0,05$ ($z_{0,975} = 1,96$). Convém referir que, estes resíduos não são completamente

estandardizados uma vez que a variância depende das observações e da sua influência na estimação dos coeficientes do modelo. Esta influência pode ser estimada pela Leverage, isto é, pelos elementos da diagonal da matriz H (matriz das segundas derivadas da função de verossimilhança em ordem aos parâmetros do modelo) (Hosmer & Lemeshow, 2000):

$$h_j = n_j \hat{\pi}_j(x_j) [1 - \hat{\pi}_j(x_j)] x_j' (X' V X)^{-1} x_j' \quad (5.16)$$

Onde $x_j' = [1, x_{1j}, x_{2j}, \dots, x_{pj}]$ é o vetor com os valores da observação j em todas as variáveis independentes, X é a matriz dos valores das variáveis independentes e V é a matriz cujos elementos da diagonal principal são $\hat{\pi}_j(1 - \hat{\pi}_j)$. A Leverage pode ser usada para avaliar a influência de cada observação no ajustamento do modelo. Uma observação j com Leverage de aproximadamente 0 é pouco importante no ajustamento do modelo, enquanto que se a Leverage for próximo de 1, essa observação é importante. Os valores de Leverage podem ser comparados com a Leverage média dada por $(p+1)/n$ mas, é de notar que se as probabilidades estimadas para uma determinada observação j forem inferiores a 0,1 ou superiores a 0,9, os valores de Leverage podem ser reduzidos e ainda assim essa observação ser influente no modelo (Norusis, 2005).

O resíduo “estudentizado” (também designados por resíduos de Pearson estandardizados) obtém-se pela expressão:

$$r_j = \frac{e_j'}{\sqrt{1-h_j}} \quad (5.17)$$

Estes resíduos apresentam agora variância constante e igual 1.

Para além de permitir identificar outliers, os resíduos estandardizados e a Leverage podem também ser utilizados para avaliar a influência de uma observação no ajustamento do modelo. Hosmer & Lemeshow (2000) demonstraram que a variação da estatística de χ^2 de Pearson (usada para testar o ajustamento do modelo) por eliminação da observação (ou célula) j é simplesmente:

$$\Delta \chi_j^2 = \frac{(e_j')^2}{1-h_j} = (r_j)^2 \quad (5.18)$$

Como regra empírica, podemos considerar que valores de $\Delta \chi_j^2$ superiores a 3,84 indicam observações influentes no ajustamento do modelo para $\alpha = 0,05$, uma vez que $\Delta \chi_j^2$ segue

uma distribuição χ^2 com um grau de liberdade e que $\chi_{0,95;1}^2 = 3,84$. Uma outra medida da influência de uma observação (ou célula) j no ajustamento do modelo é a Deviance (d_j) dessa observação:

$$d_j = \text{Sign}(e_j) \sqrt{2 \left[y_j \text{Ln} \left(\frac{y_j}{n_j \hat{\pi}_j} \right) + (n_j - y_j) \text{Ln} \left(\frac{n_j - y_j}{n_j (1 - \hat{\pi}_j)} \right) \right]} \quad (5.19)$$

Onde a função $\text{Sign}(e_j)$ devolve o sinal (+ ou -) de e_j . A estatística D para avaliar a qualidade do modelo, é simplesmente o somatório dos quadrados dos d_j (5.18). Descritas desta forma, as observações influentes no modelo são as que apresentam valores elevados (superiores a 1) de d_j . Por outro lado, observações com Deviance elevada são mal previstas pelo modelo. Se a amostra for de dimensão elevada, a Deviance apresenta distribuição normal assintótica, e valores absolutos superiores a 1,96 indicam casos extremos de Deviance para $\alpha = 0,05$. A influência de uma determinada observação (ou célula) j na estimação dos coeficientes do modelo é estimada por uma medida análoga à Distância de Cook. Esta medida indica a variação dos resíduos quando a observação j é eliminada do ajustamento do modelo. Usando a Leverage e os resíduos de Pearson estandardizados, esta medida pode calcular-se como (Pregibon, 1981):

$$DC_j = r_j^2 \frac{h_j}{(1-h_j)} \quad (5.20)$$

No SPSS, o cálculo da DC não é exato, mas uma aproximação, sendo que valores de DC_j superiores a 1 indicam observações influentes na estimação dos coeficientes do modelo (Hosmer & Lemeshow, 2000).

A influência de cada observação na estimação de cada um dos coeficientes de regressão pode ser estimada com os DfBetas que na regressão logística são calculados de forma aproximada por:

$$\text{DfBeta}_{ij} = \hat{\beta}_i - \hat{\beta}_{i(-j)} \quad (5.21)$$

Onde $\hat{\beta}_i$ é a estimativa do coeficiente de regressão, ajustado com todas as observações, e $\hat{\beta}_{i(-j)}$ é a estimativa do coeficiente de regressão ajustado sem a observação j . À semelhança da regressão linear pode-se considerar que valores de DfBetas superiores a 2 (em rigor,

superiores a $2 \times \sqrt{(p+1)/n}$ onde $p+1$ é o número de coeficientes do modelo e n é a dimensão da amostra) são observações influentes. Nesta situação a observação j deve ser examinada com precaução já que a sua presença no modelo afeta a estimativa de β_i .

5.1.1.7. Classificação da previsão do modelo

Depois de obtido o modelo de regressão logística e conhecidas as estimativas dos coeficientes do modelo torna-se possível calcular as probabilidades $\hat{\pi}_j$, de cada um dos sujeitos pertencentes ao grupo “1 – Sucessos” comparativamente ao grupo de referência “0 – Insucessos”, com base nas expressões:

$$\hat{\pi}_j = \frac{e^{\beta_0 + \beta_1 X_{1j} + \dots + \beta_m X_{mj}}}{1 + e^{\beta_0 + \beta_1 X_{1j} + \dots + \beta_m X_{mj}}} \quad (5.22)$$

equivalente a

$$\hat{\pi}_j = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{1j} + \dots + \beta_m X_{mj})}} \quad (5.23)$$

Assim, o sujeito j é classificado no grupo “1” se a probabilidade de $Y_j = 1$ for superior a 0,5 e caso contrário é classificado no grupo “0”. O valor da probabilidade 0,5 é arbitrário podendo ser escolhido outro valor de modo a tornar a classificação mais rigorosa (0,7) ou menos rigorosa (0,4). A escolha dos valores de corte pode ser ajudada pelo conhecimento existente a priori sobre tendências na população, ou decorrente, por exemplo, de outros estudos empíricos.

Para avaliar a qualidade da classificação feita pelo modelo é usual comparar a percentagem global de classificações corretas obtidas com o modelo, com a percentagem proporcional de classificações corretas por acaso. Esta percentagem é calculada a partir do número de sujeitos observados em cada uma das k classes da variável dependente (C_j), pela seguinte expressão

$$\text{Classificação correta proporcional por acaso (\%)} = 100 \times \sum_{j=1}^k \left(\frac{C_j}{N} \right)^2$$

Se a percentagem dos casos classificados corretamente pelo modelo for superior em pelo menos 25% à percentagem de classificação proporcional por acaso, considera-se que o modelo tem boas propriedades classificatórias.

A eficiência classificatória do modelo pode também ser avaliada pela sensibilidade e especificidade do modelo. Define-se sensibilidade como a percentagem de classificações corretas na classe de referência “1 – Sucessos” da variável dependente, ou seja, é quando um sujeito tem a característica que se quer modelar e o modelo prevê corretamente esta característica. Define-se especificidade como a percentagem de classificações corretas na classe de referência “0 – Insucessos” da variável dependente, ou seja, é quando um sujeito não tem a característica que se quer modelar e o modelo prevê corretamente que o sujeito não tem esta característica. Na Biomedicina, onde estas estatísticas são usadas com frequência em instrumentos de diagnóstico clínico, define-se Sensibilidade como sendo a probabilidade condicionada de um diagnóstico positivo sabendo que o paciente tem a doença, e define-se especificidade como sendo a probabilidade condicionada de um diagnóstico negativo sabendo que o paciente não tem a doença. Formalmente,

$$\text{Sensibilidade} = P[\hat{Y} = 1|Y = 1]$$

$$\text{Especificidade} = P[\hat{Y} = 0|Y = 0]$$

Um modelo com boas capacidades preditivas apresenta sensibilidade e especificidade superior a 80%. Para percentagens entre 50% e 80%, o modelo tem capacidade preditiva razoável e abaixo do 50% o modelo tem capacidade preditiva medíocre.

Uma outra medida para testar a capacidade do modelo discriminar os sujeitos com a característica de interesse versus sujeitos sem a característica de interesse é a área sob a curva ROC, Curvas de Características de Operação do Receptor (Curvas ROC- Receiver Operating Characteristic). A Curva ROC é assim uma forma mais eficiente de demonstrar a relação normalmente antagônica entre a sensibilidade e a especificidade dos exames que apresentam resultados contínuos., por isso é uma ferramenta poderosa para medir e especificar problemas no desempenho do diagnóstico em medicina por permitir estudar a variação da sensibilidade e especificidade para diferentes valores de corte.

A Curva ROC é um gráfico de sensibilidade (ou taxa de verdadeiros positivos) versus especificidade (ou taxa de falsos positivos) (Figura 4).

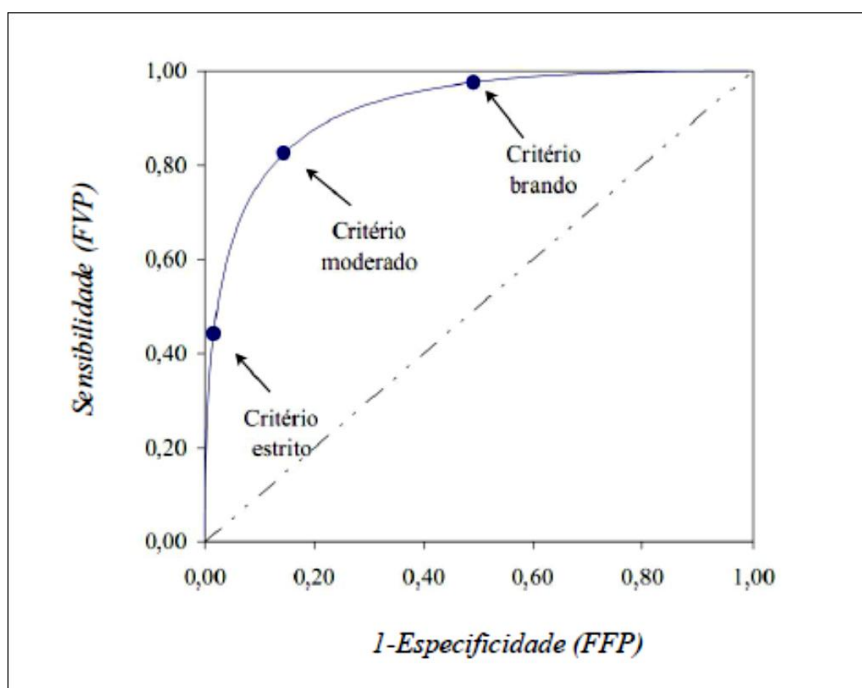


Figura 4: Gráfico da Curva ROC, para uma dada capacidade de discriminação, com a variação do critério de decisão.

Fonte: Braga(2000)

A linha diagonal pontilhada corresponde a um teste que é positivo ou negativo, aleatoriamente. A Curva ROC permite evidenciar os valores para os quais existe maior otimização da sensibilidade em função da especificidade, que corresponde ao ponto em que se encontra mais próxima do canto superior esquerdo do diagrama, uma vez que o índice de positivos verdadeiros é 1 e o de falsos positivos é zero. A curva ROC discrimina entre dois estados, onde cada ponto da curva representa um compromisso diferente entre a *Sensibilidade* e o falso positivo que pode ser definido pela adoção de um valor diferente do ponto de corte de anormalidade. Um critério restrito é aquele que traduz uma pequena fração de falsos positivos assim como uma pequena fração de verdadeiros positivos (Braga, 2000). Analiticamente, a área abaixo da curva ROC pode ser determinada através de:

- Métodos de resolução numérica, como por exemplo, a regra do trapézio;
- Métodos estatísticos: Relação com a estatística de *Wilcoxon-Mann-Witney* (Hanley, 1988, citado por Braga (2000)); e estimativa de Máxima Verosimilhança (Hanley e McNeil, 1982, citado por Braga (2000)).

Hosmer & Lemeshow (2000) apresentam valores indicativos da área sob a curva ROC que podem ser utilizados para classificar o poder discriminante do modelo de regressão e que, numa regra de aplicação geral, são dados pela tabela seguinte:

Área sob a curva ROC	Poder discriminante do modelo
0,5	Sem poder discriminativo
]0,5;0,7[Discriminação fraca
]0,7;0,8[Discriminação aceitável
]0,8;0,9[Discriminação boa
$\geq 0,9$	Discriminação excepcional

Tabela 1: Classificação do poder discriminativo do modelo de regressão

6 - ANÁLISE DE CLUSTERS

A análise de clusters consiste num conjunto de procedimentos de estatística multivariada, visando organizar um conjunto de entidades (indivíduos ou objetos), para o qual é conhecida informação detalhada, em grupos homogéneos.

6.1. Metodologia da análise de clusters

Dado um conjunto de n indivíduos, para os quais existe informação sob a forma de p variáveis, o método de análise de clusters procede ao agrupamento dos indivíduos em função da informação existente, de tal modo que os indivíduos pertencentes a um mesmo grupo sejam tão semelhantes quanto possível e sempre mais semelhantes aos elementos do mesmo grupo do que a elementos dos restantes grupos. Neste tipo de análise não há qualquer tipo de dependência entre as variáveis: os grupos definem-se por si mesmos sem que haja relação causal entre as variáveis utilizadas.

Na análise de clusters os métodos são exploratórios, pois a ideia é gerar hipóteses e não testá-las, havendo uma necessidade posterior de validação dos resultados através da aplicação de outros métodos estatísticos.

Uma dificuldade inicial é a de não existir uma única via de definição de grupos, ou seja um critério de partição ou agrupamento dos indivíduos.

A análise de cluster compreende as seguintes etapas:

- a) Selecção dos indivíduos a serem agrupados

- b) Definição do conjunto de variáveis a partir das quais se fará o agrupamento dos indivíduos
- c) Definição de uma medida de semelhança (distância) entre cada dois indivíduos
- d) Escolha de um critério de agregação, ou seja, definição de um algoritmo de partição
- e) Validação dos resultados encontrados.

6.1.1. Formulação do problema

A tarefa mais importante da formulação do problema é a seleção das variáveis em que a agregação se baseia. A inclusão de uma variável irrelevante pode distorcer os resultados. As variáveis devem ser selecionadas com base numa teoria, ou considerando as hipóteses que vão ser testadas.

6.1.2. Medida de distância

É necessária uma medida que avalie a distância ou similaridade entre os objetos de forma a poder agrupá-los. A medida de distância utilizada é a distância euclidiana: a raiz quadrada da soma dos quadrados das diferenças entre os valores para cada variável. Outra medida é a distância absoluta (city block) soma dos valores absolutos das diferenças entre os valores das variáveis.

Se as variáveis são medidas em escalas diferentes, os resultados podem ser influenciados pelas unidades de medida.

Nestes casos as variáveis devem ser normalizadas, de forma a terem média igual a zero e desvio padrão igual a 1. Esta normalização pode reduzir as diferenças entre as variáveis, contrariando os objetivos da agregação. Os outliers (casos atípicos) devem também ser removidos.

A formação de cluster deve ainda observar 2 princípios fundamentais:

Princípios Homogeneidade e Separação, caracterizados respetivamente por, elementos num mesmo cluster são próximos e elementos em clusters diferentes estão mais afastados entre si.

6.1.3. Procedimentos de Agregação

Os procedimentos de agregação podem ser hierárquicos ou não hierárquicos. O procedimento hierárquico é caracterizado por ter um desenvolvimento com uma estrutura

em árvore e compreende métodos de ligação (linkage), de variância e de centróides. Dentro destes, os métodos com melhor desempenho são o de Ward (variância) e o average linkage. Os métodos não hierárquicos, são também referidos como métodos k-médias, e possuem as seguintes desvantagens:

- a) O número de clusters tem de ser especificado à partida;
- b) A seleção de centros de cluster é arbitrária;
- c) Os resultados podem depender da seleção dos centros.

Punj. G & Stewart, DW 1983, sugeriram que os dois métodos sejam usados em sequência: primeiro um método hierárquico (Ward, Average linkage) para obter o número de clusters e centróides e depois, um método não hierárquico para encontrar os clusters ótimos.

7 - ANÁLISE DISCRIMINANTE

A análise discriminante é utilizada para discriminar grupos, analisando dados com uma variável categórica dependente e variáveis de escala intervalar independentes e tem como objetivo encontrar uma combinação linear das variáveis independentes que minimize a probabilidade de incorreta classificação dos indivíduos.

Podemos definir análise discriminante como um método estatístico para classificar indivíduos ou objetos de modo exaustivo em grupos mutuamente exclusivos, com base num conjunto de variáveis independentes, sendo para isso determinadas combinações lineares dessas variáveis que discriminam entre grupos à priori, de tal modo que seja minimizada a probabilidade de erro de incorreta classificação à posteriori. No entanto, este facto só é verdadeiro quando se verificam os seguintes pressupostos:

- Os grupos devem ser retirados de populações que seguem uma distribuição normal multivariada para as p variáveis discriminantes
- Dentro dos grupos a variabilidade deve ser idêntica, isto é as matrizes de variância e covariância devem ser iguais.
- O número de grupos deve ser $K \geq 2$
- O número de observações dentro de cada grupo é pelo menos 2.
- Nenhuma das variáveis discriminantes deve ser combinação linear das restantes.

Emprega-se para descobrir as características que distinguem os membros de um grupo dos de outro, de modo a que, conhecidas as características de um novo indivíduo, se possa prever a que grupo pertence.

O primeiro passo da análise discriminante consiste em identificar: os objetivos, a variável categórica e as variáveis independentes. A variável categórica deve consistir em duas ou mais categorias mutuamente exclusivas. Se tiver uma escala de medida intervalar tem que ser convertida em categorias. As variáveis preditivas devem ser selecionadas com base num modelo teórico ou numa pesquisa anterior, se possível.

Seguidamente deve dividir-se a amostra em 2 partes: uma amostra para análise ou estimação e uma amostra para validação. A distribuição nas duas amostras deve seguir a distribuição na amostra total.

A interpretação da análise discriminante só faz sentido se as funções discriminantes forem estatisticamente significativas. A hipótese nula indica que, na população, as médias de todas as funções discriminantes, para todos os grupos, são iguais, podendo ser testada pelo λ de Wilks. Assim a análise discriminante é levada a cabo através de uma ou mais combinações lineares das variáveis independentes utilizadas X_i . Cada combinação linear Y_i constitui uma função discriminante,

$$Y_i = a_{j0} + a_{j1}X_1 + a_{j2}X_2 + \dots + a_{jp}X_p \quad (7.1)$$

Em que os a_{ij} são coeficientes de ponderação e os X_j são as variáveis discriminantes não normalizadas. A situação ideal seria encontrar para uma dada função discriminante, valores idênticos para todos os casos pertencentes a um mesmo grupo. Simultaneamente, as funções são determinadas de modo a maximizarem a separação entre grupos. O número máximo de funções discriminantes é igual ao número de grupos menos um, ou ao número de variáveis discriminantes, sendo o critério de escolha baseado no menor destes dois valores. Uma vez estimadas as funções é possível concretizar os dois objetivos da análise discriminante: análise e classificação.

O valor do coeficiente para uma variável preditiva depende das outras variáveis preditivas incluídas na função discriminante. Pode ter-se uma ideia da importância relativa das variáveis analisando os coeficientes normalizados da função discriminante. Regra geral, variáveis preditivas com valores mais elevados nos coeficientes normalizados contribuem mais fortemente para o poder discriminante da função.

Como já foi referido, os dados são distribuídos aleatoriamente por duas amostras:

- Amostra de análise – utilizada para estimar as funções discriminantes
- Amostra de validação – utilizada para criar a matriz de classificação

Os valores das variáveis preditivas dos casos da amostra de validação são processados pelas funções discriminantes (onde se obtêm os valores previstos de validação) e é posteriormente calculada a percentagem de casos corretamente classificados (precisão de classificação).

A obtenção dos valores previstos discriminantes de validação é feita, utilizando os pesos discriminantes estimados a partir da amostra de análise que são multiplicados pelos valores das variáveis preditivas da amostra de validação.

A precisão de classificação obtida pela análise discriminante deve ser 25% superior à obtida pelo acaso, que é o inverso do número de grupos definidos pela variável categórica, isto se os grupos tiverem tamanho igual.

PARTE II

MATERIAL E RESULTADOS

Dado um determinado conjunto de dados e com o objetivo de efetuar uma análise estatística deve-se fazer uma escolha racional acerca do método mais apropriado a esta.

Neste sentido tem-se que ter em atenção algumas considerações importantes, como os objetivos definidos no contexto de um problema específico, as características matemáticas envolvidas, as hipóteses estatísticas a serem feitas sobre as variáveis em análise e ainda a forma de recolha dos dados (Braga, 1994).

1 - PARTICIPANTES E VARIÁVEIS DO ESTUDO

Foram recolhidos dados de 153 pacientes que receberam transplante de rins de dadores vivos e dadores cadáver, dos quais foram registadas 58 variáveis. Os recetores de dadores vivos foram 40 pacientes e os recetores de dadores cadáver foram 113 pacientes. Os dados recolhidos dizem respeito às seguintes **caraterísticas do recetor**: idade; sexo; ASA; Tempo em diálise; técnica dialítica; técnica anestésica; Etiologia DRC; Comorbilidades; Tipos de Comorbilidades: HTA, DCI, DAP, Tabagismo, DM, Hiperuricemia, Dislipidemia, AVC, HPT, Anemia; **caraterísticas cirurgicas**: Tempo de cirurgia; Tempo de isquemia; Enxerto safena; **caraterísticas pós-cirurgicas**: Analgesia pós-operatória; PCA, N°_pedidos_1^{as}24horas; N°_dias; BNE, N°_dias_Após; Escala_dor_R; Escala_dor_M; Tempo_internamento; Complicações_pós_oper_30_dias_ALTA; Tipos de complicações: Cardiovasculares, Respiratórias, CNS, Urorenal, Infeciosas, Hematológicas e Endócrinas; Reintervenção; Reinternamento_3^o_mês e consulta follow_up; **caraterísticas do dador no caso de dadores vivos**: Sexo_dador; Idade_dador; ASA_dador; Tipo de relação; Relação do 1^o grau; Técnica anestésica_dador; T_isquemia quente_rim_dador e diferença_idade_maior_10_anos. Para cada tipo de complicações no pós operatório foram recolhidos também os dados referentes à descrição das variantes possíveis, ao que se juntaram à base de dados mais 8 variáveis correspondentes às referidas descrições, totalizando as 58 variáveis referidas.

2 - ANÁLISE DESCRITIVA

Num estudo preliminar, que compreendeu a análise dos dados, procurou-se tomar conhecimento dos comportamentos das variáveis em análise, recorrendo-se a uma análise exploratória. Este estudo, englobou o cálculo das principais características amostrais, de forma a estudar-se os comportamentos das variáveis, e ainda, a determinação de qual ou quais as variáveis de maior interesse para o estudo. A informação assim obtida no que respeita aos dados numéricos encontra-se compilada na tabela que se segue:

Estatística Descritiva						
	N	Valor Min.	Valor Máx.	Soma	Média	Desvio Padrão
Idade	153	4	71	7140	46,67	14,772
Tempoemdiálise	153	,00	23,00	729,69	4,7692	4,34137
TCirurgia	132	63	379	16254	123,14	43,766
TIsquemia	73	27	1565	62702	858,93	463,736
Númerodepedidos1ª24h	72	0	198	2119	29,43	39,976
Númerodedias	85	1	4	151	1,78	,679
Númerodedias_A	2	3	77	80	40,00	52,326
Escala_da_dorR	83	0	2	17	,20	,435
Escala_da_dorM	83	0	2	77	,93	,513
Tempointernamento	153	5	75	2603	17,01	11,935
Idadedador	40	29	66	1836	45,90	9,870
TIsquemiaquenterimdador	34	130	780	9572	281,53	145,699

Tabela 2: Análise Descritiva dos dados numéricos

Relativamente aos dados categóricos chegou-se às seguintes conclusões:

56,2 dos transplantados do sexo masculino e 43,8 do sexo feminino. De acordo com a tabela ASA, 9,2% dos recetores foram classificados com III (Paciente com doença sistêmica severa, com limitação funcional) e 90,8% com IV (Paciente com doença sistêmica severa, representa risco de vida constante).

Relativamente aos recetores: Quanto à Etiologia DRC verificou-se que a Etiologia DRC com a maior percentagem de 28,1% (43 pacientes) era desconhecida, 22,2% (34 pacientes) referente à Glomerulonefrite, 15% (23 pacientes) referente à Uropatia malformativa-Refluxo e 14,4% (22 pacientes) referente à Nefropatia diabética. No que diz respeito às comorbilidades, 79,9% dos transplantados (122 pacientes) apresentam comorbilidades sendo as mais frequentemente verificadas a Hipertensão Arterial com 71,9% dos casos

(110 pacientes), a Anemia com 41,8% dos casos (64 pacientes), o Tabagismo com 19% dos casos (29 pacientes), a Dislipidemia com 17,6% dos casos (27 pacientes), a Hiperuricemia e HPT com 12,4% dos casos (19 pacientes) e por fim as Diabetes Mellitus com 10,5% dos casos (16 pacientes), todas as outras aparecem com uma percentagem inferior a 5%. Todos os doentes realizaram diálise com duração média de 4,8 anos e desvio padrão 4,3 anos, sendo a técnica dialítica mais utilizada a HD com 73,9% correspondentes a 113 doentes dos quais 94 são recetores de doadores cadáver e 19 recetores de doadores vivos, seguida da técnica DP com 20,9% dos casos, sendo 19 recetores de doadores cadáver e 13 recetores de doadores vivos. A técnica anestésica mais utilizada foi a AGB em 90,2% dos casos.

Debruçando sobre os procedimentos de transplante quanto ao tipo de analgesia pós-operatória foi aplicada a PCA a 58,2% dos casos (89 pacientes) e a convencional a 41,2% dos casos (63 pacientes) e apenas um paciente recebeu BNE. Em relação às complicações que surgiram após operação até 30 dias ou até à alta, ocorreram em 86,3% dos casos (132 pacientes), sendo as que ocorreram com maior frequência as Infecciosas em 62,7% dos casos, com 15,7% correspondente à infecção do líquido de preservação e 12,4% à ITU (Infecção do Trato Urinário); complicações Urorenais em 57,5% dos casos, dos quais 4,6% correspondente à AFE/HD; Hematológicas em 45,1%, sendo a Anemia/ GR a mais verificada que ocorreu em 28,8% dos pacientes; Cardiovasculares em 27,5% com maior ocorrência de HTA em 4,6% e Endócrinas em 19% dos casos, com maior ocorrência de DM pós-TR em 11,1%. O número de reintervenções corresponde ao número de reinternamentos ao 3º mês que totalizou 13,7% que correspondem no caso das reintervenções a 7 pacientes recetores de doadores vivos e 14 pacientes recetores de doadores cadáver e no caso do reinternamento correspondem a 6 pacientes recetores de doadores vivos e 15 pacientes recetores de doadores cadáver. A média do tempo de internamento foi de 17 dias e desvio padrão de 12 dias. Nenhum dos recetores de doadores vivos apresentou complicações pós operatórias respiratórias, endócrinas e do sistema nervoso central. Considerando a totalidade da base de dados mais de 95% dos pacientes não apresentaram complicações respiratórias nem de Sistema Nervoso Central.

Em relação às comorbilidades e suas diferentes tipologias, complicações pós-operatórias e suas diferentes tipologias, tempos de internamento, reintervenções e readmissões hospitalares foram já efetuados três estudos, referidos na Introdução desta dissertação,

sendo um em relação a receptores de dadores vivos, o segundo em relação a receptores de dadores cadáveres e o terceiro relativamente aos dois tipos de receptores de dadores vivos e de dadores cadáver, dos quais se obtiveram os seguintes resultados:

No caso de dadores vivos e considerando as variáveis pré-transplante e as variáveis resultado pós-transplante, os resultados mostraram que só a PAD (Doença Arterial Periférica) e a Etiologia da doença renal prevêm a um nível significativo, o tempo de hospitalização. Concluiu-se também que a causa da doença renal em estágio final (ESRD), parece ser preditiva dos diferentes resultados no transplante renal. As complicações pós-transplante foram registadas em 65% do receptores e a maioria destas complicações está relacionada com causas infecciosas, sendo as infeções do trato urinário as mais frequentemente registadas. Identificou-se ainda uma significativa correlação entre a ocorrência de complicações infecciosas ou urológicas e o tempo de hospitalização. Encontrou-se também correlação a um nível significativo entre as reintervenções e a incidência de complicações cardiovasculares e a readmissão hospitalar. Relativamente a esta última, os resultados mostraram que apenas 15% do receptores requereram readmissão totalizando quase metade dos casos descritos na literatura (Araujo AM, Santos F, Guimaraes J, Nunes CS, Casal M (2015)).

Com respeito aos receptores de dadores cadáver, outro estudo, (Santos F, Guimaraes J, Araujo AM, Nunes CS, Casal M (2015)), concluiu que a maioria dos pacientes apresentou complicações no período pós operatório. Neste estudo a regressão linear revelou três factores com uma influência significativa na duração do tempo de internamento, sendo eles a terapia de substituição renal, o tempo de isquemia e a doença Isquêmica do coração. Também neste estudo a regressão logística mostrou que a diabetes Mellitus antes do transplante e o Tempo de Isquemia foram preditores para a existência de complicações nefro-urológicas; a anemia antes do transplante renal é um fator preditor de complicações hematológicas e que a anemia e a terapia de substituição renal são factores preditores de complicações cardiovasculares. Concluiu-se assim que a anemia antes do transplante demonstrou ser um fator de risco para complicações cardiovasculares e hematológicas, pois a baixa hemoglobina antes do transplante renal aumenta o risco da necessidade de transfusão de sangue e durante o período operatório as perdas de sangue estão associadas à hipovolémia e portanto baixa na hemoglobina.

Um terceiro estudo (Guimaraes J, Araujo AM, Santos F, Nunes CS, Casal M (2015)), também referido na introdução desta tese, mostrou não haver diferenças nas reintervenções e

readmissões hospitalares entre os dois grupos. Considerando as variáveis pré-transplante e pós operatórias com diferenças significativas entre os dois grupos (Dadores vivos e Dadores cadáver), a análise de regressão logística mostrou que a idade é um fator de risco independente para a generalidade de complicações pós operatórias; a hipertensão antes do transplante e o tempo de isquemia quente são preditores de complicações cardiovasculares; o tempo de isquemia frio é um preditor de complicações nefro- urológicas e endócrinas. Este estudo concluiu ainda que 93,8% de recetores de dadores cadáver têm complicações durante o tempo de internamento com uma diferença significativa em relação aos recetores de dadores vivos. Concluiu também o seguinte: que a idade dos recetores e o tempo em diálise são fatores importantes que influenciam a longevidade de sobrevivência do enxerto nos recetores de dadores vivos e que as complicações nefro-urológicas imediatas incluindo função retardada do enxerto e rejeição aguda é mais frequente em pacientes recetores de dadores cadáver; as complicações cardiovasculares durante o tempo de internamento são maiores em recetores de dadores cadáver e que estes pacientes também apresentam mais comorbilidades incluindo uma prevalência significativamente maior de hipertensão; os pacientes recetores de dadores vivos têm menos complicações hematológicas. Este estudo demonstrou que o tipo de dador (vivo ou cadáver) são um forte preditor de complicações infecciosas e que as diferenças encontradas nos dois tipos de recetores, relativamente aos tempos de internamento podem ser justificadas pela elevada percentagem de complicações observadas nos recetores de dadores cadáver.

Nesta tese, usando a totalidade da base de dados que envolve os dois conjuntos (recetores de dadores vivos e de dadores cadáver), a análise exploratória dos dados serviu de mote para todos os estudos que se seguem.

3 - CORRELAÇÕES

O coeficiente de correlação de Spearman, chamado assim devido a Charles Spearman e normalmente denominado pela letra grega ρ (rho), é uma medida de correlação não-paramétrica, isto é, ele avalia uma função monótona arbitrária que pode ser a descrição da relação entre duas variáveis, sem fazer nenhuma suposições sobre a distribuição de frequências das variáveis.

Ao contrário do coeficiente de correlação de Pearson, não requer a suposição que a relação entre as variáveis é linear, nem requer que as variáveis sejam medidas em intervalo de classe; pode ser usado para as variáveis medidas no nível ordinal.

Neste estudo, visto que se desconhece a distribuição das variáveis e recorrendo à correlação de Spearman verificou-se a existência de correlações estatisticamente significativas entre as variáveis e que se apresentam nas tabelas presentes no ANEXO 1 – Correlações.

Por análise das tabelas verificou-se que as complicações Urorenais ($r=0.482$; $p\text{-value}<0.0001$) e a presença de HTA ($r=0.389$; $p\text{-value}<0.0001$) estão significativamente correlacionadas com o tipo de dador, tendo-se apurado que 71,7% dos 113 recetores de dadores cadáver correspondentes a 81 pacientes, tiveram complicações Urorenais e apenas 17,5% (7 pacientes) dos recetores de dadores vivos as tiveram e que 82,3% dos 113 recetores de dadores cadáver (93 pacientes) apresentavam HTA e apenas 42,5% (17 pacientes) dos recetores de dadores vivos apresentavam esta comorbilidade. Quanto à técnica anestésica do dador e a ASA do dador ($r=0.318$; $p\text{-value}<0.05$) verificou-se que dos 40 transplantes efetuados de dadores vivos 57,5% dos casos (23 dadores) usufruíram da técnica anestésica AGIN. A correlação também é estatisticamente significativa entre a ASA do recetor e o Tempo de Internamento ($r=0.206$; $p\text{-value}=0.011$) e as complicações urorenais ($r=0.278$; $p\text{-value}=0.001$) e, entre o sexo e as complicações hematológicas ($r=0.206$; $p\text{-value}=0.011$).

Com respeito às correlações entre as variáveis de saída observou-se que as complicações Cardiovasculares ($r=0.196$; $p\text{-value}=0.015$), complicações Urorenais ($r=0.553$; $p\text{-value}<0.0001$), complicações Infeciosas ($r=0.371$; $p\text{-value}<0.0001$), complicações Hematológicas ($r=0.241$; $p\text{-value}=0.003$) e complicações Endócrinas ($r=0.195$; $p\text{-value}=0.016$) têm uma correlação estatisticamente significativa com o Tempo de Internamento assim como a presença de DAP ($r=0.168$; $P\text{-value}=0.038$), o Tempo em Diálise ($r= 0.408$; $p\text{-value}<0.0001$) e o Tempo de Isquemia ($r=0.470$; $p\text{-value}<0.0001$) também apresentam correlação estatisticamente significativa com o Tempo de Internamento. As reintervenções estão significativamente correlacionadas com as complicações cardiovasculares ($r=0.223$; $p\text{-value}=0.006$). Quanto à existência de complicações pós operatórias estas estão significativamente correlacionadas com a presença de HTA ($r=0.300$; $p\text{-value}<0.0001$) e com a ASA do dador ($r=0.319$; $p\text{-value}=0.045$). Verificou-se também uma correlação estatisticamente significativa entre Tempo em diálise ($r=0.324$; $p\text{-value}<0.0001$), Tempo de Isquemia ($r=0.460$; $p\text{-value}<0.0001$) e complicações Urorenais; Tempo em Diálise ($r=0.303$; $p\text{-value}<0.0001$),

Tempo de Isquemia ($r=0.312$; $p\text{-value}=0.008$) e complicações Endócrinas. Dentro das comorbilidades, verificou-se que as que estão mais significativamente correlacionadas são: hiperuricemia e dislipidemia ($r=0.346$, $p\text{-value}<0.0001$) e que dentro das complicações pós-operatórias as que estão mais significativamente correlacionadas com as Urenais são as hematológicas ($r=0.327$, $p\text{-value}<0.0001$) e as Endócrinas ($r=0.318$, $p\text{-value}<0.0001$). Encontrou-se também uma correlação estatisticamente significativa entre o tipo_Relacao e Diferença >10Anos ($r=0.489$, $p\text{-value}<0.01$) para os recetores de dadores vivos.

De seguida e com recurso ao comando Explore do SPSS, que apresenta maiores opções de estatísticas descritivas, além de dispor de testes estatísticos para normalidade e gráficos especiais e que é utilizado exclusivamente com variáveis quantitativas e permite comparar o tipo de dados entre dois ou mais grupos, comparou-se as médias de idades dos pacientes em função das comorbilidades que apresentam. Assim obtive o gráfico e a tabela que se seguem:

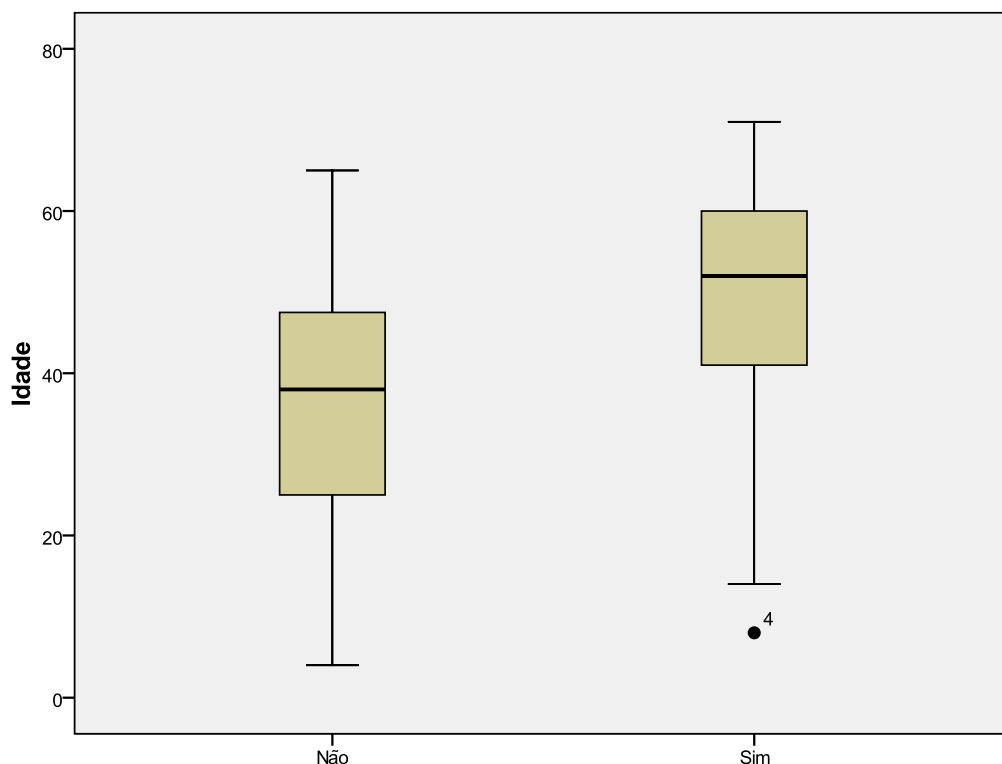


Gráfico 1: Diagrama de extremos e quartis das idades em função da existência ou não de comorbilidades

		Comorbilidades		Statistic	Std. Error
Idade	Não	Mean		36,13	2,817
		95% Confidence Interval for Lower Bound		30,38	
		Mean Upper Bound		41,88	
	Sim	Mean		49,34	1,206
		95% Confidence Interval for Lower Bound		46,96	
		Mean Upper Bound		51,73	

Tabela 3: Média de idades em função da existência ou não de comorbilidades

		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
Comorbilidades		Statistic	df	Sig.	Statistic	df	Sig.
Idade	Não	,074	31	,200	,982	31	,865
	Sim	,100	122	,004	,966	122	,004

Tabela 4: Teste de Normalidade

Pode-se assim verificar que os transplantados que não apresentam comorbilidades têm uma média de idades de 36,13 anos com erro padrão de 2,817 anos, enquanto os que apresentam comorbilidades têm média de idades de 49,34 anos e erro padrão de 1,206, significando que a média de idades dos doentes transplantados e que não apresentam comorbilidades é em média 13 anos mais baixa que a dos que apresentam comorbilidades. Com base na tabela 4 constatou-se que a idade do grupo de pacientes com comorbilidades não segue uma distribuição normal pois um valor abaixo da significância indica um desvio na distribuição dos dados em relação à Curva Normal (de preferência deve estar abaixo de 0,05 ou mesmo 0,01) o que se confirma pelos gráficos que se seguem.

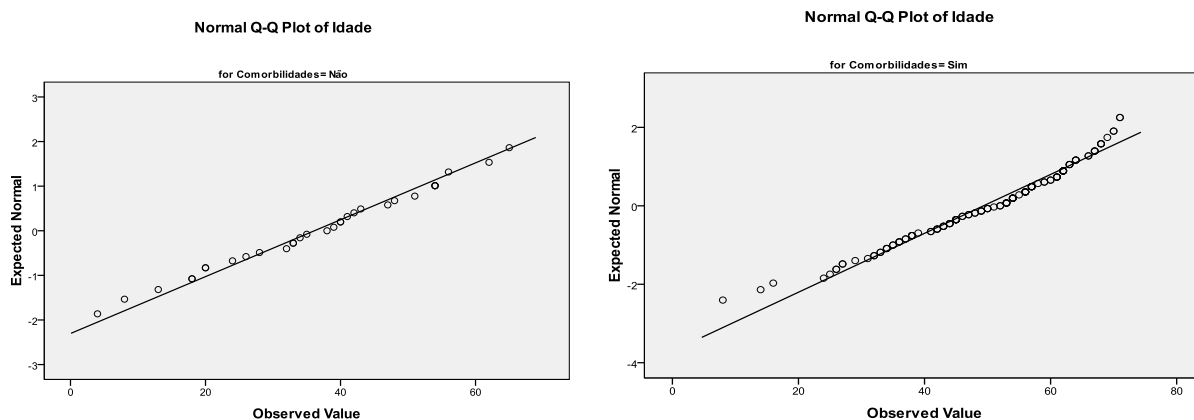


Gráfico 2: Normalidade da idade dos pacientes sem e com comorbilidades

Utilizando o teste não paramétrico U de Mann-Whitney, uma vez que a variável Idade não segue uma distribuição Normal comparou-se com as Comorbilidades dos pacientes (0 – não apresentam Comorbilidades e 1 – apresentam Comorbilidades).

	Comorbilidades	N	Mean Rank	Sum of Ranks
Idade	Não	31	47,76	1480,50
	Sim	122	84,43	10300,50
	Total	153		

Tabela 5: Comparação de Ranking

	Idade
Mann-Whitney U	984,500
Wilcoxon W	1480,500
Z	-4,117
Asymp. Sig. (2-tailed)	,000

Tabela 6: Estatística do Teste U de Mann-Whitney

O teste U de Mann-Whitney cria um *ranking* (ordenação) de todos os casos (independente do grupo) e depois compara estes *ranking* entre cada grupo. Observamos o *rank* médio e a soma dos *rankings* de cada grupo (havendo diferença significativa entre os grupos, espera-se que os *rankings* médios sejam bem diferentes), como se verifica pela tabela 5. Na tabela 6, podemos observar que há diferença estatisticamente significativa entre as idades dos pacientes com Comorbilidades e dos que não apresentam Comorbilidades (p-value<0.0001).

4 - TESTE χ^2 -QUADRADO

Neste estudo, o teste χ^2 serviu para avaliar a existência de associação significativa entre as Reintervenções, o Reinternamento ao 3º mês e as complicações pós-operatórias e entre as complicações pós-operatórias e as diferentes comorbilidades. A razão de possibilidades (odds) permitiu avaliar o grau de associação entre elas.

O intervalo de confiança mostra o limite dentro do qual existe a certeza do verdadeiro efeito entre as variáveis, estima a magnitude da associação e informa a variabilidade da estimativa, através dos limites inferior e superior.

Por aplicação do teste χ^2 e razão de possibilidades para avaliar a existência de associação e o seu grau, entre as reintervenções e as complicações pós operatórias, conclui-se de acordo com as tabelas que se seguem, obtidas no SPSS, que as reintervenções acontecem significativamente para quem tem complicações cardiovasculares. (OR= 3.584; p-value=0.006).

			Cardiovascular		Total
			Não	Sim	
Reintervenções	Não	Contagem	101	31	132
		%	76,5%	23,5%	100%
	Sim	Contagem	10	11	21
		%	47,6%	52,4%	100%
Total		Contagem	111	42	153
		%	72,5%	27,5%	100%

Tabela 7: Relação entre as Reintervenções e as Complicações Cardiovasculares

Variável	Razão de possibilidades	I.C. 95%
Reintervenções	OR = 3.584; p-value=0.006]1.391; 9.232[

Tabela 8: Valor da estimativa da razão de possibilidades e respetivo intervalo de confiança (95%)

Por procedimento análogo se concluiu que mais nenhuma das complicações pós operatórias se relaciona significativamente com as reintervenções. De seguida verificou-se que as complicações pós operatórias acontecem significativamente para quem sofre de HTA (OR= 5.525, p-value<0.0001), sendo que mais nenhuma comorbilidade se relaciona significativamente com as complicações pós operatórias.

			HTA		Total
			Não	Sim	
Complicaçõespósop30diasAlta	Não	Contagem	13	8	21
		%	61,9%	38,1%	100%
	Sim	Contagem	30	102	132
		%	22,7%	77,3%	100%
Total		Contagem	43	110	153
		%	28,1%	71,9%	100%

Tabela 9: Relação entre as Complicações pós operatórias e HTA

Variável	Razão de possibilidades	I.C. 95%
Complicaçõespósop30diasAlta	OR = 5.525; p-value<0.0001]2.094; 14.578[

Tabela 10: Valor da estimativa da razão de possibilidades e respetivo intervalo de confiança (95%)

De igual forma o reinternamento ao 3º mês acontece significativamente nos pacientes que apresentam complicações respiratórias (OR=5.333; p-value=0.02).

			Respiratório		Total
			Não	Sim	
Reinternamento 3º mês	Não	Contagem	128	4	132
		%	97,0%	3,0%	100%
	Sim	Contagem	18	3	21
		%	85,7%	14,3%	100%
Total		Contagem	146	7	153
		%	95,4%	4,6%	100%

Tabela 11: Relação entre as Reinternamento 3º mês e as Complicações Respiratórias

Variável	Razão de possibilidades	I.C. 95%
Reinternamento 3º mês	OR = 5.333; p-value=0.02]1.103; 25.794[

Tabela 12: Valor da estimativa da razão de possibilidades e respetivo intervalo de confiança (95%)

Com base nestes resultados conclui-se que os pacientes que apresentam problemas de HTA têm 5 vezes mais possibilidades de apresentar complicações pós-operatórias. Dentro dos que apresentam complicações pós-operatórias os que têm problemas cardiovasculares têm 3 vezes mais possibilidades de sofrer reintervenções e os que apresentam complicações respiratórias têm 5 vezes mais possibilidades de estar sujeitos a reinternamento ao 3º mês.

5 - REGRESSÃO LINEAR MÚLTIPLA

A análise da regressão linear múltipla foi aplicada à variável dependente “Tempo de Internamento”. A tabela 13 que se segue mostra a análise descritiva das variáveis consideradas nos diferentes modelos estudados.

	Mean	Std. Deviation	N
Tempointernamento	17,01	11,935	153
Idade	46,67	14,772	153
Sexo	,44	,498	153
ASA	3,91	,289	153
Tempoemdiálise	4,7692	4,34137	153
Técnicadialítica	1,16	,488	153
TécnicaAnestésica	,21	,645	153
HTA	,72	,451	153
DCI	,03	,178	153
DAP	,01	,114	153
Tabagismo	,19	,393	153
DM	,10	,307	153
Hiperuricemia	,12	,331	153
Dislipidemia	,18	,382	153
AVC	,01	,114	153
HPT	,12	,331	153
Anemia	,42	,495	153
Cardiovascular	,28	,449	152
Respiratório	,05	,210	152
CNS	,02	,140	152
UroRenal	,57	,496	152
Infeciosas	,63	,486	152
Hematológicas	,45	,499	152
Endocrinas	,19	,394	152

Tabela 13: Análise descritiva das variáveis usadas na Regressão Linear Múltipla

Considerando como variável dependente o Tempo de Internamento e as variáveis independentes as características demográficas do recetor (idade, sexo, ASA), a terapia pré transplante (Tempo em diálise e técnica dialítica) e a técnica anestésica aplicada ao

paciente, a análise da regressão linear múltipla revelou que apenas 11.3% da variabilidade da variável dependente é explicada pelo conjunto das variáveis independentes, como se verifica pela Tabela 14. O coeficiente de correlação ($r=0.336$) revela a relação de intensidade entre os valores observados e os estimados do tempo de internamento. O coeficiente de determinação ajustado (0.076) revela que apenas 7.6% da variabilidade do Tempo de internamento é explicada pelo modelo.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,336 ^a	,113	,076	11,470	2,081

Tabela 14: Sumário do modelo

Como se pode verificar na tabela ANOVA, $p\text{-value}<0.05$, o modelo ajusta-se aos dados, mostrando que pelo uma das variáveis independentes são significativas para o modelo.

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2441,290	6	406,882	3,093	,007 ^a
	Residual	19208,684	146	131,566		
	Total	21649,974	152			

Tabela 15: ANOVA do modelo inicial

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1 (Constant)	-14,863	12,895		-1,153	,251	-40,347	10,622		
Idade	,005	,068	,006	,075	,941	-,128	,139	,869	1,151
Sexo	,923	1,928	,039	,479	,633	-2,887	4,734	,940	1,064
ASA	7,572	3,307	,184	2,290	,023	1,037	14,108	,946	1,057
Tempoemdiálise	,735	,219	,267	3,354	,001	,302	1,168	,956	1,046
Técnicodialítica	-1,435	1,970	-,059	-,728	,468	-5,330	2,459	,935	1,069
TécnicaAnestésica	-,989	1,503	-,053	-,658	,512	-3,960	1,982	,920	1,087

Tabela 16: Significância das variáveis no modelo – Output do SPSS

De acordo com a Tabela 16, podemos concluir com segurança que não existe colinearidade entre as variáveis preditoras, uma vez que a tolerância (Tolerance) é superior a 0,1 e o Fator de Inflação da Variância (VIF) das variáveis preditoras é inferior a 10. Assim, as variáveis preditoras tem uma probabilidade maior de ser responsáveis por uma variância dos resultados relativamente à existência de um único preditor. Aplicando o método Stepwise de seleção de variáveis verifica-se na tabela seguinte que as variáveis significativas para o modelo e que influenciam o tempo de internamento são a ASA (p-value=0,023) e o tempo em diálise (p-value=0,01), confirmado pelas 2 seguintes tabelas:

Model	Variables Entered	Variables Removed	Method
1	Tempoemdiálise	.	Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).
2	ASA	.	Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).

Tabela 17: Processo de seleção das variáveis preditoras - Output do SPSS

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,276 ^a	,076	,070	11,509	
2	,324 ^b	,105	,093	11,367	2,089

Tabela 18: Sumário do modelo com as variáveis preditoras ASA e Tempo em diálise - Output do SPSS

Deste quadro verificamos que o modelo com apenas estas 2 variáveis ASA e Tempo de diálise explica agora 9,3% da variabilidade da variável dependente.

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1647,580	1	1647,580	12,438	,001 ^a
	Residual	20002,394	151	132,466		
	Total	21649,974	152			
2	Regression	2269,096	2	1134,548	8,781	,000 ^b
	Residual	19380,878	150	129,206		
	Total	21649,974	152			

Tabela 19:ANOVA do modelo com as variáveis predictoras - Output do SPSS

Como se pode verificar na tabela ANOVA, o p-value do modelo com as 2 variáveis incluídas é $p\text{-value} < 0.0001$, o que indica que as 2 variáveis são significativas para o modelo. Assim o modelo de regressão obtido é:

Tempo de Internamento = $b_0 + b_1 * \text{tempo em dálise} + b_2 * \text{ASA}$ com $b_0 = -13.937$; $b_1 = 0.761$; $b_2 = 6.990$, como se pode verificar pela tabela dos coeficientes que se segue.

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
	1(Constant)	13,396	1,385				9,674	,000	10,660
Tempo em dálise	,758	,215	,276	3,527	,001	,333	1,183	1,000	1,000
2(Constant)	-13,937	12,537		-1,112	,268	-38,710	10,836		
Tempo em dálise	,761	,212	,277	3,582	,000	,341	1,180	1,000	1,000
ASA	6,990	3,187	,169	2,193	,030	,693	13,288	1,000	1,000

Tabela 20: Tabela de coeficientes - Output do SPSS

Utilizando as capacidades gráficas do SPSS, analisou-se a importância das variáveis predictoras no modelo e obtiveram-se os seguintes gráficos

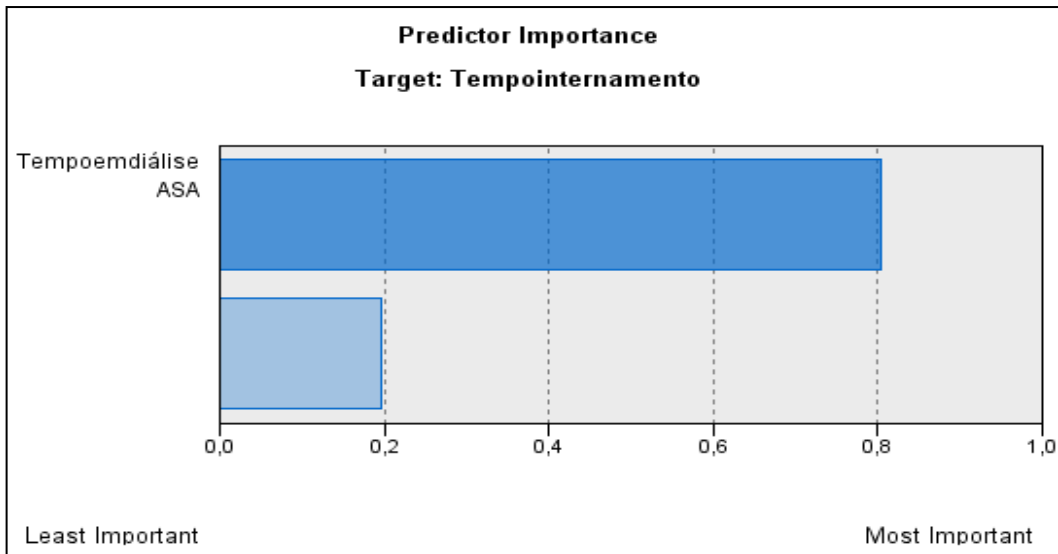


Gráfico 3: Importância dos preditores encontrados - Output do SPSS

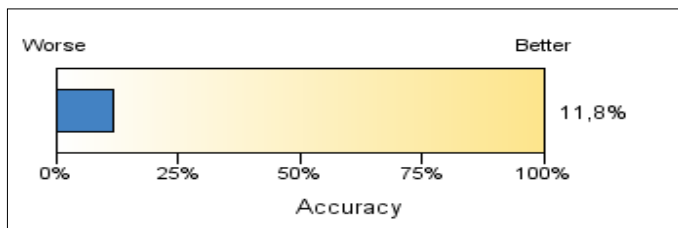


Figura 5: Acurácia do modelo - Output do SPSS

Seguidamente analisou-se a influência das variáveis relativas ao Tempo em diálise, técnica dialítica e os diferentes tipos de comorbilidades no Tempo de Internamento.

Model	Variables Entered	Variables Removed	Method
1	Tempoemdiálise	.	Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).
2	DAP	.	Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).
3	DCI	.	Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).

Tabela 21: Processo de seleção das variáveis preditoras - Output do SPSS

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,276 ^a	,076	,070	11,509	
2	,326 ^b	,106	,094	11,359	
3	,365 ^c	,133	,115	11,224	2,144

Tabela 22: Sumário do modelo com as variáveis preditoras Tempo em diálise, DAP e DCI - Output do SPSS

Deste quadro verificamos que apenas estas 3 variáveis Tempo de diálise, e as comorbilidades DAP e DCI explicam, 13,3% da variabilidade da variável dependente e que o modelo com estas 3 variáveis explica 11,5% da variabilidade do tempo de internamento.

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1647,580	1	1647,580	12,438	,001 ^a
	Residual	20002,394	151	132,466		
	Total	21649,974	152			
2	Regression	2294,920	2	1147,460	8,893	,000 ^b
	Residual	19355,054	150	129,034		
	Total	21649,974	152			
3	Regression	2877,661	3	959,220	7,614	,000 ^c
	Residual	18772,313	149	125,989		
	Total	21649,974	152			

Tabela 23: ANOVA do modelo com as variáveis predictoras - Output do SPSS

Como se pode verificar na tabela ANOVA, p-value do modelo com as 3 variáveis incluídas é $p\text{-value} < 0.0001$, o que indica que as 3 variáveis são significativas para o modelo. Assim o modelo de regressão obtido é:

Tempo de Internamento = $b_0 + b_1 \cdot \text{tempo em dálise} + b_2 \cdot \text{DAP} + b_3 \cdot \text{DCI}$ com $b_0 = 13.154$; $b_1 = 0.683$; $b_2 = 18.641$; $b_3 = 11.021$.

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
	1 (Constant)	13,396	1,385				9,674	,000	10,660
Tempo em dálise	,758	,215	,276	3,527	,001	,333	1,183	1,000	1,000
2 (Constant)	13,332	1,367		9,752	,000	10,631	16,033		
Tempo em dálise	,722	,213	,263	3,393	,001	,302	1,143	,994	1,006
DAP	18,162	8,109	,173	2,240	,027	2,140	34,184	,994	1,006
3 (Constant)	13,154	1,353		9,720	,000	10,480	15,828		
Tempo em dálise	,683	,211	,248	3,233	,002	,265	1,100	,987	1,013
DAP	18,641	8,016	,178	2,326	,021	2,803	34,480	,993	1,007
DCI	11,021	5,124	,165	2,151	,033	,895	21,147	,992	1,008

Tabela 24: Tabela de coeficientes - Output do SPSS

Este modelo mostra que há 3 fatores (Tempo em diálise, DAP e DCI) que influenciam com significância estatística a duração do tempo de internamento.

De novo utilizando as capacidades gráficas do SPSS, analisou-se a importância das variáveis preditoras no modelo e obtiveram-se os seguintes gráficos

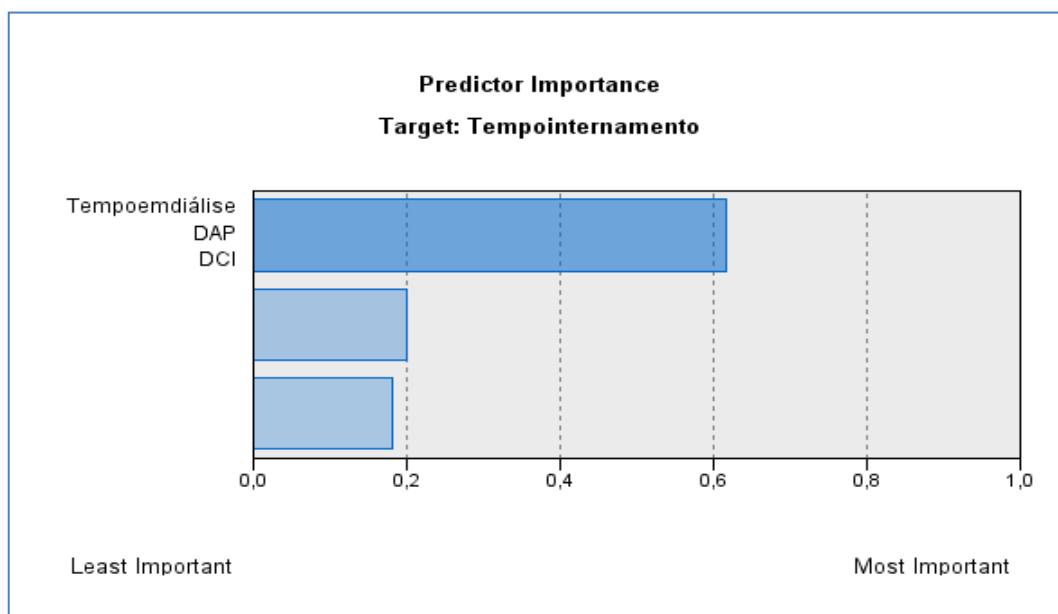


Gráfico 4: Importância dos preditores - Output do SPSS

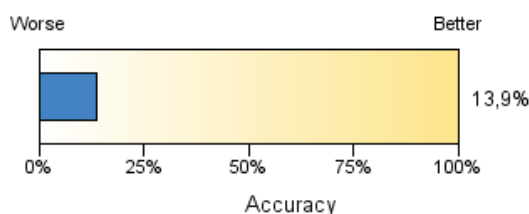


Figura 6: Acurácia do modelo - Output do SPSS

Por fim analisou-se a influência das variáveis relativas ao Tempo em diálise, técnica dialítica e as diferentes complicações pós operatórias no Tempo de Internamento.

Model	Variables Entered	Variables Removed	Method
1	UroRenal	.	Stepwise (Criteria: Probability-of-F- to-enter <= ,050, Probability-of-F- to-remove >= ,100).
2	Infeciosas	.	Stepwise (Criteria: Probability-of-F- to-enter <= ,050, Probability-of-F- to-remove >= ,100).

Tabela 25: Processo de seleção das variáveis predictoras - Output do SPSS

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,447 ^a	,200	,194	10,657	
2	,497 ^b	,247	,237	10,371	1,880

Tabela 26: Sumário do modelo com as variáveis predictoras Urorenal e Infeciosas - Output do SPSS

Deste quadro verificamos que as variáveis Urorenal e Infeciosas, explicam 23,7% da variabilidade da variável dependente.

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4251,951	1	4251,951	37,440	,000 ^a
	Residual	17035,148	150	113,568		
	Total	21287,099	151			
2	Regression	5259,641	2	2629,820	24,448	,000 ^b
	Residual	16027,458	149	107,567		
	Total	21287,099	151			

Tabela 27: Anova do modelo com as variáveis predictoras - Output do SPSS

Como se pode verificar na tabela ANOVA, p-value do modelo com as 2 variáveis incluídas é p-value < 0.0001 , o que indica que as 2 variáveis são significativas para o modelo.

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
	1 (Constant)	10,769	1,322				8,147	,000	8,157
UroRenal	10,691	1,747	,447	6,119	,000	7,238	14,143	1,000	1,000
2 (Constant)	8,139	1,547		5,261	,000	5,082	11,196		
UroRenal	9,264	1,763	,387	5,254	,000	5,780	12,748	,930	1,075
Infeciosas	5,515	1,802	,226	3,061	,003	1,954	9,075	,930	1,075

Tabela 28: Tabela dos coeficientes das variáveis preditoras - Output do SPSS

Assim o modelo de regressão obtido é:

Tempo de Internamento = $b_0 + b_1 * \text{Urorenal} + b_2 * \text{Infeciosas}$ com $b_0 = 8.139$; $b_1 = 9.264$ e $b_2 = 5.515$.

As complicações Urorenais e Infeciosas são preditoras estatisticamente significativas do tempo de internamento.

De novo, utilizando as capacidades gráficas do SPSS, analisou-se a importância das variáveis preditoras no modelo e obtiveram-se os seguintes gráficos

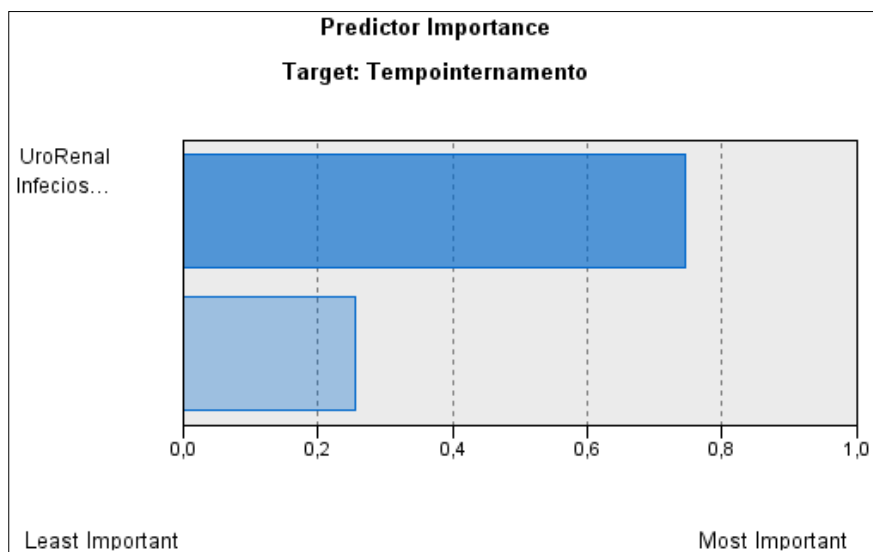


Gráfico 5: Importância dos preditores - Output do SPSS

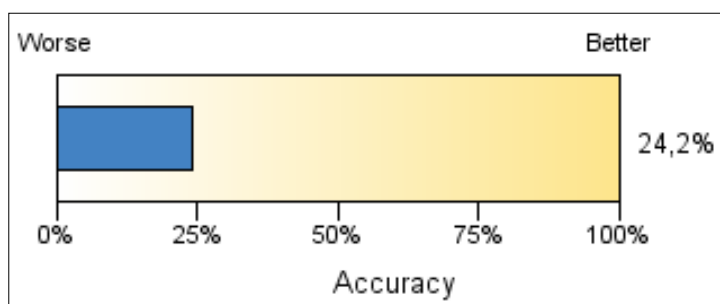


Figura 7: Acurácia do modelo - Output do SPSS

Da análise da regressão linear múltipla conclui-se que as variáveis pré transplante que são estatisticamente significativas na previsão do tempo de internamento são: ASA, T_díálise, DAP, DCI e as complicações pós operatórias que são estatisticamente significativas na previsão do tempo de internamento são as complicações Urorenais e Infeciosas.

6 - REGRESSÃO LOGÍSTICA

1 - Variável dependente: Complicações-pos-op30diasAlta

Variáveis Independentes: dador, idade; sexo; ASA; Tempo em diálise; técnica dialítica; Etiologia DRC; Tipos de Comorbilidades: HTA, DCI, DAP, Tabagismo, DM, Hiperuricemia, Dislipidemia, AVC, HPT, Anemia.

Modelo 1.1. - Para avaliar a significância das variáveis independentes sobre a probabilidade de ter complicações pós operatórias até 30 dias após ou até à alta, recorreu-se numa primeira análise à regressão logística pelo método FORWARD STEPWISE. Foram efetuados os procedimentos de validação dos pressupostos por intermédio da análise gráfica dos resíduos (Gráfico 6) e da tabela onde é visível que existem 6 observações candidatas a outlier $|r_j| > 2$ e ao diagnóstico de casos influentes (Gráfico 7) tendo-se observado que há 6 pontos com influência na qualidade do modelo e que nenhum apresenta distância de Cook superior a 1 pelo que nenhuma das observações tem grande influência sobre os coeficientes do modelo.

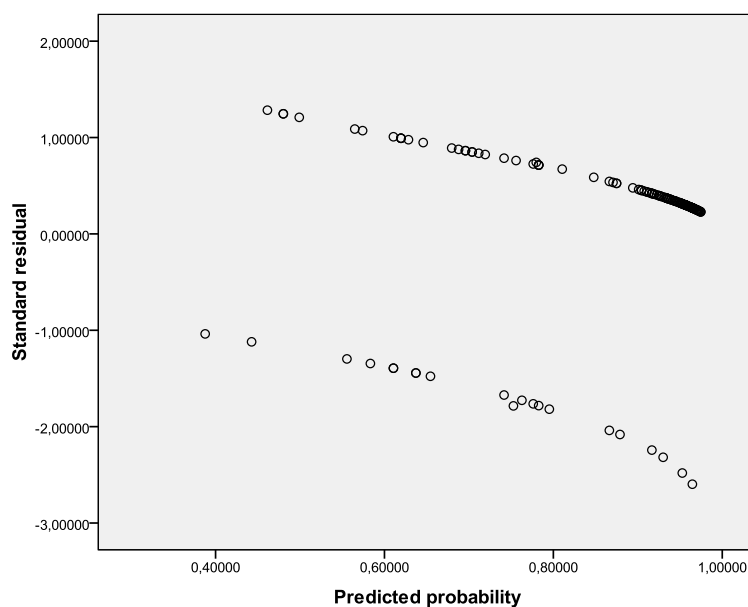


Gráfico 6: Análise gráfica dos resíduos – Output do SPSS

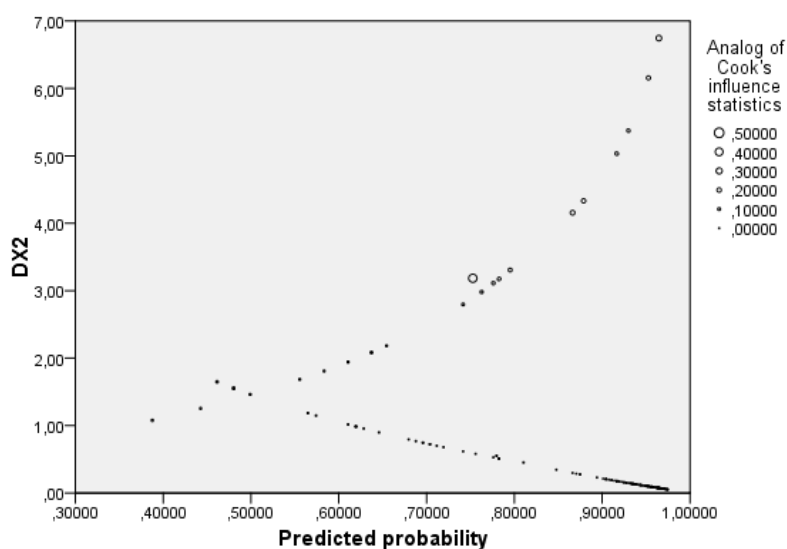


Gráfico 7: Diagnóstico de observações influentes – Output do SPSS

Case	Selected Status ^a	Observed	Predicted	Predicted Group	Temporary Variable	
		Complicação spósop30dia sALTA			Resid	ZResid
64	S	N**	,930	S	-,930	-3,645
74	S	N**	,879	S	-,879	-2,694
83	S	N**	,953	S	-,953	-4,487
92	S	N**	,965	S	-,965	-5,219
121	S	N**	,917	S	-,917	-3,316
124	S	N**	,866	S	-,866	-2,546

Tabela 29: Tabela de observações candidatas a outlier – Output do SPSS

A tabela que se segue “Omnibus Test” mostra que no passo 2, pelo menos uma das variáveis independentes é significativa para o modelo.

		Chi-square	df	Sig.
Step 1	Step	18,091	1	,000
	Block	18,091	1	,000
	Model	18,091	1	,000
Step 2	Step	4,219	1	,040
	Block	22,309	2	,000
	Model	22,309	2	,000

Tabela 30: Omnibus Test – Output do SPSS

A maior significância do modelo ocorre no 2º passo como se verifica pelo quadro seguinte:

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	104,294 ^a	,112	,203
2	100,075 ^a	,136	,246

Tabela 31: Sumário do modelo – Output do SPSS

De seguida o Hosmer Test mostrou que o modelo se ajusta bem aos dados

Step	Chi-square	df	Sig.
1	,000	0	.
2	7,732	8	,460

Tabela 32: Hosmer Test – Output do SPSS

De acordo com a seguinte tabela, podemos verificar que a percentagem de casos corretamente classificados pelo modelo é de 85% com uma especificidade de 9,5% e sensibilidade de 97%.

Observed		Predicted		Percentage Correct
		Complicaçõespósop30diasALTA		
		Não	Sim	
Step 1	Complicaçõespósop30diasA Não	0	21	,0
	LTA Sim	0	132	100,0
	Overall Percentage			86,3
Step 2	Complicaçõespósop30diasA Não	2	19	9,5
	LTA Sim	4	128	97,0
	Overall Percentage			85,0

Tabela 33: Tabela de classificação do modelo – Output do SPSS

A regressão logística revelou que as variáveis, dador (1) e Idade apresentam um efeito estatisticamente significativo sobre a probabilidade de ter complicações pós operatórias de acordo com modelo ajustado.

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	Dador(1)	-2,098	,512	16,796	1	,000	,123	,045	,335
	Constant	2,718	,390	48,492	1	,000	15,143		
Step 2 ^b	Dador(1)	-1,721	,537	10,285	1	,001	,179	,062	,512
	Idade	,038	,018	4,264	1	,039	1,039	1,002	1,076
	Constant	,962	,877	1,204	1	,273	2,617		

Tabela 34: resumo dos coeficientes do modelo e a sua significância – Output do SPSS

A função de probabilidade de ocorrer complicações pós operatórias em função do dador e da idade está ilustrada no gráfico seguinte:

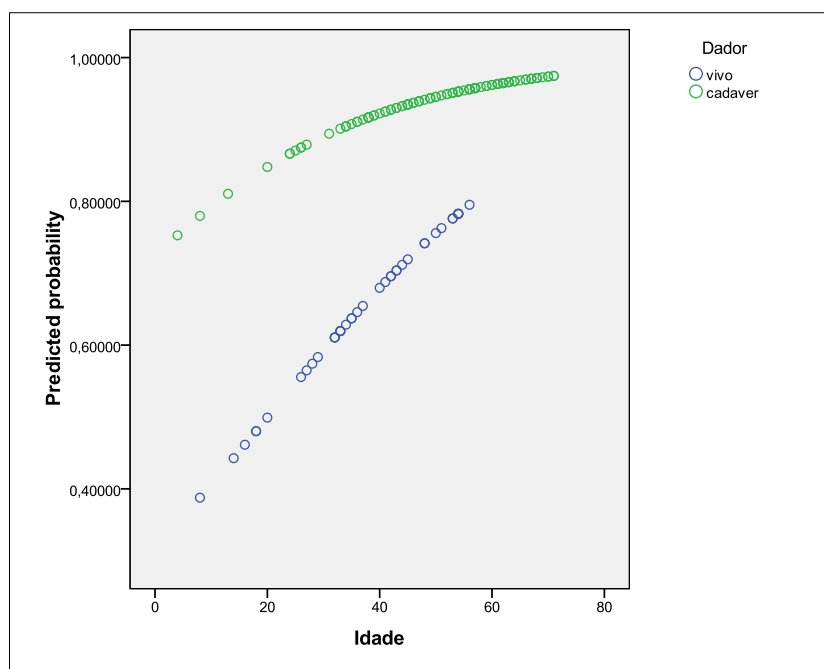


Gráfico 8: Gráfico de probabilidades – Output do SPSS

Assim a probabilidade de ter complicações pós operatórias(1) é significativamente mais elevada para os pacientes recetores de dadores cadáver, pois como se pode verificar pelo gráfico, pacientes com a mesma idade têm mais probabilidade de ter complicações pós operatórias se forem recetores de dadores cadáver e que essa probabilidade aumenta

significativamente com a idade sendo esse aumento mais significativo para os pacientes recetores de dadores vivos.

Assim sendo as variáveis predictoras de complicações operatórias são o *dador(1)* e a idade, donde se conclui que a probabilidade de ter complicações operatórias é dada pelo modelo de equação

$$\text{Logit}(\hat{p}) = 0,962 + (-1,721) * \text{Dador}(1) + 0,038 * \text{Idade}$$

Modelo 1.2. - Construiu-se de seguida um segundo modelo, mantendo o mesmo objetivo mas recorrendo ao BACKWARD STEPWISE. Assim sendo, as variáveis predictoras de complicações pós operatórias são o *dador(1)*, a *ASA* e a presença da comorbilidade *HTA(1)*, donde se conclui que a probabilidade de ter complicações operatórias é dada pelo modelo de equação

$$\text{Logit}(\hat{p}) = 0.370 + (-1.461) * \text{Dador}(1) + (-1.995) * \text{ASA} + (-1.294) * \text{HTA}(1)$$

Ambos os modelos apresentam uma boa capacidade discriminante como se pode verificar pelas respetivas curvas ROC:

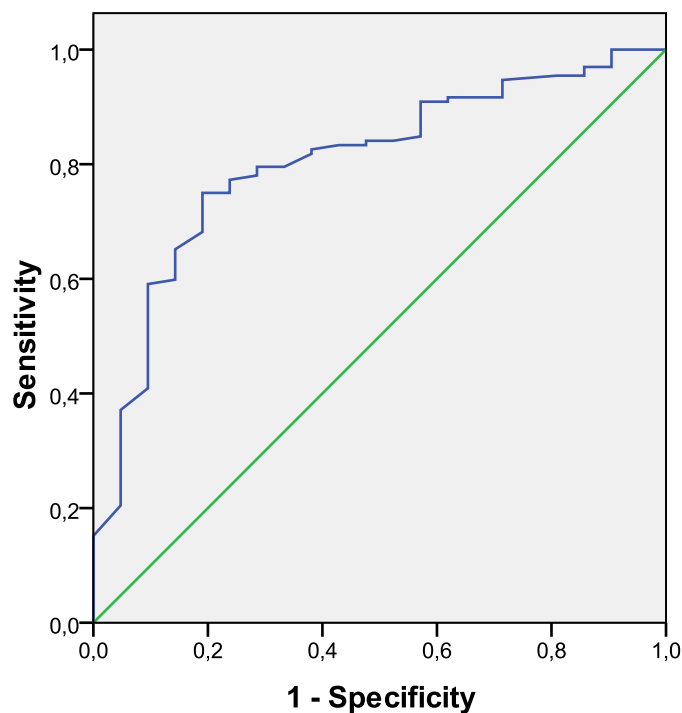


Gráfico 9: Curva Roc do Modelo 1.1 – Output do SPSS

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,804	,049	,000	,708	,899

Tabela 35: Área da Curva Roc e respectivo Intervalo de Confiança – Output do SPSS

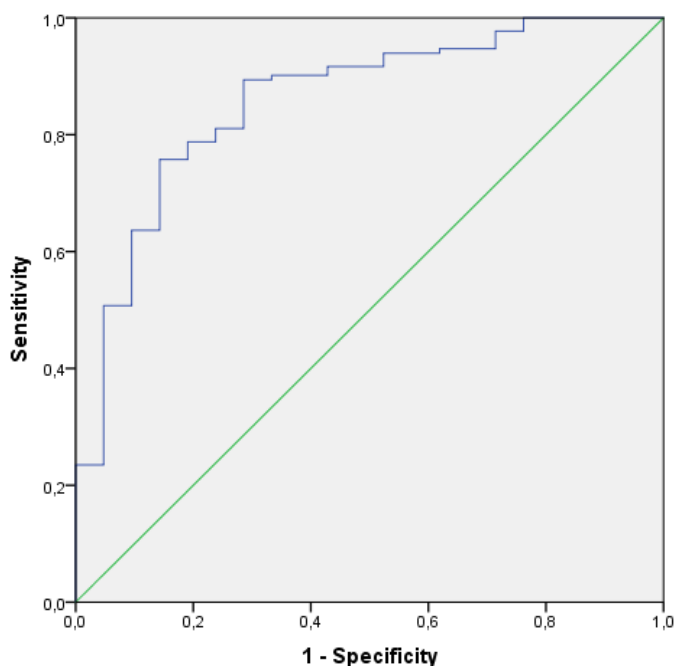


Gráfico 10: Curva Roc do Modelo 1.2 – Output do SPSS

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,858	,044	,000	,771	,944

Tabela 36: Área da Curva Roc e respectivo Intervalo de Confiança – Output do SPSS

2 – Variável dependente: Complicações-pos-op30diasAlta

Retirando as variáveis correspondentes a características demográficas construíram-se mais dois modelos usando os mesmos dois métodos aplicados anteriormente, Forward e Backward.

Varáveis Independentes: Tempo em diálise; técnica dialítica; Etiologia DRC; Tipos de Comorbilidades: HTA, DCI, DAP, Tabagismo, DM, Hiperuricemia, Dislipidemia, AVC, HPT, Anemia.

Modelo 2.1. - Para avaliar a significância das variáveis independentes sobre a probabilidade de ter complicações pós operatórias até 30 dias após ou até à alta, recorreu-se à regressão logística pelo método FORWARD:LR como descrito em Maroco(2011). Foram efetuados os procedimentos de validação dos pressupostos por intermédio da análise gráfica dos resíduos (Gráfico) onde é visível que existem 7 observações candidatas a outlier $|r_j| > 2$ e ao diagnóstico de casos influentes tendo-se observado que há 7 pontos com influência na qualidade do modelo e que nenhum apresenta distância de Cook superior a 1 pelo que nenhuma das observações tem influência sobre os coeficientes do modelo.

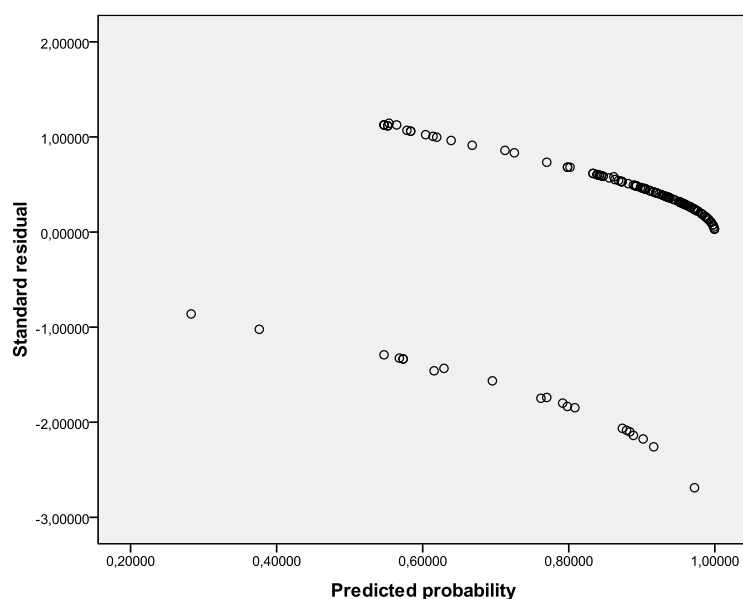


Gráfico 11: Análise gráfica dos resíduos – Output do SPSS

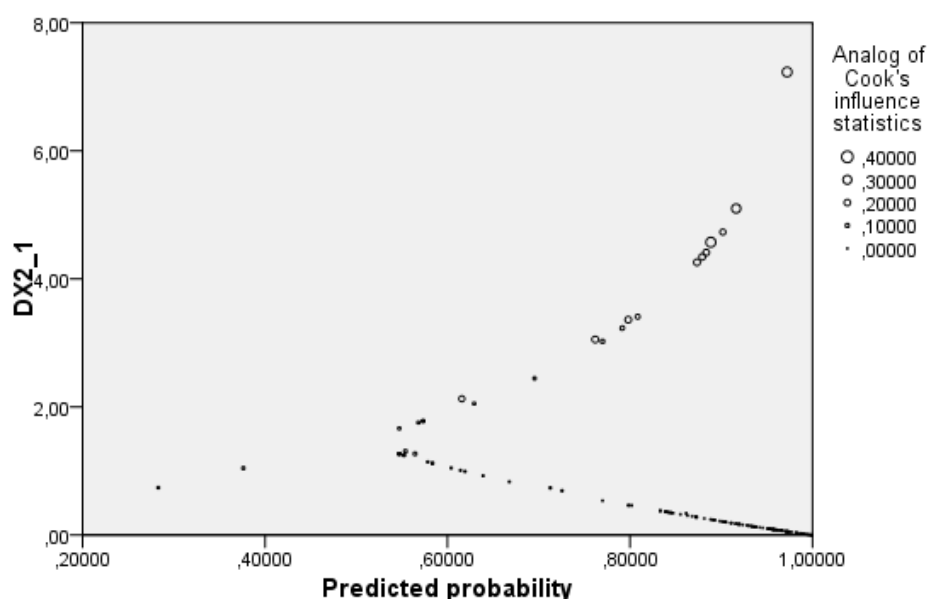


Gráfico 12: Diagnóstico de observações influentes – Output do SPSS

A tabela que se segue Omnibus Test mostra que pelo menos uma das variáveis independentes é significativa para o modelo.

		Chi-square	df	Sig.
Step 1	Step	12,342	1	,000
	Block	12,342	1	,000
	Model	12,342	1	,000
Step 2	Step	7,475	1	,006
	Block	19,816	2	,000
	Model	19,816	2	,000
Step 3	Step	3,612	1	,057
	Block	23,429	3	,000
	Model	23,429	3	,000

Tabela 37: Omnibus Test – Output do SPSS

A maior significância do modelo ocorre no 3º passo como se verifica pelo quadro seguinte:

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	110,043 ^a	,077	,141
2	102,568 ^b	,121	,221
3	98,956 ^b	,142	,258

Tabela 38: Sumário do modelo – Output do SPSS

De seguida o Hosmer Test mostrou que o modelo se ajusta bem aos dados

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	,000	0	.
2	11,982	8	,152
3	8,931	8	,348

Tabela 39: Hosmer Test – Output do SPSS

De acordo com a seguinte tabela, podemos verificar que a percentagem de casos corretamente classificados pelo modelo é de 87,6% com uma especificidade de 9,5% e sensibilidade de 100%.

Observed		Predicted		
		Complicaçõespósop30diasALTA		Percentage Correct
		Não	Sim	
Step 1	Complicaçõespósop30diasA Não	0	21	,0
	LTA Sim	0	132	100,0
	Overall Percentage			86,3
Step 2	Complicaçõespósop30diasA Não	0	21	,0
	LTA Sim	0	132	100,0
	Overall Percentage			86,3
Step 3	Complicaçõespósop30diasA Não	2	19	9,5
	LTA Sim	0	132	100,0
	Overall Percentage			87,6

Tabela 40: Tabela de classificação do modelo – Output do SPSS

A regressão logística revelou que as variáveis, Tempo em diálise, HTA(1) e Tabagismo(1) apresentam um efeito estatisticamente significativo sobre a probabilidade de ter complicações pós operatórias de acordo com modelo ajustado (Tabela – resume os coeficientes do modelo e a sua significância)

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Step 1 ^a HTA(1)	-1,709	,495	11,922	1	,001	,181	,069	,478
Constant	2,546	,367	48,068	1	,000	12,750		
Step 2 ^b Tempoemdiálise	,223	,097	5,309	1	,021	1,250	1,034	1,511
HTA(1)	-1,586	,512	9,594	1	,002	,205	,075	,559
Constant	1,679	,481	12,207	1	,000	5,359		
Step 3 ^c Tempoemdiálise	,255	,104	5,949	1	,015	1,290	1,051	1,583
HTA(1)	-1,732	,535	10,499	1	,001	,177	,062	,504
Tabagismo(1)	1,204	,620	3,772	1	,052	3,333	,989	11,229
Constant	,717	,674	1,134	1	,287	2,049		

Tabela 41: resumo dos coeficientes do modelo e a sua significância

Apesar da significância para o Tabagismo ser 0,052 o OR é 3,333 o que indicia que o Tabagismo é um fator de risco para a ocorrência de complicações pós operatórias.

A função de probabilidade de ocorrer complicações pós operatórias em função do tempo em diálise, HTA e Tabagismo está ilustrada no gráfico seguinte:

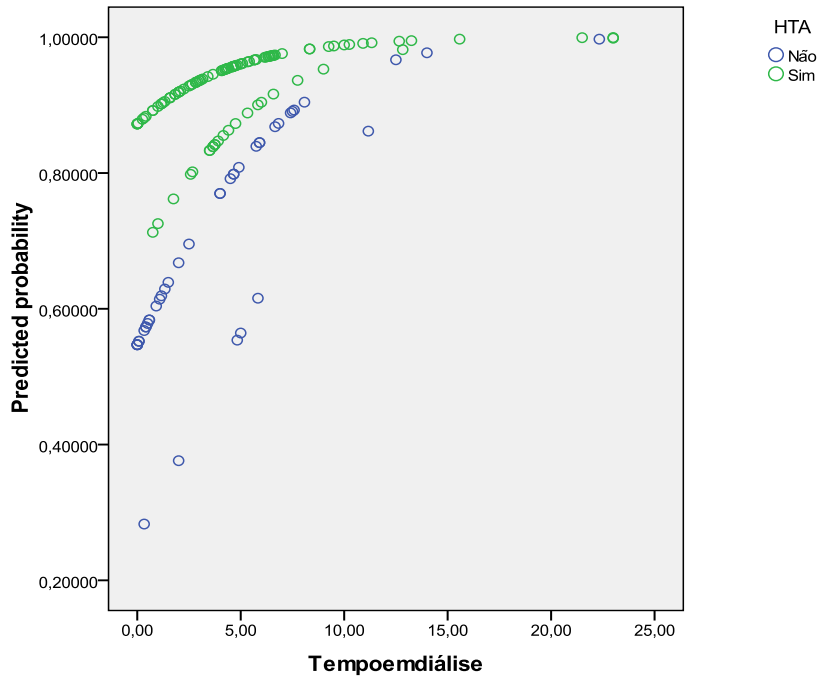


Gráfico 13: Gráfico de probabilidades – Output do SPSS

Assim a probabilidade de ter complicações pós operatórias(1) é superior nos pacientes que apresentam hábitos tabágicos comparativamente com os que não o têm, quer seja na presença de HTA ou na sua ausência. Também se verifica que a probabilidade de ocorrer complicações pós operatórias cresce significativamente com o aumento do tempo em diálise e que ela é superior para que sofre de HTA pois para recetores com o mesmo tempo de diálise a probabilidade é sempre mais elevada para os recetores que sofrem de HTA. O modelo apresenta ainda uma boa capacidade discriminante como se pode verificar pela curva ROC do modelo (ROC c=0,808: p-value<0,0001).

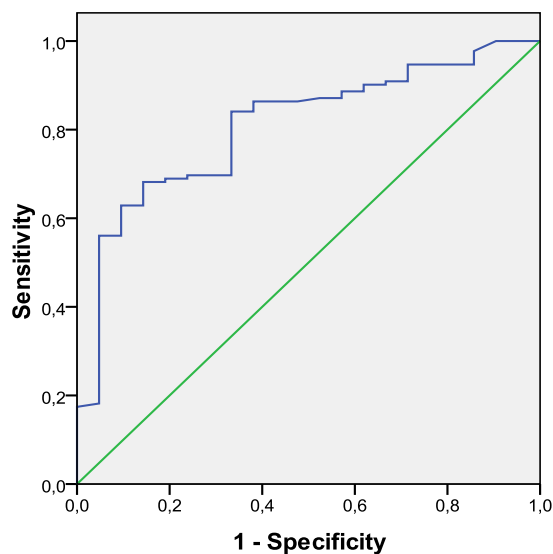


Gráfico 14: Curva Roc do Modelo 2.1. – Output do SPSS

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,808	,047	,000	,716	,899

Tabela 42: Área da Curva Roc e respetivo Intervalo de Confiança – Output do SPSS

Assim sendo as variáveis predictoras de complicações operatórias são o Tempo em diálise, a presença de HTA(1) e o Tabagismo(1) donde se conclui que a probabilidade de ter complicações pós operatórias é dada pelo modelo de equação

$$\text{Logit}(\hat{p}) = 0,717 + (-1,732) * HTA(1) + (0,255) * T_{Diálise} + 1,204 * Tabagismo(1)$$

$$\hat{p} = \frac{1}{1 + e^{-(0,717 + (-1,732) * HTA(1) + 0,255 * T_{Diálise} + 1,204 * Tabagismo(1))}}$$

Modelo 2.2. - Construiu-se de seguida um segundo modelo, mantendo o mesmo objetivo mas recorrendo ao BACKWARD STEPWISE. Assim sendo, as variáveis predictoras de complicações pós operatórias são o T_Diálise e a presença das comorbilidade HTA(1) e Tabagismo(1), donde se conclui que a probabilidade de ter complicações pós operatórias é dada pelo modelo de equação

$$\text{Logit}(\hat{p}) = 0.717 + 0.255 * T_{Diálise} + (-1.732) * HTA(1) + 1.204 * Tabagismo(1)$$

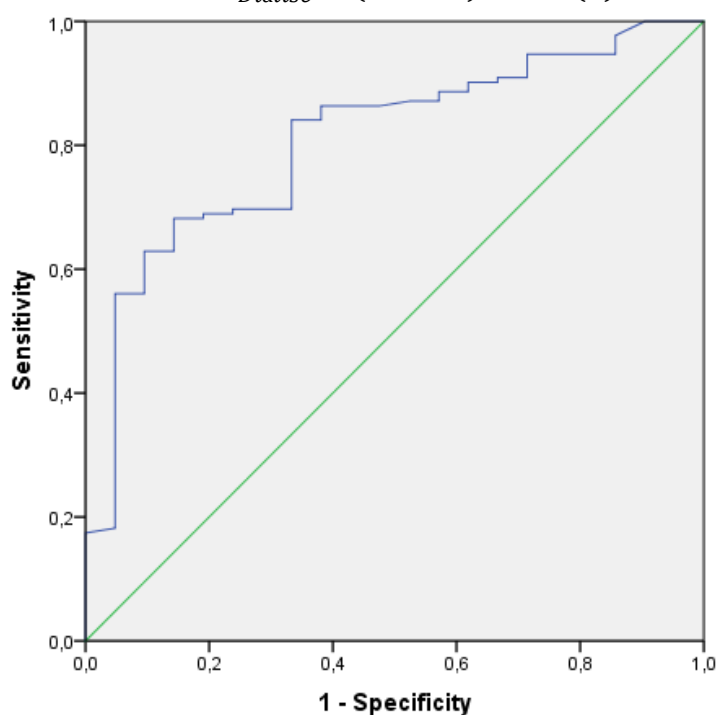


Gráfico 15: Curva Roc do Modelo 2.2. – Output do SPSS

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,808	,047	,000	,716	,899

Tabela 43: Área da Curva Roc e respetivo Intervalo de Confiança – Output do SPSS

Como se verifica pela curva ROC e respetiva área os dois modelos apresentam igual capacidade discriminante.

3 – Variável dependente Reintervensões

Variáveis Independentes: Complicações pós operatórias – Cardiovascular, Respiratório, CNS, Urorenal, Infeciosas, Hematológicas e Endócrinas.

Modelo 3.1. - Seguindo os mesmos procedimentos efetuados para a variável dependente Complicações pós operatórias, obtiveram-se os seguintes resultados.

Step	Chi-square	df	Sig.
1	,000	0	.
2	,545	2	,761
3	,883	5	,971

Tabela 44: Hosmer Test – Output do SPSS

O modelo com as variáveis independentes Cardiovascular. Urorenais e Endócrinas ajusta-se perfeitamente aos dados ao fim do 3º passo com o p-value=0,971 no teste de Hosmer.

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Step 1 ^a Cardiovascular(1)	-1,266	,483	6,879	1	,009	,282	,109	,726
Constant	-1,036	,351	8,716	1	,003	,355		
Step 2 ^b Cardiovascular(1)	-1,389	,495	7,876	1	,005	,249	,094	,658
Endocrinas(1)	1,898	1,062	3,193	1	,074	6,669	,832	53,453
Constant	-2,638	1,040	6,431	1	,011	,072		
Step 3 ^c Cardiovascular(1)	-1,183	,509	5,399	1	,020	,306	,113	,831
UroRenal(1)	-1,101	,568	3,760	1	,052	,333	,109	1,012
Endocrinas(1)	2,200	1,069	4,239	1	,040	9,026	1,111	73,301
Constant	-2,641	1,044	6,403	1	,011	,071		

Tabela 45: Resumo dos coeficientes do modelo – Output do SPSS

O modelo classifica corretamente 86,2% dos casos, com 100% de especificidade e 0% de sensibilidade.

As variáveis significativas para o modelo são Cardiovasculares, Endócrinas e urogenitais.

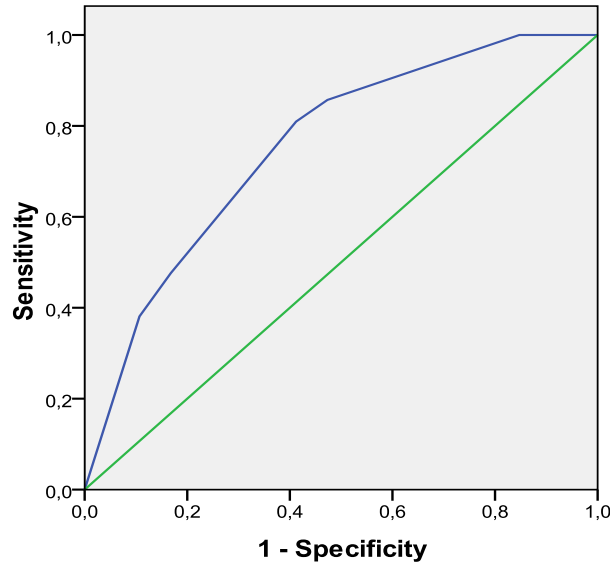


Gráfico 16: Curva Roc do Modelo

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,754	,052	,000	,652	,857

Tabela 46: Área da Curva Roc e respetivo Intervalo de Confiança – Output do SPSS

O modelo apresenta ainda uma capacidade discriminante boa como se pode verificar pela curva ROC do modelo (ROC c=0,754; p-value<0,0001).

O método Backward Stepwise gerou um modelo com as mesmas variáveis preditoras e com a mesma capacidade discriminante.

4 - Variável dependente: Cada uma das Complicações-pos-op30diasAlta

Realizei análise logística considerando como variáveis dependentes cada uma das complicações pós operatórias e variáveis independentes Tempo em diálise, Tempo de Isquemia e as diferentes comorbilidades e obtiveram-se os seguintes resultados:

Modelo 4.1. - Variáveis preditoras de complicações Urogenitais: Tempo de Isquemia

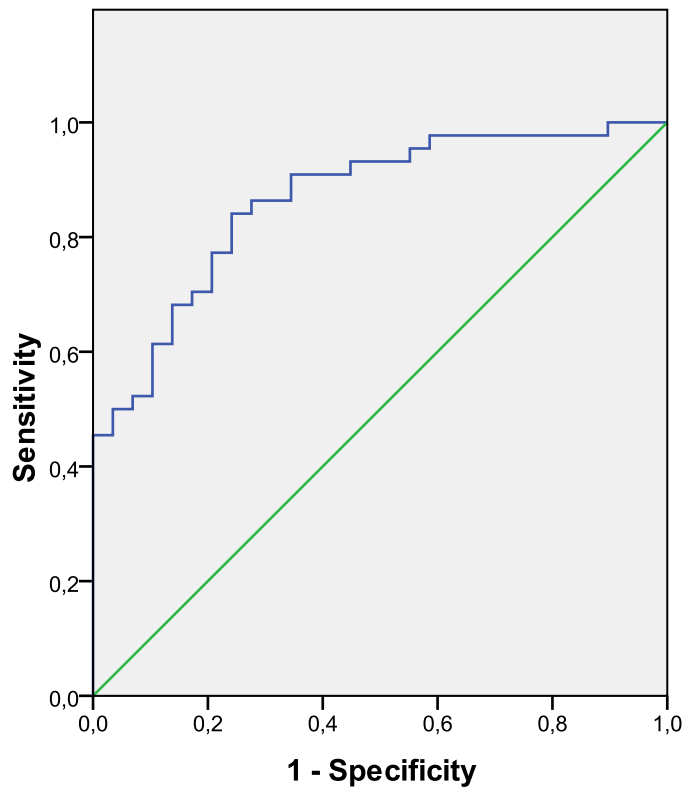


Gráfico 17: Curva Roc do Modelo 4.1. – Output do SPSS

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,865	,042	,000	,783	,947

Tabela 47: Área da Curva Roc e respetivo Intervalo de Confiança – Output do SPSS

Modelo 4.2. - Não há variáveis predictoras com coeficientes significativos na previsão de complicações hematológicas.

Modelo 4.3. - Variáveis predictoras de complicações Cardiovasculares são a presença de HTA, a Anemia e o tempo de Isquemia apresentando este modelo uma boa capacidade discriminante.

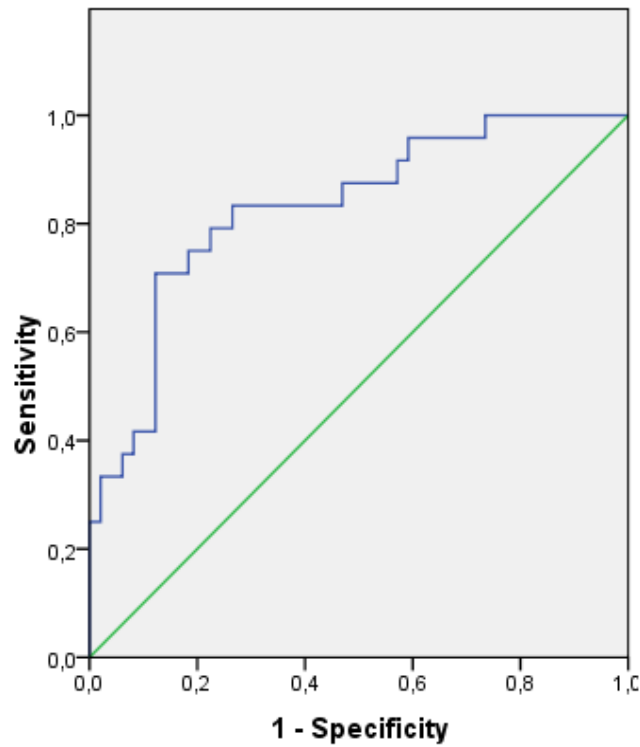


Gráfico 18: Curva Roc do Modelo 4.3. – Output do SPSS

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,830	,051	,000	,730	,930

Tabela 48: Área da Curva Roc e respetivo Intervalo de Confiança – Output do SPSS

7 - ANÁLISE DE CLUSTERS

7.1 - Para o tempo de Internamento

Permitiu encontrar 3 clusters diferenciados por médias de tempos de internamento:

Cluster 1 a média do tempo de internamento é 10,82 dias e 104 pacientes; o cluster 2 com média do tempo de internamento 25,62 dias e 42 pacientes e o cluster 3 com média igual a 57,43 dias e com 7 pacientes.

Cluster Number of Case	Mean	N	Std. Deviation
1	10,82	104	3,106
2	25,62	42	5,691
3	57,43	7	11,487
Total	17,01	153	11,935

Tabela 49: Média dos tempos de internamento e nº de casos por cluster

A tabela ANOVA de comparação das médias que se segue garante que há diferenças significativas entre os 3 grupos com p-value <0.0001.

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Tempointernamento	9268,413	2	20,754	150	446,578	,000

Tabela 50: ANOVA de comparação das médias dos cluster

7.1.1 - Análise Discriminante

A análise da tabela “Teste de igualdade de médias de grupo” permite concluir que considerando as variáveis individualmente apenas as variáveis “Complicações Cardiovasculares”, Respiratórias” e “CNS” não são significativas na diferenciação entre grupos.

	Wilks' Lambda	F	df1	df2	Sig.
Cardiovascular	,967	2,563	2	149	,080
Respiratório	,992	,573	2	149	,565
CNS	,991	,699	2	149	,499
UroRenal	,800	18,588	2	149	,000
Infeciosas	,892	9,049	2	149	,000
Hematologicas	,925	6,002	2	149	,003
Endocrinas	,955	3,492	2	149	,033

Tabela 51: Teste de igualdade de médias de grupo

A análise gerou 2 funções sendo que a primeira explica 89,8% da variância dos dados e a segunda explica 10,2% e que, o teste é significativo para ambas as funções, como se pode observar pelas tabelas que se seguem:

Função	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	,380 ^a	89,8	89,8	,525
2	,043 ^a	10,2	100,0	,204

Tabela 52: Funções discriminantes

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	,695	53,176	14	,000
2	,958	6,194	6	,402

Tabela 53: Teste de significância das funções discriminantes

A tabela seguinte mostra que as variáveis foram agrupadas da seguinte maneira: Contribuem com maior peso para a 1ª função as complicações Urorenais, Infeciosas, Hematológicas, Cardiovasculares, e CNS e para a 2ª função as variáveis Endócrinas e respiratórias.

	Function	
	1	2
UroRenal	,808*	,190
Infeciosas	,565*	-,043
Hematologicas	,459*	,097
Cardiovascular	,300*	,085
CNS	-,154*	-,087
Endocrinas	,176	,900*
Respiratório	,064	,376*

Tabela 54: Matriz de estrutura das funções

Por análise do mapa territorial o cluster 3 tem valores mais elevados para a função 1 e o cluster 1 tem os valores mais baixos. Isto significa que os pacientes com tempos de internamento mais elevados são os que apresentam em maior número as complicações Urorenais e Infeciosas e os pacientes com tempos de internamento mais baixos apresentam em menor número estas complicações. Em suma estes resultados corroboram os resultados obtidos na análise da regressão linear múltipla para a variável dependente Tempo de Internamento.

7.2 - Para o tempo de Diálise

Permitiu encontrar 2 cluster diferenciados por médias de tempos de diálise:

Cluster 1 a média do tempo de diálise é 14 anos com 18 pacientes; o cluster 2 com média do tempo de diálise 3,538 anos com 135 pacientes.

Cluster Number of Case	Mean	N	Std. Deviation
1	14,0031	18	4,96275
2	3,5380	135	2,29651
Total	4,7692	153	4,34137

Tabela 55: Média dos tempos de internamento e nº de casos por cluster

A tabela ANOVA de comparação das médias que se segue garante que há diferenças significativas entre os 2 grupos com p-value <0.0001.

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Tempoemdiálise	1739,412	1	7,453	151	233,385	,000

Tabela 56: ANOVA de comparação das médias dos cluster

7.2.1 - Análise Discriminante

A análise da tabela “Teste de igualdade de médias de grupo” permite concluir que considerando as variáveis individualmente apenas as variáveis “Complicações Urorenais” e “Hematológicas” são significativas na diferenciação entre grupos.

	Wilks' Lambda	F	df1	df2	Sig.
Cardiovascular	1,000	,000	1	150	,988
Respiratório	,994	,979	1	150	,324
CNS	,997	,407	1	150	,525
UroRenal	,963	5,824	1	150	,017
Infeciosas	,975	3,827	1	150	,052
Hematologicas	,941	9,458	1	150	,002
Endocrinas	,999	,129	1	150	,720

Tabela 57: Teste de igualdade de médias de grupo

A análise gerou 1 função que explica 100% da variância dos dados e que, o teste é significativo para a função gerada, como se pode observar nas 2 seguintes tabelas:

Função	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	,121 ^a	100,0	100,0	,329

Tabela 58: Função discriminante

Test of Função	Wilks' Lambda	Chi-square	df	Sig.
1	,892	16,741	7	,019

Tabela 59: Teste de significância da função discriminante

A tabela seguinte mostra que as variáveis foram agrupadas da seguinte maneira: Contribuem com maior peso para a função as complicações Hematológicas, Urorenais e as Infeciosas.

	Function
	1
Hematologicas	,722
UroRenal	,566
Infeciosas	,459
Respiratório	-,232
CNS	-,150
Endocrinas	,084
Cardiovascular	,003

Tabela 60: Matriz de estrutura das funções

Como temos apenas uma função e por análise dos gráficos o cluster 1 tem valores mais elevados para a função e o cluster 2 tem os valores mais baixos. Isto significa que os pacientes com tempos de diálise mais elevados são os que apresentam em maior número as complicações Hematológicas, Urorenais e Infeciosas e os pacientes com tempos de diálise mais baixos apresentam em menor número estas complicações. Estes resultados corroboram os resultados obtidos na análise da regressão logística para a variável dependente Complicações-pos-oper30diasAlta.

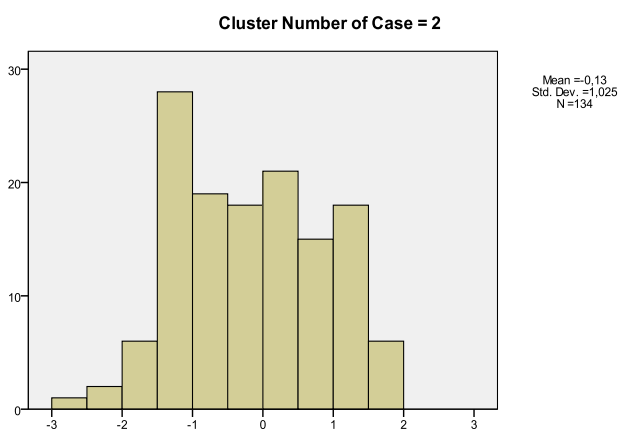


Gráfico 19: Função discriminante 1 para o cluster 2



Gráfico 20: Função discriminante 1 para o cluster 1

PARTE III

CONCLUSÕES

CONCLUSÕES

Os resultados obtidos neste trabalho vêm corroborar os resultados obtidos e já abordados nos três estudos realizados e referidos na introdução desta tese, nomeadamente "Living-donor Kidney Transplantation: Predictive Factors and Impact on Post-transplant Outcome", "Deceased-donor Kidney Transplantation: Predictive Factors and Impact on Postoperative Outcome" e "Living-donor and Deceased-donor Renal Transplantation: Differences in Early Outcome-A Single-center Experience".

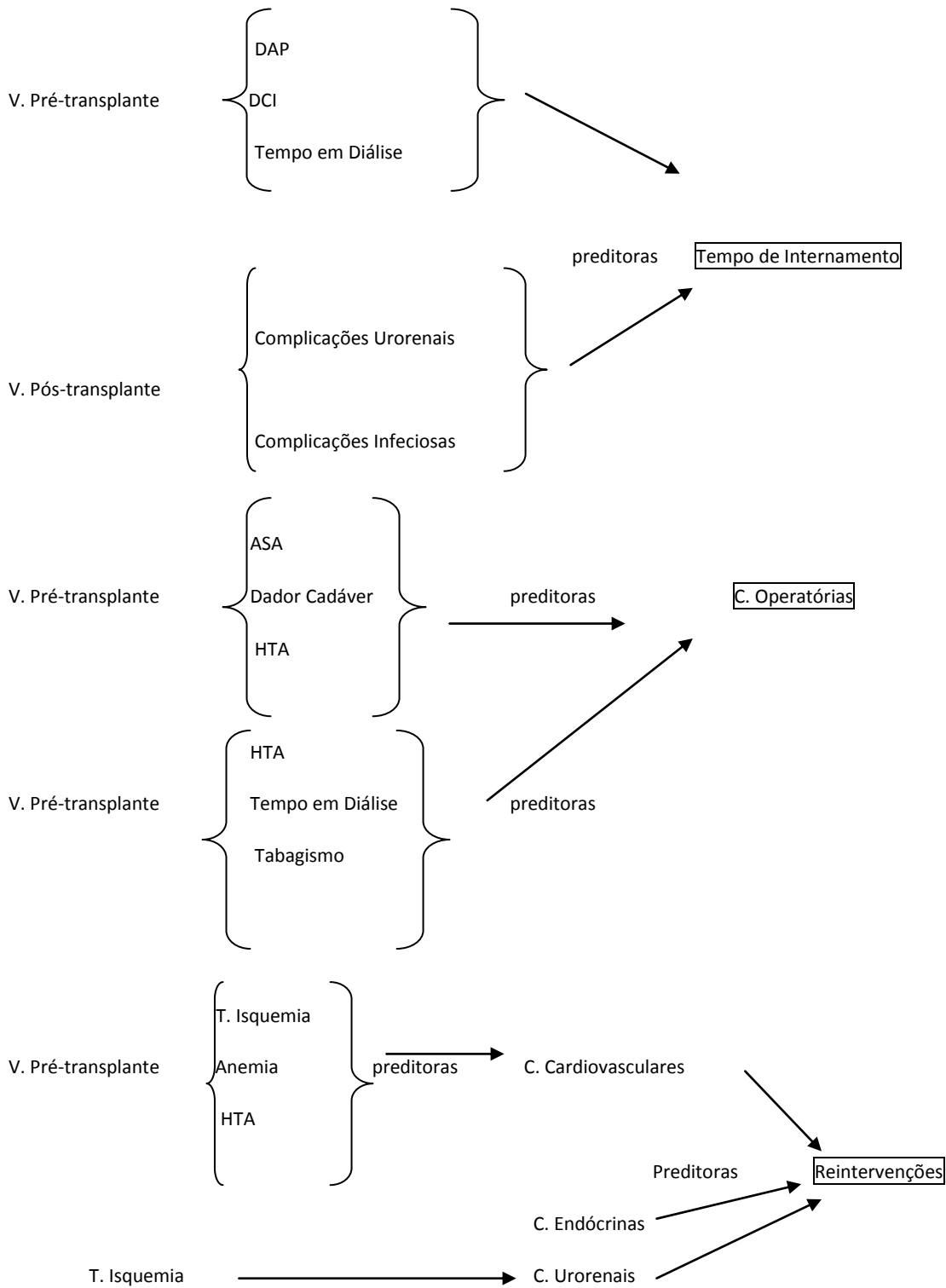
Neste estudo conclui-se que as complicações cardiovasculares são fatores de risco (OR=3,584;p-value=0,006 e IC=]1,391;9,232[) para as reintervenções, que a presença de HTA é um fator de risco (OR=5,525;p-value<0,0001; e IC=]2,094;14,578[) para as complicações pós operatórias e que as complicações respiratórias são fator de risco (OR=5,333;p-value=0,05; IC=]1,103;25,794[) para o reinternamento.

Por comparação das médias das idades verificou-se que a média de idades dos doentes transplantados e que não apresentam comorbilidades é em média 13 anos mais baixa que a dos que apresentam comorbilidades.

A análise de regressão linear mostrou que as variáveis, complicações Urenais e Infeciosas, tem grande poder preditivo na determinação do tempo de internamento com uma acurácia de 24,2% , assim com o Tempo em diálise e as comorbilidades DAP e DCI com uma acurácia de 13,9%. Estes resultados confirmam os resultados obtidos nos dois primeiros estudos que analisaram separadamente pacientes recetores de dadores vivos e de dadores cadáver, sugerindo que independentemente do tipo de dadores as comorbilidades DAP e DCI e o Tempo em Diálise são as variáveis pré-transplante com forte poder preditivo na duração do Tempo de internamento, assim como as variáveis pós-transplante Complicações Infeciosas e Urenais.

A regressão logística mostrou que o dador cadáver, a ASA e a presença de HTA são estatisticamente significativas na previsão de complicações pós operatórias com uma acurácia de 0,858 e p-value <0,0001. Ainda relativamente às complicações pós operatórias, a regressão logística mostrou que, o Tempo em Diálise, a presença de HTA e o Tabagismo, são estatisticamente significativas na sua previsão com uma acurácia de 0,808 e p-value <0,0001. Mostrou ainda que as complicações cardiovasculares, as endócrinas e urenais são estatisticamente significativas com uma acurácia de 0,754 e p-value < 0,0001 na previsão das reintervenções e que, por sua vez o tempo de Isquemia, a presença de HTA e

a Anemia são variáveis estatisticamente significativas com uma acurácia de 0,830 e p-value < 0,0001 na previsão de complicações cardiovasculares e que o Tempo de Isquemia com uma acurácia de 0,865 e p-value < 0,0001 é estatisticamente significativo na previsão de complicações urogenais. O esquema que se segue resume esta informação:



A análise de clusters permitiu concluir que as complicações cardiovasculares, respiratórias e CNS não são significativas na diferenciação dos grupos por tempos de internamento, o que confirma as conclusões obtidas da regressão linear, uma vez que estas variáveis não são estatisticamente significativas na previsão da duração dos tempos de internamento. Utilizando a análise discriminante aplicada aos 3 clusters diferenciados por médias de tempos de internamento concluiu-se que, a maior percentagem de pacientes com tempos de internamentos mais elevados são os que apresentam complicações Urorenais e Infeciosas. Utilizando a análise discriminante aplicada aos 2 clusters diferenciados por médias de tempos de diálise concluiu-se que, a maior percentagem de pacientes com tempos de diálise mais elevados são os que apresentam complicações Hematológicas, Urorenais e Infeciosas. Esta análise permitiu minimizar a probabilidade de erro de incorreta previsão em termos de duração de tempo de internamento quando estamos na presença de pacientes que estiveram em diálise por períodos prolongados, em média 14 anos.

Em função dos resultados obtidos e tendo em vista a sua aplicação atrevo-me a deixar algumas considerações e sugestões:

- É crucial o ajuste da escolha de variáveis e métodos, em função do contexto de Saúde, ouvindo e estudando os fenómenos alvos de um estudo estatístico, de modo que todas as opções a fazer sejam devidamente fundamentadas tanto na vertente estatística como na vertente da Saúde.
- É importante sensibilizar os utilizadores da Estatística na área da Saúde para a importância da correta utilização dos métodos, não só validando pressupostos mas também na seleção dos modelos (PAPOILA, A. (2012)).

Espera-se que este estudo possa, de alguma forma, contribuir, para o desenvolvimento de alterações a nível das estratégias adotadas e da clareza da informação veiculada proporcionando assim a obtenção de níveis de sucesso mais elevados nos resultados dos transplante renais ajudando a prever as situações de risco que poderão ser salvaguardadas. É importante referir que este estudo foi efetuado sobre uma amostra algo desequilibrada, pois o número de recetores de dadores vivos é muito superior ao número de recetores de dadores cadáver, com todas as características que lhes são inerentes, o que pode afetar os coeficientes que são estimados nas regressões, e não só, sendo este um aspeto que poderá ser explorado em trabalhos futuros, com vista à obtenção de resultados que poderão melhorar as conclusões obtidas.

BIBLIOGRAFIA

BIBLIOGRAFIA

1. Araujo AM, Santos F, Guimaraes J, Nunes CS, Casal M (2015). "Living-donor Kidney transplantation: Predictive Factors and Impact on Post-transplant Outcome." *Transplantation Proceedings* 47(4): 938-941
2. Santos F, Guimaraes J, Araujo AM, Nunes CS, Casal M (2015). "Deceased-donor Kidney Transplantation: Predictive Factors and Impact on Postoperative Outcome." *Transplantation Proceedings* 47(4): 933-937
3. Guimaraes J, Araujo AM, Santos F, Nunes CS, Casal M (2015). "Living-donor and Deceased-donor Renal Transplantation: Differences in Early Outcome-A Single-center Experience." *Transplantation Proceedings* 47(4): 958-962
4. Hair, JF, et al. (2014) *Multivariate Data Analysis*, 7th Edition, Pearson Education Limited
5. Kuk , AYC, et al (2014) "Variable and threshold selection to control predictive accuracy in logistic regression", *Appl. Statist.* Vol 63, Part 4, pp. 657–672
6. Braga, A. (1994). *Acidente Vascular Cerebral e seus Factores de Risco. Estudo de ocorrência de quatro tipos de AVC*. Tese de Mestrado, Universidade do Minho.
7. Pestana, M., & Gageiro, J. (2005). *Análise de dados para Ciências Sociais- A Complementaridade do SPSS*. Lisboa: Editora Sílibo, 4ª Edição.
8. Laureano, M., & Botelho, M. (2010). *SPSS o meu manual de consulta rápida*. Lisboa: Edições Sílabo, Lda, 1ª Edição.
9. <http://www.portalaction.com.br/content/an%C3%A1lise-de-regress%C3%A3o> – site credível, com informação atualizada sobre análise da regressão, com explicação muito cabal dos temas abordados.
10. MODELAÇÃO E INFERÊNCIA ESTATÍSTICA – APLICAÇÕES À EDUCAÇÃO FÍSICA ESCOLAR – Tese de Mestrado em Estatística, Matemática e Computação Especialização em Estatística Computacional - Domingos J. Lopes da Silva, 2013.
11. Modelos de Regressão: uma aplicação em Medicina Dentária, tese de Mestrado em Estatística, Matemática e Computação (Ramo – Estatística Computacional) - Maria Cristina Campos dos Santos Ferreira, 2013.
12. REGRESSÃO MÚLTIPLA: uma digressão sobre seus usos. Autores: Istvan Karoly Kasznar, PhD Professor Titular da FGV e Presidente da IBCI, Bento Mario Lages Gonçalves, MSc Consultor Senior da IBCI
13. <http://www.inf.ufsc.br/~marcelo/Cap9.pdf> - INE 7002 - Inferência Estatística – Estimação de Parâmetros

14. <http://regressao.wikispaces.com/file/view/Infer%C3%Aancia%20estat%C3%ADstica.pdf/334747526/Infer%C3%Aancia%20estat%C3%ADstica.pdf> [10 de outubro de 2015].
15. Belfiore, P. P & Fávero, L. P. “Técnicas Estatísticas Multivariadas para Análise do Comportamento de Grupos Supermercadas Brasileiros”. 2005. Universidade de São Paulo: Brasil (Artigo Científico)
16. Alexandra Pereira, Teresa Patrício, Guia Prático de Utilização de SPSS, 8ª Edição, 2013, Cap.15, Cap. 16.
17. G. O. Young, “Synthetic structure of industrial plastics,” in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64. (Capítulo de livro)
18. W.-K. Chen, *Linear Networks and Systems*. Belmont, CA: Wadsworth, 1993, pp. 123–135. (Livro)
19. H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4. (Livro)
20. B. Smith, “An approach to graphs of linear forms (Trabalho não publicado),” Não publicado.
21. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interfaces,” *IEEE Transl. J. Magn.Jpn.*, vol. 2, Aug. 1987, pp. 740–741 (Artigo Científico)
22. R. W. Lucky, “Automatic equalization for digital communication,” *Bell Syst. Tech. J.*, vol. 44, no. 4, pp. 547–588, Apr. 1965. (Artigo Científico)
23. S. P. Bingulac, “On the compatibility of adaptive controllers,” in *Proc. 4th Annu. Allerton Conf. Circuits and Systems Theory*, New York, 1994, pp. 8–16. (Artigo em Ata de Conferência)
24. G. R. Faulhaber, “Design of service systems with priority reservation,” in *Conf. Rec. 1995 IEEE Int. Conf. Communications*, pp. 3–8. (Artigo em Ata de Conferência)
25. J. Williams, “Narrow-band analyzer,” Ph.D. dissertation, Dept. Elect. Eng., Harvard Univ., Cambridge, MA, 1993. (Tese de Doutorado)
26. N. Kawasaki, “Parametric study of thermal and chemical nonequilibrium nozzle flow,” M.S. thesis, Dept. Electron. Eng., Osaka Univ., Osaka, Japan, 1993. (Tese ou Dissertação de Mestrado)
27. J. P. Wilkinson, “Nonlinear resonant circuit devices ,” U.S. Patent 3 624 12, July 16, 1990. (Patente)

28. Montgomery, D. C., Peck, E. A. and Vining, G. G. (2001). Introduction to Linear Regression Analysis. 3rd Edition, New York, New York: John Wiley & Sons.
29. http://www.apurologia.pt/publico/frameset.htm?http://www.apurologia.pt/publico/transplante_renal.htm - Associação Portuguesa de Urologia, Pedro Tiago Coelho Nunes [25 de novembro de 2015].
30. http://www.spt.pt/download/nova_lei_do_dador-20140113-212949.pdf - Nova Lei do dador [25 de novembro de 2015].
31. <http://www.spt.pt/download/rspt2014.pdf> - Registo Português de Transplantação Renal, 1980-2014; Sociedade Portuguesa de Transplantação [15 de dezembro de 2015].
32. <http://www.portaldialise.com/portal/o-que-e-o-transplante-renal> - Jornal do portal [18 de dezembro de 2015]
33. <http://www.portalaction.com.br/analise-de-regressao/33-diagnostico-de-independencia> [6 de janeiro de 2016]
34. <http://www.portalaction.com.br/analise-de-regressao/31-diagnostico-de-normalidade> [6 de janeiro de 2016]
35. <http://www.portalaction.com.br/analise-de-regressao/362-multicolinearidade> [6 de janeiro de 2016]
36. Punj. G & Stewart, DW 1983, "Cluster analysis in marketing research and suggestions for application" *Journal of Marketing Research*, vol. 20, May, pp 134-148.

ANEXO 1

Dador = vivo

Descriptive Statistics^a

	N	Range	Minimum	Maximum	Mean	Std. Deviation
Idade	40	48	8	56	37,28	12,631
Tempoemdiálise	40	14,00	,00	14,00	1,9448	3,16666
TCirurgia	36	306	73	379	136,17	66,698
TIsquemia	16	298	27	325	154,50	91,932
Númerodepedidos1º24h	22	100	0	100	29,73	24,749
Númerodedias	25	2	1	3	2,04	,676
Númerodedias_A	1	0	3	3	3,00	
Escala_da_dorR	24	2	0	2	,54	,588
Escala_da_dorM	24	2	0	2	1,08	,717
Tempointernamento	40	37	7	44	11,98	8,078
Idadedador	40	37	29	66	45,90	9,870
TIsquemiaquenterimdador	34	650	130	780	281,53	145,699
Valid N (listwise)	0					

Dador = cadaver

Descriptive Statistics^a

	N	Range	Minimum	Maximum	Mean	Std. Deviation
Idade	113	67	4	71	49,99	14,068
Tempoemdiálise	113	22,92	,08	23,00	5,7690	4,26958
TCirurgia	96	231	63	294	118,25	30,230
TIsquemia	57	1205	360	1565	1056,67	304,393
Númerodepedidos1ª24h	50	198	0	198	29,30	45,310
Númerodedias	60	3	1	4	1,67	,655
Númerodedias_A	1	0	77	77	77,00	
Escala_da_dorR	59	1	0	1	,07	,254
Escala_da_dorM	59	2	0	2	,86	,392
Tempointernamento	113	70	5	75	18,80	12,582
Valid N (listwise)	0					

TOTAIS

Descriptive Statistics

	N	Minimum	Maximum	Sum	Mean	Std. Deviation
Idade	153	4	71	7140	46,67	14,772
Tempoemdiálise	153	,00	23,00	729,69	4,7692	4,34137
TCirurgia	132	63	379	16254	123,14	43,766
TIsquemia	73	27	1565	62702	858,93	463,736
Númerodepedidos1*24h	72	0	198	2119	29,43	39,976
Númerodedias	85	1	4	151	1,78	,679
Númerodedias_A	2	3	77	80	40,00	52,326
Escala_da_dorR	83	0	2	17	,20	,435
Escala_da_dorM	83	0	2	77	,93	,513
Tempointernamento	153	5	75	2603	17,01	11,935
Idadedador	40	29	66	1836	45,90	9,870
TIsquemiaquenterimrador	34	130	780	9572	281,53	145,699
Valid N (listwise)	0					

CORRELAÇÕES

	idade	T_Diálise	HTA	T_Isque mia	T_Intern amento	C_Após Oper30 D.	Cardiov ascular e s	Urorena is	Infecios as	Hematol ógicas	Endóc ri nas
Dador	r=0,403; p<0,0001	r=0,549; p<0,0001	r=0,389; p<0,0001	r=0,717; p<0,0001	r=0,383; p<0,0001	r=0,368; p<0,0001	r=0,199; p=0,014	r=0,482; p<0,0001	r=0,218; p=0,007	r=0,210; p=0,009	r=0,290; p<0,0001

	T_Diális e	Comorbi lidades	HTA	DCI	DM	Dislipide mia	T_Isque mia	C_Após Oper30 D.	Urorena is	Infecios as	Endóc ri nas
Idad e	r=0,297; p<0,0001	r=0,334; p<0,0001	r=0,395; p<0,0001	r=0,164; p=0,043	r=0,218; p=0,007	r=0,192; p=0,018	r=0,435; p<0,0001	r=0,261; p<0,0001	r=0,193; p=0,017	r=0,160; p=0,048	r=0,163; p=0,045

	Dador	ASA	T_Diálise	DAP	T_Cirurgia	T_Isquemia	Complicações pós O p30 dias Alta	C_Cardiovasculares	C_Urorenais	C_Infeciosas	C_Hematológicas	C_Endócrinas
T_Internamento	r=0,383; p<0,00001	r=0,206; p=0,011	r=0,408; p<0,00001	r=0,168; p=0,038	r=0,178; p=0,041	r=0,470; p<0,00001	r=0,372; p<0,00001	r=0,196; p=0,015	r=0,553; p<0,00001	r=0,371; p<0,00001	r=0,241; p=0,003	r=0,195; p=0,016

	Dislipidemia
Hiperuricemia	r=0,346; p<0,0001

	C. Hematológicas
sexo	r=0,206; p=0,011

	Dif>10anos
Tipo de Relação	r=0,489; p<0,0001

	C. Cardiovasculares
Reintervenções	r=0,223; p=0,006

	Urorenais	T_Internamento
Asa Recetor	r=0,278; p=0,001	r=0,206; p=0,011

	HTA	Asa Dador
Complicações posoperat.	r=0,300; p<0,0001	r=0,319; p=0,045

	Hematológicas	Endócrinas
Urorenais	r=0,327; p<0,0001	r=0,318; p<0,0001

	T_Diálise	T_Isquemia	DAP	C. Cardiovasculares	C. Hematológicas	Urorenais	Infeciosas	Endócrinas
T_Internamento	r=0,408; p<0,0001	r=0,470; p<0,0001	r=0,168; p=0,038	r=0,196; p=0,015	r=0,241; p=0,003	r=0,553; p<0,0001	r=0,371; p<0,0001	r=0,195; p=0,016

	T_Diálise	T_Isquemia
C Urorenais	r=0,324; p<0,0001	r=0,460; p<0,0001
C. Endócrinas	r=0,303; p<0,0001	r=0,318; p<0,0001