

**UNIVERSIDADE ABERTA**



**ANÁLISE MULTIVARIADA APLICADA AO (IN)SUCESSO  
ESCOLAR A MATEMÁTICA NO 3.º CICLO - ESTUDO DE CASO**

**Tese apresentada para obtenção do grau de Mestre**

**Maria da Luz Frade Soares**

Mestrado em Estatística, Matemática e Computação

(Ramo – Estatística Computacional)

2016



**UNIVERSIDADE ABERTA**



**ANÁLISE MULTIVARIADA APLICADA AO (IN)SUCESSO  
ESCOLAR A MATEMÁTICA NO 3.º CICLO - ESTUDO DE CASO**

**Tese apresentada para obtenção do grau de Mestre**

**Maria da Luz Frade Soares**

Mestrado em Estatística, Matemática e Computação

(Ramo – Estatística Computacional)

Dissertação de tese orientada pela

**Professora Doutora Catarina S. Nunes**

2016



## RESUMO

O insucesso escolar em Portugal é preocupante e é um tema que vem despertando interesse crescente por parte dos políticos, dos professores, dos pais e do público em geral. Entre as disciplinas que mais contribuem para esse insucesso está a Matemática.

O insucesso na Matemática é uma realidade incontornável, visível não apenas pelos maus resultados alcançados pelos alunos em testes e exames, mas também pelas enormes dificuldades manifestadas por eles na resolução de problemas, no raciocínio matemático e, sobretudo, pelo seu desinteresse em relação à Matemática.

Nesta dissertação procurou aliar-se as metodologias de Estatística Multivariada, nomeadamente de Regressão Logística e de Análise de Clusters, ao insucesso escolar. As técnicas estatísticas foram aplicadas a uma base de dados construída para o efeito a partir dos resultados obtidos num inquérito aplicado aos alunos de uma escola, com o objetivo de investigar associações entre o (in)sucesso dos alunos 3.º Ciclo do Ensino Básico na disciplina de Matemática e um conjunto de variáveis referentes a dados pessoais, familiares e escolares dos alunos.

Os resultados obtidos através da Análise de Regressão Logística sugerem que a repetência a Matemática (nível inferior a 3 no final do ano letivo) está dependente da idade do aluno, das dificuldades sentidas, da participação nas aulas, do empenho e do comportamento do aluno.

A Análise de *Clusters* procurou agrupar as variáveis em grupos homogéneos relativamente a características comuns. Verificou-se que as variáveis comportamento do aluno, fazer os trabalhos de casa e a relação com a professora estão fortemente correlacionadas, bem como as variáveis dificuldades e repetência, mas estas duas últimas bastante distantes das restantes. Da aplicação do Método Duas Etapas às variáveis de opinião, fatores de insucesso escolar e medidas para colmatar esse insucesso, resultou três *Clusters*, dois grandes grupos, um em que predomina a resposta não concordo nem concordo e outro em que predomina o concordo totalmente.

Nesta dissertação conseguiu-se identificar importantes factores que podem ajudar a reduzir o insucesso, pelo que os seus resultados poderão ser usados no futuro na elaboração de medidas para melhorar o sucesso.

**Palavras Chave: Regressão Logística, Clusters, Insucesso Escolar, Matemática.**

## **ABSTRACT**

Educational failure in Portugal is a concerning issue that has been getting an increasing attention from the politicians, teachers, parents and general public. Among the subjects that more contribute to this failure, we can find Mathematics.

Failure at Mathematics is an unavoidable reality made visible, not only by the bad results reached by pupils in tests and exams, but also by the enormous difficulties revealed in the resolutions of problems, in the mathematical reasoning and, mainly, by their lack of interest towards that subject.

In this thesis, Multivariable Statistics methodologies, such as logistic regression and cluster analysis, are applied to educational failure. The statistical techniques were applied to a database built for this purpose, from the results obtained in a survey done at a school, with the aim of investigating associations between the (un)success of pupils frequenting the tenth and eleventh grades in the Mathematics subject and a set of variables related to personal, family and school data.

The results obtained using the analysis of the logistic regression indicate that the failure at Mathematics depends on the pupil's age, experienced difficulties and class participation.

The Clusters Analysis tried to group variables into homogenous groups with common characteristics. It was verified that the pupils' behaviour variables, such as 'to do the homework' and 'relationship with the teacher' are strongly correlated, as well as the variables 'difficulties' and 'year repetition', being these ones far away from the others. From the application of the two-step method to the opinion variables, 'factors of educational failure' and 'measures to overcome that failure', resulted three clusters, two big groups, one in which it predominates the answer 'I do not agree nor disagree' and other in which predominates the answer 'I totally agree'.

In this dissertation it was possible to identify important factors that might help to reduce the educational failure, so the obtained results may be used in the future to obtain measures to reduce that failure.

**Key Words: Logistic Regression, Clusters, Educational Failure, Mathematics.**

## **AGRADECIMENTOS**

À minha orientadora, Professora Doutora Catarina S. Nunes, pelo interesse que desde a primeira hora manifestou no tema que me propus tratar, pela sua disponibilidade em me apoiar em todas as tarefas necessárias para levar a bom porto esta investigação e pela confiança depositada no trabalho desenvolvido.

Aos professores do Mestrado em Estatística, Matemática e Computação, pelos conhecimentos transmitidos, em especial ao Professor Doutor Pedro Serranho, pela disponibilidade demonstrada, pelo incentivo e pelo apoio prestado no decorrer da parte curricular.

Aos meus colegas de mestrado, pela amizade e pela troca de conhecimentos.

À Diretora e ao Conselho Pedagógico do Agrupamento de Escolas Joaquim Inácio da Cruz Sobral por me terem permitido aplicar o questionário e disponibilizado a informação necessários à caracterização da Escola.

Ao meu filho, pela paciência e compreensão pelos momentos que nem sempre pude partilhar com ele.

Ao meu marido pelo apoio, companheirismo, amizade e constante incentivo.

Aos meus pais por todo o apoio e carinho que me têm dado ao longo da vida.

Aos meus avós por sempre terem acreditado em mim.

Ao Ricardo pela amizade e pela ajuda prestada.

A todos os alunos que aceitaram colaborar neste estudo.



# ÍNDICE

<b>INTRODUÇÃO.....</b>	<b>1</b>
<b>CAPÍTULO I – INSUCESSO ESCOLAR .....</b>	<b>3</b>
<b>1. INSUCESSO ESCOLAR .....</b>	<b>4</b>
1.1. Conceito de insucesso escolar .....	5
1.2. Retenção escolar em Portugal .....	9
1.3. Causas do insucesso escolar .....	12
1.4. Teorias explicativas do insucesso escolar .....	13
1.5. A realidade portuguesa em números .....	16
1.6. Matemática e o insucesso escolar.....	30
1.7. Estratégias, políticas e medidas de inclusão e promoção do sucesso educativo .....	33
1.8. Projetos e programas para combater o insucesso .....	36
<b>CAPÍTULO II – ANÁLISE MULTIVARIADA .....</b>	<b>39</b>
<b>2. ANÁLISE MULTIVARIADA .....</b>	<b>40</b>
2.1. Análise de Regressão Logística .....	40
2.1.1. Regressão Logística Univariada .....	42
2.1.1.1. Modelo de Regressão Logística Univariada .....	42
2.1.1.2. Estimacão dos parâmetros .....	43
2.1.1.3. Significância dos coeficientes .....	44
2.1.1.4. Estimacão dos Intervalos de Confiança .....	45
2.1.2. Regressão Logística Múltipla .....	46
2.1.2.1. Modelo de Regressão Logística Múltipla .....	46
2.1.2.2. Estimacão de parâmetros do modelo .....	47
2.1.2.3. Significância e qualidade do modelo .....	48
2.1.2.4. Seleção de variáveis .....	52
2.1.2.5. Análise de resíduos .....	53
2.2. Análise de <i>Clusters</i> .....	55
2.2.1. Seleção dos indivíduos .....	56
2.2.2. Seleção das variáveis .....	56
2.2.3. Medidas de semelhança ou de dissemelhança .....	57

2.2.4. Critérios de agregação ou desagregação .....	59
2.2.4.1. Agrupamento hierárquico .....	60
2.2.4.2. Agrupamento não hierárquico .....	61
2.2.4.3. Agrupamento duas etapas .....	63
2.2.5. Validação dos resultados encontrados .....	65
<b>CAPÍTULO III – OPÇÕES METODOLÓGICAS .....</b>	<b>67</b>
<b>3. OPÇÕES METODOLÓGICAS .....</b>	<b>68</b>
3.1. Objetivos e questões de investigação .....	69
3.2. Natureza do estudo .....	70
3.3. Caracterização do meio/população .....	76
3.4. Amostra .....	78
3.4.1. Dimensão da amostra .....	80
3.4.2. Seleção da amostra .....	81
3.5. Variáveis em estudo .....	85
3.6. Instrumentos de recolha.....	87
3.6.1. Elaboração de questionário.....	87
3.6.1.1. Apresentação do questionário .....	90
3.6.1.2. Pré teste do questionário .....	91
3.7. Recolha de dados .....	92
3.8. Tratamento de dados.....	93
<b>CAPÍTULO IV – APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS .....</b>	<b>93</b>
<b>4. APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS .....</b>	<b>95</b>
4.1. Caracterização da amostra .....	95
4.2. Aplicação da Análise de Regressão Logística.....	111
4.2.1. Aplicação da Análise de Regressão Logística Univariada .....	111
4.2.2. Aplicação da Análise Logística Múltipla .....	112
4.2.2.1. Análise Logística Múltipla recorrendo ao método <i>Enter</i> .....	113
4.2.2.2. Reajustamento do modelo .....	116
4.2.2.3. Comparação dos dois modelos .....	118
4.3. Aplicação da Análise de Clusters .....	118
4.3.1. Aplicação do método de agregação hierárquico .....	118

4.3.2. Aplicação do método duas etapas .....	120
<b>CONCLUSÃO</b> .....	<b>124</b>
<b>BIBLIOGRAFIA</b> .....	<b>127</b>
<b>ANEXOS</b> .....	

## ÍNDICE DE GRÁFICOS

Gráfico 1.1-Alunos matriculados por nível de educação/ensino em Portugal (2000/01-2013/14) .....	17
Gráfico 1.2 - Escolarização da população dos 3 aos 17 anos, em Portugal (2000/01- 2013/14) .....	18
Gráfico 1.3 -Taxa de abandono escolar (10-15 anos), em Portugal .....	19
Gráfico 1.4-Seis municípios que mais reduziram a taxa de abandono escolar (10-15 anos) no período 1991-2011 .....	19
Gráfico 1.5 – Taxa de abandono precoce de educação e formação (2000-2014).....	21
Gráfico 1.6 -Taxa de retenção e desistência, em Portugal, (2000/01-2013/14).....	23
Gráfico 1.7- Taxas de Atraso por Ciclo de Ensino em Portugal, 1991-2001-2011 .....	25
Gráfico 1.8 - Alunos portugueses inquiridos no PISA que ficaram retidos pelo menos uma vez, por sexo .....	26
Gráfico 1.9. Resultados de Portugal na escala Matemática ao longo dos ciclos PISA) .....	28
Gráfico 1.10- Resultados de Portugal na escala Matemática no PISA 2012, por sexo .....	28
Gráfico 1.11 – Alunos com classificação igual ou superior a 50% na prova de exame de Matemática 9.º Ano .....	30
Gráfico 1.12 – Média global dos resultados na prova de exame de Matemática de 9.º ano ....	30
Gráfico 3.1 - Sucesso por disciplina no 3.º ciclo na EBSJICS (2014-15) .....	77
Gráfico 3.2 - Sucesso a matemática no 3.º ciclo na EBSJICS (2012-13 a 2014-2015) .....	78
Gráfico 4.1 Distribuição da amostra por sexos .....	95
Gráfico 4.2-Diagrama em caixa para a variável idade .....	96
Gráfico 4.3-Repartição dos alunos que já reprovaram alguma vez .....	98
Gráfico 4.4 – Horas de sono por noite .....	99
Gráfico 4.5 – Ambiente familiar .....	99
Gráfico 4.6 – Habilitações dos pais .....	100
Gráfico 4.7 – Habilitações das mães. ....	100
Gráfico 4.8 - Profissão do pai .....	101
Gráfico 4.9 – Profissão da mãe .....	101
Gráfico 4.10 – Número de elementos do agregado familiar .....	102
Gráfico 4.11 – Realidade familiar .....	102
Gráfico 4.12 - Disciplina preferida .....	102
Gráfico 4.13 – Disciplina que menos gosta .....	102
Gráfico 4.14- Características dos alunos .....	103
Gráfico 4.15 – Dificuldades sentidas na disciplina de Matemática .....	104
Gráfico 4.16 – A quem recorrem para esclarecer dúvidas de Matemática .....	104

<b>Gráfico 4.17 - Responsáveis pela repetência .....</b>	<b>105</b>
<b>Gráfico 4.18 - Medidas para reduzir o insucesso .....</b>	<b>110</b>
<b>Gráfico 4.19 - Dendograma .....</b>	<b>119</b>
<b>Gráfico 4.20 – Qualidade de cluster .....</b>	<b>121</b>
<b>Gráfico 4.21–Importância do preditor .....</b>	<b>121</b>

## ÍNDICE DE TABELAS

Tabela 1.1 - Alunos matriculados, por nível de educação/ensino e NUTS I e II (2013/14) .....	18
Tabela 3.1 – Caracterização da População .....	83
Tabela 3.2 - Caracterização da amostra.....	84
Tabela 4.1 - Teste de sequências .....	96
Tabela 4.2- Estatísticas da variável idade .....	96
Tabela 4.3 -Tabulação cruzada Idade *Ano de escolaridade.....	97
Tabela 4.4 – Número de reprovações do aluno a Matemática a partir do 5.º ano.....	97
Tabela 4.5 – Anos em que o aluno reprovou a Matemática .....	98
Tabela 4.6 -Tabulação cruzada Ano de escolaridade * Repetência a Matemática Alguma vez .....	98
Tabela 4.7 -Tabulação cruzada Habilitações do pai * Habilitações da mãe.....	100
Tabela 4.8 - Teste qui-quadrado.....	101
Tabela 4.9 - Fatores que na opinião dos alunos contribuem para o insucesso em Matemática .....	106
Tabela 4.10 – Estatísticas das medidas/estratégias para reduzir o insucesso .....	108
Tabela 4.11 – Testes de coeficientes de modelo Omnibus modelo 1 .....	113
Tabela 4.12– Resumo do modelo 1 .....	114
Tabela 4.13 – Teste de Hosmer e Lemeshow modelo 1 .....	114
Tabela 4.14– Tabela de Classificação <sup>a</sup> modelo 1 .....	115
Tabela 4.15 – Lista entre casos <sup>b</sup> .....	116
Tabela 4.16– Resumo do modelo 2 .....	117
Tabela 4.17-Tabela de Classificação <sup>a</sup> modelo 2 .....	117
Tabela 4.18–Distribuição dos clusters .....	121
Tabela 4.19–Agrupamentos .....	123

## ÍNDICE DE FIGURAS

<b>Figura 1.1– Probabilidades de Retenção no ISCED 1 e ISCED II .....</b>	<b>27</b>
<b>Figura 1.2 – Projetos e programas de combate ao insucesso .....</b>	<b>36</b>
<b>Figura 2.1– Etapas da Análise de Clusters .....</b>	<b>56</b>
<b>Figura 2.2 – Tipos de métodos de análise de agregação .....</b>	<b>60</b>
<b>Figura 3.1 – Modelo do processo de investigação .....</b>	<b>71</b>
<b>Figura 3.2 – Métodos de amostragem.....</b>	<b>82</b>
<b>Figura 3.3 – Processo de amostragem utilizado .....</b>	<b>84</b>

## ABREVIATURAS

AIC: Critério de Informação *Akaike*

BIC: Critério de Informação *Bayesiano*

CNE: Conselho Nacional de Educação

CPCJ: Comissões de Proteção de Crianças e Jovens

DGEEC: Direção Geral de Estatísticas da Educação e Ciência

EBSJICS: Escola Básica e Secundária Joaquim Inácio da Cruz Sobral

EPIS: Empresários Pela Inclusão Social

IASE: Instituto de Apoio Sócio Educativo

IAVE: Instituto de Avaliação Educativa

INE: Instituto Nacional de Estatística

ISCED: *International Standard Classification of Education*

NUTS: Unidades Territoriais para Fins Estatísticos

OCDE: Organização para a Cooperação e Desenvolvimento Económico

PEA: Projeto Educativo do Agrupamento

PISA: *Programme for International Student Assessment*

PMSE: Programa Mais Sucesso Escolar

PNAPAE: Plano Nacional de Prevenção do Abandono Escolar

QREN: Quadro de Referência Estratégico Nacional

ROC: *Receiver Operator Characterist*

SPSS: *Statistical Package for the Social Sciences*

TEIP: Territórios Educativos de Intervenção Prioritária

UE: União Europeia

## INTRODUÇÃO

O insucesso escolar é um tema de interesse crescente no panorama do sistema educativo e suscita, cada vez mais, o interesse dos decisores políticos e organizacionais na área educativa, dos professores, dos pais e do público em geral, não podendo ser reduzido à ideia simplista de reprovação.

O conceito de insucesso escolar assumiu, ao longo dos tempos, diferentes aceções em conformidade com as conjunturas político económicas subjacentes. Aparece nos anos quarenta do século passado associado ao domínio afetivo ou psicopatológico do aluno, mas hoje é encarado como um fenómeno complexo resultante de disfuncionalidades presentes no aluno, na escola, na sociedade e na forma como estas três entidades se articulam, pelo que é necessário conhecer o funcionamento da escola, do sistema educativo e dos seus intervenientes para compreender o insucesso escolar.

Em Portugal, os primeiros estudos sobre o insucesso escolar surgem na segunda metade da década de setenta e focam-se na relação entre o insucesso ou o aproveitamento escolar e a origem social dos alunos.

Neste trabalho é abordado o insucesso escolar a Matemática, no terceiro ciclo do ensino básico, com o objetivo de conhecer as causas, os comportamentos dos alunos e encontrar possíveis medidas para reduzir esse insucesso

O estudo foi realizado com base na análise de uma base de dados, criada para o efeito no ano letivo 2014/2015, com os resultados obtidos através da elaboração e aplicação de um questionário a duzentos alunos de uma escola da região de Lisboa (Escola Básica e Secundária Joaquim Inácio da Cruz Sobral, em Sobral de Monte Agraço), bem como em informações fornecidas pelos serviços administrativos dessa escola.

Trata-se de um estudo estatístico que recorreu a técnicas como análise descritiva, regressão logística e análise de *clusters*.

O trabalho está dividido em introdução, quatro capítulos e conclusão. No capítulo um é feita a abordagem ao tema insucesso escolar, através da clarificação do conceito, da análise da situação em Portugal, do panorama na disciplina de Matemática, das causas, das teorias explicativas e das estratégias e projetos para melhorar o sucesso.

No capítulo dois consta a metodologia estatística aplicada, nomeadamente a análise de regressão logística e a análise de *clusters*. Com a análise de regressão logística visa-se

encontrar um modelo que relacione a repetência do aluno a Matemática com as suas próprias características pessoais, familiares ou escolares. Com a análise de *clusters* visa-se agrupar variáveis num pequeno número de grupos que reflitam as relações de semelhança e/ou oposição entre eles, de modo que a variabilidade nos elementos no mesmo grupo seja mínima e a variabilidade entre os grupos seja máxima.

No capítulo três apresentam-se as opções metodológicas, no que se refere a objetivos e questões de investigação, natureza do estudo, caracterização do meio/população, amostra, variáveis em estudo, instrumentos de recolha de dados, recolha e tratamento de dados.

No capítulo quatro é aplicada a metodologia abordada no capítulo dois e feita a discussão dos resultados.

Por fim na conclusão é feito o balanço do trabalho realizado.

**CAPÍTULO I**  
**INSUCESSO ESCOLAR**

## 1. INSUCESSO ESCOLAR

A noção de insucesso escolar surgiu no século passado na década de quarenta e apontava-lhe como causas aspetos do domínio afetivo ou psicopatológico do aluno (Isambert-Jamati, 1985 *apud* Duarte, 2000).

Na década de cinquenta, com a situação do pós-guerra, o insucesso escolar passa a ser visto de outra forma. Acreditando que o crescimento económico estava também dependente de um maior investimento na educação e de um aumento da escolarização, assiste-se a um forte crescimento da educação, dando a escola de elites lugar à escola de massas. Este crescimento arrastou consigo problemas como a desigualdade de acesso à educação, entre os vários grupos sociais, iniciando-se um novo campo de pesquisa com o qual se pretendia explicar as causas que estavam subjacentes ao desigual acesso à educação. Passando o insucesso escolar a ser visto como um problema social (Isambert-Jamati, 1985 *apud* Lemos 2013:156).

Em Portugal, os primeiros estudos sobre este tema surgem na segunda metade da década de setenta e focam-se na relação entre o insucesso/aproveitamento escolar e a origem dos alunos (Sebastião e Correia, 2007 *apud* Lemos 2013:156). Mas, de acordo com Benavente (1990), o insucesso escolar é um problema social que preocupa pais, professores e alunos, começando a ser mais falado em Portugal a partir de 1987.

Atualmente, o insucesso escolar tem sido alvo de inúmeros estudos, reflexões e preocupações por parte dos governantes e de todos os outros agentes sociais, sobretudo pais, professores e alunos. A problemática do insucesso escolar é complexa e multiforme e a sua abordagem implica, a clarificação do conceito, o conhecimento do funcionamento da escola, do sistema educativo e dos seus intervenientes.

Este capítulo pretende abordar a problemática do insucesso escolar e encontra-se dividido em oito secções. Na secção um procura-se clarificar o conceito, na secção dois abordar a retenção escolar em Portugal, na secção três referir-se causas, na secção quatro teorias explicativas, na secção cinco faz-se uma análise da realidade portuguesa, na secção seis aborda-se o insucesso escolar na Matemática, na secção sete referem-se as estratégias, políticas e medidas para a promoção do sucesso e por último na secção oito, os projetos e programas a decorrer em Portugal com vista a combater o insucesso.

## 1.1. Conceito de insucesso escolar

A palavra insucesso vem do latim *insucessu(m)*, o que significa de acordo com o dicionário de Língua Portuguesa da Porto Editora: “mau resultado; falta de êxito, fracasso; desastre”.

A clarificação do conceito de insucesso escolar é um passo prévio para a sua análise. O conceito é relativamente recente e surge associado à implementação da escolaridade obrigatória, decorrente das exigências da Sociedade Industrial.

Em Portugal, tal como em outros países, o conceito de insucesso escolar assumiu, ao longo dos tempos, diferentes aceções em conformidade com as conjunturas político-económicas subjacentes tendo diversos investigadores procurado defini-lo.

Para uns, o insucesso escolar designa “a não obtenção ou não realização de objetivos predeterminados pela organização escolar ou pela instrução em si” (Enciclopédia Luso Brasileira, 1980:162). Outros entendem por sucesso escolar o sucesso do aluno certificado pela escola, subentendendo-se que insucesso seja a não certificação escolar (Formosinho, 1991).

Fernandes refere que:

“insucesso escolar é a designação utilizada vulgarmente por professores, educadores, responsáveis da administração e políticos para caraterizar as elevadas percentagens de reprovações escolares verificadas no final dos anos letivos.”

(Fernandes ,1991:187)

Iturra (1990) afirma que “o insucesso escolar consiste na dificuldade que as crianças têm em aprender, em completar a escolaridade no tempo previsto, em obter notas altas ou pelo menos satisfatórias pelo seu trabalho escolar para poderem continuar os seus estudos” (Iturra, 1990:15).

Também, Rangel (1994) refere-se ao insucesso escolar como “insucesso num exame, bem como afastamento definitivo da escola provocado por repetências sucessivas” (Rangel, 1994:20). Para Landsheere (1994) o insucesso escolar define-se como um conceito teórico que integra uma situação em que não se atingiu um objetivo educativo.

Para Hutmacher:

os atores (professores, pais e alunos) consideram, em geral, que o insucesso é verdadeiramente consumado quando uma retenção é decidida: esta aparece-lhe como um sinal tangível de insucesso escolar.

(Hutmacher, 1993:33)

Muñoz (2002) refere-se ao insucesso escolar como a grande dificuldade que uma criança, com um nível de inteligência normal ou superior, pode sentir para acompanhar a formação escolar correspondente à sua idade, gerando uma inibição intelectual para com os objetivos dos conhecimentos escolares. Partindo do princípio de que a criança não sofre de nenhuma lesão cerebral, é assídua na escola e a sua família não tem um nível cultural extremamente baixo.

Benavente (1976) referia que o insucesso na escola primária, se traduzia pelas dificuldades de aprendizagem, nas reprovações e nos atrasos, sendo uma criança considerada boa ou má aluna em função dos resultados alcançados e dos progressos atingidos no cumprimento dos programas de ensino. Alguns anos mais tarde, Benavente (1999) associa a insucesso escolar a termos como reprovações, atrasos, repetências, abandono, desperdício, desadaptação, desinteresse, desmotivação, alienação e fracasso, acrescentando que a não prossecução dos estudos, após a conclusão da escolaridade obrigatória, também constitui uma forma de insucesso, uma vez que um aluno acaba por não retirar do sistema de ensino tudo o que este pode oferecer e que teria capacidade para obter.

O insucesso escolar é “entendido como a repetência ou retenção durante um ou mais anos ao longo do percurso escolar dos alunos” (CNE, 2013:40). Também Fernandes (1991) refere que a definição oficial do insucesso escolar advém do regime anual de passagem/reprovação dos alunos. Mas acrescenta que as elevadas percentagens de reprovações escolares não são por si suficientes para caracterizar o insucesso escolar, existindo insucesso ou fracasso escolar sempre que algum dos objetivos da educação escolar não é atingido. Apontando como objetivos:

“a aquisição de determinados conhecimentos e técnicas (instrução), o desenvolvimento equilibrado da personalidade do aluno (estimulação) e a interiorização de determinadas condutas ou valores com vista à vida em sociedade (socialização).”

(Fernandes, 1991:187-188)

Também, num debate com vários especialistas sobre sucesso e insucesso escolar, publicado pela revista *Ágora*, o insucesso aparece associado a retenção. Pires, afirma que o insucesso tem a ver com repetência e abandono. Considerando que uma criança com insucesso é aquela que repete sucessivamente vários anos sem uma progressão e que acaba por abandonar. Benavente, embora salvaguardando que o insucesso tem de ser definido de maneira diferente nos vários níveis de ensino, considera que existe insucesso quando o

aluno fica para trás, não atingindo alguma coisa que é suposto ser atingido por todos os alunos.

A associação entre insucesso e retenção não é consensual, Peixoto (1999) considera que o insucesso escolar ultrapassa a questão da reprovação, podendo existir insucesso mesmo que o aluno obtenha aprovação. Para este autor, insucesso escolar significa também rendimento aquém das possibilidades do aluno. Também, Cortesão e Torres (1994) defendem que além da repetência e abandono escolar, indicadores através dos quais tradicionalmente se define o insucesso escolar, existem outros aspetos reveladores do mal estar dos alunos na instituição escolar e o facto de, terminada a escolaridade, não se desencadear a mobilização dos conhecimentos adquiridos, é um indicador de que a educação não se cumpriu.

Para Sil (2004) a problemática do insucesso escolar é complexa e não pode ser reduzida à ideia simplista de reprovação, a existência de um elevado insucesso escolar não implica necessariamente elevado número de reprovações, tal como a existência de sucesso não pressupõe o contrário.

Martins (1993) refere dois tipos de insucesso: um, em que há uma redução do conceito à quantificação de um dado fenómeno observável e escolarmente determinado; e outro, mais complexo e de difícil quantificação, que se prende com o não atingir das metas pessoais e sociais de acordo com as aspirações dos indivíduos e as necessidades do sistema envolvente. O insucesso escolar ocorre quando não são atingidos os objetivos propostos ou isso não acontece no tempo previsto, ou então, apesar de terem sido conseguidos os dois pressupostos anteriores, não existe uma adequação entre os conteúdos realizados e os objetivos das partes envolvidas. Institucionalmente o insucesso escolar é usualmente atribuído ao facto de os alunos não atingirem as metas no final dos ciclos, dentro dos limites temporais estabelecidos e traduz-se por taxas de reprovação, repetência e abandono escolar. A reprovação pode ser vista como uma sanção por o aluno não ter aproveitado, por incapacidade, o saber posto à sua disposição, ou como reconhecimento de o aluno não ter apreendido os conhecimentos mínimos necessários para prosseguir para os níveis seguintes, ou ainda traduzir vontades normativas expressas (passagem ou reprovação por decreto ou circular). Para além do insucesso institucionalmente considerado existe outro tipo de insucesso, a desadequação entre os conteúdos transmitidos na escola, as aspirações dos alunos e a não conjugação destes fatores com as necessidades do sistema social

(particularmente do sistema político, cultural e económico), ou seja, se o aluno adquiriu as competências necessárias para prosseguir estudos, ingressar no mundo do trabalho, compreender o mundo que o rodeia e a construir o seu próprio saber e se existe adequação entre as aspirações dos alunos e os conteúdos transmitidos na escola, as metas socialmente propostas/impostas e as saídas/características dos empregos.

A associação de insucesso escolar à falta de preparação para a vida social e profissional também é defendida por outros autores ao referirem:

“(…) os alunos ao finalizarem a sua permanência na escola, não alcançaram os conhecimentos e as habilidades necessárias para desempenhar-se de forma satisfatória a vida social e profissional ou prosseguir os seus estudos. “

(Marchesi e Pérez, 2004:17)

Cabrita (1993) também aponta uma visão mais abrangente de insucesso escolar, ao referir-se ao insucesso educativo, o qual segundo a mesma autora manifesta-se de diversas formas, das quais destaca:

“ausência de eficiência na interiorização, deficiente apreensão e assimilação de conceitos, más notas, repetição de ano, adiamento nos exames, abandono escolar, dentro das mais objetivas e portanto mais facilmente inventariáveis pelos serviços administrativos”.

(Cabrita, 1993:32)

Segundo Eurydice (1995), em Portugal, o insucesso escolar é a incapacidade, revelada pelo aluno, em atingir os objetivos globais definidos para cada ciclo de estudos. Sendo utilizados como indicadores de insucesso escolar as taxas de retenção, de abandono e de insucesso nos exames.

No entanto, o conceito de insucesso escolar é um conceito relativo, pode assumir diversas formas ou diferentes terminologias em função do país e do sistema educativo implementado, existindo uma multiplicidade de entendimentos, diferentes países apresentam diferentes definições para o conceito. A título de exemplo, na Dinamarca “insucesso escolar é um desequilíbrio entre a aptidão do aluno, as suas capacidades naturais e o benefício que retira do ensino” (Eurydice, 1995:46), contemplando o abandono escolar no final da escolaridade obrigatória, não existindo retenções ao longo da escolaridade obrigatória. Na Inglaterra, no País de Gales e na Irlanda do Norte não existe o conceito de insucesso escolar, mas sim o de «*under-achieving*» que corresponde “à incapacidade que [o] aluno revela de concretizar o seu potencial individual” (Eurydice, 1995:46). Em França entende-se insucesso escolar fundamentalmente “em termos de saídas do sistema educativo sem qualquer qualificação” (Eurydice, 1995:47), mas também como as dificuldades escolares sentidas pelos alunos, ao longo da escolaridade, que os

impede de atingir determinadas competências e conhecimentos numa determinada idade. Na Grécia, o insucesso escolar “está associado ao nível de desenvolvimento do aluno tomando como referência as metas e objetivos curriculares definidos” (Eurydice, 1995:47). Em Itália, o insucesso escolar “representa a incapacidade do aluno adquirir os conhecimentos básicos” (Eurydice, 1995:47). Em Espanha, o insucesso escolar é entendido como as “dificuldades específicas na consecução dos objetivos fixados no ensino básico” (Eurydice, 1995:47). Na Alemanha, o insucesso escolar é percecionado através do abandono escolar e das taxas de retenção. Nos Países Baixos, o insucesso escolar está associado ao abandono precoce do sistema escolar. No Luxemburgo, o insucesso escolar entende-se “em termos de retenção, de abandono escolar ou de “outputs” educativos deficientes (significando uma taxa reduzida de diplomados, em termos comparativos internacionais)” (Eurydice, 1995:48). Na Bélgica, o insucesso escolar é definido “em termos de objetivos cognitivos não atingidos, o qual se mede pela taxa de retenção” (Eurydice, 1995:48).

## **1.2. Retenção escolar em Portugal**

Em Portugal existem atualmente mais de 150 000 alunos que ficam retidos no mesmo ano de escolaridade, pelo que a problemática da retenção no sistema educativo português assume contornos preocupantes, não só pela sua expressão, mas também pela manifesta ineficiência e ineficácia desta medida para a melhoria do desempenho escolar dos alunos. (CNE, 2015).

A retenção escolar é uma prática que consiste em manter o aluno no mesmo ano de escolaridade durante um ano adicional, em vez de avançar para o ano de escolaridade seguinte, no pressuposto de fornecer ao aluno uma nova oportunidade para ele melhorar as suas capacidades e o seu nível de aprendizagem (Pereira e Reis, 2014). Frequentemente, a retenção é associada ao insucesso académico, no entanto, pode ser resultado de doença prolongada, faltas intermitentes, ou até do desejo de um aluno (ou dos seus pais) de repetir um ano para melhor se preparar para um determinado objetivo (por exemplo, melhorar as classificações de ingresso no ensino superior).

Ao longo dos anos a nível governamental foram adotadas diversas medidas no sentido de reduzir a retenção escolar e o Conselho Nacional de Educação (CNE) tem vindo a efetuar diversas reflexões sobre a mesma.

No Parecer n.º 2/1992, o CNE pronunciou-se sobre o sistema de avaliação dos alunos do Ensino Básico, nomeadamente a cerca do Despacho n.º 98-A/1992, o qual aponta a retenção como medida de carácter excepcional, referindo:

“Se é verdade que os alunos não aprendem mais por terem repetências também é verdade que não passam a aprender mais só por serem mais dificilmente «retidos». As exigências de múltiplas justificações escritas, formais, oficiais para a “retenção” não pode significar uma pressão para a progressão facilitada. A qualidade das aprendizagens depende fundamentalmente da qualidade do ensino, da sua adequação às necessidades dos alunos e das condições em que se processa o ensino-aprendizagem. É toda a vida escolar que está em jogo. “  
(CNE,2015:29)

Referindo ainda que, embora considere o Diploma globalmente positivo, pelo relevo atribuído à avaliação formativa, pela corresponsabilização da escola pelas aprendizagens dos alunos, pela valorização da língua materna e pelo maior envolvimento dos encarregados de educação, aponta algumas carências do sistema educativo como entrave à sua concretização, nomeadamente a insuficiência de salas de aula, os tempos desadequados, as políticas de colocação de professores, a falta de recursos financeiros, as práticas pedagógicas utilizadas, a formação de professores.

Sobre o Plano Nacional de Prevenção do Abandono Escolar (PNAPAE) «Eu não desisto», o CNE emite o Parecer n.º 7/2004, onde refere que este plano “apresenta uma perspetiva que não atende suficientemente à diversidade de situações e que tende a centrar no jovem a responsabilidade do abandono”, recomendando que para a resolução do problema do abandono escolar deveria ser delimitadas linhas prioritárias de atuação, nomeadamente, apostando na generalização da educação pré-escolar; na melhoria da qualidade do 1.º Ciclo; no reforço de apoio social e económico às famílias; na criação de condições estruturais que permitam melhorar o ambiente educativo; no aperfeiçoamento da formação inicial e contínua de professores; na diversificação dos percursos de formação no Ensino Secundário; no investimento efetivo nas vias profissionalizantes; entre outras (CNE, 2015:30).

No Parecer n.º 8/2008, o CNE pronunciou-se sobre “A Educação das Crianças dos 0 aos 12 anos”, referindo que “O problema das repetições assume, em Portugal, proporções catastróficas para os alunos e para o sistema.”. Referindo que por falta de estratégias alternativas eficazes, a repetição é utilizada como estratégia pedagógica por

excelência para a regulação de problemas de aprendizagem. Referindo ainda, que a repetição não é nem um meio pedagógico adequado nem um instrumento de justiça. A repetência leva a que os alunos fiquem desintegrados nas turmas em que são colocados e perante sistemáticas avaliações negativas poderão não ter capacidade para estudar e ultrapassar os problemas. (CNE,2015).

Em 2010 o CNE refere que:

“A repetência continua a ser utilizada como uma das formas privilegiadas de superação de atrasos, o que permite inferir que as estratégias organizacionais e pedagógicas usadas não têm conseguido compensar os efeitos das diferenças de origem e as dificuldades concentradas nos percursos escolares.”

(CNE,2010)

Para muitos investigadores a repetência ou bi-repetência não contribui para a melhoria efetiva do aproveitamento escolar dos alunos, sobretudo à medida que aumenta a idade, não sendo eficaz como medida remediativa ou preventiva (Verdasca, 2007; Crahay, 2007, *apud* Moreira,2013:39).

Também Rebelo (2009) alerta para os efeitos negativos da retenção, referindo-se à mesma como:

“Medida que sanciona e que, em maior ou menor grau e dependendo do nível escolar e da idade em que os alunos se encontram, pode diminuir a sua auto estima, revoltá-los, desinteressá-los pela escola e demovê-los do empenho na aprendizagem.”

(Rebelo, 2009 *apud* CNE,2015:8)

Ferrão (2014) citando diversos autores refere como principais malefícios individuais e coletivos da retenção “o estímulo ao abandono precoce; o fomento ao autoconceito negativo; o congestionamento do sistema educativo; o desperdício de recursos.” (Almeida Júnior,1947; Teixeira de Freitas, 1947; Ferrão, Beltrão e Santos 2002b; Freitas, 2002 *apud* Ferrão, 2014).

Algumas investigações, referidas pelo CNE, demonstram que alunos retidos, nomeadamente nos anos iniciais da escolaridade, não melhoram os seus resultados e têm maior propensão para uma nova retenção, podendo estabelecer-se associação entre a retenção e o aumento dos níveis de desmotivação, indisciplina e abandono escolar. Apontando os alunos com piores condições socioeconómicas e os alunos provenientes de países estrangeiros como aqueles que apresentam uma maior probabilidade de retenção. (CNE, 2015).

No entanto, outras investigações apontam a retenção como uma medida corretiva que proporciona aos alunos a oportunidade de desenvolver a destreza necessária para o sucesso na matéria curricular mais exigente do ano escolar seguinte, tornando os grupos de

alunos mais homogêneos e colocando todos os alunos em pé de igualdade, em termos de mérito acadêmico.

Em 2015 o CNE reconhece que “o fenómeno da retenção só pode ser atenuado com a atuação no interior das escolas e em contexto de sala de aula, e envolvendo a corresponsabilização de alunos/as e das famílias” (Recomendação n.º 2/2015).

### **1.3. Causas do insucesso escolar**

Vários estudos e autores procuram explicar o facto de um número significativo de alunos reprovarem e abandonarem os estudos antes de concluírem um determinado nível de ensino, sobretudo o obrigatório.

Le Gall (1978) refere que “uma grande percentagem de insucesso escolar se relaciona com a inadaptação da personalidade da criança às exigências escolares” (Le Gall, 1978:15). Salientando que nas escolas existe uma enorme diversidade de personalidades dos alunos, e quando as escolas e as suas exigências não se sabem adaptar a essas personalidades e aos condicionamentos psicológicos de alguns alunos, tal pode levar ao insucesso escolar. Segundo este autor a adaptação da criança à escola deveria ser feita através de uma ação conjunta do professor, da escola, do aluno e da família, indo cada um ao encontro dos outros, de modo a serem ultrapassadas todas as dificuldades e possíveis falhas.

Segundo Cortesão e Torres (1999), é fácil dizer que os alunos não estudam, que vêm mal preparados ou que não se interessam. Uma vez que ninguém assume a culpa, o discurso mais comum é o do professor universitário que culpa o do secundário, enquanto que este culpa os dos níveis anteriores e o professor do primeiro ciclo atribui a culpa aos programas, às novas metodologias, à falta de inteligência dos alunos, ou à fraca cultura dos pais.

Iturra (1990) refere que em conversas com professores, apurou que existem duas razões pelas quais ocorre o insucesso escolar, uma delas é a de os estudantes não estarem interessados em aprender e os seus pais terem sobre eles expectativas que vão além das suas capacidades e a outra razão é a de inexistência de meios para ensinar e a inexistência de orientação pedagógica adequada. O mesmo autor resume estas duas razões na seguinte

afirmação: “ou os estudantes não estão aptos para aprender, ou os professores não estão aptos para ensinar. “ (Iturra, 1990:15).

Rovira (2003) refere que o termo insucesso escolar está vulgarizado e apresenta uma carga muito negativa, porque pode associar-se fracasso escolar ao fracasso do aluno na sua totalidade. Deste modo, o autor critica a simplicidade e negatividade do conceito referindo que: nem todos os insucessos são iguais, pois ninguém fracassa em tudo; às vezes o insucesso pode ocultar esforços consideráveis; e um insucesso pode ser um mal menor, quer ao nível pessoal, quer ao nível social.

O insucesso escolar deve ser visto como fenómeno sistemático e complexo, e, como tal, não pode ser explicado por uma única perspetiva, pois é resultado de um conjunto de fatores, tais como, situação social, familiar e escolar, daí que só possa ser compreendido atendendo à relação dinâmica de todos esses fatores.

Ao longo dos anos vários têm sido os autores que elaboram hipóteses e construíram teorias no sentido de explicar o insucesso escolar. Muitos autores defendem que o fenómeno do insucesso escolar iniciou-se com o ensino de massas e intensificou-se com a massificação do ensino (Martins, 1993; Mendonça, 2006). A imputação causal do insucesso escolar tem variado com a evolução do sistema de ensino e com as posturas teóricas que equacionam o problema. No entanto podem ser identificados vários fatores de ordem intrínseca e extrínseca ao próprio aluno que podem contribuir para o agravamento deste fenómeno. As causas podem centrar-se ao nível do aluno e do seu ambiente restritivo, ao nível da sociedade à qual pertence ou ao nível da própria escola e do sistema educativo (Sil, 2004). Assim, no estudo do insucesso escolar deverá ser tomada em linha de conta três realidades: o aluno, o meio social e a instituição escolar (Benavente,1976), sendo na articulação entre estas três realidades que se poderá encontrar as suas causas explicativas.

#### **1.4. Teorias explicativas do insucesso escolar**

As principais correntes de pensamento explicativas das causas de insucesso escolar, apontam para três teorias: a teoria dos fatores individuais, a teoria do *handicap* sociocultural e a teoria socioinstitucional.

##### Teoria dos fatores individuais

A teoria dos fatores individuais surge após o final da 2.<sup>a</sup> Guerra Mundial e apresenta duas correntes: a genética e a psicoafetiva (Eurydice, 1995).

A corrente genética explica o insucesso escolar por perturbações e deficiências intrínsecas ao próprio aluno, detetáveis por meio de testes, os quais permitiam identificar os alunos com necessidades específicas de apoio. Foi no âmbito desta corrente que surgiu, no século XX, as escalas métricas de inteligência e posteriormente o quociente de inteligência, com o objetivo de identificar cientificamente os alunos que não tinham capacidade para se adaptarem ao ensino primário normal e como tal, eram colocados em classes de remediação. A escola era vista como um lugar neutro, no qual eram criadas condições de acesso e sucesso a todos os alunos em igualdade de circunstâncias e onde o insucesso é considerado como insucesso do aluno, sendo a inexistência de determinadas aptidões inatas ao próprio aluno que influencia negativamente o seu rendimento escolar (Fernandes, 1991).

Para os defensores desta corrente o sucesso escolar é função da inteligência inscrita na herança genética, representada pelo quociente intelectual. Esta corrente tem sido alvo de fortes críticas por ter pouco rigor científico e por ser bastante redutora, dado que caracteriza o indivíduo com base na herança genética e não tem em linha de conta a experiência, a aprendizagem e o meio.

A outra corrente, a psicoafetiva relaciona o processo de construção da personalidade da criança com o desenrolar do seu processo de escolarização. Esta corrente identifica quatro etapas escolares que constituem momentos de ruptura na vida do indivíduo: o jardim de infância, a escola primária, o ensino secundário e o ensino superior, os quais são considerados momentos de ruptura uma vez que exigem uma adaptação a novas situações, às quais nem todos respondem da mesma forma. Algumas sentem dificuldades que afetam o seu comportamento escolar e que se refletem na sua personalidade, podendo o aluno remeter-se a uma recusa total da escola ou a uma recusa passiva que se traduz na falta de iniciativa. Também os alunos ditos “normais” com insucesso escolar podem enfrentar outras formas de disfunção manifestando atitudes reveladoras de perturbações afetivas acompanhadas por desinteresse escolar.

Alguns investigadores estabeleceram uma correlação direta entre o insucesso escolar e situações psicoafetivas específicas. O insucesso escolar é encarado como uma resposta à atitude de rejeição ou superproteção dos pais.

O insucesso é considerado como insucesso do aluno, sendo a inexistência de determinadas aptidões inatas ao próprio aluno que influencia negativamente o seu rendimento escolar, a escola é vista apenas como um espaço revelador das capacidades individuais do aluno (Fernandes, 1991).

A teoria dos fatores individuais, perdeu o seu crédito a nível científico e o surgimento de políticas educativas e as ideias de igualdade de oportunidades na educação levaram a alterações profundas no equacionar da problemática do insucesso escolar (Sil,2004). No entanto, continua a dominar no senso comum de pais e professores, pois é uma «explicação “fácil” que desculpabiliza e justifica a passividade de uns e outros» (Benavente e Correia, 1980:11).

#### Teoria do handicap sociocultural

A teoria do handicap sociocultural, em voga nos anos sessenta e setenta, explica o insucesso escolar essencialmente em termos de défices, relacionados com as condições económicas, sociais e culturais do meio envolvente. O défice de recursos culturais que caracterizam o contexto familiar e social comprometem o desenvolvimento cognitivo e verbal da criança, ou seja, causa um atraso no desenvolvimento intelectual da criança.

O (in)sucesso escolar dos alunos é justificado pela sua herança sociocultural fruto das condições de vida de que dispõem no início da sua escolaridade, pelo que uma criança proveniente de um meio familiar desfavorecido não dispõe das bases culturais necessárias ao sucesso escolar, uma vez que o défice de recursos económicos, sociais e culturais não permite proporcionar à criança as bases necessárias, linguísticas e culturais, para um desempenho bem sucedido na escola.

Como o ambiente familiar revela-se incapaz de proporcionar ao aluno bases culturais e linguísticas essenciais para a sua progressão escolar surgem uma multiplicidade de projetos pedagógicos de compensação ou de apoio educativo, que visam compensar o défice cultural da criança (organização de salas de estudo, centros de recursos educativos) e o défice educativo dos pais.

A teoria do handicap sociocultural embora ponha de parte as explicações de carácter genético e de carácter psicológico do insucesso escolar continuava a atribuir ao aluno e à sua família a causa do insucesso. Assim, apesar de reduzir o peso da responsabilidade do insucesso sobre a criança continuava a transmitir a ideia de que seria, para esta, um fenómeno inevitável e à semelhança da teoria dos fatores individuais, um indivíduo

proveniente de um meio socioeconómico desfavorável pouco ou nada podia fazer para contrariar a situação.

Esta teoria tem sido “fortemente criticada por tomar como quadro de referência da avaliação dos alunos e do seu meio um modelo de saber único e universal” (Eurydice, 1995:59). Começando por perder força, pois verificou-se que nos países ditos desenvolvidos a melhoria do nível de vida socioeconómico não fez desaparecer o insucesso escolar.

#### Teoria socioinstitucional

Nos anos setenta surge a teoria socioinstitucional que aponta a escola, os seus mecanismos, o seu funcionamento e a sua organização como causa do insucesso escolar .

Benavente (1989) refere que “a corrente socioinstitucional sublinha a necessidade de diversidade e de diferenciação pedagógica pondo em evidência o carácter ativo da escola na produção do insucesso” (Benavente, 1989 *apud* Ministério da Educação,1995:7).

Sil (2004), refere que a teoria socioinstitucional evidencia um papel mais ativo da escola em relação ao insucesso dos alunos, não da instituição em si, nem das infraestruturas, mas essencialmente das estruturas administrativas e pedagógicas, da avaliação dos alunos, da colocação dos professores, da falta de equipamentos, da abertura da escola à comunidade, da distribuição dos alunos por turma, do absentismo dos professores, da estrutura curricular, das políticas educativas e de ensino e das realidades sociais.

Fernandes (1991:189) menciona um conjunto de fatores escolares que podem estar na base do insucesso escolar: “tipo de cursos e currículos, estruturas e métodos de avaliação, formas de agrupamento dos alunos, preparação científica e pedagógica dos professores”. Pelo que apostam na transformação da escola, nomeadamente das suas estruturas, conteúdos e práticas de modo a adaptá-la à diversidade de alunos que a frequentam (Benavente, 1990:717).

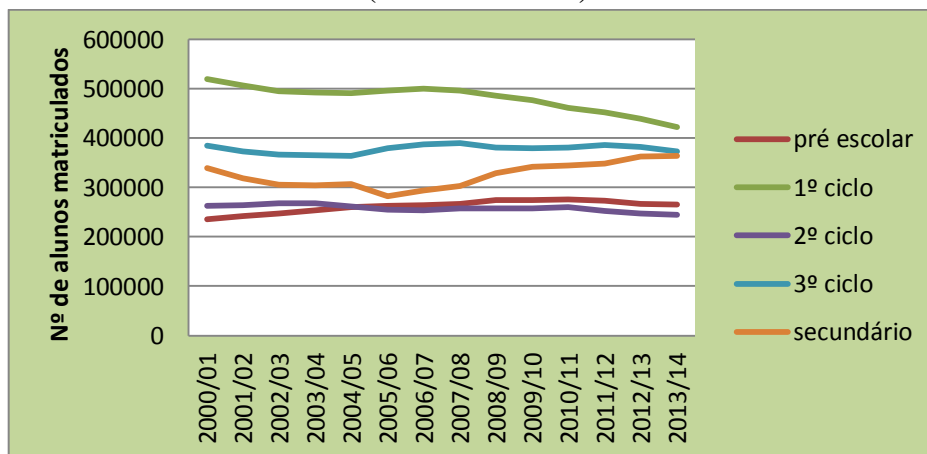
### **1.5. A realidade portuguesa em números**

Na análise da realidade portuguesa em números foram utilizados os dados provenientes da Direção Geral de Estatísticas da Educação e Ciência (DGEEC), Inquéritos ao Emprego promovidos pelo Instituto Nacional de Estatística (INE) e dos

Censos, base de dados da Pordata e estudos *Programme for International Student Assessment (PISA)*. Os números provenientes das diversas fontes nem sempre são coincidentes, no entanto, as diferenças entre eles não são muito significativas.

Começando por analisar os alunos matriculados (Gráfico 1.1), no período 2000/01 a 2013/14 é o primeiro ciclo que apresenta maior número de alunos sendo também neste ciclo de ensino que se têm registado um maior decréscimo. No ano letivo de 2000/01, Portugal apresentava um total de 1 872 509 alunos matriculados (do pré escolar ao ensino secundário), dos quais 415 778 matriculados no 3.º ciclo do ensino básico. A partir de 2000/01 e até 2005/06 o número de alunos matriculados foi sempre diminuindo em todos os níveis de ensino, com exceção do 3.º ciclo, que em 2005/06 registou um aumento. Em 2006/07 registou-se um aumento do número de alunos matriculados, continuando a aumentar nos dois anos letivos seguintes, atingindo o seu número máximo em 2008/09, com cerca de 2 014 831 alunos, ano a partir do qual se voltou a verificar decréscimo atingindo em 2013/14 o número mínimo. Nos últimos anos o ensino secundário não seguiu a tendência geral de decréscimo, vindo a aumentar desde 2006/07.

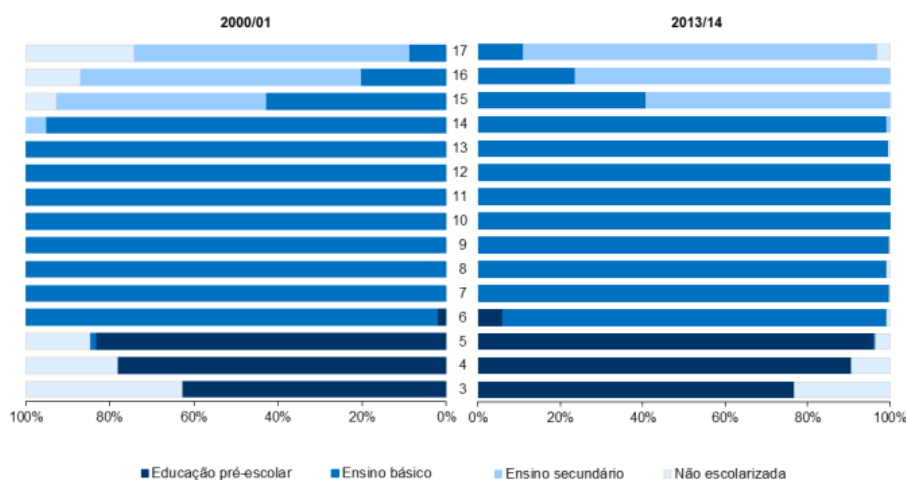
**Gráfico 1.1 – Alunos matriculados por nível de educação/ensino em Portugal (2000/01-2013/14)**



Fonte: DGEEC, 2015

Em termos de escolarização da população dos 3 aos 17 anos (Gráfico 1.2) a percentagem de população escolarizada aumentou no período 2000/01-2013/14, nomeadamente dos 3 aos 5 anos e dos 15 aos 17 anos, resultado das políticas do governo, através do alargamento do pré escolar e do aumento da idade da escolaridade obrigatória respetivamente. Em 2013/14 a percentagem de população não escolarizada dos 6 aos 17 anos é muito pouco significativa.

**Gráfico 1.2 – Escolarização da população dos 3 aos 17 anos, em Portugal (2000/01 e 2013/14)**



A região do país com maior número de alunos matriculados é a região Norte (aproximadamente um terço do total), seguida da região A.M.Lisboa (Tabela 1.1).

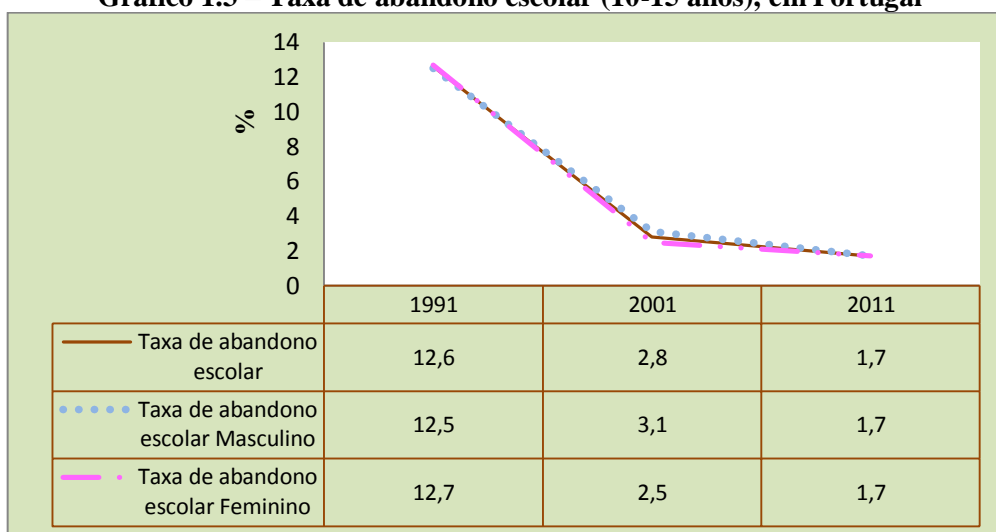
**Tabela 1.1 – Alunos matriculados, por nível de educação/ensino e NUTS I e II (2013/14)**

Nível de educação/ ensino NUTS I e II	Total	Educação pré-escolar	Ensino básico			Ensino secundário	
			Total	1.º Ciclo	2.º Ciclo		3.º Ciclo
<b>Portugal</b>	<b>1 708 083</b>	<b>265 414</b>	<b>1 057 459</b>	<b>424 284</b>	<b>249 754</b>	<b>383 421</b>	<b>385 210</b>
<b>Continente</b>	<b>1 610 770</b>	<b>251 059</b>	<b>995 294</b>	<b>399 439</b>	<b>234 625</b>	<b>361 230</b>	<b>364 417</b>
Norte	598 043	91 369	366 591	143 090	86 654	136 847	140 083
Centro	346 827	54 290	212 144	84 142	49 524	78 478	80 393
A. M. Lisboa	479 608	75 426	299 394	124 272	70 505	104 617	104 788
Alentejo	113 495	18 756	70 568	28 425	16 942	25 201	24 171
Algarve	72 797	11 218	46 597	19 510	11 000	16 087	14 982
<b>R. A. Açores</b>	<b>49 116</b>	<b>7 600</b>	<b>31 664</b>	<b>12 845</b>	<b>8 020</b>	<b>10 799</b>	<b>9 852</b>
<b>R. A. Madeira</b>	<b>48 197</b>	<b>6 755</b>	<b>30 501</b>	<b>12 000</b>	<b>7 109</b>	<b>11 392</b>	<b>10 941</b>

Fonte: DGEEC, 2015

O abandono escolar, indicador utilizado para aferir o grau de concretização da escolaridade obrigatória, reduziu significativamente nas últimas décadas. Como se pode observar no Gráfico 1.3 no período 1991-2011, a taxa de abandono passou de 12,6% para um valor quase residual de 1,7%, sendo a queda mais significativa de 9,8 pontos percentuais, registada na década de noventa. Não se pode identificar de forma inequívoca a diferença entre o abandono masculino e feminino. Ambos acompanharam o movimento geral sem grandes diferenças. Em 1991 a taxa de abandono escolar era ligeiramente maior para o sexo feminino (12,7%) ao passo que a do sexo masculino era de 12,5%, mas em 2001 a situação já se havia invertido.

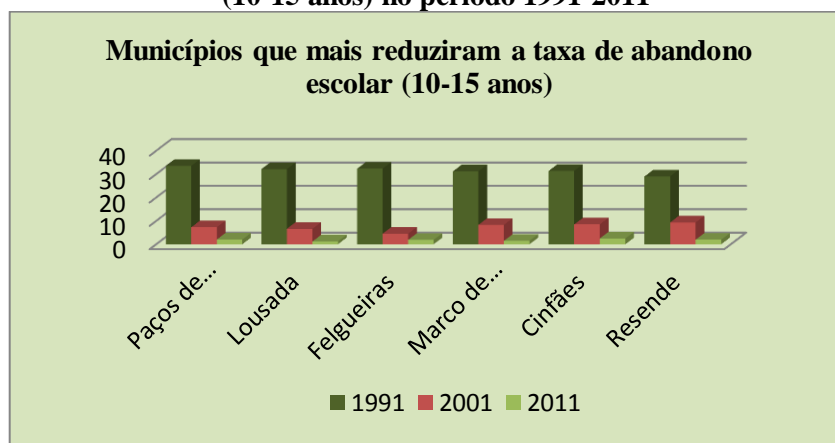
**Gráfico 1.3 – Taxa de abandono escolar (10-15 anos), em Portugal**



Fonte: INE, Censos à população

Os concelhos com maior queda na taxa de abandono escolar (Gráfico 1.4) situam-se a norte do país, em zonas onde o setor industrial recorria habitualmente ao trabalho infantil, provocando um abandono da escola antes da conclusão da escolaridade obrigatória. O concelho de Paços de Ferreira foi o que mais reduziu esta taxa, passando de 33,5% em 1991 para 2,01% em 2011, correspondendo a uma redução de 31,49 pontos percentuais.

**Gráfico 1.4 – Seis municípios que mais reduziram a taxa de abandono escolar (10-15 anos) no período 1991-2011**



Fonte: INE, Censos à população

Em 2011 o abandono escolar confina-se aos meios rurais em risco de desertificação, com bolsas de pobreza e tende a associar-se a grupos étnica e culturalmente minoritários, sendo de salientar um número considerável de concelhos da Região Autónoma dos Açores. O município que apresenta o maior valor para esta taxa é o de Gavião com 4,97% (superior em 3,27 pontos percentuais à média nacional). Os municípios que apresentam menor taxa

de abandono são os da Golegã e Vouzela com 0,31%, seguidos de Cadaval com 0,36% . Não se podendo afirmar que exista um padrão geográfico de carácter regional entre os municípios que apresentam as taxas de abandono escolar mais baixas.

Entre os potenciais fatores de abandono escolar encontram-se a escolarização dos pais, o mercado de trabalho (desemprego) e o insucesso escolar (atraso escolar). O insucesso escolar, entendido como a repetência ou retenção durante um ou mais anos ao longo do percurso escolar dos alunos, é apontado por alguns estudos como fator preditivo do abandono escolar.

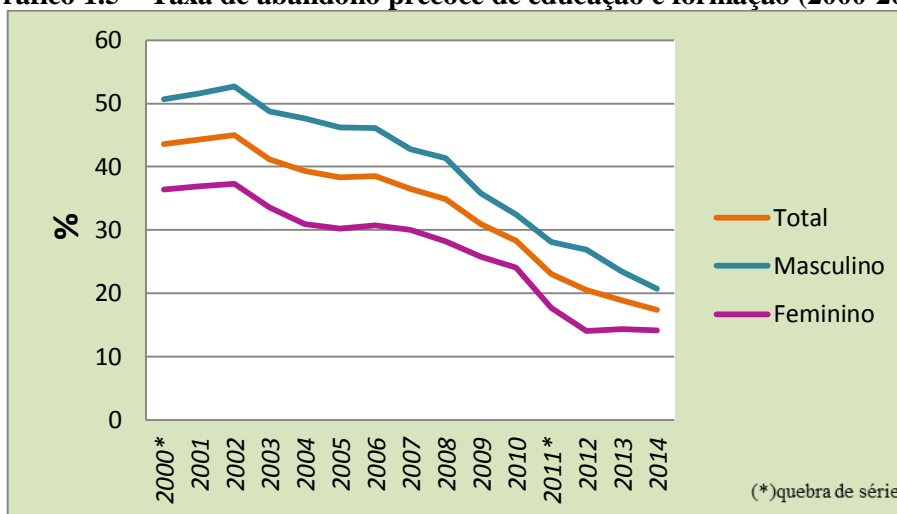
Face ao carácter residual do indicador abandono escolar (10-15 anos) é cada vez mais utilizado, ainda que visando um fenómeno relativamente diferente, o “abandono precoce”. Identifica-se como o conjunto de indivíduos que no grupo etário dos 18-24 anos já não se encontra a frequentar qualquer curso de ensino formal ou de formação, não tendo terminado o ensino secundário.

O abandono precoce ou saída precoce é um indicador que “mostra a dimensão do número de indivíduos que não concluíram a escolaridade de nível secundário.” (CNE,2013:36). Este indicador permite a comparação internacional, uma vez que os países diferem na definição de abandono e na duração da escolaridade obrigatória.

Percebendo os contextos em que devem ser analisadas as taxas de abandono precoce, é possível compreender melhor a dimensão da quebra verificada neste indicador. Como se pode observar, no Gráfico 1.5, a taxa de abandono precoce reduziu significativamente no período em análise, passando de 43,7% em 2000 para 17,4% em 2014, mas ainda assim acima da média dos 28 países da União Europeia (UE) de 11,1%. Sendo ainda de referir que segundo os Censos de 1991 esta taxa havia sido de 63,7%. Nas últimas duas décadas, Portugal protagonizou a mais relevante descida do abandono precoce na Europa. O maior contributo para esta redução deveu-se à escolarização feminina.

No período em análise a taxa de abandono precoce foi sempre significativamente superior no sexo masculino, em 2014 a diferença entre as taxas dos dois sexos foi de 6,6 pontos percentuais, evidenciando assim o impacto da dinâmica da escolarização feminina, em todos os níveis de ensino. Esta diferença não é uma característica exclusivamente nacional, de acordo com dados do Eurostat, na maior parte dos países da UE, a percentagem de homens que abandona precocemente a escola é superior à das mulheres.

**Gráfico 1.5 – Taxa de abandono precoce de educação e formação (2000-2014)**



Fonte: INE – Inquérito ao emprego

Em termos regionais, a taxa de abandono precoce diminuiu em todas as sete regiões NUTS II do país mas, em 2014, ainda nenhuma delas havia atingido o valor médio da UE. As regiões que mais se aproximaram foram a região Centro (14%) e a Área Metropolitana de Lisboa (14,4%). Ao passo que nos Açores, a taxa de abandono escolar situava-se acima dos 30% e na Madeira e Algarve nos 20%.

De acordo com um estudo do Observatório do QREN (Figueiredo, 2013), as ofertas de dupla certificação, do Ensino Básico (cursos de educação e formação) e Secundário (cursos profissionais e de aprendizagem), contribuíram de modo relevante para a redução do abandono precoce e dos níveis de retenção. Já o contributo do programa Território Educativo de Intervenção Prioritária (TEIP) é considerado moderadamente positivo, dada a fragilidade dos seus resultados.

Apesar dos progressos na redução da taxa de abandono precoce, Portugal continua a ser dos estados membros da UE com maior abandono do sistema de ensino sem o 12.º ano completo e a estar ainda longe da média definida pela estratégia Europa 2020, que coloca a redução do abandono precoce de educação e formação em 10%. Pelo que será necessário desenvolver estratégias, quer para evitar o abandono, quer para reintegrar os que já abandonaram o sistema. Neste sentido o CNE recomenda que:

“As ações para a redução dos números do abandono escolar precoce devem desenrolar-se em diferentes frentes, no sentido de prevenir o risco de abandono (assegurando uma educação de qualidade desde os primeiros anos de vida), de o evitar (reagindo aos sinais de alerta e dando o devido apoio) e de compensar os que já abandonaram (possibilitando o reingresso no sistema, reconhecendo as aprendizagens já realizadas e qualificando). As estratégias nacionais a implementar deverão, também, ter em conta as desigualdades registadas na desagregação dos dados por sexo e região.”

(CNE,2013:38)

Segundo CESNOVA - EPIS (2014) diversos estudos apontam o insucesso escolar, expresso pela acumulação de retenções, como a antecâmara do abandono. O abandono escolar pode ser resultado do insucesso escolar, pois as repetências acumuladas tendem a aumentar o risco de abandono, mas também o insucesso escolar poderá ser o resultado de uma decisão antecipada de um abandono futuro, decisão essa que leva alguns alunos a desinvestirem no esforço para o sucesso. O desinvestimento na escolarização, poderá também resultar das condições económicas familiares ou dos contextos sociais envolventes, nomeadamente da influência de grupos de amigos e colegas e da natureza da atividade económica predominante na região que pode favorecer a inserção precoce no mercado de trabalho. Também o tipo de aprendizagens pode potenciar trajetos de sucesso ou de insucesso, os alunos que apresentam retenções no primeiro ciclo têm maior probabilidade de insucesso reiterado e de abandono, dado que os conhecimentos e competências básicas indispensáveis às aprendizagens nos ciclos seguintes não foram consolidados.

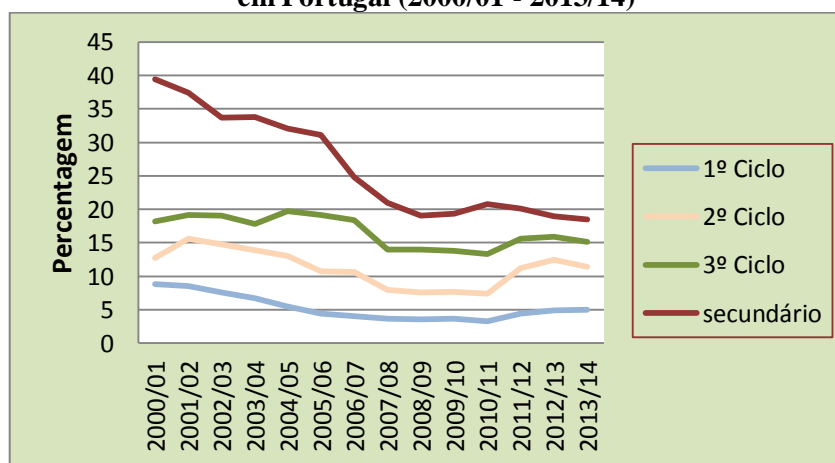
Para o estudo do insucesso utiliza-se a chamada taxa de retenção que é um indicador construído a partir das informações estatísticas do Ministério da Educação, tendo como origem os registos das escolas. No Gráfico 1.6 encontra-se a relação percentual entre o número de alunos que não podem transitar para o ano de escolaridade seguinte e o número de alunos matriculados, nesse ano letivo.

Observa-se uma descida da taxa de retenção em todos os ciclos de ensino, ainda que, de modo não constante nem uniforme. Esta taxa vai aumentando à medida que os alunos progridem nos ciclos, atingindo valores máximos no secundário. Até ao ano letivo 2010/11 verifica-se uma tendência de queda em ambos os níveis de ensino, voltando a taxa a aumentar a partir de 2011/2012. No período em análise, a queda das taxas oscila entre o mínimo de 1,3 pontos percentuais, para o caso do 2.º ciclo (de 12,7% para 11,4%), e o máximo de 20,9 pontos percentuais, para o caso do secundário (de 39,4% para 18,5%), reduzindo para cerca de metade no caso do 2.º ciclo (de 14,8% para 7,5%) e do ensino secundário (de 35,9% para 18,9%). A descida no 1.º ciclo foi de 3,8 pontos percentuais (de 8,8% para 5,0%) e a descida no 3.º ciclo de 3,1 pontos percentuais (de 18,2% para 15,1%). O padrão de comportamento também não foi o mesmo. Verifica-se que, no caso do 1.º ciclo ocorreu uma descida significativa até 2008/09, no ano seguinte subiu ligeiramente, voltando a descer em 2010/11, atingindo o valor mínimo de 3,3%, a partir daí foi sempre

crecente. No 2.º ciclo verificou-se uma subida significativa em 2001/02, decrescendo de seguida até 2008/09, sofrendo oscilações pouco significativas nos dois anos seguintes, mas em 2011/12 subiu 3,8 pontos percentuais (passou de 7,4% para 11,2%), voltando a subir 1,3 pontos percentuais no ano seguinte e a descer no ano seguinte em 2013/14. O 3.º Ciclo nos primeiros anos sofreu oscilações, com subidas e descidas não muito significativas, em 2007/08 desceu 3,6 pontos percentuais, mantendo-se constante no ano seguinte e em 2011/2012 e 2012/13 voltou a subir, descendo em 2013/14 para 15,1%, valor aquém do valor mínimo verificado em 2010/11 de 13,3%. No secundário a tendência foi de descida acentuada, sendo a maior delas em 2006/07 (de 6,3 pontos percentuais), excepto nos anos 2011/2012 e 2012/2013 em que ocorreram subidas. As descidas mais acentuadas em meados da primeira década do século XXI devem-se à introdução e alargamento de novas vias de ensino, como os cursos profissionais nas escolas públicas.

Podemos pois concluir que os percursos escolares marcados pelas retenções se iniciam em níveis educativos muito precoces e se vão acentuando à medida que avança a escolaridade.

**Gráfico 1.6 – Taxa de retenção e desistência nos ensinos básico e secundário, em Portugal (2000/01 - 2013/14)**



Fonte: DGEEC, 2015

De acordo com o relatório técnico do CNE (2015) no que se refere às taxas de retenção e desistência por NUTS II, no Ensino Básico, o Centro é, por regra, a região que revela menores taxas, sendo na Região Autónoma dos Açores que, habitualmente, se observam taxas mais elevadas, já no Ensino Secundário, o Norte é a região que revela menores taxas, com exceção do ano 2001/2002, em que a menor taxa foi atingida pela Região Autónoma da Madeira. Por outro lado, é a Região Autónoma dos Açores aquela em que se observam maiores taxas, com exceção para o período 2004-2007, em que as maiores

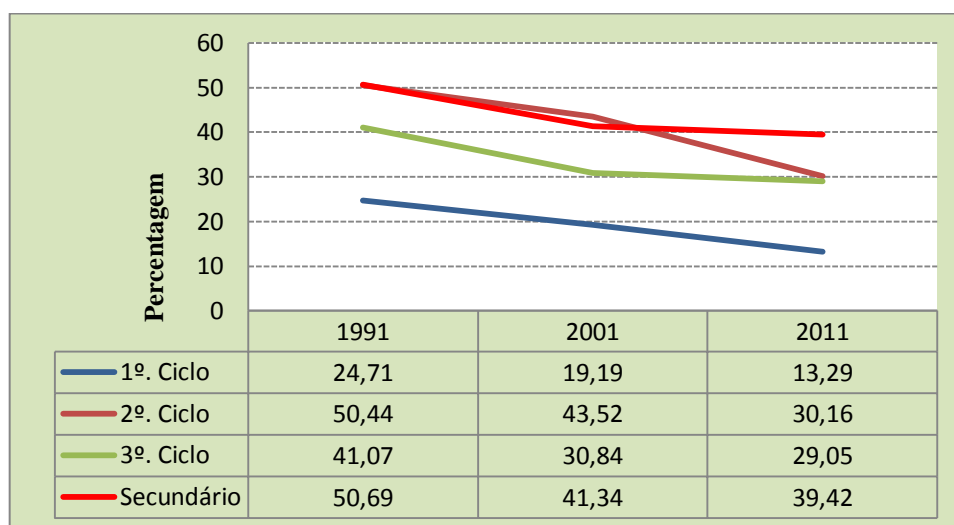
taxas foram registadas na Região Autónoma da Madeira. As taxas de retenção e desistência são sempre maiores para o sexo masculino que para o feminino e maiores para o ensino público que para o privado. No caso do Ensino Básico, as taxas do público chegam a ser mais do dobro das taxas no privado.

As taxas de atraso fornecidas pelos Censos também dão uma aproximação ao insucesso. Estas taxas medem a proporção entre os indivíduos a frequentar um determinado ciclo de ensino com idade superior à idade ajustada, no total de indivíduos com idade ajustada a esse ciclo. Embora este indicador não nos dê a dimensão da repetência, fornece-nos o número de indivíduos com, pelo menos, um ano de atraso em relação à idade ajustada à frequência do ciclo.

Analisando o Gráfico 1.7 é possível identificar algumas tendências. A percentagem de alunos a frequentar os diferentes ciclos de ensino com idade superior às idades ajustadas vem a diminuir desde 1991, sendo essa redução mais sensível no 1.º e 2.º ciclos do que no 3.º ciclo e Secundário. O 2.º ciclo apresentava em 1991 um valor anormalmente alto (50,44%), idêntico ao do Secundário (50,69%), o qual reduziu significativamente sobretudo na década seguinte. O 3.º ciclo e o Secundário, depois de quebra acentuada na década de 90, estabilizaram a proporção dos alunos com idade superior à idade ajustada, apresentando em 2011 taxas de 29,05% e 39,42%, respetivamente. Tal significa que nesses dois ciclos de ensino não se conseguiu reduzir de forma significativa a acumulação da retenção. Dos alunos atualmente a frequentar o 2.º, 3.º ciclos e Secundário cerca de um terço tem, pelo menos, uma retenção no seu trajeto escolar.

Em termos de distribuição geográfica do “atraso escolar”, as taxas de atraso do 1.º ciclo alteram a sua distribuição de 1991 para 2011 passando a maior incidência do Norte do país para o Sul, ainda que se mantenha elevada a incidência no Vale do Douro e na Beira Interior. Atualmente, a maior incidência situa-se na Lezíria do Tejo, na Área Metropolitana de Lisboa, no Alentejo e no Algarve, possivelmente pela presença de comunidades migrantes. As taxas de atraso do 2.º e 3.º ciclos apresentam um padrão geográfico semelhante, acentuando a natureza cumulativa do insucesso. No caso do atraso no Secundário, os valores seguem o mesmo padrão de distribuição, mas torna-se mais evidente a oposição litoral-interior na metade norte do país, e a generalização a zonas urbanas e rurais, quer sejam do interior quer do litoral, na metade sul (Atlas de Educação, 2015).

**Gráfico 1.7 – Taxas de Atraso por Ciclo de Ensino em Portugal, 1991-2001-2011**



Fonte: INE, Censos à população

Ao longo do tempo a medição e interpretação dos resultados escolares, incluindo do (in)sucesso tem sofrido evoluções. As práticas pedagógicas e organizacionais, no respeitante ao insucesso escolar são historicamente diferenciadas nos diversos países. Assim, se nos países do sul da Europa, tais como Portugal, França, ou Bélgica, tem sido utilizado a retenção e a repetência como instrumento pedagógico na organização dos percursos escolares dos alunos, o mesmo não tem acontecido em países do norte da Europa, tais como Dinamarca, Inglaterra ou Noruega, onde a repetência não existe ou é excecional ou residual. Pelo que, a taxa de abandono escolar precoce é um bom indicador para o sucesso ou insucesso, o mesmo não acontecendo com a taxa retenção e desistência, uma vez que apenas constitui um indicador robusto nos países do sul da Europa (Lemos, 2013).

O facto de diferentes sistemas de ensino utilizarem diferentes formas de medir os resultados escolares, cria dificuldades à realização de estudos comparativos internacionais. Para obviar esta situação, a OCDE definiu indicadores que permitem comparações entre países. Desenvolveu um programa de testagem aos alunos dos diversos países, designado PISA (*Programme for International Student Assessment*). Este programa procura testar competências dos alunos no âmbito dos desafios da sociedade do conhecimento e tem abrangido, desde 2000, as áreas de leitura, matemática e ciências, recolhendo também dados sobre as características dos próprios alunos, das famílias e de alguns fatores institucionais, que permitem contextualizar os resultados dos alunos. Através da recolha desta informação contextual o PISA tem vindo a contribuir significativamente para a base

de conhecimento disponível a partir de estatísticas oficiais nacionais sobre educação, permitindo explicar relações entre desempenho escolar e as características dos alunos, das famílias, das escolas e dos sistemas de ensino.

O PISA procura avaliar se os alunos de 15 anos, idade em que, na maioria dos países participantes, se encontram no final da escolaridade obrigatória, estão bem preparados para enfrentarem os desafios da vida quotidiana. O IAVE na sua página refere que:

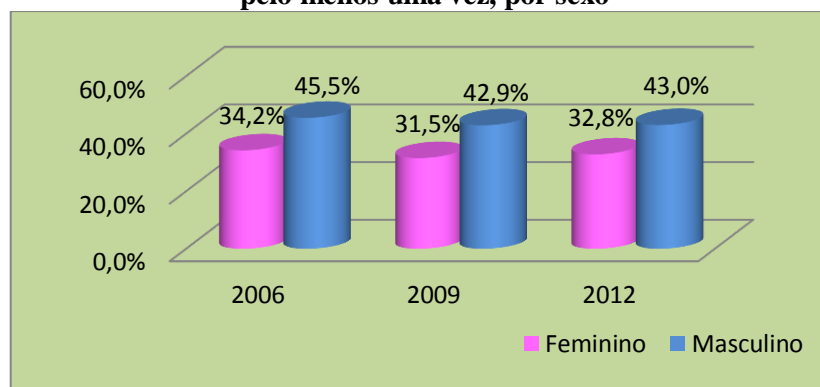
“O PISA faz um retrato de cada país e compara-o com os restantes, proporcionando uma medida da eficácia da educação de cada país face aos padrões de desempenho estabelecidos pela OCDE. Apesar de os resultados do PISA não permitirem identificar os efeitos da política educativa de cada país, permitem aos decisores reverem as suas políticas e estabelecerem metas baseadas em padrões de desempenho definidos internacionalmente.”

(IAVE, <http://iave.pt/np4/home>)

As aplicações do PISA ocorrem em ciclos de três anos e em cada ciclo é dada ênfase a uma das três áreas avaliadas (leitura, matemática ou ciências). Em 2003 e em 2012, a área principal foi a matemática.

No quadro europeu, Portugal integra o grupo de países com maior taxa de retenção. De acordo com os dados do PISA 2012, dos 31 países em análise, apenas quatro apresentam valores da retenção acima dos 30%. Em Portugal, 34,3% dos inquiridos (jovens com 15 anos) declararam já terem ficado retidos pelo menos uma vez (terceira taxa mais elevada da UE, contra a média OCDE de 13%) e desses 23,3% declararam que ficaram retidos pelo menos um ano no 1.º ou no 2.º ciclo do ensino básico. Mais de 7,5% apresentam no seu percurso mais de uma retenção. A taxa de retenção difere em função do sexo (maior incidência no sexo masculino) e vai agravando-se ao longo da escolaridade, com particular evidência no 2.º ciclo. Entre 2006 e 2012 a percentagem manteve-se relativamente estável assim como a desigualdade entre os dois sexos. Em 2012, 43% dos alunos do sexo masculino ficaram retidos e 32,8% do sexo feminino também (Gráfico 1.8).

**Gráfico 1.8 – Alunos portugueses inquiridos no PISA que ficaram retidos pelo menos uma vez, por sexo**



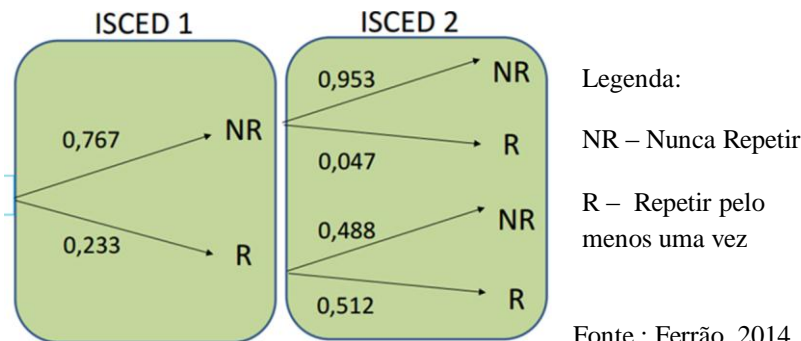
Fonte: Dados PISA

Nas provas do PISA os resultados dos alunos que apresentam pelo menos uma retenção são significativamente inferiores aos dos alunos que nunca foram retidos (Figura 1.1). Ferrão (2015) através da aplicação de métodos estatísticos, aos dados portugueses do PISA 2012, concluiu que:

“a probabilidade do aluno nunca ficar retido em algum ano de escolaridade no ISCED 1 (1.º e 2.º ciclos do ensino básico) é de 0,77 e a de ficar retido pelo menos um ano é de 0,23. Ao longo do ISCED 2 (3.º ciclo do ensino básico), o aluno que nunca ficou retido no ISCED 1 tem uma probabilidade de 0,95 de assim continuar, e uma probabilidade de 0,05 de ficar retido pelo menos um ano desse ciclo. Já o aluno que ficou retido pelo menos um ano no ISCED 1, tem probabilidade de 0,49 de não ficar retido no ISCED 2 e probabilidade de 0,51 de ficar retido pelo menos um ano nesse ciclo de ensino. Dito de outro modo, a maioria (51%) dos alunos que reprova no 1.º ou 2.º ciclos do ensino básico volta a reprovar no 3.º ciclo. Deve notar-se como a probabilidade de retenção no 3.º ciclo difere consoante se trate de um aluno que foi retido até ao 6.º ano de escolaridade ou não. Naquele caso a probabilidade é 0.51 enquanto neste é 0.05. Estas estimativas sugerem que a retenção precoce é um forte preditor da retenção tardia.”  
(Ferrão, 2015:10)

Ferrão (2015) associa ainda a probabilidade de retenção com o nível socioeconómico e com o autoconceito do aluno, concluindo da existência de correlação negativa, ou seja, quanto maior o nível socioeconómico ou o autoconceito menor a probabilidade de retenção.

**Figura 1.1 – Probabilidades de Retenção no ISCED 1 e ISCED 2 (dados PISA 2012)**

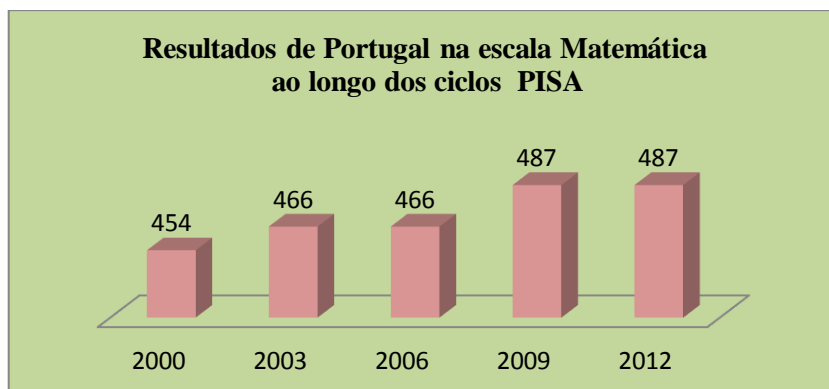


No que à Matemática diz respeito, o relatório da OCDE refere que Portugal melhorou o seu desempenho. Em 2003 estava abaixo do Luxemburgo, dos Estados Unidos, da República Checa, da França, da Suécia, da Hungria, da Espanha, da Islândia e da Noruega e em 2012 conseguiu alcançá-los. Referindo ainda, que Portugal é um dos países que conseguiu, simultaneamente, reduzir o universo dos alunos que se saem muito mal neste tipo de testes de literacia e aumentar o número dos jovens que se destacam muito pela positiva (“*top performers*”) (Relatório PISA, 2012).

Os resultados do PISA indicam que no que se refere à literacia em matemática, a média portuguesa, em 2000 era 454, a qual subiu ligeiramente para 466 em 2003, mantendo-se em 2006, tendo voltado a subir em 2009 para 487 mantendo-se este valor em

2012, o que significa uma progressão de 33 pontos relativamente a 2000 (Gráfico 1.9). Em 2012, a pontuação média dos alunos portugueses é um pouco inferior à média dos alunos da OCDE, de 494 pontos, diferença de 7 pontos.

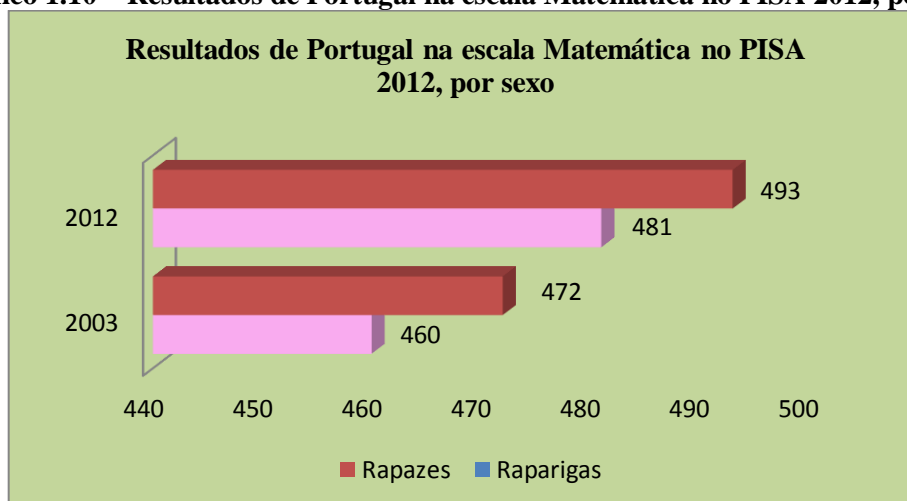
**Gráfico 1.9 – Resultados de Portugal na escala Matemática ao longo dos ciclos PISA)**



Fonte: OCDE

Em 2012, a distribuição por sexo dos resultados nacionais na escala da matemática (Gráfico 1.10) é mais favorável aos rapazes (493 pontos) do que às raparigas (481 pontos). A diferença de desempenhos entre os sexos mantém-se inalterável entre os ciclos de 2003 e 2012, embora se tenha registado um aumento expressivo de 21 pontos na média de ambos. A diferença registada em Portugal nos desempenhos de rapazes e raparigas na escala da matemática está próxima da observada para a média da OCDE, que assinala 499 pontos para os rapazes e 489 para as raparigas.

**Gráfico 1.10 – Resultados de Portugal na escala Matemática no PISA 2012, por sexo**



Fonte: PISA 2012

Também este indicador apresenta algumas limitações, o desempenho global dos alunos do país na escala de matemática é em termos de pontuação média, como tal, quando se interpreta o desempenho, só se deve ter em conta as diferenças entre países

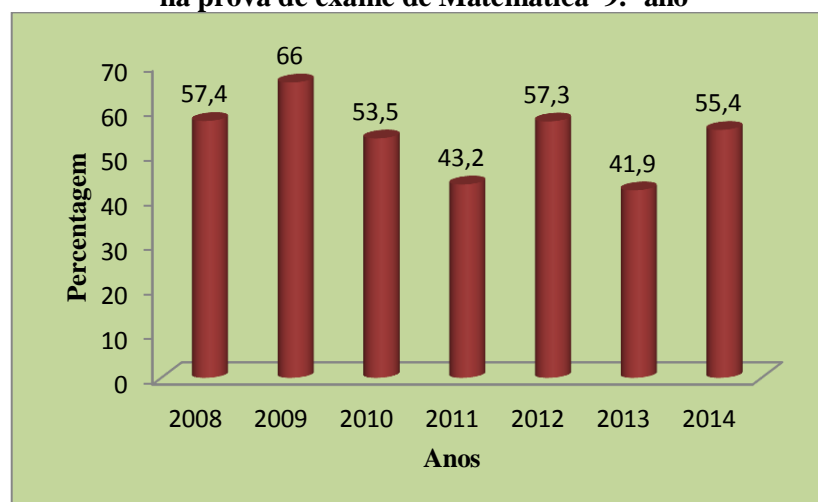
estatisticamente significativas. Como estes dados estatísticos são estimativas do desempenho nacional baseadas em amostras de alunos e não os valores que poderiam ser calculados se todos os alunos de um país participassem, existe um grau de incerteza, dado pelo erro padrão, o que não permite determinar a posição exata dos países nas comparações internacionais. As pontuações médias do desempenho são normalmente utilizadas para avaliar a qualidade das escolas e dos sistemas de ensino, no entanto, essas pontuações médias não fornecem uma visão global do desempenho dos alunos e podem esconder desigualdades no seio de uma turma, de uma escola ou de um sistema educativo (Relatório PISA, 2003).

Os exames nacionais também são instrumentos de avaliação sumativa externa, sendo a classificação em pauta um indicador de desempenho dos alunos e das escolas. Segundo o GAVE (2011), os exames contribuem para a certificação das aprendizagens e competências adquiridas pelos alunos, constituindo-se como instrumentos que contribuem para a regulação das práticas educativas, visando a melhoria sustentada das aprendizagens.

A partir de 2013 no ensino básico passaram a realizar-se exames nacionais, a Português e a Matemática, em todos os anos terminais dos três ciclos do ensino básico. De acordo com os princípios legislativos enunciados no DL n.º 139/2012, de 5 de julho, pretende-se que a avaliação interna, da responsabilidade da escola e dos seus professores, seja acompanhada de provas e exames (avaliação externa) de forma a permitir a obtenção de resultados fiáveis sobre a aprendizagem, fornecendo indicadores da consecução das metas curriculares e dos conteúdos disciplinares definidos para cada disciplina. O exame contribui em 30% para a classificação final do aluno na disciplina.

Entre 2008 e 2014, no exame nacional de matemática de 9.º ano a percentagem de alunos com classificação igual ou superior a 50%, (Gráfico 1.11), oscilou significativamente, entre os 66%, em 2009, e os 41,9%, em 2013, não sendo perceptível nenhuma linha de tendência (ascendente ou descendente). Neste período em dois anos (2011 e 2013) mais de metade dos alunos não conseguiu alcançar classificação igual ou superior a 50%, o que mostra a existência de uma percentagem elevada de alunos com dificuldades significativas nesta disciplina.

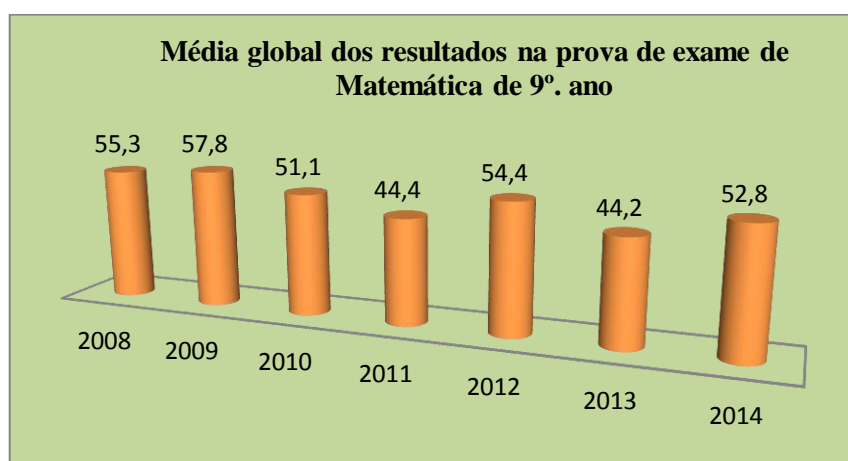
**Gráfico 1.11 – Alunos com classificação igual ou superior a 50% na prova de exame de Matemática 9.º ano**



Fonte: Pordata

A classificação média, na prova de exame de Matemática do 9.º ano, (Gráfico 1.12), ronda os 50%, no entanto em 2011 e 2012 a média foi inferior a 50%.

**Gráfico 1.12 – Média global dos resultados na prova de exame de Matemática de 9.º ano**



Fonte: Pordata

## 1.6. Matemática e o insucesso escolar

O insucesso na disciplina de Matemática reveste-se de uma importância particular, dado que esta ser uma disciplina “chave” no sistema de ensino. Todas as pessoas necessitam de uma preparação sólida em matemática, quer para desempenhar da melhor maneira a sua atividade profissional, quer para atingir realização pessoal ou para participar de forma ativa na sociedade. A Matemática está presente praticamente em todas as áreas

do saber: na Economia, na Informática, na Mecânica, na Análise Financeira, na Biologia, na Medicina, entre tantas outras.

No entanto, “A Matemática é uma das disciplinas que mais contribui para o insucesso educativo de muitos alunos” (Ponte,1988:10). A ideia de que é “normal” os alunos terem maus resultados em Matemática está muito enraizada na sociedade. Para tal concorrem as baixas expectativas dos alunos e dos outros em relação aos alunos, a ausência de uma cultura de valorização do esforço e a baixa escolarização das famílias, o que acaba por levar os alunos a desinvestir e a desistir cedo de mais da disciplina.

O insucesso na Matemática é uma realidade incontornável, visível não apenas pelos maus resultados dos alunos em testes e exames, mas também pela generalizada dificuldade dos alunos na resolução de problemas, no raciocínio matemático e até na realização das tarefas simples e, sobretudo, pelo seu desinteresse em relação à disciplina.

Ponte (1994) refere que cada um dos atores sociais que intervém ou acompanham o processo de ensino-aprendizagem apontam diferentes causas para o insucesso a Matemática. Para os professores, as causas do insucesso são a “má preparação” dos alunos em anos anteriores, o nível socioeconómico e cultural das famílias. Alunos provenientes de famílias com nível socioeconómico e cultural muito baixo, ou com um nível aceitável mas que não incentivam suficientemente os filhos têm mais insucesso. Os professores indicam que os alunos não se esforçam, não prestam atenção nas aulas nem estudam em casa. Referem também que os currículos excessivamente longos e a necessidade do seu cumprimento não permite dar maior acompanhamento aos alunos mais “fracos”, agravado pelo grau de dificuldade de alguns conteúdos. Em suma, responsabilizam os alunos, as famílias, os professores dos anos anteriores, os currículos e as características próprias da disciplina. Para os alunos, a principal razão do insucesso na disciplina resulta de dificuldades de compreensão e das metodologias usadas pelos professores. Não percebem para que serve nem porque são obrigados a estudá-la. Alguns alunos interiorizam desde muito cedo de que não são capazes de alcançar o sucesso. Dum modo geral, culpam-se a si próprios, aos professores, ou atribuem às características específicas da disciplina. Para os pais e para a opinião pública em geral, os responsáveis são os professores, que não ensinam convenientemente e os alunos que não se esforçam o suficiente. Referindo também o peso de fatores socioculturais e a dificuldade da Matemática. Pode dizer-se que as causas apontadas por todos andam todas à volta dos mesmos fatores embora com

ênfases diferentes: a disciplina, o currículo, o professor, o aluno, as razões de ordem social e cultural. Por vezes as causas misturam-se com os sintomas, isto é, alguns aspetos são reveladores de insucesso mas não o explicam por si só. Todos referem as características da disciplina e apenas os alunos admitem ter alguma responsabilidade no insucesso.

Bidarra (1983) num estudo sobre insucesso em matemática conclui que “o programa, o método, o professor, as próprias atitudes dos professores em relação à disciplina que lecionam (...) poderão estar implicados neste problema” (Bidarra, 1993 *apud* Cabrita, 1993:59).

Ponte (1988) considera que constitui obstáculo à redução do insucesso da disciplina:

“(...) uma matemática formalizada, em que se o privilegia o cálculo, o simbolismo e a abstração pura, desprezando os significados, a natureza problemática [da matemática] e os contextos de aprendizagem (...)”.

(Ponte, 1988:19)

Os resultados de um inquérito realizado em 2004, promovido pelo EDUCARE.PT e pela Porto Editora, dirigido a professores de Matemática, apontam com principais causas do insucesso a Matemática a falta de bases, a desmotivação dos alunos e o reduzido número de aulas. A propósito dos resultados deste inquérito Maria Augusta Neves (professora e autora de manuais) acrescenta que existem um conjunto de outras variáveis que contribuem para o fraco desempenho dos alunos portugueses na disciplina, que vão desde a falta de hábito de trabalho dos alunos, à indisciplina, ao reduzido número de horas de aulas, à carga horária desadequada de matemática, entre outras.

Barros (1988) sustenta que o insucesso em matemática é um fenómeno reversível, desde que exista um investimento no desenvolvimento curricular, na formação de professores e na investigação educacional.

Rosário (2004) sugere um esforço concertado entre todos os intervenientes para promover o sucesso educativo, considerando como entrave a esse sucesso “entrincheirarmo-nos como professores, alunos ou encarregados de educação, atribuindo aos demais a responsabilidade pelo que acontece, mas sobretudo pelo que deveria acontecer (...)” (Rosário, 2004:97).

Também Cabrita (1993) aponta como principal medida para reduzir o insucesso a sensibilização de todos os intervenientes no processo de modo a olharem a matemática de outra forma, desmitificando-a,

“Urge desdramatizar o carácter transcendental, inacessível, desta disciplina, consciencializando-se da sua mutualidade, do seu carácter evolutivo, da sua relação com a realidade, da sua dinâmica própria, da sua flexibilidade. Urge deixar de associar a matemática a cálculo, a

demonstração, a raciocínios dedutivos, a abstracção, a certeza, e passar a considera-la como uma disciplina essencialmente criativa que se experimenta, que se conjectura, que se questiona, que tem avanços e recuos, que é aliciante.

(Cabrita, 1993:61)

Para além dessa desmitificação, Cabrita (1993) sugere outras medidas para reduzir o insucesso, nomeadamente a formulação e resolução de problemas, como forma de alcançar o conhecimento matemático; o prolongamento da disciplina para além da dimensão curricular, através de atividades extra curriculares estruturadas de frequência livre que vão de encontro aos gostos do aluno; modernização dos planos curriculares; e mudança nos métodos de ensino.

Oliveira (2010) levou a cabo uma investigação para conhecer, analisar e compreender as causas do insucesso, na disciplina de matemática, dos alunos de uma turma, na transição do 6.º para o 7.º ano na perspectiva da professora e dos alunos. Na perspectiva dos alunos, as principais causas de insucesso escolar são aquelas que se relacionam com o próprio aluno, nomeadamente pouco estudo, falta de organização e métodos de trabalho, não apontando aspetos a ser modificados nas aulas de matemática mas reconhecendo que com a utilização de recursos didáticos diversificados aprendem mais e melhoram os resultados na disciplina. Para a professora as principais causas de insucesso escolar são, fundamentalmente, as que se relacionam com: os próprios alunos (pouco trabalho individual; a falta de atenção; de capacidades de aquisição, interpretação e relação dos conteúdos lecionados; de pré-requisitos e de interesse demonstrados pela grande maioria dos alunos face à disciplina); a organização escolar (organização das turmas); a carência de recursos didáticos e a extensão dos programas. Também para esta a utilização de recursos didáticos diversificados tornam as aulas mais entusiasmantes e motivadoras, o que se reflete na aprendizagem dos alunos.

### **1.7. Estratégias, políticas e medidas de Inclusão e Promoção do Sucesso Educativo**

O relatório conjunto Eurydice/Cedefop (2014), intitulado “Combate ao Abandono Precoce no Ensino e Formação Profissionais na Europa - Estratégias, políticas e medidas”, refere que o abandono precoce produz impacto negativo nas oportunidades dos jovens no mercado de trabalho, o que se traduz num elevado custo para os indivíduos, para a sociedade e para a economia. A conclusão da escolaridade pode, pelo contrário, conduzir a mais e melhores oportunidades de emprego e a melhores

condições de saúde para o indivíduo, melhores taxas de produtividade, menores custos sociais com o setor público, maior crescimento económico e maior coesão social.

A orientação escolar e profissional é considerada pela maioria dos países europeus como uma das medidas essenciais para combater o abandono escolar precoce, ficando a sua implementação a cargo das escolas.

Em 2011, o Conselho de Educação adotou uma recomendação sobre medidas políticas para reduzir o abandono precoce, destacando a necessidade de políticas objetivas e eficazes baseadas nas circunstâncias nacionais, e recomendando que os Estados Membros introduzissem uma estratégia abrangente para resolver o problema, a qual deveria abranger todos os níveis de ensino e envolver todas as áreas e atores políticos relevantes.

A recomendação do CNE sugere que as estratégias abrangentes de combate ao abandono precoce incluam três tipos de medidas:

- medidas de prevenção, que visam resolver os problemas de raiz que podem, eventualmente, resultar em abandono precoce;
- medidas de intervenção, que visam combater as dificuldades emergentes vivenciadas pelos alunos, através da melhoria da qualidade da educação e formação e prestação de apoio específico;
- medidas de compensação, que criam novas oportunidades para aqueles que prematuramente abandonaram o ensino.

Portugal, tal como todos os países na UE, tem em vigor políticas e medidas com vista a ajudar a reduzir o abandono precoce. A estratégia do Governo Português para promoção da aprendizagem e sucesso escolar inclui um conjunto abrangente de medidas que referem explicitamente a redução do abandono como um dos resultados esperados. O Conselho de Ministros, em 21 de junho de 2012, aprovou o DL nº 176/2012, do qual consta um capítulo específico intitulado “Medidas preventivas do insucesso e do abandono escolares” para o ensino básico e ensino secundário. Este capítulo compreende as principais medidas políticas orientadas para alunos em risco e suas famílias.

O relatório da Eurydice e Cedefop Report (2014) sugere, para Portugal, como medidas políticas abrangentes:

- rever os currículos;

- reduzir a retenção escolar, oferecendo apoio adicional aos alunos que não obtenham aprovação nos exames de modo a que tenham uma segunda oportunidade;
- permitir que as escolas introduzam um segundo professor em algumas turmas com mais insucesso; fomentar práticas colaborativas entre os alunos; construir grupos temporários e flexíveis de alunos, independentes das suas turmas, para lhes fornecer apoio, promover a sua aprendizagem e melhorar o seu rendimento;
- aumentar a flexibilidade e a permeabilidade dos percursos educativos, através de parceria entre escolas, empresas e municípios;
- desenvolver ações de formação contínua especializada para os professores;
- promover no final do secundário a transição suave dos alunos com necessidades educativas especiais para a vida ativa;
- criar uma rede nacional de centros de orientação escolar;
- apoiar os alunos em risco de abandono através da criação na escola de equipas de apoio multidisciplinares;
- tornar obrigatório as comissões municipais agirem imediatamente em casos de absentismo escolar e as escolas agirem sobre dificuldades de aprendizagem precoces em áreas de aprendizagem com a leitura, a escrita e a matemática;
- aumentar a autonomia das escolas e responsabilização dos diretores escolares;
- criar de planos de apoio pedagógico para alunos com dificuldades de aprendizagem, em cooperação com os pais, professores e psicólogos escolares.

As principais políticas e medidas desenvolvidas recentemente para combater o abandono precoce incluem:

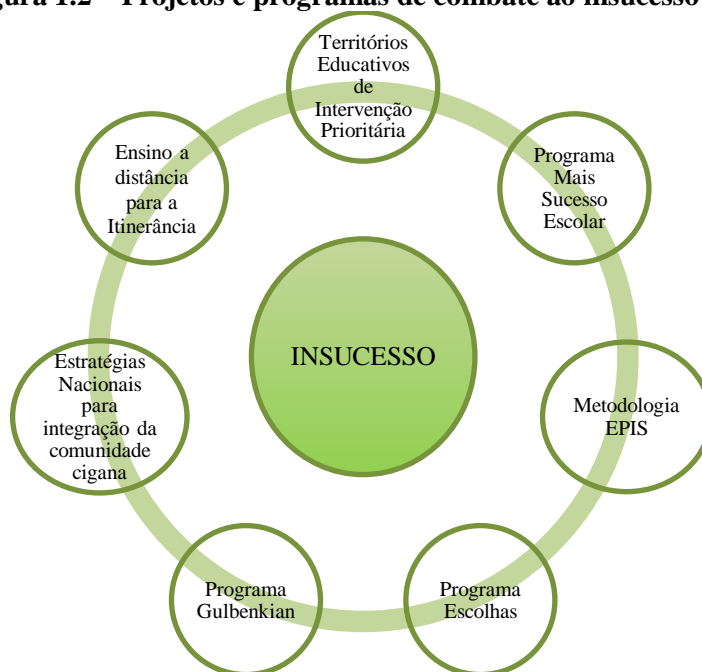
- lançamento de uma avaliação externa nacional da qualidade da educação infantil, a fim de criar diretrizes pedagógicas para 0 a 3 anos de idade e rever o currículo existente para o grupo etário dos 3 a 6 anos;
- Territórios Educativos de Intervenção Prioritária (TEIP);
- fornecimento gratuita de atividades extra-curriculares (desporto, artes, inglês) a todos os alunos do 1.º ciclo;
- incremento do programa “Mais sucesso”
- prestação de apoio e adaptação adicional do currículo aos alunos que não têm Português como língua materna;

- contratação de pessoal especializado para apoiar os professores no desenvolvimento de atividades extracurriculares;
- identificação precoce dos alunos em risco de abandono junto das Comissões de Proteção de Crianças e Jovens (CPCJ).
- obrigatoriedade das escolas comunicar à CPCJ qualquer situação de absentismo superior a duas semanas
- diversificação da oferta de ensino.

### 1.8. Projetos e programas para combater o insucesso

Em Portugal, de acordo com o CNE (2015) estão a ser implementados diversos projetos e programas no sentido de dar resposta aos problemas de aprendizagem e abandono dos jovens que frequentam o sistema educativo (Figura 1.2).

**Figura 1.2 – Projetos e programas de combate ao insucesso**



Fonte: O autor, a partir dos dados de pesquisa

O Programa Territórios Educativos de Intervenção Prioritária é uma iniciativa governamental, implementada em escolas que se localizam em territórios economicamente e socialmente desfavorecidos, marcados pela pobreza e exclusão social, onde a violência, a indisciplina, o abandono e o insucesso escolar mais se manifestam. Este programa integra todos os ciclos do ensino básico de forma articulada com a Educação Pré-Escolar e a formação profissional e tem como objetivos centrais a prevenção e redução do abandono

escolar precoce, do absentismo, da indisciplina e a promoção do sucesso educativo de todos os alunos. A promoção da inclusão de todos os alunos na escola assenta numa intervenção integradora, envolvendo professores, pessoal não docente, técnicos das áreas das ciências sociais e humanas, famílias e comunidade, procurando promover a permanência do aluno na escola e a melhoria das aprendizagens. Uma vez que um dos principais preditores do abandono escolar precoce é o número de repetências do aluno, o programa aposta na melhoria das aprendizagens, que assenta na promoção de mudanças no trabalho efetuado pelos professores na sala de aula, tais como a implementação de novas formas de organização do grupo-turma, a promoção do trabalho articulado entre professores e a implementação de parcerias pedagógicas e de pedagogias diferenciadas.

O Programa Mais Sucesso Escolar (PMSE) envolvendo três projetos: Fénix, TurmaMais, e Tipologia Híbrida. O projeto Fénix assenta num modelo organizacional de escola, que permite dar um apoio mais personalizado aos alunos na aprendizagem nas disciplinas de português e matemática, ou outra identificada pela escola. A TurmaMais caracteriza-se por utilizar pedagogias diferenciadas e formas diversificadas de organização do grupo turma, permitindo um trabalho colaborativo através de parcerias pedagógicas. A Tipologia Híbrida é criada pelas escolas a partir da sua experiência e da sua cultura escolar e visa melhorar a qualidade das aprendizagens e o sucesso dos seus alunos, através de desenhos organizacionais próprios, apoiam grupos de alunos com ritmos de aprendizagem diferentes, através de mudanças nas práticas letivas, promotoras de uma atividade mais intensa dos alunos e de processos de comunicação na sala de aula mais diversificados. Promove o trabalho colaborativo entre os professores intervenientes no programa e, em muitas situações, proporciona um ensino mais individualizado.

A Metodologia EPIS assenta em três partes fulcrais:

- Sistema de sinalização de alunos com fatores de risco de insucesso e abandono escolares, envolvendo: aluno, família, escola e território.
- Portefólio de métodos específicos de capacitação (com promoção de competências não-cognitivas) para cada um dos eixos, cujos pesos variam, dependendo de a intervenção ser focada no 1.º, 2.º ou 3.º Ciclos.
- Sistema de monitorização de resultados quantitativos, em todos os períodos e no final de cada ano letivo.

O Programa Escolhas é um programa governamental de âmbito nacional, com a missão de promover a inclusão social de crianças e jovens de contextos socioeconómicos vulneráveis, visando a igualdade de oportunidades e o reforço da coesão social. Funciona através de protocolos com os consórcios de projetos locais de inclusão social em comunidades e territórios de descendentes de imigrantes e minorias étnicas.

O Programa Gulbenkian é de médio prazo (duração prevista de 6 anos de execução) e que deverá integrar avaliação externa. Desenvolvido através do apoio à execução de projetos educativos replicáveis, à realização de estudos e à organização de seminários, workshops e outras reuniões, nacionais e internacionais.

A Estratégia Nacional para a Integração das Comunidades Ciganas (2013), tem como objetivo garantir o acesso de crianças dessas comunidades à educação pré-escolar e promover a conclusão da escolaridade obrigatória e de acesso ao Ensino Superior.

O Ensino a Distância para a Itinerância é um projeto do Ministério da Educação que tem por objetivo garantir a escolaridade a alunos que, por necessidade de deslocação dos seus pais e encarregados de educação, não podem frequentar a sua escola com a regularidade desejável. Visa assim proporcionar condições para a aprendizagem de alunos filhos de profissionais itinerantes, abrangendo também alunos hospitalizados.

**CAPÍTULO II**  
**ANÁLISE MULTIVARIADA**

## **2. ANÁLISE MULTIVARIADA**

Conforme Hair et al. (2009), a análise multivariada auxilia na formulação de questões relativamente complexas de forma específica e precisa, possibilitando a condução de pesquisas teoricamente significativas.

As técnicas multivariadas, por permitirem o tratamento simultâneo de diversas variáveis, podem oferecer ao pesquisador um material bastante robusto para a análise dos dados da pesquisa.

A escolha das técnicas de análise multivariada a utilizar neste estudo teve em linha de conta os objetivos do problema de pesquisa, o tipo de relação ou classificação entre as variáveis, o número de variáveis e o tipo de escala utilizada.

Nesta pesquisa o investigador optou pelas técnicas Regressão Logística e Análise de *Clusters*.

Este capítulo está dividido em duas secções. A secção um, referente à Regressão Logística, onde é abordada a Regressão Logística Univariada e a Regressão Logística Múltipla. A secção dois aborda a Análise de *Clusters*.

### **2.1. Análise de Regressão Logística**

A análise de regressão é uma técnica estatística cujo o objetivo é descrever a relação entre duas ou mais variáveis, através de um modelo com bom ajuste. Os modelos de regressão, quando se ajustam bem, as suas propriedades têm um elevado poder explicativo, são relativamente fáceis de interpretar e de utilizar em programas informáticos, nomeadamente no SPSS (*Statistical Package for the Social Sciences*).

Segundo Marôco (2011) quando a variável dependente é qualitativa, assumindo apenas valores de classes discretas e mutuamente exclusivas, a técnica de análise de regressão a utilizar é a regressão categorial, a qual serve os mesmos propósitos que a regressão linear em termos de inferências e de estimativas.

A regressão categorial toma diferentes designações consoante o tipo de variável dependente qualitativa. Assim, se a variável dependente é nominal dicotómica estamos perante a regressão logística, se a variável dependente é nominal policotómica estamos perante a regressão multinomial e se a variável dependente for ordinal (existe uma ordem

natural entre as possíveis categorias) deve usar-se a regressão ordinal. As variáveis independentes ou preditoras podem ser quantitativas ou qualitativas.

A análise categorial apresenta algumas vantagens relativamente à análise discriminante, nomeadamente:

- aceita variáveis independentes qualitativas e quantitativas;
- não assume relações lineares entre a variável dependente e as variáveis independentes
- não exige que as variáveis independentes apresentem distribuição normal;
- menos sensível a *outliers*.

A regressão logística é uma técnica de regressão utilizada frequentemente quando se procura descrever a relação entre uma variável dependente (resposta) e uma ou mais variáveis independentes (explicativas ou covariáveis). Permitindo ainda avaliar a significância de cada uma das variáveis independentes no modelo (Marôco, 2011).

Na regressão logística a variável dependente é categórica podendo as variáveis independentes serem qualitativas ou quantitativas (Marôco, 2011). As variáveis categóricas podem assumir apenas alguns valores particulares de resposta, podendo estes ser dicotómicos cuja resposta possui apenas dois níveis (fracasso ou sucesso) ou policotómicos no qual a resposta pode assumir mais de dois níveis mutuamente exclusivos (Hosmer e Lemeshow, 2013). Se a variável dependente é nominal dicotómica designa-se por modelo de regressão logística binário, caso a variável nominal seja policotómica designa-se por modelo de regressão logística multinomial ou apenas por regressão multinomial. A regressão logística binária constitui assim um caso particular da regressão multinomial (Marôco, 2011).

O modelo de regressão logístico binário é semelhante ao modelo de regressão linear. No entanto, no modelo logístico a variável resposta é binária, que assume dois valores, como por exemplo,  $Y_i = 0$  e  $Y_i = 1$  denominados “fracasso” e “sucesso”, respetivamente. Sendo o “sucesso” referente ao evento de interesse.

O modelo logístico permite “modelar a ocorrência, em termos probabilísticos, de uma de duas realizações da variável dependente – Avaliar a significância de cada uma das variáveis independentes no modelo.” (Marôco, 2011:804).

O objetivo principal da regressão logística é descrever a relação entre a variável dependente e as variáveis independentes.

Segundo Hair *et al* (2009), a regressão logística é uma forma de regressão que procura prever e explicar uma variável binária categórica. Estes autores referem ainda que:

“A regressão logística tem ampla aplicação em situações nas quais o principal objetivo é identificar o grupo ao qual pertence um objeto (e.g., pessoa, firma, produto). As aplicações potenciais incluem prever qualquer coisa onde o resultado é binário (por exemplo, sim/não). Essas situações incluem o sucesso ou fracasso de um novo produto, decidir se a uma pessoa deve ser concedido crédito ou predizer se uma firma será bem sucedida. Em cada instância, os objectos caem em um dos dois grupos, e o objetivo é prever e explicar as bases para cada membro do grupo de objetos através de um conjunto de variáveis independentes selecionado pelo investigador.”

(Hair *et al.*,2010:315)

Uma das desvantagens da regressão logística é a de a interpretação do modelo ser menos imediata do que na análise discriminante linear, uma vez que a visão geométrica do modelo não é tão simples (Barth, 2004).

A regressão logística univariada ou simples é usada para o caso de regressão com apenas uma variável independente e regressão logística múltipla para o caso de mais do que uma variável independente.

O método de Regressão Logística é utilizado neste estudo para avaliar quais os fatores que afetam o (in)sucesso dos alunos na disciplina de matemática. A regressão logística univariada foi utilizada para reduzir o número de variáveis independentes.

### 2.1.1. Regressão Logística Univariada

O modelo de regressão logística univariada considera apenas uma variável independente na relação com a variável dependente dicotómica.

#### 2.1.1.1. Modelo de Regressão Logística Univariada

Supondo que  $(x_i, y_i)$  é uma amostra independente com  $n$  observações em que  $x_i$  é o valor da variável independente,  $y_i$  representa o valor da variável dependente dicotómica e  $i = 0, 1, \dots, n$ . Além disso, assume-se que a variável dependente está codificada como 0 ou 1, que representa a ausência ou a presença, respetivamente.

A probabilidade de sucesso do modelo logístico univariado é dada por:

$$\pi(x_i) = P(Y = 1|X = x_i) = \frac{e^{(\beta_0 + \beta_1 x_i)}}{1 + e^{(\beta_0 + \beta_1 x_i)}} \quad (2.1)$$

e probabilidade de fracasso é dada por:

$$1 - \pi(x_i) = P(Y = 0|X = x_i) = \frac{1}{1 + e^{(\beta_0 + \beta_1 x_i)}} \quad (2.2)$$

Onde  $\beta = (\beta_0, \beta_1)^T$  é o vetor de parâmetros desconhecidos. Os parâmetros 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.

Se  $\beta_1 < 0$  então  $\pi$  será crescente e se  $\beta_1 > 0$  então  $\pi$  será decrescente. Quando  $x$  tende para infinito, então  $\pi(x)$  tende para zero se  $\beta_1 < 0$  e tende para um quando  $\beta_1 > 0$ . Caso  $\beta_1 = 0$ , a variável dependente  $Y$  é independente da variável  $X$ .

Em qualquer problema de regressão, a quantidade a ser modelada é o valor médio da variável dependente dado os valores das variáveis independentes. A esta quantidade chama-se média condicional, denotada por  $E(Y|X = x_i)$ , em que  $Y$  é a variável dependente e  $x_i$  os valores das variáveis independentes. Em regressão logística como a variável dependente é dicotômica a média condicional tem de ser maior ou igual a zero e menor ou igual que 1, ou seja  $0 \leq E(Y|X = x_i) \leq 1$ .

Em termos de distribuição condicional da variável dependente, o valor da variável dependente dado  $x$ , é expresso por  $Y_i = \pi + \varepsilon_i$ , como a quantidade  $\varepsilon_i$  pode assumir apenas um de dois possíveis valores (ou seja,  $\varepsilon_i = 1 - \pi(x_i)$ , para  $y = 1$  ou  $\varepsilon_i = -\pi(x_i)$ , para  $y = 0$ ) então  $\varepsilon_i$  tem distribuição com média zero e variância  $\pi(x_i)[1 - \pi(x_i)]$ . Isto é, a distribuição condicional da variável dependente segue uma distribuição binomial com probabilidade dada pela média condicional,  $\pi(x_i)$  (Hosmer e Lemeshow, 2009).

A transformação de  $\pi(x_i)$  é interpretada como o logaritmo da razão de chances entre  $\pi(x_i)$  e  $1 - \pi(x_i)$ , sendo definida como:

$$g(x) = \ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = \beta_0 + \beta_1 x_i. \quad (2.3)$$

### 2.1.1.2. Estimação dos parâmetros

Para ajustar o modelo de regressão logística segundo a equação (2.1) é necessário estimar os parâmetros desconhecidos  $\beta_0$  e  $\beta_1$ . O método utilizado para estimar esses parâmetros é o método da máxima verosimilhança que permite obter valores para os parâmetros desconhecidos, que maximizam a probabilidade de obter o conjunto de observações.

A função de distribuição da probabilidade de  $y_i$ , com  $y_i \sim Ber(\pi(x_i))$  é dada por:

$$\pi(x_i)^{y_i} [1 - \pi(x_i)]^{1 - y_i}. \quad (2.4)$$

Uma vez que as observações são independentes, a função de verossimilhança é dada por:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n \pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{1-y_i}. \quad (2.5)$$

Como o princípio da máxima verossimilhança é estimar o valor de  $\boldsymbol{\beta}$  que maximize  $l(\boldsymbol{\beta})$ , aplicando o logaritmo a (2.5) obtêm-se a expressão:

$$L(\boldsymbol{\beta}) = \ln[l(\boldsymbol{\beta})] = \sum_{i=1}^n \{y_i \ln[\pi(\mathbf{x}_i)] + (1 - y_i) \ln[1 - \pi(\mathbf{x}_i)]\}. \quad (2.6)$$

Para encontrar o valor de  $\boldsymbol{\beta}$  que maximiza  $L(\boldsymbol{\beta})$  diferencia-se  $L(\boldsymbol{\beta})$  em relação a  $\beta_0$  e  $\beta_1$  e iguala-se a zero as expressões resultantes. Estas equações, conhecidas como as equações de probabilidade, são dadas por:

$$\sum_{i=1}^n (y_i - \pi(\mathbf{x}_i)) = 0. \quad (2.7)$$

$$\sum_{i=1}^n \mathbf{x}_i (y_i - \pi(\mathbf{x}_i)) = \mathbf{0}. \quad (2.8)$$

Como estas equações não são lineares, são necessários métodos iterativos para resolvê-las (recurso a softwares).

Normalmente os *outputs* de computador além dos valores de  $\beta_i$  fornecem os respectivos erros padrão ( $SE_{\beta_i}$ ). Os valores dos  $SE_{\beta_i}$  serão utilizados para os testes de significância dos coeficientes e para o cálculos dos respectivos intervalos de confiança.

O valor de  $\boldsymbol{\beta}$  dado pela solução de equações (2.7) e (2.8) é designado por estimativa de probabilidade máxima e denotado por  $\hat{\boldsymbol{\beta}}$ .

### 2.1.1.3. Significância dos coeficientes

Segundo Hosmer e Lemeshow (2009) após estimar-se  $\boldsymbol{\beta}$ , é conveniente saber se a variável que teve o  $\boldsymbol{\beta}$  estimado é relevante ou não na análise, ou seja, se o modelo com a variável independente em questão explica a variável dependente melhor que o do modelo sem a variável independente.

Os métodos em regressão logística seguem o mesmo princípio que em regressão linear: comparar os valores observados com os valores preditos da variável dependente. Essa comparação é baseada na função de verossimilhança. É feita a comparação entre a função de verossimilhança dos valores observados na amostra e a função de verossimilhança do modelo saturado (aquele que contém tantos parâmetros quanto o número de pontos da amostra) é conhecida como teste da razão de verossimilhança.

$$D = -2 \times \ln \left[ \frac{\text{verossimilhança do modelo ajustado}}{\text{verossimilhança do modelo saturado}} \right] =$$

$$= -2 \times \sum_{i=1}^n \left[ y_i \times \ln \left( \frac{\hat{\pi}(x_i)}{y_i} \right) \right] + (1 - y_i) \times \ln \left( \frac{1 - \hat{\pi}(x_i)}{1 - y_i} \right). \quad (2.9)$$

Para avaliar a significância de uma variável independente, comparam-se os valores de  $D$  com e sem a variável em questão e verifica-se se é significativo. Para testar a significância da variável faz-se:

$$G = D(\text{modelo sem a variável}) - D(\text{modelo com a variável}) =$$

$$= -2 \ln \left[ \frac{\text{verossimilhança sem a variável}}{\text{verossimilhança com a variável}} \right]. \quad (2.10)$$

Na hipótese de  $\beta_1 = 0$ ,  $G$  segue uma distribuição chi-quadrado com 1 grau de liberdade

Existem dois outros testes estatisticamente equivalentes: o teste de *Wald* e o teste *Score*.

O teste de *Wald* é igual à razão entre o estimador da máxima verossimilhança  $\widehat{\beta}_1$  com uma estimativa do seu erro padrão, ou seja:

$$w = \frac{\widehat{\beta}_1}{\widehat{SE}(\widehat{\beta}_1)} \quad (2.11)$$

Onde  $\widehat{SE}(\widehat{\beta}_1)$  é a estimativa do erro padrão do parâmetro estimado. Sob a hipótese nula,  $w$  segue uma distribuição normal padrão.

O teste *Score* é baseado na teoria de distribuição das derivadas de log verossimilhança e não requer o cálculo computacional do estimador de máxima verossimilhança de  $\beta_1$ . Este teste é dado por:

$$ST = \frac{\sum_{i=1}^n x_i \times (y_i - \bar{y})}{\sqrt{\bar{y} \times (1 - \bar{y}) \times \sum_{i=1}^n (x_i - \bar{x})^2}}. \quad (2.12)$$

O teste de *Wald* e o teste *Score* falham em determinados casos pelo que o teste de verossimilhança é o mais adequado (Hauck; Domer, 1977 *apud* Hosmer e Lemeshow, 2009:14-15).

#### 2.1.1.4. Estimação dos Intervalos de Confiança

Os intervalos de confiança baseiam-se no teste de Wald. Os intervalos de confiança de  $100(1 - \alpha)\%$  para  $\beta_1$  e  $\beta_0$  são:

$$\widehat{\beta}_1 \pm z_{(1-\alpha/2)} \widehat{SE}(\widehat{\beta}_1). \quad (2.13)$$

$$\widehat{\beta}_0 \pm z_{(1-\alpha/2)} \widehat{SE}(\widehat{\beta}_0). \quad (2.14)$$

## 2.1.2. Regressão Logística Múltipla

O modelo de regressão logística múltipla é uma generalização do modelo de regressão logística simples para mais do que uma variável independente.

### 2.1.2.1. Modelo de Regressão Logística Múltipla

Considerando um conjunto com  $p$  variáveis independentes, denotadas por  $x^T = (x_1, x_2, \dots, x_p)$ ;  $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ , o vetor de parâmetros desconhecidos e  $\beta_j$  ( $j$ -ésimo parâmetro associado a variável independentes  $x_j$ ), ou seja, o vetor dos coeficientes de regressão logística.

O *logit* do modelo de regressão logística múltipla é dado pela equação:

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p. \quad (2.15)$$

Onde o rácio  $\pi(x)/(1 - \pi(x))$  designado por Rácio de Verosimilhança, Chances ou *Odds*, traduz a razão entre a probabilidade do sucesso  $\pi(x)$  face à probabilidade do insucesso  $1 - \pi(x)$ .

O modelo de regressão logística múltipla, para estimar a probabilidade de sucesso de uma determinada realização  $i$ , é dado por:

$$\pi(x) = P(Y = \mathbf{1}|x) = \frac{e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}}{\mathbf{1} + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}}. \quad (2.16)$$

Por sua vez, a probabilidade de fracasso é dada por:

$$\mathbf{1} - \pi(x) = P(Y = \mathbf{0}|x) = \frac{\mathbf{1}}{\mathbf{1} + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}}. \quad (2.17)$$

O modelo de regressão logística deve obedecer aos seguintes pressupostos:

- “- Linearidade e aditividade: a escala de *Logit*( $\pi_i$ ) é aditiva e linear (mas a de  $\pi_i$  não).
- Proporcionalidade: a contribuição para cada  $X_i$  ( $i=1, \dots, n$ ) é proporcional ao seu valor com um fator  $\beta_j$ .

- Constância de efeito: a contribuição de uma variável independente é constante, e independente da contribuição das outras variáveis independentes.
- Os erros são independentes e apresentam distribuição binomial.
- Os preditores não são multicolineares.”

(Marôco, 2011:811)

Para proceder à validação dos pressupostos do modelo pode recorrer-se à representação gráfica para análise de resíduos e para proceder ao diagnóstico da multicolinearidade calcula-se a Tolerância (T) a partir do  $R^2$  ( $T=1-R^2$ ) obtido pela regressão linear múltipla entre cada uma das variáveis independentes e as restantes variáveis independentes.

### 2.1.2.2. Estimação de parâmetros do modelo

O facto dos erros do modelo logístico não seguirem distribuição normal nem apresentarem variâncias constantes, impossibilita a obtenção de estimativas ótimas dos coeficientes do modelo pelo método dos mínimos quadrados utilizado na regressão linear. O método de ajustamento mais utilizado para estimar os parâmetros na regressão logística é o Método da Máxima Verosimilhança (*Maximum Likelihood*), o qual “estima os coeficientes de regressão que maximizam a probabilidade de encontrar realizações da variável dependente ( $y_1, y_2, \dots, y_n$ ) amostradas” (Mâroco, 2011:812).

Como cada realização de  $Y$  toma apenas dois valores (0 ou 1) sendo cada uma das observações de  $Y$  uma variável aleatória dicotómica, assume-se que a variável tem uma distribuição de Bernoulli com parâmetro de sucesso  $\pi_i$ . Assim:

$$P(Y = y_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i} . \quad (2.18)$$

Como geralmente as observações de  $Y$  são independentes, a função de verosimilhança que dá a probabilidade conjunta de se observarem  $y_1, y_2, \dots, y_n$  será dada por:

$$L(\beta) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} . \quad (2.19)$$

O princípio da máxima verosimilhança consiste em estimar o valor de  $\beta$  que maximiza a função de verosimilhança. O máximo dessa função ocorre para  $\frac{\partial L}{\partial \beta} = 0$  e  $\frac{\partial^2 L}{\partial^2 \beta} < 0$ . Aplicando o logaritmo natural, por conveniência do cálculo diferencial, a função de verosimilhança passa a ser:

$$LL = \ln[L(\beta)] = \sum_{i=1}^n [y_i x_i^T \beta - \ln(1 + e^{x_i^T \beta})]. \quad (2.20)$$

O máximo de  $LL$  ocorre com o vetor dos coeficientes para o qual  $\partial LL / \partial \beta = 0$ . O vetor que maximiza  $LL$  é o mesmo que maximiza  $L(\beta)$ . O valor de  $\beta$  que maximiza  $LL$  é encontrado após derivar-se em relação aos parâmetros  $(\beta_0, \beta_1)$ . Obtendo-se o sistema de equações não lineares nos parâmetros:

$$\sum_{i=1}^n (y_i - \pi_i) = 0 \wedge \sum_{i=1}^n x_i (y_i - \pi_i) = 0. \quad (2.21)$$

O valor de  $\beta$  é estimado iterativamente, por um algoritmo computacional que na maioria dos *softwares* é o de Newton-Raphson.

No caso do modelo logístico incluir variáveis independentes qualitativas, as variáveis, com  $k$  classes, entram no modelo recodificado em  $k - 1$  variáveis *dummy*. O cruzamento das classes dessas variáveis definem subpopulações de variáveis independentes. Sendo, a função de verosimilhança para o modelo dada por (Hosmer e Lemeshow, 2000):

$$LL = \sum_{i=1}^n [y_i \ln(\pi_i) + (n_i - y_i) \ln(1 - \pi_i)]. \quad (2.22)$$

Onde:  $n_i$  é o número de elementos de cada subpopulação;  $y_i$  é o número de sucessos de cada subpopulação;  $\pi_i$  é a probabilidade de ocorrência das subpopulações.

### 2.1.2.3. Significância e qualidade do modelo

Após o ajustamento do modelo de regressão logística “é necessário avaliar a significância e qualidade do modelo ajustado bem como a significância dos coeficientes de regressão logística.” (Marôco, 2011), recorrendo a estatísticas, tais como: Razão de Razão de Verosimilhanças (*deviance*), Qui-quadrado de *Pearson*, teste de *Hosmer e Lemeshow*, teste de *Wald* e *Pseudo-R<sup>2</sup>*.

#### i) Teste de Razão de Verosimilhanças

Para testar a significância do modelo ajustado é necessário testar as hipóteses:

$$H_0: \beta_0, \beta_1 = \dots = \beta_p = 0 \quad (\text{o modelo não é estatisticamente significativo})$$

*versus*

$$H_1: \exists_i: \beta_i \neq 0 \quad (i = 1, \dots, p) \quad (\text{o modelo é estatisticamente significativo})$$

Caso o modelo ajustado não seja significativo não é possível prever a probabilidade do “sucesso” a partir das variáveis independentes.

A estatística de teste à significância do modelo compara a verosimilhança do modelo nulo ou reduzido (modelo só com a constante, sem variáveis independentes com poder preditor) com um modelo completo.

Segundo Marôco (2011) a estatística do teste para testar a significância do modelo,  $G^2$ , é dada por:

$$G^2 = X_0^2 - X_c^2 = -2LL_0 - (-2LL_c) = -2\ln \left[ \frac{L_0}{L_c} \right] \underset{a}{\sim} \chi^2_{(p)} . \quad (2.23)$$

Onde:  $X_0^2$  é o modelo nulo ou reduzido;  $X_c^2$  é o modelo completo;  $G^2$  tem distribuição assintótica qui-quadrado com número de graus de liberdade igual ao número de variáveis independentes no modelo.

O  $G^2$  obtém-se a partir da razão das verosimilhanças dos dois modelos e é uma medida de incremento da qualidade do modelo nulo por adição das variáveis independentes. Pelo que, o modelo completo é estatisticamente significativo apenas quando a adição de uma ou mais variáveis independentes ao modelo, reduz significativamente o valor de  $-2LL$ .

No entanto, concluir que o modelo completo é significativo, não significa que o ajustamento seja bom, permite apenas afirmar que pelo menos uma variável independente incluída no modelo influencia significativamente a variável dependente como ajustado pelo modelo.

## ii) Teste da Qualidade do Ajustamento do modelo

A avaliação da qualidade do ajustamento do modelo consiste em verificar a qualidade de aderência dos valores estimados pelo modelo através da sua similitude com os valores observados.

Para testar a significância do modelo as hipóteses estatísticas são:

$H_0$ : O modelo ajusta-se aos dados *versus*  $H_1$ : O modelo não se ajusta aos dados

Para verificar a significância do modelo utilizam-se medidas como a *deviance*, a estatística Qui-quadrado de Pearson, teste de Hosmer e Lemeshow e Pseudo- $R^2$ .

A *deviance* compara a verosimilhança do modelo avaliado com a verosimilhança do modelo saturado, que prevê perfeitamente os resultados observados. Quanto mais perto estiverem os valores das duas verosimilhanças, melhor será o ajustamento do modelo (menor o valor da *deviance*).

A estatística *deviance* é dada por:

$$D = -2 \ln \left[ \frac{L_C}{L_S} \right]. \quad (2.24)$$

Onde:  $L_C$  é a verossimilhança do modelo ajustado (completou ou não);  $L_S$  é a verossimilhança do modelo saturado.

A estatística Qui-quadrado de *Pearson*,  $\chi_C^2$ , avalia a qualidade do ajustamento do modelo e é dada por:

$$\chi_C^2 = -2LL_C \overset{a}{\sim} \chi_{n-1-p}^2. \quad (2.25)$$

Onde:  $\chi_C^2$  é o modelo completo, que tem distribuição assintótica Qui-quadrado com  $n - 1 - p$  graus de liberdade.

Para que o modelo se ajuste aos dados, o  $-2LL$  deve ser o menor possível e não significativo.

O teste de Hosmer e Lemeshow avalia o modelo ajustado através das distâncias entre as probabilidades ajustadas e as probabilidades observadas. A estatística de teste é obtida com um teste de Qui-quadrado a uma tabela de contingência  $2 \times g$ . Essa tabela é construída classificando as duas classes da variável dependente dicotômica por  $g$  grupos de tamanho aproximadamente igual definidos pelos decis das probabilidades de sucesso estimadas pelo modelo. Os autores propõem, que seja utilizado  $g = 10$ , no entanto, no caso do modelo ter variáveis independentes qualitativas, pode ser inferior ou igual a 10, em função do número de células distintas resultantes do cruzamento das classes das variáveis independentes (Marôco, 2011).

A fórmula para o cálculo do teste de Hosmer e Lemeshow é dada por:

$$\chi_{HL}^2 = \sum_{i=1}^g \frac{(O_i - E_i)^2}{E_i}. \quad (2.26)$$

Onde:  $O_i$  é o número de sucessos da variável dependente em cada classe definida pelos decis;  $E_i = n_i \times \hat{\pi}_i$  é os valores esperados nessas classes, usando a média das probabilidades estimadas para cada um dos grupos e cada uma das classes;  $\chi_{HL}^2$  tem distribuição assintoticamente de  $\chi_{(g-2)}^2$  quando a dimensão da amostra é grande.

### iii) Teste à significância dos coeficientes do modelo

Se o modelo ajustado é significativo, então existe pelo menos uma variável independente linearmente relacionada com o  $Logit(\pi_i)$ . Para indentificar essa(s) variável(eis) pode recorrer-se ao Teste de Wald.

O Teste de Wald pretende testar se um determinado coeficiente é, ou não, nulo, condicionado pelos valores estimados dos outros coeficientes. As hipóteses estatísticas são:

$$H_0: \beta_i = 0 \mid \beta_0, \beta_1, \beta_{i-1}; \beta_{i+1}; \beta_p \quad \text{versus} \quad H_1: \beta_i \neq 0 \mid \beta_0, \beta_1, \beta_{i-1}; \beta_{i+1}; \beta_p$$

A estatística de teste é:

$$T_{Wald_i} = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \sim N(\mathbf{0}, \mathbf{1}) . \quad (2.27)$$

Onde:  $\hat{\beta}_i$  é o estimador de  $\beta_i$ ;  $SE(\hat{\beta}_i)$  é o estimador do erro padrão de  $\beta_i$ ;  $T_{Wald_i}$  tem distribuição t-Student, que se aproxima assintoticamente da distribuição  $N(0,1)$  quando a dimensão da amostra é grande.

#### iv) Pseudo- $R^2$

Para avaliar a qualidade do ajuste de modelos logísticos, foram desenvolvidos vários pseudo- $R^2$ , calculando o SPSS dois deles, o  $R^2$  de Cox&Snell e o  $R^2$  de Nagelkerke.

O  $R^2$  de Cox&Snell,  $R_{CS}^2$ , é calculado como:

$$R_{CS}^2 = 1 - e^{\frac{2[LL_C - LL_0]}{n}} . \quad (2.28)$$

No entanto, esta estatística mesmo que o ajustamento seja perfeito nunca atinge o valor 1.

A  $R^2$  de Nagelkerke,  $R_N^2$ , é uma correção ao  $R_{CS}^2$ , de modo que varie entre 0 e 1. A fórmula de cálculo é:

$$R_N^2 = \frac{R_{CS}^2}{1 - e^{\frac{2LL_0}{n}}} . \quad (2.29)$$

No entanto, Shtaland *et al* (2002), *apud* Marôco, 2011:821 concluíram que dos pseudo- $R^2$  o que apresenta melhor interpretabilidade é o  $R^2$  de McFadden,  $R_{MF}^2$ , dada pela fórmula:

$$R_{MF}^2 = 1 - \frac{LL_C}{LL_0} = 1 - [-2LL_C / (-2LL_0)] . \quad (2.30)$$

Onde:  $-2LL_0 = G^2 + (-2LL_C)$

Esta estatística pode ser interpretada como a proporção da redução do  $LL$  do modelo nulo, relativamente ao modelo completo.

#### 2.1.2.4. Seleção de variáveis

Para selecionar as variáveis com poder preditor pode-se recorrer a determinados algoritmos. Existem alguns métodos que auxiliam na seleção de variáveis para um modelo de Regressão Logística, assim o critério para a adição ou remoção de variáveis independentes é geralmente baseado na estatística  $F$ , comparando modelos com e sem as variáveis em questão. Estes métodos distinguem-se uns dos outros pelo que acontece à variabilidade devida ao efeito comum das variáveis independentes (quando estão correlacionadas entre si) e pelos critérios da ordem de entrada dessas variáveis na equação. Os métodos mais usados são:

i) Seleção *Forward*: método de seleção *stepwise* em que a entrada de uma variável independente no modelo é feita em função da significância da estatística “*Score*”, podendo variar a forma como a remoção da variável é feita. Essa remoção pode ser feita em função da significância do teste do rácio de verosimilhanças baseado nas estimativas condicionais dos coeficientes do modelo (*Forward*: Condicional); a partir da significância do teste do rácio de verosimilhanças baseado nas estimativas parciais de coeficientes de máxima verosimilhança do modelo (*Forward*: Rácio de Verosimilhanças); ou em função da significância do teste de *Wald* (*Forward*: *Wald*).

ii) Seleção *Backward*: num primeiro passo todas as variáveis independentes são adicionadas ao modelo, podendo variar a forma como a remoção da variável é realizada nos passos seguintes. Removidas as variáveis cuja probabilidade do rácio de verosimilhanças, baseada nas estimativas condicionais dos coeficientes do modelo, é superior ao  $p$ -value de remoção selecionado (*Backward*: Condicional). Se o  $p$ -value de remoção das variáveis são calculados a partir do teste da razão de verosimilhança baseado nas estimativas parciais de máxima verosimilhança (*Backward*: Rácio de Verosimilhanças). Se a remoção das variáveis é feita a partir da significância do teste de *Wald* (*Backward*: *Wald*).

Sendo de referir que a variável entra no modelo se o seu  $p$ -value no teste de adição for menor ou igual a 0,05 e é removida se o seu  $p$ -value no teste de remoção for superior a 0,10.

### 2.1.2.5. Análise de resíduos

Quando se realiza o ajustamento do modelo a um conjunto de dados é fundamental que as estimativas obtidas a partir desse modelo sejam resistentes a pequenas perturbações, quer no modelo quer nos dados, de modo a que o modelo ajustado apresente uma boa descrição dos dados observados. Assim, torna-se essencial realizar diagnósticos, baseados na análise de resíduos.

A análise de resíduos “permite identificar *outliers* e casos influentes na estimação do modelo.” (Marôco, 2011:823). As medidas utilizadas na análise dos resíduos são resíduos, *Leverage*, distância de Cook e *Df Betas*.

Os resíduos não standardizados podem ser definidos como a diferença entre o número de sucessos observados e o número de sucessos estimados pelo modelo. E podem ser calculados pela fórmula:

$$e_i = y_i - \hat{y}_i = y_i - n_i \hat{\pi}_i. \quad (2.31)$$

Onde:  $y_i$  é o número de sucessos observados;  $\hat{y}_i$  é o número de sucessos estimados;  $n_i$  é o número de observações da célula  $i$ ;  $\hat{\pi}_i$  é a probabilidade do sucesso estimada para a célula  $i$ .

Os resíduos standardizados ou resíduos de Pearson obtêm-se dividindo os resíduos não estandardizados pela estimativa do desvio padrão dos valores estimados, sendo a fórmula de cálculo a seguinte:

$$e'_i = \frac{e_i}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i)}}. \quad (2.32)$$

Os resíduos estandardizados têm média 0 e desvio padrão aproximadamente 1 e para amostras de grande dimensão  $e'_i \stackrel{a}{\sim} N(0,1)$ .

No entanto, estes resíduos não são completamente estandardizados uma vez que a variância depende das observações e da sua influência na estimação dos coeficientes do modelo, podendo essa influência ser estimada pela *Leverage*.

A *Leverage* é dada por:

$$h_i = n_i \hat{\pi}_i(x_i) [1 - \hat{\pi}_i(x_i)] x'_i [I(\hat{\beta})]^{-1} x'_i. \quad (2.33)$$

Onde:  $x'_i = [1, x_{1i}, x_{2i}, \dots, x_{pi}]$  é o vetor com os valores da observação  $i$  em todas as variáveis independentes;  $I(\hat{\beta}) = X'VX$  é a matriz de informação de Fisher, sendo:

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \text{ e } V = \begin{bmatrix} \hat{\pi}_1(1 - \hat{\pi}_1) & 0 & \cdots & 0 \\ 0 & \hat{\pi}_2(1 - \hat{\pi}_2) & \cdots & 0 \\ \vdots & 0 & \cdots & \vdots \\ 0 & \cdots & 0 & \hat{\pi}_n(1 - \hat{\pi}_n) \end{bmatrix}.$$

A *Leverage*, para uma observação  $i$ , se for próxima de 0 a observação é pouco importante no ajustamento do modelo, ao passo que se for próxima de 1 essa observação é importante. No entanto, “se as probabilidades estimadas para uma determinada observação  $i$  forem inferiores a 0,1 ou superiores a 0,9, os valores *Leverage* podem ser reduzidos e ainda assim essa observação ser influente no modelo” (Norusis,2005 *apud* Marôco, 2011:825).

Os resíduos “estudentizados” ou resíduos de Pearson estandardizados, apresentam variância constante e igual a 1 e obtêm-se pela fórmula:

$$r_i = \frac{e'_i}{\sqrt{1-h_i}}. \quad (2.34)$$

A *Leverage* e os resíduos “estudentizados” além de permitirem identificar *outliers* podem ser usados para avaliar a influência de uma observação no ajustamento do modelo.

A distância de Cook mede a influência da observação  $i$  sobre todos  $n$  valores ajustados  $\hat{Y}_i$ , ou seja, indica a variação dos resíduos quando a observação  $i$  é eliminada. Esta medida pode ser definida por :

$$DC_i = r_i^2 \frac{h_i}{(1-h_i)}. \quad (2.35)$$

$DC_i$  é grande quando o resíduo ou a leverage ou ambos é grande. Valores de  $DC_i$  superiores a 1 indicam observações influentes na estimação dos coeficientes do modelo.

Os *DFBetas* medem a influência da observação  $i$  na estimação de cada um dos coeficientes de regressão e pode ser calculada pela fórmula:

$$DfBeta_{ij} = \hat{\beta}_i - \hat{\beta}_{i(-j)}. \quad (2.36)$$

Onde:  $\hat{\beta}_i$  é a estimativa do coeficiente de regressão, ajustado com todas as observações;  $\hat{\beta}_{i(-j)}$  é a estimativa do coeficiente de regressão, ajustado sem a observação  $j$ .

Se os valores de *DFBetas* forem superiores a 2 as observações são influentes, devendo a observação  $j$  ser examinada com precaução, uma vez que a sua presença no modelo afeta a estimativa de  $\beta_i$ .

## 2.2. Análise de *Clusters*

A análise de *clusters* (ou de grupos) é uma técnica de análise multivariada que possibilita o agrupamento de sujeitos ou de variáveis, dos quais se conhece informação detalhada, em grupos homogêneos relativamente a uma ou mais características comuns. Todas as observações pertencentes a um mesmo *clusters* são similares e diferentes das observações pertencentes aos outros clusters, ou seja, o agrupamento é realizado de modo que os elementos de um grupo sejam o mais semelhantes possível e que os elementos de dois grupos distintos sejam o mais dissemelhantes possível (Marôco, 2011; Reis, 2001; Pereira, 2008).

Na análise de *clusters* não existem relações de dependência entre variáveis, ou seja, os *clusters* definem-se por si mesmo sem que exista relação causal entre as variáveis. Os métodos são exploratórios, a ideia é gerar hipóteses, em vez de testá-las, pelo que é necessário uma validação posterior dos resultados através da aplicação de outros métodos estatísticos.

A análise de *clusters* permite avaliar o tamanho da matriz de dados; identificar *outliers* multivariados; e levantar hipóteses relativas a relações estruturais entre variáveis.

Embora a análise de *clusters* seja bastante útil Hair et al. (2009) apontam-lhe algumas limitações, nomeadamente:

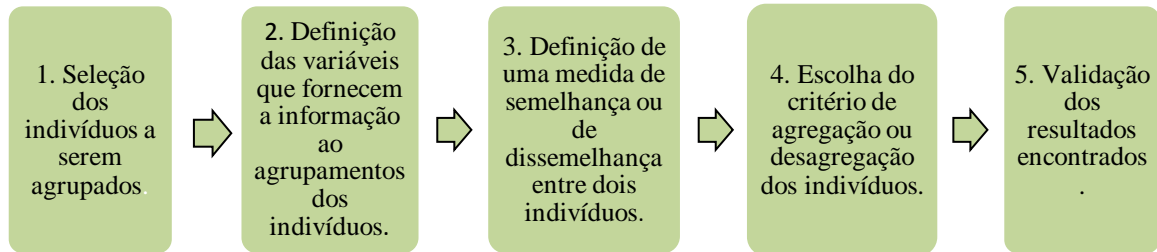
- Deverá ser usada como técnica exploratória ou descritiva e não como técnica para fazer inferências, uma vez que não tem nenhuma base estatística que permita fazer inferências a partir de uma amostra de uma população. O agrupamento obtido da aplicação da técnica não é único já que depende dos procedimentos adotados, bastando variar um ou mais elementos para obter agrupamentos diferentes.
- A análise de agrupamento cria sempre *clusters*, independentemente da existência real de qualquer estrutura nos dados. Ao usar este tipo de análise o investigador está a fazer uma suposição de alguma estrutura entre os objetos, pelo que a criação de *clusters* não valida a sua existência.
- A solução de agrupamentos não é generalizável, pois é totalmente dependente das variáveis utilizadas como base para a medida de similaridade.

Na análise de *clusters* o agrupamento dos sujeitos ou das variáveis é realizado a partir de medidas de semelhança ou de medidas de dissemelhança, numa primeira fase

entre dois sujeitos e mais tarde entre dois *clusters* recorrendo a técnicas hierárquicas ou não hierárquicas de agrupamento de *clusters*.

Segundo Reis (2011) a análise de *clusters* compreende cinco etapas (Figura 2.1):

**Figura 2.1 - Etapas da Análise de Clusters**



Fonte: Reis (2010)

### 2.2.1. Seleção dos indivíduos

A seleção dos indivíduos depende dos objetivos da análise. No caso de serem utilizados dados de análises anteriores poderá ser necessário analisá-los e retirar os indivíduos sem relevância para o estudo. No entanto, é necessário ter a preocupação de não deixar indivíduos importantes para o estudo, de fora do conjunto a analisar. No caso do conjunto dos indivíduos ser uma amostra da população é desejável que esta seja representativa para que os grupos obtidos possam ser considerados representativos dos grupos existentes na população.

### 2.2.2. Seleção das variáveis

A seleção das variáveis é a etapa mais importante, aqui o investigador tem de atender a um duplo problema:

- escolher de entre os dados disponíveis quais os mais significativos na abordagem do problema, devendo a seleção ser realizada com base em estudos anteriores, numa teoria ou nas hipóteses a ser testadas;
- atender ao tipo de variáveis utilizadas (contínuas, rácios, ordinais, nominais ou binárias), sobretudo quando estas estão definidas em diferentes unidades de medida. A aplicação da análise de *clusters* sem a standardização prévia das variáveis, leva a que qualquer medida de semelhança/distância reflecta sobretudo o peso das variáveis com maiores valores e maior dispersão, pelo que será

conveniente uma estandardização prévia. No entanto, a estandardização pode reduzir as diferenças entre os indivíduos anulando os agrupamentos naturais que possam existir dos dados, passando todas as variáveis a ter o mesmo peso, o que poderá não ser conveniente. Para obviar esta situação, caso o investigador recolha dados primários, convém que as perguntas sejam definidas na mesma unidade de medida.

A inclusão de uma variável irrelevante pode conduzir a resultados distorcidos (Pereira, 2008). O número de variáveis deve corresponder a uma situação “equilibrada”.

### 2.2.3. Medidas de semelhança ou de dissemelhança

Para realizar a análise de *clusters* é necessário uma medida que avalie a distância ou similaridade entre os sujeitos ou variáveis de modo a poder agrupá-los.

A análise teórica das relações de semelhança tem sido dominada por modelos geométricos, nos quais os objetos são representados como pontos no espaço de forma que as dissemelhanças observadas entre os objetos correspondam a distâncias métricas entre os respetivos pontos (Tversky, 1977 *apud* Reis, 2001:299).

Normalmente, os indivíduos são agrupados segundo algum tipo de distância métrica e as variáveis agrupadas com base em medidas de correlação ou associação apropriadas (Jonhson e Wichern, 2002 *apud* Marôco, 2011:531).

Existem várias medidas de dissimilaridade (quanto menor o valor mais similares são os elementos) e cada uma produz um tipo de agrupamento. A sua utilização implica conhecimento da escala de medida das variáveis (nominal, ordinal, intervalar ou razão).

O SPSS apresenta diversas medidas de semelhanças e de dissemelhanças para construção dos *clusters*, sendo as mais utilizadas:

1. Distância Euclidiana – mede o comprimento do segmento de reta que une duas observações, dada por:

$$D_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}. \quad (2.37)$$

Onde  $x_{ik}$  é o valor da variável  $k$  no individuo  $i$  e  $x_{jk}$  é o valor da variável  $k$  no individuo  $j$  e  $p$  o número de variáveis.

2. Distância de Minkowski – medida de dissemelhança métrica, considerada uma generalização da distância euclidiana, dada por:

$$D_{ij} = \sqrt[m]{\sum_{k=1}^p |x_{ik} - x_{jk}|^m} . \quad (2.38)$$

Para  $m=1$ ,  $D_{ij}$  é o módulo da distância absoluta entre os indivíduos  $i$  e  $j$  relativamente às  $p$  variáveis medidas; para  $m=2$  é a distância euclidiana.

3. Distância de Mahalanobis – medida de dissemelhança métrica que mede a distância estatística entre dois indivíduos  $i$  e  $j$ , dada por:

$$D_{ij} = \sqrt{(x_i - x_j)' S^{-1} (x_i - x_j)} . \quad (2.39)$$

Onde  $x_i$  e  $x_j$  representam os vetores  $p$  dimensionais das observações correspondentes aos indivíduos  $i$  e  $j$ ;  $S$  é uma estimativa amostral da matriz de variâncias-covariâncias  $\Sigma$  dentro dos *clusters*.

Esta medida não é aplicável na análise de *clusters*, no SPSS, por não se conhecer à partida os grupos sobre os quais se vai calcular  $S$ .

4. Medida de semelhança do coseno – medida de semelhança métrica que mede a proximidade entre dois sujeitos para  $p$  variáveis pelo menos intervalares, dada por:

$$CoSIN(i, j) = \frac{\sum_{k=1}^p x_{ik}x_{jk}}{\sqrt{\sum_{k=1}^p x_{ik}^2 \sum_{k=1}^p x_{jk}^2}} . \quad (2.40)$$

5. Coeficiente de Jaccard, de Russel & Rão e medidas de associação binárias – medidas de associação não métricas, utilizadas para variáveis nominais.

- Coeficientes de Jaccard

$$s_{ij} = \frac{a}{a+b+c} \text{ e } d_{ij} = \frac{b+c}{a+b+c} . \quad (2.41)$$

- Coeficiente de Russel & Rão

$$s_{ij} = \frac{a}{a+b+c+d} . \quad (2.42)$$

- Coeficiente de Johnson e Wichern (usado para variáveis binárias)

$$s_{ij} = \frac{a+d}{a+b+c+d} \text{ e } d_{ij} = \frac{b+c}{a+b+c+d} . \quad (2.43)$$

Onde  $a$  representa o número de atributos das  $p$  variáveis presentes em ambos os indivíduos;  $b$  representa o número de atributos presentes no indivíduo  $i$  mas ausentes no indivíduo  $j$ ;  $c$  representa o número de atributos ausentes no sujeito  $i$  mas presentes no

sujeito  $j$ ;  $d$  representa o número de atributos ausentes em ambos os indivíduos;  $s_{ij}$  medida de semelhança;  $d_{ij}$  medida de dissemelhança.

6. Medidas de semelhança para variáveis – para o agrupamento de variáveis as medidas de semelhança adequadas são os coeficientes de correlação amostrais. Para variáveis contínuas utiliza-se o coeficiente de correlação de Pearson, para variáveis ordinais o coeficiente de Spearman e para variáveis nominais o coeficiente Phi.

- Coeficiente de Pearson ( $r$ ) é dado por:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (2.44)$$

Onde:  $x_i$ ,  $y_i$  são valores amostrais de  $X$  e  $Y$ , respetivamente;  $\bar{x}$ ,  $\bar{y}$  são as médias amostrais de  $X$  e  $Y$ , respetivamente

- Coeficiente de Spearman ( $\hat{\rho}$ ) dado por:

$$\hat{\rho} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n}. \quad (2.45)$$

Onde:  $n$  é o número de pares  $(x_i, y_i)$ ;  $d_i$  é a diferença entre ordenações.

- Coeficiente Phi ( $\hat{\phi}$ ) é dado por:

$$\hat{\phi} = \frac{(ad - bc)}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}. \quad (2.46)$$

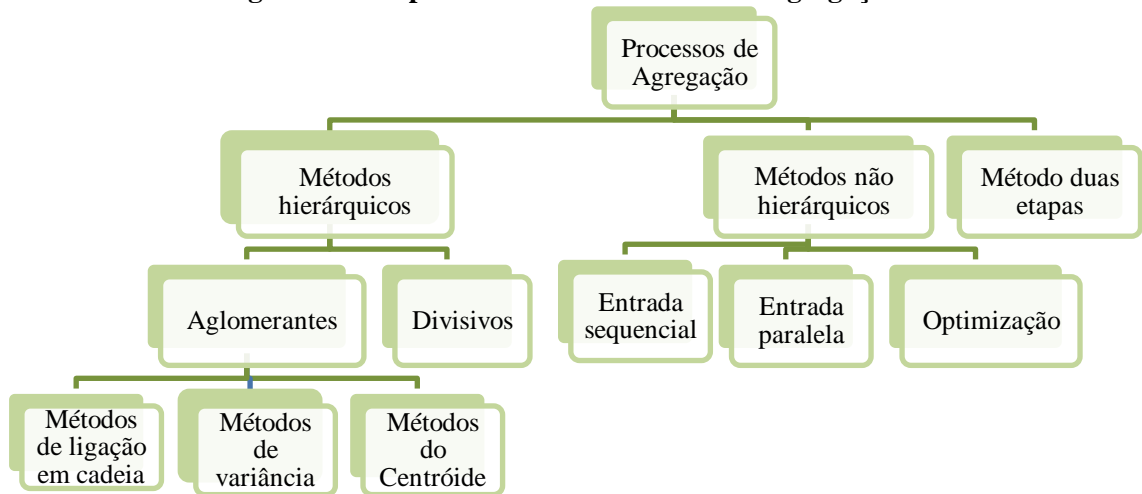
Onde:  $a, b, c, d$  são as frequências da tabela  $2 \times 2$  e  $n = (a + b + c + d)$  é o número de observações da amostra.

No entanto, o investigador deve ter presente que se as variáveis estiverem em escalas diferentes, os resultados podem ser influenciados pelas unidades medida. Pelo que será recomendável a normalização das variáveis, de modo a ficarem com média igual a zero e desvio padrão igual a um. No caso, de existirem *outliers* estes devem ser removidos.

#### 2.2.4. Critérios de agregação ou desagregação

O processo de agregação pode ser hierárquico ou não hierárquico (Figura 2.2). O agrupamento hierárquico de *clusters* permite agrupar quer indivíduos quer variáveis ao passo que o agrupamento não hierárquico apenas permite agrupar indivíduos (Marôco, 2011).

**Figura 2.2 – Tipos de métodos de análise de agregação**



Fonte: Elaborado pelo autor, segundo dados da pesquisa

#### **2.2.4.1. Agrupamento hierárquico**

Os métodos hierárquicos são caracterizados por ter um desenvolvimento com uma estrutura em árvore e podem ser de dois tipos: métodos aglomerantes ou ascendentes e métodos divisivos ou descendentes. Ambos os tipos procuram o conjunto óptimo de *clusters*, operando em cada etapa na subdivisão ou fusão progressiva dos dados com base na matriz de proximidades (semelhanças ou dissemelhanças entre casos).

Nos métodos aglomerantes parte-se de grupos com um objeto cada, que vão sendo agrupados sucessivamente de acordo com as suas proximidades até se encontrar um grupo que inclua a totalidade dos objetos. Os métodos aglomerantes compreendem métodos de ligação, de variância e de centróides.

Nos métodos divisivos parte-se de um grupo que inclui todos os objetos em estudo sendo depois divididos sucessivamente, em subgrupos de acordo com as suas medidas de distância, até se obter grupos de um elemento cada.

Os métodos de análise de *clusters* mais utilizados são os hierárquicos aglomerantes, que diferem uns dos outros essencialmente no modo como as distâncias são calculadas (após o passo inicial do algoritmo). Os métodos de ligação incluem: menor distância, maior distância, distância média entre *clusters*, distância média dentro dos clusters e distância mediana. Os métodos de variância tentam gerar *clusters* que minimizem a variância interna do *cluster* e um método comum é o de *Ward*.

No método menor distância após a formação do primeiro *cluster*, a distância deste aos restantes objetos é a menor das distâncias de cada um dos elementos constituintes desse *cluster* a cada um dos restantes objetos.

No método da maior distância após a formação do primeiro *cluster*, a distância deste aos restantes objetos é a maior das distâncias de cada um dos elementos constituintes desse *cluster* a cada um dos restantes objetos.

No método distância média entre *clusters* após a formação do primeiro *cluster*, a distância deste aos restantes objetos é a média das distâncias de cada um dos elementos constituintes desse *cluster* a cada um dos restantes objetos.

O método distância média dentro dos *clusters* combina os *clusters* de modo que a distância média entre todos os pares possíveis de objetos dentro do grupo daí resultante seja mínima.

No método distância mediana após a formação do primeiro *cluster*, a distância deste aos restantes objetos é a mediana das distâncias de cada um dos elementos constituintes desse *cluster* a cada um dos restantes objetos.

No método do centróide a distância entre dois *clusters* é definida como a distância entre os seus centróides (pontos definidos pelas médias das variáveis que caracterizam os indivíduos de cada grupo). De cada vez que se agrupam objetos, calcula-se um novo centróide.

No método de Ward para cada *cluster* calcula-se a média de todas as variáveis. Depois calcula-se a distância euclidiana de cada objecto do *cluster* à sua média. Somam-se essas médias e obtém-se a medida de comparação. Em cada estágio, agrupam-se os dois *clusters* que gerarem as somas totais mais baixas.

#### **2.2.4.2. Agrupamento não hierárquico**

Os métodos não hierárquicos de agrupamento de *clusters*, também referidos como métodos k-médias, destinam-se a agrupar indivíduos e não variáveis num conjunto de *clusters* cujo número é definido à partida pelo investigador. Incluem os métodos de entrada sequencial, de entrada paralelo e de optimização.

Os vários métodos não hierárquicos diferem essencialmente na forma como se processa a primeira agregação dos sujeitos em *clusters*, e no modo como as novas distâncias entre os centróides dos *clusters* e os indivíduos são medidas.

No método de entrada sequencial seleciona-se o centro de um *cluster* e agrupam-se todos os objectos cuja distância ao centro não ultrapasse um determinado patamar, de seguida seleciona-se outro centro e procede-se do mesmo modo.

O método de entrada paralela é semelhante método de entrada sequencial, mas selecionam-se vários centros de *clusters* de uma só vez.

O método de optimização difere dos dois anteriores no facto dos objetos poderem ser reatribuídos a *clusters* para optimizar um critério global, como a distância média dentro do *cluster* para um dado número de *clusters*.

Os métodos não hierárquicos apresentam algumas vantagens relativamente aos métodos hierárquicos, nomeadamente:

- Podem ser aplicados a matrizes de dados muito grandes, uma vez que não é preciso calcular e armazenar uma nova matriz de dissemelhança em cada passo do algoritmo.
- Conseguem reagrupar os objectos em *clusters* diferentes daqueles em que foram colocados inicialmente;
- Menor probabilidade de classificação errada de um determinado indivíduo num determinado *cluster*.

Mas apresentam também algumas desvantagens, nomeadamente:

- O número de *clusters* tem de ser estabelecido à partida;
- A seleção de centros dos *cluster* é arbitária;
- Os resultados podem depender da seleção dos centros.

Definir à partida o número de *clusters*, por vezes torna-se difícil por não ser conhecida a estrutura dos dados. No entanto, este obstáculo pode ser ultrapassado aplicando um método hierárquico aos dados para determinar o número de *clusters* (correspondente ao número de partições), ou efetuando uma escolha aleatória ou baseando-se no conhecimento do investigador.

O método K-médias genericamente consiste na transferência de um indivíduo para o *cluster* cujo centróide se encontra a menor distância. Este método é bastante utilizado em análise de *clusters* nomeadamente por se encontrar fortemente difundido na maioria dos softwares estatísticos e pela sua facilidade de aplicação quando se está perante um elevado número de observações. Este método exige que as variáveis sejam numéricas ou binárias.

O algoritmo K-médias, segundo Johson e Wichern (2007), é o seguinte:

- 1. Partição inicial dos sujeitos em  $k$  grupos definidos à partida pelo investigador;
- 2. Cálculo dos centróides para cada um dos  $k$  grupos (no SPSS, por defeito as primeiras  $k$  observações são usadas como centróides dos  $k$  grupos no primeiro passo da rotina, podendo porém o pesquisador definir qual o valor dos centróides a usar) e cálculo da distância euclidiana dos centróides a cada indivíduo na base de dados;
- 3. Agrupar os indivíduos aos grupos de cujos centróides se encontram mais próximos e voltar ao passo 2 até que não ocorra uma variação significativa na distância mínima de cada indivíduo da base de dados a cada um dos centróides dos  $k$  grupos.

O método k-médias apresenta alguns pontos fracos, nomeadamente:

- Os dados devem ser numéricos e comparáveis através de alguma forma de distância. Existe no entanto uma variante do método, chamada de K-medianas, que cuida dessas falhas;
- O algoritmo trabalha melhor com dados contidos em *clusters* esféricos, ao usarem-se outras geométricas podem não ser localizados;
- É sensível a *outliers*, ou seja, a pontos que não pertencem a nenhum *cluster*. Esses pontos podem distorcer a posição do centróide e estragar o agrupamento.

### 2.2.4.3. Agrupamento duas etapas

Os métodos clássicos (hierárquicos e não hierárquicos) nem sempre resolvem todos os problemas pertinentes à mineração de dados. Os métodos hierárquicos geralmente não funcionam bem para grandes bases de dados, e os métodos não hierárquico exigem um conhecimento prévio do número de agrupamentos, o que nem sempre é possível. Punj e Stewart (1983) sugerem que os dois métodos sejam usados em cadeia, primeiro um método hierárquico (*Ward*, distância média) para obter o número de *clusters* e centróides e depois, um método não hierárquico (otimização) para encontrar os *clusters* ótimos.

Chiu et al.(2001) propuseram como método alternativo, um algoritmo realizado em duas etapas que deteta agrupamentos em grandes bases de dados, permite a utilização de variáveis contínuas e categóricas e determina automaticamente o número ideal de clusters.

O método pode gerar os *clusters* segundo um número especificado ou caso o número de *clusters* seja desconhecido, o método vai encontrar o número ideal de grupos automaticamente.

Na primeira etapa do procedimento, formam-se os pré *clusters*, que são *clusters* dos casos originais, substituindo-os com o objetivo de reduzir o número de casos para a próxima etapa, reduzindo o tamanho da matriz que contém distâncias entre todos os casos pareados possíveis. O algoritmo pesquisa os registros um a um e decide se o registro atual deve fundir-se com os *clusters* formados anteriormente ou criar um novo *cluster* com base no critério da distância, realizando assim um agrupamento sequencial. Numa segunda etapa, por um processo hierárquico, são construídos os agrupamentos através de agregação de pré *clusters* relativamente similares. O tamanho da matriz de distâncias não depende do número de casos, mas sim do número de pré *clusters*.

O algoritmo utilizado por este procedimento tem diversas características desejáveis e que o diferencia das técnicas clássicas de aglomeração:

- manipulação de variáveis categóricas e contínuas: supõe que as variáveis são independentes, no caso de serem categóricas com distribuição multinomial e no caso de serem contínuas com distribuição normal;
- seleção automática de número de clusters: comparando os valores de um critério de escolha, através de diferentes soluções de *clusters*, o procedimento pode determinar automaticamente o número ideal de clusters.
- escalabilidade: através da construção uma árvore de recursos do cluster, que resume os registros, o algoritmo permite analisar grandes bases de dados.

Neste método existem duas medidas de distância que determinam o modo como a similaridade entre dois *clusters* é calculada:

- probabilidade de *log*: a medida de probabilidade ajusta distribuições de probabilidades às variáveis. As variáveis contínuas são ajustadas à distribuição normal e as variáveis categóricas à distribuição multinomial. Todas as variáveis são consideradas independentes;
- euclidiana: que é a distância da “linha reta” entre dois *clusters*. Só pode ser usada quando todas as variáveis são contínuas.

Esta seleção permite que se especifique o modo como é determinado o número de *clusters*:

- Determinação automática: o procedimento pode determinar automaticamente o número ótimo de *clusters*, usando o critério especificado no grupo de critério de agrupamento. Comparando os valores de um critério de escolha, através das soluções de aglomeração diferentes, o procedimento pode automaticamente determinar o número ótimo de *clusters*. O método Duas Etapas disponibiliza dois critérios de agrupamentos para a determinação automática do número de clusters para segmentar grandes bases de dados: o “Critério de Informação Bayesiano” (BIC) e o “Critério de Informação Akaike” (AIC).
- Especificação de um número fixo: permite fixar o número de *clusters* na solução. Deve-se digitar um número inteiro positivo.

Ao escolher o método Duas Etapas o SPSS fornecerá como output uma tabela de distribuição de frequência de cada *cluster*, uma tabela com as médias assumidas pelas variáveis quantitativas do modelo (quando existirem) em cada *clusters*, distribuição de frequência das categorias de cada variável categórica do estudo em cada cluster (quando existirem variáveis qualitativas), gráfico circular das frequências dos *clusters*, gráfico de barras para percentagens dentro dos *clusters* para as variáveis qualitativas, gráficos dos intervalos de confiança para as variáveis quantitativas, gráfico de significância do qui-quadrado para as variáveis qualitativas do modelo e finalmente gráfico de significância t-Student para as variáveis quantitativas

### 2.2.5. Validação dos resultados encontrados

Dada a natureza subjetiva da análise de *clusters*, o pesquisador deve tomar cuidado na validação e assegurar a significância prática da solução final do cluster. Apesar de não existir um método totalmente eficaz, existem várias metodologias que podem ser utilizadas.

A validação da solução do *cluster* inclui as tentativas feitas pelo pesquisador para assegurar que a solução do *cluster* é representativa da população, generalizável para outros objetos e estável ao longo do tempo.

Jain e Dubes (1988), apresentaram quatro tipos de critérios para validação dos resultados da Análise de *Clusters*, a saber:

- Internos: avaliam o grau de compatibilidade entre a estrutura de grupos obtida e a matriz de proximidades para os dados iniciais.
- Relativos: avaliam qual de entre as diferentes estruturas construídas utilizando métodos diferentes e a partir dos mesmos objetos é a melhor.
- Externos: avaliam o desempenho dos resultados obtidos comparando a estrutura dos dados com informação exterior não utilizada na análise.
- Replicação: avaliam os resultados obtidos com uma metade da amostra comparando-os com os resultados obtidos com a outra metade.

Hair et al. (2010) aponta duas abordagens de validação: analisar amostras separadamente e validação cruzada. Analisar amostras separadamente, comparar as soluções do *cluster* e verificar a correspondência dos resultados. No entanto, por vezes é impraticável por causa do tempo despendido, dos custos envolvidos ou por os objetos não estarem disponíveis. Nestes casos, pode dividir-se a amostra em dois grupos, analisar os clusters separadamente e depois comparar os resultados.

A validação cruzada também poderá ser usada, uma vez que os membros de qualquer *cluster* de uma solução deverão permanecer juntos noutra solução, a validação cruzada mostrará padrões de pertença ao *cluster*. Outras abordagens incluem uma forma modificada de separar as amostras, pelo que os centros de *cluster* obtidos para uma solução são utilizados para definir *clusters* a partir de outras observações e os resultados comparados.

Para qualquer destes métodos a estabilidade dos resultados do *cluster* pode ser avaliada pelo número de casos atribuídos ao mesmo *cluster* em soluções de *cluster*. Geralmente, uma solução muito estável seria produzida com menos de 10% das observações atribuídas a outro *cluster*. Uma solução estável resultaria com 10% a 20% atribuída a outro grupo e uma solução pouco estável com 20% a 25% das observações atribuídas a um *cluster* diferente do inicial.

**CAPÍTULO III.**  
**OPÇÕES METODOLÓGICAS**

### 3. OPÇÕES METODOLÓGICAS

Após a revisão bibliográfica sobre o insucesso escolar e a abordagem de técnicas de análise multivariada, apresentadas nos dois capítulos anteriores, interessa agora centrar a nossa atenção no estudo prático sobre o sucesso/insucesso escolar na Matemática no 3.º ciclo na Escola Básica e Secundária Joaquim Inácio da Cruz Sobral (EBSJICS).

A escolha do domínio “Análise multivariada aplicada ao (in)sucesso escolar a matemática no 3.º ciclo - estudo de caso” justifica-se dentro do contexto do Mestrado em Estatística, Matemática e Computação área de especialização Estatística Computacional, bem como se insere dentro da área de atividade profissional do investigador, Professor de Matemática.

Após definição da problemática há necessidade de recolher e analisar os dados obtidos. Fortin (2003) refere que o investigador deve implementar um desenho de investigação, isto é, o plano lógico criado para obter respostas válidas às questões de investigação colocadas ou às hipóteses formuladas.

O desenho de investigação determina o plano geral do investigador para a obtenção de respostas e indagações de pesquisa e para testar as hipóteses, expressa ainda as estratégias adoptadas pelo investigador para desenvolver informações precisas, objetivas e passíveis de interpretação (Polite e Hungler, 1995).

De acordo com Fortin (2003), a escolha do desenho de investigação depende do problema em causa e do estado dos conhecimentos em torno deste. Assim, o desenho de investigação inclui o meio onde o estudo será realizado, a população alvo e amostra, ou seja, a seleção dos sujeitos, o tipo de estudo, as estratégias utilizadas, os instrumentos de recolha dos dados e a sua análise. O meio é um elemento escolhido pelo investigador devendo essa escolha ser, devidamente justificada de forma a garantir a credibilidade dos resultados. A escolha da amostra tem em conta características da população que definem o grupo de sujeitos a incluir no estudo e os critérios de seleção. O controle das variáveis estranhas, isto é, das variáveis que não estão incluídas nos estudo mas que podem exercer influência sobre a variação da medida das variáveis estudadas é realizado através de diversas estratégias. Os instrumentos de colheita de dados fornecem respostas às questões de investigação ou às hipóteses. No tratamento dos dados, o método de análise a utilizar

deve ser congruente com os objetivos e com o desenho do estudo consoante este vise descrever relações, verificar relações entre variáveis ou comparar grupos.

O presente capítulo inclui oito seções. Na primeira os objetivos e as questões de investigação. Na segunda secção encontra-se a apresentação e justificação da abordagem da investigação utilizada neste estudo. A terceira secção tem como propósito caracterizar o meio/população do estudo. Na quarta secção será referenciada a amostra. Na quinta serão dados a conhecer as variáveis estudadas investigação, na sexta a referência aos instrumentos de recolha e na sétima será mencionada forma através da qual foi realizada a recolha de dados. Por fim, a última secção está destinada à referenciação de como os dados foram tratados.

### **3.1. Objetivos e questões de investigação**

Numa investigação é necessário delinear o percurso da investigação a adoptar, ou seja, “expressar o mais exatamente possível, o que procura saber, elucidar, compreender melhor” (Quivy e Campenhoudt, 1998:32).

As questões de investigação são o elemento fundamental do início de uma investigação, são elas que explicitam a área de investigação Lewis e Pamela (1987), ou seja, “as premissas sobre as quais se apoiam os resultados da investigação” (Fortin, 2003:101).

Ao estabelecer a questão de investigação, o investigador deve ter em mente:

“os três níveis de exigência que uma boa pergunta deve respeitar: primeiro, exigências de clareza; segundo exigências de exequibilidade; terceiro, exigências de pertinência, de modo a servir de primeiro fio condutor a um trabalho do domínio da investigação em ciências sociais.”  
(Quivy e Campenhoudt, 1998:44)

Moltó, 2002 *apud* Coutinho, 2014:46 refere que é conveniente que a definição do problema seja o mais específica possível contendo os aspetos essenciais do estudo, isto é, mencionar o que se estuda (objeto de investigação), quem é que vai fazer parte da investigação (sujeitos) e como se vai estudar o problema (definição das variáveis).

O propósito deste estudo é investigar se o insucesso escolar em Matemática, no 3.º ciclo, na EBSJCS, depende de determinados fatores e que medidas poderão ser tomadas para minorá-lo.

Foram formuladas à partida um conjunto de questões orientadoras do processo de recolha de dados, a saber:

- Que fatores estarão na origem do insucesso escolar a matemática?
- De quem é a responsabilidade pelo insucesso escolar em matemática?
- Quais as razões do insucesso escolar, na perspectiva dos alunos, a matemática?
- Que ações poderão ser desenvolvidas para minorar o insucesso escolar em matemática?

### **3.2. Natureza do estudo**

No presente estudo procede-se a uma investigação empírica, quantitativa e correlacional.

A investigação empírica tem por objetivo contribuir para o enriquecimento do conhecimento na área que está a ser investigada, implicando que sejam feitas escolhas, quer em termos do tema, quer das hipóteses a testar, implicando também que se realize o planeamento dos métodos de recolha de dados e que seja planeada as análises de dados antes de ser iniciada a parte empírica da investigação.

Para Hill e Hill (2012) as investigações empíricas mais vulgarmente utilizadas são de três tipos:

- Investigação pura – quando o objetivo é descobrir factos novos para testar deduções feitas a partir de uma teoria que à partida só tem interesse intelectual e que contribui essencialmente para o enriquecimento da literatura, no sentido de aprofundar os conhecimentos sobre uma determinada área do saber;
- Investigação aplicada – quando o objetivo é descobrir factos novos para testar deduções feitas a partir de uma teoria que pode ter aplicações práticas a médio prazo;
- Investigação aplicável – quando o objetivo é descobrir factos novos que possibilitem a resolução de problemas a curto prazo.

Trata-se, também, de uma investigação quantitativa a qual “tem por finalidade contribuir para o desenvolvimento e validação dos conhecimentos; oferece também a possibilidade de generalizar os resultados, de predizer e de controlar os acontecimentos.”(Freixo, 2009:145). Este tipo de investigação consiste num:

“um processo sistemático de colheita de dados observáveis e quantificáveis, baseada na observação de factos objetivos, de acontecimentos e de fenómenos que existem independentemente do investigador.”

(Fortin, 2003:22)

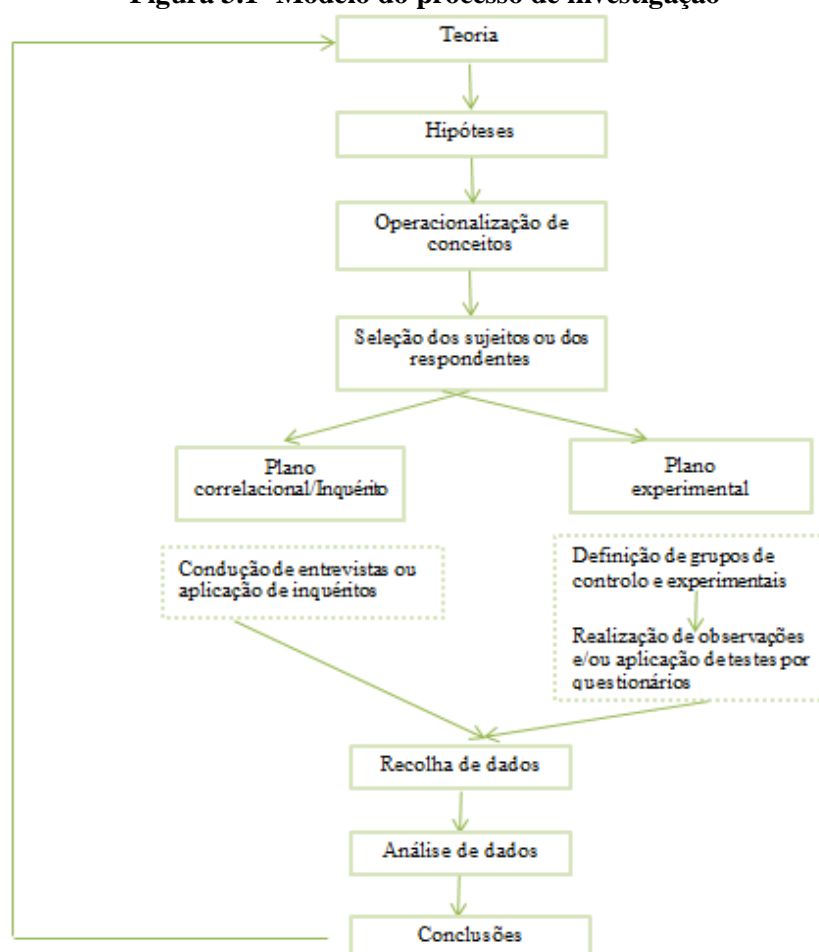
Vilelas (2009) refere que os estudos quantitativos admitem que tudo pode ser quantificável, isto é:

“possível traduzir em números as opiniões e as informações para, em seguida, poderem ser classificadas e analisadas. [...] visam a apresentação e a manipulação numérica de observações com vista à descrição e à explicação do fenómeno sobre o qual recaem as observações.”

(Vilelas, 2009:103)

Brymon e Cramer (2003) conscientes de que a investigação quantitativa nem sempre possa ser uma sequência linear bem ordenada e de que essa sequência nem sempre possa ser reproduzida numa situação real, estabeleceram um modelo do processo de investigação, procurando ilustrar os seus passos principais (Figura 3.1).

**Figura 3.1- Modelo do processo de investigação**



Fonte: Brymon e Cramer (2003:3)

Numa primeira etapa, a teoria, é ponto de partida do processo fornecendo o enquadramento teórico.

A segunda etapa, as hipóteses, após a formulação da teoria os investigadores procuram testá-la no seu todo ou em parte (deduzir a partir da teoria uma hipótese que relacione com uma parte da teoria) e proceder ao teste de investigação.

Frequentemente, as hipóteses apresentam-se como uma relação entre duas ou mais entidades denominadas de conceitos, ou seja, categorias onde se arrumam ideias e observações sobre os elementos comuns do mundo. No entanto, embora as hipóteses tenham a vantagem de forçar os investigadores a pensar de forma sistemática sobre o que pretendem estudar e a estruturar os seus planos de investigação de acordo com isso, têm uma possível desvantagem que é a de poder afastar a atenção do investigador de outras facetas interessantes dos dados recolhidos.

A terceira etapa, operacionalização de conceitos, consiste no desenvolvimento de formas para medir os conceitos de modo a avaliar a validade de uma hipótese. Trata-se da tradução dos conceitos em variáveis, ou seja, em atributos que estabeleçam diferenças entre objetos relevantes (indivíduos, empresas, países, etc.), que poderá ser conseguida através de questionários, de observação dos sujeitos ou de análise de estatísticas existentes.

A quarta etapa, seleção de inquiridos ou de sujeitos, o investigador procura encontrar os sujeitos aos quais vai aplicar o instrumento de medida que construiu. Em muitas situações não é possível contactar todos os elementos de uma população e, por isso, é necessário seleccionar uma amostra. Para que os resultados se possam generalizar a uma população mais ampla, essa amostra tem que ser representativa.

A quinta etapa, preparação de um plano de investigação, tendo presente que os dois tipos de planos de investigação mais utilizados são os planos experimentais e os correlacionais.

Nos planos de investigação experimentais, o investigador manipula alguns aspetos, em laboratório ou no campo, observando os efeitos dessa manipulação nos sujeitos experimentais. Este tipo de estudo exige, que exista um “grupo de controlo”, o qual serve de base de comparação com o grupo sujeito a manipulação experimental.

Nos planos de investigação correlacional ou inquérito, o investigador não manipula qualquer variável e todos os dados relacionados com todas as variáveis são recolhidos simultaneamente.

Na maioria das investigações, a natureza do plano de investigação - experimental ou correlacional - é conhecida no início da investigação, exercendo influência nas várias etapas do processo de investigação e na manipulação estatística que pode ser realizada com os dados obtidos.

A sexta etapa, a recolha de dados, o investigador recolhe dados a partir da realização de entrevistas, da aplicação de questionários, da observação ou de qualquer outro método.

Na sétima etapa, análise de dados, o investigador descreve os sujeitos em termos das variáveis escolhidas, ou estuda a ligação entre uma variável e cada uma das outras (análise bivariada) ou as relações entre as várias variáveis (análise multivariada).

Na oitava etapa, resultados, aqui podem ocorrer três situações:

- os resultados da análise de dados sugerem que uma hipótese é confirmada, podendo esses resultados servir de apoio à teoria que a suscitou, podendo outros investigadores vir a interessar-se por reproduzir essa mesma conclusão ou por estudar outras implicações da teoria;
- os resultados da análise de dados sugerem que uma hipótese é refutada, o que sugere que ou a teoria não é adequada ou então precisa ser revista;
- apenas alguns pontos da hipótese são confirmados, um resultado deste tipo poderá exigir uma reformulação da teoria.

O tema desta investigação e as questões de investigação que se levantaram levaram à necessidade de testar e medir as relações entre as várias variáveis em estudo. Assim, optou-se por uma investigação em que o objetivo final seja verificar a existência de algum tipo de relação entre variáveis.

Esta investigação foi desenhada segundo a lógica do estudo correlacional. Segundo Coutinho (2008), o estudo correlacional utiliza-se quando o objetivo da pesquisa é encontrar e avaliar a intensidade de relações entre variáveis, sem manipulação e sem pretensões de estabelecer relações de causa efeito. A procura de relações ou associações entre as variáveis é realizada através de procedimentos estatísticos que fornecem a quantificação da relação entre variáveis.

A investigação correlacional caracteriza-se por colocar hipóteses e objetivos de investigação sobre factos e fenómenos que não são intrinsecamente manipuláveis, com a finalidade de encontrar e analisar relações entre eles de modo a tentar explorar possíveis relações entre as hipóteses.

Pode-se apontar como características individualizadoras do modelo correlacional os seguintes aspetos: o investigador trabalha num contexto natural; não é exercida qualquer influência sobre as variáveis; e destinam-se a testar hipóteses que permitam corroborar teorias ou confrontá-las.

Para realizar uma investigação correlacional é necessário:

- Ter medições/pontuações de pelo menos duas variáveis distintas, uma variável antecedente ou preditora ( $X$ ) e uma variável critério ( $Y$ );
- Obter pares de pontuações  $X$  e  $Y$  para cada sujeito da amostra.

Ao estabelecer-se uma relação entre duas variáveis, isso significa que as pontuações das duas variáveis covariam, ou seja, mudam simultaneamente. A existência de uma relação (correlação) entre as variáveis não significa que exista uma relação causal entre elas, ou seja, não significa que alterações numa variável causem alterações na outra.

Na perspectiva de Heiman (1996) para se poder inferir de relações causais entre variáveis será necessário verificarem-se dois requisitos: primeiro, a variável antecedente terá de ocorrer temporalmente sempre antes da variável critério; segundo a variável antecedente terá de ser a única a causar a variável critério. Ora em investigação correlacional pode não se conseguir saber de forma inequívoca se a variável  $X$  ocorreu antes da  $Y$  e não é possível controlar ou eliminar outras variáveis que possam causar as mudanças ocorridas.

Na investigação correlacional é utilizado apenas um grupo em que o valor da relação entre as variáveis é um valor único calculado a partir das pontuações, dos pares  $X$ - $Y$ , de todos os sujeitos que compõem a amostra. Não sendo possível identificar qual a variável independente e qual a variável dependente, uma vez que não é possível identificar qual a variável anterior e qual a posterior. A representação gráfica das pontuações dos sujeitos em cada uma das variáveis faz-se com recurso a gráficos de pontos.

As etapas de uma investigação correlacional são:

1. Formulação do problema;
2. Formulação de hipóteses sobre as possíveis relações entre as variáveis;
3. Procedimentos metodológicos (definição de variáveis, seleção e elaboração dos instrumentos de recolha de dados, seleção da amostra);
4. Recolha de dados;
5. Análise estatística dos dados e discussão dos resultados;

## 6. Redação do relatório de investigação.

Coutinho (2008) identifica vantagens e desvantagens na aplicação de estudos correlacionais numa investigação em ciências sociais. Aponta como vantagens: permitir explorar áreas em que pouco se sabe sobre o assunto; não exigir amostras grandes; permitir trabalhar com um grande número de variáveis; permitir o estudo de variáveis em situações próximas da realidade; fornecer indicadores quantificáveis (coeficiente de correlação). E como desvantagens: valor precário dos resultados, uma vez que não é estabelecida uma relação de causalidade; não permitir um controlo rigoroso das variáveis; a precisão da informação fornecida pelos coeficientes de correlação estar muito dependente da fiabilidade dos processos de recolha de dados; possibilidade de enviesamento das conclusões por falhas na definição das variáveis, na escolha dos testes estatísticos, no tipo de análise realizada (bivariada ou multivariada) e na interpretação da significância dos coeficientes de correlação; os coeficientes de correlação não podem ser interpretados como valores absolutos, pelo que as conclusões obtidas são relativas a uma determinada população e provavelmente não se repetem noutra.

A investigação correlacional pode ser de dois tipos: estudos de relação e estudos de previsão.

Os estudos do tipo relacional têm por objetivo tentar compreender a complexidade do fenómeno, estudando as relações entre as variáveis, das quais se suspeita estarem associadas. Este tipo de análise pode envolver a relação entre duas variáveis (análise bivariada) ou envolver mais do que duas variáveis (análise multivariada). A regressão múltipla, a análise discriminante, a análise fatorial, a análise de *clusters* e a análise de sendas integram a análise correlacional multivariada.

Os estudos de previsão são desenvolvidos em áreas em que já existe algum conhecimento sobre os fenómenos, proveniente de investigações anteriores. Estes estudos procuram prever o comportamento da variável critério a partir do conhecimento do valor das variáveis predictoras. Baseiam-se na correlação e na regressão linear, múltipla, polinómica e não linear.

Nesta investigação optou-se por um estudo correlacional exploratório de relação, tendo em conta o tipo de investigação que se pretendia e as condicionantes que a rodeavam, conscientes de todas as vantagens e desvantagens ou dificuldades

que um estudo deste tipo acarretava, mas com a convicção de que era o que mais se adaptava à investigação pretendida e aos objetivos propostos.

A investigação correlacional foi realizada por amostragem numa perspetiva retrospectiva, a informação sobre a população do estudo inferida a partir de uma amostra representativa da população, em que o investigador observa essa amostra no presente e regista os eventos de interesse que ocorreram no passado, nomeadamente o sucesso ou o insucesso dos alunos.

### **3.3. Caracterização do meio/população**

O estudo realiza-se na EBSJICS integrada no Agrupamento de Escolas Joaquim Inácio da Cruz Sobral, em Sobral de Monte Agraço.

Sobral de Monte Agraço é um concelho pertencente ao distrito de Lisboa e integra a região do Oeste, distando cerca de 40 km da capital. Concelho, com 10.156 habitantes (Censos 2011), ocupa uma área de 51,95 Km<sup>2</sup> distribuída por três freguesias.

A escola dispõe de trinta e quatro salas de aula e é frequentada por alunos do 5.º ano ao 12.º ano, num total de 798 alunos.

A comunidade escolar caracteriza-se, tal como a população em geral, por assimetrias socioeconómicas e socioculturais, quase metade da população escolar beneficia de auxílios económicos no âmbito de Ação Social Escolar.

O Projeto Educativo do Agrupamento (PEA) faz referência à principal missão do agrupamento que é a de desenvolver um serviço público educativo de qualidade, contribuindo para a melhoria do ensino, definindo como metas: “A- Combate ao Insucesso; B - Promoção de um Bom Clima de Escola / Combate à Indisciplina; C- Combate ao Abandono Escolar; D- Relação com o Meio e E - Formação Contínua. Sendo os objetivos a atingir melhorar as competências básicas dos alunos e assegurar a permanência no sistema de todos os jovens até aos 18 anos, garantindo o cumprimento da escolaridade obrigatória de 12 anos (PEA, 2013-2017).

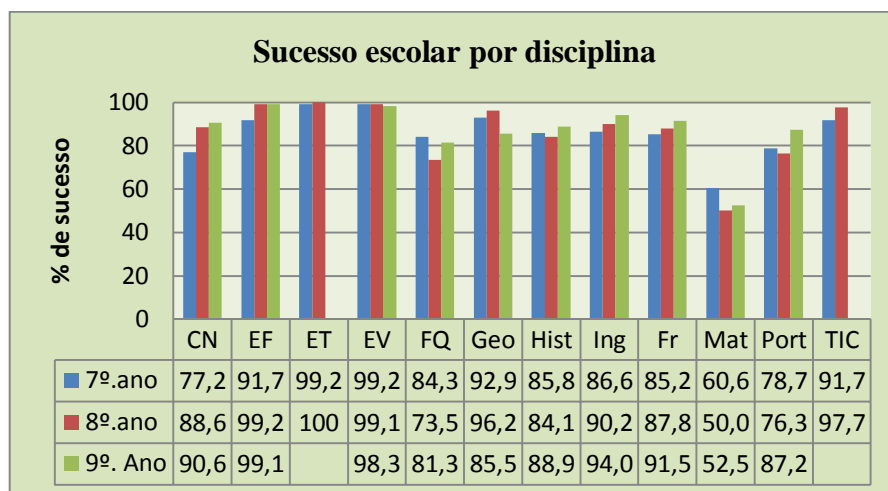
A opção do investigador de realizar o estudo neste estabelecimento de ensino prendeu-se com quatro razões: o conhecimento do funcionamento do mesmo; motivação profissional; proximidade do local que facilitou o acesso e a antevisão quer da autorização para realização estudo pelas estruturas dirigentes, quer da colaboração dos alunos.

A população consiste numa “coleção de elementos ou sujeitos que partilham características comuns, definida por um conjunto de critérios” (Fortin, 2003:102). Nesta investigação a população é todos os alunos que frequentam o 3.º ciclo do ensino básico regular na EBSJICS no ano letivo 2014/2015. A opção pelo 3.º ciclo do ensino básico prende-se com o facto de ser o ciclo em que o insucesso escolar a matemática é maior.

Optou-se por considerar como população do estudo apenas os alunos que frequentam o 3.º ciclo do ensino básico regular, excluindo-se assim do estudo os alunos do 3.º ciclo com Currículo Específico Individual e os alunos pertencentes às turmas de Percursos Curriculares Alternativos. A razão prende-se com a sua especificidade, os primeiros por terem programas curriculares diferentes e os segundos terem critérios de avaliação específicos diferentes dos restantes alunos. Assim, a dimensão da população é de 327 alunos.

Sendo um dos objetivos da EBSJICS que os alunos alcancem o sucesso escolar, pelo Gráfico 3.1 pode concluir-se que em 2014-15, no 3.º ciclo, é na disciplina de matemática que esse sucesso é menor, cerca de 60,6% no 7.º ano, 50% no 8.º ano e 52,5% no 9.º ano.

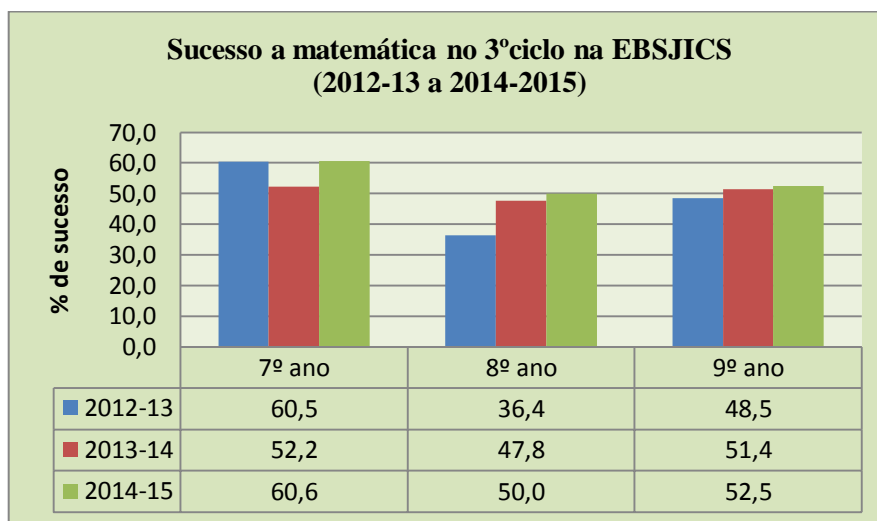
**Gráfico 3.1 – Sucesso por disciplina no 3.º ciclo na EBSJICS (2014-15)**



Fonte: AEJICS

No sentido de averiguar se a percentagem de sucesso a matemática tem sido estável analisaram-se os três últimos anos letivos (Gráfico 3.2) e verificou-se que a percentagem de sucesso no 7.º ano, embora oscilante, foi sempre superior a 50% e superior às dos restantes anos de escolaridade. No 8.ºano o sucesso aumentou consideravelmente mas ainda assim não ultrapassou os 50%, já no 9.º ano as melhorias foram mais ténues, atingindo 52,5% em 2014-15.

**Gráfico 3.2 - Sucesso a matemática no 3.º ciclo na EBSJICS (2012-13 a 2014-2015)**



Fonte: AEJICS

### 3.4. Amostra

O tipo de estudo e as questões de investigação a realizar, comandam a escolha da população a estudar, por sua vez, “a escolha da técnica de seleção da amostra, em particular, não poderá estar dissociada do questionamento que está na origem da investigação da população estudada” (Beaud,2003:201).

A maioria das investigações engloba um elevado número de sujeitos, pelo que se torna impossível estudá-los na sua totalidade, para viabilizar o seu estudo recorre-se a uma amostra. Esta é “um subconjunto de uma população ou de um grupo de sujeitos que fazem parte da mesma população. É de qualquer forma uma réplica em miniatura da população alvo” (Fortin, 2003:202).

Escolher a amostra corresponde à seleção ponderada de várias unidades de sondagem, entre os indivíduos que compõem a população. A precisão estatística, a fiabilidade e a validade da sondagem aumentam com o aumento da dimensão da amostra.

Sousa (2006) refere que as relações que se estabelecem entre a precisão estatística da sondagem e a dimensão da amostra são complexas. Para garantir a validade das estimativas, os procedimentos de seleção da amostra são tão ou mais importantes que a dimensão da amostra. A precisão estatística está ligada à dimensão absoluta da amostra e não à relação entre a dimensão da amostra e a população total. A precisão estatística de uma amostra não varia proporcionalmente em relação à dimensão da amostra, mas sim em função da raiz quadrada da dimensão dessa amostra, pelo que os ganhos de precisão

conseguidos com o aumento da dimensão de uma amostra nem sempre justificam o esforço suplementar

O objetivo principal da amostragem é o de obter uma amostra que seja uma representação honesta da população e que conduza à estimação das características da população com grande precisão.

O processo a adoptar na recolha de elementos a incluir na amostra, plano de amostragem, inclui as seguintes fases:

- definição dos objetivos do estudo;
- escolha dos dados úteis a recolher, o que significa definição da unidade de amostragem e da escala de valores para a característica em estudo;
- definição da população ou universo;
- escolha do método de amostragem;
- definição do nível de precisão ou erro de amostragem admitido.

Algumas das vantagens que se podem apontar do uso do processo de amostragem no estudo do problema são:

- redução dos custos e maior rapidez no apuramento dos resultados;
- maior profundidade na recolha de elementos;
- resolução do problema de estudar características que são destrutivas;
- minimizar os erros associados à recolha de informação (na recolha, registo e tratamento de informação há sempre erros associados). A recolha de um número menor de elementos faz, obviamente, diminuir as possibilidades deste tipo de erro.

Por questões de tempo e recursos, o investigador decidiu estudar apenas um subgrupo de indivíduos retirados da população, portanto, recorrer a uma amostra, mas com a preocupação de que o seu estudo estatístico possa fornecer informações importantes sobre a população. Analisando uma boa amostra chega-se a resultados que podem ser imputados a toda a população.

Tomada essa decisão torna-se necessário determinar o tamanho ideal de amostra, tendo presente que amostras muito grandes são dispendiosas e exigem mais tempo de manipulação e estudo e amostras pequenas são menos precisas e pouco confiáveis.

### 3.4.1. Dimensão da amostra

A determinação da dimensão da amostra é “independente da dimensão da população” (Reis e Moreira, 1993:157).

Hill e Hill (2012) referem três abordagens para encontrar o tamanho da amostra, a saber: «o caminho do esforço mínimo»; estimação por meio das «Regras do polegar»; e a estimação por meio da análise da potência.

«O caminho do esforço mínimo» consiste em obter uma amostra de tamanho tão grande quanto possível dentro dos limites dos recursos disponíveis (dinheiro, tempo, material, acesso aos respondentes ao questionário, motivação pessoal).

As «regras do polegar» procuram estimar o tamanho mínimo da amostra que possibilite uma análise estatística adequada dos dados. O tamanho mínimo da amostra depende do tipo de análise a realizar, existindo «regras do polegar» diferentes para os vários tipos de análise estatística.

A estimação por análise da potência é realizada tendo em conta: o tipo de teste, o “tamanho do efeito”, o valor do  $\alpha$  - erro tipo 1 e a potência que pretende usar. As tabelas de Cohen (1988) permitem calcular o tamanho adequado da amostra para várias técnicas estatísticas.

Nesta investigação optou-se pela técnica «regra do polegar», seguindo o conselho prático:

“Se pretender aplicar estatísticas multivariadas, é em geral suficiente usar as “regras do polegar” na escolha do tamanho da amostra. Se pretender aplicar mais do que uma análise multivariada aos mesmos dados, aplique as “regras do polegar” a cada uma das técnicas, calcule o tamanho mínimo da amostra para cada técnica e então escolha o maior destes valores como tamanho mínimo da sua amostra.

(Hill e Hill, 2012:64)

Assim, o tamanho mínimo da amostra para análises multivariadas, deve corresponder ao quádruplo do número de variáveis que pretendemos analisar (Hill & Hill, 2012). Como o estudo envolve 34 variáveis (Anexo 5) a utilizar na análise de regressão logística o tamanho mínimo da amostra deverá ser de 170 alunos.

No sentido de confirmar se a dimensão mínima da amostra encontrada pela «regra do polegar» é razoável, uma vez que é “uma regra de aproximação baseada na experiência de muitos investigadores. (...) sem nenhuma base matemática ou lógica” (Hill e Hill, 2012:54), procedeu-se a outro processo de cálculo.

Segundo Reis e Moreira (1993) o cálculo da dimensão da amostra pode ser feito matematicamente, desde que os seus elementos sejam escolhidos por um método aleatório e a dimensão da população seja finita (menos de 100 000 elementos), recorrendo à seguinte fórmula:

$$n = \frac{p(1-p)}{\frac{D^2}{2} + \frac{p(1-p)}{N}} \cdot \left(\frac{Z_{\alpha}}{2}\right)^2 \quad (3.1)$$

Onde:  $n$ —dimensão da amostra;  $p$ —proporção de indivíduos na amostra com o atributo;  $D$ — nível de precisão;  $Z_{\frac{\alpha}{2}}$ —valor da distribuição normal para um nível de confiança igual a  $\lambda = (1 - \alpha)$ ;  $\lambda$ —nível de significância;  $N$ —dimensão da população.

A determinação da dimensão da amostra exige o conhecimento da proporção da população com o atributo, ou pelo menos, de uma estimativa sua. Assim, assumiu-se a hipótese mais pessimista, ou seja, que a população apresenta dispersão máxima. Como  $p$  pode assumir todos os valores do intervalo  $[0 ; 1]$ , a função  $p(1-p)$  atinge o valor máximo (0,25) quando  $p=0,5$ . Como tal, por uma questão de prudência assumiu-se que  $p=0,5$ .

Quanto ao nível de significância e ao nível precisão, tendo em atenção a conjugação da necessidade de fiabilidade e precisão dos resultados encontrados, com controlo dos custos totais envolvidos no processo e do fator tempo foram escolhidos os seguintes valores:  $\lambda = 95\%$  e  $D = 5\%$ .

O valor da distribuição normal para um nível de confiança de  $\lambda = 95\%$  é  $Z_{0,025} = 1,96$ .

Como a dimensão da população é de 327 alunos, de acordo com os valores inerentes ao caso em estudo, a dimensão da amostra pelo método referido é de 177 alunos.

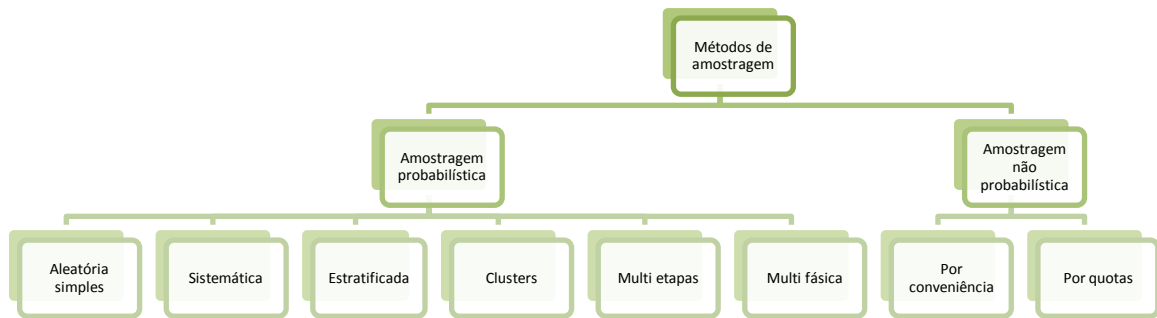
Tendo em conta os valores apresentados para a dimensão da amostra por estes dois métodos foi tomada a decisão de utilizar uma amostra de dimensão 200. Assim, a amostra será constituída por 200 alunos que frequentam o 3.º ciclo no ensino regular, no ano letivo 2014/2015 na EBSJICS.

### 3.4.2. Seleção da amostra

Escolher a amostra corresponde à seleção ponderada de várias unidades de sondagem, entre os indivíduos que compõem a população.

Hill e Hill (2012) consideram que os métodos para selecionar uma amostra podem ser agrupados em dois grandes grupos (Figura 3.2): métodos de amostragem casual (métodos probabilísticos) e métodos de amostragem não casual (métodos não probabilísticos).

**Figura 3.2 – Métodos de amostragem**



Fonte: Hill e Hill, 2012

Nos métodos probabilísticos as amostras são recolhidas de forma aleatória, e todos os elementos da população têm uma probabilidade conhecida e diferente de zero de pertencer à amostra (Fortin,2003). Os métodos mais vulgares de amostragem probabilística são: amostragem aleatória simples, amostragem sistemática, amostragem estratificada, amostragem por *clusters*, amostragem multi etapas e amostragem multi fásica (Hill & Hill, 2012).

Nos métodos não probabilísticos nem todos os elementos da população têm igual probabilidade de pertencer à amostra, correndo-se assim o risco da amostra poder não ser representativa da população (Marôco, 2011). Os métodos não aleatórios mais conhecidos são a amostragem por conveniência e a amostragem por quotas (Hill e Hill, 2012).

Os métodos de amostragem probabilísticos “são preferíveis quando o investigador pretende extrapolar com confiança para o Universo os resultados obtidos a partir da amostra” (Hill e Hill, 2012:45). Os mesmos autores referem que a utilização destes métodos de amostragem apresenta como vantagens: permitir demonstrar a representatividade da amostra; e permitir estimar, por processos estatísticos, o grau de confiança com o qual as conclusões tiradas da amostra se aplicam à População.

No entanto, “em muitos cenários de investigação social não é possível, prático ou mesmo desejável (por limitações de tempo e ou custos)” (Marôco, 2011:11), utilizar métodos de amostragem aleatórios. Os métodos de amostragem não probabilísticos podem ser úteis no início de uma investigação, no entanto, não são aconselháveis no caso de se

pretender extrapolar para a População os resultados e conclusões retirados da amostra (Hill e Hill, 2012).

Na investigação foi utilizado um dos métodos de amostragem casual, a amostragem estratificada, por:

- oferecer maior garantia de representatividade;
- permitir obter estimativas com uma dada precisão para a variável de interesse em cada estrato;
- permitir um aumento de precisão nas estimativas; essa precisão é tanto maior quanto mais homogêneos forem os estratos;
- conveniências administrativas de organização do trabalho de recolha da informação.

Segundo Hill e Hill (2012) para obter uma amostra estratificada é necessário seguir os seguintes passos:

1. Decidir o número e a natureza dos estratos.
2. Obter uma lista com todos os elementos da população.
3. A partir dessa lista, construir um quadro que caracterize a população.
4. Decidir qual o tamanho da amostra.
5. Selecionar uma amostra aleatória.

A população foi dividida em três grupos ou estratos mutuamente exclusivos. Cada ano de escolaridade do 3.º ciclo foi considerado um estrato (Tabela 3.1).

**Tabela 3.1 - Caracterização da população**

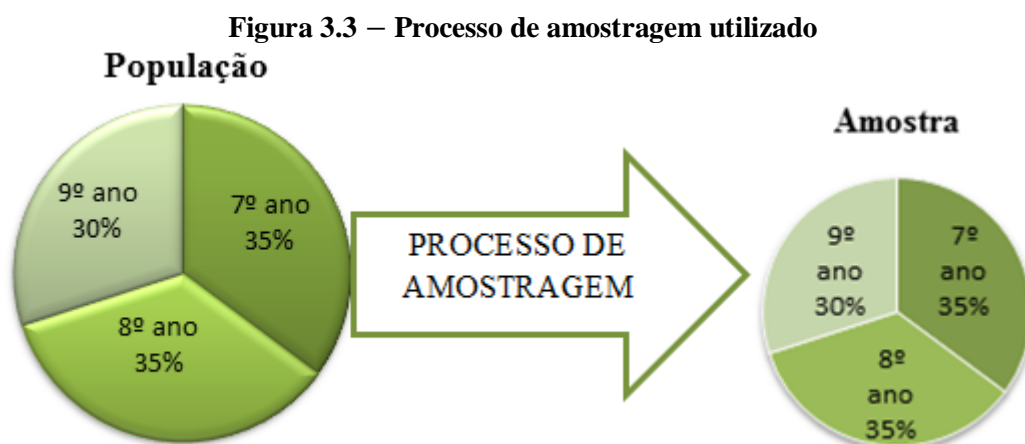
		Anos de escolaridade		
		7.º ano	8.º ano	9.º ano
Turmas	A	26	26	19
	B	25	22	20
	C	20	27	20
	D	18	19	19
	E	15	20	20
	F	11	.....	.....
	Total	115	114	98

Fonte: Elaborado pelo autor, segundo dados da pesquisa

Com o intuito de selecionar a amostra foi associado um número convencional sequencial a cada elemento da população.

De cada um dos estratos extraiu-se, aleatoriamente (pelo método da amostragem aleatória simples) uma amostra. O conjunto de todas as unidades amostrais das amostras por estratos recolhidas constitui a amostra pretendida.

A população de 327 alunos, dividida em três estratos: 7.º ano – 115 alunos (35%); 8.º ano - 114 alunos (35%) e 9.º ano – 98 alunos (30%). A dimensão da amostra pretendida é de 200, pelo que foram selecionadas amostras de dimensão 70, 70 e 60, do conjunto de alunos do 7.º ano, 8.º ano e 9.º ano respetivamente. Esta seleção teve em conta a manutenção da igualdade da proporção do tamanho da amostra em cada estrato (amostragem estratificada proporcional), como se pode observar na Figura 3.3.



Fonte: Elaborado pelo autor, segundo dados da pesquisa

Dentro de cada um dos estratos (anos de escolaridade) a seleção dos elementos foi realizada de forma aleatória, com recurso às funções “sort” e “sample” do software R Project. A Tabela 3.2 mostra os resultados do processo de amostragem.

**Tabela 3.2 - Caracterização da amostra**

		Anos de escolaridade		
		7.º ano	8.º ano	9.º ano
Turmas	A	16	19	11
	B	13	13	13
	C	13	14	11
	D	13	9	11
	E	10	15	14
	F	5	.....	.....
	Total	70	70	60

Fonte: Elaborado pelo autor, segundo dados da pesquisa

### 3.5. Variáveis em estudo

As variáveis são “qualidades, propriedades ou características de objetos, de pessoas ou de situações que são estudadas numa investigação” (Fortin, 2003:36), ou seja, são algo que o investigador “precisa medir, controlar ou manipular durante o processo de investigação” (Marôco, 2011:7).

As variáveis estatísticas podem ser classificadas em variáveis qualitativas e em variáveis quantitativas (McCall, 1998 *apud* Marôco, 2011).

As variáveis qualitativas são as características que não possuem valores quantitativos, mas, ao contrário, são definidas por várias categorias, ou seja, representam uma classificação dos indivíduos. Podem ser medidas em escala nominal ou ordinal. A nominal é medida em classes discretas, não sendo possível estabelecer à partida algum tipo de qualificação ou ordenação. A ordinal é medida em classes discretas, sendo possível definir uma determinada ordem, segundo uma relação descritível mas não quantificável.

As variáveis quantitativas são aquelas cuja escala de medida permite ordenação e quantificação. Podem ser medidas em escala intervalar ou razão. Na intervalar a variável assume valores quantitativos, os dados são expressos numa escala numérica com origem arbitrária, não possuindo uma medida de ausência de atributo (zero absoluto). Na razão a variável assume valores quantitativos sendo possível definir a relação exata entre eles, uma vez que possui um zero absoluto.

Tendo em conta que se pretende fazer uma investigação correlacional, na qual não há lugar à manipulação de variáveis, nem a existência de um grupo de controlo que permita testar o seu efeito na amostra, a escolha das variáveis deve ser realizada tendo em conta os objetivos do estudo, as questões de investigação e não resultado de uma escolha livre e arbitrária do investigador (Richardson, 1989).

As variáveis que constam desta investigação (Anexo 5) são:

- ✓ Sexo: caracterizar os alunos inquiridos em termos de género
- ✓ Idade: conhecer a idade dos inquiridos
- ✓ Ano de escolaridade: conhecer o ano de escolaridade do aluno;
- ✓ Possui internet: saber se o aluno possui ou não internet em casa;
- ✓ Tempo de sono: conhecer o número de horas que o aluno dorme por noite;
- ✓ Número de pessoas do agregado familiar: saber quantas pessoas vivem com o aluno;

- ✓ Realidade familiar: saber com quem o aluno vive;
- ✓ Encarregado de educação: saber quem é o encarregado de educação do aluno;
- ✓ Habilitações escolares do pai: conhecer o grau de instrução do pai;
- ✓ Habilitações escolares da mãe: conhecer o grau de instrução da mãe;
- ✓ Situação profissional do pai: saber se o pai está empregado; desempregado, em formação, reformado ou outra;
- ✓ Situação profissional da mãe: saber se a mãe está empregada; desempregada, em formação, reformada ou outra;
- ✓ Profissão do pai: conhecer a profissão do pai de acordo com a lista nacional;
- ✓ Profissão da mãe: conhecer a profissão da mãe de acordo com a lista nacional;
- ✓ Qualidade do ambiente familiar: saber o que o aluno pensa do seu ambiente familiar;
- ✓ Meio de transporte: saber o meio de transporte utilizado pelo aluno nas deslocações para a escola;
- ✓ Tempo de deslocações: saber o tempo que o aluno leva a chegar à escola;
- ✓ Universidade: saber se o aluno pretende de ir para a universidade;
- ✓ Disciplina que mais gosta: conhecer a disciplina preferida do aluno;
- ✓ Disciplina que menos gosta: conhecer a disciplina que o aluno menos gosta;
- ✓ Características do aluno: saber se o aluno se considera pontual, participativo, empenhado, distraído ou trabalhador;
- ✓ Dificuldades: saber se o aluno sente dificuldades na disciplina de matemática;
- ✓ Quem esclarece as dúvidas: saber com quem o aluno esclarece as dúvidas de matemática;
- ✓ Periodicidade do estudo: saber com que frequência o aluno estuda matemática;
- ✓ Tempo de estudo: saber o tempo semanal de estudo do aluno para matemática;
- ✓ Comportamento da turma: saber como o aluno classifica o comportamento da turma nas aulas de matemática;
- ✓ Comportamento do aluno: saber como o aluno classifica o seu comportamento nas aulas de matemática;
- ✓ Relacionamento do aluno-professora: conhecer como o aluno se relaciona com a professora de matemática;

- ✓ Trabalhos de casa de matemática: saber se o aluno realiza regularmente os trabalhos de casa de matemática;
- ✓ Interesse: pela disciplina de matemática;
- ✓ Reprovações: saber o número de reprovações do aluno a matemática desde o 5.º ano;
- ✓ Anos de escolaridade da reprovação: saber que anos o aluno reprovou a matemática;
- ✓ Responsáveis pela repetência: saber na opinião do aluno de quem é a responsabilidade da repetência;
- ✓ Factores de insucesso: saber a opinião do aluno se os factores seguintes influenciam o insucesso (desinteresse pela disciplina, falta de hábitos e métodos de trabalho; falta de pré requisitos necessários; falta de atenção e concentração nas aulas; manual desadequado, indisciplina na sala de aula, carga horária global excessiva, dificuldades de interpretação e de compreensão, falta de apoio por parte do professor; falta de apoio familiar; rapidez com que o professor dá a matéria; falta de oportunidade para esclarecer dúvidas) ;
- ✓ Medidas para reduzir o insucesso: saber a opinião do aluno se as medidas seguintes podem contribuir para diminuir o insucesso (utilização de tecnologias; diminuição do número de aulas; aumento do número de aulas; mudança das estratégias de ensino; promoção de aulas de apoio; diminuição do número de alunos por turma; mudança nos conteúdos; maior articulação das matérias com aspectos da vida diária).

### **3.6. Instrumentos de recolha**

#### **3.6.1. Elaboração de questionário**

Para a concretização deste estudo, foram utilizadas as seguintes técnicas: a análise de documentação (da base de dados e de pautas finais) e a elaboração de questionário.

Sousa (2006) refere que os inquéritos são instrumentos de pesquisa que visam colher informação sobre ideias, afetos e comportamentos dos indivíduos, baseando-se em questionários. Identificando três tipos de inquérito:

- a) Inquéritos descritos – os quais têm por objetivo documentar e descrever o que existe num determinado momento.
- b) Inquéritos analíticos – os quais tentam descrever e explicar quais as razões para a ocorrência de determinados fenómenos.
- c) Inquéritos mistos – os quais misturam características dos inquéritos descritivos com as dos inquéritos analíticos.

Na presente investigação optou-se por um inquérito misto, dado que se pretende traçar o perfil de um aluno com (in)sucesso a Matemática, conhecer a perceção dos alunos a cerca do(s) responsáveis pela falta de sucesso, os fatores que contribuem para o insucesso e avaliar até que ponto a implementação de determinadas medidas poderão contribuir para aumentar o sucesso.

O questionário é um instrumento de recolha de dados/informação, utilizado em investigação quantitativa. A sua estrutura e o tipo de construção é fundamental para conseguir recolher os dados pretendidos e realizar análises diversificadas de modo a extrair o máximo de informação e resultados conclusivos.

Após a seleção da amostra e da escolha das variáveis a estudar e antes da elaboração do questionário foi necessário tomar determinadas decisões, tais como: modo de aproximação aos respondentes; comprimento do questionário; aspetos como confidencialidade e anonimato; sequência de questões; e tipo de questão a utilizar.

A utilização do questionário apresenta vantagens e desvantagens. Como vantagens pode-se referir a possibilidade de ser repetido para uma amostra grande, exigindo menores recursos financeiros e evitando o condicionamento psicológico, garantindo o anonimato, comparativamente com a realização de entrevistas. No que se refere a desvantagens, o investigador não consegue controlar a honestidade das respostas e poderá não ser aplicável a todo o tipo de populações.

Hill e Hill (2012) referem que aquando da elaboração do questionário deve-se pegar nas hipóteses gerais e decidir o tipo de perguntas a utilizar, para medir as variáveis que se pretendem estudar, bem como o tipo de resposta mais adequado para cada pergunta, o tipo de escala de medida associado às respostas e os métodos corretos para analisar os dados.

A formulação das perguntas que compõem o questionário é um dos pontos cruciais para que o inquérito seja bem sucedido, sendo necessário ponderar em termos de extensão do questionário e característica das perguntas. Em termos de extensão do questionário, o

número de perguntas deve ser o estritamente necessário, tendo presente que se o questionário for longo menor número de pessoas aceitarão responder ao e torna-se fastidioso para as que aceitam responder. As características gerais das perguntas, estas devem ser claras, simples, não ambíguas, acessíveis aos inquiridos e pertinentes sem conteúdos demasiadamente pessoais.

O inquérito pode incluir um único tipo de perguntas ou, pelo contrário, vários tipos. Os tipos de perguntas mais utilizados são:

- Perguntas fechadas – são apresentadas as alternativas de resposta ao inquirido, limitando-se este a escolher entre as respostas alternativas. Este tipo de pergunta permite um apuramento rápido dos resultados, facilita o tratamento estatístico dos dados, facilmente perceptível pelo inquirido e de resposta simples. No entanto, as respostas podem não contemplar todas as possibilidades e a informação obtida pode ser escassa.
- Perguntas abertas – o inquirido responde livremente à pergunta, permitindo assim riqueza de informação, no entanto pode ser difícil o tratamento dos dados, dada a variedade de respostas e possível falta de objectividade.
- Perguntas semi abertas – é uma junção de pergunta aberta e pergunta fechada. Numa primeira fase o inquirido escolhe uma das respostas e numa segunda fase justifica a sua resposta. Este tipo de pergunta permite facilidade e rapidez no apuramento dos dados e informação suficiente, uma vez que é apresentada uma justificação da resposta.
- Perguntas dicotómicas – são um caso especial de perguntas fechadas, em que as resposta apenas pode ser uma de duas hipóteses apresentadas.
- Perguntas encadeadas/dependentes – em que a resposta a uma segunda pergunta está dependente da resposta à primeira.
- Pergunta com matriz de respostas – este tipo de pergunta permite obter diversas informações na mesma pergunta. É apresentado um quadro para facilitar a resposta, limitando-se o inquirido a assinalar a opção com a qual se identifica.
- Perguntas que implicam ordem de preferência – o inquirido atribui uma ordem de preferência às várias alternativas de resposta apresentadas.

- Perguntas de avaliação de opinião – baseada numa escala de classificação concebida para avaliar atitudes ou opiniões, visam o conhecimento quantificado e direto do comportamento do inquirido.
- Perguntas de múltipla escolha – o inquirido escolhe mais de uma opção de resposta de entre as alternativas apresentadas, mas não tem oportunidade de manifestar a sua opinião. Têm como desvantagem o facto das alternativas de resposta poderem não contemplar todas as situações possíveis.

A cada questão está associada a escala de medida que será indicador do tipo de análise estatística a efetuar. Os tipos de escala que podem ser utilizadas num questionário são a escala nominal, ordinal e métrica, sendo as duas primeiras as mais frequentes (Hill e Hill, 2012).

A escala nominal consiste numa característica do indivíduo em que a resposta é qualitativamente diferente e mutuamente exclusiva. As escalas ordinais aplicam-se, também, a questões qualitativas mas que admitem uma ordenação numérica, não permitindo no entanto, medir as diferenças entre as categorias. A escala métrica consiste na quantificação numérica de uma variável e apresenta duas formas: escala de intervalo e escala de rácio.

Os blocos temáticos que estruturam o questionário que foi elaborado são os seguintes: dados pessoais, dados do agregado familiar e dados escolares. As questões colocadas têm por objetivo obter dados sobre as variáveis que se pretendem estudar.

A maioria das questões efetuadas no questionário são fechadas e tem uma escala ordinal ou uma escala nominal com um número variável de respostas alternativas.

### **3.6.1.1. Apresentação do questionário**

O questionário elaborado (Anexo II), com trinta e duas questões, pretendeu recolher dados pessoais, familiares e escolares do aluno, sobretudo referentes à disciplina de matemática e foi construído essencialmente com base em perguntas fechadas. Esta escolha deve-se ao facto de ser mais fácil aplicar um estudo estatístico na análise deste tipo de perguntas. No entanto, no questionário existem perguntas dicotómicas, fechadas, dependentes, abertas, com matriz de resposta, múltipla escolha e de avaliação de opinião.

O inquérito envolve a recolha simultânea de dados sobre diferentes variáveis, o que torna impossível estabelecer uma ordem temporal entre elas, não sendo, no entanto, nenhuma destas variáveis manipulada pelo investigador.

Em termos de dados pessoais foi solicitada informação quanto ao género, a idade, o ano de escolaridade, se tinham internet em casa, o número de horas que dormiam por noite.

Em termos de dados familiares foi solicitada informação sobre o número de pessoas do agregado familiar, pessoas com quem reside, quem é o seu encarregado de educação, habilitações literárias, situação profissional e profissão do pai e da mãe, e qualidade do ambiente familiar.

Em termos de dados escolares foi questionado o meio de transporte, o tempo de deslocação, intenção de ir para a universidade, a disciplina que gosta mais e a que gosta menos, características do aluno (pontual, participativo, empenhado, distraído, trabalhador), e questões específicas em relação à disciplina de matemática, nomeadamente: se sente dificuldades, a quem recorre para esclarecer dúvidas, com que periodicidade estuda e quanto tempo semanalmente, comportamento da turma e do aluno, relacionamento com a professora de matemática, realização dos trabalhos de casa, interesse pelos conteúdos, número de retenções à disciplina desde o 5.º ano. Foram também colocadas três questões de opinião sobre quem o aluno considera responsável pelas retenções, sobre fatores que contribuem para as retenções e sobre medidas para melhorar os resultados escolares.

### **3.6.1.2. Pré teste do questionário**

Após a construção de uma versão “provisória” do questionário, para avaliar a adequação do mesmo, foi solicitado a quatro pessoas a leitura e opinião sobre a clareza e compreensão do mesmo e de seguida realizado o pré teste.

A aplicação do pré teste do questionário teve como objetivos:

- identificar potenciais problemas com o questionário, nomeadamente identificar perguntas que justifiquem uma modificação da sua redação ou a sua eliminação e/ou modificações no formato do questionário;
- verificar o nível de compreensão das perguntas e a adequabilidade da sequência das mesmas;
- Aferir o tempo de preenchimento;
- Avaliar a obtenção da informação pretendida.

O pré teste foi aplicado com caráter anónimo a quinze indivíduos que fazem parte da população do estudo. Aos inquiridos foi-lhes explicado, pela investigadora, os objetivos do questionário e dada a informação de que se tratava de um pré teste e da importância do mesmo. De seguida, foi-lhes solicitada indicação das principais dificuldades no preenchimento do questionário e a formulação de críticas e sugestões.

No decorrer da aplicação do pré teste, foram colocadas algumas dúvidas pelos respondentes, nomeadamente na pergunta 21 “Quando tens dúvidas ou dificuldades em Matemática, a quem recorres para esclarecê-las?” e na pergunta 30 “No que se refere à repetência pensas que a responsabilidade é de quem? (Assinala apenas uma opção)”.

O pré teste do questionário foi aplicado no dia 16 de março de 2015 e os respondentes demoraram entre 15 e 20 minutos a responder ao questionário, não se demonstrando cansados nem enfastiados com o preenchimento do questionário. Nenhum dos inquiridos apresentou críticas nem sugestões ao questionário.

Após analisadas as dificuldades sentidas pelos respondentes, dado que na pergunta 21 alguns dos respondentes afirmaram recorrer a mais do que uma entidade para esclarecer as suas dúvidas. Na versão final do questionário foi adicionada à pergunta a nota “(Podes assinalar mais do que uma opção)”. Na pergunta 30, dado que alguns respondentes nunca haviam reprovado, e por isso questionaram se deveriam responder à questão, a mesma foi reformulada passando o texto a ser o seguinte: “Quando um aluno reprova a Matemática pensas que a responsabilidade é de quem? (Assinala apenas uma opção)”. Assim, o questionário pré teste com as referidas alterações deu lugar ao questionário final da investigação (Anexo 2).

### **3.7. Recolha de dados**

A aplicação do questionário decorreu no período de 17 de Abril a 23 de Abril de 2015. O questionário foi aplicado a 200 alunos do 3.º ciclo do Ensino Básico, da EBSJICS no ano letivo 2014/2015 provenientes de seis turmas de sétimo ano, cinco turmas de oitavo ano e cinco turmas de nono ano, todas elas de ensino regular.

A aplicação do questionário foi feita pela pesquisadora decorreu numa aula do horário da turma com a autorização prévia do docente que a lecionava.

No dia da aplicação do questionário, antes da sua entrega aos respondentes, foi-lhes explicado que a finalidade do questionário era a de recolher informações para a realização

de um estudo no âmbito de uma dissertação de Mestrado em Estatística, Matemática e Computação, cujo tema é: “Análise de regressão aplicada ao (in)sucesso escolar a Matemática no 3.º ciclo - Estudo de caso” e garantido o anonimato e a confidencialidade dos dados.

Todos os alunos selecionados pelo processo de amostragem responderam o questionário e as respostas dos inquiridos anotadas diretamente na folha do questionário.

A recolha decorreu com normalidade, na presença do investigador e de um docente. Os respondentes demoraram entre 15 a 20 minutos a preencherem o questionário, tal como havia acontecido na aplicação do pré teste do questionário.

### **3.8. Tratamento de dados**

Após a recolha de dados, procedeu-se à codificação e inserção dos mesmos num ficheiro para análise estatística.

No tratamento dos dados foi utilizado o software SPSS (*Statistical Package for Social Sciences*), versão 22.

Este *software* teve a sua primeira versão em 1968 criada, por Norman H. Nie, C. Hadlai Hull e Dale H. Bent, para grandes computadores, surgindo em 1984 uma versão para computadores pessoais.

O SPSS é um software de análise de dados que contempla um conjunto variado de ferramentas e módulos suplementares, amplamente utilizado nas mais variadas áreas científicas e empresariais, com maior incidência nas aplicações em Ciências Sociais, Educação, Ciências da Saúde e Medicina, Gestão e Marketing. É um software de utilização relativamente fácil e de grande aplicabilidade, possuindo um vasto leque de recursos importantes no tratamento de dados estatísticos. A opção por este software prende-se com o facto de ele permitir elaborar gráficos variados e atrativos; relatórios; realizar uma análise, recorrendo às opções disponíveis nas várias janelas; aplicar métodos de estatística descritiva, visualização de dados univariados e multivariados; inferências estatísticas; técnicas de estatística multivariada.

O SPSS é constituído por duas janelas principais: a janela SPSS Editor de Dados, na qual são apresentados os dados em estudo e as características das variáveis definidas e pela janela SPSS Visualizador, na qual aparecem os resultados dos estudos realizados.

**CAPÍTULO IV**  
**APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS**

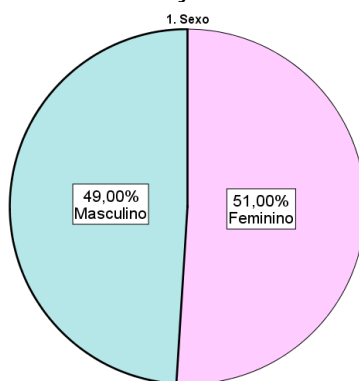
## 4. APRESENTAÇÃO E DISCUSSÃO DE RESULTADOS

Neste capítulo apresentam-se e discutem-se os resultados da pesquisa e é constituído por três secções. Na secção um, aplica-se uma Análise Descritiva para caracterização da amostra, na secção dois apresenta-se uma Análise de Regressão Logística para encontrar um modelo de previsão que estime a probabilidade de um aluno obter sucesso ou insucesso a matemática. Na secção três faz-se aplicação da Análise de *Clusters* de forma a agrupar variáveis.

### 4.1. Caracterização da amostra

A amostra é constituída por 200 alunos (70 do 7.º ano, 70 do 8.º ano e 60 do 9.º ano), de ambos os sexos, sendo 98 (49%) do sexo masculino e 102 (51%) do sexo feminino (Gráfico 4.1), estando 70 a frequentar o 7.º ano, 70 o 8.º ano e 60 o 9.º ano.

**Gráfico 4.1 – Distribuição da amostra por sexos**



Fonte: O autor, com base nos dados

Para testar se a amostra é de facto aleatória recorreu-se ao teste dos *runs* (sequências), o qual pode ser aplicado a qualquer tipo de dados, no entanto, o SPSS processa este teste apenas para variáveis do tipo numérico. Assim, para verificar se a amostra é constituída por indivíduos selecionados aleatoriamente escolhemos a variável numérica “idade” (Tabela 4.1). A variável deve ser dividida em dois grupos, para tal foi indicado como o ponto de corte, a mediana (14 anos). Qualquer valor menor do que a mediana irá pertencer a um grupo e qualquer valor maior ou igual à mediana irá pertencer a outro grupo. Como o valor de  $p$  é 0,000, portanto significativo, podemos dizer que a variável idade é aleatória.

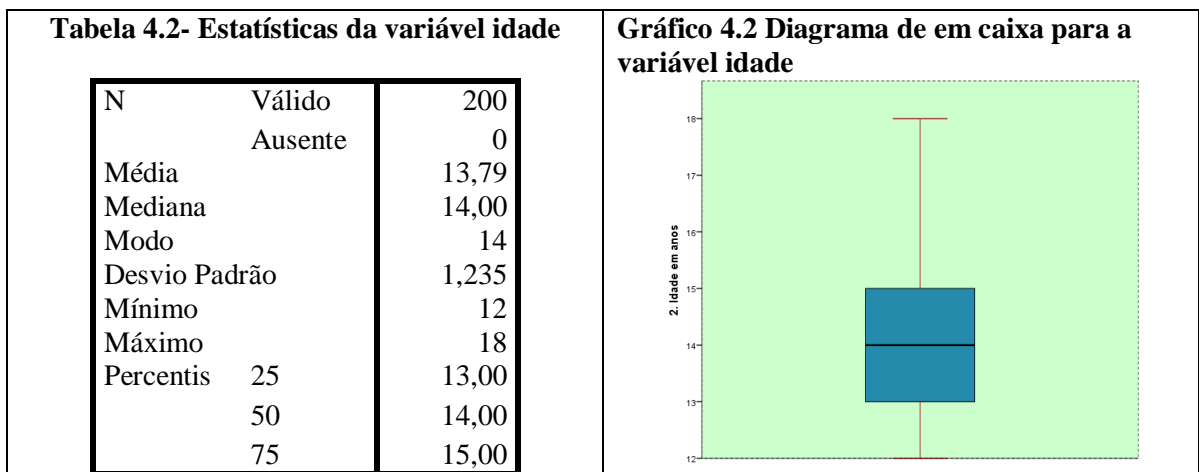
**Tabela 4.1 – Teste de seqüências**

	2. Idade
Valor de teste <sup>a</sup>	14
Casos < Valor de Teste	87
Casos >= Valor de Teste	113
Total de casos	200
Número de Sequências	50
Z	-7,112
Significância Sig. (2 extremidades)	,000
Sig exata (2 extremidades)	,000
Probabilidade de ponto	,000

a. Mediana

Fonte: O autor, com base nos dados

Na Tabela 4.2 pode observar-se que os inquiridos têm idades compreendidas entre os 12 (mínimo) e os 18 (máximo) anos, sendo a idade média de 13,79 anos e a moda de 14 anos. Pelo Gráfico 4.2 pode concluir-se que 25% dos inquiridos têm idades menores ou iguais a 13 anos, 25% têm idades superiores ou iguais a 15 anos, metade têm idades compreendidas entre os 13 e 15 anos, existindo enviesamento do lado direito, isto é, os dados estão mais dispersos, ou seja, menos concentrados na parte superior do que na parte inferior.



Fonte: O autor, com base nos dados

Fazendo a comparação entre as idades dos alunos e o ano de escolaridade que frequentam, Tabela 4.3, verifica-se que no 7.º ano existem alunos com idades entre os 12 e os 16 anos, quando a idade normal seria 12/13 anos. No 8.º ano existem alunos com idades entre os 13 e os 16 anos quando a idade normal seria 13/14 anos e no 9.º ano existem alunos com idades entre os 14 e os 18 anos, quando a idade normal seria 14/15 anos. A existência de alunos num determinado ano de escolaridade com idades diferentes da idade normal indicia repetência e em alguns casos múltiplas repetências. Repetências essas

também na disciplina de Matemática, como se que pode comprovar pelas respostas dadas pelos inquiridos à questão “ 29. Desde o 5.º ano já obtiveste nota inferior a 3 no 3.º período na disciplina de Matemática?”

**Tabela 4.3 -Tabulação cruzada Idade \*Ano de escolaridade**

Contagem	3. Ano de escolaridade			Total	
	7.º ano	8.º ano	9.º ano		
2. Idade em anos	12	32	0	0	32
	13	22	33	0	55
	14	9	26	23	58
	15	6	10	23	39
	16	1	1	10	12
	17	0	0	3	3
	18	0	0	1	1
	Total	70	70	60	200

Fonte: O autor, com base nos dados

Quanto ao facto de já terem tido alguma repetência a partir do 5.º ano (início do 2.º ciclo) na disciplina de Matemática (Tabela 4.4) 102 alunos (51%) responderam não e 98 alunos (48,5%) responderam sim e um aluno (0,5%) não respondeu à questão. Cerca de 29 alunos (14,5%) reprovaram uma vez, 25 alunos (12,5%) reprovaram 2 vezes, 20 alunos (10%) reprovaram 3 vezes e 23 alunos (11,5%) mais de 3 vezes.

**Tabela 4.4 – Número de reprovações do aluno a Matemática a partir do 5.º ano**

Reprovações a Matemática		Frequência	%
Válido	Não	102	51,0
	1 vez	29	14,5
	2 vezes	25	12,5
	3 vezes	20	10,0
	Mais que 3 vezes	23	11,5
	Total	199	99,5
Ausente	NR	1	,5
	Total	200	100,0

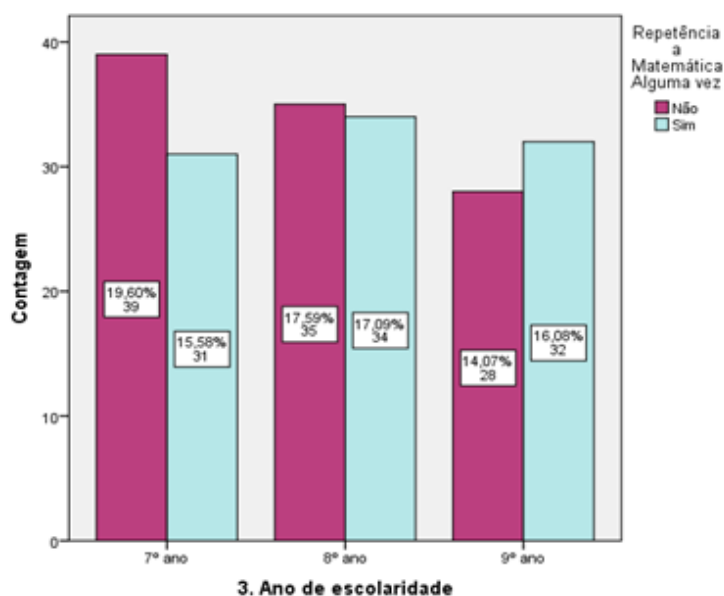
Analisando o(s) ano(s) em que os alunos reprovaram a Matemática (obtenção nível inferior a 3), Tabela 4.5, conclui-se que dos alunos inquiridos 1 não respondeu à questão, 37 (18,5%) reprovaram no 5.º ano, 44 (22%) reprovaram no 6.º ano, 62 (31%) no 7.º ano, 43 (21,5%) no 8.º ano e 17 (8,5%) no 9.º ano.

**Tabela 4.5 – Anos em que o aluno reprovou a Matemática**

		5.º ano		6.º ano		7.º ano		8.º ano		9.º ano	
		N.º	%	N.º	%	N.º	%	N.º	%	N.º	%
Válido	Não	60	30,0	53	26,5	35	17,5	54	27,0	80	40,0
	Sim	37	18,5	44	22,0	62	31,0	43	21,5	17	8,5
	Total	97	48,5	97	48,5	97	48,5	97	48,5	97	48,5
Ausente	NSA	101	50,5	101	50,5	101	50,5	101	50,5	101	50,5
	NR	2	1,0	2	1,0	2	1,0	2	1,0	2	1,0
	Total	103	51,5	103	51,5	103	51,5	103	51,5	103	51,5
Total		200	100	200	100	200	100	200	100	200	100

Dos alunos inquiridos encontram-se a frequentar o 7.º ano 16%, o 8.º ano 17% e o 9.º ano 16% que já reprovaram alguma vez a matemática (Gráfico 4.3). Dos alunos que responderam que já reprovaram a matemática alguma vez, cerca de 32% referem que estão no 7.º ano, 35% no 8.º ano e 33% no 9.º ano (Tabela 4.6).

**Gráfico 4.3 – Repartição dos alunos que já reprovaram alguma vez**

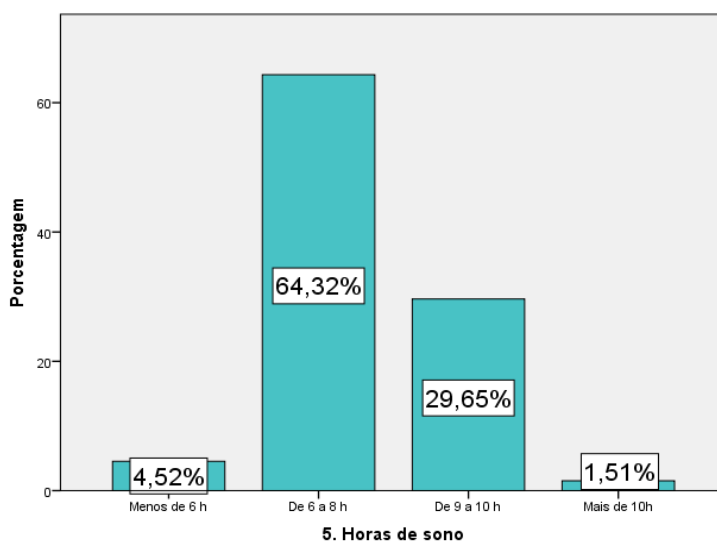


**Tabela 4.6 – Tabulação cruzada Ano de escolaridade \* Repetência a Matemática Alguma vez**

		Repetência a Mat.		Total
		Não	Sim	
3. Ano de escolaridade	7.º ano	39	31	70
	Contagem			
	% em Repetência a Mat.	38,2%	32,0%	35,2%
8.º ano	8.º ano	35	34	69
	Contagem			
	% em Repetência a Mat.	34,3%	35,1%	34,7%
9.º ano	9.º ano	28	32	60
	Contagem			
	% em Repetência a Mat.	27,5%	33,0%	30,2%
Total	Contagem	102	97	199
	% em Repetência a Mat.	100,0%	100,0%	100,0%

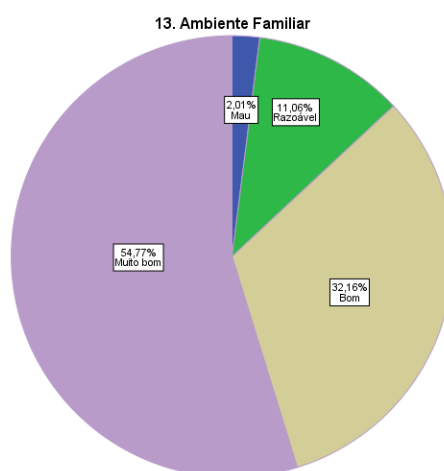
A grande maioria dos inquiridos (94,5%) possui acesso a internet em casa. Cerca de 64% dos inquiridos dorme em média entre 6 e 8 horas por noite, 30% dorme entre 9 e 10 horas e 5% menos de 6 horas (Gráfico 4.4).

**Gráfico 4.4 – Horas de sono por noite**



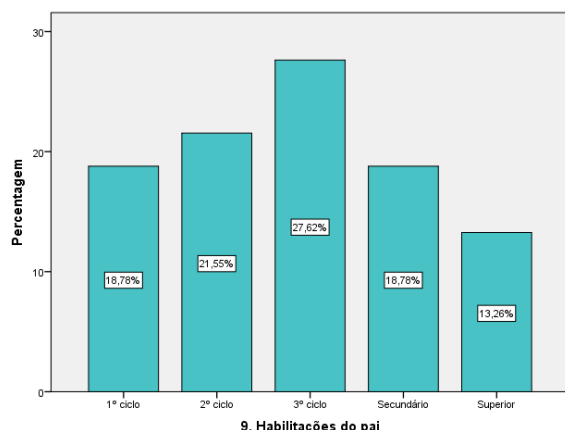
Na generalidade os inquiridos referiram ter um ambiente familiar favorável (Gráfico 4.5), 55% referiu ser muito bom, 32% bom e apenas 2% considera ter mau ambiente.

**Gráfico 4.5 – Ambiente familiar**

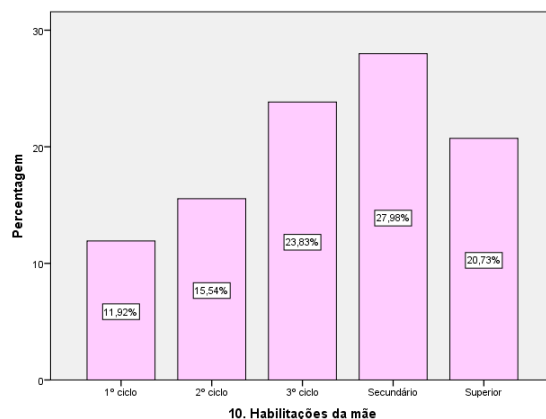


Em termos de habilitações académicas dos pais dos inquiridos (Gráfico 4.6 e Gráfico 4.7) nenhum inquirido referiu ter pais analfabetos existindo maior percentagem de mães do que de pais com o ensino secundário e com o ensino superior. Dos pais 28% têm o 3.º ciclo, 19% o 1.º ciclo e apenas 13% possui curso superior, das mães 28% têm o secundário, 24% o 3.º ciclo e 21% curso superior.

**Gráfico 4.6 – Habilitações dos pais**



**Gráfico 4.7 – Habilitações das mães**



Dos alunos cujas mães têm o 1.º ciclo, a maioria, cerca de 55%, refere que o pai tem igualmente o 1.º ciclo; dos alunos cujas mães têm o 2.º ciclo, metade, cerca de 50%, refere que o pai tem o mesmo nível de ensino; dos alunos cujas mães têm o 3.º ciclo, cerca de 44%, refere que o pai tem a mesma habilitação, dos alunos cujas mães têm o secundário, 34% refere que o pai tem a mesma habilitação, dos alunos cujas mães têm o ensino superior, cerca de 34%, refere que o pai também tem o mesmo nível de ensino (Tabela 4.7).

**Tabela 4.7 – Tabulação cruzada Habilitações do pai \* Habilitações da mãe**

			10. Hab. da mãe					Total
			1.º ciclo	2.º ciclo	3.º ciclo	Secund.	Superior	
9. Hab. do pai	1.º ciclo	Contagem	12	6	9	5	2	34
		% em hab. mãe	54,5%	21,4%	22,0%	9,4%	5,7%	19,0%
	2.º ciclo	Contagem	6	14	6	10	3	39
		% em hab. mãe	27,3%	50,0%	14,6%	18,9%	8,6%	21,8%
	3.º ciclo	Contagem	2	6	18	12	11	49
		% em hab. mãe	9,1%	21,4%	43,9%	22,6%	31,4%	27,4%
	Secundário	Contagem	1	2	5	18	7	33
		% em hab. mãe	4,5%	7,1%	12,2%	34,0%	20,0%	18,4%
	Superior	Contagem	1	0	3	8	12	24
		% em hab. mãe	4,5%	0,0%	7,3%	15,1%	34,3%	13,4%
Total		Contagem	22	28	41	53	35	179
		% em hab. mãe	100	100	100	100	100	100

Para averiguar se existe uma relação entre as habilitações dos progenitores, ou seja, se o facto de um aluno responder que ambos os progenitores têm a mesma habilitação não é uma resposta independente, recorreu-se ao teste do Qui quadrado de Pearson. No entanto,

não se verifica um dos pressupostos deste teste, uma vez que mais do que 20% das observações são inferiores a 5, pelo que o teste não pode ser aplicado (Tabela 4.8).

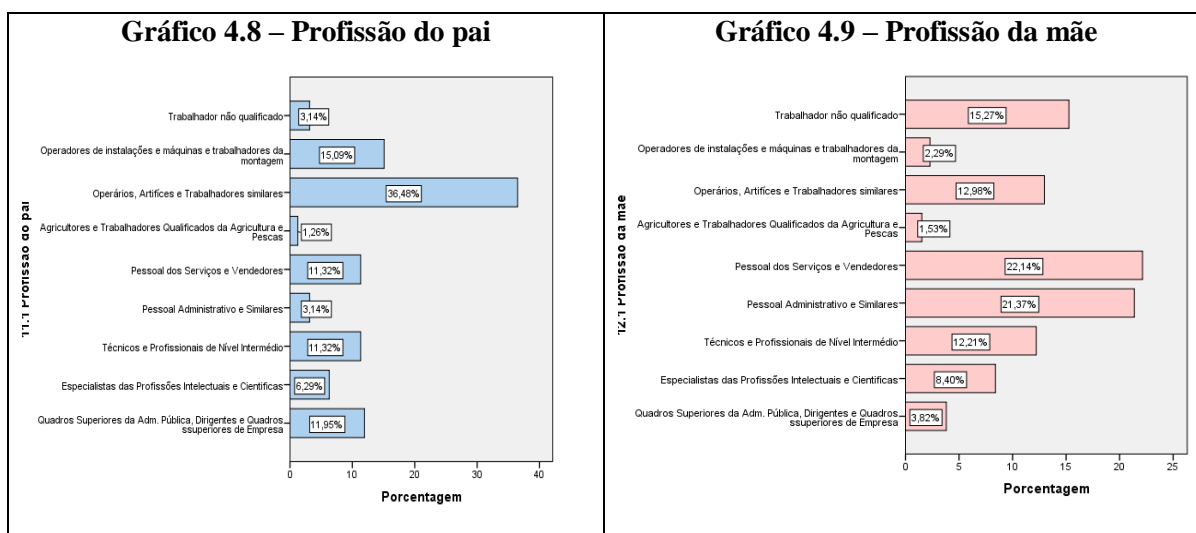
**Tabela 4.8 – Teste qui-quadrado**

	Valor	df	Significância Sig. (2 lados)
Qui-quadrado de Pearson	72,758 <sup>a</sup>	16	,000
Razão de verossimilhança	69,449	16	,000
Associação Linear por Linear	43,750	1	,000
N de Casos Válidos	179		

a. 6 células (24,0%) esperavam uma contagem menor que 5.  
A contagem mínima esperada é 2,95

Já os inquiridos 75% têm aspiração de prosseguir estudos de forma a irem para a universidade. Dos alunos cuja mãe têm como habilitações o ensino superior, a maioria, cerca de 85%, refere que pretende ir para a universidade e dos alunos cujo pai têm o ensino superior, 83% refere que pretende ir para a universidade.

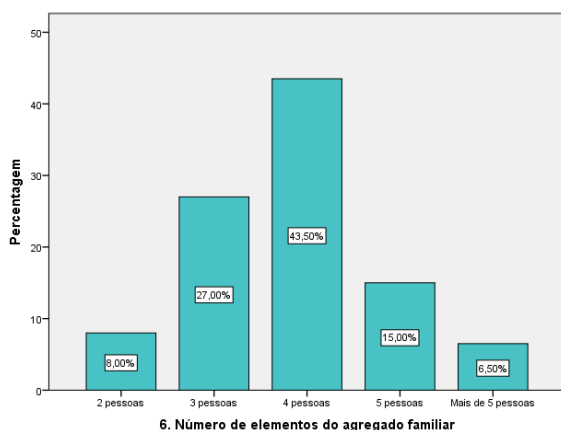
Em termos profissionais, no caso dos pais predomina os grupos profissionais operários, artífices e trabalhadores similares (36,48%), operadores de instalações e máquinas e trabalhadores de montagem (15,09%) e quadros superiores da administração pública, dirigentes e quadros superiores de empresa (11,95%) (Gráfico 4.8), no caso das mães predomina os grupos profissionais pessoal dos serviços e vendedores (22,14%), pessoal administrativo e similares (21,37%) e trabalhadores não qualificados (15,27%) (Gráfico 4.9).



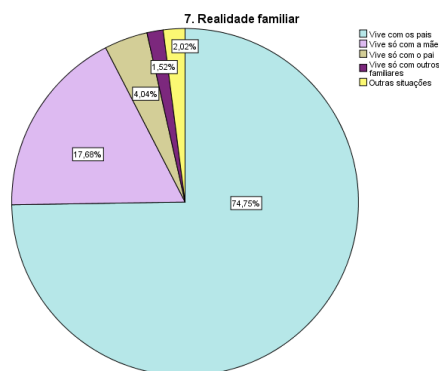
Em termos de agregado familiar (Gráfico 4.10) 43,5% dos inquiridos fazem parte de agregados familiares com 4 pessoas, 27% de agregados com 3 pessoas, 15% de agregados

com 5 pessoas, 8% de agregados com 2 pessoas e 6,5% de agregados com mais de 5 pessoas. Em termos de composição familiar (Gráfico 4.11) 75% vive com ambos os progenitores e aproximadamente 22% fazem parte de famílias monoparentais, em que 18% vive apenas com a mãe e 4% apenas com o pai.

**Gráfico 4.10 – Número de elementos do agregado familiar**

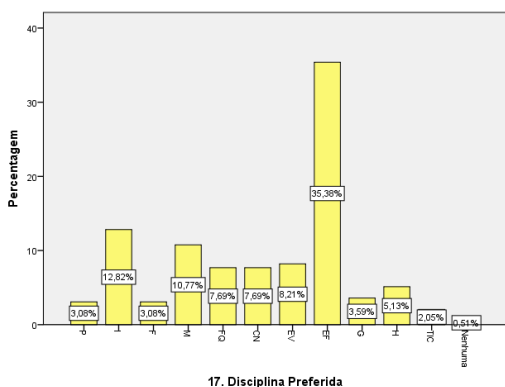


**Gráfico 4.11 – Realidade familiar**

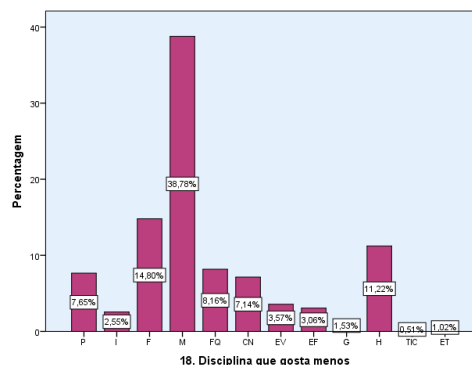


Apesar da heterogeneidade em termos de “preferências” pelas disciplinas a que reúne maior número de preferências por parte dos inquiridos (Gráfico 4.12) é Educação Física (35%), sendo a Matemática a preferida por 11%, a qual também se revelou como sendo a disciplina da qual os alunos menos gostam (39%), (Gráfico 4.13).

**Gráfico 4.12 – Disciplina preferida**

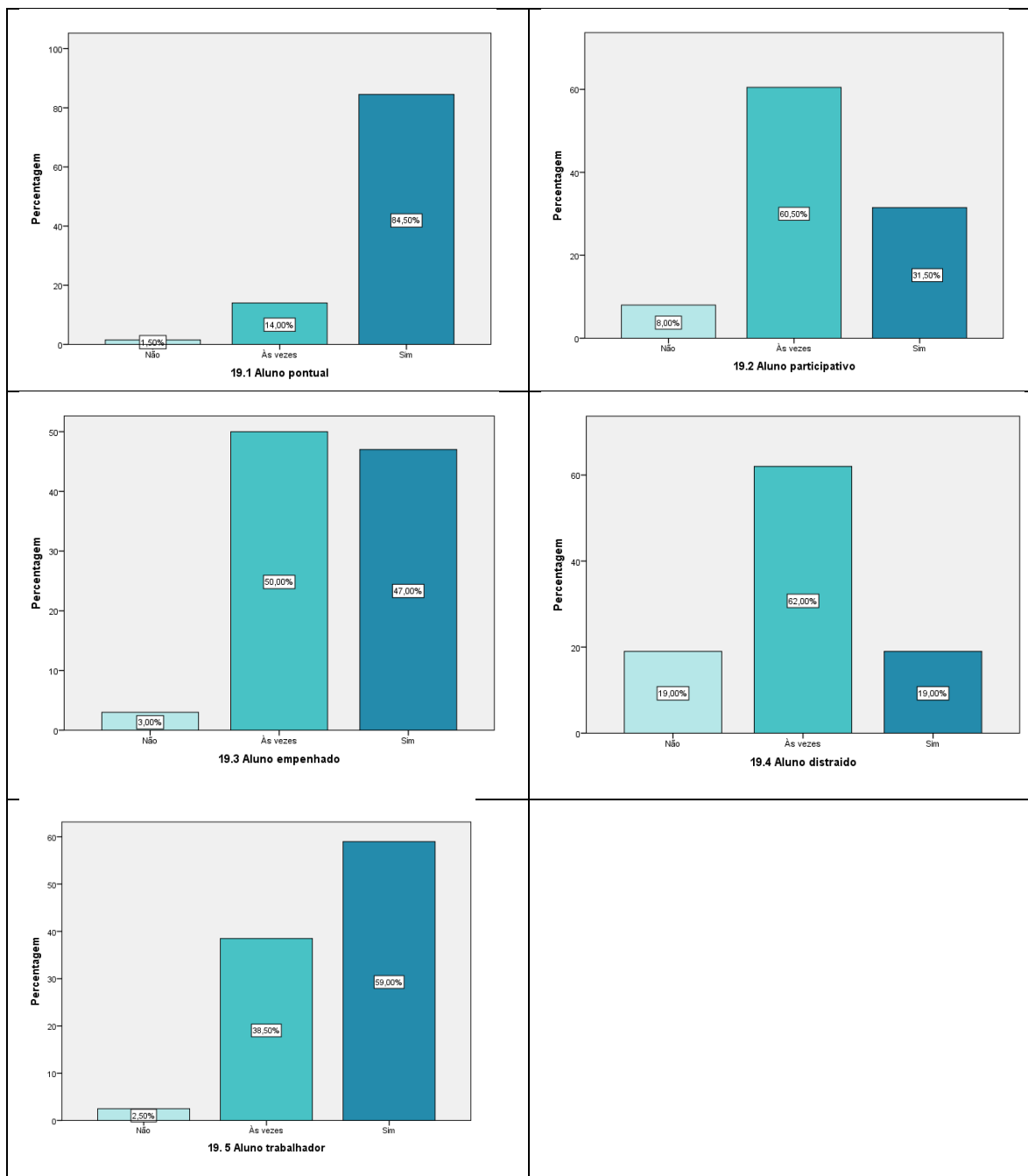


**Gráfico 4.13 – Disciplina que menos gosta**



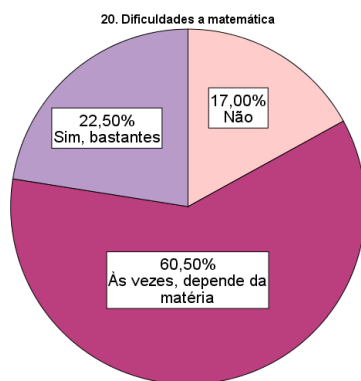
Os inquiridos foram questionados se se consideram pontuais, participativos, empenhados, distraídos ou trabalhadores (Gráfico 4.14) e obteve-se os seguintes resultados: 84,5% consideram-se pontuais, 59% trabalhadores, 47% empenhados (50% só o é às vezes), 31,5% participativos (60,5% só o é às vezes) e 19% distraídos (50% só o é às vezes).

**Gráfico 4.14 – Características dos alunos**



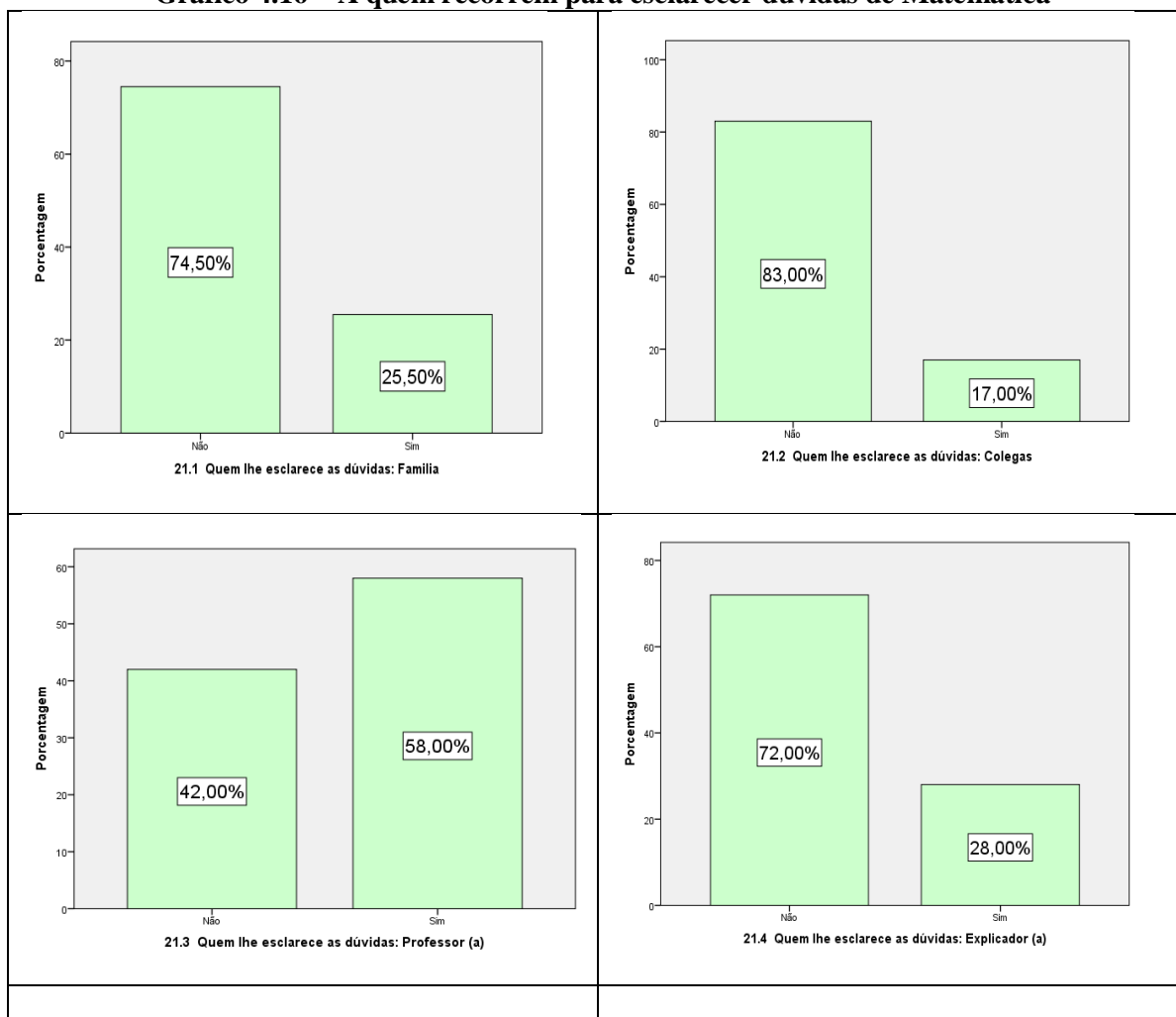
Dos inquiridos, 22,5% referiu sentir bastantes dificuldades na disciplina Matemática, 60,5% referiu sentir às vezes, dependendo das matérias e apenas 17% referiu não sentir dificuldades na disciplina (Gráfico 4.15).

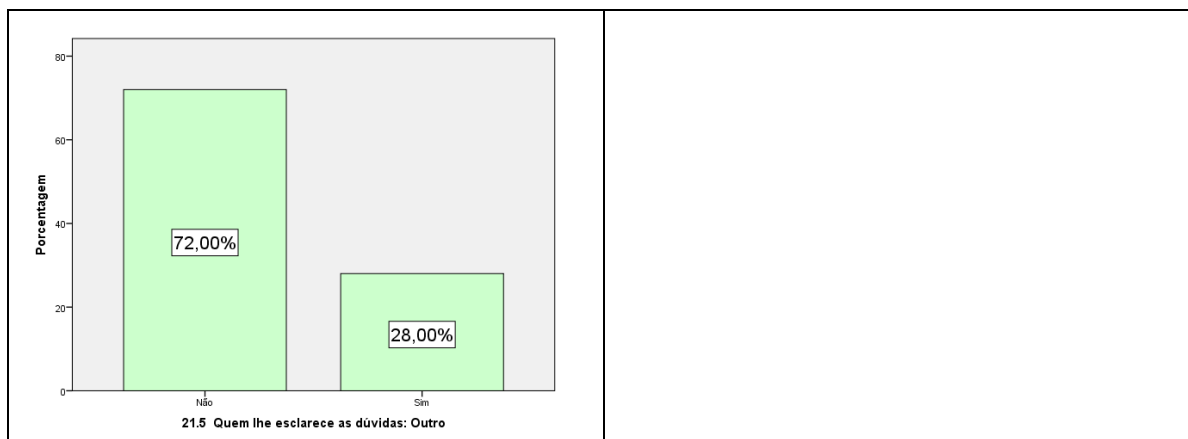
**Gráfico 4.15 – Dificuldades sentidas na disciplina de Matemática**



Quando questionados a quem recorrem para esclarecer dúvidas de Matemática alguns inquiridos responderam a mais do que uma entidade (Gráfico 4.16) 25,5% respondeu que recorre à família, 17% a colegas, 58% ao professor, 28% ao explicador e 5% a outros, entre os quais a internet e a ninguém.

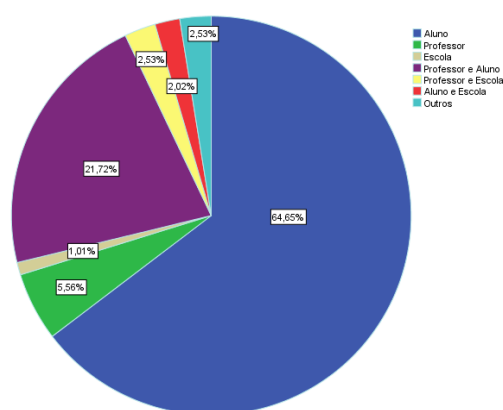
**Gráfico 4.16 – A quem recorrem para esclarecer dúvidas de Matemática**





Os inquiridos apontam o aluno como o principal responsável pela sua própria repetência (Gráfico 4.17), 64,65% apontam como responsável apenas o aluno, 21,72% apontam o aluno e o professor e ainda 2,02% o aluno e a escola e mesmo os 2,53% dos inquiridos que apontam como responsável outros acabam por referir como responsáveis o aluno e os pais; o aluno, o professor e a escola, e o aluno e a situação familiar.

**Gráfico 4.17 – Responsáveis pela repetência**



Na tentativa de conhecer quais os fatores que na opinião dos alunos causam insucesso foi-lhes apresentado doze fatores e procurou-se conhecer o seu grau de concordância com eles (escala de Likert - ordenados em valor de concordância). Assim, na opinião dos respondentes os fatores que mais contribuem para o insucesso na disciplina de Matemática são o desinteresse pela disciplina, falta de hábitos de trabalho e de estudo, falta de atenção/concentração nas aulas, dificuldades na interpretação e na compreensão e os fatores que merecem maior discordância são manual desadequado, falta de apoio por parte do professor, falta de apoio familiar, rapidez com que o professor dá a matéria e a falta de oportunidade para esclarecer dúvidas (Tabela 4.9) .

**Tabela 4.9 – Fatores que na opinião dos alunos contribuem para o insucesso em Matemática**

Fatores	Discordo totalmente (1) %	Discordo em algum grau (2) %	Não concordo nem discordo (3) %	Concordo em algum grau (4) %	Concordo totalmente (5) %	Ausentes %	Mediana
1. Desinteresse pela disciplina	7,0	5,5	18,0	33,0	36,0	0,5	4
2. Falta de hábitos de trabalho e de estudo	5,0	3,5	20,0	33,0	38,5	----	4
3. Falta de pré requisitos necessários	10,5	13,5	33,5	27	12,5	3,0	3
4. Falta de atenção/concentração nas aulas	3,5	7,5	17,5	31,5	39	1,0	4
5. Manual desadequado	33,5	19,0	30,5	7,5	7,5	2,0	2
6. Indisciplina na sala de aula	11,5	11,0	30,5	27,0	19,5	0,5	3
7. Carga horária global excessiva	13,0	6,5	32,0	19,0	28,5	1,0	3
8. Dificuldades de interpretação/compreensão	4,5	11,0	23,5	28,0	32,5	0,5	4
9. Falta de apoio por parte do professor	26,0	20,0	27,0	12,5	14,5	---	3
10. Falta de apoio familiar	32,0	16,0	25,0	13,0	14,0	---	3
11. Rapidez com que o professor dá a matéria	18,0	16,0	26,0	20,0	20,0	---	3
12. Falta de oportunidade para esclarecer dúvidas	24,5	11,5	32,5	19,5	12,0	---	3

Através da determinação da mediana de cada um dos fatores é possível verificar se estamos perante indiferença perante o fator ou se existe uma tendência para a concordância ou discordância. Apenas o fator “manual desadequado” apresenta tendência para a discordância (mediana 2), os fatores “desinteresse pela disciplina”, “falta de hábitos de trabalho e de estudo”, “falta de atenção/concentração nas aulas” e “dificuldades de interpretação/compreensão” apresentam tendência para a concordância (mediana 4), nos restantes fatores estamos perante indiferença (mediana 3).

A resposta concordo totalmente é a mais frequente para os fatores “desinteresse pela disciplina” (36% das respostas), “falta de hábitos de trabalho e de estudo” (38,5%), “falta de atenção/concentração nas aulas” (39% das respostas), “dificuldades de interpretação/compreensão” (32,5%). O discordo totalmente é a resposta mais frequente nos fatores “manual desadequado” (33,5% das respostas) e “falta de apoio familiar” (32%

das respostas). O não concordo nem discordo predomina nos fatores “falta de pré requisitos” (33,5% das respostas), “indisciplina na sala de aula” (30,5% das respostas), “carga horária global excessiva” (32% das respostas), “falta de apoio por parte do professor” (27% das respostas), “rapidez com que o professor dá a matéria” (36% das respostas) e “falta de oportunidade para esclarecer dúvidas” (32,5% das respostas).

Para analisar se existe associação entre estes doze fatores explicativos do insucesso, utilizou-se o coeficiente de correlação de Spearman ( $\rho$ ) verificando-se que este apresenta correlações estatisticamente significativas e positivas, embora fracas ou moderadas, entre quase todos os fatores. A associação entre a “falta de hábitos de trabalho e estudo” e a “falta de apoio por parte do professor” apresenta uma correlação negativa muito baixa, estatisticamente não significativa ( $\rho=-0.007$  com  $p>0,05$ ). O fator que apresentam menor correlação com os restantes é “carga horária global excessiva” e o que aparece correlacionado significativamente com maior número de fatores é “dificuldades de interpretação/compreensão” (com  $p<0,01$ ). Os maiores valores do coeficiente de correlação são entre os fatores “falta de apoio por parte do professor” e “rapidez com que o professor dá a matéria” ( $\rho=0,605$  com  $p<0,01$ ); “falta de apoio por parte do professor” e “falta de oportunidade para esclarecer dúvidas” ( $\rho=0,580$  com  $p<0,01$ ), “falta de atenção/concentração” e “falta de hábitos de trabalho e estudo” ( $\rho =0,557$  com  $p<0,01$ ) (Anexo 6).

Analisando as respostas dos alunos inquiridos quanto a possíveis medidas/estratégias para reduzir o insucesso (Tabela 4.10) pode concluir-se que das oito medidas apresentadas, a “utilização das tecnologias” e “diminuição do número de aulas” são aquelas que merecem maior consenso. Já o “aumento do número de aulas” e a “diminuição do número de alunos por turma” são as medidas com que os inquiridos menos concordam.

Através da mediana de cada uma das estratégias é possível verificar que “utilização de tecnologias” e “diminuição do número de aulas” apresentam tendência para a concordância (mediana=4), a estratégia “aumento do número de aulas” apresenta tendência para a discordância (mediana=2) e as restantes estratégias apresentam tendência para a indiferença (mediana=3).

Nas estratégias “utilização de tecnologia” e “diminuição do número de aulas” predomina a resposta concordo totalmente (moda 5), no “aumento do número de aulas” e

na “diminuição do número de alunos por turma” predomina o discordo totalmente (moda=1) nas restantes estratégias predomina o não concordo nem discordo (moda=3).

**Tabela 4.10 – Estatísticas das medidas/estratégias para reduzir o insucesso**

	Utilização de tecnologia	Diminuição do número de aulas	Aumento do número de aulas	Mudança das estratégias de ensino	Promoção de aulas de apoio	Diminuição do número de alunos por turma	Mudanças nos conteúdos	Maior articulação das matérias com aspetos da vida diária
N Válido	198	200	200	200	196	198	200	200
Ausente	2	0	0	0	4	2	0	0
Mediana	4,00	4,00	2,00	3,00	3,00	3,00	3,00	3,00
Moda	5	5	1	3	3	1	3	3

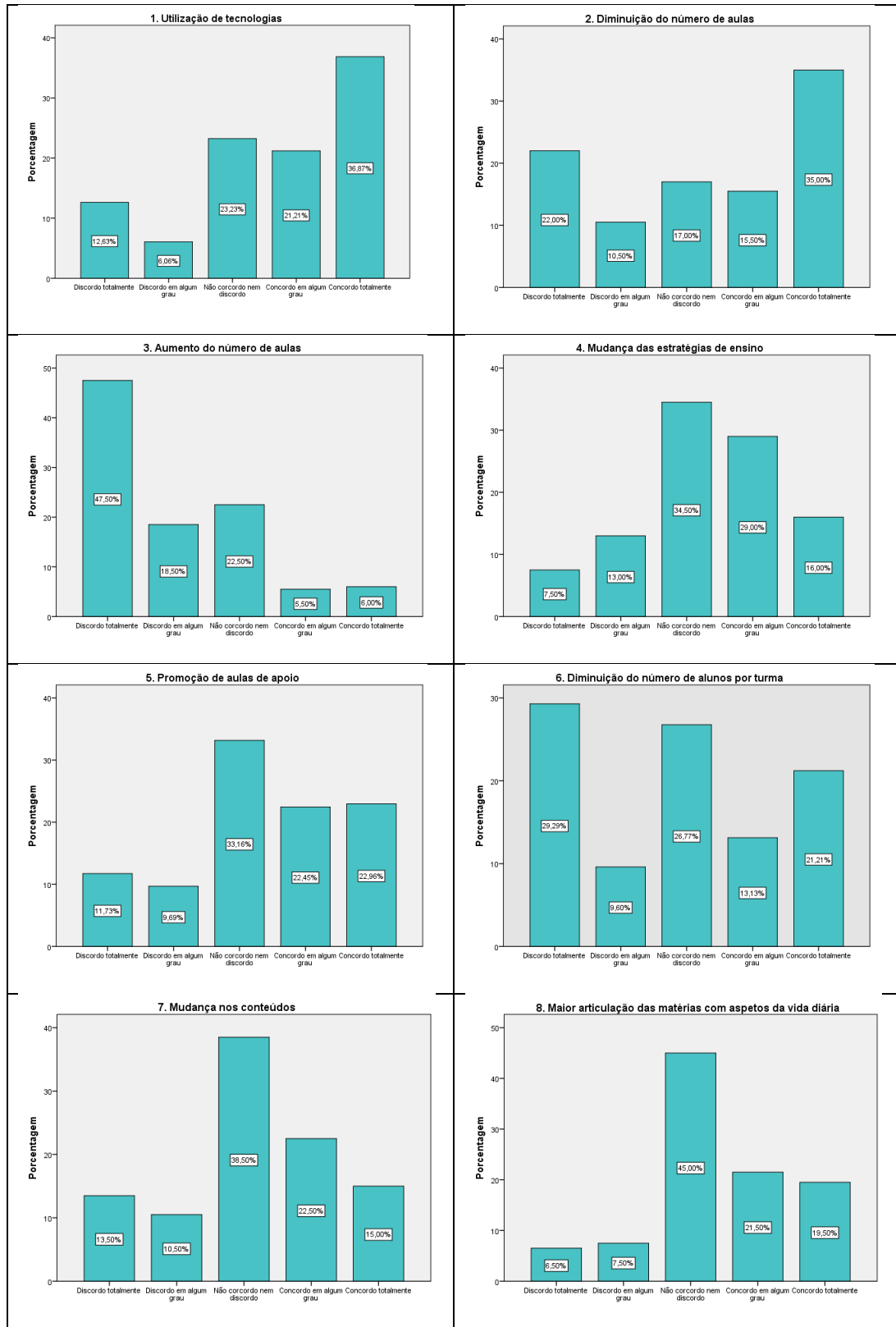
O recurso ao coeficiente de correlação de Spearman entre as diversas estratégias (Anexo 7) permite observar correlações significativas, associações positivas e negativas, entre algumas dessas estratégias, embora fracas ou moderadas.

O maior valor do coeficiente de correlação verifica-se entre a “utilização de tecnologias” e a “diminuição do número de aulas” ( $\rho=0,459$ ). Além desta associação existe correlação significativa,  $p<0,01$ , entre as seguintes estratégias: “utilização de tecnologias” e “mudança de estratégias de ensino” ( $\rho=0,198$ ); “utilização de tecnologias” e “mudança de conteúdos” ( $\rho=0,237$ ); “utilização de tecnologias” e “maior articulação das matérias com aspetos da vida diária” ( $\rho =0,315$ ); “diminuição do número de aulas” e “aumento do número de aulas” ( $\rho=-0,251$ ); “diminuição do número de aulas” e “mudança nos conteúdos” ( $\rho=0,292$ ); “diminuição do número de aulas” e “maior articulação das matérias com aspetos da vida diária” ( $\rho=0,192$ ); “aumento do número de aulas” e “promoção de aulas de apoio” ( $\rho =0,313$ ); “mudança das estratégias de ensino” e “promoção de aulas de apoio” ( $\rho =0,252$ ); “Mudança das estratégias de ensino” e “diminuição do número de alunos por turma” ( $\rho =0,268$ ), “mudança das estratégias de ensino” e “mudança nos conteúdos” ( $\rho=0,382$ ); “Mudança das estratégias de ensino” e “maior articulação das matérias com aspetos da vida diária” ( $\rho=0,269$ ); “diminuição do número de alunos por turma” e “mudança nos conteúdos” ( $\rho=0,375$ ); e “diminuição do número de alunos por turma” e “maior articulação das matérias com aspetos da vida diária” ( $\rho=0,254$ ); “Mudança nos conteúdos” e “maior articulação das matérias com aspetos da vida diária” ( $\rho =0,455$ ). Existe correlação significativa,  $p<0,05$ , entre as seguintes estratégias:

“utilização de tecnologias e “aumento do número de aulas” ( $\rho = -0,145$ ); “diminuição do número de aulas e “mudança das estratégias de ensino” ( $\rho = 0,142$ ); “diminuição do número de aulas” e “promoção de aulas de apoio” ( $\rho = -0,142$ ); “aumento do número de aulas” e “mudança das estratégias de ensino” ( $\rho = 0,161$ ).

Da observação do Gráfico 4.18 pode concluir-se que cerca de 36,87% dos inquiridos concordam totalmente que a “utilização de tecnologias” pode contribuir para reduzir o insucesso, 21,21% concorda em algum grau, 12,63% discorda totalmente, 6,06% discorda em algum grau e 23,23% não concorda nem discorda. Dos inquiridos 35% concorda totalmente com a medida “diminuição do número de aulas”, 15,5% concorda em algum grau, 22% discorda totalmente, 10,5% discorda em algum grau e 17% não concorda nem discorda. Da medida “aumento do número de aulas” como promotora de sucesso, apenas 6% dos inquiridos concordam totalmente e 5,5% concorda em algum grau. A grande maioria discorda, 47,5% discorda totalmente, 18,5% discorda em algum grau e 22,5% não concorda nem discorda. Quanto ao impacto da medida “mudança das estratégias de ensino” 16% dos inquiridos responderam que concordavam totalmente, 29% concorda em algum grau, 7,5% discorda totalmente e 13% discorda em algum grau, sendo a tendência da resposta (34,5%) não concordo nem discordo. A medida “promoção de aulas de apoio” mereceu a concordância total de 22,96% dos inquiridos e a concordância em algum grau de 22,45%. Dos inquiridos 11,73% discordam totalmente do efeito benéfico desta medida e 9,69% discorda em algum grau e 33,16% não concorda nem discorda. A medida “diminuição do número de alunos por turma” regista maior discordância do que concordância, merecendo a concordância total de 21,21% dos inquiridos, a concordância em algum grau de 13,13%, a discordância total de 29,29%, a discordância em algum grau de 9,6% e a não concordância nem discordância de 26,77%. A “mudança dos conteúdos da disciplina” recebeu a concordância de 15% dos alunos, a concordância em algum grau de 22,5%, a discordância total de 13,5%, a discordância em algum grau de 10,5%, predominando o não concordo nem discordo com 38,5%. Com a medida “maior articulação das matérias com aspetos da vida diária”, 19,5% concordam totalmente, 21,5% concordam em algum grau, 6,5% discordam totalmente, 7,5% discordam em algum grau, predominando claramente o não concordo nem discordo com 45%.

**Gráfico 4.18 - Medidas para reduzir o insucesso**



## 4.2. Aplicação da Análise de Regressão Logística

O problema a ser resolvido, procura construir um modelo para prever o resultado do sucesso/insucesso de um aluno na disciplina de matemática e analisar os fatores que influenciam esse resultado. A variável dependente do problema é “Reprovou alguma vez”, sendo o resultado “sim” ou “não”. O “sim” significa que o aluno já reprovou a matemática pelo menos uma vez desde o 5.º ano (nível inferior a 3 no 3.º período) e o “não” que nunca reprovou a matemática desde o 5.º ano. Em termos de variáveis independentes não foram consideradas no estudo de regressão as variáveis de opinião, nomeadamente as que dizem respeito a responsáveis pela repetência, fatores que contribuem para o insucesso e medidas para reduzir o insucesso.

Dado que a variável dependente é uma variável dicotômica, para construção do modelo será usada a seguinte codificação: 0 para “não”, que representa o fracasso e 1 para “sim”, que representa o sucesso. Neste caso, o sucesso é já ter reprovado, uma vez que é esta a probabilidade que se pretende modelar.

A partir de uma regressão logística procura-se encontrar um modelo de previsão que estime a probabilidade de um aluno obter sucesso ou insucesso a matemática.

Tendo presente que o melhor modelo é aquele que melhor explica a realidade, envolvendo as variáveis independentes mais interessantes do ponto de vista estatístico foram escolhidas para esta análise as variáveis que se revelaram estatisticamente interessantes nas associações com a variável dependente (com *p-value* relativamente baixo).

Dos duzentos elementos da amostra foram considerados cento e noventa e nove por ter ocorrido uma resposta não válida.

### 4.2.1. Aplicação da Análise de Regressão Logística Univariada

Dado o elevado número de variáveis foi realizada uma seleção das variáveis candidatas ao modelo de regressão logística múltipla através de uma análise de regressão logística univariada. Este procedimento foi realizado para averiguar a existência de associação de cada variável independente com a variável dependente.

O critério utilizado para verificar a significância da dependência entre cada variável independente com a variável dependente foi o nível habitual de significância ( $p \leq 0,05$ ).

No Anexo 8 são apresentados os resultados da análise de regressão logística univariada. Das 30 variáveis independentes candidatas a fazer parte do modelo de regressão logístico múltiplo, apenas 16 apresentam significância estatística ( $p \leq 0,05$ ), a saber: “idade”, “ter internet em casa”, “ambiente familiar”, “desejo de ir para a universidade”, “aluno participativo”, “aluno empenhado”, “aluno distraído”, “aluno trabalhador”, “dificuldades a matemática”, “periodicidade de estudo para matemática”, “horas de estudo semanal para matemática”, “comportamento da turma nas aulas de matemática”, “comportamento do aluno nas aulas de matemática”, “relação com a professora de matemática”, “realiza os trabalhos de casa de matemática” e o “interesse pela disciplina de matemática”.

Ao minimizar o número de variáveis no modelo espera-se que o modelo resultante seja numericamente mais estável e mais facilmente generalizado. Quanto mais variáveis forem incluídas no modelo maior serão os erros padrão e mais dependente o modelo se torna dos dados observados. Alguns investigadores sugerem a inclusão no modelo de todas as variáveis que sejam clinicamente e intuitivamente relevantes, independentemente do seu significado estatístico, no entanto, isso poderá levar a que o modelo se torne “*overfit*”, produzindo estimativas numericamente instáveis, com grandes coeficientes estimados e /ou erros padrão estimados irrealistas. Isto pode ser especialmente problemático nos problemas em que o número de variáveis no modelo é grande em relação ao número de indivíduos e / ou em que a percentagem global de respostas é perto de 0 ou 1.

#### **4.2.2. Aplicação da Análise Logística Múltipla**

As hipóteses a estudar no modelo representam o facto da variável dependente “reprovação a matemática alguma vez desde o 5.º ano” ser influenciada pelas variáveis independentes: “idade”, “ter internet em casa”, “ambiente familiar”, “desejo de ir para a universidade”, “aluno participativo”, “aluno empenhado”, “aluno distraído”, “aluno trabalhador”, “dificuldades a matemática”, “periodicidade de estudo para matemática”, “horas de estudo semanal para matemática”, “comportamento da turma nas aulas de matemática”, “comportamento do aluno nas aulas de matemática”, “relação com a

professora de matemática”, “realiza os trabalhos de casa de matemática” e o “interesse pela disciplina de matemática”.

Uma vez que em regressão logística não há um modelo único, neste sentido, e com o objetivo de procurar o modelo que mais se ajuste foram aplicados três métodos na seleção a essas dezasseis variáveis independentes: *Enter*, *Forward Stepwise Wald* e *Backward Stepwise Condicional* com recurso SPSS.

Da comparação dos *outputs* obtidos pelo três métodos de seleção de variáveis referidos anteriormente conclui-se que o método que apresentava melhores resultados, ou seja, que melhor caracteriza os dados é o *Enter*. Nesta comparação foi tida em linha de conta o que apresenta melhor percentagem de classificação de sucessos, melhor balanço entre erros do tipo I e II e melhor R quadrado.

#### 4.2.2.1. Análise Logística Múltipla recorrendo ao método *Enter*

Optou-se, pelas razões mencionadas anteriormente, pelo método *Enter*, no qual as dezasseis variáveis independentes são colocadas no modelo de regressão num único bloco e as estimativas dos parâmetros calculadas para cada bloco (Field:2009).

Na análise foram incluídos 187 casos e verificaram-se 12 casos ausentes. As variáveis categóricas foram categorizadas e escolhida a primeira categoria como referencia em todas as variáveis (Anexo 9).

A Tabela 4.11 apresenta o teste do rácio de verosimilhanças (Testes de coeficientes de modelo *Omnibus*) entre o modelo nulo e os modelos em cada uma das etapas, bloco e modelo final. Dado que se recorreu ao método *Enter* os três valores são coincidentes, sendo  $G^2(43) = 150,267$ , com  $p=0,001 < 0,05$ , o que indica que existe pelo menos uma variável independente no modelo com poder preditivo sobre a variável dependente.

**Tabela 4.11 - Testes de coeficientes de modelo Omnibus modelo 1**

		Qui-quadrado	df	Sig.
Etapa 1	Etapa	150,267	43	,000
	Bloco	150,267	43	,000
	Modelo	150,267	43	,000

Na tabela 4.12, pode observar-se o valor *Nagelkerke* R quadrado que indica que o modelo explica 73,6% da variância nos dois grupos da variável dependente Repetencia2: 0 (não), 1(sim). A verossimilhança de log-2 (-2LL) apresenta o valor de 108,965, que pode

ser usado para avaliar a qualidade do ajustamento do modelo. O p value correspondente ao  $-2LL$ , dado por  $\chi^2_C (187 - 16 - 1) = 108,965$ , é 0,9999 pelo que não se rejeita a hipótese nula de que o modelo se ajusta aos dados. Na tabela encontra-se ainda o R quadrado Cox & Snell ( $R^2_{CS} = 0,552$ ).

O pseudo- $R^2$  de McFadden é:  $R^2_{MF} = 1 - 108,965 / 259,232 = 0,580$ , [dado que,  $-2Ln(L_0) = 150,267 + 108,965 = 259,232$ ].

Os valores destas estatísticas revelam um modelo com qualidade adequada.

**Tabela 4.12- Resumo do modelo 1**

Etapa	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke
1	108,965 <sup>a</sup>	,552	,736

a. Estimação finalizada no número de iteração 30 porque o máximo de iterações foi atingido. Não é possível encontrar a solução final.

O teste Hosmer e Lemeshow (Tabela 4.13) apresenta  $\chi^2_{HL} (8) = 1,151$  com  $p=0,997 > 0,05$ , o que indica que não se verifica significância estatística, ou seja, existe um bom ajustamento entre os dados e o modelo.

**Tabela 4.13 -Teste de Hosmer e Lemeshow modelo 1**

Etapa	Qui-quadrado	df	Sig.
1	1,151	8	,997

A Tabela 4.14 apresenta a classificação dos sujeitos observada e prevista pelo modelo ajustado, verificando-se que existem 9 sujeitos que não repetiram alguma vez mas que o modelo prediz que repetiram e 13 que repetiram alguma vez e o modelo prediz que não repetiram. A sensibilidade do modelo é de 86%, ou seja, classifica corretamente 86% dos sujeitos que reprovaram-sucessos e a especificidade do modelo é de 90,4%, ou seja, o modelo classifica corretamente 90,4% dos sujeitos que não repetiram alguma vez-insucessos. A percentagem de casos corretamente classificados é de 88,2% (um acréscimo de aproximadamente 38% relativamente ao modelo nulo) valor superior em mais de 38% à percentagem proporcional de classificações corretas [ $((94/187)^2 + (93/187)^2) \times 100\% = 50\%$ ].

**Tabela 4.14-Tabela de Classificação<sup>a</sup> modelo 1**

		Previsto		
		Repetência a Matemática Alguma vez		Porcentagem correta
Observado		Não	Sim	
Etapa 1	Repetência a Matemática Não	85	9	90,4
	Alguma vez Sim	13	80	86,0
Porcentagem global				88,2

a. O valor de recorte é ,500

As variáveis que fazem parte da equação, pela extensão da tabela, encontram-se em anexo (Anexo 9) e foram todas introduzidas na etapa 1 devido ao tipo de método utilizado.

As variáveis explicativas que evidenciam significância estatística ( $p \leq 0,05$ ) e que para além da constante parecem explicar o modelo, verificadas pelo teste *Wald*, foram as seguintes: Idade ( $\chi^2_{Wald}(1) = 6,398, p = 0,011$ ); Participativo ( $\chi^2_{Wald}(2) = 11,105, p = 0,004$ ), para os casos específicos Participativo(1) ( $\chi^2_{Wald}(1) = 7,748, p = 0,005$ ), Participativo(2) ( $\chi^2_{Wald}(1) = 10,918, p = 0,001$ ); Empenhado ( $\chi^2_{Wald}(2) = 8,555, p = 0,014$ ) especificamente para Empenhado(1) ( $\chi^2_{Wald}(1) = 4,659, p = 0,031$ ); DificMat ( $\chi^2_{Wald}(2) = 7,599, p = 0,022$ ), especificamente DificMat(2) ( $\chi^2_{Wald}(1) = 7,543, p = 0,006$ ); CompAluno no global apresenta ( $\chi^2_{Wald}(4) = 4,701, p = 0,393$ ) valor não significativo, no entanto, CompAluno(2) ( $\chi^2_{Wald}(1) = 3,970, p = 0,046$ ) e CompAluno(3) ( $\chi^2_{Wald}(1) = 4,636, p = 0,031$ ) portanto, significativas; FazTPC no global é significativa, apresenta ( $\chi^2_{Wald}(3) = 9,171, p = 0,027$ ), mas FazTPC(1) apresenta ( $\chi^2_{Wald}(1) = 0,316, p = 0,574$ ) e FazTPC(2) apresenta ( $\chi^2_{Wald}(1) = 0,069, p = 0,793$ ).

No caso das variáveis qualitativas com mais de duas classes, o SPSS Statistics produz um teste à significância do modelo, como é o caso das variáveis Participativo, Empenhado, DificMat e CompAluno. Os coeficientes do modelo relativos às variáveis qualitativas indicam as diferenças entre as médias das classes presentes no modelo relativamente à classe de referência.

O modelo obtido é:

$$\text{Logit}(\hat{\pi}) = 14,380 + 0,903 \text{ Idade} - 4,112 \text{ Participativo}(1) - 5,481 \text{ Participativo}(2) + 6,55 \text{ Empenhado}(1) + 3,806 \text{ DificMat}(2) + 4,462 \text{ CompAluno}(2) + 5,140 \text{ CompAluno}(3)$$

Onde Participativo(1) refere-se a “às vezes”, Participativo(2) a “sim”, Empenhado(1) a “às vezes”, DificMat(2) a “sim bastantes”, CompAluno(2) a “pouco satisfatório” e CompAluno(3) a “satisfatório”.

O modelo final que permite estimar a probabilidade de reprovar a matemática alguma vez é então:

$$\hat{\pi} = \frac{1}{1+e^{-(\text{Logit}(\hat{\pi}))}}$$

Pela Tabela 4.15, verifica-se que existem 16 observações que apresentam valores elevados de resíduos estandardizados (*ZResid*) o que indica que estes 16 casos foram mal classificados pelo modelo. No entanto, dois deles (189 e 192) figuram na lista mas apresentam “N” quer no “observado” quer no “Grupo previsto” e os *ZResid* significativamente inferiores aos restantes do grupo.

A qualidade do modelo pode ser melhorada reavaliando a pertinência dos casos mal classificados no modelo.

**Tabela 4.15 – Lista entre casos<sup>b</sup>**

Caso	Status selecionado <sup>a</sup>	Observado	Previsto	Grupo previsto	Variável temporária	
		Repetência a Matemática Alguma vez			Resid	ZResid
25	S	N**	,771	S	-,771	-1,834
56	S	S**	,098	N	,902	3,038
59	S	N**	,765	S	-,765	-1,806
64	S	S**	,347	N	,653	1,373
109	S	N**	,882	S	-,882	-2,740
126	S	S**	,182	N	,818	2,118
132	S	S**	,054	N	,946	4,168
137	S	S**	,250	N	,750	1,731
139	S	S**	,226	N	,774	1,852
155	S	S**	,182	N	,818	2,122
178	S	N**	,644	S	-,644	-1,344
183	S	S**	,061	N	,939	3,919
184	S	N**	,581	S	-,581	-1,177
189	S	N	,389	N	-,389	-,798
191	S	N**	,914	S	-,914	-3,252
192	S	N	,351	N	-,351	-,735

a. S = Selecionado, U = Casos não selecionados, e \*\* = Casos classificados incorretamente.

b. São listados os casos com resíduos estudentizados maiores que 2,000.

#### 4.2.2.2. Reajustamento do modelo

Foram retirados os casos mal classificados da base de dados, exceto os casos 189 e 192 (pelos motivos anteriormente apresentados), bem como as variáveis que apresentam associações não significativas no modelo anterior. Utilizando-se apenas as variáveis “Idade”, “Participativo”, “DificMat”, “CompAluno” e “Empenhado” procedeu-se ao reajustamento do modelo. Como os valores do teste de *Wald* para cada um dos coeficientes do modelo são condicionados pelos valores dos outros coeficientes, novas combinações de

variáveis independentes podem apresentar diferentes significâncias. Segundo Marôco (2011) é aconselhável usar um procedimento *Stepwise* que a cada passo reavalie a significância das variáveis, mas recorreu-se ao método *Enter*, por comparativamente este ter voltado a apresentar melhores resultados.

A análise contou com 185 casos verificando-se 184 casos incluídos.

O valor *Nagelkerke* R quadrado é  $R_N^2 = 0,566$ . Este valor indica que o modelo explica 56,6% da variância nos dois grupos da variável dependente Repetencia2: 0 (não), 1(sim). A verossimilhança de log-2 (-2LL) apresenta o valor de 153,275. O R quadrado Cox & Snell ( $R_{CS}^2 = 0,425$ ) (Tabela 4.16).

**Tabela 4.16 – Resumo do modelo 2**

Etapa	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke
1	153,275 <sup>a</sup>	,425	,566

a. Estimação finalizada no número de iteração 6 porque as estimativas de parâmetro mudaram foram alteradas para menos de ,001.

Há 17 alunos que nunca reprovaram a matemática, mas o modelo prediz que reprovaram (falso-positivo) e 19 que reprovaram a matemática e o modelo prediz que não reprovaram (falso-negativos). Este modelo classifica corretamente 78,9% dos alunos que reprovaram a matemática (sucessos), ou seja, a sensibilidade do modelo é de 78,9%. E classifica corretamente 81,9% dos alunos que nunca reprovaram (insucessos), isto é, a especificidade do modelo é 81,9%. A percentagem de casos corretamente classificados é de 80,4% (Tabela 4.17).

**Tabela 4.17 – Tabela de Classificação<sup>a</sup> modelo 2**

	Observado	Previsto		
		Repetência a Matemática Alguma vez		Porcentagem correta
		Não	Sim	
Etapa 1	Repetência a Matemática Não	77	17	81,9
	Alguma vez Sim	19	71	78,9
	Porcentagem global			80,4

a. O valor de recorte é ,500

As variáveis preditoras significativas, para  $p \leq 0,05$  e que para além da constante parecem explicar o modelo, verificadas pelo teste *Wald*, foram as seguintes: Idade, Participativo(1), Participativo(2), DificMat(1) e DificMat(2) (Anexo 10).

O modelo é:

$$\text{Logit}(\hat{\pi}) = -6,939 + 0,634 \text{ Idade} + 1,518 \text{ DificMat}(1) + 4,213 \text{ DificMat}(2) + \\ -2,232 \text{ Participativo}(1) - 2,64 \text{ Participativo}(2)$$

O modelo final que permite estimar a probabilidade de reprovar a matemática alguma vez é então:

$$\hat{\pi} = \frac{1}{1 + e^{-(6,939 + 0,634 \text{ Idade} + 1,518 \text{ DificMat}(1) + 4,213 \text{ DificMat}(2) - 2,232 \text{ Participativo}(1) - 2,642 \text{ Participativo}(2))}}$$

#### 4.2.2.3. Comparação dos dois modelos

Apesar de tudo o primeiro modelo apresentado é melhor do que o segundo, dado que apresenta boas capacidades preditivas, uma vez que quer a sensibilidade quer a especificidade é superior a 80%, o que já não acontece no segundo. Apresenta maior de valor Nagelkerke R e a área da curva ROC (*Receiver Operating Characteristic*) que estuda a variação da sensibilidade e especificidade, para diferentes valores de corte, apresenta o valor de 0,945 ao passo que a área no segundo modelo é de 0,885.

Assim, os resultados obtidos Análise de Regressão Logística Múltipla sugerem que a probabilidade de a repetência a Matemática aumenta exponencialmente com a idade do aluno, as dificuldades sentidas, a não participação nas aulas, a falta de empenho e com o comportamento menos positivo do aluno.

### 4.3. Aplicação da Análise de *Clusters*

Realizou-se uma aplicação da análise de *clusters* a algumas das variáveis do inquérito com vista a encontrar possíveis agrupamentos.

Aplicou-se o método de agregação hierárquico às variáveis consideradas significativas pela Análise de Regressão Logística Univariada, bem como a variável “Repetência” (indica o número de reprovações do aluno a matemática desde o 5.º ano). Não foram escolhidas todas as variáveis, por serem demasiadas poderiam dificultar a formação e interpretação dos *clusters*. Aplicou-se o método duas etapas às vinte variáveis que na opinião dos alunos constituem fatores de insucesso e possíveis medidas para minorar esse insucesso.

#### 4.3.1. Aplicação do método de agregação hierárquico

O número de variáveis considerado foi 17, a saber: idade; dificuldades; reprovações a matemática; ter internet em casa; ambiente familiar; desejo de ir para universidade;

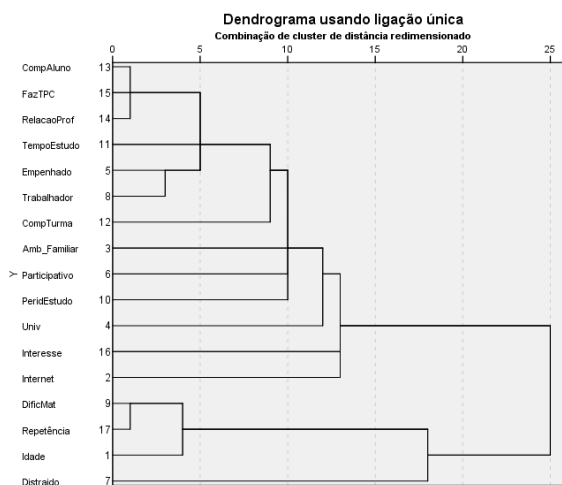
empenhado; participativo; trabalhador; periodicidade de estudo; horas de estudo; comportamento da turma; comportamento do aluno; relação com a professora; realiza os trabalhos de casa, interesse pelos conteúdos, e distraído. O número de casos válidos foi 187 e ausentes 12.

Recorreu-se ao método hierárquico, a medida de proximidade foi o coeficiente de correlação de Pearson e o método do vizinho mais próximo (por tender a maximizar a conectividade entre *clusters*). As variáveis por estarem em diferentes escalas de resposta foram padronizadas através de Pontuações Z.

Segundo o Dendograma (Gráfico 4.19) as variáveis comportamento do aluno, faz os trabalhos de casa de matemática e relação com a professora de matemática estão muito próximas (fortemente correlacionadas), bem como as variáveis dificuldades na disciplina de matemática e repetência em matemática. Por sua vez as variáveis dificuldades a matemática”, Repetência e idade, juntam-se tardiamente à variável “distraído” e são muito dissemelhantes de todas as outras, só se juntando às restantes no final.

Considerando o corte em 10 o Dendograma sugere que existem 3 ou 4 *clusters*.

**Gráfico 4.19 - Dendograma**



Colocando a hipótese de 3 *clusters*, o *cluster* 1 é constituído pelas variáveis idade, dificuldades a matemática e reprovações a matemática. O *cluster* 2 pelas ter internet em casa, ambiente familiar, desejo de ir para universidade, empenhado, participativo, trabalhador, periodicidade de estudo, horas de estudo, comportamento da turma, comportamento do aluno, relação com a professora, realiza os trabalhos de casa e interesse. O *cluster* 3 pela variável distraído.

Colocando a hipótese de 4 *clusters*, o *cluster* 1 seria constituído pelas variáveis idade, dificuldades a matemática e reprovações a matemática. O *cluster* 2 pela variável ter internet. O *cluster* 3 pelas variáveis ambiente familiar, desejo de ir para universidade, empenhado, participativo, trabalhador, periodicidade de estudo, horas de estudo, comportamento da turma, comportamento do aluno, relação com a professora, interesse e realiza os trabalhos de casa. O *cluster* 4 pela variável distraído. Verificando-se assim, dois *clusters* com uma única variável.

Pela leitura dos dendograma e do síncelos aparece-nos mais adequada a opção de 3 *clusters*, a diferença está basicamente na variável 4 que na opção de 4 *clusters* aparece sozinha num *cluster*, dado só se juntar às do *cluster* 3 após o corte em 10, mas a sua agregação está relativamente próxima da agregação da variável universidade que faz parte do *cluster* 3.

Assim, podemos considerar três grupos de variáveis para classificar os alunos. Sendo as variáveis mais fortemente relacionadas: comportamento do aluno, faz os trabalhos de casa e relação com a professora, e as variáveis dificuldades a matemática e repetência. Para avaliar a solução de *clusters* escolhida e o número de *clusters* a reter poderia se recorrer, por exemplo, ao software SAS (*PROD CLUSTER*) que disponibiliza vários índices.

#### **4.3.2. Aplicação do método duas etapas**

Uma vez que este estudo também tem por objetivo avaliar os fatores que na opinião dos alunos contribuem para o insucesso na matemática, bem como as medidas, que segundo os alunos, poderão contribuir para reduzir esse mesmo insucesso decidiu-se aplicar a análise de *clusters* às questões de opinião e que não fazem parte do grupo de variáveis às quais se aplicou o método hierárquico.

Assim, aplicou-se o método duas etapas, por ser o mais adequado, a 20 variáveis categóricas: “Desinteresse”, “FaltaHabTrab”, “FaltaReq”, “FaltaAtenção”, “Manual” “Indisciplina”, “CargaHorária”, “DificInterp”, “FApoioProf”, “FApoioFAMIL”, “Rapidez, FEclarecimento”, “Tecnologias”, “MenosAulas”, “MaisAulas”, “Estratégias”, “AulasApoio”, “TurmaPequena”, “Conteúdos”, “VidaDiária”.

O critério de informação utilizado foi o AIC.

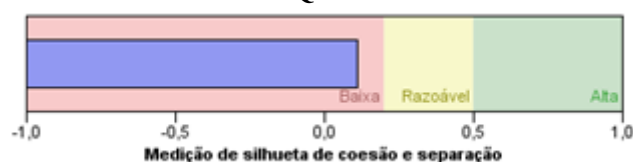
Verificaram-se 178 casos combinados e 21 casos não enquadrados em nenhum *cluster*. Este tipo de análise constrói três *clusters*: *cluster 2* com 51,1%, *cluster 1* com 39,9% e *cluster 3* com 9%, sendo a proporção de tamanhos entre o maior e o menor de 5,69 (Tabela 4.18).

**Tabela 4.18 – Distribuição dos *clusters***

<i>Clusters</i> formados	Tamanho do <i>clusters</i>	% do total
2	91	51,1%
1	71	39,9%
3	16	9%
Casos combinados	178	
Casos excluídos	21	

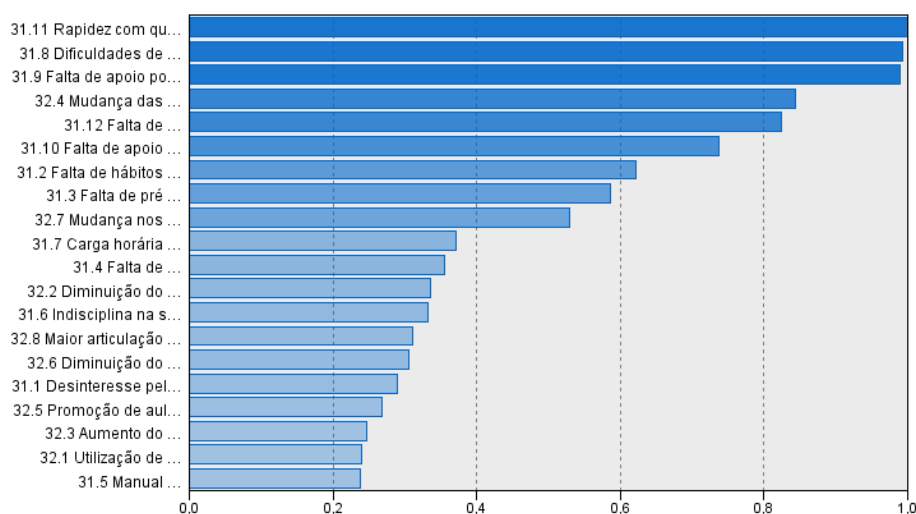
Em termos de qualidade do modelo (Gráfico 4.20.) é baixa (silhueta média=0,1).

**Gráfico 4.20 – Qualidade de *cluster***



O Gráfico 4.21 apresenta as variáveis utilizadas no processo de identificação dos *clusters*, ordenadas por importância na clusterização. No entanto, cada *cluster* apresenta importância diferente para cada variável com variação entre 0 e 1. A variável “Rapidez com que o professor dá a matéria” foi a mais importante no processo de clusterização, seguida das variáveis “dificuldades de interpretação e de compreensão” e “falta de apoio por parte do professor”.

**Gráfico 4.21 – Importância do preditor**



Na Tabela 4.21 pode observar-se o comportamento das variáveis dentro dos *clusters*, isto é, como cada variável se comporta dentro dos *clusters* e no Anexo 11 pode observar-se com maior pormenor a importância da variável e qual a sua categoria mais frequente, o que permite traçar o perfil de cada *cluster*.

No *cluster 1* predomina a resposta concordo totalmente (12), seguida de concordo em algum grau (5), não concordo nem discordo (2) e discordo totalmente (1). No *cluster 2* predomina como resposta não concorda nem discorda (12), seguida de concordo em algum grau (3) e discordo totalmente (1). No *cluster 3* predomina a resposta discordo totalmente (14), seguida de concordo totalmente (5) e não concordo nem discordo (1).

O *cluster 1* foi formado pelos alunos que concordam totalmente que as causas do insucesso sejam a rapidez com que o professor dá a matéria, falta de apoio do professor, falta de oportunidade para esclarecer dúvidas, dificuldades de interpretação, falta de apoio familiar, carga horária global excessiva, falta de atenção/concentração nas aulas, desinteresse pela disciplina, falta de hábitos de trabalho e de estudo. Concordam totalmente com as estratégias diminuição do número de aulas, diminuição do número de alunos por turma, utilização de tecnologias, promoção de aulas de apoio. Concordam em algum grau com as causas falta de pré requisitos necessários, indisciplina na sala de aula. Concordam em algum grau com as estratégias: mudanças nas estratégias de ensino e mudança nos conteúdos. Não concordam nem discordam com a causa: manual desadequado, nem com a estratégia maior articulação das matérias com aspectos da vida diária. Discordam totalmente com a estratégia de aumento do número de aulas.

O *cluster 2* foi formado pelos alunos que não concordam nem discordam que as causas do insucesso sejam falta de apoio familiar, rapidez com que o professor dá a matéria, falta de apoio do professor, dificuldades de interpretação, falta de oportunidade para esclarecer dúvidas, falta de pré requisitos necessários, carga horária global excessiva, indisciplina na sala de aula, manual desadequado. Nem concordam nem discordam com as estratégias: mudanças nas estratégias de ensino, mudança nos conteúdos, diminuição do número de aulas, promoção de aulas de apoio, utilização de tecnologias, maior articulação das matérias com aspectos da vida diária, diminuição do número de alunos por turma. Concordam em algum grau que as causas sejam falta de hábitos de trabalho e de estudo, desinteresse pela disciplina, falta de atenção/concentração nas aulas. Discordam totalmente com a estratégia de aumento do número de aulas.

O *cluster* 3 é formado por aqueles que discordam totalmente que as causas do insucesso sejam: dificuldades de interpretação, falta de apoio do professor, rapidez com que o professor dá a matéria, falta de oportunidade para esclarecer dúvidas, falta de pré requisitos necessários, falta de apoio familiar, indisciplina na sala de aula, carga horária global excessiva. Discordando totalmente também das estratégias: mudanças nas estratégias de ensino, mudança nos conteúdos, aumento do número de aulas, diminuição do número de alunos por turma, manual desadequado e diminuição do número de aulas. Concordam totalmente com as causas de insucesso: falta de hábitos e métodos de trabalho, falta de atenção e concentração nas aulas, desinteresse pela disciplina. Concordam totalmente com as estratégias: promoção de aulas de apoio, utilização de tecnologias. Não concordam nem discordam com a estratégia: maior articulação das matérias com aspectos da vida diária.

**Tabela 4.21 – Agrupamentos**

**Agrupamentos**

Importância da Entrada (preditor)

■ 1,0 ■ 0,8 ■ 0,6 ■ 0,4 ■ 0,2 ■ 0,0

Cluster	2	1	3
Tamanho	51,1%	39,9%	9,0%
Entradas	FApoioFAMIL	Rapidez	DificInterp
	Rapidez	FApoioProf	FaltaHabTrab
	Estratégias	Fesclarecimento	FApoioProf
	FApoioProf	DificInterp	Rapidez
	DificInterp	FApoioFAMIL	Estratégias
	Fesclarecimento	Estratégias	Fesclarecimento
	FaltaReq	CargaHorária	FaltaReq
	Conteúdos	Conteúdos	Conteúdos
	MenosAulas	FaltaReq	FApoioFAMIL
	CargaHorária	MenosAulas	FaltaAtenção
	AulasApoio	TurmaPequena	Indisciplina
	Tecnologias	FaltaAtenção	MaisAulas
	FaltaHabTrab	Desinteresse	TurmaPequena
	VidaDiária	VidaDiária	VidaDiária
	Desinteresse	Indisciplina	Manual
	Indisciplina	FaltaHabTrab	Desinteresse
	MaisAulas	Manual	CargaHorária
	FaltaAtenção	Tecnologias	AulasApoio
	Manual	AulasApoio	Tecnologias
	TurmaPequena	MaisAulas	MenosAulas

## CONCLUSÃO

Em Portugal, o insucesso escolar, e em particular o insucesso a Matemática no 3.º ciclo do ensino básico, é uma realidade preocupante. Não apenas pelos números (classificações de final de período e de exame, resultados de estudos), mas sobretudo pelas dificuldades reveladas pelos jovens nas aulas, na ausência de raciocínio matemático, na dificuldade em aplicar conhecimentos básicos de matemática no seu dia a dia, na desmotivação pela disciplina, na atitude passiva perante o insucesso. Como professora defronto-me diariamente com esta realidade, daí ter escolhido abordar o tema insucesso, na sua face mais visível, a repetência.

O estudo decorreu no ano letivo 2014/15 e baseou-se na aplicação de um questionário a uma amostra aleatória, de 200 alunos do 3.º ciclo de uma escola e teve como objetivo central traçar o perfil do aluno com repetência, encontrar factores explicativos do insucesso e medidas para promover o sucesso.

Dos inquiridos 51% responderam não terem tido nenhuma repetência matemática 14,5% reprovou uma vez, 12,5% reprovou 2 vezes, 10% reprovou 3 vezes e 11,5% mais de 3 vezes. Apontando-se o a si próprios como o principal responsável pela sua repetência. Apenas 17% referiu não sentir dificuldades na disciplina.

Neste estudo colocou-se a hipótese da variável dependente reprovação a matemática alguma vez desde o 5.º ano (nível inferior a 3 no final do ano letivo) ser influenciada por dezasseis variáveis independentes. Através da Análise de Regressão Logística Múltipla foi possível modelar a ocorrência, em termos probabilísticos, de uma de duas realizações da variável dependente (sim e não) e avaliar a significância de cada uma das variáveis independentes no modelo. Assim, os resultados obtidos através da Análise de Regressão Logística Múltipla sugerem que a probabilidade de repetência a Matemática aumenta exponencialmente com a idade do aluno, as dificuldades sentidas, a não participação nas aulas, o empenho e com o comportamento menos positivo do aluno. O modelo encontrado classifica corretamente 86% dos sujeitos que já reprovaram e corretamente 90,4% dos sujeitos que nunca reprovaram.

Realizou-se uma aplicação da análise de *clusters* a algumas das variáveis do inquérito com vista a encontrar possíveis agrupamentos.

Aplicou-se o método de agregação hierárquico às variáveis consideradas significativas pela Análise de Regressão Logística Univariada, bem como a variável “Repetência” (indica o número de reprovações do aluno a matemática desde o 5.º ano). E verificou-se que as variáveis “comportamento do aluno”, “faz os trabalhos de casa de matemática” e “relação com a professora de matemática” estão muito próximas (fortemente relacionadas), bem como as variáveis “dificuldades na disciplina de matemática” e “repetência em matemática”. Por sua vez as variáveis “dificuldades a matemática”, “repetência” e “idade”, juntam-se tardiamente à variável “distraído” e são muito dissemelhantes de todas as outras, só se juntando às restantes no final.

Foi também aplicado o método duas etapas às vinte variáveis que na opinião dos alunos podem constituir fatores de insucesso e possíveis medidas para minorar esse insucesso. Obtendo-se três *clusters*, mas um deles com tamanho superior a 50% e dando a entender que os alunos não têm uma ideia muito clara sobre os fatores que causam insucesso nem sobre as medidas que poderão aumentar o sucesso.



## BIBLIOGRAFIA

Anderberg, Michael R. (1973). *Cluster Analysis for Applications*. New York and London: Academic Press.

Avanzini, G. (s.d.). *O Insucesso Escolar*. Coleção para viver melhor. Lisboa: Editorial Pórtico.

Barros, M.G. (1988). “*Insucesso em Matemática Irreversível*”, in *Medidas que Promovam o Sucesso Educativo*: textos das comunicações e conclusões do Seminário realizado em Braga, Lisboa, C.R.S.E, GEP/ME, pp. 169-189.

Barth, Nelson Lerner (2004). *Inadimplência: Construção de modelos de previsão*. São Paulo: Nobel.

Beaud, Jean Pierre (2003). *A amostragem*. In Gauthier, Benôit – *Investigação Social: da problemática à colheita de dados*. 3.<sup>a</sup> Ed. Loures: Lusociência, pp 201.

Benavente, Ana (1976). *A Escola na sociedade de classes – o insucesso escolar*. Lisboa: Livros Horizonte.

Benavente, Ana (1999). *Escola, professores e processos de mudança*. Lisboa: Livros Horizonte.

Benavente, Ana; Correia, Adelaide (1980). *Obstáculos ao sucesso na escola primária*. Lisboa: Instituto de Estudos para o Desenvolvimento.

Bryman, Alan; Cramer, Duncan (2003). *Análise de Dados em Ciências Sociais: Introdução às Técnicas Utilizando o SPSS para Windows*. 3.<sup>a</sup> Ed. Oeiras: Celta Editora.

Cabrita, I. (1993). *A problemática do insucesso educativo em matemática no 3.º ciclo do ensino básico*. In Martins, António; Cabrita, Isabel. *A Problemática do Insucesso Escolar*. (pp. 30-66) Aveiro: Universidade de Aveiro

Cortesão, L.; Torres, M. A. (1990). *Avaliação Pedagógica I – Insucesso Escolar*. Porto: Porto Editora.

Costa, Giovanni (2014). *Uso da Técnica Two Step Cluster para Segmentação de Aparelhos de Ar Condicionado Split Segundo Desempenho, Segurança e Ruído: Um Estudo de Caso*. Centro Universitário Tupy – UNISOCIESC. V. 04, n. 01: pp. 203-227. Disponível em: [https://www.researchgate.net/publication/274238212\\_Uso\\_da\\_Tecnica\\_Two\\_Step\\_Cluster\\_para\\_Segmentacao\\_de\\_Aparelhos\\_de\\_Ar\\_Condicionado\\_Split\\_Segundo\\_Desempenho\\_Seguranca\\_e\\_Ruido\\_Um\\_Estudo\\_de\\_Caso](https://www.researchgate.net/publication/274238212_Uso_da_Tecnica_Two_Step_Cluster_para_Segmentacao_de_Aparelhos_de_Ar_Condicionado_Split_Segundo_Desempenho_Seguranca_e_Ruido_Um_Estudo_de_Caso) [20 de fevereiro de 2016].

Coutinho, Clara P. (2014). *Metodologia de Investigação em Ciências Sociais e Humanas: Teoria e Prática*. Coimbra: Edições Almedina, S.A.

- Cesca, Cleuza (2006). *Relações Públicas e suas interfaces*. São Paulo: Summus.
- Chiu, Tom; Fang, DongPing; Chen, John; Wang, Yao; Jeris, Christopher (2001). *A robust and scalable clustering algorithm for mixed type attributes in large database environment*. São Francisco, California : Anais de ACM KDD 01, pp. 263-268.
- Duarte, Maria (2000). *Alunos e insucesso escolar*. Lisboa: Instituto de Inovação Educacional.
- Educare (2004). *As razões do insucesso a Matemática*. Disponível em: <http://www.educare.pt/noticias/noticia/ver/?id=18300&langid=1> [1 março 2014].
- Enciclopédia Luso Brasileira de Cultura, 20.º Ed. (1980). Lisboa: Editorial o Livro.
- European Commission/EACEA/Eurydice/Cedefop (2014). *Tackling Early Leaving from Education and Training in Europe: Strategies, Policies and Measures*. Eurydice and Cedefop Report. Luxembourg: Publications Office of the European Union. Pp 193-195. Disponível em: [http://eacea.ec.europa.eu/education/eurydice/documents/thematic\\_reports/175en.pdf](http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/175en.pdf) [15 de dezembro de 2015].
- Eurydice (1995). *A luta contra o insucesso escolar: um desafio para a construção europeia*. Lisboa: Departamento de Programação e Gestão Financeira do Ministério da Educação e do Programa Educação para todos.
- Fernandes, António S. (1998). *O Insucesso Escolar*, in *A Construção Social da Educação Escolar*, Col. Biblioteca Básica de Educação e Ensino. Rio Tinto: Edições ASA/ Clube do Professor, 2.ª Ed. pp.187-213.
- Ferrão M. E. (2015). *Tópicos sobre retenção escolar em Portugal através do PISA: qualidade e equidade*. Arquivos Analíticos de Políticas Educativas, 23 (114). Disponível em <http://epaa.asu.edu/ojs/article/viewFile/2091/1693> [11 de fevereiro de 2016]
- Field, A. *Descobrendo a estatística usando SPSS* (2009), 2.ª ed. Porto Alegre: Artmed.
- Formosinho, João. (1998). *A Igualdade em Educação*. In *A Construção Social da Educação Escolar*, Col. Biblioteca Básica de Educação e Ensino. Rio Tinto: Edições ASA/ Clube do Professor, 2.ª Ed. pp. 169-186.
- Fortin, Marie Fabienne (2003) – *O processo de investigação: da concepção à realização*. 3.ª Ed. Loures : Lusociência.
- Freixo, Manuel J. (2009). *Metodologia científica: fundamentos, métodos e técnicas*. Lisboa: Instituto Piaget.
- Gil, António C. (1995). *Métodos e técnicas de pesquisa social*. 4.ª Ed. São Paulo: Atlas

- Heiman, Gary (1996). *Basic statistics for the behavioral sciences*. New York: Houghton Mifflin Harcourt Publishing Company.
- Hair, J.; Black, B. Babin, B.; Anderson, R. e Tatham, R. (2010). *Multivariate Data Analysis* (7th edition). Upper Saddle River, NJ: Prentice-Hall.
- Hoswer, D.; Lemeshow, S. (2009). *Applied Logistic Regression*, 2.<sup>a</sup> Ed. Wiley – Interscience.
- Hill, M. e Hill, A. (2012). *Investigação por Questionário*. 2.<sup>a</sup> Ed. Lisboa, Sílabo.
- Hutmacher, W. (1993). *Quand la réalité résiste à la lutte contre l'échec scolaire. Analyse du reboulement dans l'enseignement primaire genevois*. Genebra: Service de la Recherche Sociologique, Caderno n.º 36.
- IAVE. *PISA* (s.d). Disponível em: <http://iave.pt/np4/12.html> [1 de abril de 2015].
- IBM Corporation (1989, 2011). *TwoStep Cluster Analysis*. IBM Knowledge Center. Disponível em: [http://www-01.ibm.com/support/knowledgecenter/SSLVMB\\_20.0.0/com.ibm.spss.statistics.help/idh\\_t\\_wostep\\_main.htm](http://www-01.ibm.com/support/knowledgecenter/SSLVMB_20.0.0/com.ibm.spss.statistics.help/idh_t_wostep_main.htm) [16 de fevereiro de 2016]
- Iturra, R. (1990). *A Construção Social do Insucesso Escolar – Memória e Aprendizagem em Vila Ruiva*. Lisboa: Editorial Escher.
- Jain, A.K. ; Dubes, R.C. (1988) *Algorithms for Clustering Data*, Englewood cliffs. New Jersey: Prentice Hall.
- Johnson, Richard; Wichern, Dean (2007). *Applied Multivariate Statistical Analysis*. Sixth edition. New Jersey: Pearson Education.
- Landsheere, Vivianne (1994). *Educação e Formação*. Porto: Asa.
- Le Gall, A. (1978). *O Insucesso Escolar*. Lisboa: Editorial Estampa.
- Lewis, I.; Pamela, M. (1987). *So You Want to do Research: A Guide for Teachers on How to Formulate Research Questions*. Edinburgh: The Scottish Council for Research in Education.
- Marchesi, A.; Pérez, E. (2004). *A Compreensão do Fracasso Escolar*, in Marchesi, A. E Gil, C. (org). *Fracasso Escolar. Uma perspectiva multicultural*. Porto Alegre: Editorial Artmed, pp. 17-32.
- Marôco, J. (2011), *Análise Estatística com utilização do SPSS*, 5.<sup>a</sup> Ed. Pero Pinheiro: ReportNumber.

Martins, A. (1993). *Insucesso Escolar e Apoio Educativo*. In Martins, A. e Cabrita, I. In A Problemática do Insucesso Educativo em matemática no 3.º Ciclo do Ensino Básico. Aveiro: Universidade de Aveiro.

Muñiz, B. (1993). *A Família e o Insucesso Escolar*. Lisboa: Porto Editora.

Oliveira, Carla (2010). *(In)Sucesso na Matemática e a utilização de recursos didácticos no 7.º ano de escolaridade: estudo de caso*. Porto: Universidade Portucalense D. Henrique. Disponível em: <http://repositorio.uportu.pt/jspui/bitstream/11328/54/2/TME%20413.pdf> [11 de junho de 2015].

Peixoto, L. (1999). *Auto-Estima, Inteligência e Sucesso Escolar*. Braga: Edições APPACDM Distrital de Braga.

Pereira, Alexandre (2008). *SPSS. Guia prático de utilização. Análise de dados para ciências sociais e psicologia*. 7.ª Ed., Lisboa: Edições Sílabo.

Ponte, João P. (1988). *Matemática, Insucesso e Mudança: Problema Possível, Impossível ou Indeterminado?* Aprender, nº6, 10-19.

Ponte, João P. (1994). *Uma Disciplina Condenada Ao Insucesso*. Noesis, 32. pp 24-26.

Polit, D.; Hungler, B. (1995). *Fundamentos de pesquisa em enfermagem*. 3.ª Ed., Porto Alegre, Artes Médicas.

Punj, G.; Stewart, D. (1983): *Cluster Analysis in Marketing Research: Review and Suggestions for Application*. Journal of Marketing Research, Vol. XX, pp 134- 148.

Quivy, R.; Campenhoudt, L. (1998). *Manual de investigação em ciências sociais*. 2.ª Ed., Lisboa: Gradiva Publicações.

Rangel, Annamaria (1994). *Insucesso escolar*. Lisboa: Instituto Piaget/Horizontes Pedagógicos.

Reis, Elizabeth; Moreira, Raul (1993). *Pesquisa de Mercados*. Lisboa: Edições Sílabo.

Reis, Elizabeth. (2001). *Estatística Multivariada Aplicada*. 2.ª Ed. Lisboa: Edições Sílabo.

Richardson, Roberto J.et al (1989). *Pesquisa social: métodos e técnicas*. 2.ª Ed. São Paulo: Atlas.

Rosário, P. (2004). *Estudar o estudar: As (Des)venturas do Testas*. Porto: Porto Editora.

Rovira, J. (2003). *Educación en valores y fracaso escolar*. In A. Marchesi & C. H. Gil (coords.). El fracaso escolar: Una perspectiva internacional. Madrid: Alianza Editorial, pp. 83-97.

Sil, Vítor (2004). *Alunos em situação de insucesso escolar*. Lisboa: Instituto Piaget.

SPSS. (s.d). *The SPSS TwoStep Cluster Component*. SPSS Inc. Disponível em: [http://www.spss.ch/upload/1122644952\\_The%20SPSS%20TwoStep%20Cluster%20Component.pdf](http://www.spss.ch/upload/1122644952_The%20SPSS%20TwoStep%20Cluster%20Component.pdf) [16 de fevereiro de 2016].

Vilelas, José (2009). *Investigação: o processo de construção do conhecimento*. Lisboa: Edições Sílabo.



## **ANEXOS**

## **ANEXO 1 – Pedido de autorização**

Exma. Sra  
Dra. Joaquina Lourenço  
Diretora do Agrupamento de Escolas Joaquim  
Inácio da Cruz Sobral

Assunto: Recolha de dados estatísticos para Tese de Mestrado

O meu nome é Maria da Luz Soares e exerço funções de professora de Matemática. na Escola Básica e Secundária Joaquim Inácio da Cruz Sobral.

Em 2013, com o objetivo de adquirir novos conhecimentos e melhorar as minhas competências profissionais iniciei o Mestrado em Estatística, Matemática e Computação, Área de Especialização em Estatística computacional, na Universidade Aberta, tendo concluído já a parte curricular do mesmo.

No sentido de aliar o meu gosto pessoal à minha área profissional, estou a desenvolver a tese com o seguinte tema: “Análise de regressão aplicada ao (in)sucesso escolar a Matemática no 3.º ciclo - estudo de caso”.

Para tal solicito autorização para:

- Aplicar um questionário a alguns alunos das turmas do 3.º ciclo;
- Utilizar classificações dos alunos do 3.º ciclo na disciplina de Matemática;

de modo a recolher dados que permitam a realização posterior da análise de regressão.

Comprometo-me a utilizar os dados apenas para o fim indicado.

Grata pela atenção dispensada.

Sobral de Monte Agraço, 08 de janeiro de 2015

---

(Maria da Luz Soares)

## **ANEXO 2 -Questionário aos alunos**

## Questionário aos alunos

Este questionário tem por finalidade recolher informações para a realização de um estudo no âmbito de uma dissertação de Mestrado em Estatística, Matemática e Computação, cujo tema é: “Análise de regressão aplicada ao (in)sucesso escolar a Matemática no 3.º ciclo - Estudo de caso”. A tua colaboração é fundamental para este estudo. O questionário é anónimo e as tuas respostas são totalmente confidenciais e só serão rigorosamente utilizadas para o fim acima mencionado.

### A - Dados pessoais

1. Qual o teu género?

- Masculino       Feminino

2. Qual a tua idade?

- 11     12     13     14     15     16     17     18

3. Qual o ano de escolaridade que frequentas?

- 7.º ano       8.º ano       9.º ano

4. Tens internet em casa?

- Sim                                       Não

5. Quantas horas dormes, em média, por noite?

- Menos de 6 horas     De 6 a 8 horas       De 9 a 10 horas     Mais de 10 horas

### B- Dados do Agregado Familiar

6. Quantas pessoas tem o teu agregado familiar?

- 2       3       4       5       Mais de 5

7. Selecciona a situação que mais se aproxima da tua realidade familiar:

- Vivo com os meus pais (incluindo ou não outros familiares)  
 Vivo só com a minha mãe (incluindo ou não outros familiares)  
 Vivo só com o meu pai (incluindo ou não outros familiares)  
 Vivo só com outros familiares  
 Outras situações

8. Quem é o teu Encarregado de Educação?

- Pai       Mãe       Outro. Qual? \_\_\_\_\_

9. Quais as habilitações do teu pai?

- Analfabeto     1.º Ciclo     2.º Ciclo     3.º Ciclo     Secundário     Curso Superior

10. Quais as habilitações da tua mãe?

- Analfabeta     1.º Ciclo     2.º Ciclo     3.º Ciclo     Secundário     Curso Superior

11. Qual a situação profissional do teu pai?

- Empregado     Desempregado     Em formação     Reformado     Outra

11.1. No caso de teres respondido Empregado, indica a profissão. \_\_\_\_\_

12. Qual a situação profissional da tua mãe?

- Empregada     Desempregada     Em formação     Reformada     Outra

12.1. No caso de teres respondido Empregada, indica a profissão. \_\_\_\_\_

13. Como classificas o teu ambiente familiar?

- Muito Mau     Mau     Razoável     Bom     Muito Bom

### C- Dados escolares

14. Qual o meio de transporte que usas mais vezes para vires para a escola?

- A pé     Autocarro     Carro     Outro. Qual? \_\_\_\_\_

15. Quanto tempo demoras a chegar à escola?

- Menos de 15 min     Entre 15 e 29 min     Entre 30 e 45 min     Mais de 45 min

16. Desejas ir para a universidade?

- Sim     Não

17. Qual a tua disciplina preferida? \_\_\_\_\_

18. Qual a disciplina que menos gostas? \_\_\_\_\_

19. Consideras-te um aluno:

	Sim	Não	Às vezes
Pontual	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Participativo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Empenhado	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Distraído	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trabalhador	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

20. Sentes dificuldades na disciplina de Matemática?

- Sim, bastantes       Às vezes, depende das matérias       Não

21. Quando tens dúvidas ou dificuldades em Matemática, a quem recorres para esclarecê-las?

(Podes assinalar **mais do que uma** opção)

- Família       Colegas       Professora       Explicador(a)       Outro. Quem? \_\_\_\_\_

22. Com que periodicidade costumavas estudar Matemática?

- Nunca       Diariamente       Fins de semana  
 Apenas na altura dos testes       Outra. Qual? \_\_\_\_\_

23. Quantas horas estudas semanalmente para Matemática ?

(realização dos TPC e estudo da matéria dada)

- 0 horas       Menos de 1 hora       De 1 hora a 2 horas       De 3 horas a 4 horas  
 Mais de 4 horas

24. Como classificas o comportamento da turma nas aulas de Matemática?

- Mau       Pouco Satisfatório       Satisfatório       Bom       Muito Bom

25. Como classificas o teu comportamento nas aulas de Matemática?

- Mau       Pouco Satisfatório       Satisfatório       Bom       Muito Bom

26. Como classificas a tua relação com a professora de Matemática?

- Má       Pouco Satisfatória       Satisfatória       Boa       Muito Boa

27. Realizas os trabalhos de casa de Matemática?

- Sempre       Quase sempre       Às vezes       Nunca

28. Qual o interesse dos conteúdos da disciplina de Matemática?

- Sem interesse       Pouco interessante       Interessante  
 Muito interessante       Sem opinião

29. Desde o 5.º ano já obtiveste nota inferior a 3 no 3.º período na disciplina de Matemática?

- Não       1 vez       2 vezes       3 vezes       Mais do que 3

29.1. Se já reprovaste a Matemática, assinala o(s) ano(s) de escolaridade:

- 5.º ano       6.º ano       7.º ano       8.º ano       9.º ano

30. Quando um aluno reprova a Matemática pensas que a responsabilidade é de quem? (Assinala **apenas uma** opção)

- Do aluno
  Do professor
  Da escola  
 Do professor e do aluno
  Do professor e da escola
  Do aluno e da escola  
 Outros. Quais? \_\_\_\_\_

31. Avalia os seguintes aspetos, relativos aos fatores, que na tua opinião, contribuem para as más notas em Matemática. (Assinala cada linha com um X)

Fatores	Discordo totalmente	Discordo em algum grau	Não concordo nem discordo	Concordo em algum grau	Concordo totalmente
Desinteresse pela disciplina					
Falta de hábitos de trabalho e de estudo					
Falta de pré-requisitos necessários					
Falta de atenção/concentração					
Manual desadequado					
Indisciplina existente na sala de aula					
Carga horária global excessiva					
Dificuldades de interpretação e de compreensão					
Falta de apoio por parte do professor					
Falta de apoio familiar					
Rapidez com que o professor dá a matéria					
Falta de oportunidade para esclarecer as dúvidas					

32. Avalia os seguintes aspetos, relativos às medidas, que em tua opinião, poderiam ser tomadas para melhorar os resultados escolares na disciplina de Matemática. (Assinala cada linha com um X)

Medidas a implementar	Discordo totalmente	Discordo em algum grau	Não concordo nem discordo	Concordo em algum grau	Concordo totalmente
Utilização de tecnologias					
Diminuição do número de aulas semanais da disciplina					
Aumento do número de aulas semanais da disciplina					
Mudança das estratégias de ensino					
Promoção de aulas de apoio					
Diminuição do número de alunos por turma					
Mudança nos conteúdos da disciplina					
Maior articulação das matérias da disciplina com aspetos da vida diária					

**Obrigado pela tua colaboração!**

### **ANEXO 3 -Classificação Nacional das Profissões**

A Classificação Nacional das Profissões é o conjunto de todas as profissões existentes em Portugal e da sua respectiva descrição funcional, apresentando-se agregadas por grupos profissionais.

**1. Quadros Superiores da Administração Pública, Dirigentes e Quadros Superiores de Empresa**

**1.1 Quadros Superiores da Administração Pública**

- *Quadros Superiores da Administração Pública*
- *Dirigentes e Quadros Superiores de Organizações Especializadas*

**1.2 Directores de Empresa**

- *Directores Gerais*
- *Directores de Produção, Exploração e Similares*
- *Outros directores de Empresas*

**1.3 Directores e Gerentes de Pequenas Empresas**

- *Directores e Gerentes da Agricultura, Silvicultura e da Pesca*
- *Directores e Gerentes da Produção industrial*
- *Directores e Gerentes da construção civil*
- *Directores e gerentes do comércio grossista e retalhista*
- *Directores e gerentes de restauração e hotelaria*
- *Directores e gerentes de transportes, entreposto e telecomunicações*
- *Directores e gerentes de empresas de mediação e prestação de serviços*
- *Directores e gerentes de empresas de serviços pessoais, limpeza e similares*
- *Directores e gerentes não classificados em outra parte*

**2. Especialistas das Profissões Intelectuais e Científicas**

**2.1 Especialistas das ciências físicas, matemáticas e engenharia**

- *Físicas, químicas e especialistas similares*
- *Matemáticas, estaticistas e especialistas similares*
- *Especialistas da Informática*
- *Arquitectos, Engenheiros e Especialistas similares*

**2.2 Especialistas das ciências da vida e profissionais da saúde**

- *Especialistas das ciências e da vida*
- *Médicos e profissões similares - à excepção dos enfermeiros*
- *Enfermeiros*

**2.3 Docentes do ensino secundário, superior e profissões similares**

- *Docentes do ensino universitário e de estabelecimentos de ensino superior*
- *Docentes do ensino básico (2.º e 3.º ciclo) e secundário*
- *Docentes do ensino superior, básico e secundário e similares não classificados em outra parte*

**3. Técnicos e Profissionais de nível intermédio**

**3.1 Técnicos e profissionais de nível intermédio das ciências físicas e químicas, da engenharia e trabalhadores similares**

- *Técnicos de ciências físico-químicas*
- *Técnicos de engenharia civil*
- *Técnicos de electricidade*
- *Técnicos de electrónica e telecomunicações*
- *Técnicos de relojoaria*
- *Técnicos intermédios de química industrial*
- *Desenhadores e trabalhadores similares*
- *Técnicos de investigação física e química, do fabrico industrial e trabalhadores similares não classificados em outra parte*

**3.2 Profissionais de nível intermédio das ciências da vida e da saúde**

- *Técnicos das ciências da vida e da saúde*

- *Profissionais técnicos de medicina - à excepção dos enfermeiros*
- *Parteiras*
- *Especialistas da medicina tradicional*

### **3.3 Profissionais de nível intermédio do ensino**

- *Docentes do ensino básico, primário e pré-primário*
- *Educadores de infância*
- *Docentes de Educação especial*
- *Profissionais do ensino não classificado em outra parte*

### **3.4 Outros técnicos e profissionais de nível intermédio**

- *Profissionais de nível intermédio de finanças e serviços comerciais*
- *Agentes comerciais e correctores*
- *Profissionais de nível intermédio de gestão e administração*
- *Profissionais de nível intermédio da administração pública, das alfândegas, dos impostos e trabalhos similares*
- *Inspectores da policia judiciária e detectives*
- *Profissionais da criação artística do espectáculo e do desporto*

## **4. Pessoal Administrativo e Similares**

### **4.1 Empregados de escritório**

- *Secretárias e Operadores de equipamentos e tratamento de informação*
- *Empregados dos serviços de contabilidade e dos serviços financeiros*
- *Empregados do aprovisionamento, de planeamento e dos transportes*
- *Empregados de biblioteca, carteiros e trabalhadores similares*
- *Empregados de escritório não classificados em outra parte*

### **4.2 Empregados de recepção, caixas, bilheteiras e similares**

- *Caixas e Bilheteiras*
- *Caixas e Estabelecimentos bancários*
- *Empregados da banca e casinos e similares*
- *Penhoristas e trabalhadores similares*
- *Cobreadores e trabalhadores similares*

## **5. Pessoal dos Serviços e Vendedores**

### **5.1 Pessoal dos Serviços Directos e Particulares, de protecção e segurança**

- *Assistentes, cobreadores, guias e trabalhadores similares*
- *Ecónomos e pessoal do serviço de restauração*
- *Vigilantes, assistentes médicos e trabalhadores similares*
- *Outro pessoal dos serviços directos e particulares*

### **5.2 Manequins, vendedores e demonstradores**

- *Manequins e outros modelos*
- *Vendedores e demonstradores*
- *Vendedores de quiosque e mercado*

## **6. Agricultores e Trabalhadores Qualificados da Agricultura e Pescas**

### **6.1 Agricultores e trabalhadores qualificados da agricultura, criação de animais e pescas**

- *Agricultores e trabalhadores qualificados de culturas agrícolas*
- *Criadores e trabalhadores qualificados do tratamento de animais*
- *Agricultores e trabalhadores qualificados da policultura, criação e tratamento de animais*
- *Trabalhadores florestais e similares*
- *Trabalhadores da aquacultura e pescas*

### **6.2 Agricultores e pescadores - agricultura e pesca de subsistência**

- *Agricultor - agricultura de subsistência*
- *Pescador - Pesca de subsistência*
- *Outros agricultores e pescadores - agricultura e pesca de subsistência*

## **7. Operários, Artífices e Trabalhadores Similares**

### **7.1 Operários, artífices e trabalhadores similares das indústrias extractivas e da construção civil**

- *Mineiros, canteiros, carregadores de fogo e trabalhadores de pedra*
- *Trabalhadores da construção civil e obras públicas*
- *Trabalhadores da construção civil e similares - acabamentos*
- *Pintores, limpadores de fachadas e trabalhadores similares*

#### **7.2 Trabalhadores da metalurgia e da metalomecânica e trabalhadores similares**

- *Moldadores, soldadores, bate-chapas, caldeireiros, montadores de estruturas e trabalhadores similares*
- *Forjadores, serralheiros mecânicos e trabalhadores similares*
- *Mecânicos e ajustadores de máquinas*
- *Mecânicos e ajustadores de equipamentos eléctricos e electrónicos*

#### **7.3 Mecânicos de precisão, oleiros e vidreiros, artesãos, trabalhadores das artes gráficas e trabalhadores similares**

- *Mecânicos de precisão em metal e materiais similares*
- *Oleiros, vidreiros e trabalhadores similares*
- *Artesãos de madeira, tecido, couro e materiais similares*
- *Compositores tipográficos e trabalhadores similares*

#### **7.4 Outros operários, artífices e trabalhadores similares**

- *Trabalhadores de preparação e confecção de alimentos e bebidas e trabalhadores similares*
- *Trabalhadores das madeiras e similares*
- *Trabalhadores dos têxteis e confecções e trabalhadores similares*
- *Trabalhadores de peles, couro e calçado*
- *Trabalhadores de artigos de pirotecnia*

### **8. Operadores de instalações e máquinas e trabalhadores da montagem**

#### **8.1 Operadores de instalações fixas e similares**

- *Operadores e condutores de máquinas e instalações mineiras de extracção e tratamento de minerais*
- *Operadores de instalações de transformação de metais*
- *Operadores de instalações de fabricação de vidro, cerâmica e trabalhadores similares*
- *Operadores de instalações de tratamentos químicos*
- *Operadores de instalações de produção de energia e trabalhadores similares*
- *Operadores de cadeias de montagem automatizadas e de "robots" industriais*

#### **8.2 Operadores de máquinas e trabalhadores da montagem**

- *Operadores de máquinas para trabalhar metais e produtos minerais*
- *Operadores de máquinas para trabalhar metais e produtos químicos*
- *Operadores de máquinas para fabricar produtos de borracha e matéria plástica*
- *Operadores de máquinas para fabricar produtos de madeira*
- *Operadores de máquinas de impressão, encadernação e fabricação de produtos de papel*
- *Operadores de máquinas para fabricar produtos têxteis e artigos em pele e couro*
- *Operadores de máquinas para fabricar alimentos e produtos similares*
- *Trabalhadores de montagem*
- *Outros operadores de máquinas e trabalhadores similares*

#### **8.3 Condutores de veículos e embarcações e operadores de equipamentos pesados móveis**

- *Maquinistas de locomotivas e trabalhadores similares*
- *Condutores de veículos a motor*
- *Operadores de maquinaria agrícola móvel e de outras máquinas móveis*
- *Mestres, marinheiros e trabalhadores similares*

### **9. Trabalhadores não qualificados**

#### **9.1 Trabalhadores não qualificados dos serviços e comércio**

- *Vendedores ambulantes e trabalhadores similares*
- *Engraxadores e trabalhadores similares*
- *Pessoal de limpeza, lavadeiras, engomadores de roupa e trabalhadores similares*
- *Porteiros de prédios urbanos, lavadores de vidros e veículos e trabalhadores similares*

**9.2 Trabalhadores não qualificados da agricultura e pescas**

**9.3 Trabalhadores não qualificados das minas, da construção e obras públicas, da indústria transformadora e dos transportes**

- *Trabalhadores não qualificados das minas e da construção civil e obras públicas*
- *Trabalhadores não qualificados da indústria transformadora*
- *Trabalhadores não qualificados dos transportes*

## **ANEXO 4- Solicitação de dados estatísticos**

Exma. Sra.

Dra. Joaquina Lourenço

Diretora do Agrupamento Joaquim Inácio da Cruz Sobral

Assunto: Pedido de dados estatísticos

O meu nome é Maria da Luz Frade Soares e exerço funções como professora do Grupo 500-Matemática neste agrupamento.

Na sequência do pedido do pedido por mim realizado de autorização para aplicação da parte prática da minha tese de mestrado, intitulada “Análise de regressão aplicada ao (in)sucesso escolar a Matemática no 3.º ciclo - estudo de caso”, neste agrupamento, o qual foi autorizado pelo Conselho Pedagógico de 14 de janeiro de 2015, venho solicitar a seguinte informação:

Em relação aos anos letivos **2014-2015; 2013-2014 e 2012-2013:**

\* número inscritos por ano de escolaridade

\* taxas globais de sucesso ou número de alunos retidos ou número de alunos transitados por ano de escolaridade

\* número de alunos NEE por ano de escolaridade

\* número de alunos com ASE por ano de escolaridade

Comprometendo-me a utilizar estes dados apenas para o fim indicado.

Sobral de Monte Agraço, 28 de agosto de 2015

---

(Maria da Luz Soares)

## **ANEXO 5-Classificação das Variáveis**

Questão	Variável	Designação	Tipo de variável	Valores assumidos
1	Género	Sexo	Nominal	-1 – “NR” 0 – “Masculino” 1 – “Feminino”
2	Idade	Idade	Escala	-1 -”NR”; 12,13,14,15,16,17,18
3	Ano de escolaridade	Ano	Ordinal	-1 – “NR” 1 – “7.ºano” 2 – “8.ºano” 3 – “9.ºano”
4	Possuir internet em casa	Internet	Nominal	-1 – “NR” 1 – “Sim” 0 – “Não”
5	Horas de sono	HoraSono	Ordinal	-1 – “NR” 1 – “Menos de 6h” 2 – “De 6h a 8h” 3 – “De 9 a 10h” 4 – “Mais de 10h”
6	Elementos do agregado familiar	ElemsFamilia	Ordinal	-1 – “NR” 1 – “2 pessoas” 2 – “3 pessoas” 3 – “4 pessoas” 4 – “5 pessoas” 5 – “Mais de 5 pessoas”
7	Realidade familiar	NucleoFamiliar	Nominal	-1 – “NR” 1 – “Vive com os pais” 2 – “Vive só com a mãe” 3 – “Vive só com o pai” 4 – “Vive só com outros familiares” 5 – “Outras situações”
8	Encarregado de educação	EE	Nominal	-1 – “NR” 1 – “Pai” 2 – “Mãe” 3 – “Outro”
8.1.	Outros encarregados de educação	EE_outro	Nominal	-1 – “NR” -2 – “NSA” 1 – “Tia” 2 – Madrasta” 3 – “Irmã” 4 – “Avó”
9	Habilitações do Pai	HabilPai	Nominal	-1 – “NR” 1 – “Analfabeto” 2 – “1.º ciclo” 3 – “2.º ciclo” 4 – “3.º ciclo” 5 – “Secundário” 6 – “Superior”
10	Habilitações da mãe	HabilMae	Nominal	-1 – “NR” 1 – “Analfabeto” 2 – “1.º ciclo” 3 – “2.º ciclo” 4 – “3.º ciclo” 5 – “Secundário” 6 – “Superior”

<b>Questão</b>	<b>Variável</b>	<b>Designação</b>	<b>Tipo de variável</b>	<b>Valores assumidos</b>
11	Situação Profissional do pai	SitProfPai	Nominal	-1 – “NR” 1 – “Empregado” 2 – “Desempregado” 3 – “Em formação” 4 – “Reformado” 5 – “Outra”
11.1	Profissão do pai	ProfPai	Nominal	-1 – “NR” 1 – “Quadros Superior da Adm. Pública, Dirigentes e Quadros Superiores de Empresa” 2 – “Especialistas das Profissões Intelectuais e Científicas” 3 – “Técnicos e Profissionais de Nível Intermédio” 4 – “Pessoal Administrativo e Similares” 5 – “Pessoal dos Serviços e Vendedores” 6 – “Agricultores e Trabalhadores Qualificados da Agricultura e Pescas” 7 – “Operários, Artífices e Trabalhadores similares” 8 – “Operadores de instalações e máquinas e trabalhadores da montagem” 9 – “Trabalhador não qualificado”
12	Situação Profissional da mãe	SitProfMae	Nominal	-1 – “NR” 1 – “Empregado” 2 – “Desempregado” 3 – “Em formação” 4 – “Reformado” 5 – “Outra”

Questão	Variável	Designação	Tipo de variável	Valores assumidos
12.1	Profissão da mãe	ProfMae	Nominal	-1 – “NR” 1 – “Quadros Superior da Adm. Pública, Dirigentes e Quadros Superiores de Empresa” 2 – “Especialistas das Profissões Intelectuais e Científicas” 3 – “Técnicos e Profissionais de Nível Intermédio” 4 – “Pessoal Administrativo e Similares” 5 – “Pessoal dos Serviços e Vendedores” 6 – “Agricultores e Trabalhadores Qualificados da Agricultura e Pescas” 7 – “Operários, Artífices e Trabalhadores similares” 8 – “Operadores de instalações e máquinas e trabalhadores da montagem” 9 – “Trabalhador não qualificado”
13	Ambiente familiar	Amb_Familiar	Nominal	-1 – “NR” 1 – “Muito mau” 2 – “Mau” 3 – “Razoável” 4 – “Bom” 5 – “Muito bom”
14	Meio de transporte	Transporte	Nominal	-1 – “NR” 1 – “A pé” 2 – “Autocarro” 3 – “Carro” 4 – “Outro”
14.1	Outro transporte	Transporte_outro	Nominal	-1 – “NR” -2 – “NSA”
15	Duração da viagem	TempoTransporte	Ordinal	-1 – “NR” 1 – “Menos de 15 min” 2 – “Entre 15 e 29 min” 3 – “Entre 30 e 45 min” 4 – “Mais de 45 min”
16	Deseja ir para universidade	Univ	Nominal	-1 – “NR” 1 – “Sim” 0 – “Não”

Questão	Variável	Designação	Tipo de variável	Valores assumidos
17	Disciplina preferida	DiscPreferida	Nominal	-3 – “NV” -1 – “NR” 1 – “P” 2 – “I” 3 – “F” 4 – “M” 5 – “FQ” 6 – “CN” 7 – “EV” 8 – “EF” 9 – “G” 10 – “H” 11 – “TIC” 12 – “ET” 13 – “Nenhuma”
18	Disciplina que menos gosta	DiscGostaMenos	Nominal	-3 – “NV” -1 – “NR” 1 – “P” 2 – “I” 3 – “F” 4 – “M” 5 – “FQ” 6 – “CN” 7 – “EV” 8 – “EF” 9 – “G” 10 – “H” 11 – “TIC” 12 – “ET” 13 – “Nenhuma”
19.1	Aluno pontual	AlunoPontual	Ordinal	-1 – “NR” 0 – “Não” 1 – “Às vezes” 2 – “Sim”
19.2	Aluno participativo	AlunoParticipativo	Ordinal	-1 – “NR” 0 – “Não” 1 – “Às vezes” 2 – “Sim”
19.3	Aluno empenhado	AlunoEmpenhado	Ordinal	-1 – “NR” 0 – “Não” 1 – “Às vezes” 2 – “Sim”
19.4	Aluno distraído	AlunoDistraido	Ordinal	-1 – “NR” 0 – “Não” 1 – “Às vezes” 2 – “Sim”
19.5	Aluno trabalhador	AlunoTrabalhador	Ordinal	-1 – “NR” 0 – “Não” 1 – “Às vezes” 2 – “Sim”

Questão	Variável	Designação	Tipo de variável	Valores assumidos
20	Dificuldades em Matemática	DificMat	Ordinal	-1 – “NR” 0 – “Não” 1 – “Às vezes, depende das matérias” 2 – “Sim, bastantes”
21	A quem o aluno recorre para esclarecer dúvidas			
21.1	Quem lhe esclarece as dúvidas: Família	AjudaMat_1	Nominal	-1 – “NR” 1 – “Sim” 0 – “Não”
21.2	Quem lhe esclarece as dúvidas: Colegas	AjudaMat_2	Nominal	-1 – “NR” 1 – “Sim” 0 – “Não”
21.3	Quem lhe esclarece as dúvidas: Professor	AjudaMat_3	Nominal	-1 – “NR” 1 – “Sim” 0 – “Não”
21.4	Quem lhe esclarece as dúvidas: Explicador	AjudaMat_4	Nominal	-1 – “NR” 1 – “Sim” 0 – “Não”
21.5	Quem lhe esclarece as dúvidas: Outro	AjudaMat_5	Nominal	-1 – “NR” 1 – “Sim” 0 – “Não”
21.5	Que outro lhe esclarece as dúvidas	AjudaMat_outro	Nominal	-1 – “NR” -2 – “NSA” 1 – “Ninguém” 2 – “Internet”
22	Periodicidade do estudo de matemática	PeridEstudo	Nominal	-1 – “NR” 1 – “Nunca” 2 – “Diariamente” 3 – “Fins de semana” 4 – “Apenas na altura dos testes” 5 – “Outra”
22	Periodicidade do estudo outra qual	PeridEstudo_outro	Nominal	-1 – “NR” -2 – “NSA” 1 – “Quando tem dificuldades e altura de teste” 2 – “Quando tem dificuldades” 3 – “Dias da explicação” 4 – “Quando tem TPC” 5 – “Altura dos testes e fins semana” 6 – “Às vezes” 7 – “Dias em que tem aula” 8 – “Dias de sala de estudo” 9 – “Semanalmente” 10 – “Dois dias por semana”
23	Horas de estudo semanal para matemática	TempoEstudo	Ordinal	-1 – “NR” 1 – “0 h” 2 – “Menos de 1 h” 3 – “De 1 a 2 h” 4 – “De 3 a 4 h” 5 – “Mais de 4 h”

Questão	Variável	Designação	Tipo de variável	Valores assumidos
24	Comportamento da turma	CompTurma	Ordinal	-1 – “NR” 1 – “Mau” 2 – “Pouco Satisfatório” 3 – “Satisfatório” 4 – “Bom” 5 – “Muito Bom”
25	Comportamento do aluno	CompAluno	Ordinal	-1 – “NR” 1 – “Mau” 2 – “Pouco Satisfatório” 3 – “Satisfatório” 4 – “Bom” 5 – “Muito Bom”
26	Relação com a professora	RelacaoProf	Ordinal	-1 – “NR” 1 – “Má” 2 – “Pouco Satisfatória” 3 – “Satisfatória” 4 – “Boa” 5 – “Muito Boa”
27	Realização dos trabalhos de casa de Mat.	FazTPC	Ordinal	-1 – “NR” 1 – “Nunca” 2 – “Às vezes” 3 – “Quase sempre” 4 – “Sempre”
28	Interesse pelos conteúdos de Mat.	IntCont	Nominal	-1 – “NR” 1 – “Sem interesse” 2 – “Pouco interessante” 3 – “Interessante” 4 – “Muito interessante” 5 – “Sem opinião”
29	Reprovações a Matemática	Repetência	Ordinal	-1 – “NR” 0 – “Não” 1 – “1 vez” 2 – “2 vezes” 3 – “3 vezes” 4 – “Mais do que 3 vezes”
29.1	Anos em que reprovou: 5.º	AnosRep_5	Nominal	-1 – “NR” -2 – “NSA” 0 – “Não” 1 – “Sim”
29.1	Anos em que reprovou: 6.º	AnosRep_6	Nominal	-1 – “NR” -2 – “NSA” 0 – “Não” 1 – “Sim”
29.1	Anos em que reprovou: 7.º	AnosRep_7	Nominal	-1 – “NR” -2 – “NSA” 0 – “Não” 1 – “Sim”

Questão	Variável	Designação	Tipo de variável	Valores assumidos
29.1	Anos em que reprovou: 8.º	AnosRep_8	Nominal	-1 – “NR” -2 – “NSA” 0 – “Não” 1 – “Sim”
29.1	Anos em que reprovou: 9.º	AnosRep_9	Nominal	-1 – “NR” -2 – “NSA” 0 – “Não” 1 – “Sim”
30	Responsável pela repetência	RespRepet	Nominal	-1 – “NR” 1 – “Aluno” 2 – “Professor” 3 – “Escola” 4 – “Professor e aluno” 5 – “Professor e escola” 6 – “Aluno e escola” 7 – “Outros”
30.1	Responsável pela repetência outro qual	RespRepet_outro	Nominal	-1 – “NR” -2 – “NSA” 1 – “Aluno e pais” 2 – “Ninguém” 3 – “Aluno, professor e escola” 4 – “Aluno e situação familiar”
31	Causas do insucesso			
31.1	Desinteresse pela disciplina	DesintDiscipl	Ordinal	1-Discordo totalmente 2-Discordo em algum grau 3-Não concordo nem discordo 4-Concordo em algum grau 5-Concordo totalmente
31.2	Falta de hábitos de trabalho e de estudo	FaltaHabTrab	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”

<b>Questão</b>	<b>Variável</b>	<b>Designação</b>	<b>Tipo de variável</b>	<b>Valores assumidos</b>
31.3	Falta de pré-requisitos necessários	FaltaReq	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”
31.4	Falta de atenção/concentração nas aulas	FaltaAtenc	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”
31.5	Manual desadequado	Manual	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”
31.6	Indisciplina na sala de aula	Indisciplina	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”
31.7	Carga horária global excessiva	CargaHoraria	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”
31.8	Dificuldades de interpretação e de compreensão	DificInterp	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”

<b>Questão</b>	<b>Variável</b>	<b>Designação</b>	<b>Tipo de variável</b>	<b>Valores assumidos</b>
31.9	Falta de apoio por parte do professor	FApoioProf	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”
31.10	Falta de apoio familiar	FApoioFAMIL	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”
31.11	Rapidez com que o professor dá a matéria	Rapidez	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”
31.12	Falta de oportunidade para esclarecer as dúvidas	FEsclarecim	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”
32	Medidas para promover o sucesso			
32.1	Utilização das tecnologias	Tecnologias	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”
32.2	Diminuição do número de aulas	MenosAulas	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”

<b>Questão</b>	<b>Variável</b>	<b>Designação</b>	<b>Tipo de variável</b>	<b>Valores assumidos</b>
32.2	Diminuição do número de aulas	MenosAulas	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”
32.3	Aumento do número de aulas	MaisAulas	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”
32.4	Mudança das estratégias de ensino	Estratégias	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”
32.5	Promoção de aulas de apoio	AulasApoio	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”
32.6	Diminuição do número de alunos por turma	TurmaPequena	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”

<b>Questão</b>	<b>Variável</b>	<b>Designação</b>	<b>Tipo de variável</b>	<b>Valores assumidos</b>
32.7	Mudança nos conteúdos	Conteúdos	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”
32.8	Maior articulação das matérias com a realidade da vida diária	VidaDiária	Ordinal	-1 – “NR” 1 – “Discordo totalmente” 2 – “Discordo em algum grau” 3 – “Não concordo nem discordo” 4 – “Concordo em algum grau” 5 – “Concordo totalmente”

**ANEXO 6 - Correlações – Coeficiente de Correlação de Spearman**  
**Entre os fatores que contribuem para o insucesso**

### Fatores que contribuem para o insucesso

Coeficiente de Spearman	31.1 Desinteresse pela disciplina	31.2 Falta de hábitos de trabalho e de estudo	31.3 Falta de pré requisitos necessários	31.4 Falta de atenção/concentração nas aulas	31.5 Manual desadequado	31.6 Indisciplina na sala de aula	31.7 Carga horária global excessiva	31.8 Dificuldades de interpretação e de compreensão	31.9 Falta de apoio por parte do professor	31.10 Falta de apoio familiar
31.1 Desinteresse	1,000	,520**	,219**	,463**	,120	,369**	,139	,370**	,172*	,162*
	.	,000	,002	,000	,093	,000	,052	,000	,015	,022
31.2 Falta de hábitos de trabalho e de estudo	,520**	1,000	,366**	,557**	,140*	,283**	,062	,383**	-,007	,172*
	,000	.	,000	,000	,050	,000	,389	,000	,923	,015
31.3 Falta de pré requisitos necessários	,219**	,366**	1,000	,297**	,238**	,220**	,001	,367**	,184*	,223**
	,002	,000	.	,000	,001	,002	,986	,000	,010	,002
31.4 Falta de atenção/concentração nas aulas	,463**	,557**	,297**	1,000	,149*	,395**	,062	,391**	,015	,125
	,000	,000	,000	.	,038	,000	,385	,000	,830	,079
31.5 Manual desadequado	,120	,140*	,238**	,149*	1,000	,277**	,179*	,271**	,323**	,257**
	,093	,050	,001	,038	.	,000	,013	,000	,000	,000
31.6 Indisciplina na sala de aula	,369**	,283**	,220**	,395**	,277**	1,000	,071	,240**	,210**	,308**
	,000	,000	,002	,000	,000	.	,324	,001	,003	,000
31.7 Carga horária global excessiva	,139	,062	,001	,062	,179*	,071	1,000	,305**	,319**	,068
	,052	,389	,986	,385	,013	,324	.	,000	,000	,343
31.8 Dificuldades de interpretação e de compreensão	,370**	,383**	,367**	,391**	,271**	,240**	,305**	1,000	,319**	,188**
	,000	,000	,000	,000	,000	,001	,000	.	,000	,008
31.9 Falta de apoio por parte do professor	,172*	-,007	,184*	,015	,323**	,210**	,319**	,319**	1,000	,257**
	,015	,923	,010	,830	,000	,003	,000	,000	.	,000
31.10 Falta de apoio familiar	,162*	,172*	,223**	,125	,257**	,308**	,068	,188**	,257**	1,000
	,022	,015	,002	,079	,000	,000	,343	,008	,000	.

\*\* . A correlação é significativa no nível 0,01 (2 extremidades).

\* . A correlação é significativa no nível 0,05 (2 extremidades).

**ANEXO 7 - Correlações – Coeficiente de Correlação de Spearman**  
**Entre as medidas para reduzir o insucesso**

**Coefficientes de correlação de Spearman entre as estratégias para reduzir o insucesso**

		Utilização de tecnologias	Diminuição do número de aulas	Aumento do número de aulas	Mudança das estratégias de ensino	Promoção de aulas de apoio	Diminuição do número de alunos por turma	Mudanças nos conteúdos	Maior articulação das matérias com aspetos da vida diária
Utilização de tecnologias	Coef. Correlação	1,000	<b>,459**</b>	<b>-,145*</b>	<b>,198**</b>	,089	-,024	<b>,237**</b>	<b>,315**</b>
	Sig. (2 extremidades)	.	,000	,041	,005	,218	,737	,001	,000
Diminuição do número de aulas	Coef. Correlação	<b>,459**</b>	1,000	<b>-,251**</b>	<b>,142*</b>	<b>-,142*</b>	-,046	<b>,292**</b>	<b>,192**</b>
	Sig. (2 extremidades)	,000	.	,000	,045	,048	,522	,000	,006
Aumento do número de aulas	Coef. Correlação	<b>-,145*</b>	<b>-,251**</b>	1,000	<b>,161*</b>	<b>,313**</b>	,069	-,137	-,054
	Sig. (2 extremidades)	,041	,000	.	,023	,000	,334	,053	,443
Mudança das estratégias de ensino	Coef. Correlação	<b>,198**</b>	<b>,142*</b>	<b>,161*</b>	1,000	<b>,252**</b>	<b>,268**</b>	<b>,382**</b>	<b>,269**</b>
	Sig. (2 extremidades)	,005	,045	,023	.	,000	,000	,000	,000
Promoção de aulas de apoio	Coef. Correlação	,089	<b>-,142*</b>	<b>,313**</b>	<b>,252**</b>	1,000	,038	-,114	,103
	Sig. (2 extremidades)	,218	,048	,000	,000	.	,602	,111	,151
Diminuição do número de alunos por turma	Coef. Correlação	-,024	-,046	,069	<b>,268**</b>	,038	1,000	<b>,375**</b>	<b>,254**</b>
	Sig. (2 extremidades)	,737	,522	,334	,000	,602	.	,000	,000
Mudança nos conteúdos	Coef. Correlação	<b>,237**</b>	<b>,292**</b>	-,137	<b>,382**</b>	-,114	<b>,375**</b>	1,000	<b>,455**</b>
	Sig. (2 extremidades)	,001	,000	,053	,000	,111	,000	.	,000
Maior articulação das matérias com aspetos da vida diária	Coef. Correlação	<b>,315**</b>	<b>,192**</b>	-,054	<b>,269**</b>	,103	<b>,254**</b>	<b>,455**</b>	1,000
	Sig. (2 extremidades)	,000	,006	,443	,000	,151	,000	,000	.

\*\* . A correlação é significativa no nível 0,01 (2 extremidades).

\* . A correlação é significativa no nível 0,05 (2 extremidades).

**ANEXO 8 – Resultados da Regressão Logística Univariada para as  
variáveis significativas**

**Resultado da análise de regressão logística univariada para as variáveis significativas**

Variáveis independentes	Teste de coeficientes de modelo Omnibus P valor	Teste de ajuste da qualidade do Modelo HL P valor	Variáveis na equação			
			P valor	OR	95% CI para EXP(B)	
					Inferior	Superior
idade	0,000	0,125	0,000	2,091	1,578	2,771
Internet(1)	0,002	---	0,021	11,609	1,457	92,510
HabilPai	0,057	1,000	0,077			
HabilPai(1)			0,018	4,048	1,276	12,840
HabilPai(2)			0,009	4,533	1,461	14,068
Amb_Familiar	0,001	1,000	0,036			
Amb_Familiar(2)			0,004	4,836	1,663	14,063
Univ	0,000		0,000	3,769	1,839	7,725
Participativo	0,001	1,000	0,002			
Participativo(1)			0,002	8,667	2,224	33,777
Participativo(2)			0,014	2,211	1,172	4,170
Empenhado	0,000	1,000	0,000			
Empenhado(2)			0,000	5,476	2,952	10,157
Distraido	0,034	1,000	0,039			
Distraido(1)			0,013	0,303	0,118	0,777
Distraido(2)			0,047	0,464	0,217	0,990
Trabalhador	0,000	1,000	0,000			
Trabalhador(2)			0,000	3,376	1,846	6,173
DificMat	0,000	1,000	0,000			
DificMat(1)			0,000	0,019	0,004	0,078
DificMat(2)			0,000	0,048	0,014	0,162
PeridEstudo	0,022	1,000	0,040			
TempoEstudo	0,007	1,000	0,013			
CompTurma	0,028	1,000	0,037			
CompAluno	0,002	1,000	0,004			
RelProfessora	0,000	1,000	0,006			
RelacaoProf(2)			0,016	13,650	1,632	114,191
FazTPC	0,000	1,000	0,000			
FazTPC(1)			0,000	12,500	3,647	42,843
FazTPC(2)			0,000	9,943	4,121	23,992
FazTPC(3)			0,014	2,534	1,208	5,315
Interesse	0,000	1,000	0,000			
Interesse(1)			0,044	3,215	1,031	10,024
Interesse(3)			0,002	0,297	0,139	0,635

**ANEXO 9 – Outputs do SPSS para Regressão logística com  
16 variáveis independentes**

GET

```
FILE='D:\GoogleDriveNuvem\TESE\Inquéritos Insucesso_CN2roc.sav'.  
DATASET NAME Conjunto_de_dados1 WINDOW=FRONT.  
LOGISTIC REGRESSION VARIABLES Repetencia2  
/METHOD=ENTER Idade Internet Amb_Familiar Univ Empenhado Participativo Distraido Trabalhador DificMat PeridEstudo  
TempoEstudo CompTurma CompAluno RelacaoProf FazTPC Interesse  
/CONTRAST (Internet)=Indicator(1)  
/CONTRAST (Amb_Familiar)=Indicator(1)  
/CONTRAST (Univ)=Indicator(1)  
/CONTRAST (Empenhado)=Indicator(1)  
/CONTRAST (Participativo)=Indicator(1)  
/CONTRAST (Distraido)=Indicator(1)  
/CONTRAST (Trabalhador)=Indicator(1)  
/CONTRAST (DificMat)=Indicator(1)  
/CONTRAST (PeridEstudo)=Indicator(1)  
/CONTRAST (TempoEstudo)=Indicator(1)  
/CONTRAST (CompTurma)=Indicator(1)  
/CONTRAST (CompAluno)=Indicator(1)  
/CONTRAST (RelacaoProf)=Indicator(1)  
/CONTRAST (FazTPC)=Indicator(1)  
/CONTRAST (Interesse)=Indicator(1)  
/CLASSPLOT  
/CASEWISE OUTLIER(2)  
/PRINT=GOODFIT CI(95)  
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

### Regressão logística

[Conjunto\_de\_dados1] D:\GoogleDriveNuvem\TESE\Inquéritos Insucesso\_CN2roc.sav

Resumo de processamento do caso			
Casos não ponderados <sup>a</sup>		N	Porcentagem
Casos selecionados	Incluído na análise	187	94,0
	Casos ausentes	12	6,0
	Total	199	100,0
Casos não selecionados		0	,0
Total		199	100,0

a. Se a ponderação estiver em vigor, veja a tabela de classificação para o número total de casos.

### Codificação de variável dependente

Valor original	Valor interno
Não	0
Sim	1

**Codificações de variáveis categóricas**

		Frequência	Codificação de parâmetro			
			(1)	(2)	(3)	(4)
28. Interesse pelos conteúdos	Sem interesse	25	,000	,000	,000	,000
	Pouco interessante	34	1,000	,000	,000	,000
	Interessante	80	,000	1,000	,000	,000
	Muito interessante	6	,000	,000	1,000	,000
26. Relação com a professora	Sem opinião	42	,000	,000	,000	1,000
	Má	8	,000	,000	,000	,000
	Pouco Satisfatória	12	1,000	,000	,000	,000
	Satisfatória	47	,000	1,000	,000	,000
25. Comportamento do aluno	Boa	79	,000	,000	1,000	,000
	Muito Boa	41	,000	,000	,000	1,000
	Mau	8	,000	,000	,000	,000
	Pouco Satisfatório	13	1,000	,000	,000	,000
24. Comportamento da turma	Satisfatório	62	,000	1,000	,000	,000
	Bom	88	,000	,000	1,000	,000
	Muito Bom	16	,000	,000	,000	1,000
	Mau	18	,000	,000	,000	,000
23. Horas de estudo semanal	Pouco Satisfatório	33	1,000	,000	,000	,000
	Satisfatório	89	,000	1,000	,000	,000
	Bom	41	,000	,000	1,000	,000
	Muito Bom	6	,000	,000	,000	1,000
22. Periodicidade de estudo	0 h	24	,000	,000	,000	,000
	Menos de 1 h	52	1,000	,000	,000	,000
	De 1 a 2 h	94	,000	1,000	,000	,000
	De 3 a 4 h	12	,000	,000	1,000	,000
27. Realização dos TPC de matemática	Mais de 4 h	5	,000	,000	,000	1,000
	Nunca	26	,000	,000	,000	,000
	Diariamente	23	1,000	,000	,000	,000
	Fins semana	39	,000	1,000	,000	,000
13. Ambiente Familiar	Apenas na altura dos testes	86	,000	,000	1,000	,000
	Outra	13	,000	,000	,000	1,000
	Nunca	18	,000	,000	,000	,000
	Às vezes	44	1,000	,000	,000	,000
19.3 Aluno empenhado	Quase sempre	62	,000	1,000	,000	,000
	Sempre	63	,000	,000	1,000	,000
	Mau	4	,000	,000	,000	,000
	Razoável	19	1,000	,000	,000	,000
19.2 Aluno participativo	Bom	61	,000	1,000	,000	,000
	Muito bom	103	,000	,000	1,000	,000
	Não	5	,000	,000	,000	,000
	Às vezes	91	1,000	,000	,000	,000
19.4 Aluno distraído	Sim	91	,000	1,000	,000	,000
	Não	15	,000	,000	,000	,000
	Às vezes	114	1,000	,000	,000	,000
	Sim	58	,000	1,000	,000	,000
19.5 Aluno trabalhador	Não	36	,000	,000	,000	,000
	Às vezes	116	1,000	,000	,000	,000
	Sim	35	,000	1,000	,000	,000
	Não	4	,000	,000	,000	,000
20. Dificuldades a matemática	Às vezes	71	1,000	,000	,000	,000
	Sim	112	,000	1,000	,000	,000
	Não	30	,000	,000	,000	,000
	Às vezes, depende da matéria	115	1,000	,000	,000	,000
16. Deseja ir para universidade	Sim, bastantes	42	,000	1,000	,000	,000
	Não	47	,000	,000	,000	,000
	Sim	140	1,000	,000	,000	,000
	Não	10	,000	,000	,000	,000
4. Ter Internet em casa	Sim	177	1,000	,000	,000	,000

**Bloco 0: Bloco Inicial**

**Tabela de Classificação<sup>a,b</sup>**

	Observado	Previsto			
		Repetência a Matemática Alguma vez		Porcentagem correta	
		Não	Sim		
Etapa 0	Repetência a Matemática Alguma vez	Não	94	0	100,0
		Sim	93	0	,0
	Porcentagem global				50,3

a. A constante está incluída no modelo.

b. O valor de recorte é ,500

**Variáveis na equação**

	B	S.E.	Wald	df	Sig.	Exp(B)	
Etapa 0	Constante	-,011	,146	,005	1	,942	,989

**Variáveis não presentes na equação**

Etapa 0	Variáveis	Pontuação	df	Sig.
	Idade	29,897	1	,000
	Internet(1)	6,852	1	,009
	Amb_Familiar	16,411	3	,001
	Amb_Familiar(1)	10,056	1	,002
	Amb_Familiar(2)	,043	1	,836
	Amb_Familiar(3)	7,357	1	,007
	Univ(1)	15,365	1	,000
	Empenhado	35,312	2	,000
	Empenhado(1)	30,081	1	,000
	Empenhado(2)	35,135	1	,000
	Participativo	15,520	2	,000
	Participativo(1)	1,666	1	,197
	Participativo(2)	9,690	1	,002
	Distraido	6,912	2	,032
	Distraido(1)	,657	1	,418
	Distraido(2)	6,113	1	,013
	Trabalhador	16,798	2	,000
	Trabalhador(1)	14,625	1	,000
	Trabalhador(2)	16,715	1	,000
	DificMat	47,510	2	,000
	DificMat(1)	11,317	1	,001
	DificMat(2)	44,867	1	,000
	PeridEstudo	13,128	4	,011
	PeridEstudo(1)	1,179	1	,278
	PeridEstudo(2)	2,504	1	,114
	PeridEstudo(3)	,051	1	,821
	PeridEstudo(4)	,072	1	,789
	TempoEstudo	17,278	4	,002
	TempoEstudo(1)	1,050	1	,306
	TempoEstudo(2)	5,138	1	,023
	TempoEstudo(3)	3,138	1	,077
	TempoEstudo(4)	,195	1	,659
	CompTurma	11,344	4	,023
	CompTurma(1)	3,099	1	,078
	CompTurma(2)	,913	1	,339
	CompTurma(3)	5,103	1	,024
	CompTurma(4)	,711	1	,399
	CompAluno	20,598	4	,000
	CompAluno(1)	10,130	1	,001
	CompAluno(2)	2,576	1	,109
	CompAluno(3)	13,990	1	,000

CompAluno(4)	,001	1	,982
RelacaoProf	17,614	4	,001
RelacaoProf(1)	9,020	1	,003
RelacaoProf(2)	,300	1	,584
RelacaoProf(3)	7,564	1	,006
RelacaoProf(4)	,019	1	,890
FazTPC	41,394	3	,000
FazTPC(1)	14,694	1	,000
FazTPC(2)	,067	1	,796
FazTPC(3)	28,763	1	,000
Interesse	33,701	4	,000
Interesse(1)	3,727	1	,054
Interesse(2)	21,777	1	,000
Interesse(3)	2,711	1	,100
Interesse(4)	1,190	1	,275
Estatísticas globais	103,378	43	,000

**Bloco 1: Método = Enter**

**Testes de coeficientes de modelo Omnibus**

		Qui-quadrado	df	Sig.
Etapa 1	Etapa	150,267	43	,000
	Bloco	150,267	43	,000
	Modelo	150,267	43	,000

**Resumo do modelo**

Etapa	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke
1	108,965 <sup>a</sup>	,552	,736

a. Estimação finalizada no número de iteração 20 porque o máximo de iterações foi atingido. Não é possível encontrar a solução final.

**Teste de Hosmer e Lemeshow**

Etapa	Qui-quadrado	df	Sig.
1	1,151	8	,997

**Tabela de contingência para teste de Hosmer e Lemeshow**

		Repetência a Matemática Alguma vez = Não		Repetência a Matemática Alguma vez = Sim		Total
		Observado	Esperado	Observado	Esperado	
Etapa 1	1	19	18,895	0	,105	19
	2	18	18,439	1	,561	19
	3	17	17,406	2	1,594	19
	4	16	15,365	3	3,635	19
	5	13	12,133	6	6,867	19
	6	7	7,650	12	11,350	19
	7	3	3,314	16	15,686	19
	8	1	,746	18	18,254	19
	9	0	,052	19	18,948	19
	10	0	,001	16	15,999	16

Variáveis na equação

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. para EXP(B)	
							Inferior	Superior
Etapa 1ª								
Idade	,903	,357	6,398	1	,011	2,467	1,225	4,966
Internet(1)	-,539	2,121	,064	1	,800	,584	,009	37,275
Amb_Familiar			,235	3	,972			
Amb_Familiar(1)	-20,971	16858,683	,000	1	,999	,000	,000	.
Amb_Familiar(2)	-21,338	16858,683	,000	1	,999	,000	,000	.
Amb_Familiar(3)	-20,980	16858,683	,000	1	,999	,000	,000	.
Univ(1)	-,037	,773	,002	1	,961	,963	,212	4,383
Empenhado			8,555	2	,014			
Empenhado(1)	6,551	3,035	4,659	1	,031	699,601	1,827	267898,704
Empenhado(2)	4,801	3,076	2,437	1	,119	121,635	,293	50463,526
Participativo			11,105	2	,004			
Participativo(1)	-4,112	1,477	7,748	1	,005	,016	,001	,296
Participativo(2)	-5,481	1,659	10,918	1	,001	,004	,000	,108
Distraido			,264	2	,876			
Distraido(1)	-,189	,934	,041	1	,840	,828	,133	5,165
Distraido(2)	-,578	1,223	,224	1	,636	,561	,051	6,160
Trabalhador			4,826	2	,090			
Trabalhador(1)	-3,314	2,709	1,496	1	,221	,036	,000	7,355
Trabalhador(2)	-4,543	2,720	2,790	1	,095	,011	,000	2,198
DificMat			7,599	2	,022			
DificMat(1)	1,726	,975	3,130	1	,077	5,616	,830	37,984
DificMat(2)	3,806	1,386	7,543	1	,006	44,983	2,974	680,390
PeridEstudo			3,092	4	,543			
PeridEstudo(1)	-2,172	1,811	1,438	1	,231	,114	,003	3,968
PeridEstudo(2)	-1,741	1,660	1,100	1	,294	,175	,007	4,535
PeridEstudo(3)	-1,292	1,599	,653	1	,419	,275	,012	6,307
PeridEstudo(4)	-,397	1,845	,046	1	,830	,672	,018	25,016
TempoEstudo			2,407	4	,661			
TempoEstudo(1)	-1,051	1,639	,411	1	,521	,350	,014	8,677
TempoEstudo(2)	-1,118	1,632	,469	1	,493	,327	,013	8,015
TempoEstudo(3)	-2,065	2,147	,924	1	,336	,127	,002	8,534
TempoEstudo(4)	1,145	2,369	,234	1	,629	3,143	,030	326,452
CompTurma			8,361	4	,079			
CompTurma(1)	-,155	1,380	,013	1	,910	,856	,057	12,808

CompTurma(2)	-1,629	1,366	1,423	1	,233	,196	,013	2,851
CompTurma(3)	-2,806	1,670	2,824	1	,093	,060	,002	1,595
CompTurma(4)	1,810	2,118	,730	1	,393	6,113	,096	388,423
CompAluno			4,701	4	,319			
CompAluno(1)	4,392	4,759	,852	1	,356	80,776	,007	907701,756
CompAluno(2)	4,462	2,239	3,970	1	,046	86,644	1,076	6978,228
CompAluno(3)	5,140	2,388	4,636	1	,031	170,797	1,586	18396,443
CompAluno(4)	4,528	2,541	3,176	1	,075	92,596	,636	13473,105
RelacaoProf			8,935	4	,063			
RelacaoProf(1)	-,086	2,850	,001	1	,976	,917	,003	244,775
RelacaoProf(2)	-4,712	2,459	3,672	1	,055	,009	,000	1,113
RelacaoProf(3)	-4,590	2,457	3,490	1	,062	,010	,000	1,253
RelacaoProf(4)	-2,807	2,336	1,444	1	,230	,060	,001	5,883
FazTPC			9,171	3	,027			
FazTPC(1)	1,090	1,939	,316	1	,574	2,975	,067	132,890
FazTPC(2)	,511	1,946	,069	1	,793	1,667	,037	75,557
FazTPC(3)	-1,734	2,021	,736	1	,391	,177	,003	9,266
Interesse			3,069	4	,546			
Interesse(1)	-2,056	1,563	1,731	1	,188	,128	,006	2,736
Interesse(2)	-,836	1,421	,346	1	,556	,434	,027	7,018
Interesse(3)	-,644	2,300	,079	1	,779	,525	,006	47,603
Interesse(4)	-,295	1,448	,041	1	,839	,745	,044	12,724
Constante	14,380	16858,684	,000	1	,999	1759423,327		

a. Variável(is) inserida(s) na etapa 1: Idade, Internet, Amb\_Familiar, Univ, Empenhado, Participativo, Distruido, Trabalhador, DificMat, PeridEstudo, TempoEstudo, CompTurma, CompAluno, RelacaoProf, FazTPC, Interesse.



Lista entre casos<sup>b</sup>

Caso	Status selecionado <sup>a</sup>	Observado	Previsto	Grupo previsto	Variável temporária	
		Repetência a Matemática Alguma vez			Resid	ZResid
25	S	N**	,771	S	-,771	-1,834
56	S	S**	,098	N	,902	3,038
59	S	N**	,765	S	-,765	-1,806
64	S	S**	,347	N	,653	1,373
109	S	N**	,882	S	-,882	-2,740
126	S	S**	,182	N	,818	2,118
132	S	S**	,054	N	,946	4,168
137	S	S**	,250	N	,750	1,731
139	S	S**	,226	N	,774	1,852
155	S	S**	,182	N	,818	2,122
178	S	N**	,644	S	-,644	-1,344
183	S	S**	,061	N	,939	3,919
184	S	N**	,581	S	-,581	-1,177
189	S	N	,389	N	-,389	-,798
191	S	N**	,914	S	-,914	-3,252
192	S	N	,351	N	-,351	-,735

a. S = Selecionado, U = Casos não selecionados, e \*\* = Casos classificados incorretamente.

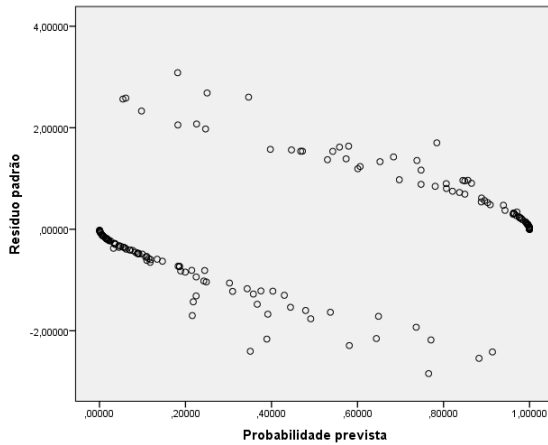
b. São listados os casos com resíduos estudentizados maiores que 2,000.

GRAPH  
/SCATTERPLOT(BIVAR)=PRE\_1 WITH SRE\_1  
/MISSING=LISTWISE

Gráfico

Observações

Saída criada		29-FEB-2016 09:28:45
Comentários		
Entrada	Dados	D:\GoogleDriveNuvem\TESE\Inquéritos Insucesso_CN2roc.sav
	Conjunto de dados ativo	Conjunto_de_dados1
	Filtro	<none>
	Ponderação	<none>
	Arquivo Dividido	<none>
	N de linhas em arquivo de dados de trabalho	199
Sintaxe		GRAPH /SCATTERPLOT(BIVAR)=PRE_1 WITH SRE_1 /MISSING=LISTWISE.
Recursos	Tempo do processador	00:00:02,14
	Tempo decorrido	00:00:03,55



ROC PRE\_1 BY Repetencia2 (1)  
 /PLOT=CURVE(REFERENCE)  
 /PRINT=SE  
 /CRITERIA=CUTOFF(INCLUDE) TESTPOS(LARGE) DISTRIBUTION(FREE) CI(95)  
 /MISSING=EXCLUDE.

**Curva ROC**

**Resumo de processamento do caso**

Repetência a Matemática Alguma vez	N válido (de lista)
Positivo <sup>a</sup>	93
Negativo	94
Ausente	12

Os valores maiores da(s) variável(eis) de resultado de teste indicam uma evidência mais forte de um estado real positivo.

a. O estado real positivo é Sim.

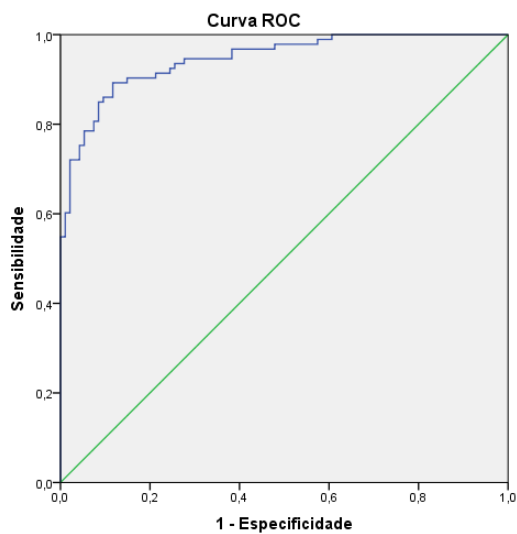
**Área sob a curva**

Variável(eis) de resultado de teste: Probabilidade prevista

Área	Erro Padrão <sup>a</sup>	Sig. assintótico <sup>b</sup>	Intervalo de Confiança 95% Assintótico	
			Limite inferior	Limite superior
,945	,015	,000	,916	,975

a. Sob a suposição não paramétrica

b. Hipótese nula: área real = 0,5



**ANEXO 10 – Outputs do SPSS para Regressão logística com  
5 variáveis independentes**

LOGISTIC REGRESSION VARIABLES Repetencia2  
 /METHOD=ENTER Idade Participativo Empenhado DificMat CompAluno  
 /CONTRAST (Participativo)=Indicator(1)  
 /CONTRAST (Empenhado)=Indicator(1)  
 /CONTRAST (DificMat)=Indicator(1)  
 /CONTRAST (CompAluno)=Indicator(1)  
 /CLASSPLOT  
 /CASEWISE OUTLIER(2)  
 /PRINT=GOODFIT CI(95)  
 /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

**Regressão logística**

**Observações**

**Resumo de processamento do caso**

Casos não ponderados <sup>a</sup>		N	Porcentagem
Casos selecionados	Incluído na análise	184	99,5
	Casos ausentes	1	,5
	Total	185	100,0
Casos não selecionados		0	,0
Total		185	100,0

a. Se a ponderação estiver em vigor, veja a tabela de classificação para o número total de casos.

**Codificação de variável dependente**

Valor original	Valor interno
Não	0
Sim	1

**Codificações de variáveis categóricas**

		Frequência	Codificação de parâmetro			
			(1)	(2)	(3)	(4)
25. Comportamento do aluno	Mau	8	,000	,000	,000	,000
	Pouco Satisfatório	13	1,000	,000	,000	,000
	Satisfatório	62	,000	1,000	,000	,000
	Bom	85	,000	,000	1,000	,000
19.3 Aluno empenhado	Muito Bom	16	,000	,000	,000	1,000
	Não	5	,000	,000		
	Às vezes	90	1,000	,000		
20. Dificuldades a matemática	Sim	89	,000	1,000		
	Não	32	,000	,000		
	Às vezes, depende da matéria	111	1,000	,000		
19.2 Aluno participativo	Sim, bastantes	41	,000	1,000		
	Não	15	,000	,000		
	Às vezes	108	1,000	,000		
	Sim	61	,000	1,000		

**Bloco 0: Bloco Inicial**

**Tabela de Classificação<sup>a,b</sup>**

		Previsto			Porcentagem correta
		Repetência a Matemática Alguma vez			
Observado		Não	Sim		
Etapa 0	Repetência a Matemática Alguma vez Não	94	0	100,0	
	Repetência a Matemática Alguma vez Sim	90	0	,0	
Porcentagem global				51,1	

a. A constante está incluída no modelo.

b. O valor de recorte é ,500

**Variáveis na equação**

	B	S.E.	Wald	df	Sig.	Exp(B)
Etapa 0 Constante	-,043	,147	,087	1	,768	,957

**Variáveis não presentes na equação**

			Pontuação	df	Sig.
Etapa 0	Variáveis	Idade	26,133	1	,000
		Participativo	15,550	2	,000
		Participativo(1)	1,563	1	,211
		Participativo(2)	9,497	1	,002
		Empenhado	33,461	2	,000
		Empenhado(1)	28,132	1	,000
		Empenhado(2)	33,227	1	,000
		DificMat	49,800	2	,000
		DificMat(1)	7,849	1	,005
		DificMat(2)	45,080	1	,000
		CompAluno	17,368	4	,002
		CompAluno(1)	7,135	1	,008
		CompAluno(2)	3,134	1	,077
		CompAluno(3)	11,726	1	,001
		CompAluno(4)	,187	1	,665
Estatísticas globais			79,294	11	,000

**Bloco 1: Método = Enter**

**Testes de coeficientes de modelo Omnibus**

		Qui-quadrado	df	Sig.
Etapa 1	Etapa	101,716	11	,000
	Bloco	101,716	11	,000
	Modelo	101,716	11	,000

**Resumo do modelo**

Etapa	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke
1	153,275 <sup>a</sup>	,425	,566

a. Estimação finalizada no número de iteração 6 porque as estimativas de parâmetro mudaram foram alteradas para menos de ,001.

**Teste de Hosmer e Lemeshow**

Etapa	Qui-quadrado	df	Sig.
1	11,571	8	,171

**Tabela de Classificação<sup>a</sup>**

	Observado	Previsto			
		Repetência a Matemática Alguma vez		Porcentagem correta	
		Não	Sim		
Etapa 1	Repetência a Matemática Alguma vez	Não	77	17	81,9
		Sim	19	71	78,9
Porcentagem global					80,4

a. O valor de recorte é ,500



**Lista entre casos<sup>b</sup>**

Caso	Status selecionado <sup>a</sup>	Observado	Previsto	Grupo previsto	Variável temporária	
		Repetência a Matemática Alguma vez			Resid	ZResid
38	S	S**	,104	N	,896	2,932
62	S	S**	,104	N	,896	2,932
145	S	S**	,083	N	,917	3,323
177	S	N**	,983	S	-,983	-7,592

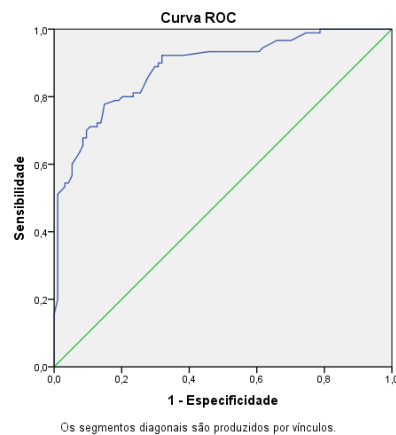
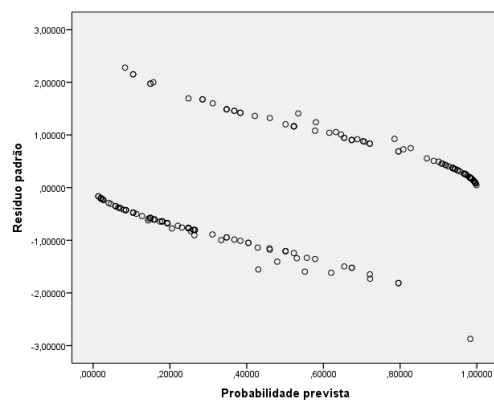
a. S = Selecionado, U = Casos não selecionados, e \*\* = Casos classificados incorretamente.

b. São listados os casos com resíduos estudentizados maiores que 2,000.

**GRAPH**

/SCATTERPLOT(BIVAR)=PRE\_1 WITH SRE\_1  
/MISSING=LISTWISE.

**Gráfico**



ROC PRE\_1 BY Repetencia2 (1)

/PLOT=CURVE(REFERENCE)

/PRINT=SE

/CRITERIA=CUTOFF(INCLUDE) TESTPOS(LARGE) DISTRIBUTION(FREE) CI(95)

/MISSING=EXCLUDE

**Curva ROC**

**Resumo de processamento do caso**

Repetência a Matemática Alguma vez	N válido (de lista)
Positivo <sup>a</sup>	90
Negativo	94
Ausente	1

Os valores maiores da(s) variável(eis) de resultado de teste indicam uma evidência mais forte de um estado real positivo.

a. O estado real positivo é Sim.

**Área sob a curva**

Variável(eis) de resultado de teste: Probabilidade prevista

Área	Erro Padrão <sup>a</sup>	Sig. assintótico <sup>b</sup>	Intervalo de Confiança 95% Assintótico	
			Limite inferior	Limite superior
,885	,024	,000	,838	,933

A variável ou variáveis de resultado de teste: Probabilidade prevista possuem pelo menos um vínculo entre o grupo de estado real positivo e o grupo de estado real negativo. As estatísticas podem ser enviesadas.

a. Sob a suposição não paramétrica

b. Hipótese nula: área real = 0,5

**ANEXO 11 – AGRUPAMENTO DUAS ETAPAS**

## Agrupamento duas etapas

Cluster	2	1	3
Entradas	FApoioFAMIL Importância=0,74 Categoria mais frequente: Não concordo nem discordo (41,8%)	Rapidez Importância=1,00 Categoria mais frequente: Concordo totalmente (45,1%)	DificInterp Importância=0,99 Categoria mais frequente: Discordo totalmente (37,5%)
	Rapidez Importância=1,00 Categoria mais frequente: Não concordo nem discordo (40,7%)	FApoioProf Importância=0,99 Categoria mais frequente: Concordo totalmente (36,6)	FaltaHabTrab Importância=0,62 Categoria mais frequente: Concordo totalmente (43,8%)
	Estratégias Importância=0,84 Categoria mais frequente: Não concordo nem discordo (57,1%)	FEsclarecimento Importância=0,82 Categoria mais frequente: Concordo em algum grau (39,4%)	FApoioProf Importância=0,99 Categoria mais frequente: Discordo totalmente (100%)
	FApoioProf Importância=0,99 Categoria mais frequente: Não concordo nem discordo (38,5%)	DificInterp Importância=0,99 Categoria mais frequente: Concordo totalmente (62%)	Rapidez Importância=1,00 Categoria mais frequente: Discordo totalmente (81,2%)
	DificInterp Importância=0,99 Categoria mais frequente: Não concordo nem discordo (36,3%)	FApoioFAMIL Importância=0,74 Categoria mais frequente: Concordo totalmente (33,8%)	Estratégias Importância=0,84 Categoria mais frequente: Discordo totalmente (43,8%)
	FEsclarecimento Importância=0,82 Categoria mais frequente: Não concordo nem discordo (51,5%)	Estratégias Importância=0,84 Categoria mais frequente: Concordo em algum grau (38,0%)	FEsclarecimento Importância=0,82 Categoria mais frequente: Discordo totalmente (87,5%)
	FaltaReq Importância=0,59 Categoria mais frequente: Não concordo nem discordo (47,3%)	Carga Horária Importância=0,37 Categoria mais frequente: Concordo totalmente (49,3%)	FaltaReq Importância=0,59 Categoria mais frequente: Discordo totalmente (56,2%)
	Conteúdos Importância=0,53 Categoria mais frequente: Não concordo nem discordo (53,8%)	Conteúdos Importância=0,53 Categoria mais frequente: Concordo em algum grau (31,0%)	Conteúdos Importância=0,53 Categoria mais frequente: Discordo totalmente (56,2%)
	MenosAulas Importância=0,34 Categoria mais frequente: Não concordo nem discordo (28,6%)	FaltaReq Importância=0,59 Categoria mais frequente: Concordo em algum grau (32,4%)	FApoioFAMIL Importância=0,74 Categoria mais frequente: Discordo totalmente (87,5%)
	CargaHorária Importância=0,37 Categoria mais frequente: Não concordo nem discordo (48,4%)	MenosAulas Importância=0,34 Categoria mais frequente: Concordo totalmente (52,1%)	FaltaAtenção Importância=0,36 Categoria mais frequente: Concordo totalmente (43,8%)

Cluster	2	1	3
Entradas	AulasApoio Importância=0,27 Categoria mais frequente: Não concordo nem discordo (49,5%)	TurmaPequena Importância=0,31 Categoria mais frequente: Concordo totalmente (36,6%)	Indisciplina Importância=0,33 Categoria mais frequente: Discordo totalmente (43,8%)
	Tecnologias Importância=0,24 Categoria mais frequente: Não concordo nem discordo (35,2%)	FaltaAtenção Importância=0,36 Categoria mais frequente: Concordo totalmente (57,7%)	MaisAulas Importância=0,25 Categoria mais frequente: Discordo totalmente (62,5%)
	FaltaHabTrab Importância=0,62 Categoria mais frequente: Concordo em algum grau (38,5%)	Desinteresse Importância=0,29 Categoria mais frequente: Concordo totalmente (52,1%)	TurmaPequena Importância=0,31 Categoria mais frequente: Discordo totalmente (81,2%)
	VidaDiária Importância=0,31 Categoria mais frequente: Não concordo nem discordo (57,1%)	VidaDiária Importância=0,31 Categoria mais frequente: Não concordo nem discordo (36,6%)	VidaDiária Importância=0,31 Categoria mais frequente: Não concordo nem discordo (31,2%)
	Desinteresse Importância=0,29 Categoria mais frequente: Concordo em algum grau (38,5%)	Indisciplina Importância=0,33 Categoria mais frequente: Concordo em algum grau (33,8%)	Manual Importância=0,24 Categoria mais frequente: Discordo totalmente (81,2%)
	Indisciplina Importância=0,33 Categoria mais frequente: Não concordo nem discordo (41,8)	FaltaHabTrab Importância=0,62 Categoria mais frequente: Concordo totalmente (52,1%)	Desinteresse Importância=0,29 Categoria mais frequente: Concordo totalmente (50,0%)
	MaisAulas Importância=0,25 Categoria mais frequente: Discordo totalmente (40,7%)	Manual Importância=0,24 Categoria mais frequente: Não concordo nem discordo (28,2%)	CargaHorária Importância=0,37 Categoria mais frequente: Discordo totalmente (43,8%)
	FaltaAtenção Importância=0,36 Categoria mais frequente: Concordo em algum grau (37,4%)	Tecnologias Importância=0,24 Categoria mais frequente: Concordo totalmente (49,3%)	AulasApoio Importância=0,27 Categoria mais frequente: Concordo totalmente (50,0%)
	Manual Importância=0,24 Categoria mais frequente: Não concordo nem discordo (37,4%)	AulasApoio Importância=0,27 Categoria mais frequente: Concordo totalmente (33,8%)	Tecnologias Importância=0,24 Categoria mais frequente: Concordo totalmente (62,5%)
	TurmaPequena Importância=0,31 Categoria mais frequente: Não concordo nem discordo (36,3%)	MaisAulas Importância=0,25 Categoria mais frequente: Discordo totalmente (52,1%)	MenosAulas Importância=0,34 Categoria mais frequente: Discordo totalmente (50,0%)

