

Instituto Superior de Ciências do Trabalho e da Empresa



UMA ANÁLISE MULTINÍVEL DO EFEITO DA
HETEROGENEIDADE DAS ESCOLAS SECUNDÁRIAS NO
1º ANO DE ENGENHARIA: UM ESTUDO DE CASO

Carla Cristina Augusto Patrocínio

Tese submetida como requisito parcial para obtenção do grau de
Mestre em Prospecção e Análise de Dados

Orientador:

Prof. Doutor José Gonçalves Dias, Prof. Auxiliar,
Departamento de Métodos Quantitativos, ISCTE Business School

Co-Orientador:

Prof. Doutor Eduardo Manuel Baptista Ribeiro Pereira, Prof. Associado,
Departamento de Engenharia Civil e Arquitectura, Instituto Superior Técnico

Setembro de 2008

Agradecimentos

A elaboração desta tese foi um percurso longo, constituído por várias etapas, qualquer uma exigindo extensas horas de estudo, reflexão e um grande sentido crítico. Na concretização deste objectivo tive o prazer de contar com um grupo de pessoas que se destacaram nas mais variadas áreas e a quem desejo expressar os meus sinceros agradecimentos:

Ao meu orientador, Prof. José Dias, pela sua capacidade de transmissão de saberes fulcrais para a realização desta tese, pela sua direcção e apoio na superação dos diversos obstáculos, e, sobretudo, pela sua aptidão na promoção de longas e produtivas discussões durante a elaboração da dissertação.

Ao meu co-orientador, Prof. Eduardo Pereira, pelo seu incansável encorajamento à realização deste projecto, pelos seus sábios conselhos, recomendações e contagioso entusiasmo na discussão desta temática, e, especialmente, pelo seu espírito crítico e capacidade de analisar um problema por vários ângulos.

A ambos, que conquistaram a minha profunda admiração ao longo deste projecto, obrigada por terem aceite a incumbência de (co)orientar este trabalho, pela vossa valiosa disponibilidade, compreensão, críticas e sugestões feitas durante estes dois anos.

Ao Instituto Superior Técnico e ao Gabinete de Estatística e Planeamento da Educação do Ministério da Educação, por terem facultado os dados e outros documentos necessários à concretização deste estudo de caso.

Aos amigos e colegas que me ajudaram, deram apoio e tiveram paciência para me ouvir e discutir comigo as ideias aqui concretizadas. Um agradecimento especial às minhas amigas Isabel, Marta, Sofia e Margarida, pelo seu calor humano e incansável apoio moral, e à minha coordenadora, Marta Pile, pelo seu interesse e apoio neste trabalho.

À minha família, em particular à minha querida irmã Maria João, pelo seu apoio, compreensão e carinho quando estava a trabalhar neste projecto e não tinha tempo para mais nada.

Finalmente, à pessoa que mais me incentivou e apoiou a iniciar e a concretizar este projecto, o meu marido. Não há palavras que consigam expressar o meu agradecimento pelo seu apoio incondicional e sem reservas, e que, mesmo quando as maiores adversidades surgiram, me deu a coragem e a confiança necessária para continuar.

Resumo

O insucesso escolar no Ensino Superior é hoje uma realidade bem conhecida no panorama educativo nacional. Os factores que estarão na origem desta realidade são diversos e nem sempre passíveis de mensurar, seja pelo seu carácter de excepção, seja pela sua natureza. Uns (factores) serão conhecidos no momento de ingresso na universidade, podendo ser intrínsecos ao aluno, outros resultarão da frequência do Ensino Superior.

Este trabalho incide, em particular, sobre a problemática do insucesso no 1º ano, nos cursos da área da Ciência, Tecnologia e Engenharia, tendo por base o manancial de informação conhecido *a priori* sobre o aluno, a par com elementos do seu percurso anterior, evidenciando-se a influência da escola secundária frequentada. O objectivo genérico é aferir o contributo da conjugação de alguns destes factores para os processos de construção do sucesso do aluno enquanto estudante do Ensino Superior.

Para tal é usada uma modelação multinível, considerando dois níveis de análise. Um primeiro nível que inclui características intrínsecas ao aluno e, um segundo nível, com uma caracterização de segmentos de escolas secundárias de proveniência dos alunos.

Palavras-chave: Modelos Misturas Finitas, Modelos Multinível, Insucesso Escolar, Ensino Secundário, Ensino Superior, Engenharia.

JEL Classification: I21, C21

Abstract

Nowadays school failure in Higher Education is a well-known reality in the Portuguese educational system. The factors underlying this reality are numerous and not always measurable, either due to their exceptional quality or nature. Some factors may be known at the moment the student is admitted to the University, either being intrinsic to the student himself or derived from the student's path in Higher Education.

This research focuses, in particular, on the problem of failure in the 1st year, in the Science, Technology and Engineering courses, on the basis of an a priori known array of information on the student, alongside with elements of their previous path, of which the influence of secondary school should be highlighted. The overall purpose is to determine how these factors together contribute to the construction of students' success processes as Higher Education students.

To attain this, a multilevel model is used, considering two levels of analysis. The first includes characteristics that are intrinsic to the student, and a second level that involves characteristics of secondary school sectors from where students come.

Keywords: Finite Mixture Models, Multilevel Models, School Failure, High School, Higher Education, Engineering

JEL Classification: I21, C21

Sumário Executivo

O insucesso escolar no Ensino Superior é hoje uma realidade bem conhecida no panorama educativo nacional. Os factores que estarão na origem desta realidade são diversos e nem sempre passíveis de mensurar, seja pelo seu carácter de excepção, seja pela sua natureza. Uns (factores) serão conhecidos no momento de ingresso na universidade, podendo ser intrínsecos ao aluno, outros resultarão da frequência do Ensino Superior.

Este trabalho incide, em particular, sobre a problemática do insucesso no 1º ano, nos cursos da área da Ciência, Tecnologia e Engenharia, tendo por base o manancial de informação conhecido *a priori* sobre o aluno, a par com elementos do seu percurso anterior, evidenciando-se a influência da escola secundária frequentada. O objectivo genérico é aferir o contributo da conjugação de alguns destes factores para os processos de construção do sucesso do aluno enquanto estudante do Ensino Superior.

Para tal é usada uma modelação multinível, considerando dois níveis de análise. Um primeiro nível que inclui características intrínsecas ao aluno, e um segundo nível com uma caracterização de segmentos de escolas secundárias de proveniência dos alunos.

O vasto trabalho de investigação desenvolvido na área de discussão deste trabalho revela muitas directrizes sobre o caminho a percorrer, especialmente no sentido de completar a compreensão do fenómeno ou dos factores que poderão contribuir para o mesmo. Com a presente investigação pretende-se ir um passo mais adiante da já desenvolvida a nível nacional, na medida em que se contemplam indicadores relacionados com a escola secundária que o aluno frequentou aquando do seu percurso escolar anterior.

É importante referir que informação desta natureza ao dispor dos investigadores é relativamente recente. De facto, só a partir de 2002 a Direcção-Geral de Inovação e de Desenvolvimento Curricular¹ (Ministério da Educação) começou a disponibilizar tal informação.

¹ <http://www.dgidec.min-edu.pt/jneweb/estat.htm>, consultado em 10/10/2006.

O aprofundamento do conhecimento desta temática justifica-se não só como ferramenta de apoio a processos de intervenção e promoção do sucesso escolar, como pela sua relevância em potenciais modificações a introduzir no Ensino Superior como forma de colmatar o problema. Uma ferramenta essencial para programas de intervenção precoce é um modelo de predição dos resultados do 1º ano do Ensino Superior, visto que só desta forma é possível identificar os estudantes que potencialmente deverão ser abrangidos por tais programas.

Os dados que são usados para a experimentação em hipótese foram facultados por uma Instituição de Ensino Superior particularmente vocacionada para o ensino de Engenharia, Ciência e Tecnologia e referem-se ao ano lectivo de 2004/05.

Este trabalho contemplará um enquadramento da problemática em estudo, uma análise dos estabelecimentos de Ensino Secundário, mais concretamente do processo de construção dos perfis-tipo de escolas secundárias que possam contribuir para a explicação do fenómeno em causa, e uma análise multinível do percurso escolar entre o Ensino Secundário e o Superior. Finalmente serão apresentadas as principais conclusões, a descrição das limitações encontradas neste estudo e potenciais linhas de investigação futura.

Índice

1.	Introdução	1
1.1	A problemática do sucesso escolar no Ensino Superior	2
1.2	Estrutura da tese.....	7
2.	Caracterização das escolas secundárias	9
2.1	Objectivo.....	11
2.2	Fontes de informação.....	13
2.3	Dimensões de análise.....	13
2.4	Análise descritiva.....	17
2.5	Metodologia: modelos de mistura finita com variáveis concomitantes.....	24
2.6	Estimação e selecção do modelo	29
2.6.1	Descrição dos dados utilizados	29
2.6.2	Seleccção do número de segmentos	29
2.6.3	Diagnóstico dos resíduos bivariados.....	31
2.7	Perfis de escolas secundárias	32
2.8	Análise complementar	36
2.9	Conclusões e notas finais.....	38
3.	Percurso escolar entre o Ensino Secundário e o Superior: uma análise multinível	41
3.1	Modelo conceptual.....	41
3.2	Metodologia.....	43
3.2.1	Modelos multinível lineares.....	43
3.2.2	Modelos multinível binomiais	48
3.3	Caracterização do objecto empírico.....	50
3.3.1	Fontes de informação.....	53
3.3.2	Medição do sucesso escolar	53
3.3.3	Dimensões preditivas do sucesso escolar	56
3.4	Estimação do modelo.....	62
3.5	Discussão dos resultados	66
4.	Conclusões	71
4.1	Limitações do estudo	72
4.2	Investigação futura.....	74
5.	Bibliografia	77

Lista de Figuras

Figura 1: Questões de investigação.....	7
Figura 2: Dimensões em análise.....	14
Figura 3: Distribuição dos estabelecimentos de Ensino Secundário (Público e Privado) em Portugal Continental em 2004.....	18
Figura 4: Distribuição das freguesias do continente por tipologia.....	19
Figura 5: Distribuição das escolas secundárias de acordo com a tipologia das freguesias a que pertencem.....	20
Figura 6: Distribuição das escolas por natureza.....	20
Figura 7: Histograma da variável percentagem de matriculados em cursos gerais.	21
Figura 8: Distribuição do número de escolas por média na disciplina de Matemática. ...	22
Figura 9: Distribuição da percentagem de docentes com menos de 40 anos.	23
Figura 10: Algoritmo EM.....	27
Figura 11: Resultados dos critérios de informação.	30
Figura 12: Distribuição escolas por segmento.	32
Figura 13: Medidas de discriminação.	37
Figura 14: Projecção das categorias em estudo.....	38
Figura 15: Perfis de escolas secundárias.....	39
Figura 16: Modelo conceptual.....	43
Figura 17: Distribuição da variável rendimento académico (Taxa de aprovação).....	56
Figura 18: Eixos dimensionais em análise no nível do aluno.	57
Figura 19: Distribuição das classificações dos exames de Matemática e dos exames contabilizados como prova de ingresso para ingresso no IST.	67

Lista de Quadros

Tabela 1: Descrição das variáveis incluídas na modelação por misturas finitas (\mathbf{u}).	29
Tabela 2: Resíduos bivariados (variáveis).	31
Tabela 3: Resíduos bivariados (covariáveis).	31
Tabela 4: Estimativas para as variáveis dependentes.	33
Tabela 5: Estimativas para as variáveis concomitantes.	33
Tabela 6: Caracterização das variáveis dependentes por segmento.	34
Tabela 7: Caracterização das variáveis concomitantes por segmento.	34
Tabela 8: Variáveis codificadas, e respectivas categorias, para a ACM.	36
Tabela 9: Descrição das variáveis incluídas na análise multinível – nível do aluno.	60
Tabela 10: N° e proporção de alunos ingressados do IST em 2004/05 por segmento.	61
Tabela 11: Descrição das variáveis incluídas na análise multinível – nível da escola. ...	61
Tabela 12: Resumo dos modelos testados.	63
Tabela 13: Estatísticas de ajustamento (<i>fit statistics</i>) dos modelos binomiais standard.65	
Tabela 14: Estatísticas de ajustamento (<i>fit statistics</i>) dos modelos binomiais multinível.66	

1. INTRODUÇÃO

O insucesso escolar no Ensino Superior é hoje uma realidade bem conhecida no panorama educativo nacional: a taxa de insucesso em 2004/05 rondou os 35% no Ensino Superior Público (GPEAR, 2007). Os factores que estarão na origem desta realidade são diversos e nem sempre passíveis de quantificar, seja pelo seu carácter de excepção, seja pela sua natureza. Uns (factores) serão conhecidos no momento de ingresso na universidade, sendo intrínsecos ao aluno, outros estarão relacionados com a frequência do Ensino Superior (Almeida *et al.*, 2004; Martins e Campos, 2005; Nasser e Hagtvet, 2006; Pike e Saupe, 2002; Taveira, 2000).

Este trabalho incidirá, em particular, sobre a problemática do sucesso no 1º ano nos cursos da área da Ciência e Engenharia. Tendo por base o manancial de informação conhecido *a priori* sobre o aluno, incluindo e evidenciando a influência da escola secundária frequentada. O objectivo genérico é aferir o contributo da conjugação de alguns destes factores para os processos de construção do sucesso do aluno enquanto estudante do Ensino Superior.

Na concretização deste objectivo é usada uma modelação multinível, considerando dois níveis de análise. Um primeiro nível que inclui características intrínsecas ao aluno, e um segundo nível que inclui características de segmentos de escolas secundárias de proveniência dos alunos.

O trabalho de investigação desenvolvido na área de discussão deste trabalho revela muitas directrizes sobre o caminho a percorrer, especialmente no sentido de completar a compreensão do fenómeno ou dos factores que poderão contribuir para o mesmo. Com o presente trabalho pretende-se contribuir para a discussão deste tema através da análise de um caso empírico, incorporando-se um elemento inovador na investigação levada a cabo a nível nacional na medida em que se contemplam indicadores relacionados com a escola secundária que o aluno frequentou². Do ponto de vista metodológico esta investigação também se destaca pela utilização de um modelo hierárquico binomial

² É importante referir que a disponibilização de informação desta natureza é relativamente recente (só a partir de 2002 a Direcção-Geral de Inovação e de Desenvolvimento Curricular/Ministério da Educação começou a disponibilizar tal informação).

aplicado ao rendimento escolar do aluno medido pelo número de disciplinas concluídas face às que se inscreveu.

O aprofundamento desta temática justifica-se não só como ferramenta de apoio a processos de intervenção e promoção do sucesso escolar, como pela sua relevância em potenciais modificações a introduzir nos estabelecimentos de Ensino Superior. Uma ferramenta que se poderia revelar como essencial para programas de intervenção seria um modelo de predição dos resultados do 1º ano no Ensino Superior, visto que só desta forma se conseguem identificar precocemente os estudantes que deverão ser abrangidos por tais programas (Pike e Saupe, 2002).

1.1 A problemática do sucesso escolar no Ensino Superior

O estudo da problemática do sucesso deverá começar pela sua definição: como se mede o sucesso escolar e como se classifica um aluno como sendo um aluno de sucesso?

Aprofundando esta problemática, outras questões se levantam: Serão estes fenómenos aleatórios? Dependerão exclusivamente da capacidade do estudante? Da sua experiência académica anterior? Será que o *background* de um aluno se reflecte na sua capacidade de ultrapassar as adversidades com que se depara quando entra na faculdade? Será uma conjugação de todos estes factores e de outros não passíveis de quantificar?

Obviamente que muitos dos factores que poderão estar subjacentes a estes fenómenos não são passíveis de ser mensuráveis, mas tal dúvida leva-nos a outra questão: o que é de facto mensurável? Será que é possível prever este fenómeno, mediante o conhecimento das características de cada aluno? Será que as características conhecidas num aluno à entrada para o Ensino Superior poderão ajudar a detectar precocemente as dificuldades com que este se deparará na universidade, ou será que o contexto no qual ele se insere após o ingresso no Ensino Superior é determinante para o seu sucesso?

O insucesso escolar é um problema actual, mas não novo, sendo importante o seu diagnóstico, compreensão e predição com vista ao seu combate e desmistificação. No contexto português o combate ao insucesso tem sido desenvolvido em várias frentes por políticos, gestores, investigadores e professores, preocupados com o impacto que problemas como a baixa taxa de escolarização e o insucesso académico poderão afectar a sociedade em geral (Taveira, 2000).

A problemática do sucesso escolar no Ensino Superior tem sido alvo de estudos e trabalho de investigação há mais de um século (Pike e Saupe, 2002). No contexto nacional esta temática é relativamente recente, o flagelo social que é o insucesso escolar despertou os investigadores portugueses para esta temática há cerca de 2 décadas (Alarcão, 2000), ainda que os projectos de diagnóstico e intervenção desenvolvidos em torno de problemas desta natureza sejam mais recentes. Um dos primeiros projectos sobre esta temática, desenvolvido por José Tavares da Universidade de Aveiro no âmbito do programa Praxis XXI, intitulou-se “Factores de sucesso/insucesso no 1º ano das licenciaturas em Ciências e Engenharia do Ensino Superior” (Taveira, 2000).

A consciencialização da sociedade em geral, mas sobretudo dos actores das instituições portuguesas de Ensino Superior, para esta problemática resultou das profundas alterações observadas no panorama do Ensino Superior Português nos últimos 30 anos, destacando-se em particular a massificação e a diversidade da população estudantil que tem acedido a este grau de ensino. Estas mudanças trouxeram ao contexto do Ensino Superior, um sector extremamente importante na constituição da sociedade do futuro pelas suas múltiplas responsabilidades (Taveira, 2000), uma população diferenciada e também índices de insucesso mais elevados.

O facto deste fenómeno só tão recentemente ter sido objecto de estudo aprofundado poderá também estar relacionado com a falta de séries estatísticas que atestassem a dimensão do problema. Na realidade, até há relativamente pouco tempo (10-20 anos), não existiam indicadores disponíveis que traduzissem a realidade observada (e.g. as séries estatísticas disponibilizadas quer ao nível do Ensino Secundário³, quer ao nível do Ensino Superior⁴, têm o seu início em 1997/98). A evolução das novas tecnologias de informação e a sua utilização ao serviço da sociedade para a gestão da informação deram (Missão para a Sociedade da Informação, 1997), certamente, um contributo significativo para a obtenção de indicadores que sustentem e apoiem a identificação do problema ao nível universitário.

No contexto da problemática em discussão têm-se desenvolvido vários trabalhos neste contexto, sobretudo nas áreas da Psicologia e das Ciências da Educação, sendo exemplo

³ <http://www.dgidec.min-edu.pt/estatistica/estatistica.asp>, consultado em 10/06/2006.

⁴ <http://www.estatisticas.gpeari.mctes.pt/>, consultado em 15/07/2008.

disso a publicação “Ensino superior - (in)sucesso académico” (Tavares e Santiago, 2000; Nóvoa e *et al.*, 2005). Na mesma obra é analisada esta problemática mediante várias dimensões: científica, pedagógica, organizacional e de gestão. Esta publicação, datada de 2000, colige um conjunto de trabalhos desenvolvidos no âmbito nacional por investigadores portugueses, nos campos de diagnóstico, reflexão, intervenção e meta-análise das questões do sucesso e insucesso escolar. Deste trabalho evidenciam-se quatro grandes vectores que contribuem para a promoção do sucesso académico no Ensino Superior: aluno, professor, currículo académico e instituição (Alarcão, 2000).

Grande parte dos trabalhos desenvolvidos tem por base dados sociográficos (exemplos: sexo, idade, habilitações e escolaridade dos pais, se o aluno se encontra ou não na residência habitual ou tempo que demora no percurso para a mesma), o percurso passado do aluno (exemplo: nota de seriação⁵, que reflecte não só o desempenho do aluno no Ensino Secundário como também o seu desempenho nas provas de ingresso), a sua motivação para o par curso/instituição em que ingressou, traduzido pela prioridade com que o aluno foi colocado (vulgarmente conhecida como opção/preferência de colocação) (Almeida *et al.*, 2004, Lencastre, *et al.*, 2000, Soares *et al.*, 2006). Outros factores têm também sido tomados nas modelações, nomeadamente factores relacionados com a satisfação, as expectativas, as vivências académicas, a capacidade de socialização e a perspectiva dos alunos perante os métodos de estudo e a auto-regulação (Alarcão, 2000; Almeida *et al.*, 2004).

No trabalho de investigação desenvolvido em Portugal sobre a temática em estudo, verifica-se que uma das áreas mais estudadas é, sem dúvida, a dos cursos de Ciências e Engenharia, especialmente pelo insucesso académico verificado nesta área (Alarcão, 2000). Tal facto é facilmente perceptível visto que subjacente a estes cursos está a disciplina de Matemática onde se verificam taxas de reprovação bastante elevadas no final do Ensino Secundário (ver, a título exemplificativo, os resultados dos exames de Matemática no final do Ensino Secundário⁶).

⁵ A nota de seriação é a classificação com que o aluno é seriado para o ingresso no Ensino Superior e baseia-se na classificação obtida no Ensino Secundário, nos resultados das provas exigidas para o acesso ao curso pretendido (variam consoante os cursos a que se candidatam), e, eventualmente, em pré-requisitos (consoante os cursos a que se candidatam).

⁶ <http://sitio.dgicd.min-edu.pt/JNE/Paginas/relatorio.aspx>, consultado em 10/06/2006.

Este alvo justifica-se ainda por se verificar que as variáveis que contribuem para a explicação do fenómeno do insucesso são diferentes de área para área (ou têm pesos diferentes). No caso da área de Ciência e Tecnologia, o percurso anterior, sobretudo a nota de seriação, revela capacidade explicativa no rendimento académico, mais do que variáveis relacionadas com as vivências académicas e/ou expectativas/comportamentos de envolvimento e ainda os métodos de estudo (Almeida *et al.*, 2004; Soares *et al.*, 2006).

Adicionalmente, alguns trabalhos desenvolvidos no contexto nacional (e.g. Almeida *et al.*, 2004; Soares *et al.*, 2006) revelam que os métodos e hábitos de estudo são igualmente factores determinantes para o processo de transição, como tal é legítimo assumir que este legado pode ter sido, em parte, dado pela formação anterior assim como pelo contexto educativo em que este decorreu (escola secundária onde obteve a formação anterior).

A nível internacional este fenómeno tem sido modelado não só pela introdução destes factores mas também pela inclusão de outro tipo de elementos; relacionados com o percurso anterior do aluno (Amelink, 2005; Windham, 2005), em particular a escola onde o aluno fez o Ensino Secundário (Pike e Saupe, 2002); relacionados com o enquadramento institucional da universidade onde ingressam (Porter, 2006; Amelink, 2005; Bailey *et al.*, 2006; Montmarquette *et al.*, 2000); ou ainda factores relacionados com carga horária e o grau de dificuldade do curso onde ingressaram (Szafran, 2001). Note-se que alguns destes trabalhos não reflectem a realidade portuguesa na medida em que o acesso ao Ensino Superior é feito mediante um processo diferente.

Independentemente do contexto onde é desenvolvido, é legítimo aceitar que o processo de construção desenvolvido pelo aluno ao longo do Ensino Secundário irá influenciar o seu percurso no Ensino Superior e tomá-lo como hipótese de trabalho. Existem algumas variáveis contempladas em estudos internacionais que, ou não fazem sentido (diferentes processos de recrutamento), ou não existem no contexto nacional (diferentes modelos de selecção dos estudantes); devendo as mesmas ser substituídas por outras que, no âmbito nacional, façam sentido e que se possam revelar promissoras na explicação do fenómeno do sucesso.

A temática tem, assim, sido alvo de estudo por várias áreas do saber e modelada de várias formas, desde a sua caracterização por simples análises descritivas, relacionais e de variância, passando pela regressão linear, pela modelação por equações estruturais e

pela sua modelação através de modelos multinível. Exemplo desta última metodologia de modelação é um trabalho desenvolvido por Pike e Saupe (2002) a propósito da predição do desempenho escolar dos estudantes no seu primeiro ano no Ensino Superior, no qual foram desenvolvidos três tipos de modelação: regressão linear múltipla; modelo de efeitos fixos (*high school effects model*); e modelo hierárquico linear. Um aspecto que todos os trabalhos referidos têm em comum é a incorporação das classificações do 1º ano, e não apenas a relação entre o que o aluno se propôs fazer face ao que fez efectivamente, independentemente da classificação com que o fez.

Na revisão de literatura levada a cabo por Pike e Saupe (2002), os modelos que incluem o desempenho no Ensino Secundário a par com resultados em exames de admissão (ACT) revelaram uma capacidade explicativa de 25% a 30% da variação das classificações do 1º ano e que a incorporação de medidas de análise do Ensino Secundário melhoraram a capacidade preditiva. Também em Tai *et al.* (2006), é evidenciada a capacidade preditiva dos resultados do Ensino Secundário, numa ordem grandeza semelhante.

Todo o enquadramento do sucesso escolar é, como já foi referido várias vezes, complexo e pode ser não só analisado em várias áreas do saber, como também de vários prismas. O sucesso é um fenómeno de natureza multifacetada, na medida em que nos remete para uma noção mais abrangente que o mero sucesso académico, e subjectiva, na medida em que existem diversas formas de percepcionar o sucesso, dependendo muitas vezes dos objectivos e ponto de vista do aluno (Lencastre *et al.*, 2000).

Com a preocupação de combater o insucesso e o abandono escolar, a par do intuito de apoiar esta investigação, uma instituição de Ensino Superior – o Instituto Superior Técnico acordou em facultar a informação sobre uma geração de alunos ingressados no seu estabelecimento (ano lectivo de 2004/05). Esta instituição, particularmente vocacionada para o ensino de Engenharia, Ciência e Tecnologia, tem actualmente alguns mecanismos de apoio à integração dos alunos através do programa de tutorado⁷.

Todavia, actualmente, só é possível identificar os alunos em situação de risco no fim do primeiro semestre, nomeadamente quando se conhecem os primeiros resultados do seu

⁷ Para mais informações consultar <http://tutorado.ist.utl.pt/>, consultado em 16/07/2008.

desempenho nas disciplinas que funcionaram, e às quais o aluno esteve inscrito. Nesta fase, obviamente tardia, verificam-se já algumas situações de abandono e outras de insucesso extremo (número mínimo de disciplinas com aprovação) que põem em causa o sucesso do percurso curricular do aluno.

Com esta preocupação presente, e tendo por base a literatura consultada (estudos empíricos e argumentos teóricos apresentados na literatura), os dados disponibilizados e as metodologias de modelação estatística que melhor se adequarem ao fenómeno a modelar, procura-se no presente trabalho de investigação dar resposta a questões que se mantêm em aberto neste contexto (Figura 1).

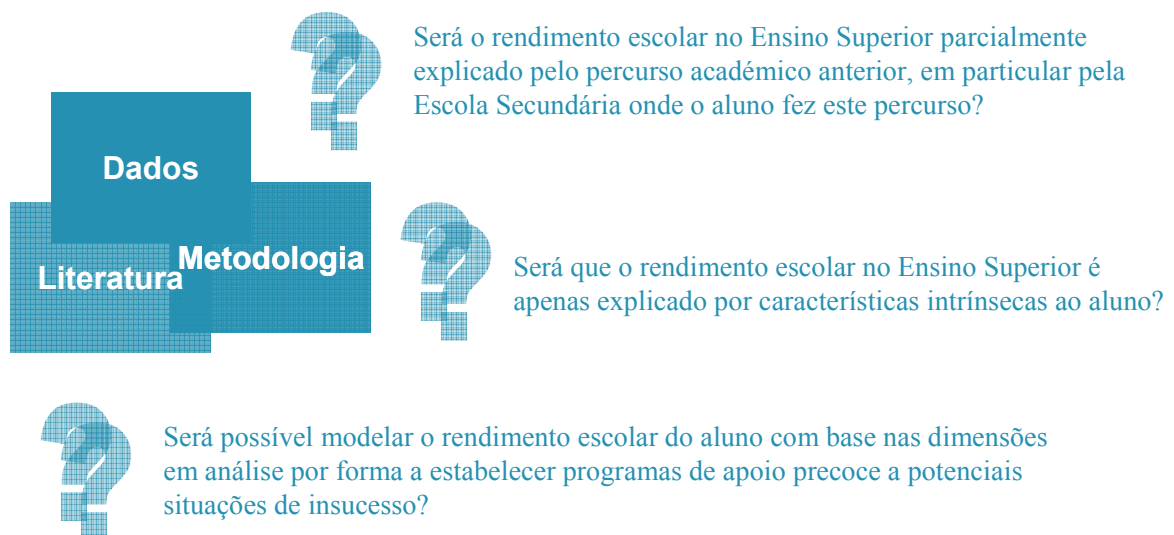


Figura 1: Questões de investigação.

1.2 Estrutura da tese

Este trabalho contemplará, para além da presente introdução, onde se desenvolveu um enquadramento da problemática em estudo; uma análise dos estabelecimentos de Ensino Secundário, mais concretamente do processo de construção dos perfis-tipo de escolas secundárias que possam contribuir para a explicação do fenómeno em causa; uma análise multinível do percurso escolar entre o Ensino Secundário e o Superior, onde se apresenta desde o modelo conceptual desenvolvido neste trabalho, aos resultados obtidos no mesmo, passando por uma exploração desta metodologia. Finalmente serão

apresentadas as principais conclusões, a descrição das limitações encontradas neste estudo e potenciais linhas de investigação futura.

2. CARACTERIZAÇÃO DAS ESCOLAS SECUNDÁRIAS

A hipótese de que as características do estabelecimento de Ensino Secundário onde o aluno realizou o percurso escolar anterior poderão contribuir parcialmente para a explicação dos seus resultados no Ensino Superior, tem sido particularmente explorada a nível internacional. Historicamente têm sido tomadas várias linhas de investigação nesta matéria, com resultados nem sempre convergentes (Zvoch e Stevens, 2006):

- aquelas que sugerem que as características e percurso (*background*) dos estudantes são determinantes para o seu sucesso no Ensino Superior;
- as que aludem à estrutura organizacional e ao clima da escola como relacionados com o desempenho no Ensino Superior;
- e aquelas onde se considera que as principais componentes para a explicação do fenómeno (desempenho académico no Ensino Superior) resultam de um binómio entre as características intrínsecas aos estudantes a par com o contexto escolar.

Alguns dos factores directamente relacionados com as escolas que são documentados como pertinentes para a explicação do desempenho dos estudantes são o contexto social e clima organizacional das escolas, as políticas e as práticas do corpo docente (Zvoch e Stevens, 2006; Tai *et al.*, 2006). As estruturas educacionais do corpo docente revelam também um papel crucial na melhoria dos resultados obtidos pelos estudantes, em particular nos docentes com experiência na área da ciência (Caldas e Bankston, 1999; Zvoch e Stevens, 2006).

A eficácia da escola pode dever-se aos recursos e aos modelos educacionais – a perspectiva da organização da comunidade enfatiza a importância de valores partilhados, relações positivas e estruturas democráticas de governação dentro das escolas (Caldas e Bankston, 1999).

No contexto nacional apenas se encontrou um trabalho onde foi estudada especificamente a questão em investigação. Em Teixeira *et al.* (2002), foi afluída esta temática mas apenas para um pequeno conjunto de escolas. Os resultados evidenciaram

que apenas em algumas das escolas analisadas existia uma relação entre o sucesso académico e as características do estabelecimento.

Ainda a nível nacional foram desenvolvidas duas investigações promovidas pela SEDES (Associação para o Desenvolvimento Económico e Social⁸), nas quais se exploraram as assimetrias regionais do desempenho educativo e as características das escolas com melhores resultados nos exames nacionais de Matemática (Tavares *et al.*, 2002; Tavares e Almeida, 2006). Ambos os trabalhos tiveram como base os resultados nos exames nacionais, sendo que no primeiro trabalho se focou a diversidade de resultados escolares face à diversidade do contexto social económico e cultural das escolas ao nível concelhio. As principais conclusões deste trabalho foram que as influências do contexto socioeconómico são relevantes mas não determinantes para a explicação dos resultados dos exames e que, dentro de concelhos com características semelhantes, existe uma enorme variabilidade de resultados, e, portanto, as características intrínsecas (e.g. clima organizacional, liderança, recursos e qualidade do corpo docente, etc.) às escolas são factores determinantes (Tavares *et al.*, 2002).

No segundo trabalho analisaram-se as características de escolas onde se desenvolvem as melhores práticas e onde se atingem os objectivos mais inovadores, sempre medindo o sucesso da escola através das classificações no exame de Matemática. Foram também analisadas as escolas que revelaram uma evolução positiva nas classificações no exame de Matemática nos anos em análise no estudo (2001-2004). Os resultados obtidos sugerem que a gestão, cultura, e o envolvimento da comunidade académica na prossecução dos objectivos estabelecidos são, de uma forma geral, factores que contribuem para estes fenómenos de sucesso do ponto de vista dos resultados dos exames de Matemática (Tavares e Almeida, 2006).

Os resultados dos exames do Ensino Secundário estão, habitualmente, relacionados com a publicação de algumas notícias subordinadas ao tema *rankings versus* resultados dos exames nacionais do Ensino Secundário.

De acordo com Justino (2005), a polémica que envolve a publicação dos resultados dos exames do 12º ano, e a sua utilização para a construção de *rankings*, assenta em vários

⁸ <http://www.sedes.pt/>, consultado em 16/07/2008.

aspectos: por um lado, os que defendem que os resultados não devem ser publicados, na medida em que a sua publicação, e conseqüente construção de *rankings*, para além de se tratar de uma análise unidimensional e simplista, causa estigmas e diferenciação entre os vários estabelecimentos de ensino de acordo com os seus resultados; por outro lado, os que defendem a publicação dos mesmos, por se tratar de informação produzida pela administração educativa, por não estar provado que a publicação desta informação tenha efeitos nefastos nos vários estabelecimentos, e em nome do igual acesso à informação que poderá ser importante na escolha do potencial estabelecimento de Ensino Secundário a frequentar.

Na realidade, independentemente da publicação destes resultados é um facto que há muito existe uma “percepção social da qualidade da oferta”, determinante da procura por determinadas escolas, em particular em grandes centros urbanos, onde existe maior oferta educativa. Naturalmente que o processo de construção desta percepção tem subjacente o grupo social e a valorização do investimento em educação familiar e subsequente preocupação com as trajetórias escolares dos filhos (Justino, 2005).

A selectividade de uma escola secundária e a sua localização geográfica conduz ao agrupamento de estudantes de determinado estrato socioeconómico e com desempenhos semelhantes. Este agrupamento poderá tornar difícil a distinção entre o contributo institucional resultante das boas práticas da escola e das características da população que a frequenta no desenvolvimento académico dos estudantes (Zvoch e Stevens, 2006).

Sumariamente, a diversidade de características das escolas secundárias que poderá contribuir para a trajetória académica futura dos seus estudantes pode incluir factores como a sua dimensão, natureza, perfil educativo, resultados académicos dos seus alunos, multiplicidade de percursos e qualificações dos professores, e contexto envolvente. Esta pluralidade pode traduzir-se em diferentes níveis de preparação dos estudantes, com contributos diferentes no seu futuro académico.

2.1 Objectivo

Em cada ano lectivo existem grupos de alunos que ingressam no Ensino Superior provenientes de determinado estabelecimento de Ensino Secundário. Centrando-se este estudo apenas numa geração ingressada numa única instituição de Ensino Superior, isto

significa que podem ingressar nesse estabelecimento de ensino um número diminuto ou mesmo nulo de alunos de algumas escolas. Tal revela-se problemático para a estimação do efeito da relação entre a escola secundária e o sucesso do aluno, visto que, se apenas existe um pequeno número de alunos provenientes de determinada escola, não é possível analisar a variabilidade das características do estabelecimento de ensino e, conseqüentemente, estabelecer a sua relação com o futuro sucesso dos alunos. Assim, do ponto de vista metodológico, considerou-se que o agrupamento das escolas com características semelhantes poderia resolver tal problema, na medida em que para efeitos dessa estimação se consideraria o grupo de escolas a que o aluno pertence, e respectivas características, e não a escola de proveniência. Este exercício permitirá ainda a aplicação do modelo desenvolvido em situações futuras, nas quais provenham alunos de escolas não contempladas na geração de alunos em estudo ou novas escolas entretanto criadas.

O passo tomado no sentido de analisar esta hipótese conduziu a uma análise de *clusters* com vista a tentar encontrar grupos diferentes entre si (heterogeneidade inter grupos) e cujos membros sejam tão semelhantes entre si quanto possível (homogeneidade intra grupos). O processo a aplicar para este fim será a modelação por misturas finitas, por todas as vantagens que esta apresenta face aos métodos tradicionais de *clustering* (métodos hierárquicos e de optimização). Entre as inúmeras vantagens desta metodologia destacam-se as seguintes pela sua importância para o problema em estudo: apuramento objectivo do número de grupos, hipótese de congregar num único modelo variáveis com diferentes escalas de medida e inclusão de variáveis preditivas (Dias, 2004, Magdison e Vermunt, 2002).

Com base nesta análise espera-se obter perfis de estabelecimentos, cujas características permitam analisar o seu contributo para o sucesso académico pós secundário dos alunos que ali concluem a sua formação.

Importa salientar que com este estudo não se pretende a avaliação do desempenho ou a ordenação (construção de *rankings*) das escolas, mas sim agrupar escolas com características semelhantes em várias dimensões, que se julgam pertinentes para a investigação em causa.

2.2 Fontes de informação

Actualmente existem mais de 600 instituições de Ensino Secundário em Portugal. A recolha de informação sobre os vários estabelecimentos de Ensino Secundário não é simples e os dados necessários para caracterizar a maioria das dimensões analisadas na literatura como potenciais caracterizadoras dos estabelecimentos não é de fácil acesso.

O período de referência dos dados usados neste projecto será o ano de 2004 e a informação recolhida para caracterização das escolas secundárias teve três fontes de informação:

- Gabinete de Estatística e Planeamento da Educação (GEPE) do Ministério da Educação, que gentilmente facultou um conjunto de dados relativos a cada escola secundária em funcionamento em Portugal Continental, nomeadamente, dados sobre a oferta formativa, dimensão e características do corpo docente das escolas com oferta de Ensino Secundário;
- Júri Nacional de Exames (JNE/GIASE, 2004), de onde se extraiu informação sobre os resultados dos exames nacionais do 12º ano de cada escola secundária;
- Publicação "O País em Números" (Instituto Nacional de Estatística, 2007), onde se recolheram dados sócio-demográficos dos municípios aos quais pertencem as escolas secundárias em estudo.

2.3 Dimensões de análise

Os estabelecimentos de Ensino Secundário podem caracterizar-se por um elevado número de dimensões. Esta multiplicidade pode incluir elementos sobre o corpo docente, infra-estruturas disponíveis, dimensão da escola em termos de área coberta e capacidade, localização e enquadramento socioeconómico da população servida, estrutura de níveis oferecida (Ensino Básico – 1º/2º ciclo e/ou Secundário), factores relacionados com a natureza da escola em causa, isto é, se se trata de um estabelecimento de cariz público ou privado, e resultados médios dos seus alunos.

No âmbito desta investigação consideram-se seis dimensões: contexto geográfico, natureza, perfil da oferta educativa, orientação para os resultados, características do

corpo docente e tamanho (Figura 2), que poderão caracterizar uma escola e influir no percurso académico futuro de um estudante.

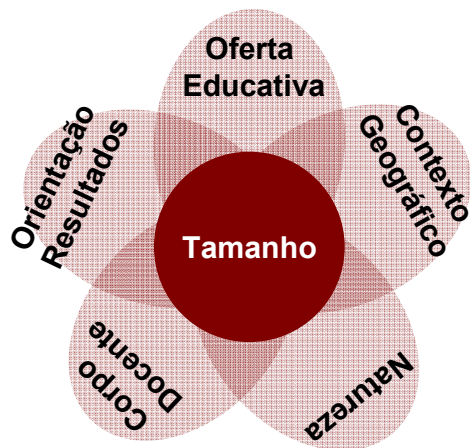


Figura 2: Dimensões em análise.

A perspectiva subjacente à adopção destas dimensões é a seguinte: o contexto geográfico reflecte a diversidade de oferta escolar e de conjunto de acessibilidades disponíveis; o tamanho, natureza e o corpo docente indicam qual a matriz identificadora da escola; o perfil de oferta educativa e a orientação para os resultados a métrica do cumprimento dos objectivos estabelecidos para a escola.

Partindo desta base discute-se de seguida a importância destas dimensões e o potencial contributo para o fenómeno em estudo.

A localização de um estabelecimento de ensino poderá determinar vários aspectos, nomeadamente a facilidade de acesso a bens culturais e sociais, bem como a fixação e estabilização do corpo docente (Justino, 2005, p. 30). De acordo com Tavares *et al.* (2002), a dimensão regional traduz assimetrias regionais, que, por sua vez, integram desequilíbrios sociais, culturais e económicos a par com uma diferenciação da oferta escolar.

Os factores envolventes da escola podem também reflectir o nível socioeconómico da população que serve, o nível educacional dos pais e os recursos culturais disponíveis em casa. Este *background* pode aumentar ou diminuir a qualidade do ensino de uma escola por mera concentração de alunos nestas condições e níveis de preparação, padrões de

comportamento e atitudes perante a aprendizagem, todos relacionados com o estatuto socioeconómico (Caldas e Bankston, 1999).

A natureza do estabelecimento, pública ou privada, pode evidenciar resultados diferentes, tipicamente mais elevados em colégios privados, na performance dos estudantes no percurso académico pré-universitário (Pike e Saupe, 2002). A nível nacional podemos analisar esta dimensão à luz da selectividade dos estabelecimentos de ensino, particularmente visível nos grandes centros urbanos e em escolas privadas. A maioria das escolas desta natureza (privada), especialmente as que não têm contratos de associação com o Estado, encontra-se nos grandes centros urbanos (Justino, 2005).

Naturalmente que o ensino privado terá uma estratégia de preparação e acompanhamento dos alunos para o Ensino Superior numa perspectiva de lógica de mercado. As escolas privadas, na sua grande maioria dependentes exclusivamente de receitas próprias, orientam-se por uma estratégia comercial e tendem a estar localizadas nos grandes centros urbanos, onde existe viabilidade (estratos sociais médios/altos em elevada concentração) para a sua manutenção. Acresce que estas escolas tendem a adoptar uma política de selecção dos alunos do Ensino Secundário virada para o sucesso no ingresso no Ensino Superior, utilizando este factor como ferramenta de marketing e promoção. Todavia existem algumas escolas estatais que, face a uma elevada procura, têm subjacente um processo de selecção semelhante (Justino, 2005).

Por outro lado, o tamanho da escola poderá influenciar o desempenho do aluno, na medida em que quanto maior for a escola melhor será a sua adaptação a um *campus* universitário, habitualmente grande (Pike e Saupe, 2002). Importa referir que, em Portugal, a natureza do estabelecimento está involuntariamente relacionada com a dimensão da escola, sendo que os estabelecimentos de ensino privado têm habitualmente um menor número de alunos. A questão subjacente à inclusão desta dimensão é, assim, a avaliação do incremento ou mais-valia que estas diferenças de percurso trazem verdadeiramente ao rendimento académico no Ensino Superior.

Anteriormente foi referida a discussão entre os que são a favor e contra a publicação dos resultados dos exames nacionais por escola secundária. Se por um lado pode ser nocivo para a imagem da escola, por outro pode conduzir à orientação do seu funcionamento e a uma estratégia de desenvolvimento educacional em função dos mecanismos de concorrência. Afinal, e não se pretendendo apenas uma análise unidimensional do problema, os resultados dos exames dos alunos que frequentam uma escola podem ser

encarados como uma medida de qualidade geral do ensino ministrado na escola (Pike e Saupe, 2002).

De acordo com Justino (2005, pp. 15 e 16), os exames também servem para avaliar, aferir e orientar, e, como tal, os seus resultados podem ser utilizados como indicadores de sucesso, prossecução de objectivos e progressão, sendo exactamente esta a dimensão pretendida nesta investigação. Se estes são determinantes para o acesso ao Ensino Superior (contando mais de 35% a 50% da nota de seriação), e tendo em conta que em investigações anteriores (e.g. Almeida *et al.*, 2004) se concluiu que os resultados dos exames explicam parte do sucesso dos alunos no 1º ano da universidade, considera-se pertinente a inclusão de variáveis desta índole na medida em que os resultados da escola e o empenho na preparação dos alunos para os exames nacionais pode ter impacto no seu desempenho futuro. Acresce ainda que este é um dos únicos parâmetros disponibilizados que permite comparar as diferentes escolas em termos de resultados à saída.

O sucesso escolar e aproveitamento escolar estão ainda relacionados com uma maior probabilidade de dar continuidade ao investimento escolar (Grácio, 1997, *in* Martins e Campos, 2005), logo a orientação das escolas para os resultados poderá incutir motivações diferentes nos alunos.

Para o acesso ao Ensino Superior os candidatos precisam de ter uma classificação superior ou igual a 95 pontos (escala 0-200) nos exames nacionais das disciplinas exigidas como provas de ingresso, podendo este critério traduzir igualmente um investimento da escola na qualificação dos seus alunos e, por conseguinte, em médias mais elevadas.

Relacionada com os resultados está também a oferta formativa. A orientação da escola em termos de oferta formativa poderá influenciar a preparação dos alunos que pretendem seguir estudos superiores, especialmente porque essas escolhas poderão estar condicionadas pela oferta existente nas escolas (Pike e Saupe, 2002; Martins e Campos, 2005).

Em Portugal existem escolas secundárias/colégios mais vocacionados para determinado tipo de ensino, seja ele geral, tecnológico e/ou profissional. Existem, em particular, colégios privados especialmente vocacionados para a preparação dos alunos para os exames nacionais (Justino, 2005 p. 23), mas também escolas públicas que, por

determinação do seu público-alvo, apostam na preparação dos alunos para os exames no Ensino Secundário, como catapulta para o acesso aos cursos pretendidos no Ensino Superior.

Os perfis de oferta educativa poderão ter subjacente uma associação entre os objectivos da sua população e a gestão do projecto educativo, reflectindo-se em diferentes dinâmicas educacionais. Procurando este estudo a associação entre o percurso anterior e o sucesso do Ensino Superior, a vocação da escola em torno da preparação para o prosseguimento de estudos poderá ser um factor importante.

O corpo docente actua como gestor e dinamizador do projecto educativo e do clima organizacional de uma escola. A estabilidade e a qualificação do corpo docente, o seu envolvimento na definição e prossecução dos objectivos e subsequente qualidade educativa, são também factores a pesar na medida em que estes podem assegurar a continuidade pedagógica e o acompanhamento do trajecto escolar dos alunos (Justino, 2005). Em Tavares e Almeida (2006), um dos factores que se destacou na discriminação de escolas com bons resultados foi, exactamente, o corpo docente, sendo que a mudança de professor trouxe em algumas das escolas analisadas uma melhoria dos resultados na disciplina de Matemática.

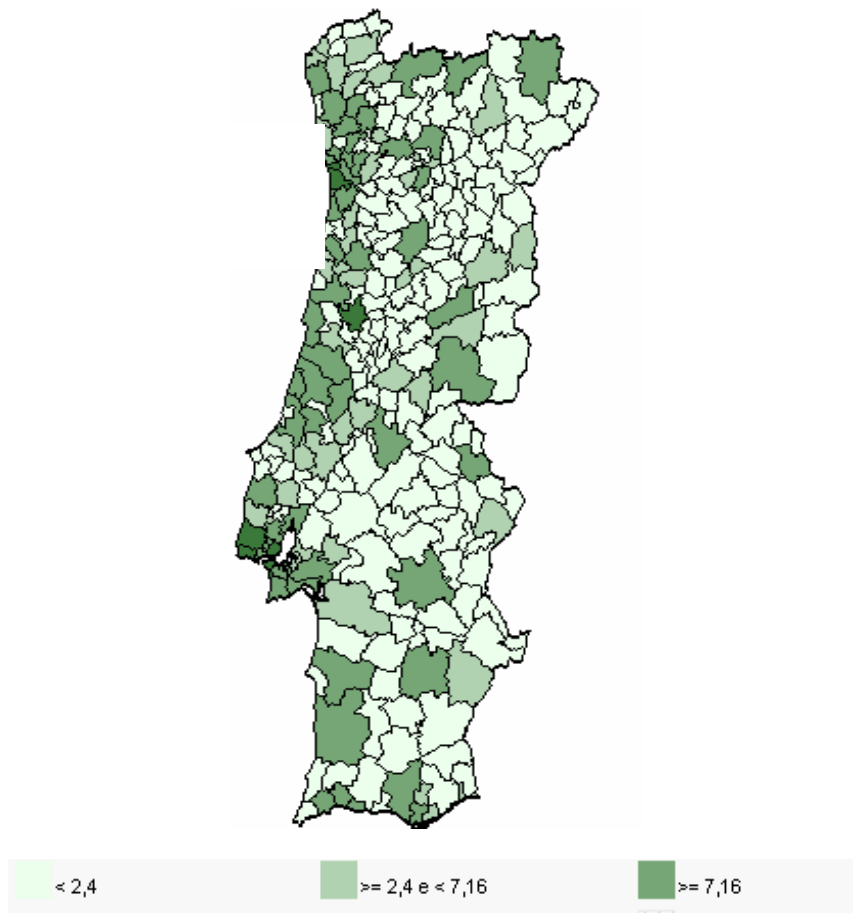
Acredita-se que a conjugação destas dimensões permitirá encontrar grupos de escolas secundárias, cujas características permitirão analisar a contribuição do estabelecimento pré-universitário para o complexo processo de ensino e aprendizagem dos alunos no Ensino Superior.

2.4 Análise descritiva

A caracterização de estabelecimentos de Ensino Secundário em Portugal pode começar pela análise da distribuição espacial das 666 escolas⁹ em funcionamento em 2004, ano de referência deste trabalho, em Portugal Continental. A Figura 3 ilustra a distribuição das escolas que oferecem Ensino Secundário pelos vários municípios de Portugal em

⁹ O universo de análise cingiu-se às escolas secundárias, não incluindo as escolas profissionais. Não obstante, importa salientar que existem escolas secundárias que oferecem cursos profissionais, sendo que essas foram contempladas neste estudo.

2004, constatando-se que a maior concentração de escolas está associada aos grandes centros urbanos.



Fonte: País em Números – Edição 2007, INE.

Figura 3: Distribuição dos estabelecimentos de Ensino Secundário (Público e Privado) em Portugal Continental em 2004¹⁰.

¹⁰ Medido em termos do n.º de escolas por concelho.

O contexto geográfico pode analisar-se por várias vias, sendo a escolhida neste trabalho a tipologia de áreas urbanas (Instituto Nacional de Estatística, 1998). De acordo com esta tipologia, existem três tipos de freguesias:

- Freguesia urbana – freguesia que possua uma densidade populacional superior a 500 hab./km² ou cuja população residente seja superior ou igual a 5000 habitantes;
- Freguesia semi-urbana – freguesia não urbana, que possua uma densidade populacional compreendida entre 100 hab./km² e 500 hab./km², ou cuja população residente esteja compreendida entre 2000 e 5000 habitantes;
- Freguesia rural – todas as freguesias que não se enquadravam nas duas categorias anteriores.

As freguesias de Portugal Continental distribuem-se por estas categorias de acordo com as proporções apresentadas na Figura 4.



Fonte: O País em Números 2004/05. INE

Figura 4: Distribuição das freguesias do continente por tipologia.

Na Figura 5 apresenta-se a distribuição das escolas secundárias consideradas na análise desenvolvida neste capítulo, contextualizadas por esta tipologia de acordo com a freguesia a que pertencem¹¹. Estes resultados permitem concluir que a grande maioria dos estabelecimentos de Ensino Secundário (75%) se encontra em freguesias urbanas, sendo que apenas 3% estão em freguesias rurais.



Figura 5: Distribuição das escolas secundárias de acordo com a tipologia das freguesias a que pertencem.

A natureza do estabelecimento de ensino é caracterizada por uma variável dicotómica, que tomará valores consoante se trata de um estabelecimento de ensino privado ou público (Figura 6). No que toca ao cariz da instituição, nomeadamente, se esta é pública ou privada, verifica-se que a maioria das escolas em Portugal Continental é de natureza pública (81%).

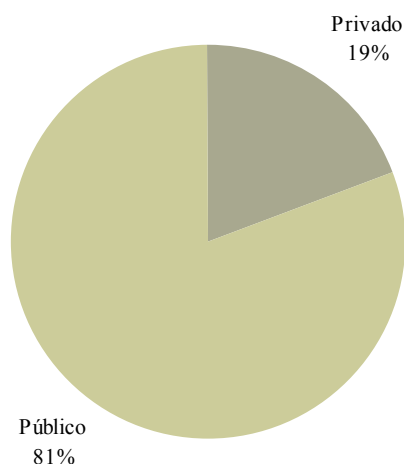


Figura 6: Distribuição das escolas por natureza.

O perfil do estabelecimento em termos da oferta educativa ao nível do Ensino Secundário está relacionado com a vocação da escola em termos de oferta de cursos. Em 2004 um aluno matriculado num estabelecimento de Ensino Secundário podia frequentar um curso geral, um curso tecnológico e/ou profissional para concluir este nível de ensino. Os cursos gerais e tecnológicos, ainda que com objectivos de

¹¹ Embora existissem 666 escolas em 2004, apenas para 567 existe informação sobre todas as variáveis incorporadas na análise levada a cabo. Assim, será sobre este conjunto de escolas que incidirá toda a análise.

aprendizagem diferentes, visam preparar os alunos para o acesso ao Ensino Superior; no caso dos cursos profissionais os alunos podiam fazer os exames nacionais para aceder ao Ensino Superior, mas o conteúdo programático dos cursos não estava vocacionado para tal preparação, nem a conclusão dos cursos o requeria.

Assim, optou-se por analisar esta vertente de acordo com a oferta apenas de cursos gerais, considerando a variável percentagem de matriculados em cursos gerais face ao total de alunos inscritos no nível secundário (Figura 7).

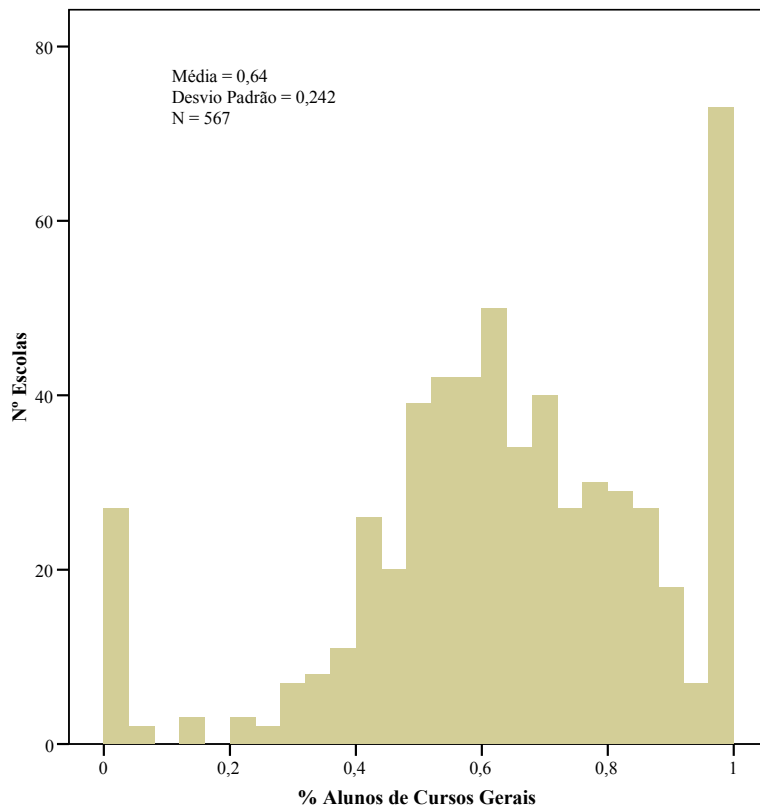


Figura 7: Histograma da variável percentagem de matriculados em cursos gerais.

Não considerando os resultados dos exames do Ensino Secundário como um instrumento de avaliação educacional de cada estabelecimento de ensino, mas sim como um indicador da orientação para os resultados (preparação dos alunos para o acesso ao Ensino Superior), inclui-se aqui a variável média no exame nacional da disciplina de Matemática. Na realidade, a inclusão desta variável, especialmente no contexto em que decorre esta investigação, justifica-se pelo facto desta análise não ser unidimensional e integrar vários factores relacionados com os vários estabelecimentos de ensino.

Embora se tenham observado variações nos resultados absolutos e relativos dos vários estabelecimentos de ensino ao longo dos anos, apenas se usaram os resultados referentes ao ano de 2004, por uma questão de coerência com o ano de referência dos dados incluídos nesta investigação. Note-se que, apesar de se verificarem estas variações, em termos relativos não se verifica uma alteração significativa nos resultados das escolas (Justino, 2005; pp. 23 e 32).

A orientação para os resultados do ensino será assim caracterizada através da média dos alunos internos na disciplina de Matemática na 1ª fase dos exames nacionais em 2004 (Figura 8).

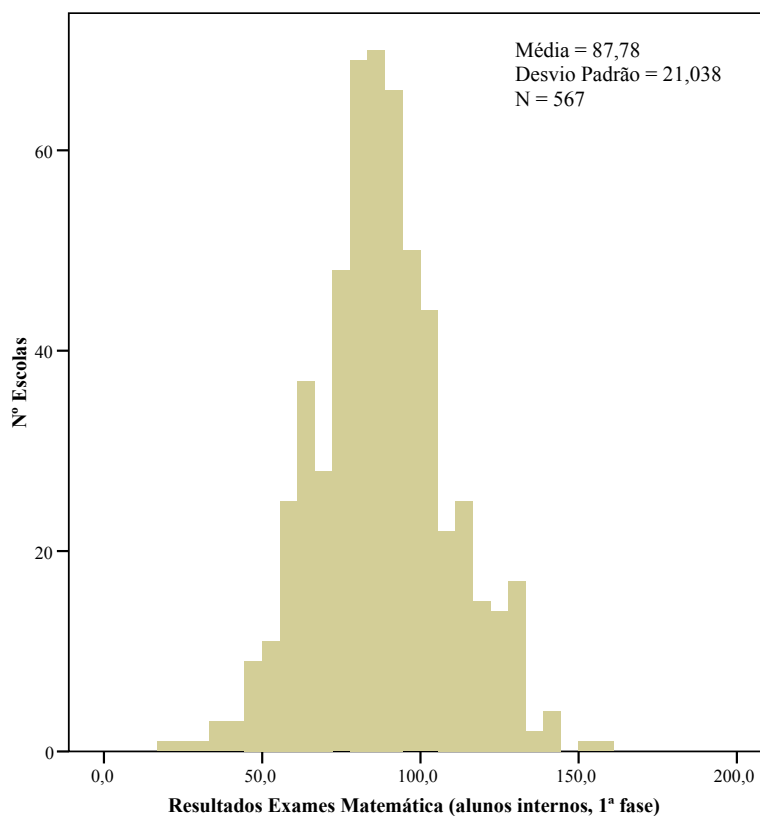


Figura 8: Distribuição do número de escolas por média na disciplina de Matemática.

A escolha da Matemática justifica-se não só pelo facto deste trabalho se centrar na área de Ciência e Tecnologia, como também porque alguma da investigação desenvolvida em Portugal sobre a caracterização dos factores determinantes do sucesso escolar dos alunos evidenciou a importância dos resultados anteriores em Matemática (Almeida *et al.*, 2004).

A incidência sobre a Matemática assume ainda outro interesse na medida em que os resultados deste exame permitem o acesso à generalidade dos cursos com elevada procura, e, como tal, poderá ser uma condicionante das escolhas dos cursos de Ensino Superior (Tavares *et al.*, 2002). Assim, revela-se importante explorar os resultados médios dos alunos de cada escola secundária nesta área, traduzidos pelo desempenho médio nesta disciplina, como potencial contributo para o desenvolvimento de bases e competências que perdurem no aluno.

A análise das características associadas ao corpo docente centrar-se-á numa variável: a proporção de docentes com menos de 40 anos. Esta variável procura medir, por ausência de um indicador mais preciso, a experiência, estabilidade (proximidade local de residência) e maturidade do corpo docente. Tipicamente quanto maior a proporção de docentes com mais de 40 anos, maior será a sua estabilidade na prossecução do projecto educativo e maior probabilidade de existir um acompanhamento da trajectória escolar dos alunos de uma escola; por outro lado, importa realçar que poderá também ser o grupo etário com um menor grau de actualização da formação. Na Figura 9 encontra-se a distribuição desta variável.

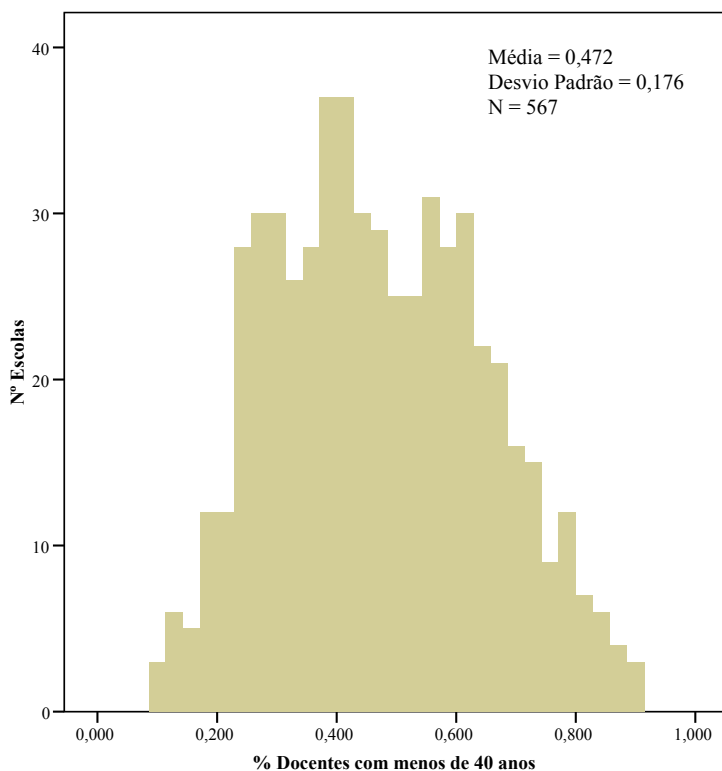


Figura 9: Distribuição da percentagem de docentes com menos de 40 anos.

2.5 Metodologia: modelos de mistura finita com variáveis concomitantes

Na definição dos segmentos a incluir neste nível de análise identificou-se a necessidade de agrupar as escolas secundárias com características conhecidas semelhantes, tarefa levada a cabo através da aplicação de modelos de mistura finita. Este ponto centra-se na apresentação e descrição desta metodologia cujos objectivos são, entre outros, estimar a distribuição das variáveis representativas das dimensões em análise e permitir o delineamento de perfis dos vários segmentos obtidos.

Os modelos de mistura finita têm mais de um século de existência, ainda que apenas com a evolução tecnológica mais recente tenham sido mais desenvolvidos (Dias, 2004). Este tipo de modelação tem várias aplicações, entre as quais a estimação de densidades, detecção de observações *outliers* e o *clustering* probabilístico (*model-based clustering*), salientando-se esta última como uma das áreas aplicacionais onde se tem tornado muito conhecida (Dias, 2004; Magdison e Vermunt, 2002).

A ideia subjacente a este tipo de modelos é que as observações recolhidas na amostra provêm de duas ou mais populações desconhecidas – classes latentes, daí as várias designações pelas quais esta metodologia também é conhecida – Modelos de Segmentos/Classes Latentes (Dias, 2004; Wedel e DeSarbo, 1994).

Nos modelos de misturas finitas a pertença a um grupo (*cluster*) não é conhecida *a priori*, sendo o objectivo identificar os grupos subjacentes, através da decomposição da amostra nas suas componentes de mistura (Wedel e DeSarbo, 1994). Neste tipo de modelos, as variáveis observadas (base de segmentação) podem ser de natureza mista, modeladas através de uma mistura de distribuições, o que torna este tipo de modelos muito flexível¹².

A inclusão de covariáveis activas ou inactivas no modelo permite transformar a natureza descritiva dos modelos de mistura finita num modelo explicativo e preditivo da alocação dos indivíduos às componentes (Dias, 2004).

Se escolhermos introduzir covariáveis activas no modelo, para além de classificarmos os segmentos obtidos, aquelas (covariáveis) também poderão ser preditivas dos

¹² As variáveis contínuas são modeladas através de uma distribuição Normal e, de um modo geral, as variáveis categóricas (discretas) são modeladas por uma distribuição Multinomial.

segmentos (explicando a pertença aos mesmos); se escolhermos introduzir covariáveis inactivas no modelo, estas apenas servirão para a caracterização dos segmentos, não sendo introduzidas na modelação e não influenciando na maximização da verosimilhança. De referir, que ao introduzir covariáveis activas no modelo, este torna-se mais complexo, aumentando o número de parâmetros a estimar.

Considerando uma amostra de n indivíduos com J indicadores, assuma-se a seguinte notação:

- i – indivíduo, $i = 1, \dots, n$;
- j – indicador, em que $j = 1, \dots, J$;
- S – número de componentes;
- \mathbf{v} – variáveis dependentes (para as quais a distribuição de probabilidade é condicionada às s componentes);
- \mathbf{w} – variáveis concomitantes (que explicam a probabilidade *a priori* da observação i pertencer a cada uma das s componentes);
- \mathbf{z} – segmento de pertença do indivíduo, representado por um vector $\mathbf{z} = (z_1, \dots, z_S)$, para o qual se propõe uma distribuição multinomial: $\mathbf{z} \sim M_{S-1}(\boldsymbol{\pi})$, $\boldsymbol{\pi} = (\pi_1, \dots, \pi_{S-1})$;
- π_s – peso relativo de cada componente (pesos da mistura), isto é, probabilidade *a priori* de um indivíduo provir da componente s , podendo ser interpretado como o tamanho relativo das componentes e satisfazendo: $\pi_s > 0$, $s = 1, \dots, S$ e
$$\sum_{s=1}^S \pi_s = 1;$$
- $f_s(\cdot)$ – distribuição do vector \mathbf{v}_i (função massa de probabilidade ou densidade, consoante a natureza das variáveis que compõem \mathbf{v}_i) de cada componente s ;
- $f_{sj}(\cdot)$ – função massa de probabilidade ou densidade da variável j para a componente s ;
- $\boldsymbol{\theta}_s$ – vector de parâmetros de cada componente.

O modelo de mistura finita com S componentes com variáveis concomitantes é definido por:

$$f(\mathbf{v}_i; \mathbf{w}_i; \varphi) = \sum_{s=1}^S \pi_s(\mathbf{w}_i) f_s(\mathbf{v}_i; \boldsymbol{\theta}_s) \quad (1)$$

onde, partindo do pressuposto da independência local,

$$f_s(\mathbf{v}_i; \boldsymbol{\theta}_s) = \prod_{j=1}^J f_s(v_{ij}; \theta_{js}). \quad (2)$$

A densidade $f(\mathbf{v}_i; \mathbf{w}_i; \varphi)$ diz-se identificada se, pelo menos, 2 conjuntos distintos de parâmetros não têm a mesma distribuição.

A estimação deste tipo de modelos visa responder aos seguintes objectivos: ou maximizar a função de verosimilhança ou maximizar a função (densidade) de probabilidade *a posteriori*¹³. Em ambos os casos, pretende-se obter um modelo mais ajustado aos dados. Geralmente a estimação é efectuada via maximização da função de verosimilhança (que será a via utilizada neste trabalho):

$$L(\varphi | \mathbf{v}) = \prod_{i=1}^n f(\mathbf{v}_i; \varphi) = \prod_{i=1}^n \left[\sum_{s=1}^S \pi_s(\mathbf{w}_i) f_s(\mathbf{v}_i; \boldsymbol{\theta}_s) \right] \quad (3)$$

com

$$\pi_s(\mathbf{w}_i) = \frac{e^{\gamma_{0s} + \sum_{l=1}^L \gamma_{ls} \mathbf{w}_{il}}}{\sum_{r=1}^S e^{\gamma_{0r} + \sum_{l=1}^L \gamma_{lr} \mathbf{w}_{il}}} \quad (4)$$

onde γ representam os coeficientes das variáveis concomitantes (\mathbf{w}).

Assim, é necessário estimar os pesos dos diversos segmentos, bem como os parâmetros das distribuições de cada segmento (associadas às diversas variáveis base \mathbf{v}).

O algoritmo mais utilizado para obter as estimativas de máxima verosimilhança nos modelos de mistura finita é o *Expectation-Maximization* (EM). Trata-se de um algoritmo numericamente estável mas lento devido à convergência linear, cujo processo

¹³ Esta abordagem tem a vantagem de conseguir ultrapassar alguns problemas verificados na obtenção de estimativas de máxima verosimilhança (Vermunt e Magdison, 2005).

iterativo permite, em cada iteração, o incremento da função de verosimilhança em duas etapas (Dias, 2003), de acordo com o princípio esquematizado na Figura 10.

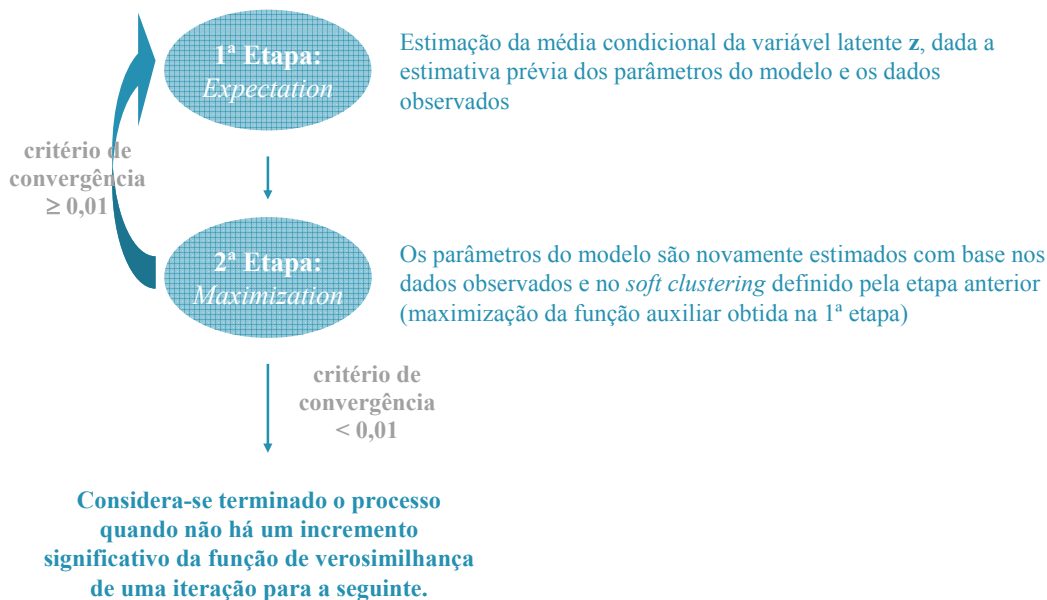


Figura 10: Algoritmo EM.

Este processo tende a terminar em máximos locais, não atingindo o valor máximo de verosimilhança. A sua repetida aplicação com valores iniciais diferentes poderá ser uma solução para este problema.

O *Latent Gold* (software utilizado no desenvolvimento do presente trabalho) integra uma variante do algoritmo EM (Vermunt e Magdison, 2005). Neste software, o EM é utilizado numa fase inicial (até atingir o número máximo de iterações estabelecidas ou o critério de convergência) e o algoritmo Newton-Raphson é usado no final quando se está próximo da convergência para a solução final (até atingir o número máximo de iterações estabelecidas ou o critério de convergência geral).

O ajustamento do modelo pode ser verificado não só pelas medidas globais de ajustamento como também pela análise do quanto o modelo consegue reproduzir as interações observadas entre pares de indicadores. Um dos pressupostos do modelo é a existência de independência local, isto é, que as variáveis observadas sejam independentes dentro de cada componente da mistura (Dias, 2004). A violação deste pressuposto provoca a falta de ajustamento do modelo, algo que pode ser transposto de uma de duas formas: aumentando o número de *clusters* ou relaxando o pressuposto,

permitindo uma relação directa entre os indicadores e/ou efeitos directos entre covariáveis nos indicadores (Vermunt e Magdison, 2005). A segunda opção é frequentemente uma alternativa mais viável, visto que evita que se termine a modelação com um elevado número de grupos, permitindo uma melhor classificação dos objectos nos grupos (Vermunt e Magdison, 2002)

O software usado na estimação foi, conforme referido, o *Latent Gold 4.0*. Este software permite a análise dos resíduos bivariados, usados para detectar a similaridade das associações estimadas e as observadas entre os pares de variáveis (Vermunt e Magdison, 2005).

A estimação dos modelos é realizada via maximização da função de verosimilhança. A selecção do número de *clusters* baseia-se no melhor compromisso possível entre a maximização e a penalização da complexidade do modelo. Ou seja, para termos um modelo melhor ajustado aos dados poderemos estar a complexificá-lo de tal forma (aumentando o número de segmentos, e conseqüentemente de parâmetros a estimar, com o intuito de explicar as associações entre as variáveis base) que não justifica o aumento da verosimilhança. Portanto, quando se analisam os resultados provenientes deste processo, no sentido de decidir qual o número ideal de *clusters*, há que ter em atenção os critérios de informação que traduzem a relação entre a complexidade do modelo e a verosimilhança obtida.

Os critérios de informação, cuja fórmula base é:

$$-2 \ln L + \alpha P, \quad (5)$$

em que P é o número de parâmetros a estimar e α a constante de penalização, tomam em consideração este compromisso. Neste trabalho serão considerados os critérios de informação BIC (*Bayesian Information Criterion*), com $\alpha = \ln(n)$, AIC (*Akaike's Information Criterion*), com $\alpha = 2$ e AIC3, com $\alpha = 3$, todos penalizando a maximização em função da complexidade (Dias, 2004). Escolhe-se o número de segmentos que minimiza estes critérios.

2.6 *Estimação e selecção do modelo*

2.6.1 *Descrição dos dados utilizados*

Um primeiro aspecto a ter em consideração será, naturalmente, a escolha das variáveis a incorporar na modelação e a natureza das mesmas. Como discutido em 2.4, para a classificação em grupos homogéneos (*clusters*) consideraram-se as variáveis constantes na Tabela 1, que se julga poderem contribuir para uma caracterização da heterogeneidade das escolas.

Tabela 1: Descrição das variáveis incluídas na modelação por misturas finitas (U).

Nome variável	Tipo variável
Natureza	Dicotómica: Pública; Privada
Localização	Dicotómica: rural/semi-urbana; urbana
Resultados Matemática	Contínua
% Docentes com menos 40 anos	Contínua
% Alunos Cursos Gerais	Contínua

Adicionalmente incluiu-se ainda a covariável contínua tamanho do estabelecimento de ensino (w), que se considera como promissora na predição e explicação da pertença aos grupos. A inclusão desta variável prende-se não só com a facilidade na obtenção da mesma, mas também pela sua solidez, na medida em que é uma variável estável e caracterizadora do ponto de vista socio-demográfico.

2.6.2 *Seleção do número de segmentos*

O objectivo da segmentação é, como referido anteriormente, traçar perfis de escolas. Contudo, não existe informação disponível que nos indique quantos perfis poderemos esperar à partida (tendo em consideração as variáveis escolhidas). Assim, e com vista à estimação do número de grupos ideal, foram ensaiadas soluções para um número de segmentos entre 1 e 15.

Na parametrização do algoritmo (EM) foi, também, aumentado o número de inicializações: das 10, por defeito, para 200 (ver 2.5 para analisar a utilidade de tal procedimento). As variáveis base de segmentação foram devidamente caracterizadas como nominais ou contínuas. O número de grupos a considerar foi identificado, tendo

por base as medidas BIC, AIC e AIC3, critérios de informação para penalizar o ajustamento e parcimónia do modelo (quanto mais baixo melhor o modelo). Recorde-se que estas medidas, que permitem a comparação entre as várias soluções (número de grupos), têm por base um compromisso entre a maximização da função de verosimilhança e a penalização por complexidade do modelo (número de parâmetros a estimar).

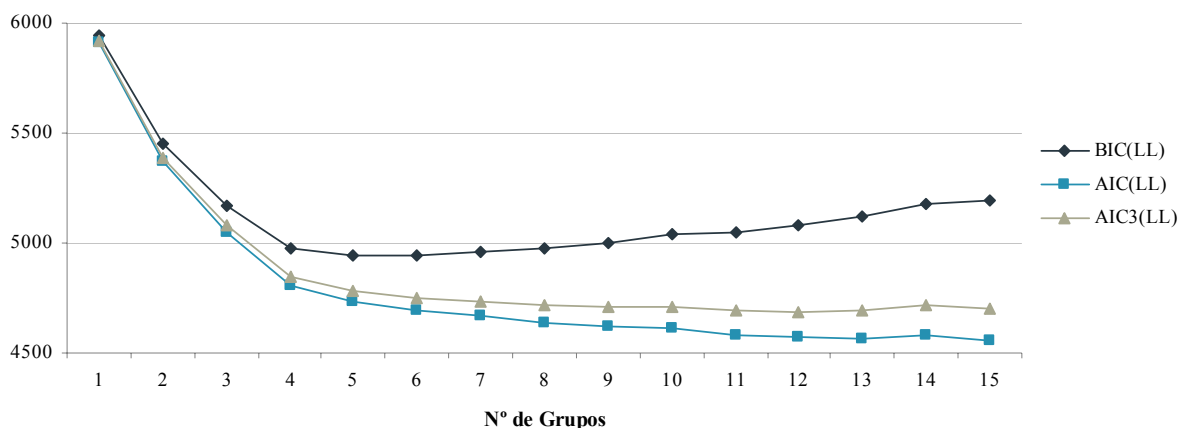


Figura 11: Resultados dos critérios de informação.

A análise destes critérios (Figura 11), cujo objectivo é a selecção da melhor estrutura de segmentos, sugere, numa primeira leitura, a adopção de uma estrutura de 6 segmentos visto que é nestes valores que o critério BIC apresenta a existência de um mínimo. Os critérios de informação AIC e AIC3 não apresentam mínimos para um número de segmentos entre 1 e 15.

A opção tomada, em função não só dos resultados do critério BIC¹⁴, como também da interpretabilidade do modelo, foi a solução com 6 segmentos por se revelar a que melhor espelha a heterogeneidade da amostra.

¹⁴ Em caso de divergência entre critérios deve optar-se pelo número de segmentos que minimiza o critério BIC, visto que o mesmo é consistente (Bozdogan, 1987 *in* Pontes e Dias, 2007).

2.6.3 Diagnóstico dos resíduos bivariados

No que se refere aos resíduos bivariados ($\chi^2 / g.l.$), após a estimação do modelo (e seu ajustamento) há que verificar se existe ou não associação entre as variáveis (se elas são condicionalmente independentes). Visto o valor-esperto do χ^2 corresponder ao número de graus de liberdade, os resíduos não deverão ultrapassar o valor um (e não sendo aconselhável ultrapassar o 2).

Tabela 2: Resíduos bivariados (variáveis).

	Natureza	Localização	% Docentes com menos 40 anos	Resultados Matemática	% Alunos Cursos Gerais
Natureza	.				
Localização	3,21	.			
% Docentes com menos 40 anos	6,80	0,57	.		
Resultados Matemática	3,83	0,02	3,18	.	
% alunos Cursos Gerais	0,46	0,01	0,88	2,08	.

Tabela 3: Resíduos bivariados (covariáveis).

	Natureza	Localização	% Docentes com menos 40 anos	Resultados Matemática	% Alunos Cursos Gerais
Nº Alunos Ens. Secundário	0,00	1,37	0,01	2,66	0,10

A análise destes valores para a solução dos 6 segmentos revelou que, na maioria dos resultados, estes se encontravam dentro dos valores “normais”. Em algumas situações, assinaladas a cinzento na Tabela 2 e na Tabela 3, evidencia-se algum desajustamento local (e.g. % Docentes com menos 40 anos *vs.* Resultados Matemática).

A análise destes valores revela um desajustamento local, ou seja, verifica-se a existência de resíduos acima de 1, indicadores que o modelo pode não explicar bem a associação observada sob o pressuposto da independência local (subjacentes à metodologia aplicada). O caso mais notório é na relação da variável natureza com as variáveis localização, % Docentes com menos 40 anos e Resultados Matemática.

Como referido no ponto 2.5, este problema poderia ser solucionado de uma de duas formas: aumentando o número de segmentos ou relaxando o pressuposto da independência local. Todavia, e porque, na primeira opção se aumentaria a complexidade do modelo, sem uma aumento da interpretabilidade do mesmo; no

segundo caso existe um risco de se perderem *clusters* relevantes para o problema (Magdison e Vermunt, 2002); e os resíduos em causa não serem muito elevados, optou-se por manter o modelo estimado.

Foram desenvolvidas algumas experiências, incorporando outras variáveis, por forma a tentar ultrapassar este problema. Numa dessas experiências obteve-se uma solução que não tinha o problema dos resíduos bivariados acima do valor limite, contudo, o modelo em causa era mais complexo, sobretudo do ponto de vista interpretativo, existindo ainda uma concentração grande de escolas em dois grupos. Assim, e pautando a escolha pela interpretabilidade e simplificação do modelo, optou-se pela solução aqui apresentada e discutida no próximo ponto.

2.7 Perfis de escolas secundárias

Obtiveram-se 6 segmentos de escolas com a seguinte distribuição por grupo (Figura 12).

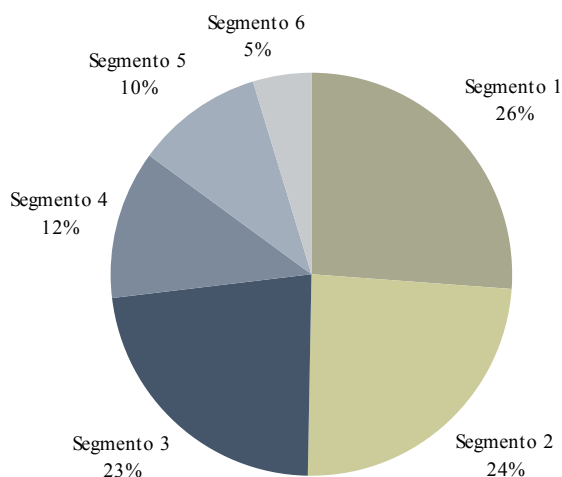


Figura 12: Distribuição escolas por segmento.

Tendo por base a estimativa da probabilidade *a priori* das várias componentes ($\hat{\pi}_s$), verifica-se que o primeiro segmento é o maior, contemplando 26% das escolas secundárias e que o 6º segmento é um pequeno grupo de escolas composto por apenas 5% de estabelecimentos.

Nas tabelas seguintes encontram-se as estimativas para as variáveis dependentes (Tabela 4), e para as variáveis concomitantes (Tabela 5), salientando que foi usada uma codificação por efeitos¹⁵ para o apuramento das estimativas de variáveis nominais.

Tabela 4: Estimativas para as variáveis dependentes.

Variáveis		Teste Wald	Solução com 6 segmentos					
		Estimativa	1	2	3	4	5	6
Natureza	Pública	55,46***	-0,377	2,410	-0,717	2,386	-2,333	-1,369
	Privada		0,377	-2,410	0,717	-2,386	2,333	1,369
Localização	rural/semi-urbana	107,64***	-0,066	-1,004	1,813	-0,632	0,128	-0,240
	urbana		0,066	1,004	-1,813	0,632	-0,128	0,240
Resultados Matemática		41,41***	-4,558	2,946	-9,352	6,271	9,687	-4,994
% Docentes com menos 40 anos		593,24***	0,023	-0,166	0,192	-0,159	0,084	0,027
% Alunos Cursos Gerais		41345,10***	-0,007	-0,093	0,078	0,237	0,390	-0,604

*** p < 0,001

Tabela 5: Estimativas para as variáveis concomitantes.

Variáveis	Teste Wald	Solução com 6 segmentos					
	Estimativa	1	2	3	4	5	6
Nº Alunos Ens. Secundário	114,91***	0,006	0,009	-0,003	0,005	-0,009	-0,007

*** p < 0,001

A estatística de Wald permite-nos perceber se existem diferenças significativas entre os coeficientes das várias classes obtidas (segmentos), neste caso dos 6 segmentos. Com um nível de significância de 1% rejeita-se a hipótese nula (inexistência de diferenças de parâmetros entre segmentos), logo todos os indicadores revelam capacidade discriminante entre os segmentos.

No caso da variável concomitante *Nº Alunos Ensino Secundário* (Tabela 5), a análise desta estatística permite-nos rejeitar a hipótese de que não existe diferença neste parâmetro entre os vários segmentos e que, portanto, a mesma permitirá a alocação de outras escolas não contempladas nesta análise nos segmentos encontrados.

¹⁵ Ou seja, os resultados devem ser interpretados em termos dos desvios da média.

Na Tabela 6 apresenta-se a probabilidade de ocorrer um nível do atributo, ou, no caso de variáveis contínuas o valor médio do indicador, dado que pertence a um dado segmento, a par com os resultados para a totalidade da amostra (modelo agregado).

Tabela 6: Caracterização das variáveis dependentes por segmento.

Variáveis		Modelo Agregado	Solução com 6 segmentos					
			1	2	3	4	5	6
Natureza	Pública	81%	88%	100%	79%	100%	13%	50%
	Privada	19%	12%	0%	21%	0%	87%	50%
Localização	rural/semi-urbana	25%	11%	2%	84%	4%	15%	8%
	urbana	75%	89%	98%	16%	96%	85%	92%
Resultados Matemática		87,8	84,3	91,8	79,5	95,1	98,6	83,9
% Docentes com menos 40 anos		47%	49%	30%	66%	31%	55%	50%
% Alunos Cursos Gerais		64%	60%	52%	69%	85%	100%	0%

Tabela 7: Caracterização das variáveis concomitantes por segmento.

Variáveis	Modelo Agregado	Solução com 6 segmentos					
		1	2	3	4	5	6
Nº Alunos Ens. Secundário	554	623	1008	248	573	147	171

Na Tabela 7 apresenta-se o valor médio da variável explicativa *Nº Alunos Ens. Secundário*, ou seja, o indicador que permite predizer a pertença a cada segmento.

De uma forma geral os estabelecimentos podem caracterizar-se da seguinte forma:

Cluster 1 (26%) - é uma escola pública, localizada em zona urbana, vocacionada maioritariamente para cursos gerais, mediana em termos de resultados de Matemática e com um corpo docente relativamente estável. O seu tamanho em termos de número de alunos é grande. Designa-se este segmento por **Escola Pública Urbana**.

Cluster 2 (24%) - é uma escola pública, localizada em zona urbana, cuja vocação divide-se entre a oferta de cursos gerais e de cursos de outra natureza, com classificações em Matemática acima da média e com um corpo docente muito estável. Neste segmento enquadram-se escolas localizadas nos grandes centros urbanos (mais de 60% das escolas deste grupo estão situadas nos distritos de Lisboa, Porto e Setúbal) e

muito grandes em termos de número de alunos. Designa-se este segmento por **Escola Pública Urbana Grande**.

Cluster 3 (23%) - é uma escola pública localizada numa zona rural a semi-urbana (mais de 50% destas escolas encontram-se situadas no interior do país), vocacionada sobretudo para a oferta de cursos gerais, com um corpo docente muito pouco estável, e com classificações em Matemática abaixo da média. Neste segmento enquadram-se escolas pequenas em termos de número de alunos, o que é coerente com a sua localização e classificação geográfica. Designa-se este segmento por **Escola Pública do Interior**.

Cluster 4 (12%) - é uma escola pública localizada numa zona urbana, com classificações em Matemática bastante acima da média, com um corpo docente bastante estável, particularmente vocacionada para a preparação de jovens para o seguimento de estudos. Neste segmento enquadram-se escolas de média dimensão. Designa-se este segmento por **Escola Pública Urbana Proficiente**.

Cluster 5 (10%) - é uma escola privada, localizada em zona urbana, particularmente orientada para o sucesso na Matemática (com classificações em Matemática acima da média), com um corpo docente relativamente estável, exclusivamente vocacionada para a preparação de jovens para o seguimento de estudos. Neste segmento enquadram-se escolas localizadas em grandes centros urbanos, onde tipicamente coabitam as escolas privadas, com uma componente de oferta educativa centrada apenas em cursos gerais. Todos os elementos que caracterizam este grupo, incluindo o facto de se tratar de uma escola de pequena dimensão, são coerentes com a óptica de mercado de estabelecimentos desta natureza. Designa-se este segmento por **Escola Privada Urbana Proficiente**.

Cluster 6 (5%) – é uma escola, que independentemente da natureza (pública ou privada) está localizada numa zona urbana, com uma componente de oferta educativa centrada em cursos técnico profissionais, e com um corpo docente relativamente estável. Este é claramente um *cluster outlier* de escolas onde se enquadram escolas de pequena dimensão, com vocação para a preparação de alunos para a entrada precoce na vida activa. Designa-se este segmento por **Escola Técnico-Profissional**.

2.8 *Análise complementar*

Como forma de complementar a análise desenvolvida, e mostrar graficamente os elementos caracterizadores dos segmentos obtidos, foi desenvolvida uma aplicação da técnica de Análise de Correspondências Múltiplas (ACM). Em linhas gerais, este método permite reduzir a dimensionalidade do espaço de partida, otimizando a homogeneidade das variáveis e potenciando uma representação gráfica (por via de planos) da multidimensionalidade daquele espaço.

As variáveis dicotómicas já incluídas na modelação anterior foram incorporadas com a mesma estrutura, todavia, para este efeito foi necessário categorizar as variáveis contínuas. Numa fase inicial a codificação das variáveis contínuas foi feita através do método de agrupamento disponível no SPSS (software estatístico), mais concretamente especificando um número de categorias e distribuindo os valores pelas categorias com base na distribuição normal. Numa segunda fase, agruparam-se algumas categorias com poucos registos com as adjacentes. O resultado é o constante na Tabela 8.

Tabela 8: Variáveis codificadas, e respectivas categorias, para a ACM.

Nome variável	Categorias
Média Matemática	0-73 pontos; 73-101 pontos; 101-118 pontos; 118-200 pontos
% Docentes com menos 40 anos	<35%; 35-59%; >59%
% Alunos Cursos Gerais	<30%; 30-48%; 48-64%; 64-80%; 80-98%; >98%
Nº Alunos Ens. Secundário	<280; 280-551; 558-830; >830

Do máximo de dimensões possíveis (21), decidiu-se que somente duas bastariam para reproduzir a multidimensionalidade da informação recolhida. Estas duas dimensões explicam cerca de 27% da variância total (16% na primeira dimensão e 11% na segunda).

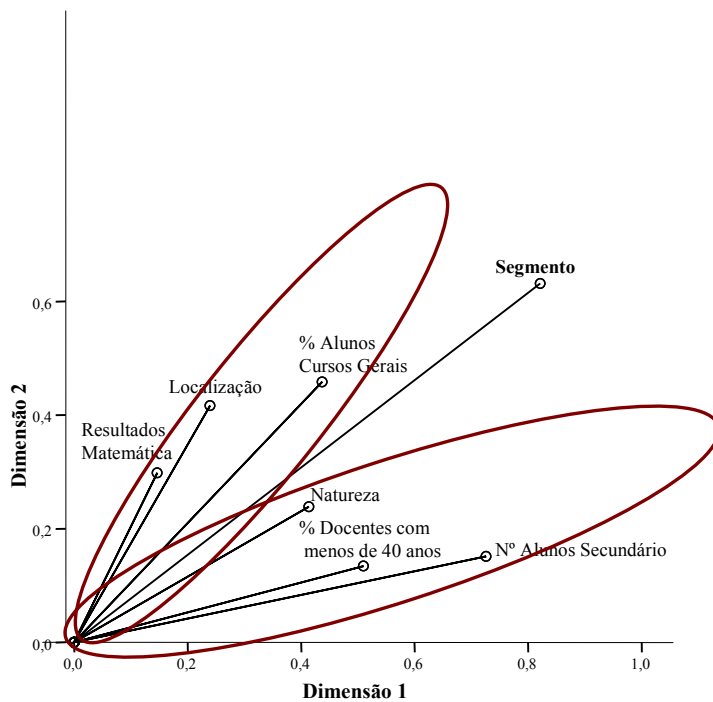


Figura 13: Medidas de discriminação.

A Figura 13 mostra-nos quais as variáveis que melhor discriminam cada dimensão, evidenciando-se aquelas que estão associadas à natureza e corpo docente (dimensão 1); e aquelas que estão relacionadas com a vocação e capacidade de orientação para os resultados (dimensão 2).

Na Figura 14 encontra-se a projecção das categorias num plano (que corresponde às duas dimensões apresentadas), permitindo identificar as distâncias e oposições entre as categorias. Desta forma, pretende identificar grupos de escolas que tendencialmente partilham as mesmas características, contribuindo-se, assim, para a visualização dos perfis de escolas estabelecidos.

Destacam-se 6 perfis, confirmados pela associação entre os resultados por este método (*scores*) e os segmentos obtidos através da modelação por misturas finitas (o segmento de pertença foi projectado suplementarmente). Destacando-se, por exemplo, os segmentos *Escola Privada Urbana Proficiente* (segmento 5) e *Escola Técnico-Profissional* (segmento 6) no 3º quadrante.

Note-se que a associação entre a segmentação obtida em 2.7 e a realizada nesta análise complementar (via ACM) não é directa, até porque para a segunda houve a necessidade

de categorizar as variáveis contínuas. O objectivo desta complementaridade foi apenas a visualização gráfica dos aspectos que mais evidenciam cada segmento.

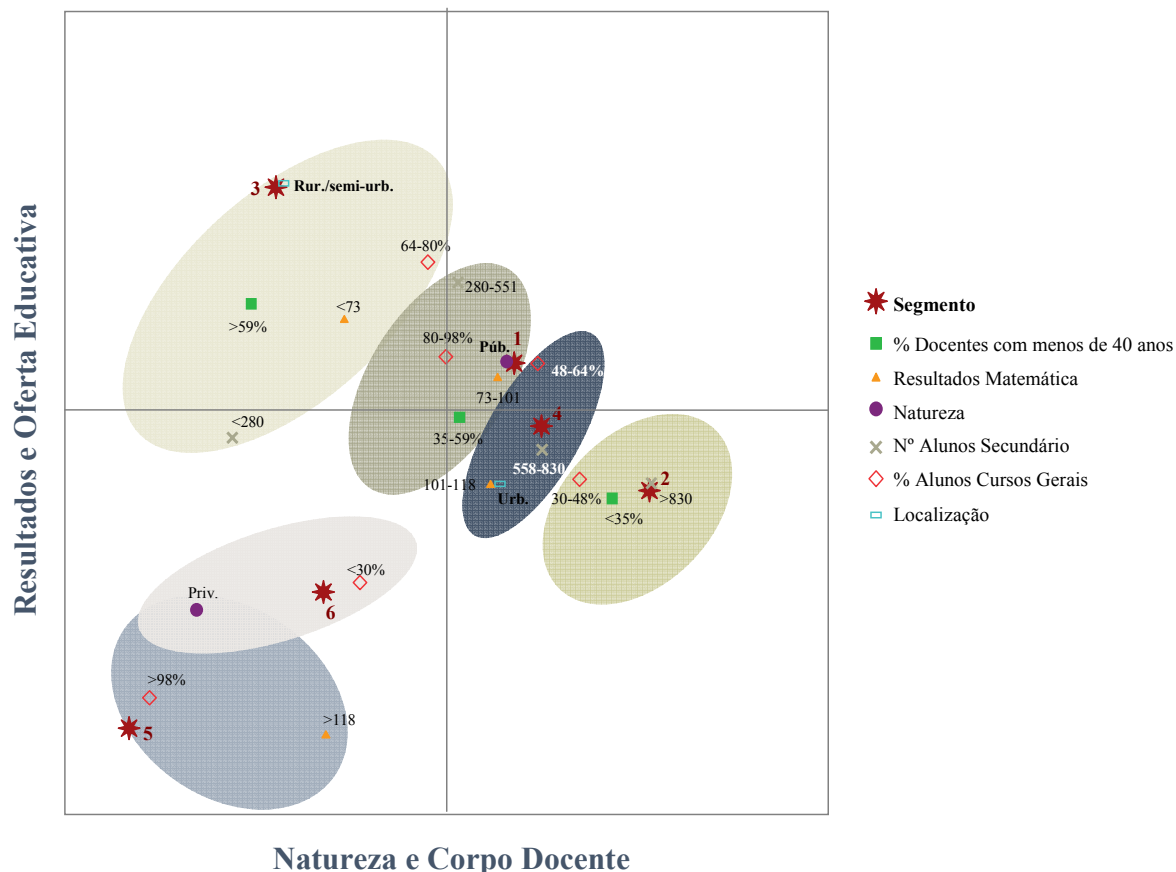


Figura 14: Projecção das categorias em estudo.

2.9 Conclusões e notas finais

Com vista a compreender a heterogeneidade entre os estabelecimentos de ensino, foi efectuada uma análise de *clusters* com base num modelo de mistura finita com variáveis concomitantes. A diversidade de características das escolas secundárias em análise incluiu factores em 6 dimensões, tendo os resultados permitido distinguir 6 segmentos no universo de escolas secundárias em análise.

No contexto da presente investigação espera-se que os (6) segmentos obtidos, sintetizados na Figura 15, permitam caracterizar a influência dos Estabelecimento de Ensino Secundário, e das várias dimensões que os distinguem, no percurso académico dos alunos, designadamente no impacto do seu sucesso académico no Ensino Superior.

Segmentos	π_s	Natureza	Localização	Resultados Matemática	% Docentes com menos 40 anos	% Alunos Cursos Gerais	Nº Alunos Ens. Secundário	
	1	26%	Pub.	Urb.	~	~	~	~
2	24%	Pub.	Urb.	↑	↓	↓	↑↑	→ Escola Pública Urbana Grande
3	23%	Pub.	Semi-Urb./Rur.	↓	↑↑	~	↓	→ Escola Pública do Interior
4	12%	Pub.	Urb.	↑↑	↓	↑	~	→ Escola Pública Urbana Proficiente
5	10%	Priv.	Urb.	↑↑	↑	↑↑	↓↓	→ Escola Privada Urbana Proficiente
6	5%	Pub./Priv.	Urb.	~	~	↓↓	↓	→ Escola Técnico-profissional

Legenda:

- ↑↑ muito acima da média
- ↑ acima da média
- ~ próximo da média
- ↓ abaixo da média
- ↓↓ muito abaixo da média

Figura 15: Perfis de escolas secundárias.

Os resultados obtidos são coerentes com as análises desenvolvidas na bibliografia consultada a nível nacional, em particular Tavares *et al.* (2002), Tavares e Almeida (2006) e Justino (2005), nomeadamente quanto ao facto dos resultados dos exames de Matemática estarem associados não só com a estabilidade do corpo docente, como também com a localização e a natureza da escola.

Os resultados obtidos poderão servir de base a investigações futuras, nomeadamente:

- desenvolvimento de medidas de combate ao insucesso escolar, tanto no Ensino Secundário, com a regulação da oferta educativa a nível nacional, como no Ensino Superior, através da identificação das lacunas na preparação dos alunos ingressados;
- recrutamento e divulgação de informação sobre os cursos de Ensino Superior.

Para tal será vantajoso a recolha de informação adicional sobre os estabelecimentos de Ensino Secundário, no sentido de os caracterizar e contextualizar com maior grau de detalhe.

No Capítulo 4 serão incluídos alguns aspectos adicionais relativos às limitações identificadas no decorrer deste trabalho.

3. PERCURSO ESCOLAR ENTRE O ENSINO SECUNDÁRIO E O SUPERIOR: UMA ANÁLISE MULTINÍVEL

3.1 *Modelo conceptual*

Um aluno quando ingressa numa instituição de Ensino Superior traz consigo um historial académico resultante da sua passagem pelo Ensino Secundário. A sua passagem por este ciclo de estudos desenvolveu-se num determinado contexto escolar que contribui de alguma forma para a preparação do aluno para o passo seguinte da sua vida académica, a entrada no Ensino Superior. Outros factores poderão estar envolvidos no processo de adaptação de um aluno ao Ensino Superior, sendo estes de carácter estrutural da própria instituição onde o aluno ingressa, da conjugação entre as suas preferências (pares de cursos/instituições escolhidos pelo aluno) e a realidade (curso em que efectivamente foi colocados).

A adaptação ao contexto universitário e os resultados académicos neste nível de ensino resultam de um processo complexo e diversificado que envolve factores de várias naturezas, em particular de natureza intra pessoal e contextual, nem sempre fáceis de integrar (Soares *et al.*, 2006).

A modelação do fenómeno tem incluindo também factores relacionados com o percurso pré-universitário do aluno (Amelink, 2005; Windham, 2005; Almeida *et al.*, 2004), incorporando os efeitos da escola onde o aluno fez o Ensino Secundário (Pike e Saupe, 2002); relacionados com o enquadramento institucional da universidade e da estrutura curricular do curso onde ingressam (Porter, 2006, Amelink, 2005; Szafran, 2001).

No contexto do Ensino Superior nacional, destacam-se alguns trabalhos:

- Em Martins e Campos (2005) foi desenvolvida uma caracterização social do recrutamento e trajectórias escolares, a par com as escolhas de formação, no qual se exploraram as biografias e vivências de alunos recém-chegados ao Instituto Politécnico de Santarém;
- Em Soares *et al.* (2006) foi proposto um modelo conceptual (Modelo multidimensional de ajustamento de jovens ao contexto Universitário) compreensivo e integrativo que explica a adaptação ao Ensino Superior. Neste modelo presumia-se que as características pré-universitárias dos estudantes

(sócio-demográficas: sexo, nível socioeconómico e residência habitual; e académica: nota ingresso, opção de colocação, área do curso) afectam directamente ou indirectamente o rendimento académico no Ensino Superior. Os resultados deste trabalho sugeriam o desenvolvimento de modelos holísticos de análise que combinem factores académicos e não académicos no processo de adaptação ao Ensino Superior;

- Em Nóvoa (2005), foi feita uma caracterização dos factores de sucesso e insucesso escolar na Universidade de Lisboa, incluindo um levantamento bibliográfico bastante completo sobre esta temática a nível nacional.

Com a convicção que cada caso é um caso, e que um aluno não se resume a um conjunto de dados quantitativos, com a presente investigação pretende-se construir um modelo preditivo que traduza as características do aluno, conhecidas *a priori*, num nível de propensão ao insucesso escolar. Obviamente que a explicação do sucesso do aluno passa certamente pela conjunção de vários factores, uns mensuráveis e previamente disponíveis, outros não mensuráveis e intrínsecos ao aluno. Assim, é necessário ter presente que nenhum modelo poderá explicar integralmente tal conceito, contudo, espera-se que alguns factores em análise possam contribuir para a definição de grupos de risco.

Importa realçar que, com um sistema desta índole, não se pretende criar elites de estudantes ou tão pouco contribuir para a marginalização dos alunos com resultados potencialmente menos satisfatórios. Pretende-se, sim, construir um instrumento que, com base nas características mensuráveis e disponíveis relativamente a cada aluno, permita identificar padrões de insucesso de forma a agir precocemente junto do aluno, podendo igualmente contribuir para uma maior personalização do ensino.

Conceptualmente, considera-se como hipótese de trabalho que o processo de construção desenvolvido pelo aluno ao longo do seu percurso escolar no Ensino Secundário influenciará o seu desempenho. Ou seja, que os alunos provenientes de determinado estabelecimento de ensino terão um padrão de comportamento (em termos de sucesso escolar) resultante da conjugação de factores como as experiências de ensino e aprendizagem, o contexto social, as infra-estruturas disponíveis, a vivência dos alunos no estabelecimento de ensino frequentado.

Do ponto de vista do aluno assumir-se-á que são várias as vertentes que poderão contribuir para a explicação do seu sucesso: aptidões académicas, capital familiar, nível socioeconómico, expectativas, motivações e interesse no curso, e outros factores contextuais.

Ao nível do segmento de escolas secundárias de proveniência, assumem-se as dimensões de análise estabelecidas no Capítulo 2.

A Figura 16 sintetiza o modelo conceptual proposto neste trabalho.

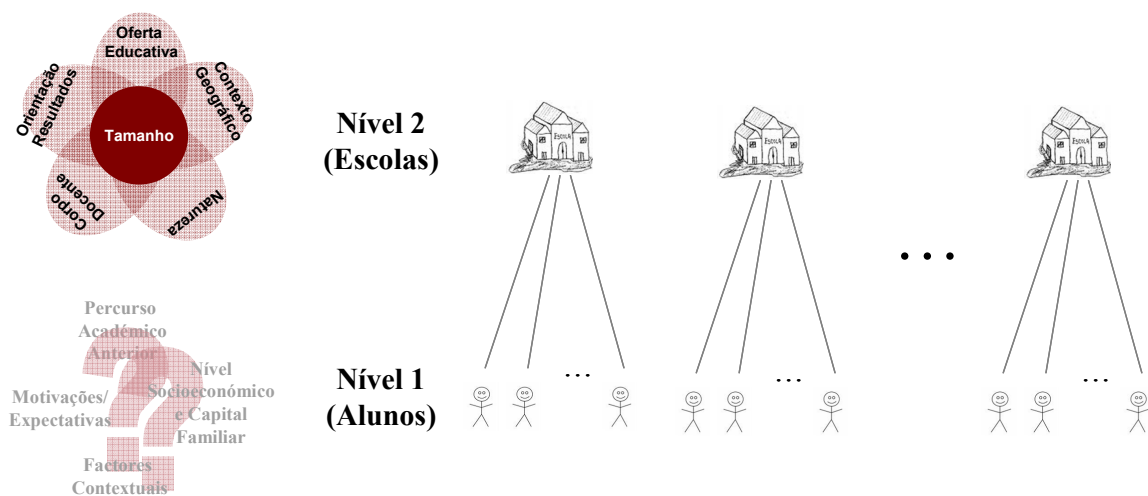


Figura 16: Modelo conceptual.

3.2 Metodologia

3.2.1 Modelos multinível lineares

Os modelos multinível, também conhecidos por modelos hierárquicos ou de coeficientes aleatórios, permitem modelar estruturas hierárquicas com complexos padrões de variabilidade (Snijders e Bosker, 1999; Goldstein, 1999; Hox, 1995). Um exemplo clássico de aplicação desta metodologia é na investigação em educação, mais concretamente a relação entre os alunos e as turmas; até que ponto e de que forma o rendimento escolar de um aluno poderá ser influenciado pelo contexto que o rodeia, em particular a turma onde este se insere? Os alunos que partilham a mesma turma dificilmente se podem considerar como observações independentes (por exemplo, Carvalho e Lourenço, 2006). Na prática, o ignorar que existe variabilidade não só entre

os alunos mas também entre turmas/escolas pode traduzir-se numa fonte de erro. A agregação ou desagregação de indicadores subjacente a um dos níveis da hierarquia, neste caso do aluno ou da turma, pode, igualmente, ser uma fonte de erro.

Os modelos multinível emergiram no início da década de 80 através de duas correntes: análise contextual (estudo dos efeitos do contexto social no comportamento do indivíduo) e análise de efeitos mistos (modelos estatísticos de análise de variância e de regressão, onde se assume a existência de efeitos fixos e aleatórios) (Snijders e Bosker, 1999).

Na essência deste tipo de modelação está a possibilidade de aferir a dimensão da variabilidade em cada nível e entre os níveis da hierarquia em estudo, a par com o facto desta variabilidade poder ser contabilizada através da existência de variáveis preditoras incluídas em cada um destes níveis hierárquicos. Os resultados do modelo permitem obter informações sobre a composição da variabilidade total. A análise multinível tem também como vantagem a apresentação de estimativas mais fidedignas na medida em que estas, ao contrário do modelo clássico de regressão linear, não assume que as observações são independentes entre si, pelo contrário, tem em conta que as observações do nível mais baixo estão agrupadas nos grupos caracterizados pelos níveis superiores (Hox, 1995)

À semelhança dos modelos de regressão linear, o objectivo subjacente à aplicação dos modelos multinível é a construção de um modelo que exprima a forma como um fenómeno (traduzido por uma variável - variável dependente ou a explicar) é explicado em função de determinados indicadores (variáveis preditoras). A variável a explicar estará no primeiro nível dado que este tipo de modelo é um modelo de explicação de um fenómeno no nível mais baixo (mais detalhado). A ideia subjacente à modelação é que a variável a explicar poderá ter um efeito individual assim como um efeito do grupo onde se insere.

O modelo a seguir explicitado corresponde a uma especificação teórica constituída por dois níveis hierárquicos – nível um ou inferior (nível individual) e nível dois ou superior (nível do grupo).

Considere-se a seguinte notação:

- N – número de grupos;
- n_j – número de indivíduos no grupo j ;
- j – índice usado para a referência aos grupos ($j = 1, \dots, N$);
- i – índice usado para a referência aos indivíduos dentro de cada grupo ($i = 1, \dots, n_j$);
- k – índice da variável explicativa do nível 1 ($k = 1, \dots, K$);
- l – índice da variável explicativa do nível 2 ($l = 1, \dots, L$);
- por cada grupo (j) e indivíduo (i) têm-se as seguintes variáveis: Y_{ij} – variável dependente; X_{kij} – variável explicativa k no nível 1;
- por cada grupo (j) têm-se as seguintes variáveis: Z_{lj} – variável explicativa l no nível do grupo.

De uma forma genérica, o modelo multinível linear pode ser representado da seguinte forma:

$$Y_{ij} = \beta_{0j} + \sum_{k=1}^K \beta_{kj} X_{kij} + e_{ij} \quad (6)$$

onde e_{ij} é a componente de erro aleatório no nível individual ($e_{ij} \sim N(0, \sigma^2)$). As constantes (β_{0j}) e os coeficientes de regressão (β_{kj}), dependentes dos grupos a que pertencem (j), podem ser explicitados da seguinte forma:

$$\beta_{0j} = \gamma_{00} + \sum_{l=1}^L \gamma_{0l} Z_{lj} + u_{0j} \quad (7)$$

e

$$\beta_{kj} = \gamma_{k0} + \sum_{l=1}^L \gamma_{kl} Z_{lj} + u_{kj}, \quad (8)$$

onde u_{0j} e u_{kj} são as componentes de erro aleatório (resíduos) do nível dois¹⁶.

Assume-se que os erros do nível um, que são independentes e identicamente distribuídos entre grupos, são independentes dos do nível dois, e que, dados todos os valores das variáveis explicativas, ambas as variáveis aleatórias (v.a.) residuais (nível um e dois) têm média zero. Substituindo as equações que definem as constantes e os coeficientes de regressão na equação que define o modelo hierárquico linear (1) obtém-se a seguinte equação:

$$Y_{ij} = \underbrace{\gamma_{00} + \sum_{l=1}^L \gamma_{0l} Z_{lj} + \sum_{k=1}^K \gamma_{k0} X_{kij} + \sum_{k=1}^K \sum_{l=1}^L \gamma_{kl} Z_{lj} X_{kij}}_{(a)} + \underbrace{u_{0j} + \sum_{k=1}^K u_{kj} X_{kij}}_{(b)} + e_{ij} \quad (9)$$

onde a primeira parte da equação (a) é denominada por componente fixa do modelo e a segunda parte (b) por componente aleatória. A componente fixa do modelo engloba a média global, os efeitos das variáveis explicativas do nível individual (nível um) e do nível do grupo (nível dois) na variável a predizer e os efeitos de interação entre as variáveis explicativas dos dois níveis. A componente aleatória é composta pelos erros dos dois níveis considerados na modelação, assim como uma interação aleatória entre o grupo e as variáveis preditoras do nível individual.

Como qualquer modelo estatístico, o modelo multinível linear tem subjacente um conjunto de pressupostos essenciais para a sua estimação e validade. Alguns dos pressupostos deste modelo já foram aflorados aquando da apresentação da especificação teórica, nomeadamente a independência dos erros aleatórios. Existem, porém, outros pressupostos que necessitam de ser validados para garantir que todas as relações entre os dados em estudo estão bem representadas, a saber: dependência linear da variável dependente, não só das variáveis independentes incluídas no modelo, como também dos efeitos aleatórios; a normalidade da distribuição dos resíduos; a especificação das variáveis que têm coeficientes aleatórios (*random slopes*). Em suma, os pressupostos a validar no âmbito de um modelo multinível são os seguintes: os erros do segundo nível (nível do grupo) são independentes entre os grupos; são também independentes dos erros do primeiro nível (individual); dados os valores de todas as variáveis explicativas,

¹⁶ Na presente aplicação assume-se a variação da constante entre os grupos do nível 2, mas não se assume

ambos os resíduos (nível um e dois) têm média populacional zero; os erros do primeiro nível têm uma distribuição normal e variância constante; os resíduos do segundo nível têm uma distribuição normal multivariada com uma matriz de covariância constante¹⁷ (Snijders e Bosker, 1999).

Intuitivamente, o processo de estimação dos parâmetros funciona de uma forma iterativa, alternando entre a estimação dos coeficientes de regressão (parte fixa) e dos resíduos (parte aleatória), até existir convergência do processo. Esta descrição sucinta corresponde, de grosso modo, ao método IGLS (*Iterated Generalized Least Squares*), usado para o cálculo de estimativas por Máxima Verossimilhança (Goldstein, 1999).

A capacidade explicativa do modelo traduz-nos, na regressão linear múltipla (RLM), a proporção de variação explicada da variável alvo (predita) pelo conjunto das variáveis predictoras. Tal conceito, expresso através do R^2 (coeficiente de determinação múltipla) na RLM, é mais complexo de conceptualizar nos modelos multinível, visto que existem vários níveis hierárquicos e que em cada um existe uma componente de variância.

Em Snijders e Bosker (1999) é proposto que se usem os mesmos princípios da RLM, ou seja, que a capacidade explicativa do modelo tenha presente a proporção de diminuição de variação não explicada do modelo; e que se considere que podemos prever os valores da variável a explicar (ao nível do indivíduo) e a média dessa mesma variável em cada um dos grupos (nível grupo). Partindo desta base, dois conceitos de proporção de variância explicada emergem: proporção da diminuição do erro de predição da variável alvo de um indivíduo e proporção da diminuição do erro de predição da média do grupo, definidos da seguinte forma:

$$R_{\text{indiv\u00edduo}}^2 = 1 - \frac{\text{Var}(Y_{ij} - \sum_{k=1}^K \gamma_{k0} X_{kij})}{\text{Var}(Y_{ij})} \quad (10)$$

e

o mesmo para os coeficientes de regressão.

¹⁷ A propriedade de variância constante é designada por homocedasticidade, ou, no caso de variância não constante, de heterocedasticidade.

$$R_{grupo}^2 = 1 - \frac{Var(\bar{Y}_{\bullet j} - \sum_{k=1}^K \gamma_{k0} \bar{X}_{k\bullet j})}{Var(\bar{Y}_{\bullet j})} \quad (11)$$

A análise multinível é recomendada quando as correlações intra classes (medida da proporção da variância no nível inferior que é explicada pela variação do nível superior) são substancialmente grandes (Muthén, 1991, *in* Nasser e Hagtvet, 2006).

Actualmente, existem vários softwares para a estimação de modelos multinível lineares (SAS, HLM/2L, SPSS, etc.), sugerindo-se por exemplo Sullivan *et al.* (1999) para uma análise comparativa de resultados obtidos entre dois programas.

3.2.2 Modelos multinível binomiais

Os modelos multinível descritos no ponto anterior, e que podemos considerar como os modelos multinível standard ou normais, assumem a variável dependente com uma distribuição Gaussiana. Nos casos em que a variável dependente não tem esta distribuição, mais concretamente quando a distribuição em causa é discreta – número de sucessos em n provas, e se mantém a estrutura hierárquica, o modelo multinível binomial é mais adequado (Guo e Zao, 2000; Kim *et al.*, 2007).

No ponto de vista conceptual, o modelo multinível binomial é equivalente ao modelo multinível linear com excepção da natureza da variável resposta, a qual terá a distribuição $Y_{ij} \sim Bin(\theta_{ij}, n_{ij})$, e, assumindo apenas a variação da constante entre os grupos do nível 2 (modelo de efeitos fixos), obtêm-se reescrevendo a equação (9) da seguinte forma

$$\text{logit}(\theta_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \gamma_{00} + \sum_{l=1}^L \gamma_{0l} Z_{lj} + \sum_{k=1}^K \gamma_{k0} X_{kij} + u_{0j} \quad (12)$$

O p_{ij} é a probabilidade de resposta igual a um, γ_{00} é uma constante, Z_{lj} o vector de características do nível 2, X_{kij} é o vector de características do nível 1, γ_{k0} e γ_{0l} vectores dos coeficientes de regressão associados, respectivamente, às características do nível 1 e 2, $u_{0j} \sim N(0, \sigma_u^2)$ é o efeito aleatório do 2º nível. A equação (12) representa o

modelo multinível, sendo que as equações correspondentes a cada nível são as seguintes:

$$\text{Nível 1: } \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{0j} + \sum_{k=1}^K \beta_{kj} X_{kij} \quad (13)$$

$$\text{Nível 2: } \beta_{0j} = \gamma_{00} + \sum_{l=1}^L \gamma_{0l} Z_{lj} + u_{0j} \quad (14)$$

$$\text{e } \beta_{kj} = \gamma_{k0}. \quad (15)$$

No caso de um modelo multinível linear existem duas componentes de variação: o erro aleatório no nível individual - $e_{ij}(\sigma^2)$, com a qual se obtém a variância dentro dos grupos, e a do erro aleatório do nível dois - $u_{0j}(\sigma_u^2)$, com a qual se obtém a variabilidade entre grupos. A partir destas duas variâncias é possível calcular o

$$\text{Coeficiente de correlação intra-classes (Goldstein, 1999): } \rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma^2}. \quad (16)$$

No caso dos modelos multinível binomiais, a primeira componente de erro não é incluída na equação (12), devendo, para o apuramento deste coeficiente, ser substituída pela variância de uma distribuição logística standard: $\sigma_e^2 = \frac{\pi^2}{3}$ (Bryk e Raudenbush, 1992).

No caso de modelos não lineares, como é o caso das binomiais, a estimação é semelhante ao caso linear, usando um estimador *pseudo-likelihood* (Goldstein, 1999; Breslow e Clayton, 1993).

Os modelos foram estimados usando o software SAS, mais concretamente a macro GLIMMIX que permite a estimação e inferência estatística de modelos lineares generalizados. O GLIMMIX permite o ajustamento de dados não normais e a modelação de estruturas hierárquicas de efeitos aleatórios, desde que estes tenham uma distribuição normal (Schabenberger, sd.).

Este procedimento emprega, no caso da modelação com efeitos aleatórios, um estimador *pseudo-likelihood*, proposto por Wolfinger e O'Connell (1993) e Breslow e Clayton (1993). Para o ajustamento do modelo assume-se que:

$$E[\mathbf{Y} | \boldsymbol{\gamma}] = g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}), \quad (17)$$

onde $g(\cdot)$ é uma função de ligação monótona e diferenciável, $g^{-1}(\cdot)$ a sua inversa, e os efeitos aleatórios assumem uma distribuição normal com média 0 e matriz de variância G . O procedimento contém um modelo linear misto dentro da inversa da função de ligação ($g^{-1}(\cdot)$), sendo esta componente referida como o preditor linear (SAS, 2006):

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} \quad (18)$$

A técnica de optimização usada por defeito é o método Quasi-Newton (SAS, 2006).

A sintaxe usada para modelar a variável resposta foi *events/trials* (nº aprovações em nº inscrições). Ao assumir esta sintaxe, específica para dados binomiais, o software/macro assume, por defeito, a distribuição binomial e a função de ligação *logit*. Na estimação dos modelos multinível foi usada a opção “between/within” para contabilização dos graus de liberdade nos testes para efeitos fixos.

A modelação multinível apresentada revela-se apropriada para testar as hipóteses de investigação, na medida em que respeita a estrutura de agrupamento em estudo (escolas secundárias/aluno) de uma forma parcimoniosa (Ferrão *et al.*, 2001).

Existem algumas linhas de investigação para aplicação destes modelos numa perspectiva longitudinal (ver por exemplo: Singer, 1998; Ferrão, 2002), que poderá ser interessante numa investigação futura, em que se tome em conta o acompanhamento dos alunos ao longo do seu percurso académico no Ensino Superior.

3.3 Caracterização do objecto empírico

Uma das questões que se coloca no contexto nacional é o quadro de conhecimento que as instituições de Ensino Superior dispõem *a priori* em relação aos seus novos alunos, e como essa informação poderá (ou não) pesar no seu desempenho.

Actualmente, quando um aluno ingressa no Ensino Superior leva consigo um vasto leque de informação, maioritariamente quantitativa, sobre o seu percurso anterior. Essa informação justifica a sua colocação, o seu mérito para o acesso ao curso desejado, mas não só, traduz também um percurso influenciado pela escola secundária onde

desenvolveu a sua preparação académica para o acesso, e que poderá influenciar o seu percurso como aluno do Ensino Superior.

Os dados disponíveis para esta investigação referem-se a uma instituição de Ensino Superior conhecida como uma das mais antigas escolas de Engenharia de Portugal – Instituto Superior Técnico (IST), integrado na Universidade Técnica de Lisboa.

A amostra em causa contempla 1253¹⁸ alunos ingressados naquela instituição em 2004/05 pelo concurso nacional de acesso ao Ensino Superior. Note-se que embora existam outros alunos ali ingressados em cada ano lectivo, para estes não está disponível o mesmo leque de informações sobre o percurso passado, em particular os elementos considerados no modelo a desenvolver, para além dos mesmos representarem uma proporção marginal dos alunos admitidos nesta instituição.

O facto do universo de análise estar circunscrito a apenas a uma instituição de Ensino Superior Universitária (IST) com missão e população singular¹⁹ pode, por um lado, ser considerado uma limitação deste estudo, na medida em que não pode ser generalizado a todas as instituições de Ensino Superior na área de Ciência e Tecnologia; por outro lado, tratando-se de uma população com características específicas permite a definição de um modelo particular para actuação precoce em casos de insucesso de alunos desta instituição, e que foi o móbil para este trabalho.

É importante realçar que dificilmente se podem extrapolar os resultados obtidos empiricamente numa instituição de Ensino Superior para as demais, ainda que no mesmo âmbito (nacional), visto que a envolvente, dinâmica e o público ingressado serão diferentes, até pela imposição de critérios de ingresso mais selectivos (Porter, 2006, p. 523).

A análise de (apenas) uma geração de alunos ingressados poderia também constituir um problema, na medida em que esta pode não ser representativa das gerações vindouras, e desta forma inviabilizar a aplicação de um modelo preditivo. Ora perante este facto, será

¹⁸ Dos quais foram retirados 16 alunos por pertencerem a um *cluster* de escolas que foi excluído da análise e 4 observações *outliers* (alunos com um nº de inscrições muito pequeno/grande face à média).

¹⁹ O “filtro de acesso ao IST” é bastante exigente, não só do ponto de vista do elenco de provas de ingresso exigidas como também pelas classificações mínimas exigidas, tratando-se de uma população com características próprias.

que o universo de alunos usado é uma amostra, correspondente a uma geração entre muitas, ou a população, ainda que ingressada num ano lectivo?

Em Snijders e Bosker (1999) esta questão é abordada, isto é, porquê usar um modelo probabilístico se a população toda foi observada, ou seja se não foi recolhida uma amostra. A utilização de tal modelo (probabilístico) que assume variação estatística, mesmo quando a população toda foi investigada, pode ser justificada por se entender que as conclusões se aplicam não só à população observada como também a uma população mais vasta. Assume-se que a população investigada é representativa de uma população mais vasta, por exemplo anos anteriores ou seguintes. A aplicabilidade do modelo a esta população mais vasta não deve ser automática e tem que ser cuidadosamente discutida, considerando se a população observada é ou não representativa da mesma. A inferência não é assim em relação aos indivíduos mas em relação aos mecanismos e processos sociais, comportamentais.

Neste sentido, foram analisadas várias gerações de alunos ingressados no IST, constatando-se que estas são muito semelhantes do ponto de vista das dimensões de análise que serão consideradas neste trabalho (tais resultados podem ser confirmados por relatórios anuais desenvolvidos no âmbito do Gabinete de Estudos e Planeamento do IST²⁰, nomeadamente nos estudos do ingresso e a na caracterização socioeconómica dos alunos ingressados), salientando que a construção das mesmas teve em conta a coerência dos indicadores entre gerações.

Por exemplo, um dos factores que difere entre as várias gerações é a distribuição de classificações no exame de Matemática, diferenças essas observadas a nível nacional e já referidas no contexto da análise via *clustering* probabilístico. Assim, e embora a classificação no exame de Matemática permita aferir a capacidade do aluno perante uma prova de abrangência nacional, sendo referida por vários autores como pertinentes para o caso de estudo presente nesta investigação:

- disciplina base para qualquer curso ligado a Ciências e Engenharia (Lencastre *et al.*, 2000);

²⁰ <http://gpep.ist.utl.pt/>, consultado em 13/07/2007.

- *background* académico em Matemática pode ser considerado como “trampolim” de acesso a cursos de “banda larga” do Ensino Superior, visto que esta é condicionante da entrada em cursos diversificados e determinante para o seguimento de determinadas áreas vocacionais, tal como referido em Tavares *et al.* (2002) e em Martins e Campos (2005);
- uma boa preparação em Matemática facilita o estudo nas áreas científicas Física, Química e Biologia, como referido em Sadler e Tai (2007),

a mesma não será considerada.

A operacionalização do modelo multinível será assim levada a cabo através de uma realidade concreta com especificidades próprias, mais concretamente através do objecto empírico de uma geração de alunos ingressados no 1º ano do IST.

3.3.1 *Fontes de informação*

As fontes de informação para a amostra em estudo (alunos ingressados no IST em 2004/05) foram as seguintes: dados facultados anualmente pelos serviços de acesso ao Ensino Superior/DGES para cada uma das instituições, dados recolhidos através de um inquérito aos alunos ingressados levado a cabo anualmente pela instituição que cedeu os dados; secretaria de graduação. Os dados facultados revestiram-se de anonimato, sem qualquer alusão a qualquer elemento que possa identificar cada um dos alunos.

3.3.2 *Medição do sucesso escolar*

A aplicação de um modelo multinível para a explicação da hipótese em estudo pressupõe antes de mais a definição do conceito de sucesso.

Mas como se define o que é o sucesso ou insucesso escolar? Este conceito pode ser interpretado de formas variadas. Pode-se considerar como insucesso um aluno que: não transite de ano; ou que não obtenha aprovação em nenhuma disciplina; ou que apenas obtenha aprovação num número mínimo de disciplinas; ou que tenha obtido aprovação num número de disciplinas com uma determinada média. Numa perspectiva longitudinal, pode-se ainda considerar que o insucesso está associado à conclusão do

curso num período temporal superior ao previsto para o efeito. Ou seja, esta problemática começa pela definição do conceito de sucesso escolar e da métrica a utilizar na sua quantificação.

O trabalho de investigação desenvolvido (nacional e internacionalmente) tem incidido sobre o desempenho/rendimento dos alunos no 1º ano do Ensino Superior. As razões são bastante óbvias: o processo de transição e adaptação ao contexto do Ensino Superior são fulcrais para o sucesso do aluno ao longo do curso.

Mas então como medir o desempenho dos alunos? Interessa que concluem todas as disciplinas às quais se inscrevem? Ou interessa que as concluem com boas classificações?

A escolha da variável que traduza o nível do sucesso de um aluno no primeiro ano, como se pretende com este trabalho, não é consensual na bibliografia consultada. Na realidade este conceito tem sido alvo de varias definições, sendo exemplo as seguintes:

- somatório das classificações obtidas no primeiro ano (Pike e Saupe, 2002);
- média, ponderada por um peso, das disciplinas aprovadas no primeiro ano vezes a proporção de disciplinas concluídas face às disciplinas a concluir (Almeida *et al.*, 2004 e Soares *et al.*, 2006);
- variável dicotómica em que o sucesso e o insucesso são ditados pela conclusão, respectivamente, de mais ou menos 50% das disciplinas previstas (Lencastre *et al.*, 2000).

No caso da primeira definição identificam-se vantagens na consideração da quantidade de disciplinas feitas a par com o nível de resultados (quanto mais elevado ou mais disciplinas feitas, melhores os resultados das classificações), contudo, no presente objecto de estudo coexistem estudantes de vários cursos e nos quais estão previstos números de disciplinas diferentes (8, 10, 11), pelo que esta variável não permite comparar níveis de sucesso similares em cursos diferentes.

No segundo caso, a definição adoptada tem a vantagem de combinar a proporção de disciplinas concretizadas com o nível académico do aluno. Na realidade, o sucesso de um aluno, nos termos aqui definidos, traduzir-se-á pelo compromisso entre a quantidade de disciplinas feitas, relativizado pelas que estava inscrito, e as classificações com que as fez.

No terceiro caso, embora se trate de uma interpretação simples do sucesso escolar, perde-se informação sobre o padrão de rendimento académico.

Considerando o sucesso como a razão entre aquilo que se pretende atingir – objectivos, e o que se consegue efectivamente – resultados, estes últimos com níveis de complexidade/qualidade diferentes (Lencastre, *et al.*, 2000), podemos assumir que a medição do sucesso através de variável de natureza mista poderá ser uma opção legítima. Assim, uma hipótese de trabalho para a explicação do sucesso poderia ser uma variável determinada pela razão/distribuição entre a média das disciplinas aprovadas e a taxa de aprovação, com um peso de 50% cada uma, traduzindo de uma forma eficiente o rendimento de um aluno no primeiro ano. Na realidade, desta forma tanto beneficiaria um aluno que faça todas as disciplinas com uma classificação baixa, como um aluno que não faça todas as disciplinas mas que as faça com classificações mais altas.

No entanto, e face ao objectivo estabelecido neste trabalho, pretende-se analisar o sucesso do aluno medido apenas pelos objectivos atingidos face aos propostos do ponto de vista quantitativo, e, como tal, usar-se-á a taxa de aprovação medida pelo número de disciplinas em que o aluno foi aprovado sobre o número de disciplinas em que o aluno se inscreveu.

O conceito de sucesso escolar aqui adoptado será assim a razão de disciplinas concluídas, independentemente da classificação média obtida nas mesmas. A escolha de um indicador apenas com esta componente tem particular importância no actual contexto do Ensino Superior. Actualmente, um aluno necessita de cumprir requisitos mínimos de desempenho sob risco de prescrição, não sendo estes requisitos relacionados com o nível das classificações obtidas nas disciplinas. Por outro lado, o nível de classificações pode ser melhorado através de melhorias de notas no ano seguinte, enquanto que uma baixa taxa de aprovações é mais difícil de recuperar.

A taxa de aprovação, no caso da amostra considerada neste trabalho, é uma variável de resposta discreta, conforme se pode observar na Figura 17, logo a sua modelação deverá ter em conta a sua distribuição.

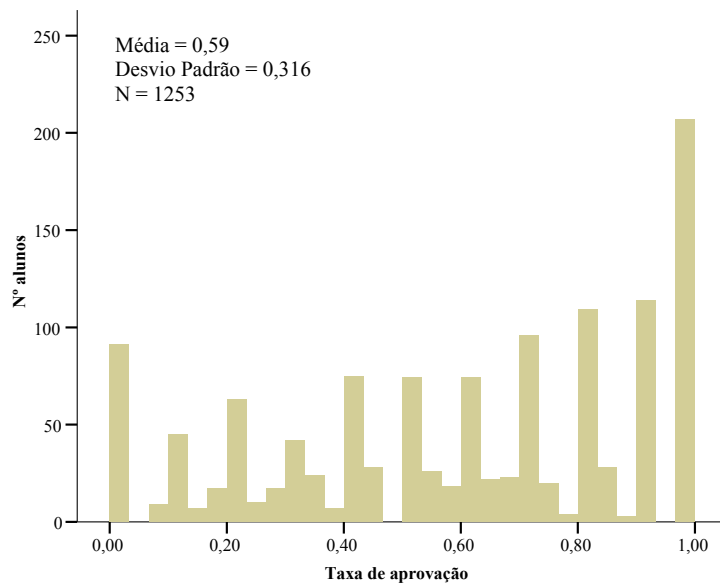


Figura 17: Distribuição da variável rendimento académico (Taxa de aprovação).

Importa referir que muitas das situações em que a taxa de aprovação é baixa, ou mesmo nula, resultam em casos de abandono escolar.

3.3.3 Dimensões preditivas do sucesso escolar

Estabelecida a forma como se vai avaliar o desempenho do aluno no 1º ano, importa agora explorar os eixos dimensionais de análise no nível do aluno. Na prática, proceder à análise das variáveis disponíveis que se considerem importantes/pertinentes para a caracterização do aluno do ponto de vista do seu desempenho escolar no primeiro ano.

Os eixos dimensionais em análise no nível do aluno serão os seguintes (Figura 18): percurso académico anterior, nível socioeconómico e capital familiar, motivações/expectativas, e factores contextuais.



Figura 18: Eixos dimensionais em análise no nível do aluno.

No eixo de análise do *percurso académico anterior* pretende-se avaliar o contributo das aptidões académicas do aluno para o seu sucesso académico no Ensino Superior, o qual será realizado através dos seguintes indicadores:

- **Classificação do 10-12º ano** – esta variável traduz a habilidade académica do aluno do ponto de vista contínuo, isto é nos 3 anos anteriores à entrada no Ensino Superior (Lencastre, *et al.*, 2000), reflectindo ainda uma bagagem de experiências formativas anteriores (Martins e Campos, 2005); A análise do desempenho do aluno no Ensino Superior tem que ter presente que os estudantes chegam a este patamar com 18 anos ou mais de experiência de vida, dos quais 3 de estudo no Ensino Secundário (Tai *et al.*, 2006);
- **Frequência da disciplina de Física no Ensino Secundário** – variável que indica, parcialmente, a vocação do curso frequentado no Ensino Secundário, que poderá traduzir a aptidão para a frequência de cursos ligados à área de Ciência e Tecnologia (Pike e Saupe, 2002).

No eixo de análise *nível socioeconómico e capital familiar*, pretende-se aferir o contributo da herança social e cultural legada pela família, a par com as condições e contextos da sua proveniência social, caracterizados pelas seguintes variáveis:

- **Nível de escolaridade dos pais** – indicador crucial para aferição do acesso a recursos culturais e orientações escolares (Almeida *et al.*, 2004; Amelink, 2005; Martins e Campos, 2005; Cadete e Cardoso, 2006);
- **Sexo** – indicador fiável na predição do sucesso na medida em que as raparigas e os rapazes da faixa etária predominante no acesso ao Ensino Superior (17-19 anos) não só têm vivências académicas diferentes, como também têm posturas

diferentes face ao estudo e à motivação, evidenciando-se a importância do género na determinação das escolhas educativas (Almeida *et al.*, 2004; Amelink, 2005; Lencastre, *et al.*, 2000; Soares *et al.*, 2006; Herzog, 2005).

- **Nível de rendimento do agregado familiar** – a análise do nível socioeconómico familiar tem uma particular preponderância na caracterização dos contextos e recursos e no reconhecimento das trajectórias de vida dos estudantes (Amelink, 2005; Martins e Campos, 2005; Cadete e Cardoso, 2006; Herzog, 2005).

O eixo dimensional associado às *motivações e expectativas* permite uma caracterização no nível de compromisso na prossecução dos estudos, aqui medido através dos seguintes indicadores:

- **Opção de Colocação** – este indicador evidencia o resultado da análise de decisão do ingresso no Ensino Superior no curso pretendido, podendo traduzir um elevado empenho por parte do estudante no caso deste ser o curso de primeira escolha. Por vezes, esta variável pode não traduzir a verdadeira preferência do aluno, visto que esta pode ter sido influenciada por familiares ou simplesmente determinada por um fraco desempenho no Ensino Secundário (Alarcão, 2000, Lencastre, *et al.*, 2000);
- **Momento da escolha do curso e Expectativas de desempenho** – o momento em que o aluno escolheu o curso e as expectativas que tem sobre o rendimento académico pode determinar o grau de interesse na prossecução de um projecto/trajectória escolar, que se traduzirá em diferentes níveis de rendimento académico.

Por último, na dimensão *contextual*, incorporam-se as seguintes variáveis:

- **Fase de ingresso** – este indicador caracteriza o momento em que o aluno iniciou a sua vida académica do 1º ano. No caso dos alunos colocados em 2ª fase esta foi iniciada mais tarde do que os demais;
- **Aluno deslocado da sua residência habitual** – esta variável traduz a estabilidade afectiva dada pelo ambiente familiar e contributo que tal poderá ter no rendimento académico (Herzog, 2005);

- **Tempo de trajecto** – este indicador relacionado com o acesso a uma rede viária facilitadora de deslocação e/ou com a proximidade do estabelecimento de Ensino Superior poderá estar relacionado com o tempo disponível para os estudos e a restrições horárias de permanência na instituição.

Na análise das variáveis disponíveis para inclusão em cada dimensão houve outras que se consideraram como importantes para explicação do fenómeno em estudo, contudo, as mesmas revelavam uma relação forte com as consideradas. As variáveis nesta situação são as seguintes: curso/agrupamento frequentado no Ensino Secundário, que apresenta uma forte relação com a variável frequência de Física no Ensino Secundário; o estatuto de bolseiro dos Serviços de Acção Social que apresenta uma forte relação com o nível de rendimento do agregado familiar; e a intenção de mudar de curso, muito relacionada com a opção de colocação. Estas associações provocavam níveis elevados de multicolinearidade, o que poderia afectar as estimativas, e, subsequentemente, a interpretação do modelo.

A Tabela 9 sintetiza os eixos dimensionais em análise no nível do aluno e as variáveis que os caracterizam.

Tabela 9: Descrição das variáveis incluídas na análise multinível – nível do aluno.

Dimensão	Variável	Medida(s)	Descrição/Distribuição
Percursos acadêmicos anteriores	Classificação 10º-12º ano	Quantitativa	Média – 156 pontos; Desvio padrão – 16,49
	Frequência Física Ensino Secundário	Nominal, dicotómica	Sim (0) – 53%; Não (1) – 47%
Nível socioeconómico e Capital familiar	Sexo	Nominal, dicotómica	Masculino (0) – 76%; Feminino (1) – 24%
	Nível de escolaridade dos pais ²¹	Nominal, dicotómica	Formação não superior (0) – 42%; Formação superior (1) – 58%
	Nível de rendimentos agregado familiar	Nominal, dicotómica	Inferior média nacional ²² (0) – 54%; Superior média nacional (1) – 46%;
Motivações e expectativas	Opção de colocação	Nominal, dicotómica	1ª opção (0) -80%; Outras (1) – 20%
	Expectativas de desempenho	Nominal, dicotómica	Passar a todas as disciplinas com boa média (0) – 62%; Outras (1) - 31%; NR- 7%
	Momento da escolha do curso	Nominal, dicotómica	No ano ingresso (0) – 43%; Outro momento (1) – 57%
Factores contextuais	Fase de ingresso	Nominal, dicotómica	1ª fase (0) -94%; 2ª fase (1) -6%
	Deslocado	Nominal, dicotómica	Não (0) – 66%; Sim (1) – 33%; NR- 1%
	Tempo de trajeto	Nominal, dicotómica	≤ 1h (0) – 67%; >1h (1) - 32%; NR- 1%

Importa salientar que nas áreas de Engenharia e Tecnologia a proporção de mulheres é relativamente pequena (Soares *et al.*, 2006). Neste estudo de caso, numa escola vocacionada para o ensino de Engenharia, tradicionalmente associado ao sexo masculino, constatamos que apenas 24% da população estudantil é do género feminino, o que é coerente.

No nível de análise das escolas importa fazer a integração do *clustering* realizado no capítulo anterior (Capítulo 2- Caracterização das escolas secundárias) com o universo de estudo desta investigação. Na Tabela 10 apresenta-se o número de alunos

²¹ À semelhança do que foi feito em Martins e Campos (2005), a escolaridade dos pais corresponde à categoria mais alta dos cônjuges.

²² Média nacional, cerca de 1500€, estabelecida pelo indicador “rendimento médio anual líquido (€) dos agregados domésticos privados. Fonte: INE.

ingressados no IST em 2004/05 que provieram de cada um dos segmentos de escolas secundárias.

Tabela 10: N° e proporção de alunos ingressados do IST em 2004/05 por segmento.

Segmento	Alunos Ingressados	
	N	%
Escola Pública Urbana	187	15%
Escola Pública Urbana Grande	549	44%
Escola Pública do Interior	43	3%
Escola Pública Urbana Proficiente	331	26%
Escola Privada Urbana Proficiente	127	10%
Escola Técnico-Profissional	16	1%
Total	1253	100%

Verifica-se que o segmento *Escola Pública Urbana Grande* é o maior, com 44% da amostra, e que os segmentos proficientes (público e privado) congregam cerca de 35% dos alunos ingressados.

As variáveis em análise no nível de análise Escola serão as incorporadas na análise levada a cabo para a construção dos segmentos e para este efeito tomarão os valores obtidos através do *clustering* probabilístico (Tabela 11).

Tabela 11: Descrição das variáveis incluídas na análise multinível – nível da escola.

Variáveis	Escola Pública Urbana	Escola Pública Urbana Grande	Escola Pública do Interior	Escola Pública Urbana Proficiente	Escola Privada Urbana Proficiente	Escola Técnico-Profissional
Natureza	Púb. (0)	Púb. (0)	Púb. (0)	Púb. (0)	Priv. (1)	<i>Púb/Priv</i>
Localização	Urb. (0)	Urb. (0)	Rur./semi-urb. (1)	Urb. (0)	Urb. (0)	<i>Urb. (0)</i>
Resultados Matemática	84,3	91,8	79,5	95,1	98,6	83,9
% Docentes com menos 40 anos	49%	30%	66%	31%	55%	50%
% Alunos Cursos Gerais	60%	52%	69%	85%	100%	0%
N° Alunos Ens. Secundário	623	1008	248	573	147	171

O segmento *Escola Técnico-Profissional* foi excluído da análise por ter apenas 16 alunos ingressados, o que causa constrangimentos na estimação do modelo multinível. A exclusão de grupos com menos de 20 observações foi igualmente uma opção tomada em Pike e Saupe (2002).

3.4 *Estimação do modelo*

Com base nos 5 segmentos de escolas e nas características dos alunos delas provenientes procedeu-se à estimação de um modelo multinível, com vista à validação das hipóteses estabelecidas, particularmente a presença de correlação dentro dos segmentos de escolas, causada por características não observadas dos mesmos.

Conforme sugerido, por exemplo, em Hox (1995), o processo de estimação do modelo para a previsão do rendimento académico dos alunos foi realizado em várias etapas, nomeadamente:

- **Modelo binomial “tradicional”**, apenas congregando o nível do aluno e onde é ignorada uma característica fundamental dos dados, que os alunos estão agrupados de acordo com o segmento de escolas de proveniência (princípio de independência das observações);
- **Modelo binomial multinível**, analisando a variabilidade dos parâmetros entre escolas e identificação das características das escolas que estejam relacionadas com esta variabilidade:
 - sem variáveis explicativas do aluno e da escola (teste à existência de variabilidade no desempenho no 1º ano entre escolas);
 - apenas com variáveis explicativas do aluno (analisando se as variáveis de nível do estudante são suficientemente relacionadas com o desempenho no 1º ano);
 - com variáveis explicativas do aluno e da escola (analisando se as variáveis de nível do estudante e da escola são suficientemente relacionadas com o desempenho no 1º ano).

Só desta forma se tem um referencial de comparação e se garante que não existe um acréscimo de complexidade desnecessário.

Os resultados de aplicação estão sintetizados na Tabela 12, sendo apresentadas as estimativas, entre parêntesis os erros padrão, *odds ratio* e o nível de significância das variáveis. O *odds ratio* é usado para interpretar os efeitos reais dos coeficientes estimados e é obtido pela exponenciação do valor do coeficiente associado.

Tabela 12: Resumo dos modelos testados.

	Modelo binomial nulo		Modelo binomial (modelo 1)		Modelo binomial multinível nulo (modelo 2)		Modelo binomial multinível com variáveis explicativas nível 1 (modelo 3)	
	Estimativa	odds ratio	Estimativa	odds ratio	Estimativa	odds ratio	Estimativa	odds ratio
Efeitos Fixos								
Constante	0,453 (0,018)***	1,57	-5,376 (0,225)***	0,01	0,456 (0,027)***	1,58	-5,434 (0,229)***	0,00
Classificação 10º-12º ano	-	-	0,039 (0,001)***	1,04	-	-	0,04 (0,001)***	1,04
Frequência Física Ensino Secundário	-	-	-0,322 (0,042)***	0,73	-	-	-0,324 (0,042)***	0,73
Sexo	-	-	0,091 (0,05)*	1,10	-	-	0,086 (0,05)*	1,09
Nível de escolaridade dos pais	-	-	-0,031 (0,046)	0,97	-	-	-0,028 (0,046)	0,97
Nível de rendimentos agregado familiar	-	-	-0,082 (0,046)*	0,92	-	-	-0,083 (0,046)*	0,92
Opção de colocação	-	-	-0,177 (0,053)***	0,84	-	-	-0,173 (0,053)***	0,84
Momento da escolha do curso	-	-	0,196 (0,042)***	1,22	-	-	0,197 (0,042)***	1,22
Expectativas de desempenho	-	-	-0,096 (0,044)**	0,91	-	-	-0,095 (0,044)**	0,91
Fase de ingresso	-	-	-0,47 (0,287)	0,63	-	-	-0,465 (0,286)	0,63
Deslocado	-	-	-0,185 (0,045)***	0,83	-	-	-0,187 (0,047)***	0,83
Tempo de trajecto	-	-	-0,107 (0,044)**	0,90	-	-	-0,104 (0,044)**	0,90
Efeitos Aleatórios								
Constante ($\hat{\sigma}_u^2$)	-	-	-	-	0,001 (0,003)		0,004 (0,005)	

*** p<0,01; ** p<0,05; * p<0,10;

Os resultados obtidos através do modelo 1 permitem as seguintes constatações:

- As variáveis nível de escolaridade dos pais e fase de ingresso não se revelaram significativas;
- As variáveis incorporadas na dimensão *percurso académico anterior* revelaram-se todas significativas, chamando particular atenção para o impacto da classificação 10º-12º ano no rendimento académico (40% de melhoria do desempenho por cada 10 pontos²³); da frequência de Física no Ensino

²³ Escala da classificação 10º- 12º ano: 0-200 pontos.

Secundário (acréscimo 72% na probabilidade de sucesso escolar, face aos que não tiveram);

- Na dimensão *nível socioeconómico e capital familiar*, as variáveis sexo e nível de rendimentos revelaram-se significativas, verificando-se que as mulheres têm uma probabilidade de sucesso superior à dos homens em 10% e o nível de rendimentos inferior à média aumenta em 8% o rendimento escolar;
- No campo das *motivações e expectativas*, verifica-se que a opção de colocação diferente da primeira (-16%) e o facto do aluno não se empenhar na passagem a todas a todas as disciplinas com uma boa média (-9%) tem um impacto negativo no rendimento académico. As escolhas premeditadas do curso, antes do ano do ingresso, têm um impacto positivo na taxa de aprovação (+22%);
- *Contextualmente*, observa-se que o facto de um aluno ser deslocado da sua residência habitual tem reflexos negativos no seu rendimento académico (um aluno deslocado diminui em 17% o seu rendimento escolar) e que o tempo de trajecto também se reflecte negativamente (um aluno que demora mais de 1h em cada deslocação para o IST diminui em 10% o seu rendimento escolar).

Através dos resultados obtidos através do modelo 2: componente aleatório (constante) não significativa ($|estimativa / erro\ padrão| < 1,96$) e coeficiente de correlação intra classes inferior a 0,001; verifica-se que não existe um efeito do grupo de escolas nos resultados académicos entre grupos de escolas, ou seja, que uma das hipóteses equacionadas nesta investigação não se verifica. Os segmentos de escolas secundárias de proveniência dos alunos não contribuem para a explicação do fenómeno em estudo, revelando que o padrão de rendimento dos alunos não é heterogéneo entre grupos.

As estimativas obtidas através do modelo 3 não diferem muito das obtidas no modelo 1, visto que não se observa heterogeneidade entre escolas (resultados do modelo 2).

Não foi possível incorporar na modelação binomial multinível variáveis explicativas do nível escola, visto que tal incorria num problema de matriz de correlações nula. De acordo com informação disponível na página de suporte técnico do SAS²⁴, este

²⁴ <http://support.sas.com/kb/22/614.html>, consultado em 01/09/2008.

problema sugere que uma ou mais componentes de variância dos efeitos aleatórios é zero, e, como tal, que deve ser excluída do modelo.

Conclui-se assim que o modelo binomial “tradicional” (modelo 1) é suficiente para analisar e prever o rendimento escolar dos alunos. As estatísticas de ajustamento do modelo confirmam estes aspectos (Tabela 13 e Tabela 14) e fornecem outras pistas.

Tabela 13: Estatísticas de ajustamento (*fit statistics*) dos modelos binomiais standard.

	Modelo binomial nulo	Modelo binomial (modelo 1)
-2 Log Likelihood	8431,22	6385,45
AIC (smaller is better)	8433,22	6409,45
AICC (smaller is better)	8433,22	6409,73
BIC (smaller is better)	8438,34	6469,87
CAIC (smaller is better)	8439,34	6481,87
HQIC (smaller is better)	8435,15	6432,27
Pearson Chi-Square	4973,66	3658,07
Pearson Chi-Square / DF	4,04	3,26

Os valores *-2 Log Likelihood* são medidas da qualidade do modelo estimado aos dados e são úteis para comparação dos modelos estimados. Os critérios de informação revelam-se importantes para comparar modelos não hierárquicos (SAS, 2006).

No caso particular dos modelos não hierárquicos (Tabela 13) verifica-se, pela análise dos critérios de informação (AIC, AICC, BIC, CAIC e HQIC), que o modelo que inclui variáveis explicativas relacionadas com o aluno (modelo 1) é melhor para a explicação do fenómeno do que o modelo nulo (modelo que inclui apenas a constante).

Um teste de razão de verosimilhança²⁵ foi usado para determinação do modelo mais ajustado para o conjunto de dados estudado, tanto no caso dos modelos “tradicional” como nos multinível, comparando-se o aumento na função de máxima verosimilhança (*-2 Log Likelihood*) no caso de adição de parâmetros ao modelo a uma distribuição de Qui-quadrado com *g* graus de liberdade²⁶ e probabilidade de erro de 5%. Este teste

²⁵ *Likelihood ratio test – LRT.*

²⁶ Sendo *g* a diferença de graus de liberdade entre os modelos comparados.

comprovou, em ambos os casos, a existência de uma diferença significativa ($p < 0,05$) entre os modelos estudados.

Tabela 14: Estatísticas de ajustamento (*fit statistics*) dos modelos binomiais multinível.

	Modelo binomial multinível nulo (modelo 2)	Modelo binomial multinível com variáveis explicativas nível 1 (modelo 3)
-2 Res Log Pseudo-Likelihood	6166,64	4963,25
Generalized Chi-Square	4971,48	3652,38
Gener. Chi-Square / DF	4,04	3,25

O rácio da estatística Qui-quadrado pelo nº de graus de liberdade (*Gener. Chi-Square / DF*) deveria ser, idealmente, 1. Este rácio é uma medida da variabilidade residual da distribuição marginal dos dados, sendo que valores superiores a 1 significam uma sobre dispersão dos mesmos (SAS, 2006).

Neste caso particular verifica-se que apesar do modelo 3 ter um rácio inferior ao modelo 2, o valor continua a ser superior a 1. Ou seja, os dados apresentam uma dispersão superior ao esperado no caso de um modelo binomial multinível.

As razões para tal dispersão podem ser variadas: os dados podem não seguir uma distribuição binomial; um ou mais efeitos importantes podem não estar a ser considerados no modelo; os dados estão correlacionados positivamente (SAS, 2006:181).

Conforme referido anteriormente, os modelos de previsão do rendimento académico não vão além de uma capacidade explicativa de 30%. Assim, existem alguns efeitos fixos que não estão a ser considerados no modelo, por não ser conhecidas *a priori* ou por não serem mensuráveis, sendo esta a fonte de dispersão dos resultados. Acresce ainda a possibilidade de existirem outras fontes de correlação dos dados não testadas.

3.5 Discussão dos resultados

Os resultados obtidos evidenciam que, no caso da geração analisada, não existe um efeito do grupo de escolas de proveniência no rendimento escolar dos alunos ingressados no IST, não confirmando os resultados obtidos na literatura consultada (Pike e Saupe, 2002; Zvoch e Stevens, 2006). Na realidade, os resultados sugerem que

não existe evidência de existência de uma estrutura hierárquica nos dados, e que o rendimento académico dos alunos poderia ter sido obtido em qualquer um dos segmentos em análise.

Este resultado poderá ser uma consequência da selecção da amostra (*selection bias*), visto que, no caso particular em estudo, os critérios de seriação para o ingresso são bastante restritivos. Por exemplo, a existência de uma classificação mínima nas provas de ingresso e de seriação, implica um filtro selectivo *a priori*. Comparando a distribuição das classificações dos exames de Matemática das escolas secundárias em análise com a distribuição dos exames contabilizados como prova de ingresso pelos alunos da amostra, observa-se a dimensão do filtro aplicado (Figura 19).

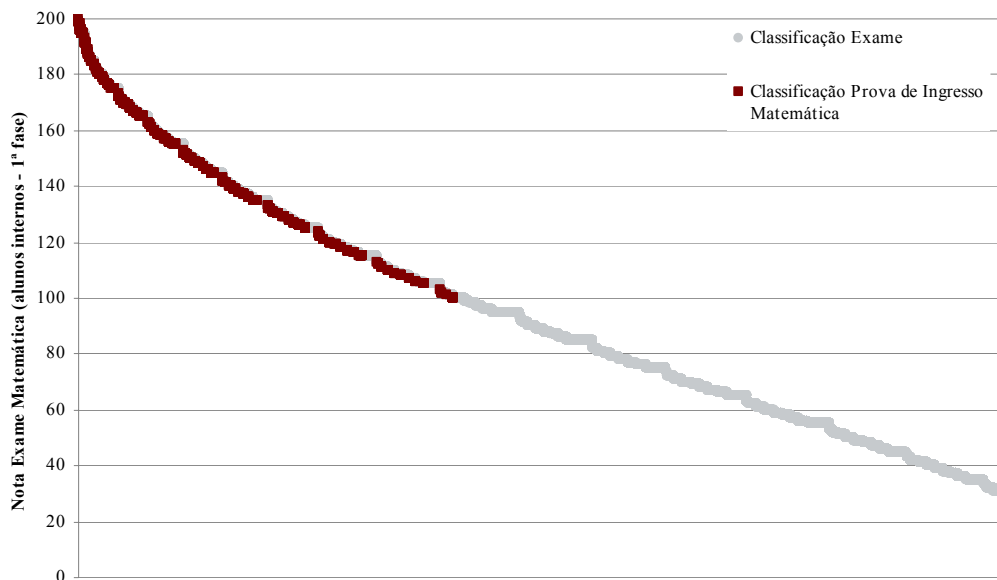


Figura 19: Distribuição das classificações dos exames de Matemática²⁷ e dos exames contabilizados como prova de ingresso para ingresso no IST.

O facto de termos uma amostra seleccionada e de esta não se distribuir proporcionalmente pelos vários grupos de escolas também poderá contribuir para estes resultados (um dos grupos congrega quase metade dos alunos ingressados). Esta questão

²⁷ das escolas secundárias em análise.

(*selection bias*) não é necessariamente um problema visto que um dos objectivo limite desta investigação englobava a construção de um modelo para uma instituição de Ensino Superior específica, a qual a amostra representa.

Existem ainda outros aspectos que poderão estar na origem da ausência do efeito escola, nomeadamente se a construção das dimensões dos segmentos não teve em conta factores importantes para este trabalho e não disponíveis, e, como tal, o número de segmentos obtidos não permite conseguir captar a singularidade dos (grupos) estabelecimentos de Ensino Secundário (Lam *et al.*, 2002).

Os resultados obtidos ao nível das várias dimensões de análise no nível do aluno são na sua maioria coincidentes com os trabalhos desenvolvidos nesta área, evidenciando-se os seguintes: influência positiva do percurso académico anterior, motivações e expectativas, nível socioeconómico e o sexo (Soares *et al.*, 2006; Almeida *et al.*, 2004).

As divergências encontram-se por exemplo na variável habilitações dos pais, que contrariamente ao obtido em Almeida *et al.* (2004) não revelara significância na explicação do sucesso escolar no 1º ano dos alunos do IST. Aqui importa salientar que embora o constructo seja o mesmo – sucesso escolar, a métrica usada foi diferente, não contemplando no presente trabalho, a componente relacionada com as classificações obtidas. Tal poderá sugerir, ou uma diferença nas características da amostra em estudo, ou que existe uma distinção entre os factores que contribuem para o rácio de disciplinas feitas e para o nível de classificações com que os alunos as fazem.

Os resultados obtidos contrariam as hipóteses equacionadas relativamente a algumas das dimensões/variáveis, carecendo de alguma discussão em torno das mesmas:

- Em relação às variáveis que não se revelaram significativas:
 - nível de escolaridade dos pais – será que a capacidade do aluno, medida pelas variáveis que quantificam o percurso académico anterior, não incorporará este efeito? No sentido de validar esta hipótese repetiu-se o primeiro modelo excluindo as variáveis referentes à dimensão percurso académico anterior e esta variável torna-se significante;
 - fase de ingresso – esta variável tinha por objectivo caracterizar o momento de entrada no curso/instituição, e o impacto que isso teria no rendimento académico do aluno. Todavia, é preciso evidenciar que muitos dos alunos que ingressaram em 2ª fase no IST já o tinham feito na

1ª fase noutra curso do IST, e, como tal, não perderam as primeiras semanas de aulas. Outro aspecto importante a evidenciar é que, comparativamente com os alunos ingressados em 1ª fase, existe uma maior proporção de alunos inscritos em menos de 10 disciplinas (média do IST), logo com uma menor carga de trabalho;

- O nível de rendimentos do agregado familiar inferior à média tem um contributo maior para o rendimento escolar. Conforme referido anteriormente, existe uma relação entre o nível de rendimentos e o estatuto de bolseiro (dos serviços de Acção Social), sugerindo que uma menor capacidade financeira, e a necessidade de obter um rendimento escolar mínimo²⁸ para a manutenção dos apoios, poderá funcionar como um incentivo aos resultados académicos.

²⁸ A métrica do sucesso académico considerada neste trabalho é semelhante à dos serviços de Acção Social, isto é baseia-se apenas no rácio de disciplinas concluídas, sem levar em conta as classificações.

4. CONCLUSÕES

A análise dos percursos escolares dos alunos assume actualmente uma grande importância. Graças a esta análise será possível construir programas de intervenção que ajudem os alunos no seu percurso escolar. Somente com programas ajustados às necessidades específicas de cada aluno/escola se poderá intervir no sentido de melhorar as classificações escolares, aumentar o sucesso do aluno (e do estabelecimento de ensino) e diminuir os níveis de abandono escolar. Torna-se, assim, imprescindível não só a realização de relatórios de diagnóstico que possibilitem aferir acerca desta realidade, como também a determinação de factores que possam antecipar/prever este fenómeno.

Nos últimos anos, as instituições de Ensino Superior em Portugal têm tido alguma preocupação com o combate ao insucesso e abandono escolar desenvolvendo trabalhos de investigação em torno do tema, assim como programas de tutorado, que englobam, entre outras, ferramentas de monitorização do desempenho académico. Esta preocupação também se estende ao Governo, visto que em 2006 lançou um concurso para financiar projectos desta natureza²⁹.

Como base de trabalho para estes programas que têm revelado efeitos positivos no rendimento académico dos alunos, é essencial o desenvolvimento de instrumentos que possibilitem a identificação precoce de potenciais situações de insucesso/abandono (Pike e Saupe, 2002).

Nesta dinâmica e no âmbito do presente trabalho estabeleceram-se como perguntas de investigação:

- será o rendimento escolar no Ensino Superior parcialmente explicado pelo percurso académico anterior, em particular pela Escola Secundária onde o aluno fez este percurso?
- será que o rendimento escolar no Ensino Superior é apenas explicado por características intrínsecas ao aluno?

²⁹ Despacho conjunto N° 295/2005, publicado no Diário da República - II série, N° 69, em 8 de Abril de 2005.

- será possível modelar o rendimento escolar do aluno com base nas dimensões em análise por forma a estabelecer programas de apoio precoce a potenciais situações de insucesso?

Na primeira questão a resposta é afirmativa, o percurso académico anterior contribui para o sucesso académico do aluno, mas, no caso da amostra em estudo, o agrupamento de escolas de onde provêm não contribui para essa explicação. Mesmo maximizando a heterogeneidade entre as escolas (Capítulo 2), não se detectou diferenças de efeito da escola. Este facto mostra a grande homogeneidade de percursos na população em estudo.

De entre as dimensões estabelecidas para a explicação do rendimento do aluno, muitas são as que contribuem para o seu sucesso académico, na sua grande maioria intrínsecas ao aluno, as quais serão usadas para prever o rendimento escolar do aluno apenas com base nas características conhecidas no momento do ingresso.

4.1 Limitações do estudo

As limitações apresentadas dizem respeito aos dois níveis de análise realizada: a referente aos *clustering* efectuado sobre as escolas secundárias e a que se refere aos resultados do caso empírico que esteve no centro deste trabalho, o modelo multinível.

A análise da heterogeneidade dos estabelecimentos de Ensino Secundário, via modelos de *clustering* probabilístico, incorporou apenas dados recolhidos junto de fontes secundárias. Se por um lado se tratam de elementos objectivos de caracterização dos estabelecimentos, que poderão ser recolhidos com relativa facilidade para afectação de futuras observações (escolas); por outro lado a sua recolha não envolveu estratégias metodológicas de tipo extensivo, como por exemplo inquéritos por questionário ou entrevistas às escolas secundárias.

Ao nível da escola existem factores que poderão influenciar a preparação dos alunos para o Ensino Superior, como o bom ambiente (não hostil) e empenho dos professores, associações de pais activas e dinâmicas e interessadas, e que não foram considerados. A recolha de elementos desta natureza implicaria o desenvolvimento de trabalho de campo afim de melhor captar a realidade, condições, dinâmicas, processos, estrutura e

estabilidade do corpo docente de cada escola, à semelhança do realizado em Tavares e Almeida (2006).

Esta opção (recolha de informação adicional por questionário/entrevista) estenderia o tempo previsto para este projecto, sem garantia de resposta por parte da totalidade de estabelecimentos incorporados nesta análise. Todavia, esta opção poderá ser tomada para um aprofundamento da investigação nesta temática.

Outra limitação prende-se com o facto de apenas terem sido consideradas escolas em funcionamento em Portugal Continental. Esta delimitação deveu-se à inexistência de informação disponível para os arquipélagos dos Açores e da Madeira.

No contexto da modelação por *clustering* probabilístico houve ainda um factor que não foi tomado em conta (porque não foi alvo de recolha) e que poderá ser relevante para análises futuras: a oscilação dos resultados dos exames de Matemática de um ano para outro (apenas foram considerados os resultados do ano em análise). Embora tal possa ser considerado como uma limitação, na realidade a oscilação de resultados dos exames do Ensino Secundário de um ano para outro não incorporaria alterações na caracterização dos *clusters*, sendo esta uma mais-valia, por oposição a considerar as escolas nominalmente, na análise levada a cabo.

No âmbito do trabalho empírico realizado no contexto da análise multinível identificam-se as seguintes limitações:

- universo de análise circunscrito a apenas a uma instituição de Ensino Superior Universitária (IST) e análise de (apenas) uma geração de alunos ingressados, conforme discutido no ponto 3.3 (p. 50);
- o facto de não se ter feito uma validação do modelo estimado com uma nova geração de alunos ingressados ou com um subconjunto da amostra. Em relação à segunda hipótese, tal não seria possível visto que tal diminuiria o tamanho da amostra e que poderia levantar problemas de estimação;
- Este modelo apenas se aplica a alunos ingressados pelo concurso nacional de acesso ao Ensino Superior com informação da Escola Secundária.

4.2 *Investigação futura*

Com vista à concretização do objectivo limite deste trabalho (desenvolvimento de um sistema de alerta precoce para o insucesso dos alunos ingressados no IST), o próximo passo deverá ser a aplicação do modelo a gerações subsequentes à utilizada para a construção do modelo, no sentido de o validar.

Os resultados permitem inferir que, tendo em conta os “filtros” de seriação adoptados pelo IST, os sistemas de alerta precoce para alunos ingressados nesta instituição poderão subavaliar a influência da escola secundária de origem em detrimento de outros factores caracterizadores não só dos estudantes a ingressar (desempenho anterior, proveniência socioeconómica, etc.), como dos cursos em que estão a ingressar (carga horária, estrutura dos currículos do 1º ano, dimensão da população do curso, etc.).

Um próximo passo no sentido de dar continuidade à investigação desenvolvida, deverá ser uma análise do contexto académico em que o aluno se insere quando ingressa, considerando, por exemplo, o efeito do curso, e respectivas características, na predição dos resultados académicos dos alunos.

Esta hipótese tem sido alvitrada em algumas das referências bibliográficas consultadas, sendo exemplo disso:

- a dimensão curricular, em Martins e Campos (2005) foi identificada como uma das razões para o insucesso curricular;
- o curso, em Lencastre, *et al.* (2000) concluiu-se que a realidade entre cursos da mesma instituição, ainda que dentro da mesma área, pode ter subjacente populações com vivências diferentes. Em Porter (2006) foi estudada a relação entre as motivações dos alunos e as características institucionais/cursos;
- o modelo pedagógico, em Nasser e Hagtvvet (2006) foi usado um modelo multinível para analisar a relação entre as características dos estudantes, docentes e cursos e o seu impacto nos resultados pedagógicos, e no qual foram incluídas variáveis como o tamanho das turmas.

Sendo o universo de estudo deste trabalho apenas os alunos ingressados num estabelecimento de Ensino Superior, com todas as limitações/vantagens (já discutidas) que tal implica, outra linha de investigação poderia ser a extensão deste estudo ao

universo de alunos ingressados no Ensino Superior em todos os cursos de Engenharia. Desta forma, seria possível analisar a importância da Escola Secundária no desempenho dos alunos no Ensino Superior na área de Engenharia, permitindo confirmar ou infirmar os resultados obtidos neste projecto para este estudo de caso.

5. BIBLIOGRAFIA

- Alarcão, I. (2000), Para uma conceptualização dos fenómenos insucesso/sucesso escolares no ensino superior, em Tavares, J., Santiago, R.A (2000), *Ensino Superior (in)sucesso académico*, Porto Editora, 11-24.
- Almeida, L.S., Gonçalves, A., Soares, A.P., Marques, A.P., Machado, C., Fernandes, E., Machado, J.C., Casal, M., Vasconcelos, R. (2004), *Transição, Adaptação e Rendimento Académico de Jovens no Ensino Superior*, manuscrito policopiado.
- Amelink, C. (2005), *Predicting Academic Success among First-Year, First Generation Students*. Tese de Doutoramento em Educational Leadership and Policy Studies, Virginia Polytechnic Institute and State University.
- Bailey, T., Calcagno, J., Jenkins, D., Leinbach, T., Kienzl, G. (2006), Is student-right-to-know all you should know? an analysis of community college graduation rates, *Research in Higher Education*, Vol. 47 N° 5, 491-519.
- Bozdogan, H. (1987), Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions, *Psychometrika*, Vol. 52, N° 3, 345-370 [citado em Pontes e Dias (2007)].
- Breslow, N., Clayton, D. (1993), Approximate Inference in Generalized Linear Mixed Models, *Journal of the American Statistical Association*, Vol. 88, N. 421, 9-25.
- Bryk, A., Raudenbush, S. (1992), *Hierarchical linear models: Applications and data analysis methods*, Newbury Park, CA: Sage Publications.
- Cadete, N.P., Cardoso, M.G.M.S. (2006), Algoritmo Cart: Previsão do desempenho na Matemática do Secundário, *Revista de Ciências de Computação*, Vol. I, Ano I, N° 1, 11-27.
- Caldas, S., Bankston, C. (1999), Multilevel examination of student, school, and district-level effects on academic achievement. *The Journal of Educational Research*, Vol. 93 N° 2, 91-100
- Carvalho, H., Lourenço, C. (2006). Modelos Multinível e Dados Hierárquicos, Actas do XIII Congresso da Sociedade Portuguesa de Estatística.
- Dias, José G. (2003), Introdução aos Modelos de Mistura Finita em *Temas em Métodos Quantitativos 3*, Edições Silabo, 101-116.
- Dias, José G. (2004), *Finite Mixture Models: Review, Applications, and Computer-intensive Methods*. Research School Systems, Organisation and Management (SOM), Groningen University, The Netherlands.
- Ferrão, M.E., Leite, I., Beltrão, K. (2001), Introdução à modelagem multinível em Avaliação Educacional, Escola Nacional de Ciências Estatísticas - Instituto Brasileiro de Geografia e Estatística (Ministério do Planejamento, Orçamento e Gestão), Rio de Janeiro, Consultado em: <http://www.mepsyd.es/cide/espanol/investigacion/rieme/documentos/files/varios/tutorial.pdf>, em 18/10/2007.
- Ferrão, M.E. (2002). Modelo multinível de resposta discreta para dados longitudinais: uma aplicação aos dados da Pesquisa Mensal de Emprego. Actas do IX Congresso da Sociedade Portuguesa de Estatística.

- GPEARI, Gabinete de Planeamento, Estratégia, Avaliação e Relações Internacionais/Ministério da Ciência e Ensino Superior (2007), *Sucesso Escolar no Ensino Superior: Diplomados em 2004/2005*, Consultado em: http://www.estatisticas.gpeari.mctes.pt/docs/ficheiros/SUCESSO_ESCOLAR_ES04_05.pdf, em 15/09/2007.
- Grácio, S. (1997), *Dinâmicas da escolarização e das oportunidades individuais*. Lisboa: EDUCA – Formação [citado em Martins e Campos (2005)].
- Goldstein, H. (1999), *Multilevel Statistical Models*, London: Institute of Education, Multilevel Models Project, Consultado em: http://www.ats.ucla.edu/stat/examples/msm_goldstein/goldstein.pdf, em 18/10/2007.
- Guo, G., Zhao, H. (2000), Multilevel modeling for binary data, *Annual Review Sociology*, N° 26, 441–462.
- Herzog, S. (2005), Measuring determinants of student return vs. dropout/stopout vs. transfer: a first-to-second year analysis of new freshmen, *Research in Higher Education*, Vol. 46 N° 8, 883-928.
- Hox, J. (1995), *Applied Multilevel Analysis*, Amsterdam: TT Publishers, Consultado em: <http://www.geocities.com/joophox/publist/amaboek.pdf>, em 19/12/2007.
- Instituto Nacional de Estatística (1998), *Tipologia de Áreas Urbanas*. Lisboa.
- Instituto Nacional de Estatística (2007), *O País em Números*. Lisboa.
- JNE/GIASE (2004), Base de Dados ENES - 2005 contendo os resultados dos exames nacionais do ensino secundário (2004/2005), Consultado em: <http://www.dgidec.min-edu.pt/jneweb/estat.htm>, em 01/06/2007.
- Justino, D. (2005), *No silêncio somos todos iguais*. Gradiva Publicações.
- Kim, D., Lee, Y., Washington, S., Choi, K. (2007), Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models. *Accident Analysis and Prevention*, N° 39, 125–134.
- Lam, Y., Wong, K., Ho, L. (2002). School effectiveness of a streamed-school system: a multilevel modelling of the Hong Kong secondary schools, *Australian Journal of Education*, Vol. 46, 287-304.
- Lencastre, L., Guerra, M.P., Lemos, M.S., Pereira, D.C. (2000), Adaptação dos alunos do 1º ano das licenciaturas da Faculdade de Ciências da Universidade do Porto, em Tavares, J., Santiago, R.A (2000), *Ensino Superior (in)sucesso académico*, Porto Editora, 73-106.
- Magdison, J., Vermunt, J. (2002), Latent class models for clustering: a comparison with K-means, *Canadian Journal of Marketing Research*, Vol. 20, 37-44.
- Martins, S., Campos, J. (2005), Entre o secundário e o superior: trajetórias e orientações escolares dos estudantes recém chegados ao IPS, *Interações*, N° 1, 125-148. Consultado em: <http://www.eses.pt/interaccoes>, em 06/12/2007.
- Missão para a Sociedade da Informação (1997), *Livro Verde para a Sociedade da Informação em Portugal*, Ministério da Ciência e da Tecnologia Português.
- Montmarquette, C., Mahseredjian, S., Houle, R. (2000), The determinants of university dropouts: a bivariate probability model with sample selection, *Economics of Education Review*, Vol. 20, 475–484.

- Muthén, B. (1991), Multilevel factor analysis of class and student achievement components, *Journal of Educational Measurement*, Vol. 28, Nº 4, 338-354, em Nasser e Hagtvvet (2006).
- Nasser, F., Hagtvvet, K. (2006), Multilevel analysis of the effects of student and instructor/course characteristics on student ratings, *Research in Higher Education*, Vol. 47 Nº 5, 559-590.
- Nóvoa, A. (coord.), Curado, A., Machado, J. (2005), Percursos Escolares dos Estudantes da Universidade de Lisboa - Estudo 1 - Factores de Sucesso e Insucesso na Universidade de Lisboa, Lisboa, Consultado em: <http://www.ul.pt/pls/portal/docs/1/86780.PDF>, em 12/08/2007.
- Pike, G., Saupé, J. (2002), Does high school matter? An analysis of three methods of predicting first-year grades. *Research in Higher Education*, Vol. 43, Nº 2, 187-207.
- Pontes, L., Dias, J. (2007), Análise da heterogeneidade nas motivações para o empreendedorismo: Uma aplicação de modelos de classes latentes com variáveis concomitantes. Actas do XV Congresso Anual da SPE, Lisboa, 451-460.
- Porter, S. (2006), Institutional structures and student engagement. *Research in Higher Education*, Vol. 47 Nº 5, 521-558.
- Sadler, P., Tai, R. (2007), The two high-school pillars supporting college science, *Science*, Vol. 317, Nº 5837, 457-458.
- SAS (2006), *The GLIMMIX Procedure*, Consultado em: <http://support.sas.com/rnd/app/papers/glimmix.pdf>, em 18/05/2008.
- Schabenberger, O. (sd.), Introducing the GLIMMIX Procedure for Generalized Linear Mixed Models, NC SUGI 30 Proceedings, Paper 196-30, Philadelphia, Pennsylvania, SAS Institute Inc., Consultado em: <http://www2.sas.com/proceedings/sugi30/196-30.pdf>, em 18/05/2008.
- Singer, J. (1998), Using SAS Proc Mixed to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models, *Journal of Educational and Behavioral Statistics*, Vol. 23, Nº 4, 323-355.
- Snijders, T., Bosker, R. (1999), *Multilevel Analysis: An Introduction to Basic and Advanced, Multilevel Modelling*. Sage Publishers, London.
- Soares, A.P., Almeida, L.S., Diniz, A.M., Guisande, M. A. (2006), Modelo multidimensional de ajustamento de jovens ao contexto Universitário (MMAU): Estudo com estudantes de Ciências e Tecnologias versus Ciências Sociais e Humanas, *Análise Psicológica*, Vol. 24, Nº 1 (XXIV), 15-27.
- Sullivan, L., Dukes, K., Losina, E. (1999). Tutorial in Biostatistics an introduction to Hierarchical Linear Modelling, *Statistics in Medicine*, Vol. 18, 855-888.
- Szafran, R. (2001), The effect of academic load on success for new college students: is lighter better? *Research in Higher Education*, Vol. 42, Nº 1, 27-50.
- Tai, R., Sadler, P., Mintzes, J. (2006), Factors influencing college science, Success, *Journal of College Science Teaching*, Vol. 36, Nº 1, 52-56.
- Tavares, J., Santiago, R.A (2000), *Ensino Superior (in)sucesso académico*. Porto Editora.

- Tavares, L.V., Almeida, A. (2006), *Estudo da SEDES sobre caso de sucesso: Melhor aprendizagem da Matemática em Portugal*. Observatório de Prospectiva da Engenharia e da Tecnologia e Centro de Sistemas Urbanos e Regionais do IST, manuscrito policopiado.
- Tavares, L.V., Graça, P.M., Tavares, M.M.V. (2002), *Estudo da SEDES: Assimetrias Regionais do Desempenho Educativo*, Observatório de Prospectiva da Engenharia e da Tecnologia e Centro de Sistemas Urbanos e Regionais do IST, manuscrito policopiado.
- Taveira, M.C. (2000), Sucesso no ensino superior: uma questão de adaptação e de desenvolvimento vocacional em Tavares, J., Santiago, R.A (2000), *Ensino Superior (in)sucesso académico*, Porto Editora, 49-72.
- Teixeira, F. (coord.), Pile, M., Patrocínio, C., Lourenço, L., Mendes, R. (2002), *Análise do Desempenho Escolar em Contextos Educativos Diferenciados*, Projecto apoiado pelo Instituto de Inovação Educacional (Medida 2 - Projecto N° 48/2000, Ref.ª SIQE/02.01), manuscrito policopiado.
- Vermunt, J.K., Magdison, J. (2002), Latent class cluster analysis, em J.Hagenaars and A.McCutcheon (eds.), *Applied latent class analysis*, 89-106. Cambridge University Press, Consultado em: <http://spitswww.uvt.nl/~vermunt/hagenaars2002b.pdf>, em 15/09/2007.
- Vermunt, J.K., Magdison, J. (2005), *Technical Guide for Latent Gold 4.0: Basic and Advanced*. Belmont Massachusetts: Statistical Innovations Inc.
- Wedel, M., DeSarbo, W. S. (1994), A Review of Recent Developments in Latent Class Regression Models em Latent Class Regression Models, R.P. Bagozzi (ed.), *Advanced methods of Marketing Research*, 352-388, Cambridge: Blackwell Publishers.
- Windham, P. (2005), The importance of work and other factors to attrition: a comparison of significance and odds ratio for different outcomes. *Paper presented at the Annual Conference of the Southeastern Association for Community College Research (24th, Asheville, NC, August 6-9, 1995)*, Consultado em: http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/14/18/2d.pdf, em 15/09/2007.
- Wolfinger, R., O'Connell, M. (1993), Generalized Linear Mixed Models: A Pseudo-Likelihood Approach”, *Journal of Statistical Computation and Simulation*, Vol. 4, 233–243 [citado em SAS (2006)].
- Zvoch, K., Stevens, J.J. (2006), Longitudinal effects of school context and practice on middle school mathematics achievement. *The Journal of Educational Research*, Vol. 99 N° 6, 347-356.