Universidade Aberta

Modelos Multi-nível: Fundamentos e Aplicações

Cláudia Catarina Mendes Silva da Cruz

Dissertação apresentada na Universidade Aberta para obtenção do grau de Mestre em Matemática, Estatística e Computação (especialização em Estatística Computacional)

Orientadora:

Prof.ª Doutora Teresa Paula Costa Azinheira Oliveira

Lisboa 2010

UNIVERSIDADE ABERTA

Modelos Multi-nível: Fundamentos e Aplicações

Cláudia Catarina Mendes Silva da Cruz

Aluno nº 802403

Dissertação apresentada na Universidade Aberta para obtenção do grau de Mestre em Matemática, Estatística e Computação (especialização em Estatística Computacional)

Orientadora:

Prof.ª Doutora Teresa Paula Costa Azinheira Oliveira

Lisboa 2010

Resumo

Os modelos multi-nível são uma resposta à necessidade de analisar a relação entre os indivíduos e o meio que os rodeia. Através destes modelos podemos separar o papel de cada uma das características de uma estrutura interactiva complexa, com o intuito de melhorar o conhecimento da realidade, permitindo uma intervenção mais eficiente.

Neste trabalho procuramos explorar os fundamentos dos modelos multi-nível, ilustrando uma aplicação em educação, área pioneira de pesquisa destas metodologias.

Em concreto, apresentamos primeiro uma abordagem teórica e uma revisão histórica, seguidas de uma explicação mais prática da construção deste tipo de modelos, na qual se evidenciam os comandos do SPSS a utilizar para ajustar e interpretar os modelos multi-nível.

Aplicamos esta teoria aos dados de uma escola profissional localizada em Sines, com o intuito de estudar as diferenças entre as classificações médias dos alunos nos cinco cursos, tendo em conta um conjunto de variáveis, quer do nível 1 (alunos), quer do nível 2 (cursos). As observações foram registadas entre 2004 e 2010, sendo a análise de dados e o modelo obtidos com recurso ao SPSS versão 16.0.

Pudemos constatar que não existem diferenças significativas entre os cursos nem entre os professores de Português e Matemática. Foram, no entanto, construídos os modelos, com a introdução das co-variáveis estatisticamente significativas. Verificou-se a existência de diferenças significativas entre as classificações médias dos alunos e as variáveis género, a zona de proveniência, o facto de ter obtido ou não sucesso no curso e as interacções (ano lectivo × sexo) e (ano lectivo × zona de proveniência). Concluímos ainda a não existência de diferenças significativas entre a idade de entrada no curso e o sucesso do aluno.

Não se verificaram diferenças entre o modelo com três níveis e o modelo de classificação cruzada. As co-variáveis significativas foram a idade, o total de módulos em atraso e o professor de Português.

Recomendamos a utilização de uma amostra de maior dimensão e a comparação com outras escolas com características semelhantes à escola aqui em estudo.

Palavras – chave:

- Modelo multi-nível
- Modelo linear hierárquico
- Modelos multi-nível cruzados
- SPSS nos modelos multi-nível
- Modelo multi-nível em educação

Abstract

Multilevel models answer the need to analyze the relation between individuals and surrounding environment. Through these models we can separate each one of the characteristics of a complex interactive structure, aiming to improve knowledge of reality, enabling a more effective intervention.

With this work we intend to explore the fundaments of multilevel models, illustrating an application in education, the pioneering area for these methodologies.

We start with a theoretical approach and an historical revision, followed by a more practical specification of the conception of this type of models, on which it's highlighted the SPSS commands to use to adjust and interpret the multilevel models.

We apply this theory to the data of a professional school located in Sines, in order to study the differences between the average of the students scores in the five courses, taking into account a set of variables, both for level 1 (students) as for level 2 (courses). The observations were registered between 2004 and 2010, having used SPSS version 16.0 to perform data analysis and obtaining the model.

We can establish that there are no significant differences between the courses nor between the Portuguese and Mathematics teachers, nonetheless the models have been developed with the introduction of the statistically significant co-variables. There were, however, significant differences between the average student test scores and the gender variables, the provenance/district , the success rate in the course and the interactions (school year sex) and (school year provenance). We also concluded that there are no significant differences between the age of admission to the course and the student success rate.

There were no differences between the model with three levels and the crossed classification model. The significant variables were age, total modules in arrear and the Portuguese teacher.

We recommend a wider sample to be used and comparison with other schools with similar characteristics to the school under study.

Key words:

- Multilevel model

- Hierarchical linear model

- Crossed multilevel models

- SPSS for multilevel models

- Multilevel model in education

Agradecimentos

Na apresentação do presente trabalho exprimo o meu profundo agradecimento a todos os familiares e amigos que me incentivaram e apoiaram na sua realização.
De uma forma muito especial quero apresentar um obrigado muito grande à minha mãe, ao meu pai e ao meu marido, que muito me apoiaram para que este trabalho se tornasse realidade.
Um agradecimento muito especial aos meus colegas da ETLA pela disponibilidade em ajudar-me em tudo o que foi necessário.
Um muito obrigado à minha amiga Arabela pela preciosa ajuda que me prestou.
Em particular manifesto um vivo e profundo agradecimento à professora doutora Teresa Oliveira pela força, dedicação e orientação dadas ao desenvolvimento deste trabalho, e pelo seu contributo no meu enriquecimento científico e profissional.
Os meus sinceros agradecimentos ao Vítor Valente pelas suas valiosas dicas.

Simbologia e notações

γ_{00} é a grande média, ou seja, a ordenada média para os elementos do nível 2.

γ_{10} é a média dos declives de todos para os elementos do nível 2.

u_{0k} é o erro aleatório para cada elemento do nível 2 (afastamento em relação à ordenada média).

u_{1k} é o erro aleatório de cada elemento do nível 2 (afastamento em relação ao declive médio).

$\tau_{00} = \sigma_{u0}^2$ é a variância populacional das ordenadas.

$\tau_{11} = \sigma_{u1}^2$ é a variância populacional dos declives.

τ_{01} é a co-variância entre as ordenadas e os declives.

ÍNDICES

Índice Geral

Índice de figuras

Índice de tabelas

Índice de gráficos

Capítulo I

1 Enquadramento

1.1 Introdução

Podemos definir um modelo multi-nível como sendo um modelo de regressão, que se diferencia dos modelos tradicionais, por ter em conta os contextos (designados por níveis) em que os indivíduos estão inseridos, considerando as variáveis predictoras[1] dos diversos níveis. Estes modelos permitem ao investigador verificar a adequação do hipotético modelo explicativo, ou seja, estes modelos são adequados para estudar fenómenos cuja compreensão está dependente quer das características dos indivíduos, quer da sua própria organização.

Os modelos multi-nível são portanto uma resposta à necessidade de analisar a relação entre os indivíduos e o meio que os rodeia. Através desta metodologia podemos separar o papel de cada um das características de um contexto, com o intuito de melhorar o conhecimento da realidade, podendo, caso necessário, intervir de forma mais eficiente.

A adopção de modelos multi-nível para o estudo e compreensão de fenómenos que ocorrem no contexto organizacional começa a ser uma prática cada vez mais frequente, quer em Portugal quer no estrangeiro. Já há mais de 30 anos que se discutem as contribuições deste tipo de modelos. Foi em 1988 que se organizou um dos primeiros encontros internacionais sobre pesquisa multi-nível, com o objectivo de comemorar 10 anos de existência do Grupo de Pesquisa Multi-nível da Holanda (Klein e Kozlowski (2000).

Apesar destes modelos se terem desenvolvido nos anos oitenta do século passado, foi com o desenvolvimento de programas computacionais capazes de resolvê-los que o seu uso se expandiu a muitas áreas das ciências sociais (Catalán e tal, 2003), entre outras.

Tradicionalmente, quando se ignorava o uso dos modelos multi-nível, os estudos sobre comportamento organizacional eram realizados sob a forma de uma perspectiva micro ou macro. Isto é, os modelos macro debruçavam-se sobre aspectos gerais da organização, não tendo em conta as diferenças existentes no nível de indivíduos. Por outro lado, os modelos micro davam ênfase às diferenças individuais, não contemplando

[1] Ou variáveis explicativas

as características do contexto macro. A contribuição principal dos modelos multi-nível é o reconhecimento do papel do contexto (nível macro) na compreensão de comportamentos relativos do nível micro.

Em conversas informais é cada vez mais comum ouvirmos os pais e educadores comentarem que cada vez há mais desinteresse e, consequentemente, insucesso por parte dos alunos. Mas será que a idade do aluno tem influência neste desinteresse? E o sexo? E o facto de residir em zonas mais rurais? E o professor, poderá ter alguma influência?

A abordagem do tema da modelação multi-nível neste trabalho deve-se ao facto de se reconhecer que, grande parte das investigações em educação, proporciona uma base de dados que se podem agrupar em estruturas hierárquicas, decompondo-se em níveis. Assim, os indivíduos em estudo (unidades de nível 1) podem pertencer a grupos ou unidades maiores (unidades de nível 2, 3 ou 4). Por exemplo, a escola a que os alunos pertencem, a turma em que estão inseridos, o professor que lhes lecciona determinada disciplina.

Para se proceder ao estudo do insucesso e desempenho escolar dos alunos, será necessário que nos debrucemos sobre este fenómeno no seu conjunto, conhecendo, de forma sucinta as suas causas e consequências e as suas implicações no comportamento do aluno, de forma a, através de estudos práticos a realizar no terreno, diga-se escola, podermos tirar as nossas próprias conclusões.

Pretende-se assim, neste trabalho, saber em que medida a escolha do curso e as características da turma e do próprio aluno contribuem para o sucesso escolar e rendimento escolar do aluno.

Esta dissertação está organizada em cinco capítulos. O capítulo I faz o enquadramento e a revisão da literatura mais importante. No capítulo II indicam-se alguns estudos consultados e desenvolvem-se os conteúdos teóricos e as respectivas metodologias estatísticas. O capítulo III faz referência a algumas extensões do Modelo Linear Hierárquico (MLH) com dois níveis. No capítulo IV fazemos a análise descritiva dos dados. Finalmente, no capítulo V ajustam-se os modelos através do software estatístico SPSS versão 16, interpretando-se os respectivos resultados.

1.2 Revisão de literatura e retrospectiva histórica

Muitos são os factores que podem influenciar o desempenho escolar dos alunos. Segundo Ferrão e Fernandes (2000) e Soares (2004), estas influências podem ser classificadas em três grandes grupos: os associados à família (características socioeconómicas e culturais), os associados a factores escolares (infra-estruturas, práticas didácticas, características dos professores) e aqueles relacionados ao próprio aluno (habilidade, motivação).

Diversos sociólogos realizaram estudos nesta área. Durante os anos de 1950 e 1960, em países como os Estados Unidos, Inglaterra e França, concluiu-se em vários trabalhos de investigação que os factores que mais influência têm no desempenho escolar dos alunos são os factores extra-escolares, ou seja, os resultados escolares encontram-se directamente relacionados com as características socioeconómicas e culturais dos alunos.

Bourdieu e Coleman (1966) introduziram o conceito de capital, no sentido em que se refere não apenas à sua forma económica, mas também à sua forma cultural e social. O termo da área económica "capital" foi utilizado por estes sociólogos no estudo das desigualdades escolares, como metáfora para analisar as vantagens culturais e sociais que indivíduos ou famílias possuem e, consequentemente, os conduzem a um nível socioeconómico mais elevado. Bourdieu (1977) concluiu nos seus estudos que a origem social dos alunos se traduz em desigualdades escolares.

No entanto, novos estudos – sendo pioneiro o "*Fifteen Thousand Hours*" (Rutter et al., 1979), questionam a decisão de se concluir que o factor escola praticamente não influencia no desempenho escolar. Neste sentido, estudos recentes mostram a grande importância da escola nos resultados dos alunos, surgindo desta forma o conceito de **escola eficaz**.

Um estudo realizado por Soares, Sátyro e Mambrini (2000) acerca dos factores da eficácia escolar conclui que uma escola é eficaz quando possui uma equipa de professores qualificados. Estes autores consideram ainda que a formação do professor está directamente relacionada com a formação dos alunos e afirmam que "uma escola se configura como eficaz quando possui características que garantem a efectividade e a

eficácia do seu ensino, produzindo reflexo positivo no progresso académico e no desempenho escolar".

Ferrão e Fernandes (2003) definem escola eficaz como uma escola que assume uma parcela de responsabilidade nos resultados escolares atingidos pelos alunos. Também Murillo (2003) define que "uma escola é eficaz se consegue um desenvolvimento integral de todos e cada um dos seus alunos maior do que seria esperado tendo em conta o seu rendimento prévio e a situação social, económica e cultural das famílias". Desta forma, nota-se um grande interesse em investigar os factores que distinguem as escolas eficazes das demais.

Como já referimos, a investigação multi-nível teve a sua origem no campo da educação. Nomeadamente teve por base um estudo bastante conhecido sobre alunos de uma escola primária, levado a cabo em finais dos anos 70 (Bennett, 1976), em que se pretendia investigar se os alunos sujeitos ao chamado estilo formal (tradicional) de ensino apresentavam maior progresso do que os alunos que não estavam sujeitos a este método de ensino (método progressivo). Os dados foram analisados com auxílio das técnicas tradicionais de regressão múltipla, que consideraram os alunos como unidades individuais de análise e não foram agrupados considerando, por exemplo, professores ou turmas. Os resultados foram estatisticamente significativos. Posteriormente, Aitkin et al (1981) demonstraram que quando a análise considerava agrupar os alunos em turmas, as diferenças significativas desapareciam e os alunos sujeitos ao ensino formal não demonstravam diferenças relativamente aos outros alunos. Esta "reanálise" é o primeiro exemplo importante de uma análise de níveis múltiplos de dados nas ciências sociais. Assim, só a partir desta altura é que se começou a ter em conta a estrutura organizativa dos dados. Foi então que Aitkin e Longford (1986), dois matemáticos ingleses, escreveram um artigo que veio revolucionar o mundo da investigação educativa. Neste, foi demonstrado que os modelos de regressão linear usados para estudar a forma como um conjunto de variáveis explicavam uma variável produto, só poderiam ser empregue num caso muito especial: quando as observações eram independentes (Gelman e Hill, 2007; Goldstein, 2003; Heck e Thomas, 2000; Hox, 1998). No entanto, a nossa realidade, na qual os estudantes são agrupados em turmas ou cursos, diferentes turmas estão agrupadas em escolas e as escolas em distritos, ou regiões, ou países, não é compatível com a imposição de independência das observações.

A partir desta análise crítica, Aitkin e Longford (1986), propuseram uma técnica de análise que marcou a investigação educativa desde então: os modelos multi-nível. Apesar dos esforços e das diferentes propostas para solucionar de forma correcta problemas de modelação com dados organizados em diferentes níveis, foi apenas no final dos anos 80 que alguns estatísticos ingleses e americanos, nomeadamente Harvey Goldstein e Stephen Raudenbush, propuseram soluções de extrema importância para este tipo de modelação: *softwares* fáceis de manusear (*HLM* ou *MLwiN*).

Desde então tem havido um crescente desenvolvimento destas técnicas. Por exemplo, no espaço de 11 anos houve quatro revistas internacionais prestigiadas que destinaram números inteiros a aplicações e desenvolvimento da Modelação Multi-nível (ver *International Journal of Education Research*, 1990; *Journal of Education and Behavioral Statistics*, 1995; *Counseling Psychologist*, 1999; *Multivariate Behavioral Research*, 2001).

Segundo Hox (2002), o modelo de regressão multi-nível é também conhecido por modelo linear hierárquico - HLM – (por Raudenbush e Bryk, 1986, 1992), Modelo de componentes de variância (Longford, 1987) ou Modelo de coeficientes Aleatórios (por Leeuw e Kreft, 1986; Longford, 1993). Mais adiante veremos a descrição de cada um destes modelos.

Na opinião de Bryk e Raudenbush (1992), estes modelos não são uma solução para todos os problemas, no entanto, representam um grande passo para o auxílio das análises, uma vez que são estatisticamente correctos e não desperdiçam informação. Actualmente, a análise multi-nível é utilizada nas mais diversas áreas, como iremos investigar em seguida.

Neste trabalho pretende-se investigar uma aplicação, com estudo da influência dos factores escolares numa escola profissional do Litoral Alentejano, isto é, verificar se a ETLA é uma escola eficaz. Assim estudaremos a contribuição da escola, nomeadamente das infra-estruturas, das práticas didácticas e das características dos professores, no desempenho escolar dos alunos. Não descurando, obviamente, as características intrínsecas do próprio aluno. Para a realização deste estudo, utilizaremos uma modelação multi-nível, tendo como base, no nível 1 os alunos e no nível 2 os cursos em que os alunos estão inseridos.

Capítulo II: Desenvolvimento de conteúdos teóricos envolvidos. Detalhe das Metodologias estatísticas a usar e indicação de alguns estudos consultados

2 Modelos multi-nível: generalidades

2.1 Modelos multi-nível: áreas de aplicação e fundamentos

2.1.1 Áreas de aplicação

O desenvolvimento de modelos multi-nível tem viabilizado a análise de estudos que integram indivíduos dentro dos seus grupos ou contextos sociais, examinando os efeitos combinados tanto das variáveis individuais como das de grupos. Assim, a principal importância de utilizar os modelos multi-nível é a possibilidade de investigação da interacção das variáveis nos diferentes níveis. Estruturas hierárquicas podem ser encontradas, por exemplo, em estudos internacionais, nos quais os indivíduos são aninhados nas suas unidades nacionais (países); na pesquisa organizacional, em que os indivíduos são agrupados em departamentos, e estes, por sua vez, em organizações; na pesquisa familiar, com os membros agrupados em famílias; e na pesquisa metodológica, quanto aos efeitos do entrevistador, com os entrevistados agrupados por entrevistadores.

Assim, actualmente, a análise multi-nível está a ser utilizada, entre outras, em áreas tão distintas como a demografia, a sociologia, a saúde e epidemiologia, a educação e o desporto, conforme passamos a exemplificar.

- Na **demografia** por exemplo, se os especialistas desejarem examinar de que forma as diferenças no desenvolvimento da economia nacional podem interferir na relação entre o grau educacional dos adultos e a taxa de fertilidade. Estas pesquisas combinam indicadores económicos recolhidos a nível nacional com informações recolhidas a nível domiciliar sobre a educação e a fertilidade. Desta forma, domicílios e países são unidades na pesquisa, estando os domicílios aninhados nos países.

- Na **sociologia**, por exemplo, Soares e Alves (2007) analisaram o impacto dos processos escolares sobre os resultados dos alunos com o intuito de diminuir a diferença de resultados escolares entre grupos sociais.

- Na **saúde**, a análise multi-nível também se encontra bem presente. Por exemplo, Cruz (2008) evidencia uma investigação **epidemiológica**, na qual a

estrutura dos indivíduos está organizada hierarquicamente. Nesta investigação, salienta as relações mais ou menos evidentes entre a saúde dos indivíduos e a zona geográfica onde habitam, ou entre o tratamento recebido pelos pacientes e as características do médico e/ou do serviço de saúde no qual são atendidos. Ainda na área da **saúde**, distingue-se o estudo de Gómez (2009) no qual relaciona a despesa na farmácia com o médico de clínica geral, com aplicação da análise multi-nível. O estudo foi realizado com dados individuais da população que integra o sistema público de saúde espanhol na comunidade Autónoma de Aragón no ano de 2004. Conclui-se que, para a cidade de Saragoça, os resultados mostram a existência de um "efeito aleatório" significativo do médico de clínica geral na despesa que os pacientes realizam na farmácia. Salienta-se ainda a investigação realizada por Sichieri e Moura (2009), na qual se faz uma análise multi-nível das variações no índice de massa corporal entre adultos, segundo factores individuais e características ambientais das cidades. A conclusão desta investigação é que, embora haja grande discrepância nas médias de IMC entre as cidades brasileiras, a existência de local para actividade física, características económicas e de consumo alimentar pouco explicaram a variação no IMC. As mesmas autoras, em 2010, fizeram ainda um estudo sobre a relação entre a baixa estatura e a obesidade no Brasil, utilizando mais uma vez, uma análise multi-nível.

- Na **educação** as análises hierárquicas também são muito frequentemente utilizadas. Por exemplo, na medida da classificação escolar dos alunos que compõe as turmas. Estas, por sua vez, compõem as escolas. E as escolas são dirigidas por órgãos administradores. Assim, podemos considerar, por exemplo, quatro níveis hierárquicos, associando-se os alunos ao 1.º nível, as turmas de alunos ao 2.º, as escolas ao 3.º e os órgãos administradores ao 4.º nível. Isto também pode acontecer, por exemplo numa escola, onde o pesquisador pode estar interessado em investigar de que forma diversas variáveis (por exemplo, professor, disciplina, local onde vive) influenciam a classificação escolar do aluno. Neste contexto, Albernaz, Ferreira e Franco (2002) avaliaram, através dos dados do SAEB-1997 (dados do Sistema Nacional de Avaliação da Educação Básica), o efeito de variáveis escolares, tais como a escolaridade do professor e a qualidade da infra-estrutura física no desempenho dos estudantes.

Ainda na área da educação, Valente (2007) utiliza os modelos lineares hierárquicos para fazer um estudo sobre a relevância do apoio da escola nas perspectivas profissionais dos alunos do 10º ano de escolaridade, tendo utilizado os MLH com dois níveis: aluno e escola. Outro estudo de importância relevante na área da educação é o estudo de Murillo (2008), no qual faz referência aos modelos multi-nível como ferramenta para a investigação educativa para estudar quais os factores escolares associados ao rendimento dos estudantes.

- Nas **ciências do desporto** temos o estudo de Maia et al (2003), no qual salienta a necessidade de considerar a natureza hierárquica da informação. Outro estudo relevante em ciências do desporto foi realizado em 2008, no qual Lopes *et al* fizeram uma investigação cujo objectivo foi analisar a associação entre a coordenação motora, a aptidão física e a actividade física, numa amostra seguida longitudinalmente desde os 6 aos 10 anos de idade.

Exemplos de dados hierárquicos menos comuns são pesquisas longitudinais e pesquisas de curvas de crescimento, nas quais, séries de observações distintas são vistas como aninhadas por indivíduos. De grande relevância, destacamos ainda a metanálise, em que os objectos de análise são agrupados em diferentes estudos (Hox, 2002).

Ferrão (2002) realizou um estudo no qual investiga um modelo multi-nível de resposta discreta para dados longitudinais, no qual faz uma aplicação aos dados da Pesquisa Mensal de Emprego. Esta análise de dados longitudinais é realizada através de modelos de regressão multi-nível onde as observações são unidades de nível 1 e os indivíduos são unidades de nível 2.

2.1.2 Fundamentos da análise multi-nível

Como Ferrão (2002) salienta, uma das vantagens de utilizar os modelos multi-nível é a possibilidade de melhor compreensão do processo, devido à decomposição da variância do erro nos diversos níveis. Para além disto, na existência de correlação intra-classe, através da regressão clássica obtemos estimativas de erro padrão muito reduzidas, enquanto a regressão multi-nível dá estimativas mais conservadoras.

Segundo Murillo (2008), os modelos multi-nível representam um dos métodos de análise mais interessantes em investigação quantitativa desenvolvidos nos últimos anos. Representam uma extensão do modelo de regressão tradicional quando as variáveis são analisadas dispostas em vários níveis de agregação. Esta técnica é um tipo de análise de regressão que, simultaneamente, tem em consideração múltiplos níveis de agregação, tornando assim correctos os erros padrão, intervalos de confiança e testes de hipóteses.

Ao ignorarmos o padrão hierárquico que a informação contém, podemos ser confrontados com algumas insuficiências. Considerando um possível exemplo de aplicação na área da educação (no qual se consideram alunos aninhados em escolas), temos diversos aspectos a ter em conta na utilização da modelação multi-nível.

- Heterogeneidade das rectas de regressão

É de esperar que o desempenho médio seja diferente entre escolas (cada escola terá a sua recta de regressão, distintas que são umas das outras). Dado que, em princípio, cada escola está sujeita a um conjunto variado de factores que contribuem para explicar as diferenças encontradas, existe um declive diferente nas respectivas rectas de regressão. Esta variabilidade não deve ser ignorada, qualquer que seja o tipo de análise.

- Ausência de independência nas observações

Facilmente se reconhece que, em geral, os indivíduos pertencentes a um mesmo contexto tendem a ser mais semelhantes no seu comportamento do que os que pertencem a contextos diferentes. A fim de ilustrar a dependência entre observações, pode-se citar o exemplo de pesquisas educacionais nas quais alunos são *aninhados* ou *agrupados* em escolas. Os alunos de uma mesma escola tendem a ser similares, em razão do processo de selecção por esta empregue (por exemplo, devido à sua localização, algumas escolas podem “atrair” alunos de níveis socioeconómicos mais

elevados, enquanto outras atraem alunos de níveis socioeconómicos mais baixos), e do ambiente e da história comuns que os alunos compartilham por frequentar a mesma escola. Assim, ao lidar com variáveis em diferentes níveis, o modelo de regressão tradicional pode não ser o mais adequado, pois não tem em consideração a correlação entre indivíduos associados a um mesmo nível de agregação. Quanto maior for essa correlação maior a inadequação do modelo de regressão tradicional, ou seja, quanto maior essa dependência, mais a análise multi-nível se torna necessária.

A dependência entre as observações é indicada pela chamada correlação intra-classe, a qual representa a homogeneidade num mesmo grupo, e, ao mesmo tempo, a heterogeneidade entre grupos distintos. Por exemplo, a correlação média (expressa na referida correlação intra-classe) nas variáveis medidas nos alunos da mesma escola tenderá a ser mais elevada do que a correlação média das variáveis medidas nos alunos das diferentes escolas. Assim, a homogeneidade entre as observações conduz a estimativas erradas para os erros-padrão. Este tipo de erro é conhecido como o "efeito do delineamento" (Kish, 1965, 1987). Um procedimento de correcção normalmente aplicado consiste em calcular os erros padrão através de métodos de análise tradicionais, estimar a correlação intra-classe entre os inquiridos dentro dos *clusters*, e, finalmente, empregar uma fórmula de correcção para os erros padrão. Hox (2002) aponta que uma amostra de 200 alunos agrupados em 10 classes, com correlação intra-classe igual a 0,10, resulta, após a correcção, numa amostra de tamanho real igual a 69 alunos, ou seja, o tamanho factual da amostra diminuiu 66%.

Sumariamente, como as análises estatísticas assumem a independência das observações, se esta suposição é violada (o que acontece habitualmente nos dados multi-nível), os estimadores dos erros padrão das análises estatísticas convencionais são muito menores, tornando os resultados falsamente significativos. Isto é, uma consequência da dependência entre as observações é a subestimação dos erros padrão dos coeficientes da regressão. Por conseguinte, apesar da regressão múltipla ser uma das técnicas de análise de dados mais utilizadas nas áreas de ciências sociais e humanas, o problema maior em usar a regressão múltipla nestas áreas é o facto de que, muitas vezes, um dos pressupostos centrais, a independência das observações, é violado, pois nas ciências sociais e humanas, os dados recolhidos são frequentemente de pessoas agrupadas em *clusters*. Torna-se, assim, impreterível que qualquer procedimento de análise considere,

em simultâneo, as diferenças inter-individuais dos alunos (nível 1[2]) e as características diversificadas das escolas (nível 2[3]).

- Agregação

O problema da agregação acontece quando num estudo:

- caso 1: os dados são agrupados ao nível das escolas (ignorando a variação inter-individual dos alunos);

- caso 2: apenas são considerados os dados ao nível das diferenças entre sujeitos (como ocorre em estudos de regressão linear simples ou múltipla), ignorando os efeitos da variação encontrada ao nível das próprias escolas

No caso 1, existe uma perda substancial de informação útil, pois a informação acerca dos alunos não é tida em conta na análise (de Leeuw, 2005; Hox, 2002). Tem, no entanto a vantagem de se conseguir estudar de que forma é que as relações num nível de análise varia através doutro nível (Nezlek, 2001).

No caso 2 consideram-se dependências nos dados que, na realidade não existem, pois os resultados dos alunos são os mesmos, considerando ou não a escola desagregada (de Leeuw, 2005; Hox, 2002). Neste caso, não é possível estudar a forma como variam as relações entre as variáveis através dos alunos (Nezlek, 2001). Por outro lado, as inferências podem ser erróneas, por se considerar que os dados desagregados são independentes entre si (Fox e Glas, 2002; Hox, 2002).

Desta forma, torna-se claro que, ainda que corramos o risco de repetição, nunca será de mais salientar a importância do recurso a modelos com estrutura hierárquica ou multi-nível, que considerem, numa única estrutura de análise, a informação contida nos dois níveis da hierarquia – alunos e escolas.

Contudo, é necessário ter em atenção que a forma como os dados são tratados podem induzir-nos em erro. Tradicionalmente, nas investigações quantitativas, por exemplo na área da educação, analisavam-se conjuntamente as variáveis respeitantes aos alunos e as variáveis respeitantes à turma. Neste caso, havia duas alternativas, ambas erróneas. Por um lado, se se recolhiam os dados de cada sujeito de forma independente, daqui

[2] ou nível micro da informação
[3] ou nível macro da informação

resultavam uma série de variáveis, entre as quais as que diziam respeito à turma, ou seja, recolhiam-se os dados dos indivíduos e, em seguida, agrupavam-se de forma a tirar conclusões do grupo a que pertenciam (Valente (2007)). Neste caso, caímos na chamada falácia atomística (Hox, 1995). Por outro lado, se se considerava que a unidade de análise fosse a turma, ou a escola, as variáveis respeitantes aos alunos incluíam-se nos dados da escola. Neste caso, podíamos cair na chamada falácia ecológica, na qual se atribuem incorrectamente as características do contexto aos sujeitos (Hill & Rowe, 1996; Hox, 1998; Goldstein, 2003).

2.1.3 Tipos de modelos segundo a característica da variável de resposta

Tipo de modelo	Variável de resposta
Linear com variável contínua	Onde há uma variável de resposta contínua
Logístico	Para variável de resposta dicotómica
Longitudional	Para uma resposta medida em diferentes ocasiões, no qual o objectivo é estudar a evolução no tempo
Medidas repetidas	Onde a resposta se mede em diferentes ocasiões, mas o fundamental não é o tempo
Multinomial	A resposta tem mais de duas categorias
Poisson	Para contagem de acontecimentos
Multivariada	Quando se estudam várias respostas
Multivariada Longitudional	Para várias respostas medidas em diferentes tempos
Sobrevivência	Onde a resposta é a ocorrência de um acontecimento
Classificação cruzada	Para unidades que pertencem a diferentes grupos
Metanálise	Para a análise da dimensão do efeito de vários estudos
Metodologia	Quando o interesse principal da investigação é dar a conhecer a metodologia multi-nível
Categoria não específica	Utiliza-se quando o resumo e as referências não proporcionam informação sobre o tipo de modelo

Tabela 1: Tipo de modelos segundo a característica da variável de resposta

Fonte: Adaptado de Catalán et al (2003)

2.1.4 Amostragem na análise multi-nível

Como salienta Cruz (2008), podemos afirmar que uma amostra de uma população multi-nível é “faseada”, isto é, tomando como exemplo a área da educação, a obtenção de uma amostra de uma população organizada em níveis faz-se através da selecção aleatória entre unidades do nível macro (por exemplo, escolas). Uma vez seleccionadas estas unidades, o segundo passo seria escolher, também de modo aleatório, as unidades do nível micro (por exemplo, alunos dentro das escolas). Este procedimento descreve sumariamente a técnica adequada para se conseguir uma amostra na qual o pressuposto de independência entre os sujeitos não é violado. Porém, por diversos motivos (práticos, de natureza financeira ou logística), o que é feito frequentemente é proceder a uma amostragem de todos os indivíduos disponíveis, depois de escolhidas as unidades do nível macro. Por vezes, nem mesmo a escolha das unidades do nível macro é feita de forma aleatória, mas sim de acordo com a conveniência ou disponibilidade de recursos. Obviamente, nestes casos, podemos obter resultados falsificados, pois ocorre a violação do pressuposto de independência das observações. (Laros e Marciano, 2008). Nestas amostras as observações individuais não são independentes.

2.1.5 Vantagens e desvantagens dos modelos multi-nível

Existem uma série de razões para utilizar modelos multi-nível. Quando utilizados correctamente, estes modelos permitem-nos obter melhores estimativas dos coeficientes de regressão e da sua variação do que com os métodos tradicionais. A grande flexibilidade dos modelos multi-nível diz respeito à modelação da estrutura da variância dos dados em função das variáveis explicativas, o que nos permite analisar os dados nos quais a variância não é homogénea. Além disso explora, com grande detalhe, o comportamento da variância, a existência de variáveis correlacionadas; características comuns dentro de cada nível; dissemelhança entre níveis; aninhamento entre níveis ou cruzamento entre níveis.

No entanto, existem alguns factores que poderão limitar o uso deste tipo de modelos. Há que ter em conta que a sua teoria é muito mais complexa, tornando difícil a compreensão do investigador que não possua uma forte base de estatística. Outra limitação para o uso deste método são os softwares disponíveis, pois embora alguns modelos multi-nível possam ser ajustados com pacotes estatísticos como o SAS, o Stata, o SPSS ou R, os *softwares* especializados, como o MLwiN (Rasbash et al., 2000) ou o HLM (Raudenbush, Bryk, Cheong e Congdon, 2000) são capazes de ajustar todo o tipo de modelos, mesmo os mais complexos, no entanto, para além de obrigarem o utilizador a familiarizar-se com um novo programa, estes pacotes estatísticos apresentam uma dificuldade de transferência de dados e incapacidade para trabalhar automaticamente com os valores em falta (*missing values*).

2.2 Regressão clássica e regressão multi-nível

2.2.1 Diferenças entre regressão clássica e regressão multi-nível

O modelo hierárquico (Bryk, Raudenbush, 1992) tem em consideração a estrutura de agrupamento dos dados. Concretamente, isto reflecte-se na especificação do modelo. Para o modelo de regressão clássico, a ordenada na origem [4]e o declive[5] são parâmetros fixos, enquanto para o modelo multi-nível estes podem ser considerados parâmetros aleatórios, dependentes da influência do nível hierárquico mais alto.

Para melhor ilustrar as diferenças entre a regressão clássica e a multi-nível utilizaremos um exemplo baseado em Ferrão (2001). Consideremos um conjunto de dados hipotéticos sobre os resultados escolares (classificação escolar) dos alunos provenientes de um certo número de escolas e o respectivo rendimento familiar.

Pretende-se saber se existe relação entre o rendimento familiar e a classificação escolar do aluno.

2.2.2 Modelo de regressão clássica

Através da figura 1 podemos observar uma possível recta de regressão clássica, na qual não se leva em conta a dependência dos alunos nas escolas. Ou seja, todas as escolas estão representadas por uma única recta.

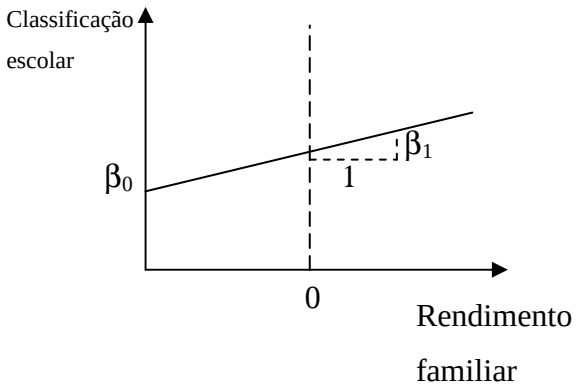


Figura 1: Recta ajustada do modelo de regressão clássico

[4] Também referido como intercepto
[5] Também referido como coeficiente de inclinação

No eixo das abcissas está representado o rendimento familiar que, neste exemplo, é uma variável que está centrada na média[6]. Nesta situação a ordenada na origem é interpretada como o valor médio da classificação escolar de um aluno cujo rendimento familiar é igual à média do rendimento familiar dos alunos de todas as escolas. O eixo dos Y's representa o rendimento escolar. Sendo o declive da recta positivo, verifica-se que, em média, alunos com rendimento familiar mais elevado também tendem a ter melhores resultados escolares.

O modelo de regressão clássico apresentado na equação (1) especifica a relação entre estas duas variáveis e é o modelo que está subjacente ao gráfico apresentado na figura 1.

$$classificação\ escolar_{ik} = \beta_0 + \beta_1 * rendimento\_familiar_{ik} + e_{ik} \qquad (1)$$

Ou mais formalmente, $Y_{ik} = \beta_0 + \beta_1 X_{ik} + e_{ik}$

Neste caso, utilizando a notação dos modelos multi-nível, $\sigma^2_{u_0} = \sigma^2_{u_1} = 0$.[7]

Logo

$$\beta_0 = \gamma_0$$

$$\beta_1 = \gamma_1$$

onde a variável Y_{ik} representa a classificação escolar, β_0 e β_1, são a ordenada na origem e o declive, respectivamente. Estes parâmetros são desconhecidos e devem ser estimados a partir dos dados. O parâmetro β_0 pode ser interpretado como o valor esperado da classificação escolar para os alunos que têm valor nulo de rendimento. O declive, β_1, representa o efeito do rendimento familiar no desempenho escolar do aluno. Por definição de declive, por cada unidade adicional no rendimento familiar, a média do desempenho do aluno observará uma variação de β_1 unidades.

O termo e_{ik} é o erro do modelo, associado aos efeitos individuais do aluno não captados pela componente determinística do modelo, e o pressuposto usual é que tenha uma distribuição normal com média nula e variância, σ^2_e, constante entre os grupos, e que sejam não correlacionados entre si, isto é, $e_{ik} \sim NID(0, \sigma^2_e)$.

[6] Denominam-se variáveis centradas as que incorporam algum tipo de correcção

[7] $\sigma^2_{u_0}$ corresponde à variância do termo u_{0k} e $\sigma^2_{u_1}$ à variância do termo u_{1k}

2.3 Modelo linear hierárquico com dois níveis

Os modelos multi-nível são, na sua essência, ampliações dos modelos de regressão linear clássicos, através dos quais se elaboram vários modelos de regressão para cada nível de análise (Reise e Duan, 2003; Bickel, 2007). Assim, os modelos do nível 1 estão relacionados através de um modelo de nível 2, no qual os coeficientes de regressão do nível 1 se "incorporam" num 2º nível de variáveis explicativas, e assim sucessivamente para os diferentes níveis.

Antes de nos debruçarmos nos métodos mais formais da análise multi-nível, importa ter em conta três conceitos fundamentais e respectivas implicações: correlação intra-classe; coeficiente fixo e aleatório e interacção inter-nível.

- Correlação intra-classe:

É a medida do grau de dependência dos indivíduos. Desta forma, é uma estimação do que têm em comum os alunos, pelo facto de estudarem na mesma escola. Caso o valor desta correlação seja baixo (próximo de zero), então os indivíduos dentro do mesmo grupo são tão diferentes entre si como os que pertencem a outros grupos. Neste caso, não há necessidade de agrupar os dados, pois os grupos não são homogéneos internamente e as observações são independentes[8]. Ao ignorarmos a presença desta correlação intra-classe, os modelos resultantes são falsamente complexos, dado que aparecem relações significativas inexistentes.

Assim, uma das questões de maior interesse é estudar o valor da variância dos resíduos representada por σ_{u0}^2. Se, relativamente à variância total, este valor é pequeno então podemos concluir que a escola tem pouco efeito, ou seja, saber qual é a escola onde o aluno estuda não ajuda a explicar os resultados escolares atingidos pois ele poderia tê-los atingido em qualquer outra escola.

O coeficiente de correlação intra-escola é uma estatística que permite aferir sobre a magnitude do efeito-escola, isto é, mede a correlação entre duas unidades do nível 1 dentro de um mesmo grupo do nível 2, num modelo com dois níveis. Desta forma, expressa a variância total devido ao nível 2.

[8] Podemos assim utilizar os modelos lineares tradicionais

Assumindo que u_{0k} e e_{ik} variam independentemente, o coeficiente de correlação intra-grupo[9] define-se:

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{e0}^2 + \sigma_{u0}^2}$$

Sendo σ_{u0}^2 a variância dos resíduos u_{ok} do nível 2 e σ_e^2 a variância dos resíduos e_{ik} do nível 1.

- Coeficiente fixo e coeficiente aleatório:

Nos modelos de regressão clássicos, os parâmetros que se estimam são o declive[10] e a ordenada na origem. Numa perspectiva clássica, estes coeficientes assumem-se como fixos, isto é, comuns a todos os sujeitos e são estimados a partir dos dados. Por outro lado, os coeficientes aleatórios são variáveis e distribuem-se segundo uma função de distribuição de probabilidade. Numa estrutura multi-nível, os coeficientes do nível 1 (alunos) são tratados como aleatórios no 2º nível (escolas). Isto é, as unidades que definem os níveis são vistos como efeitos aleatórios. Assim, numa amostra aleatória de uma população, as unidades são efeitos aleatórios e traduzem-se num modelo de coeficientes aleatórios que tem em conta a variabilidade entre grupos, desde formas mais simples – através da variabilidade a nível da ordenada na origem - a formas mais complexas – através da variabilidade a nível do declive das rectas. Nestes modelos é possível que os grupos se desviem da solução central ou global, tanto no declive como na ordenada. Ou seja, os modelos multi-nível são compostos por duas partes: uma geral, comum a todos os grupos[11], que é a chamada parte fixa; e outra que representa o específico de cada grupo, que varia e se estima através da variância nos diferentes níveis.

Cruz (2008) refere um exemplo na área da epidemiologia no qual podemos facilmente distinguir os efeitos aleatórios dos efeitos fixos. Estudam-se efeitos específicos de níveis de categorias de factores de risco. Por exemplo, para estudar o estado nutricional das crianças, pretendemos saber o efeito da classe social no índice de peso, para a idade. Para isso, comparamos as médias do índice de peso nas classes A, B, C e D, usando, por exemplo, análise de variância. Podemos também comparar crianças amamentadas "ao

[9] No caso da variável resposta não ser contínua, designa-se por coeficiente de partição de variância, Variance Partition Coefficient (VPC) em inglês
[10] Ou intercepto ou ponto de corte
[11] Ou contextos

peito” até aos 4 meses e crianças amamentadas “ao peito” durante menos tempo, utilizando um teste T. Em ambos os casos, estamos interessados em variáveis específicas de cada grupo. Suponhamos que queremos estudar a variação do índice antropométrico entre centros de saúde. Obviamente que não há grande interesse em fazer a comparação em centros de saúde de localidades adjacentes, a menos que se evidenciem realidades muito bem definidas em cada localidade. É mais interessante saber se existe variabilidade entre centros, independentemente da sua localização. Mais especificamente, podemos tentar identificar a razão da variabilidade (a existir), através de características dos centros. Neste caso, podemos considerar os centros de saúde em estudo como uma amostra da população dos centros de saúde e considerar o efeito dos centros de saúde como um efeito aleatório (em contraste com o efeito fixo da variável classe social), medido por um parâmetro que indica a variabilidade entre grupos, e visto como representativo da população de origem.

Podemos assim concluir que decidir se o efeito de uma variável é fixo ou aleatório depende, em grande parte, do contexto ou dos objectivos do estudo.

- Interacção inter-nível:

Representa a interacção entre variáveis medidas em diferentes níveis de uma estrutura hierárquica de dados. Faz referência à interacção que pode existir entre variáveis de diferentes níveis. Por exemplo, uma determinada metodologia pode ser mais benéfica com alguns alunos. A comprovação deste tipo de hipóteses necessita de um modelo de análise no qual seja evidente a estrutura hierárquica dos dados e que permita estudar as interacções inter-nível.

2.3.1 Especificação formal do modelo multi-nível com dois níveis

Estudos de natureza contextual, hierárquica ou multi-nível implicam, necessariamente, a especificação de duas equações, uma para cada um dos níveis em estudo, alunos (micro nível) e cursos (macro nível), por exemplo. Ao incluirmos, por exemplo, o professor da disciplina, obtemos uma situação de cruzamento de dados, pois o mesmo professor pode leccionar em vários cursos. Mais adiante voltaremos aos modelos de classificação cruzada.

Para ilustrar a forma como o modelo multi-nível trata a estrutura hierárquica dos dados, utilizaremos um exemplo de um modelo a dois níveis: alunos e escolas. Suponhamos que se pretende explicar a classificação escolar dos alunos através de duas variáveis explicativas, uma de cada nível:

- o rendimento familiar, no nível do aluno;

- o tipo de escola (pública ou privada), no nível da escola.

Os alunos são identificados pelo índice i e as escolas pelo índice k. O índice k varia de 1 a K (sendo K o número total de escolas em estudo) e o índice i varia de 1 a n_k (sendo n_k o número de alunos que pertence à escola k). A variável resposta do modelo é a classificação escolar do aluno i pertencente à escola k - $classificação\ escolar_{ik}$ - e a variável explicativa associada a este aluno, é o respectivo rendimento familiar, $rendimento\_familiar_{ik}$.

No modelo de dois níveis (alunos e escolas) tanto a ordenada na origem como o declive podem ser considerados variáveis aleatórias que variam de escola para escola. Em seguida estudaremos, através do exemplo em educação de alunos aninhados em escolas, modelos multi-nível, considerando o modelo de coeficientes aleatórios que tem em conta a variabilidade entre grupos, desde a variabilidade a nível da ordenada na origem passando pela variabilidade a nível do declive das rectas, até à variabilidade de ambos.

2.3.2 Modelo multi-nível de componentes de variância[12]

Quando as rectas têm diferentes médias ($\sigma^2_{u_0} > 0$, isto é, existe variância nas ordenadas) mas o mesmo declive ($\sigma^2_{u_1} = 0$, ou seja, não existe variância nos declives), dizemos que temos um modelo multi-nível de componentes de variância. Para ilustrarmos as rectas ajustadas segundo este modelo multi-nível, temos a figura 2.

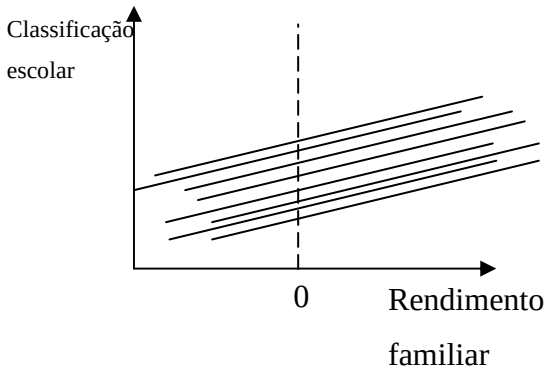


Figura 2: Rectas ajustadas do modelo multi-nível com declive aleatório

Agora, cada uma das rectas está associada a uma escola. Considera-se a ordenada na origem como uma variável aleatória, uma para cada escola, (o que significa que a média da classificação escolar varia de escola para escola) e o declive fixo, isto é, o efeito do rendimento familiar do aluno na sua classificação escolar não varia de escola para escola.

Na equação (2) apresenta-se o modelo em que apenas a ordenada na origem varia aleatoriamente entre as escolas. Este é o modelo subjacente ao gráfico apresentado na figura 2.

$$classificação\ escolar_{ik} = \beta_{0k} + \beta_1 * rendimento\_familiar_{ik} + e_{ik} \qquad (2)$$
$$\beta_{0k} = \gamma_{00} + u_{0k}$$
$$e_{ik} \sim NID\ (0, \sigma^2_e)$$

Ou seja, $Y_{ik} = \beta_{0k} + \beta_1 X_{ik} + e_{ik}$

[12] Random intercept model

A primeira característica a ser observada é que neste modelo o parâmetro da ordenada na origem, β_{0k}, tem o índice *k*, indicando a existência de um parâmetro para cada escola, conforme enunciado previamente. Ou seja, o valor médio da classificação escolar está dividido na contribuição da escola (β_{0k}) e no erro (e_{ik}) de cada estudante à contribuição da escola. Na segunda linha de (2) pode observar-se que a contribuição de cada escola foi decomposta na média global da classificação escolar (envolvendo todas as escolas), γ_{00}, e no afastamento de cada escola, u_{0k}, a essa média global. Este é o efeito individual da escola *k* (componente aleatória do nível 2 associada à ordenada na origem).

O coeficiente constante de regressão β_1 será agora denotado por γ_{10}, para indicar que é um parâmetro em todo o modelo.

Substituindo, obtemos

$$Y_{ik} = \gamma_{00} + \gamma_{10} X_{ik} + e_{ik} + u_{0k}$$

Esta fórmula pode ser interpretada de duas formas:

O parâmetro u_{0k} pode ser fixo ou aleatório.

Os parâmetros desconhecidos do modelo: γ_{10}, γ_{00}, σ^2_{uo} e σ^2_e são estimados a partir dos dados, sendo os primeiros dois parâmetros designados por parâmetros fixos e os dois últimos por parâmetros aleatórios. A componente aleatória associada à ordenada na origem tem variância σ^2_{uo}, representando a variabilidade da ordenada na origem entre escolas. O erro de nível 1, e_{ik}, tem variância σ^2_e e representa a variabilidade intra-escola.

Generalizando, consideremos Y_{ij} uma variável resposta contínua, X_1,…, X_p são as variáveis aleatórias do nível 1 e W_1, …, W_q as variáveis do nível 2, obtemos o modelo (Snijders, 1999):

$$Y_{ik} = \beta_0 + \beta_{10} X_{1ik} + \cdots + \beta_{p0} X_{pik} + \beta_{01} W_{1k} + \cdots + \beta_{0q} W_{qk} + e_{ik} + u_{ok}$$

Os parâmetros fixos β_{h0} e β_{0h} nesta equação, dos níveis 1 e 2 respectivamente, têm a mesma interpretação dos coeficientes no modelo de regressão múltipla. A parte aleatória do modelo é formada pelos termos de erro e_{ik} *e* u_{ok} do nível 1 e 2, respectivamente, mutuamente independentes, com média zero e variância σ^2_{u0} e σ^2_{e0}, respectivamente (Snijders, 1999).

2.3.3 Modelo com declives aleatórios

Na prática podemos ainda encontrar situações em que as rectas de regressão variam apenas no declive. Neste caso, $\sigma_{u_0}^2 = 0$ e $\sigma_{u_1}^2 > 0$.

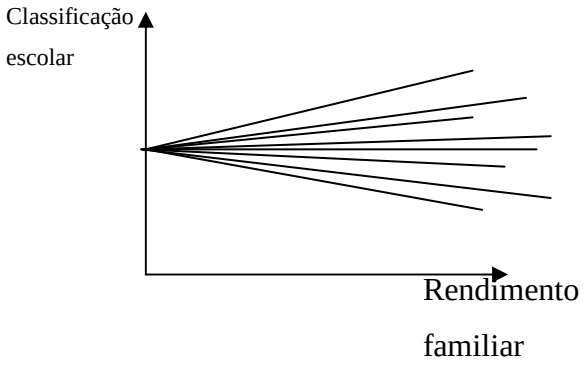


Figura 3: Rectas ajustadas do modelo multi-nível com declives aleatórios

Neste caso, como o valor da ordenada na origem é igual para todas as escolas, significa que o efeito do rendimento familiar na classificação escolar não depende da escola que o aluno frequenta. Ao analisarmos este gráfico, podemos concluir que umas rectas têm declive positivo e outras têm declive negativo. Neste caso hipotético, nas rectas com declive positivo, os alunos com um rendimento familiar mais elevado atingem melhores resultados escolares, enquanto nas rectas com declive negativo o rendimento familiar e os resultados escolares variam de forma inversa. Por outro lado, as escolas cuja recta tem menor declive, são escolas que promovem a igualdade social, nas quais o efeito do rendimento familiar nos resultados escolares do aluno é pequeno. As rectas com grande declive representam escolas onde o rendimento familiar do agregado familiar influencia fortemente os resultados escolares.

2.3.4 Modelo multi-nível de coeficientes aleatórios

A figura 4, representa uma generalização do modelo multi-nível acima apresentado, onde, tanto a ordenada na origem como o declive são aleatórios, isto é, ambos variam entre as escolas. Neste caso, $\sigma_{u_0}^2 > 0$ e $\sigma_{u_1}^2 > 0$.

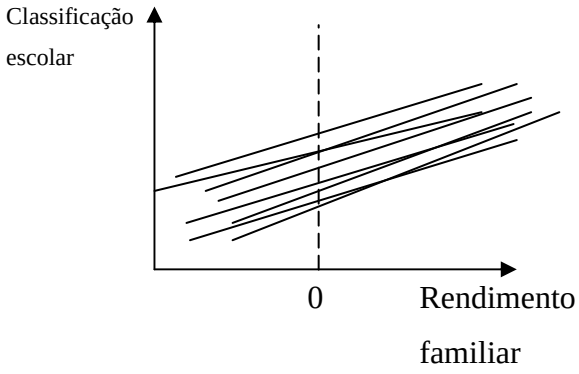


Figura 4: Rectas ajustadas do modelo multi-nível com ordenada na origem e declive aleatórios

Ao analisarmos este gráfico, podemos concluir que todas as rectas têm declive positivo, o que significa que alunos com um rendimento familiar mais elevado atingem melhores resultados escolares, no entanto, sendo a ordenada na origem diferente em todas elas, o efeito do rendimento familiar na classificação escolar depende da escola que o aluno frequenta.

Em suma, a escola que procuramos é aquela que tem uma ordenada na origem elevada (classificação escolar média/alta) e declive tão próximo de zero quanto possível (rendimento familiar não influencia os resultados escolares).

O modelo especificado em (3) (correspondente ao gráfico da figura 4) além da ordenada na origem aleatória também tem o declive aleatório.

$$classificação\ escolar_{ik} = \beta_{0k} + \beta_{1k} * rendimento\_familiar_{ik} + e_{ik} \qquad (3)$$

$$\beta_{0k} = \gamma_{00} + u_{0k}$$
$$\beta_{1k} = \gamma_{10} + u_{1k}$$

$$e_{ik} \sim NID\ (0, \sigma_e^2)$$
$$u_{0k} \sim NID\ (0, \sigma_{u0}^2), ou\ seja, N(0, \tau_{00})$$
$$u_{1k} \sim NID\ (0, \sigma_{u1}^2), ou\ seja, N(0, \tau_{11})$$

Daqui $\beta_{0k} \sim N(\gamma_{00}, \tau_{00})$ e $\beta_{1k} \sim N(\gamma_{11}, \tau_{11})$

Mais genericamente, $Y_{ik} = \beta_{0k} + \beta_{1k}X_{ik} + e_{ik}$

em que

γ_{00} é a grande média, ou seja, a ordenada média para todas as escolas

γ_{10} é a média dos declives de todas as escolas.

u_{0k} é o erro aleatório para cada escola (afastamento em relação à ordenada média)

u_{1k} é o erro aleatório de cada escola (afastamento em relação ao declive médio)

τ_{00} é a variância populacional das ordenadas

τ_{11} é a variância populacional dos declives

Geralmente estima-se também a co-variância entre as ordenadas e os declives τ_{01}.

Similarmente à inclusão do parâmetro da ordenada na origem, a inclusão do parâmetro de declive específico para cada escola indica que a relação entre a classificação escolar e o nível socioeconómico varia de escola para escola. Este modelo, que inclui os termos aleatórios u_{0k} e u_{1k}, correspondentes ao nível 2, descreve a heterogeneidade contextual através das escolas. No entanto, este modelo ainda não inclui a heterogeneidade entre alunos nas escolas. Pelo que ao assumir a homocedasticidade dos resíduos, os modelos com estimadores MQO (Método de Mínimos Quadrados Ordinários) estão a fazê-lo erradamente, já que o efeito *rendimento_familiar* pode variar de aluno para aluno, dentro da mesma escola e a variância não tem que ser obrigatoriamente constante. Neste caso, existe a necessidade de modelar a heterogeneidade quer a nível dos alunos (individual) – neste caso através do parâmetro γ_{10} associado ao *rendimento_familiar* -, que a nível das escolas (contextual).

2.3.5 Modelo com mais de uma variável explicativa

Até ao momento, considerou-se apenas uma variável explicativa, rendimento do agregado familiar do aluno, medido no nível 1. De seguida, apresentar-se-á o modelo (2) acrescentando uma variável explicativa medida ao nível da escola. Esta variável é o tipo de escola, *tipo_escola*$_k$, uma variável binária que designa se a escola é pública ou privada. A sua inclusão no nível 2 do modelo dá-se substituindo a equação (2.a) na segunda linha do modelo (2),

$$\beta_{0k} = \gamma_{00} + \gamma_{01} tipo\_escola_k + u_{ok} \qquad (2.a)$$

$$classificação\ escolar_{ik} = (\gamma_{00} + \gamma_{01} tipo\_escola_k + u_{ok}) + \beta_1 * rendimento\_familiar_{ik} + e_{ik} \qquad (4)$$

$$= (\gamma_{00} + \beta_1 * rendimento\_familiar_{ik} + \gamma_{01} tipo\_escola_k) + (e_{ik} + u_{ok})$$

Ficando então

$$Y_{ik} = (\gamma_{00} + \beta_1 X_{ik} + \gamma_{01} X_k) + (e_{ik} + u_{ok}) \qquad (4a)$$

Finalmente, substituindo a equação (2.a) na primeira linha da equação (2), podemos identificar duas componentes distintas no modelo. A componente determinística ou sistemática do modelo é dada pela expressão

$$(\gamma_{00} + \beta_1 * rendimento\_familiar_{ik} + \gamma_{01} tipo\_escola_k)$$

Relativamente à componente aleatória ou estocástica, é dada por

$$(e_{ik} + u_{ok})$$

Não é demais reforçar a ideia de que a parte aleatória ou estocástica do modelo representa os diversos efeitos aleatórios que influenciam a classificação escolar do aluno, actuando tanto ao nível do aluno como ao nível da escola, e que não são captados pela parte determinística do modelo.

A componente aleatória (e_{ik} *e* u_{0k}) do modelo está decomposta no erro de nível 1, e_{ik}, e no erro de nível 2, u_{0k}. As estimativas destes erros são os resíduos (o que no modelo fica por explicar). Assim, a variância residual do modelo é dada por $(\sigma_e^2 + \sigma_{u0}^2)$.

Com a decomposição da variância residual do modelo, torna-se fácil avaliar o efeito de cada variável (seja esta medida ao nível do aluno ou da escola) na explicação da classificação escolar. Além disso, no modelo nulo (modelo sem variáveis explicativas), é a proporção da estimativa da variância entre escolas, σ_{u0}^2, face à variância total (variância entre-escolas e variância intra-escolas), que mostra a presença do "efeito-escola" no desempenho escolar do aluno.

Finalmente apresenta-se o modelo (4) incluindo a variável explicativa *tipo_escola* na equação do declive. A equação resultante é a seguinte (5):

$$classificação\ escolar_{ik} = (\gamma_{00} + \gamma_{01} tipo\_escola_k + u_{ok}) + (\gamma_{10} + \gamma_{11} tipo\_escola_k + u_{1k}) * rendimento\_familiar_{ik} + e_{ik} \quad (5)$$

$$= \gamma_{00} + \gamma_{10} rendimento\_familiar_{ik} + \gamma_{01} tipo\_escola_k + \gamma_{11} tipo\_escola_k * rendimento\_familiar_{ik} + u_{1k} * rendimento\_familiar_{ik} + e_{ik} + u_{ok}$$

Ou mais genericamente,

$$Y_{ik} = \gamma_{00} + \gamma_{10} X_{ik} + \gamma_{01} W_k + \gamma_{11} W_k * X_{ik} + u_{1k} * X_{ik} + e_{ik} + u_{ok} \quad (5a)$$

Seguindo este raciocínio podemos escrever o modelo para p variáveis do nível 1 e q variáveis do nível 2. Obtemos assim o modelo (notação utilizada por Hox (2002)):

$$Y_{ik} = \gamma_{00} + \gamma_{p0} X_{pik} + \gamma_{0q} W_{qk} + \gamma_{pq} W_{qk} * X_{pik} + u_{pk} * X_{pik} + e_{ik} + u_{ok}^{13} \quad (6)$$

Este modelo tem variáveis que contém efeitos fixos e aleatórios, sendo designado por modelo misto (Monette et al, 2002).

Temos então a parte fixa

$$\gamma_{00} + \gamma_{p0} X_{pik} + \gamma_{0q} W_{qk} + \gamma_{pq} W_{qk} * X_{pik}$$

Que, para além das variáveis explicativas do nível 1 (X_{pik}) e do nível 2 (W_{qk}), contém ainda um termo entre níveis ($W_{qk} * X_{pik}$) que representa a interacção de nível cruzado, tendo como coeficientes γ_{00}, γ_{p0} e γ_{pq}.

[13] Modelo combinado

E a parte aleatória

$$u_{pk} * X_{pik} + e_{ik} + u_{ok}$$

Sendo também interpretado como um modelo linear com um forma complexa do erro (Kreft e De Leeuw, 1998).

Estamos agora em condições de fazermos um estudo mais formal dos modelos hierárquicos.

2.3.6 Forma matricial do modelo multi-nível com dois níveis

A forma matricial do modelo é útil sobretudo para simplificar o processo de estimação (Valente, 2007). Assim, serão utilizados vectores e matrizes (Bergamo, 2002; Monette et al, 2002; Sullivan et al, 1999; Natis, 2000; Bryk e Raudenbush, 1992).

Modelo do nível 1 (ver (10), exemplo da página 44)

$$Y_k = X_k\beta_k + e_k$$

Sendo

$$Y_k = \begin{bmatrix} Y_{1k} \\ Y_{2k} \\ \vdots \\ Y_{n_k k} \end{bmatrix} \quad X_j = \begin{bmatrix} 1 & X_{1k} \\ 1 & X_{2k} \\ \vdots & \vdots \\ 1 & X_{3k} \end{bmatrix} \qquad \beta_k = \begin{bmatrix} \beta_{0k} \\ \beta_{1k} \end{bmatrix} \quad e_k = \begin{bmatrix} e_{1k} \\ e_{2k} \\ \vdots \\ e_{n_k k} \end{bmatrix}$$

Modelo do nível 2 (ver (9), exemplo da página 42)

$$\beta_k = W_k\gamma + u_k$$

Sendo

$$\beta_k = \begin{bmatrix}\beta_{0k}\\ \beta_{1k}\end{bmatrix} \quad W_k = \begin{bmatrix}1 & W_k & 0 & 0\\ 0 & 0 & 1 & W_k\end{bmatrix} \gamma = \begin{bmatrix}\gamma_{00}\\ \gamma_{01}\\ \gamma_{10}\\ \gamma_{11}\end{bmatrix} u_j = \begin{bmatrix}u_{0k}\\ u_{1k}\end{bmatrix}$$

Modelo combinado (ver (11), exemplo da página 47)

$$Y_k = X_k W_k \gamma + X_k u_k + e_k$$

Considerando $A_k = X_k W_k = \begin{bmatrix}1 & W_k & X_{1k} & W_k X_{1k}\\ 1 & W_k & X_{2k} & W_k X_{2k}\\ \vdots & \vdots & \vdots & \vdots\\ 1 & W_k & X_{n_k k} & W_k X_{n_k k}\end{bmatrix}$

Podemos escrever:

$Y_k = A_k \gamma + X_k u_k + e_k$

Em que:

A_k e W_k são as matrizes do delineamento

γ é o vector dos efeitos fixos

$u_k \sim N(0, G), sendo\ G = \begin{bmatrix}\tau_{00} & \tau_{01}\\ \tau_{10} & \tau_{11}\end{bmatrix}$ é um vector dos efeitos aleatórios

$e_k \sim N(0, R), sendo\ R = \sigma^2 I_n$ é o vector dos erros ou resíduos

Temos ainda $cov(u_k, e_k) = 0$

Agrupando os elementos do nível 2 em novas matrizes podemos generalizar este modelo e obter um modelo mais complexo, com mais variáveis explicativas em cada nível e entre níveis. Obtemos então o modelo, que incorpora mais especificações da sua flexibilidade (Gill, 2004).

:

$$Y = A\gamma + Xu + e$$

Sendo

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_k \end{bmatrix};\ \ A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_k \end{bmatrix};\ \gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_k \end{bmatrix};\ X = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & X_k \end{bmatrix};\ u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{bmatrix} \text{ e } e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_k \end{bmatrix}$$

E

$$u_k \sim N\left(0, \begin{bmatrix} G & 0 & \cdots & 0 \\ 0 & G & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & G \end{bmatrix}\right)$$

$$e_k \sim N(0, \sigma^2 I)$$

A matriz de variância-covariância é

$$Var\ (Y) = X \sum X^T + \sigma^2 I = V\sigma^2$$

Sendo

$$\sum = \begin{bmatrix} G & 0 & \cdots & 0 \\ 0 & G & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & G \end{bmatrix}$$

Assume-se que V é uma matriz não singular[14].

[14] Matriz singular é aquela que não tem inversa. Neste caso, o seu determinante é nulo. A matriz não singular é aquela que admite inversa. É identificável pelo determinante não ser nulo.

2.3.7 Centralização das variáveis

Nos modelos multi-nível é fundamental a centralização de variáveis para que β_{0k} (média das observações dos indivíduos) possa ser interpretado a fim de mostrar o efeito do contexto na variável resposta.

Segundo Nezlek (2001), existem "três opções básicas: não haver centralização, centralização na grande média e centralização na média do grupo".

As interpretações de β_{0k} nas várias situações mencionadas por Nezlek (2001) são:

- caso não se centralize a variável de nível 1, X_{ik}, sendo assim considerada na sua medida original, então β_{0k} é o valor esperado da variável resposta Y_{ik} quando X_{ik} for igual a zero.

- se X_{ik} for centrada na sua média geral, então β_{0k} representa a média da k-ésima escola do nível 2 ajustada para a variável X

- por fim, se a variável X_{ik} está centrada na média da respectiva escola do nível 2, então β_{0k} é interpretada como a média não ajustada da variável resposta Y_{ik}. (Bergamo, 2002; Sullivan et al, 1999; Bryk e Raudenbush, 1992; Kreft e De Leeuw, 1998)

"*Nenhuma regra simples cobre todos os casos*" (Bryk e Raudenbush, 1992), portanto a decisão que tomamos acerca da centralização depende da estrutura dos dados e do que se pretende estudar.

2.3.8 Análise multi-nível e Análise de co-variância (ANCOVA)

A relação entre o modelo multi-nível e a análise de co-variância é análoga à diferença entre um teste t de Student para grupos independentes e outro teste t para provas repetidas. A diferença consiste na matriz de variâncias co-variâncias da distribuição das observações. Se a distribuição das observações segundo o modelo linear for formulada da seguinte forma

$$\begin{bmatrix} y_{1j} \\ y_{2j'} \end{bmatrix} \approx N_2 \left(\mu_y = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, V_y = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right)$$

Sendo j e j' dois indivíduos sujeitos a condições diferentes. Este modelo é um modelo típico de efeitos fixos, uma vez que apenas existe uma componente de variância σ^2, assumindo que $\sigma_1^2 = \sigma_2^2$, correspondendo às diferenças individuais de cada grupo em estudo.

Por outro lado, num modelo de provas repetidas, a distribuição será

$$\begin{bmatrix} y_{1j} \\ y_{2j} \end{bmatrix} \approx N_2 \left(\mu_y = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, V_y = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \right)$$

Sendo y_{1j} e y_{2j} resultados do mesmo individuo, considerando o antes e o depois de determinada prova; μ_1 e μ_2 são as médias populacionais em cada caso (efeitos fixos). Assume-se que $\sigma_1^2 = \sigma_2^2$ e designamos abreviadamente por σ^2.

Basicamente, a diferença entre os dois modelos encontra-se na associação entre variáveis. Enquanto no primeiro caso as medições realizadas são independentes, no segundo caso a matriz de variâncias co-variâncias contém um elemento adicional σ_{12}. É precisamente esta característica que torna necessário o uso de diferentes estratégias de análise, considerando-se o modelo de componentes de variância ou de efeitos aleatórios, já descritos anteriormente.

2.4 Construção do MLH (modelo linear hierárquico) para dois níveis

2.4.1 Construção do modelo para dois níveis

Para a análise de dados experimentais aplica-se, normalmente, a Análise de Variância (ANOVA), devido à simplicidade do seu algoritmo para estimar os parâmetros de modelos com efeitos fixos, aleatórios e mistos[15]. Os modelos lineares associados à ANOVA são apropriados quando se ajustam conjuntos de dados equilibrados, no entanto apresentam dificuldades no ajuste de dados não equilibrados. Por outro lado, permitem explicar a variação das observações quando o modelo contém variáveis explicativas discretas, não admitindo, no entanto, a incorporação de variáveis explicativas contínuas (Raudenbush, 1993). Para superar estas limitações, sugeriu-se ajustar um modelo linear de regressão múltipla, que pode aplicar-se a dados equilibrados e não equilibrados e permite incluir variáveis explicativas discretas e contínuas, o que dá origem à análise da co-variância e a modelos de regressão com variáveis indicadoras. No entanto, as vantagens da análise de regressão apenas são imediatas para modelos de efeitos fixos (Kirk, 1982; Montgomery, 2005), ainda que a generalização a situações mais complexas, que já existe desde a década de 50, tenha proporcionado o desenvolvimento do modelo linear geral misto (Sahai, Ageel, 2000).

Como alternativa para solucionar os problemas de análise de dados que continuaram a não resultar satisfatoriamente, foi proposta a formulação de modelos lineares hierárquicos (Raudenbush, 1993; Goldstein, 1995). Estes modelos proporcionam as vantagens do modelo misto da ANOVA e da análise de regressão, que considera variáveis explicativas a nível de grupos de unidades experimentais (Ojeda et al., 1999). Segundo esta linha, é possível reformular um modelo de ANOVA e expressá-lo como um modelo linear hierárquico com dois níveis, que permite considerar efeitos fixos e aleatórios, incorporar variáveis explicativas discretas e contínuas e ajustar-se a dados equilibrados e não equilibrados.

Bryk e Raudenbush (2002), Longford (1993) e Goldstein (1995) apresentam a metodologia geral da modelação linear hierárquica.

[15] Os chamados modelos mistos são modelos em geral lineares, que contém efeitos fixos (constantes), além da média, e aleatórios (sujeitos a uma distribuição de probabilidade), além do resíduo.

Segundo Raudenbush e Bryk (2002), os pressupostos da modelação hierárquica são a normalidade, linearidade e homocedasticidade dos resíduos de ambos os níveis. Os mesmos autores sugerem uma estratégia de complexidade crescente cujas etapas da análise multi-nível são as seguintes:

- Em primeiro lugar realiza-se uma análise de variância com efeitos aleatórios[16], de modo a providenciar informação acerca de quanta variação observada no desempenho existe no seio de cada escola (i.e. ao nível dos alunos – nível 1) e entre escolas (i.e. ao nível 2).

As questões que aqui podem ser colocadas são as seguintes:

(1) Haverá ou não variação suficiente entre alunos no seu desempenho que exige interpretação adequada, desde que sejam identificados as suas variáveis preditoras?

(2) Quanta variação observada no desempenho é devida à circunstância dos alunos pertencerem a escolas diferentes? (a resposta a esta questão, considerada fundamental na modelação hierárquica, é dada pela magnitude do coeficiente de correlação intra-classe);

(3) Existirá, ou não, variação suficiente ao nível do desempenho médio das escolas?

- Em segundo lugar especifica-se um modelo de coeficientes aleatórios[17] para examinar as equações de regressão no seio de cada escola e entre escolas. Pretende-se obter resposta para as seguintes questões:

(1) Quais são os valores médios dos coeficientes de regressão das escolas, incluindo valores na ordenada e declives?

(2) Qual é a magnitude da variação dos coeficientes de regressão entre escolas? Espera-se, nesta situação, identificar aspectos relativos às diferenças no desempenho entre escolas.

- Em terceiro lugar, assumindo que os coeficientes de regressão são diferentes entre escolas e que uma "reduzida" quantidade de variância pode ser explicada ao nível dos alunos, deve ser utilizado um modelo mais complexo para determinar o porquê de determinadas escolas possuírem médias mais elevadas no desempenho.

[16] do inglês *random effects anova*
[17] do inglês *random coefficient model*

2.4.2 Utilização do SPSS para a construção do modelo a dois níveis

Pardo et al (2007) faz referência à forma de ajustar e interpretar modelos multi-nível através do SPSS. A metodologia apresentada vai de encontro ao proposto por Raudenbush e Bryk (2002).

Já sabemos que existem programas específicos para ajustar modelos multi-nível (como o MLwiN ou o HLM). No entanto, este procedimento pode ser realizado com programas de análise de dados mais generalistas (SPSS, SAS, R, S-Plus). Apesar de estes programas não serem específicos para o tratamento deste tipo de dados, incluem procedimentos que podem ser adaptados para ajustar modelos multi-nível. A vantagem do uso destes programas mais generalistas é que o utilizador, em princípio, já os conhece. No entanto, os programas específicos para o tratamento de dados multi-nível permitem ajustar praticamente todos os tipos de modelos multi-nível, por complexos que sejam. Para além disso, o ajuste de modelos multi-nível poderá resultar numa tarefa complexa, não havendo muita documentação relativa ao ajuste deste tipo de modelos com programas generalistas.

Leyland (2004) publicou um trabalho em formato electrónico, através do *Centre for Multilevel Modelling*, no qual dá a sintaxe necessária para ajustar vários modelos multi-nível, não prestando atenção à interpretação dos resultados.

Norusis (2005) dá-nos uma interpretação acessível de alguns modelos multi-nível, no entanto não dá muita atenção aos seus aspectos formais nem à descrição dos seus elementos.

Pardo et al (2007) dá-nos um guia completo, o qual ajuda a eleger o modelo apropriado, a ajustá-lo através do SPSS e a interpretar correctamente os resultados obtidos.

2.4.3 Sub-modelos simplificados do MLH

É possível obter alguns sub-modelos mais simplificados a partir do modelo (6):

$$Y_{ik} = \gamma_{00} + \gamma_{p0}X_{pik} + \gamma_{0q}W_{qk} + \gamma_{pq}W_{qk} * X_{pik} + u_{pk} * X_{pik} + e_{ik} + u_{ok}$$

Para tal basta anular alguns termos desta equação, o que muitas vezes é possível e desejável. Surgem assim novos modelos a partir deste, nos quais é possível aplicar os métodos de análise de dados mais comuns. De seguida apresentam-se alguns dos modelos obtidos desta forma, por ordem crescente de complexidade (Bryk e Raudenbush, 1992 referido em Valente, 2007):

1) Modelo ANOVA *one-way* com efeitos aleatórios
2) Modelo de regressão de médias como respostas
3) Modelo ANCOVA *one-way* com efeitos aleatórios
4) Modelo de Regressão com coeficientes aleatórios
5) Modelo com Ordenadas na origem e Declives como respostas
6) Modelo com Declives a variar não aleatoriamente

A metodologia geral da modelação linear hierárquica apresentada por Bryk e Raudenbush (2002), Longford (1993) e Goldstein (1995) considera os cinco passos, que se descrevem em seguida, apresentando também os comandos do SPSS, para qualquer versão a partir da 11.

Passo 1: Modelo ANOVA *one-way* com efeitos aleatórios[18]

Este modelo representa o ponto de partida do estudo. Serve de base para a estimação da variância explicada, a partir da qual se avaliam as contribuições de modelos mais elaborados. Contém apenas a variável resposta e a constante γ_{00}, não incluindo nenhuma variável explicativa[19]. Desta forma, o modelo contém efeitos aleatórios nos dois níveis e não inclui variáveis explicativas em nenhum deles.

A equação que representa este modelo é:

$$Y_{ik} = \beta_{0k} + e_{ik} = \gamma_{00} + u_{ok} + e_{ik} \qquad (7)$$

[18] Também conhecido por modelo Nulo ou Vazio
[19] Ou predictora

Nesta equação $\beta_{0k} = \gamma_{00} + u_{ok}$ representa a média de Y na k-ésima escola; γ_{00} representa a ordenada na origem da regressão, u_{ok} e e_{ik} são os resíduos (erro) do nível 2 e do nível 1, respectivamente.

Neste modelo são estimados três parâmetros – a ordenada β_{0k}, a variância dos resíduos do nível 1 e a variância dos resíduos do nível 2 e a razão de verosimilhança: -2Ln (verosimilhança), que servirá para avaliar as diferentes contribuições do modelo. O número de parâmetros é utilizado para determinar o número de graus de liberdade para a comparação do ajuste do modelo.

Como já foi referido, este modelo não envolve variáveis explicativas de nenhum dos níveis, servindo apenas como ponto de partida da análise hierárquica, pois proporciona a estimação pontual e o intervalo de confiança para γ_{00} - grande média.

Desta forma, apesar de não explicar nenhuma variância em Y, decompõe-na em dois termos independentes, a variabilidade intra-grupo[20] σ_e^2e a variabilidade entre-grupos[21] σ_{u0}^2, sendo

$$var\ (Y_{ik}) = var(u_{0k} + e_{ik}) = \sigma_{u0}^2 + \sigma_e^2$$

Este modelo permite-nos definir o coeficiente de correlação intra-classe[22].

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{e0}^2 + \sigma_{u0}^2}$$

Sendo σ_{u0}^2 a variância dos resíduos u_{ok} do nível 2 e σ_e^2 a variância dos resíduos e_{ik} do nível 1.

No modelo nulo (modelo sem variáveis explicativas), o coeficiente de correlação representa o valor relativo da variância entre escolas (também é a correlação da classificação escolar entre dois alunos da mesma escola). O coeficiente varia entre 0 e 1.

- O seu valor é zero, quando os grupos do nível 2 são semelhantes e, neste caso, o agrupamento não é relevante, o que significa que os alunos estão homogeneamente distribuídos entre as escolas e que o desempenho do aluno não depende da escola que

[20] Variância dos erros e_{ik}
[21] Variância dos erros u_{0k}
[22] Também interpretado como correlação intra-unidade, considerando o aluno como unidade, é possível estudar a correlação esperada entre duas unidades do mesmo grupo (Goldstein, 1995; Browne e Rasbash, 2001; Hox, 2002).

frequenta. Portanto, não existe uma estrutura multi-nível, pelo que uma análise no nível individual (nível 1) será mais apropriada. Nesta situação, σ_{u0}^2 seria estatisticamente igual a zero, significando que toda a variância da classificação escolar seria devido à variabilidade entre alunos e, por conseguinte, o efeito-escola seria inexistente. No entanto, um valor pequeno de correlação intra-grupo não impede a existência de associações significativas entre variáveis no nível de contexto[23] e nível individual[24] (Merlo, 2005). É importante avaliar o comportamento das variâncias na presença destas variáveis contextuais, na estrutura multi-nível.

- Quando o coeficiente de correlação intra-escola toma o valor 1, o que quer dizer que os grupos do nível 2 são muito importantes na compreensão das diferenças individuais, significa que toda a variabilidade no desempenho dos alunos é devida à diferença entre as escolas. Nesta situação, as características individuais do aluno em nada afectariam o seu desempenho escolar ficando este a dever-se inteiramente às características da escola que ele frequenta.

Comandos do SPSS

Analyze > Mixed Models > Linear

Continue

Escolher a *Dependent Variable* e o *Factor*

Em *Random* seleccionar o *Factor* e *Add*

Continue

Em *Statistics* seleccionar

- *Descriptive Statistics*

- *Parameter Estimates*

- *Tests of Covariance Parameters*

Continue

[23] Nível 2
[24] Nível 1

Passo 2 (Nível 2) Modelo de regressão de médias como respostas

Neste passo acrescentam-se as variáveis explicativas do nível 2 e estimam-se as *Means-as-outcomes regression*. Pretende-se assim predizer o rendimento médio das escolas através de características de grupo. Neste sentido, a única diferença relativamente ao modelo anterior é que se agregam as variáveis no nível 2, através do qual é possível quantificar a proporção da variância explicada de forma individual pelos predictores do nível 2.

Quando se pretendem como resposta as médias de cada grupo, a ser explicada pelas características de grupo, temos um dos mais comuns problemas estatísticos (Bryk e Raudenbush, 1992).

Para simplificar este problema, consideramos no nível 1

$$Y_{ik} = \beta_{0k} + e_{ik}$$

E no nível 2

$$\beta_{0k} = \gamma_{00} + \gamma_{01} W_k + u_{ok}$$

O que dá origem a

$$Y_{ik} = \gamma_{00} + \gamma_{01} W_k + u_{ok} + e_{ik}$$

Mais genericamente, com mais variáveis explicativas, a equação que descreve este modelo é:

$$Y_{ik} = \gamma_{00} + \gamma_{0q} W_{qk} + e_{ik} + u_{ok} \qquad (9)$$

Sendo W_{qk} as q variáveis explicativas do nível 2.

É neste passo que se estuda a contribuição de cada variável explicativa deste nível.

Neste caso, a variância

$$var\,(Y_{ik}) = var(u_{0k} + e_{ik})$$

é condicional devido à variável do nível 2, W.

Também u_{0k} é agora definido por $u_{0k} = \beta_{ok} - \gamma_{00} - \gamma_{01} W_k$

Pelo que σ_{u0}^2 é agora a variância condicional em β_{ok} depois de corrigida por W_k.

Para a proporção de variância explicada no nível 2 temos a equação:

$$R_2^2 = \left(\frac{\sigma_{u0|b}^2 - \sigma_{u0|m}^2}{\sigma_{u0|b}^2} \right)$$

Sendo $\sigma_{u0|b}^2$ a variância residual do nível 2 para o modelo vazio e $\sigma_{u0|m}^2$ a variância residual do nível 2 no modelo de comparação.

Comandos do SPSS

Analyze > Mixed Models > Linear

Em *Subjects* colocar a variável do Factor

Continue

Escolher a *Dependent Variable* e o *Factor*

Em *Covariate* colocar a variável do nível 2

Em *Fixed* seleccionar essa co-variável e *Add*

Continue

Em *Random* seleccionar *Include Intercept*

Passar a variável que corresponde ao factor para a lista de *Combinations*

Continue

Em *Statistics* seleccionar

- *Parameter Estimates*

- *Tests of Covariance Parameters*

Continue

Passo 3 (Nível 1) Modelo ANCOVA *one-way* com efeitos aleatórios[25]

Neste passo acrescentam-se as variáveis explicativas fixas do nível 1 (nível mais baixo). Modela-se o rendimento dos alunos através das variáveis do nível 1, as quais se centram através da grande média (populacional) das variáveis seleccionadas no nível do aluno. Além disso, supõe-se que estas variáveis são fixas ou não variam entre escolas, ou seja, sem componentes aleatórios entre as unidades do nível 2. Desta forma, este modelo constrói-se a partir do modelo nulo, mas incluindo tanto na parte fixa como na aleatória as variáveis consideradas. Neste caso, os componentes de variância correspondentes aos coeficientes são fixados em zero.

Sendo W_{qk} as q variáveis explicativas do nível 2 e X_{pik} as p variáveis explicativas do nível 1, obtemos o modelo

$$Y_{ik} = \gamma_{00} + \gamma_{p0}X_{pik} + \gamma_{0q}W_{qk} + u_{0k} + e_{ik} \qquad (10)$$

Neste modelo, os efeitos do nível 2 são considerados aleatórios.

Com $\begin{bmatrix} u_{0k} \\ \cdots \\ u_{pk} \end{bmatrix} \sim N(0, \Omega u)$

$$\Omega u = \begin{bmatrix} \sigma^2_{u0} & & & \\ \sigma_{u10} & \sigma^2_{u1} & & \\ \vdots & \vdots & \ddots & \\ \sigma_{up0} & \ldots & \ldots & \sigma^2_{up} \end{bmatrix}$$

$[e_{0ik}] \sim N(0, \Omega_e)$

$\Omega_e = \lfloor \sigma^2_{e0} \rfloor$

A variância explicada no nível 1 é dada pela equação:

$$R^2_1 = \left(\frac{\sigma^2_{e|b} - \sigma^2_{e|m}}{\sigma^2_{e|b}} \right)$$

Sendo $\sigma^2_{e|b}$ a variância residual do nível 1 para o modelo vazio e $\sigma^2_{e|m}$ a variância residual do nível 1 no modelo de comparação[26].

[25] Conhecido na literatura anglo-saxónica por One-Way Ancova com efeitos aleatórios

Comandos do SPSS

Analyze > Mixed Models > Linear

Em *Subjects* colocar a variável do Factor

Continue

Escolher a *Dependent Variable* e o *Factor*

Em *Covariate* colocar a variável do nível 2 e a variável do nível 1

Em *Fixed* seleccionar essas co-variáveis e *Add*

Continue

Em *Random* seleccionar *Include Intercept*

Passar a variável que corresponde ao factor para a lista de *Combinations*

Continue

Em *Statistics* seleccionar

- *Parameter Estimates*

- *Tests of Covariance Parameters*

Continue

Os modelos dos passos 2 e 3 são denominados modelos de componentes de variância, uma vez que decompõem a variância da ordenada em diferentes componentes de variância, para cada nível hierárquico. Nestes modelos considera-se a ordenada variável mas os coeficientes de regressão são considerados fixos.

Os três sub-modelos até agora apresentados são modelos de ordenada aleatória, não tendo sido considerada a existência de declives, à excepção do terceiro modelo, no qual foi considerado γ_{p0}, sendo no entanto fixo, o que significa que não varia de grupo para grupo. São os modelos os mais simplificados. Segundo Bryk e Raudenbush (1992) é

[26] Modelo do passo 2

mais usual encontrarmos aplicações dos modelos lineares hierárquicos com declives no nível 1 que variam aleatoriamente através de todas as unidades do nível 2, como acontece nos sub-modelos que seguem.

Passo 4: Modelo de Regressão com coeficientes aleatórios

Neste passo faz-se uma avaliação para verificar se algum dos coeficientes de regressão das variáveis explicativas do nível micro tem um componente significativo de variância (isto é, diferente de zero) entre os valores do nível macro.

Neste tipo de modelo, os coeficientes do nível 1 podem variar aleatoriamente (Bryk e Raudenbush, 1992).

Temos então o modelo do nível 1

$$Y_{ik} = \beta_{0k} + \beta_{1k}X_{ik} + e_{ik}$$

Sendo o modelo do nível 2

$$\beta_{0k} = \gamma_{00} + u_{0k}$$

$$\beta_{1k} = \gamma_{10} + u_{1k}$$

De onde resulta

$$Y_{ik} = \underbrace{\gamma_{00} + \gamma_{10}X_k}_{\text{Parte fixa}} + \underbrace{u_{0k} + u_{1k}X_{ik} + e_{ik}}_{\text{Parte aleatória}}$$

Nesta equação, temos a resposta Y_{ik} como função de uma equação de regressão média $\gamma_{00} + \gamma_{10}X_{ik}$ e do erro aleatório com três componentes: u_{0k} (efeito aleatório da unidade k sobre a média); $u_{1k}X_{ik}$ (o efeito aleatório da unidade k, u_{1k}, sobre o declive β_{1k}) e e_{ik} o erro do nível 1.

Sendo

γ_{00} a ordenada média das unidades do nível 2

γ_{10} o declive médio das unidades do nível 2

u_{0k} o incremento único para a ordenada associada à unidade *k* do nível 2

u_{1k} o incremento para o declive associado à unidade k do nível 2

A matriz de variância-covariância representa a variabilidade dos efeitos aleatórios do nível 2

$$var\left(\begin{bmatrix} u_{ok} \\ u_{1k} \end{bmatrix}\right) = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix}$$

Onde

var(u_{ok}) = τ_{00} é a variância incondicional nas ordenadas do nível 1

var(u_{1k}) = τ_{11} é a variância incondicional nos declives do nível 1

cov(u_{ok}, u_{1k}) = τ_{01} é a covariância incondicional entre as ordenadas e os declives do nível 1.

Com variáveis explicativas dos dois níveis, obtemos

$$Y_{ik} = \gamma_{00} + \gamma_{p0}X_{pik} + \gamma_{0q}W_{qk} + u_{pk} * X_{pik} + e_{ik} + u_{ok} \qquad (11)$$

Sendo u_{pk} os resíduos do nível 2 dos coeficientes das p variáveis explicativas X_{pik} do nível 1.

Este tipo de modelo não atribui importância à categorização do grupo k para determinar o efeito esperado em Y_{ik}. No entanto, defende que existe uma fonte adicional de erro oriundo dessas categorias. Desta forma, ao especificar o modelo apenas com uma fonte de erro, perde-se um efeito de heterocedasticidade (Gill, 2004).

Comandos do SPSS

Analyze > Mixed Models > Linear

Em *Subjects* colocar a variável do Factor

Continue

Escolher a *Dependent Variable* e o *Factor*

Em *Covariate* colocar a variável do nível 1

Em *Fixed* seleccionar essa co-variável para a lista *Model* e *Add*

Continue

Em *Random* seleccionar *Unstrutured** do menu *Covariance Type* e marcar *Include Intercept*

Passar a co-variável para a lista *Model*

Passar a variável que corresponde ao factor para a lista de *Combinations*

Continue

Em *Statistics* seleccionar

- *Parameter Estimates*
- *Tests of Covariance Parameters*

Continue

* ao utilizarmos factores de efeitos aleatórios, impõe-se uma estrutura de co-variância dos dados. Nos modelos estudados até agora utilizou-se a estrutura de co-variância que o SPSS utiliza por defeito: componentes de variância. Apesar de esta ser a estrutura de co-variância utilizada normalmente nos modelos de intersecções aleatórias, no modelo de coeficientes aleatórios (no qual não se assume independência entre β_{0k} e β_{1k}) é necessário definir que tipo de estrutura de co-variância se deve utilizar. Como normalmente não temos informação sobre esta estrutura, devemos utilizar um tipo de

co-variância não estruturada, que significa que não estamos a impor nenhum tipo de estrutura pré-definida e, assim, é o procedimento que a estima, a partir dos dados.

Passo 5: Modelo com Ordenadas na origem e Declives como respostas

Este sub-modelo representa o modelo completo do modelo apresentado anteriormente - Modelo de coeficientes aleatórios. São acrescentadas as interacções entre níveis e as variáveis explicativas do nível 1 que tiveram variância significativa no passo anterior.

Temos então para o nível 1,

$$Y_{ik} = \beta_{0k} + \beta_{1k}X_{ik} + e_{ik}$$

E para o nível 2

$$\beta_{0k} = \gamma_{00} + \gamma_{01}W_k + u_{0k}$$

$$\beta_{1k} = \gamma_{10} + \gamma_{11}W_k + u_{1k}$$

Da conjugação destas equações resulta

$$Y_{ik} = \underbrace{\gamma_{00} + \gamma_{01}W_k + \gamma_{10}X_{ik} + \gamma_{11}W_kX_{ik}}_{\text{Parte fixa}} + \underbrace{u_{0k} + u_{1k}X_{ik} + e_{ik}}_{\text{Parte aleatória}}$$

Neste caso, aparece-nos o termo cruzado $W \times X$, chamado de interacção de nível cruzado. Este termo permite ao declive médio relacionado com X variar com W.

Assim, neste modelo acrescentam-se as interacções entre níveis entre variáveis explicativas do nível 2 e aquelas variáveis explicativas do nível 1 que apresentaram variância significativa de coeficientes no passo 4. Obtemos assim o modelo completo:

$$Y_{ik} = \gamma_{00} + \gamma_{p0}X_{pik} + \gamma_{0q}W_{qk} + \gamma_{pq}W_{qk} * X_{pik} + u_{pk} * X_{pik} + e_{ik} + u_{ok} \quad (12)$$

Comandos do SPSS

Analyze > Mixed Models > Linear

Em *Subjects* colocar a variável do Factor

Continue

Escolher a *Dependent Variable* e o *Factor*

Em *Covariate* colocar as variáveis do nível 1 e as variáveis do nível 2

Em *Fixed* seleccionar essas co-variáveis e as respectivas interacções e passar para lista de *Model* e *Add*

Continue

Em *Random* seleccionar *Unstrutured* do menu *Covariance Type* e marcar *Include Intercept*

Passar a variável que corresponde ao factor para a lista de *Combinations* e a primeira variável de nível 1 que se utilizou no passo 3 para a lista *Model*

Continue

Em *Statistics* seleccionar

- *Parameter Estimates*

- *Tests of Covariance Parameters*

Continue

2.4.4 Modelo com Declives a variar não aleatoriamente

Este modelo não é utilizado para chegar ao modelo combinado, no entanto surge porque, por vezes, é possível justificar quase completamente a variabilidade dos declives de regressão β_{1k}. Neste caso, deve-se encontrar a variância residual de β_{1k} depois do controlo por W_k, ou seja, a variância dos resíduos u_{1k}, deve ser muito próxima de zero. Isto é, após o controlo por W_k, praticamente não resta variância nos declives por explicar (Bryk e Raudenbush, 1992). Desta forma, considera-se $\tau_{11} = 0$, e consequentemente, $\tau_{01} = 0$, dado que G é uma matriz de variância (Valente, 2007).

Neste sub-modelo, a variação em β_{ik}, de grupo para grupo, é completamente consistente com a variação esperada intra-grupo, não havendo necessidade de postular que var(u_{1k}) = $\tau_{11} > 0$ (Monette, 2002:30).

Mantendo a equação do nível 1,

$$Y_{ik} = \beta_{0k} + \beta_{1k}X_{ik} + e_{ik}$$

Temos para o nível 2

$$\beta_{0k} = \gamma_{00} + \gamma_{01}W_k + u_{0k}$$

$$\beta_{1k} = \gamma_{10} + \gamma_{11}W_k$$

Da conjugação destas equações resulta

$$\underbrace{Y_{ik} = \gamma_{00} + \gamma_{01}W_k + \gamma_{10}X_{ik} + \gamma_{11}W_kX_{ik}}_{\text{Parte fixa}} + \underbrace{u_{0k} + e_{ik}}_{\text{Parte aleatória}}$$

Podemos observar que os declives variam de grupo para grupo de forma não aleatória, como podemos verificar na equação de β_{1k}, na qual só existe variação em função de W_k.

Daqui tiramos que, neste modelo, a ordenada varia aleatoriamente através das unidades do nível 2.

2.4.5 Inferência estatística

Antes de apresentarmos os métodos de estimação, convém esclarecer, três termos importantes – modelo, técnica estatística e algoritmo.

- Um modelo é uma representação simplificada da realidade. Relativamente ao modelo estatístico, pode ser considerado um instrumento capaz de transformar uma realidade complexa num sistema mais simples. É composto por um conjunto de equações que descrevem as relações entre quantidades aleatórias. Apesar dos modelos multi-nível possuírem predictores fixos nos seus diferentes níveis hierárquicos, o modelo contém, sempre, um termo estocástico ou aleatório. Os modelos possuem, na sua generalidade, um conjunto de parâmetros desconhecidos e que são utilizados para descrever aspectos que se consideram fundamentais.

- Uma técnica estatística é uma função ou programa, que através dos dados (*input*), produz valores para os parâmetros desconhecidos do modelo. Muitas vezes, a técnica estatística provém da aplicação de um princípio estatístico ao modelo e que pode ser, por exemplo, a máxima verosimilhança ou os mínimos quadrados.

Um algoritmo corresponde à implementação das técnicas estatísticas que se utilizam para realizar os cálculos, optimizando a solução de um dado modelo, principalmente relativamente aos valores dos parâmetros relevantes.

2.4.6 Métodos de estimação

A estrutura dos modelos multi-nível baseia-se num conjunto de pressupostos que, para não obter estimativas enviesadas dos parâmetros do modelo, não devem ser violados.

No entanto, nem sempre é possível cumprir esses pressupostos. Por este motivo, em muitos estudos, é referido o que acontece quando os dados não são normalmente distribuídos ou a dimensão da amostra nos níveis é pequena. Nestes casos, pode referir-se que a precisão das estimativas dos parâmetros depende do que se está a estimar (parâmetros fixos e respectivos erros-padrão, ou parâmetros aleatórios e respectivos erros-padrão), e sobretudo da dimensão das amostras.

De um modo geral, podemos dizer que relativamente:

- às estimativas dos parâmetros fixos, qualquer que seja o método utilizado (mínimos quadrados, mínimos quadrados generalizados, máxima verosimilhança), nunca são enviesadas;

- à qualidade das estimativas e à sua eficiência, o método mais fiável é o método da máxima verosimilhança.

- à estimativa da componente de variância do modelo, não é conhecida a vantagem de qualquer um dos métodos. No entanto, em alguns estudos evidencia-se o método da máxima verosimilhança (Maia et al, 2003).

2.5 Componentes de variância

A existência de co-variância entre as observações origina um viés na estimação dos erros-padrão, no modelo de regressão por mínimos quadrados ordinários, o que produz erros de inferência. Este enviesamento está dependente da grandeza da co-variância intra-grupos, tanto da variável dependente como da predictora, assim como da quantidade de unidades experimentais dos dois níveis. Goldstein (1995) concluiu que, em circunstâncias normais de amostragem, o valor do erro padrão calculado através de uma regressão ordinária pode ser metade do valor real, produzindo assim intervalos de confiança demasiado pequenos e um aumento da probabilidade de Erro Tipo I, nos testes de hipóteses. Para contornar esta situação, devemos recorrer a procedimentos que nos permitam estimar todos os parâmetros do modelo.

Existem soluções analíticas para a estimação de componentes de co-variância, mas somente para os casos onde haja equilíbrio entre o número de observações por condição, tal como acontece nos delineamentos experimentais equilibrados. Num contexto de regressão, isto implicaria que apenas o número de unidades do Nível 1 seja o mesmo para cada unidade do nível 2 e que a distribuição dos valores da variável predictora seja também a mesma para todos os grupos de nível superior (Bryk, Raudenbush, 1992). Na prática, nos estudos de educação, esta situação raramente é cumprida. Nestes casos é aconselhável recorrer a procedimentos de Máxima Verosimilhança (MV) iterativos, uma vez que produzem estimadores com propriedades desejáveis com amostras grandes, tais como consistência e eficiência, ou seja, com uma grande quantidade de dados, o estimador será imparcial e com variância mínima (Bryk, Raudenbush, 1992). Foram Hartley e Rao (1967) que desenvolveram estes métodos, os quais propuseram uma série de equações de MV, muito gerais e aplicáveis a modelos de efeitos mistos e aleatórios, com ou sem co-variáveis e para dados equilibrados ou desequilibrados.

2.5.1 Coeficientes de regressão fixos

Os procedimentos de MV aplicados à estimação de coeficientes de regressão fixos de tendência central geram expressões contidas nas equações do modelo misto (Hocking, 1995; Henderson, 1984; Diggle, Liang e Zeger, 1994). Estas são equivalentes às equações dos mínimos quadrados generalizados, que representam uma extensão natural das equações dos mínimos quadrados ordinários na regressão clássica, quando o modelo contém estruturas de co-variância mais complexas. A estimação dos coeficientes de regressão de efeitos fixos, através do procedimento dos mínimos quadrados generalizados, necessita dos resultados da fase anterior de estimação das componentes de variância, tendo propriedades tais como a de ser melhor estimador linear imparcial.

2.5.2 Coeficientes de regressão aleatórios

Além dos coeficientes de regressão de tendência central fixos, também os desvios aleatórios na média ou declive de cada unidade hierárquica de nível superior podem ter um grande interesse.

Quanto à estimação dos coeficientes dos modelos multi-nível, podemos considerar dois métodos de estimação: máxima verosimilhança (ML) ou máxima verosimilhança restrita (REML), que são muito semelhantes para a estimação dos coeficientes fixos, mas diferem na estimação dos coeficientes aleatórios. A diferença, é que o método de máxima verosimilhança restrita tem em conta os graus de liberdade perdidos na estimação dos coeficientes fixos e, o método de máxima verosimilhança, não. A literatura sugere que o REML é preferível para a estimação de variâncias (Snijders, 1999). O cálculo destes estimadores requer um processo iterativo, realizado com auxílio de um programa informático.

Os algoritmos utilizados pelos modelos hierárquicos são (Goldstein, 2003, Rasbash, 2005):

- Mínimos Quadrados Iterativos Generalizados (IGLS)

-Mínimos Quadrados Iterativos Generalizados Restritos ou Reponderados (RIGLS)

De uma forma feral, o IGLS utiliza o facto de que se as variâncias dos efeitos aleatórios são conhecidas então os coeficientes fixos podem ser estimados através do Método de Mínimos Quadrados Generalizado. Desta forma, este algoritmo alterna entre processar os coeficientes fixos (dadas as variâncias) e as variâncias (dados os coeficientes fixos).

As variâncias dos dois níveis são estimadas separadamente. Uma parte da variância residual entre as unidades do nível 2 pode ser explicada pelas variáveis predictoras no nível contextual. Da mesma forma, parte da variância explicada entre as unidades do nível 1 pode ser explicada pelas variáveis predictoras do nível individual (Jarvelin, 1997). Portanto, as variâncias dos dois níveis podem ser modeladas em função das co-variáveis predictoras, resultando daqui um modelo com variância complexa. A modelação é semelhante à da parte fixa do modelo. Poderá portanto haver uma alteração da variância em cada nível, em função dos factores em estudo, através de associações lineares ou quadráticas (Barros, 2002).

No nível 1, no qual a variância se encontra no nível básico, apenas tem interesse a função que especifica a variância como um todo. Não existe interpretação para os parâmetros individuais (variância e co-variância). Já nos níveis superiores de hierarquia, σ_{u0}^2 é a variância populacional das ordenadas das várias rectas de regressão, cada uma associada a um grupo do nível 2. O parâmetro σ_{u1}^2 é a variância populacional dos declives e τ_{01} é a co-variância entre as ordenadas e os declives (Barros, 2002).

Também os erros padrão podem ser obtidos para os coeficientes fixos e aleatórios, no entanto, para os efeitos aleatórios são menos fiáveis (Snijders, 1999).

As significâncias estatísticas dos coeficientes da parte fixa e da parte aleatória do modelo multi-nível são avaliadas pelo teste de Wald[27], que passamos a descrever.

[27] Teste à significância individual de cada variável

2.5.3 Teste de Wald e Teste da Razão de Verosimilhanças

Após a estimação dos parâmetros deve-se proceder à investigação da significância estatística dos mesmos. O teste de Wald é utilizado para avaliar se o parâmetro é estatisticamente significativo. A estatística do teste utilizada é obtida através da razão do coeficiente pelo seu respectivo erro padrão. Esta estatística teste tem distribuição Normal, sendo seu valor comparado com valores tabelados de acordo com o nível de significância definido. A estatística teste, para avaliar se o parâmetro β é igual a zero, é assim especificada:

$$W = \frac{\hat{\beta}}{\sqrt{var(\hat{\beta})}}$$

No entanto, por vezes, o teste de Wald é impreciso, e rejeita coeficientes que são estatisticamente significativos (Hauck e Donner, 1977). Sendo assim, é aconselhável que os coeficientes, identificados pelo teste de Wald como sendo estatisticamente não significativos, sejam testados novamente pelo teste da razão de verosimilhanças[28].

A razão de verosimilhança permite testar uma H_0 contra uma H_1 que contém um maior número de parâmetros. O teste da razão de verosimilhança é obtido através da comparação entre os modelos com e sem as varáveis testadas. A estatística teste (D), definida abaixo, compara o valor obtido com o de uma distribuição qui-quadrado com número de graus de liberdade igual à diferença no número de parâmetros (Goldstein, 1995; Raudenbush e Bryk, 1992).

$$D = -2\log\left(\frac{\lambda_0}{\lambda_1}\right)$$

Onde λ_0 e λ_1 representam os valores de verosimilhança dos modelos correspondentes à hipótese nula e à hipótese alternativa.

Por exemplo, considerando a área da educação, um investigador pode analisar dados sobre o nível socioeconómico dos alunos e o seu rendimento escolar, com o intuito de constatar se existe uma relação funcional entre as duas variáveis. O objectivo da análise multi-nível é testar se existe variabilidade nas rectas de regressão entre as diferentes

[28] Likelihood ratio test

escolas. Assim, para testar a hipótese nula de que existe variabilidade no rendimento académico entre escolas, temos as hipóteses

$$H_0: \sigma_0^2 = 0$$

versus

$$H_1: \sigma_0^2 > 0$$

O nível de significância estatística indicará, através da distribuição qui-quadrado, a evidência para concluir se existe ou não variabilidade no rendimento académico entre escolas. A não rejeição da hipótese nula indicará que não há necessidade de um modelo multi-nível, sendo suficiente uma análise de regressão por mínimos quadrados ordinários.

O procedimento da razão de verosimilhanças pode ser também utilizado de forma análoga para testes de hipóteses aos efeitos fixos do modelo.

Outra vantagem destes métodos de inferência é a sua flexibilidade para realizar testes sobre a combinação linear das variâncias do modelo. Voltando à área da educação, é possível determinar se as variâncias dos alunos (nível 1) se alteram com o estatuto socioeconómico, se há diferença na variabilidade entre o sexo, ou se estas diferenças dependem da idade ou do facto da escola ser pública ou privada. Nestes casos, as componentes da co-variância têm uma interpretação análoga aos efeitos de interacção com efeitos fixos. A descrição destas diferenças nas componentes da variância é muito útil para melhorar a especificação do modelo a ajustar (Goldstein et al., 1998; Goldstein, 1995).

2.5.4 Análise de resíduos

Os resíduos representam os afastamentos das estimativas médias em relação à média geral predita (Merlo, 2005, Rasbash, 2005).

Na análise de regressão linear, assumimos que os erros satisfazem os seguintes pressupostos:

- seguem uma distribuição normal

Esta condição pode ser verificada usando um gráfico de probabilidade normal (Normal Probability Plot). Se os erros possuírem distribuição Normal, todos os pontos dos gráficos devem posicionar-se mais ou menos sobre uma recta.

- têm média zero;

- homocedasticidade (variância constante);

- são independentes.

Estes três últimos pressupostos podem ser verificados graficamente, representando os resíduos em função dos valores estimados da variável dependente (gráfico residual) ou em função dos valores de uma das variáveis independentes. Os pontos do gráfico devem distribuir-se de forma aleatória em torno da recta que corresponde ao resíduo zero, formando uma mancha de largura uniforme. Desta forma será de esperar que os erros sejam independentes, de média nula e de variância constante.

A verificação das hipóteses é fundamental, visto que toda a inferência estatística no modelo de regressão linear (testes de hipóteses) se baseia nesses pressupostos. Nesse sentido, se houver violação dos mesmos, a utilização do modelo deve ser posta em causa. A análise dos resíduos é uma ferramenta popular para detectar violações de tais pressupostos.

2.5.5 Deviance

A medida *deviance*[29] mede o grau de desajuste do modelo e permite comparar modelos: quanto maior for o valor do *deviance*, maior é o desajuste do modelo.

Assim, o melhor modelo é identificado pela menor *deviance* pela equação:

$$Deviance = -2\log(L_0) - [-2\log(L_1)] = -2LL$$

Sendo L_0 a verosimilhança do modelo nulo (sem co-variáveis) e L_1 a verosimilhança do modelo completo.

O SPSS fornece-nos modificações de -2LL que incrementam o seu valor através de alguma função do número de parâmetros.

Temos então

Critério de informação de Akaike (Akaike, 1973)

$$AIC = -2LL + 2d$$

é uma estatística frequentemente utilizada para a escolha da especificação óptima de uma equação de regressão.

Este critério pode utilizar-se quando a estrutura de co-variâncias de um dos modelos não constitui uma forma reduzida do outro (Natis, 2000).

Critério de informação de Akaike corrigido (Hurvich e Tsai, 1989)

$$AICC = -2LL + \frac{2dn}{n-d-1}$$

Critério de informação de Akaike consistente (Bozdogan, 1987)

$$CAIC = -2LL + d[\log(n) + 1]$$

Critério de informação bayesiano (Schwarz, 1978)

$$BIC = -2LL + dlog(n)$$

[29] O *deviance* não pode ser interpretado isoladamente, mas sim comparado com os modelos posteriores. Espera-se que o ajuste melhore com a introdução de variáveis explicativas.

Utilizando o método MV[30]

LL representa o logaritmo da verosimilhança

d é o número de parâmetros associados aos efeitos fixos mais o número de parâmetros associdos aos efeitos aleatórios

n é o número total de casos

Se se utiliza o método MVR[31]

LL representa o logaritmo da verosimilhança restrita

d é o número de parâmetros associados aos efeitos aleatórios

n é o número total de casos menos o número de parâmetros associados aos efeitos fixos

Os autores acordaram que, quando se lida com modelos hierarquicamente contidos uns nos outros (do inglês *nested models*), devemos comparar o valor da *Deviance* entre modelos. As diferenças entre *Deviances* possuem uma distribuição de qui-quadrado cujos graus de liberdade são iguais às diferenças entre o número de parâmetros testados em cada modelo.

Desta forma, também é possível comparar dois modelos através do teste do qui-quadrado. Por exemplo para os modelos M1 e M5, podemos calcular:

$$\frac{Deviance\ M1 - Deviance\ M5}{Parâmteros\ M5 - Parâmetros\ M1}$$

Caso este resultado seja notoriamente superior ao valor crítico 1,96 (para o nível de significância de 5%), concluímos que o modelo M5 se ajusta melhor aos dados do que o modelo M1.

[30] Máxima verosimilhança (ML na literatura anglosaxónica), formalizada e estimulada a partir do artigo de Hartley e Rao (1967). Mais tarde, a modificação proposta por Patterson e Thompson (1971), hoje conhecida máxima verossimilhança restrita ou residual, tornou-se uma opção ainda mais atractiva.

[31] Máxima verosimilhança restrita (REML na literatura anglo-saxónica)

Capítulo III – Extensões do MLH com 2 níveis

3 Extensões do Modelo LH com 2 níveis

3.1 Modelo linear hierárquico com 3 níveis

Facilmente se verifica que os sistemas escolares são um exemplo típico de uma estrutura hierárquica. Por exemplo, alunos, turmas e escolas constituem uma sequência de agrupamentos naturalmente aninhados. Assim, podemos considerar um modelo com três níveis, estando as unidades do primeiro nível (alunos) agrupadas conforme as unidades do nível dois (turmas) e as unidades do segundo nível agrupadas nas unidades do terceiro nível (escolas). Segundo a notação de Bryk, Raudenbush (1992), cada aluno é representado pelo índice i, o índice j representa cada turma e o índice k representa cada escola. Supomos ainda que *x* representa, genericamente, uma variável do aluno, *w* uma variável da turma e *z* uma variável da escola. O modelo terá a seguinte expressão geral:

$y_{ijk} = \beta_{0jk} + \sum_{f=1}^{F} \beta_{fjk} x_{fijk} + e_{ijk}$

$\beta_{fjk} = \gamma_{f0k} + \sum_{s=1}^{S} \gamma_{fsk} w_{sjk} + u_{fjk},$ f = 0, …, F

$\gamma_{fsk} = \pi_{fs0} + \sum_{t=1}^{T} \pi_{fst} z_{tk} + r_{fsk},$ f = 0, …, F e s = 0, …, S

Nestas expressões F representa o número de variáveis do primeiro nível, S o número de variáveis do segundo nível e T o número de variáveis do terceiro nível.

y_{ijk} representa a classificação escolar do i-ésimo aluno da j-ésima turma da k-ésima escola.

β_{fjk} são os coeficientes do nível 1

x_{fijk} são as variáveis predictoras do nível 1

e_{ijk} é o efeito aleatório do nível 1. Temos $e_{ijk} \sim N(0, \sigma^2)$

γ_{fsk} são os coeficientes do nível 2

w_{sjk} são as varáveis predictoras do nível 2

u_{fjk} é o efeito aleatório do nível 2. Considerando este efeito como um vector, assume-se que este segue uma distribuição Normal Multivariada com média 0 e matriz de covariância T_γ de dimensão $\sum_{f=0}^{F}(S_f + 1) \times \sum_{f=0}^{F}(S_f + 1)$

π_{fst} são os coeficientes do nível 3

z_{tk} são as variáveis predictoras do nível 3

r_{fsk} é o efeito aleatório do nível 3. Considerando este efeito como um vector, assume-se que este segue uma distribuição Normal Multivariada com média 0 e matriz de covariância T_β de dimensão $(T + 1) \times (T + 1)$.

As hipóteses de inexistência de correlação dos erros dos diferentes níveis, também adoptadas nos modelos com dois níveis, são mantidas, isto é, E($eu_{f.}$) = 0, sendo $eu_{f.}$ o erro do nível 2, E($er_{fs.}$) = 0 , onde $er_{fs.}$ representa erro do nível 3 e E($r_{fs.}u_{f.}$) = 0, onde $r_{fs.}$ representa o efeito aleatório do nível 3 e $u_{f.}$ o efeito aleatório do nível 2. Os erros do modelo são todos normais.σ_e^2, $\sigma_{u_f}^2$ e $\sigma_{r_{fs}}^2$ são as variâncias dos erros do modelo nos diferentes níveis, e também são denominadas por componentes aleatórias do modelo. Os parâmetros π_{fst} são parâmetros fixos do modelo.

O modelo nulo[32] de três níveis é, assim, representado por:

$$y_{ijk} = \beta_{0jk} + e_{ijk}$$

$$\beta_{0jk} = \gamma_{00k} + u_{0jk}$$

$$\gamma_{00k} = \pi_{000} + r_{00k}$$

Com π_{000} representando a grande média e r_{00k} representando o erro aleatório do nível 3 e e_{ijk} o erro aleatório do nível 1.

Neste caso, tendo em conta a inexistência de correlação dos erros dos diferentes níveis do modelo, a proporção da variância explicada devida a cada nível é dada por:

$\frac{\sigma_e^2}{\sigma_e^2 + \sigma_{u_0}^2 + \sigma_{r_{00}}^2}$, para o nível 1

[32] Sem variáveis explicativas

$\frac{\sigma_{u_0}^2}{\sigma_e^2+\sigma_{u_0}^2+\sigma_{r_{00}}^2}$, para o nível 2

$\frac{\sigma_{r_{00}}^2}{\sigma_e^2+\sigma_{u_0}^2+\sigma_{r_{00}}^2}$, para o nível 3

Recordamos que σ_e^2 representa a variância do nível 1, $\sigma_{u_0}^2$ a variância do nível 2e $\sigma_{r_{00}}^2$ a variância do nível 3.

3.1.1 Hipóteses e método de estimação do modelo de três níveis

A estimação dos coeficientes fixos e a estimação das componentes de variância pode ser realizada através do método de máxima verosimilhança (Bryk, Raudenbush, 1992).

A medida de ajuste do modelo, a chamada estatística de *deviance*, é definida por:

$D = -2Ln\ (L)$

Sendo L o valor da função de verosimilhança avaliada no seu valor. Esta medida será utilizada para avaliar o grau de explicação alcançado pelos modelos construídos a partir o modelo nulo.

3.1.2 Comandos do SPSS para a construção de um modelo com três níveis

Analyze > Mixed Models > Linear

Em *Subjects* colocar os Factores dos níveis 2 e 3

Continue

Escolher a *Dependent Variable* e os *Factores* definidos anteriormente

Em *Covariate* colocar as co-variáveis

Em *Fixed* seleccionar essas co-variáveis e as respectivas interacções e passar para lista de *Model* e *Add*

Continue

Em *Random* marcar *Include Intercept.* Em *Subject Groupings* escolher o factor do nível 3. Em *Random effect* clicar em next para adicionar o 2º efeito aleatório. Passar as variáveis que correspondem aos factores para a lista de *Combinations*

Continue

Em *Statistics* seleccionar

- *Parameter Estimates*

- *Tests of Covariance Parameters*

Continue

3.2 Modelos multi-nível de classificação cruzada

Uma outra característica relevante dos modelos multi-nível é a possibilidade de modelar efeitos em que há classificação cruzada no mesmo nível (Raudenbush e Bryk, 2002; Snijders e Bosker, 1999), bastante importante, por exemplo, em aplicações de problemas relacionados com indústrias e países.

Meyers (2004) analisa o impacto do tratamento inadequado de classificações cruzadas em modelos multi-nível e conclui que, entre outras consequências, as estimativas de componentes de variância são enviesadas.

3.2.1 A origem da classificação cruzada e suas consequências

Particularmente em educação, os exemplos descritos até aqui fazem referência a estruturas onde os alunos estão aninhados em escolas que, por sua vez, estão aninhadas em agrupamentos.

No entanto, muitas vezes as estruturas de dados são mais complexas. Isto conduziu a uma necessidade de aprofundar a metodologia para lidar com a análise dos efeitos nessas estruturas.

Goldstein (2003), dá um exemplo esclarecedor. Por exemplo, os alunos podem ser aninhados em escolas, mas também no bairro onde moram. Uma possível representação esquemática desta situação pode ser observada na figura 5 para quatro escolas e três bairros. Temos um total de trinta e três alunos, organizados entre um e seis alunos dentro de cada célula da escola. A cruz representa a classificação no nível 2 de estudantes no nível 1. Neste sentido, temos um modelo de nível dois, sendo o segundo nível uma combinação de escolas e bairros.

	Escola 1	Escola 2	Escola 3	Escola 4
Bairro 1	XXXX	XX	X	X
Bairro 2	X	XXXXX	XXX	XX
Bairro 3	XX	XX	XXXX	XXXXXX

Figura 5: Um modelo de classificação cruzada (nível 2) de alunos por bairro e escola (adaptado de Fielding and Goldstein, 2006)

A figura 6 ajuda a esclarecer um pouco melhor este tipo de estruturas. Estes diagramas poderão tornar-se demasiado elaborados para estruturas mais complexas, podendo tornar-se mais confusos do que esclarecedores. Mais adiante veremos representações pictóricas alternativas, mais simples no seu conteúdo, mas que necessitam ser articuladas com algum detalhe. Na figura 6 estão apenas a ser considerados doze dos trinta e três alunos. Neste caso, os alunos 1 e 2 frequentam a mesma escola 1, mas vêm de diferentes bairros. Por outro lado, os alunos 6 e 10 são provenientes do mesmo bairro mas frequentam diferentes escolas.

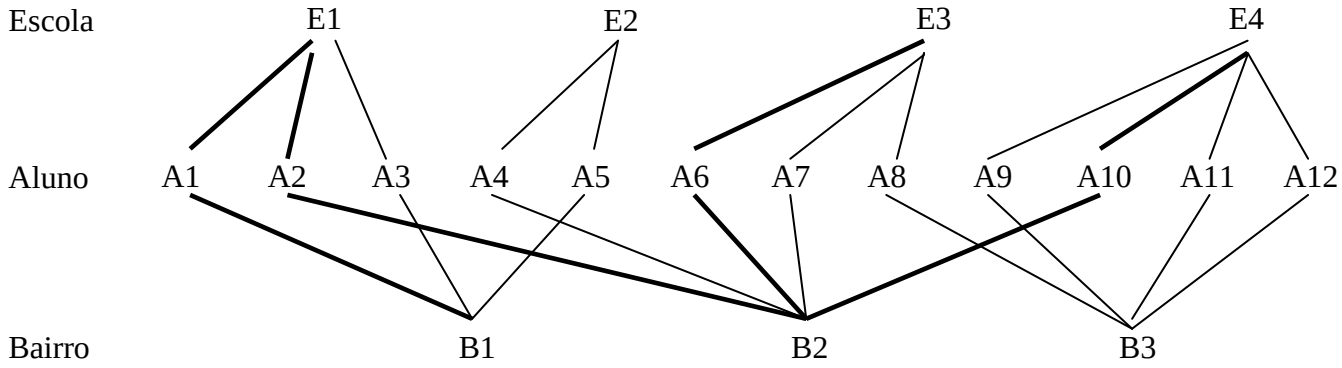


Figura 6: Doze estudantes no nível 1 aninhados por bairro e escola, com cruzamento no nível 2 (adaptado de Fielding and Goldstein, 2006)

A metodologia destas estruturas leva-nos ao reconhecimento de que os efeitos do nível 2 são agora mais complexos, podendo surgir a partir de duas hierárquicas, de forma transversal. Além disso, pretendemos diferenciar os efeitos das escolas e bairros. Isto é particularmente importante se houver um grau de associação entre os bairros e a escola frequentada pelo aluno. Caso os efeitos do bairro sejam importantes e não estejam especificados no modelo, o efeito-escola poderá criar alguns efeitos que se atribuem aos bairros correspondentes.

Assim, a utilização de estruturas com cruzamento e a inclusão dos efeitos dos diferentes factores no modelo estatístico é uma das estratégias possíveis para a realização da investigação. Ao não reconhecermos que existem alguns efeitos derivados do bairro, para além dos efeitos da escola, sobre os resultados do estudante, estamos perante aquilo que se chama de "*underspecified model*".

Por exemplo, se concluirmos que determinada escola obteve resultados fracos, pode ser, em parte, devido ao facto de grande parte dos alunos desta escola serem provenientes de bairros com características que eventualmente poderiam prejudicar os estudantes. Por outro lado, se obtemos um modelo que apenas analisou as diferenças provenientes do bairro, não incluindo o efeito escola, alguns bairros podem ter melhores resultados, do que seria de esperar. Isto pode acontecer, em parte, devido à escola onde grande parte desses alunos está inserida.

3.2.2 Alguns objectivos da análise de modelos multi-nível de classificação cruzada

No estudo das estruturas de dados cruzadas pode haver um conjunto de objectivos na construção do modelo estatístico que contém vários níveis de efeitos aleatórios para ambos os factores do nível 2 da estrutura. O exemplo anterior foi conduzido no contexto de escola e bairro, porém, muitas vezes aparecem-nos estruturas ainda mais complexas.

Os pontos que pretendemos optimizar e dar resposta são:

i) Melhorar as estimativas dos efeitos das variáveis explicativas

Em primeiro lugar poderíamos tentar melhorar as estimativas dos efeitos das variáveis explicativas do modelo garantindo que a estrutura do erro aleatório é adequada e inclui características essenciais dos efeitos aleatórios.[33] No entanto, muitas vezes a precisão das estimativas dos coeficientes dos efeitos fixos do modelo é subestimada, através dos erros padrão, embora haja consistência estatística. Isto leva a que os testes estatísticos, muitas vezes levem à conclusão de que os efeitos das variáveis explicativas sejam estatisticamente significativos, quando num modelo bem especificado, não o sejam. No exemplo descrito anteriormente, esta dificuldade pode ser superada ao reconhecermos os efeitos da escola num modelo com dois níveis. No entanto, reconhecer a importância do bairro de proveniência do aluno poderá melhorar a qualidade das inferências.

[33] Greene (2003) desenvolve de forma aprofundada este tema.

ii) Identificar as componentes da variância nos resultados

Num modelo cruzado, antes de introduzirmos as variáveis explicativas, podemos procurar analisar de que forma a variação dos resultados pode ser atribuída à escola, ao bairro e ao aluno individualmente. Desta forma, podemos depois estender o modelo para identificar de que forma a escola, o bairro e as características do aluno poderão explicar os componentes da variância. Em seguida, podemos introduzir as variáveis, de forma a estudar os componentes da variância residual. Estes dão-nos uma ideia do grau de variação nos resultados que pode ser atribuída às influências de cada um dos três tipos de unidade do modelo.[34]

iii) Estudar o efeito diferencial

Podemos querer investigar se existe relação entre as características dos alunos, da escola e do bairro. Por exemplo, saber se o efeito "capacidade inicial" do aluno à entrada da escola secundária é mais importante numas escolas do que noutras. Ou se os efeitos das características da escola variam entre os bairros, ou se as características dos bairros influenciam de forma diferente em escolas diferentes. O efeito dos estudantes irem para uma escola mais selectiva pode, por exemplo, ser mais acentuado nuns bairros do que noutros. A este tipo de influências chamamos efeitos diferenciais. [35]

iv) Estimar os efeitos em modelos de 2 níveis

A maioria dos modelos multi-nível e de outros modelos de efeitos aleatórios permitem estimar os efeitos aleatórios associados a escolas em particular ou bairros depois de ajustar as características explicativas. Esta é a abordagem adoptada em muita literatura sobre a eficácia e vantagens dos modelos hierárquicos (por exemplo, O'Donoghue et al (1997)). Pode também ser

[34] Caso o efeito bairro seja importante e não o tenhamos incluído no modelo de acordo com o especificado no nível 2, apenas com a escola como efeito aleatório, os componentes da variância podem ser atribuídos ao efeito escola quando, devido à imprecisão, parte podem ser devido a características do bairro. Rasbash e Browne (2001) dão um exemplo, na área da saúde, onde se pretende avaliar a importância das práticas gerais e do hospital no prognóstico do paciente. Devido à natureza dos dados, nos quais existe cruzamento, a construção de modelos distintos de pacientes dentro de hospitais e de pacientes nas práticas gerais, é insuficiente e pode ser enganosa. Deste modo, o ideal é fazer uma avaliação da variação da rede hospitalar tendo em conta as práticas gerais e vice-versa. Para isso terá de se recorrer a um modelo de classificação cruzada para estimar os componentes da variância.

[35] differencial effects, na literatura anglo-saxónica

usada como dispositivo de triagem para identificar escolas com resultados ajustados, para uma investigação mais profunda (Goldstein e Spiegelhalter (1997)).

3.2.3 Alguns exemplos em educação de estruturas com cruzamento e sua análise

Já vimos um exemplo de aplicação de modelo multi-nível de classificação cruzada. Exemplos mais complicados ocorrem em estudos de medidas repetidas. Esses estudos, mostram que a leitura das medidas em ocasiões diferentes ao longo do tempo pode ser considerada como aninhar alunos em particular. No entanto, num quadro alargado a vários níveis, podemos também querer examinar o efeito dos professores, da turma ou da escola. Isto pode tornar-se complicado, pois todos estes factores podem sofrer alterações ao longo do tempo e por diversas vezes. No entanto, os modelos de medidas repetidas podem ter valor acrescentado para ilustrar a riqueza das estruturas que podem ser formuladas com a ideia de efeitos cruzados. Neste caso, não é necessário dispor de dados equilibrados em todos os conjuntos de medições. Esta é uma vantagem de uma abordagem multi-nível para estudos de medidas repetidas, amplamente discutida por Goldstein (2003).

Fielding e Goldstein (2006) dá um exemplo deste tipo de aplicações. Supondo que se pretende fazer uma medição, em 7 ocasiões, por dois professores. No entanto, por diversas razões, as medidas nem sempre são completas e, por vezes, temos apenas a medida de um professor. Para um determinado aluno, a situação pode ser ilustrada na figura 7, onde X indica as medidas que forem tomadas.

	Ocasião						
Professor	1	2	3	4	5	6	7
A	X	X			X	X	X
B	X		X	X	X	X	

Figura 7: Classificação cruzada para professores e ocasião, para um aluno (adaptado de Fielding and Goldstein, 2006)

Esta situação pode ser considerada de classificação cruzada, considerando um par de professores e as ocasiões. Nas células são evidenciadas várias unidades de nível 2, com medições no nível 1, aninhadas na combinação ocasião / professor. Neste caso, temos uma situação especial de classificação cruzada de nível 2 com, no máximo, uma unidade por célula. Suponhamos que as medições em cada estudante são feitas isoladamente dos outros estudantes, isto é, por um momento, os professores A e B são exclusivos de um estudante. Desta forma, as unidades celulares são medições hierarquicamente organizadas, de nível 3. O diagrama da figura 8 representa, de forma mais simplificada e completa, esta situação, para três alunos em particular. No diagrama estão evidenciadas (a sombreado) os blocos de cruzamento de professor e ocasião. Esta situação pode ter a vantagem de facilitar a estimação do respectivo modelo (Rasbash e tal, 2004). Podemos observar uma falta de equilíbrio, uma vez que nem todos os alunos têm observação em todas as ocasiões. Por exemplo, o aluno 3 tem observações em apenas três das sete ocasiões.

	Aluno 1							Aluno 2						Aluno 3		
Ocasião	1	2	3	4	5	6	7	1	2	3	4	5	6	1	4	7
Professor A	X	X			X	X	X									
Professor B	X		X	X	X	X										
Professor C								X	X	X	X	X				
Professor D								X	X	X	X	X	X			
Professor E														X	X	X
Professor F																X

Figura 8: Medições na classificação de um cruzamento de nível 2, apenas com uma unidade por célula, com alunos no nível 3 (adaptado de Fielding and Goldstein, 2006)

Suponhamos agora que, em todos os estudantes está envolvido o mesmo conjunto de professores, ou seja, cada examinador vai avaliar mais que um estudante. A situação da figura 8, com blocos separados já não ilustra esta situação. Consideremos que em cada ocasião existe apenas uma medição ao invés de duas, para simplificação. Temos agora três professores envolvidos, A, B e C. a figura 9 ilustra esta situação para três alunos. Neste caso, os professores são cruzados com os alunos de nível 2, nos quais cada ocasião está aninhada. Esta figura 9 é um caso especial da figura 5 podendo ser representado por um diagrama semelhante à figura 6, havendo agora muitas células vazias e, no máximo, uma observação por célula.

	Aluno 1				Aluno 2		Aluno 3				
Ocasião	1	2	3	4	1	2	1	2	3	4	5
Professor A	X	X	X		X						X
Professor B				X			X	X			
Professor C						X			X	X	

Figura 9: Classificação cruzada dos professores com alunos no nível 2 e observações por ocasião no nível 1 (adaptado de Fielding and Goldstein, 2006)

Uma extensão desta situação poderá ocorrer ao considerarmos várias medições por ocasião, sendo cada uma realizada pelo mesmo professor. Nesta situação, cada X no diagrama representa várias medições (nível 1) dentro de um nível 2 (ocasiões). O cruzamento de alunos e professores passa assim a ser de nível 3.

Outras estruturas de classificação cruzada também podem ocorrer quando existe uma ruptura numa estrutura hierárquica. Consideremos, por exemplo, uma base de medidas repetidas que criamos com uma amostra de alunos pertencentes a várias turmas de uma escola, que são observados ao longo de três anos escolares. Neste caso, temos uma estrutura com três níveis hierárquicos, sendo os anos aninhados nos alunos e estes, por sua vez, aninhados em turmas da escola. Consideremos a hipótese (muitas vezes válida) de um aluno mudar de turma durante o período de tempo considerado. Nesta situação, os alunos já não podem ser considerados aninhados dentro de uma turma em particular. Para três alunos, três turmas e três ocasiões, podemos ter a situação da figura 10. Formalmente temos o mesmo tipo estrutural da figura 9. Temos assim uma estrutura com dois níveis, mas com classificação cruzada das turmas pelos alunos no nível 2 e as observações das ocasiões no nível 1.[36]

	Aluno 1			Aluno 2		Aluno 3		
Ano	1	2	3	1	2	1	2	3
Turma A	X	X		X				X
Turma B			X					
Turma C					X	X	X	

Figura 10: Alunos que mudam de turma (adaptado de Fielding and Goldstein, 2006)

[36] Para além do campo das medidas repetidas, Fielding (2002, 2004) dá exemplos da estrutura formais semelhantes às figuras 9 e 10

Vejamos agora uma situação que traz um pouco mais de complexidade à situação da figura 10, incluindo várias escolas num nível superior. Se os alunos permanecerem à mesma escola, esta será considerada unidade de nível 3, abaixo do qual estão aninhados o aluno e a turma. No entanto os alunos podem mudar de escola no decurso do estudo. Neste caso, os alunos devem ser cruzados com a escola, no nível 3. As turmas são depois aninhadas no nível 2 e as observações no nível 1. Os alunos passaram a ter cruzamento na escola e não na turma, pois desta forma são automaticamente cruzados com a turma, ao mudar de escola. Assim não é necessário especificar separadamente o cruzamento de turmas e alunos. Esta característica poderá ser estudada mais aprofundadamente em Goldstein (2003).

Outra alteração pode ser feita, tendo em conta que as turmas são "re-formadas" em cada ano lectivo. Consideremos ainda que cada turma é ensinada pelo mesmo professor durante um ano e que, cada nova turma em cada ano lectivo tem um professor diferente.[37]A figura 11 ilustra esta situação para três alunos, com quatro professores diferentes ao longo de dois anos. Para simplificar a ilustração, consideremos apenas uma escola. Temos agora uma classificação cruzada, no nível 2, de professores e alunos, com as observações em cada ocasião no nível 1. Novamente a maioria das células está vazia, não havendo mais do que uma observação por célula. Raudenbush (1993) mostra um exemplo de investigação para estudar a influência das características do aluno e do efeito professor no desenvolvimento em Matemática. Raudenbush e Bryk (2003) elaboraram outro exemplo para mostrar a flexibilidade deste tipo de estudos. Neste estudo pretendia investigar-se o efeito dos professores para o progresso dos alunos. Estes estudos mostram uma forma eficaz de resolver este problema, através de uma classificação cruzada do conjunto de professores numa ocasião com eles próprios noutras ocasiões.

[37] Note-se que as turmas, em muitas estruturas, são ministradas pelo mesmo professor e esse professor tem apenas uma turma. Neste caso, o conjunto de turmas e os professores representam o mesmo. Sendo esta a situação da figura 10, podemos chamar as unidades de turma/professor. Apesar de na figura 11 as turmas e os professores estarem em relação um para um, desde que uma turma inteira passe de professor para professor, pode-se considerar este efeito como sendo do professor.

		Aluno 1		Aluno 2		Aluno 3	
	Ano						
Professor A	1	X		X			
Professor B	1					X	
Professor C	2		X				X
Professor D	2				X		

Figura 11: Alunos que mudam de professor / turma em cada ano (adaptado de Fielding and Goldstein, 2006)

Estes exemplos foram apresentados apenas para ilustrar algumas situações no contexto de medidas repetidas. Podemos ter situações semelhantes em estudos longitudinais quando, por exemplo, os indivíduos se deslocam de uma localidade para outra, ou trabalhadores que mudam de local de trabalho. Exemplos mais complexos são proporcionados por Goldstein (2003) e Raudenbush e Bryk (2002).

3.2.4 Notação para alguns modelos cruzados

Já foi referida a notação utilizada em modelos multi-nível. Num modelo básico com dois níveis, a k-ésima unidade do segundo, era dada por u_{0k} e a variância por $\sigma^2_{u_0}$. Os efeitos de nível 2 são agora mais complexos e exigem uma extensão da notação. Em primeiro lugar, é necessário considerar a combinação de células de dois factores. Se usarmos k_1 para indicarmos uma determinada unidade da primeira classificação, k_2 representa a segundo factor. Assim, identificamos a combinação de duas unidades de cada factor do nível 2 por (k_1, k_2). Nos modelos básicos de nível 2, obtemos o efeito através da soma dos efeitos aleatórios que agora denotamos por $\left\{u^{(1)}_{0_{k_1}} + u^{(2)}_{0_{k_2}}\right\}$, sendo $u^{(1)}_{0_{k_1}}$ o erro aleatório do nível 2 correspondente ao factor k_1 e $u^{(2)}_{0_{k_2}}$ o erro aleatório do nível 2 correspondente ao factor k_2. Usando uma formulação semelhante à do modelo básico, mas recorrendo agora a uma notação de célula, para a unidade i do nível 1 e (k_1, k_2) para o nível 2, o modelo pode ser escrito como

$$y_{i(k_1,k_2)} = X_{i(k_1,k_2)}\beta + u^{(1)}_{0_{k_1}} + u^{(2)}_{0_{k_2}} + e_{0i(k_1,k_2)}$$

Onde

$y_{i(k_1,k_2)}$ é a variável resposta para a unidade i do nível 1 e k_1 e k_2 do nível 2.

$X_{i(k_1,k_2)}$ é a i-ésima variável explicativa do nível 1 correspondente aos dois factores do nível 2.

β é o coeficiente da variável $X_{i(k_1,k_2)}$

$u_{0_{k_1}}^{(1)}$ é o erro aleatório do nível 2 correspondente ao factor k_1

$u_{0_{k_2}}^{(2)}$ é o erro aleatório do nível 2 correspondente ao factor k_2

$e_{0i(k_1,k_2)}$ é o erro padrão do nível 1

Este modelo é válido, mesmo não havendo mais do que uma observação por célula. Ora, a variância de nível 1 é dada por $\sigma_{e_0}^2$, sendo agora a soma dos dois componentes de cada factor na classificação cruzada no nível 2 dada por $\left(\sigma_{u_0^{(1)}}^2 + \sigma_{u_0^{(2)}}^2\right)$, sendo $\sigma_{u_0^{(1)}}^2$ a variância do erro aleatório do nível 2 correspondente ao factor k_1 e $\sigma_{u_0^{(2)}}^2$ a variância do erro aleatório do nível 2 correspondente ao factor k_2. Para modelos de coeficientes aleatórios para uma ou ambas as classificações de cruzamento ou estruturas com funções de variância mais complexas, podemos fazer uma extensão da estrutura do modelo e da respectiva notação. Esta extensão também pode ser feita para outras formas de classificação e para mais níveis de hierarquia, com possibilidade de cruzamento em níveis mais elevados. No entanto, quanto mais complexo é o modelo, mais elaboradas são as notações. Rasbash e Brown (2001) dão alguns exemplos detalhados e regras mais complexas. Mais adiante vamos introduzir uma notação mais geral, que pode ser usada com um método de estimação recente conhecido por MCMC,[38]ou Método de Cadeia de Markov de Monte Carlo.

Apenas a título informativo e sem entrarmos em muitos detalhes, vamos caracterizar a variância devido aos efeitos cruzados, pela incorporação de uma expressão de interacção, semelhante ao referido nos procedimentos da *two-way* ANOVA. Tomando como referência os exemplos anteriores, suponhamos que o efeito marginal de

[38] Método da Cadeia de Markov de Monte Carlo (*MarkovChain Monte Carlo* (MCMC)), que representa uma alternativa ao procedimento baseado na verosimilhança (Goldstein, 2003)

residência num determinado bairro pode variar de acordo com a escola que o aluno frequenta, ou vice-versa. Por outras palavras, existe uma relação entre o bairro de residência e a escola frequentada que pode contribuir para um efeito bairro e efeito-escola para determinada célula. Este efeito é agora caracterizado pela adição de três componentes:

$$\left\{u_{0_{k_1}}^{(1)} + u_{0_{k_2}}^{(2)} + u_{0_{(k_1,k_2)}}^{(3)}\right\}$$

Onde

$u_{0_{k_1}}^{(1)}$ é o erro aleatório do nível 2 correspondente ao factor k_1

$u_{0_{k_2}}^{(2)}$ é o erro aleatório do nível 2 correspondente ao factor k_1

$u_{0_{(k_1,k_2)}}^{(3)}$ é o erro aleatório do nível 2 correspondente ao cruzamento (k_1, k_2)

Descrevemos a seguir o significado habitual dos efeitos de interacção, através de um exemplo. O efeito de k_2 deixou de contribuir directamente no k_2, independentemente da unidade k_1 em que o aluno está inserido, como acontecia se $u_{0_{k_2}}^{(2)}$ fosse adicionado a $u_{0_{k_1}}^{(1)}$. Em vez disso, a contribuição de adicionar a $u_{0_{k_1}}^{(1)}$ unidades k_2 é agora $\left\{u_{0_{k_2}}^{(2)} + u_{0_{(k_1,k_2)}}^{(3)}\right\}$ que depende da unidade k_1.

A variância correspondente no nível de classificação cruzada é dada pela soma dos componentes $\left(\sigma_{u_0^{(1)}}^2 + \sigma_{u_0^{(2)}}^2 + \sigma_{u_0^{(3)}}^2\right)$, sendo $\sigma_{u_0^{(3)}}^2$ a variância do efeito interacção, conhecida por variância da interacção.

3.2.5 Comandos do SPSS para a construção de um modelo de classificação cruzada

Analyze > Mixed Models > Linear

Em *Subjects* colocar os Factores dos níveis 2 e 3

Continue

Escolher a *Dependent Variable* e os *Factores* definidos anteriormente

Em *Covariate* colocar as co-variáveis

Em *Fixed* seleccionar essas co-variáveis e as respectivas interacções e passar para lista de *Model* e *Add*

Continue

Em *Random* marcar *Include Intercept.* Em *Subject Groupings* escolher o factor do nível 3 e passar para a lista de *Combinations*. Em *Random effect* clicar em next para adicionar o 2º efeito aleatório. Clicar novamente em *next* e colocar agora em *Subject Groupings* os dois factores em simultâneo.

Continue

Em *Statistics* seleccionar

- *Parameter Estimates*

- *Tests of Covariance Parameters*

Continue

3.3 Modelo logístico para variáveis dependentes binárias

A modelação multi-nível possui um carácter de forte generalização, sendo aplicada, também, a dados binários e/ou categorizados.

Desta forma, um modelo linear multi-nível generalizado com dois níveis de hierarquia pode ser usado para identificar predictores de dados binários (y_{ik}), tendo em conta a variação entre os indivíduos (nível 1) e entre os grupos (nível 2):

$$y_{ik} = \frac{exp\ (\beta_0 + \beta_1 X + u_k)}{1 + exp\ (\beta_0 + \beta_1 X + u_k + e_{ik})}$$

A variável resposta deste modelo y_{ik} tem distribuição binomial, tomando o valor 1 se o i-ésimo indivíduo apresentar a característica ou 0, caso contrário. Sendo β_0 e β_1 a ordenada na origem e o declive para a variável X, no nível 2. e_{ik} a componente aleatória do nível 1 e u_k é a componente aleatória do nível 2. O termo e_{ik} (nível 1) tem média 0 e variância σ_e^2. Esta variância é normalmente conhecida por parâmetro de dispersão ou parâmetro extra-binomial e a sua estimativa é 1 (Ferrão, 2000). Assim, uma parte da variabilidade do modelo é binomial (nível 1) e outra parte é normal (nível 2) (Diez-Roux, 1998, Goldstein, 2003).

A relação entre as variáveis predictoras (X) e a variável resposta é assegurada pela função de "relação" (logit), que é uma função linear de X. Temos então:

$$log\left(\frac{\pi_i}{1-\pi_i}\right) = logit(\pi_i) = \beta_0 + \beta_1 X + u_k$$

π_i representa a probabilidade associada a $y_i = 1$.

Onde $u_k \sim N(0, \Omega_u)$ é o efeito aleatório no nível 2.

Os coeficientes são:

$$\beta_0 = log\left(\frac{\pi_0}{1-\pi_0}\right)$$

$$\beta_1 = log\left(\frac{\pi_1(1-\pi_0)}{\pi_0(1-\pi_1)}\right)$$

3.3.1 Métodos de Estimação

Se utilizássemos a estimação de máxima verosimilhança nos modelos multi-nível de resposta discreta, obtínhamos um processo computacionalmente intensivo, pelo que se utiliza o método de estimação de quase-verosimilhança. Este método baseia-se na expansão da série de Taylor, transformando um modelo de resposta discreta num modelo de resposta contínua. Após a linearização, o modelo é estimado através dos Mínimos Quadrados Iterativos Generalizados (IGLS) ou Mínimos Quadrados Iterativos Generalizados Restritos ou Reponderados (RIGLS) (Goldstein, 2003, Rasbash, 2005).

Na regressão multi-nível linear a estimação dos parâmetros pode ser realizada através do método de máxima verosimilhança (ML), assumindo-se normalidade dos erros, ou alternativamente a máxima verosimilhança restrita (REML) (Goldstein, 1995). Para os modelos logísticos multi-nível as estimativas são obtidas através da maximização da função de verosimilhança marginal ou de verosimilhança penalizada.

A verosimilhança marginal, obtida pela integração da distribuição conjunta em relação ao efeito aleatório, é analiticamente intratável, porém várias aproximações têm sido propostas na literatura. As aproximações numéricas disponíveis na literatura para linearizar a função de ligação incluem aproximação de Laplace, PQL e integrações numéricas. A aproximação de Laplace lineariza a parte não linear usando expansão em série de Taylor até a segunda ordem com base na maximização de uma função conhecida e unimodal. A aproximação realizada através do PQL, por sua vez, consiste na linearização da parte não linear usando expansão em série de Taylor com base nos coeficientes de regressão estimados e nos resíduos actuais, podendo ser de primeira ou segunda ordem. Já a integração aproximada através do MQL considera a linearização apenas baseado na parte fixa do predictor linear (Hox, 2002; Goldstein e Rasbash, 1996).

O procedimento MQL de 1ª ordem produz, por vezes, estimativas enviesadas, sendo o procedimento PQL de 2ª ordem mais aperfeiçoado, no entanto menos estável.

Para detalhes técnicos consultar, por exemplo, Hox (2002) ou Goldstein e Rasbash (1996).

3.3.2 Correlação intra-grupo

Como já foi referido, a correlação intra-grupo[39] expressa a proporção da variância total "provocada" pelo nível 2. Na regressão logística multi-nível, podemos estimar o VPC através de vários procedimentos, sendo um deles o método da variável latente:

$$VPC = \frac{\sigma_{u0}^2}{\frac{\pi^2}{3} + \sigma_{u0}^2}$$

Sendo σ_{u0}^2 a variância entre as unidades do nível 2 e $\frac{\pi^2}{3}$ a variância entre unidades do nível 1, isto é, a variância de uma distribuição logística padrão. Desta forma, as duas variâncias ficam num escala contínua (Goldstein, 2003, Rasbash, 2005).

As significâncias estatísticas dos coeficientes da parte fixa e da parte aleatória do modelo multi-nível são avaliadas pelo teste de Wald.

3.3.3 Métodos de avaliação do modelo

O ajuste do modelo pode ser avaliado através *Deviance Information Criterion* (DIC = $2\overline{D} - D(\bar{\theta})$, sendo $\overline{D}$ a *deviance* média do número de iterações e $D(\bar{\theta})$ a *deviance* para o valor esperado dos parâmetros desconhecidos), que é uma generalização do Critério de Informação de Akaike (AIC = D + 2q, sendo D a *deviance* e q o número de parâmetros estimados) e é calculado a partir de uma estimação Bayesiana, através do método da Cadeia de Markov de Monte Carlo (*MarkovChain Monte Carlo* (MCMC)), que representa uma alternativa ao procedimento baseado na verosimilhança (Goldstein, 2003). Quanto menor for o valor do DIC, melhor o modelo.

A análise dos resíduos do nível 2 pode ser avaliada através do gráfico dos resíduos reduzidos ordenados com IC 95% (Merlo, 2005).

[39] *Variance Partition Coefficient* (VPC)

Capítulo IV

4 Caso prático – Análise Descritiva

4.1 Enquadramento Geográfico e Institucional

A ETLA (Escola Tecnológica do Litoral Alentejano) é uma escola profissional que comemorou dia 1 de Outubro de 2010 o seu vigésimo aniversário. Situa-se em pleno complexo petroquímico da REPSOL, em Sines. Surgiu de um acordo com a empresa que se encontrava na altura no complexo petroquímico, a NESTE POLÍMEROS. O seu aparecimento deveu-se à necessidade de técnicos especializados nas áreas de Química, Electrónica e Instrumentação e Informática para integrar os quadros das empresas da região.

Até ao ano lectivo de 2000 / 2001 manteve-se apenas com os três cursos referidos. Foi neste ano que se iniciou um dos cursos mais procurados da escola: o curso de Mecatrónica. Passados 6 anos, no ano lectivo de 2006 / 2007 começou então o curso mais recente da escola: o curso de Higiene e Segurança do Trabalho e Ambiente.

Actualmente estão em funcionamento os cinco cursos, estando a sua abertura dependente do número de candidatos em cada ano lectivo.

Sendo estes cursos de carácter tecnológico, tem havido a necessidade de se ajustar os programas, de forma a poderem estar o mais actualizado possível. Para além destes ajustes, houve uma alteração dos currículos de todos os cursos profissionais, por parte do Ministério da Educação, de forma a tornar este tipo de ensino mais homogéneo. São estes anos lectivos, cujos cursos já foram abrangidos pelos novos currículos, que têm interesse para o nosso estudo, encontrando-se os cursos dos antigos currículos já obsoletos.

Assim sendo, a nossa amostra será constituída pelos cursos de:

- Química Tecnológica
    - 2004 / 2007 (11 alunos)
    - 2005 / 2008 (14 alunos)
    - 2007 / 2010 (20 alunos)

- Electrónica, Automação e Instrumentação
    - 2005 / 2008 (16 alunos)
    - 2006 / 2009 (19 alunos)
    - 2007 / 2010 (20 alunos)
- Informática de Gestão
    - 2005 / 2008 (16 alunos)
- Mecatrónica
    - 2005 / 2008 (16 alunos)
    - 2006 / 2009 (20 alunos)
    - 2007 / 2010 (20 alunos)
- Higiene e Segurança do Trabalho e Ambiente
    - 2006 / 2009 (18 alunos)
    - 2007 / 2010 (15 alunos)
    -

4.2 Recolha, análise e tratamento de dados

Para ser possível a realização deste estudo, foi necessário recolher as avaliações das turmas seleccionadas, bem como alguns dados pessoais dos alunos.

Esta recolha foi efectuada directamente na secretaria da escola, com prévia autorização da direcção da mesma.

No que refere à análise, esta foi quantitativa.

O tratamento de dados foi feito no programa estatística SPSS. Foram tratados os dados de 205 alunos dos quais 5 tinham falta de dados.

4.3 Questões a serem respondidas com este estudo

Através deste estudo pretendemos obter resposta para um conjunto de questões, nomeadamente,

1) A idade de entrada está relacionada com o sucesso escolar?
2) O facto da localidade de proveniência do aluno ser rural ou urbana tem influência na classificação média?
3) Existem diferenças significativas nas classificações dos alunos relativamente ao género?
4) Existe diferença significativa no rendimento escolar entre alunos dos vários cursos?
5) A média final do curso está relacionada com o sexo, idade ou proveniência do aluno?
6) O ano lectivo do curso tem influência na média das classificações dos alunos?

4.4 Definição dos níveis

Neste trabalho utilizaram-se inicialmente dois níveis. O nível 1, que corresponde aos alunos, e o nível 2, que corresponde ao curso.

Para a realização do modelo de três níveis, utilizámos para nível 1 os alunos, para nível 2 os cursos e para nível 3 os professores.

Para o modelo de classificação cruzada, temos o nível 1 os alunos e no nível 2 (de cruzamento) os professores e os cursos.

4.5 Variáveis a analisar

Com o intuito de obter resposta para as questões referidas, foram recolhidas diversas variáveis. Assim, para cada aluno em estudo, foram recolhidas as seguintes informações:

- Curso frequentado (Electrónica, Mecatrónica, Química, HSTA, Informática)
- Ano lectivo do início do curso
- Sexo
- Idade no início do curso
- Zona de residência (urbana, se reside em Sines, Santo André e Santiago do Cacém e rural em todos os outros casos)
- Sucesso (sim, se o aluno concluiu o curso até Julho do 3º ano do curso e não, caso contrário)
- Médias das componentes sociocultural, científica e técnica
- Média final
- Número de módulos em atraso das componentes sociocultural, científica e técnica
- Número total de módulos em atraso
- Rendimento (sim, se a média é superior ou igual a 14 e não, caso contrário)

Durante a recolha de dados, também se recolheu a variável que indicava se o aluno entrou na sua 1ª opção, para estudar se esse facto poderia ter influência no seu sucesso e média. No entanto, constatou-se que apenas 3 dos alunos não entraram na 1ª opção, logo esta não era uma variável com importância.

A tabela seguinte apresenta a descrição das variáveis utilizadas neste estudo:

Variável	Descrição	Nível da variável
Curso	Curso frequentado (Electrónica, Mecatrónica, Química, HSTA, Informática)	
Anolect	Ano lectivo do início do curso	Nível 2
Sexo	Sexo	Nível 1
Idade	Idade no início do curso	Nível 2
Idade_centrada	Idade centrada, que se obteve subtraindo o valor da média geral das idades a cada valor da variável Idade	Nível 2
Zona	Zona de residência (urbana, se reside em Sines, Santo André e Santiago do Cacém e rural em todos os outros casos)	Nível 1
Sucesso	Sucesso (sim, se o aluno concluiu o curso até Julho do 3º ano do curso e não, caso contrário)	Nível 1
Sociocult; cientif e tecnica (respectivamente)	Médias das componentes sociocultural, científica e técnica	Nível 2
Media_total	Média final	Nível 2
Modatrasc; modatraci e modatratec	Número de módulos em atraso das componentes sociocultural, científica e técnica	Nível 2
Totalmodatra	Número total de módulos em atraso	Nível 2
rendimento	Rendimento (sim, se a média é superior ou igual a 14 e não, caso contrário)	Nível 1
Profs	É a conjugação de professores de Português e Matemática que existe nas turmas em estudo	
Prof_mat	Professores de Matemática (Cl e Co)	Nível 3
Prof_port	Professores de Português (Mm e It)	Nível 3

Tabela 2: Descrição das variáveis utilizadas no estudo

4.6 Estatística Descritiva dos dados

Para proceder à estatística descritiva dos dados em questão, foi utilizado o programa SPSS, versão 17.0.

4.6.1 Dados referentes aos alunos

Tendo em conta a generalidade dos alunos, no que diz respeito ao género, temos a maioria de rapazes (68,8%) e apenas 31,2% de raparigas.

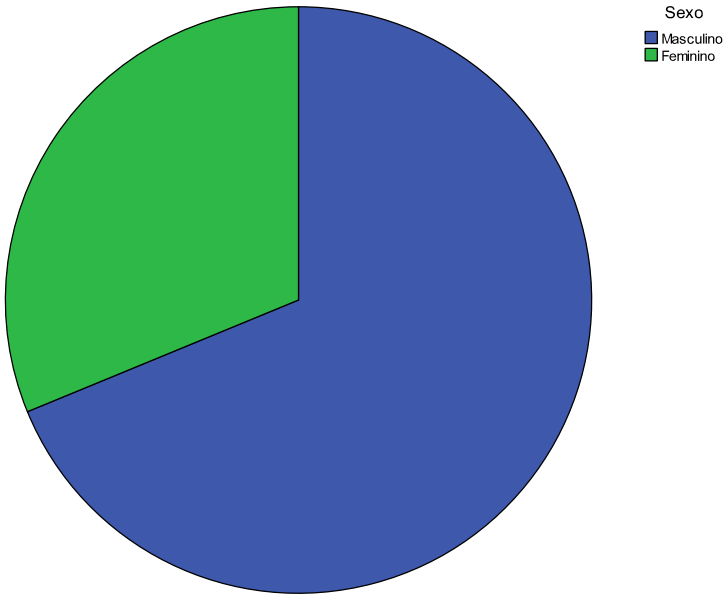


Gráfico 1: Distribuição por género

No que refere à distribuição por ano lectivo, temos sempre uma maioria de rapazes, à excepção do ano lectivo de 2004/2005, em que apenas se está a considerar a turma de Química, pois os restantes cursos ainda funcionavam segundo os currículos antigos.

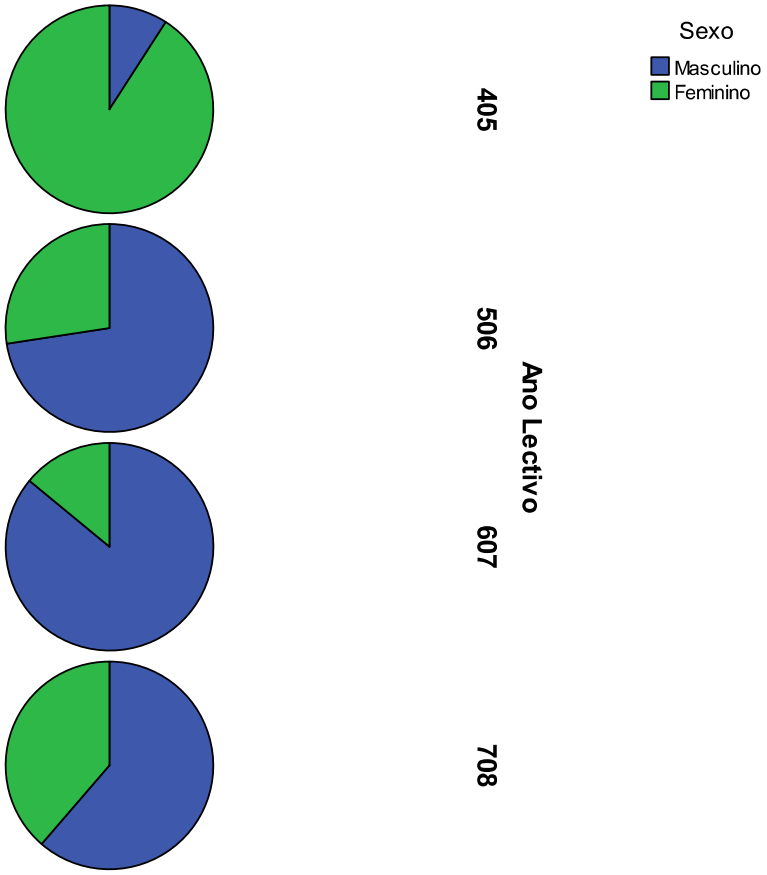
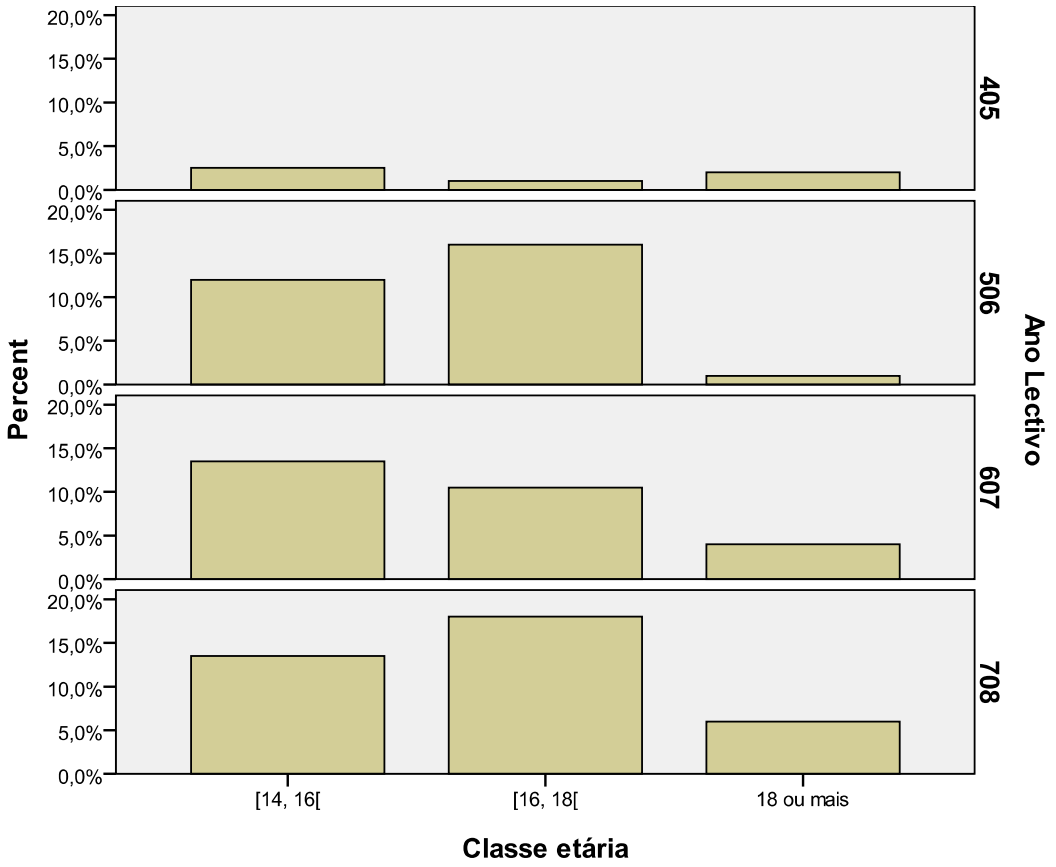


Gráfico 2: Género por ano lectivo

Relativamente à classe etária dos alunos, foram utilizadas 3 classes: dos 14 aos 16 anos, doa 16 aos 18 anos e 18 anos ou mais. No geral, 40% dos alunos têm entre 14 e 16 anos, cerca de 45% entre 16 e 18 e apenas 13 % têm 18 anos ou mais, sendo que existem 5 valores em falta. Relativamente à distribuição etária dos alunos, por ano lectivo, a maioria dos alunos encontra-se nas classes etárias [14, 16[e [16, 18[, sendo que apenas no ano lectivo de 2004 / 2005 existem mais de 36% dos alunos com 18 anos ou mais.

Gráfico 3: Classe etária por ano lectivo



De um modo geral, metade dos alunos obteve sucesso. Fazendo uma breve análise ao sucesso dos alunos segundo o ano lectivo, apesar não haver diferenças muito significativas entre as percentagens de alunos com e sem sucesso, verificamos um pequeno decréscimo do número de alunos com sucesso, exceptuando o ano lectivo de 2006 / 2007, no qual cerca de 60% dos alunos obtiveram sucesso.

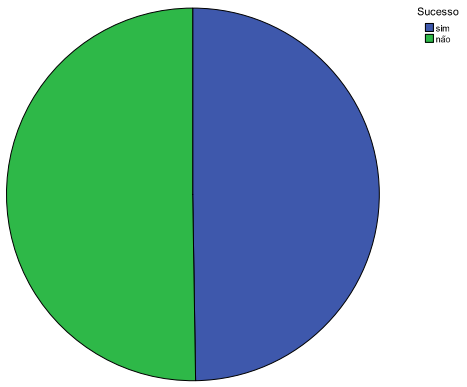


Gráfico 4: Distribuição da variável Sucesso

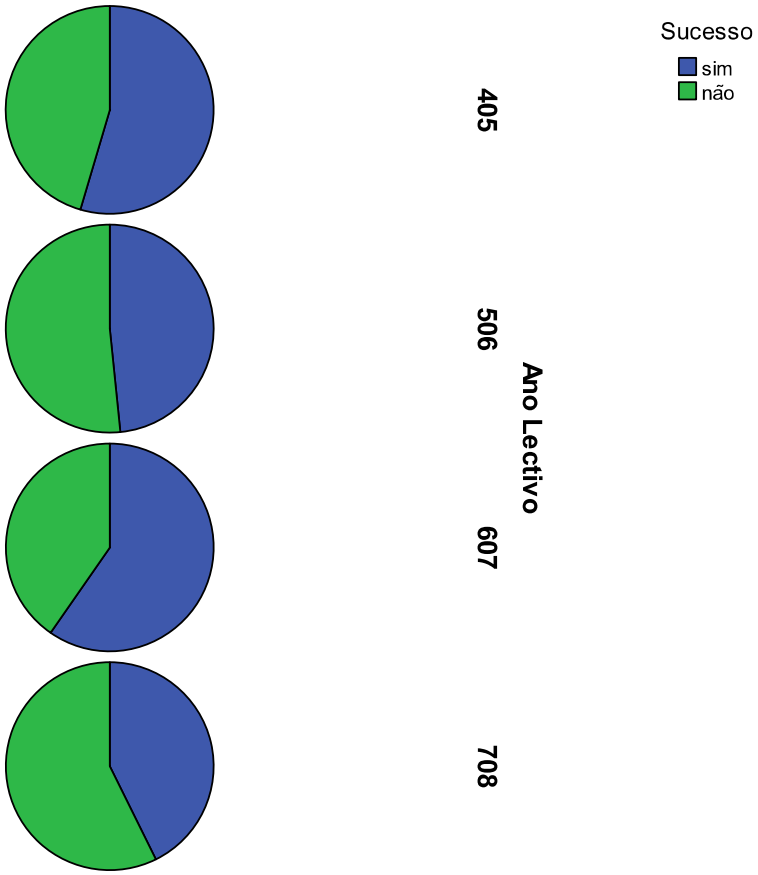


Gráfico 5: Sucesso por ano lectivo

De um modo geral, existem mais alunos (62,3%) residentes em região urbana. É no ano lectivo de 2005 / 2006 que esta diferença mais se evidencia. (ver anexo 1)

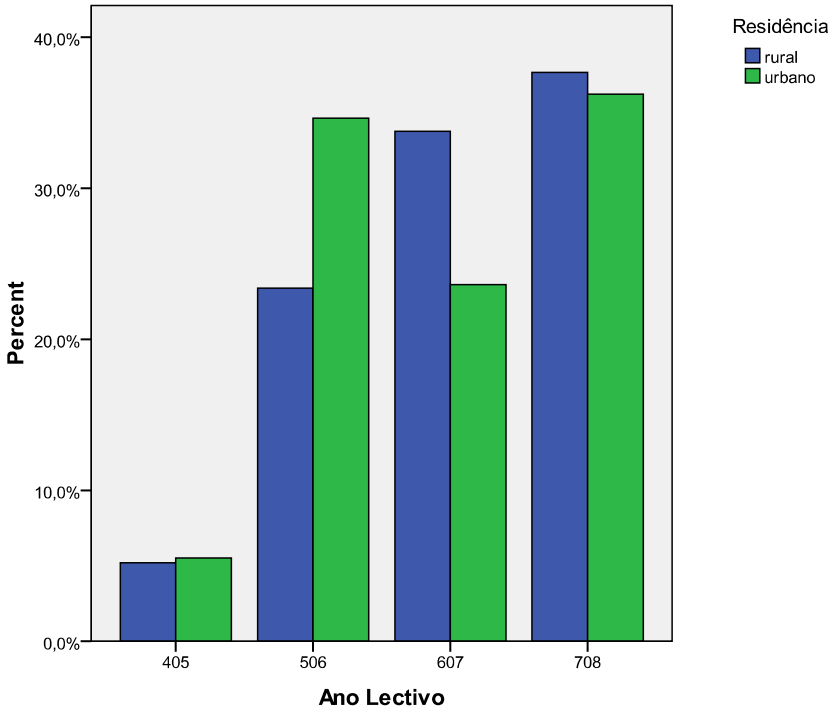


Gráfico 6: Residência por ano lectivo

4.6.2 Dados referentes ao curso

No conjunto das turmas seleccionadas, verifica-se que 27,3% são do curso de Mecatrónica, 26,8% do curso de Electrónica, 22% do curso de Química 7,8% do curso de Informática e os restantes 16,1% do curso de HSTA. Analisando as turmas segundo a dimensão, verifica-se que as turmas com mais alunos são as dos cursos de Mecatrónica e Electrónica.

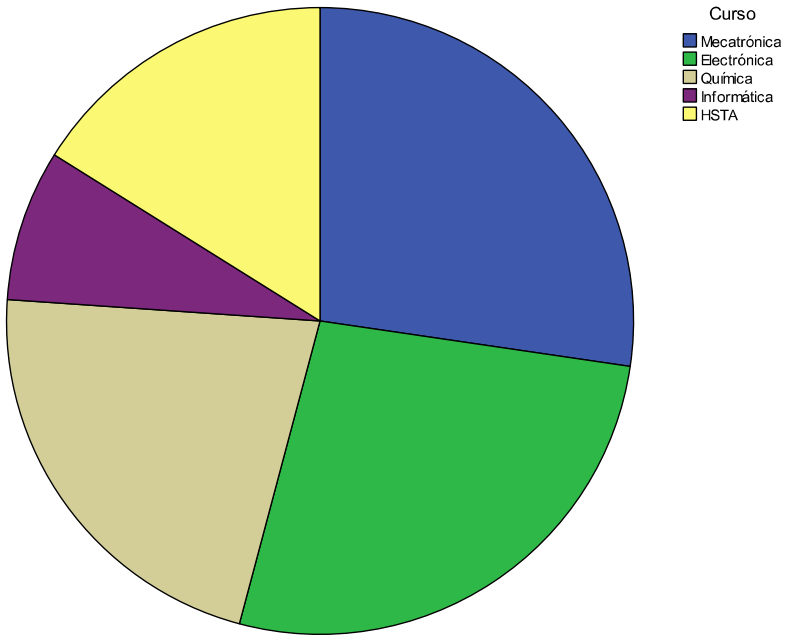


Gráfico 7: Distribuição de alunos por curso

Nos cursos de Mecatrónica e Electrónica temos uma clara maioria de rapazes (mais de 90%), também no curso de Informática existem mais rapazes (cerca de 70%). No curso de Química a tendência inverte-se (mais de 80% de raparigas) e no curso de HSTA já existe um maior equilíbrio (54,5% de rapazes e 45,5% de raparigas).

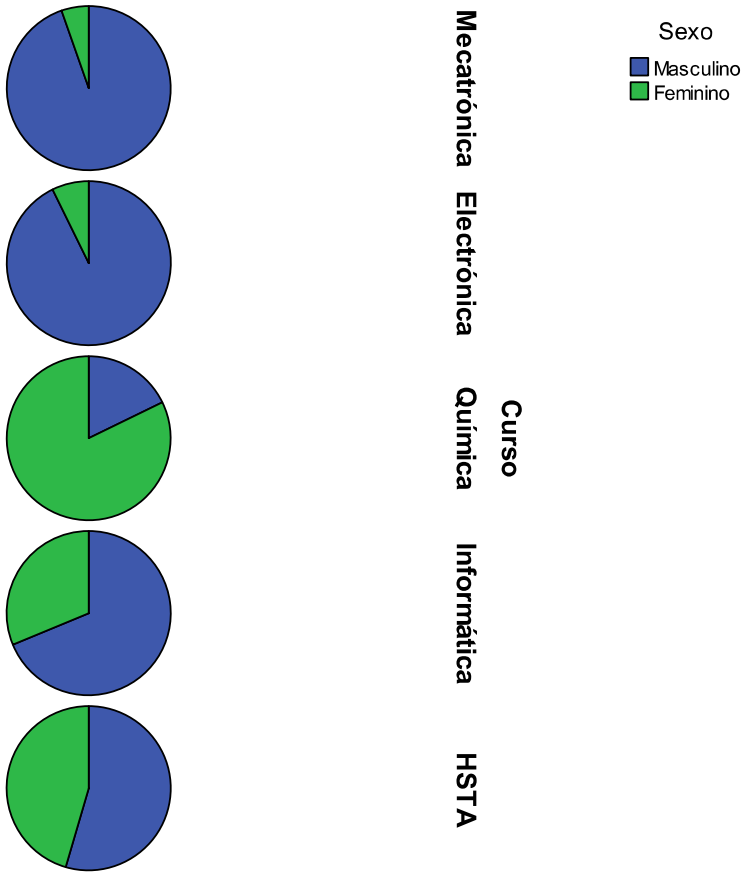


Gráfico 8: Género por curso

No que refere à distribuição por curso, de um modo geral a maioria dos alunos tem entre 16 e 18 anos, à excepção do curso de Electrónica, no qual se verifica que cerca de 50% dos alunos tem entre 14 e 16 anos. Os cursos que têm mais alunos mais velhos são os de Electrónica (17%) e HSTA (18,2%).

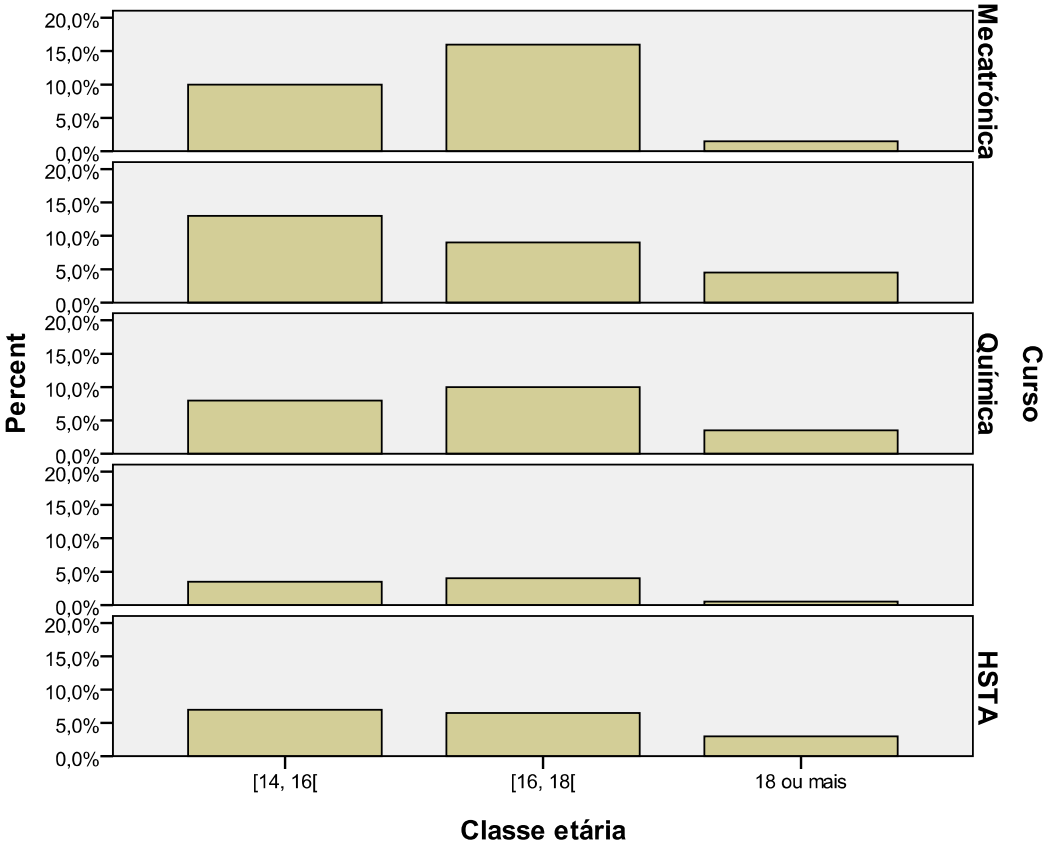


Gráfico 9: Classe etária por curso

Quando fazemos esta análise por curso, verificamos que no curso de Mecatrónica existe mais insucesso (cerca de 60%), Nos cursos de Electrónica e Química as percentagens de sucesso e insucesso são equivalentes e nos cursos de Informática e HSTA, existem mais alunos com sucesso.

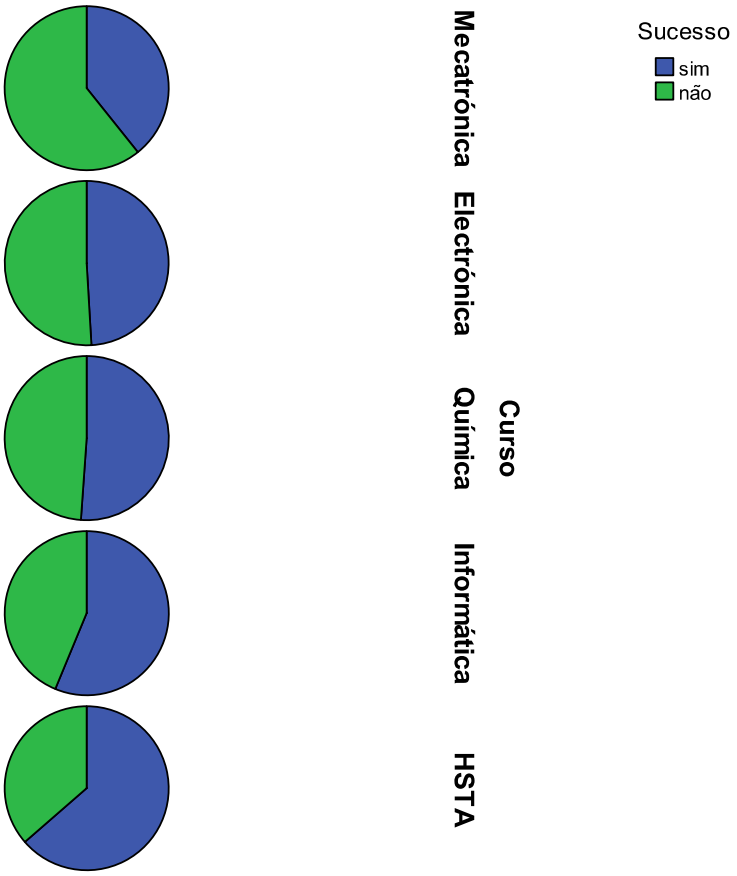


Gráfico 10: Variável sucesso por curso

Relativamente à proveniência dos alunos, segundo o curso a maioria dos alunos são residentes em zona urbana, sendo no curso de informática que esta tendência é mais marcada (mais de 90%).

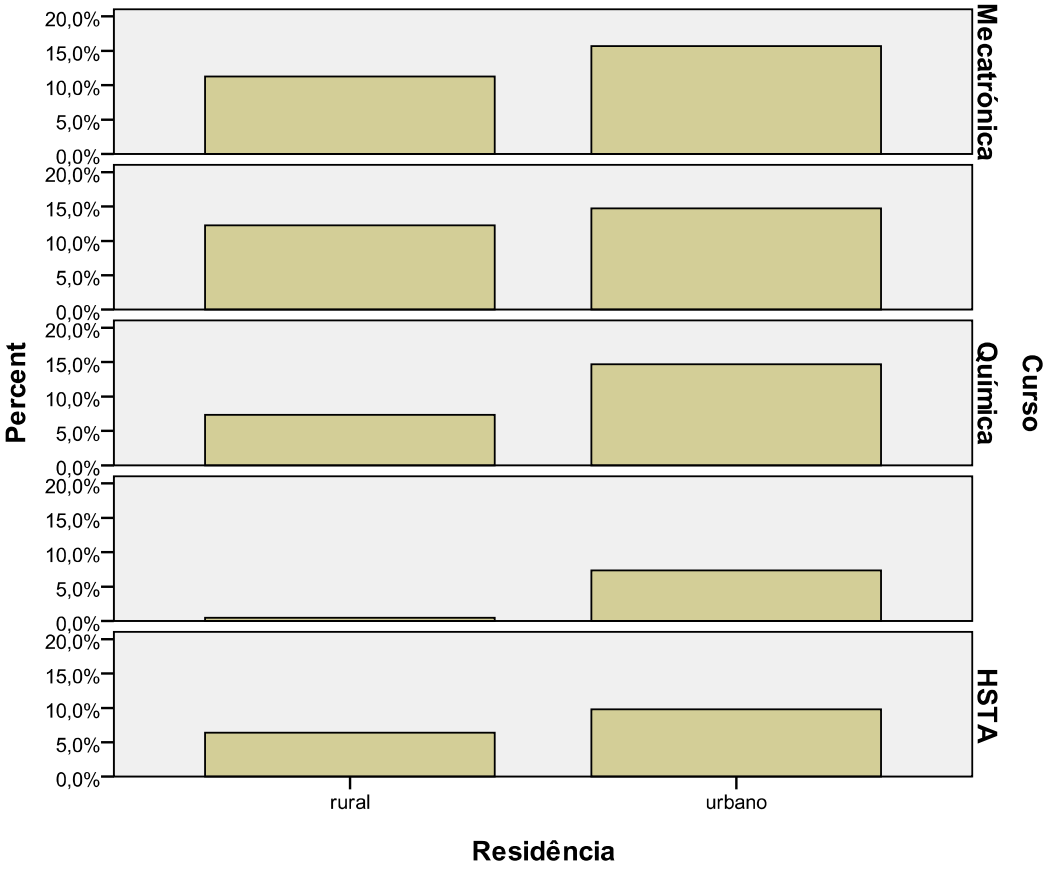


Gráfico 11: Residência por curso

Através da análise da média dos alunos segundo o curso, verifica-se que não existe grandes diferenças entre as médias finais entre cursos. No que respeita à média das componentes sociocultural e técnica, os cursos de Química e Informática revelam médias ligeiramente superiores. O número de módulos em atraso, que está directamente relacionado com o insucesso dos alunos, é claramente superior nos cursos de Mecatrónica, Electrónica e Química. (ver anexo 2)

Capítulo V

5 Construção dos modelos estatísticos

5.1 O modelo adoptado

Este modelo pretende dar resposta às seguintes questões:

- Existem diferenças significativas na média final do aluno, entre os vários cursos?

- A idade do aluno, o género sexual, a zona de residência, o sucesso, o ano lectivo e o número de módulos em atraso poderão ter influência na média final do aluno, tendo em conta o factor curso?

O modelo de regressão multi-nível para dois níveis é dado pela equação

$$Y_{ik} = \gamma_{00} + \gamma_{p0}X_{pik} + \gamma_{0q}W_{qk} + \gamma_{pq}W_{qk}X_{pik} + u_{pk}X_{pik} + u_{0k} + e_{ik}.$$

Neste trabalho, o método utilizado para a elaboração do modelo é o método dos cinco passos referido em 2.4.1. para a elaboração do modelo usou-se o *software* SPSS versão 16.0.

1) **ANOVA[40]: um factor de efeitos aleatórios**

   Modelo vazio ou modelo nulo

media_total como variável dependente e *curso* como factor

Este modelo indica:

1) A variância dentro de cada curso, ou seja, a diferença entre as médias dos alunos no mesmo curso (variância de nível 1)
2) A variância entre as médias dos diferentes cursos, ou seja, a diferença entre as médias dos cursos (variância do nível 2)

Como já foi mencionado, neste caso, o modelo é

$Y_{ik} = \beta_{0k} + e_{ik} = \gamma_{00} + u_{ok} + e_{ik}$

[40] Análise de variância

Nesta equação $\beta_{0k} = \gamma_{00} + u_{ok}$ representa a média de Y no k-ésimo curso; γ_{00} representa a ordenada na origem da regressão, u_{ok} e e_{ik} são os resíduos (erro) do nível 2 e do nível 1, respectivamente.

Neste caso, a média do aluno (Y_{ik}) é interpretado com sendo o resultado entre a media do curso a que pertence (β_{0k}) e os resíduos (e_{ik}). Assumimos que os erros se distribuem normalmente, com média zero e variância σ_e^2 , igual em todos os cursos.

No nível 2 (nível do curso), a média de cada curso (β_{0k}) interpreta-se como a combinação entre a média na população dos cursos (γ_{00}) e a variação aleatória de cada centro (u_{0k}) em torno da média.

$$\beta_{0k} = \gamma_{00} + u_{0k}$$

Assume-se que o componente aleatório u_{0k} tem média zero e variância σ_{u0}^2. Obtemos o modelo combinado

$$Y_{ik} = \gamma_{00} + u_{0k} + e_{ik}$$

Que corresponde ao modelo ANOVA com um factor de efeitos aleatórios, donde podemos usar a notação convencional dos modelos ANOVA:

$$Y_{ik} = \mu + \alpha_k + e_{ik}$$

Através do SPSS, obtemos no output as tabelas, que se seguem.

Descriptive Statistics

Media final				
Curso	Count	Mean	Standard Deviation	Coefficient of Variation
Mecatrónica	56	13,0071	1,02209	7,9%
Electrónica	55	12,9727	1,29049	9,9%
Química	45	13,8378	1,53583	11,1%
Informática	16	13,5312	1,53350	11,3%
HSTA	33	12,7727	1,02691	8,0%
Total	205	13,1834	1,31198	10,0%

Tabela 3: Informação descritiva modelo curso

Através desta tabela concluímos que o número de alunos por curso varia entre 16 e 56, num total de 205. A média final obtida na avaliação (de 0 a 20) não é igual em todos os cursos: o curso de Electrónica tem a média mais baixa (12,97) e o curso de Química a média mais alta (13,84), desta forma, parece que a média final do aluno possa estar relacionada com o curso.

Information Criteria[a]	
-2 Restricted Log Likelihood	686,312
Akaike's Information Criterion (AIC)	690,312
Hurvich and Tsai's Criterion (AICC)	690,372
Bozdogan's Criterion (CAIC)	698,948
Schwarz's Bayesian Criterion (BIC)	696,948
The information criteria are displayed in smaller-is-better forms.	
a. Dependent Variable: Media final.	

Tabela 4: Estatísticas de ajuste global (modelo nulo)

Através dos dados da tabela 4 podemos estudar em que medida é que o modelo proposto é capaz de representar a variabilidade observada nos dados (o ajuste do modelo é tanto melhor quanto menor é o valor destas estatísticas).

O primeiro destes valores é a *deviance* [41](-2LL). Os restantes são modificações de -2LL que incrementam o seu valor através de alguma função do número de parâmetros.

Temos então

$$AIC = -2LL + 2d$$ [42]

$$AICC = -2LL + \frac{2dn}{n - d - 1}$$ [43]

[41] Possibilita a comparação do grau de ajuste de modelos alternativos.

[42] Critério de informação de Akaike (Akaike, 1973)

[43] Critério de informação de Akaike corrigido (Hurvich e Tsai, 1989)

$$CAIC = -2LL + d[\log(n) + 1]$$[44]

$$BIC = -2LL + dlog(n)$$[45]

Utilizando o método MV[46]

LL representa o logaritmo da verosimilhança

d é o número de parâmetros associados aos efeitos fixos mais o número de parâmetros associados aos efeitos aleatórios

n é o número total de casos

Se se utiliza o método MVR[47]

LL representa o logaritmo da verosimilhança restrita

d é o número de parâmetros associados aos efeitos aleatórios

n é o número total de casos menos o número de parâmetros associados aos efeitos fixos

Para os valores aqui apresentados utilizou-se a MVR.

Estes critérios não têm uma interpretação directa, contudo são muito úteis para comparar modelos alternativos sempre que um deles inclua todos os termos do anterior. A diferença entre -2LL correspondentes a dois modelos distintos, segue uma distribuição qui-quadrado, com o número de graus de liberdade igual ao número de parâmetros em que diferem os dois modelos comparados, obtendo assim o ganho que se obtém ao acrescentar os efeitos em que diferem ambos os modelos.

[44] Critério de informação de Akaike consistente (Bozdogan, 1987)

[45] Critério de informação bayesiano (Schwarz, 1978)

[46] Máxima verosimilhança (ML na literatura anglosaxónica), formalizada e estimulada a partir do artigo de Hartley e Rao (1967). Mais tarde, a modificação proposta por Patterson e Thompson (1971), hoje conhecida máxima verossimilhança restrita ou residual, tornou-se uma opção ainda mais atractiva.

[47] Máxima verosimilhança restrita (REML na literatura anglo-saxónica)

Apesar da avaliação de um efeito concreto ser parte dos resultados dados no SPSS, a estratégia baseada na alteração da *deviance* é mais fiável do que o teste de Wald para amostras pequenas, pois a Razão de Verosimilhança (RV) é menos conservadora que o teste de Wald, que algumas vezes pode falhar em rejeitar H_0. Isto significa que os coeficientes de regressão de algumas variáveis podem apresentar *p-values* descritivos nos testes de Wald > 0,05 (não significantes) sinalizando para a possibilidade de exclusão dessas variáveis dos modelos, enquanto tal exclusão não será permitida quando utilizado o teste da razão de verosimilhança. Esta constatação indica que a estatística de Wald constitui um bom teste durante a triagem inicial das variáveis (análises univariadas), servindo para apontar, nesta etapa, quais as variáveis que deverão compor os modelos multivariados. Uma vez composto o elenco de variáveis para os modelos multivariados, o critério de exclusão a partir de então deverá estar baseado no valor obtido para a razão de verosimilhança.

A tabela 5 indica o valor estimado da ordenada na origem, que é o único parâmetro de efeitos fixos no modelo. Esta estimação representa a média populacional dos cinco cursos na variável dependente media_total. Temos a estimação $\hat{\mu} = 13{,}2$ e o respectivo erro padrão 0,20, bem como o número de graus de liberdade, o valor estabelecido, que se obtém dividindo a estimação pelo erro padrão, e o p-value , para testar a hipótese de que o parâmetro é zero.

$$H_0: \gamma_{00} = 0$$

$$H_1: \gamma_{00} > 0$$

Neste caso, como p-value = 0 < 0,0005, podemos concluir que a ordenada na origem é diferente de zero, com uma probabilidade de erro de 0,05%. Desta forma concluímos que a média da população de alunos é maior que zero (como seria de esperar, pois todas as notas que aparecem são superiores a 10, as negativas aparecem como valor em falta).

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1,321047E1	,200326	4,007	65,945	,000	12,654639	13,766308

a. Dependent Variable: Media final.

Tabela 5: Estimação dos efeitos fixos (modelo nulo)

Na tabela 6 temos as estimações dos parâmetros associados aos efeitos aleatórios do modelo. A variância do factor curso (0,155343) indica quanto varia a variável dependente entre os cursos. A variância dos resíduos ($\hat{\sigma}_e^2$ =1,600443) indica quanto varia a variável dependente dentro de cada curso. Segundo estas estimações, a variabilidade entre os centros representa $\rho = \frac{0,155}{0,155+1,60} = 0,088$, apenas 8,8% ≈ 9% da variabilidade total. Este quociente denomina-se por coeficiente de correlação intra-classe[48] e representa o grau de variabilidade existente entre os diferentes cursos em comparação com a variabilidade existente entre os alunos do mesmo curso. Neste caso, significa que aproximadamente 9% da variância das classificações médias podem ser atribuídos ao nível do curso.

A tabela 6 dá-nos ainda informação que nos permite testar a significância de cada estimação. A hipótese que pretendemos testar no modelo é se o efeito do factor é nulo.

$$H_0: \sigma_\beta^2 = 0$$

versus

$$H_1: \sigma_\beta^2 > 0$$

Para fazer este teste, recorremos à estatística Z de Wald. Este teste tem um *p-value* de 0,267 > 0,05, pelo que não rejeitamos a hipótese nula, de que a variância populacional do factor curso é zero, podendo a média não diferir de curso para curso. No entanto, dado que o teste Wald é muito conservador para amostras pequenas, talvez seja prudente pensarmos que fica por explicar parte das diferenças entre os cursos.

Os parâmetros de co-variância estimaram-se assumindo que o factor curso é independente dos resíduos.

[48] Ver pág. 40

Estimates of Covariance Parameters[a]						
Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval
						Lower Bound / Upper Bound
Residual		1,600443E0	,159991	10,003	,000	1,315674 / 1,946848
curso	Variance	,155343	,140008	1,110	,267	,026553 / ,908798

a. Dependent Variable: Media final.

Tabela 6: Estimação dos parâmetros de co-variância (modelo nulo)

Obtemos assim o modelo nulo, tal como se segue:

Modelo Nulo

$$Y_{ik} = 13{,}2 + u_{0k} + e_{ik}$$

2) Análise de regressão: ordenadas na origem como resultados

Inclusão da co-variável *idade_centrada*

Depois de se verificarem diferenças entre as médias dos cursos, o passo seguinte é averiguar se há alguma variável capaz de justificar essas diferenças. Comecemos por incluir a variável de nível 2 *idade*.

Relativamente ao modelo nulo apresentado anteriormente, o modelo actual apenas acrescenta uma co-variável do nível 2. Assim, o modelo de nível 1 continua a ser

$$Y_{ik} = \beta_{0k} + e_{ik}$$

E o modelo do nível 2 passará a ser

$$\beta_{0k} = \gamma_{00} + \gamma_{01} w_k + u_{0k}$$

Com $w_k = W_k - \overline{W}$, com W_k a representar a k-ésima observação da variável e $\overline{W}$ a média de todas as observações da variável W (para que a constante γ_{00} tenha um significado claro, utilizam-se os diferenciais w em vez dos valores directos W).

Substituindo, obtemos o modelo combinado

$$Y_{ik} = \gamma_{00} + \gamma_{01} w_k + (u_{0k} + e_{ik})$$

Onde, como se sabe,

Y_{ik} é a variável resposta para o elemento i do nível 1 e k do nível 2.

w_k é a variável explicativa do nível 2.

e_{ik} é o erro aleatório relativo ao nível 1

u_{0k} é o erro aleatório para cada elemento do nível 2 (afastamento em relação à ordenada média).

Este modelo pretende predizer a média de cada curso a partir da idade média dos seus alunos.

Como a constante (ordenada na origem) do nível 1, β_{0k}, que representa a média da variável dependente quando se utilizam variáveis independentes centradas), é função dos coeficientes e variáveis do nível 2, chamamos este modelo de *médias*[49] *como resultados*.

É de notar que o termo u_{0k} não se refere exactamente ao efeito do factor curso, mas ao efeito do factor curso depois de incluída a co-variável w. Da mesma forma, a variância que exprime a variabilidade entre os cursos, $\hat{\sigma}^2_{u_0}$, é agora uma variância condicional: indica como variam os cursos ao incluir as diferenças atribuídas à co-variável w.

Da tabela 7 obtemos o valor da ordenada na origem ($\hat{\gamma} = 13{,}2$) e o coeficiente associado à co-variável idade ($\hat{\gamma}_{01} = -0{,}12$). Sabendo que a co-variável *idade_centrada* é centrada[50], o valor da ordenada na origem é uma estimação da média na população de centros. O valor do coeficiente associado à co-variável indica que por cada ano que aumenta a idade média num curso, a média final dos alunos diminui 0,12 valores. Como este coeficiente tem associado uma estatística t, cujo p-value = 0,038 < 0,05, podemos afirmar que a idade dos estudantes está relacionada com a média final de curso.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1,322810E1	,204270	4,039	64,758	,000	12,663109	13,793097
idade_centrada	-,124740	,059759	194,716	-2,087	,038	-,242599	-,006881

a. Dependent Variable: Media final.

Tabela 7: Estimação dos parâmetros dos efeitos fixos (passo 2)

Através da tabela 8 podemos observar a estimação da variância dos resíduos ($\hat{\sigma}^2_e = 1{,}57$), semelhante à do modelo nulo ($\hat{\sigma}^2_e = 1{,}60$), logo a presença da co-variável do nível 2 idade não parece ter afectado a variabilidade do nível 1. Também na estimação da variabilidade entre os centros ($\hat{\sigma}^2_{u_0}$) aumentou um pouco. No modelo vazio era de 0,155 e agora passou a ser 0,163, logo a variabilidade do nível 2 ficou ligeiramente afectada pela presença da co-variável do nível 2. O *p-value* do teste de Wald (0,26)

[49] Ou constantes, ou intersecções
[50] Variável centrada: $z_k = Z_k - \bar{Z}$

mostra que depois de introduzir a idade dos alunos, não parece que os cursos diferem na média. No entanto, mais uma vez alertamos que sendo este teste pouco adequado para amostras pequenas, poderá ficar por explicar parte das diferenças entre os cursos. De facto, por análise da estatística -2LL nos dois modelos, chegamos à conclusão que a variância entre os cursos é diferente de zero. Como podemos observar, no modelo nulo obtivemos -2LL = 686,312 e quando incluímos a variável idade, obtivemos -2LL = 668,507 (ver figura 9). A diferença entre ambos os valores (17,805) segue uma distribuição qui-quadrado com 1 grau de liberdade (os dois modelos apenas diferem de um parâmetro - γ_{01}). A probabilidade de encontrar valores maiores ou iguais a 17,805 na distribuição qui-quadrado com um grau de liberdade é inferior a 0,005. Daqui podemos concluir que, depois de inserir o efeito da idade, a média não é a mesma em todos os cursos, isto é, a variância das médias dos cursos é maior que zero.

Estimates of Covariance Parameters[a]

Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Residual	1,567046E0	,159047	9,853	,000	1,284368	1,911940
Intercept [subject = curso]	Variance ,163399	,145209	1,125	,260	,028629	,932594

a. Dependent Variable: Media final.

Tabela 8: Estimação dos parâmetros de co-variância (passo 2)

Information Criteria[a]

-2 Restricted Log Likelihood	668,507
Akaike's Information Criterion (AIC)	672,507
Hurvich and Tsai's Criterion (AICC)	672,568
Bozdogan's Criterion (CAIC)	681,083
Schwarz's Bayesian Criterion (BIC)	679,083

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: Media final.

Tabela 9: Estatísticas de ajuste global (passo 2)

Para determinar qual a proporção da variância total que se deve às diferenças entre os cursos, calculemos o coeficiente de correlação intra-classe:

$$\rho = \frac{\hat{\sigma}^2_{u_0}}{\hat{\sigma}^2_{u_0}+\hat{\sigma}^2_{e}} = \frac{0{,}163399}{0{,}163399+1{,}567046} = 0{,}094$$

Este valor indica que, ao acrescentar o efeito atribuível à idade média, 9,4% da variância total (variância da variável dependente) ainda se atribui às diferenças entre as médias dos cursos. Este coeficiente, que agora está condicionado, informa o que ocorre nos cursos em relação à sua média quando se acrescenta a variável idade.

No modelo nulo, ρ = 8,8%, pelo que, neste modelo, aumentou um pouco.

Comparando as estimações dos parâmetros da co-variância do modelo nulo e deste modelo, ficamos a conhecer a proporção de variância explicada no nível 1:

$$R_1^2 = \frac{1{,}600443 - 1{,}567046}{1{,}600443} = 0{,}02$$

E no nível 2:

$$R_2^2 = \frac{1{,}600443-0{,}163399}{1{,}600443} = 0{,}898$$

Este valor significa que cerca de 90% das diferenças observadas entre os cursos (diferenças na classificação média) são diferenças atribuíveis à idade dos alunos.

Modelo de análise de regressão: ordenadas na origem como resultados

$$Y_{ik} = 13{,}2 - 0{,}12 idade\_centrada + (u_{0k} + e_{ik})$$

3) ANCOVA[51]: um factor de efeitos aleatórios

Inclusão da co-variável *zona*

Uma co-variável do nível 2, como era o caso da idade, permite explicar as diferenças existentes entre as médias dos cursos, isto é, a variabilidade do nível 2. Para estudar a variabilidade do nível 1, ou seja, as diferenças entre os alunos do mesmo curso, é necessária uma co-variável do nível 1.

Para tal, vamos usar a variável *zona,* uma variável dicotómica que indica a zona de proveniência do aluno: urbano / rural. A variável *zona* toma o valor 0 para "rural" e 1 para "urbano" (sendo uma variável dicotómica pode ser incluída nas co-variáveis). Vamos assim verificar se a zona de proveniência do aluno está relacionada com a média final dos alunos. Se sim, a zona de proveniência poderia ajudar a explicar, pelo menos em parte, as diferenças observadas entre os alunos de um mesmo curso.

Ao incluir a co-variável do nível 1, o modelo nesse nível ficará

$Y_{ik} = \beta_{ok} + \beta_{1k}x_{ik} + e_{ik}$, com $x_{ik} = X_{ik} - \bar{X}$

No nível 2 o termo, $\beta_{0k} = \gamma_{00} + \gamma_{01}z_k + u_{0k}$ permanece inalterado e o termo $\beta_{1k} = \gamma_{10}$ é igual em todos os cursos, pois apenas se relacionam duas variáveis do nível 1.

O coeficiente γ_{10} representa o declive médio que relaciona a média dos alunos com a zona de proveniência.

Através de substituição, obtemos o modelo combinado

$$Y_{ik} = \gamma_{00} + \gamma_{01}z_k + \gamma_{10}x_{ik} + (u_{0k} + e_{ik})$$

Ao incluir esta nova co-variável, obtemos os resultados das tabelas 10, 11 e 12.

A tabela 10 indica-nos as estimações dos efeitos fixos do modelo:

1) A constante ou ordenada na origem ($\hat{\gamma}_{00} = 13{,}3$), que é uma estimação da média, na população dos cursos
2) O coeficiente associado à variável idade ($\hat{\gamma}_{01} = -0{,}13$), que é sensivelmente igual relativamente ao obtido antes de incluir a co-variável *zona*

[51] Análise de covariância

3) O coeficiente associado à variável *zona* ($\hat{\gamma}_{10} = -0{,}07$), que indica que os alunos da zona 1 (urbana) têm uma média 0,07 valores inferior à dos alunos da zona 0 (rural).

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1,327878E1	,238698	6,974	55,630	,000	12,713921	13,843645
idade_centrada	-,125506	,060065	192,683	-2,089	,038	-,243976	-,007036
zona	-,072640	,186250	195,369	-,390	,697	-,439960	,294679

a. Dependent Variable: Media final.

Tabela 10: Estimação dos efeitos fixos (passo 3)

A tabela 11 dá-nos as estimações dos parâmetros da co-variância. A estimação da variabilidade entre os cursos ($\hat{\sigma}^2_{u_0}$) aumentou um pouco e a variância dos resíduos ($\hat{\sigma}^2_e$) diminuiu em relação ao modelo nulo. A variabilidade intra-curso, dada por

$$R_1^2 = \frac{1{,}600443 - 1{,}577715}{1{,}600443} = 0{,}014$$

Sendo a variabilidade do nível 2 (entre-cursos) dada por:

$$R_2^2 = \frac{1{,}600443 - 0{,}166205}{1{,}600443} = 0{,}896$$

O ρ é agora $\rho = \frac{\hat{\sigma}^2_{u_0}}{\hat{\sigma}^2_{u_0} + \hat{\sigma}^2_e} = \frac{0{,}166205}{0{,}166205 + 1{,}577715} = 0{,}095 \approx 9{,}5\%$

Como podemos observar, este valor aumentou um pouco, pelo que uma parte das diferenças observadas nos cursos está explicada pela zona de proveniência do aluno.

Estimates of Covariance Parameters[a]						
Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Residual	1,577715E0	,160968	9,801	,000	1,291766	1,926962
Intercept [subject = curso]	Variance ,166205	,148232	1,121	,262	,028940	,954545
a. Dependent Variable: Media final.						

Tabela 11: Estimação dos parâmetros de co-variância (passo 3)

Na tabela 12 podemos constatar que a qualidade do modelo aumentou ligeiramente, uma vez que -2LL diminuiu.

Information Criteria[a]	
-2 Restricted Log Likelihood	667,094
Akaike's Information Criterion (AIC)	671,094
Hurvich and Tsai's Criterion (AICC)	671,156
Bozdogan's Criterion (CAIC)	679,650
Schwarz's Bayesian Criterion (BIC)	677,650
The information criteria are displayed in smaller-is-better forms.	
a. Dependent Variable: Media final.	

Tabela 12: Estatísticas de ajuste global (passo 3)

Modelo de ANCOVA: um factor de efeitos aleatórios

$$Y_{ik} = 13{,}3 - 0{,}13 idade\_centrada - 0{,}07 zona + (u_{0k} + e_{ik})$$

4) Análise de regressão: coeficientes aleatórios

Até agora os modelos encontrados são os chamados modelos de constantes ou intersecções aleatórias porque, em todos eles, o único coeficiente que varia aleatoriamente de um curso para outro é a constante de intersecção do nível 1, β_{ok}.

Nestes modelos, o declive, β_{1k}, ou não existe (como é o caso da ANOVA com um factor de efeitos aleatórios e na regressão com médias como resultados) ou toma um valor fixo (como é o caso da ANCOVA de um factor de efeitos aleatórios).

No último modelo apresentado, foi assumida uma relação homogénea[52] em todos os cursos entre a co-variável (*zona*) e a variável dependente (*media_total*). No entanto, para dizer que parte da variabilidade intra-curso (variabilidade de nível 1) pode ser explicada pela zona de residência, ou seja, para avaliar correctamente a relação existente entre a média e a zona de proveniência do aluno, é necessário obter uma equação de regressão para cada curso e analisar como variam as ordenadas na origem e os declives dessas equações. Pois, poderá haver diferenças entre as médias dos cursos (médias diferentes) e, também, a relação entre as médias e a zona pode não ser a mesma em todos os cursos (diferentes declives).

Este novo modelo denomina-se por modelo de coeficientes aleatórios, já que ambos os coeficientes (ordenada na origem e declive) podem variar aleatoriamente de curso para curso.

No nível 1, o modelo é semelhante ao anterior (ANCOVA de um factor aleatório):

$$Y_{ik} = \beta_{ok} + \beta_{1k} x_{ik} + e_{ik}$$

No nível 2, o termo β_{ok} também se define de modo semelhante ao anterior modelo:

$$\beta_{0k} = \gamma_{00} + u_{0k}$$

(Obviamente é possível introduzir uma ou mais co-variáveis de nível 2.)

A diferença este modelo e o anterior está na forma de definir o declive β_{1k}. No modelo anterior (ANCOVA) é interpretado como uma constante (estima-se apenas um declive

[52] $\beta_{1k} = \gamma_{10}$para todo o k

para todos os cursos: $\beta_{1k} = \gamma_{10}$). No modelo de regressão com coeficientes aleatórios interpreta-se como uma variável ($\beta_{1k} = \gamma_{10} + u_{1k}$).

Logo, cada curso terá o seu próprio declive (estimam-se tantos declives como cursos). Substituindo, obtemos o modelo combinado:

$$Y_{ik} = \gamma_{00} + \gamma_{10}x_{ik} + \gamma_{01}w_k + (u_{0k} + u_{1k}x_{ik} + e_{ik})$$

Sendo

γ_{00} a média na população de cursos

γ_{10} o declive médio que relaciona a variável dependente (média) com a co-variável (nível1)

γ_{10} o declive médio que relaciona a variável dependente (média) com a co-variável (nível 2)

u_{0k} é o efeito do k-ésimo curso sobre a ordenada na origem

u_{1k} é o efeito do k-ésimo curso sobre os declives

e_{ik} é o erro do nível 1

Assume-se que e_{ik} se distribuem normalmente com média zero e igual variância (σ_e^2) em todos os cursos e u_{0k} e u_{1k} se distribuem normalmente com valor médio zero e variâncias σ_{u0}^2 e σ_{u1}^2, respectivamente.

Neste caso, incluímos as co-variáveis *sexo*, *zona* e *sucesso* do nível 1 e *anolect*, *idade_centrada* e *totalmodatra* do nível 2. Neste caso, o modelo ficará:

$$Y_{ik} = \gamma_{00} + \gamma_{p0}X_{pik} + \gamma_{0q}W_{qk} + u_{pk} * X_{pik} + e_{ik} + u_{ok}$$

Ao ajustarmos este modelo de regressão com coeficientes aleatórios, obtemos as tabelas 13, 14 e 15.

A tabela 13 dá-nos as estimações dos parâmetros de efeitos fixos que, neste modelo, são: o valor da ordenada na origem $\hat{\gamma}_{00} = 14{,}27$, que indica a média dos alunos na população de cursos, o valor do coeficiente associado às variáveis *sexo* $\hat{\gamma}_{10} = -0{,}379$,

zona $\hat{\gamma}_{20} = -0{,}014$, *sucesso* $\hat{\gamma}_{30} = -0{,}014$, *anolect* $\hat{\gamma}_{01} = -0{,}0002$, *idade_centrada* $\hat{\gamma}_{02} = -0{,}036$ e *totalmodatra* $\hat{\gamma}_{03} = -0{,}041$, que são uma estimação do declive médio. Em cada curso estimou-se uma equação de regressão que relaciona cada variável com a média do aluno. Os valores obtidos são uma estimação da média de todos esses declives. Neste caso, o teste t

$$H_0: \gamma_{0q} = 0 \; ou \; \gamma_{p0} = 0$$

versus

$$H_1: \gamma_{0q} \neq 0 \; ou \; \gamma_{p0} \neq 0$$

dá-nos apenas as variáveis *totalmodatra* e *sucesso* como significativamente diferentes de zero, pois o valor do *p-value* no teste t (0 < 0,05).

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1,426852E1	,486174	20,855	29,349	,000	13,257037	15,280004
anolect	-,000166	,000689	183,047	-,240	,810	-,001526	,001195
idade_centrada	-,036067	,040827	187,148	-,883	,378	-,116607	,044473
sexo	-,378763	,406028	3,734	-,933	,407	-1,538414	,780889
totalmodatra	-,041015	,008849	186,937	-4,635	,000	-,058471	-,023559
sucesso	-1,257357E0	,233132	700,232	-5,393	,000	-1,715077	-,799636
zona	-,114117	,214219	4,358	-,533	,620	-,690114	,461880

a. Dependent Variable: Media final.

Tabela 13: Estimação dos efeitos fixos (passo 4)

A tabela 14 indica-nos as estimações dos quatro parâmetros de co-variância:

- A variância dos erros $\hat{\sigma}_e^2 = 0{,}704$

Esta variância diz-nos em que medida variam os alunos em torno da recta de regressão do respectivo curso. O valor estimado é inferior ao modelo estimado pelo modelo nulo (1,600443). Para conhecer a proporção de variância explicada no nível 1, calculamos

$\frac{1,600443-0,703885}{1,600443} = 0,56$, o que significa que ao incluir as variáveis do nível 1 no modelo de regressão, utilizando uma equação separada para cada curso, a variabilidade intra-curso passa a ser de 56%. Recorde-se que utilizando apenas uma equação de regressão para todos os centros (modelo 3), a variável de nível 1 representava uma variabilidade intra-curso de apenas 1,4%.

- A variância das ordenadas na origem, representada por $UN(1,1) = \hat{\sigma}_{u0}^2 = 0,248$

Pelo valor de *p-value* do teste de Wald, não rejeitamos a hipótese (H_0: $\sigma_{u0}^2 = 0$) de que a variância das ordenadas na origem seja zero, pois o valor crítico é 0,477 > 0,005. Portanto podemos concluir que poderá haver igualdade nas intersecções das rectas de regressão dos diferentes cursos. Por outras palavras, poderá não haver diferenças significativas entre as médias dos cursos.

- A variância dos declives, representada por $UN(2,1) = \hat{\sigma}_{u1}^2 = -0,200$

Através do valor crítico do teste de Wald, também não rejeitamos H_0: $\sigma_{u1}^2 = 0$, pois *p-value* = 0,763 > 0,05. Desta forma, concluímos que os também os declives das equações de regressão poderão ser iguais em todos os cursos. Isto é, poderá não existir diferença na relação entre o sexo e a média final, nos vários cursos.

- A co-variância entre as ordenadas na origem e os declives, representada por *UN(2, 2)* =0,669

Não parece haver relação entre as ordenadas na origem e os declives (*p-value* = 0,257). Assim, a relação intra-curso entre o sexo e a média final não parece aumentar nem diminuir, conforme o que acontece na ordenada na origem.

O mesmo acontece relativamente às restantes variáveis, pois em todos os casos, o valor de p-value > 0,05.

Estimates of Covariance Parameters[b]							
Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		,703885	,071959	9,782	,000	,576079	,860046
Intercept + sexo + sucesso + zona [subject = curso]	UN (1,1)	,248016	,349123	,710	,477	,015713	3,914581
	UN (2,1)	-,199731	,662723	-,301	,763	-1,498644	1,099183
	UN (2,2)	,668758	,589728	1,134	,257	,118757	3,765996
	UN (3,1)	-1,245903E-1[a]	,000000	.	.	.	.
	UN (3,2)	-5,098915E-2[a]	,000000	.	.	.	.
	UN (3,3)	,142988[a]	,000000	.	.	.	.
	UN (4,1)	,175671	,117138	1,500	,134	-,053915	,405257
	UN (4,2)	-1,349235E-1[a]	,000000	.	.	.	.
	UN (4,3)	-,060776	,173441	-,350	,726	-,400713	,279162
	UN (4,4)	,149823	,155234	,965	,334	,019662	1,141614

a. This covariance parameter is redundant. The test statistic and confidence interval cannot be computed.

b. Dependent Variable: Media final.

Tabela 14: Estimação dos parâmetros de co-variância (passo 4)

Information Criteria[a]	
-2 Restricted Log Likelihood	549,447
Akaike's Information Criterion (AIC)	571,447
Hurvich and Tsai's Criterion (AICC)	572,914
Bozdogan's Criterion (CAIC)	618,279
Schwarz's Bayesian Criterion (BIC)	607,279
The information criteria are displayed in smaller-is-better forms.	
a. Dependent Variable: Media final.	

Tabela 15: Estatísticas de ajuste global (passo 4)

Relativamente ao modelo anterior, também o valor de *deviance* decresceu, pelo que a qualidade de ajustamento do modelo aumentou.

Considerando as variáveis significativamente diferentes de zero, obtemos o modelo:

Modelo de análise de regressão: coeficientes aleatórios

$$Y_{ik} = 14{,}27 - 1{,}26 sucesso_{ik} - 0{,}04 totalmodatra_k + (u_{0k} + u_{2k} sucesso_{ik} + e_{ik})$$

5) Análise de regressão: ordenadas na origem e declives como resultados

Depois de chegarmos à conclusão que as médias e os declives variam de curso para curso, o passo seguinte é averiguar que variáveis podem estar relacionadas com esta variabilidade.

A diferença entre este modelo e o anterior é a presença do efeito interacção entre as variáveis do nível 1 e as do nível 2. Miles e Shevlin (2001) descrevem este efeito de interacção como "efeitos diferentes para grupos diferentes". Por exemplo, a interacção $sexo \times totalmodatra$ indica que a influência do sexo do aluno é diferente entre cursos com alunos com mais ou menos módulos em atraso.

Neste caso, vamos ainda assim, fazer a estimação dos parâmetros, utilizando as co-variáveis *sexo, zona, sucesso, anolect, totalmodatra* e *idade_centrada*. No modelo de ordenadas na origem como resultado verificámos que a idade dos alunos explica 90% das diferenças observadas nas médias dos cursos, ou seja, 90% da variabilidade entre as médias. Pretendemos agora verificar que variáveis podem ter influência nesta variabilidade observada entre os declives.

O modelo de regressão que interpreta as médias e os declives como resultados é semelhante ao modelo de coeficientes aleatórios, no nível 1:

$$Y_{ik} = \beta_{ok} + \beta_{1k}x_{ik} + e_{ik}$$

No entanto, no nível 2, inclui as variáveis que se pretendem utilizar para explicar a variabilidade das médias e dos declives:

$$\beta_{ok} = \gamma_{00} + \gamma_{01}z_k + \gamma_{02}w_k + u_{0k}$$

$$\beta_{1k} = \gamma_{10} + \gamma_{11}z_k + \gamma_{12}w_k + u_{1k}$$

Considerando duas variáveis no nível 2: *z* e *w*.

Substituindo, obtemos o modelo combinado:

$$Y_{ik} = \gamma_{00} + \gamma_{01}z_k + \gamma_{02}w_k + \gamma_{10}x_{ik} + \gamma_{11}x_{ik}z_k + \gamma_{12}x_{ik}w_k + (u_{0k} + u_{1k}x_{ik} + e_{ik})$$

Sendo

γ_{00} = média das classificações de todos os cursos

γ_{01} = efeito principal da variável 1 do nível 2.

z_k é a variável explicativa do nível 2correspondente ao k-ésimo curso

γ_{02} = efeito principal da variável 2 do nível 2.

w_k é a variável explicativa do nível 2correspondente k-ésimo curso

γ_{10} = declive médio que relaciona a média das classificações com a variável 1 do nível 1.

x_{ik} é a variável explicativa do nível 1correspondente ao i-ésimo aluno do k-ésimo curso

u_{0k} = efeito do k-ésimo curso sobre as medias (ordenadas na origem).

u_{1k} = efeito do k-ésimo curso sobre os declives.

e_{ik} = erro ou resíduo aleatório do nível 1.

Neste modelo são ainda incluídas duas interacções entre variáveis de diferentes níveis (*sexo*, *zona, sucesso* do nível 1 e *anolect, totalmodatra* e *idade_centrada* do nível 2)

γ_{11} = efeito conjunto das variáveis 1 do nível 1 e 1 do nível 2.

γ_{12} = efeito conjunto das variáveis 1 do nível 1 e 2 do nível 2.

Assume-se que os erros e_{ik} distribuem-se normalmente com média zero e igual variância σ_e^2 em todos os cursos. E u_{0k} e u_{1k} distribuem-se normalmente com valor esperado zero e variâncias σ_{u0}^2 e σ_{u1}^2, respectivamente.

O output do SPSS fornece-nos as tabelas seguintes.

Information Criteria[a]	
-2 Restricted Log Likelihood	580,759
Akaike's Information Criterion (AIC)	602,759
Hurvich and Tsai's Criterion (AICC)	604,303
Bozdogan's Criterion (CAIC)	649,063
Schwarz's Bayesian Criterion (BIC)	638,063
The information criteria are displayed in smaller-is-better forms.	
a. Dependent Variable: Media final.	

Tabela 16: Estatísticas de ajuste global (passo 5)

A tabela 17 fornece-nos a estimativa dos 16 parâmetros dos efeitos fixos: a ordenada na origem, os seis efeitos principais e as nove interacções.

Temos $\hat{\gamma}_{00} = 17{,}49$ que corresponde à média das classificações na população dos cursos. O valor crítico associado ao teste t é 0 < 0,0005, pelo que podemos afirmar que essa média é diferente de zero.

As variáveis cujos coeficientes são significativamente diferentes de zero (p-value < 0,05) são:

Anolect

Sexo

Sucesso

Zona

E as interacções

Anolect × *sexo*

Anolect × *zona*

Tendo em conta os efeitos *sexo*, *sucesso* e *zona*, o *anolect* está em relação negativa $\hat{\gamma}_{01} = -0{,}006$, com a média das classificações. O que significa que, tendo em conta a influência das variáveis do nível 1, por cada ano lectivo que se incrementa, a classificação média dos alunos diminui 0,006 valores.

Considerando a variável *anolect*, obtemos

$\hat{\gamma}_{10} = -3{,}35$ (variável sexo)

$\hat{\gamma}_{20} = -1{,}92$ (variável zona)

$\hat{\gamma}_{30} = -2{,}34$ (variável sucesso)

Todos estes valores são negativos, o que indica que:

- a classificação média dos alunos do sexo feminino (valor 1) é inferior em 3,35 valores à dos alunos do sexo masculino.

- a classificação média dos alunos provenientes do centro urbano (zona 1) é inferior em 1,92 valores relativamente aos alunos da zona rural.

- a classificação dos alunos sem sucesso (valor 1) é inferior em 2,34 valores, relativamente aos alunos com sucesso.

Relativamente às interacções, temos

$$\hat{\gamma}_{11} = 0{,}005 \text{ (Anolect} \times \text{sexo)}$$

$$\hat{\gamma}_{21} = 0{,}003 \text{ (Anolect} \times \text{zona)}$$

Estas interacções têm coeficientes positivos e significativos, pelo que o ano lectivo relaciona-se positivamente tanto com o género como com a zona de proveniência do aluno. Isto é, indica que a relação entre o sexo e a média das classificações é tanto maior quanto mais recente for o ano lectivo dos cursos. Por outro lado, também a relação entre a zona de residência do aluno e a sua classificação média também é tanto maior quanto mais recente for o ano lectivo do curso.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1,748609E1	,956040	87,093	18,290	,000	15,585889	19,386297
anolect	-,005533	,001521	130,860	-3,637	,000	-,008543	-,002523
idade_centrada	-,020934	,083644	173,813	-,250	,803	-,186023	,144155
sexo	-3,354994E0	,932850	222,985	-3,596	,000	-5,193324	-1,516663
totalmodatra	,812393	,834695	183,057	,973	,332	-,834467	2,459254
sucesso	-2,342367E0	,867913	161,539	-2,699	,008	-4,056286	-,628448
zona	-1,921370E0	,866141	63,890	-2,218	,030	-3,651743	-,190997
anolect * sexo	,005138	,001438	168,928	3,573	,000	,002299	,007977
anolect * zona	,002937	,001351	84,771	2,174	,032	,000251	,005624

anolect * sucesso		,001816	,001388	159,607	1,308	,193	-,000926	,004558
idade_centrada * sexo	*	-,063642	,087263	183,675	-,729	,467	-,235809	,108524
idade_centrada * zona	*	,073428	,088992	178,695	,825	,410	-,102183	,249038
idade_centrada * sucesso	*	-,003739	,090730	182,627	-,041	,967	-,182752	,175274
sexo * totalmodatra		-,030700	,018112	30,913	-1,695	,100	-,067643	,006243
totalmodatra * zona		,003627	,015080	159,180	,241	,810	-,026156	,033410
totalmodatra * sucesso	*	-,851327	,834954	183,121	-1,020	,309	-2,498695	,796040
a. Dependent Variable: Media final.								

Tabela 17: Estimação dos efeitos fixos (passo 5)

A tabela 18 dá-nos as estimações dos parâmetros de variância e co-variância: a variância dos erros ou resíduos, $\hat{\sigma}_e^2$, a variância das médias, UN (1, 1) = $\hat{\sigma}_{u0}^2$, a variância dos declives da variável sexo, UN (2, 2) = $\hat{\sigma}_{u1}^2$ e a covariância entre as médias e os declives da variável sexo, UN (2, 1). Da mesma forma, temos as variâncias e co-variâncias para as restantes variáveis.

- $\hat{\sigma}_e^2$ = 0,66 indica-nos em que medida variam os alunos em torno da recta de regressão do respectivo curso. Este valor é inferior ao do modelo anterior, pelo que as interacções contribuíram para reduzir este erro.
- Variância das ordenadas na origem, UN (1, 1) = $\hat{\sigma}_{u0}^2$=0,094 é também um valor inferior ao obtido no modelo anterior.
- Variância dos declives, UN (2, 2) = $\hat{\sigma}_{u1}^2$=0,38, sendo o erro padrão zero.
- Co-variância entre as ordenadas e os declives, UN (2, 1) = 0,009. Também com erro padrão zero.

Estimates of Covariance Parameters[b]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		,662278	,069330	9,553	,000	,539427	,813108
Intercept + sexo + sucesso + zona [subject = curso]	UN (1,1)	,093729	,247596	,379	,705	,000529	16,611699
	UN (2,1)	,008725[a]	,000000	.	.	.	.
	UN	,380207[a]	,000000	.	.	.	.

(2,2)						
UN (3,1)	-6,240061E-2[a]	,000000	.	.	.	.
UN (3,2)	-1,447684E-1[a]	,000000	.	.	.	.
UN (3,3)	,094534[a]	,000000	.	.	.	.
UN (4,1)	,116373	,190507	,611	,541	-,257013	,489759
UN (4,2)	-,066841	,283126	-,236	,813	-,621757	,488076
UN (4,3)	-,059820	,143829	-,416	,677	-,341720	,222081
UN (4,4)	,323355	,300803	1,075	,282	,052221	2,002233

a. This covariance parameter is redundant. The test statistic and confidence interval cannot be computed.

b. Dependent Variable: Media final.

Tabela 18: Estimação dos parâmetros de co-variância (passo 5)

Verificação dos pressupostos: Análise dos resíduos

Os pressupostos de regressão são: os erros são independentes e identicamente distribuídos com distribuição Normal de média zero e variância σ^2. Uma vez que não conhecemos os erros temos que analisar a sua estimativa que é dada pelos resíduos:

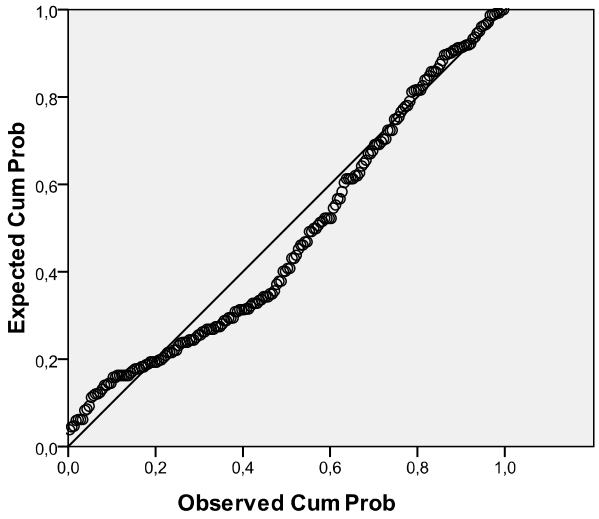


Gráfico 13: P-P Plot (nível 1)

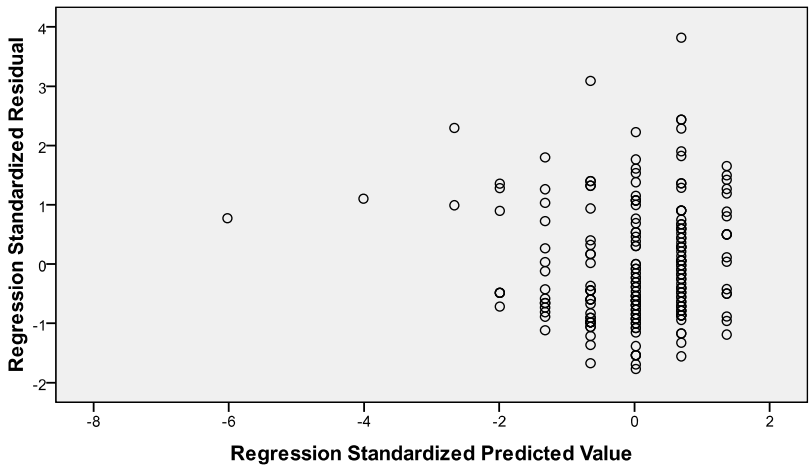


Gráfico 12: Gráfico de dispersão dos resíduos (nível 1)

O PP-plot não nos dá qualquer indicação que contrarie o pressuposto da normalidade dos resíduos. Por outro lado, também não evidencia a existência de outliers.

O gráfico de dispersão dos resíduos em função dos valores preditos estandardizados mostra-se aleatório.

Mostra-se assim que os pressupostos não são violados pelo modelo gerado (Raudenbush, Bryk, 2002).

Obtemos assim o modelo seguinte, considerando as variáveis cujo coeficiente seja significativamente diferente de zero.

Modelo de análise de regressão: ordenadas na origem e declives como resultados

$$Y_{ik} = 17{,}49 - 3{,}35sexo_{ik} - 1{,}92zona_{ik} - 2{,}34sucesso_{ik} + 0{,}005Anolect_k\ sexo_{\ ik} + 0{,}003Anolect_k\ zona_{\ ik} + (u_{0k} + u_{1k}sexo_{ik} + u_{2k}zona_{ik} + u_{3k}sucesso_{ik} + e_{ik})$$

5.2 Modelo que relaciona a idade de entrada com o sucesso escolar

Factor: sucesso

Variável dependente: media_total

Covariavel: idade_centrada

Modelo vazio

Através da tabela 19 concluímos que o número de sucessos e de insucessos é sensivelmente o mesmo. Temos que a média final dos alunos com sucesso é de cerca de 14 valores e dos alunos com insucesso é de 12,4 valores. A média total dos alunos é de 13,2 valores.

Descriptive Statistics				
Media final				
Sucesso	Count	Mean	Standard Deviation	Coefficient of Variation
sim	102	13,9951	1,23444	8,8%
não	103	12,3796	,78894	6,4%
Total	205	13,1834	1,31198	10,0%

Tabela 19: Estatísticas descritivas (modelo sucesso)

Information Criteria[a]	
-2 Restricted Log Likelihood	603,053
Akaike's Information Criterion (AIC)	609,053
Hurvich and Tsai's Criterion (AICC)	609,173
Bozdogan's Criterion (CAIC)	622,008
Schwarz's Bayesian Criterion (BIC)	619,008
The information criteria are displayed in smaller-is-better forms.	
a. Dependent Variable: Media final.	

Tabela 20: Estatísticas de ajuste global (passo 1)

A tabela 21 indica o valor estimado da ordenada na origem, que é o único parâmetro de efeitos fixos no modelo. Esta estimação representa a média populacional dos alunos com e sem sucesso, na variável dependente *media_total*. Temos a estimação $\hat{\mu} = 13{,}2$ e o respectivo erro padrão 0,81 e o *p-value* , para testar a hipótese de que o parâmetro é zero.

$$H_0: \gamma_{00} = 0$$

$$H_1: \gamma_{00} > 0$$

Neste caso, como *p-value* = 0,039 < 0,05, podemos concluir que a ordenada na origem é diferente de zero, com uma probabilidade de erro de 0,05%. Desta forma concluímos que a média da população de alunos é maior que zero.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1,318732E1	,807743	1,000	16,326	,039	2,923973	23,450674

a. Dependent Variable: Media final.

Tabela 21: Estimação dos efeitos fixos (passo 1)

Na tabela 22 temos as estimações dos parâmetros associados aos efeitos aleatórios do modelo. A variância do factor sucesso (1,29) indica quanto varia a variável dependente no factor sucesso. A variância dos resíduos ($\hat{\sigma}_e^2 = 1{,}07$) indica quanto varia a variável dependente no factor sucesso. Segundo estas estimações, a variabilidade do factor sucesso representa $\rho = \frac{1{,}29}{1{,}29+1{,}07} = 0{,}547$, cerca de 55% da variabilidade total. Este quociente denomina-se por coeficiente de correlação intra-classe e representa o grau de variabilidade existente entre os alunos com e sem sucesso.

A tabela 22 dá-nos o *p-value* do teste de Wald para testar a hipótese de que o efeito do factor é nulo.

$$H_0: \sigma_\beta^2 = 0$$

$$H_1: \sigma_\beta^2 > 0$$

Este teste tem um *p-value* de 0,483 > 0,05, pelo que não rejeitamos a hipótese nula, de que a variância populacional do factor sucesso é zero, podendo a média não diferir significativamente entre os alunos com e sem sucesso. No entanto, dado que o teste Wald é muito conservador para amostras pequenas, talvez seja prudente pensarmos que fica por explicar parte das diferenças entre o grupo de alunos com e sem sucesso.

Os parâmetros de co-variância estimaram-se assumindo que o factor sucesso é independente dos resíduos.

Estimates of Covariance Parameters[b]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		1,070910E0	,106297	10,075	,000	,881585	1,300894
Intercept	Variance	0,000000E0[a]	,000000	.	.	.	.
sucesso	Variance	1,294450E0	1,845405	,701	,483	,079175	21,163259

a. This covariance parameter is redundant. The test statistic and confidence interval cannot be computed.

b. Dependent Variable: Media final.

Tabela 22: Estimação dos parâmetros de covariância (passo 1)

Obtemos assim o modelo nulo.

Modelo Nulo

$$Y_{ik} = 13{,}2 + u_{0k} + e_{ik}$$

Análise de regressão: ordenadas na origem como resultados

A comparamos a qualidade de ajustamento nos dois modelos, observamos que houve uma pequena melhoria com a inclusão da co-variável idade. De facto, no modelo nulo obtivemos -2LL = 603,053 e quando incluímos a variável idade, obtivemos -2LL = 591,559. A diferença entre ambos os valores (11,494) segue uma distribuição qui-quadrado com 1 grau de liberdade (os dois modelos apenas diferem de um parâmetro - γ_{01}). Este valor é consideravelmente superior ao valor crítico de 1,96. Daqui podemos concluir que, depois de inserir o efeito da idade, a média não é a mesma, tendo em conta o factor sucesso, isto é, a variância das médias dos dois grupos de alunos é maior que zero.

Information Criteria[a]	
-2 Restricted Log Likelihood	591,559
Akaike's Information Criterion (AIC)	595,559
Hurvich and Tsai's Criterion (AICC)	595,620
Bozdogan's Criterion (CAIC)	604,135
Schwarz's Bayesian Criterion (BIC)	602,135
The information criteria are displayed in smaller-is-better forms.	
a. Dependent Variable: Media final.	

Tabela 23: Estatísticas de ajuste global (passo 2)

Da tabela 24 obtemos o valor da ordenada na origem ($\hat{\gamma} = 13{,}2$) e o coeficiente associado à co-variável *idade_centrada* ($\hat{\gamma}_{01} = -0{,}08$). Sabendo que a co-variável *idade_centrada* é centrada[53], o valor da ordenada na origem é uma estimação da média na população dos dois grupos de alunos. O valor do coeficiente associado à co-variável indica que por cada ano que aumenta a idade média no grupo, a média final dos alunos diminui 0,08 valores. Como este coeficiente tem associado uma estatística t, cujo *p-value* = 0,121 > 0,05, não rejeitamos H_0 de que a idade não influencia a média dos alunos com e sem sucesso.

[53] Variável centrada: $z_k = Z_k - \bar{Z}$

Estimates of Fixed Effects[a]							
Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1,319114E1	,793087	1,000	16,633	,038	3,110864	23,271410
idade_centrada	-,076961	,049415	197,026	-1,557	,121	-,174410	,020489
a. Dependent Variable: Media final.							

Tabela 24: Estimação dos efeitos fixos (passo 2)

Estimates of Covariance Parameters[a]						
Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Residual	1,070620E0	,107874	9,925	,000	,878758	1,304371
Intercept [subject = Variance sucesso]	1,247268E0	1,779164	,701	,483	,076167	20,424661
a. Dependent Variable: Media final.						

Tabela 25: Estimação dos parâmetros de co-variância (passo 2)

Para determinar qual a proporção da variância total que se deve às diferenças entre os cursos, calculemos o coeficiente de correlação intra-classe:

$$\rho = \frac{\hat{\sigma}^2_{u_0}}{\hat{\sigma}^2_{u_0} + \hat{\sigma}^2_e} = \frac{1,25}{1,25+1,07} = 0,539$$

Este valor indica que, ao acrescentar o efeito atribuível à idade média, 54% da variância total (variância da variável dependente) ainda se atribui às diferenças entre as médias dos dois grupos (sucesso/insucesso). Este coeficiente, que agora está condicionado, pois informa o que ocorre nos grupos em relação à sua média quando se acrescenta a variável idade.

No modelo nulo, ρ = 55%, pelo que, neste modelo, diminuiu ligeiramente.

Comparando as estimações dos parâmetros da co-variância do modelo nulo e deste modelo, ficamos a conhecer a proporção de variância explicada no nível 2:

$$R^2_2 = \frac{1,29-1,25}{1,29} = 0,03$$

Logo, apenas cerca de 3% das diferenças observadas nos dois grupos são atribuídas à idade dos alunos.

Obtemos assim o modelo seguinte, tendo em conta que o coeficiente da variável *idade_centrada* não é significativamente diferente de zero:

Modelo de análise de regressão: ordenadas na origem como resultados

$$Y_{ik} = 13{,}2 - 0{,}08 idade\_centrada + (u_{0k} + e_{ik})$$

5.3 O modelo que relaciona o professor com a classificação média do aluno

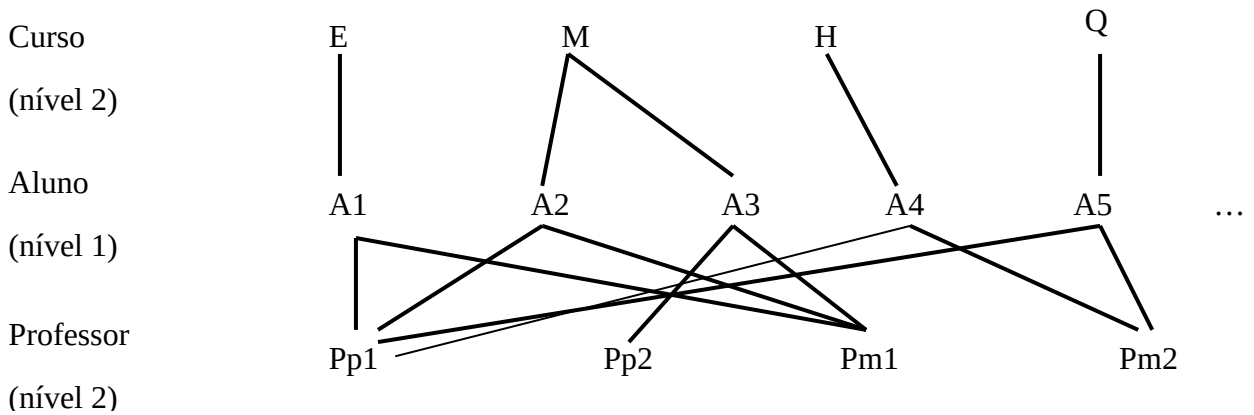
Modelos de classificação cruzada

Para aprofundar um pouco mais o nosso estudo, resolvemos verificar se o professor poderá ter influência na classificação média escolar.

Para tal, usaram-se apenas as turmas dos dois últimos anos lectivos, isto é, 2006 / 2007 e 2007 / 2008 para as disciplinas de Português e Matemática. Foram apenas utilizadas estas disciplinas por serem disciplinas comuns a todos os cursos e por haver diferentes professores a leccioná-las. Obtivemos, assim, duas variáveis dicotómicas, que foram codificadas de 0 e 1 para os dois diferentes professores em cada uma destas disciplinas (Pp1 e Pp2 para a disciplina de Português e Pm1 e Pm2 para a disciplina de Matemática).

Desta forma, obtemos um modelo multi-nível de classificação cruzada no nível 2, pois diferentes professores podem leccionar ao mesmo curso.

O esquema que se segue ilustra este modelo.



Consideramos a variável que inclui os quatro professores considerados (*profs*), que pode tomar três valores distintos: 0, 1 e 2. Toma o valor 0 para a conjugação de professores *mm* e *cl*; o valor 1 para *mm* e *co* e o valor 2 para *it* e *cl*, que são as conjugações existentes.

Modelo de três níveis simples

Comecemos por construir o modelo com três níveis simples, para comparação do modelo cruzado.

Como já foi referido, a notação utilizada para o modelo a três níveis é a seguinte:

$$y_{ijk} = \beta_0 + \beta_1 x_{1ijk} + r_{0k} + u_{0jk} + e_{0ijk}$$

$y_{ijk} = \beta_{0jk} + \sum_{f=1}^{F} \beta_{fjk} x_{fijk} + e_{ijk}$

$\beta_{fjk} = \gamma_{f0k} + \sum_{s=1}^{S} \gamma_{fsk} w_{sjk} + u_{fjk}$, f = 0, …, F

$\gamma_{fsk} = \pi_{fs0} + \sum_{t=1}^{T} \pi_{fst} z_{tk} + r_{fsk}$, f = 0, …, F e s = 0, …, S

Nestas expressões F representa o número de variáveis do primeiro nível, S o número de variáveis do segundo nível e T o número de variáveis do terceiro nível.

y_{ijk} representa a classificação escolar do i-ésimo aluno da j-ésima turma da k-ésima escola.

β_{fjk} são os coeficientes do nível 1

x_{fijk} são as variáveis predictoras do nível 1

e_{ijk}é o efeito aleatório do nível 1. Temos $e_{ijk} \sim N(0, \sigma^2)$

γ_{fsk} são os coeficientes do nível 2

w_{sjk} são as varáveis predictoras do nível 2

u_{fjk} é o efeito aleatório do nível 2. Considerando este efeito como um vector, assume-se que este segue uma distribuição Normal Multivariada com média 0 e matriz de covariância T_γ de dimensão $\sum_{f=0}^{F}(S_f + 1) \times \sum_{f=0}^{F}(S_f + 1)$

π_{fst} são os coeficientes do nível 3

z_{tk} são as variáveis predictoras do nível 3

r_{fsk} é o efeito aleatório do nível 3. Considerando este efeito como um vector, assume-se que este segue uma distribuição Normal Multivariada com média 0 e matriz de co-variância T_β de dimensão $(T+1) \times (T+1)$.

De forma simplificada,

$$r_{0k} \sim N(0, \sigma^2_{r0})$$

$$u_{0jk} \sim N(0, \sigma^2_{u0})$$

$$e_{0ijk} \sim N(0, \sigma^2_{e0})$$

Comecemos pelo modelo nulo, isto é,

$$y_{ijk} = \beta_0 + r_{0k} + u_{0jk} + e_{0ijk}$$

Para acrescentar este terceiro nível no modelo, basta usar mais um sub-comando RANDOM (Leyland, 2004).

```
MIXED media_total BY curso profs
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(5) SCORING(1) SINGULAR(0.000000000001)
HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERG
   E(0.000001, ABSOLUTE)
  /FIXED=| SSTYPE(3)
  /METHOD=ML
  /PRINT=SOLUTION TESTCOV
  /RANDOM=INTERCEPT | SUBJECT(curso) COVTYPE(VC)

  /RANDOM=INTERCEPT | SUBJECT(profs) COVTYPE(VC).
```

Neste caso, o primeiro sub-comando RANDOM indica o SUBJECT (prof). A inclusão de outro sub-comando com o SUBJECT (curso), indica que o outro nível.

Obtemos as tabelas seguintes no output

Information Criteria[a]	
-2 Log Likelihood	422,381
Akaike's Information Criterion (AIC)	430,381
Hurvich and Tsai's Criterion (AICC)	430,696
Bozdogan's Criterion (CAIC)	445,912
Schwarz's Bayesian Criterion (BIC)	441,912
The information criteria are displayed in smaller-is-better forms.	
a. Dependent Variable: Media final.	

Tabela 26: Critérios de selecção (modelo com 3 níveis)

Estimates of Fixed Effects[a]							
Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1,304650E1	,219399	3,357	59,465	,000	12,388484	13,704511
a. Dependent Variable: Media final.							

Tabela 27: Estimação dos efeitos fixos (modelo com 3 níveis)

Estimates of Covariance Parameters[a]							
Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		1,378188E0	,173355	7,950	,000	1,077061	1,763504
Intercept [subject = curso]	Variance	,032988	,080302	,411	,681	,000279	3,894104
Intercept [subject = profs]	Variance	,086355	,108575	,795	,426	,007346	1,015097
a. Dependent Variable: Media final.							

Tabela 28: Estimação dos parâmetros de co-variância (modelo com 3 níveis)

Para testarmos a variância do nível 1,

$$H_0: \sigma_{e0}^2 = 0$$

$$H_1: \sigma_{e0}^2 > 0$$

Como *p-value* = 0, rejeitamos H_0 para qualquer nível de significância. Logo existe variabilidade na classificação média dos vários alunos dentro do mesmo curso e professor.

Por outro lado, ao testarmos a variância de *curso*, sendo

$$H_0: \sigma_{u0}^2 = 0$$

$$H_1: \sigma_{u0}^2 > 0$$

Como o valor crítico é 0,681, também não rejeitamos H_0, pelo que com uma probabilidade de erro de 5%, podemos afirmar que não existem diferenças significativas entre os cursos considerados.

Relativamente ao factor *profs*, ao testarmos as hipóteses

$$H_0: \sigma_{r0}^2 = 0$$

$$H_1: \sigma_{r0}^2 > 0$$

concluímos que, sendo p-value = 0,426 > 0,05, não rejeitamos H_0, pelo que com uma probabilidade de erro de 5%, podemos afirmar que não existem diferenças significativas entre os conjuntos de professores considerados.

A proporção da variância explicada devida a cada nível é dada por:

$\frac{\sigma_e^2}{\sigma_e^2+\sigma_{u_0}^2+\sigma_{r_{00}}^2} = \frac{1{,}38}{1{,}38+0{,}032+0{,}086} = 0{,}92 = 92\%$, para o nível 1

$\frac{\sigma^2_{u_0}}{\sigma^2_e+\sigma^2_{u_0}+\sigma^2_{r_{00}}} = \frac{0,032}{1,38+0,032+0,086} = 0,021 = 2,1\%$, para o nível 2

$\frac{\sigma^2_{r_{00}}}{\sigma^2_e+\sigma^2_{u_0}+\sigma^2_{r_{00}}} = \frac{0,086}{1,38+0,032+0,086} = 0,057 = 5,7\%$, para o nível 3

Desta forma, a variabilidade total da variável independente é quase na sua totalidade devida ao nível 1 (aluno).

Obtemos

$$y_{ijk} = 13,05 + r_{0k} + u_{0jk} + e_{0ijk}$$

Acrescentando as co-variáveis *idade_centrada, prof_mat, prof_port, sexo, zona* e *totalmodatra,* obtemos o output

A qualidade do teste melhorou (o valor de *deviance* diminuiu).

Information Criteria[a]	
-2 Log Likelihood	328,914
Akaike's Information Criterion (AIC)	388,914
Hurvich and Tsai's Criterion (AICC)	407,702
Bozdogan's Criterion (CAIC)	504,940
Schwarz's Bayesian Criterion (BIC)	474,940
The information criteria are displayed in smaller-is-better forms.	
a. Dependent Variable: Media final.	

Tabela 29: Critérios de selecção (modelo com 3 níveis, incluindo variáveis)

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1,423927E1	,374237	5,179	38,049	,000	13,287161	15,191374
sexo	1,834000E0	2,055421	122,340	,892	,374	-2,234797	5,902798
zona	-,444067	,274440	123,073	-1,618	,108	-,987301	,099166
totalmodatra	-,104502	,023393	123,980	-4,467	,000	-,150802	-,058201
idade_centrada	-,262237	,101715	122,337	-2,578	,011	-,463587	-,060886
Prof_mat	-,804307	,532148	64,216	-1,511	,136	-1,867327	,258713
Prof_port	-1,471260E0	,459821	129,932	-3,200	,002	-2,380965	-,561554
sexo * totalmodatra	-,154303	,150459	122,583	-1,026	,307	-,452137	,143530
sexo * idade_centrada	1,367895E0	1,752351	122,428	,781	,437	-2,100938	4,836727
sexo * Prof_mat	-1,240296E0	2,087445	122,721	-,594	,553	-5,372359	2,891767
sexo * Prof_port	-1,533299E0	2,127813	123,133	-,721	,473	-5,745128	2,678531
zona * totalmodatra	-,009094	,033265	122,598	-,273	,785	-,074943	,056755
zona * idade_centrada	,369554	,139347	124,010	2,652	,009	,093747	,645360
zona * Prof_mat	,471628	,484998	123,471	,972	,333	-,488358	1,431615
zona * Prof_port	,582168	,499338	124,211	1,166	,246	-,406145	1,570480
totalmodatra * Prof_mat	-,026480	,059372	124,993	-,446	,656	-,143985	,091025
totalmodatra * Prof_port	,056230	,049580	125,751	1,134	,259	-,041889	,154349
idade_centrada * Prof_mat	,735004	,259990	122,686	2,827	,005	,220356	1,249651
idade_centrada * Prof_port	,196179	,341166	128,218	,575	,566	-,478864	,871223
sexo * totalmodatra * Prof_mat	,083445	,154016	122,570	,542	,589	-,221431	,388320
sexo * totalmodatra * Prof_port	,153566	,166860	122,532	,920	,359	-,176737	,483868
sexo * idade_centrada * Prof_mat	-1,489575E0	1,762123	122,429	-,845	,400	-4,977750	1,998600
sexo * idade_centrada * Prof_port	-1,526431E0	1,788740	122,627	-,853	,395	-5,067238	2,014376

zona * totalmodatra * Prof_mat	,089100	,062186	122,753	1,433	,154	-,033997	,212196
zona * totalmodatra * Prof_port	,023329	,057276	127,689	,407	,684	-,090004	,136662
zona * idade_centrada * Prof_mat	-,843959	,272809	123,889	-3,094	,002	-1,383929	-,303989
zona * idade_centrada * Prof_port	-,220374	,352880	125,169	-,625	,533	-,918758	,478010

a. Dependent Variable: Media final.

Tabela 30: Estimação dos parâmetros dos efeitos fixos (modelo com 3 níveis, incluindo variáveis)

Considerando as variáveis e interacções estatisticamente significativas, obtemos o modelo seguinte:

$$y_{ijk} = 14{,}24 - 0{,}26 idade\_centrada_j - 1{,}47 prof\_port_k - 0{,}11 totalmodatra_j + 0{,}74 idade\_centrada * Prof\_mat_{jk} + 0{,}37 idade\_centrada * zona_{ij} - 0{,}844 idade\_centrada * Prof\_mat * zona_{ijk} + r_{0k} + u_{0jk} + e_{0ijk}$$

O valor

- 14,24 indica-nos a média geral das classificações.
- -0,26 significa que por cada ano na idade do aluno, a sua média diminui 0,26 valores.
- -1,47 indica que a classificação média dos alunos que têm o professor *it* (*prof_port* = 1) é 1,47 valores inferior à dos alunos que têm o professor *mm* (*prof_port* = 0).
- -0,11 diz que por cada módulo em atraso, a classificação media do aluno diminui 0,11 valores.
- 0,74 significa que a relação entre a idade e a classificação média é significativamente maior no professor *co* (*prof_mat* = 1) do que no professor *cl* (*prof_mat* = 0).
- 0,37 indica que a relação entre a idade e a classificação média é significativamente maior nos alunos que provêm de centro urbano (*zona* = 1) do que rural (*zona* = 0).

- -0,844 significa que a relação entre a idade e a classificação média é significativamente menor nos alunos de centro urbano (*zona* = 1) do que rural (*zona* = 0) e menor nos alunos que têm o professor *co* (*prof_mat* = 1) do que os que têm o professor *cl* (*prof_mat* = 0).

Estimates of Covariance Parameters[b]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		,678384	,086738	7,821	,000	,528008	,871587
Intercept [subject = curso]	Variance	,270631	,289068	,936	,349	,033357	2,195671
Intercept [subject = profs]	Variance	0,000000E0[a]	,000000	.	.	.	.

a. This covariance parameter is redundant. The test statistic and confidence interval cannot be computed.

b. Dependent Variable: Media final.

Tabela 31: estimação dos parâmetros de co-variância (modelo com 3 níveis, incluindo variáveis)

Tal como no modelo nulo, apenas a variância dos resíduos do nível 1 são estatisticamente diferentes de zero (*p-value* = 0). Comparando com o modelo nulo, concluímos que o valor da variância dos resíduos do nível 1 diminuiu bastante, logo a introdução das co-variáveis diminuiu a variabilidade deste nível. No nível 2 houve um aumento da variância. No nível 3, a introdução das co-variáveis diminuiu ao mínimo a variabilidade entre professores.

Modelo com dois níveis de classificação cruzada

Para adaptarmos os modelos de classificação cruzada no SPSS, através do comando MIXED basta acrescentar outro sub-comando RANDOM e declarar uma hierarquia (não aninhada) adicional (Leyland, 2004).

Como já vimos, a notação utilizada para os modelos de classificação cruzada é

$$y_{i(k_1,k_2)} = X_{i(k_1,k_2)}\beta + u^{(1)}_{0_{k_1}} + u^{(2)}_{0_{k_2}} + e_{0i(k_1,k_2)}$$

Onde

$y_{i(k_1,k_2)}$ é a variável resposta para a unidade i do nível 1 e k_1 e k_2 do nível 2.

$X_{i(k_1,k_2)}$ é a i-ésima variável explicativa do nível 1 correspondente aos dois factores do nível 2.

β é o coeficiente da variável $X_{i(k_1,k_2)}$

$u^{(1)}_{0_{k_1}}$ é o erro aleatório do nível 2 correspondente ao factor k_1

$u^{(2)}_{0_{k_2}}$ é o erro aleatório do nível 2 correspondente ao factor k_2

$e_{0i(k_1,k_2)}$ é o erro padrão do nível 1

Mais especificamente, vamos considerar a notação seguinte para um modelo de classificação cruzada com dois níveis:

$$y_{i(Prof,curso)} = \beta_{0i}X_{i(Prof,curso)} + u^{(2)}_{0_{Prof(i)}} + u^{(1)}_{0_{curso(i)}} + e_{0i}$$

$$\left[u^{(2)}_{0_{Prof(i)}}\right] \sim N(0, {\sigma^2}^{(2)}_u)$$

$$\left[u^{(1)}_{0_{curso(i)}}\right] \sim N(0, {\sigma^2}^{(1)}_u)$$

$$[e_{oi}] \sim N(0, \sigma^2_e)$$

Modelo Nulo

```
MIXED media_total BY curso profs
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(5) SCORING(1) SINGULAR(0.000000000001)
HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERG
   E(0.000001, ABSOLUTE)
  /FIXED=| SSTYPE(3)
  /METHOD=ML
  /PRINT=SOLUTION TESTCOV
  /RANDOM=INTERCEPT | SUBJECT(curso) COVTYPE(UN)
  /RANDOM=INTERCEPT | SUBJECT(profs) COVTYPE(VC)

  /RANDOM=INTERCEPT | SUBJECT(curso*profs) COVTYPE(VC).
```

Information Criteria[a]	
-2 Log Likelihood	422,381
Akaike's Information Criterion (AIC)	432,381
Hurvich and Tsai's Criterion (AICC)	432,857
Bozdogan's Criterion (CAIC)	451,795
Schwarz's Bayesian Criterion (BIC)	446,795
The information criteria are displayed in smaller-is-better forms.	
a. Dependent Variable: Media final.	

Tabela 32: Critérios de selecção (modelo de classificação cruzada)

Estimates of Fixed Effects[a]							
Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1,304650E1	,219399	3,357	59,465	,000	12,388484	13,704511
a. Dependent Variable: Media final.							

Tabela 33: Estimação dos coeficientes dos efeitos fixos (modelo de classificação cruzada)

Como parâmetro de efeito fixo, temos a média geral das classificações, $\beta_{0i} = 13{,}05$.

Estimates of Covariance Parameters[b]							
Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		1,378188E0	,173355	7,950	,000	1,077061	1,763504
Intercept [subject = curso]	Variance	,032988	,080302	,411	,681	,000279	3,894105
Intercept [subject = profs]	Variance	,086355	,108575	,795	,426	,007346	1,015098
Intercept [subject = curso * profs]	Variance	0,000000E0[a]	,000000	.	.	.	.

a. This covariance parameter is redundant. The test statistic and confidence interval cannot be computed.

b. Dependent Variable: Media final.

Tabela 34: Estimação dos parâmetros de co-variância (modelo de classificação cruzada)

Como podemos constatar, não existe o efeito marginal de professor a pode variar de acordo com o curso que o aluno frequenta, pois o valor de $u_{0_{(Prof,curso)}}^{(3)}$, correspondente aos efeitos de interacção, é nula.

Para testarmos a variância do nível 1,

$$H_0: \sigma_{e0}^2 = 0$$

$$H_1: \sigma_{e0}^2 > 0$$

Como *p-value* = 0, rejeitamos H_0 para qualquer nível de significância. Logo existe variabilidade na classificação média dos vários alunos dentro do mesmo curso e professor.

A variância do nível de classificação cruzada é agora dada por $\left(\sigma_{u_0^{(1)}}^2 + \sigma_{u_0^{(2)}}^2\right)$. Neste caso temos 0,032988 + 0,086355 = 0,119343.

$$H_0: \sigma_{u_0^{(1)}}^2 + \sigma_{u_0^{(2)}}^2 = 0$$

$$H_1: \sigma_{u_0^{(1)}}^2 + \sigma_{u_0^{(2)}}^2 > 0$$

Como o valor crítico tanto para $\sigma^2_{u_0^{(1)}}$ como para $\sigma^2_{u_0^{(2)}}$ é superior a 0,05, logo estes valores não são significativamente diferentes de zero, isto é, não rejeitamos H_0, com uma probabilidade de erro de 5%.

A proporção da variância explicada devida a cada nível é dada por:

$\frac{\sigma^2_{e0}}{\sigma^2_{e0}+\sigma^2_{u_0^{(1)}}+\sigma^2_{u_0^{(2)}}} = \frac{1{,}38}{1{,}38+0{,}1193} = 0{,}92 = 92\%$, para o nível 1

$\frac{\sigma^2_{u_0^{(1)}}+\sigma^2_{u_0^{(2)}}}{\sigma^2_{e0}+\sigma^2_{u_0^{(1)}}+\sigma^2_{u_0^{(2)}}} = \frac{0{,}1193}{1{,}38+0{,}1193} = 0{,}08 = 8\%$, para o nível de classificação cruzada

Desta forma, a variabilidade total da variável independente é quase na sua totalidade devida ao nível 1 (aluno), tal como aconteceu no modelo com três níveis simples. Obtemos

$$y_{ijk} = 13{,}05 + u^{(2)}_{0_{Prof(i)}} + u^{(1)}_{0_{curso(i)}} + e_{0i}$$

Modelo de classificação cruzada com as variáveis *idade_centrada, prof_mat, prof_port, sexo, zona* e *totalmodatra*

Information Criteria[a]	
-2 Log Likelihood	329,507
Akaike's Information Criterion (AIC)	391,507
Hurvich and Tsai's Criterion (AICC)	411,752
Bozdogan's Criterion (CAIC)	511,401
Schwarz's Bayesian Criterion (BIC)	480,401
The information criteria are displayed in smaller-is-better forms.	
a. Dependent Variable: Media final.	

Tabela 35: Critérios de selecção (modelo de classificação cruzada, incluindo variáveis)

Relativamente ao modelo nulo, a qualidade deste modelo melhorou (o valor do *deviance* passou de 422,381 para 329,507).

Estimates of Fixed Effects[a]							
Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1,399175E1	,325980	7,337	42,922	,000	13,228061	14,755447
Prof_port	-1,325582E0	,573209	16,370	-2,313	,034	-2,538505	-,112659
Prof_mat	-,357147	,565660	15,451	-,631	,537	-1,559764	,845469
idade_centrada	-,262234	,102264	122,102	-2,564	,012	-,464675	-,059794
totalmodatra	-,104662	,023495	124,058	-4,455	,000	-,151164	-,058159
zona	-,445305	,275797	122,981	-1,615	,109	-,991230	,100619
sexo	1,833418E0	2,066506	122,105	,887	,377	-2,257402	5,924238
sexo * totalmodatra	-,153914	,151248	122,396	-1,018	,311	-,453314	,145487
sexo * idade_centrada	1,365137E0	1,761710	122,211	,775	,440	-2,122283	4,852556
sexo * Prof_mat	-1,191196E0	2,098208	122,525	-,568	,571	-5,344629	2,962238
sexo * Prof_port	-1,671140E0	2,138800	122,884	-,781	,436	-5,904803	2,562523

zona * totalmodatra	-,009005	,033440	122,414	-,269	,788	-,075200	,057189
zona * idade_centrada	,370515	,139953	124,094	2,647	,009	,093512	,647518
zona * Prof_mat	,491041	,487139	123,444	1,008	,315	-,473187	1,455268
zona * Prof_port	,530608	,501279	124,330	1,059	,292	-,461537	1,522752
totalmodatra * Prof_mat	-,022648	,059563	125,140	-,380	,704	-,140529	,095234
totalmodatra * Prof_port	,049105	,049710	126,073	,988	,325	-,049269	,147480
idade_centrada * Prof_mat	,740731	,261341	122,480	2,834	,005	,223400	1,258062
idade_centrada * Prof_port	,129402	,342493	127,820	,378	,706	-,548288	,807093
sexo * totalmodatra * Prof_mat	,082854	,154824	122,384	,535	,594	-,223625	,389334
sexo * idade_centrada * Prof_mat	-1,488315E0	1,771540	122,210	-,840	,402	-4,995195	2,018565
sexo * totalmodatra * Prof_port	,156458	,167711	122,413	,933	,353	-,175531	,488447
sexo * idade_centrada * Prof_port	-1,477439E0	1,798503	122,331	-,821	,413	-5,037658	2,082780
zona * totalmodatra * Prof_mat	,087595	,062500	122,593	1,402	,164	-,036124	,211315
zona * totalmodatra * Prof_port	,034063	,057442	127,607	,593	,554	-,079598	,147725
zona * idade_centrada * Prof_mat	-,856480	,273894	123,965	-3,127	,002	-1,398594	-,314366
zona * idade_centrada * Prof_port	-,174722	,354031	125,367	-,494	,623	-,875374	,525929

a. Dependent Variable: Media final.

Tabela 36: Estimação dos coeficientes dos efeitos fixos (modelo de classificação cruzada, incluindo variáveis)

Considerando as co-variáveis e interacções significativas (p-value < 0,05), obtemos o modelo:

$$y_{i(Prof,curso)} = 13{,}99 - 0{,}26 idade\_centrada_j - 1{,}33 prof\_port_k - 0{,}105 totalmodatra_j + 0{,}741 idade\_centrada * Prof\_mat_{jk} + 0{,}371 idade\_centrada * zona_{ij} - 0{,}856 idade\_centrada * Prof\_mat * zona_{ijk} + u^{(2)}_{0_{Prof(i)}} + u^{(1)}_{0_{curso(i)}} + e_{0i}$$

Não se detectam diferenças significativas entre este modelo e o modelo com três níveis.

Estimates of Covariance Parameters[b]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		,685724	,087761	7,814	,000	,533592	,881230
Intercept [subject = curso]	Variance	0,000000E0[a]	,000000	.	.	.	.
Intercept [subject = profs]	Variance	0,000000E0[a]	,000000	.	.	.	.
Intercept [subject = curso * profs]	Variance	,131464	,110305	1,192	,233	,025387	,680785

a. This covariance parameter is redundant. The test statistic and confidence interval cannot be computed.

b. Dependent Variable: Media final.

Tabela 37: Estimação dos parâmetros de co-variância (modelo de classificação cruzada, incluindo variáveis)

Podemos observar a estimação da variância dos resíduos ($\hat{\sigma}_e^2 = 0{,}686$), que diminuiu relativamente ao modelo nulo, pelo que concluímos que a presença destas co-variáveis diminuiu um pouco a variabilidade do nível 1. No nível de classificação cruzada, o valor da variância é agora 0,131464, correspondente apenas à variância da interacção. Este valor aumentou ligeiramente, pelo que a introdução das co-variáveis teve uma pequena influência na variabilidade deste nível, além desta fraca variação, o valor desta variância não é significativo. Em suma, a introdução destas co-variáveis praticamente não teve influência na variabilidade do nível de classificação cruzada, tendo esta variabilidade diminuído um pouco no nível de classificação cruzada.

6 Conclusões

Na sua vertente de aplicação, com este trabalho procurámos identificar a importância que o curso, idade, género sexual, sucesso, zona de proveniência e o número de módulos em atraso têm na classificação média dos alunos. Para tal, utilizámos um modelo multi-nível com três níveis. Foram obtidos dois modelos significativos, sendo que, segundo os resultados obtidos, o modelo seguinte,

Modelo de análise de regressão: coeficientes aleatórios

$$Y_{ik} = 14{,}27 - 1{,}26 sucesso_{ik} - 0{,}04 totalmodatra_k + (u_{0k} + u_{2k} sucesso_{ik} + e_{ik})$$

ajusta-se melhor aos dados, sendo preferencial comparativamente com o modelo de análise regressão: ordenadas na origem e declives como resultados.

Modelo de análise de regressão: ordenadas na origem e declives como resultados

$$\begin{aligned} Y_{ik} = {} & 17{,}49 - 3{,}35 sexo_{ik} - 1{,}92 zona_{ik} - 2{,}34 sucesso_{ik} + 0{,}005 Anolect_k \; sexo_{\;ik} \\ & + 0{,}003 Anolect_k \; zona_{\;ik} \\ & + (u_{0k} + u_{1k} sexo_{ik} + u_{2k} zona_{ik} + u_{3k} sucesso_{ik} + e_{ik}) \end{aligned}$$

No primeiro modelo apresentado, concluímos que as variáveis sucesso e total de módulos em atraso têm uma influência significativa na classificação média do aluno.

Porém, ao incluirmos as interacções das variáveis, concluímos que as variáveis que influenciam significativamente a classificação média do aluno são o género, a zona de proveniência, o facto de ter obtido ou não sucesso no curso e as interacções (ano lectivo × sexo) e (ano lectivo × zona de proveniência).

Estudámos ainda se a idade de entrada no curso teria alguma influência no sucesso do aluno. Ao nível de significância de 5% verificou-se que a idade não influencia significativamente o sucesso do aluno.

Ao incluirmos o factor professor no nosso estudo, decidimos comparar as diferenças entre o modelo multi-nível com três níveis e o modelo multi-nível de classificação cruzada no nível 2. Não foram detectadas diferenças entre um e o outro modelo.

O modelo com três níveis obtido foi

$$y_{ijk} = 14{,}24 - 0{,}26idade\_centrada_j - 1{,}47prof\_port_k - 0{,}11totalmodatra_j + 0{,}74idade\_centrada * Prof\_mat_{jk} + 0{,}37idade\_centrada * zona_{ij} - 0{,}844idade\_centrada * Prof\_mat * zona_{ijk} + r_{0k} + u_{0jk} + e_{0ijk}$$

Sendo o modelo de classificação cruzada o seguinte

$$y_{i(Prof,curso)} = 13{,}99 - 0{,}26idade\_centrada_j - 1{,}33prof\_port_k - 0{,}105totalmodatra_j + 0{,}741idade\_centrada * Prof\_mat_{jk} + 0{,}371idade\_centrada * zona_{ij} - 0{,}856idade\_centrada * Prof\_mat * zona_{ijk} + u^{(2)}_{0_{Prof(i)}} + u^{(1)}_{0_{curso(i)}} + e_{0i}$$

Em ambos os modelos concluímos que as variáveis influentes na classificação média do aluno foram a idade, o professor de português e o número de módulos em atraso. Por outro lado, concluímos que as interacções significativas são (idade × professor de Matemática), (idade ×zona de proveniência) e (idade × professor de Matemática × zona de residência).

Na interacção (idade × professor de Matemática) temos um coeficiente positivo, pelo que quanto mais velho for o aluno, maior é a relação entre a sua classificação média e o professor de Matemática. Também na interacção (idade ×zona de proveniência) temos uma relação positiva entre a idade e a zona de proveniência do aluno, isto é, quanto menos jovem for o aluno maior é a relação entre a sua classificação média e a zona de residência. Já na interacção (idade × professor de Matemática × zona de residência), temos um coeficiente negativo, pelo que quanto mais jovem for o aluno maior é a relação da sua classificação média com a conjugação professor de Matemática e zona de residência.

É de notar que os resultados apresentados possuem importantes limitações. A amostra não é probabilística, e, portanto, a análise é frágil em termos de validade externa.

Procuraremos contornar estas limitações em trabalhos futuros, nomeadamente com ampliação das amostras em estudo e comparação com outras escolas com as mesmas características gerais.

7 Perspectivas para o futuro

Um dos principais desafios que emergem a partir das conclusões deste trabalho, consiste em verificar que influência é que os restantes professores poderão ter na variável dependente classificação média, quer da classificação média geral quer da classificação média da disciplina. Este novo tipo de estudo terá que recorrer a modelos multi-nível de classificação cruzada.

Por outro lado, poderemos utilizar dados de outras escolas profissionais e centros de formação, nos quais existam os mesmos cursos que na ETLA, de forma que seja possível fazer uma análise comparativa entre escolas. Neste caso poderemos recorrer a uma análise multi-nível com 3 níveis, ou de classificação cruzadas, caso haja alunos transferidos entre os estabelecimentos de ensino em estudo.

Pretendemos ainda em trabalhos futuros alargar os horizontes de aplicação dos modelos multi-nível a outras áreas da ciência, nomeadamente às Ciências da Saúde. Ambicionamos explorar os softwares disponíveis para tratamentos de certos casos, como por exemplo modelos multi-nível de classificação cruzada, e desenvolver *packages* adequados a casos especiais recorrendo à linguagem R.

8 Referências bibliográficas.

Aguerre, T. F. (2003): "Métodos Estadísticos de Estimación de los Efectos de la Aplicación al Estudio de las Escuelas Eficaces". REICE – Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, Vol. I, nº 2. Disponível em: http://www.ice.deusto.es/RINACE/reice/vol1n2/Tabare.pdf

Aitkin, M.; Longford, N. (1986). "Statistical modeling issues in school effectiveness studies". Journal of the Royal Statistical Society.

Akaike H. (1973) "Information theory and an extension of the maximum likelihood principle". In: Petrov BN, Csaki F, eds. *2nd International Symposium on Information Theory*. Budapest: Akademiai Kiado, 267-281.

Albright, J.; Marinova, D. M. (2010). "Estimating Multilevel Models using SPSS, Stata, SAS and R"

Bauer D. J. (2004). "Mixed Models and Hierarchical Data. Summer Programme in Data Analysis". SPIDA 2005, June 8-9. Institute for Social Research, York University, Toronto, Canada.
Disponível em: http://www.yorku.ca/isr/spida2005/courses.html

Bennet, N. (1976). "Teaching styles and pupil progress". Open Books, London.

Bergamo, G. C. (2002). "Aplicação de modelos multinível na análise de dados de medidas repetidas no tempo". Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz, Tese de Mestrado.

Bickel, R. 2007. "*Multilevel Analysis for Applied Research: It's Just Regression*". Guilford Press. Disponível em http://books.google.pt/books?id=TmkXmmytO9kC&printsec=frontcover&dq=Bickel,+2007+multilevel&source=bl&ots=Nmgu46BlRh&sig=1ivIb8jMwQfPR6FUI80ly9uj79o&hl=pt-PT&ei=56K0TJ-

mKNS5jAfOoIi2Aw&sa=X&oi=book_result&ct=result&resnum=1&ved=0CBUQ6AEwAA#v=onepage&q=Bickel%2C%202007%20multilevel&f=false

BOURDIEU, P.(1977). "Cultural Reproduction and Social Reproduction". In: Power and Ideology in Education, edited by Jerome Karabel and A. H. Halsey, New York: Oxford University Press,

Bozdogan, H. (1987). "Model selection and Akaike's selection criterion (AIC): The general theory and its analytical extensions". Psychometrika, 52, 345-370.

Bryk, A. e Raudenbush, S. (1992). "Hierarchical Linear Models: applications and data analysis methods". London: Sage Publications.

Bryman, A. & Cramer, D. (2003). "Análise de Dados em Ciências Sociais – introdução às técnicas utilizando o SPSS para Windows". 3ª Edição. Oeiras, Celta Editora.

BRYK, A.S.; RAUDENBUSH, S.W. (1992). "Hierarchical linear models". Chicago: Sage Publications.

Catalán-Reyes; Billardón, G. (2003) "Utilización de modelo multinivel en investigación sanitária". Gaceta Sanitaria, 2003; 17 (supl 3): 35-52

Colbourn, C. and Dinitz, J. (2006). "The CRC Handbook of Combinatorial Designs"

COLEMAN, J. S. (1988) "Social capital in the creation of human capital". American Journal of Sociology, v.94, 95-120.

CRUZ, F. (2008), "Modelos Multinível", Revista per. Epidemiol., Vol. 12 nº 3.

Delaunay, D (2003a). « Présentation générale de l'analyse multiniveau. Atelier de formation doctorale à la recherche en démographie » (INED, IRD, Univ. Paris 1, Univ. Paris 5, Univ. Paris 10). Disponível em : http://ceped.cirad.fr/formation/formation%20reciproque/presentationgene.pdf

Delaunay, D (2003b). « Analyse des données d'observatoire: Séries chronologiques,

données de panel, analyse multiniveau ». Atelier de formation réciproque à l'analyse des données d'observatoire, 1-4 Décembre 2003, Nogent-sur-Marne. Disponível em: http://ceped.cirad.fr/formation/formation%20reciproque/Series%20chronologiques%20et%20panel_version%20en%20ligne.pdf

De Leeuw, J.; Meijer, E. (2008). "Handbook of multilevel analysis". New York, NY: Springer Science +Business Media.

Diggle, P.J., Liang, K.-Y. and Zeger, S.L. (1994). Analysis of Longitudinal Data, Oxford: Oxford Science

Ferrão, M.E., Beltrão, K.I.; Santos, D.P. (2002). "Modelo de regressão multinível: Aplicação ao estudo do impacto da política de não-repetência no desempenho escolar dos alunos da 4ª série."

Ferrão, M. E.; Fernandes, C. (2003). "A Contribuição da Escola no Desempenho Escolar do Aluno: Evidências do SAEB". Revista Brasileira de Economia, Artigo submetido.

Ferrão, M.E. (2003). "Introdução aos modelos de regressão multi-nível em educação". São Paulo: Komedi,.

Fielding, A. (2002). "Ordered category responses and random effects in multilevel and other complex structures: scored and generalised models". *Multilevel Modelling: Methodological Advances, Issues and Applications*. N. Duan and S.Reise. New York, Erlbaum.

Fielding, A. (2004). "The role of the Hausman test and whether higher level effects should be treated as fixed or random". Centre for Multilevel Modelling, Institute of Education, University of London, *Multilevel Modelling Newsletter,* **16**, 2, 3-9.

Fox, J.-P.; Glas, C.A.W. (2002). "Modeling measurement error in a structural multilevel model". In G.A. Marcoulides & I. Moustaki (Eds.), Latent Variable and Latent Structure Models (pp. 245-269), London: Lawrence Erlbaum Associates, Publishers

Gelman, A.;J. Hill. (2007). "*Data Analysis Using Regression and Multilevel/Hierarchical Models*". Cambridge University Press: New York.

Goldstein, H. (1986): «Multilevel mixed linear model analysis using iterative generalized least square». Biometrika, **73**, 43-56.

Goldstein, H. (1991a*):"*Non linear multilevel models, with an application to discrete response data*"*. *Biometrika*, **78**, 45-51.

Goldstein, H. (1991b). "Computational Algorithms for Random Cross Classifications".

Goldstein, H. (1995). "Multilevel Statistical Models". 2nd Edition. London: Edward Arnold; New York: Wiley. Disponível em: http://www.mlwin.com/hgpersonal/index.html

Goldstein, H. and Browne, W.J. (2002). Multilevel factor analysis modeling using Markov Chain Monte Carlo estimation. In *Latent Variable and Latent Structure Models* edited by G. Marcoulides and I. Moustaki, London, Lawrence Erlbaum, 225-44.

Goldstein, H. (2003). "Multilevel Statistical Models". 3rd Edition. London, Edward Arnold: New York, Wiley. Parcialmente disponível em: http://www.ioe.ac.uk/hgpersonal/multmodels-edition3/multilevel_statistical_models-third_edition.htm

Goldstein, H. and Fielding, A. (2006). "Cross-classified and Multiple Membership Structures in Multilevel Models: An Introduction and Review", University of Birmingham, ISBN 1 84478 797 2

Gómez, M. H., (2009). "Gasto en farmácia y médico de atención primaria. Um enfoque multinivel", Revista estadística española, vol. 51, núm. 171, 331 - 361.

Gutierrez, G. C. (2005), "Estimação das escalas dos construtos capital social, capital cultural e capital econômico e análise do efeito escola nos dados do Peru-PISA 2000". Tese de mestrado. Disponível em http://www2.dbd.puc-rio.br/pergamum/tesesabertas/0321214_05_cap_07.pdf

Hartley, H. O.; Rao, C. R. (1967). “Maximum-likelihood estimation for the mixed analysis of variance model”. Biometrika, Oxford, v. 54.

Heck, R. H.; Thomas, S.L. (2000). “An introduction to multilevel modeling techniques”. Mahwah: Lawrence Erlbaumm Associates, Publishers.

Henderson, CR (1984). "ANOVA, MIVQUE, REML, and ML Algorithms for Estimation of Variances and Covariances". in David, H. A. and David, H. T.. *Statistics: An Appraisal*. Iowa State University.

Hill, P. W. and Goldstein, H. (1998). “Multilevel modelling of educational data with cross classification and missing identification of units”. *Journal of Educational and Behavioral statistics* **23**: 117-128.

Hill, M. M. & Hill, A. (2002). Investigação por Questionário. Lisboa, Edições Sílabo

Hocking G., 1995. “Supercritical with drawal from a two-layer fluid through a line sink”, *J. Fluid Mech.*, 297

Hox, J. (1998). “Multilevel modeling: When and why”. New York: Springer Verlag.

Hox, J. (2002). “Multilevel analysis: techniques and applications”. Mahwah, NJ: Lawrence Erlbaum Associates.

Hurvich, C. M.; Tsai, C.L. (1989). “Regression and time series model selection in small samples”. *Biometrika* **76**, 297-307.

Jayasinghe, U. W.,;Marsh, H. W.;Bond, N. (2003). “A multilevel cross classified modelling approach to peer review of grant proposals: the effects of assessor and researcher attributes on assessor ratings”. *Journal of the Royal Statistical Society, A.* 166: 279-300.

Kish, L. (1965). “Survey sampling”. New York: Wiley.

Kish, L. (1987). "Statistical design for research". New York: Wiley.

Kirk, R. E. (1982) "Experimental Design: Procedures for the Behavioral Sciences". Brooks Cole Publishing Company. 2ª edição. California.

Klein, K. J.; Kozlowski, S. W. (2000) "Multilevel Theory, Research and methods in organizations: foundations, extensions and new directions". San Francisco: Jossey-Bass.

Kreft, I.G.G.;Leeu, W.J.; Dleeden, R.V.D. (1994). "Review of five multilevel analysis programs: *BMDP-5V, GENMOD, HLM, MLN3, VARCL"*. The American Statistician, v. 48, n.4, nov.

Kreft, I.G.G.; Leeu, W.J. (1998) "Introducing multilevel modeling". California: Sage Publications.

Laros, J, Marciano, J., (2008). "Análise multinível aplicada aos dados do NELS:88". Estudos em Avaliação Educacional, Brasília.

Leyland, A. H. (2004) "A review of multilevel modelling in SPSS", University of Glasgow.

Llanos, A.A, Salas, M.M., (2007). "La conveniência de análisis multinivel para la investigación em salud: uma aplicación para Costa Rica", Revista electtrónica Población Y Salud em Mesoamérica, Vol. 4, num. 2,

Maia, J.A.; Lopes, V.P.; Silva, R.G.; Seabra, A.; Ferreira, J.; Cardoso, M. (2003). "Modelação hierárquica ou multi-nível. Uma metodologia estatística e um instrumento útil de pensamento na investigação em Ciências do Desporto". Revista Portuguesa de Ciências do Desporto, vol. 3, nº 1 [92–107]

Maroco, J. (2003). "Análise Estatística – com utilização do SPSS". 2ª Edição. Lisboa, Edições Sílabo.

Maroco, J.; Bispo, R. (2003). "Estatística Aplicada às Ciências Sociais e Humanas". Lisboa, Climepsi Editores.

Miles, J.; Shevlin, M. (2001). “Applying regression and correlation: a guide for students and researchers.” London. Sage Publications.

Monette, G.; Shao, Q. e Kwan, E. (2002). “A First Look at Multilevel Models”. Institute for Social Research. Statistical Consulting Service. October – November , 2001. York University. Disponível em: http://www.math.yorku.ca/~georges/OptPortFontDeflts.pdf

Montgomery, D. C. (2005). “Design and Analysis of Experiments”. John Wiley & Sons, 6th Ed,.

Murillo, F.J. (1999). “Los Modelos Jerárquicos Lineares aplicados a la Investigación sobre Eficacia Escolar”. *Revista de Investigación Educativa, 17*(2), 453-460

Murillo, F.J. (2008). “Hacia Un Modelo De Eficacia Escolar. Estudio Multinivel Sobre Los Factores De Eficacia En Las Escuelas Españolas”, Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación 2008, Vol. 6, No. 1

Murillo Torrecilla, F. J. (2008). “Los modelos multinivel como herramienta para la investigación educativa”. Magis, Revista Internacional de Investigación en Educación, 1, 45-62,

Natis, L. (2000). “Modelos lineares hierárquicos”. Instituto de Matemática e Estatística,

Universidade de São Paulo, São Paulo. Dissertação de Mestrado.

Nezlek, J. B. (2001). “Multilevel Random Coefficient Analyses of Event and Interval Contingent Data in Social and Personality Psychology Research”. *Personality and Social Psychology Bulletin*, Vol. 27, nº 7, Julho 2001, 771-785.

Norusis, M.J. (2005). “SPSS 13.0 Advanced statistical procedures companion”. Upper Saddle River, NJ: Prentice Hall.

O’Donoghue, C.; Thomas, S.; Goldstein, H.; and Knight, T. (1997). “1996 Study of Value Added for 16-18 Year Olds in England”. London, Department for Education and Employment.

Ojeda, M.M.; H. Sahai; S.F.Juárez-Cerrillo (1999) “Multilevel data analysis with hierarchical linear models”. Statistica Applicata

Oliveira, T.A.(2004). “Estatística Aplicada”, Edições Universidade Aberta.

Pardo, A.; Ruiz, M.A.; San Martin, R. (2007). “Cómo ajustar e interpreter modelos multinivel con SPSS”. Psicotherma 2007, vol. 19, nº2. Universidad Autónoma de Madrid.

Patterson, H. D.; Thompson, R. (1971). “Recovery of inter-block information when block sizes are unequal”. Biometrika, Oxford, v. 58, p. 545-554,

Paterson, L. ;Raudenbush, S. W.; Willms, J.D. (1990). “An introduction to multilevel modeling”. *Schools, Classrooms and Pupils.* San Diego, Academic Press.

Rasbash, J. and Browne, W. (2001). “Non hierarchical multilevel models”. *Multilevel Modelling of Health Statistics.* A. Leyland and H. Goldstein. Chichester, Wiley.

Rasbash, J., Steele, F, Browne, W. and Prosser, B. (2004). “A User’s Guide to MLwiN Version 2.0”. Centre for Multilevel Modelling, Institute of Education, University of London..

Raudenbush, S.W., Bryk A.S. (2002). “Hierarchical linear models. Applications and data analysis methods”. Second Edition. Thousand Oaks: Sage Publications, Ltd.

Raudenbush, S.W. (1993). “Hierarchical linear models and experimental design”. In Edwards, L.K. (ed.) Applied Analysis of variance in the Behavioral Sciences, Mercel Dekker, New York

Raudenbush, S., Bryk, A., Cheonh, Y., Congdon, R. (2000). “HLM5 – Hierarchical linear and nonlinear modeling”. Illinois: Scientific Software International.

Reis, E. (2001). “Estatística Multivariada Aplicada”. 2ª Edição. Lisboa, Edições Sílabo.

Reise, S. P., & Duan, N. D. (2003). *Multilevel Modeling:Methodological Advances, Issues, and Applications*. Mahwah, NJ: Erlbaum.

Rutter, M.; Maughan, B.; Mortimore, P.; Ouston, J. (1979). “Fifteen Thousand Hours”. EUA

Santos, C., Ferreira, L, Oliveira, N., Dourado, M., Barreto, M., (2000). “Modelagem Multi-nível”, Sitientibus, Feira de Santana, num.22, pp. 89-98

Sahai, H., Ageel, M. (2000). “The analysis of variance: fixed, Random and mixed models”. Birkhäuser, Boston, USA

Schwarz, G. (1978). “Estimating the dimension of a model”. Annals of Statistics, 6, 461-464.

Sichieri, R., Moura, E.C., (2009). “Análise multinível das variações no índice de massa corporal entre adultos”. Brasil, 2006, Ver. Saúde Pública, vol.43, Supl. 2, São Paulo.

Sichieri, R., Barbosa, F.; Moura, E.C., (2010). “Relationship between short stature and obesity in Brasil: a multilevel analysis”. Full papers. Behaviour, appetite and Obesity. Rio de Janeiro

Soares, J. F.; Sátyro, N. G. D.; Mambrini, J. (2000). “Modelo explicativo do desempenho escolar dos alunos e análise dos fatores do SAEB – 1997”. Universidade Federal de Minas Gerais: Instituto de ciências exatas.

Soares, J. F. (2004). “O efeito da escola no desempenho cognitivo de seus alunos”. Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, v. 2, n. 2.

Soares, T. M., (2005). “Modelo de três níveis hierárquicos para a proficiência dos alunos de 4ª série avaliados no teste de língua portuguesa do SIMAVE/PROEB – 2002”, Revista Brasileira de Educação.

Soares, J. F.; Alves, M. T. G. (2007). “As pesquisas sobre o efeito das escolas: contribuições metodológicas para a sociologia da educação”. Brasília

Spiegelhalter, D.J., Thomas, A., Best, N. G.; Gilks, N. R. (1997). "BUGS: Bayesian inference using Gibbs sampling, Version 0.60". Cambridge, Medical Research Council, Biostatistics Unit.

Subramanian, S. V. (2004). "Multilevel methods, theory and analysis". Em N. Anderson (Ed.), *Encyclopedia on health and Behavior* (602-608). Thousand Oaks, CA: Sage Publications. Disponível em: http://www.hsph.harvard.edu/shdh/svsubramanian/SVSPapers/ Encyclopedia%20of%20Health%20and%20Behavior%20-%20M.pdf

Sullivan, L. M.; Dukes, K. A. ; Losina, E. (1999) "Tutorial in Biostatistics: an introduction to hierarchical linear modeling" *Statistics in Medicine*, 18, 855-888. Disponível em http://stat.gamma.rug.nl/snijders/sullivan_tutorial.pdf

Torrecilla, J. M. (coordinador) (2006). "Estudios sobre eficácia escolar en Iberoamerica – 15 buenas investigaciones", Convenio Andres Bello

Valente, V.(2007). "Estudo da «*Relevância do apoio da Escola nas perspectivas profissionais dos alunos do 10ºAno de escolaridade*» com aplicação dos Modelos Lineares Hierárquicos, Tese de Mestrado em Ensino das Ciências, Universidade Aberta.

Valente, V., Oliveira, T.A. (2007). "Modelos Lineares Hierárquicos na Educação: Uma aplicação", abstract publicado no livro de resumos da SPE 2006, Publicações INE, pg 159; artigo publicado em Ferrão, M. E., Nunes, C. e Braumann, C. A., eds, 2007. Estatística Ciência Interdisciplinar, Actas do XIV Congresso Anual da SPE, Covilhã, p. 827-837.

Valente, V., Oliveira, T.A. (2009). "Hierarquical Linear Models in Education Sciences: an application". Biometrical Letters Vol. 46, nº1, 71-86.

ANEXOS

Anexo 1: Análise dos dados referentes ao aluno (nível 1)

Índice de tabelas

Sexo					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Masculino	141	68,8	68,8	68,8
	Feminino	64	31,2	31,2	100,0
	Total	205	100,0	100,0	

Tabela A1 1 Distribuição do género sexual

Ano Lectivo * Sexo Crosstabulation					
			Sexo		Total
			Masculino	Feminino	
Ano Lectivo	405	Count	1	10	11
		% within Ano Lectivo	9,1%	90,9%	100,0%
		% within Sexo	,7%	15,6%	5,4%
		% of Total	,5%	4,9%	5,4%
	506	Count	45	17	62
		% within Ano Lectivo	72,6%	27,4%	100,0%
		% within Sexo	31,9%	26,6%	30,2%
		% of Total	22,0%	8,3%	30,2%
	607	Count	49	8	57
		% within Ano Lectivo	86,0%	14,0%	100,0%
		% within Sexo	34,8%	12,5%	27,8%
		% of Total	23,9%	3,9%	27,8%
	708	Count	46	29	75
		% within Ano Lectivo	61,3%	38,7%	100,0%
		% within Sexo	32,6%	45,3%	36,6%
		% of Total	22,4%	14,1%	36,6%
Total		Count	141	64	205
		% within Ano Lectivo	68,8%	31,2%	100,0%
		% within Sexo	100,0%	100,0%	100,0%
		% of Total	68,8%	31,2%	100,0%

Tabela A1 2 Distribuição do género sexual por ano lectivo

Classe etária					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	[14, 16[	82	40,0	41,0	41,0
	[16, 18[	92	44,9	46,0	87,0
	18 ou mais	26	12,7	13,0	100,0
	Total	200	97,6	100,0	
Missing	System	5	2,4		
Total		205	100,0		

Tabela A1 3 Distribuição da classe etária

Ano Lectivo * Classe etária Crosstabulation						
			Classe etária			Total
			[14, 16[	[16, 18[	18 ou mais	
Ano Lectivo	405	Count	5	2	4	11
		% within Ano Lectivo	45,5%	18,2%	36,4%	100,0%
		% within Classe etária	6,1%	2,2%	15,4%	5,5%
		% of Total	2,5%	1,0%	2,0%	5,5%
	506	Count	24	32	2	58
		% within Ano Lectivo	41,4%	55,2%	3,4%	100,0%
		% within Classe etária	29,3%	34,8%	7,7%	29,0%
		% of Total	12,0%	16,0%	1,0%	29,0%
	607	Count	27	21	8	56
		% within Ano Lectivo	48,2%	37,5%	14,3%	100,0%
		% within Classe etária	32,9%	22,8%	30,8%	28,0%
		% of Total	13,5%	10,5%	4,0%	28,0%
	708	Count	26	37	12	75
		% within Ano Lectivo	34,7%	49,3%	16,0%	100,0%
		% within Classe etária	31,7%	40,2%	46,2%	37,5%
		% of Total	13,0%	18,5%	6,0%	37,5%
Total		Count	82	92	26	200
		% within Ano Lectivo	41,0%	46,0%	13,0%	100,0%
		% within Classe etária	100,0%	100,0%	100,0%	100,0%
		% of Total	41,0%	46,0%	13,0%	100,0%

Tabela A1 4 Distribuição da classe etária por ano lectivo

	Sucesso				
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	sim	102	49,8	49,8	49,8
	não	103	50,2	50,2	100,0
	Total	205	100,0	100,0	

Tabela A1 5 Distribuição do sucesso

Ano Lectivo * Sucesso Crosstabulation					
			Sucesso		Total
			sim	não	
Ano Lectivo	405	Count	6	5	11
		% within Ano Lectivo	54,5%	45,5%	100,0%
		% within Sucesso	5,9%	4,9%	5,4%
		% of Total	2,9%	2,4%	5,4%
	506	Count	30	32	62
		% within Ano Lectivo	48,4%	51,6%	100,0%
		% within Sucesso	29,4%	31,1%	30,2%
		% of Total	14,6%	15,6%	30,2%
	607	Count	34	23	57
		% within Ano Lectivo	59,6%	40,4%	100,0%
		% within Sucesso	33,3%	22,3%	27,8%
		% of Total	16,6%	11,2%	27,8%
	708	Count	32	43	75
		% within Ano Lectivo	42,7%	57,3%	100,0%
		% within Sucesso	31,4%	41,7%	36,6%
		% of Total	15,6%	21,0%	36,6%
Total		Count	102	103	205
		% within Ano Lectivo	49,8%	50,2%	100,0%
		% within Sucesso	100,0%	100,0%	100,0%
		% of Total	49,8%	50,2%	100,0%

Tabela A1 6 Distribuição do Sucesso por ano lectivo

Anexo 2: Dados referentes ao curso (nível 2)

Índice de tabelas

Curso					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Mecatrónica	56	27,3	27,3	27,3
	Electrónica	55	26,8	26,8	54,1
	Química	45	22,0	22,0	76,1
	Informática	16	7,8	7,8	83,9
	HSTA	33	16,1	16,1	100,0
	Total	205	100,0	100,0	

Tabela A2 1 Distribuição de alunos por curso

Ano Lectivo * Curso Crosstabulation							
Count							
		Curso					Total
		Mecatrónica	Electrónica	Química	Informática	HSTA	
Ano Lectivo	405	0	0	11	0	0	11
	506	16	16	14	16	0	62
	607	20	19	0	0	18	57
	708	20	20	20	0	15	75
Total		56	55	45	16	33	205

Tabela A2 2 Distribuição segundo ano lectivo e curso

Curso * Sexo Crosstabulation					
			Sexo		Total
			Masculino	Feminino	
Curso	Mecatrónica	Count	53	3	56
		% within Curso	94,6%	5,4%	100,0%
		% within Sexo	37,6%	4,7%	27,3%
		% of Total	25,9%	1,5%	27,3%
	Electrónica	Count	51	4	55
		% within Curso	92,7%	7,3%	100,0%
		% within Sexo	36,2%	6,2%	26,8%
		% of Total	24,9%	2,0%	26,8%
	Química	Count	8	37	45
		% within Curso	17,8%	82,2%	100,0%
		% within Sexo	5,7%	57,8%	22,0%
		% of Total	3,9%	18,0%	22,0%
	Informática	Count	11	5	16

		% within Curso	68,8%	31,2%	100,0%
		% within Sexo	7,8%	7,8%	7,8%
		% of Total	5,4%	2,4%	7,8%
	HSTA	Count	18	15	33
		% within Curso	54,5%	45,5%	100,0%
		% within Sexo	12,8%	23,4%	16,1%
		% of Total	8,8%	7,3%	16,1%
Total		Count	141	64	205
		% within Curso	68,8%	31,2%	100,0%
		% within Sexo	100,0%	100,0%	100,0%
		% of Total	68,8%	31,2%	100,0%

Tabela A2 3 Distribuição de género sexual por curso

Ano Lectivo * Residência Crosstabulation

			Residência		Total
			rural	urbano	
Ano Lectivo	405	Count	4	7	11
		% within Ano Lectivo	36,4%	63,6%	100,0%
		% within Residência	5,2%	5,5%	5,4%
		% of Total	2,0%	3,4%	5,4%
	506	Count	18	44	62
		% within Ano Lectivo	29,0%	71,0%	100,0%
		% within Residência	23,4%	34,6%	30,4%
		% of Total	8,8%	21,6%	30,4%
	607	Count	26	30	56
		% within Ano Lectivo	46,4%	53,6%	100,0%
		% within Residência	33,8%	23,6%	27,5%
		% of Total	12,7%	14,7%	27,5%
	708	Count	29	46	75
		% within Ano Lectivo	38,7%	61,3%	100,0%
		% within Residência	37,7%	36,2%	36,8%
		% of Total	14,2%	22,5%	36,8%
Total		Count	77	127	204
		% within Ano Lectivo	37,7%	62,3%	100,0%
		% within Residência	100,0%	100,0%	100,0%

Tabela A2 4 Distribuição de residência do aluno por ano lectivo

Curso * Sucesso Crosstabulation					
			Sucesso		Total
			sim	não	
Curso	Mecatrónica	Count	22	34	56
		% within Curso	39,3%	60,7%	100,0%
		% within Sucesso	21,6%	33,0%	27,3%
		% of Total	10,7%	16,6%	27,3%
	Electrónica	Count	27	28	55
		% within Curso	49,1%	50,9%	100,0%
		% within Sucesso	26,5%	27,2%	26,8%
		% of Total	13,2%	13,7%	26,8%
	Química	Count	23	22	45
		% within Curso	51,1%	48,9%	100,0%
		% within Sucesso	22,5%	21,4%	22,0%
		% of Total	11,2%	10,7%	22,0%
	Informática	Count	9	7	16
		% within Curso	56,2%	43,8%	100,0%
		% within Sucesso	8,8%	6,8%	7,8%
		% of Total	4,4%	3,4%	7,8%
	HSTA	Count	21	12	33
		% within Curso	63,6%	36,4%	100,0%
		% within Sucesso	20,6%	11,7%	16,1%
		% of Total	10,2%	5,9%	16,1%
Total		Count	102	103	205
		% within Curso	49,8%	50,2%	100,0%
		% within Sucesso	100,0%	100,0%	100,0%
		% of Total	49,8%	50,2%	100,0%

Tabela A2 5 Distribuição de sucesso por curso

Curso * Residência Crosstabulation

			Residência		Total
			rural	urbano	
Curso	Mecatrónica	Count	23	32	55
		% within Curso	41,8%	58,2%	100,0%
		% within Residência	29,9%	25,2%	27,0%
		% of Total	11,3%	15,7%	27,0%
	Electrónica	Count	25	30	55
		% within Curso	45,5%	54,5%	100,0%
		% within Residência	32,5%	23,6%	27,0%
		% of Total	12,3%	14,7%	27,0%
	Química	Count	15	30	45
		% within Curso	33,3%	66,7%	100,0%
		% within Residência	19,5%	23,6%	22,1%
		% of Total	7,4%	14,7%	22,1%
	Informática	Count	1	15	16
		% within Curso	6,2%	93,8%	100,0%
		% within Residência	1,3%	11,8%	7,8%
		% of Total	,5%	7,4%	7,8%
	HSTA	Count	13	20	33
		% within Curso	39,4%	60,6%	100,0%
		% within Residência	16,9%	15,7%	16,2%
		% of Total	6,4%	9,8%	16,2%
Total		Count	77	127	204
		% within Curso	37,7%	62,3%	100,0%
		% within Residência	100,0%	100,0%	100,0%
		% of Total	37,7%	62,3%	100,0%

Tabela A2 6 distribuição de residência do aluno por curso

Descriptives

	Curso			Statistic	Std. Error
Media final	Mecatrónica	Mean		13,0071	,13658
		95% Confidence Interval for Mean	Lower Bound	12,7334	
			Upper Bound	13,2809	
		5% Trimmed Mean		12,9512	
		Median		12,8000	
		Variance		1,045	
		Std. Deviation		1,02209	
		Minimum		11,00	
		Maximum		16,50	
		Range		5,50	
		Interquartile Range		1,45	
		Skewness		,955	,319
		Kurtosis		1,279	,628
	Electrónica	Mean		12,9727	,17401
		95% Confidence Interval for Mean	Lower Bound	12,6239	
			Upper Bound	13,3216	
		5% Trimmed Mean		12,9409	
		Median		12,7000	
		Variance		1,665	
		Std. Deviation		1,29049	
		Minimum		10,60	
		Maximum		15,70	
		Range		5,10	
		Interquartile Range		2,00	
		Skewness		,490	,322
		Kurtosis		-,538	,634
	Química	Mean		13,8378	,22895
		95% Confidence Interval for Mean	Lower Bound	13,3764	
			Upper Bound	14,2992	
		5% Trimmed Mean		13,7870	
		Median		13,7000	
		Variance		2,359	
		Std. Deviation		1,53583	
		Minimum		10,90	
		Maximum		18,30	
		Range		7,40	
		Interquartile Range		2,30	
		Skewness		,459	,354
		Kurtosis		,191	,695
	Informática	Mean		13,5312	,38338
		95% Confidence Interval for Mean	Lower Bound	12,7141	

		Mean	Upper Bound	14,3484	
		5% Trimmed Mean		13,4625	
		Median		13,4500	
		Variance		2,352	
		Std. Deviation		1,53350	
		Minimum		11,20	
		Maximum		17,10	
		Range		5,90	
		Interquartile Range		1,97	
		Skewness		,792	,564
		Kurtosis		,623	1,091
	HSTA	Mean		12,7727	,17876
		95% Confidence Interval for	Lower Bound	12,4086	
		Mean	Upper Bound	13,1369	
		5% Trimmed Mean		12,7485	
		Median		12,5000	
		Variance		1,055	
		Std. Deviation		1,02691	
		Minimum		11,20	
		Maximum		14,90	
		Range		3,70	
		Interquartile Range		1,95	
		Skewness		,470	,409
		Kurtosis		-,866	,798

Tabela A2 7 Estatísticas descritivas da média final por curso

Descriptives

	Curso			Statistic	Std. Error
Média Sociocultural	Mecatrónica	Mean		12,6179	,12536
		95% Confidence Interval for Mean	Lower Bound	12,3666	
			Upper Bound	12,8691	
		5% Trimmed Mean		12,5631	
		Median		12,4500	
		Variance		,880	
		Std. Deviation		,93810	
		Minimum		11,30	
		Maximum		15,50	
		Range		4,20	
		Interquartile Range		1,17	
		Skewness		,971	,319
		Kurtosis		,626	,628
	Electrónica	Mean		12,9145	,13485
		95% Confidence Interval for Mean	Lower Bound	12,6442	
			Upper Bound	13,1849	
		5% Trimmed Mean		12,8894	
		Median		12,7000	
		Variance		1,000	
		Std. Deviation		1,00008	
		Minimum		10,70	
		Maximum		15,50	
		Range		4,80	
		Interquartile Range		1,40	
		Skewness		,383	,322
		Kurtosis		,094	,634
	Química	Mean		13,5622	,18291
		95% Confidence Interval for Mean	Lower Bound	13,1936	
			Upper Bound	13,9309	
		5% Trimmed Mean		13,5031	
		Median		13,4000	
		Variance		1,506	
		Std. Deviation		1,22702	
		Minimum		11,30	
		Maximum		17,30	
		Range		6,00	
		Interquartile Range		1,75	
		Skewness		,771	,354
		Kurtosis		,733	,695
	Informática	Mean		13,3312	,33933
		95% Confidence Interval for Mean	Lower Bound	12,6080	

		for Mean	Upper Bound	14,0545	
		5% Trimmed Mean		13,2514	
		Median		13,1500	
		Variance		1,842	
		Std. Deviation		1,35731	
		Minimum		11,80	
		Maximum		16,30	
		Range		4,50	
		Interquartile Range		1,58	
		Skewness		1,030	,564
		Kurtosis		,250	1,091
	HSTA	Mean		13,0061	,17451
		95% Confidence Interval	Lower Bound	12,6506	
		for Mean	Upper Bound	13,3615	
		5% Trimmed Mean		12,9811	
		Median		12,9000	
		Variance		1,005	
		Std. Deviation		1,00248	
		Minimum		11,50	
		Maximum		15,00	
		Range		3,50	
		Interquartile Range		1,60	
		Skewness		,385	,409
		Kurtosis		-,824	,798
Média Técnica	Mecatrónica	Mean		13,2464	,14941
		95% Confidence Interval	Lower Bound	12,9470	
		for Mean	Upper Bound	13,5459	
		5% Trimmed Mean		13,2079	
		Median		13,0500	
		Variance		1,250	
		Std. Deviation		1,11811	
		Minimum		11,20	
		Maximum		16,50	
		Range		5,30	
		Interquartile Range		1,47	
		Skewness		,605	,319
		Kurtosis		,447	,628
	Electrónica	Mean		13,2945	,19659
		95% Confidence Interval	Lower Bound	12,9004	
		for Mean	Upper Bound	13,6887	
		5% Trimmed Mean		13,2606	
		Median		13,0000	
		Variance		2,126	

		Std. Deviation		1,45798	
		Minimum		10,60	
		Maximum		16,60	
		Range		6,00	
		Interquartile Range		2,20	
		Skewness		,315	,322
		Kurtosis		-,683	,634
	Química	Mean		14,2267	,26083
		95% Confidence Interval	Lower Bound	13,7010	
		for Mean	Upper Bound	14,7523	
		5% Trimmed Mean		14,1951	
		Median		13,9000	
		Variance		3,062	
		Std. Deviation		1,74973	
		Minimum		11,00	
		Maximum		18,70	
		Range		7,70	
		Interquartile Range		2,60	
		Skewness		,234	,354
		Kurtosis		-,445	,695
	Informática	Mean		14,1438	,48296
		95% Confidence Interval	Lower Bound	13,1144	
		for Mean	Upper Bound	15,1731	
		5% Trimmed Mean		14,1208	
		Median		14,0000	
		Variance		3,732	
		Std. Deviation		1,93183	
		Minimum		10,70	
		Maximum		18,00	
		Range		7,30	
		Interquartile Range		2,83	
		Skewness		,242	,564
		Kurtosis		-,071	1,091
	HSTA	Mean		13,1030	,18897
		95% Confidence Interval	Lower Bound	12,7181	
		for Mean	Upper Bound	13,4880	
		5% Trimmed Mean		13,0678	
		Median		13,0000	
		Variance		1,178	
		Std. Deviation		1,08555	
		Minimum		11,50	
		Maximum		15,30	
		Range		3,80	

		Interquartile Range		2,00	
		Skewness		,506	,409
		Kurtosis		-,912	,798
Média Científica	Mecatrónica	Mean		12,7446	,18941
		95% Confidence Interval	Lower Bound	12,3651	
		for Mean	Upper Bound	13,1242	
		5% Trimmed Mean		12,6488	
		Median		12,4000	
		Variance		2,009	
		Std. Deviation		1,41741	
		Minimum		10,50	
		Maximum		17,00	
		Range		6,50	
		Interquartile Range		1,78	
		Skewness		1,135	,319
		Kurtosis		1,069	,628
	Electrónica	Mean		12,7091	,23238
		95% Confidence Interval	Lower Bound	12,2432	
		for Mean	Upper Bound	13,1750	
		5% Trimmed Mean		12,6551	
		Median		12,3000	
		Variance		2,970	
		Std. Deviation		1,72340	
		Minimum		10,00	
		Maximum		16,70	
		Range		6,70	
		Interquartile Range		2,60	
		Skewness		,580	,322
		Kurtosis		-,486	,634
	Química	Mean		13,7489	,28968
		95% Confidence Interval	Lower Bound	13,1651	
		for Mean	Upper Bound	14,3327	
		5% Trimmed Mean		13,7012	
		Median		13,5000	
		Variance		3,776	
		Std. Deviation		1,94324	
		Minimum		10,50	
		Maximum		18,80	
		Range		8,30	
		Interquartile Range		3,05	
		Skewness		,320	,354
		Kurtosis		-,277	,695
	Informática	Mean		13,1375	,41381

		95% Confidence Interval	Lower Bound	12,2555	
		for Mean	Upper Bound	14,0195	
		5% Trimmed Mean		13,0361	
		Median		12,7000	
		Variance		2,740	
		Std. Deviation		1,65524	
		Minimum		11,00	
		Maximum		17,10	
		Range		6,10	
		Interquartile Range		2,40	
		Skewness		,865	,564
		Kurtosis		,647	1,091
	HSTA	Mean		12,2303	,24024
		95% Confidence Interval	Lower Bound	11,7409	
		for Mean	Upper Bound	12,7197	
		5% Trimmed Mean		12,1825	
		Median		11,9000	
		Variance		1,905	
		Std. Deviation		1,38010	
		Minimum		10,00	
		Maximum		15,40	
		Range		5,40	
		Interquartile Range		1,60	
		Skewness		,791	,409
		Kurtosis		,293	,798
Total Mód atraso	Mecatrónica	Mean		7,89	1,537
		95% Confidence Interval	Lower Bound	4,81	
		for Mean	Upper Bound	10,97	
		5% Trimmed Mean		6,28	
		Median		4,50	
		Variance		132,243	
		Std. Deviation		11,500	
		Minimum		0	
		Maximum		47	
		Range		47	
		Interquartile Range		10	
		Skewness		2,005	,319
		Kurtosis		3,860	,628
	Electrónica	Mean		6,78	1,569
		95% Confidence Interval	Lower Bound	3,64	
		for Mean	Upper Bound	9,93	
		5% Trimmed Mean		4,85	
		Median		2,00	

		Variance		135,322	
		Std. Deviation		11,633	
		Minimum		0	
		Maximum		59	
		Range		59	
		Interquartile Range		11	
		Skewness		2,779	,322
		Kurtosis		8,862	,634
	Química	Mean		6,67	1,454
		95% Confidence Interval	Lower Bound	3,74	
		for Mean	Upper Bound	9,60	
		5% Trimmed Mean		5,48	
		Median		,00	
		Variance		95,136	
		Std. Deviation		9,754	
		Minimum		0	
		Maximum		44	
		Range		44	
		Interquartile Range		11	
		Skewness		1,845	,354
		Kurtosis		3,845	,695
	Informática	Mean		3,00	1,133
		95% Confidence Interval	Lower Bound	,59	
		for Mean	Upper Bound	5,41	
		5% Trimmed Mean		2,67	
		Median		,00	
		Variance		20,533	
		Std. Deviation		4,531	
		Minimum		0	
		Maximum		12	
		Range		12	
		Interquartile Range		7	
		Skewness		1,268	,564
		Kurtosis		,022	1,091
	HSTA	Mean		2,97	,995
		95% Confidence Interval	Lower Bound	,94	
		for Mean	Upper Bound	5,00	
		5% Trimmed Mean		2,09	
		Median		,00	
		Variance		32,655	
		Std. Deviation		5,714	
		Minimum		0	
		Maximum		23	

		Range	23	
		Interquartile Range	4	
		Skewness	2,402	,409
		Kurtosis	5,683	,798

Tabela A2 8 estatísticas descritivas das médias das componentes e numero de módulos em atraso, por curso