

# Alguns contributos para a qualidade da formação do CENFIM: Caraterização dos formandos, satisfação e sucesso escolar

Loide Ascenso

MMEAD/ECT da Universidade de Évora, [loyderaquel@hotmail.com](mailto:loyderaquel@hotmail.com)

Gonçalo Jacinto

DMAT/ECT e CIMA-UE da Universidade de Évora, [gjcj@uevora.pt](mailto:gjcj@uevora.pt)

Paulo Infante

DMAT/ECT e CIMA-UE da Universidade de Évora, [pinfante@uevora.pt](mailto:pinfante@uevora.pt)

**Palavras-chave:** Sucesso escolar, regressão logística e regressão ordinal

**Abstract:** O CENFIM é um centro de formação profissional, associado à indústria Metalúrgica e Metalomecânica, para empresas, adultos e jovens. Para o sucesso da formação torna-se importante conhecer os fatores que levam os seus formandos a escolher este centro de formação, quais as suas motivações e as razões para o sucesso escolar. Para tal, construiu-se um inquérito por questionário que se distribuiu pelos centros do CENFIM. Neste trabalho analisamos os resultados desse inquérito.

Começamos por realizar uma análise exploratória de dados visando caracterizar o tipo de alunos que frequenta o Centro. De seguida identificamos alguns fatores que podem influenciar o sucesso escolar recorrendo para tal à regressão logística e à regressão ordinal. Concluimos que os alunos do sexo feminino, que residem num meio rural, com pouca dedicação ao estudo e falta de bases relativamente a anos anteriores manifestam mais dificuldades na aprendizagem. Além disso, concluimos também que os alunos com um mau ambiente familiar tendem a apresentar dificuldades que podem levar ao insucesso escolar.

## 1 Introdução

O CENFIM, Centro de Formação Profissional da Indústria Metalúrgica e Metalomecânica, compreende uma diversidade de formações direcionadas para empresas, para adultos e para jovens, através dos 12 núcleos de formação distribuídos pelo país. A forma como um estudante tem conhecimento dos cursos ministrados pelo CENFIM e os motivos que o levam a ingressar num curso profissional são fatores muito importantes para o centro determinar o que pode cativar mais alunos. Neste contexto de grandes esforços para captação de novos alunos e manutenção dos já existentes, assume grande importância o estudo da satisfação discente. A avaliação dada não só a formadores como à oferta que o centro dispõe, tanto de formação

como de instalações e serviços, pode ser encarada como indicador do grau de satisfação dos formandos que frequentam as ações do CENFIM.

Em [7] é realizado um estudo no âmbito da transição do ensino secundário para o ensino superior, onde foi analisada a influência do grau de satisfação no rendimento académico, utilizando um modelo de regressão logístico multinomial. Os autores agruparam em 6 categorias aquilo que os estudantes consideram ser determinantes para o rendimento académico: aspetos motivacionais, atitudes face ao estudo, métodos de ensino, aspetos psico-sociais, organização do processo de ensino e infra-estruturas e recursos da instituição de ensino.

Neste trabalho pretendemos conhecer os fatores que levam os formandos do CENFIM a escolher este centro de formação, quais as suas motivações e as razões para o sucesso escolar. Com esse intuito, construiu-se um inquérito que se distribuiu pelos centros do CENFIM de forma a poder avaliar esta temática. A sua análise permitiu caracterizar o aluno que frequenta o Centro, estudar algumas associações entre variáveis relacionadas com o sucesso escolar e identificar fatores potenciadores e protetores desse mesmo sucesso. Foram considerados 11 dos 12 núcleos de formação, por serem aqueles que tinham, no período da recolha de dados, a decorrer formação profissional para jovens, onde estão incluídos os Cursos de Educação e Formação (CEF), Cursos de Aprendizagem Nível 4 (APZ) e Cursos de Especialização Tecnológica (CET). Os 11 núcleos de formação considerados foram: Amarante, Arcos de Valdevez, Caldas da Rainha, Ermesinde, Lisboa, Marinha Grande, Oliveira de Azeméis, Porto, Santarém, Torres Vedras e Trofa.

Neste trabalho considerou-se um universo de pesquisa com 658 formandos pertencentes aos 11 núcleos de formação do CENFIM já descritos. Como instrumento de recolha de dados utilizou-se um questionário dividido em duas partes, aplicado entre Dezembro de 2011 e Abril de 2012. O questionário foi construído com base em alguma revisão da literatura, justificando a escolha da escala e de algumas questões de carácter geral [5].

A primeira parte no nosso inquérito aborda as questões de identificação do formando, onde são colocadas questões sobre o sexo, a idade, o curso que frequenta, habilitações literárias, situação perante o trabalho, meio de residência e país de origem. Numa segunda parte foram feitas questões relacionadas com os dados escolares dos formandos, nomeadamente o núcleo de formação a que pertence, forma como tomou conhecimento do curso, objetivos após a conclusão do curso e questões relacionadas com a satisfação dos formandos em relação ao centro e aos formadores. O inquérito foi validado com a aplicação a um grupo de 20 alunos e mediante as suas respostas e dificuldades de interpretação, ajustou-se a versão final do inquérito.

Recorrendo a modelos de regressão logística e regressão ordinal estudamos a influência de algumas covariáveis sobre a variável resposta que descreve as dificuldades em compreender os conteúdos programáticos, permitindo identificar fatores significativos para modelar o sucesso escolar.

Este artigo está organizado da seguinte forma. Na Secção 2 fazemos uma descrição da amostra de forma a caracterizar os estudantes inquiridos, como por exemplo ao nível da sua idade, zona onde vivem, os seus objetivos e motivações. Na Secção 3 fazemos uma descrição sucinta sobre os métodos estatísticos utilizados. Na Secção 4 apresentamos e discutimos os resultados obtidos com os modelos de regressão logística e ordinal. Finalmente, na Secção 5 salientamos algumas conclusões retirados com este estudo.

## 2 Caracterização da base de dados

No presente estudo participaram 658 formandos pertencentes ao centro de formação CENFIM sendo quase todos do sexo masculino (96.5%). Na Tabela 1 podemos observar algumas das variáveis que foram recolhidas.

Tabela 1: Tabela de variáveis analisadas.

Variável	Categorias
Idade	-
Sexo	feminino, masculino
Curso	CEF, APZ, CET
Núcleo de formação	Amarante, Arcos de Valdevez, Caldas da Rainha, Lisboa, Marinha Grande, Oliveira de Azeméis, Porto, Santarém, Torres Vedras, Trofa
Situação perante o trabalho	estudante, desempregado, trabalhador, trabalhador/estudante, trabalhador por conta própria, trabalhador por conta de outrem, outra
Freguesia de residência	rural, urbana
Como tomou conhecimento do curso	pais, amigos, professores, jornal/rádio, outros formandos, outro
Boa relação com os formadores	Discordo totalmente, discordo, concordo, concordo totalmente
Objectivos pós curso	ensino superior, trabalho, curso CET, outro curso, outros
Motivos de entrada no curso	curso acessível, acesso a profissão bem remunerada, acesso a profissão com empregabilidade, aconselhado, adequado às minhas pretensões
Dificuldade de compreensão	nenhuma, alguma, muita
Dificuldade de concentração	nenhuma, alguma, muita
Dificuldade em memorizar	nenhuma, alguma, muita
Dificuldade de interpretação de textos	nenhuma, alguma, muita
Falta de dedicação às atividades de estudo	nenhuma, alguma, muita
Falta de bases relativamente a anos anteriores	nenhuma, alguma, muita
Problemas familiares	nenhuma, alguma, muita
Turma agitada	nenhuma, alguma, muita
Condições do Centro: Criação de biblioteca/ sala de estudo	Muito Importante, Importante, Razoavelmente importante, Pouco Importante Sem Importância
Localização geográfica	Muito Importante, Importante, Razoavelmente importante, Pouco Importante Sem Importância

A distribuição dos formandos por núcleo de formação pode ser observada na

Tabela 2: Distribuição dos formandos por núcleo de formação e curso

Núcleo	Inquiridos (%)	Curso (%)		
		CEF	APZ	CET
Amarante	9,0	0,0	37,3	62,7
Arcos de Valdevez	4,0	0,0	0,0	100,0
Caldas da Rainha	9,9	0,0	68,8	31,2
Ermesinde	16,3	0,0	67,3	32,7
Lisboa	10,9	18,3	39,4	42,3
Marinha Grande	6,7	0,0	100,0	0,0
Oliveira de Azeméis	8,8	0,0	100,0	0,0
Porto	8,4	0,0	43,6	56,4
Santarém	5,5	0,0	25,0	75,0
Torres Vedras	10,6	17,1	82,9	0,0
Trofa	10,0	0,0	45,5	54,5

Tabela 2. Os núcleos de formação com maior predominância de formandos inquiridos são os de Ermesinde e de Lisboa, seguindo-se o núcleo de Torres Vedras. A amostra é constituída por 59,3% de formandos que frequentam cursos de Aprendizagem (APZ), correspondentes ao ensino secundário, 36,9% dos quais frequentam os Cursos de Especialização Tecnológica (CET) e uma percentagem residual de 3,8% frequenta os Cursos de Educação e Formação (CEF), correspondentes ao ensino básico. Note-se que relativamente ao período de tempo em que foram aplicados os inquéritos, Dezembro de 2011 a Abril de 2012, foram inquiridos todos os alunos que se encontram numa destas três formações do CENFIM.

A curva de distribuição das idades dos formandos inquiridos é uma curva assimétrica à direita, visto que a população de formandos adultos é consideravelmente menor que a população de jovens. Verifica-se que a idade mediana do sexo masculino (19.0 anos) não difere muito da idade mediana do sexo feminino (20.0 anos). Verifica-se ainda que cerca de 2/3 dos inquiridos é apenas estudante, enquanto a restante minoria se divide em trabalhador estudante e desempregado. A esmagadora maioria dos inquiridos (90.9%) tem nacionalidade Portuguesa, seguindo-se os formandos de outros países da Europa (4.3%), África (2.4%), América (2.1%) e Ásia (0.3%).

No que diz respeito ao distrito de residência dos formandos do CENFIM, existe uma predominância do distrito do Porto (38.9%), registando-se também uma representação relevante dos distritos de Lisboa (19,8%) e Leiria (15,8%) e cada um dos restantes distritos de residência (Aveiro, Santarém, Braga, Viana do Castelo, Setúbal e Viseu) tem menos de 10% dos inquiridos. A maioria dos inquiridos (54.6%) tomou conhecimento do respetivo curso através de outros formandos que já frequentavam as ações do CENFIM. Muitos foram aconselhados por professores (10.7%) ou pela família (14.5%). É ainda importante salientar os 11.4% de inquiridos que tiveram conhecimento através de publicidade. Um pouco mais de 2/3 dos inquiridos

que frequentam as ações de formação ministradas no CENFIM pretendem concluir o curso para integrar o mundo do trabalho, enquanto 19.4% dos formandos pretendem ingressar no ensino superior após a conclusão do curso.

### 3 Metodologia utilizada

O teste de qui-quadrado de Pearson foi utilizado para testar a associação entre o núcleo de formação e as variáveis associadas às dificuldades escolares, e também para avaliar a associação entre o núcleo de formação e os motivos de entrada no curso. O nível de significância foi fixado em 5%.

Os interesses do CENFIM conduziram a um grau de especificidade do questionário que nos impôs algumas restrições. Das variáveis que dispunhamos para analisar, e de acordo com um instrumento previamente estudado e adaptado ao caso português [3], aquela variável que pensamos traduzir melhor as dificuldades escolares do formando é a que avalia as dificuldades em compreender os conteúdos programáticos e, conseqüentemente, optou-se por considerar esta a nossa variável resposta de interesse. Esta é uma variável ordinal com três categorias: nenhuma, algumas e muitas dificuldades. Assim, para conhecer os fatores que podem condicionar o sucesso escolar, ajustámos um modelo de regressão ordinal usando esta variável como variável resposta e usando as restantes variáveis da Tabela 1 como variáveis explicativas. Optámos por ajustar o modelo das chances proporcionais por ser mais simples de interpretar e porque se verificaram os pressupostos subjacentes, mas outros modelos poderiam ter sido ajustados, como se pode ver, por exemplo, em [2] e [4].

No modelo das chances proporcionais, de acordo com a parametrização usada pelo R, compara-se a probabilidade de uma resposta ser maior ou igual a uma dada categoria da variável,  $Y \geq k$ , com a probabilidade de se obter uma resposta de uma categoria inferior,  $Y < k$ . A função logit para este modelo é dada por

$$\begin{aligned} g_k(\mathbf{x}) &= \ln \left( \frac{P(Y \geq k|\mathbf{x})}{P(Y < k|\mathbf{x})} \right) \\ &= \ln \left( \frac{\phi_k(\mathbf{x}) + \phi_{k+1}(\mathbf{x}) + \dots + \phi_K(\mathbf{x})}{\phi_0(\mathbf{x}) + \phi_1(\mathbf{x}) + \dots + \phi_{k-1}(\mathbf{x})} \right) = \beta_{0k} + \mathbf{x}'\beta \end{aligned}$$

onde  $k = 0, 1, \dots, K - 1$ , representa as  $K$  categorias da variável resposta,  $\phi_k(\mathbf{x}) = P(Y = k|\mathbf{x})$  representa a probabilidade da variável resposta tomar o valor  $k$  condicional ao vetor  $\mathbf{x}$  das  $p$  covariáveis, e onde  $\beta$  representa o vetor dos coeficientes das  $p$  covariáveis. Tal como na regressão logística, a exponencial dos coeficientes do modelo traduz a razão de chances (Odds Ratio) de pertencer a uma das categorias superiores ou iguais a  $k$  relativamente a pertencer a categorias antes de  $k$ . Mais informações e especificações sobre estes modelos podem ser consultadas por exemplo em [4].

A verificação da adequabilidade do modelo de chances proporcionais foi verificada recorrendo aos resíduos score e resíduos parciais [1] e também usando a recomendação de Hosmer e Lemeshow ([4], pag. 304), que consiste em comparar este modelo com o modelo sem restrições (modelo multinomial) através do teste de razão de verossimilhanças.

Com o objetivo de comparar os resultados do modelo ordinal com os resultados obtidos através de um modelo de regressão logística, agrupámos a variável resposta em duas categorias para ajustar os modelos de regressão logística, neste caso, os procedimentos utilizados para o ajustamento do modelo final foram os mesmos que os usados para ajustar o modelo de regressão ordinal.

Para ajustar os modelos seguimos a metodologia recomendada por Hosmer e Lemeshow, [4]: (1) para o modelo inicial foram selecionadas todas as variáveis que se revelaram significativas na fase univariada (valor  $p < 0.20$ ); (2) a partir deste modelo foram eliminadas sucessivamente, e por ordem decrescente dos valores  $p$  de Wald, todas as variáveis não significativas (valor  $p > 0.10$ ); (3) verificámos se alguma(s) das variáveis que não foram incluídas no modelo inicial se mostra(m) agora significativa(s) em presença das que estão no modelo, caso em que são adicionadas ao modelo; (4) adequabilidade do número de categorias de cada variável e verificação, para as variáveis contínuas, do pressuposto da linearidade com o *logit*; (5) foram testadas, via teste de razão de verossimilhanças, as interações entre as variáveis que faziam sentido no contexto do estudo (valor  $p < 0.05$ ); (6) análise de resíduos para pesquisa de observações influentes ou outliers. A qualidade do ajustamento dos modelos de regressão logística foi verificada pelo teste de Hosmer e Lemeshow, enquanto para o modelo de regressão ordinal foi verificada pelo teste de Cessie-van Houwelingen. A capacidade discriminativa dos modelos foi avaliada pela área sob a curva ROC (Area under Curve, AUC). O modelo final foi validado usando a técnica de validação por bootstrap. A análise estatística foi realizada no software R project, versão 2.15.3 [6], e para ajustamento dos modelos foi usado o pacote *rms*.

## 4 Análise e discussão dos resultados

Dos formandos que afirmam que querem novas oportunidades de trabalho a maioria (85,7%) reside num meio rural e apenas 14.3% residem num meio urbano. Os meios rural e urbano também apresentam diferenças significativas ( $\chi^2_{(3)} = 9.3$ , valor  $p=0.03$ ) relativamente aos objetivos pós curso. Os inquiridos residentes num meio rural preferem as novas oportunidades de trabalho e integração no mundo do trabalho relativamente aos do meio urbano. Estes últimos preferem o ingresso no ensino superior ou aquisição de novos conhecimentos.

Verificou-se uma associação significativa entre o núcleo de formação e os motivos de entrada nos cursos (valor  $p < 0.049$  para todos os testes) concluindo-

se que a região geográfica onde cada centro se localiza influencia as razões para a escolha de um curso de formação no CENFIM. Além disso verifica-se que existe associação entre as variáveis afetas às dificuldades escolares e o núcleo de formação (valor  $p < 0.02$  para todos os testes), com exceção da variável falta de bases relativamente a anos anteriores (valor  $p = 0.09$ ).

Concluiu-se que existem centros que têm alunos com maiores dificuldades escolares: para os alunos de Arcos de Valdevez, Marinha Grande, Porto e Trofa, a escolha do curso por ser acessível/fácil é menos importante, em contraste com os alunos dos centros de Oliveira de Azeméis e Torres Vedras, registando-se muitos mais alunos que o esperado (se houvesse independência) a acharem importante o fato do curso ser acessível/fácil para o frequentarem. Na etapa seguinte estudaram-se os fatores que contribuem para o sucesso escolar. Como já foi referido optou-se por utilizar a variável *dificuldades em compreender os conteúdos programáticos* como variável resposta para modelar o insucesso escolar do aluno. Esta é uma variável ordinal com três categorias: nenhuma, algumas e muitas dificuldades. Devido à natureza desta variável ajustou-se um modelo de regressão logística ordinal.

Usando o método de seleção de variáveis referido na Secção 3, pudemos concluir que as variáveis sexo, freguesia de residência, núcleo de formação, dedicação às atividades de estudo, falta de bases e problemas familiares, têm efeito significativo no modelo multivariado. Note-se que as variáveis relevantes para o próprio centro, como turma agitada, criação de biblioteca/sala de estudo, objetivos de entrada no curso, relação com os formadores e localização geográfica, não se mostraram significativas para modelar as dificuldades dos alunos. Por outro lado, apenas a variável potenciadora do sucesso escolar, dedicação às atividades de estudo, se revelou significativa, enquanto das variáveis que se podem considerar de risco para o sucesso escolar apenas as variáveis falta de bases e problemas familiares se revelaram significativas.

Como é feito em [4], para testar a igualdade entre dois coeficientes do modelo é aplicado um teste de Wald. Caso não haja diferenças entre os coeficientes, o resultado sugere que é possível utilizar um modelo mais simples agrupando as categorias cujos coeficientes não diferem, caso no contexto da modelação tal faça sentido. Obtido o modelo com menos coeficientes, pode usar-se o teste de razão de verosimilhanças para o comparar com o modelo inicial (sem junção de categorias), uma vez que se tratam de modelos encaixados, para verificar que não existem diferenças significativas na deviance dos dois modelos. Utilizámos esta abordagem para a variável núcleo de formação e constatou-se que vários coeficientes eram idênticos, traduzindo diferenças não significativas entre esses centros, pelo que no modelo final aparecem apenas dois grupos: grupo 1 - Lisboa; Porto; Torres Vedras; Caldas da Rainha; Santarém (298 alunos); grupo 2 - Amarante; Arcos de Valdevez; Marinha Grande; Ermesinde; Trofa; Oliveira de Azeméis; (360 alunos).

Testou-se de seguida a existência de interações entre as variáveis, não se verificando nenhuma interação significativa. A análise dos resíduos score

e resíduos parciais permitiu validar o pressuposto dos riscos proporcionais, verificando-se que este modelo não é significativamente diferente do modelo sem restrições (modelo multinomial, valor  $p=0.43$ ). A capacidade discriminativa do modelo é aceitável, com uma AUC de 0.76. A aplicação do teste de Cessie-van Houwelingen permitiu validar a qualidade do ajustamento do modelo ordinal (valor  $p=0.21$  e valor  $p=0.14$ ). As estimativas finais dos parâmetros do modelo de regressão ordinal encontram-se na Tabela 3. Para valores fixos das restantes variáveis podemos concluir que:

Tabela 3: Estimativas dos parâmetros ( $\hat{\beta}$ ) e respetivos desvios-padrão ( $SE(\hat{\beta})$ ), valores  $p$  associados, razão das chances (RC) e respetivos intervalos de confiança a 95% ( $IC_{95\%}$ ) do modelo de regressão logística ordinal para as dificuldades em compreender os conteúdos programáticos.

Variável	$\hat{\beta}$	$SE(\hat{\beta})$	valor $p$	RC ( $IC_{95\%}$ )
<b>Constantes:</b>				
$y \geq 1$	0.61	0.50	0.204	
$y \geq 2$	-3.67	0.53	<0.001	
<b>Sexo:</b>				
Feminino				
Masculino	-1.16	0.47	0.013	0.3 (0.1, 0.8)
<b>Freguesia de residência:</b>				
Rural				
Urbana	-0.72	0.18	<0.001	0.5 (0.3, 0.7)
<b>Núcleo de formação:</b>				
grupo 1				
grupo 2	0.58	0.18	0.001	1.8 (1.3, 2.5)
<b>Falta de dedicação às atividades de estudo:</b>				
Nenhuma				
Alguma	0.82	0.19	<0.001	2.3 (1.6, 2.5)
Muita	1.58	0.34	<0.001	4.9 (2.5, 9.4)
<b>Falta de bases relativamente a anos anteriores:</b>				
Nenhuma				
Alguma	1.04	0.20	<0.001	2.8 (1.9, 4.2)
Muita	1.82	0.30	<0.001	6.3 (3.4, 11.1)
<b>Problemas familiares:</b>				
Nenhuns				
Alguns	0.52	0.22	0.014	1.7 (1.1, 2.6)
Bastantes	1.40	0.43	0.001	4.6 (1.8, 9.4)

- a possibilidade de um aluno ter mais dificuldades nos centros de Amaranate, Arcos de Valdevez, Marinha Grande, Ermesinde, Trofa e Oliveira de Azeméis, aumenta aproximadamente 1.8 vezes em relação aos centros de Lisboa, Porto, Torres Vedras, Caldas da Rainha e Santarém.
- Um formando do sexo feminino tem o triplo (1/0.3) das possibilidades de ter mais dificuldades que um do sexo masculino.



- Os formandos residentes em freguesias rurais têm o dobro (1/0.5) das possibilidades de ter mais dificuldades que os que residem em freguesias urbanas.
- Quanto às causas das dificuldades escolares, verifica-se que os alunos que afirmam que têm muita falta de dedicação às atividades de estudo têm aproximadamente 5 vezes mais possibilidades de terem dificuldades em compreender os conteúdos programáticos do que os alunos que afirmam que têm falta de dedicação às actividades de estudo.
- Verifica-se ainda que os alunos que afirmam que a falta de bases tem muita influência no seu desempenho escolar têm aproximadamente 6 vezes mais possibilidades de apresentarem dificuldades em compreender os conteúdos programáticos do que os alunos que acham que a falta de bases não tem nenhuma influência.
- Relativamente ao ambiente familiar os alunos que afirmam que os problemas familiares têm muita influência no seu desempenho escolar têm 5 vezes mais possibilidades de apresentarem dificuldades em compreender os conteúdos programáticos do que os alunos que acham que os problemas familiares não têm nenhuma influência.

Para perceber se a variável dificuldade em compreender os conteúdos programáticos poderia ser explicada de forma diferente consoante uma abordagem como resposta binária ou como resposta ordinal, construíram-se dois modelos distintos em que a variável resposta se codificou da seguinte forma: Modelo 1 - pouca ou nenhuma dificuldade em compreender os conteúdos programáticos (0) e muita dificuldade em compreender os conteúdos programáticos (1); Modelo 2 - nenhuma dificuldade em compreender os conteúdos programáticos (0) e muita dificuldade em compreender os conteúdos programáticos (1). Note-se que ao passo que no modelo 1 separámos os alunos com muitas dificuldades daqueles que não as têm ou têm algumas, no modelo 2 simplesmente eliminámos a categoria *pouca dificuldade* da variável para se poder estudar os fatores que levam os alunos a estarem em condições completamente opostas.

Na fase de modelação constatou-se que algumas variáveis apresentavam coeficientes semelhantes nas diferentes categorias, pelo que foi possível agrupar algumas categorias dessas variáveis por não serem significativamente diferentes. Verificou-se que ambos os modelos apresentaram uma muito boa capacidade discriminativa (ambos os modelos apresentaram uma AUC de 0.82), registando-se no teste de bondade de ajustamento de Hosmer e Lemeshow valores de prova de 0.86 ( $\chi^2_{(7)} = 3.5$ ) para o modelo 1 e de 0.41 ( $\chi^2_{(7)} = 7.2$ ) para o modelo 2. Na Tabela 4 encontram-se as estimativas finais para os dois modelos de regressão logística considerados. Em ambos os modelos a variável núcleo não se mostrou significativa, sendo substituída pela localização geográfica como fator determinante na frequência dos cursos do centro (por exemplo, para o modelo 2 que compara os dois extremos do

Tabela 4: Estimativas dos parâmetros ( $\hat{\beta}$ ) e respetivos desvios-padrão ( $SE(\hat{\beta})$ ) e valores  $p$  associados dos dois modelos de regressão logística para as dificuldades em compreender os conteúdos programáticos.

Variável	Modelo 1			Modelo 2		
	$\hat{\beta}$	$SE(\hat{\beta})$	valor $p$	$\hat{\beta}$	$SE(\hat{\beta})$	valor $p$
Constante	-2.14	0.81	0.008	-1.37	1.05	0.191
<b>Sexo</b>						
Feminino						
Masculino	-1.66	0.64	0.010	-1.97	0.93	0.034
<b>Freg. Residência</b>						
Rural						
Urbana	-0.89	0.38	0.018	-1.32	0.48	0.006
<b>Falta ded. estudo</b>						
Nenhuma ou alguma						
Muita	1.19	0.41	0.004	1.79	0.62	0.004
<b>Falta de bases</b>						
Nenhuma						
Alguma	1.37	0.56	0.02	2.22	0.62	<0.001
Muita	2.30	0.60	<0.001	3.43	0.71	<0.001
<b>Probl. familiares</b>						
Nenhuns						
Alguns ou bastantes	1.43	0.54	0.008	2.07	0.77	0.008
<b>Loc. geográfica</b>						
Nada ou pouco imp.						
Razoavel. imp.	-1.02	0.42	0.016	-0.46	0.52	0.39
Imp. ou muito imp.	0.42	0.42	0.314	1.16	0.60	0.051

tipo de aluno, o aluno que considera que a localização geográfica é importante ou muito importante tem 3 vezes mais possibilidades de ser um aluno com muitas dificuldades do que um aluno que considera que a localização geográfica é nada ou pouco importante). Todas as restantes variáveis se mantiveram no modelo (embora por vezes com menos categorias) quando comparadas com o modelo ordinal.

Comparando os dois modelos verifica-se que quando a classe intermédia do grau de dificuldade é eliminada, isto é, quando comparamos os dois extremos do tipo de alunos, os efeitos de praticamente todas as variáveis acentuam as dificuldades dos alunos, tal como seria de esperar.

## 5 Conclusão

Neste trabalho pretendeu-se estudar os fatores que levam os formandos do CENFIM a escolher este centro de formação, quais as suas motivações e as razões para o sucesso escolar. Para tal, construiu-se um inquérito por questionário que se distribuiu pelos centros do CENFIM. Da sua análise foi

possível caracterizar os formandos e testar algumas associações que contêm informação importante para a direção do Centro.

Usando um modelo de regressão logística ordinal estudaram-se os fatores que podem condicionar o sucesso escolar dos formandos do CENFIM. Este modelo permitiu concluir que os alunos do sexo feminino, que residem num meio rural, com falta de dedicação ao estudo e falta de bases relativamente a anos anteriores manifestam mais dificuldades na aprendizagem. Além disso, concluímos também que os alunos com um mau ambiente familiar tendem a apresentar dificuldades que podem levar ao insucesso escolar.

Para perceber se a variável dificuldade em compreender os conteúdos programáticos poderia ser explicada de forma diferente consoante uma abordagem como resposta binária ou como resposta ordinal, construíram-se dois modelos logísticos. Num dos modelos separámos os alunos com muitas dificuldades, daqueles que não as têm ou têm algumas dificuldades, e no outro simplesmente eliminámos a categoria *pouca dificuldade* para se poder estudar os fatores que levam os alunos a estarem em condições opostas. Verificámos que quer usando um modelo ordinal ou os modelos logísticos, as variáveis significativas são idênticas (embora por vezes com menos categorias) e apenas a variável núcleo foi substituída pela variável localização geográfica.

## Referências

- [1] Abreu, M., Siqueira, A., Caiaffa, W. (2009). Regressão logística ordinal em estudos epidemiológicos. *Rev. Saúde Pública*, 43(1),183-194.
- [2] Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. John Wiley and Sons, New Jersey, 2nd Ed..
- [3] Barros, A. (1992). Atribuições Causais e Expectativas de Controlo do desempenho na matemática. Tese de Doutoramento. Universidade do Minho.
- [4] Hosmer, D., Lemeshow, S. (2000). *Applied Logistic Regression*. John Wiley and Sons, New York.
- [5] Neves, L. (2002). Um estudo sobre as relações entre a percepção e as expectativas dos professores e dos alunos e o desempenho em matemática. Tese de Mestrado. Universidade Estadual de Campinas, Faculdade de Educação.
- [6] R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- [7] Rebelo, H., Bonito, J., Oliveira, M., Cid, M., Saraiva, M. (2009). Do Rendimento académico ao grau de satisfação dos estudantes: a visão sobre a qualidade na transição do ensino secundário para o ensino superior. *Actas do X Congresso Internacional Galego-Português de Psicopedagogia*, 2971-2932, Universidade do Minho.