



ESCOLA DE CIÊNCIAS E TECNOLOGIA - DEPARTAMENTO DE MATEMÁTICA

LITERACIA MATEMÁTICA DO ALUNO PORTUGUÊS SEGUNDO PISA

Sónia Carla Cabrita da Silva de Vasconcelos Barbosa

Dissertação apresentada à Universidade de Évora para obtenção do grau de
Mestre em Modelação Estatística e Análise de Dados

Orientador: Prof. Doutor Paulo Infante

Universidade de Évora
Évora, setembro de 2012



ESCOLA DE CIÊNCIAS E TECNOLOGIA - DEPARTAMENTO DE MATEMÁTICA

LITERACIA MATEMÁTICA DO ALUNO PORTUGUÊS SEGUNDO PISA

Sónia Carla Cabrita da Silva de Vasconcelos Barbosa

Dissertação apresentada à Universidade de Évora para obtenção do grau de
Mestre em Modelação Estatística e Análise de Dados

Orientador: Prof. Doutor Paulo Infante

Universidade de Évora
Évora, setembro de 2012

Agradecimentos:

Esta dissertação de mestrado não é resultado de um trabalho solitário, só foi possível graças ao contributo e colaboração de várias pessoas às quais gostaria de exprimir palavras de agradecimento:

ao meu orientador Prof. Doutor Paulo Infante pela disponibilidade durante todo o trabalho, pela exigência de método e rigor, pela revisão muito crítica de todos os textos e pelo apoio amigo nas horas de dúvidas e incertezas;

à minha amiga Elsa Barbosa pela preciosa ajuda na escolha do objecto de estudo, e pela sua incansável ajuda nos últimos dois anos;

à Prof. Doutora Ana Ferreira do Gabinete de Estatística e Planeamento de Educação pela disponibilização de material indispensável à realização desta dissertação;

aos Analista Maciej Jakubowski e Ikeda Miyako da Organização para a Cooperação e Desenvolvimento por todos os esclarecimentos prestados;

aos meus amigos por existirem para todas as horas;

à minha colega de Mestrado Vera que juntas ultrapassamos dificuldades, e quem sempre me apoiou nas horas difíceis;

à minha querida avó, mãe, Zé Luís e irmãos por acreditarem em mim e apoiarem durante todo o processo;

por último, mas não menos importante, aos três homens da minha vida, Miguel, Henrique e Francisco pelo apoio incondicional em todas as horas e pelos diversos sacrifícios suportados.

Índice

Índice de Figuras	vii
Índice de Tabelas	ix
Resumo:	xiii
Abstract – Mathematical Literacy Student Portuguese by PISA.....	xv
Capítulo 1: Introdução do Tema	1
1.1. Enquadramento do tema	1
1.2. Objetivo da dissertação	4
Capítulo 2: Modelação Estatística de respostas nominais e ordinais.....	5
2.1. Modelos Lineares Generalizados	5
2.1.1. Modelo de regressão logística	6
2.1.2. Seleção das covariáveis	9
2.1.3. Diagnóstico do modelo de Regressão logística múltipla.....	10
2.1.4. Avaliação do modelo de Regressão logística múltipla	13
2.1.5. Interpretação dos coeficientes do modelo de Regressão logística múltipla	16
2.2. Modelo de Regressão logística ordinal	18
2.2.1. Modelos de categorias adjacentes	18
2.2.2. Modelos de razão contínua	19
2.2.3. Modelos de odds proporcionais	19
2.3. Modelo de Regressão logística multinomial.....	20
2.4. Teste de Cochran- Mantel-Haenszel e teste Breslow-Day	23
Capítulo 3: Análise de dados - área espaço e forma e área da quantidade.....	25
3.1. Análise da percentagem de acertos dos alunos às questões - questionário dos alunos.....	25
3.1.1. Distribuição das respostas pelo País	26
3.1.2. Tipologia das escolas	26
3.1.3. Ano de escolaridade.....	27
3.1.4. Ciclo de estudos que frequenta	28
3.1.5. Género	28
3.1.6. Frequência do ensino pré-escolar.....	30
3.1.7. Repetências durante o percurso de estudo	30
3.1.8. Material de apoio ao estudo	31
3.1.9. Hábitos de leitura fora da escola	33

3.1.10. Metodologia de estudo	34
3.1.11. Frequência de aulas de enriquecimento ou recuperação a Matemática	35
3.1.12. Situação profissional dos progenitores	36
3.2. Análise da percentagem de acertos dos alunos às questões - questionário aos Pais	37
3.2.1. Tempo dedicado à leitura pelos pais	37
3.2.2. Atividades com os filhos	38
3.2.3. Habilitações literárias dos pais	40
3.2.4. Rendimento familiar do agregado familiar	42
3.2.5. Número de filhos	43
3.2.6. Participação dos pais em atividades escolares.....	43
3.3. Análise da percentagem de acerto dos alunos às questões - por Países.....	44
3.3.1. Género	46
3.3.2. Frequência do ensino pré-escolar.....	47
3.3.3. Repetências durante o percurso de estudo	48
3.3.4. Uso de software de programas educativos	49
3.3.5. Hábitos de leitura fora da escola	50
3.4. Comparação entre os alunos Portugueses e os alunos dos países Europeus	53
Capítulo 4: Definição de perfis – área de conteúdos espaço e forma e área de conteúdos quantidade	54
4.1. Perfil do aluno: área espaço e forma	54
4.1.1. Consideração ao perfil do aluno – área espaço e forma.....	62
4.2. Perfil do aluno : área da quantidade.....	64
4.3. Perfil do agregado familiar: área espaço e forma.....	67
4.3.1. Consideração ao perfil dos agregados familiares – área espaço e forma.....	71
4.4. Perfil do agregado familiar: área da quantidade	73
4.5. Modelação conjunta das respostas dadas nas duas questões – área espaço e forma e área da quantidade.....	75
4.5.1. Análise da percentagem de acertos com questionário dos alunos.....	75
4.5.2. Perfil do aluno - área espaço e forma e área da quantidade	77
4.5.3. Considerações do modelo logístico ordinal - inquérito ao aluno.....	81
Capítulo 5: Considerações Finais.....	82
<i>Referências Bibliograficas.....</i>	<i>84</i>
ANEXOS.....	86
<i>Anexo 1 - Descritores por nível de proficiência na área espaço e forma e descritores por nível de proficiência na área da quantidade</i>	<i>86</i>
<i>Anexo 2 - Exemplos de questões teste PISA - área da incerteza.....</i>	<i>90</i>

<i>Anexo 3 - Questionário do aluno e questionário dos pais</i>	<i>93</i>
<i>Anexo 4 – Passos da modelação modelo logístico</i>	<i>107</i>
<i>Anexo 5 - Modelo alunos na área espaço e forma - univariada</i>	<i>119</i>
<i>Anexo 6 - Análise resíduos do modelo dos alunos na área espaço e forma.....</i>	<i>122</i>
<i>Anexo 7 - Análise das observações modelo logístico área espaço e forma - modelo do aluno.....</i>	<i>123</i>
<i>Anexo 8 - Validação do modelo dos alunos na área espaço e forma</i>	<i>124</i>
<i>Anexo 9 - Interpretação das variáveis modelo alunos na área espaço e forma.....</i>	<i>125</i>
<i>Anexo 10 - Exemplo perfil dos alunos na área espaço e forma</i>	<i>130</i>
<i>Anexo 11 - Modelo alunos na área da quantidade - univariada</i>	<i>131</i>
<i>Anexo 12 - Modelo pais na área espaço e forma - univariada</i>	<i>134</i>
<i>Anexo 13 - Análise resíduos do modelo dos pais na área espaço e forma.....</i>	<i>138</i>
<i>Anexo 14 - Análise das observações modelo logístico área espaço e forma - modelo dos pais</i>	<i>139</i>
<i>Anexo 15 - Validação do modelo dos pais na área espaço e forma</i>	<i>140</i>
<i>Anexo 16 - Interpretação das variáveis modelo pais na área espaço e forma.....</i>	<i>141</i>
<i>Anexo 17 - Exemplo perfil dos pais na área espaço e forma</i>	<i>143</i>
<i>Anexo 18 - Modelo agregado familiar na área da quantidade - univariada</i>	<i>144</i>
<i>Anexo 19 – Passos da modelação modelo ordinal.....</i>	<i>148</i>
<i>Anexo 20 - Análise dos resíduos – possibilidades proporcionais</i>	<i>150</i>
<i>Anexo 21 - Teste qui-quadrado de Pearson - Países</i>	<i>152</i>

Índice de Figuras

<i>Figura 3.1 - Proporção de alunos que responderam corretamente à questão por zona geográfica.....</i>	<i>26</i>
<i>Figura 3.2 - Proporção de alunos que responderam corretamente à questão por escola pública e privada.....</i>	<i>27</i>
<i>Figura 3.3 - Proporção de alunos que responderam corretamente à questão por ano de escolaridade</i>	<i>27</i>
<i>Figura 3.4 - Proporção de alunos que responderam corretamente à questão por tipologia de ensino .</i>	<i>28</i>
<i>Figura 3.5 - Proporção de alunos que responderam corretamente à questão por gênero.....</i>	<i>28</i>
<i>Figura 3.6 - Proporção de alunos que responderam corretamente à questão por frequência no ensino pré-escolar</i>	<i>30</i>
<i>Figura 3.7 - Proporção de alunos que responderam corretamente à questão por repetências no ensino básico.....</i>	<i>30</i>
<i>Figura 3.8 - Proporção de alunos que responderam corretamente à questão na questão da área de espaço e forma.....</i>	<i>31</i>
<i>Figura 3.9 - Proporção de alunos que responderam corretamente à questão na área da quantidade</i>	<i>32</i>
<i>Figura 3.10 - Proporção de alunos que responderam corretamente à questão por número de livros em casa.....</i>	<i>33</i>
<i>Figura 3.11 - Proporção de alunos que responderam corretamente à questão por tempo diário dedicado à leitura.....</i>	<i>33</i>
<i>Figura 3.12 - Proporção de alunos que responderam corretamente à questão por metodologia de estudo na questão da área espaço e forma</i>	<i>34</i>
<i>Figura 3.13 - Proporção de alunos que responderam corretamente à questão por metodologia de estudo da área da quantidade.....</i>	<i>34</i>
<i>Figura 3.14 - Proporção de alunos que responderam corretamente à questão por frequência de aulas suplementares.....</i>	<i>35</i>
<i>Figura 3.15 - Proporção de alunos que responderam corretamente à questão por situação profissional da mãe.....</i>	<i>36</i>
<i>Figura 3.16 - Proporção de alunos que responderam corretamente à questão por situação profissional do pai.....</i>	<i>36</i>
<i>Figura 3.17 - Proporção de alunos que responderam corretamente à questão por tempo dedicado á leitura</i>	<i>37</i>

<i>Figura 3.18 - Proporção de alunos que responderam corretamente à questão por atividades com os filhos na área espaço e forma.....</i>	<i>39</i>
<i>Figura 3.19 - Proporção de alunos que responderam corretamente à questão por atividades com os filhos na área da quantidade.....</i>	<i>40</i>
<i>Figura 3.20 - Proporção de alunos que responderam corretamente à questão por habilitações literárias do pai</i>	<i>41</i>
<i>Figura 3.21 - Proporção de alunos que responderam corretamente à questão por habilitações literárias da mãe</i>	<i>41</i>
<i>Figura 3.22 - Proporção de alunos que responderam corretamente à questão por rendimento bruto anual do agregado familiar</i>	<i>42</i>
<i>Figura 3.23 - Proporção de alunos que responderam corretamente à questão por gastos anual com o filho.....</i>	<i>43</i>
<i>Figura 3.24 - Proporção de alunos que responderam corretamente à questão por número de filhos ..</i>	<i>43</i>
<i>Figura 3.25 - Proporção de alunos que responderam corretamente à questão por participação na escola dos encarregados de educação:.....</i>	<i>44</i>
<i>Figura 3.26 - Proporção de acertos na questão da área espaço e forma</i>	<i>45</i>
<i>Figura 3.27 - Proporção de acertos na questão da área da quantidade</i>	<i>46</i>
<i>Figura 3.28 - Proporção de alunos que responderam corretamente à questão por países segundo o seu género na área espaço e forma</i>	<i>46</i>
<i>Figura 3.29 - Proporção de alunos que responderam corretamente à questão por países segundo o seu género na área da quantidade</i>	<i>47</i>
<i>Figura 3.30 - Proporção de alunos que responderam corretamente por frequência do ensino pré-escolar na questão da área espaço e forma.....</i>	<i>47</i>
<i>Figura 3.31 - Proporção de alunos que responderam corretamente por frequência do ensino pré-escolar na questão da área da quantidade</i>	<i>48</i>
<i>Figura 3.32 - Proporção de alunos que responderam corretamente por repetências na questão da área espaço e forma.....</i>	<i>48</i>
<i>Figura 3.33 - Proporção de alunos que responderam corretamente por repetências na questão da área da quantidade.....</i>	<i>49</i>
<i>Figura 3.34 - Proporção de alunos que responderam corretamente por uso de programas software educativos na questão da área de espaço e forma</i>	<i>49</i>

<i>Figura 3.35 - Proporção de alunos que responderam corretamente por uso de programas software educativos na questão da área da quantidade</i>	<i>50</i>
<i>Figura 3.36 - Proporção de alunos que responderam corretamente por número de livros em casa à questão da área de espaço e forma.....</i>	<i>50</i>
<i>Figura 3.37 - Proporção de alunos que responderam corretamente por tempo diário dedicado à leitura à questão da área espaço e forma.....</i>	<i>51</i>
<i>Figura 3.38 - Proporção de alunos que responderam corretamente por número de livros em casa à questão da área da quantidade.....</i>	<i>51</i>
<i>Figura 3.39 - Proporção de alunos que responderam corretamente por tempo diário dedicado à leitura à questão da área da quantidade.....</i>	<i>52</i>
<i>Figura 3.40 - Proporção de alunos que responderam corretamente por tempo dedicado à leitura por país</i>	<i>52</i>
<i>Figura 4.1. - Curva de ROC modelo área da quantidade – questionário aos alunos</i>	<i>66</i>
<i>Figura 4.2. - Curva de ROC modelo área da quantidade – questionário aos país</i>	<i>74</i>
<i>Figura 4.3 - Percentagem de respostas às questões</i>	<i>75</i>
<i>Figura 1 - Curva de ROC modelo ajustado – alunos área espaço e forma</i>	<i>124</i>
<i>Figura 2 - Curva de ROC para validação do modelo – alunos área espaço e forma.....</i>	<i>124</i>
<i>Figura 3 - Curva de ROC modelo ajustado – país área espaço e forma</i>	<i>140</i>
<i>Figura 4 - Curva de ROC para validação do modelo – país área espaço e forma.....</i>	<i>140</i>
<i>.....</i>	<i>140</i>
<i>Figura 5 - Resíduos score das covariável modelo logística ordinal</i>	<i>150</i>
<i>Figura 6 - Resíduos parcial das covariável modelo logística ordinal.....</i>	<i>151</i>

Índice de Tabelas

<i>Tabela 2.1 - Modelos lineares generalizados</i>	<i>7</i>
<i>Tabela 2.2 - Funções de ligação.....</i>	<i>7</i>
<i>Tabela 3.1 - Resultados obtidos por alguns países da união europeia a nível do teste PISA a matemática.....</i>	<i>44</i>
<i>Tabela 3.2 - Amostra do estudo às questões na área espaço e forma e na área da quantidade por países.....</i>	<i>45</i>

<i>Tabela 4.1 - Coeficientes do modelo para o acerto à questão na área espaço e forma-questionário aos alunos, desvio padrão e valor p teste Wald correspondentes.....</i>	<i>56</i>
<i>Tabela 4.2 - Coeficientes do modelo para o acerto na área espaço e forma – questionário aos pais, desvio padrão e valor p teste Wald correspondentes.....</i>	<i>68</i>
<i>Tabela 4.3 - Coeficientes do modelo ordinal para o acerto às questão na área espaço e forma e na área da quantidade, desvio padrão e valor p teste Wald correspondentes.....</i>	<i>77</i>
<i>Tabela 4.4 - Odds ratios e intervalos de confiança a 95% do modelo ordinal para o acerto às questão espaço e forma e na área da quantidade.....</i>	<i>79</i>
<i>Tabela 1 - Os descritores de cada nível de proficiência na escala de matemática para a área de espaço e forma.....</i>	<i>86</i>
<i>Tabela 2 - Os descritores de cada nível de proficiência na escala de matemática para a área da quantidade.....</i>	<i>88</i>
<i>Tabela 3 - Coeficientes das covariáveis no modelo logístico univariado, o seu desvio padrão, valores p associados à estatística de Pearson, de Wald e da razão de verossimilhanças, odds-Ratios (OR) e respectivos intervalos de confiança a 95%. Questão na área espaço e forma- inquérito ao aluno.....</i>	<i>119</i>
<i>Tabela 4 - Coeficientes, deviance, coeficiente de determinação e teste Homer e Lemeshow do modelo sem as observações indicadas – questão área espaço e forma – questionário alunos.....</i>	<i>123</i>
<i>Tabela 5 - Coeficientes das covariáveis no modelo logístico univariado, o seu desvio padrão, valores p associados à estatística de Pearson, de Wald e da razão de verossimilhanças, odds-Ratios (OR) e respectivos intervalos de confiança a 95%. Questão na área da quantidade - inquérito ao aluno.....</i>	<i>131</i>
<i>Tabela 6 - Coeficientes das covariáveis no modelo logístico univariado, o seu desvio padrão, valores p associados à estatística de Pearson, de Wald e da razão de verossimilhanças, odds-Ratios (OR) e respectivos intervalos de confiança a 95%. Questão na área espaço e forma- inquérito aos pais.....</i>	<i>134</i>
<i>Tabela 7- Coeficientes, deviance, coeficiente de determinação e teste Homer e Lemeshow do modelo sem as observações indicadas – questão área espaço e forma – questionário pais.....</i>	<i>139</i>
<i>Tabela 8 - Coeficientes das covariáveis no modelo logístico univariado, o seu desvio padrão, valores p associados à estatística de Pearson, de Wald e da razão de verossimilhanças, odds-Ratios (OR) e respectivos intervalos de confiança a 95%. Questão na área da quantidade- inquérito aos pais.....</i>	<i>144</i>
<i>Tabela 9 - Teste qui-quadrado – acerto à questão na área espaço e forma versus número de livros em casa.....</i>	<i>152</i>
<i>Tabela 10 - Teste qui-quadrado – acerto à questão na área da quantidade versus número de livros em casa.....</i>	<i>152</i>

<i>Tabela 11 - Teste qui-quadrado – acerto à questão na área espaço e forma versus tempo dedicado à leitura</i>	<i>152</i>
<i>Tabela 12 - Teste qui-quadrado – acerto à questão na área da quantidade versus tempo dedicado à leitura</i>	<i>152</i>

Resumo:

O Projeto PISA (*Programme for International Student Assessment*) foi introduzido pela Organização para o Desenvolvimento e Cooperação Económica (OCDE) em 1997 e tem como objetivo avaliar a capacidade dos jovens de 15 anos no uso dos seus conhecimentos, de forma a enfrentarem os desafios da vida real. Na matemática a literacia é definida como a capacidade de um aluno identificar e compreender o papel da disciplina nos desafios da vida real, de fazer julgamentos bem fundamentados enquanto cidadão construtivo, preocupado e reflexivo. Neste trabalho procuramos caracterizar os alunos com nível de proficiência elevado na área espaço e forma e na área da quantidade. Para tal, recorreremos aos dados dos questionários de alunos e pais, relacionando as variáveis destes questionários com o desempenho dos alunos, usando a regressão logística como ferramenta estatística principal. Também foi elaborado um estudo recorrendo à regressão logística ordinal no sentido de explorar mais os alunos que responderam corretamente a uma das questões, às duas questões ou pelo contrário não obtiveram sucesso a nenhuma das questões.

Este estudo permitiu encontrar fatores potenciadores de um bom desempenho nas áreas em estudo, como o fato do aluno viver na zona Norte ou Centro de Portugal, ser rapaz ou mesmo possuir acesso a livros. Constatou-se também que diferentes metodologias de estudo implicavam diretamente no desempenho. Apesar dos itens colocados aos alunos não incidirem diretamente sobre conteúdos programáticos, os alunos que apresentaram repetências durante o percurso escolar apresentaram dificuldades acrescidas.

Foram estudados os resultados obtidos pelos alunos de alguns países da União Europeia nas duas questões, e foram comparados os resultados a nível global e individual, tendo-se constatado que, os fatores que influenciam positivamente o aluno em Portugal não são significativos nos diferentes países.

Pretendeu-se, como objetivo final, caracterizar o aluno de 15 anos que estuda em Portugal, no nível de proficiência elevado na área espaço e forma e na área da quantidade e assim poder encontrar um instrumento de trabalho útil para a definição e/ou refinamento das políticas educativas no sentido de melhorar a preparação dos alunos para a vida futura.

Palavras-chave: teste de PISA, literacia matemática, regressão logística, regressão ordinal.

Abstract – Mathematical Literacy Student Portuguese by PISA

The PISA (Programme for International Student Assessment) Project was introduced by OCDE in the year 1997, with the objective of assess young people (15 years old) capacity of using their knowledge to face real life challenges. In Mathematics, literacy is defined as the student capacity to identify and understand the importance of this Subject in the world performance, to make well based judgements as a constructive citizen, concerned and reflexive.

In this work, we try to characterize Students with higher proficiency in space and shape area, also in quantity area. We used Parents and Students questionnaire data, relating variables from these questionnaires with Students performances, using logistic regression as a main statistical tool.

We also carried out a study, based on ordinal logistic regression, with the objective of analyse Students who had answered correctly none, one, and two questions.

This work also allowed finding some factors of a good performance in the areas under study, like the fact of Students living in the North or Center of Portugal, be a boy, or even if they have access to books.

Also it was found that different study methodologies implied directly in student's performance. Despite of the question didn't focus directly programmatic contents, Students presented greater difficulties.

We also studied the performance, at the two questions, of students from some European Union countries, and compared the global and individual results. We conclude some that factors that positively influence student's performance in Portugal are not significant in different countries.

As a final objective, we purposed to characterize the 15 years old Portuguese student, in order to find an useful work tool to define and refine the educative policies, trying to improve students preparation for his life in the future.

Keywords: PISA test, mathematical literacy, logistic regression, ordinal regression.

Capítulo 1: Introdução do Tema

O estudo Pisa foi introduzido pela Organização para a Cooperação e Desenvolvimento Económico (OCDE). Os resultados obtidos permitem monitorizar regularmente os sistemas educativos tendo em conta o desempenho dos alunos. Assim, irá ser dada uma introdução acerca do aparecimento do mesmo, áreas em estudo, e objetivos.

1.1. Enquadramento do tema

Em 30 de Setembro de 1961 foi criada a Organização para a Cooperação e Desenvolvimento Económico, sucedendo à Organização para a Cooperação Económica Europeia (OECE). Esta organização é composta por 30 países, sendo Portugal membro desde o início da sua fundação. Trata-se de uma organização onde os Governos podem comparar e trocar experiências, identificar boas práticas e promover decisões e recomendações. O seu grande objetivo é o de auxiliar Governos e Sociedade a aproveitarem plenamente as vantagens da globalização, fazendo face aos desafios económicos, sociais e de governação que acompanham aquele fenómeno.

Segundo HENRY(2001), a organização OCDE adquiriu o estatuto de “internacional mediator of knowledge and global policy actor”, uma vez que a mesma é globalizadora tendo um papel fundamental nas políticas educativas dos estados e um papel fundamental na corrente de pensamento sobre a educação à escala internacional.

Na área da educação a OCDE introduziu o Programa para a Avaliação Internacional dos Alunos, designado por PISA - *Programme for International Student Assessment*, que fornece instrumentos para auxiliar os países a desenvolver políticas educativas. Os resultados obtidos permitem monitorizar os diferentes sistemas educativos num contexto de um enquadramento conceptual aceite internacionalmente.

Este estudo realiza-se em ciclos de 3 anos e tem como objetivo, avaliar de que forma os alunos de 15 anos adquiriram conhecimentos e destrezas essenciais para uma participação na sociedade durante um percurso escolar de 9 anos e não de simplesmente avaliar o domínio que detêm sobre os conteúdos de um currículo escolar.

O primeiro ciclo dos testes PISA realizou-se no ano 2000, tendo o seu foco incidido na literacia em contexto de leitura. O segundo ciclo realizou-se em 2003, tendo este dado maior incidência à literacia Matemática, mas também foram abordados os domínios da literacia de leitura,

literacia científica e a resolução de problemas. Em 2006 realizou-se o terceiro ciclo, onde a maior incidência recaiu sob o domínio da literacia científica. No ano 2009 realizou-se o quarto ciclo, onde a predominância foi dada à literacia da leitura.

O primeiro estudo do Pisa envolveu 31 países, este número tem vindo a aumentar substancialmente, tendo participado 65 países no PISA 2009. Este aumento da adesão dos países ao estudo vem reforçar a importância do mesmo como instrumento de trabalho para os países.

O estudo do PISA possui uma área que tem como base a literacia matemática em PISA (2003), a literacia matemática é definida como a capacidade de um indivíduo identificar e compreender o papel que a matemática desempenha num mundo real, elaborar opiniões bem fundamentadas, usar e desenvolver-se com a matemática, de maneira a satisfazer as necessidades da sua vida, enquanto cidadão construtivo, preocupado e reflexivo.

A definição de literacia matemática no PISA provém da teoria sobre a estrutura e o uso da língua, realizada em estudos socioculturais sobre literacia. O termo literacia refere o uso humano da linguagem. A capacidade de ler, escrever, ouvir e falar uma língua é o instrumento mais importante na medição de qualquer atividade social humana.

Após investigação efetuada, PISA (2003), a avaliação a Matemática foi estabelecida tendo como base quatro áreas de conteúdos que apresentamos de seguida.

Espaço e forma - área relacionada com:

- a) os conteúdos lecionados no currículo de geometria, envolvendo relações espaciais e geométricas;
- b) são abordadas temáticas relacionadas com a procura de semelhanças e diferenças em formas e o seu reconhecimento em diferentes representações e dimensões;
- c) a compreensão das propriedades dos objetos e as suas posições relativas.

Transformações e relações - área relacionada com:

- a) os conteúdos lecionados no currículo de Álgebra;
- b) as questões envolvem transformações, relações funcionais e dependência entre variáveis;
- c) as relações matemáticas exprimem-se sob a forma de equações, inequações, equivalência, divisibilidade entre outros;
- d) representações como as simbólicas, algébricas, gráficas, tabulares e geométricas;
- e) mudança de representação.

Quantidade - área relacionada com:

- a) os conteúdos lecionados no currículo da aritmética;
- b) fenómenos numéricos, relações e padrões quantitativos;
- c) o tratamento do raciocínio quantitativo, que envolve o sentido de número, a representação do número, a compreensão do significado das operações, o cálculo mental e a estimativa.

Incerteza - área relacionada com:

- a) os conteúdos lecionados no currículo da estatística e das probabilidades;
- b) fenómenos probabilísticos e estatísticos;
- c) relações probabilísticas e estatísticas.

Em 2003 e 2009 foram utilizados 85 itens da Matemática com diferentes níveis de dificuldade. Aos itens foram atribuídos pontos tendo em consideração a dificuldade do item e a percentagem de respostas dadas pelos alunos. A escala do Pisa ficou, desta forma, subdividida em seis níveis de proficiência (Anexo 1) consoante as áreas de conteúdo em estudo.

Os itens foram subdivididos em cadernos, onde cada aluno recebeu quatro cadernos, organizados em blocos de meia hora, perfazendo um total de duas horas de prova. Na área da Matemática cada aluno respondeu a um subconjunto de tarefas retirado das 85 tarefas matemáticas disponíveis, existindo 13 conjuntos de tarefas diferenciados.

Com esta metodologia o nível de competência relativa de um aluno que respondeu a um determinado caderno pode ser estimado tendo como base a proporção de itens que respondeu corretamente. Assim, o modelo matemático utilizado para analisar os dados resultantes do teste de PISA, foi obtido através de processos iterativos que permitiram estimar simultaneamente a probabilidade do aluno responder corretamente a um item e a probabilidade de um item ser respondido por um conjunto de alunos. Este processo gerou 5 variáveis de valores plausíveis que permitem obter estimadores da população bastante robustos. No entanto, esses dados não podem ser considerados para a avaliação individual por aluno, como está descrito por Zoanetti (2001).

Com as limitações encontradas no manuseamento da base de dados de não ser possível encontrar a cotação total do teste por aluno, optou-se por realizar o perfil dos alunos que responderam às questões, *M406Q01 - nível 5 (634,5 pontos)* de proficiência dentro da área espaço e forma e a questão *M603Q02T - nível 4 (602,6 pontos)* de proficiência dentro da área quantidade.

Outras limitações encontradas prenderam-se com o fato de nem todos os alunos responderam aos mesmos itens, uma vez que os testes estão divididos em conjuntos com diferentes questões, originando a existência de alunos que não responderam a questões com este grau de dificuldade.

Para situar as tipologias das questões das áreas temáticas, foram disponibilizados pelo grupo de trabalho do PISA a título de exemplo algumas atividades (Anexo 2), uma vez que as atividades realizadas pelos alunos são sigilosas.

1.2. Objetivo da dissertação

O principal objetivo da dissertação, é a construção de modelos estatísticos que relacionem o desempenho obtido pelos alunos em questões de nível elevado de proficiência na escala de matemática segundo o relatório técnico do PISA 2009 na área de conteúdos espaço e forma e na área da quantidade, com as variáveis do questionário do aluno e do questionário dos pais, procurando encontrar as variáveis estatisticamente significativas e quantificar o seu impacto individual e global.

Para traçar o perfil dos alunos que responderam às questões teve-se em conta as respostas dadas ao questionário do aluno (Anexo 3) sobre si próprio, sobre hábitos de aprendizagem e as suas perceções no contexto de aprendizagem. Numa segunda fase, traçou-se o perfil dos agregados familiares, recorrendo a um questionário elaborado pelos pais (Anexo 3) sobre as suas características, hábitos de leitura, perceção e envolvimento destes na escola.

A dissertação será subdividida em 5 capítulos, no primeiro capítulo é elaborada uma introdução aos testes PISA e a sua aplicação, bem como os objetivos da dissertação. O apoio teórico ao trabalho é desenvolvido no capítulo dois. No capítulo três é realizada uma análise aos dados referentes aos questionários dos alunos e dos pais em Portugal e realização de um estudo comparativo dos resultados obtidos às questões em Portugal, Espanha, Grécia, Reino Unido, França e Itália. Seguindo-se o capítulo quatro, onde é elaborado um estudo no sentido de encontrar o perfil de alunos e agregados familiares para o sucesso nas questões das áreas em estudo. Por último no capítulo cinco serão tecidas as considerações finais.

As bases de dados utilizadas na tese foram obtidas em Database – PISA 2009 (*for professional researchers*) em <http://pisa2009.acer.edu.au/>.

Capítulo 2: Modelação Estatística de respostas nominais e ordinais

O trabalho a ser desenvolvido nesta dissertação, irá em grande parte recair na construção de modelos adequados e parcimoniosos que estabeleçam relações entre os acertos às questões em estudo e variáveis aleatórias predictoras que constam nos questionários realizado pelos alunos e pelos pais, analisando a influência de cada uma por si e em conjunto. O trabalho desenvolvido está assente na modelação estatística de respostas nominais e ordinais. Serão desenvolvidas quatro secções como apoio teórico: regressão logística, regressão ordinal, regressão multinomial e testes de comparação para grupos estratificados. Para outros desenvolvimentos e apoio nestas matérias veja-se, por exemplo, *McCullagh e Nelder (1983)*, *Hosmer e Lemeshow (2000)*, *Turkman e Silva (2000)* e *Agresti (2002)*.

Para aplicação das regressões recorreu-se a métodos iterativos com recurso a *software*. Optou-se por usar o programa R Project (R Development Core Team, 2011).

2.1. Modelos Lineares Generalizados

Um modelo de regressão pretende relacionar uma variável resposta com variáveis explicativas.

Desde o início do século XIX até meados do século XX a aplicação dos modelos lineares normais criados por Legendre e Gauss dominaram a modelação estatística, no entanto no sentido de dar resposta a situações que não eram satisfatórias para os modelos lineares normais foram desenvolvidos modelos não lineares ou não normais, entre eles por exemplo os modelos *complementares log-log* por McCullagh e Nelder (1989) e Lindsey (1997), os modelos *probit* (Bliss, 1935) e *logit* Berkson (1944) usados para proporções, Dyke and Patterson (1952), Rasch (1960), os modelos *log-lineares* Birch (1963) usados para dados de contagem, os modelos de regressão para análise de sobrevivência Feigl e Zelen (1965), Zippin e Armitage (1966) e Glasser (1967).

Os modelos possuem uma estrutura de regressão linear, seguindo a variável resposta uma distribuição dentro de uma famílias de distribuições, a família exponencial.

Definição:

Diz-se que, uma variável Y tem **distribuição** pertencente à **família exponencial** se a sua função densidade de probabilidade se puder escrever na forma:

$$f(y|\theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\}$$

com θ e ϕ parâmetros escalares, e $a(\cdot)$, $b(\cdot)$ e $c(\cdot, \cdot)$ funções reais conhecidas.

θ – forma canónica do parâmetro de localização.

ϕ – parâmetro de dispersão.

E com $E(Y) = b'(\theta)$ e $var(Y) = a(\phi)b''(\theta)$

Algumas distribuições que pertencem à família exponencial são: as distribuições normal, binomial, poisson e gama.

Para a distribuição binomial:

Seja Y tal que mY segue uma distribuição binomial com parâmetros m e π ,

$$Y \sim B(m, \pi)/m$$

a função massa de probabilidade dada por:

$$\begin{aligned} f(y|\pi) &= \binom{m}{ym} = \pi^m (1 - \pi)^{m-ym} \\ &= \exp \left\{ ym \ln(\pi) + m(1 - y) \ln(1 - \pi) + \ln \binom{m}{ym} \right\} \\ &= \exp \left\{ m \left(y \ln \left(\frac{\pi}{1 - \pi} \right) - \ln(1 - \pi) \right) + \ln \binom{m}{ym} \right\} \end{aligned}$$

Com $y \in \left\{ 0, \frac{1}{m}, \frac{2}{m}, \dots, 1 \right\}$.

Está-se perante a família exponencial, logo a esperança matemática e a variância é dada por:

$$E(Y) = \pi \quad e \quad var(Y) = \frac{\pi(1 - \pi)}{m}$$

Sendo o parâmetro canónico a função *logit*, $\ln \left(\frac{\pi}{1 - \pi} \right)$.

A introdução dos modelos lineares generalizados por Nelder e Wedderburn (1972), veio reunir toda a informação teórica e conceptual acerca da modelação estatística, sendo estes modelos caracterizados por uma estrutura com componente aleatória e componente estrutural.

2.1.1. Modelo de regressão logística

Alguns exemplos de modelos lineares generalizados tendo em conta a componente aleatória e a componente da estrutura.

Tabela 2.1 - Modelos lineares generalizados

Componente aleatória	Componente estrutural		Modelo final
	Função ligação	Covariáveis	
Normal	identidade	Contínuas	Regressão linear
Normal	identidade	categorizadas	Análise de variância
Normal	identidade	Mistas	Análise de covariância
Binomial	logit	Mistas	Regressão logística
Poisson	logarítmica	Mistas	Log-linear

Acerca da componente estrutural:

O valor esperado μ_i está relacionado com o preditor linear $\eta_i = z_i^T \beta$ através da relação

$$\mu_i = h(\eta_i) = h(z_i^T \beta), \eta_i = g(\mu_i);$$

Onde h é uma função monótona e diferenciável, $g = h^{-1}$ é a função de ligação, β é um vetor de parâmetros de dimensão p e z_i um vetor de especificação de dimensão p função do vetor de covariáveis x_i .

A função ligação depende do tipo de variável resposta, sendo de considerar os seguintes exemplos de funções de ligação:

Tabela 2.2 - Funções de ligação

identidade	μ
recíproca	$\frac{1}{\mu}$
quadrática inversa	$\frac{1}{\mu^2}$
Raíz quadrada	$\sqrt{\mu}$
expoente	$(\mu + c_1)^2$
logarítmica	$\ln(\mu)$
logit	$\ln\left(\frac{\mu}{1-\mu}\right)$
Complementar log-log	$\ln[-\ln(1-\mu)]$
probit	$\Phi^{-1}(\mu)$

Acerca da componente de aleatoriedade:

Dado o vetor da covariáveis x_i as variáveis Y_i são independentes com distribuição pertencentes à família exponencial com $E(Y_i|x_i) = \mu_i = b'(\theta), i = 1, \dots, n$ e um parâmetro de dispersão ϕ .

Começamos por abordar a modelação de uma variável resposta dicotómica (1 – acerto à questão, 0 – não acerto à questão).

Sendo $Y_i \sim B(1, \pi_i)$, isto é,

$f(y_i|\pi_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}$ com $y_i = 0, 1$ onde cada y_i está associado a um vetor z_i , construído com as covariáveis explicativas x_i com $i = 1, \dots, n$.

Sabendo que $E(Y_i) = \pi_i$ e $\theta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$, ao fazer

$$\theta_i = \eta_i = Z_i^T \beta$$

Conclui-se que a função de ligação canónica é a função *logit*. Logo a probabilidade de sucesso $P(X = 1)$, está relacionada com o vetor z_i da seguinte forma:

$$P(X = 1) = \pi_i = \frac{\exp(Z_i^T \beta)}{1 + \exp(Z_i^T \beta)}$$

Sendo a função distribuição logística, definida pelo modelo binomial com função de ligação canónica *logit* dada por,

$$F: \mathfrak{R} \rightarrow [0,1]$$

$$F(x) = \frac{\exp(x)}{1 + \exp(x)}$$

No entanto poder-se-ia estar presentes a qualquer outra função de ligação devido ao fato de neste modelo se ter $E(Y_i) = \mu_i \in [0,1]$.

Por exemplo, uma vez que se poderá supor que a relação existente entre as probabilidades de sucesso π_i e o vetor de covariáveis é da forma $\pi_i = \phi(\eta_i) = \phi(Z_i^T \beta)$, onde $\phi(\cdot)$ é uma função de distribuição $N(0,1)$.

Obtendo-se assim uma função de ligação $g(\mu_i) = \phi^{-1}(\mu_i)$, designada por função de ligação *probit*, conduzindo a um modelo de regressão *probit*.

2.1.2. Seleção das covariáveis

A seleção das covariáveis a introduzir no modelo será feita inicialmente através de uma análise univariada de cada variável recorrendo à significância via *Pearson* (tabelas de contingência) ou via Razão de Verossimilhanças (modelo logístico univariado).

Uma atenção especial é dada às tabelas de contingência com alguma célula igual a zero. Neste caso proceder-se-á à junção ou eliminação das categorias, ou se se tratar de uma variável ordinal procede-se a sua modelação como se fosse contínua.

Ao recorrer à significância via *Pearson* pretende-se testar se duas variáveis estão associadas.

$$\begin{cases} H_0: \text{As variáveis são independentes} \\ H_1: \text{As variáveis não são independentes} \end{cases}$$

Com a estatística teste dada por,

$X^2 = \sum_{i=1}^L \sum_{j=1}^C \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, onde C representa o número de colunas, L o número de linhas da tabela e $E_{ij} = \frac{\sum_{i=1}^C O_{ij} \times \sum_{j=1}^L O_{ij}}{N}$ as frequências esperadas, sendo O_{ij} os valores observados.

Com, $X^2 \sim \chi_{(C-1)(L-1)}^2$.

Como condições de aplicabilidade do teste deverá considerar-se que as frequências esperadas em cada classe não devem ser inferiores a 5 unidades sempre que o número total de observações for inferior a 20. Caso o número de observações forem superiores a 20 não deverá existir mais do que 20% das células com frequências esperadas inferiores a 5 nem deverá existir nenhuma célula com frequência esperada inferior a 1.

Relativamente ao teste razão de verossimilhança de *Wilks*, é realizado para comparar o modelo com cada uma das covariáveis e o modelo sem essa covariável (modelo só com a ordenada na origem - modelo nulo).

$H_0: \beta_i = 0$ (o modelo não é estatisticamente significativo)

$H_1: \beta_i \neq 0$, com $i = 1 \dots n$ (o modelo é estatisticamente significativo)

Onde a estatística teste $G^2 = -2 \ln \left[\frac{L_0}{L_1} \right] \sim \chi^2(1)$ sendo que L_0 representa a verossimilhança do modelo nulo e L_1 representa a verossimilhança do modelo com a covariável a testar.

Para as variáveis que apresentam um nível de associação moderado serão estimadas as suas possibilidades bem como os respetivos intervalos de confiança.

Será tido em conta a interpretação dos valores p relativos ao teste de Wald, que pretende testar cada parâmetro β ,

$H_0: \beta_j = 0$ (o coeficiente não é estatisticamente significativo)

$H_1: \beta_j \neq 0$ (o coeficiente é estatisticamente significativo)

Onde a estatística $W = \frac{\hat{\beta}_j^2}{\sigma_{jj}} \sim N(0,1)$

Após análise univariada das variáveis, foram selecionadas todas as que apresentaram um valor de prova p inferior a 0,25, sendo estas incluídas num modelo de efeitos principais. As variáveis que não contribuem para o modelo, teste de *Wald*, com valor $p > 0,10$ serão eliminadas uma a uma e novos modelos ajustados. Estes novos modelos foram comparados sucessivamente com os seus anteriores utilizando o teste de razão de verosimilhanças. Este processo de exclusão é efetuado até a inclusão de todas as variáveis importantes e estatisticamente significativas.

Após o processo anterior são alvo de estudo as categorias das variáveis discretas, tendo em conta os valores dos seus coeficientes bem como o conhecimento que se possui acerca das mesmas, assim como a verificação do pressuposto de linearidade das variáveis contínuas.

Procede-se de seguida à interação entre os diferentes pares de variáveis, testando uma de cada vez no modelo de efeitos principais, considera-se estatisticamente significativas as que apresentaram um valor p inferior a 0,05. De seguida, cada variável não selecionada inicialmente foi reintroduzida a um novo modelo, na tentativa de encontrar variáveis que por si só, não são significativas mas dão uma contribuição importante na presença de outras variáveis.

2.1.3. Diagnóstico do modelo de Regressão logística múltipla

Ao encontrar o modelo que se ajuste aos dados em estudo, é realizado um diagnóstico ao modelo para despiste de possíveis observações discordantes no modelo. Recorrendo à revisão bibliográfica, segundo *Hosmer and Lemeshow, (2000)*, e segundo *McCullagh and Nelder (1983)*, encontramos medidas que auxiliam no estudo, assim, sucintamente:

- Resíduos de *Pearson*

A medida dos resíduos de *Pearson* corresponde à contribuição de cada observação para o modelo, sendo definida pela diferença entre as observações e as suas estimativas. Num modelo binomial este vêm dados por:

$$R_i^P = \frac{m_i^{1/2}(Y_i - \hat{\mu}_i)}{\sqrt{\widehat{var}(Y_i)}}$$

onde m_i é o número de indivíduos.

Para análise destes resíduos serão representados os valores dos resíduos de *Pearson* contra os valores estimados, permitindo identificar as observações discordantes.

Comando *R* utilizado:

```
rp<-rstandard (modelo, type="pearson")
```

- Resíduos *Deviance*

À semelhança dos resíduos de *Pearson*, os resíduos *deviance* permitem analisar as observações no sentido de encontrar as que se encontram mais afastadas.

São definidos pela expressão:

$$R_i^D = \delta_i [y_i \ln(\hat{\pi}_i) + (1 - y_i) \ln(1 - \hat{\pi}_i)]$$

Com $\delta_i = \text{sign}(y_i - \hat{\pi}_i)$

Para análise, os resíduos *deviance* são usualmente representadas contra os valores estimados, identificando as observações que se destacam.

Comando *R* utilizado:

```
rd<-residuals (modelo, type="deviance")
```

- *Leverage*

Com a análise dos resíduos *Leverage* pretende-se medir o efeito que a observação tem sobre os valores preditos, verificando-se a influência da mesma no modelo.

Seja $X_{(j)(p+1)}$ a matriz dos valores das combinações de níveis das covariáveis do modelo.

$$H = V^{\frac{1}{2}}X((X'VX))^{-1}X'V^{\frac{1}{2}}, \text{ sendo } V \text{ a matriz diagonal } J \times J$$

$$v_j = m_j \hat{\pi}(x_j) (1 - \hat{\pi}(x_j)), \quad j = 1, 2, \dots, J$$

$$h_j = m_j \hat{\pi}(1 - \hat{\pi}_j) x_j' (X'VX)^{-1} x_j' = v_j \times b_j \text{ onde } b_j = x_j' (X'VX)^{-1} x_j' \text{ e}$$

$$\text{tra}(H) = \sum_{j=1}^n h_{jj} = p$$

Se $h_j > \frac{2p}{n}$ ou $\frac{nh_j}{p} > 2$, então a observação j possui uma repercussão elevada.

Para análise dos resíduos *Leverage* serão representados graficamente os valores preditos *versus* observação, identificando as observações que se destacam.

Comando *R* utilizado:

```
h <- influence(modelo)
```

- Resíduos *DfBetas*

Para medir a influência de cada observação na estimação de cada coeficiente dos modelos calcula-se os *DfBetas*, calculado através da diferença entre a estimativa do coeficiente de regressão ajustado com todas as observações e a estimativa do coeficiente de regressão ajustado sem a observação j ,

$$DfBetas_{ij} = \hat{\beta}_i - \hat{\beta}_{i(-j)}$$

Sendo $\hat{\beta}_{i(-j)}$ a estimativa do coeficiente de regressão ajustado sem a observação j .

Para análise dos resíduos *DfBetas* serão representados graficamente os *DfBetas* de cada coeficiente *versus* as observações

Comando *R* utilizado:

```
rdf <- dfbetas(modelo)
```


- Distância de Cook

Esta medida permite não só a detecção de *outliers* como também a de pontos que contribuam de forma significativa para o ajuste do modelo encontrado, sendo estes pontos as i -ésimas observações (y_i, z_i) no vetor estimado $\hat{\beta}$, podendo ser calculado pela diferença $\hat{\beta} - \hat{\beta}_{(-i)}$ onde: $\hat{\beta}_{(-i)}$ representa a estimativa de máxima verosimilhança do vetor β obtidas da amostra sem a observação (y_i, z_i) .

Como descrito por Turkman e Silva (2000), a medida é calculada pela expressão,

$$D_i = (\hat{\beta} - \hat{\beta}_{(-i)})^T (X^T V X) (\hat{\beta} - \hat{\beta}_{(-i)})$$

Para análise da Distância de Cook serão representados graficamente cada distância de Cook *versus* observações, identificando as observações que se mais se distanciam das restantes.

Comando *R* utilizado:

```
plot(cooks.distance(modelo), xlab="Observações", ylab="Distância de Cook")
```

2.1.4. Avaliação do modelo de Regressão logística múltipla

Para a avaliação dos modelos encontrados serão verificadas medidas que nos permitiram avaliar os modelos.

- Teste de Hosmer e Lemeshow

No caso em estudo recorre-se ao teste de Hosmer e Lemeshow. Propõe-se para o teste dois tipos de agrupamento dos resultados previstos pelos modelos, assim supondo que $J = n$ onde existe n probabilidades estimadas. Para realizar o teste primeiro ordena-se as n probabilidades estimadas, agrupando-as da seguinte forma:

1º Tendo em conta os decis das probabilidades estimadas - Considera-se $g = 10$ grupos em que os primeiros $n'_1 = \frac{n}{10}$ são os que apresentam as menores probabilidades estimadas e $n'_{10} = \frac{n}{10}$ são os que apresentam as maiores probabilidades estimadas.

2º Tendo em conta pontes de corte definidos - Método este descrito em Hosmer e Lemeshow (2000).

Comando R utilizado:

```
hosmerlem = function(y, yhat, g=10) {cutyhat = cut(yhat,
breaks = quantile(yhat, probs=seq(0,1, 1/g)), include.lowest=TRUE)
obs = xtabs(cbind(1 - y, y) ~ cutyhat)
expect = xtabs(cbind(1 - yhat, yhat) ~ cutyhat)
chisq = sum((obs - expect)^2/expect)
P = 1 - pchisq(chisq, g - 2) return(list(chisq=chisq,p.value=P))}
hosmerlem(variável resposta, fitted(modelo))
```

- Coeficiente de determinação R^2

Outra medida que permite avaliar o modelo, é o coeficiente de determinação R^2 de *Cox & Snell* e o coeficiente de *Nagelkerke*.

O coeficiente de *Cox & Snell* não é uma medida da variabilidade explicada pelo modelo uma vez que é baseado na comparação do modelo ajustado (*LLC*) com o modelo nulo (*LLo*).

$$R_N^2 = \frac{1 - e^{-\frac{2[LLC-LLo]}{n}}}{1 - e^{-\frac{2LLo}{n}}}$$

Esta medida assume valores inferiores a 1.

Como forma de ultrapassar o fato do anterior coeficiente assumir valores inferiores a 1, utilizou-se o R^2 de *Nagelkerke* que modificou o coeficiente anterior de forma a atingir o valor 1.

Comando R utilizado:

```
R2N<-(1-exp(-(modelo$dev-modelo$null)/n))/(1-exp(-modelo$null/n))
```

- Curva de ROC

A curva de ROC é um teste diagnóstico que é construído com base na sensibilidade e na especificidade, sendo a sensibilidade dada através da proporção de verdadeiros positivos, determinando a capacidade do modelo prever corretamente os eventos (no nosso caso, os acertos dos alunos). Já a especificidade é dada através da proporção de verdadeiros negativos, sendo determinada a capacidade do modelo prever corretamente as não ocorrências (no nosso caso, os não acertos).

Existem algumas precauções a tomar quando analisada a sensibilidade e a especificidade. Assim, quando o objetivo em estudo é evitar a ocorrência de falsos positivos (prever uma ocorrência incorretamente) então o ponto de corte deve tomar máxima a especificidade. Se a maior preocupação for evitar falsos negativos (prever uma não ocorrência incorretamente) o ponto de corte deve tomar máxima a sensibilidade.

A curva de ROC, construída com base nas duas medidas anteriores sendo estas calculadas para vários pontos de corte e dispostas num gráfico especificidade (eixo xx) *versus* sensibilidade (eixo yy). A uma maior área de curva de ROC delimitada pela reta que contém os pontos de (0,0) e (1,1), e os pontos anteriormente encontrados corresponde maior poder discriminativo do modelo em estudo.

Sendo os valores que poderá assumir a área da curva e a classificação da capacidade discriminativa de acordo com o seguinte:

$\text{área} = 0,5$ – não existe discriminação

$0,5 \leq \text{área} < 0,7$ – pobre ou fraca

$0,7 \leq \text{área} < 0,8$ – aceitável

$0,8 \leq \text{área} < 0,9$ – excelente

$0,9 \leq \text{área} < 1$ – excepcional

Comando R utilizado:

```
library(Epi)
ROC(form=dados$base de dados1~modelo,dat=dados,plot="ROC",PV=T,MX=T,AUC=T)
```

2.1.5 Interpretação dos coeficientes do modelo de Regressão logística múltipla

Para a interpretação dos coeficientes do modelo de regressão logística, sabe-se que:

$$\pi(x) = E(Y|x) = \frac{\exp^{\beta_0 + \beta_1 X}}{1 + \exp^{\beta_0 + \beta_1 X}}$$

Após ser linearizada através da função *logit* obtém-se:

$$\text{logit} = g(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 X$$

$$\text{Odds} = \frac{\pi(x)}{1 - \pi(x)}$$

Para interpretar o coeficiente é introduzido o termo da medida *odds ratio*, para um intervalo $[a, b]$:

$$\hat{g}(a) - \hat{g}(b) = \hat{\beta}_0 + a \times \hat{\beta}_1 - [\hat{\beta}_0 + b \times \hat{\beta}_1] = (a - b)\hat{\beta}_1$$

Sendo odds ratio dado por:

$$\widehat{OR}(a, b) = \exp[(a - b)\hat{\beta}_1]$$

com intervalo de confiança de nível $(1 - \alpha) \times 100\%$:

$$\left(\hat{\beta}_1 \pm Z_{1-\alpha/2} \hat{\sigma}_{\hat{\beta}_1}\right)$$

com $\hat{\sigma}_{\hat{\beta}_1}$ o desvio padrão do coeficiente $\hat{\beta}_1$ e $z_{1-\alpha/2}$ o valor crítico da distribuição normal centrada e reduzida na percentil $1 - \alpha/2$.

Caso a interpretação seja feita por unidades c obtém-se:

$$g(x + c) - g(x) = c\hat{\beta}_1$$

$$OR = \exp(c\hat{\beta}_1)$$

$$IC_{1-\alpha} = \left(c\hat{\beta}_1 \pm Z_{1-\alpha/2} c\hat{\sigma}_{\hat{\beta}_1}\right)$$

$$\widehat{Var}[g(x)] = \widehat{Var}(\hat{\beta}_0) + x^2 \widehat{Var}(\hat{\beta}_1) + 2xcov(\hat{\beta}_0, \hat{\beta}_1)$$

Então o intervalo de confiança de nível $(1 - \alpha) \times 100\%$ para $\pi(x)$ é dado por

$$\frac{\exp\left(\hat{g}(x) \pm Z_{1-\alpha/2} \hat{\sigma}_{\hat{g}(x)}\right)}{1 + \exp\left(\hat{g}(x) \pm Z_{1-\alpha/2} \hat{\sigma}_{\hat{g}(x)}\right)}$$

onde $z_{1-\alpha/2}$ o valor crítico da distribuição normal centrada e reduzida na percentil $1 - \alpha/2$.

Quando se está na presença de uma variável com mais do que uma categoria, a interpretação do *odds ratio* é feita a partir da categoria de referência.

Para realizar a estimativa dos *odds ratio* na presença de interação, primeiro escreve-se as expressões do *logit* para os dois níveis de fator de risco que se pretende comparar, de seguida calcula-se a exponencial da diferença entre os dois *logit* encontrados.

Por exemplo, considera-se um modelo com duas variáveis e a sua interação, denota-se $F = f$ o fator de risco, $X = x$ a covariável e $F \times X$ a sua interação.

O modelo
$$g(f, x) = \hat{\beta}_0 + \hat{\beta}_1 f + \hat{\beta}_2 x + \hat{\beta}_3 f \times x$$

Para o cálculo dos *odds ratio* da comparação de dois níveis de F , isto é $F = f_1$ contra $F = f_0$ para $X = x$.

$$g(f_1, x) = \hat{\beta}_0 + \hat{\beta}_1 f_1 + \hat{\beta}_2 x + \hat{\beta}_3 f_1 \times x \quad \text{e} \quad g(f_0, x) = \hat{\beta}_0 + \hat{\beta}_1 f_0 + \hat{\beta}_2 x + \hat{\beta}_3 f_0 \times x$$

A diferença entre os *logit* encontrados é dada por:

$$\begin{aligned} \ln[OR(F = f_1, F = f_0, X = x)] &= g(f_1, x) - g(f_0, x) \\ &= \hat{\beta}_1(f_1 - f_0) + \hat{\beta}_3 x(f_1 - f_0) \end{aligned}$$

Logo *odds ratio* é igual a $OR = \exp[\hat{\beta}_1(f_1 - f_0) + \hat{\beta}_3 x(f_1 - f_0)]$

Sendo o estimador da variância dado por:

$$\begin{aligned} \widehat{Var}\{\ln[\widehat{OR}(F = f_1, F = f_0, X = x)]\} &= (f_1 - f_0)^2 \times \widehat{Var}(\hat{\beta}_1) + [x(f_1 - f_0)]^2 \times \widehat{Var}(\hat{\beta}_3) + \\ &+ 2x(f_1 - f_0) \times \widehat{Cov}(\hat{\beta}_1, \hat{\beta}_3) \end{aligned}$$

E o intervalo de confiança a nível $(1 - \alpha) \times 100\%$ para o estimador *odds ratio* igual a:

$$\exp \left[\hat{\beta}_1(f_1 - f_0) + \hat{\beta}_3 x(f_1 - f_0) \pm z_{1-\alpha/2} \sqrt{\widehat{Var}\{\ln[\widehat{OR}(F = f_1, F = f_0, X = x)]\}} \right]$$

onde $z_{1-\alpha/2}$ o valor crítico da distribuição normal centrada e reduzida na percentil $1 - \alpha/2$.

2.2. Modelo de Regressão logística ordinal

Os modelos de regressão logística ordinal permitem modelar dados quando a variável resposta é ordinal. Quando a variável resposta possui categorias ordenadas, segundo Agresti, (2007), o uso da regressão logística para respostas ordinais possui uma interpretação mais potente e simplificada.

A regressão logística ordinal está assente no uso da probabilidade acumulada, onde a probabilidade é um valor que está delimitado por uma determinada categoria ou por outra inferior a ela.

Como descreve Hosmer e Lemeshow, (2000), existem três tipos de modelos logísticos usados com mais frequência: modelos de categorias adjacentes, modelos de razão contínua e modelos de *odds* proporcionais.

Serão brevemente apresentados nas três secções seguintes os tipos de modelos logísticos mais usuais, no entanto será dada mais ênfase ao modelo de *odds* proporcionais aplicado aos dados deste trabalho.

Na modelação da regressão logística ordinal inicialmente foram seguidas as mesmas etapas realizadas no modelo de regressão logística.

2.2.1. Modelos de categorias adjacentes

O modelo de regressão ordinal de categorias adjacentes é utilizado, quando se pretende comparar a probabilidade de uma resposta com outra imediatamente superior. Assumindo que as probabilidades de cada observação não dependem da resposta e as hipóteses de *logit* são lineares para os coeficientes, então o *logit* para categorias adjacentes é dado por:

$$a_k(X) = \ln \left[\frac{\phi_k(X)}{\phi_{k-1}(X)} \right] = \alpha_k + X' \beta, \quad \text{com } k = 1, 2, \dots, K$$

com $\phi_k(X) = P(Y_j = k|X)$.

Conclui-se também que as categorias adjacentes não são mais do que uma versão restrita dos *logit* lineares, uma vez que:

$$\begin{aligned} \ln \left[\frac{\phi_k(X)}{\phi_0(X)} \right] &= \ln \left[\frac{\phi_1(X)}{\phi_0(X)} \right] + \ln \left[\frac{\phi_2(X)}{\phi_1(X)} \right] + \dots + \ln \left[\frac{\phi_k(X)}{\phi_{k-1}(X)} \right] \\ &= a_1(X) + a_2(X) + \dots + a_k(X) \\ &= (\alpha_1 + X' \beta) + (\alpha_2 + X' \beta) + \dots + (\alpha_k + X' \beta) = (\alpha_1 + \alpha_2 + \dots + \alpha_k) + kX' \beta \end{aligned}$$

2.2.2. Modelos de razão contínua

No modelo ordinal de razão contínua pretende-se comparar a probabilidade de uma resposta igual a uma categoria k com a probabilidade de uma resposta menor, $Y < k$, definindo o *logit* do modelo como:

$$\begin{aligned} r_k(X) &= \ln \left[\frac{P(Y = k|X)}{P(Y < k|X)} \right] \\ &= \ln \left[\frac{\phi_k(x)}{\phi_0(x) + \phi_1(x) + \dots + \phi_{k-1}(x)} \right] \\ &= \theta_k + X'\beta_k \quad \text{com } k = 1, 2, \dots, K \end{aligned}$$

O modelo obtido irá ter diferentes constantes e coeficientes para cada uma das comparações efetuadas. Como é demonstrado por Hosmer e Lemeshow, (2000), o modelo pode ser ajustado através de K modelos de regressão logística binária.

2.2.3. Modelos de odds proporcionais

O modelo de *odds* proporcionais compara a probabilidade da resposta não exceder uma dada categoria, ($k = 0, 1, \dots, K - 1$), com a probabilidade de uma resposta exceder essa categoria.

$$C_k(x) = \ln \left[\frac{P(Y \leq k|x)}{P(Y > k|x)} \right] = \ln \left[\frac{\phi_0(x) + \phi_1(x) + \dots + \phi_k(x)}{\phi_{k+1}(x) + \phi_{k+2}(x) + \dots + \phi_K(x)} \right] = \tau_k - X'\beta$$

O modelo encontrado terá $k - 1 + p$ parâmetros, possui os valores das interseções α_j variando para cada equação, tal que, $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_{k-1}$ e p coeficientes β .

Este modelo apresenta uma característica descrita por McCullagh e Nelder (1983), trata-se da possibilidade de fornecer uma única estimativa de *odds ratio* para todas as categorias comparadas, tornando-se mais fácil a interpretação.

Para aplicação dos modelos de *odds* proporcionais, terá que se ter em atenção a verificação do pressuposto subjacente devendo para o efeito realizar-se uma análise de resíduos.

Dois tipos de resíduos serão aplicados, os resíduos escores (*score.binary*) e os resíduos parcial (*partial*).

Relativamente aos resíduos escore, se a suposição de odds proporcionais for válida é esperado que, para cada covariável a tendência em torno das categorias da variável resposta tenha um comportamento horizontal constante. Já a análise dos resíduos parciais, espera-se que para um modelo com um bom ajuste os gráficos sejam paralelos.

Exemplo de comando R utilizado:

```
library (rms)

modelo<-lrm(variável resposta~covariáveis explocativas,data= modo, x=TRUE, y=TRUE)

residuals (modelo, type="partial", pl=TRUE)

residuals (modelo, type="score.binary", pl=TRUE)
```

Para o diagnóstico do modelo recorre-se à análise de resíduos aplicados na regressão logística anteriormente descrita. Esta análise será efetuada a todos os pares de respostas possíveis da variável resposta, assim, se a variável resposta admite os valores 0, 1 e 2 a análise é realizada a *0 versus 1*, *0 versus 2* e *1 versus 2*.

A qualidade do ajuste dos modelos ordinais é verificada através do teste de *Pearson* ou *deviance* que são baseadas nas aproximações da estatística para a distribuição qui-quadrado com $(L - 1)(K - 1)p$ graus de liberdade, sendo L o número de linhas, K o número de colunas da tabela de contingência e p o número de covariáveis do modelo, sob o teste:

$$\begin{cases} H_0: o \text{ modelo ajusta - se aos dados} \\ H_1: o \text{ modelo não se ajusta aos dados} \end{cases}$$

Para o cálculo dos *odds ratios* recorre-se à equação do modelo de odds proporcional, comparando valores menores ou iguais a uma dada categoria a valores superiores, assim:

$$\begin{aligned} \ln \left[\frac{P(Y \leq k|x_1)}{P(Y > k|x_1)} \right] - \ln \left[\frac{P(Y \leq k|x_0)}{P(Y > k|x_0)} \right] &= (\tau_{k+1} - x_1 \times \beta) - (\tau_{k+1} - x_0 \times \beta) \\ &= -\beta(x_1 - x_0) \end{aligned}$$

Concluindo que os $OR = \exp[-\beta(x_1 - x_0)]$.

2.3. Modelo de Regressão logística multinomial

O modelo de regressão logístico multinomial é aplicado quando a variável resposta é representada por mais do que duas categorias nominais.

Suponha-se que se tem p covariáveis e um termo constante, representadas pelo vetor X , de comprimento $p + 1$ com $x_0 = 1$ e que a variável resposta tem três categorias (0, 1 e 2).

Neste caso temos duas funções *logit* que são dadas por,

$$g_1(X) = \ln \left[\frac{P(Y = 1|X)}{P(Y = 0|X)} \right] = X' \beta_1$$

$$g_2(X) = \ln \left[\frac{P(Y = 2|X)}{P(Y = 0|X)} \right] = X' \beta_2$$

e as probabilidades estimadas são obtidas por

$$P(Y = 0|X) = \frac{e^{X\beta_0}}{e^{X\beta_0} + e^{X\beta_1} + e^{X\beta_2}}$$

$$P(Y = 1|X) = \frac{e^{X\beta_1}}{e^{X\beta_0} + e^{X\beta_1} + e^{X\beta_2}}$$

$$P(Y = 2|X) = \frac{e^{X\beta_2}}{e^{X\beta_0} + e^{X\beta_1} + e^{X\beta_2}}$$

Desta forma o modelo multinomial é constituído por 3 modelos logísticos. Para implementação do mesmo foram seguidos os passos descritos anteriormente para o modelo logístico.

O diagnóstico ao modelo multinomial é realizado com recurso à análise de resíduos (descrita para o modelo logístico) de cada um dos modelos encontrados.

Acerca da significância do modelo ou a avaliação do ajuste do modelo será realizado o teste de *Pearson e Deviance*.

A capacidade discriminativa do modelo de regressão logístico multinomial é em tudo equivalente ao modelo logístico, no entanto uma observação i , pode ser classificada em mais do que duas categorias da variável dependente. A probabilidade de observar uma determinada categoria j comparativamente à categoria de referência (por exemplo 0) numa observação i é,

$$P(Y_i = j|X_i = x_i) = \frac{e^{x_i \beta_j}}{1 + \sum_{k=1}^n e^{x_i \beta_k}}$$

e para a classe de referência

$$P(Y = 0|X_i = x_i) = \frac{1}{1 + \sum_{k=1}^n e^{x_i \beta_k}}$$

Na prática, para uma observação i , calcula-se a probabilidade de pertencer a cada uma das j categorias e classifica-se-a na categoria onde a probabilidade é superior.

Exemplo de comando *R* utilizado:

```
library(mlogit)

mldata<-mlogit.data(modos, varying=NULL, choice="Variável resposta", shape="wide")

mlogit1<- mlogit(Variável resposta ~1 | variáveis explicativas, data = mldata)
```

2.4. Teste de Cochran- Mantel-Haenszel e teste Breslow-Day

No sentido de comparar dois grupos estratificados X e Y por uma variável de controlo Z, irá ser aplicado o teste de *Cochran-Mantel-Haenszel*, senda as hipótese a testar:

H_0 : X é condicionalmente independente da Y, em qualquer estrato Z.

O teste de *Cochran-Mantel-Haenszel* encontra-se descrito em Agresti (2002), sendo aplicado o teste em tabelas de contingência $2 \times 2 \times k$ estratos condicionado aos totais de linhas e colunas em cada tabela parcial.

Neste teste garante-se que qualquer associação entre as variáveis dicotómicas não é afetada pela terceira variável.

Comando R utilizado:

```
mantelhaen.test (variável 1, variável 2, variável estratificadora, correct=T)
```

Após interpretação do teste *Mantel-Haenszel*, no caso de garantia de existência de associação entre as variáveis dentro de cada estrato, aplicar-se-á o teste *Breslow-Day*.

O teste *Breslow-Day* testa a hipótese nula de homogeneidade dos *odds ratio*, isto é, se o *odds ratio* entre duas variáveis X e Y é igual para as diferentes categorias Z (variável estratificadora). No caso da não rejeição da hipótese nula é possível sintetizar a associação condicional por um valor de *odds ratio*. O teste aplicado tem a forma da estatística qui-quadrado de Pearson, comparando as contagens de células observadas às estimadas, procurando relações de *odds ratio* comuns.

Comando R utilizado:

```
source ("breslowday.R")  
breslowday.test (tabela)
```


Capítulo 3: Análise de dados - área espaço e forma e área da quantidade

Neste capítulo será realizado um estudo dos alunos que responderam às questões selecionadas na área de conteúdos de espaço e forma e na área da quantidade em Portugal. Para o efeito, inicialmente foi realizado um estudo das amostras recolhidas, caracterizando os alunos que responderam às duas questões e os seus agregados familiares.

Serão também analisados os resultados dos alunos às mesmas questões em países europeus, procurando realizar um estudo comparativo entre o comportamento dos alunos desses países e os alunos portugueses.

3.1. Análise da percentagem de acertos dos alunos às questões - questionário dos alunos

Dos alunos que responderam no ano de 2009 à questão da temática espaço e forma, verificou-se que 27% responderam corretamente. Se for considerado apenas os alunos portugueses então a percentagem de respostas corretas sobe para 35%. Comparando as proporções de acertos à questão dos alunos portugueses com os restantes países através de um teste de hipóteses para a diferença de proporções (valor de $p < 0,001$) concluímos estatisticamente a diferença entre as proporções obtidas.

Relativamente à questão na área temática quantidade a percentagem de acertos dos alunos portugueses foi de 54%, enquanto dos restantes países situou-se nos 35%, registando-se estatisticamente a diferença entre a proporção de acertos (valor de $p < 0,001$).

Para o estudo foram selecionadas as questões com interesse relevante no questionário dos alunos e pais (Anexo 3) e cujas respostas invalidadas não colocassem em causa os resultados finais.

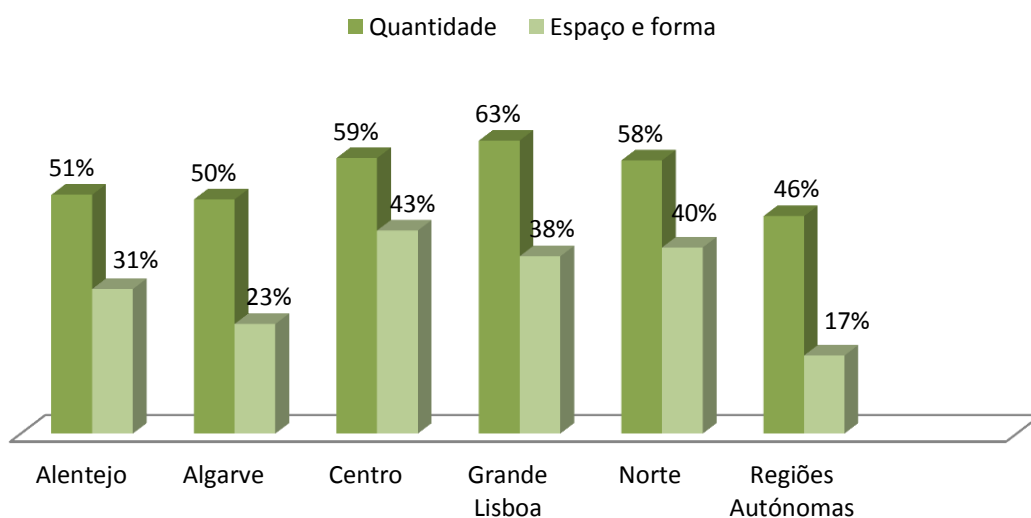
O número de alunos portugueses que responderam à questão na área espaço e forma foi de 1426 passando para 1206 após análise da base de dados, tendo a proporção de acerto da amostra passado para 37%. Na área da quantidade a amostra passou de 1284 para 1081, sendo a proporção de acertos de 57%.

3.1.1. Distribuição das respostas pelo País

Na Figura 3.1 observa-se a percentagem da distribuição dos acertos às duas questões em estudo tendo em conta diferentes regiões de Portugal. Na associação entre as proporções de acertos por região à questão relacionada com espaço e forma, verificou-se que existem evidências estatísticas suficientes (valor $p < 0,001$) para afirmar que as respostas corretas dadas à questão, dependem das zonas geográficas onde os alunos estudam. A maior predominância de acertos registou-se na região do Centro de Portugal (43%) sendo a menor registada na Madeira e Açores (17%).

Na questão da área da quantidade não existe evidência estatística suficiente para garantir a dependência nos acertos nas diferentes regiões (valor $p = 0,131$).

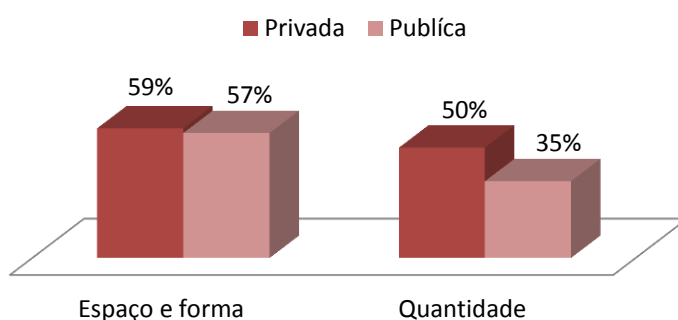
Figura 3.1 - Proporção de alunos que responderam corretamente à questão por zona geográfica



3.1.2. Tipologia das escolas

Verificou-se que metade dos alunos que estudam no ensino privado, responderam corretamente à questão na área espaço e forma enquanto que, no ensino público apenas 35% responderam corretamente, Figura 3.2. Também se pode concluir que os alunos das escolas privadas obtêm melhores resultados na questão em estudo (valor $p < 0,001$). Na área da quantidade a proporção de acertos é semelhante sendo de 59% no ensino público e 57% no privado, não existindo diferenças significativas entre as proporções de acertos (valor $p = 0,637$).

Figura 3.2 - Proporção de alunos que responderam corretamente à questão por escola pública e privada



3.1.3. Ano de escolaridade

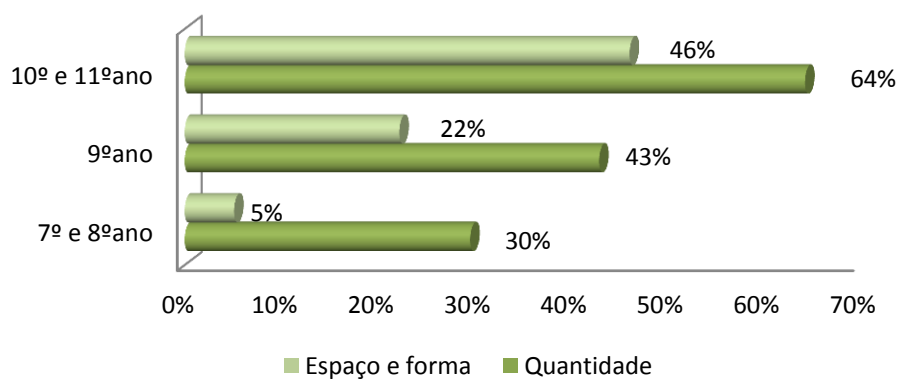
Na Figura 3.3 apresenta-se, a distribuição de acertos dos alunos por ano de frequência do 3º Ciclo e do Secundário. Como os alunos de 11º ano constituem uma percentagem residual resolveu-se agrupá-los aos de 10º ano, e o mesmo raciocínio se fez para os alunos do 7º ano onde foram agrupados aos do 8º ano.

Ambas as amostras são constituídas maioritariamente por alunos do 10º ano e 11º ano (70%), sendo que cerca de 1/4 dos alunos é do 9º ano.

Na questão espaço e forma, os alunos que frequentam o 10º/11ºano obtiveram a percentagem mais alta de acertos (46%), enquanto os alunos do 7º/8ºano a percentagem mais baixa (5%), analisada a proporção de acertos por ano de escolaridade verifica-se que, a mesma é estatisticamente diferente (valor $p < 0,001$).

Na área da quantidade para o 10º/11º ano obtiveram-se as percentagens mais elevadas (64%) e as mais baixas para o 7º/8º ano (30%), existindo diferenças significativas (valor $p < 0,001$).

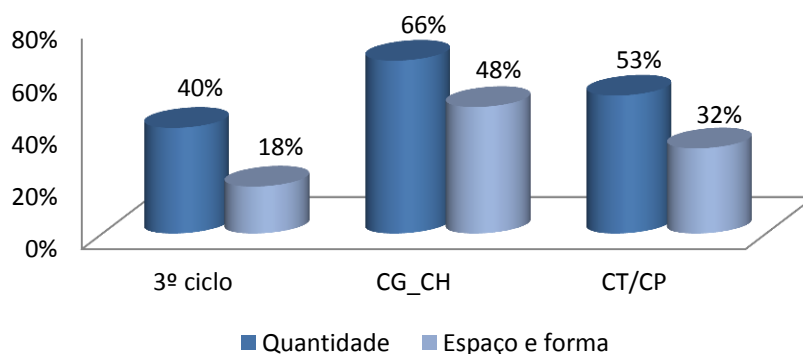
Figura 3.3 - Proporção de alunos que responderam corretamente à questão por ano de escolaridade



3.1.4. Ciclo de estudos que frequenta

Há uma maior proporção de acertos nos alunos que frequentam o ensino secundário-curso geral/científico humanístico, sendo de 48% para a área espaço e forma e de 66% para a área da quantidade (Figura 3.4). A desigualdade de distribuição das respostas dadas às duas questões é significativa (valores $p < 0,001$).

Figura 3.4 - Proporção de alunos que responderam corretamente à questão por tipologia de ensino

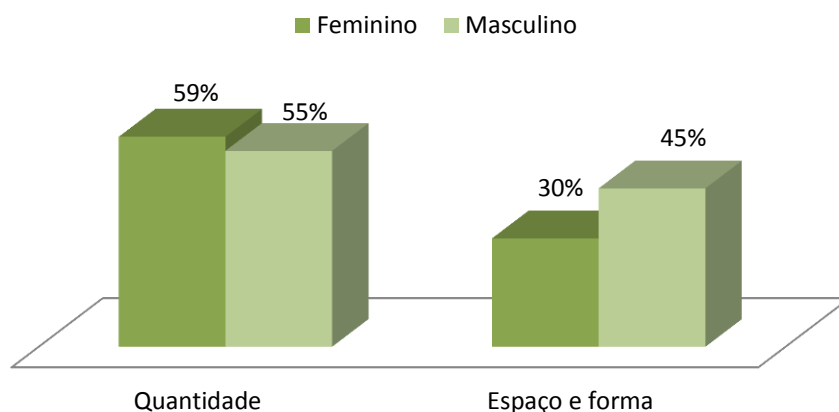


CG_CH – Ensino Secundário, curso geral, científico humanístico; CT/CP – Ensino Secundário, curso tecnológico/Ensino Secundário, curso profissional

3.1.5. Género

A distribuição dos acertos dos alunos por sexo encontra-se representada na Figura 3.5., verifica-se que 45% dos rapazes responde corretamente à questão na área espaço e forma, sendo este valor significativamente superior ao das raparigas (valor $p < 0,001$). Na questão na área da quantidade não se regista diferenças significativas entre os dois sexos (valor $p = 0,191$).

Figura 3.5 - Proporção de alunos que responderam corretamente à questão por género

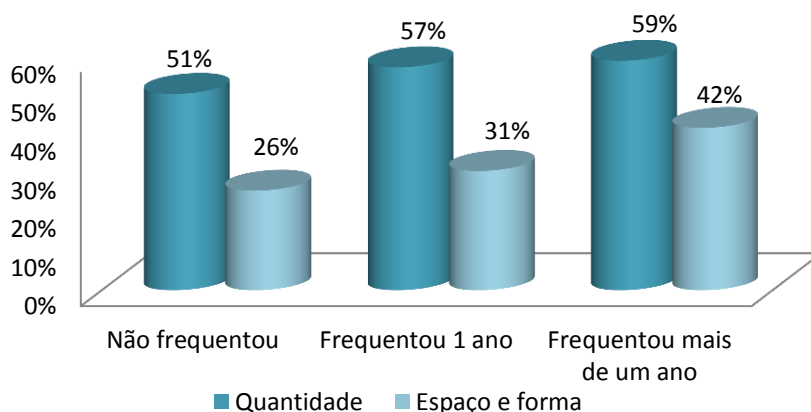


3.1.6. Frequência do ensino pré-escolar

Representada a distribuição de acertos por frequência de ensino pré-escolar (Figura 3.6), a maioria dos acertos foi para os alunos que o frequentaram mais do que um ano, 42% na área espaço e forma e 59% na área quantidade.

Verificou-se a existência de evidências estatísticas suficientes para afirmar que o desempenho do aluno à questão na área espaço e forma, depende da frequência do ensino pré-escolar (valor $p < 0,001$). Relativamente à questão na área da quantidade, não se garante a diferença entre as proporções de acertos dos alunos que frequentam ou não o ensino pré-escolar (valor $p = 0,108$).

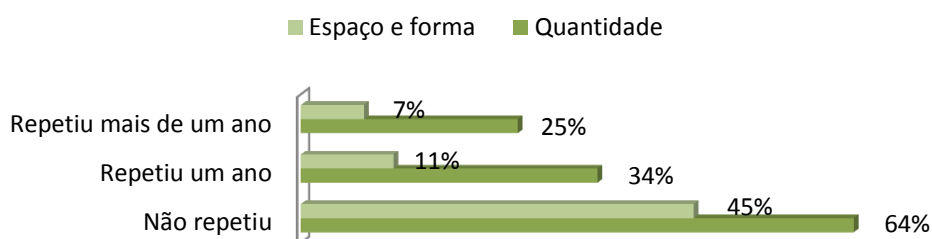
Figura 3.6 - Proporção de alunos que responderam corretamente à questão por frequência no ensino pré-escolar



3.1.7. Repetências durante o percurso de estudo

Na figura 3.7 encontra-se a distribuição de acertos tendo em conta as repetências no ensino básico, foi construída uma variável com base nas respostas dadas pelos alunos ao questionário nas questões, “número de repetências no 1º e 2º ciclos” e “número de repetências no 3º ciclo”, resultando numa única variável com três opções: nunca ficou retido, já ficou retido um ano e já ficou retido mais do que um ano.

Figura 3.7 - Proporção de alunos que responderam corretamente à questão por repetências no ensino básico



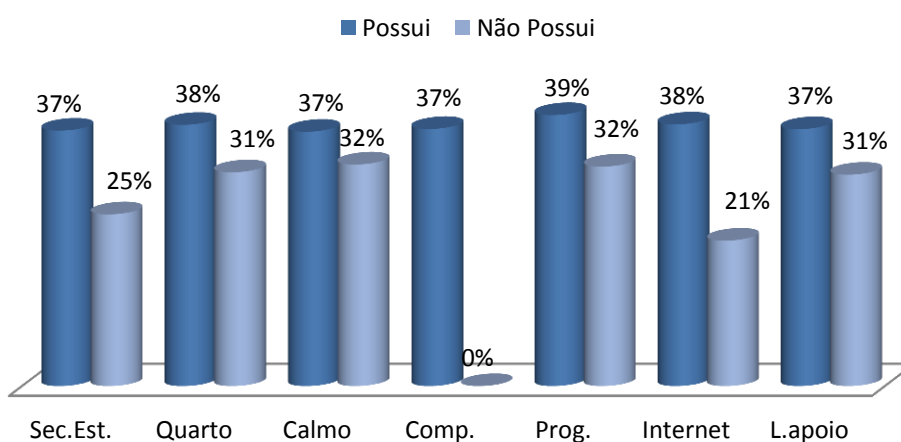
Verificou-se que dentro do universo dos alunos que já repetiram um ano ou mais, durante os três primeiros ciclos de ensino, a percentagem de respostas corretas às questões é muito baixa sendo de 7% na área espaço e forma e de 25% na área da quantidade. Nos alunos que nunca ficaram retidos, a percentagem de acertos é de 45% para a área espaço e forma e 64% para a área da quantidade. Estatisticamente há evidência para afirmar que existem diferenças significativas para as proporções de acertos nas duas áreas em estudos (valores $p < 0,001$).

3.1.8. Material de apoio ao estudo

No questionário do aluno, existe uma secção acerca de algumas das condições que os alunos possuem em casa no auxílio ao estudo, estando os resultados representados na Figura 3.8. Relativamente à questão na área espaço e forma, verifica-se existir evidência estatística para afirmar que a proporção de acertos não depende de ter uma secretária (valor $p = 0,187$) ou um local calmo para estudar (valor $p = 0,587$), no entanto, os alunos que possuem um quarto individual apresentam uma maior proporção de acertos (valor $p = 0,067$), sendo esta associação ligeira.

Outras das questões colocadas aos alunos prenderam-se com o possuir um computador para estudar em casa, programas educativos e ligação à internet. Nestes casos, verificou-se que existem evidências estatísticas suficientes (valor $p < 0,001$, valor $p = 0,011$ e valor $p = 0,002$ respetivamente) para afirmar que a percentagem de respostas corretas à questão não é independente dos alunos terem acesso às três condições anteriores. Pode-se concluir que o acerto à questão é independente do acesso a livros de apoio ao estudo (valor $p = 0,292$).

Figura 3.8 - Proporção de alunos que responderam corretamente à questão na questão da área de espaço e forma



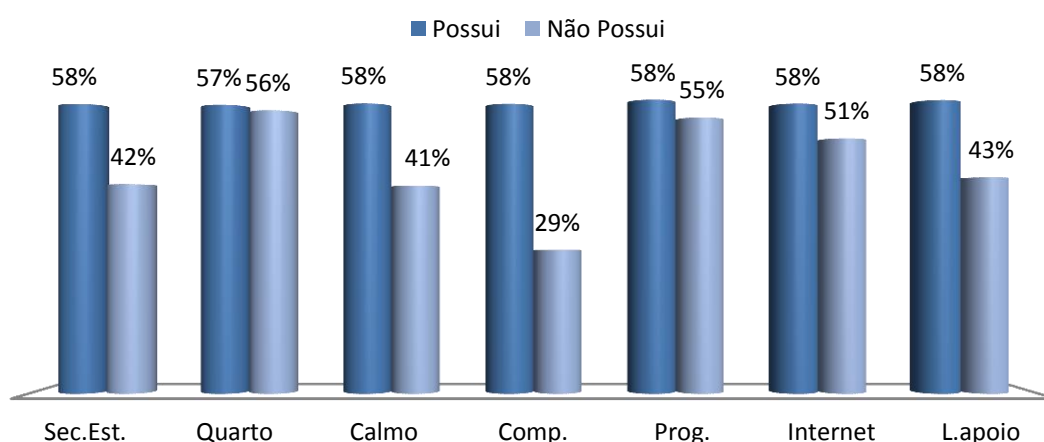
Sec.Est - Uma secretária onde estudar ; Quarto - Um quarto só para si; Calmo - Um sítio calmo para estudar ; Comp.- Um computador que possa usar para os estudos; Prog. - Programas de computador educativos; Internet - Ligação à Internet; L. apoio – livros de apoio ao estudo.

Relativamente à área da quantidade (Figura 3.9), ao se analisar os acertos dos alunos que possuíam uma secretária para estudar ou um local calmo para o efeito, verificou-se não existir dependência (valor $p=0,118$, valor $p=0,787$ respetivamente). No entanto, os alunos que possuem um quarto individual voltaram a apresentar tendencialmente uma maior proporção de acertos (valor $p=0,080$).

A proporção de acertos não depende dos alunos possuírem programas educativos (valor $p=0,265$) ou ligação à internet (valor $p=0,275$).

Nesta questão, o acesso a livros auxiliares de estudo está associado a uma maior proporção de acertos (valor $p=0,008$).

Figura 3.9 - Proporção de alunos que responderam corretamente à questão na área da quantidade

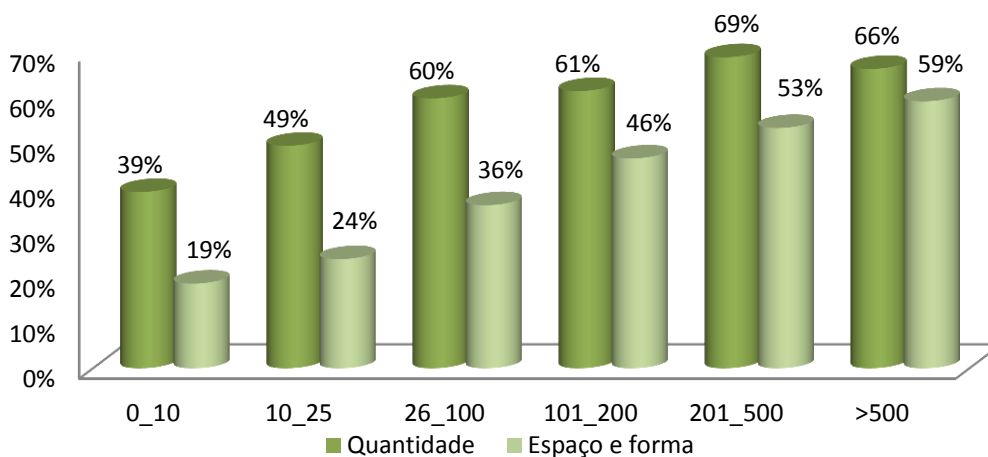


Sec.Est - Uma secretária onde estudar , Quarto - Um quarto só para si; Calmo - Um sítio calmo para estudar ; Comp.- Um computador que possa usar para os estudos; Prog. - Programas de computador educativos; Internet - Ligação à Internet; L. apoio – livros de apoio ao estudo.

Na Figura 3.10 estão representadas as proporções de acerto às duas questões tendo em conta o número de livros que os alunos possuem em casa. A associação entre o número de livros e o desempenho dos alunos é significativa na área espaço e forma (valor $p<0,001$) e na área da quantidade (valor $p<0,001$). Ao analisar os alunos que possuem entre 201 e 500 livros, verifica-se que 3 em cada 5 respondem corretamente, enquanto se a análise for efetuada a alunos que possuem somente entre 10 e 25 livros apenas 1 em cada 4 responde corretamente.

Quanto à questão na área da quantidade verificou-se que 7 em cada 10 alunos que possuem em casa entre 201 e 500 livros responderam corretamente, no entanto os alunos que possuem de 10 a 25 livros apenas 1 em cada 2 o fizeram.

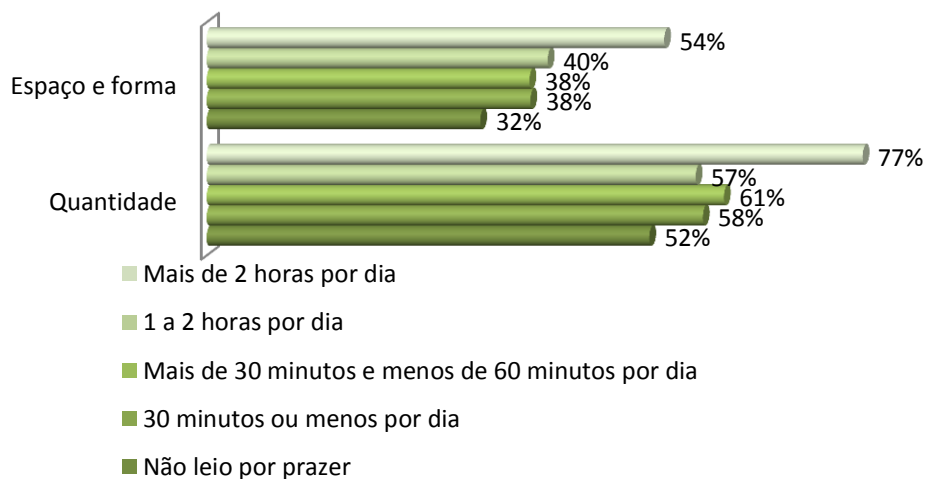
Figura 3.10 - Proporção de alunos que responderam corretamente à questão por número de livros em casa



3.1.9. Hábitos de leitura fora da escola

Na Figura 3.11 apresenta-se a distribuição dos alunos que acertaram corretamente à questão de acordo com os hábitos de leitura verifica-se que, não há igual distribuição das respostas corretas pelos diferentes hábitos de leitura quer na questão espaço e forma (valor $p=0,049$), quer na questão da quantidade (valor $p=0,024$), sendo de destacar um maior percentagem de acertos para o alunos que leem mais de 2 horas por dia.

Figura 3.11 - Proporção de alunos que responderam corretamente à questão por tempo diário dedicado à leitura



3.1.10. Metodologia de estudo

O questionário contempla uma secção onde os alunos são questionados acerca das metodologias de estudo usadas, classificando-as de acordo com o tipo de frequência que as aplicam. Os resultados para a área espaço e forma são apresentados na Figura 3.12, e os resultados na área da quantidade na Figura 3.13.

Figura 3.12 - Proporção de alunos que responderam corretamente à questão por metodologia de estudo na questão da área espaço e forma

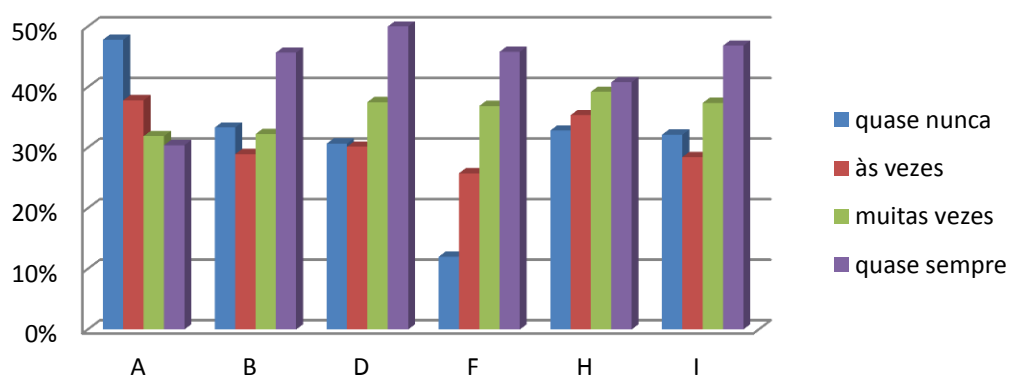
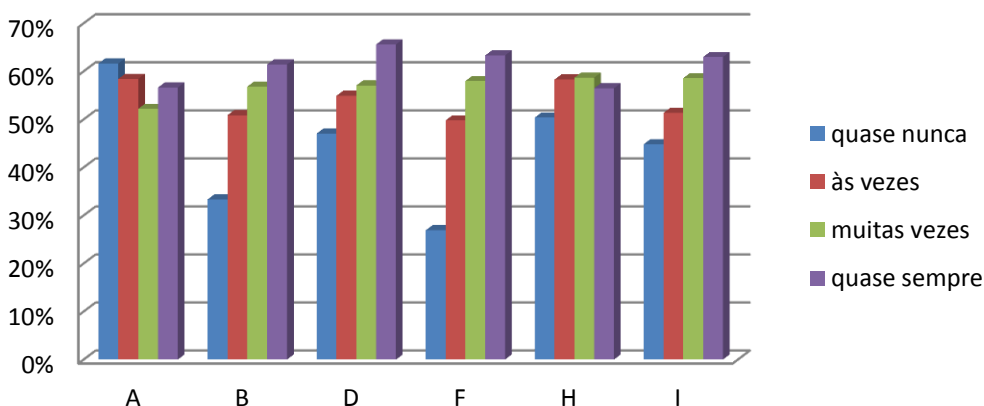


Figura 3.13 - Proporção de alunos que responderam corretamente à questão por metodologia de estudo da área da quantidade



A- Quando estudo, tento decorar tudo o que está no texto ; B- Quando estudo, começo por tentar perceber o que é mesmo preciso aprender; D- Quando estudo, tento relacionar a informação nova com conhecimentos prévios, que aprendi noutras disciplinas; F - Quando estudo, verifico se estou a perceber o que já li ; H - Quando estudo, percebo de que maneira a informação pode ser útil fora da escola ; I - Quando estudo, tento identificar os conceitos que ainda não percebo bem.

Quando analisadas as respostas corretas às questões em estudo nas duas áreas, para alunos que:

- quando estudam decoram os conteúdos (A) – verificou-se estatisticamente que, não existe igual distribuição de acertos(valor $p=0,003$) na área espaço e forma e na área da quantidade

constatou-se a não existência de relação entre esta metodologia e a proporção de acertos (valor $p=0,240$);

- quando estudam, começam por tentar perceber o que é mais importante (B) – na área de espaço e forma verificou-se estatisticamente a diferença entre as proporções de acertos, (valor $p<0,001$) e na área da quantidade garante-se também essa diferença (valor $p=0,012$);

- quando estudam, tentam relacionar conhecimentos novos com os já adquiridos (D) – verificando-se estatisticamente, que existe diferenças entre os acertos, na área espaço e forma e na área da quantidade (valor $p<0,001$ e valor $p=0,013$ respetivamente);

- quando estudam, verificam se estão a perceber o que já leram (F) – não havendo igual distribuição entre os acertos nas duas área em estudo com (valores $p<0,001$);

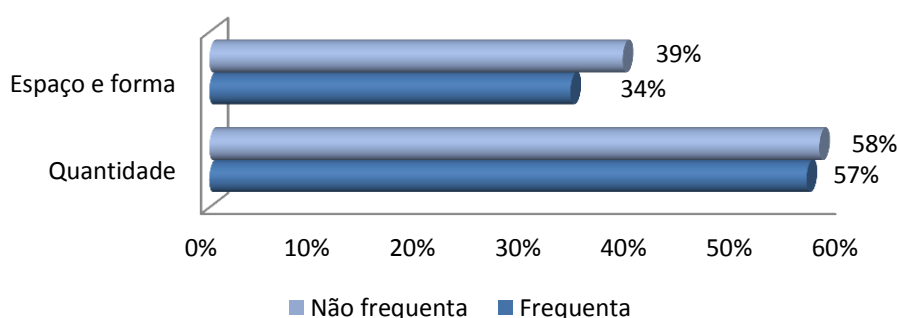
- Quando estudam, verificam de que forma poderão aplicar os conhecimentos fora da escola (H) – não existência de relação entre a aplicação desta metodologia de estudo e a proporção de acertos na área espaço e forma e na área da quantidade (valor $p=0,356$ e valor $p=0,376$ respetivamente);

- Quando estudam, tentam identificar os conceitos que ainda não aprenderam (I) – na área espaço e forma verificou-se estatisticamente que não existia igual distribuição entre os acertos, (valor $p<0,001$) igualmente na área da quantidade(valor $p=0,020$).

3.1.11. Frequência de aulas de enriquecimento ou recuperação a Matemática

Como se pode observar na Figura 3.14 é muito semelhante a proporção de acertos à questão da área da quantidade, tendo em conta a frequência de aulas suplementares a matemática, não existindo diferenças entre a proporção de acertos (valor $p=0,698$). Na área espaço e forma a não frequência às aulas suplementares está ligeiramente associada a uma maior proporção de acertos (valor $p=0,077$).

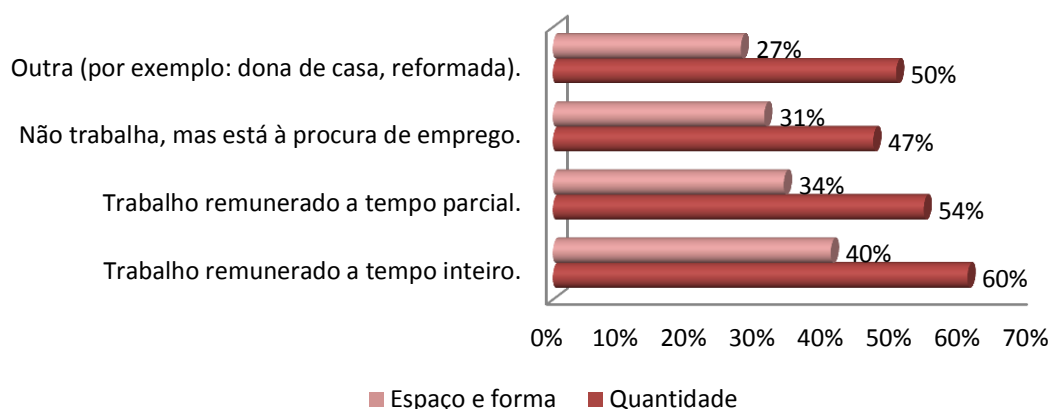
Figura 3.14 - Proporção de alunos que responderam corretamente à questão por frequência de aulas suplementares



3.1.12. Situação profissional dos progenitores

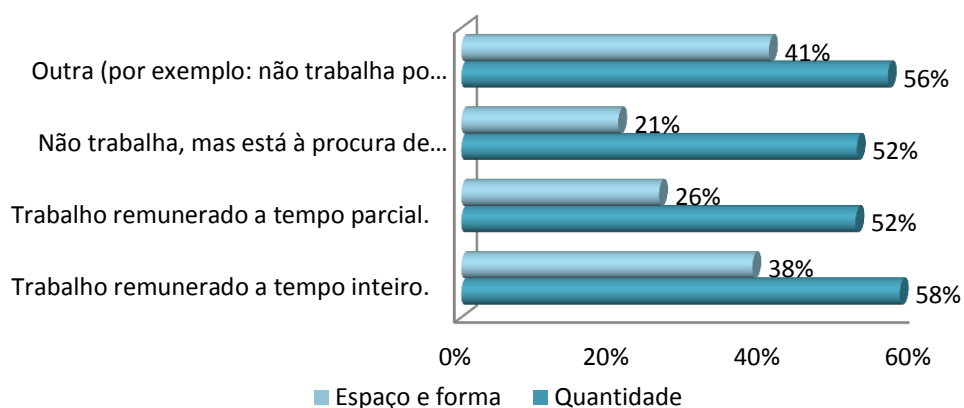
Quando analisada a situação profissional dos pais dos alunos Figura 3.15, destaca-se que a maior proporção de acertos surge nos alunos cujas mães tinham um trabalho remunerado a tempo inteiro, sendo de 40% na questão sobre espaço e forma e de 60% para a questão da quantidade, constatando-se a existência de diferenças significativas entre os acertos às questões tendo em conta a situação profissional das mães dos alunos (valor $p=0,003$ e valor $p=0,014$ respetivamente).

Figura 3.15 - Proporção de alunos que responderam corretamente à questão por situação profissional da mãe



Na Figura 3.16 apresentam-se os resultados de acordo com a situação profissional dos pais dos alunos, neste caso, a distribuição de acertos na área da quantidade é idêntica para as diferentes situações profissionais dos pais (valor $p=0,702$) enquanto na questão da área espaço e forma se regista uma associação significativa (valor $p=0,008$), tendo os melhores resultados sido obtidos pelos alunos cujos pais tinham um trabalho remunerado a tempo inteiro (60%).

Figura 3.16 - Proporção de alunos que responderam corretamente à questão por situação profissional do pai



3.2. Análise da percentagem de acertos dos alunos às questões - questionário aos Pais

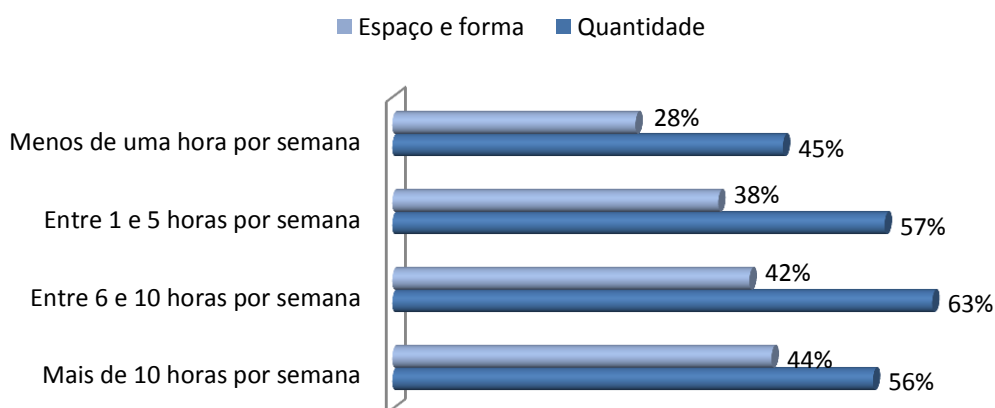
Segundo os dados apurados dos alunos que realizaram as questões e que os seus pais preencheram e entregaram os respetivos inquéritos verificou-se, que no caso da área espaço e forma a amostra inicial de 1426 respostas com uma proporção de acertos de 35% passou para 886 com uma proporção de acertos de 37%, e na questão da área da quantidade a amostra inicial era de 1284 com uma proporção de acertos de 53%, passou para 808 com uma proporção de acertos de 54%.

3.2.1. Tempo dedicado à leitura pelos pais

Na proporção de acertos dos alunos tendo em conta o tempo que um dos pais dedica à leitura semanal Figura 3.17, constatou-se que a maior proporção de acertos foi para os alunos que um dos pais liam mais de 10 horas por semana (44%) e as mais baixas para o caso de um dos pais ler menos de uma hora por semana (28%), verificando-se estatisticamente a existência de diferenças significativas entre a proporção de acertos e o tempo de leitura semanal de um dos progenitores (valor $p=0,019$).

Na questão da área da quantidade a maior proporção de acertos foi para os alunos onde um dos pais leem entre 6 e 10 horas (63%), e as proporções mais baixas para o caso onde um dos pais lê menos de uma hora semanalmente (45%). Analisada a proporção de acertos por tempo dedicado à leitura de um dos pais garante-se a existência de diferenças significativas (valor $p=0,015$).

Figura 3.17 - Proporção de alunos que responderam corretamente à questão por tempo dedicado á leitura



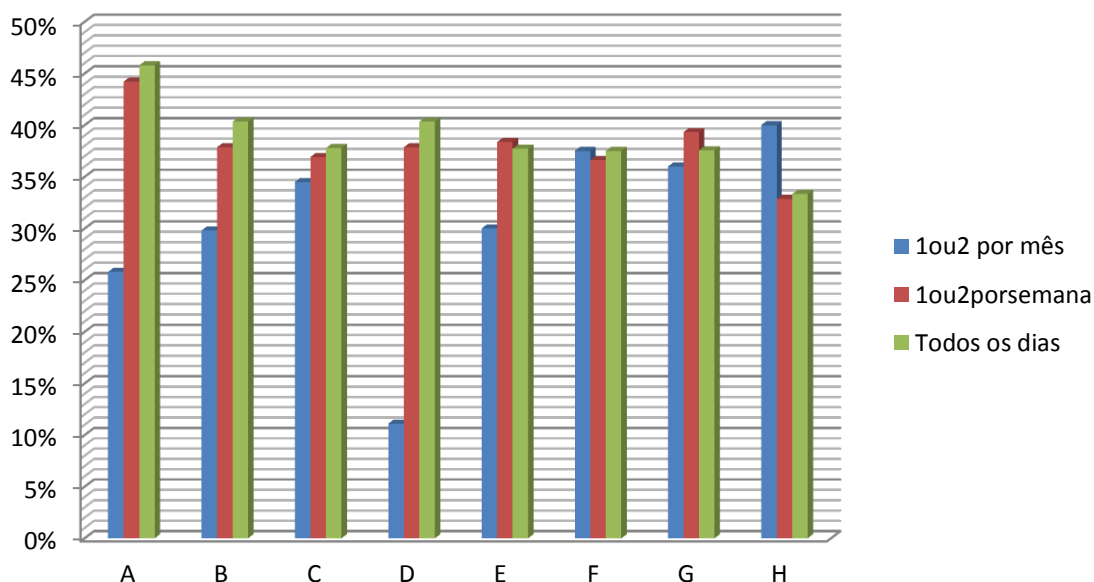
3.2.2. Atividades com os filhos

Quando analisadas algumas das atividades, que os pais realizam com os filhos Figura 3.18, **constata-se que a maior proporção de acertos** na questão da área espaço e forma verificou-se quando:

- os pais dos alunos apenas uma ou duas vezes por mês ajudam os seus filhos no trabalhos de casa (40%), existindo ligeiras diferenças entre a ajudada dada pelos pais e a proporção de acertos (valor $p=0,088$);
- os pais dos alunos conversam regularmente com os filhos acerca do que estão a ler (39%), não se encontrando diferenças significativas entre a proporção de acertos e as conversas mantidas pelos pais com os filhos (valor $p=0,690$);
- os pais dos alunos que frequentam livrarias ou bibliotecas apenas uma ou duas vezes por mês ou todos os dias (38%), não sendo estatisticamente diferente as proporções de acertos (valor $p=0,987$);
- os pais dos alunos que mantém diálogos regulares com os filhos (40%), não existindo diferenças significativas entre os diálogos e o acerto à questão (valor $p=0,468$);
- os pais dos alunos jantam à mesa com o filho diariamente (40%), existindo diferenças significativas entre os acertos à questão e o jantar diariamente com os filhos (valor $p=0,064$);
- os pais dos alunos conversam diariamente acerca do aproveitamento escolar (38%), não existindo diferenças significativas entre a frequência das conversas e os acertos à questão (valor $p=0,873$);
- os pais dos alunos conversam sobre livros, filmes ou programas de televisão diariamente (40%), verificando-se ligeira diferença entre a proporção de acertos e a frequência das conversas (valor $p=0,062$);
- os pais dos alunos, conversam sobre questões políticas ou sociais uma ou duas vezes por semana (46%), verificando que é estatisticamente diferente a proporção de acertos consoante a frequência (valor $p<0,001$).

Relativamente à **menor proporção de acertos** à a realçar a percentagem de apenas 11% para os alunos que não jantam diariamente com os pais e 26% para os alunos que raramente conversam com os pais acerca de questões políticas ou sociais.

Figura 3.18 - Proporção de alunos que responderam corretamente à questão por atividades com os filhos na área espaço e forma



A- Conversar sobre questões políticas ou sociais; B- Conversar sobre livros, filmes ou programas de televisão; C- Conversar sobre o aproveitamento escolar do seu filho; D- Jantar à mesa com o seu filho; E- Passar tempo a conversar simplesmente com o seu filho; F- Ir a uma livraria ou a uma biblioteca com o seu filho. G- Conversar com o seu filho sobre o que ele está a ler; H- Ajudar o seu filho a fazer os trabalhos de casa.

Na área da quantidade, Figura 3.19, **verificámos que a maior proporção** de acertos verificou-se quando:

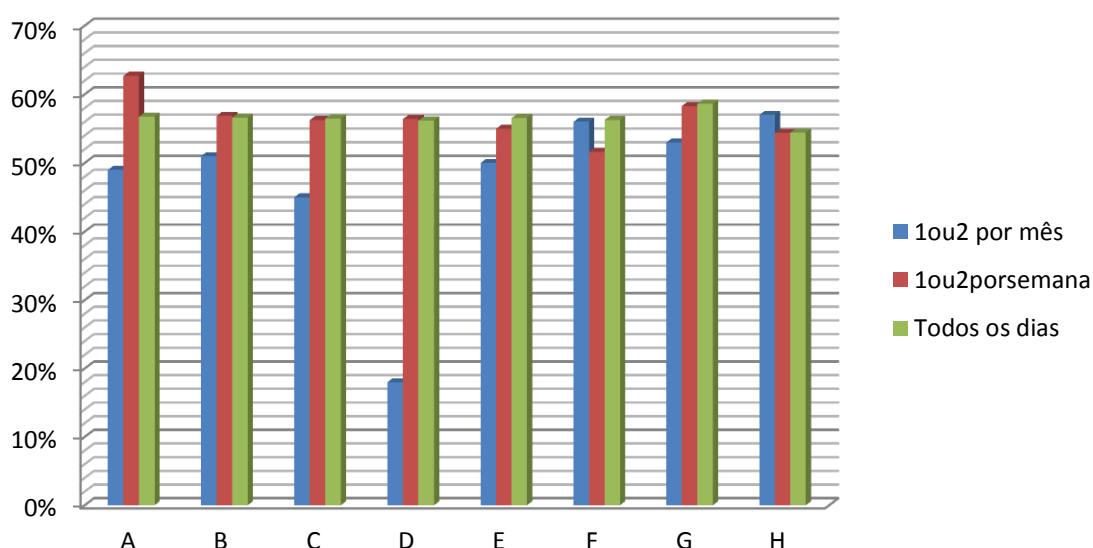
- os pais dos alunos apenas uma ou duas vezes por mês ajudam os seus filhos no trabalhos de casa (57%), não existindo diferenças significativas entre o acerto à questão e a frequência da ajuda dos pais nos trabalhos de casa (valor $p=0,838$);
- os pais dos alunos conversam regularmente com os filhos acerca do que estão a ler (59%), não existindo diferenças significativas com o acerto à questão (valor $p=0,317$);
- os pais dos alunos frequentam livrarias ou bibliotecas apenas uma ou duas vezes por mês ou todos os dias (56%), não sendo estatisticamente diferente com a proporção de acertos (valor $p=0,702$);
- os pais dos alunos mantêm diálogos diários com os filhos (57%), não existindo diferenças significativas entre a frequência dos diálogos e o acerto à questão (valor $p=0,634$);
- os pais dos alunos jantam à mesa com o filho regularmente (56%), existindo diferenças significativas entre o acerto á questão e a frequência dos jantares à mesa (valor $p=0,042$);
- os pais dos alunos conversam regularmente acerca do aproveitamento escolar (56%), não existindo diferenças significativas entre a frequência das conversas e o acerto á questão (valor $p=0,287$);

- os pais do aluno que conversam sobre livros, filmes ou programas de televisão regularmente (57%), não existindo diferenças significativas entre a frequência e o acerto à questão (valor $p=0,474$);

- os pais dos alunos, que conversam sobre questões políticas ou sociais uma ou duas vezes por semana (63%), verificando também que é estatisticamente diferente a proporção de acertos (valor $p=0,005$).

Relativamente à **menor proporção de acertos** realça-se a percentagem de apenas 18% para os alunos que não jantam diariamente com os pais.

Figura 3.19 - Proporção de alunos que responderam corretamente à questão por atividades com os filhos na área da quantidade.



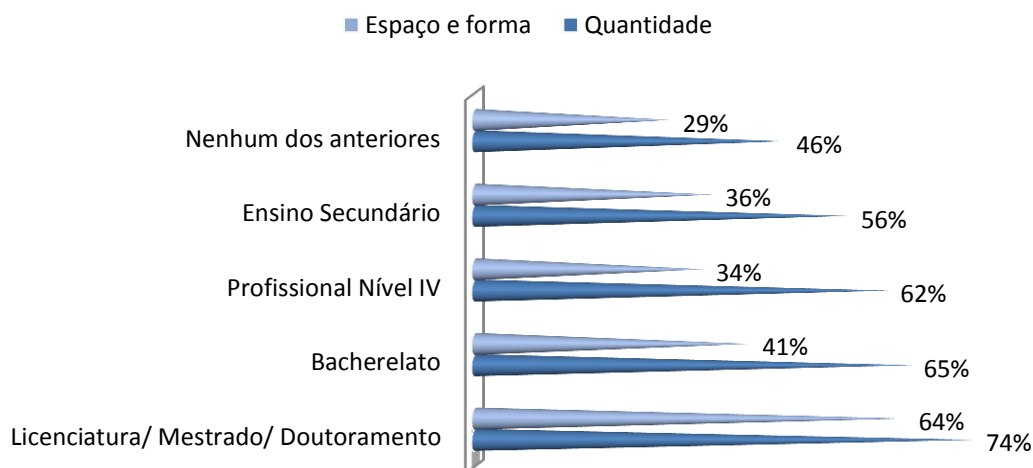
A- Conversar sobre questões políticas ou sociais; B- Conversar sobre livros, filmes ou programas de televisão; C- Conversar sobre o aproveitamento escolar do seu filho; D- Jantar à mesa com o seu filho; E- Passar tempo a conversar simplesmente com o seu filho; F- Ir a uma livraria ou a uma biblioteca com o seu filho; G- Conversar com o seu filho sobre o que ele está a ler; H- Ajudar o seu filho a fazer os trabalhos de casa.

3.2.3. Habilitações literárias dos pais

Ao analisar as habilitações literárias do Pai tendo em conta os acertos à questão espaço e forma Figura 3.20, constatou-se que a maior proporção de acertos (64%) registou-se para os filhos cujo o pai possuem um nível superior de educação, e a menor (29%) para os alunos que o pai possuem habilitações literárias inferiores ao nível secundário, existindo diferenças entre as habilitações literárias dos pais e os acertos (valor $p<0,001$). Relativamente à questão na área da quantidade Figura 3.21, constatou-se a mesma situação, isto é, a maior proporção de acertos para alunos cujo o pai possui habilitações literários de nível superior (74%) e a menor para os alunos que o pai possui um nível de escolaridade abaixo do ensino secundário (46%), a

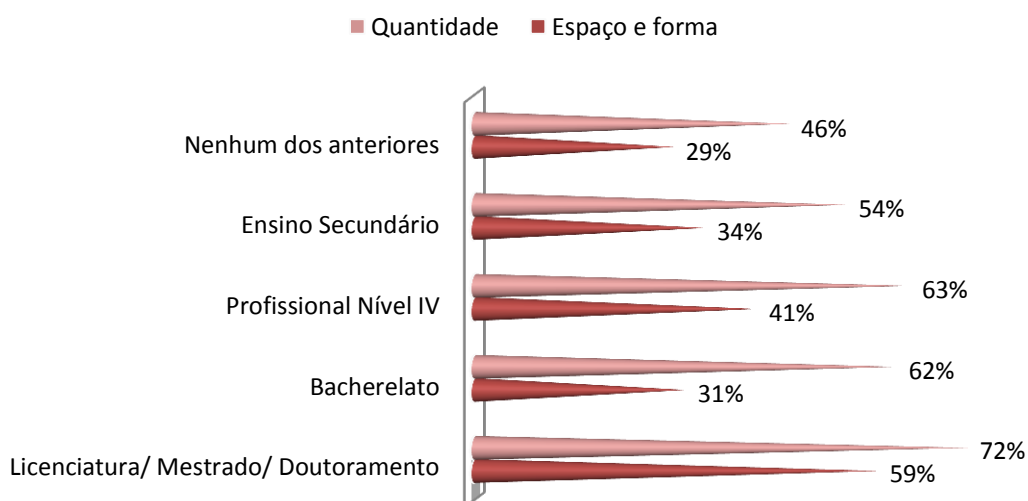
diferença entre a proporção de acerto é significativamente diferente para os diferentes níveis de habilitações literárias do pai (valor $p < 0,001$).

Figura 3.20 - Proporção de alunos que responderam corretamente à questão por habilitações literárias do pai



As mesmas diferenças significativas de proporções de acertos às duas questões registaram-se para o nível de habilitações literárias das mães (valores $p < 0,001$), sendo a maior percentagem de acertos na questão da área da quantidade (72%) e da área espaço e forma (59%) para os alunos cujas mães possuem um grau de habilitações literárias igual ou superior à Licenciatura, e a menor percentagem para os alunos que as mães apenas possuem um nível inferior ao do ensino secundário, com 46% para a área da quantidade e 29% para a área espaço e forma.

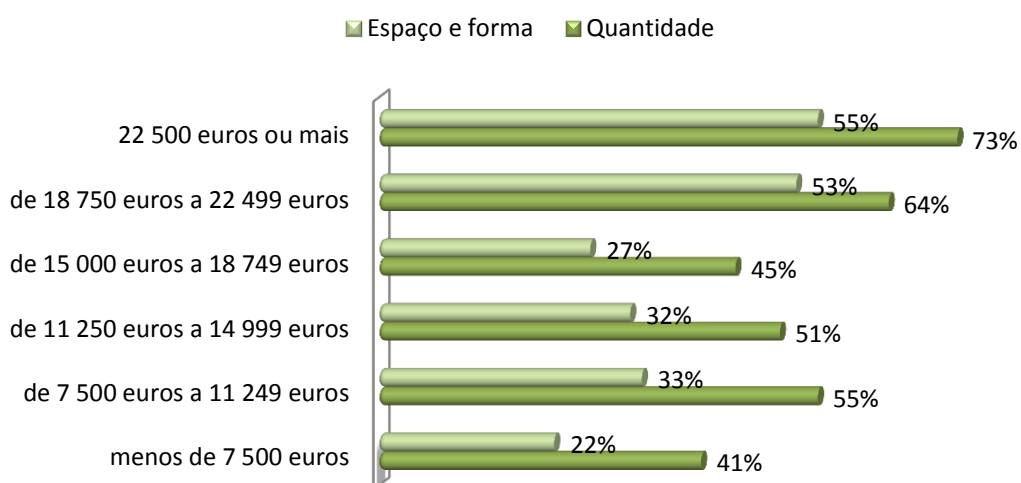
Figura 3.21 - Proporção de alunos que responderam corretamente à questão por habilitações literárias da mãe



3.2.4. Rendimento familiar do agregado familiar

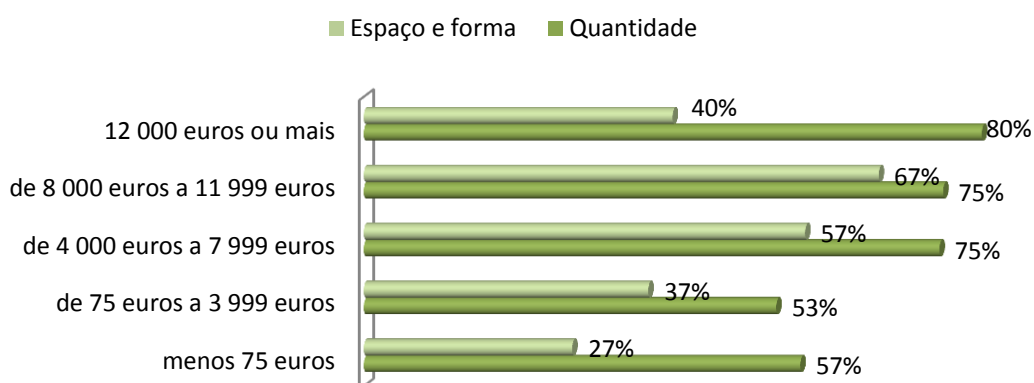
De entre os rendimento familiar dos agregados Figura 3.22, as proporções mais baixa de acertos foram obtidas pelos alunos cujos rendimentos familiares anuais são inferiores a 7500€ na área espaço e forma (22%) e na área da quantidade (41%) sendo as mais altas para os alunos com rendimentos familiares mais elevados, 55% na área espaço e forma e 73% na área da quantidade. Colocando-se a hipótese das proporções serem estatisticamente diferentes, obteve-se valores $p < 0,001$, garantindo uma diferença estatisticamente significativa entre o rendimento do agregado familiar e a proporção de acertos nas duas questões.

Figura 3.22 - Proporção de alunos que responderam corretamente à questão por rendimento bruto anual do agregado familiar



Quando analisado a proporção de acertos na área espaço e forma tendo em conta os gastos anuais do aluno constata-se que a maior proporção de acerto foi para os que apresentam gastos variáveis de 8000€ a 11999€ anuais (67%) , enquanto que a menor proporção foi para os gastos inferiores a 75€ anuais (27%) , garantindo as diferenças significativas entre os gastos individuais por aluno e o acerto à questão (valor $p=0,006$). Na área da quantidade a maior proporção verificou-se para os alunos com gastos anuais em educação superiores a 12 000 € (80%) e a menor proporção, para os gastos anuais entre 75€ e 3999 (75%), onde também é garantida a diferença significativa (valor $p=0,044$).

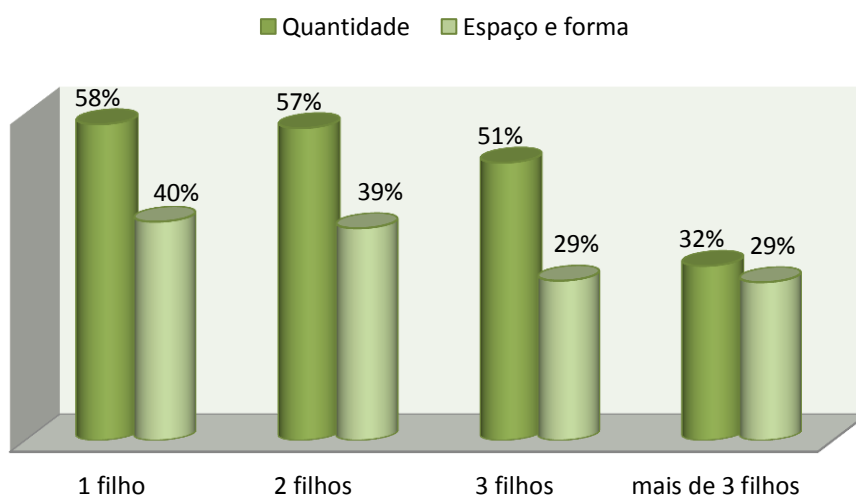
Figura 3.23 - Proporção de alunos que responderam corretamente à questão por gastos anual com o filho



3.2.5. Número de filhos

A proporção de acertos condicionalmente ao número de filhos do agregado familiar (Figura 3.24), não se revelou estatisticamente diferente na área espaço e forma (valor $p=0,194$), mas foi garantida essa diferença na área da quantidade (valor $p=0,044$).

Figura 3.24 - Proporção de alunos que responderam corretamente à questão por número de filhos



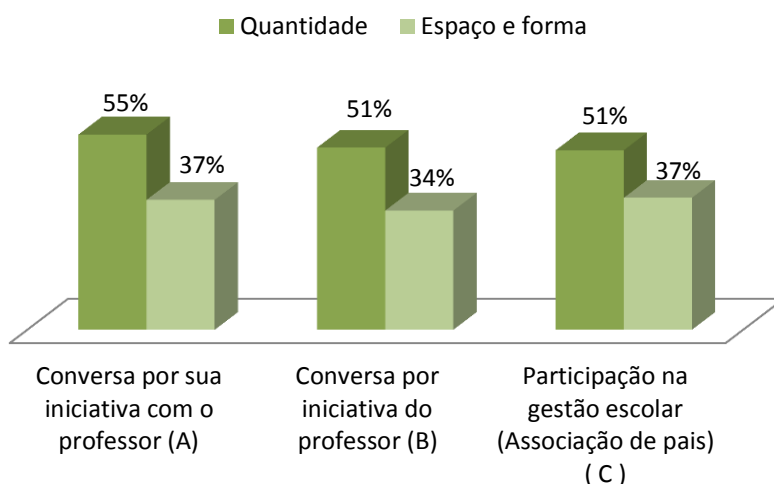
3.2.6. Participação dos pais em atividades escolares

Na área espaço e forma o acerto à questão por parte do aluno, Figura 3.25, não é significativamente diferente quando existe contactos com os professores por iniciativa dos encarregados de educação (valor $p=0,511$), no entanto se esse contato for por iniciativa dos professores a proporção de acertos é significativamente diferente (valor $p=0,009$), resultando uma menor proporção de acertos para os alunos cujos seus encarregados de educação são

convocados à escola, o mesmo se verifica para a questão na área da quantidade (valores $p=0,621$ e $p=0,006$).

Tendo em conta a participação dos encarregados de educação na gestão escolar e o acerto às questões em estudo, não se registaram associações nos acertos na questão da área de espaço e forma (valor $p=0,979$) e na questão da área da quantidade (valor $p=0,183$).

Figura 3.25 - Proporção de alunos que responderam corretamente à questão por participação na escola dos encarregados de educação:



- (A) Conversar por sua iniciativa, com um professor sobre o comportamento ou os progressos do seu filho;
 (B) Conversa a pedido de um dos professores, sobre o comportamento ou os progressos do seu filho;
 (C) Participar em atividades ligadas à gestão da escola, por exemplo, associação de pais.

3.3. Análise da percentagem de acerto dos alunos às questões - por Países

Foram selecionados para o estudo cinco países da União Europeia sendo os resultados alcançada em termos globais a Matemática dada na Tabela 3.1. Como foi referido anteriormente, os dados divulgados pela OCDE dos testes PISA derivam de variáveis plausíveis construídos através de resultados obtidos pelos alunos estando a sua construção descrita no PISA (2009).

Tabela 3.1 - Resultados obtidos por alguns países da união europeia a nível do teste PISA a matemática

País	média	Desvio padrão	Ranking OCDE cima / baixo	Ranking global cima / baixo
França (2)	497	3,1	13/22	19/28
Reino Unido (1)	492	2,4	17/25	23/31
Portugal (3)	487	2,9	22/29	28/36
Espanha (3)	483	2,1	26/29	32/36
Itália (3)	483	1,9	26/29	32/36
Grécia (3)	466	3,9	30/30	38/40

Média OCDE 496- fonte: PISA 2009 Result: What Students Know and can do, Volume I , OCDE .

- (1) Estatisticamente significativo acima da média da OCDE
 (2) Não é Estatisticamente significativo da média da OCDE
 (3) Estatisticamente significativo abaixo da média da OCDE

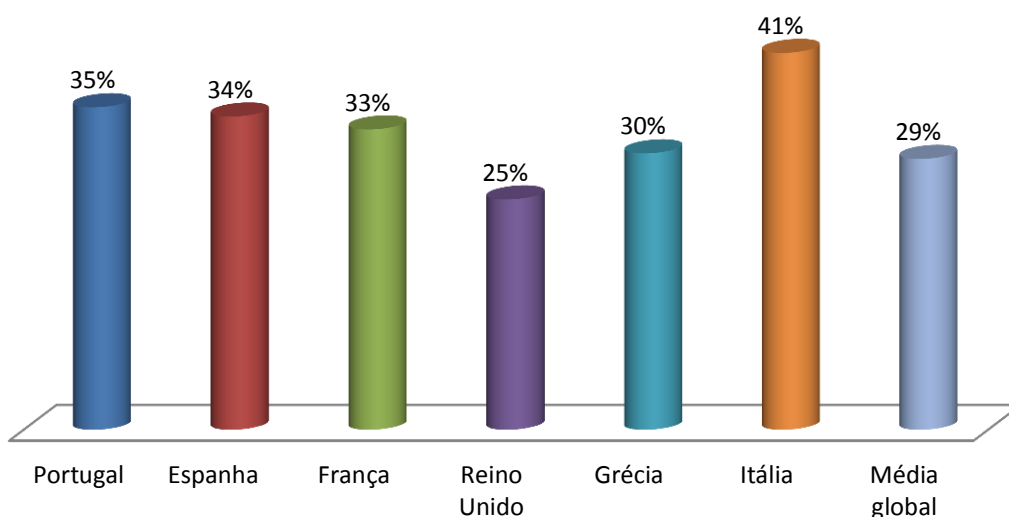
Foram contabilizadas as respostas válidas às duas questões em estudo, dos países abaixo referidos, Tabela 3.2.

Tabela 3.2 - Amostra do estudo às questões na área espaço e forma e na área da quantidade por países

País	Amostra inicial questão na área espaço e forma	Amostra após tratamento questão na área espaço e forma	Amostra inicial questão quantidade	Amostra após tratamento questão quantidade
França	986	937	861	817
Reino Unido	2524	2369	2131	2007
Espanha	6103	5851	4568	4347
Itália	6966	6672	5446	5206
Grécia	1121	1058	813	764

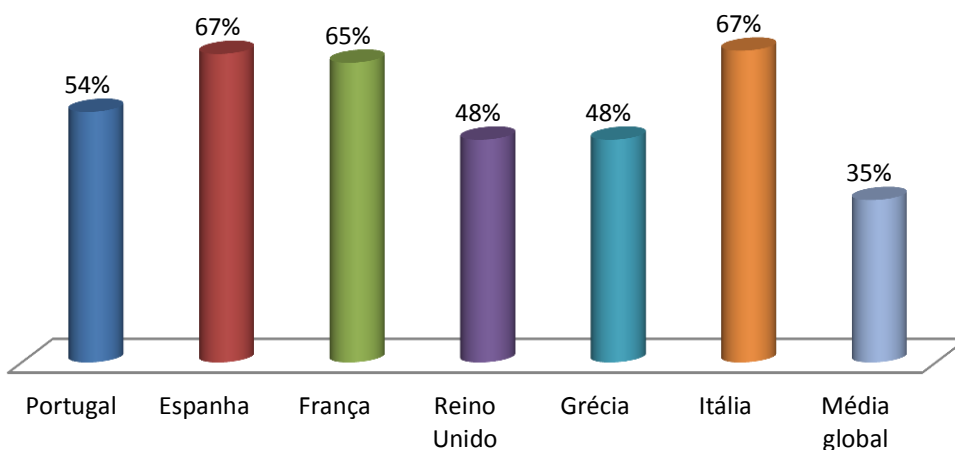
Quando analisadas as proporções dos acertos à questões na área espaço e forma nos diferentes países na questão (Figura 3.26), observa-se que o país que possui maior proporção de acertos é a Itália (41%) sendo o Reino Unido quem apresenta menor proporção de acertos (25%), abaixo do registado na média europeia. As proporções dos acertos obtidos na questão pelos países são significativamente diferentes, concluindo que a distribuição de acertos não é igual de acordo com os países (valor $p < 0,001$)

Figura 3.26 - Proporção de acertos na questão da área espaço e forma



Nos resultados obtidos na questão da área da quantidade verificou-se que, todos os países obtiveram uma proporção de acertos acima do alcançado na média global da OCDE (Figura 3.27). Constatando-se que a proporção de acertos à questão apresenta diferenças significativas entre os países (valor $p < 0,001$).

Figura 3.27 - Proporção de acertos na questão da área da quantidade



3.3.1. Género

O comportamento dos alunos nas questões das duas áreas em estudo nos diferentes países segundo o género encontra-se representado na Figuras 3.28 e Figura 3.29.

Ao se aplicar o teste de *Mantel-Haenszel* testando a independência entre os acertos em cada questão e o sexo dos alunos estratificado pelos Países, garante-se estatisticamente a não existência de uma associação entre os acertos e o sexo dos alunos (valores $p=0,197$ e $p=0,136$, nas áreas espaço e forma e área da quantidade respetivamente).

Figura 3.28 - Proporção de alunos que responderam corretamente à questão por países segundo o seu género na área espaço e forma

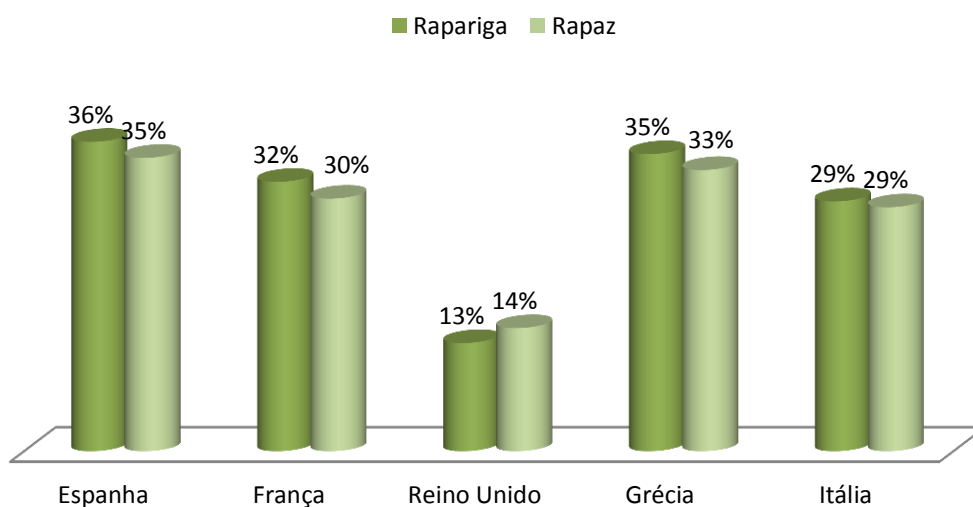
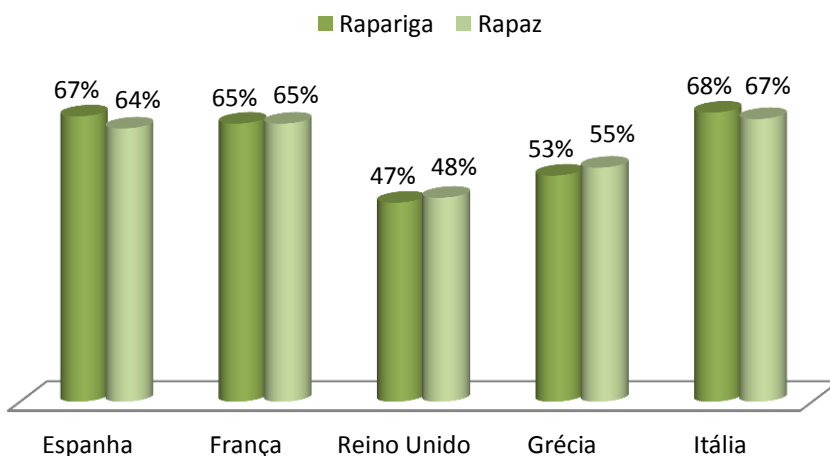


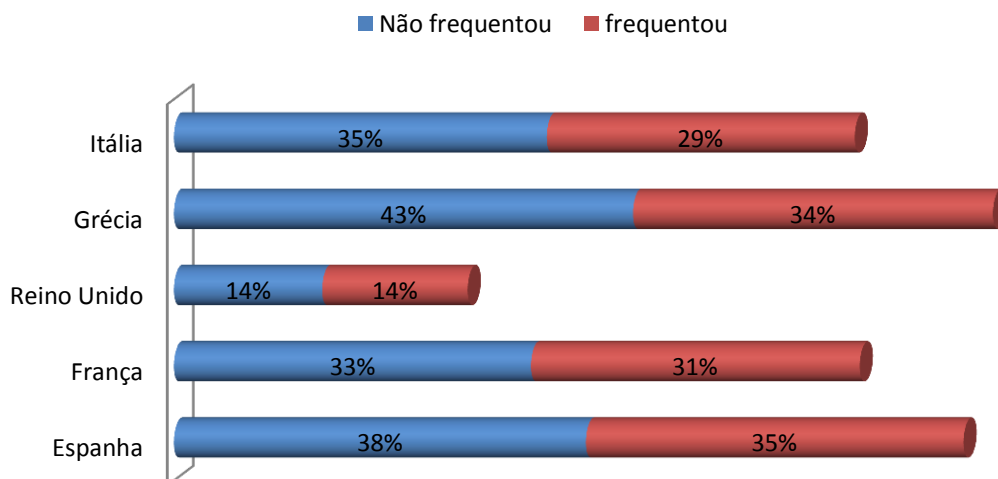
Figura 3.29 - Proporção de alunos que responderam corretamente à questão por países segundo o seu género na área da quantidade



3.3.2. Frequência do ensino pré-escolar

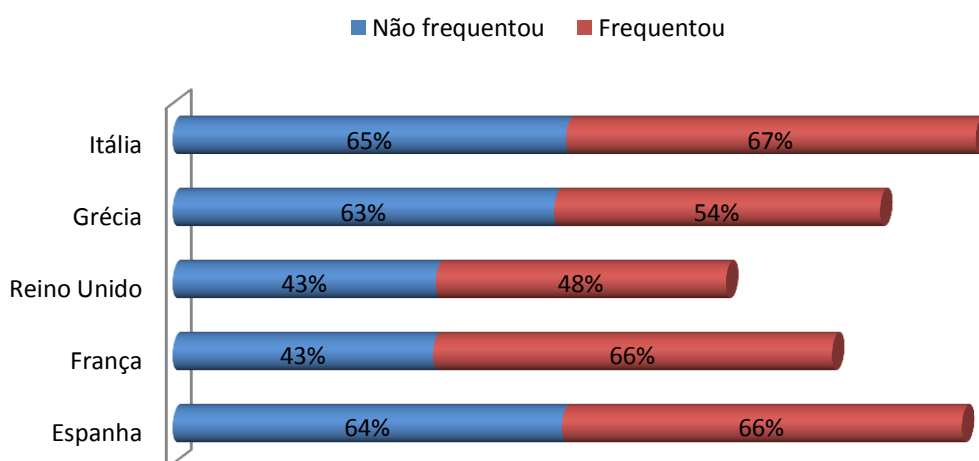
Quando analisadas as proporções de acertos na questão espaço e forma Figura 3.30, conclui-se que estratificando por países existe uma associação entre os acertos à questão e a frequência do ensino pré-escolar (valor $p=0,010$). No sentido de se verificar a existência de *odds ratios* comuns entre os países foi aplicado o teste *breslow-day* (valor $p=0,802$), que permite garantir a existência dos mesmos, $OR=0,82$ com $IC_{95\%} =]0,71; 0,95[$, logo um aluno que frequente o ensino pré-escolar possui cerca de 20% menos possibilidades de acerto à questão na área espaço e forma independentemente do país.

Figura 3.30 - Proporção de alunos que responderam corretamente por frequência do ensino pré-escolar na questão da área espaço e forma



Na questão da área da quantidade Figura 3.31, conclui-se que estratificando por países existe uma associação entre os acertos e a frequência do ensino pré-escolar (valor $p=0,053$), garantindo-se estatisticamente a existência de odds ratios comuns entre os países (valor $p=0,367$) iguais a $OR=1.151$ com $IC_{95\%} =]1,00; 1,32[$, logo um aluno que frequente o ensino pré-escolar possui 15% mais possibilidades de acerto á questão na área de quantidade independentemente do país.

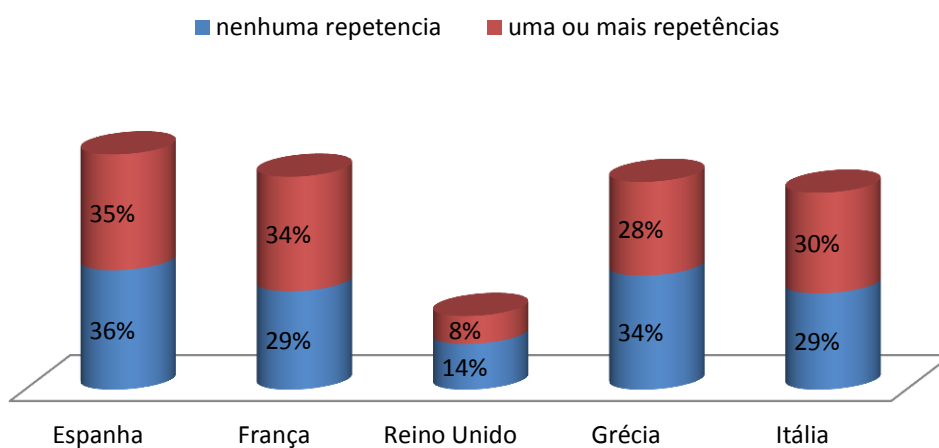
Figura 3.31 - Proporção de alunos que responderam corretamente por frequência do ensino pré-escolar na questão da área da quantidade



3.3.3. Repetências durante o percurso de estudo

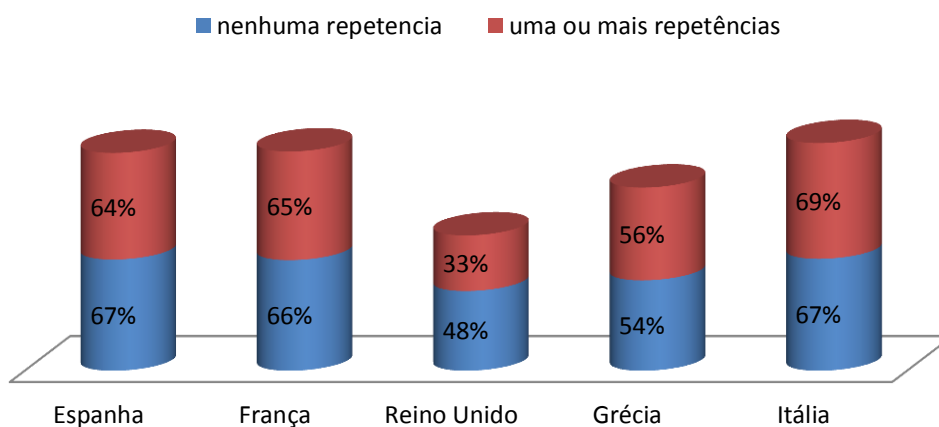
Na proporção de acertos à questão na área espaço e forma condicionalmente ao percurso escolar (Figura 3.32), verificou-se que os acertos são independentes do percurso escolar do aluno em cada país (valor $p=0,875$).

Figura 3.32 - Proporção de alunos que responderam corretamente por repetências na questão da área espaço e forma



Analisada a questão na área da quantidade (Figura 3.33) , também é verificada a mesma tendência garantindo-se estatisticamente que estratificando por país, os acertos à questão são independentes do percurso escolar do aluno ($p=0,296$).

Figura 3.33 - Proporção de alunos que responderam corretamente por repetências na questão da área da quantidade

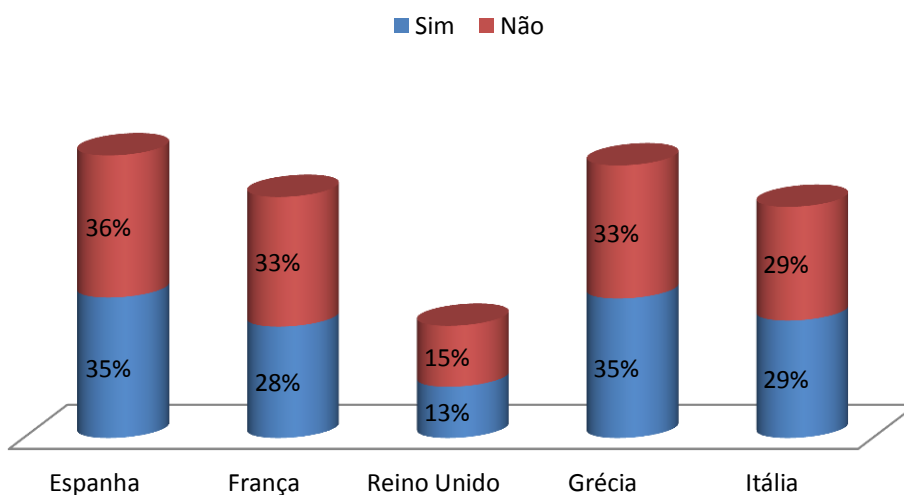


3.3.4. Uso de software de programas educativos

Nos últimos anos a existência de programas educativos, nomeadamente a Matemática, permitiu aos alunos uma exploração diferente de conteúdos, deixando assim unicamente o “lápiz e papel” .

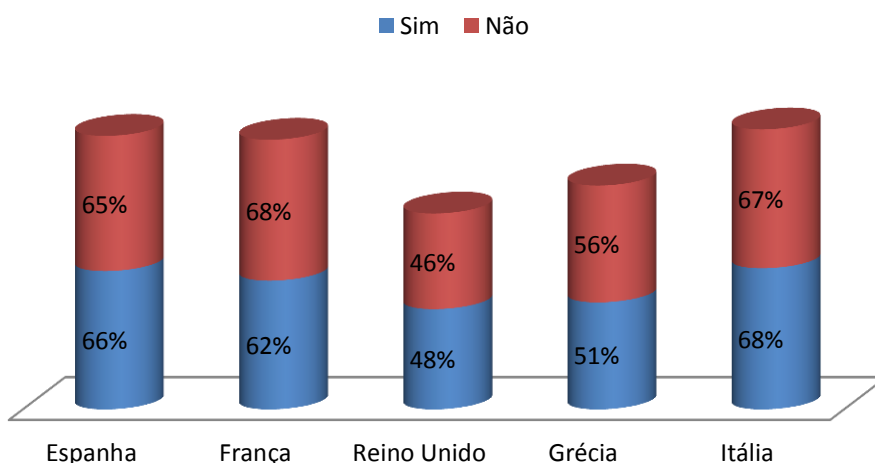
Quando comparadas as proporções acertos à questão na área espaço e forma com o recurso aos *software* educativos condicionalmente aos países (Figura 3.34), garante-se estatisticamente a não existência de uma associação (valor $p=0,704$).

Figura 3.34 - Proporção de alunos que responderam corretamente por uso de programas *software* educativos na questão da área de espaço e forma



Quando analisadas na mesma perspetiva a proporção de acertos na área da quantidade (Figura 3.35), não existe estisticamente associação entre a proporção de acertos e o uso do *software* em cada país estudado (valor $p=0,469$).

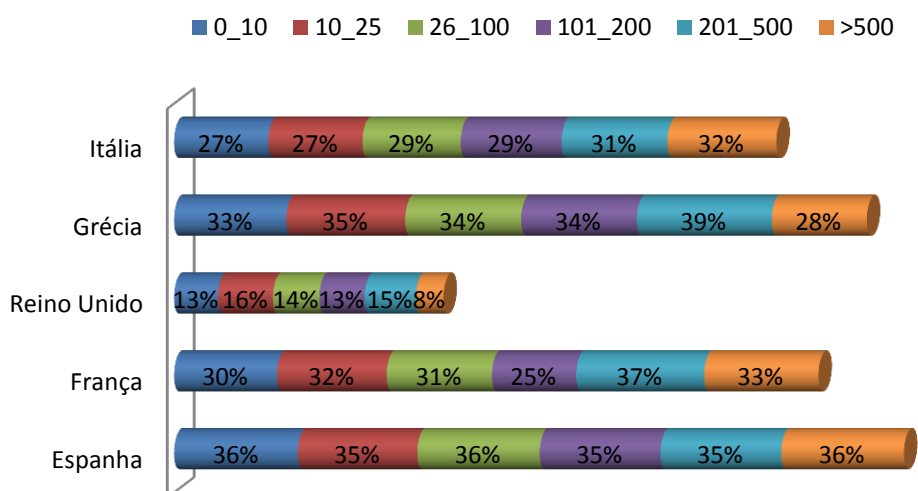
Figura 3.35 - Proporção de alunos que responderam corretamente por uso de programas *software* educativos na questão da área da quantidade



3.3.5. Hábitos de leitura fora da escola

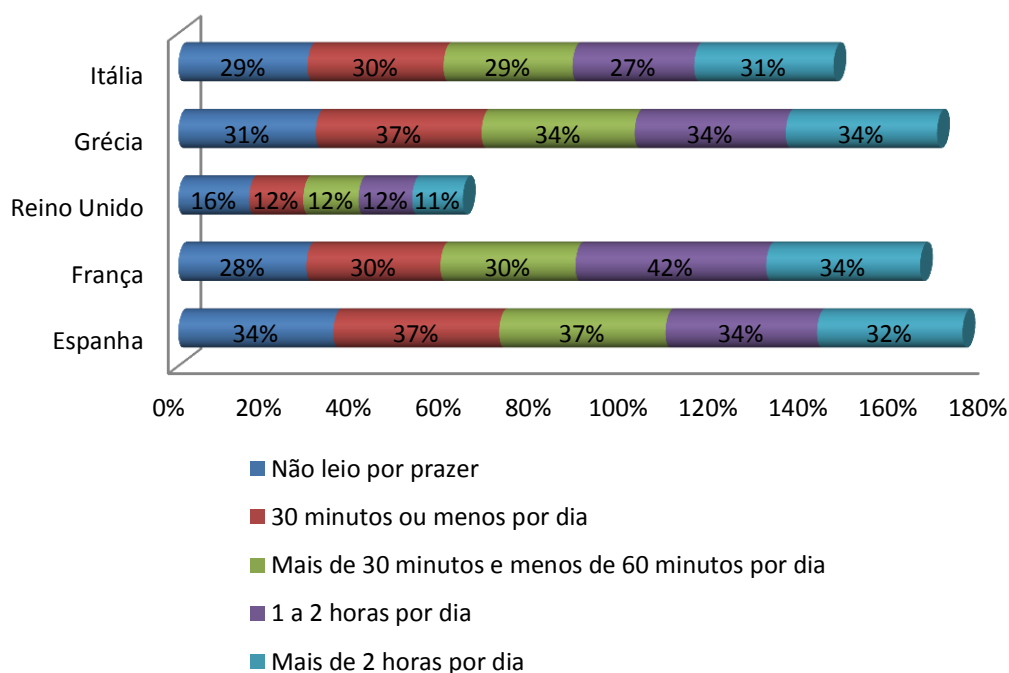
A proporção de acerto à questão na área espaço e forma nos países em estudo (Figura 3.36), não revelou ser diferente tendo em conta o número de livros que os alunos possuem em casa (Anexo 21).

Figura 3.36 - Proporção de alunos que responderam corretamente por número de livros em casa à questão da área de espaço e forma



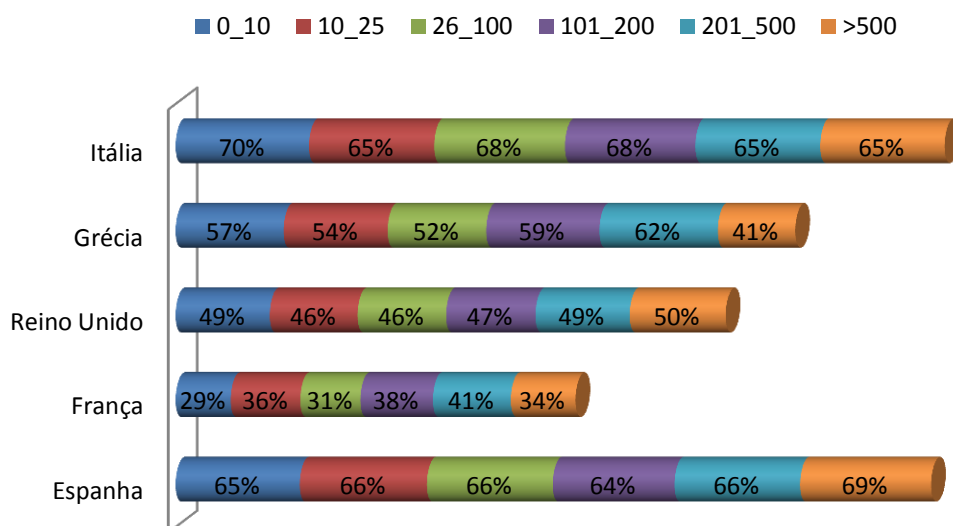
Da mesma forma ao analisar o tempo dedicado à leitura diária dos alunos que responderam à questão (Figura 3.37) não se encontraram associações com a proporção de acertos (Anexo 21).

Figura 3.37 - Proporção de alunos que responderam corretamente por tempo diário dedicado à leitura à questão da área espaço e forma



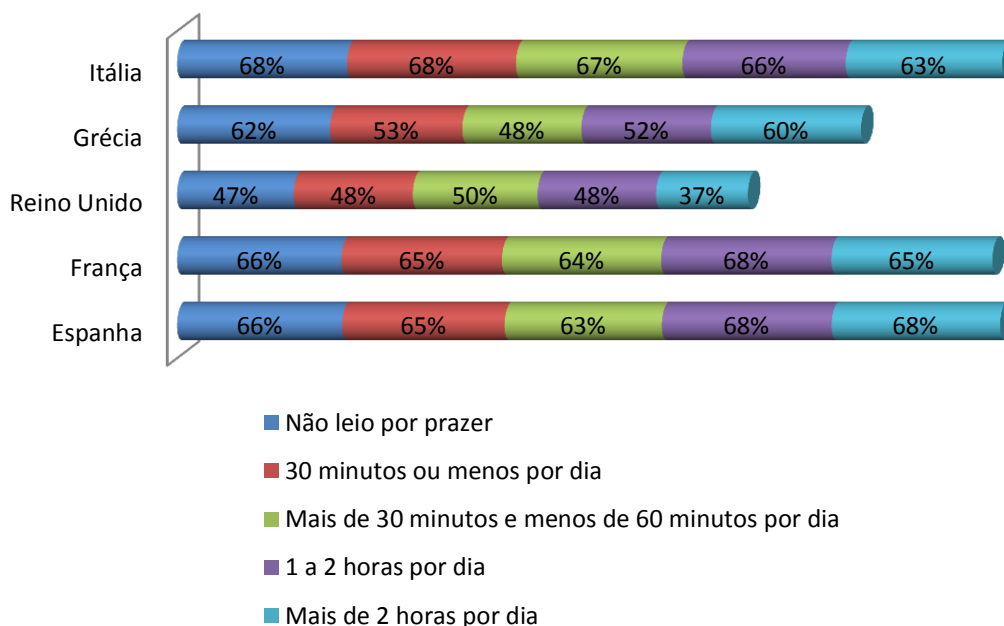
Na questão da área da quantidade (Figura 3.38), à semelhança do verificado anteriormente não se registou uma associação entre o número de livros em casa e o desempenho à questão dos alunos nos diferentes países (Anexo 21), assim como o tempo dedicado à leitura pelos alunos (Figura 3.39), não registou associação com os acertos à questão (Anexo 21).

Figura 3.38 - Proporção de alunos que responderam corretamente por número de livros em casa à questão da área da quantidade



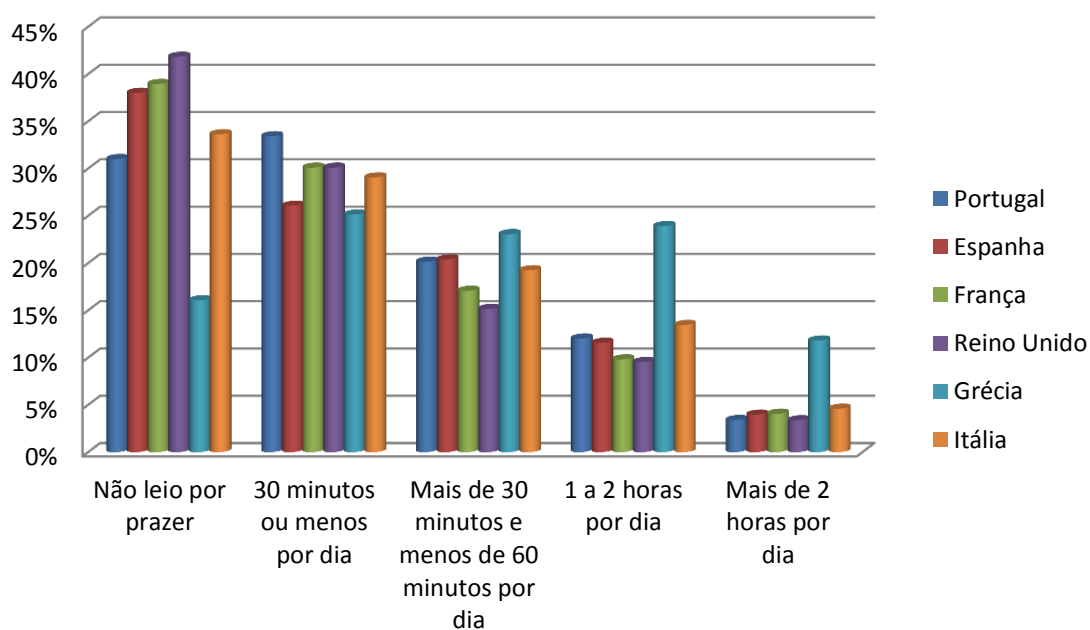
Relativamente ao tempo dedicado à leitura dos alunos nos diferentes países, este fato não revelou estar associado à proporção de acertos à questão na área da quantidade (Anexo 21).

Figura 3.39 - Proporção de alunos que responderam corretamente por tempo diário dedicado à leitura à questão da área da quantidade



Analisando o comportamento dos hábitos de leitura dos alunos de 15 anos nos diferentes países Figura 3.40, verificamos que o Reino Unido(42%), França (39%), Espanha (38%) Itália (34%) e Portugal (31%) apresentam percentagens altas de alunos que não lêem, contra a Grécia que apresenta uma percentagem mais reduzida de 16%.

Figura 3.40 - Proporção de alunos que responderam corretamente por tempo dedicado à leitura por país



3.4. Comparação entre os alunos Portugueses e os alunos dos países Europeus

O estudo comparativo entre os diferentes países e as variáveis seleccionadas, teve como objetivo verificar qual o comportamento dos alunos relativamente às duas questões.

Os rapazes, apenas em Portugal revelaram um aumento na proporção de acertos na questão da área de espaço e forma relativamente aos restantes países do estudo. No entanto, no global o teste Pisa identificou 35 países, dos 65 participantes, onde o comportamento dos rapazes era significativa superior e somente 5 países onde essa vantagem era atribuída às raparigas, PISA (2009) verificando-se assim a maior predominância dos rapazes face às raparigas em questões ligadas à vida real.

A frequência do ensino pré-escolar nos países em estudo, é à semelhança do registado em Portugal um fator que aumenta as possibilidades de acerto na questão da área de quantidade, revelando a importância do ensino no processo de construção do individuo, no entanto é de realçar que a frequência deste nível de escolaridade diminuiu as possibilidades de acerto à questão na área espaço e forma nos países estudados.

As repetências dos alunos Portugueses demonstraram ser negativas para o sucesso às questões não sendo acompanhadas pelos restantes países, este é um fator que é dado como adquirido pelos professores Portugueses, sendo necessária uma mudança de mentalidades e por conseguinte a recuperação dos jovens no sentido de adquirirem as competências necessárias para a resolução de questões onde seja necessária uma elevada literacia matemática nas áreas em estudo.

Os resultados obtidos, relativamente ao acesso a livros em casa e tempo dedicado à leitura dos alunos, não revelaram associações com os acertos nos países em estudo, contrariamente ao registado em Portugal.

Capítulo 4: Definição de perfis – área de conteúdos espaço e forma e área de conteúdos quantidade

O objetivo desta secção será identificar algumas características que expliquem o sucesso às questões na área espaço e forma e na área da quantidade e assim, traçar perfis mais prováveis de alunos e agregados familiares.

Construíram-se modelos de regressão logística (Hosmer e Lemeshow (2000)) cujos coeficientes, variável resposta e algumas especificidades relativas à sua avaliação e bondade do ajustamento serão apresentados.

4.1. Perfil do aluno: área espaço e forma

Para construção do modelo a variável resposta assume o valor 0 se o aluno respondeu incorretamente e o valor 1 se o aluno respondeu corretamente (Anexo 4).

No Anexo 5 apresentam-se os coeficientes das covariáveis nos modelos logísticos univariados, o seu desvio padrão, valores p associados às estatísticas de Pearson, de Wald e da razão de verossimilhanças, *odds ratios* e respetivos intervalos de confiança a 95% para a questão de nível 5 na área espaço e forma relativamente às variáveis do inquérito realizado aos alunos.

Numa primeira fase alguns fatores revelaram ser potenciadores ou não, para o aumento das possibilidades de acerto.

As variáveis que se revelaram potenciadoras foram:

- ◆ as regiões do Centro, Norte e grande Lisboa comparativamente com a região do Alentejo;
- ◆ os alunos que frequentam o 9º, 10º ou 11º anos;
- ◆ os alunos do sexo masculino;
- ◆ os alunos que frequentaram o ensino pré escolar mais do que dois anos;
- ◆ os progenitores trabalharem a tempo inteiro;
- ◆ o aluno possuir uma secretária para estudar;
- ◆ ter acesso em casa a livros;
- ◆ possuir hábitos de leitura diários;
- ◆ ao estudarem não decorarem os conteúdos;
- ◆ quando estudam, reflectem acerca do que estão a ler;
- ◆ quando estudam, começam por tentar perceber o que é mesmo preciso aprender;

- ◆ quando estudam, tentam identificar os conceitos que ainda não percebem bem;
- ◆ não frequência de aulas suplementares a Matemática.

Enquanto as variáveis significativas que se revelaram protetoras logo diminuem as possibilidades de acerto são:

- ◆ as regiões do Algarve e Ilhas comparativamente com o Alentejo;
- ◆ o ensino na escola pública;
- ◆ a frequência do 7º e 8º anos de escolaridade;
- ◆ as repetências durante o percurso escolar;
- ◆ não possuir um quarto individual;
- ◆ não possuir programas de computadores educativos;
- ◆ não possuir ligação de Internet em casa
- ◆ quando estudam não relacionam a informação com os conhecimentos já adquiridos.

Seguindo a metodologia descrita no capítulo 2 subsecção 2.1 foi construído um modelo de regressão logística, cujos os seus coeficientes, desvio padrão e valor p do teste *Wald* se encontram representados na Tabela 4.1, validação e bondade do ajustamento.

Tabela 4.1 - Coeficientes do modelo para o acerto à questão na área espaço e forma-questionário aos alunos, desvio padrão e valor p teste Wald correspondentes

Variável	Coeficientes	Desvio Padrão Coeficientes	Valor p (Wald)
Constante	-3.66	0.621	0.000
Zona Geográfica (1)			
Centro	0.642	0.198	0.0013**
Grande Lisboa/ Península de Setúbal	0.272	0.215	0.205
Norte	0.541	0.185	0.003**
Ano de escolaridade do aluno (2)			
9º ano, 10ºe 11º ano	1.113	0.507	0.028*
Sexo do aluno			
Masculino	1.234	0.190	0.001
Frequência do ensino pré escolar (3)			
Sim	0.465	0.196	0.018
Repetências (4)			
Sim	-1.228	0.339	0.001
Programas de computador educativos em casa			
Não	-0.249	0.148	0.092
Quantidade de livros em casa (5)			
de 26 a 100 livros	0.437	0.177	0.013 **
de 101 a 500 livros	0.846	0.186	0.001 ***
mais de 500 livros	1.161	0.278	0.001 ***
Estudar a decorar (6)			
muitas vezes	-0.329	0.146	0.024*
Tentar perceber o que é preciso aprender (6)			
muitas vezes	0.696	0.359	0.052
Verificas se está a perceber o que já leu (6)			
muitas vezes	0.956	0.247	0.001
Sexo do aluno X Tentar perceber o que é preciso aprender			
masculino X muitas vezes	-0.7286	0.247	0.009**
Tentar perceber o que é preciso aprender X Verificas se está a perceber o que já leu			
muitas vezes X muitas vezes	-0.823	0.353	0.0198*
Tentar perceber o que é preciso aprender X repetências			
muitas vezes X já repetiu	-1.089	0.544	0.045*

- (1) Categoria de referência – Alentejo, Algarve e Ilhas
 (2) Categoria de referência – 7º e 8ºano de escolaridade
 (3) Categoria de referência – Não frequentou o ensino pré escolar
 (4) Categoria de referência – não ter nenhuma repetências nos três primeiros ciclos
 (5) Classe de referência – de 0 a 25 livros
 (6) Classe de referência – quase nunca , às vezes e quase sempre

Para o modelo utilizaram-se as covariáveis:

- ◆ Zona geográfica (Norte, Centro, grande Lisboa e uma categoria com Alentejo, Algarve e Ilhas);
- ◆ Ano de escolaridade (foram agrupados os alunos do 7º e 8º anos e os alunos do 9º, 10º e 11º anos);
- ◆ Sexo;

- ◆ Frequência do ensino pré-escolar (foram agrupados os alunos com um ano e mais anos de frequência);
- ◆ Repetências (foram agrupados os alunos que apresentam uma ou mais repetências no 1º, 2º e 3º ciclos);
- ◆ Uso de programas informáticos educativos em casa;
- ◆ Número de livros em casa (sendo divididos em 4 classes, de 0 a 24, de 25 a 100, de 100 a 500 e mais de 500 livros)
- ◆ Quando estuda, tenta decorar tudo o que está no texto;
- ◆ Quando estuda, começa por tentar perceber o que é mesmo preciso aprender; e
- ◆ Quando estuda, verifica se está a perceber o que já leu.

Elaborada uma verificação de pressupostos e uma análise dos resíduos do modelo encontrado recorrendo aos resíduos *deviance* e aos resíduos de *Pearson* assim como aos resíduos *dfbetas* e distância de Cook (Anexo 6) para avaliação de *outliers* e de observações influentes. Na análise efetuada foram identificadas seis observações que se destacavam, tendo sido analisada a sua influência no modelo como se apresenta no Anexo 7 As maiores diferenças encontradas nos coeficientes apresentavam valores próximos de 20%, não justificando uma análise mais profunda.

Posteriormente, foi realizada uma validação cruzada por *bootstrap*, onde foram ajustados modelos a 10 amostras aleatórias constituídas por 90% das observações da amostra original, registando-se para cada um dos modelos, os valores estimados para os 10% de observações que ficaram de fora em cada modelo e registando os valores das estatísticas qui-quadrado do teste de ajustamento, *hosmer e Lemeshow* realizado a cada modelo, bem como a estatística *do R de Nagelkerque*.

O modelo ajustado revelou um bom ajustamento aos dados pelo teste de bondade de ajustamento de Hosmer e Lemeshow ($\chi^2_8=3,21$ valor $p = 0,92$) e valor de *R de Nagelkerque* igual a 0,29, registando-se um valor médio do qui-quadrado para os 10 modelos obtidos através da validação cruzada por *bootstrap*, de 4,4 com um mínimo de 1,62 e um máximo de 10,1. O modelo ajustado tem uma capacidade discriminativa aceitável (AUC=0,77; IC95% =]0,75; 0,80]) com uma sensibilidade igual a 70,3% e uma especificidade igual a 69,0% para um ponto de corte igual a 0,39, (Anexo 8). Para a validação do modelo, obtemos uma AUC igual a 0,78 (IC_{95%} =]0,75; 0,79]) que é reveladora da consistência interna do mesmo (Anexo 8).

$$\begin{aligned}
g(x) = & -3,66 + 0,642(\text{zona centro}) + 0,272(\text{zona grande lisboa e Setúbal}) \\
& + 0,541 (\text{zona norte}) + 1,113(9, 10^{\text{º}} \text{ ou } 11^{\text{º}}\text{ano}) + 1,234(\text{Masculino}) \\
& + 0,465(\text{frequentou pré escolar}) - 1,228(\text{Repetências}) \\
& - 0,249(\text{programas comp. educativos}) + 0,437(\text{livros [26; 100]}) \\
& + 0,846(\text{livros [101; 500]}) + 1,161(\text{ mais de 500}) \\
& - 0,329(\text{metodologia estudo A}) + 0,696(\text{metodologia estudo B}) \\
& + 0,956(\text{metodologia estudo F}) \\
& - 0,7286(\text{Masculino} * \text{metodologia de estudo B}) \\
& - 0,823(\text{metodologia de estudo B} * \text{metodologia estudo F}) \\
& - 1,089(\text{metodologia de estudo B} * \text{Repetências})
\end{aligned}$$

Com,

metodologia estudo A – Quando estudo, tento decorar tudo o que está no texto.

metodologia estudo B – Quando estudo, começo por tentar perceber o que é mesmo preciso aprender.

metodologia estudo F – Quando estudo, verifico se estou a perceber o que já li.

A partir do modelo obtido é possível retirar as seguintes conclusões para um aluno que respondeu corretamente à questão que traduz as competências nível 5 pelo estudo PISA na área espaço e forma.

Os fatores que aumentam a probabilidade de acerto são:

- Viver na zona norte e centro de Portugal;
- Possuir um nível de escolaridade acima do 8º ano;
- Ser do sexo masculino;
- Ter frequentado o ensino pré-escolar;
- Não ter repetido nenhum ano;
- Usar programas de computadores educativos em casa;
- Ter acesso a livros em casa;
- Estudar não decorando conteúdos;
- Estudar tentando muitas vezes perceber o que é mesmo preciso aprender;
- Estudar não verificando muitas vezes se está a perceber o que está a ler.

Fixando as restantes covariáveis do modelo, as conclusões a retirar são as seguintes (Anexo 9):

- ◆ Um aluno que estude na **zona Centro** tem cerca do **dobro das possibilidades de acerto**, relativamente a um aluno que estude **na zona do Alentejo, Algarve ou Ilhas**.

(IC_{95%} =]1,3; 2,8[]); já se estudar na **zona Norte**, essas possibilidades passam **para uma vez e meia mais** (IC_{95%} =]1,2; 2,5[]);

- ◆ Um aluno que **frequente mais do 9º ano** de escolaridade tem **três vezes mais possibilidade** de acerto à questão do que um aluno que **frequente o 7º ou 8º ano** (IC_{95%} =]2,1; 4,4[]);
- ◆ Um aluno que iniciou os seus estudos no **ensino pré-escolar** tem cerca de **uma vez e meia mais possibilidades** de responder corretamente à questão (IC_{95%} =]1,1; 2,4[]);
- ◆ Um aluno que tenha acesso em casa a um **número de livros entre 26 e 100** tem cerca de **uma vez e meia mais** possibilidade do que um aluno que possua **menos de 25 livros** (IC_{95%} =]1,1; 2,2[]); se possuir entre **101 e 500 livros** as suas possibilidades de acerto **aumentam para cerca de duas vezes e meia mais** (IC_{95%} =]1,6; 3,4[]), e se possuir **mais do que 500 livros** tem **três vezes mais** possibilidades de acerto (IC_{95%} =]1,9; 5,5[]);
- ◆ Um aluno que dispõe de **programas de computador educativos** em casa tem **mais 1/4 de possibilidades** de acerto (IC_{95%} =]0,96; 1,71[]);
- ◆ Caso o aluno **não use como metodologia de estudo decorar muitas vezes** tudo o que está no texto tem **40% mais de possibilidades** de vir a responder corretamente à questão (IC_{95%} =]4%; 85%[]);
- ◆ **Um rapaz** que ao **estudar tente inicialmente perceber o que é necessário aprender** tem **uma vez e meia mais** possibilidades de acerto do que se **for rapariga** usando a **mesma metodologia de estudo** (IC_{95%} =]1,1; 2,5[], (Nota 1);
- ◆ **Um rapaz** que ao estudar **inicialmente não comece por tentar perceber o que é necessário aprender** tem **3 vezes e meia mais** possibilidades de acerto do que uma **rapariga** que não usa a **mesma metodologia de estudo** (IC_{95%} =]2,4; 5,0[], (Nota 1);
- ◆ **Um aluno** que ao estudar não tente **inicialmente perceber o que é necessário aprender** e que **se questiona se esta a entender o que está a ler**, tem **duas vezes e meia mais** possibilidades de acertos comparativamente com um **aluno que não se questiona** (IC_{95%} =]1,6; 4,2[]);
- ◆ **Um aluno** que ao estudar tente **inicialmente perceber o que é necessário aprender** e é **repetente**, tem **sete vezes e meia menos** possibilidades de acertos comparativamente com um aluno que **não repetente** (IC_{95%} =]3,3; 17[]); No caso do aluno ao estudar **não tente inicialmente perceber o que é necessário aprender**, o ser **repetente** leva a **três vezes e meia menos** de possibilidade de acerto comparativamente com **um aluno que nunca repetiu** (IC_{95%} =]2,1; 6,2[]);

- ◆ Um **rapaz, repetente** que ao estudar **verifica muitas vezes se está a aprender**, o fato de **quando inicia o estudo tenta perceber o que é necessário aprender** tem associado uma **redução de 7 vezes na possibilidade** de acerto à questão comparativamente com um rapaz que não o faz ($IC_{95\%} =]2,4; 20[$); já no caso de ser **rapariga, repetente** e ao **estudar verificar muitas vezes se está a aprender**, o fato de **quando inicia o estudo tenta perceber o que é necessário aprender** tem associado uma **redução de cerca de três vezes e meia** na possibilidade de acerto à questão comparativamente com uma rapariga que não o faz ($IC_{95\%} =]1,1; 10[$);
- ◆ Uma **rapariga, não repetente** que ao estudar **verifica muitas vezes se está a aprender**, o fato de **quando inicia o estudo tenta perceber o que é necessário aprender** aumenta em **5 vezes a possibilidade** de acerto à questão comparativamente com uma rapariga que não o faz ($IC_{95\%} =]1,8; 15[$); Já no caso de ser rapaz na mesma situação o fato de **quando inicia o estudo tenta perceber o que é necessário aprender** diminui em cerca de **2 vezes e meia a possibilidade** de acerto à questão comparativamente com um rapaz que não o faz ($IC_{95\%} =]1,5; 3,7[$);

Nota 1:

Cálculo para obtenção dos OR e respetivos intervalos de confiança no caso de interação significativa sexo do aluno *versus* metodologia de estudo “Quando estudo, começo por tentar perceber o que é mesmo preciso aprender.”

Calculo das possibilidades na comparação entre os rapazes e as raparigas mantendo constante a metodologia de estudo.

$$\hat{g}_1(x) = \ln(OR) = g(\text{Sexo} = 1, \text{Metodologia B} = 1) - g(\text{Sexo} = 0, \text{Metodologia B} = 1)$$

$$\begin{aligned} \ln(OR) &= -3,66 + 1,234(\text{Masculino}) + 0,696(\text{metodologia estudo B}) \\ &\quad - 0,7286(\text{Masculino} * \text{metodologia de estudo B}) \\ &\quad - (-3,66 + 1,234(\text{Feminino}) + 0,696(\text{metodologia estudo B}) \\ &\quad - 0,7286(\text{Feminino} * \text{metodologia de estudo B})) \end{aligned}$$

$$\begin{aligned} \ln(OR) &= -3,66 + 1,234 \times 1 + 0,696 \times 1 - 0,7286 \times 1 \times 1 - (-3,66 + 1,234 \times 0 + 0,696 \\ &\quad \times 1 - 0,7286 \times 0 \times 1) \end{aligned}$$

$$\ln(OR) = 0,507$$

$$OR = \exp(0,507) = 1,66$$

Com intervalo de confiança a 95% dado por:

$$IC_{95\%} = \exp\left(\hat{g}_1(x) \pm 1,96\sqrt{V\hat{a}r\hat{g}_1(x)}\right)$$

Com,

$$\begin{aligned} Var(\hat{g}_1(x)) &= V\hat{a}r(Masculino) + V\hat{a}r(metodologia estudo B) \\ &+ 2Cov(Masculino, metodologia de estudo B) \end{aligned}$$

Sendo que $V\hat{a}r(\text{nome da variável})$ significa que se estás a referir à variância do coeficiente dessa variável no modelo.

Com recurso ao *software R-GUI* foi construída a matriz de variâncias e covariâncias do modelo, obtendo-se assim,

$$V\hat{a}r(\hat{g}_1(x)) = 0,0361 + 0,0784 + 2 \times (-0,0357) = 0,0431$$

$$\hat{g}(x) \pm 1,96\sqrt{Var(\hat{g}_1(x))} = (0,507 \pm 1,96 \times \sqrt{0,0431}) = (0,100; 0,914)$$

$$IC_{95\%} = (\exp(0,100); \exp(0,914)) = (1,10 ; 2,50)$$

$$\hat{g}_2(x) = \ln(OR) = g(\text{Sexo} = 1, \text{Metodologia B} = 0) - g(\text{Sexo} = 0, \text{Metodologia B} = 0)$$

$$\begin{aligned} \ln(OR) &= -3,66 + 1,234(Masculino) + 0,696(metodologia estudo B) \\ &- 0,7286(Masculino * metodologia de estudo B) \\ &- (-3,66 + 1,234(Feminino) + 0,696(metodologia estudo B) \\ &- 0,7286(Feminino * metodologia de estudo B)) \end{aligned}$$

$$\begin{aligned} \ln(OR) &= -3,66 + 1,234 \times 1 + 0,696 \times 0 - 0,7286 \times 1 \times 0 - (-3,66 + 1,234 \times 0 + 0,696 \\ &\times 0 - 0,7286 \times 0 \times 0) \end{aligned}$$

$$\ln(OR) = 1,234$$

$$OR = \exp(1,234) = 3,43$$

Com intervalo de confiança a 95% dado por:

$$IC_{95\%} = \exp\left(\hat{g}_2(x) \pm 1,96\sqrt{Var\hat{g}_2(x)}\right)$$

Com, $V\hat{a}r(\hat{g}_2(x)) = V\hat{a}r(Masculino) = 0,0361$

$$\hat{g}(x) \pm 1,96\sqrt{V\hat{a}r(\hat{g}_2(x))} = (1,234 \pm 1,96 \times \sqrt{0,0361}) = (0,862; 1,606)$$

$$IC_{95\%} = (\exp(0,862) ; \exp(1,606)) = (2,37 ; 4,99)$$

4.1.1. Consideração ao perfil do aluno – área espaço e forma

As competências atingidas por um aluno que respondeu corretamente à questão em estudo na área espaço e forma visaram a capacidade de resolver problemas executando processos múltiplos e sequenciais, que impliquem o uso de hipóteses, usando o raciocínio espacial, argumentação e perspicácia para identificar informação relevante e para interpretar e ligar diferentes representações.

Observou-se uma assimetria de conhecimentos revelados pelos nossos alunos, tendo em conta a sua situação geográfica, sendo as zonas menos populosas desfavorecidas.

Uma vez que o teste PISA não pretende testar currículos, mas sim os conhecimentos dos alunos perante situações da vida real, após um ensino de mais ou menos 9 anos de escolaridade, verifica-se que os alunos que frequentam os níveis mais baixos de escolaridade não conseguem responder às exigências pedidas.

A frequência do ensino pré-escolar também se mostrou significativa para o sucesso dos alunos, desta forma, a aposta dos últimos governos numa educação pré-escolar obrigatória terá resultados no desenvolvimento dos jovens durante o seu percurso escolar.

Um dos fatores que se evidenciou como fator protetor, foi o acesso a livros em casa, uma vez que os alunos que os possuem têm mais possibilidades de adquirirem competências para a resolução da questão, revelando a importância dos livros no processo desenvolvimento dos jovens.

As repetências revelaram-se um fator muito negativo para o acerto da questão, afetando mais os rapazes, o que nos leva a pensar que são alunos repetentes onde a aplicabilidade da matemática em questões mais complexas não são entendidas.

Os alunos que ao estudarem muitas vezes se questionam acerca do que estão a aprender apresentaram mais dificuldades, revelando que a insegurança durante o estudo afetou a capacidades destes de conseguirem atingir um nível de compreensão esperado.

A terminar refira-se que o perfil mais provável de um aluno de 15 anos que atinge um nível 5 de competência na área espaço e forma é do sexo masculino, residente na zona norte ou centro de Portugal, que iniciou o percurso de ensino no pré-escolar e não teve repetências durante o seu percurso escolar, que frequente mais do que o 8º ano de escolaridade, com acesso a livros em casa e material informático didático, que ao estudar não decora, começando por perceber o que é necessário aprender e não se questionando sempre se está a perceber o que está a estudar.

O modelo ajustado permite-nos, ainda estimar probabilidades de acerto associadas a determinados perfis de alunos (Anexo 10). Vejamos os dois caso seguintes.

Um aluno do sexo masculino, do Porto, a frequentar o 10º ano de escolaridade, que iniciou o seu percurso escolar na pré primária, sem repetências, com acesso a livros em casa que varia de 100 a 500, que ao estudar não usa programas informáticos educativos tem por hábito quando estuda ir percebendo se está a entender o que está a ler, não decora os conteúdos e não começa a estudar a tentar perceber o que é necessário aprender, apresenta um probabilidade estimada de 74% de ter atingido o nível 5 na área espaço e forma ($IC_{95\%} =] 63\% ; 82\% [$) (ver Nota 2):

No caso de um aluno na mesma situação, mas a frequentar o 9º ano de escolaridade e com uma repetência a probabilidade de acerto cai para 44% ($IC_{95\%} =] 28\% ; 61\% [$).

Nota 2:

$$\begin{aligned}
 g(1, 1, 1, 1, 0, 1, 0, 1, 0, 0, 0, 1, 1 * 0, 0 * 1, 0 * 0) = \\
 &= -3,66 + 0,541 \times 1 + 1,113 \times 1 + 1,234 \times 1 + 0,465 \times 1 - 1,228 \times 0 \\
 &- 0,249 \times 1 + 0,437 \times 0 + 0,846 \times 1 + 1,161 \times 0 - 0,329 \times 0 + 0,696 \times 0 \\
 &+ 0,956 \times 1 - 0,7286 \times 0 - 0,823 \times 0 - 1,089 \times 0 = 1,246
 \end{aligned}$$

$$\hat{\pi} = \frac{\exp(1,246)}{1 + \exp(1,246)} \approx 0,74$$

$$IC_{95\%} = \left(\frac{\exp(\hat{g}(x) \pm 1,96\sqrt{V\hat{a}r(\hat{g}(x))})}{1 + \exp(\hat{g}(x) \pm 1,96\sqrt{V\hat{a}r(\hat{g}(x))})} \right) =]0,63 ; 0,82[$$

4.2. Perfil do aluno : área da quantidade

Para construção do modelo a variável resposta assume o valor 0 se o aluno não responde corretamente à questão e o valor 1 se responde corretamente.

No Anexo 11 apresentam-se os coeficientes das covariáveis nos modelos logísticos univariados, o seu desvio padrão, valores p associados à estatística de *Pearson*, de *Wald* e da razão de verosimilhanças, *Odds ratios* e respetivos intervalos de confiança a 95% para a questão de nível 4 na área da quantidade com o inquérito realizado aos alunos.

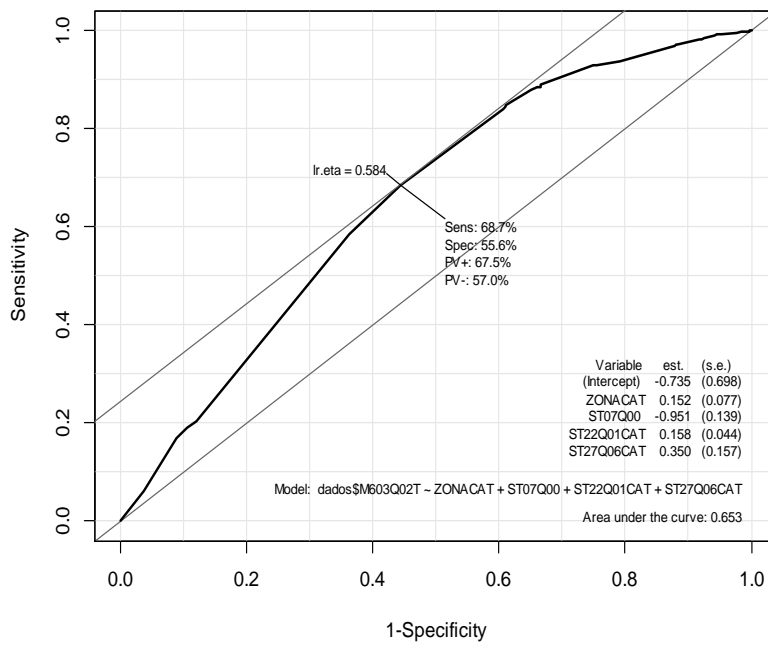
Foram encontrados numa primeira fase fatores que se revelaram protetores ou potenciadores do acerto à questão.

- ◆ As regiões **do Cento** de Portugal revelaram **as mesmas possibilidades** de responder corretamente à questão colocada quando comparadas com a **região do Alentejo**, ($IC_{95\%} =]0,9; 2,1[$, no entanto as regiões do **Norte e Grande Lisboa** apresentam cerca de **uma vez e meia mais hipótese de acerto**, $IC_{95\%} =]1,1; 2,5[$ e $IC_{95\%} =]0,9; 2,0[$ respetivamente.
- ◆ Os alunos do **10º e 11º** ano de escolaridade apresentam **quatro vezes mais possibilidades** de vir a responder corretamente à questão comparativamente com os alunos do **7º e 8º anos**, $IC_{95\%} =]2,5; 7,8[$. No entanto os alunos do **9º ano**, têm cerca de **duas vezes mais hipótese** de acerto $IC_{95\%} =]1,0; 3,3[$.
- ◆ **Os rapazes** apresentam **as mesmas possibilidades** de responder corretamente à questão do que **as raparigas**, $IC_{95\%} =]0,7; 1,1[$.
- ◆ Os alunos que frequentaram **mais do que um ano o ensino pré-escolar** apresentam cerca de **uma vez e meia mais de possibilidades** de vir a responder corretamente à questão, comparativamente com os que não frequentaram $IC_{95\%} =]1,0; 2,0[$.
- ◆ Os alunos que **ficaram retidos um ano** durante o seu percurso escolar apresentam cerca de **três vezes e meia menos possibilidades** de acerto, **comparativamente** aos alunos que **nunca ficaram retidos**, $IC_{95\%} =]2,4; 4,8[$; no entanto se as **repetências** forem **superiores a uma**, as possibilidades passam **para 5 vezes** menos, $IC_{95\%} =]2,6; 12,5[$.
- ◆ Quando as mães **estão desempregadas ou estão em casa por qualquer razão**, os alunos apresentam cerca de **75% menos de possibilidades** de responder corretamente à questão comparativamente com os alunos cujas **mães trabalham a tempo inteiro**, $IC_{95\%} =]10\%; 190\%[$; se a mãe estiver em **casa por outra razão** como por exemplo domestica, as possibilidades **caem para metade**, $IC_{95\%} =]50\%; 110\%[$.

- ◆ Os alunos que **não têm livros de apoio ao estudo em casa**, apresentam **cerca 85% menos de possibilidades** de atingir o patamar esperado no acerto à questão $IC_{95\%} =]20\%; 190\%$.
- ◆ Quando comparamos os alunos que **apenas possuem menos de 10 livros em casa** com diferentes situações verifica-se que as possibilidades de acertarem é muito diferente, constatando-se que quanto **mais livros, mais hipótese tem o aluno de responder corretamente** podendo atingir as **três vezes e meia mais para alunos** que possuam entre **201 e 500 livros em casa**, $IC_{95\%} =]2,1; 5,9\%$.
- ◆ Quando analisado o **tempo de leitura diários** dos alunos, observou-se que os alunos que leem **diariamente mais de 120 minutos**, apresentam cerca de **três vezes mais possibilidades** de responder corretamente, do que os que **nunca leem** $IC_{95\%} =]1,5; 7,1\%$, passando para cerca de **uma vez e meia mais**, no caso do aluno ler **entre meia hora e duas horas** diárias, $IC_{95\%} =]1,0; 2,0\%$.
- ◆ Para os alunos que ao estudarem **decoram a matéria** as suas possibilidades de acerto são cerca de **metade** comparativamente com os que **nunca o fazem**, $IC_{95\%} =]1,0; 2,2\%$.
- ◆ Os alunos que **refletem quase sempre acerca do que estão ao ler** apresentam **3 vezes menos** possibilidades de acerto do que os que **não o fazem** $IC_{95\%} =]1,3; 8,3\%$.
- ◆ Os alunos que **quando estudam quase nunca relacionam a informação nova com conhecimentos prévios** apresentam 2 vezes menos de possibilidades de acertos comparativamente aos que o **fazem sempre**, $IC_{95\%} =]1,3; 3,6\%$. Caso o **faça às vezes ou muitas vezes** essa possibilidade cai para metade, $IC_{95\%} =]1,1; 2,2\%$ e $IC_{95\%} =]1,0; 2,0\%$ respetivamente.
- ◆ Os alunos que quando estudam, **quase nunca tentam identificar os conceitos que ainda não percebem bem**, apresentam, **4 vezes e meia menos** possibilidades de **acerto**, comparativamente com os alunos que o **fazem sempre**, $IC_{95\%} =]2,0; 12,5\%$. Caso o **façam às vezes** essa possibilidade **passa para 70% menos**, $IC_{95\%} =]30\%; 140\%$.

Após tentativa para obter um modelo para a área da quantidade não foi possível, uma vez que o poder de discriminação obtido revelou-se muito baixo, optando-se pela não inclusão no estudo, como se poderá observar na curva de ROC que representa a melhor discriminação, Figura 4.1.

Figura 4.1. - Curva de ROC modelo área da quantidade – questionário aos alunos



4.3. Perfil do agregado familiar: área espaço e forma

Para construção do modelo a variável resposta assume o valor 0 se o aluno não responde corretamente á questão e o valor 1 se responde corretamente.

No Anexo 12, apresentam-se os coeficientes das covariáveis nos modelos logísticos univariados, o seu desvio padrão, valores p associados à estatística de *Pearson*, de *Wald* e da razão de verossimilhanças, e *Odds ratios* e respetivos intervalos de confiança a 95% para a questão de nível 5 na área espaço e forma com o inquérito realizado aos pais dos alunos.

Numa primeira fase alguns fatores revelaram ser potenciadores ou não, para o aumento das possibilidades de acerto.

As variáveis que revelaram aumentar a probabilidade de acerto foram:

- ◆ hábitos de leitura de um dos pais superior a uma hora semanal;
- ◆ conversas com os filhos acerca de questões políticas e sociais;
- ◆ jantar regularmente com os filhos à mesa;
- ◆ Rendimento anual do agregado familiar superior a 7500€;
- ◆ Gastos anuais com educação do filho superior a 4000€ comparativamente com os agregados sem gastos com a educação;
- ◆ A não solicitação do professor para idas á escola.

Sendo os fatores que revelaram diminuir as probabilidades de acerto, os seguintes:

- ◆ Ajuda regular dos pais na resolução dos trabalhos de casa;
- ◆ Habilitações académicas do Pai inferior à Licenciatura;
- ◆ Habilitações académicas da mãe inferior à Licenciatura;
- ◆ Agregados familiares com mais do que 3 filhos;

Foi construído um modelo de regressão logístico, cujos seus coeficientes, desvio padrão e valor p do teste *Wald* se encontram representados na Tabela 4.2, validação e bondade do ajustamento.

Tabela 4.2 - Coeficientes do modelo para o acerto na área espaço e forma – questionário aos pais, desvio padrão e valor p teste Wald correspondentes

Variável	Coeficientes	Desvio Padrão Coeficientes	Valor p (Wald)
Constante	-0.286	0.308	0.352
Conversas sobre questões política ou sociais(1)			
Uma ou várias vezes por semana	0.358	0.197	0.070
Ajudar os filhos a fazer os trabalhos de casa (1)			
Uma ou várias vezes por semana	-0.911	0.295	0.002
Habilitações literárias Pai (2)			
Outra	-0.875	0.229	0.001***
Rendimento anual do agregado familiar (3)			
7500 € a 11249 €	0.494	0.233	0.034*
11250 € a 18749 €	0.253	0.238	0.287
18750 € a 22499 €	1.089	0.289	0.001***
22500 € ou mais	0.969	0.235	0.001***
Número de filhos (4)			
Três ou mais	-0.471	0.203	0.020*
Conversas sobre questões política ou sociais X Ajudar os filhos a fazer os trabalhos de casa			
Uma ou várias vezes por semana X Uma ou várias vezes por semana	0.684	0.348	0.049*

(1) Categoria de referência – Nunca, uma ou duas vezes por mês

(2) Categoria de referência – Licenciatura (ou equivalente) / Mestrado / Doutoramento

(3) Categoria de referência – menos de 7500 euros

(4) Categoria de referência – Um filho

Para o modelo foram usadas as covariáveis:

- ◆ Zona geográfica (Norte, Centro, grande Lisboa e uma categoria com Alentejo, Algarve e Ilhas);
- ◆ Ano de escolaridade (alunos do 9º ano, foram agrupados os alunos do 7º e 8º anos e os alunos do 10º e 11º anos);
- ◆ Sexo;
- ◆ Conversas com os pais acerca de questões políticas ou sociais (sendo agrupados em duas categorias: nunca, uma ou duas vezes por mês e a categoria uma ou 2 vezes por semana ou todos os dias ou quase todos os dias);
- ◆ Ajudar os filhos a fazer os trabalhos de casa (sendo agrupados em duas categorias: nunca, uma ou duas vezes por mês e a categoria uma ou 2 vezes por semana ou todos os dias ou quase todos os dias)
- ◆ Habilitações literárias do pai (sendo agrupadas em duas categoria: licenciatura, mestrado ou doutoramento e as restantes);
- ◆ Rendimento do agregado familiar (quatro classes: de 7 500 € a 11 249 € , de 11 250 € a 18 749 €, de 18750 € a 22499 € e 22500 € ou mais);
- ◆ Número de filhos (duas categorias: dois ou menos e mais de dois filhos).

Elaborada uma verificação de pressupostos e uma análise dos resíduos do modelo encontrado recorrendo aos resíduos *deviance* e aos resíduos de *Pearson* assim como aos resíduos *dfbetas* e distância de Cook (Anexo 13) para avaliação de outliers e de observações influentes. Na análise efetuada foram identificadas cinco observações que se destacavam, tendo sido analisada a sua influência no modelo como se apresenta no Anexo 14. As maiores diferenças encontradas nos coeficientes apresentam valores próximos de 10%, não justificando uma análise mais profunda.

O modelo ajustado revelou um bom ajustamento aos dados pelo teste de bondade de ajustamento de Hosmer e Lemeshow ($\chi^2_8=10,56$ valor $p = 0,23$) e valor de *R de nagelkerque* igual a 0,16, registando-se um valor médio do qui-quadrado para os 10 modelos obtidos através da validação cruzada por *bootstrap*, de 11,1 com um mínimo de 2,68 e um máximo de 15,2. O modelo ajustado tem uma capacidade discriminativa aceitável (AUC=0,70; IC95% =]0,67; 0,74[) com uma sensibilidade igual a 53,3% e uma especificidade igual a 76% maximizando-as para um ponto de corte de 0,43 (Anexo 15). Para a validação do modelo, obtemos uma AUC igual a 0,69 (IC_{95%} =]0,65; 0,73[) que é reveladora da consistência interna do mesmo (Anexo 15).

$$\begin{aligned}
 &g(x) \\
 &= -0.286 + 0,358(\text{conversas sobre politica ou sociais frequentes}) \\
 &\quad - 0,911(\text{ajuda frequente dos pais nos trabalhos de casa}) \\
 &\quad - 0,875(\text{habilitações literaria nível secundário ou inferior}) \\
 &\quad + 0,4941(\text{rendimento anual agregado familiar 7500 a 11249}) \\
 &\quad + 0,2534(\text{rendimento anual agregado familiar 11250 a 18749}) \\
 &\quad + 1,089(\text{rendimento anual agregado familiar 18750 a 22499}) \\
 &\quad + 0,9689(\text{rendimento anual agregado familiar igual ou superior a 22500}) \\
 &\quad - 0,471(\text{número de filhos agregado familiar igual ou superior a 3}) \\
 &\quad + 0,684(\text{conversas sobre politica ou sociais frequentes}) \\
 &\quad \times (\text{ajuda frequente dos pais nos trabalhos de casa})
 \end{aligned}$$

A partir do modelo obtido é possível retirar as seguintes conclusões para o agregado familiar de um aluno que respondeu corretamente à questão que traduz as competências de nível 5 pelo estudo PISA na área espaço e forma.

Os fatores potenciadores do acerto foram:

- ◆ Conversas frequentes com os filhos sobre questões políticas ou sociais;
- ◆ Ajuda por parte dos pais na realização dos trabalhos de casa;
- ◆ Habilitações académicas de nível superior dos pais dos alunos;
- ◆ Rendimentos mais elevados do agregado familiar;
- ◆ Agregado familiar com um ou dois filhos.

Fixando as restantes covariáveis do modelo, as conclusões a retirar são as seguintes (Anexo16):

- ◆ Um aluno cujo **pai tenha habilitações académicas inferior à licenciatura** tem **2 vezes e meia menos** de possibilidades de acerto, comparativamente com um aluno que o seu **pai possua formação académica ao nível superior** ($IC_{95\%} =]1,5; 3,8[$);
- ◆ Um aluno cujo agregado familiar **aufere anualmente entre 7500 e 11249 euros** tem **uma vez e meia mais de possibilidades** de acerto do que um aluno cujo **agregado familiar** tenha um rendimento anual **inferior a 7500 euros** ($IC_{95\%} =]1,0; 2,6[$); Caso o rendimento anual se **situe entre 18750 e 22499 euros** essa possibilidade passa **três vezes** ($IC_{95\%} =]1,7; 5,2[$); no entanto se for **igual ou superior a 22500 euros** a possibilidade é de **duas vezes e meia mais** ($IC_{95\%} =]1,7; 4,2[$);
- ◆ Um aluno que tenha **dois ou mais irmãos** tem **menos 60%** de possibilidades de acerto do que um aluno que **seja filho único ou tenha só com um irmão** ($IC_{95\%} =]8\%; 138\%[$);
- ◆ Um aluno que **frequentemente converse com os pais acerca de questões política ou sociais**, o fato dos **pais ajudarem frequentemente a realizar os trabalhos de casa**, **umenta em cerca de três vezes** as possibilidades de acerto à questão ($IC_{95\%} =]1,6; 5,0[$) ao passo que **se não existir essa ajuda**, as possibilidades de acerto **caem para cerca de metade** ($IC_{95\%} =]1,0; 2,1[$).
- ◆ Um aluno que tenha a **ajuda dos pais na realização dos trabalhos de casa**, o fato de **manter diálogos com estes acerca de questões políticas e sociais** apresenta cerca de **1/4 menos possibilidade de acerto** à questão ($IC_{95\%} =]0,9; 1,8[$, no entanto se os pais **não manterem esses diálogos com os filhos** as possibilidades de acerto **caem duas vezes e meia** ($IC_{95\%} =]1,4; 4,4[$).

4.3.1. Consideração ao perfil dos agregados familiares – área espaço e forma

Na relação entre o aluno e a família é de esperar que sejam vários os fatores que expliquem o desenvolvimento de capacidades cognitivas. No caso em estudo salienta-se que um aluno que esteja envolvido num ambiente onde o pai investiu na sua formação académica, revela capacidades de acerto superiores, bem como a capacidade financeira do agregado familiar sendo que quanto maior o rendimento do agregado familiar, maior será as possibilidades do aluno responder corretamente à questão.

Quando analisado o número de filhos do agregado familiar, a diferença de possibilidade de acerto entre as famílias numerosas (3 ou mais filhos) e as outras é significativa, estando os alunos inseridos num contexto de família numerosa em desvantagem.

As relações e os diálogos entre pais e filhos são um aspeto a ter em conta no sucesso. Assim, os alunos oriundos de famílias onde existe diálogo acerca de assuntos da atualidade, quer políticos quer sociais, apresentam maiores possibilidades de atingir o nível 5 na área espaço e forma, porém se os pais habitualmente não ajudam os seus filhos na realização dos trabalhos de casa, torna-se uma desvantagem para o aluno comparativamente com os pais que o fazem, este fato poderá indiciar que os alunos ao atingirem os 15 anos de idade ainda necessitam do apoio em casa na realização das tarefas escolares.

A terminar refira-se que o perfil mais provável do agregado familiar do aluno que atinge um nível 5 de competência na área espaço e forma é uma família com um ou dois filhos com um rendimento anual superior a 7500€, cujo pai possui habilitações académicas de nível superior, que ajuda o aluno na realização dos trabalhos de casa e mantém diálogos acerca de assuntos políticos e sociais com os filhos.

O modelo ajustado permite-nos, ainda, estimar probabilidades de acerto associadas a determinados perfis de agregados familiares dos alunos (Anexo 16), assim:

um aluno que vive numa família com rendimentos anuais superiores a 22500 euros, tem um irmão, o pai é licenciado, conversa frequentemente com os filhos acerca de assuntos políticos e sociais, e realiza os trabalhos de casa sem a ajuda dos pais.

Este alunos apresenta um probabilidade estimada de 74% de ter atingido o nível 5 na área espaço e forma (IC_{95%} =] 65% ; 81% [).

No entanto se o aluno estiver na mesma situação mas o pai possui o 12º ano então a probabilidade cai para 54%,(IC_{95%} =]45 % ;63 % [).

Se no entanto o aluno provir de uma família numerosa com baixos recursos, cerca de 14000 euros anuais, cujo pai possui o 9ºano de escolaridade, não existam conversas sobre questões políticas ou sociais nem os pais ajudam nos trabalhos de casa, este aluno tem uma probabilidade estimada de 20% de acerto à questão, isto é, apenas 2 em cada 10 alunos consegue atingir o nível 5 na escala de espaço e forma ($IC_{95\%} =]13\% ; 30\% [$).

4.4. Perfil do agregado familiar: área da quantidade

Para construção do modelo a variável resposta assume o valor 0 se o aluno não responde corretamente à questão e o valor 1 se responde corretamente.

No Anexo 18 apresentam-se os coeficientes das covariáveis nos modelos logísticos univariados, o seu desvio padrão, valores p associados à estatística de Pearson, de Wald e da razão de verosimilhanças, *Odds ratios* e respetivos intervalos de confiança a 95% para a questão de nível 4 na área da quantidade com o inquérito realizado aos pais dos alunos.

Foram encontrados numa primeira fase fatores que se revelaram protetores ou potenciadores do acerto à questão.

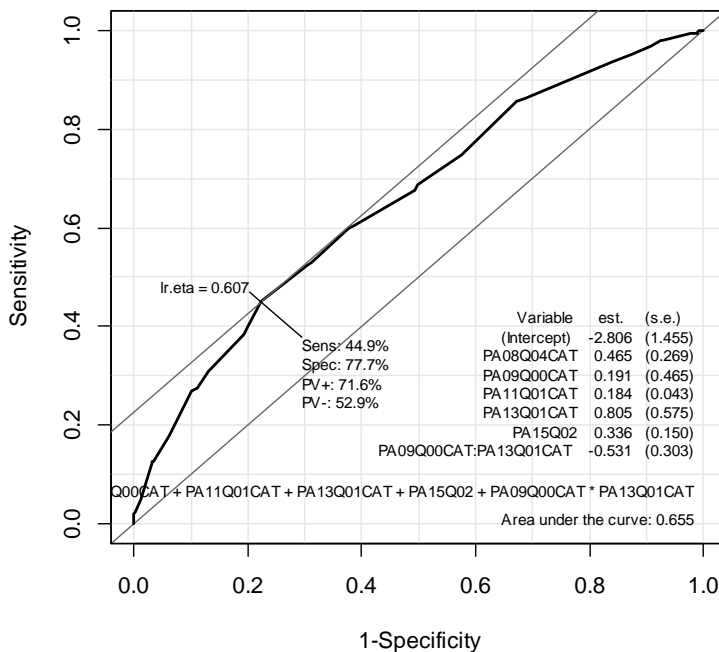
- ◆ Os alunos cujos Pais revelam **hábitos de leitura superiores a 10 horas semanais** apresentam cerca de **uma vez e meia vezes mais possibilidades** de acerto à questão quando comparados com ao alunos que os pais lêem **apenas uma hora por semana**, ($IC_{95\%} =]0,9 ; 2,5 [$), no entanto se leem **entre 6 e 10 horas semanais** essa possibilidade chega até **duas vezes mais** ($IC_{95\%} =]1,3 ; 3,2 [$).
- ◆ Os alunos de 15 anos que **conversam regularmente acerca do seu aproveitamento escolar com os pais** apresentam cerca de **uma vez e meia mais** possibilidades de acerto à questão do que os alunos que **não o fazem** ($IC_{95\%} =]0,9 ; 2,9 [$).
- ◆ Ao alunos que regularmente **jantam à mesa com os pais**, apresentam cerca de **seis vezes mais** possibilidades de acerto à questão **comparativamente** com os que **não o fazem** ($IC_{95\%} =]1,3 ; 41,6 [$).
- ◆ Os filhos que o **pai possui o nível secundário** de habilitações literárias apresenta cerca de **duas vezes e meia menos** possibilidade de acerto à questão comparativamente com um aluno que o pai possua **formação superior** ($IC_{95\%} =]1,4 ; 3,7 [$), no entanto se as habilitações literárias do pai forem inferiores ao **nível secundário** essa possibilidade passa **para três vezes e meia menos** ($IC_{95\%} =]2,2 ; 5,3 [$). Para a situação da mãe se esta possuir o **nível secundário** as possibilidades de acerto do filho **diminuem duas vezes** comparativamente com as mães que possuem o nível de **habilitações superiores** ($IC_{95\%} =]1,4 ; 3,3 [$), caso a habilitação literária da mãe seja **inferior ao nível secundário** as possibilidades de acerto do seu filho passam para **3 vezes menos** ($IC_{95\%} =]2,0 ; 4,5 [$).
- ◆ Os alunos, cujo o seu **agregado familiar** tenha um **rendimento anual de 7500 euros a 11249 euros** apresentam cerca de **duas vezes mais possibilidades** de acerto que os alunos que apresentam um **rendimento anual inferior a 7500 euros** ($IC_{95\%} =]1,2 ; 2,8 [$). Se o rendimento for de **18750 euros a 44 499 anuais** essa possibilidade passa para

duas vezes e meia mais ($IC_{95\%} =]1,5 ; 4,7 [$), chegando a **quatro vezes mais** se o rendimento for **superior a 22 500 euros anuais** ($IC_{95\%} =]2,6 ; 6,0 [$).

- ◆ Um aluno com **três ou mais irmãos**, apresenta cerca de **três vezes menos de hipóteses** de vir a responder corretamente à questão comparativamente com um aluno que **não tenha irmãos** ($IC_{95\%} =]1,3 ; 7,1 [$).
- ◆ Um aluno que o seu **encarregado de educação não seja solicitado pelo professor** para contatos acerca do comportamento e do aproveitamento apresenta **cerca de 50% mais de possibilidades de acerto** que um aluno que o seu encarregado de educação **seja convocado pelo professor** ($IC_{95\%} =]12\% ; 96\% [$).

À semelhança da tentativa de montar um modelo para a questão na área da quantidade com os questionário realizado pelos alunos, não é possível mais uma vez realizar um modelo para o questionário dos pais, uma vez que o grau de discriminação é baixo (Figura 4.1).

Figura 4.2. - Curva de ROC modelo área da quantidade – questionário aos pais



4.5. Modelação conjunta das respostas dadas nas duas questões – área espaço e forma e área da quantidade

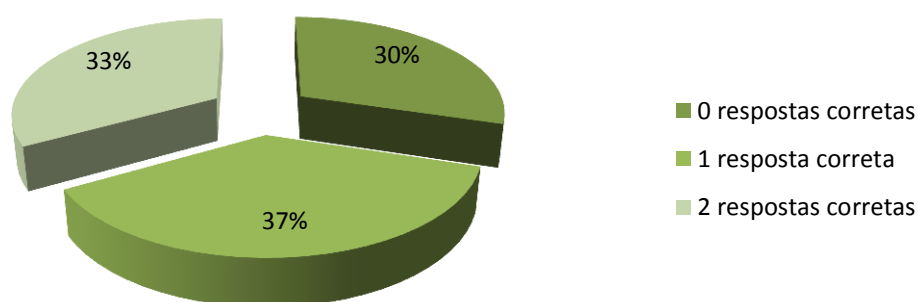
Para o perfil do aluno a amostra considerada é de 895 respostas válidas, tendo em consideração que nem todos os alunos responderam às duas questões em estudo, derivando este fato das características do teste PISA, já o perfil do agregado familiar não constará do trabalho por opção, uma vez que o poder discriminativo do melhor modelo encontrado foi reduzido.

Para construção do modelo seguindo a metodologia descrita no capítulo 2 subsecção 2.2 a variável resposta assume o valor 0 se o aluno não responde corretamente a nenhuma das questões na área espaço e forma ou na área da quantidade, valor 1 se responde a uma das questões corretamente e valor 2 se responde corretamente às duas questões colocadas

4.5.1. Análise da percentagem de acertos com questionário dos alunos

A percentagem de alunos que responderam corretamente às duas questões foi de 33%, (Figura 4.3). Verificando-se anteriormente que a percentagem de acertos à questão de espaço e forma foi de 35% seria espectável que a percentagem de acertos às duas questões fosse inferior.

Figura 4.3 - Percentagem de respostas ás questões



Na distribuição de resposta pelo País verificou-se uma ligeira associação entre os acertos e a região (valor $p=0,082$).

A proporção de acertos às duas questões está estatisticamente associada ao sexo dos alunos, apresentando os rapazes uma maior proporção de acertos (valor $p=0,011$), assim como para os

alunos que frequentaram o ensino pré escolar, tendo associado um maior sucesso nos acertos às duas questões (valor $p=0,009$).

A repetência dos alunos ao longo dos ciclos apresenta-se como uma grande desvantagem na aquisição de competências para a resolução de ambas as questões, não existindo igual distribuição entre o acerto às duas questões, significando que um aluno que não possui repetências no seu percurso escolar possui mais possibilidades de sucesso (valor $p<0,001$).

Quando analisada a proporção de acertos às duas questões tendo em conta o acesso a livros em casa, verificou-se que estas são estatisticamente diferentes (valor $p<0,001$), assim como para o tempo de leitura diária dos alunos (valor $p=0,031$).

Uma das metodologias de estudo que revelou individualmente em cada questão ser mais eficiente para os alunos foi a dos alunos que ao estudarem verificam se estão a perceber o que estão a ler, em relação à proporção de acertos às duas questões existe diferenças significativas na frequência da aplicação da metodologia (valor $p<0,001$).

4.5.2. Perfil do aluno - área espaço e forma e área da quantidade

Foi ajustado um modelo de regressão logístico ordinal, cujos seus coeficientes, desvio padrão e valor p do teste *Wald* se encontram representados na Tabela 4.3. , Anexo 19.

Tabela 4.3 - Coeficientes do modelo ordinal para o acerto às questão na área espaço e forma e na área da quantidade, desvio padrão e valor p teste *Wald* correspondentes

Variável	Coeficientes	Desvio Padrão	Coeficientes	Valor p (Wald)
$\gamma \geq 1$	-0.512	0.286		0.073
$\gamma \geq 2$	-2.322	0.296		0.001***
Zona Geográfica (1)				
Grande Lisboa/ Península de Setúbal	0.354	0.198		0.073
Centro /Norte	0.537	0.158		0.001***
Sexo do aluno (2)				
Masculino	0.553	0.138		0.001***
Frequência do ensino pré escolar (3)				
Sim	0.363	0.178		0.041*
Repetências (4)				
Sim um ano	-1.351	0.193		0.001***
Sim mais do que um ano	-2.180	0.482		0.001***
Quantidade de livros em casa (5)				
de 26 a 200 livros	0.435	0.155		0.005**
mais 201 livros	0.787	0.196		0.001***
Verificas se está a perceber o que já leu (6)				
muitas vezes	0.353	0.165		0.032*
Tempo de leitura por prazer (7)				
Menos de 120 minutos	0.212	0.154		0.169
Mais de 120 minutos	0.673	0.352		0.055

- (1) Categoria de referência - Alentejo, Algarve e Ilhas
 (2) Categoria de referência – sexo Feminino
 (3) Categoria de referência – Não frequentou o ensino pré escolar
 (4) Categoria de referência – não ter nenhuma repetências nos três primeiros ciclos
 (5) Categoria de referência – de 0 a 25 livros
 (6) Categoria de referência – quase nunca , às vezes e quase sempre
 (7) Categoria de referência- não lê por prazer

Para este modelo utilizaram-se as covariáveis:

- ◆ Zona geográfica (Norte, Centro, grande Lisboa e uma categoria com Alentejo, Algarve e Ilhas);
- ◆ Ano de escolaridade (alunos do 9º ano, foram agrupados os alunos do 7º e 8º anos e os alunos do 10º e 11º anos);
- ◆ Sexo;
- ◆ Frequência do ensino pré-escolar (foram agrupados os alunos com um ano e mais anos de frequência);
- ◆ Repetências (foram agrupados os alunos que apresentam uma ou mais repetências em no 1º, 2º e 3º ciclos);
- ◆ Uso de programas informáticos educativos em casa;

- ◆ Número de livros em casa (sendo divididos em 4 classes, de 0 a 24, de 25 a 100, de 100 a 500 e mais de 500 livros)
- ◆ Quando estuda, tenta decorar tudo o que está no texto;
- ◆ Quando estuda, começa por tentar perceber o que é mesmo preciso aprender; e
- ◆ Quando estuda, verifica se está a perceber o que já leu.

Verificada a homogeneidade dos declives de cada covariável através da análise dos gráficos de resíduos score e resíduos parcial (Anexo 20), não existindo discrepâncias nos declives apresentados à exceção de uma das categorias da variável ST23Q01CAT=1-120m referente ao tempo de leitura diário dos alunos, decidindo que apesar de sugerir a violação do pressuposto de *odds proporcionais* optou-se pela não exclusão da variável.

Foram realizadas as análises de resíduos *deviance* e aos resíduos de *Pearson* assim como aos *dfbetas* de distância de *Cook*, para os três modelos logísticos binários associados, isto é, resposta 0 e 1, resposta 1 e 2 e por ultimo resposta 0 e 2, não se identificando observações discrepantes que justifiquem uma análise mais profunda.

O modelo ajustado revelou um bom ajustamento aos dados pelo teste de *Pearson* ($\chi^2_{415}=374,7$ valor $p=0,923$) e teste *deviance* ($\chi^2_{415}=403,5$ com valor $p=0,649$).

Para verificar a qualidade do ajuste do modelo cumulativo encontrado foi calculado o coeficiente R^2 Nagelkerque = 0,19.

Foi ajustado um modelo de regressão logística multinomial e comparadas as *deviance* dos dois modelos.

$$D = -2\ln\left(\frac{\text{verossimilhança modelo nulo}}{\text{verossimilhança do modelo saturado ou completo}}\right)$$

$$D_{\text{modelo ordinal}} = -2\ln\left(\frac{1792,423}{1958,979}\right) = 0,1777$$

$$D_{\text{modelo multinomial}} = -2\ln\left(\frac{1786,747}{1962,979}\right) = 0,1881$$

Realizada a análise de variância entre o modelo multinomial e o modelo ordinal conclui-se que não diferem significativamente optando assim pelo modelo menos complexo (modelo ordinal).

O modelo ordinal de *odds* proporcionais ajustado pode ser escrito pela expressão:

$$\begin{aligned} \ln \left[\frac{P(Y \leq k)}{P(Y > k)} \right] = & \beta_k + 0,354(\text{zona da grande Lisboa}) + 0,537(\text{zona Norte/Centro}) \\ & + 0,553(\text{Masculino}) + 0,363(\text{pré escolar}) - 1,351(\text{uma repetência}) \\ & - 2,180(\text{mais de uma repetência}) + 0,435(\text{livros [26; 200]}) \\ & + 0,787(\text{ mais de 201}) + 0,353(\text{metodologia estudo F}) + 0,212(\text{leitura} \\ & < 120 \text{ min }) + 0,672(\text{leitura } > 120 \text{min}) \end{aligned}$$

Com $k = 0, 1, 2$

Realizando as estimativas das hipóteses em estudo (0, 1 ou duas respostas corretas) , foram calculados os odds ratios e respetivos intervalos de confiança das variáveis significativas do modelo encontrado, Tabela 4.4.

Tabela 4.4 - Odds ratios e intervalos de confiança a 95% do modelo ordinal para o acerto às questão espaço e forma e na área da quantidade

Variável	OR	IC 95%	
		LI	LS
Zona Geográfica (1)			
Centro /Norte	1,71	1,26	2,33
Sexo do aluno(2)			
Masculino	1,74	1,32	2,28
Frequência do ensino pré escolar (3)			
Sim	1,44	1,02	2,04
Repetências (4)			
Sim um ano	0,26	0,18	0,38
Sim mais do que um ano	0,11	0,04	0,29
Quantidade de livros em casa (5)			
de 26 a 200 livros	1,55	1,14	2,09
mais de 201 livros	2,20	1,50	3,22
Verificas se está a perceber o que já leu (6)			
muitas vezes	1,42	1,03	1,97
Tempo dedicado à leitura			
Mais de 120m	1,96	0,98	3,91

(1)referência: Alentejo, Algarve e Ilhas

(2) referência: Sexo feminino

(3) referência: não frequentou o ensino pré escolar

(4) referência: não apresenta repetências

(5) referência: possui menos de 25 livros em casa

(6) referência: não verifica regularmente o que já leu quando estuda, ou faz sempre.

Algumas conclusões poderão ser retiradas do modelo encontrado, admitindo como válido o pressuposto dos *odds* proporcionais, e admitindo como constantes as restantes covariáveis:

- ◆ As possibilidades de responder **corretamente a mais uma questão aumenta cerca de 3/4 vezes** quando estamos perante alunos da **zona Norte** comparativamente com alunos da **zona do Alentejo, Algarve e Ilhas** ($IC_{95\%} =]1,3;2,3[$);

- ◆ O fato de ser **rapaz aumenta em quase 3/4** as possibilidades de responder **corretamente a mais uma questão** ($IC_{95\%} =] 1,3 ; 2,3 [$);
- ◆ A **frequência do ensino pré-escolar aumenta** as possibilidades de responder corretamente a mais uma questão em **cerca de metade das vezes** ($IC_{95\%} =]1,0;2,0[$);
- ◆ O aluno que **já ficou retido um ano** durante o seu percurso escolar tem **cerca de quatro vezes menos** possibilidades de acerto a **mais uma questão** ($IC_{95\%} =] 2,6; 5,6 [$); no entanto verificando-se **mais do que uma repetência** as possibilidades são **9 vezes inferiores** ($IC_{95\%} =] 3,4 ; 25 [$);
- ◆ O aluno que possui entre **26 e 200 livros em casa** tem aproximadamente **uma vez e meia mais** possibilidades de responder corretamente a **mais uma questão** ($IC_{95\%} =]1,1;2,1[$); no entanto se os livros que o aluno possui em casa forem **superiores a 201** essa possibilidade **aumenta para cerca de duas vezes** ($IC_{95\%} =] 1,5; 3,2 [$).
- ◆ O aluno que ao estudar vai **verificando sempre se está a perceber o que está a ler** tem cerca de **uma vez e meia mais** possibilidades de acerto a **mais uma questão** do que o aluno que **não o faz sempre** ($IC_{95\%} =] 1,0 ; 2,0[$);
- ◆ O aluno que **lê por prazer mais de 120 minutos** por dia tem cerca de **duas vezes mais** possibilidades de acerto a **mais uma questão** do que o aluno que **não lê** ($IC_{95\%} =]1,0;3,9[$);

4.5.3. Considerações do modelo logístico ordinal - inquérito ao aluno

A importância de encontrar um perfil de aluno que consegue ir mais além do que a resposta correta a uma questão foi o objetivo do modelo encontrado.

Sai reforçada a existência de assimetrias geográficas, sendo as zonas mais populosas as que apresentam melhores resultados nos acertos a mais uma questão.

Fatores como o aluno ser do sexo masculino, não repetente, frequentar o ensino pré-escolar e ter acesso a livros em casa foram também diagnosticados como potenciadores de acertos a mais uma questão, bem como a metodologia de estudo adotada pelo aluno de quando estuda vai verificando se está a perceber o que já leu, realça-se que no perfil do aluno ao acerto a uma questão estes fatores já foram identificados.

Capítulo 5: Considerações Finais

Este estudo pretende ser uma ferramenta para professores e educadores no sentido de perceber os fatores que levam ao sucesso dos alunos portugueses nas competências nível 5 na área espaço e forma e nas competências nível 4 área da quantidade. Revela-se uma abordagem inicial com o intuito de fornecer instrumentos de medida para o auxílio a políticas educativas que vão de encontro às necessidades reveladas pelos alunos e encarregados de educação.

Algumas conclusões podem ser retiradas do estudo, destacando-se de seguida as mais relevantes.

A zona do Alentejo, Algarve e Ilhas apresentaram piores resultados, revelando que as zonas mais populosas são onde os alunos se encontram mais bem preparados quando confrontados com questões ligadas à vida real e com um nível de complexidade elevado.

Um dos fatores mais penalizadores foi as repetências dos alunos durante o percurso escolar, este dado só por si é bastante desanimador, uma vez que comparativamente com os países em estudo, Portugal apresentara percentagens de acertos muito baixas para alunos com repetências, não esquecendo que o teste PISA não testa os conteúdos programáticos mas sim a capacidade de um aluno após cerca de 9 anos em contato com o ensino escolar adquirir conhecimentos e destrezas essenciais para uma participação na sociedade, os alunos que apresentam repetências ficam desde logo com as suas capacidades comprometidas, levando a que o sistema de ensino se tenha que adequar para a resolução desta problemática.

Políticas educativas foram tomadas, O *Programa Educação 2015* que foi lançado no ano letivo 2010-2011, pretende aprofundar o envolvimento das escolas e das comunidades educativas na concretização dos compromissos nacionais e internacionais em matéria de política educativa, sendo um dos objetivos a redução do abandono escolar tendo como indicador privilegiado a redução da taxa de repetências, assim espera-se uma franca melhoria nos resultados obtidos por Portugal.

A frequência do ensino pré-escolar também se mostrou significativa para o sucesso dos alunos, a aposta dos últimos governos numa educação pré-escolar obrigatória terá resultados no desenvolvimento das crianças durante o seu percurso escolar, estes aumento de oferta de educação pré-escolar que mobilizou os Ministérios da Educação e do Trabalho e da Solidariedade Social, as autarquias locais, as autoridades gestoras dos fundos comunitários e

as comunidades educativas permitiu aumentar progressivamente as taxas de frequência deste nível educativo, resultando numa melhoria dos resultados obtidos pelos alunos.

Um outro fator que se evidenciou como potenciador para os alunos atingirem níveis elevados de proficiência nas áreas em estudo foi o acesso a livros em casa, uma vez que os alunos que os possuem têm mais possibilidades de adquirirem competências, revelando a importância dos livros no processo de desenvolvimento dos jovens e no aumento das capacidades de análise de problemáticas da vida real. Os programas da “ Rede de Bibliotecas Escolares” e “Plano Nacional de Leitura”, poderão ser uma das chaves para o acesso global ao livro, e assim melhorar as capacidades dos alunos. Também na mesma vertente quando os pais revelam possuir hábitos de leitura, os seus filhos apresentaram melhores resultados e consequentemente atingiram os níveis de competência mais elevados. O exemplo de hábitos de leitura revelados pelos pais revela-se assim importante no desenvolvimento e aumento das competências reveladas pelos filhos.

O nível de rendimento global dos agregados familiares, bem como os gastos por estes despendidos por aluno revelou ser um dos fatores que melhorara o desempenho dos alunos, verificando-se que a rendimentos mais elevados está associada uma maior proporção de acertos, num sistema de ensino gratuito e universal, espera-se que este fatores não exista.

Como seria de esperar as dinâmicas familiares são de grande importância para o sucesso dos alunos, não se poderá esperar só por si, que a escola prepare os alunos para a sociedade, revelando-se que quando nos agregados familiares habitualmente os pais jantam à mesa com os seus filhos e existem conversas acerca de questões políticas, sociais, sobre o aproveitamento escolar as competências para a vidas nas áreas em estudo são superiores. Ao acompanhamento dos estudos pelos pais, também a ajuda nos trabalhos de casa mostrou-se importante. Assim, cabe às famílias também um papel importante na construção das competências e do saber fazer dos seus filhos.

Com esta amostra de duas questões retiradas do teste PISA, revela-se que está ao alcance das escolas e das famílias dos aluno de 15 anos em Portugal melhorar significativamente as competências que estes adquiriram ao longo do seu percurso escolar, e formar o cidadão que possua uma elevada capacidade e sentido crítico para dar resposta aos desafios que irão ser colocados durante a vida futura.

Neste momento está a decorrer o estudo PISA 2012, que este ano tem como enfoque a literacia matemática, originando um leque de possibilidades alargadas para futuros estudos.

Referências Bibliográficas

- ABREU MNS, SIQUEIRA AL, CAIAFFA WY , *Regressão logística ordinal em estudeosepidemiológicos.*(pp.183-194) Revista Saúde Pública 2009 ;43 (Retirado de <http://www.scielo.br/pdf/rsp/v43n1/6604.pdf> em 17/05/201).
- AGRESTI, A. ,(2002), *An introduction to categorical data analysis. (2nd Edition)* . New York . Jonh Wiley & Sons, Inc.
- BERKSON, J. , (1944), *Application of the logistic function to bioassay.* Journal of the American Statistical Association, **39**, 357-365.
- BIRCH, M.W., (1963), *Maximum likelihood in three-way contingency tables.* Journal of the Royal Statistical Society , **B52**, 220-233.
- BLISS, C.I. (1935). *The calculation of the dosage-mortality curve.* Annals of Applied Biology , **22**, 134-167.
- DYKE & PATTERSON , (1952), *Analysis of factorial arrangements when the data are proportions.* Biometrics, **8**, 1-12.
- FARAWAY, J.(2006), *Extending the Linear Model with R – Generalized Linear, Mixed Effects and Nonparametric Regression Models.* London, New York. Chapman & Hall / CRC. Taylor & Francis Group.
- FEIGL & ZELEN, (1965), *Estimation of exponential survival probabilities with concomitant information.* Biometrics, **21**, 826-838 .
- GAVE, PISA (2004), *Literacia Matemática*, Portugal: Ministério da Educação.
- GLASSER, M., (1967), *Exponential survival with covariance.* Journal of the American Statistical Association, **62**, 561-568.
- HENRY, L. R. T. (2001), *The OECD, Globalisation and Education Policy.* Amesterdam: International Association of Universities Press, Pergamon, (pp. 188-189). Revista Lusofona, **8**, 2006 (Retirado de <http://revistas.ulusofona.pt/index.php/rleducacao/article/view/684/576> em 26/11/2011).
- HOSMER & LEMESSHOW (2000), *Applied Logistic Regression.* New York . Jonh Wiley & Sons, Inc (2nd Edition).
- LINDSEY, J.K. (1997). *Applying Generalized Linear Models.* Springer, New York.

- MCCULLAGH & NELDER (1989), *Generalized Linear Models* London, New York. (2nd Edition). Chapman and Hall.
- MCCULLOCH & SEARLE (2001), *Generalized, Linear, and Mixed Models*. New York . Jonh Wiley & Sons, Inc.
- ME (2010), *Programa Educação 2015 – elevar as competências básicas dos alunos Portugueses. Assegurar o cumprimento da escolaridade obrigatória de 12 anos. Reforçar o papel das escolas, Portugal: Ministério da Educação GAVE, PISA. Itens libertos, Portugal (Retirado de <http://www.gave.min-edu.pt/np3/134.html> em 10/12/2011).*
- NELDER, J.A. (1966). *Inverse polynomials, a useful group of multi-factor response functions*. International Biometric Society.
- OCDE(2003), *O rendimento dos alunos em matemática*. Edições Santillana.
- OCDE (2009), *PISA 2006 Technical Report*, PISA, OECD Publishing.
- OCDE (2010). *PISA 2009 Results: What Students Know and Can Do Student Performance in Reading, Mathematics and Science, Volume I*. OECD Publishing.
- OECD (2012), *PISA 2009 Technical Report*, PISA, OECD Publishing.
- PISA(2003), *Assessment Framework – Mathematics, Reading, Science and Problem Solving, Knowledge and Skill*. OECD Publishing.
- PISA(2009), *Assessment Framework - Key Competencies in Reading, Mathematics and Science* . OECD Publishing.
- 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/>.
- TURKMAN, A. & SILVA ,G. (2000) *Modelos Lineares Generalizados*. FCT – PRAXIS XXI – FEDER
- ZIPPIN & ARMITAGE, (1966), *Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter*. *Biometrics*, **22**, 665-672.
- ZOANETTI, G. A. M.(2001) , *Applications of item response theory to identify and account for suspect rater data*. University of Melbourne (Retirado de <http://www.aare.edu.au/06pap/zoa06310.pdf> em 15/12/2011).

ANEXOS

Anexo 1 - Descritores por nível de proficiência na área espaço e forma e descritores por nível de proficiência na área da quantidade

Os descritores de cada nível de proficiência na escala de matemática para a espaço e forma, tabela 1 (PISA, 2003):

Tabela 1 - Os descritores de cada nível de proficiência na escala de matemática para a área de espaço e forma

Escala	Nível	Competências gerais	Tarefas capazes de executar
Pontos	6	5% dos alunos da OCDE são capazes de executar tarefas de nível 6 na escala de espaço e forma	
Mais de 668,7		Resolver problemas complexos que envolvam representações múltiplas e, frequentemente, processos de cálculo sequencial; identificar e extrair informação relevante e ligar informação diferente, mas relacionada; usar raciocínio, perspicácia (insight) e reflexão; generalizar resultados e descobertas, comunicar soluções e fornecer explicações e argumentação..	<ul style="list-style-type: none"> – Interpretar descrições textuais complexas e relacioná-las com outras representações (muitas vezes múltiplas); – Usar o raciocínio, envolvendo proporções em situações não familiares e complexas; – Revelar perspicácia (insight) para conceptualizar situações geométricas complexas ou para interpretar representações complexas e não familiares; – Identificar e combinar múltiplas partes da informação para resolver problemas; – Criar uma estratégia para ligar um contexto geométrico com rotinas e procedimentos matemáticos conhecidos; – Executar, de um modo preciso e completo, uma sequência complexa de cálculos, por exemplo, cálculos de volume ou outros procedimentos de rotina num determinado contexto; – Fornecer explicações e argumentos escritos, com base em reflexão, perspicácia (insight) e generalização da compreensão.
Escala	Nível	Competências gerais	Tarefas capazes de executar
	5	15% dos alunos da OCDE são capazes de executar tarefas pelo menos de nível 5 na escala de espaço e forma	
Mais de 606,6		Resolver problemas que exijam a formulação de hipóteses adequadas ou que impliquem trabalhar com hipóteses fornecidas; usar raciocínio espacial, argumentação e perspicácia (insight) bem desenvolvidos, para identificar	<ul style="list-style-type: none"> – Utilizar, a nível espacial/geométrico, raciocínio, argumentação, reflexão e perspicácia (insight), aplicados a objetos bi- e tridimensionais, familiares ou não; – Formular hipóteses ou trabalhar com hipóteses para simplificar e resolver um problema geométrico, num contexto da vida real, que implique, por exemplo, a estimativa de quantidades numa situação da vida real; e comunicar as justificações; – Interpretar representações múltiplas de fenómenos geométricos;

		informação relevante e para interpretar e ligar representações diferentes; trabalhar estrategicamente e executar processos múltiplos e sequenciais.	<ul style="list-style-type: none"> – Utilizar construções geométricas; – Conceptualizar e conceber estratégias de múltiplas etapas para resolver problemas geométricos; – Usar algoritmos geométricos bem conhecidos em situações não familiares, tais como o teorema de Pitágoras; e cálculos que envolvam perímetro, área e volume.
Escala	Nível	Competências gerais	Tarefas capazes de executar
	4	32% dos alunos da OCDE são capazes de executar tarefas pelo menos de nível 4 na escala de espaço e forma	
Mais de 544,4		Resolver problemas que envolvam raciocínio visual e espacial, bem como argumentação em contextos não familiares; ligar e integrar diferentes representações; executar processos sequenciais; aplicar capacidades bem desenvolvidas em visualização e interpretação espaciais.	<ul style="list-style-type: none"> – Interpretar um texto complexo para resolver problemas geométricos; – Interpretar instruções sequenciais; seguir uma sequência de etapas; – Usar representações múltiplas para resolver um problema prático; – Interpretar, usando a perspicácia (insight) espacial em situações geométricas não padronizadas; – Usar um modelo bidimensional para trabalhar com representações tridimensionais de uma situação geométrica não familiar; – Ligar e integrar duas representações visuais diferentes de uma situação geométrica; – Desenvolver e implementar uma estratégia que envolva cálculos em situações geométricas; – Raciocinar e argumentar acerca de relações numéricas num contexto geométrico; – Executar cálculos simples (por exemplo, multiplicação de números decimais com múltiplos dígitos por um inteiro; conversões numéricas usando proporção e escala; cálculo de áreas de formas familiares).
Escala	Nível	Competências gerais	Tarefas capazes de executar
	3	54% dos alunos da OCDE são capazes de executar tarefas pelo menos de nível 3 na escala de espaço e forma	
Mais de 482,4		Resolver problemas que envolvam raciocínio elementar, visual e espacial, em contextos familiares; ligar representações diferentes de objetos familiares; usar capacidades elementares de resolução de problemas (conceção de estratégias simples); aplicar algoritmos simples.	<ul style="list-style-type: none"> – Interpretar descrições textuais de situações geométricas não familiares; – Utilizar capacidades básicas de resolução de problemas, tais como a conceção de uma estratégia simples; – Usar a percepção visual e as capacidades elementares de raciocínio espacial numa situação familiar; – Trabalhar com um modelo matemático familiar dado; – Executar cálculos simples, tais como conversões de escalas (usando multiplicações e raciocínio proporcional básico); – Aplicar algoritmos de rotina para resolver problemas geométricos (por exemplo, calcular comprimentos de formas familiares).
Escala	Nível	Competências gerais	Tarefas capazes de executar
	2	73% dos alunos da OCDE são capazes de executar tarefas pelo menos de nível 2 na escala de espaço e forma	
Mais de 420,4		Resolver problemas que envolvam uma representação matemática única, em que o conteúdo matemático é direto e apresentado com clareza; usar convenções e pensamento matemáticos básicos, em contextos familiares.	<ul style="list-style-type: none"> – Reconhecer padrões geométricos simples; – Usar definições e termos técnicos básicos e aplicar conceitos geométricos básicos (por exemplo, simetria); – Aplicar a interpretação matemática de um termo relacional da linguagem comum (por exemplo, «maior») num contexto geométrico; – Criar e usar a imagem mental de um objeto, quer bidimensional quer tridimensional; – Compreender uma representação visual bidimensional de uma situação familiar da vida real; – Aplicar cálculos simples (por exemplo, subtração, divisão por um número de dois dígitos) para resolver problemas num contexto geométrico.
Escala	Nível	Competências gerais	Tarefas capazes de executar

	1	87% dos alunos da OCDE são capazes de executar tarefas pelo menos de nível 1 na escala de espaço e forma	
Mais de 358,3		Resolver problemas simples num contexto familiar, usando figuras ou desenhos familiares de objetos geométricos e aplicar capacidades de contagem ou de cálculo básico. entre duas variáveis familiares.	– Usar uma representação bidimensional dada para contar ou calcular elementos de um objeto tridimensional simples.

Os descritores de cada nível de proficiência na escala de matemática para a área de quantidade, tabela 2 (PISA 2003):

Tabela 2 - Os descritores de cada nível de proficiência na escala de matemática para a área da quantidade

Escala	Nível	Competências gerais	Tarefas capazes de executar
Pontos	6	5% dos alunos da OCDE são capazes de executar tarefas de nível 6 na escala de quantidade	
Mais de 668,7		Conceptualizar e trabalhar com modelos de relações e processos matemáticos complexos; trabalhar com expressões formais e simbólicas; usar capacidades de raciocínio avançado para gizar estratégias de resolução de problemas e para ligar múltiplos contextos; usar processos de cálculo sequencial; formular conclusões, argumentos e explicações precisas.	<ul style="list-style-type: none"> – Conceptualizar processos matemáticos complexos, tais como o crescimento exponencial, a média ponderada, bem como propriedades do número e relações numéricas; – Interpretar e compreender informação complexa e ligar múltiplas fontes de informação complexa; – Usar o raciocínio avançado relativamente a proporções, representações geométricas de quantidades e relações numéricas combinatórias e de números inteiros; – Interpretar e compreender expressões matemáticas puras e formais, de relações entre números, inclusive em contexto científico; – Executar cálculos sequenciais num contexto complexo e não familiar, incluindo trabalhar com números grandes; – Formular conclusões, argumentos e explicações precisas; – Gizar uma estratégia (desenvolver a heurística) para trabalhar com processos matemáticos complexos.
Escala	Nível	Competências gerais	Tarefas capazes de executar
	5	15% dos alunos da OCDE são capazes de executar tarefas pelo menos de nível 5 na escala de quantidade	
Mais de 606,6		Trabalhar, eficazmente, com modelos de situações mais complexas para resolver problemas; usar capacidades de raciocínio bem desenvolvidas, de perspicácia (insight) e de interpretação, no âmbito de diferentes representações; executar processos sequenciais; comunicar o raciocínio e o argumento.	<ul style="list-style-type: none"> – Interpretar informação complexa sobre situações da vida real (incluindo gráficos, figuras e tabelas complexas); – Ligar fontes de informação diferentes (tais como gráficos, dados tabulares e texto relacionado); – Extrair dados relevantes da descrição de uma situação complexa e executar cálculos; – Usar capacidades de resolução de problemas (por exemplo, interpretar, gizar uma estratégia, usar o raciocínio; contagem sistemática) em contextos da vida real e que envolvam um nível substancial de matematização; – Comunicar o raciocínio e o argumento; – Fazer uma estimativa, recorrendo ao conhecimento da vida quotidiana; – Calcular mudança relativa e/ou absoluta.
Escala	Nível	Competências gerais	Tarefas capazes de executar
	4	32% dos alunos da OCDE são capazes de executar tarefas pelo menos de nível 4 na escala de quantidade	
Mais de 544,4		Trabalhar, eficazmente, com modelos simples de situações complexas; usar capacidades de raciocínio em vários contextos,	<ul style="list-style-type: none"> Aplicar, com precisão, um determinado algoritmo numérico, que envolva várias etapas; – Interpretar descrições textuais complexas de um processo sequencial; – Extrair informação de uma tabela e comunicar um

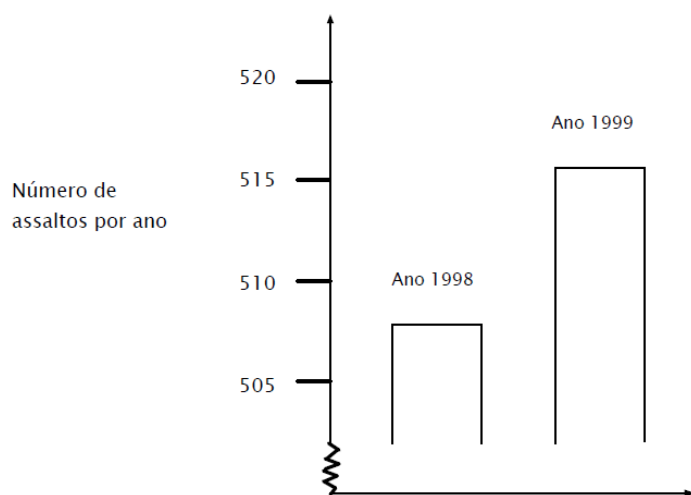
		interpretar representações diferentes da mesma situação; analisar e aplicar relações quantitativas; usar várias capacidades de cálculo para resolver problemas.	argumento simples com base nessa informação. – Relacionar informação baseada em texto a uma representação sob forma de gráfico; – Executar cálculos que impliquem raciocínio proporcional, divisibilidade ou percentagens, em modelos simples de situações complexas; – Desenvolver listagens sistemáticas e contagem de resultados combinatórios; – Identificar e utilizar informação de múltiplas fontes; – Analisar e aplicar um sistema simples; – Interpretar um texto complexo para produzir um modelo matemático simples.
Escala	Nível	Competências gerais	Tarefas capazes de executar
	3	54% dos alunos da OCDE são capazes de executar tarefas pelo menos de nível 3 na escala de quantidade	
Mais de 482,4		Interpretar informação e dados estatísticos e ligar diferentes fontes de informação; raciocínio básico relacionado com conceitos, convenções e símbolos probabilísticos simples e comunicação do raciocínio.	– Interpretar uma descrição textual de um processo de cálculo sequencial e implementar corretamente o processo; – Utilizar capacidades básicas de resolução de problemas (gizar uma estratégia simples, procurar relações, compreender e trabalhar com constrangimentos dados, utilizar o método tentativa/erro, raciocínio simples); – Executar cálculos, incluindo trabalhar com números grandes, cálculos com velocidade e tempo e conversão de unidades (por exemplo, da taxa anual para a taxa diária); – Interpretar informação tabular, localizar dados relevantes numa tabela; – Conceptualizar relações que impliquem movimento circular e tempo; – Interpretar textos e diagramas que descrevam um modelo simples.
Escala	Nível	Competências gerais	Tarefas capazes de executar
	2	73% dos alunos da OCDE são capazes de executar tarefas pelo menos de nível 2 na escala de quantidade	
Mais de 420,4		Interpretar tabelas simples para identificar e extrair informação relevante; executar cálculos aritméticos básicos; interpretar e trabalhar com relações quantitativas simples..	– Interpretar um modelo quantitativo simples (por exemplo, uma relação proporcional) e aplicá-lo, recorrendo a cálculos aritméticos básicos; – Interpretar informação tabular simples, ligar informação textual a dados tabulares relacionados; – Identificar o cálculo simples requerido para a resolução de um problema explícito; – Executar cálculos simples, que impliquem as operações aritméticas básicas, bem como ordenar números.
Escala	Nível	Competências gerais	Tarefas capazes de executar
	1	87% dos alunos da OCDE são capazes de executar tarefas pelo menos de nível 1 na escala de quantidade	
Mais de 358,3		Resolver problemas do tipo mais básico, em que toda a informação relevante é explicitamente apresentada, a situação é explícita e de âmbito muito limitado, a atividade computacional requerida é óbvia e a tarefa matemática é básica, como, por exemplo, uma simples operação aritmética..	– Interpretar uma relação matemática simples e explícita e aplicá-la diretamente, recorrendo ao cálculo.

Anexo 2 - Exemplos de questões teste PISA - área da incerteza

Assaltos

Questão 1:

Num programa de televisão, um jornalista apresentou este gráfico e disse:



«O gráfico mostra que, de 1998 para 1999, houve um aumento muito grande do número de assaltos.»

Considera que a afirmação do jornalista é uma interpretação aceitável do gráfico? Dê uma explicação que justifique a sua resposta.

Possíveis respostas:

- Respostas que indiquem “ Não, não é aceitável” e que se assentem no fato de que apenas uma pequena parte do gráfico é apresentada, ou que contenha argumentos

corretos em termos da razão ou percentagem de aumento, ou que mencionem que devia haver dados sobre os assaltos ao longo do tempo para ser possível dar uma opinião.

O Item é classificado com 694 pontos, obtendo um nível de proficiência 6 na escala matemática de Incerteza.

- Se, as respostas indiquem, “ Não, não é aceitável”, mas que contenham uma explicação pouco pormenorizada (foque apenas o aumento dado indicando o número exato de assaltos, mas não comparando com o total) ou que usem um método correto mas com erros de cálculo menores.

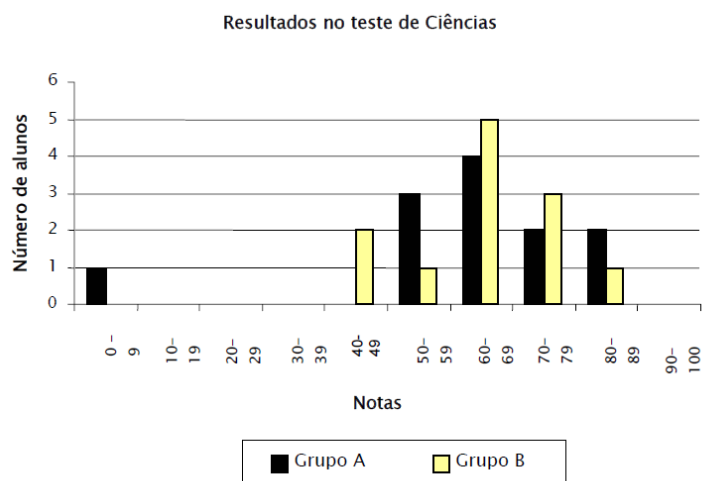
O Item é classificado com 577 pontos, obtendo um nível de proficiência 4 na escala matemática de Incerteza.

Resultado de um teste

Questão 2:

O gráfico seguinte mostra os resultados de um teste de Ciências obtidos por dois grupos de alunos, designados por «Grupo A» e «Grupo B».

A nota média no grupo A é de 62,0 e no grupo B de 64,5. Os alunos passam neste teste se tiverem uma nota igual ou superior a 50.



Com base neste gráfico, o professor concluiu que o grupo B teve melhores resultados neste teste do que o grupo A.

Os alunos do grupo A não estão de acordo com o professor. Tentam convencer o professor de que o grupo B não teve necessariamente melhores resultados.

Utilizando o gráfico, apresente um argumento matemático que possa ser utilizado pelos alunos do grupo A.

Possíveis respostas:

- Respostas que apresentam um argumento válido. Argumentos válidos podem referir-se ao número de alunos que passam, à influência desproporcionada do aluno mais fraco ao número de alunos que obtiveram os resultados mais altos.

O Item é classificado com 620 pontos, obtendo um nível de proficiência 5 na escala matemática de Incerteza.

Anexo 3 - Questionário do aluno e questionário dos pais

Anexo 4 – Passos da modelação modelo logístico

```
dados<-read.table("406alunoscopia.csv", header = T, sep = ";", dec = ".")
attach (dados)
#recodificar factores
dados$TIPO<-factor(dados$TIPO,labels=c("Privada","Pública"))
dados$CNT<-factor(dados$CNT)
dados$SCHOOLID<-factor(dados$SCHOOLID)
dados$ZONA<-factor(dados$ZONA)
dados$ST01Q01Agrup<-factor(dados$ST01Q01Agrup,labels=c("7º/8ºano","9ºano,10ºano"))
dados$ST04Q01<-factor(dados$ST04Q01,labels=c("Feminino","Masculino"))
dados$ST05Q01<-factor(dados$ST05Q01,labels=c("Não","Sim um ano","Sim mais de um ano"))
dados$ST07Q00<-factor(dados$ST07Q00,labels=c("Não","Sim ","Sim dois ou mais"))
dados$ST12Q01<-factor(dados$ST12Q01,labels=c("tempo inteiro","tempo parcial","desempregada","outra"))
dados$ST16Q01<-factor(dados$ST16Q01,labels=c("tempo inteiro","tempo parcial","desempregada","outra"))
dados$ST20Q01<-factor(dados$ST20Q01,labels=c("Sim","Não"))
dados$ST20Q02<-factor(dados$ST20Q02,labels=c("Sim","Não"))
dados$ST20Q03<-factor(dados$ST20Q03,labels=c("Sim","Não"))
dados$ST20Q04<-factor(dados$ST20Q04,labels=c("Sim","Não"))
dados$ST20Q05<-factor(dados$ST20Q05,labels=c("Sim","Não"))
dados$ST20Q06<-factor(dados$ST20Q06,labels=c("Sim","Não"))
dados$ST20Q10<-factor(dados$ST20Q10,labels=c("Sim","Não"))
dados$ST22Q01<-factor(dados$ST22Q01,labels=c("0-10","11-25","26-100","101-200","201-500","mais 500"))
dados$ST23Q01<-factor(dados$ST23Q01,labels=c("0m","1-30m","31-60m","61-120m","mais de 120m"))
dados$ST27Q01<-factor(dados$ST27Q01, labels=c("quase nunca","às vezes","muitas vezes", "quase sempre"))
dados$ST27Q02<-factor(dados$ST27Q02, labels=c("quase nunca","às vezes","muitas vezes", "quase sempre"))
dados$ST27Q04<-factor(dados$ST27Q04, labels=c("quase nunca","às vezes","muitas vezes", "quase sempre"))
dados$ST31Q00<-factor(dados$ST31Q00,labels=c("Sim","Não"))
dados$ST27Q06<-factor(dados$ST27Q06, labels=c("quase nunca","às vezes","muitas vezes", "quase sempre"))
dados$ST27Q08<-factor(dados$ST27Q08, labels=c("quase nunca","às vezes","muitas vezes", "quase sempre"))
dados$ST27Q09<-factor(dados$ST27Q09, labels=c("quase nunca","às vezes","muitas vezes", "quase sempre"))
##### ajuste a um modelo logístico
#### análise univariada - #####
fit.0<-glm( M603Q01T ~1, family=binomial("logit"), data=dados) ### modelo nulo
#Zona
fit0<-glm( M603Q01T ~ZONA, family=binomial("logit"), data=dados)
summary (fit0)
OR<-data.frame(OR=round(exp(fit0$coef), 2))
OR
exp(confint(fit0))
table(ZONA,dados$M603Q01T)
```

```

chisq.test(table(ZONA,dados$M603Q01T)) #Qui-quadrado de Pearson
anova (fit.0, fit0, test="Chisq")
#tipologia
fit<-glm(M603Q01T ~TIPO, family=binomial("logit"), data=dados)
summary (fit)
OR<-data.frame(OR=round(exp(fit$coef), 2))
OR
exp(confint(fit))
table(TIPO,dados$M603Q01T)
chisq.test(table(TIPO,dados$M603Q01T)) #Qui-quadrado de Pearson
anova (fit.0, fit, test="Chisq")
##Ano que frequenta
fit0.1<-glm(M603Q01T~ST01Q01A, family=binomial("logit"),data=dados)
summary (fit0.1)
OR<-data.frame(OR=round(exp(fit0.1$coef), 2))
OR
exp(confint(fit0.1))
table(ST01Q01A,dados$M603Q01T )
chisq.test(table(ST01Q01,dados$M603Q01T)) #Qui-quadrado de Pearson
anova (fit.0, fit0.1, test="Chisq")
# Sexo
fit1<-glm(M603Q01T~ST04Q01, family=binomial("logit"),data=dados)
summary (fit1)
OR<-data.frame(OR=round(exp(fit1$coef), 2))
OR
exp(confint(fit1))
table(ST04Q01,dados$M603Q01T )
chisq.test(table(ST04Q01,dados$M603Q01T)) #Qui-quadrado de Pearson
anova (fit.0, fit1, test="Chisq")
#pré
fit2<-glm(M603Q01T~ST05Q01, family=binomial("logit"),data=dados)
summary (fit2)
OR<-data.frame(OR=round(exp(fit2$coef), 2))
OR
exp(confint(fit2))
table(ST05Q01,dados$M603Q01T )
chisq.test(table(ST05Q01,dados$M603Q01T)) #Qui-quadrado de Pearson
anova (fit.0, fit2, test="Chisq")
# repetencias
dados$ST07Q00 <- releval(dados$ST07Q00, ref="Não")
fit3<-glm(M603Q01T~ST07Q00, family=binomial("logit"),data=dados)

```



```

summary (fit3)
OR<-data.frame(OR=round(exp(fit3$coef), 2))
OR
exp(confint(fit3))
table(ST07Q00,dados$M603Q01T )
chisq.test(table(ST07Q00,dados$M603Q01T)) #Qui-quadrado de Pearson
anova (fit.0, fit3, test="Chisq")
# tipo de actividade da mãe
fit6<-glm(M406Q01~ST12Q01, family=binomial("logit"),data=dados)
summary (fit6)
OR<-data.frame(OR=round(exp(fit6$coef), 2))
OR
exp(confint(fit6))
table(ST12Q01,dados$M406Q01 )
chisq.test(table(ST12Q01,dados$M406Q01))
anova (fit.0, fit6, test="Chisq")
# tipo de actividade do pai
fit12<-glm(M406Q01~ST16Q01, family=binomial("logit"),data=dados)
summary (fit12)
OR<-data.frame(OR=round(exp(fit12$coef), 2))
OR
exp(confint(fit12))
table(ST16Q01,dados$M406Q01 )
chisq.test(table(ST16Q01,dados$M406Q01))
anova (fit.0, fit12, test="Chisq")
# o que dispoe em casa
fit13<-glm(M406Q01~ST20Q01, family=binomial("logit"),data=dados)
summary (fit13)
OR<-data.frame(OR=round(exp(fit13$coef), 2))
OR
exp(confint(fit13))
table(ST20Q01,dados$M406Q01 )
chisq.test(table(ST20Q01,dados$M406Q01))
anova (fit.0, fit13, test="Chisq")
fit14<-glm(M406Q01~ST20Q02, family=binomial("logit"),data=dados)
summary (fit14)
OR<-data.frame(OR=round(exp(fit14$coef), 2))
OR
exp(confint(fit14))
table(ST20Q02,dados$M406Q01 )
chisq.test(table(ST20Q02,dados$M406Q01))

```

```

anova (fit.0, fit14, test="Chisq")
fit15<-glm(M406Q01~ST20Q03, family=binomial("logit"),data=dados)
summary (fit15)
OR<-data.frame(OR=round(exp(fit15$coef), 2))
OR
exp(confint(fit15))
table(ST20Q03,dados$M406Q01 )
chisq.test(table(ST20Q03,dados$M406Q01))
anova (fit.0, fit15, test="Chisq")
fit17<-glm(M406Q01~ST20Q05, family=binomial("logit"),data=dados)
summary (fit17)
OR<-data.frame(OR=round(exp(fit17$coef), 2))
OR
exp(confint(fit17))
table(ST20Q05,dados$M406Q01 )
chisq.test(table(ST20Q05,dados$M406Q01))
anova (fit.0, fit17, test="Chisq")
fit18<-glm(M406Q01~ST20Q06, family=binomial("logit"),data=dados)
summary (fit18)
OR<-data.frame(OR=round(exp(fit18$coef), 2))
OR
exp(confint(fit18))
table(ST20Q06,dados$M406Q01 )
chisq.test(table(ST20Q06,dados$M406Q01))
anova (fit.0, fit18, test="Chisq")
fit19<-glm(M406Q01~ST20Q10, family=binomial("logit"),data=dados)
summary (fit19)
OR<-data.frame(OR=round(exp(fit19$coef), 2))
OR
exp(confint(fit19))
table(ST20Q10,dados$M406Q01 )
chisq.test(table(ST20Q10,dados$M406Q01))
anova (fit.0, fit19, test="Chisq")
# livros em casa
fit20<-glm(M406Q01~ST22Q01, family=binomial("logit"),data=dados)
summary (fit20)
OR<-data.frame(OR=round(exp(fit20$coef), 2))
OR
exp(confint(fit20))
table(ST22Q01,dados$M406Q01 )
chisq.test(table(ST22Q01,dados$M406Q01))

```

```

anova (fit.0, fit20, test="Chisq")
# tempo de leitura
fit21<-glm(M406Q01~ST23Q01, family=binomial("logit"),data=dados)
summary (fit21)
OR<-data.frame(OR=round(exp(fit21$coef), 2))
OR
exp(confint(fit21))
table(ST23Q01,dados$M406Q01 )
chisq.test(table(ST23Q01,dados$M406Q01))
anova (fit.0, fit21, test="Chisq")
# tipo de estudo
fit23<-glm(M406Q01~ST27Q01, family=binomial("logit"),data=dados)
summary (fit23)
OR<-data.frame(OR=round(exp(fit23$coef), 2))
OR
exp(confint(fit23))
table(ST27Q01,dados$M406Q01 )
chisq.test(table(ST27Q01,dados$M406Q01))
anova (fit.0, fit23, test="Chisq")
dados$ST27Q02 <- relevel(dados$ST27Q02, ref="quase sempre")
fit24<-glm(M406Q01~ST27Q02, family=binomial("logit"),data=dados)
summary (fit24)
OR<-data.frame(OR=round(exp(fit24$coef), 2))
OR
exp(confint(fit24))
table(ST27Q02,dados$M406Q01 )
chisq.test(table(ST27Q02,dados$M406Q01))
anova (fit.0, fit24, test="Chisq")
dados$ST27Q04 <- relevel(dados$ST27Q04, ref="quase sempre")
fit25<-glm(M406Q01~ST27Q04, family=binomial("logit"),data=dados)
summary (fit25)
OR<-data.frame(OR=round(exp(fit25$coef), 2))
OR
exp(confint(fit25))
table(ST27Q04,dados$M406Q01 )
chisq.test(table(ST27Q04,dados$M406Q01))
anova (fit.0, fit25, test="Chisq")
dados$ST27Q06 <- relevel(dados$ST27Q06, ref="quase sempre")
fit25.6<-glm(M406Q01~ST27Q06, family=binomial("logit"),data=dados)
summary (fit25.6)
OR<-data.frame(OR=round(exp(fit25.6$coef), 2))

```

```

OR
exp(confint(fit25.6))
table(ST27Q06,dados$M406Q01 )
chisq.test(table(ST27Q06,dados$M406Q01))
anova (fit.0, fit25.6, test="Chisq")
dados$ST27Q08 <- relevel(dados$ST27Q08, ref="quase sempre")
fit25.8<-glm(M406Q01~ST27Q08, family=binomial("logit"),data=dados)
summary (fit25.8)
OR<-data.frame(OR=round(exp(fit25.8$coef), 2))
OR
exp(confint(fit25.8))
table(ST27Q08,dados$M406Q01 )
chisq.test(table(ST27Q08,dados$M406Q01))
anova (fit.0, fit25.8, test="Chisq")
dados$ST27Q09 <- relevel(dados$ST27Q09, ref="quase sempre")
fit25.9<-glm(M406Q01~ST27Q09, family=binomial("logit"),data=dados)
summary (fit25.9)
OR<-data.frame(OR=round(exp(fit25.9$coef), 2))
OR
exp(confint(fit25.9))
table(ST27Q09,dados$M406Q01 )
chisq.test(table(ST27Q09,dados$M406Q01))
anova (fit.0, fit25.9, test="Chisq")
# aulas suplementares
fit26<-glm(M406Q01~ST31Q00, family=binomial("logit"),data=dados)
summary (fit26)
OR<-data.frame(OR=round(exp(fit26$coef), 2))
OR
exp(confint(fit26))
table(ST31Q00,dados$M406Q01 )
chisq.test(table(ST31Q00,dados$M406Q01))
anova (fit.0, fit26, test="Chisq")
### modelo multivariado simples
dados$ST27Q02 <- relevel(dados$ST27Q02, ref="quase sempre")
dados$ST27Q04 <- relevel(dados$ST27Q04, ref="quase sempre")
dados$ST27Q06 <- relevel(dados$ST27Q06, ref="quase sempre")
dados$ST27Q08 <- relevel(dados$ST27Q08, ref="quase sempre")
dados$ST27Q09 <- relevel(dados$ST27Q09, ref="quase sempre")
fit29<-glm(M406Q01
~ZONA+TIPO+ST01Q01Agrup+ST04Q01+ST05Q01+ST07Q00+ST12Q01+ST16Q01+ST20Q01+ST20Q02+ST20Q05+ST20Q06
+ST20Q10+ST22Q01+ST23Q01+ST27Q01+ST27Q02+ST27Q04+ST27Q06+ST27Q08+ST27Q09+ST31Q00,

```

```

family=binomial("logit"),data=dados)
summary (fit29)
# retirar ST16Q01
fit30<-glm(M406Q01 ~ZONA+TIPO+ST01Q01Agrup+ST04Q01+ST05Q01+ST07Q00+ST12Q01+ST20Q01+ST20Q02+ST20Q05+ST20Q06
+ST20Q10+ST22Q01+ST23Q01+ST27Q01+ST27Q02+ST27Q04+ST27Q06+ST27Q08+ST27Q09+ST31Q00,
family=binomial("logit"),data=dados)
summary (fit30)
anova (fit29, fit30, test="Chisq")
# retirar ST20Q01
fit31<-glm(M406Q01
~ZONA+TIPO+ST01Q01Agrup+ST04Q01+ST05Q01+ST07Q00+ST12Q01+ST20Q02+ST20Q05+ST20Q06+ST20Q10+ST22Q01+
ST23Q01+ST27Q01+ST27Q02+ST27Q04+ST27Q06+ST27Q08+ST27Q09+ST31Q00, family=binomial("logit"),data=dados)
summary (fit31)
anova (fit30, fit31, test="Chisq")
# retirar ST20Q02
fit32<-glm(M406Q01 ~ZONA+TIPO+ST01Q01Agrup+ST04Q01+ST05Q01+ST07Q00+ST12Q01+ST20Q05+ST20Q06+ST20Q10+ST22Q01+
ST23Q01+ST27Q01+ST27Q02+ST27Q04+ST27Q06+ST27Q08+ST27Q09+ST31Q00, family=binomial("logit"),data=dados)
summary (fit32)
anova (fit31, fit32, test="Chisq")
# retirar ST20Q06
fit34<-glm(M406Q01 ~ZONA+TIPO+ST01Q01Agrup+ST04Q01+ST05Q01+ST07Q00+ST12Q01+ST20Q05+ST20Q10+
ST22Q01+ST23Q01+ST27Q01+ST27Q02+ST27Q04+ST27Q06+ST27Q08+ST27Q09+ST31Q00, family=binomial("logit"),data=dados)
summary (fit34)
anova (fit32, fit34, test="Chisq")
#### retirar 27.08
fit35<-glm(M406Q01~ZONA+TIPO+ST01Q01Agrup+ST04Q01+ST05Q01+ST07Q00+ST12Q01+ST20Q05+ST20Q10+ST22Q01+
ST23Q01+ST27Q01+ST27Q02+ST27Q04+ST27Q06+ST27Q09+ST31Q00, family=binomial("logit"),data=dados)
summary (fit35)
anova (fit34, fit35, test="Chisq")
## retirar 27.09
fit36<-glm(M406Q01~ZONA+TIPO+ST01Q01Agrup+ST04Q01+ST05Q01+ST07Q00+ST12Q01+ST20Q05+ST20Q10+ST22Q01
+ST23Q01+ST27Q01+ST27Q02+ST27Q04+ST27Q06+ST31Q00, family=binomial("logit"),data=dados)
summary (fit36)
anova (fit35, fit36, test="Chisq")
#### retirar 31.00
fit37<-glm(M406Q01~ZONA+TIPO+ST01Q01Agrup+ST04Q01+ST05Q01+ST07Q00+ST12Q01+ST20Q05+ST20Q10+ST22Q01+
ST23Q01+ST27Q01+ST27Q02+ST27Q04+ST27Q06, family=binomial("logit"),data=dados)
summary (fit37)
anova (fit36, fit37, test="Chisq")
#### retirar 20.10
fit38<-glm(M406Q01 ~ZONA+TIPO+ST01Q01Agrup+ST04Q01+ST05Q01+ST07Q00+ST12Q01+ST20Q05+ST22Q01+ST23Q01+ST27Q01+
ST27Q02+ST27Q04+ST27Q06, family=binomial("logit"),data=dados)

```

```

summary (fit38)
anova (fit37, fit38, test="Chisq")
#### retirar 12.01
fit39<-glm(M406Q01 ~ZONA+TIPO+ST01Q01Agrup+ST04Q01+ST05Q01+ST07Q00+ST20Q05+ST22Q01+ST23Q01+ST27Q01+ST27Q02+
ST27Q04+ST27Q06, family=binomial("logit"),data=dados)
summary (fit39)
anova (fit38, fit39, test="Chisq")
### retirar 23.01
fit39.1<-glm(M406Q01
~ZONA+TIPO+ST01Q01Agrup+ST04Q01+ST05Q01+ST07Q00+ST20Q05+ST22Q01+ST27Q01+ST27Q02+ST27Q04+
ST27Q06, family=binomial("logit"),data=dados)
summary (fit39.1)
anova (fit39, fit39.1, test="Chisq")
## retirar TIPO
fit39.2<-glm(M406Q01 ~ZONA+ST01Q01Agrup+ST04Q01+ST05Q01+ST07Q00+ST20Q05+ST22Q01+ST27Q01+ST27Q02+ST27Q04+
ST27Q06, family=binomial("logit"),data=dados)
summary (fit39.2)
anova (fit39.1, fit39.2, test="Chisq")
## retirar ST27Q04
fit39.3<-glm(M406Q01 ~ZONA+ST01Q01Agrup+ST04Q01+ST05Q01+ST07Q00+ST20Q05+ST22Q01+ST27Q01+ST27Q02
+ST27Q06, family=binomial("logit"),data=dados)
summary (fit39.3)
anova (fit39.2, fit39.3, test="Chisq")
### voltar a incluir a que foi retirada ST20Q03
fit40<-glm(M406Q01 ~ST20Q03+ZONA+ST01Q01Agrup+ST04Q01+ST05Q01+ST07Q00+ST20Q05+ST22Q01+ST27Q01+
ST27Q02+ST27Q06, family=binomial("logit"),data=dados)
summary (fit40)
anova (fit39.3, fit40, test="Chisq")
### retirar 20.3
fit41<-glm(M406Q01 ~ST20Q10+ZONA+ST01Q01Agrup+ST04Q01+ST05Q01+ST07Q00+ST20Q05+ST22Q01+ST27Q01+ST27Q02+
ST27Q06, family=binomial("logit"),data=dados)
summary (fit41)
anova (fit39, fit41, test="Chisq")
summary (fit39.3)
### categorizar pré
dados$ST05Q01CAT<-factor(dados$ST05Q01CAT,labels=c("Não","Sim"))
fit43<-glm(M406Q01 ~ZONA+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00+ST20Q05+ST22Q01+ST27Q01+
ST27Q02+ST27Q06, family=binomial("logit"),data=dados)
summary (fit43)
###categoriza repetencias
dados$ST07Q00CAT<-factor(dados$ST07Q00CAT,labels=c("Não repetiu","Já repetiu"))

```

```

fit44<-glm(M406Q01 ~ZONA+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01+ST27Q01+
ST27Q02+ST27Q06, family=binomial("logit"),data=dados)
summary (fit44)
### categoriza número de livros
dados$ST22Q01CAT<-factor(dados$ST22Q01CAT,labels=c("1-25","26-100","101-500"," >500"))
fit45<-glm(M406Q01 ~ZONA+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01+ST27Q02
+ST27Q06, family=binomial("logit"),data=dados)
summary (fit45)
#### categoriza 27.1
dados$ST27Q01CAT<-factor(dados$ST27Q01CAT,labels=c("quase nunca","muitas vezes"))
fit46<-glm(M406Q01 ~ZONA+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02+
ST27Q06, family=binomial("logit"),data=dados)
summary (fit46)
#### categoriza27.2
dados$ST27Q02CAT<-factor(dados$ST27Q02CAT,labels=c("outros","muitas vezes"))
fit46.1<-glm(M406Q01 ~ZONA+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+
ST27Q02+ST27Q06CAT, family=binomial("logit"),data=dados)
summary (fit46.1)
#### categoriza27.6
dados$CST27Q06<-factor(dados$CST27Q06,labels=c("quase nunca","muitas vezes"))
fit47<-glm(M406Q01
~ZONA+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+
ST27Q06CAT, family=binomial("logit"),data=dados)
summary (fit47)
## categorizar ZONA
dados$ZONACAT<-factor(dados$ZONACAT)
fit47.1<-glm(M406Q01
~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+
ST27Q06CAT, family=binomial("logit"),data=dados)
summary (fit47.1)
##### interações ZONACAT
fit48<-
glm(M406Q01~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+
ST27Q06CAT+ZONACAT*ST01Q01Agrup, family=binomial("logit"),data=dados)
summary (fit48)
fit49<-
glm(M406Q01~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+
ST27Q06CAT+ZONACAT*ST04Q01, family=binomial("logit"),data=dados)
summary (fit49)
fit50<-glm(M406Q01
~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+
ZONACAT*ST05Q01CAT, family=binomial("logit"),data=dados)

```

```

summary (fit50)
fit51<-
glm(M406Q01~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+
ST27Q06CAT+ZONACAT*ST07Q00CAT, family=binomial("logit"),data=dados)
summary (fit51)
fit52<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+
ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ZONACAT*ST20Q05, family=binomial("logit"),data=dados)
summary (fit52)
fit53<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+
ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ZONACAT*ST22Q01CAT, family=binomial("logit"),data=dados)
summary (fit53)
fit54<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+
ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ZONACAT*ST27Q01CAT, family=binomial("logit"),data=dados)
summary (fit54)
fit55<-glm(M406Q01~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+
ST27Q02CAT+ST27Q06CAT+ZONACAT*ST27Q02CAT, family=binomial("logit"),data=dados)
##### interações ano de escolaridade
fit58<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+
ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST01Q01Agrup*ST04Q01, family=binomial("logit"),data=dados)
summary (fit58)
fit59<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+
ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST01Q01Agrup*ST05Q01CAT, family=binomial("logit"),data=dados)
summary (fit59)
fit60<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+
ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST01Q01Agrup*ST07Q00CAT, family=binomial("logit"),data=dados)
summary (fit60)
fit61<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+
ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST01Q01Agrup*ST20Q05, family=binomial("logit"),data=dados)
summary (fit61)
fit62<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+
ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST01Q01Agrup*ST22Q01CAT, family=binomial("logit"),data=dados)
summary (fit62)
fit63<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+
ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST01Q01Agrup*ST27Q01CAT, family=binomial("logit"),data=dados)
summary (fit63)
fit64<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+
ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST01Q01Agrup*ST27Q02CAT, family=binomial("logit"),data=dados)
summary (fit64)
fit65<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+
ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST01Q01Agrup*ST27Q06CAT, family=binomial("logit"),data=dados)
summary (fit65)

```


interações sexo do aluno

```
fit66<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST05Q01CAT, family=binomial("logit"),data=dados)
```

summary (fit66)

```
fit67<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST07Q00CAT, family=binomial("logit"),data=dados)
```

summary (fit67)

```
fit68<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST20Q05, family=binomial("logit"),data=dados)
```

summary (fit68)

```
fit69<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST22Q01CAT, family=binomial("logit"),data=dados)
```

summary (fit69)

```
fit70<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q01CAT, family=binomial("logit"),data=dados)
```

summary (fit70)

```
fit71<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT, family=binomial("logit"),data=dados)
```

summary (fit71) **** significativa a 5%

```
fit72<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT+ST04Q01*ST27Q06CAT, family=binomial("logit"),data=dados)
```

summary (fit72)

interações com frequencia pré escolar

```
fit73<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT+ST05Q01CAT*ST07Q00CAT, family=binomial("logit"),data=dados)
```

summary (fit73)

```
fit74<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT+ST05Q01CAT*ST20Q05, family=binomial("logit"),data=dados)
```

summary (fit74)

```
fit75<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT+ST05Q01CAT*ST22Q01CAT, family=binomial("logit"),data=dados)
```

summary (fit75)

```
fit76<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT+ST05Q01CAT*ST27Q01CAT, family=binomial("logit"),data=dados)
```

summary (fit76)

```
fit77<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT+ST05Q01CAT*ST27Q02CAT, family=binomial("logit"),data=dados)
```

summary (fit77)

```
fit78<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT+ST05Q01CAT*ST27Q06CAT, family=binomial("logit"),data=dados)
```

summary (fit78)

interações repetencias

```
fit79<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT+ST07Q00CAT*ST20Q05, family=binomial("logit"),data=dados)
```

summary (fit79)

```
fit80<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT+ST07Q00CAT*ST22Q01CAT, family=binomial("logit"),data=dados)
```

summary (fit80)

```
fit81<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT+ST07Q00CAT*ST27Q01CAT, family=binomial("logit"),data=dados)
```

summary (fit81)

```
fit82<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT+ST07Q00CAT*ST27Q02CAT, family=binomial("logit"),data=dados)
```

summary (fit82)

```
fit83<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT+ST07Q00CAT*ST27Q06CAT, family=binomial("logit"),data=dados)
```

summary (fit83)

interações com nº de livros

```
fit84<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT+ST22Q01CAT*ST27Q01CAT, family=binomial("logit"),data=dados)
```

summary (fit84)

```
fit85<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT+ST22Q01CAT*ST27Q02CAT, family=binomial("logit"),data=dados)
```

summary (fit85)

```
fit86<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT+ST22Q01CAT*ST27Q06CAT, family=binomial("logit"),data=dados)
```

summary (fit86)

interações com decorar

```
fit87<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT+ST27Q01CAT*ST27Q02CAT, family=binomial("logit"),data=dados)
```

summary (fit87)

```
fit88<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+ST27Q06CAT+ST04Q01*ST27Q02CAT+ST27Q01CAT*ST27Q06CAT, family=binomial("logit"),data=dados)
```

summary (fit88)

modelo final a 10%

```
fit89<-glm(M406Q01 ~ZONACAT+ST01Q01Agrup+ST04Q01+ST05Q01CAT+ST07Q00CAT+ST20Q05+ST22Q01CAT+ST27Q01CAT+ST27Q02CAT+CST27Q06+ST04Q01*ST27Q02CAT+ST27Q02CAT*CST27Q06+ST07Q00CAT*ST27Q02CAT, family=binomial("logit"),data=dados)
```

summary (fit89)

Anexo 5 - Modelo alunos na área espaço e forma - univariada

Tabela 3 - Coeficientes das covariáveis no modelo logístico univariado, o seu desvio padrão, valores p associados à estatística de Pearson, de Wald e da razão de verossimilhanças, odds-Ratios (OR) e respetivos intervalos de confiança a 95%. Questão na área espaço e forma- inquérito ao aluno

Variável	Coeficientes	Desvio Padrão Coeficientes	Valor p (Wald)	Valor p (RR)	Valor p (Pearson)	OR	IC a 95%	
							LI	LS
Zona Geográfica (2)				0.001 ***	0.001 ***			
Algarve	-0.379	0.294	0.197			0.68	0.38	1.20
Centro	0.540	0.203	0.008 **			1.72	1.16	2.56
Grande Lisboa/ Península de Setúbal	0.311	0.216	0.150			1.36	0.90	2.09
Norte	0.391	0.193	0.043 *			1.48	1.02	2.17
Regiões Autónomas	-0.802	0.445	0.071			0.45	0.17	1.02
Tipologia da escola				0.001 ***	0.001 ***			
Pública	-0.633	0.175	0.001 ***			0.53	0.38	0.75
Ano de escolaridade do aluno (3)				0.001 ***	0.001 ***			
9º ano	1.668	0.480	0.001 ***			5.30	2.27	15.51
10ºe 11º ano	2.763	0.464	0.001 ***			15.85	7.06	45.3
Sexo do aluno				0.001 ***	0.001 ***			
Masculino	0.605	0.121	0.001 ***			1.83	1.45	2.32
Frequência do ensino pré escolar (4)				0.001 **	0.001 ***			
Sim um ano	0.252	0.214	0.24			1.29	0.85	1.96
Sim mais de um ano	0.733	0.178	0.001 ***			2.08	1.48	2.97
Repetências(5)				0.001 ***	0.001 ***			
Sim um ano	-1.938	0.222	0.001 ***			0.14	0.09	0.22
Sim dois ou mais	-2.360	0.523	0.001 **			0.09	0.03	0.23
Tipo de atividade-mãe(6)				0.002 **	0.003 **			
Tempo parcial	-0.295	0.209	0.159			0.74	0.49	1.12
Desempregada	-0.425	0.238	0.074 *			0.65	0.40	1.03
Outra	-0.591	0.174	0.001 ***			0.55	0.39	0.77
Tipo de atividade-pai** (6)				0.007 **	0.010 **			
Tempo parcial	-0.569	0.251	0.031 **			0.57	0.34	0.91
Desempregado	-0.870	0.345	0.012 **			0.42	0.20	0.86
Outra	0.092	0.272	0.735					
Secretária onde estudar				0.174	0.262			
Não	-0.574	0.441	0.192			0.55	0.22	1.25
Quarto só para o aluno				0.064	0.080			
Não	-0.306	0.167	0.068			0.74	0.53	1.02
Um sítio calmo para estudar				0.584	0.723			

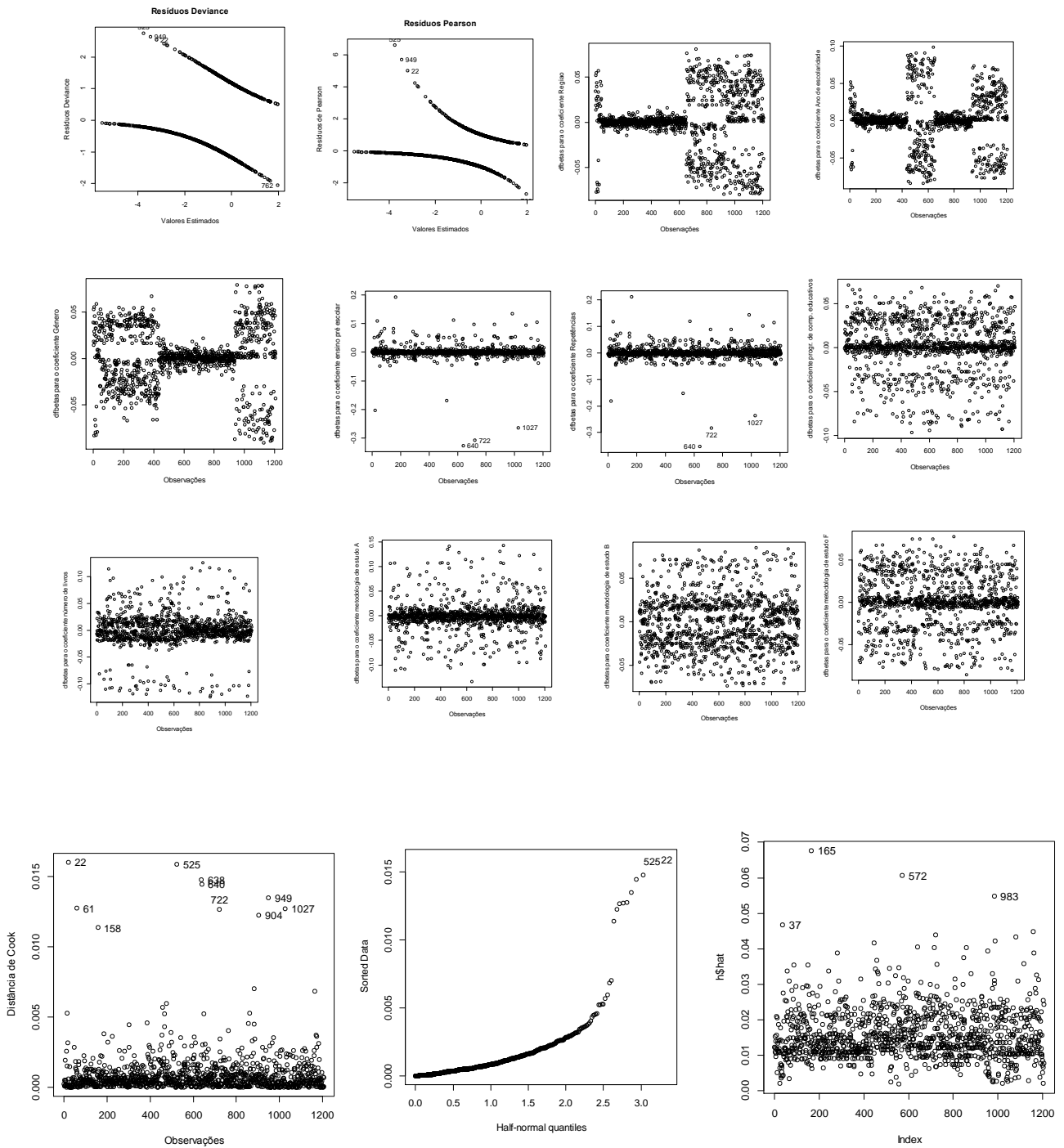
Variável	Coeficiente s	Desvio Padrão Coeficientes	Valor p (Wald)	Valor p (RR)	Valor p (Pearson)	OR	IC a 95%	
							LI	LS
em casa								
Não	-0.211	0.389	0.558					
Programas de computador educativos em casa				0.010 **	0.013**			
Não	-0.327	0.128	0.011**			0.72	0.56	0.93
Ligação à Internet em casa				0.002**	0.002**			
Não	-0.829	0.273	0.002**			0.44	0.25	0.73
Livros de apoio ao estudo em casa				0.202	0.251			
Não	-0.296	0.235	0.209			0.74	0.46	1.17
Quantidade de livros em casa (7)				0.001***	0.001***			
De 11 a 25 livros	0.323	0.261	0.230					
De 26 a 100 livros	0.895	0.237	0.001 ***			2.45	1.56	3.96
De 101 a 200 livros	1.324	0.260	0.001 ***			3.76	2.28	6.34
De 201 a 500 livros	1.594	0.272	0.001 ***			4.92	2.92	8.50
Mais de 500 livros	1.834	0.304	0.001 ***			6.26	3.49	11.5
Tempo de leitura por prazer por dia (8)				0.051*	0.049*			
Entre 11 e 30 min	0.259	0.151	0.087			1.30	0.96	1.74
Entre 31 e 60min	0.259	0.172	0.140			1.29	0.92	1.81
Entre 61 e 120min	0.344	0.203	0.089			1.41	0.95	2.10
Mais de 120min	0.896	0.332	0.007**			2.45	1.28	4.74
Estudar a decorar (9)				0.003**	0.003**			
Às vezes	-0.410	0.178	0.021 *			0.66	0.47	0.94
Muitas vezes	-0.671	0.200	0.001***			0.51	0.34	0.76
Quase sempre	-0.741	0.250	0.003 **			0.48	0.29	0.77
Tentar perceber o que é preciso aprender (10)				0.001***	0.001**			
Quase nunca	-0.520	0.472	0.270					
Às vezes	-0.726	0.183	0.001***			0.48	0.34	0.69
Muitas vezes	-0.569	0.132	0.001***			0.57	0.44	0.73
Relacionar a informação nova com conhecimentos prévios (10)				0.001***	0.001***			
Quase nunca	-0.082	0.026	0.002**			0.44	0.26	0.74
Às vezes	-0.084	0.017	0.001***			0.43	0.31	0.60

Continuação

Variável	Coeficientes	Desvio Padrão Coeficientes	Valor p (Wald)	Valor p (RR)	Valor p (Pearson)	OR	IC a 95%	
							LI	LS
Muitas vezes	-0.051	0.016	0.002**			0.60	0.44	0.82
Verificar se está a perceber o que já leu (10)				0.001***	0.001***			
Quase nunca	-1.825	0.623	0.003**			0.16	0.04	0.47
Às vezes	-0.891	0.174	0.001***			0.41	0.29	0.58
Muitas vezes	-0.370	0.136	0.006**			0.69	0.53	0.90
Perceber de que maneira a informação pode ser útil fora da escola (10)				0.356	0.356			
Quase nunca	-0.343	0.259	0.185			0.71	0.43	1.18
Às vezes	-0.231	0.202	0.255			0.79	0.54	1.18
Muitas vezes	-0.065	0.209	0.755					
Identificar os conceitos que ainda percebe*** (10)				0.001***	0.001***			
Quase nunca	-0.621	0.423	0.142			0.54	0.22	1.20
Às vezes	-0.796	0.172	0.001***			0.45	0.32	0.63
Muitas vezes	-0.389	0.150	0.009**			0.68	0.51	0.91
Aulas Suplementares a Matemática								
Não	0.213	0.124	0.078	0.127	0.142	1.24	0.98	1.57

- (1) taxas de probabilidade
- (2) categoria de referência – Alentejo
- (3) Categoria de referência – 7º e 8ºano de escolaridade
- (4) Categoria de referência – Não frequentou o ensino pré-escolar
- (5) Categoria de referência – não ter nenhuma repetência nos três primeiros ciclos
- (6) Categoria de referência – Tempo inteiro
- (7) Classe de referência – de 0 a 10 livros
- (8) Classe de referência – não lê por prazer
- (9) Classe de referência – quase nunca
- (10) Classe referência- quase sempre

Anexo 6 - Análise resíduos do modelo dos alunos na área espaço e forma



Anexo 7 - Análise das observações modelo logístico área espaço e forma - modelo do aluno

Tabela 4 - Coeficientes , deviance, coeficiente de determinação e teste Homer e Lemeshow do modelo sem as observações indicadas – questão área espaço e forma – questionário alunos

Observações		Zona3	Zona 4	Zona 5	9ºano	10º/1º ano	Sexo	Pré escolar	Repet.	Mat.di d.	Livros 26-100	Livros 101-500	Livros >500	Met. A	Met. B	Met. F	1ºint	2ºint	3ºint	Deviance	R ²	Hosmer e Lemeshow. (p)
	Coef.	0.642	0.272	0.541	1.097	1.163	1.235	0.465	-1.228	-0.249	0.437	0.846	1.161	-0.329	0.696	0.956	-0.729	-0.823	-1.089	1302	0.29	0.90
640	Coef.	0.638	0.244	0.539	1.332	1.428	1.237	0.455	-1.168	-0.259	0.447	0.859	1.178	-0.319	0.699	0.949	-0.746	-0.829	-1.091	1297	0.29	0.98
	Dif.%	0.471	10.224	0.331	-21.47	-22.83	-0.197	2.190	4.839	-4.116	-2.396	-1.548	-1.453	3.020	-0.308	0.701	-2.346	-0.782	-0.237			
525	Coef.	0.639	0.247	0.540	1.334	1.388	1.260	0.456	-1.272	-0.264	0.449	0.861	1.180	-0.342	0.687	0.940	-0.748	-0.806	-1.031	1294	0.29	0.96
	Dif.%	0.498	9.266	0.199	-21.66	-19.35	-1.992	1.931	-3.594	-6.064	-2.753	-1.772	-1.646	-3.874	1.346	1.714	-2.630	2.037	5.330			
762	Coef.	0.658	0.289	0.560	1.314	1.369	1.221	0.457	-1.260	-0.241	0.449	0.860	1.177	-0.322	0.667	0.928	-0.714	-0.796	-1.046	1297	0.29	0.91
	Dif.%	-2.548	-6.122	-3.494	-19.87	-17.74	1.118	1.832	-2.663	2.990	-2.827	-1.716	-1.403	2.117	4.214	2.938	1.952	3.308	3.945			
722	Coef.	0.622	0.273	0.541	1.307	1.366	1.223	0.494	-1.257	-0.242	0.420	0.839	1.152	-0.325	0.673	0.933	-0.716	-0.801	-1.045	1298	0.29	0.92
	Dif.%	3.081	-0.285	-0.000	-19.20	-17.53	0.984	-6.144	-2.373	2.610	3.876	0.806	0.738	1.034	3.367	2.364	1.757	2.564	4.056			
22	Coef.	0.659	0.290	0.561	1.329	1.383	1.232	0.498	-1.254	-0.239	0.415	0.837	1.149	-0.326	0.727	0.992	-0.724	-0.861	-1.052	1295	0.29	0.86
	Dif.%	-2.674	-6.546	-3.674	-21.18	-18.93	0.252	-7.057	-2.225	3.689	5.097	1.107	0.983	0.729	-4.403	-3.727	0.618	-4.695	3.339			
949	Coef.	0.644	0.276	0.543	1.053	1.114	1.234	0.459	-1.238	-0.246	0.438	0.848	1.172	-0.327	0.703	0.955	-0.725	-0.830	-1.089	1301	0.29	0.79
	Dif.%	0.315	-1.619	-0.369	3.998	4.220	0.090	1.270	-0.864	0.923	-0.334	-0.202	-0.931	0.548	-0.963	0.055	0.482	-0.924	-0.048			

Anexo 8 - Validação do modelo dos alunos na área espaço e forma

Figura 1 - Curva de ROC modelo ajustado – alunos área espaço e forma

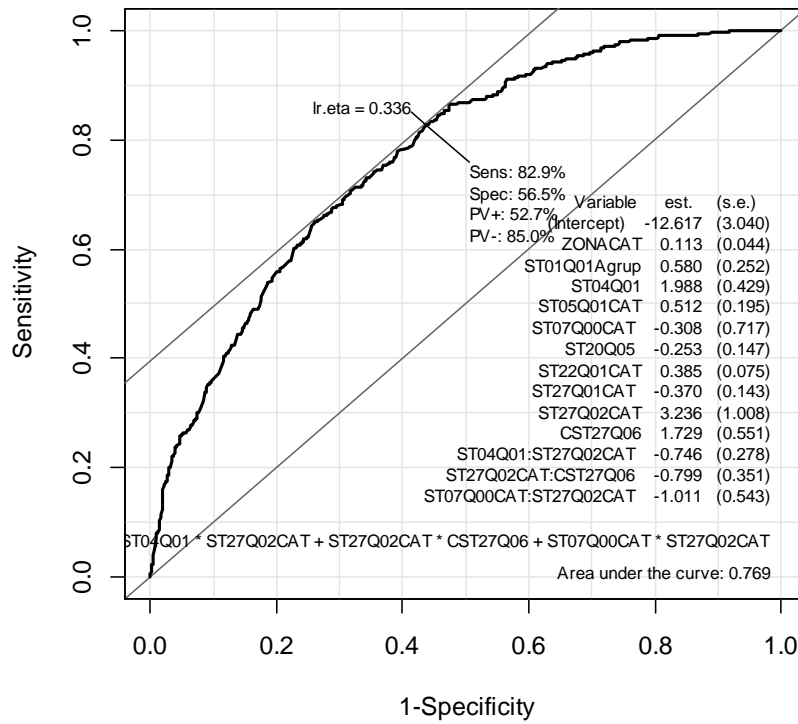
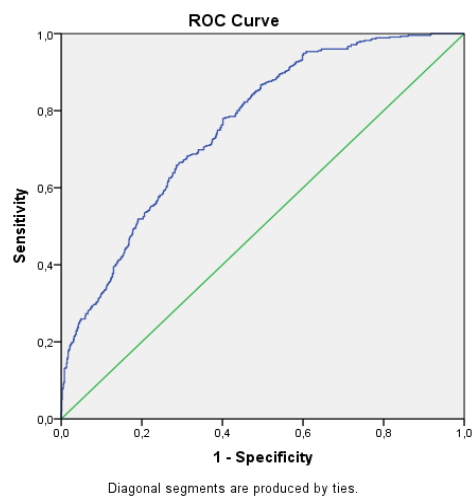


Figura 2 - Curva de ROC para validação do modelo – alunos área espaço e forma



Area Under the Curve

Test Result Variable(s): Predit

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,759	,014	,000	,733	,786

Anexo 9 - Interpretação das variáveis modelo alunos na área espaço e forma

```
#Odds Ratio da variável Zona
OR<-exp(modf$coef[2]) #Odds Ratio da variável Zona Centro
OR
ZONACAT3
1.905224
mc<-summary(modf)$cov.scaled # matriz de covariâncias do modelo
### Intervalo confiança 95%
exp(c(modf$coef[2]-qnorm (0.975) * sqrt(mc[2,2]), modf$coef[2] +qnorm (0.975) *sqrt(mc[2,2])))
ZONACAT3 ZONACAT3
1.293557 2.806121

#Odds Ratio da variável Norte
OR<-exp(modf$coef[4])
OR
ZONACAT5
1.722138
ZONACAT5 ZONACAT5
1.198324 2.474921

###Odds Ratio da variável Ano escolaridade
OR<-exp(modf$coef[5]) #Odds Ratio da variável 9ºano escolaridade
OR
ST01Q01Agrup9ºano,10ºano
3.044358
### Intervalo confiança 95%
exp(c(modf$coef[5]-qnorm (0.975) * sqrt(mc[5,5]), modf$coef[5] +qnorm (0.975) *sqrt(mc[5,5])))
ST01Q01Agrup9ºano,10ºano ST01Q01Agrup9ºano,10ºano
1.127716 8.218486

###Odds Ratio da variável frequência pré escolar
OR<-exp(modf$coef[7])
OR
ST05Q01CATSim
1.588304
### Intervalo confiança 95%
exp(c(modf$coef[7]-qnorm (0.975) * sqrt(mc[7,7]), modf$coef[7] +qnorm (0.975) *sqrt(mc[7,7])))
ST05Q01CATSim ST05Q01CATSim
1.081533 2.332531

###Odds Ratio da variável frequência número de livros

OR<-exp(modf$coef[10]) #Odds Ratio da variável número de livros entre 26 e 100
OR
ST22Q01CAT26-100
1.54858

OR<-exp(modf$coef[11]) #Odds Ratio da variável número de livros entre 101 e 500
OR
ST22Q01CAT101-500
2.335693
OR<-exp(modf$coef[12]) #Odds Ratio da variável número de livros mais de 500
OR
ST22Q01CAT >500
3.189253

### Intervalo confiança 95%
exp(c(modf$coef[10]-qnorm (0.975) * sqrt(mc[10,10]), modf$coef[10] +qnorm (0.975) *sqrt(mc[10,10])))
ST22Q01CAT26-100 ST22Q01CAT26-100
1.094437 2.191171
exp(c(modf$coef[11]-qnorm (0.975) * sqrt(mc[11,11]), modf$coef[11] +qnorm (0.975) *sqrt(mc[11,11])))
```

```

ST22Q01CAT101-500 ST22Q01CAT101-500
  1.622302    3.362791
exp(c(modf$coef[12]-qnorm (0.975) * sqrt(mc[12,12]), modf$coef[11] +qnorm (0.975) *sqrt(mc[12,12])))
ST22Q01CAT >500 ST22Q01CAT101-500
  1.850271    4.025959

### Odds Ratio da variável programa computadores
OR<-exp(modf$coef[9])
OR
ST20Q05Não
  0.7794415
### Intervalo confiança 95%
exp(c(modf$coef[9]-qnorm (0.975) * sqrt(mc[9,9]), modf$coef[9] +qnorm (0.975) *sqrt(mc[9,9])))
ST20Q05Não ST20Q05Não
  0.5837795  1.0406823

### Odds Ratio da variável metodologia de estudo A - decorar
OR<-exp(modf$coef[13])
OR
ST27Q01CATmuitas vezes
  0.7172961

### Intervalo confiança 95%
exp(c(modf$coef[13]-qnorm(0.975)*sqrt(mc[13,13]),modf$coef[13]+qnorm(0.975)*sqrt(mc[13,13])))
ST27Q01CATmuitas vezes ST27Q01CATmuitas vezes
  0.5398743    0.9530250

###Odds Ratios para Masculino versus feminino
Metodologia B ST27Q02 - preceber o q preciso aprender
# metodologia de estudo B =1 ( muitas vezes )
OR<-exp(modf$coef[6]+modf$coef[16])
OR
ST04Q01Masculino
  1.654173
### respectivos intervalos de confiança a 95%
exp(modf$coef[6]+modf$coef[16]-qnorm(0.975)*sqrt(mc[6,6]+mc[16,16]+2*mc[6,16]))
ST04Q01Masculino
  1.101219
exp(modf$coef[6]+modf$coef[16]+qnorm(0.975)*sqrt(mc[6,6]+mc[16,16]+2*mc[6,16]))
ST04Q01Masculino
  2.484781

# metodologia de estudo B =0 ( nunca , quase nunca e quase sempre )
OR<-exp(modf$coef[6])
OR
ST04Q01Masculino
  3.43447
### Intervalo de confiança a 95%
exp(modf$coef[6]-qnorm(0.975)*sqrt(mc[6,6]))
ST04Q01Masculino
  2.366585
exp(modf$coef[6]+qnorm(0.975)*sqrt(mc[6,6]))
ST04Q01Masculino
  4.98422
##Odds Ratios para comparar ST27Q06CAT - metodologia F - Quando estuda, verifica se está a perceber o que já
leu.

### ST27Q02CAT=1
OR<-exp(modf$coef[15]+modf$coef[17])
OR
CST27Q06muitas vezes
  1.145913
exp(modf$coef[15]+modf$coef[17]-qnorm(0.975)*sqrt(mc[15,15]+mc[17,17]+2*mc[15,17]))### Intervalo de

```

```

confiança a 95%
CST27Q06muitas vezes
  0.7004205
exp(modf$coef[15]+modf$coef[17]+qnorm(0.975)*sqrt(mc[15,15]+mc[17,17]+2*mc[15,17]))
CST27Q06muitas vezes
  1.874756
### ST27Q02CAT=0
OR<-exp(modf$coef[15])
OR
CST27Q06muitas vezes
  2.608204
exp(modf$coef[15]-qnorm(0.975)*sqrt(mc[15,15]))### Intervalo de confiança a 95%
CST27Q06muitas vezes
  1.608875
exp(modf$coef[15]+qnorm(0.975)*sqrt(mc[15,15]))
CST27Q06muitas vezes
  4.228252
## compara metodologia ST07Q00- repetencias
### ST27Q02CAT=1
OR<-exp(modf$coef[8]+modf$coef[16])
OR
ST07Q00CATJá repetiu
  0.133458
exp(modf$coef[8]+modf$coef[16]-qnorm(0.975)*sqrt(mc[8,8]+mc[16,16]+2*mc[8,16]))### Intervalo de confiança a
95%
ST07Q00CATJá repetiu
  0.05912199
exp(modf$coef[8]+modf$coef[16]+qnorm(0.975)*sqrt(mc[8,8]+mc[16,16]+2*mc[8,16]))
ST07Q00CATJá repetiu
  0.3012591

### ST27Q02CAT=0
OR<-exp(modf$coef[8])
OR
ST07Q00CATJá repetiu
  0.2770916
exp(modf$coef[8]-qnorm(0.975)*sqrt(mc[8,8]))### Intervalo de confiança a 95%
ST07Q00CATJá repetiu
  0.1614818
exp(modf$coef[8]+qnorm(0.975)*sqrt(mc[8,8]))
ST07Q00CATJá repetiu
  0.4754699

##Odds Ratios para comparar ST27Q02CAT - metodologia B preceber o q preciso aprender
###Sexo =1 Metodologia F = 1 Repetente =1
OR<-exp(modf$coef[14]+modf$coef[16]+modf$coef[17]+modf$coef[18])
OR
ST27Q02CATmuitas vezes
  0.1429634
### Intervalo de confiança a 95%
exp(modf$coef[14]+modf$coef[16]+modf$coef[17]+modf$coef[18]-qnorm(0.975)*sqrt(mc[14,14]+
mc[16,16]+mc[17,17]+mc[18,18]+2*mc[14,16]+2*mc[14,17]+2*mc[14,18]+2*mc[16,17]+2*mc[16,18]+2*mc[17,18]
))
ST27Q02CATmuitas vezes
  0.04878994
exp(modf$coef[14]+modf$coef[16]+modf$coef[17]+modf$coef[18]+qnorm(0.975)*sqrt(mc[14,14]+mc[16,16]+mc[1
7,17]+mc[18,18]+2*mc[14,16]+2*mc[14,17]+2*mc[14,18]+2*mc[16,17]+2*mc[16,18]+2*mc[17,18]))
ST27Q02CATmuitas vezes
  0.4189087
###Sexo =1 Metodologia F = 1 Repetente =0
OR<-exp(modf$coef[14]+modf$coef[16]+modf$coef[17])
OR
ST27Q02CATmuitas vezes

```

```

0.4242933
exp(modf$coef[14]+modf$coef[16]+modf$coef[17]-qnorm(0.975)*sqrt(mc[14,14]+mc[16,16]
+mc[17,17]+2*mc[14,16]+2*mc[14,17]+2*mc[16,17]))### Intervalo de confiança a 95%
ST27Q02CATmuitas vezes
0.2731117
exp(modf$coef[14]+modf$coef[16]+modf$coef[17]+qnorm(0.975)*sqrt(mc[14,14]+mc[16,16]+mc[17,17]+2*mc[14,
16]+2*mc[14,17]+2*mc[16,17]))
ST27Q02CATmuitas vezes
0.659162
###Sexo =1 Metodologia F = 0 Repetente =1
OR<-exp(modf$coef[14]+modf$coef[16]+modf$coef[18])
OR
ST27Q02CATmuitas vezes
0.3253978
### Intervalo de confiança a 95%
exp(modf$coef[14]+modf$coef[16]+modf$coef[18]-qnorm(0.975)*sqrt(mc[14,14]+mc[16,16]+
mc[18,18]+2*mc[14,16]+2*mc[14,18]+2*mc[16,18]))
ST27Q02CATmuitas vezes
0.10421
exp(modf$coef[14]+modf$coef[16]+modf$coef[18]+qnorm(0.975)*sqrt(mc[14,14]+
mc[16,16]+mc[18,18]+2*mc[14,16]+2*mc[14,18]+2*mc[16,18]))
ST27Q02CATmuitas vezes
1.016061
###Sexo =1 Metodologia F = 0 Repetente =0
OR<-exp(modf$coef[14]+modf$coef[16])
OR
ST27Q02CATmuitas vezes
0.9657306
### Intervalo de confiança a 95%
exp(modf$coef[14]+modf$coef[16]-qnorm(0.975)*sqrt(mc[14,14]+mc[16,16]+2*mc[14,16]))
ST27Q02CATmuitas vezes
0.4960313
exp(modf$coef[14]+modf$coef[16]+qnorm(0.975)*sqrt(mc[14,14]+mc[16,16]+2*mc[14,16]))
ST27Q02CATmuitas vezes
1.880195
###Sexo =0 Metodologia F = 1 Repetente =1
OR<-exp(modf$coef[14]+modf$coef[17]+modf$coef[18])
OR
ST27Q02CATmuitas vezes
0.2968271
### Intervalo de confiança a 95%
exp(modf$coef[14]+modf$coef[17]+modf$coef[18]-qnorm(0.975)*sqrt(mc[14,14]+mc[17,17]+mc[18,18]+2*mc[14,1
7]+2*mc[14,18]+2*mc[17,18]))
ST27Q02CATmuitas vezes
0.1003411
exp(modf$coef[14]+modf$coef[17]+modf$coef[18]+qnorm(0.975)*sqrt(mc[14,14]+mc[17,17]+mc[18,18]+2*mc[14,
17]+2*mc[14,18]+2*mc[17,18]))
ST27Q02CATmuitas vezes
0.8780678

###Sexo =0 Metodologia F = 0 Repetente =1
OR<-exp(modf$coef[14]+modf$coef[18])
OR
ST27Q02CATmuitas vezes
0.6756057
### Intervalo de confiança a 95%
exp(modf$coef[14]+modf$coef[18]-qnorm(0.975)*sqrt(mc[14,14]+mc[18,18]+2*mc[14,18]))
ST27Q02CATmuitas vezes
0.2060428
exp(modf$coef[14]+modf$coef[18]+qnorm(0.975)*sqrt(mc[14,14]+mc[18,18]+2*mc[14,18]))
ST27Q02CATmuitas vezes

```

```

2.215283
###Sexo =0 Metodologia F = 1 Repetente =0
OR<-exp(modf$coef[14]+modf$coef[15])
OR
ST27Q02CATmuitas vezes
5.229694
### Intervalo de confiança a 95%
exp(modf$coef[14]+modf$coef[15]-qnorm(0.975)*sqrt(mc[14,14]+ mc[15,15]+2*mc[14,15]))
ST27Q02CATmuitas vezes
1.781432
exp(modf$coef[14]+modf$coef[15]+qnorm(0.975)*sqrt(mc[14,14]+ mc[15,15]+2*mc[14,15]))
ST27Q02CATmuitas vezes
15.35265
###Sexo =0 Metodologia F = 0 Repetente =0
OR<-exp(modf$coef[14])
OR
ST27Q02CATmuitas vezes
2.005094
### Intervalo de confiança a 95%
> exp(modf$coef[14]-qnorm(0.975)*sqrt(mc[14,14]))
ST27Q02CATmuitas vezes
0.9929173
> exp(modf$coef[14]+qnorm(0.975)*sqrt(mc[14,14]))
ST27Q02CATmuitas vezes
4.04908

```

Anexo 10 - Exemplo perfil dos alunos na área espaço e forma

```
# 1ª situação
x0 <- c(1,0,1,0,1,1,1,0,1,0,1,0,0,0,1,1*0,0*1,0*0) ### vetor do aluno
eta0 <- sum(x0*coef(modf))
ilogit(eta0) #Probabilidade estimada
[1] 0.7366873
c<-summary(modf)$cov.scaled
se<-sqrt(t(x0) %*% c %*% x0)
se
      [,1]
0.2622585
ilogit(c(eta0-1.96*se,eta0+1.96*se)) #Intervalo de confiança para a probabilidade estimada
0.6259297 0.8238783

# 2ª situação
x1 <- c(1,0,1,0,1,1,1,1,0,1,0,0,0,0,1,1*0,0*1,0*1) ### vetor do aluno
eta1 <- sum(x1*coef(modf))
ilogit(eta1) #Probabilidade estimada
[1] 0.436695
c<-summary(modf)$cov.scaled
se<-sqrt(t(x1) %*% c %*% x1)
se
      [,1]
[1,] 0.3495547
ilogit(c(eta1-1.96*se,eta1+1.96*se)) #Intervalo de confiança para a probabilidade estimada
[1] 0.2809582 0.6060029
```

Anexo 11 - Modelo alunos na área da quantidade - univariada

Tabela 5 - Coeficientes das covariáveis no modelo logístico univariado, o seu desvio padrão, valores p associados à estatística de Pearson, de Wald e da razão de verossimilhanças, odds-Ratios (OR) e respetivos intervalos de confiança a 95%. Questão na área da quantidade - inquérito ao aluno

Variável	Coeficientes	Desvio Padrão Coeficientes	Valor p (Wald)	Valor p (RR)	Valor p (Pearson)	OR (1)	IC a 95%	
							LI	LS
Zona Geográfica (2)				0.132	0.131			
Algarve	-0.040	0.268	0.882					
Centro	0.317	0.207	0.125			0.96	0.91	2.06
Grande Lisboa/ Península de Setúbal	0.474	0.216	0.026*			1.37	1.05	2.46
Norte	0.297	0.196	0.130			1.61	0.92	1.98
Regiões Autónomas	-0.183	0.414	0.658					
Tipologia da escola				0.637	0.705			
Pública	-0.088	0.186	0.637					
Ano de escolaridade do aluno (3)				0.001***	0.001***			
9º ano	0.583	0.308	0.059			1.79	0.99	3.34
10º e 11º ano	1.455	0.291	0.001***			4.29	2.46	7.75
Sexo do aluno				0.191	0.213			
Masculino	-0.162	0.124	0.191			0.85	0.67	1.08
Frequência do ensino pré escolar (4)				0.110	0.108			
Sim um ano	0.276	0.206	0.181			1.32	0.88	1.98
Sim mais de um ano	0.345	0.164	0.035*			1.41	1.02	1.95
Repetências (5)				0.001***	0.001***			
Sim um ano	-1.223	0.167	0.001***			0.29	0.21	0.41
Sim dois ou mais	-1.667	0.391	0.001***			0.19	0.08	0.39
Tipo de atividade- mãe (6)				0.015*	0.014*			
Tempo parcial	-0.261	0.228	0.252					
Desempregada	-0.556	0.243	0.022*			0.57	0.35	0.92
Outra	-0.4223	0.169	0.0125*			0.66	0.47	0.91
Tipo de atividade- pai** (6)				0.704	0.702			
Tempo parcial	-0.237	0.247	0.338					
Desempregado	-0.230	0.309	0.457					
Outra	-0.065	0.280	0.816					
Secretária onde estudar				0.121	0.176			
Não	-0.644	0.419	0.124			0.53	0.22	1.18
Quarto só para o aluno				0.787	0.851			
Não	-0.044	0.164	0.787					
Um sítio calmo para estudar				0.082	0.118			

Variável	Coeficiente s	Desvio Padrão Coeficientes	Valor p (Wald)	Valor p (RR)	Valor p (Pearson)	OR (1)	IC a 95%	
							LI	LS
em casa								
Não	-0.659	0.382	0.085			0.52	0.24	1.09
Programas de computador educativos em casa				0.266	0.294			
Não	-0.144	0.130	0.265					
Ligação à Internet em casa				0.277	0.336			
Não	-0.279	0.256	0.279					
Livros de apoio ao estudo em casa				0.008**	0.011*			
Não	-0.619	0.236	0.009**			0.54	0.34	0.84
Quantidade de livros em casa (7)				0.001***	0.001***			
De 11 a 25 livros	0.416	0.233	0.075			1.52	0.96	2.40
De 26 a 100 livros	0.841	0.212	0.001***			2.32	1.53	3.53
De 101 a 200 livros	0.909	0.241	0.001***			2.48	1.55	4.00
De 201 a 500 livros	1.240	0.269	0.001***			3.45	2.05	5.90
Mais de 500 livros	1.121	0.297	0.001***			3.07	1.73	5.55
Tempo de leitura por prazer por dia (8)				0.024*	0.020*			
Entre 11 e 30 min	0.254	0.153	0.096			1.29	0.96	1.74
Entre 31 e 60min	0.357	0.178	0.045*			1.43	1.01	2.03
Entre 61 e 120min	0.220	0.204	0.282					
Mais de 120min	1.128	0.395	0.004**			3.09	1.48	7.09
Estudar a decorar (9)				0.241	0.240			
Às vezes	-0.134	0.184	0.467					
Muitas vezes	-0.386	0.206	0.061			0.68	0.45	1.02
Quase sempre	-0.207	0.250	0.408					
Tentar perceber o que é preciso aprender (10)				0.012*	0.012*			
Quase nunca	-1.159	0.473	0.014*			0.31	0.12	0.77
Às vezes	-0.431	0.181	0.018*			0.65	0.46	0.93
Muitas vezes	-0.191	0.137	0.164			0.83	0.63	1.08
Relacionar a informação nova com conhecimentos prévios (10)				0.012*	0.013*			
Quase nunca	-0.764	0.259	0.003**			0.47	0.28	0.77
Às vezes	-0.449	0.175	0.010*			0.64	0.45	0.90

Variável	Coeficientes	Desvio Padrão Coeficientes	Valor p (Wald)	Valor p (RR)	Valor p (Pearson)	OR (1)	IC a 95%	
							LI	LS
Muitas vezes	-0.613	0.175	0.039*			0.70	0.49	0.98
Verificar se está a perceber o que já leu (10)				0.001***	0.001***			
Quase nunca	-1.544	0.455	0.001***			0.21	0.08	0.50
Às vezes	-0.554	0.169	0.001**			0.57	0.41	0.80
Muitas vezes	-0.223	0.144	0.122			0.80	0.60	1.06
Perceber de que maneira a informação pode ser útil fora da escola (10)				0.380	0.376			
Quase nunca	-0.246	0.260	0.345					
Às vezes	0.074	0.212	0.728					
Muitas vezes	0.090	0.224	0.687					
Identificar os conceitos que ainda percebe*** (10)				0.020*	0.020*			
Quase nunca	-0.740	0.395	0.061			0.48	0.22	1.03
Às vezes	-0.478	0.174	0.006**			0.62	0.44	0.87
Muitas vezes	-0.184	0.159	0.246					
Aulas Suplementares a Matemática				0.698	0.746			
Não	0.049	0.127	0.698					

(1)taxas de probabilidade

(2)categoria de referência – Alentejo

(3)Categoria de referência – 7º e 8ºano de escolaridade

(4)Categoria de referência – Não frequentou o ensino pré escolar

(5)Categoria de referência – não ter nenhuma repetências nos três primeiros ciclos

(6) Categoria de referência – Tempo inteiro

(7) Classe de referência – de 0 a 10 livros

(8) Classe de referência – não lê por prazer

(9) Classe de referência – quase nunca

(10) Classe referência- quase sempre

Anexo 12 - Modelo país na área espaço e forma - univariada

Tabela 6 - Coeficientes das covariáveis no modelo logístico univariado, o seu desvio padrão, valores p associados à estatística de Pearson, de Wald e da razão de verossimilhanças, odds-Ratios (OR) e respetivos intervalos de confiança a 95%. Questão na área espaço e forma- inquérito aos pais.

Variável	Coeficientes	Desvio Padrão Coeficientes	Valor p (Wald)	Valor p (RR)	Valor p (Pearson)	OR (1)	IC a 95%	
							LI	LS
Tempo de leitura (2)				0.017*	0.019*			
Mais de 10 horas por semana	0.693	0.247	0.005**			2.00	1.23	3.26
Entre 6 e 10 horas por semana	0.589	0.230	0.010*			1.80	1.15	2.83
Entre 1 e 5 horas por semana	0.437	0.192	0.023*			1.55	1.07	2.27
Conversar sobre questões políticas ou sociais (3)				0.001***	0.001***			
Uma ou duas vezes por semana	0.821	0.170	0,001***			2.27	1.63	3.18
Todos os dias ou quase todos os dias	0.885	0.176	0,001***			2.42	1.72	3.43
Conversar sobre livros, filmes ou programas de televisão (3)				0.058	0.062			
Uma ou duas vezes por semana	0.363	0.202	0.072			1.44	0.97	2.15
Todos os dias ou quase todos os dias	0.467	0.199	0.191*			1.59	1.08	2.37
Conversar sobre o aproveitament o escolar do seu filho(3)								
Uma ou duas vezes por semana	-0.143	0.289	0.620					
Todos os dias ou quase todos os dias	-0.037	0.164	0.821					
Jantar à mesa com o seu filho (3)				0.038*	0.064			
Uma ou duas vezes por semana	1.674	0.817	0.040*			5.33	1.28	37
Todos os dias ou quase todos os dias	1.582	0.753	0.036*			4.86	0.37	31
Passar tempo				0.458	0.468			

Continuação

Variável	Coeficientes	Desvio Padrão Coeficientes	Valor p (Wald)	Valor p (RR)	Valor p (Pearson)	OR (1)	IC a 95%	
							LI	LS
a conversar simplesmente com o seu filho (3)								
Uma ou duas vezes por semana	0.373	0.312	0.232					
Todos os dias ou quase todos os dias	0.345	0.294	0.240					
Ir a uma livraria ou a uma biblioteca com o seu filho (3)				0.987	0.987			
Uma ou duas vezes por semana	-0.035	0.220	0.873					
Todos os dias ou quase todos os dias	0.002	0.522	0.997					
Conversar com o seu filho sobre o que ele está a ler (3)				0.691	0.690			
Uma ou duas vezes por semana	0.136	0.159	0.391			1.15	0.84	1.56
Todos os dias ou quase todos os dias	0.062	0.188	0.740			1.06	0.73	1.54
Ajudar o seu filho a fazer os trabalhos de casa (3)				0.086	0.088			
Uma ou duas vezes por semana	-0.323	0.172	0.060			0.72	0.51	1.01
Todos os dias ou quase todos os dias	-0.302	0.193	0.118			0.74	0.50	1.08
Habilitações literárias do Pai (4)				0.001***	0.001***			
Bacharelato	-0.953	0.374	0.011*			0.39	0.18	0.80
Formação profissional de nível IV	-1.249	0.291	0.001***			0.29	0.16	0.50
Ensino secundário	-1.187	0.221	0.001***			0.31	0.20	0.47
1º,2º ou 3ºciclo	-1.480	0.219	0.001***			0.23	0.15	0.35
Habilitações literárias do Mãe (4)				0.001***	0.001***			
Bacharelato	-1.165	0.379	0.002**			0.31	0.14	0.64
Formação profissional de nível IV	-0.711	0.290	0.014*			0.49	0.28	0.86

Continuação

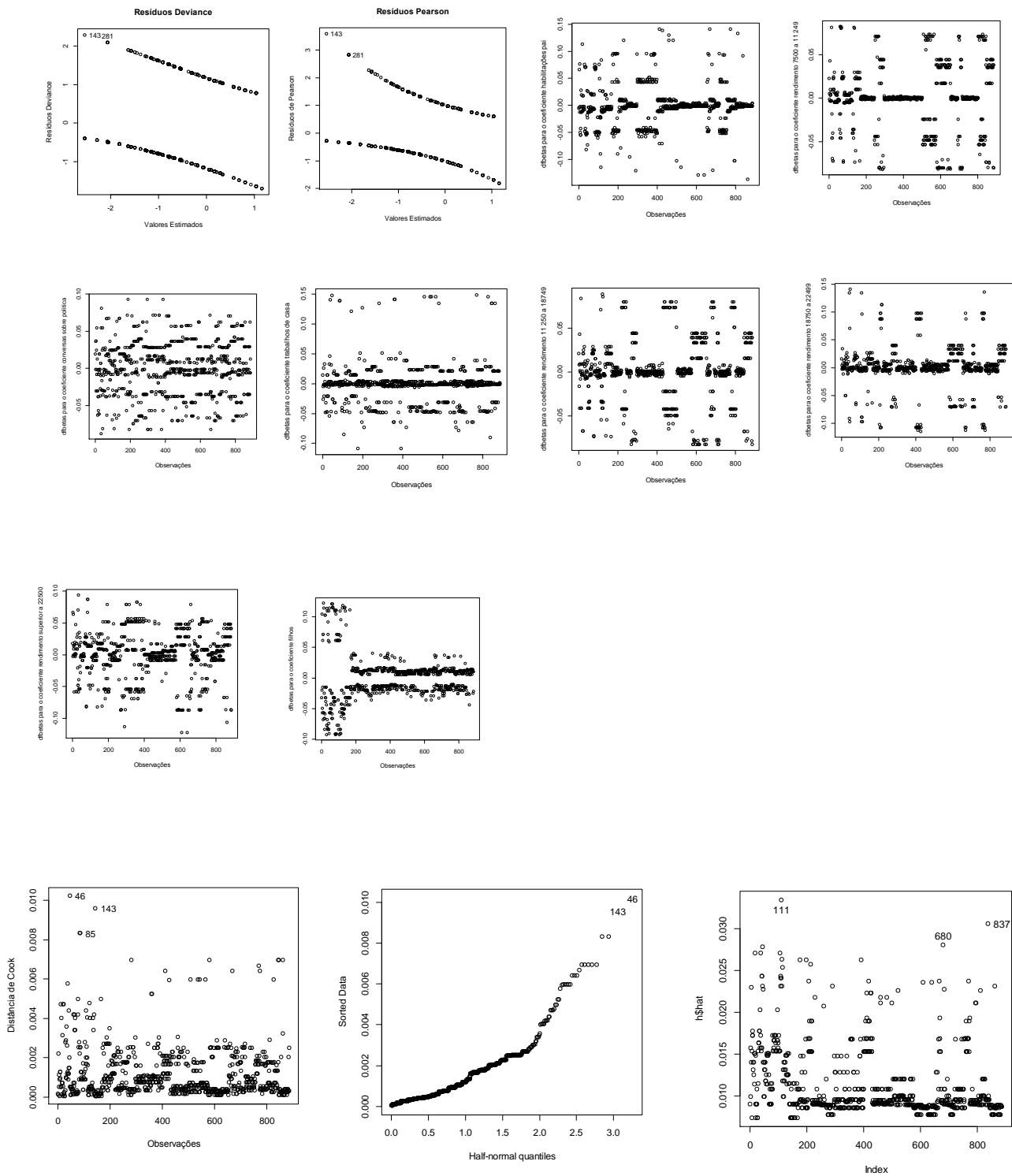
Variável	Coeficientes	Desvio Padrão Coeficientes	Valor p (Wald)	Valor p (RR)	Valor p (Pearson)	OR (1)	IC a 95%	
							LI	LS
Ensino secundário	-1.036	0.200	0.001***			0.35	0.24	0.52
1º,2º ou 3ºciclo	-1.249	0.195	0.001***			0.29	0.20	0.42
rendimento anual agregado familiar (5)				0.001***	0.001***			
de 7 500 euros a 11 249 euros	0.557	0.228	0.015*			1.75	1.12	2.74
de 11 250 euros a 14 999 euros	0.490	0.269	0.069			1.63	0.96	2.76
de 15 000 euros a 18 749 euros	0.247	0.308	0.423					
de 18 750 euros a 22 499 euros	1.361	0.280	0.001***			3.90	1.26	6.79
22 500 euros ou mais	1.470	0.209	0.001***			4.35	2.90	6.60
Gastos anuais educação do aluno (6)				0.007**	0.005**			
mais de 0 e menos de 75 euros	-0.211	0.481	0.662					
de 75 euros a 3 999 euros	0.155	0.440	0.725					
de 4 000 euros a 7 999 euros	0.981	0.510	0.055			2.67	1.00	7.55
de 8 000 euros a 11 999 euros	1.138	0.829	0.095			4.00	0.83	23
12 000 euros ou mais	0.288	1.010	0.778					
Número de filhos (7)				0.102	0.109			
Dois	-0.050	0.163	0.759					
Três	-0.480	0.229	0.036*			0.62	0.39	0.97
Quatro ou cinco	-0.514	0.439	0.242					
Conversar, por sua iniciativa, com um professor sobre o comportamento ou os progressos do seu filho				0.512	0.563			
Não	0.102	0.155	0.512					
Conversar, a pedido de um dos professores, sobre o comportamento ou os progressos do seu filho				0.009**	0.011*			

Continuação

Variável	Coeficientes	Desvio Padrão Coeficientes	Valor p (Wald)	Valor p (RR)	Valor p (Pearson)	OR (1)	IC a 95%	
							LI	LS
Não	0.366	0.140	0.008**			1.44	1.10	1.90
Participar em atividades ligadas à gestão da escola, por exemplo, associação de pais.				0.979	0.949			
Não	0.005	0.180	0.979					

- (1) Taxas de probabilidades
- (2) Categoria de referência - Menos de uma hora por semana
- (3) Categoria de referência – Nunca, uma ou duas vezes por mês
- (4) Categoria de referência – Licenciatura (ou equivalente) / Mestrado / Doutorado
- (5) Categoria de referência – menos de 7500 euros
- (6) Categoria de referência- sem gastos
- (7) Categoria de referência – Um filho

Anexo 13 - Análise resíduos do modelo dos pais na área espaço e forma



Anexo 14 - Análise das observações modelo logístico área espaço e forma - modelo dos pais

Tabela 7- Coeficientes , deviance, coeficiente de determinação e teste Homer e Lemeshow do modelo sem as observações indicadas – questão área espaço e forma – questionário pais

Observações		Conv. Políticas	Trabalhos de casa	Habilitações literárias Pai	Rendimento 7500 11249	Rendimento 11250 18749	Rendimento 18750 22499	Rendimento mais de 22500	Filhos	Interação	AIC	R ²	Hosmer e Lemeshow. (p)
todas	Coef.	0.358	-0.911	-0.875	0.494	0.253	1.089	0.969	-0.471	0.684	1078.2	0.16	0.23
46	Coef.	0.360	-0.967	-0.882	0.492	0.249	1.038	0.963	-0.500	0.738	1074.8	0.17	0.40
	Dif.%	-0.67	-6.07	-0.85	0.40	1.57	4.65	0.59	-6.17	-7.85			
143	Coef.	0.354	-0.974	-0.880	0.518	0.276	1.110	0.991	-0.502	0.746	1072.9	0.17	0.26
	Dif.%	0.89	-6.86	-0.57	-4.88	-8.84	-1.95	-2.33	-6.44	-9.01			
111	Coef.	0.369	-0.903	-0.857	0.493	0.253	1.064	0.973	-0.484	0.675	1077.1	0.16	0.32
	Dif.%	-3.09	0.97	2.11	0.15	0.22	2.28	-0.39	-2.66	1.38			
281	Coef.	0.355	-0.971	-0.874	0.520	0.277	1.111	0.992	0.463	0.744	1073.8	0.17	0.21
	Dif.%	0.58	-6.57	0.07	-5.14	-9.47	-2.08	-2.43	1.70	-8.77			
837	Coef.	0.358	-0.889	-0.893	0.503	0.253	1.088	0.963	-0.476	0.661	1077.4	0.16	0.23
	Dif.%	-0.20	2.49	-2.07	-1.79	0.15	0.02	0.59	-0.89	3.35			

Anexo 15 - Validação do modelo dos pais na área espaço e forma

Figura 3 - Curva de ROC modelo ajustado – pais área espaço e forma

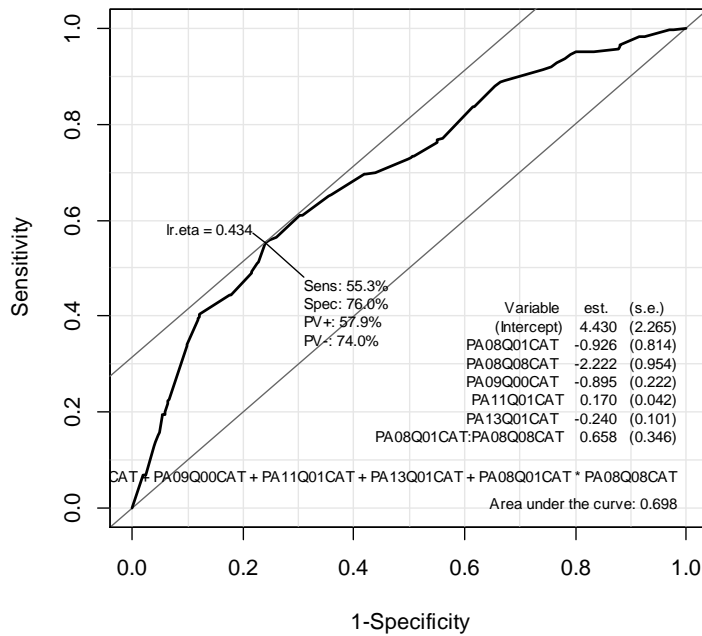
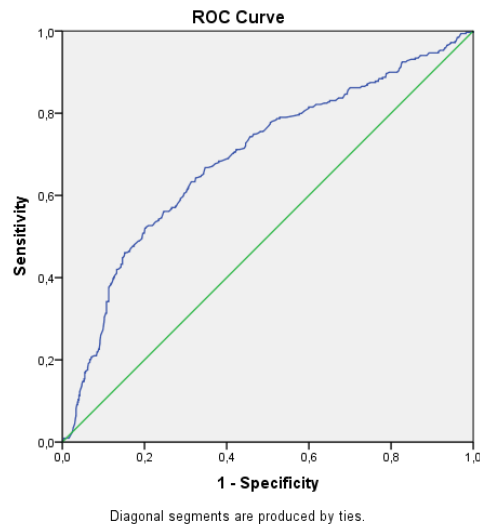


Figura 4 - Curva de ROC para validação do modelo – pais área espaço e forma



Area Under the Curve

Test Result Variable(s): predict

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,690	,019	,000	,653	,728

Anexo 16 - Interpretação das variáveis modelo pais na área espaço e forma

```
#Odds Ratio da variável Habilitações literárias do Pai

OR<-exp(modf$coef[4]) #Odds Ratio da variável habilitações literárias pai
OR
PA09Q00CATOutra
  0.4168513
mc<-summary(modf)$cov.scaled # matriz de covariâncias do modelo
### Intervalo confiança 95%
exp(c(modf$coef[4]-qnorm (0.975) * sqrt(mc[4,4]), modf$coef[4] +qnorm (0.975) *sqrt(mc[4,4])))
PA09Q00CATOutra PA09Q00CATOutra
0.2659793    0.6533030

#Odds Ratio da variável rendimento

OR<-exp(modf$coef[5]) # Odds Ratio da variável rendimento de 7500 a 11249
OR
PA11Q01CATde 7 500 euros a 11 249 euros
  1.639048
### Intervalo confiança 95%
exp(c(modf$coef[5]-qnorm (0.975) * sqrt(mc[5,5]), modf$coef[5] +qnorm (0.975) *sqrt(mc[5,5]))) PA11Q01CATde
7 500 euros a 11 249 euros PA11Q01CATde 7 500 euros a 11 249 euros
  1.037999          2.588132

OR<-exp(modf$coef[7]) # Odds Ratio da variável rendimento de 18750 a 22499
OR
PA11Q01CATse 18750 euraos a 22499
  2.970089
### Intervalo confiança 95%
exp(c(modf$coef[7]-qnorm (0.975) * sqrt(mc[7,7]), modf$coef[7] +qnorm (0.975) *sqrt(mc[7,7])))
PA11Q01CATse 18750 euraos a 22499 PA11Q01CATse 18750 euraos a 22499
  1.685895          5.232490

OR<-exp(modf$coef[8]) # Odds Ratio da variável rendimento de mais de 22500
OR
PA11Q01CAT22 500 euros ou mais
  2.634973
### Intervalo confiança 95%
exp(c(modf$coef[8]-qnorm (0.975) * sqrt(mc[8,8]), modf$coef[8] +qnorm (0.975) *sqrt(mc[8,8])))
PA11Q01CAT22 500 euros ou mais PA11Q01CAT22 500 euros ou mais
  1.663713          4.173244

#Odds Ratio da variável número de filhos

OR<-exp(modf$coef[9]) # Odds Ratio da variável número de filhos
OR
PA13Q01CATtrês ou mais
  0.6241705
### Intervalo confiança 95%
exp(c(modf$coef[9]-qnorm (0.975) * sqrt(mc[9,9]), modf$coef[9] +qnorm (0.975) *sqrt(mc[9,9])))
PA13Q01CATtrês ou mais PA13Q01CATtrês ou mais

#Odds Ratios para Conversas frequente sobre politica/sociais versus quase inexistentes sobre politica/sociais

# Ajuda nos TPC =1 ( uma ou várias vezes por semana )
OR<-exp(modf$coef[2]+modf$coef[10])
OR
PA08Q01CATuma ou várias vezes por semana
  2.834372
```

```

# Intervalo de confiança a 95%
exp(modf$coef[2]+modf$coef[10]-qnorm(0.975)*sqrt(mc[2,2]+mc[10,10]+2*mc[2,10]))
PA08Q01CATuma ou várias vezes por semana
1.600016
exp(modf$coef[2]+modf$coef[10]+qnorm(0.975)*sqrt(mc[2,2]+mc[10,10]+2*mc[2,10]))
PA08Q01CATuma ou várias vezes por semana
5.020991

# Ajuda nos TPC =0 ( quase nunca )

OR<-exp(modf$coef[2])
OR
PA08Q01CATuma ou várias vezes por semana
1.42974

# Intervalo de confiança a 95%
exp(modf$coef[2]-qnorm(0.975)*sqrt(mc[2,2]))
PA08Q01CATuma ou várias vezes por semana
0.9716211
exp(modf$coef[2]+qnorm(0.975)*sqrt(mc[2,2]))
PA08Q01CATuma ou várias vezes por semana
2.103863

#Odds Ratios ajuda dos pais nos TPC frequentemente versus raramente

Conversas sobre politica/sociais =1 ( uma ou várias vezes por semana )

OR<-exp(modf$coef[3]+modf$coef[10])
OR
PA08Q08CATuma ou várias vezes por semana
0.7969301

# Intervalo de confiança a 95%
exp(modf$coef[3]+modf$coef[10]-qnorm(0.975)*sqrt(mc[3,3]+mc[10,10]+2*mc[3,10]))
PA08Q08CATuma ou várias vezes por semana
0.5559179
exp(modf$coef[3]+modf$coef[10]+qnorm(0.975)*sqrt(mc[3,3]+mc[10,10]+2*mc[3,10]))
PA08Q08CATuma ou várias vezes por semana
1.142431

# Conversas sobre politica/sociais =0 ( quase nunca )

OR<-exp(modf$coef[3])
OR
PA08Q08CATuma ou várias vezes por semana
0.4019949

# Intervalo de confiança a 95%
exp(modf$coef[3]-qnorm(0.975)*sqrt(mc[3,3]))
PA08Q08CATuma ou várias vezes por semana
0.2254371
exp(modf$coef[3]+qnorm(0.975)*sqrt(mc[3,3]))
PA08Q08CATuma ou várias vezes por semana
0.7168291

```

Anexo 17 - Exemplo perfil dos pais na área espaço e forma

```
### 1ª situação
x0 <- c(1,1, 0, 0,0,0,0,1,0,1*0) ### vetor do aluno
eta0 <- sum(x0*coef(modf))
ilogit(eta0) #Probabilidade estimada
[1] 0.7388477
c<-summary(modf)$cov.scaled
se<-sqrt(t(x0) %*% c %*% x0)
se
      [,1]
[1,] 0.2203134
ilogit(c(eta0-1.96*se,eta0+1.96*se)) #Intervalo de confiança para a probabilidade estimada
[1] 0.6475241 0.8133311

### 2ª situação
x0 <- c(1,1, 0, 1,0,0,0,1,0,1*0) ### vetor do aluno
eta0 <- sum(x0*coef(modf))
ilogit(eta0) #Probabilidade estimada
[1] 0.5411473
c<-summary(modf)$cov.scaled
se<-sqrt(t(x0) %*% c %*% x0)
se
      [,1]
[1,] 0.1861207
ilogit(c(eta0-1.96*se,eta0+1.96*se)) #Intervalo de confiança para a probabilidade estimada
[1] 0.4502070 0.6294269

### 3ª situação
x0 <- c(1,0,0, 1,0,1,0,0,1,0*0) ### vetor do aluno
eta0 <- sum(x0*coef(modf))
ilogit(eta0) #Probabilidade estimada
[1] 0.2011158
c<-summary(modf)$cov.scaled
se<-sqrt(t(x0) %*% c %*% x0)
se
      [,1]
[1,] 0.2706839
ilogit(c(eta0-1.96*se,eta0+1.96*se)) #Intervalo de confiança para a probabilidade estimada
[1] 0.1289948 0.2996857
```

Anexo 18 - Modelo agregado familiar na área da quantidade - univariada

Tabela 8 - Coeficientes das covariáveis no modelo logístico univariado, o seu desvio padrão, valores p associados à estatística de Pearson, de Wald e da razão de verossimilhanças, odds-Ratios (OR) e respetivos intervalos de confiança a 95%. Questão na área da quantidade- inquérito aos pais.

Variável	Coeficientes	Desvio Padrão Coeficientes	Valor p (Wald)	Valor p (RR)	Valor p (Pearson)	OR	IC a 95%	
							LI	LS
Tempo de leitura (2)				0.0153*	0.0152*			
Mais de 10 horas por semana	0.420	0.248	0.090			1.52	0.94	2.48
Entre 6 e 10 horas por semana	0.706	0.230	0.002**			2.03	1.29	3.19
Entre 1 e 5 horas por semana	0.475	0.1883	0.0116*			1.31	1.11	2.33
Conversar sobre questões políticas ou sociais (3)				0.005**	0.005**			
Uma ou duas vezes por semana	0.544	0.169	0.001***			1.72	0.23	2.40
Todos os dias ou quase todos os dias	0.296	0.175	0.090			1.35	0.96	1.90
Conversar sobre livros, filmes ou programas de televisão (3)				0.475	0.474			
Uma ou duas vezes por semana	0.224	0.198	0.258					
Todos os dias ou quase todos os dias	0.214	0.194	0.270					
Conversar sobre o aproveitamento escolar do seu filho (3)				0.290	0.287			
Uma ou duas vezes por semana	0.441	0.309	0.154			1.55	0.85	2.87
Todos os dias ou quase todos os dias	0.450	0.289	0.119			1.57	0.89	2.78
Jantar à mesa com o seu filho (3)				0.036*	0.041*			
Uma ou duas vezes por semana	1.762	0.846	0.037*			5.82	1.29	41.6
Todos os dias ou quase todos os dias	1.753	0.785	0.026*			5.77	1.48	38.0
Passar tempo				0.636	0.634			

Continuação

Variável	Coeficientes	Desvio Padrão Coeficientes	Valor p (Wald)	Valor p (RR)	Valor p (Pearson)	OR	IC a 95%	
							LI	LS
a conversar simplesmente com o seu filho (3)								
Uma ou duas vezes por semana	0.020	0.030	0.513					
Todos os dias ou quase todos os dias	0.026	0.029	0.357					
Ir a uma livraria ou a uma biblioteca com o seu filho (3)				0.703	0.702			
Uma ou duas vezes por semana	-1.186	0.221	0.401					
Todos os dias ou quase todos os dias	0.001	0.510	0.998					
Conversar com o seu filho sobre o que ele está a ler (3)				0.317	0.317			
Uma ou duas vezes por semana	0.210	0.162	0.196			1.23	0.90	1.70
Todos os dias ou quase todos os dias	0.223	0.195	0.253			1.25	0.85	1.84
Ajudar o seu filho a fazer os trabalhos de casa (3)				0.838	0.838			
Uma ou duas vezes por semana	-0.086	0.175	0.620					
Todos os dias ou quase todos os dias	-0.086	0.195	0.660					
Habilitações literárias do Pai (4)				0.001***	0.001***			
Bacharelato	-0.443	0.380	0.244			0.64	0.31	1.37
Formação profissional de nível IV	-0.592	0.316	0.061			0.55	0.30	1.03
Ensino secundário	-0.836	0.243	0.001***			0.43	0.27	0.69
1º,2º ou 3ºciclo	-1.224	0.233	0.001***			0.29	0.19	0.46
Habilitações literárias do Mãe (4)				0.001***	0.001***			
Bacharelato	-0.468	0.373	0.210			0.63	0.30	1.32
Formação profissional de nível IV	-0.385	0.315	0.222			0.68	0.37	1.27

Continuação

Variável	Coeficientes	Desvio Padrão Coeficientes	Valor p (Wald)	Valor p (RR)	Valor p (Pearson)	OR	IC a 95%	
							LI	LS
Ensino secundário	-0.768	0.217	0.001***			0.46	0.30	0.71
1º,2º ou 3ºciclo	-1.106	0.211	0.001***			0.33	0.22	0.50
rendimento anual agregado familiar (5)				0.001***	0.001***			
de 7 500 euros a 11 249 euros	0.592	0.215	0.005**			1.81	1.19	2.76
de 11 250 euros a 14 999 euros	0.401	0.250	0.108			1.49	0.92	2.44
de 15 000 euros a 18 749 euros	0.177	0.282	0.532					
de 18 750 euros a 22 499 euros	0.963	0.293	0.001***			2.62	1.49	4.71
22 500 euros ou mais	1.365	0.211	0.001***			3.92	2.60	5.95
Gastos anuais educação do aluno (6)				0.034*	0.044*			
mais de 0 e menos de 75 euros	-0.223	0.496	0.653					
de 75 euros a 3 999 euros	-0.349	0.456	0.444					
de 4 000 euros a 7 999 euros	0.587	0.552	0.288					
de 8 000 euros a 11 999 euros	0.612	0.932	0.511					
12 000 euros ou mais	0.901	1.205	0.455					
Número de filhos (7)				0.042*	0.042*			
Dois	-0.029	0.168	0.863					
Três	-0.284	0.236	0.228			0.75	0.47	1.19
Quatro ou cinco	-1.068	0.428	0.013*			0.34	0.14	0.78
Conversar, por sua iniciativa, com um professor sobre o comportamento ou os progressos do seu filho				0.621	0.678			
Não	0.078	0.158	0.621					
Conversar, a pedido de um dos professores, sobre o comportamento ou os progressos do seu filho				0.006**	0.008**			

Continuação

Variável	Coeficientes	Desvio Padrão Coeficientes	Valor p (Wald)	Valor p (RR)	Valor p (Pearson)	OR	IC a 95%	
							LI	LS
Não	0.392	0.144	0.006**			1.48	1.12	1.96
Participar em atividades ligadas à gestão da escola, por exemplo, associação de pais.				0.184	0.215			
Não	0.245	0.184	0.183			1.28	0.89	1.83

- (1) Taxas de probabilidades
- (2) Categoria de referência - Menos de uma hora por semana
- (3) Categoria de referência – Nunca, uma ou duas vezes por mês
- (4) Categoria de referência – Licenciatura (ou equivalente) / Mestrado / Doutorado
- (5) Categoria de referência – menos de 7500 euros
- (6) Categoria de referência- sem gastos
- (7) Categoria de referência – Um filho

Anexo 19 – Passos da modelação modelo ordinal

```
modo<-read.table("alunosordinal.csv", header = T, sep = ";", dec = ".")
attach (modo)
modo$ZONA<-factor(modo$ZONA)
modo$ST01Q00<-factor(modo$ST01Q00)
modo$ST04Q01<-factor(modo$ST04Q01,labels=c("Feminino","Masculino"))
modo$ST05Q01<-factor(modo$ST05Q01,labels=c("Não","Sim um ano","Sim mais de um ano"))
modo$ST07Q00<-factor(modo$ST07Q00,labels=c("Não","Sim ","Sim dois ou mais"))
modo$ST20Q05<-factor(modo$ST20Q05,labels=c("Sim","Não"))
modo$ST22Q01<-factor(modo$ST22Q01,labels=c("0-10","11-25","26-100","101-200","201-500","mais 500"))
modo$ST27Q02<-factor(modo$ST27Q02, labels=c("quase nunca","às vezes","muitas vezes", "quase sempre"))
modo$ST27Q04<-factor(modo$ST27Q04, labels=c("quase nunca","às vezes","muitas vezes", "quase sempre"))
modo$ST27Q06<-factor(modo$ST27Q06, labels=c("quase nunca","às vezes","muitas vezes", "quase sempre"))
modo$ST05Q01CAT<-factor(modo$ST05Q01CAT,labels=c("Não","Sim"))
modo$ST22Q01CAT<-factor(modo$ST22Q01CAT,labels=c("1-25","26-200",">201"))
modo$ST27Q06CAT<-factor(modo$ST27Q06CAT, labels=c("quase nunca","muitas vezes"))
modo$ST12Q01<-factor(modo$ST12Q01,labels=c("tempo inteiro","tempo parcial","desempregada","outra"))
modo$ST16Q01<-factor(modo$ST16Q01,labels=c("tempo inteiro","tempo parcial","desempregada","outra"))
modo$ST20Q01<-factor(modo$ST20Q01,labels=c("Sim","Não"))
modo$ST20Q02<-factor(modo$ST20Q02,labels=c("Sim","Não"))
modo$ST20Q03<-factor(modo$ST20Q03,labels=c("Sim","Não"))
modo$ST20Q04<-factor(modo$ST20Q04,labels=c("Sim","Não"))
modo$ST20Q06<-factor(modo$ST20Q06,labels=c("Sim","Não"))
modo$ST20Q10<-factor(modo$ST20Q10,labels=c("Sim","Não"))
modo$ST23Q01<-factor(modo$ST23Q01,labels=c("0m","1-30m","31-60m","61-120m","mais de 120m"))
modo$ST27Q01<-factor(modo$ST27Q01, labels=c("quase nunca","às vezes","muitas vezes", "quase sempre"))
modo$ST27Q08<-factor(modo$ST27Q08, labels=c("quase nunca","às vezes","muitas vezes", "quase sempre"))
modo$ST27Q09<-factor(modo$ST27Q09, labels=c("quase nunca","às vezes","muitas vezes", "quase sempre"))
modo$ST23Q01CAT<-factor(modo$ST23Q01CAT,labels=c("0m","1-120m","mais de 120m"))
modo$ST27Q01CAT<-factor(modo$ST27Q01CAT, labels=c("por nunca","quase sempre"))
modo$ST27Q09CAT<-factor(modo$ST27Q09CAT, labels=c("quase nunca","às vezes","muitas vezes"))
modo$ZONACAT<-factor(modo$ZONACAT)
library (rms)
modO1<-lrm(RESPOSTA~ZONACAT+ST01Q00+ST04Q01+ST05Q01+ST07Q00+ST20Q05+ST22Q01+
ST27Q02+ST27Q04+ST27Q06,data= modo, x=TRUE, y=TRUE)
modO1
##retirei ST20Q05
modO2<-lrm(RESPOSTA~ST01Q00+ST04Q01+ST05Q01+ST07Q00+ST22Q01+ST27Q02+ST27Q04+ST27Q06,data=
modo, x=TRUE, y=TRUE)
modO2
##retirei ST27Q02
modO3<-lrm(RESPOSTA~ST01Q00+ST04Q01+ST05Q01+ST07Q00+ST22Q01+ST27Q04+ST27Q06,data= modo,
x=TRUE, y=TRUE)
modO3
## retirei ST27Q04
modO4<-lrm(RESPOSTA~ST01Q00+ST04Q01+ST05Q01+ST07Q00+ST22Q01+ST27Q06,data= modo, x=TRUE, y=TRUE)
modO4
## retirei ST01Q00
##acrescentar as não colocadas inicialmente
modO5<-lrm(RESPOSTA~ST27Q09CAT+ST27Q01CAT+ST04Q01+ST05Q01CAT+ST07Q00+
ST22Q01CAT+ST27Q06CAT,data= modo, x=TRUE, y=TRUE)
modO5
## interações
modO6<-
lrm(RESPOSTA~ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+ST27Q09CAT+ST27Q01CAT+ST23Q01
CAT+ST04Q01*ST05Q01CAT,data= modo, x=TRUE, y=TRUE)
modO6
modO7<-lrm(RESPOSTA~ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+ST27Q09CAT+
ST27Q01CAT+ST23Q01CAT+ST04Q01*ST07Q00,data= modo, x=TRUE, y=TRUE)
modO7
```



```

modO8<-lrm(RESPOSTA~ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+ST27Q09CAT+ST27Q01CAT
+ST23Q01CAT+ST04Q01*ST27Q06CAT,data= modo, x=TRUE, y=TRUE)
modO8
modO8.1<-lrm(RESPOSTA~ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+ST27Q09CAT+
ST27Q01CAT+ST23Q01CAT+ST04Q01*ST27Q09CAT,data= modo, x=TRUE, y=TRUE)
modO8.1
modO8.2<-lrm(RESPOSTA~ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+ST27Q09CAT+
ST27Q01CAT+ST23Q01CAT+ST04Q01*ST23Q01CAT,data= modo, x=TRUE, y=TRUE)
modO8.2
modO9<-lrm(RESPOSTA~ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+
ST27Q09CAT+ST27Q01CAT+ST23Q01CAT+ST05Q01CAT*ST07Q00,data= modo, x=TRUE, y=TRUE)
modO9
modO10<-lrm(RESPOSTA~ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+
ST27Q09CAT+ST27Q01CAT+ST23Q01CAT+ST05Q01CAT*ST27Q06CAT,data= modo, x=TRUE, y=TRUE)
modO10
modO10.1<-lrm(RESPOSTA~ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+
ST27Q09CAT+ST27Q01CAT+ST23Q01CAT+ST05Q01CAT*ST27Q09CAT,data= modo, x=TRUE, y=TRUE)
modO10.1
modO10.2<-lrm(RESPOSTA~ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+ST27Q09CAT+
ST27Q01CAT+ST23Q01CAT+ST05Q01CAT*ST27Q01CAT,data= modo, x=TRUE, y=TRUE)
modO10.2
modO10.3<-lrm(RESPOSTA~ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+ST27Q09CAT+
ST27Q01CAT+ST23Q01CAT+ST05Q01CAT*ST23Q01CAT,data= modo, x=TRUE, y=TRUE)
modO10.3
modO11<-
lrm(RESPOSTA~ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+ST27Q09CAT+ST27Q01CAT+
ST23Q01CAT+ST07Q00*ST27Q06CAT,data= modo, x=TRUE, y=TRUE)
modO11
modO11.1<-lrm(RESPOSTA~ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+ST27Q09CAT+
ST27Q01CAT+ST23Q01CAT+ST07Q00*ST27Q09CAT,data= modo, x=TRUE, y=TRUE)
modO11.1
modO11.2<-
lrm(RESPOSTA~ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+ST27Q09CAT+ST27Q01CAT+
ST23Q01CAT+ST07Q00*ST23Q01CAT,data= modo, x=TRUE, y=TRUE)
modO11.2
modO12<-lrm(RESPOSTA~ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+ST27Q09CAT+
ST27Q01CAT+ST23Q01CAT+ST22Q01CAT*ST27Q06CAT,data= modo, x=TRUE, y=TRUE)
modO12
modO12.1<-lrm(RESPOSTA~ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+ST27Q09CAT+
ST27Q01CAT+ST23Q01CAT+ST27Q09CAT*ST27Q01CAT,data= modo, x=TRUE, y=TRUE)
modO12.1
modO13<-lrm(RESPOSTA~ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+ST27Q09CAT+
ST27Q01CAT+ST23Q01CAT+ST27Q09CAT*ST23Q01CAT,data= modo, x=TRUE, y=TRUE)
modO13
modO14<-
lrm(RESPOSTA~ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+ST27Q09CAT+ST27Q01CAT+
ST23Q01CAT+ST27Q01CAT*ST23Q01CAT,data= modo, x=TRUE, y=TRUE)
modO14
modO5<-lrm(RESPOSTA~ZONACAT+ST04Q01+ST05Q01CAT+ST07Q00+ST22Q01CAT+ST27Q06CAT+
ST23Q01CAT,data= modo, x=TRUE, y=TRUE)
modO5

```

Anexo 20 - Análise dos resíduos – possibilidades proporcionais

Figura 5 - Resíduos score das covariável modelo logística ordinal

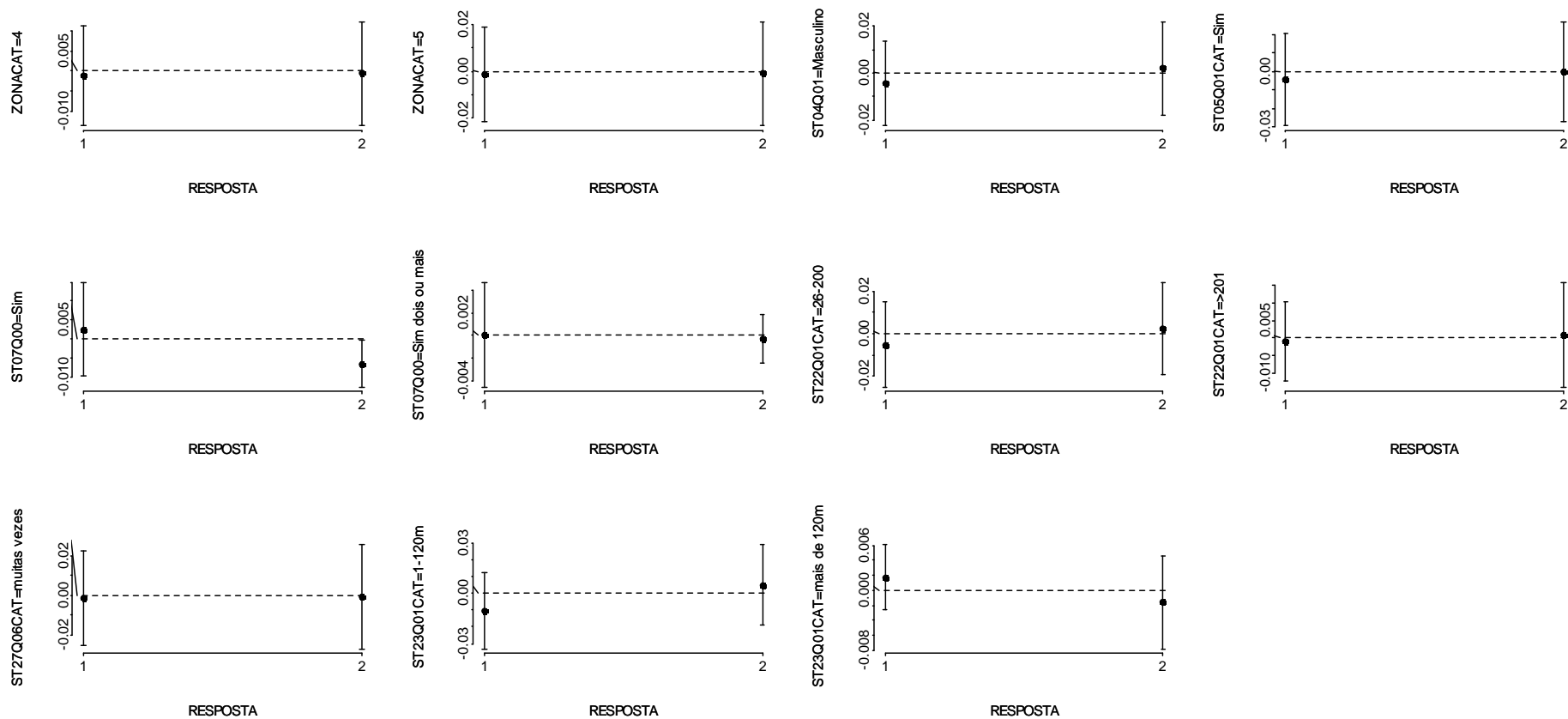
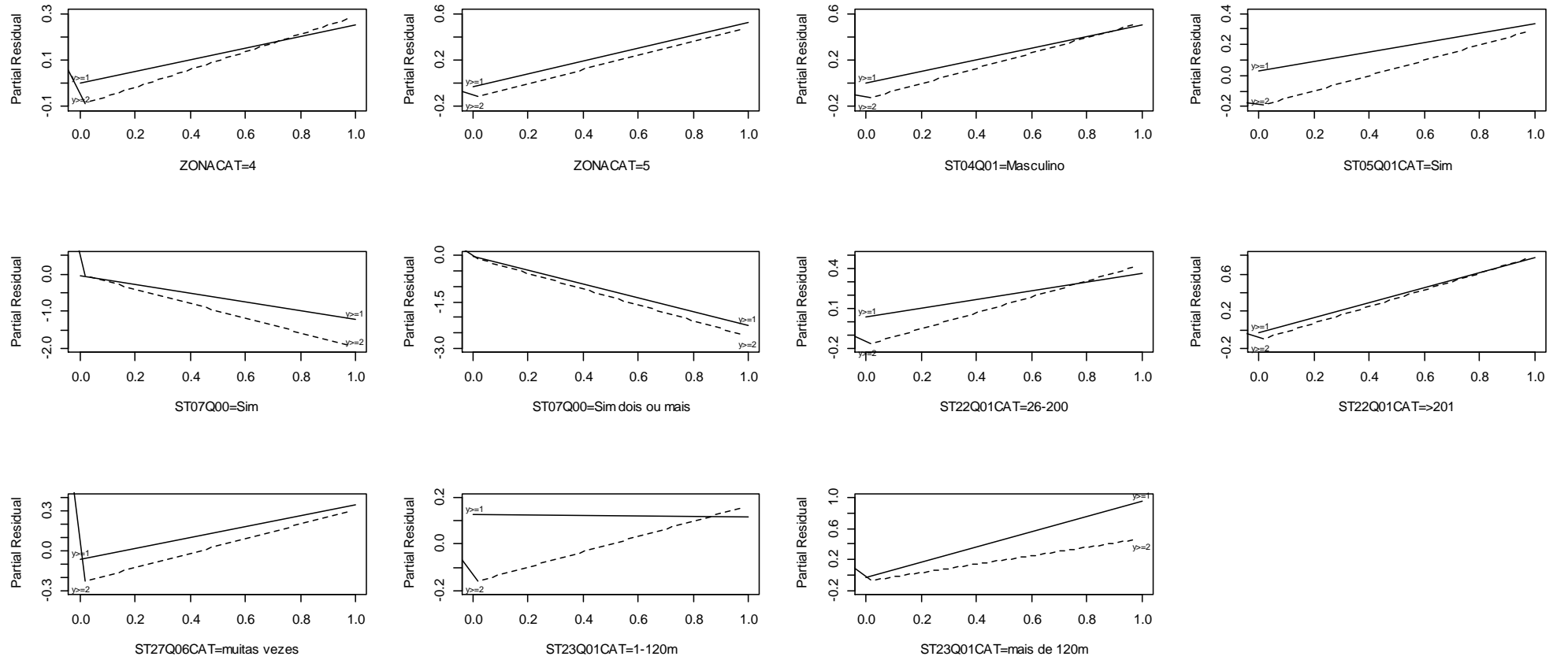


Figura 6 - Resíduos parcial das covariável modelo logística ordinal



Anexo 21 - Teste qui-quadrado de Pearson - Países

Tabela 9 - Teste qui-quadrado – acerto à questão na área espaço e forma versus número de livros em casa

País	Valor de p
Espanha	0,997
França	0,303
Reino Unido	0,194
Grécia	0,665
Itália	0,319

Tabela 10 - Teste qui-quadrado – acerto à questão na área da quantidade versus número de livros em casa

País	Valor de p
Espanha	0,612
França	0,363
Reino Unido	0,853
Grécia	0,104
Itália	0,280

Tabela 11 - Teste qui-quadrado – acerto à questão na área espaço e forma versus tempo dedicado à leitura

País	Valor de p
Espanha	0,255
França	0,130
Reino Unido	0,149
Grécia	0,760
Itália	0,441

Tabela 12 - Teste qui-quadrado – acerto à questão na área da quantidade versus tempo dedicado à leitura

País	Valor de p
Espanha	0,360
França	0,957
Reino Unido	0,406
Grécia	0,111
Itália	0,372

