

**Universidade do Minho**  
Instituto de Educação

Delson Alexandre Alberto Zioto Mugabe

**Evolução dos significados pessoais de  
associação estatística em estudantes  
universitários moçambicanos de  
Matemática através do ensino**



**Universidade do Minho**  
Instituto de Educação

Delson Alexandre Alberto Zioto Mugabe

**Evolução dos significados pessoais de  
associação estatística em estudantes  
universitários moçambicanos de  
Matemática através do ensino**

Dissertação de Mestrado  
Mestrado em Ciências da Educação  
Área de Especialização em Supervisão Pedagógica na  
Educação Matemática

Trabalho realizado sob a orientação do  
**Doutor José António Fernandes**

Julho de 2012

## **DECLARAÇÃO**

Nome: Delson Alexandre Alberto Zioto Mugabe

Endereço eletrónico: [delsonmugabe@yahoo.com.br](mailto:delsonmugabe@yahoo.com.br)

Tel. 925199688

Número do Bilhete de Identidade: 030100242068F

Título da tese: Evolução dos significados pessoais de associação estatística em estudantes universitários moçambicanos de Matemática através do ensino

Orientador:

Doutor José António Fernandes

Ano de conclusão: 2012

Mestrado em Ciências da Educação, Área de especialização em Supervisão Pedagógica na Educação Matemática

É autorizada a reprodução integral desta tese para efeitos de investigação, mediante declaração escrita do interessado, que a tal se compromete.

Universidade do Minho 15 de Julho de 2012

## **AGRADECIMENTOS**

A realização deste trabalho não teria sido possível sem o esforço coletivo e conjugado de algumas pessoas e instituições. Quero destacar o papel do meu supervisor, Doutor José António Fernandes, pelos seus sábios conselhos, pela disponibilidade em satisfazer algumas dúvidas inerentes ao trabalho e pelo encorajamento que sempre me deu no sentido de prosseguir com a pesquisa. Agradeço também pela amizade que sempre demonstrou para comigo ao longo da concretização de todo trabalho.

A todos os meus professores do curso de Mestrado em Ciências da Educação, área de especialização em Supervisão Pedagógica na Educação Matemática, em especial ao Doutor Floriano Viseu e à Doutora Maria Helena Martinho, que ao longo do ano curricular do mestrado me proporcionaram conhecimentos e habilidades científicas, e mais do que isso, transmitiram-me valores, de modo a tornarmo-nos Homens e excelentes profissionais de Educação.

Aos meus colegas do curso, pelo apoio e ajuda tanto em termos materiais como também pelo encorajamento que sempre e incondicionalmente me deram nos momentos mais difíceis.

Agradeço imensamente à Direção da Universidade Pedagógica por ter autorizado a efetivação da investigação na sua instituição, aos Docentes, especialmente ao Mestre Abudo Atumane Ossofo e à Mestre Margarida, que puseram à minha disposição de forma incondicional as suas turmas para a aplicação dos instrumentos de recolha de dados. Não poderia deixar de agradecer também aos estudantes do 2º ano do curso de Matemática da Universidade Pedagógica (Delegação de Nampula e Maputo), do ano letivo de 2011/2012, pois sem eles a realização desta pesquisa teria sido impossível.

Finalmente, não encontro palavras para agradecer aos meus pais e aos meus irmãos o apoio que me têm dado para o sucesso dos meus estudos.



EVOLUÇÃO DOS SIGNIFICADOS PESSOAIS DE ASSOCIAÇÃO ESTATÍSTICA EM ESTUDANTES  
UNIVERSITÁRIOS MOÇAMBICANOS DE MATEMÁTICA ATRAVÉS DO ENSINO

Delson Alexandre Alberto Zioto Mugabe

Mestrado em Ciências da Educação, Área de especialização em Supervisão Pedagógica na  
Educação Matemática

Universidade do Minho, 2012

**RESUMO**

O presente estudo teve por propósito fundamental investigar os significados que estudantes universitários atribuem à associação estatística antes e de depois do ensino do tema da correlação e regressão linear, e centrou-se nas duas seguintes questões de investigação: 1– Que significados atribuem os estudantes à associação estatística? Esses significados alteram-se com o ensino do tema da correlação e regressão linear?; 2– Quais as aquisições conceituais sobre a associação estatística entre duas variáveis que os estudantes adquirem com o ensino do tema da correlação e regressão linear?

O estudo foi conduzido segundo uma abordagem de investigação qualitativa, de natureza interpretativa, centrada na análise das respostas e estratégias adotadas pelos estudantes em tarefas sobre a temática em estudo. No estudo participaram 57 estudantes do 2º ano do curso de Matemática da Universidade Pedagógica de Moçambique. Os dados usados no estudo resultaram da aplicação de um questionário e um teste aos estudantes. Com o questionário, aplicado imediatamente antes depois do ensino do tema da correlação e regressão linear, pretendeu-se avaliar sobretudo as aquisições intuitivas e com o teste, aplicado apenas depois do ensino, pretendeu-se avaliar sobretudo as aquisições conceituais.

Em termos de resultados obtidos, evidencia-se um impacto diferente do ensino sobre as conceções e sobre as aquisições conceituais. Algumas conceções incorretas persistiram no pós-ensino, sobretudo em relação aos tipos de problemas menos enfatizados durante o ensino, como foi o caso do estudo da associação numa tabela de contingência e de variáveis dicotómicas. Já no caso das aquisições conceituais, em geral, elas foram satisfatórias, sobretudo quando comparadas com as aquisições mais intuitivas reveladas pelos estudantes no questionário.



THE EVOLUTION OF THE PERSONAL MEANINGS OF STATISTICAL ASSOCIATION IN  
MOZAMBICAN UNIVERSITY MATHEMATICS STUDENTS THROUGH TEACHING

Delson Alexandre Alberto Zioto Mugabe

Master of Education, Area of Specialization in Pedagogical Supervision in Mathematics Education

University of Minho, 2012

**ABSTRACT**

The present study had as its fundamental purpose to investigate the meanings that the university students attribute to the statistical association before and after the teaching of the topic of correlation and linear regression, and it focused on the two following research questions: 1 - What meanings do students attach to the statistical association? Do these meanings change with the teaching of the topic of correlation and linear regression? 2 - What conceptual acquisitions on the statistical association between two variables do the students acquire with the teaching of the topic of linear correlation and regression?

The study was conducted according to a qualitative research approach, with an interpretative nature, focused on the analysis of responses and strategies adopted by the students in tasks on the subject under investigation. In the study, participated 57 students from the 2nd year of Mathematics Course of the Pedagogical University of Mozambique. The data used in this study resulted from the application of a questionnaire and a test to students. With the questionnaire, being applied immediately before and after the teaching of the topic of linear correlation and regression, it was mainly intended to assess issues related to the intuitive acquisitions and with the test applied only after the teaching, it was primarily intended to evaluate the conceptual acquisition.

In terms of the results obtained, it is evident a different impact from the teaching about the conceptions and the conceptual acquisitions. Some incorrect conceptions persisted in the post-teaching stage, especially in relation to the types of problems less stressed during the teaching process, such as the case of the study of the association in a contingency table and dichotomous variables. In the case of the conceptual acquisitions, in general, they were satisfactory, particularly when compared with the more intuitive acquisitions revealed by the students involved in the study.



## ÍNDICE

<b>AGRADECIMENTOS</b> .....	<b>iii</b>
<b>RESUMO</b> .....	<b>v</b>
<b>ABSTRACT</b> .....	<b>vii</b>
<b>ÍNDICE</b> .....	<b>ix</b>
<b>LISTA DE TABELAS</b> .....	<b>xii</b>
<b>LISTA DE QUADROS</b> .....	<b>xiv</b>
<b>LISTA DE FIGURAS</b> .....	<b>xv</b>
<b>CAPÍTULO I – INTRODUÇÃO</b> .....	<b>1</b>
1.1. Problema e questões de investigação.....	1
1.2. Importância do estudo.....	3
1.3. Organização do estudo .....	6
<b>CAPÍTULO II – REVISÃO DE LITERATURA</b> .....	<b>9</b>
2.1. Significado institucional e pessoal dos objetos matemáticos .....	9
2.2. Ensino da estatística em Moçambique .....	12
2.3. Crenças e concepções: o caso da associação estatística .....	18
2.3.1. Crenças e concepções .....	18
2.3.2. Crenças e concepções sobre associação estatística .....	22
Influência das teorias prévias.....	24
Associação e causalidade .....	26
Influência do contexto e a representação da informação .....	27
2.4. Estratégias de avaliação da associação estatística .....	28
2.4.1. Estratégias de avaliação da associação estatística em tabelas de contingência .....	28
2.4.2. Estratégias de avaliação da associação estatística na comparação de amostras .....	33
2.4.3. Estratégias de avaliação da associação estatística em gráficos de dispersão.....	35
<b>CAPÍTULO III – METODOLOGIA</b> .....	<b>39</b>
3.1. Opções metodológicas.....	39
3.2. Participantes .....	40
3.3. Métodos de recolha e análise de dados.....	42
3.3.1. Descrição dos instrumentos de recolha de dados.....	42
3.3.2. Recolha de dados.....	43

3.3.3. Análise de dados .....	45
<b>CAPÍTULO IV – APRESENTAÇÃO DOS RESULTADOS.....</b>	<b>47</b>
4.1. Evolução dos significados pessoais de associação estatística com o ensino .....	47
4.1.1. Avaliação da associação estatística numa tabela de contingência de $2 \times 2$ .....	47
Questão 1 .....	48
4.1.2. Avaliação da associação estatística numa tabela de duas amostras.....	59
Questão 2 .....	59
Questão 3 .....	62
Questão 4 .....	66
4.1.3. Avaliação da associação estatística num diagrama de dispersão .....	69
Questão 5 .....	69
Questão 6 .....	74
4.1.4. Evolução dos significados sobre associação estatística com o ensino .....	80
4.2. Aquisições conceituais da associação estatística com o ensino .....	82
4.2.2. Análise das respostas dos estudantes às questões do teste.....	82
Questão 1 .....	82
Questão 2 .....	84
Questão 3 .....	86
Questão 4 .....	88
Questão 5 .....	89
Questão 6 .....	95
Questão 7 .....	100
Questão 8 .....	102
2.2.2. Síntese sobre as aquisições conceituais dos estudantes.....	106
<b>CAPÍTULO V – CONCLUSÃO, IMPLICAÇÕES E RECOMENDAÇÕES .....</b>	<b>111</b>
5.1. Síntese do estudo.....	111
5.2. Que significados atribuem os estudantes à associação estatística? Esses significados alteram-se com o ensino do tema correlação e regressão linear? .....	113
5.3. Quais as aquisições conceituais sobre a associação estatística entre duas variáveis que os estudantes adquirem com o ensino do tema correlação e regressão linear?.....	118
5.4. Implicações do estudo para o ensino .....	121
5.5. Limitações e sugestões de futuras investigações .....	122

<b>BIBLIOGRAFIA.....</b>	<b>123</b>
<b>ANEXO .....</b>	<b>129</b>
Questionário.....	131
Teste.....	139

## LISTA DE TABELAS

Tabela 1. Formato típico duma tabela de contingência $2 \times 2$ .....	29
Tabela 2. Frequência absoluta (percentagem) de estudantes segundo o número de células a que recorreram nas suas respostas .....	48
Tabela 3 Frequência absoluta (percentagem) dos diferentes tipos de resposta segundo o número de células envolvidas.....	49
Tabela 4 Frequência absoluta (percentagem) de estudantes nas diferentes estratégias quando foi usada informação de uma célula .....	50
Tabela 5 Frequência absoluta (percentagem) de estudantes nas diferentes estratégias quando foi usada informação de duas células.....	51
Tabela 6 Frequência absoluta (percentagem) de estudantes nas diferentes estratégias quando foi usada informação de quatro células. ....	53
Tabela 7 Frequência absoluta (percentagem) de estudantes nas diferentes estratégias da questão 2.....	59
Tabela 8 Frequência absoluta (percentagem) de estudantes nas diferentes estratégias da questão 3.....	62
Tabela 9 Frequência absoluta (percentagem) de estudantes nas diferentes estratégias da questão 4.....	66
Tabela 10 Frequência absoluta (percentagem) de estudantes nas diferentes estratégias da questão 5c).....	72
Tabela 11 Frequência absoluta (percentagem) de estudantes nas diferentes respostas da questão 6b.....	75
Tabela 12 Frequência absoluta (percentagem) de estudantes nas respostas à questão 1 .....	83
Tabela 13 Frequência absoluta (percentagem) de estudantes nas respostas à questão 2.....	85
Tabela 14. Frequência absoluta (percentagem) de estudantes nas respostas à questão 3.....	87
Tabela 15. Valores dos coeficientes de correlação esperados e médias, desvios padrão e amplitudes totais dos valores dos coeficientes de correlação estimados pelos estudantes. ....	90

Tabela 16. Frequências (percentagens) das respostas corretas e incorretas nas  
diferentes questões do teste .....107

## **LISTA DE QUADROS**

QUADRO 1 – Organização do ensino básico e secundário moçambicano por ciclos e classes .....	13
--	----

## LISTA DE FIGURAS

Figura 1. Resolução da questão 1 pelo estudante E12 no pré-ensino. ....	52
Figura 2. Resolução da questão 1 pelo estudante E2 no pós-ensino.....	54
Figura 3. Resolução da questão 1 pelo estudante E20 no pré-ensino. ....	54
Figura 4. Resolução da questão 1 pelo estudante E40 no pós-ensino. ....	54
Figura 5. Resolução da questão 1 pelo estudante E23 no pós-ensino. ....	55
Figura 6. Resolução da questão 1 pelo estudante E34 no pós-ensino. ....	56
Figura 7. Resolução da questão 1 pelo estudante E40 no pré-ensino. ....	57
Figura 8. Resolução da questão 1 pelo estudante E50 no pós-ensino. ....	58
Figura 9. Resolução da questão 3 pelo estudante E29 no pós-ensino. ....	63
Figura 10. Resolução da questão 3 pelo estudante E29 no pós-ensino. ....	64
Figura 11. Resolução da questão 3 pelo estudante E29 no pós-ensino. ....	67
Figura 12. Resolução da questão 5b pelo estudante E7 no pós-ensino.....	71
Figura 13. Resolução da questão 6c) pelo estudante E16 no pós-ensino.....	79
Figura 14. Resolução da questão 5d) pelo estudante E11. ....	95
Figura 15. Resolução da questão 6a) pelo estudante E26.....	96
Figura 16. Resolução da questão 6a) pelo estudante E50.....	97
Figura 17. Resolução da questão 6b) pelo estudante E39. ....	98
Figura 18. Resolução da questão 6b) pelo estudante E45. ....	99
Figura 19. Resolução da questão 6c) pelo estudante E9.....	99
Figura 20. Resolução da questão 6c) pelo estudante E55.....	100

Figura 21. Resolução da questão 7a) pelo estudante E42.....	101
Figura 22. Resolução da questão 7b) pelo estudante E2.....	102
Figura 23. Resolução da questão 8a) pelo estudante E27.....	103
Figura 24. Resolução da questão 8b) pelo estudante E16. ....	104
Figura 25. Resolução da questão 8b) pelo estudante E48. ....	105

# **CAPÍTULO I**

## **INTRODUÇÃO**

Este capítulo encontra-se organizado em três secções: a primeira secção é dedicada à apresentação do problema e das questões de investigação; na segunda secção apresentam-se as principais razões da importância do estudo; e, finalmente, na terceira secção apresenta-se uma breve organização do estudo.

### **1.1. Problema e questões de investigação**

Atualmente, o ensino de estatística tem aumentado substancialmente em muitos países devido ao seu reconhecimento na educação integral dos cidadãos. A ênfase dada à estatística nos currículos escolares requer uma preparação intensiva dos professores em ordem a permitir a consecução dos seus objetivos educacionais com sucesso (Batanero, Godino, Vallecillos, Green & Holmes, 1994). Esta preparação deve incluir o conhecimento das conceções e das dificuldades e erros apresentados pelos alunos durante a aprendizagem dos conteúdos estatísticos. Batanero et al. (1994) consideram que “a análise dos erros deve ser considerada como uma estratégia promissora para clarificar algumas questões fundamentais da aprendizagem matemática” (p. 2).

Ainda em relação aos erros dos alunos em estatística, Batanero (2001) advoga que se queremos formar cidadãos competentes na interpretação da informação estatística que se encontra na vida profissional e quotidiana, será importante que o seu ensino tenha em conta as possíveis dificuldades dos alunos na compreensão dos conceitos estatísticos. Consubstanciando esta ideia, Batanero et al. (1994) afirmam que falar sobre os erros e as dificuldades dos alunos em relação aos conceitos estatísticos é urgente porque muitos deles poderiam ser evitados se fossem seguidas indicações didáticas adequadas. Consequentemente, a exploração dos erros e das dificuldades mais frequentes e persistentes dos alunos contribui para a compreensão dos seus processos de aprendizagem.

Uma importante parte das investigações que têm sido realizadas na educação matemática centra-se em factos observados nas respostas dos alunos durante a realização de uma determinada tarefa. Às vezes, as suas respostas apresentam discrepâncias ou simplesmente não

são capazes de dar respostas corretas na sua totalidade quando comparadas com uma resposta standard (Batanero, et al., 1994). Nesse sentido, tentando encontrar explicações ou respostas sobre tais dificuldades, as investigações didáticas de matemática interessam-se por explicar as razões que fazem com que o aluno se equivoque quando se lhe solicita a realização de uma determinada tarefa.

Segundo Batanero (2001), referindo-se particularmente à associação estatística, os erros que os estudantes apresentam não aparecem de forma aleatória ou imprevisível, antes, pelo contrário, os estudantes possuem intuições corretas em relação a algumas propriedades da associação estatística que coexistem com outras concepções erradas, o que os leva a fazer interpretações incorretas ou parcialmente corretas da associação estatística.

Um princípio amplamente assumido na psicologia educativa, enunciado por Ausubel (1983), é que “o fator mais importante que influencia a aprendizagem é o conhecimento prévio dos alunos” (p. 66). Como consequência deste princípio, tem-se verificado um interesse crescente na realização de investigações centradas na identificação e exploração das ideias prévias dos estudantes.

Este autor salienta ainda que algumas destas concepções permitem resolver corretamente determinados problemas, mas são inapropriadas para resolver outros problemas mais gerais, e que muitas vezes o sujeito mostra resistência em alterar estas concepções. Portanto, estas concepções erradas necessitam de serem identificadas pelos professores de modo a planificar uma estratégia de ensino adequada (Batanero et al., 1994). Daí, a importância de os professores, nas suas aulas, tomarem em consideração as concepções e as dificuldades dos alunos, com o propósito de os ajudar a vencer essas dificuldades e erros.

A correlação e regressão são conceitos fundamentais para muitos métodos estatísticos que são usados para fazer predições e tomar decisões em várias áreas, designadamente na política, na economia e na educação (Estepa & Batanero, 1996). Estudos destes conceitos, muitas vezes, envolvem a procura de explicações causais para compreender o meio que nos rodeia. Contudo, a associação não implica necessariamente uma relação causal, dado que, por vezes, pode existir um alto coeficiente de correlação entre as variáveis sem que haja efetivamente um vínculo causal (Estepa, Batanero & Sánchez, 1999).

Além destas dificuldades, as investigações psicológicas têm mostrado que a habilidade de emitir juízos de associação não se desenvolve intuitivamente e de uma forma geral é pobre. (Nilsbett & Ross, 1980). De acordo com Chapman e Chapman (1969), por vezes, as pessoas

adultas baseiam os seus juízos de associação nas suas crenças prévias acerca o tipo de associação que deveria existir entre variáveis, em vez de analisar os dados empíricos apresentados no problema. Portanto, a existência de tais preconceções constitui outra dificuldade do ensino da associação.

Não obstante a importância deste tema, acima referida, e da constatação feita por psicólogos e pesquisadores de educação matemática sobre a fraca capacidade de estabelecer uma correta avaliação da associação, este tema não tem sido objeto de muitas pesquisas de modo a determinar e aprofundar a natureza dos erros dos alunos nesta temática. Nesta ordem de ideias, revela-se de grande importância a realização desse estudo, especificamente no que respeita a determinar as concepções dos estudantes universitários no conceito de associação estatística.

Para Estepa et al. (1999), o julgamento correto de associação entre duas variáveis não é suficiente para determinar as preconceções que os estudantes têm acerca deste tópico, pois pode-se obter uma solução correta para um determinado problema usando um procedimento incorreto. Para estes autores, na perspectiva da educação matemática, tanto os julgamentos corretos como incorretos de associação são necessários para uma adequada compreensão da ideia que os estudantes têm sobre associação estatística. Dai, no nosso estudo privilegiar-se não só a avaliação do “correto” ou “incorreto” mas sobretudo às estratégias de resolução adotadas pelos estudantes, como forma de inferir as suas concepções.

Neste contexto, o objeto do estudo que se pretende desenvolver foca-se nas ideias dos estudantes sobre a associação estatística e correlação linear. Mais especificamente, com o estudo pretende-se analisar as concepções e estratégias de estudantes do 2.º ano do curso de Matemática na avaliação da associação estatística, antes e depois da lecionação do tema. No âmbito desta problemática, procura-se responder as duas seguintes questões de investigação:

- 1 – Que significados atribuem os estudantes à associação estatística? Esses significados alteram-se com o ensino do tema correlação e regressão linear?
- 2 – Quais as aquisições conceituais sobre a associação estatística entre duas variáveis que os estudantes adquirem com o ensino do tema correlação e regressão linear?

## **1.2. Importância do estudo**

O interesse pelo estudo dos significados que os estudantes têm acerca da associação estatística resulta do pressuposto de que a atuação e o comportamento dos professores na sala

de aula, muitas vezes, está associado às concepções que eles próprios têm sobre a natureza da matemática e do seu ensino e aprendizagem.

Thompson (1992) afirma que o interesse pelo estudo das crenças e concepções dos professores foi motivada por uma mudança nos paradigmas no que concerne à investigação no ensino. Esta mudança de paradigma, do processo-produto, cujo objeto de estudo era o comportamento do professor, para um paradigma do pensamento e processo de tomada de decisão, levou com que os investigadores se interessassem em identificar e compreender melhor a composição e estrutura do sistema de crenças e concepções dos professores. É nesse sentido que Dougherty (1990, citado em Thompson, 1992) afirma que as crenças dos professores acerca da matemática e do seu ensino jogam um papel significativo na formação de padrões característicos do seu comportamento na instrução.

De acordo com Thompson (1992), a abordagem que os professores fazem ao ensino, depende fundamentalmente dos seus sistemas de crenças, em particular das suas concepções sobre a natureza e significado da matemática e dos seus modelos mentais relativos ao ensino e aprendizagem.

Portanto, apesar da popularidade que os estudos sobre as crenças dos professores vêm tendo, a natureza dessas crenças acerca do ensino e aprendizagem e a sua influência na prática de ensino não têm sido muito exploradas na literatura da educação matemática. De acordo com os autores supra citados, estudos sobre a influência dos sistemas de crenças não se restringem apenas a professores e ao ensino, afirmando que abrange também aspetos ligados à aprendizagem do estudante, isto é, procura analisar o papel desempenhado pelas crenças ou concepções da matemática dos estudantes na interpretação e resolução de problemas ou tarefas.

Ora, uma vez que os sujeitos de pesquisa do presente estudo são estudantes e potenciais futuros professores de matemática do ensino secundário e atendendo à tendência para incorporar no currículo do ensino secundário aspetos ligados à associação e regressão dos currículos do ensino secundário da maioria dos países desenvolvidos, achamos pertinente saber quais são as concepções, crenças e erros mais frequentes que os estudantes (futuros professores) apresentam nesta temática. Subsequentemente, com base neste conhecimento, podem ser desenvolvidas atividades com vista a minimizar tais concepções limitadas e tais erros, de modo a contribuir para a melhoria do processo de ensino e aprendizagem deste tema.

Sobre a influência das crenças na *performance* de resolução de problemas, Estepa e Batanero, nos seus variados estudos sobre a associação estatística, mostraram que os

indivíduos, com particular destaque para os adultos, tomavam decisões acerca de associação estatística com base nas suas crenças ou ideias preconcebidas, sem terem em conta os dados empírico, ou seja, mantinham as suas crenças apesar da evidência fornecida sobre as variáveis em estudo.

Outro aspeto que contribui para formação das crenças, que são muitas vezes usadas na interpretação de dados e factos à nossa volta, é a experiência de vida e o ambiente cultural. Jennings, Amabile e Ross (1982) estudaram o efeito que a crença acerca do contexto do problema tem no julgamento da associação, concluindo que quando existe concordância entre os dados e as crenças pessoais, as pessoas aumentam a sua confiança na atribuição da covariação, enquanto no caso contrário se produzirá um conflito cognitivo.

Chapman e Chapman (1969) nos seus estudos apontaram as “crenças” ou preconceções com origem no contexto do problema como um fator que influencia os julgamentos de associação entre duas variáveis, pois muitas vezes as pessoas, em função da natureza do problema, mantinham as suas crenças sobre a associação que devia existir, em vez de analisarem os dados empíricos apresentados.

Inhelder e Piaget (1976) realizaram um estudo sobre o raciocínio acerca da associação estatística e consideraram a compreensão da associação como a última etapa do desenvolvimento da ideia de probabilidade. De acordo com estes autores, o conceito de associação e probabilidade estão relacionados, pois perceber a associação requerer *a priori* a compreensão da proporcionalidade, de probabilidades e de combinações. Estes autores investigaram a compreensão da associação estatística em crianças de 13 anos de idade, propondo problemas de associação entre duas variáveis qualitativas dicotômicas em tabelas de contingência de  $2 \times 2$ , e concluíram que o conceito é adquirido apenas na última fase do estágio das operações formais, por volta dos 15 anos de idade.

Os resultados de estudos desenvolvidos nesta área, no âmbito da psicologia e da didática de matemática, têm mostrado que o raciocínio correlacional dos estudantes, tendo em vista estabelecer um correto julgamento de associação entre duas variáveis, é pobre e, por vezes, alguns significados que os alunos têm sobre o mesmo são discrepantes com os significados institucionais, que se pretende que eles adquiram (Nisbett & Ross, 1980).

As evidências referidas pelos autores demonstram, por um lado, a multiplicidade de significados que os alunos mantêm sobre a associação estatística, as quais se constituem em dificuldades na aprendizagem deste tópico; e, por outro, a persistência das conceções que

muitas vezes influenciam a tomada de decisões por parte dos estudantes na resolução dos problemas. Assim, julgamos necessária a realização deste estudo para conhecer os diversos significados pessoais atribuídos pelos estudantes à associação estatística e para estudar a sua alteração com o ensino.

Acreditamos que os resultados desta investigação constituirão uma informação da maior utilidade para os professores de matemática, ao nível das concepções erróneas, das ideias intuitivas, assim como sobre a evolução das mesmas com o ensino. Permitirão também analisar estratégias adotadas pelos estudantes no processo de resolução de tarefas sobre associação estatística, as quais permitirão compreender melhor a aprendizagem dos alunos e adequar o ensino.

### **1.3. Organização do estudo**

O estudo que aqui se apresenta encontra-se organizado em cinco capítulos. No capítulo 1 – Introdução, organizado em três secções, apresentam-se o problema e as questões de investigação, bem como a importância do estudo e a sua organização.

No capítulo 2 – Revisão de literatura, apresenta-se o quadro teórico construído a partir de resultados de várias investigações relacionadas com a temática em estudo. Especificamente são abordados os tópicos seguintes: significado institucional e pessoal dos objetos matemáticos; ensino da estatística em Moçambique; conceito de crenças e concepções em geral, e em estatística em particular; e, finalmente, estratégias de avaliação da associação estatística em situações apresentadas em tabelas de contingência, em tabelas envolvendo a comparação de duas amostras e em gráficos de dispersão.

No capítulo 3 – Metodologia, explicitam-se as opções metodológicas adotadas no estudo, o processo de seleção dos participantes, bem como a sua caracterização. A descrição dos instrumentos de recolha de dados, assim como a os métodos de recolha e análise de dados são também abordados neste capítulo.

No capítulo 4– Apresentação dos resultados, descrevem-se os resultados da análise levada a cabo, tanto numa vertente qualitativa como quantitativa. A análise quantitativa compreende a comparação de percentagens de respostas corretas e incorretas decorrentes das distintas estratégias adotadas pelos estudantes nas várias questões com que foram confrontados. Em relação a componente qualitativa são apresentados os erros e estratégias dos estudantes e as concepções que daí advieram.

Por último, no capítulo 5 – Conclusão, implicações e recomendações, apresenta-se a síntese e a discussão dos principais resultados do estudo, estruturadas segundo as duas questões de investigação estabelecidas na investigação. Ainda neste capítulo referem-se algumas implicações para o ensino que derivam da investigação, assim como algumas limitações do estudo realizado e recomendações para futuras investigações.

Finalmente, incluem-se também no relatório de investigação a bibliografia utilizada e dois anexos relativos aos instrumentos de recolha de dados usados no estudo, um com o questionário e outro com o teste.



## CAPÍTULO II

### REVISÃO DE LITERATURA

Neste capítulo apresenta-se uma síntese de estudos que fundamentam o nosso objeto de estudo, nomeadamente estudos relacionados com o significado pessoal e institucional dos objetos matemáticos e sobre as concepções e estratégias de avaliação da associação estatística e regressão linear. Faz-se também uma abordagem geral do ensino de estatística em Moçambique.

#### 2.1. Significado institucional e pessoal dos objetos matemáticos

Para Godino e Batanero (1994) o significado dos objetos matemáticos é concebido como um sistema de práticas (operatórias e discursivas) ligadas a um determinado campo de problemas, em que três tipos de elementos de significado necessitam ser considerados: os elementos *extensivos*, os elementos *intensivos* e os representacionais. Os elementos *extensivos* são as diferentes situações prototípicas que induzem a atividade matemática, como por exemplo as situações-problema. Os elementos *intensivos* do significado são as diferentes propriedades e relações com outras entidades, ou seja, ideias matemáticas, abstrações, generalizações (conceitos e proposições). Por último, identificam-se os elementos *representacionais ou ostensivos* do significado, que são as representações materiais utilizadas na atividade matemática, como por exemplo: os termos e expressões, símbolos, tabelas e gráficos.

Estes autores consideram práticas matemáticas toda a atividade ou processos seguidos pelos alunos com vista à resolução de um determinado problema ou tarefa. Nestas práticas intervêm objetos ostensivos (símbolos, gráficos, etc.) e não ostensivos, ou seja, aqueles que evocamos ao fazer matemática, que são representados de forma textual ou oral.

Godino e Batanero (1994) consideram que para os distintos objetos matemáticos, como sejam os conceitos, proposições, termos, etc., tem que se ter em conta duas facetas: a institucional (ou epistemológica) e a pessoal (ou cognitiva). Estes autores definem o significado institucional de um conceito como sendo o sistema de práticas que é compartilhado dentro de uma instituição e o significado pessoal como um sistema de práticas que uma pessoa utiliza para resolver o campo de problemas de que emerge o objeto num determinado momento.

Assim, o significado pessoal é relativo a uma pessoa particular e o significado institucional é relativo a uma instituição, que, por exemplo, pode ser formada pelos matemáticos, pelos professores de matemática ou pelos manuais escolares de matemática.

Este modelo teórico parte da noção de situação-problema como noção primitiva, considerando essa situação-problema como qualquer circunstância em que se devem realizar atividades de matemática, as quais incluem, segundo Freudenthal (1991): construir ou procurar soluções que não são imediatamente acessíveis; justificar as soluções propostas (validar ou argumentar) e generalizar a solução a outros contextos, situações-problema e procedimentos.

Em consequência disso, quando o sujeito enfrenta um problema e se propõe a resolvê-lo, comunica a solução a outras pessoas, valida e generaliza a solução a outros contextos e problemas está realizando distintos tipos de prática. Particularmente, na presente investigação centramo-nos nos campos de problemas de associação estatística, como por exemplo determinar a intensidade da associação entre duas variáveis, representadas gráfica ou numericamente, determinar o valor do coeficiente de correlação e construir um diagrama de dispersão.

De acordo com Godino e Batanero (1994), ao estudar os processos cognitivos e didáticos, a partir das noções de situações-problema e das práticas descritas anteriormente, desenvolvem-se as noções derivadas de práticas significativas e o significado de objeto, para os quais se postulam duas dimensões interdependentes, uma pessoal e outra institucional.

De acordo com os autores, uma prática considera-se significativa para uma pessoa ou para uma instituição, se ela cumpre uma função para resolver o problema, comunicar, validar ou estender a sua solução a outras situações. No caso de associação estatística, haveria práticas significativas vinculadas à sua análise, tais como no cálculo do coeficiente de correlação para determinar o sinal e a intensidade da associação, mas também ao determinar a covariância ou ao construir o diagrama de dispersão.

Como se pode ver, para um mesmo problema é possível apresentar-se uma diversidade de práticas significativas, o que nos leva a diferenciar as práticas institucionais das práticas pessoais. Segundo Batanero (2001), as práticas institucionais são aquelas que são aceites dentro de uma instituição, ao passo que práticas pessoais são aquelas que são realizadas por uma pessoa, como por exemplo um só aluno, e têm um carácter subjetivo.

Em consequência desta prática, o objeto matemático imana como um ente abstrato que emerge progressivamente do sistema de práticas socialmente compartilhadas, ligadas à

resolução de um certo campo de problemas. Este processo emergente é progressivo ao longo do tempo, sofrendo transformações progressivas conforme se vai aplicando ao campo de problemas associados, ou seja, os objetos matemáticos vão progressivamente emergindo no sujeito como consequência da experiência e da aprendizagem, e são estes objetos os constituintes do conhecimento subjetivo (ou pessoal).

Paralelamente à noção de significado que temos estado a descrever, deduz-se uma teoria da compreensão (Godino, 1996). A compreensão pessoal dum conceito é, neste modelo, a “captação” do significado desse conceito. Este autor afirma que desde que o significado de um objeto não se conceba como uma entidade absoluta e unitária mas composta e relativa aos contextos institucionais, a compreensão de um conceito pelo sujeito num dado momento e circunstâncias implicará a aquisição dos distintos elementos que compõem os significados institucionais correspondentes.

Por exemplo, numa turma o

professor segue determinadas diretrizes curriculares, livros de texto e materiais didáticos, que estabelecem um significado particular restringido para uma determinada temática. Portanto, ao realizar a avaliação, o professor considera que o aluno “conhece” ou “compreende” determinado objeto matemático, no nosso caso concreto a associação estatística, se houver concordância entre o significado institucional e pessoal construído pelo sujeito. Se não há concordância entre estes dois significados, dizemos que não há compreensão por parte do aluno e que o tema é difícil para o aluno. Assim, a noção de compreensão, na perspectiva destes autores, estabelece-se à custa dos significados pessoal e institucional.

Nas instituições escolares, os processos educativos têm por finalidade última ajudar os estudantes a adquirirem o conhecimento escolar, tornando-o tanto quanto possível coincidente com o seu conhecimento pessoal.

Godino (1996) distingue compreensão pessoal de significado pessoal pela possibilidade de o conhecimento ser observável ou não. Para ele, a compreensão pessoal corresponderia à parte não observável do conhecimento, isto é, trata-se de um constructo psicológico, ao passo que o significado pessoal corresponderia a um conjunto de práticas observáveis.

Neste modelo, a avaliação centra-se no estudo da correspondência entre os significados pessoais e institucionais. É com este sentido que na presente investigação avaliamos o significado pessoal construído pelos estudantes antes e depois do ensino do temática da

associação estatística e o comparamos com o significado de referência, ou seja, os significados institucionais.

Este autor adverte que a avaliação da compreensão de um sujeito deve ser relativa aos contextos institucionais em que este sujeito participa. Assim, numa certa instituição (escola ou de outro tipo) diz-se-á que um sujeito “compreende” o significado de um objeto se o sujeito é capaz de realizar as distintas práticas que configuram o significado institucional desse objeto. Além disso, que seja capaz de fundamentar e refletir sobre o processo seguido.

Quando se quer caracterizar o significado de um objeto numa instituição ou para uma pessoa, as práticas observáveis são os indicadores empíricos que permitem essa caracterização. Em consequência, a compreensão pessoal de um indivíduo sobre um certo objeto matemático deverá ser inferida mediante a análise das práticas por ele realizadas na resolução das tarefas problemáticas. Assim, no nosso estudo, com o intuito de perceber a compreensão pessoal dos estudantes sobre a associação estatística, privilegiar-se-á a análise das suas respostas às tarefas propostas.

## **2.2. Ensino da estatística em Moçambique**

Atualmente, a grande quantidade de informação com que lidamos (nos jornais, nas revistas, na rádio e na televisão, assim como em livros e relatórios) e a necessidade do seu tratamento e análise para a tomada de decisões esclarecidas realça a importância que a estatística vem tendo nos nossos dias. De facto, a estatística é uma área da matemática em que tem havido uma grande evolução, não só no próprio tema como nas suas aplicações na sociedade.

Perante a importância crescente do pensamento estatístico e probabilístico na sociedade, coloca-se a qualquer cidadão o desafio de gerir e utilizar a informação que lhe chega para tomar as suas decisões conscientemente. Portanto, torna-se imprescindível que o cidadão adquira competências para atingir esse fim. Nessa perspetiva, reconhecendo a importância do estudo da estatística na melhoria das capacidades de intervenção do aluno como cidadão, dotando-o de ferramentas apropriadas para avaliar afirmações de carácter estatístico, muitos países incluem o estudo de conceitos e procedimentos da estatística básica nos seus currículos de educação elementar (Batanero et al., 1994). Igualmente, face às exigências de uma sociedade em crescente mudança, o National Council of Teachers of Mathematics (1994) preconiza a

necessidade de formação em estatística em todos os níveis de ensino para dar respostas a essas exigências.

Já no século XX, mais precisamente nos anos cinquenta e sessenta, o movimento que começa então a registar-se em alguns países, com objetivo de modificar os programas e métodos de ensino da matemática nas escolas secundárias, chega também a Portugal (Fernandes, Sousa & Ribeiro, 2004). É precisamente em resultado deste movimento que o tema de estatística e probabilidades é introduzido pela primeira vez nos programas de matemática do ensino secundário (antigos liceus). De acordo com Sebastião e Silva (citado em Branco, 2000, p.13), que foi responsável pela reforma do ensino da matemática no ensino liceal,

a modificação dos programas com vista a adaptá-los às exigências da revolução científica e tecnológica levam à introdução, pela primeira vez, nos liceus portugueses de vários temas entre os quais elementos de cálculo de probabilidades e estatística.

No entanto, este ramo da matemática está longe de ter o lugar que merece nos currículos atuais do ensino básico e secundário. Além disso, continua a constituir um tema que é frequentemente relegado para “segundo plano” (Ponte & Fonseca, 2000), ou mesmo esquecido quando chegam os momentos de estabelecer prioridades. Por outro lado, Fernandes et al. (2004), baseando-se em Batanero (2000), afirmam que:

apesar dos novos currículos da educação básica e secundária incluírem de forma generalizada recomendações sobre o ensino da estatística, são poucos os professores que ensinam este tema e noutros casos tratam-no muito brevemente ou de forma excessivamente formalizada. (p. 169)

Numa visão geral dos programas de ensino de matemática do ensino geral em Moçambique (da 1ª à 12ª classe) constata-se que este está estruturado por ciclos conforme os níveis de ensino, como se mostra no Quadro 1.

Quadro 1 – Organização do ensino básico e secundário moçambicano por ciclos e classes

Ensino Básico		Ensino Secundário	
1º ciclo	1ª–2ª classe	1º ciclo	8ª–10ª classe
2º ciclo	3ª–5ª classe	2º ciclo	11ª–12ª classe
3º ciclo	6ª–7ª classe		

Fonte: Quadro feito pelo autor adaptada dos Programas do Ensino Básico (2008), e Programas de Matemática do ensino Secundário (2010)

O ensino de estatística em Moçambique inicia-se na 3ª classe, isto é, na 1ª classe do 2º ciclo do ensino básico, tendo por objetivo geral desenvolver nos alunos a capacidade de ler e interpretar tabelas, esquemas e gráficos. Nesta fase, são tratados especificamente aspetos sobre a construção e interpretação de tabelas e gráficos simples (de barras e circular), com base em dados reais que representam, por exemplo, o decurso de um determinado facto (social ou económico).

No que se refere ao 3º ciclo do ensino básico (6ª–7ª classe), o objetivo geral do ensino da estatística neste ciclo é desenvolver as capacidades técnicas de recolha, organização e representação de dados estatísticos para casos simples. Assim, na 6ª classe faz-se apenas a consolidação dos aspetos de construção e interpretação de gráficos e tabelas iniciados no 2º ciclo, enquanto na 7ª classe, para além de se tratar conteúdos sobre interpretação de gráficos e tabelas, introduzem-se outros conteúdos, tais como a recolha, organização e registo de dados, e dá-se início ao ensino de algumas medidas de tendência central, como é o caso da média aritmética e da moda.

Como sugestão metodológica para esta unidade, o Programa de Ensino Básico (2008) sugere que o professor deve criar situações didáticas simples, que permitam a discussão e reflexão sobre os problemas e desafios apresentados aos alunos, de modo que eles mesmos possam construir o seu conhecimento. Recomenda-se ainda que os problemas a serem propostos devem estar, se possível, próximos da realidade do aluno, isto é, devem ter sentido para ele. Segundo Fernandes et al. (2011),

Os educadores acreditam que o uso de dados reais em tópicos de interesse dos alunos, o que não acontece apenas em Estatística, contribui para a motivação dos alunos em aprenderem Estatística e para gostarem de o fazer. (p. 588).

Outros conteúdos tratados referem-se à construção e interpretação de gráficos e ao cálculo e análise das medidas de tendência central (média, moda e mediana). Exige-se que o aluno nesta fase saiba interpretar informações estatísticas da vida representadas em gráficos ou tabelas e tire conclusões para compreender fenómenos da sociedade e da natureza. Para Fernandes et al. (2007) a abordagem da estatística a partir de situações da vida real do aluno e a ênfase na interpretação dos dados em detrimento da aplicação de fórmulas e procedimentos constituem duas orientações atuais para o ensino desta temática.

Outro aspeto que o programa de Matemática da 9ª e 10ª classes (MINED, 2010) sugerem é a elaboração de projetos de investigação com dados reais, envolvendo os próprios alunos de

forma individual ou coletiva, de forma a permitir-lhes apreciar a importância do trabalho, interessando-os pela estatística como meio de abordar problemas variados da vida real, que poderão estar próximos das experiências e necessidades dos alunos. O trabalho de projeto serve neste contexto como uma atividade complementar às outras, nomeadamente a análise de dados, coleção de dados e exercícios estatísticos do tipo tradicional, que permite aos alunos conectarem diferentes conhecimentos para solucionar um problema.

O trabalho de projeto constitui outra das orientações atuais para o ensino e a aprendizagem da estatística, o qual pode ser usada numa dupla vertente: como forma de síntese e do estabelecimento de conexões entre conhecimentos já adquiridos (como é enfatizado no programa) ou como forma de aprender novos conceitos (Fernandes et al., 2007). Por outro lado, este tipo de tarefa é particularmente adequado ao trabalho dos alunos em grupo, o qual, segundo Fernandes et al. (2011),

permite aos professores desenvolverem tarefas mais compreensivas, capacita os alunos a adquirirem um insight sobre as dinâmicas e os processos de grupo, possibilita aos alunos o desenvolvimento de competências interpessoais, permite expor os alunos aos pontos de vista dos outros membros do grupo, encoraja os alunos a prepararem-se para o ponto de vista real (que cada vez mais envolve trabalho em equipa) e promove a reflexão e a discussão como parte essencial do processo de se tornarem práticos competentes e reflexivos. (pp. 588-589)

O ensino de estatística volta a ser tratado no nível secundário, mais particularmente na 12<sup>a</sup> classe com a introdução ao estudo do cálculo combinatório e das probabilidades.

Como sabemos, as probabilidades fornecem conceitos e métodos para estudar situações de incerteza e para efetuar e interpretar previsões em situações incerteza. Portanto, um dos objetivos específicos do seu estudo é o de fornecer uma base concetual que capacite o aluno a interpretar, de forma crítica, toda a comunicação que utiliza a linguagem das probabilidades, bem como a linguagem estatística.

Para o ensino do cálculo combinatório, o programa da 12<sup>a</sup> classe recomenda que se comece por ensinar as técnicas de contagem como forma de se desenvolver mais as capacidades do raciocínio combinatório e as conexões matemáticas. Apela também que se explore experiências que permitam tirar partido de materiais lúdicos e de simulações com a calculadora, de modo a contribuir para esclarecer conceitos através da experimentação e para dinamizar discussões de tipo científico, bem como para incentivar o trabalho cooperativo.

Um facto constatado nos programas do ensino da estatística, desde o ensino básico ao secundário, é que a maior parte deste conteúdo é inserido no último capítulo do programa. Este

facto leva a que muitas das vezes os professores não ensinem este tema por falta de tempo e pelo facto de não estar muitas vezes presente nos exames.

Outro aspeto que achamos interessante é o da recomendação da utilização dos recursos tecnológicos, o que acontece apenas no último ano do ensino secundário e de forma não obrigatória. Julgamos que este seja um dos fatores que leva a que os alunos cheguem ao ensino superior com fraco ou nenhum domínio destas ferramentas, que são muitos úteis na aprendizagem de alguns conteúdos de estatística e de probabilidades. Segundo Jolliffe (2007), é a chamada revolução tecnológica que está na origem das maiores alterações no ensino da Estatística, que por sua vez influenciam o modo como é avaliada. Apesar das potencialidades dos recursos tecnológicos no ensino da estatística, ao permitir resolver problemas reais com dados reais, que antes se tornavam impraticáveis de abordar, nem sempre eles fazem parte da rotina das salas de aula. No estudo de Fernandes et al. (2007) as professoras atribuíram o não uso destes recursos ao facto de não as saberem usar, à sua ausência nas escolas e à falta de tempo.

O ensino da estatística e probabilidades no ensino superior, particularmente na Licenciatura em Ensino de Matemática da Universidade Pedagógica de Moçambique, começa a ser feito a partir do 2º ano, na disciplina de Estatística Descritiva. Os objetivos gerais desta disciplina são recolher, resumir e interpretar dados de diversas variáveis através de tabelas de distribuição de frequências e medidas estatísticas. Faz-se aqui uma breve revisão dos conceitos básicos aprendidos nas classes anteriores e incluem-se alguns elementos novos, tais como o agrupamento de dados em classes, bem como as medidas de assimetria e de achatamento. Enfatiza-se também a utilização de meios informáticos, tais como pacotes informáticos (SPSS e Excel), na construção e manuseamento de bases de dados. Os trabalhos práticos constituem outras atividades que são desenvolvidas neste ano, cujo foco incide basicamente na realização de pesquisas de opinião, as quais envolvem a definição da amostra, a elaboração de instrumentos de recolha de dados, o processamento e análise de dados e a elaboração de relatórios dos resultados obtidos.

A correlação e regressão simples é um conteúdo que é abordado pela primeira vez no currículo da Licenciatura de Ensino de Matemática da Universidade Pedagógica. Quanto à correlação, são abordados aspetos que vão desde o conceito de correlação, tipos de correlação, diagrama de dispersão, medidas de correlação, como é o caso do coeficiente de correlação de Pearson e da covariância, e fala-se também do coeficiente de determinação e das propriedades

do coeficiente de correlação. No que diz respeito a regressão, trata-se basicamente do ajustamento linear, do conceito de regressão, das retas de regressão de  $y$  sobre  $x$  e de  $x$  sobre  $y$ , assim como do coeficiente de regressão. É fundamentalmente neste conteúdo que se enquadra o nosso trabalho, em que procuramos conhecer as concepções e estratégias usadas pelos estudantes na avaliação da associação estatística, antes do ensino e depois do ensino desta temática.

Ainda no 2º ano, inclui-se a disciplina de Teoria de Probabilidades, cujos objetivos gerais são: i) aplicar o conceito de probabilidade e suas propriedades na resolução de problemas; e ii) desenvolver no estudante a capacidade de resolver problemas utilizando modelos de distribuição de probabilidade. Os conteúdos desenvolvidos vão desde a introdução ao conceito de probabilidade, métodos de contagem, soma e produto de probabilidades, acontecimentos independentes, variáveis aleatórias e, por fim, as distribuições de probabilidades.

O estudo da estatística neste nível de ensino culmina com a disciplina de Inferência Estatística. Um dos objetivos dessa disciplina é aplicar os métodos de inferência estatística na sistematização e análise de dados, bem como desenvolver a capacidade dos estudantes de analisar bases de dados e elaborar relatórios de pesquisa aplicando diferentes tipos de testes. Para o efeito, desenvolvem-se e aplicam-se temas sobre amostragem, métodos de estimação (pontual e intervalar), testes de hipóteses, testes não paramétricos e análise de variância (ANOVA). Para cada tema deve apresentar-se o contexto de aplicação prática, com exercícios de aplicação usando software informático (SPSS, Stata, Excel, ...). Recomenda-se também a inclusão nos exercícios de situações concretas em que os estudantes identificam os vários procedimentos necessários para a sua resolução, como sejam a definição da amostra, os métodos de seleção da amostra e as técnicas estatísticas necessárias à análise de dados com a devida justificação.

Fazendo uma avaliação dos conteúdos de estatística tratados no ensino geral e no ensino superior, mais especificamente na Licenciatura em Ensino de Matemática, constatámos que eles estão de acordo com as orientações para o ensino de estatística para avaliar o raciocínio estatístico (Garfield, 1998):

*Raciocinar sobre os dados:* reconhecer ou categorizar os dados como qualitativos ou quantitativos, discretos ou contínuos e saber que o tipo de dados conduz a um tipo particular de tabela, gráfico ou medida estatística;

*Raciocinar sobre representações dos dados:* compreender como se lê e interpreta um gráfico, modificar um gráfico de maneira a que represente melhor uma série de dados;

*Raciocinar sobre medidas estatísticas:* compreender as medidas de tendência central e de dispersão, localizá-las na distribuição dos dados e saber quais devem ser aplicadas em circunstâncias diferentes;

*Raciocinar sobre amostras:* saber que as amostras estão relacionadas com a população, que a amostra deve ser representativa da população;

*Raciocinar sobre a associação:* saber julgar e interpretar o relacionamento de duas variáveis e a sua correlação.

Sobretudo nos países mais desenvolvidos, nos últimos tempos, tem-se assistido a um aprofundamento do ensino da estatística no ensino não superior, que se traduz na introdução do tema de estatística logo nos primeiros anos da escolaridade básica, na utilização de recursos tecnológicos, na realização de trabalhos de projeto e no estudo de conteúdos que tradicionalmente eram abordados apenas muito mais tarde. No último caso, encontram-se conteúdos sobre variação estatística no ensino básico e sobre associação e inferência estatística no ensino secundário (Fernandes, 2009). Assim, face aos atuais currículos moçambicanos, no que se refere à estatística, estas alterações curriculares recentes podem ser vistas como um desafio para o aprofundamento do seu ensino num futuro próximo.

### **2.3. Crenças e concepções: o caso da associação estatística**

Nesta secção começamos por analisar a problemática das crenças e concepções em geral, para de seguida abordarmos a questão mais específica das crenças e concepções no âmbito da associação estatística.

#### **2.3.1. Crenças e concepções**

A noção de concepção é na verdade difícil de definir, pois ela é estudada em diferentes áreas e como tal resulta numa variedade de significados, donde a comunidade de investigadores de Educação Matemática não tem sido capaz de adotar uma definição específica deste termo. (Pajares, 1992). Guimarães (2010) associa a noção de concepção “a um sentido de construção ou criação de algo, num ato onde ocorrem elementos interiores (da pessoa) e elementos exteriores (da coisa)” (p. 84). Ainda sobre a noção de concepção, Guimarães, citando Lallande (1976), afirma que concepção no sentido amplo pode designar “todo o ato do pensamento que se aplica a um objeto” (2010, p. 84). Por sua vez, Dewey (1991, citado em Guimarães, 2010) “associa fortemente a ideia de concepção, que também designa por noção, com a ideia de

significado, considerando que qualquer ‘significado padrão’ ou ‘qualquer significado suficientemente individualizado’ (...) é uma concepção” (p. 85).

Portanto, Dewey salienta que as concepções não são apenas meros significados comuns de objetos ou situações que se vão constituindo no confronto da pessoa com esses objetos ou situações. Ele defende que as concepções se constroem através duma atitude ativa do sujeito, ou seja, a partir da evocação que o sujeito faz de representações que repousam em suas experiências anteriores, que o levam a antecipar determinadas interpretações do objeto ou situação com que se depara. Assim, este autor considera as concepções como algo dinâmico e que vai ganhando corpo e consistência à medida que as representações se vão confrontando com a experiência, e assim validadas ou rejeitadas através de processos de pressuposição e de experimentação.

O interesse do estudo das crenças surge inicialmente em 1920 no âmbito da psicologia, na tentativa de explicar a natureza das crenças e a sua influência nas ações das pessoas. Nas décadas seguintes, sobretudo com a emergência do behaviorismo em 1930, este tópico perdeu interesse na literatura de psicologia. Em 1970, com o surgimento da ciência da cognição, abre-se um espaço para o estudo do sistema de crenças na relação com outros aspetos da cognição humana (Thompson, 1992). Em 1980 testemunha-se um ressurgimento do interesse do estudo das crenças e sistemas de crenças nas várias áreas, tais como ciência política, educação, antropologia.

Segundo Guimarães (2010), em estudos da literatura anglo-saxónica, pioneiros na área do pensamento e conhecimento do professor de Matemática, os conceitos de concepção e crença aparecem frequentemente associados e algumas vezes equivalentes. Tal é o caso de estudos de Thompson (1982, 1992), onde se consideram sistemas de concepções como o conjunto das crenças que os professores possuem sobre a Matemática e o seu ensino e aprendizagem. Outros investigadores, tais como Clark e Peterson (1986, citado em Guimarães, 2010), dão maior relevância ao conceito de crenças do que ao conceito de concepção.

O estudo das concepções dos professores insere-se, no que se refere à investigação educacional, numa área mais ampla, habitualmente reconhecida como estudo do pensamento ou do conhecimento do professor. No que diz respeito ao ensino da Matemática, trata-se de uma área de investigação em desenvolvimento sensivelmente desde o início dos anos oitenta, do século passado, e que desde então foi merecendo uma atenção crescente (Guimarães, 2010). Esta área de investigação, particularmente em Portugal, iniciou-se na segunda metade dos anos

oienta por influência dos trabalhos de Thompson (1982). Os trabalhos de Abrantes (1986) e de Guimarães (1988) terão sido os primeiros nesta área de investigação, cujo foco incidirá essencialmente no conceito de concepções gerais dos professores de Matemática e do seu ensino. Mais tarde, com os trabalhos de Canavarro (1993) e Rodrigues (1993), a investigação prosseguiu na mesma linha de pesquisa, incidindo quer sobre concepções gerais relativas à Matemática e ao seu ensino, quer sobre concepções relativas a aspetos mais específicos, tais como a resolução de problemas (Oliveira, 1993), utilização de tecnologia (Azevedo, 1993), comunicação na sala de aula (Meneses, 1995), entre outros.

Tendo em conta que o professor é um elemento fulcral no processo de ensino-aprendizagem e que as suas concepções sobre a Matemática e o seu ensino são responsáveis, em parte, pela construção dos modelos de ensino que adota nas suas práticas (Thompson, 1992), é pertinente identificar e analisar as concepções para se poder compreender a ação pedagógica do professor. Guimarães (2010) referindo-se ao estudo das concepções dos professores sobre a matemática e o seu ensino, salienta que:

Existe, na verdade, um consenso cada vez crescente sobre a importância em ter acesso à 'vida mental' dos professores, em conhecer e compreender os vários aspetos do seu pensamento e conhecimento, bem como as relações desses aspetos com a sua atuação ou comportamento (p. 82).

Para este autor, subjacente a esse interesse está a convicção de que aquilo que o professor pensa influencia de forma significativa aquilo que o professor faz na sala de aula. Thompson acrescenta que a abordagem que os professores fazem ao ensino depende fundamentalmente dos seus sistemas de crenças, em particular das suas concepções sobre a natureza e significado da matemática e dos seus modelos mentais relativos ao ensino e à aprendizagem.

Guimarães (2010) entende que os sistemas conceituais, como conjuntos de concepções de uma pessoa, se estabelecem com diferentes graus de convicção e possuem diferentes níveis de conexão entre os seus elementos, o que faz com que existam concepções mais facilmente modificáveis do que outras. Segundo este autor, quanto mais cedo se formam, maior estabilidade adquirem e, conseqüentemente, mais difícil se torna modificá-las. Apesar disso, tem-se reconhecido que esses sistemas não são estáticos e que com a experiência podem sofrer modificações mais ou menos profundas.

Thompson (1992), recorrendo a outras fontes, caracteriza os sistemas de crenças referindo-se a três dimensões. Na primeira dimensão descreve que as crenças se caracterizam

pelo modo como estão relacionadas umas com as outras, em termos de uma “hierarquia de precedência lógica”, destacando as *crenças primárias* e as *crenças derivadas* dos sistemas de crenças. Por exemplo, se um professor acredita ser importante apresentar uma matemática “clara” para o aluno, enquanto crença primária, pode em consequência desta crença gerar outras crenças, crenças derivadas, como sejam: acreditar que é importante preparar as aulas por forma a ter uma apresentação coerente e sequencial, mas também acreditar que é importante estar preparado para responder às questões levantadas pelos estudantes.

A segunda dimensão estabelece uma tipologia de crenças com base no grau de convicção com que as mantemos. Nesta dimensão é possível encontrar crenças *centrais* e *crenças periféricas*, sendo que as primeiras são psicologicamente mais firmemente estabelecidas do que as segundas e, por isso, mais dificilmente modificáveis. Rokeach (1976, citado em Guimarães, 2010), referindo-se à organização das crenças, afirma que “nem todas as crenças têm a mesma importância para a pessoa” e que “quanto mais central for a crença mais resistirá à mudança” (p. 93).

Ainda de acordo com este autor, a importância das crenças, que ele chamou de *centralidade*, é definida em termos das relações que a crença tem com outras crenças do sistema. Para ele, quanto mais relações tiver uma crença, mais central é, o que implica que se uma crença tem relações com outras crenças, qualquer mudança naquela vai ter consequências em todas estas com que está relacionada, o que faz com se torne mais difícil modificá-la. Ou seja, quanto maior for a centralidade da crença, maior é a sua estabilidade.

Ainda sobre as modificações das crenças, Pajares (1992) considera que as crenças mais recentemente adquiridas são as que mais facilmente podem ser modificadas, sendo as mais antigas aquelas que mais dificilmente se podem alterar. Este autor afirma ainda que,

Com o tempo, as crenças se fortalecem, estabelecem mais relações no sistema de crenças e tornam-se mais centrais, o que conduz (...) a que as pessoas possam muitas vezes não modificar as suas crenças, mesmo quando confrontadas com algo que as contraria ou põe em causa. (p. 317)

Pajares (1992) afirma que uma vez estabelecidas, as crenças tendem de certa forma a “auto perpetuar-se”, não significando isso que elas não sofram mudanças nas pessoas ao longo das suas vidas.

Thompson, no entanto, distancia-se da ideia de que os sistemas de crenças são “entidades estáticas” e de que a relação entre as crenças e a prática é de simples causalidade. Segundo esta autora, a investigação sobre a relação entre as crenças e a prática sugere que os

sistemas de crenças são “dinâmicos, estruturas mentais permeáveis, suscetíveis de mudança com a experiência”, e que a sua relação com a prática é uma “relação dialética e não uma simples relação de causa e efeito” (p. 140).

Importa referir que os estudos sobre a influência dos sistemas conceituais não se restringem apenas a professores e ao ensino, abrangendo também aspetos ligados à aprendizagem do estudante, como seja analisar o papel desempenhado pelas crenças ou conceções da matemática dos estudantes na interpretação e resolução de problemas ou outras tarefas. É assim que no presente estudo tentamos identificar crenças ou conceções dos estudantes sobre a associação e independência estatística de duas variáveis nos diferentes contextos em que são avaliadas.

### **2.3.2. Crenças e conceções sobre associação estatística**

O conceito de associação tem grande importância na educação matemática porque amplia o conceito de dependência funcional e é fundamental para muitos métodos estatísticos, permitindo modelar numerosos fenómenos nas diversas ciências (Batanero, 2001), mantendo este tema também conexões importantes com o pensamento funcional e o raciocínio proporcional.

A finalidade de muitas dessas aplicações é encontrar explicações causais que nos permitam compreender o meio que nos rodeia. Contudo, a associação estatística não estabelece necessariamente uma relação causal. Por vezes, existe uma “correlação enganadora” entre as variáveis devido à influência de outros fatores concorrentes, sem que haja um vínculo causal.

Para além destas dificuldades epistemológicas, as investigações psicológicas têm mostrado que a habilidade para emitir juízos de associação não se desenvolve de forma intuitiva. Muitas vezes as pessoas adultas baseiam os seus juízos nas crenças prévias sobre o tipo de associação apresentada nos dados (Batanero, 2001).

Estudos sobre problemas de associação estatística, tais como o de associação em tabelas de contingência, associação entre duas variáveis numéricas a partir de dados representados num diagrama de dispersão e a correlação numérica tendo em vista a comparação de uma variável numérica em duas ou mais amostras (que podem ser independentes ou relacionadas), têm sido levados a cabo por investigadores e educadores matemáticos, particularmente da Universidade de Granada. Destacaremos aqui os estudos de Batanero et al. (1994), Batanero,

Estepa, Godino e Green (1996), Estepa e Batanero (1996), Batanero, Godino & Estepa (1998) e Estepa et al. (1999).

Nestes estudos visava-se, fundamentalmente, determinar os significados pessoais que os estudantes atribuíam aos conceitos de associação nos três tipos de problemas anteriormente referidos, antes de terem estudado o tema na escola. No estudo de Batanero et al. (1998) estiveram envolvidos 213 alunos do final do ensino secundário (17-18 anos) e nele foi aplicado um conjunto de tarefas sobre os diferentes tipos de problema de associação estatística. Um dos problemas era relativo à associação estatística em tabelas de contingência de  $2 \times 2$ , tendo sido apresentado sumariamente o número de instâncias de presença e ausência da variável  $X$  (por exemplo, uma doença particular), supostamente associada com a presença e ausência da variável  $Y$  (por exemplo, um sintoma particular). A questão principal que se propunha analisar era o tipo de relação existente entre as duas variáveis (associação positiva, negativa ou independência) e a relação entre o contexto do problema e as crenças prévias.

Um segundo tipo de problema de associação proposto no estudo de Batanero et al. (1998) foi o da análise da relação entre duas variáveis numéricas a partir de dados representados num diagrama de dispersão. Nesse tipo de problema analisam-se as estratégias intuitivas dos alunos para resolver estes problemas sobre associação e determinam-se as propriedades matemáticas da associação em questão.

O terceiro tipo de problema diz respeito à correlação numérica, onde se faz a avaliação da intensidade da correlação entre duas variáveis quantitativas, para o qual se pode calcular a covariância, o coeficiente de correlação ou ainda estudar a qualidade de ajuste de uma linha de regressão ao diagrama de dispersão. Ainda neste caso, outro problema consiste em averiguar se uma variável numérica tem a mesma distribuição em duas amostras diferentes, o qual pode ser resolvido comparando a diferença entre as médias ou medianas ou as representações gráficas ou tabulares de ambas distribuições.

Cobo (1998) afirma que as estratégias utilizadas pelos estudantes nos problemas sobre juízos de associação não são o único fator que influencia a solução que realizam, afirmando que as soluções dependem das crenças e concepções (umas corretas e outras incorretas ou parcialmente corretas) que os estudantes têm sobre associação estatística. Assim, nos trabalhos de Estepa e Batanero (1996, 1999) ao comparar as estratégias e os juízos de associação dos alunos (em tabela de contingência, na comparação de duas amostras e no gráfico de dispersão),

através da análise dos seus argumentos, estes autores identificaram e descreveram algumas concepções incorretas sobre a associação estatística, que se referem a seguir.

- (I) *Concepção determinista de associação*: alguns estudantes não admitem a existência de mais de um valor da variável dependente para cada valor da variável independente. Quando isto não ocorre, consideram que não há dependência. Por outras palavras, para eles a relação entre as variáveis deve ser uma função de um ponto de vista matemático, ou seja, uma relação funcional.
- (II) *Concepção unidirecional de associação*: alguns estudantes concebem a associação apenas quando ela é positiva (associação direta), considerando a associação inversa como independência.
- (III) *Concepção local de associação*: os estudantes utilizam apenas parte dos dados fornecidos no problema para emitir juízos de associação. Se essa parte dos dados confirma um tipo de associação, então eles adotam esse tipo de associação na sua resposta.
- (IV) *Concepção causal de associação*: alguns estudantes consideram a existência de associação entre variáveis quando se pode atribuir uma relação de causa e efeito entre elas.

Para além destas concepções, Batanero et al. (1998), Batanero et al. (1996) e Cobo (1998), entre outros autores, identificaram outros tipos de concepções que influenciam igualmente o tipo de respostas dos estudantes, de entre os quais destacaremos as seguintes: influências das teorias prévias, a causalidade e o contexto e representação da informação. Descreve-se a seguir cada uma dessas concepções.

### **Influência das teorias prévias**

Uma questão muito importante nas investigações da associação estatística consiste em verificar se os sujeitos baseiam as suas avaliações nos dados do problema ou se se orientam pelas suas crenças acerca do tipo de relação que deve existir entre as variáveis. Segundo Cobo (2008, p. 65), “as experiências e ambientes onde o sujeito se desenvolve contribuem para a formação de uma série de teorias que eles utilizam para interpretar os factos do quotidiano”. Nestas situações, mais do que os dados do problema, são as teorias prévias que determinam as avaliações de associação estatística dos estudantes.

Sobre a influência das crenças na *performance* de resolução de problemas, Estepa e Batanero, nos seus variados estudos sobre associação estatística, mostraram que os indivíduos,

com particular destaque para os adultos, tomavam decisões acerca da associação estatística com base nas suas crenças ou ideias preconcebidas, sem terem em consideração os dados empíricos, ou seja, as crenças manifestavam uma resistência em detrimento da evidência empírica sobre a variável ou variáveis em estudo.

Estepa et al. (1999) referem que têm sido realizados por alguns autores estudos sobre o efeito que as teorias acerca do contexto de um problema têm no julgamento de associação. Estes estudos concluíram que em presença de teorias em concordância com os dados, os sujeitos reforçam uma boa aproximação da correlação entre as variáveis, mesmo quando ela é baixa. Ao contrário, quando estas teorias não existem, a estimação da associação é mais variável e afasta-se mais da realidade objetiva. Por outro lado, estes autores afirmam que

para perceber o grau de covariação entre dois acontecimentos são relevantes duas fontes de informação: a informação sobre a situação de contingência objetiva entre os acontecimentos proporcionada pelo meio e as teorias e crenças prévias do sujeito em relação à covariação dos acontecimentos em questão (p.18).

Assim, se a percepção da covariação coincide com as teorias prévias e com a informação da situação, aumenta a confiança de covariação entre as duas variáveis. Se pelo contrário, estiver em desacordo, gera-se um conflito cognitivo. Nestas situações, o que determina a natureza e a exatidão da percepção de covariância são as forças das duas fontes de informação em jogo, prevalecendo a mais forte.

Um fenómeno relacionado com estas teorias prévias é o de “correlação ilusória”, descrito por Chapman e Chapman (1969). Este fenómeno sugere que as pessoas partilham expectativas e crenças acerca da relação entre variáveis, as quais causam a impressão de contingências empíricas. Este fenómeno é conhecido por correlação ilusória porque as pessoas mantêm as suas crenças apesar da evidência de independência das variáveis. Estas teorias informais têm origem na experiência e no contexto do sujeito e são usadas na interpretação de dados e factos que o rodeiam.

De entre as heurísticas de julgamento probabilístico de Tversky e Kahneman (1982), o fenómeno da correlação ilusória também pode ser visto como manifestação da heurística da acessibilidade. A adesão a esta heurística significa que as pessoas avaliam a frequência de uma classe ou a probabilidade de um acontecimento em função da facilidade com que exemplificações ou ocorrências podem ser construídas, evocadas ou associadas. Especialmente da facilidade de evocação resulta que a avaliação errada de probabilidade é influenciada pela experiência do indivíduo.

## **Associação e causalidade**

Frequentemente, as pessoas tendem a procurar explicações do tipo causal para justificar a correlação entre duas variáveis. Ora, como é sabido, nem sempre a associação e a causalidade coincidem, sendo que duas variáveis podem estar fortemente associadas, mas não haver uma relação de causa e efeito entre elas (Poza, 1987). No mesmo sentido, Batanero et al. (1996) revelaram que uma percentagem muito grande de alunos tem muitas dificuldades em distinguir associação e causalidade. Afirmam que muitas vezes quando os alunos admitem existir associação entre as variáveis, tendem a pensar que uma delas provoca ou é causa da variação da outra.

Para Tversky e Kahneman (1982), a diferença significativa entre a correlação e causalidade reside na relação de simetria que existe entre as variáveis. Assim, na correlação a relação entre as variáveis é simétrica, porque mesmo permutando as variáveis o valor do coeficiente de correlação mantém-se; ao passo que na causalidade esta relação é assimétrica, pois nas variáveis causa e efeito a causa precede o efeito. Os autores referem ainda que na hora de se fazer a inferência, os alunos optam preferencialmente no sentido das causas para os efeitos do que dos efeitos para as causas. Desta maneira, torna-se mais fácil aceitar, por exemplo, a previsão da estatura de um filho a partir da do pai do que o contrário. O mesmo pode acontecer no caso em que ambas as variáveis não são assumidas como causa uma da outra, como por exemplo ao relacionarmos o peso e a altura, o que acontece “sempre que a primeira variável parece explicar melhor a segunda do que a explicação que a segunda oferece da primeira” (Tversky & Kahneman, 1982, p. 120). Por conseguinte, uma concepção errada que pode ocorrer é que, embora o peso e a altura de uma pessoa não sejam causa uma da outra, os sujeitos acham mais fiável explicar o peso de uma pessoa a partir da sua altura.

Especialmente no caso de probabilidades condicionadas, as pessoas avaliam a probabilidade de  $p(A|B)$  em função de possíveis relações entre os acontecimentos A e B. Mesmo quando tais relações são simplesmente sugeridas (não verificadas), as pessoas utilizam-nas para fazer os seus julgamentos (Fernandes, 1999).

Tversky e Kahneman (1982) identificaram quatro tipos de relações entre os acontecimentos A e B que as pessoas usam na avaliação de  $p(A|B)$ :

(1): se B é entendida como uma causa da ocorrência de A, referimo-nos a B como uma referência causal de A;

(2): se A é visto como uma possível causa de B, referimo-nos a B como referência diagnóstica de A;

(3): se B não é entendida como causa nem efeito de A, mas são ambos vistos como consequência de um outro fator, referimo-nos a B como indicador;

(4): quando acreditamos que não existe relação entre B e A através de qualquer ligação causal, direta ou indireta, referimo-nos a B como incidental.

Segundo Estepa e Batanero (1996), embora no estudo normativo da probabilidade condicional e da associação não seja necessário distinguir esses tipos de relações, elas podem ser importantes de um ponto de vista psicológico. No estudo realizado por esses autores, constatou-se que os estudantes não atribuíam um papel simétrico a ambas as variáveis, tal como na associação; em vez disso, utilizaram as suas expectativas em relação ao tipo de correlação que deveria haver entre as variáveis, manifestando uma concepção causal de associação.

### **Influência do contexto e a representação da informação**

Embora tenham menor importância que as teorias prévias, algumas investigações têm demonstrado que a representação da informação e os contextos utilizados nas tarefas propostas aos sujeitos sobre juízos de associação pode também influenciar nos resultados e conclusões obtidas (Cobo, 1998). Neste sentido, alguns autores estudaram a influência de três tipos de variáveis sobre a capacidade dos indivíduos de analisar a relação de associação: (i) a forma em que a informação era apresentada; (ii) a intensidade real da correlação apresentada na informação; e (iii) as expectativas dos sujeitos sobre a relação em questão. Os resultados indicam que as estimativas dos estudantes sobre a correlação eram significativamente influenciadas por estes três fatores. Eles constataram que os estudantes faziam estimativas muito mais elevadas da correlação quando eles acreditavam que as variáveis estavam efetivamente correlacionadas do que no caso contrário.

Por último, segundo Beyth-Marón (1982), a representação da informação por meio de variáveis simétricas ou assimétricas e o tipo de instrução que é dada também influenciam nos juízos sobre a correlação emitidas pelos estudantes. Uma variável simétrica é aquela em que os valores que podem tomar têm o mesmo peso para o indivíduo, como por exemplo a variável sexo (masculino, feminino). Uma variável assimétrica é aquela cujos valores não tem o mesmo peso para o indivíduo, como por exemplo padecer de uma enfermidade (padece da enfermidade, não padece da enfermidade). Nas variáveis assimétricas discrimina-se entre a ocorrência e não

ocorrência dos acontecimentos, entre os casos positivos e os negativos. Beyth-Maron (1982) constatou que os estudantes tinham uma melhor percepção da correlação quando se utilizam variáveis simétricas.

Como consequência dos resultados desses estudos, apresentam-se seguidamente alguns exemplos de estratégias adotadas pelos estudantes na avaliação da associação estatística nos diferentes tipos de tarefas.

## **2.4. Estratégias de avaliação da associação estatística**

Nesta secção, para além dos estudos de Inhelder e Piaget (1976) sobre associação, abordam-se outros estudos que se lhes seguiram sobre associação estatística desenvolvidos, sobretudo, por Batanero, Estepa e Godino nas diferentes vertentes.

### **2.4.1. Estratégias de avaliação da associação estatística em tabelas de contingência**

Antes de nos focarmos propriamente nos estudos de Batanero e seus colaboradores, que nos últimos tempos têm dado enormes contributos à didática do ensino da estatística no âmbito das estratégias intuitivas dos estudantes sobre a associação em tabelas de contingência, importa referir alguns estudos iniciais sobre esta temática, com particular realce aos de Inhelder e Piaget (1976) e de Pérez Echeverría (1990).

Inhelder e Piaget (1976) são considerados pioneiros nos estudos do desenvolvimento cognitivo do conceito de associação estatística, particularmente em crianças e adolescentes. Estes autores consideraram a compreensão da associação como a última etapa do desenvolvimento da ideia de probabilidade. De acordo com estes autores, os conceitos de associação e probabilidade estão relacionados, pois perceber a associação requerer *a priori* a compreensão da proporcionalidade, de probabilidades e combinações. Estes autores investigaram a compreensão da associação em crianças de 13 anos de idade, propondo problemas de associação entre duas variáveis qualitativas dicotómicas representadas em tabelas de contingência de  $2 \times 2$ .

Considerando as variáveis dicotómicas *cor do cabelo* (loiro e castanho) e *cor dos olhos* (azul e castanho), Inhelder e Piaget (1976) questionaram os sujeitos sobre a associação entre estas variáveis, partindo dos dados relativos a um conjunto de cartões (inferior a 40), representando rostos com olhos e cabelos coloridos, sob a forma de uma tabela de contingência

como a apresentada a seguir, previamente elaborada pelo sujeito ou fornecida pelo experimentador.

Na Tabela 1 representam-se as frequências absolutas, de uma população ou amostra, classificadas em relação a duas variáveis. Neste caso,  $a$ ,  $b$ ,  $c$  e  $d$  são as frequências absolutas.

Tabela 1 – Formato típico duma tabela de contingência  $2 \times 2$ .

	$B$	Não B	Total
$A$	$a$	$b$	$a + b$
Não A	$c$	$d$	$c + d$
Total	$a + c$	$b + d$	$a + b + c + d$

Nesse estudo, eles constataram que alguns adolescentes eram capazes de calcular probabilidades simples, considerando apenas os casos favoráveis positivos da associação (célula  $[a]$  na Tabela 1) em relação aos casos totais.

No estágio IIIA de desenvolvimento cognitivo do adolescente, fase que ocorre por volta dos 11-13 anos, os sujeitos analisavam apenas a relação existente entre os casos favoráveis positivos (célula  $[a]$  na Tabela 1) e os casos totais, sem terem em conta os casos favoráveis negativos (célula  $[d]$  na Tabela 1). Portanto eles utilizavam apenas uma das quatro células da tabela. Constataram ainda que estes adolescentes comparavam também as células duas a duas. Por exemplo, na Tabela 1, admitindo que também os casos  $[d]$  (negativo – negativo) eram favoráveis à existência de associação, os sujeitos não calculavam a relação entre os casos que confirmam a associação  $a + d$  e os restantes casos  $b + c$  (Inhelder & Piaget, 1976). De acordo com estes autores, este tipo de raciocínio produziu-se apenas a partir dos 15 anos, ou seja, no estágio IIIB. Neste estágio, a maior parte dos estudantes baseiam os seus juízos na célula  $[a]$  ou comparam  $[a]$  com  $[b]$ .

Estes autores definem a fórmula de associação como sendo a diferença entre os casos que confirmam a associação  $a + d$  e casos opostos  $b + c$ , comparados com todas as possibilidades  $a + b + c + d$ , isto é ,  $R = \frac{(a + d) - (b + c)}{a + b + c + d}$ . Se as probabilidades

$\frac{a + d}{(a + d) + (b + c)}$  e  $\frac{b + c}{(a + d) + (b + c)}$  são iguais, então  $R = 0$ , o que implica que a correlação

seria zero, sendo mais forte quanto maior for o valor da diferença  $(a + d) - (b + c)$ . Em

oposição à estratégia da diagonal, proposta por Piaget e Inhelder (1976), a qual só era válida se as frequências marginais forem iguais, Jenkins e Ward (1965, citado em Estepa & Batanero, 1996) propuseram para o caso geral como estratégia correta a comparação da diferença entre duas probabilidades condicionais,  $P(B|A)$  e  $P(B|\bar{A})$ , isto é,  $R = \frac{a}{a+b} - \frac{c}{c+d}$ , sendo que a associação é positiva se  $R > 0$ , negativa se  $R < 0$  e não existe associação se  $R = 0$ .

Os resultados de estudos de Inhelder e Piaget (1976) mostraram que as pessoas, geralmente, exibem um fraco desempenho em tarefas que envolvem tabelas de contingência. Em termos explicativos, são apontadas como razões das dificuldades o facto de os sujeitos ignorarem uma ou mais das quatro células da tabela. Por exemplo, uma falha comum envolve a consideração exclusiva do valor da célula correspondente às categorias positivas em ambas variáveis. Para estes autores, o desenvolvimento do conceito de associação em tabelas de contingência não ocorre de forma espontânea, sem que se realize uma instrução formal do tema.

Ainda no que respeita às estratégias de julgamento de associação em tabelas de contingência, Pérez Echeverría (1990) constatou que os sujeitos adotavam uma variedade de estratégias centradas no recurso a uma ou mais células da tabela de contingência, que se referem a seguir.

- A estratégia  $[a]$ , que consistia em utilizar isoladamente a célula  $[a]$  (positiva – positiva). Neste caso, a relação seria positiva, negativa ou nula consoante o valor de  $a$  fosse maior, menor ou igual que o das restantes células. Inhelder e Piaget (1976) veem esta estratégia como precursora das operações formais, embora alguns investigadores a tenham verificado em sujeitos adultos.
- A estratégia  $[a-b]$ , que consistia em comparar a diferença de frequências absolutas entre a célula  $a$  (positivo – positivo) e a célula  $b$  (positivo – negativo). Esta estratégia também foi descrita por Inhelder e Piaget (1976) e Smedlund (1963). Também na estratégia  $[a-c]$ , similar à anterior, os alunos usavam a diferença das frequências absolutas das células  $a$  e  $c$  (negativo – positivo).
- A estratégia  $[a, b, c]$ , em termos de juízos causais, determina a necessidade e suficiência das causas a partir das diferenças encontradas sem cálculos matemáticos, inferindo apenas de maneira subjetiva a partir das diferenças encontradas.

- A estratégia  $\frac{[a-b]}{[c-d]}$  foi utilizada para calcular a razão entre os casos que confirmam a possível relação e os que a rejeitam. Uma variante desta estratégia foi estudada por Allan e Jenkins (1983), e consistiu em determinar a diferença da soma das diagonais, isto é:  $R = [(a+d) - (b+c)]$ . Allan e Jenkins (1983) e Ward (1965) (citados em Batanero et al., 1994) observaram limitações desta estratégia na medida em que ela só é válida nos casos em que as frequências marginais são todas iguais, já que em caso de serem diferentes se pode detetar associação quando existe de facto independência.
- A última estratégia mencionada por Pérez Echeverría diz respeito a qualquer método multiplicativo entre as quatro células, o que significa que os sujeitos relacionavam as frequências absolutas das quatro células através da multiplicação, sendo uma dessas estratégias a determinação de  $R = \frac{a}{a+b} - \frac{c}{c+d}$ , já antes referida.

Pérez Echeverría (1990) agrupou as estratégias acima descritas nos seus trabalhos de psicologia em em cinco níveis de complexidade, a saber:

Nível 1: uso de apenas uma célula na tabela, usualmente a célula [a];

Nível 2: comparação da célula [a] com a célula [b] ou da célula [a] com a célula [c];

Nível 3: comparação da célula [a] com a célula [b] e da célula [a] com a célula [c];

Nível 4: uso de todas as quatro células da tabela empregando comparações aditivas;

Nível 5: uso de todas as quatro células da tabela empregando comparações multiplicativas.

Nas suas investigações sobre associação estatística, envolvendo estudantes pré-universitários, Batanero et al. (1996), classificaram as estratégias dos estudantes recorrendo à classificação dos níveis de dificuldades propostos por Echeverría (1990) para tabela de contingência  $2 \times 2$ , mas também ampliaram a sua análise para tabelas com mais de duas linhas ou colunas. Estes autores realizaram também uma classificação das estratégias sob ponto de vista matemático, tendo em conta os teoremas e conceitos implícitos nas mesmas, e agruparam-nas em estratégias corretas, parcialmente corretas e incorretas. Das estratégias corretas encontradas, destacamos as seguintes:

- a) *Comparação de todas as distribuições de frequências relativas condicionadas  $h(B_j | A_i)$  de todos os valores  $B_j$  para dois ou mais valores de  $A_i$ .* Os estudantes que usaram esta estratégia implicitamente basearam-se no facto de que a

dependência entre uma variável B e uma variável A implica que a frequência relativa condicional  $h(B_j | A_i)$  varia quando A varia.

- b) *Comparação de apenas uma frequência relativa condicional  $h(B_j | A_i)$  para um valor fixo de  $B_j$  para cada valor possível de  $A_i$ , com a frequência relativa marginal  $h_j$ .* Os estudantes que seguiram esta estratégia implicitamente usaram a propriedade da invariância da distribuição de B quando condicionada pelos valores de A.
- c) *Comparação de casos a favor e contra B para cada dos valores de A (odds).* Os estudantes que utilizaram esta estratégia, intuitivamente estavam empregar a razão de possibilidades que é uma medida de associação para variáveis dicotômicas.

No que diz respeito às estratégias parcialmente corretas, foram identificadas as seguintes:

- d) *Comparação de duas ou mais distribuições de frequências absolutas.* Nesta estratégia os estudantes utilizaram toda a informação pertinente, mas a comparação fazia-se em termos de frequências absolutas em vez da comparação de frequências relativas.
- e) *Comparação de uma frequência absoluta com a frequência marginal correspondente.*
- f) *Comparação das somas das diagonais da tabela.* O sujeito reconhece os casos favoráveis e desfavoráveis de associação, mas não chega a relacioná-los com o total de casos, isto é, utilizaram a estratégia  $[(a + d) - (b + c)]$ .

Outras estratégias dos estudantes, particularmente as que se basearam nas frequências absolutas ou apenas aplicaram uma parte dos dados na tabela de contingência, eram inadequadas, conduzindo-os a juízos incorretos de associação. De entre essas, destacam-se as seguintes:

- g) *Utilização de apenas uma célula, geralmente a de maior valor.* Esta estratégia está relacionada com a heurística de representatividade (Kahneman & Tversky, 1982), porque o estudante não utiliza toda informação relevante para resolver o problema, mas apenas a informação mais saliente.
- h) *Utilização de apenas uma distribuição condicional.* Os estudantes baseiam os seus julgamentos com base nas frequências de uma linha ou coluna.
- i) *Comparação de frequências absolutas de algumas células com o número total de casos.* Neste caso, os estudantes não chegam a comparar as probabilidades de um valor de A com diferentes valores de B.

j) *Uso das frequências marginais.* Alguns estudantes consideram o problema impossível de resolver por causa das diferenças das frequências marginais nas diferentes linhas ou colunas da tabela.

Os resultados encontrados pelos autores no estudo vão de encontro aos obtidos por Inhelder e Piaget (1976), na medida em que a maior parte dos estudantes envolvidos na pesquisa recorreram preferencialmente a estratégia g), *utilização de apenas uma célula, geralmente a de maior valor.* Constataram também que uma percentagem significativa de estudantes não sentiu a necessidade de comparar as diferentes amostras implicadas no problema para fazer um julgamento de associação, não atingindo, portanto, o nível IIIB de compreensão de associação definido por Inhelder e Piaget. Não obstante, eles não terem atingido o nível IIIB, houve um número significativo de estudantes que escolheram uma resposta correta ou parcialmente correta. Alguns estudantes não só compararam as distribuições relativas condicionais (estratégia a), como também compararam estas distribuições com as frequências marginais. Outros ainda utilizaram a comparação das possibilidades (*odds*) para resolver o problema.

#### **2.4.2. Estratégias de avaliação da associação estatística na comparação de amostras**

Nas estratégias intuitivas dos estudantes sobre associação estatística na comparação de duas amostras, Estepa et al. (1999) distinguiram dois tipos de problemas de comparação: o primeiro consistiu na comparação de duas amostras independentes (uma variável numérica medida em duas amostras independentes) e o segundo consistiu em fazer comparações de amostras relacionadas, isto é, comparar uma variável numérica medida duas vezes na mesma amostra. Os resultados desse estudo, realizados com 213 estudantes do último ano do ensino secundário (17-18 anos) revelaram que um número significativo de estudantes deu uma resposta correta e a capacidade em determinar a existência de associação em problemas de comparação de duas amostras. A seguir apresentaremos as estratégias corretas, parcialmente corretas e incorretas apresentadas pelos estudantes:

No que diz respeito as estratégias corretas, salientam-se as seguintes:

a) *Comparação de médias.* Alguns estudantes compararam as duas médias amostrais para decidir sobre a existência de associação entre as variáveis, enquanto outros usavam

implicitamente a ideia correta de que uma diferença nas médias implica associação entre variáveis.

b) *Comparação dos totais*. Alguns estudantes compararam os valores totais da variável resposta em cada amostra, partindo do pressuposto de que a dependência implica a variação da soma dos valores das duas distribuições.

c) *Comparação de percentagens*. Neste caso, alguns estudantes baseavam os seus julgamentos na comparação das proporções ou percentagens das duas amostras, partindo do pressuposto de que se nas duas amostras a percentagem dos casos em que os valores da variável aumentam é igual à percentagem dos casos em que os valores diminuem, então não existe associação.

d) *Comparar as duas distribuições*. Outros estudantes compararam a distribuição completa de frequências em ambas amostras. Assume-se que os estudantes podem ter seguido o seguinte critério: “existe independência, quando ambas as amostras têm a mesma distribuição de frequências” (Estepa et al., 1999, p. 23).

Entre as estratégias parcialmente corretas, destacamos a seguintes:

e) *Comparação de valores da variável explicativa para cada caso nas amostras relacionadas*. Estes estudantes compararam os valores correspondentes para um mesmo sujeito em ambas amostras. Esta estratégia é similar a estratégia c), *Comparação de percentagens*, embora não tenham usado percentagens ou proporções. Quando usaram esta estratégia, alguns estudantes esperavam que as diferenças dos valores de ambas as variáveis tivessem o mesmo sinal em todos os casos, o que indica uma conceção funcional ou determinista de associação.

f) *Cálculo das diferenças*. Os estudantes calcularam as diferenças entre os valores da variável explicativa em todos sujeitos da amostra relacionada e baseavam os seus julgamentos tendo em conta essas diferenças. Presume-se que os estudantes utilizaram esta estratégia partiram do pressuposto que “para determinar a associação entre uma variável contínua e uma variável dicotómica, a diferença nos correspondentes valores da variável contínua devem ser bastante significativas” (Estepa et al., 1999, p. 24).

g) *Comparação global*. Neste caso, o julgamento de associação era realizado através de uma comparação qualitativa global de ambas as distribuições sem especificar que estatísticas foram utilizadas.

Finalmente, no que diz respeito as estratégias incorretas, constataram-se as seguintes:

h) *Esperar obter valores similares.* Por vezes, alguns alunos esperavam obter valores similares em todos os casos para cada amostra, donde quando isso não acontecia, por causa da aleatoriedade dos dados, eles concluíam que as amostras eram independentes.

g) *Comparação dos valores mais altos e mais baixos de ambas as distribuições.* Os estudantes que optaram por esta estratégia possuíam uma conceção local de associação.

h) *Avaliação de coincidências.* Alguns estudantes justificaram a independência por causa da coincidência de alguns casos.

### **2.4.3. Estratégias de avaliação da associação estatística em gráficos de dispersão**

Há um consenso geral de que uma pessoa culta deve poder ler criticamente tabelas e gráficos estatísticos que se encontram na imprensa, na internet, nos meios de comunicação e no trabalho profissional. Isto supõe não só a leitura literal das tabelas ou gráficos, mas identificar as tendências, variabilidades e possíveis associações entre os dados, assim como detetar os possíveis erros, conscientes ou inconscientes, que podem distorcer a informação representada (Schield, 2006). Sobre as habilidades dos estudantes na leitura e compreensão de tabelas e gráficos estatísticos, diversos autores têm-se dedicado ao seu estudo. De entre eles, talvez a classificação mais conhecida seja a de Curcio (1989), na qual se definem os seguintes níveis de leitura e interpretação: *ler os dados*, que consiste em fazer uma leitura literal do gráfico ou tabela e que se realiza através da leitura dos factos que nele estão representados sem interpretar a informação contida no mesmo; *ler entre os dados*, em que é necessário combinar e integrar a informação e identificar relações matemáticas através de algum conhecimento prévio sobre o assunto tratado no gráfico; e, finalmente, *ler para além dos dados*, que pressupõe a capacidade de efetuar previsões a partir da informação do gráfico e um conhecimento prévio aprofundado sobre o assunto referente aos dados do gráfico.

Friel, Curcio e Bright (2001) ampliam a classificação anterior, definindo um novo nível *ler por detrás dos dados*, que consiste em analisar criticamente o método de recolha de dados, ou sua validade e fiabilidade, assim como as possibilidades de extensão das conclusões.

Fernandes e Morais (2011) verificaram que a grande maioria dos 108 alunos do 9º ano de escolaridade, que participaram no seu estudo, foram capazes de *ler os dados* corretamente dos gráficos (nível 1 de Curcio), enquanto apenas cerca de um terço desses alunos foi capaz de *ler entre os dados* e *ler além dos dados* (níveis 2 e 3 de Curcio).

Para o caso das estratégias intuitivas dos estudantes sobre a associação nos gráficos de dispersão, Estepa e Batanero (1996) constataram, tal como nos casos anteriormente descritos, estratégias corretas, parcialmente corretas e incorretas. No caso das estratégias corretas, estes autores descreveram as seguintes:

a) *Análise da variação da função.* Os estudantes baseavam os seus julgamentos acerca do tipo de associação (positiva, negativa ou independência) na ideia intuitiva da monotonia da função linear, ou seja, se as variáveis variavam no mesmo sentido, em sentido oposto ou nem num sentido nem noutro. Esta estratégia demonstra que os estudantes em questão tinham uma conceção correta de associação, segundo a qual “*independência significa a invariância da distribuição da variável dependente Y quando condicionada com o valor de X*” (Estepa & Batanero, 1996, p. 33).

b) *Padrão de ajustamento entre a nuvem de pontos e uma reta.* Os estudantes justificaram a associação entre as variáveis através da análise do ajustamento da nuvem de pontos com uma função linear. Segundo Estepa e Batanero (1996), os estudantes podem ter implicitamente assumido que se “*encontramos um bom ajustamento da nuvem de pontos a uma reta, então existe uma correlação forte ou moderada entre as duas variáveis*” (p.33).

No que diz respeito as estratégias parcialmente corretas, destacam-se as seguintes:

c) *Interpretação correta dos pontos isolados.* Embora a partir dessa informação os estudantes tenham deduzido o tipo correto de associação, eles não usaram a informação completa providenciada pelo diagrama de dispersão, o que evidencia uma conceção local de associação.

Finalmente, relativamente às estratégias incorretas, descrevem-se as seguintes:

d) *Teorias prévias.* Os estudantes basearam os seus julgamentos nas crenças prévias acerca da relação entre as duas variáveis, negligenciando a evidência dos dados concernente ao tipo de correlação. É o tipo de fenómeno a que Chapman e Chapman (1969) chamaram “*correlação ilusória*”.

e) *Outras variáveis.* Quando além da variável independente existem outras variáveis que podem afetar o valor da variável dependente, alguns estudantes consideraram que não existia correlação entre as duas variáveis. Estes estudantes interpretaram a questão concernente à dependência sob um ponto de vista causal, mostrando uma conceção causal de associação.

f) *Uniformidade*. Os estudantes esperavam ter valores singulares na correspondência entre as variáveis para afirmarem a dependência, revelando portanto uma concepção determinista de associação.

g) *Causalidade*. Para além da correlação empírica, os estudantes argumentaram que havia correlação pelo fato da variável independente não causar uma influência direta na variável dependente. Os autores consideraram esta estratégia como indicador duma concepção causal.



## **CAPÍTULO III**

### **METODOLOGIA**

No âmbito de um trabalho de investigação, a metodologia refere-se ao conjunto de opções tomadas no que concerne ao desenho da investigação, à abordagem utilizada para o estudo, às técnicas e instrumentos de recolha de dados e à análise de dados, bem como aos aspetos éticos que se necessita de ter em consideração para levar a cabo a investigação.

Este capítulo está dividido em quatro secções, onde se apresentam as opções metodológicas que definem e orientam o conjunto de métodos e técnicas utilizadas neste estudo, os participantes, os métodos de recolha das informações e, por fim, os procedimentos utilizados para o tratamento e análise dos dados recolhidos.

#### **3.1. Opções metodológicas**

Em termos de questões de investigação, pretendeu-se no estudo responder às duas seguintes questões de investigação:

- 1- Que significados atribuem os estudantes à associação estatística? Esses significados alteram-se com o ensino do tema correlação e regressão linear?
- 2- Quais as aquisições conceituais sobre a associação estatística entre duas variáveis que os estudantes adquirem com o ensino do tema correlação e regressão linear?

Neste estudo analisaram-se os significados pessoais de estudantes universitários sobre a associação estatística antes da lecionação e a sua evolução depois da lecionação deste conceito, analisando-se assim não só as suas ideias intuitivas sobre a temática como também as estratégias adotadas pelos estudantes, de modo a perceber se tais estratégias estavam de acordo ou não com os significados institucionais. Para tal elaborou-se um questionário envolvendo seis questões sobre associação estatística, as quais se inserem em três contextos: a) tabela de contingência de  $2 \times 2$ ; b) comparação de uma variável numérica em duas amostras e c) gráficos de dispersão. Este questionário foi aplicado em dois momentos: antes do ensino do tema da correlação e regressão linear e depois do ensino deste tema, como forma de avaliar a evolução dos significados de associação estatística adquiridos pelos estudantes em resultado das aulas lecionadas sobre o tema da correlação e regressão linear. Além do questionário, foi

também usado um teste para avaliar as estratégias e as aquisições conceituais dos estudantes sobre a associação estatística, que foi aplicado apenas depois dos estudantes terem passado pelo ensino do tema da correlação e regressão linear.

Dado que o estudo procura descrever os argumentos dos estudantes e estabelecer uma correspondência com os seus juízos de associação para identificação das conceções corretas e incorretas dos estudantes, tendo em conta os significados institucionais dos objetos matemáticos abordados, optou-se por realizar uma investigação de carácter qualitativo e interpretativo. Este tipo de investigação parece-nos pertinente tendo em conta a problemática do estudo, pois a compreensão dos significados atribuídos pelos estudantes implica descrição detalhada e envolve subjetividade, o que é tido em conta numa metodologia do tipo da que é aqui adotada (Bodgan & Biklen, 1994).

Tendo em conta o objetivo da pesquisa — analisar os significados pessoais dos estudantes universitários sobre a associação estatística —, antes da lecionação e a sua evolução depois da lecionação deste conceito, analisou-se não só as suas ideias intuitivas sobre a temática como também as estratégias adotadas pelos estudantes, de modo a perceber se as estratégias adotadas estão de acordo ou não com os significados institucionais.

Por outro lado, complementarmente, para estudar o efeito que a lecionação teve nas conceções e estratégias dos estudantes na avaliação da associação estatística, recorreremos também a alguns elementos quantitativos, particularmente para comparar as percentagens de estudantes nas diferentes estratégias adotadas nas duas etapas do estudo (pré-ensino e pós-ensino), segundo as respostas corretas, parcialmente corretas e incorretas.

### **3.2. Participantes**

Na presente pesquisa estudaram-se os significados que os estudantes universitários atribuem à associação estatística antes e depois do seu ensino, donde se impunha o envolvimento de estudantes no decurso de tempo entre o antes e o depois do ensino do tema da correlação e regressão linear. Para tal, participaram neste estudo um total de 57 estudantes, que frequentavam o 2º ano do curso universitário para formação de futuros professores de Matemática da Universidade Pedagógica de Moçambicanos (UP), de entre os quais 30 estudavam na delegação de Nampula e 27 na delegação de Maputo. Duas razões especiais ditaram a escolha dos estudantes do 2º ano do curso de Matemática da (UP): a primeira é que a maioria desses estudantes estão a ser preparados para exercerem futuramente a atividade do

professorado, pelo que é importante identificar e descrever as suas estratégias e concepções sobre associação de modo a contemplar esses aspetos ainda na formação inicial dos professores; a segunda razão é que interessava trabalhar com um grupo de estudantes que ainda não tinham tido aulas sobre associação estatística e regressão linear, para que se pudesse avaliar o impacto do ensino nas suas concepções iniciais.

Por uma questão de englobar no estudo um maior número de participantes, o que por sua vez aumentaria a probabilidade de obter uma maior diversificação de estratégias, optou-se por trabalhar com estudantes de duas delegações distintas da Universidade Pedagógica de Moçambique, da delegação de Maputo e da delegação de Nampula.

Do total dos 57 estudantes que participaram no estudo, 5 (8,8%) eram do sexo feminino e 52 (91,2%) do sexo masculino e apresentavam uma média de classificações na disciplina de matemática à entrada na universidade de 10,4 valores.

A maioria dos estudantes envolvidos na pesquisa 37 (64,9%) concluiu o nível médio do Ensino Secundário Geral, sendo que os restantes 20 (35,1%) concluíram o nível médio em outros outros subsistemas de ensino, especificamente 6 (10,5%) no Ensino Técnico Profissional, 12 (21,1%) na Formação de Professores e 2 (3,5%) concluíram o nível médio no seminário.

Quanto às noções de estatística adquiridas em anos anteriores, a maior parte dos participantes 44 (77,2%) afirmaram que tiveram aulas de estatística nos anos anteriores e 13 (22,8%) afirmaram que não aprenderam na escola quaisquer noções de estatística. Dos 44 (77,2%) estudantes que aprenderam noções de estatística, 15 (26,3%) afirmaram que estudaram essa temática no ensino básico e 29 (50,9%) no ensino médio.

Relativamente ao interesse em estatística, quase todos os estudantes (53, ou seja, 93,0%) afirmaram ter muito interesse em estatística, apenas 3 (5,3%) estudantes disseram ter algum interesse e só 1 (1,8%) estudante afirmou que não tem nenhum interesse.

Os estudantes responderam a um questionário e a um teste. O questionário foi aplicado em dois momentos: antes de ter sido lecionado o tema de correlação e regressão linear e imediatamente depois de ter sido lecionado esse tema, de modo a avaliar a evolução das concepções e estratégias de avaliação da associação dos estudantes. O teste foi aplicado imediatamente após o ensino do tema de correlação e regressão linear e teve por propósito principal avaliar os ganhos conceituais dos estudantes no âmbito da correlação e regressão linear com o ensino. Estes dois instrumentos de recolha de dados são descritos na próxima secção.

### **3.3. Métodos de recolha e análise de dados**

No estudo foram usados dois instrumentos de recolha de dados: um questionário (ver Anexo) e um teste (ver Anexo), que foram elaborados de modo a incluírem os diferentes significados de associação estatística. Seguidamente descreve-se cada um destes instrumentos.

#### **3.3.1. Descrição dos instrumentos de recolha de dados**

Os instrumentos de recolha de dados foram elaborados tendo em conta os conteúdos temáticos oficialmente adotados pela Universidade Pedagógica para o tratamento do tema associação estatística e regressão linear, especificamente para os estudantes que frequentam o 2º ano do curso de Ensino de Matemática.

O questionário, para além de contemplar algumas questões para obter informação de natureza pessoal dos estudantes (o sexo, o subsistema em que concluiu o ensino médio, o estudo de estatística no ensino médio, a classificação obtida na disciplina de matemática no exame de admissão à universidade e o interesse da estatística para a formação do estudante), inclui seis questões sobre a avaliação da associação estatística. Destas últimas, tem-se: a questão 1 diz respeito à avaliação da associação estatística (independência/dependência) numa tabela de contingência de  $2 \times 2$ ; as questões, 2, 3 e 4 dizem respeito à avaliação da associação estatística de uma variável numérica em duas amostras ou de duas variáveis numéricas numa amostra, representadas na forma tabular; e, finalmente, as questões 5 e 6 dizem respeito à avaliação da associação estatística entre duas variáveis numéricas a partir de dados representados num diagrama de dispersão. Além disso, na questão 6, os estudantes são também inquiridos sobre a previsão do valor de uma variável a partir do conhecimento do valor da outra variável.

O teste envolve 8 questões, das quais as questões 1, 2 e 3 são questões do tipo verdadeiro/falso e versam, fundamentalmente, sobre aspetos relativos à interpretação de propriedades do coeficiente de correlação, designadamente a intensidade da correlação, o sentido da variação e o tipo de correlação (linear ou não linear). Na questão 4 avalia-se a relação entre o valor do coeficiente de correlação entre duas variáveis e a intensidade da associação das variáveis. As questões 5 e 6 focalizam-se na utilização do diagrama de dispersão para representar uma distribuição bidimensional e avaliar a associação entre as variáveis. Nas quatro alíneas da questão 5 pede-se aos estudantes para estimarem o valor do coeficiente de

correlação linear a partir de um diagrama de dispersão dado, enquanto no caso das três alíneas da questão 6 é pedido aos estudantes para representarem um diagrama de dispersão que mostre a variação conjunta, ou não variação, dos pontos relativos a duas variáveis previamente estabelecidas. A questão 7 tem por objetivo avaliar a capacidade dos estudantes de calcularem o valor do coeficiente de correlação linear e de o interpretarem quanto ao sinal e à intensidade da correlação. Finalmente, a questão 8 centra-se na regressão linear, questionando-se os estudantes sobre estimação do valor de uma variável a partir do conhecimento do valor da outra variável. Ainda nesta questão, avalia-se se os estudantes são conscientes dos perigos de fazer determinadas previsões através de um modelo de regressão linear, especificamente para valores que se afastam consideravelmente dos valores efetivamente observados.

Globalmente, as questões do questionário, assim como as do teste, foram retiradas ou adaptadas de diferentes estudos, sempre tendo em conta o seu ajustamento aos objetivos do nosso estudo. Relativamente ao questionário, as questões 1, 3 e 5 foram retiradas de Batanero et al. (1998), a questão 6 foi adaptada de Landwehr e Watkins (1986), a questão 2 foi adaptada de Estepa et al. (1999) e, por último, a questão 4 foi retirada de Fernandes (2004) No caso do teste, as questões de 1 a 6 foram adaptadas da tese de doutoramento de Cobo (1998) e as questões 7 e 8 foram adaptadas de Fernandes (2004).

No caso do questionário, antes da sua aplicação aos estudantes que fizeram parte da amostra final, optou-se por submetê-lo a uma amostra reduzida, constituída por 12 estudantes, com características semelhantes às dos estudantes da amostra final. Com esta fase de pilotagem do questionário pretendeu-se verificar em que medida as questões formuladas despoletavam ideias intuitivas dos alunos sobre a associação estatística, pois na primeira aplicação do questionário (antes do ensino) os estudantes nunca tinham estudado o tema na escola. Por outro lado, estávamos também interessados em estudar possíveis dificuldades de compreensão e interpretação dos enunciados das questões, tendo em vista possíveis alterações. Desta fase de pilotagem do questionário não resultaram quaisquer alterações significativas.

### **3.3.2. Recolha de dados**

Os instrumentos de recolha de dados foram aplicados em dois momentos distintos da investigação. O intervalo de tempo que separou esses dois momentos foi de um mês, tendo o primeiro ocorrido em finais do mês de agosto e o segundo nos finais do mês de setembro do ano letivo de 2011/2012. Em ambos os momentos, os instrumentos de recolha de dados foram

aplicados pelos docentes da disciplina de Estatística Descritiva, sem qualquer interferência do investigador. Embora a recolha de dados não se tenha realizado no mesmo dia em ambos os grupos de estudantes das delegações de Maputo e Nampula, todos os estudantes de cada delegação iniciaram a resolução do questionário em simultâneo. Por outro lado, para que se pudesse estudar os raciocínios que os estudantes utilizavam na resolução dos problemas, foi-lhes solicitado que explicitassem convenientemente e de forma clara os seus raciocínios, quer envolvessem linguagem natural escrita, cálculos matemáticos ou ambas as formas.

No primeiro momento, imediatamente antes da leção do tema da correlação e regressão linear, foi aplicado o questionário, que teve a duração de 120 minutos (sem limite de tempo), com objetivo de analisar as conceções prévias dos estudantes em relação à associação estatística. Para tal, analisaram-se as respostas e estratégias adotadas pelos estudantes na resolução dos problemas para determinar as suas conceções e estratégias de avaliação da associação estatística. Nesta fase deu-se particular ênfase à questão de juízos de associação em que os dados são apresentados em tabelas de contingência de  $2 \times 2$ , diagramas de dispersão e através de tabelas de comparação de uma ou duas amostras.

No segundo momento do estudo, imediatamente depois da leção do tema de correlação e regressão linear, foi aplicado novamente o questionário com o intuito de avaliar o impacto do ensino na evolução dessas conceções e estratégias de avaliação da associação estatística. Tal como no primeiro momento, a aplicação do questionário teve a duração de 120 minutos. Dado que alguns estudantes não estiveram presentes simultaneamente nos dois momentos de aplicação do questionário, respondendo apenas uma vez, quando estiveram presentes, eles não foram considerados no nosso estudo. Nestas circunstâncias, na delegação de Maputo houve 12 estudantes que responderam apenas no pré-ensino e 6 estudantes que responderam apenas no pós-ensino, enquanto na delegação de Nampula 3 estudantes reponderam apenas no pré-ensino e 2 responderam apenas no pós-ensino.

Ainda no segundo momento foi aplicado também um teste com a mesma duração de tempo, ou seja, 120 minutos, para avaliar as aquisições conceituais dos alunos no âmbito da correlação e regressão linear depois do tema ter sido ensinado. Neste caso, diferentemente do questionário, esse tempo foi estabelecido como o limite máximo permitido para responder ao teste. As questões contempladas no teste estão relacionadas com a avaliação e a determinação da correlação e regressão linear, em cuja representação intervêm as formas numérica, gráfica e verbal.

### **3.3.3. Análise de dados**

Após a recolha dos dados, passámos ao seu tratamento e análise tendo em vista obter os resultados do estudo e, posteriormente, responder às questões de investigação. Para tal, e com base na análise das respostas escritas dos estudantes, classificaram-se as respostas em corretas, parcialmente corretas e incorretas e, em cada um destes tipos de resposta, agruparam-se as suas justificações em diferentes categorias, sendo determinadas frequências absolutas e percentagens dos vários tipos de respostas e categorias que resumimos em tabelas, segundo os dois momentos em que foi administrado do questionário, ou seja, no pré-ensino e no pós-ensino. Com este procedimento compararam-se as frequências (percentagens) das respostas corretas, parcialmente corretas e incorretas do pré-ensino com as do pós-ensino, com vista a estudar o impacto do ensino sobre o tipo de respostas. No caso do teste procedeu-se de modo análogo, naturalmente sem se efetuarem comparações uma vez que ele foi aplicado apenas uma vez, imediatamente após a lecionação do tema da correlação e regressão linear.

A apresentação dos resultados foi acompanhada, para além do uso das frequências absolutas e relativas das respostas e estratégias, também da apresentação de exemplos ilustrativos dos raciocínios subjacentes às mesmas, tal como foram referidos pelos estudantes, de modo a tornar mais descritivo e vivo o relatório da investigação (Gall, Gall & Borg, 2003).

Por último, em toda a análise estatística usou-se o programa Statistical Packedge for the Social Sciences (SPSS), versão 18 para Windows.



## **CAPÍTULO IV**

### **APRESENTAÇÃO DOS RESULTADOS**

#### **4.1. Evolução dos significados pessoais de associação estatística com o ensino**

Nesta secção apresentam-se os significados sobre associação estatística atribuídos pelos estudantes do 2º ano do ensino universitário em Moçambique. Os significados foram estudados a partir das respostas dos estudantes a seis questões sobre associação estatística, as quais se inserem em três contextos: a) tabela de contingência de  $2 \times 2$ ; b) comparação de uma variável numérica em duas amostras e c) gráficos de dispersão. Em cada um dos contextos analisou-se, fundamentalmente, o tipo de associação percebida pelos estudantes e a independência. Para classificar e compreender as estratégias de avaliação da correlação utilizadas intuitivamente pelos estudantes, analisamos os conceitos matemáticos e teoremas em ação que poderiam estar implicados nos diferentes procedimentos, assim como os tipos de erros relacionados com os mesmos. Entenda-se que o teorema em ação a que nos referimos diz respeito a um conjunto de estratégias, propriedades que são tidas como verdadeiros pelos estudantes na medida em que executam uma determinada tarefa (Vergnaud, 1998).

As questões foram aplicadas aos estudantes antes e depois da leção do tema Correlação e Regressão Lineares, com o propósito de avaliar o impacto do ensino do tema nos significados atribuídos pelos estudantes à associação estatística.

##### **4.1.1. Avaliação da associação estatística numa tabela de contingência de $2 \times 2$**

Neste contexto insere-se apenas a questão 1, cujos resultados da análise efetuada se apresentam a seguir.

### Questão 1

Num centro médico, foram observadas 250 pessoas para determinar se o hábito de fumar tem alguma relação com a doença da bronquite, tendo-se obtido os resultados da tabela seguinte.

	Tem bronquite	Não tem bronquite	Total
Fuma	90	60	150
Não fuma	60	40	100
Total	150	100	250

Usando a informação da tabela, achas que nesta amostra de pessoas a doença da bronquite depende de fumar? Explica a tua resposta.

Esta questão diz respeito à relação entre duas variáveis qualitativas, especificamente “fumar” ( $F$ ) e “doença de bronquite” ( $B$ ), representadas numa tabela de contingência  $2 \times 2$ . O objetivo desta questão era analisar as estratégias de resolução utilizadas pelos estudantes sobre a associação entre as variáveis em estudo antes e depois de lecionadas as aulas do tema Correlação e Regressão Lineares. Para tal, analisaram-se as respostas dos estudantes e definiram-se estratégias de resolução segundo a informação envolvida nas respostas dizia respeito a: 1.1: uma única célula; 1.2: duas células e 1.3: quatro células. Finalmente, na categoria 1.4: outra incluíram-se as estratégias distintas das anteriores.

Na tabela 1 podem observar-se as frequências absolutas e percentagens de estudantes que recorreram à informação de um número diferente de células.

Tabela 2. Frequência absoluta (percentagem) de estudantes segundo o número de células a que recorreram nas suas respostas

Número de células envolvidas na resposta	Frequências (%)	
	Pré-ensino	Pós-ensino
1.1: utilização de uma única célula	14 (24,6)	10 (17,5)
1.2: utilização de duas células	22 (38,6)	14 (24,6)
1.3: utilização de quatro células	13 (22,8)	25 (43,9)
1.4: outra	8 (14,0)	8 (14,0)

A sofisticação da resolução do estudante aumenta à medida que ela envolve informação de uma, duas ou quatro células da tabela de contingência. Pela Tabela 1 verifica-se que em ambos os momentos de aplicação do questionário houve uma maior percentagem de respostas com base na estratégia 1.2 e 1.3, que envolvem a utilização de duas e quatro células respetivamente. Salienta-se ainda, do pré-ensino para o pós-ensino, um aumento do número de

estudantes que utilizaram informação de quatro células, que é a estratégia mais sofisticada, acompanhada por uma diminuição semelhante de estudantes a utilizarem informação de duas células.

Na Tabela 2 apresentam-se as frequências absolutas e as percentagens dos vários tipos de resposta (correta e incorreta) segundo o número de células envolvidas na resposta.

Tabela 3. Frequência absoluta (percentagem) dos diferentes tipos de resposta segundo o número de células envolvidas

Número de células envolvidas na resposta	Frequências (%) dos diferentes tipos de resposta			
	Pré-ensino		Pós-ensino	
	Correta	Incorreta	Correta	Incorreta
1.1: utilização de uma única célula	6 (10,5)	8 (14,0)	5 (8,8)	5 (8,8)
1.2: utilização de duas células	4 (7,0)	18 (31,6)	1 (1,7)	13 (22,8)
1.3: utilização de quatro células	11 (19,3)	2 (3,5)	20 (35,1)	5 (8,8)
1.4: outra	3 (5,3)	5 (8,8)	0 (0,0)	8 (14,0)
Total	24 (42,1)	33 (57,9)	26 (45,6)	31 (54,4)

Por observação da tabela verifica-se que é quando os estudantes recorrem à informação de duas células que é menor a proporção entre as frequências de respostas corretas e incorretas, aumenta quando recorrem à informação de uma célula e, mais ainda, quando recorrem à informação de quatro células. Assim, é no caso em que os estudantes recorrem à informação das quatro células que se obtém respostas mais fiáveis.

Por outro lado, entre o pré-ensino e o pós-ensino não se observam grandes discrepâncias, o que leva a concluir que o ensino da associação estatística não produziu alterações importantes ao nível das respostas (corretas e incorretas).

Finalmente, na categoria “outra”, que inclui um pequeno número de estudantes, verificou-se que eles basearam as suas respostas em ideias prévias acerca da associação entre as variáveis em causa (4 alunos no pré-ensino e 3 alunos no pós-ensino) ou apresentaram justificações ininteligíveis (4 alunos no pré-ensino e 5 alunos no pós-ensino). Por exemplo, o estudante E6 referiu no pré-ensino: “Na minha opinião, acho que nesta amostra de pessoas a doença de bronquite não depende de fumar porque o facto de o maior número de fumadores terem bronquite não significa que é devido ao fumo do cigarro” e o estudante E22 mencionou no pós-ensino: “Eu penso que o fumar pode influenciar que a pessoa tenha bronquite [pois] é um dos fatores da bronquite”. Em ambos estes casos, para além do fumar, os estudantes admitem

a existência de outras causas que podem estar na origem da bronquite, mobilizando assim outros conhecimentos/experiências prévias para efetuarem julgamentos na situação com que se depararam. Já no caso dos argumentos não suficientemente claros para serem interpretados, o estudante E33, pré-ensino, respondeu do seguinte modo: “A bronquite depende de fumar devido à intoxicação dos pulmões por causa do fumo que as pessoas inalam”.

Em ordem a aprofundar a forma como foi usada pelos estudantes a informação da célula ou células envolvidas nas suas estratégias de resolução criaram-se estratégias referidas de acordo com os conceitos matemáticos envolvidos e os teoremas implicitamente aplicados nos diferentes procedimentos.

No caso do recurso de uma única célula definiram-se três estratégias de resposta e determinadas as respetivas frequências e percentagens, que se apresentam na Tabela 3.

Tabela 4. Frequência absoluta (percentagem) de estudantes nas diferentes estratégias quando foi usada informação de uma célula

Estratégias	Frequências (%)	
	Pré-ensino	Pós-ensino
1.1.a): célula relativa a fumar e ter bronquite (célula de maior frequência, célula $a$ )	7 (12,3)	6 (10,5)
1.1.b): célula relativa a não fumar e ter bronquite (célula $c$ )	7 (12,3)	4 (7,0)

Quando os estudantes recorreram à informação de uma célula, identificaram-se as duas estratégias 1.1.a) e 1.1.b). Estas duas estratégias conduziram, quase sempre, a tipos de respostas diferentes. No caso da estratégia 1.1.a), os estudantes afirmaram a dependência entre as variáveis pelo facto de ser maior a frequência da célula relativa a fumar e ter bronquite (célula  $a$ ), conduzindo assim à resposta errada. É disso exemplo a seguinte resposta, apresentada pelo estudante E31, pós-ensino: “Doença da bronquite depende de fumar porque o número das pessoas que fumam e têm bronquite é maior em relação aos outros casos”. Segundo Batanero et al. (1996), este tipo de estratégia constitui um indicador de uma conceção local de associação na medida em que os sujeitos nos seus julgamentos utilizam apenas parte dos dados disponíveis no problema. Se esta parte de informação lhes permitiu confirmar um determinado tipo de associação, então eles generalizaram-na para a totalidade dos dados.

Os estudantes que recorreram à estratégia 1.1.b) sustentaram os seus julgamentos tendo em conta a célula correspondente ao caso não fumar e ter bronquite (célula  $c$ ), partindo do pressuposto de que para haver dependência entre as variáveis em estudo a célula devia ser

nula; caso contrário, não haveria dependência. Por outras palavras, eles esperavam a existência de uma correspondência que atribuísse um único valor à variável “bronquite” para cada valor da variável “fumar”. Assim, porque a frequência desta célula não era nula, os estudantes afirmaram a independência entre as variáveis, que é a resposta correta. A resposta do estudante E13, no pré-ensino, elucida tal facto: “Ter bronquite não depende de fumar porque existem pessoas que não fumam e têm bronquite”. Este tipo de estratégia é, segundo Batanero et al. (1996), um indicador de uma conceção determinística de associação porque os estudantes não admitem a presença de qualquer exceção para a existência de associação entre duas variáveis. Neste caso, apesar de o estudante em referência (E13) ter feito um julgamento correto (independência), ele fê-lo com base numa conceção incorreta.

No caso da utilização da informação de duas células foram estabelecidas também duas estratégias de resposta e determinadas as respetivas frequências e percentagens, tal como se mostra na Tabela 4.

Tabela 5. Frequência absoluta (percentagem) de estudantes nas diferentes estratégias quando foi usada informação de duas células

Estratégias	Frequências (%)	
	Pré-ensino	Pós-ensino
1.2a): comparação das frequências absolutas de duas células	15 (26,3)	11 (19,3)
1.2b): comparação das frequências relativas de duas células	7 (12,3)	3 (5,3)

Qualquer das estratégias 1.2a) e 1.2b) levaram os estudantes a selecionarem a resposta incorreta quase sempre.

Na estratégia 1.2a) os estudantes compararam as frequências absolutas em linha, em coluna destacando-se as comparações entre as células da primeira coluna, células *a* e *c* (8 estudantes no pré-ensino e 4 estudantes no pós-ensino), e entre as células da primeira linha, células *a* e *b* (4 estudantes no pré-ensino e 2 estudantes no pós-ensino). Em ambas estas estratégias apenas se relacionam parcialmente as duas variáveis, pois no primeiro caso não se consideram os casos sem bronquite e no segundo caso não se consideram os casos dos não fumadores.

No recurso às células da primeira coluna, os estudantes compararam as frequências de fumar e não fumar entre as pessoas que têm bronquite. Nesta comparação, perante uma maior frequência de fumar e ter bronquite do que de não fumar e ter bronquite, os estudantes afirmaram a dependência entre fumar e ter bronquite. O exemplo a seguir, apresentado pelo

estudante E53, no pré-ensino, ilustra esta situação: “a doença depende de fumar devido a [existir] um maior número de pessoas que fumam com bronquite do que as que não fumam com a mesma doença”.

No caso do recurso às células da primeira linha, os estudantes compararam as frequências de ter e não ter bronquite entre as pessoas que fumam. Nesta comparação, observando uma maior frequência de fumar e ter bronquite do que fumar e não ter bronquite, os estudantes afirmaram igualmente a dependência entre fumar e ter bronquite. A este respeito, o estudante E51, no pós-ensino, refere: “A doença da bronquite, com base nesta amostra, depende de fumar porque, para 150 fumadores, 90 têm bronquite e 60 não têm. Portanto, existe mais fumadores com bronquite do que fumadores sem bronquite”.

Relativamente à estratégia 1.2b), verificou-se que os estudantes compararam, quase sempre, as frequências relativas das duas células da primeira coluna, células *a* e *c* (6 estudantes no pré-ensino e 3 estudantes no pós-ensino), tendo em conta a dimensão total da amostra. Neste caso, os estudantes calcularam as frequências relativas de forma isolada e não relacionaram a frequência relativa condicional de uma variável com a da outra variável. Assim, estes estudantes afirmaram existir dependência entre fumar e doença de bronquite porque a frequência ou a percentagem de pessoas com bronquite e que fumam é maior do que a frequência ou percentagem dos que não fumam e têm bronquite, tal como mostra a seguinte resposta de um dos estudantes:

Explica a tua resposta: Sim, porque o número maior das pessoas que fuma tem bronquite ou maior percentagem das pessoas que fuma tem a doença de bronquite em relação das pessoas que fumam e não tem bronquite.

$$\frac{90}{250} = 0,36 \cdot 100\% = 36\% \text{ pessoas que fumam e tem bronquite.}$$
$$\frac{60}{250} = 0,24 \cdot 100\% = 24\% \text{ pessoas que fumam e não tem bronquite.}$$

Logo a doença do bronquite depende de fumar.

Figura 1. Resolução da questão 1 pelo estudante E12 no pré-ensino.

Tal como nas estratégias envolvendo o recurso à informação de apenas uma célula, também nestas estratégias os argumentos dos estudantes mostram a sua adesão a uma conceção local de associação, que foi descrita anteriormente (Batanero et al., 1996).

Finalmente, no caso da utilização da informação de quatro células foram estabelecidas sete estratégias de resposta e determinadas as respectivas frequências e percentagens, tal como se mostra na Tabela 5.

Tabela 6. Frequência absoluta (percentagem) de estudantes nas diferentes estratégias quando foi usada informação de quatro células

Estratégias	Frequências (%)	
	Pré-ensino	Pós-ensino
1.4a): comparação das frequências absolutas das quatro células	1 (1,8)	3 (5,3)
1.4b): comparação das proporções em linha e/ou coluna	4 (7,0)	5 (8,8)
1.4c): comparação de duas ou mais frequências relativas condicionadas	6 (10,5)	11 (19,3)
1.4d): comparação da probabilidade de cada célula com o produto das respetivas probabilidades marginais (critério de independência de dois eventos)	1 (1,8)	2 (3,5)
1.4e): comparação das diferenças $a - b$ e $c - d$	0 (0,0)	1 (1,8)
1.4f): comparação de uma frequência relativa condicionada com a correspondente frequência na amostra total	1 (1,8)	1 (1,8)
1.4g): determinação do valor do coeficiente de correlação de Pearson	0 (0,0)	2(3,5)

No que concerne à estratégia 1.4a), verificou-se que os estudantes estabeleceram comparações entre as frequências absolutas das células em linha ou em coluna. Esta estratégia conduziu à seleção da resposta correta ou errada com base numa avaliação subjetiva do estudante. Por exemplo, o estudante E18, no pós-ensino, referiu:

Nesta amostra de pessoas, a doença de bronquite não depende de fumar. Olhando para o número de pessoas que fumam e para as pessoas que não fumam sempre o número de pessoas que têm bronquite é maior em relação ao número de pessoas que não têm bronquite.

Já na estratégia 1.4b), que consistiu na comparação das proporções em linha e/ou em coluna, a partir das respetivas frequências absolutas, os estudantes selecionaram a resposta correta. Nesta estratégia comparam-se os casos a favor de uma variável com os casos contra a outra variável (*odds*). Recorrendo a esta estratégia, o estudante E2, no pós-ensino, apresentou o seguinte argumento:

A doença de bronquite não depende do fumo, mas vez que observando na tabela, a razão entre os que fumam e os que não fumam é constante, isto é, é igual ( $\frac{90}{60} = \frac{60}{40}$ ).

Figura 2. Resolução da questão 1 pelo estudante E2 no pós-ensino.

Embora este estudante tenha determinado proporções em linha, houve outros estudantes que determinaram proporções em coluna ou simultaneamente em linha e em coluna. Por exemplo, o estudante E20, no pré-ensino, definiu a proporcionalidade em coluna, como se mostra na figura seguinte, e concluiu que as variáveis não estão relacionadas.

O número de pacientes com bronquite é directamente proporcional aos dos que não tem bronquite,  
 $K = \frac{90}{60} = 1,5$ , e mesmo acontece para os que não fumam  
 $K = \frac{60}{40} = 1,5$ .

Figura 3. Resolução da questão 1 pelo estudante E20 no pré-ensino.

Na estratégia 1.4c) comparam-se as frequências relativas condicionadas  $h(B_j | F_i)$  de cada valor de  $B_j$  com dois valores diferentes de  $F_i$ . Portanto, os estudantes que usaram esta estratégia, implicitamente, basearam-se no seguinte teorema: "A dependência de uma variável  $B$  de outra  $F$  implica a variação da frequência relativa condicional  $h(B_j | F_i)$  quando  $F_i$  varia" (Batanero et al., 1996, p. 159). É exemplo desta estratégia o argumento apresentado pelo estudante E40 no pós-ensino.

PARA FUMADORES: 90 TEM BRONQUITE:  $\frac{90}{150} = 0,6 \cdot 100\% = 60\%$   
 PARA OS NÃO FUMADORES: 60 TEM BRONQUITE:  $\frac{60}{100} = 0,6 \cdot 100\% = 60\%$   
 COM ESTE RESULTADO, MOSTRA QUE A DOENÇA DE BRONQUITE NÃO DEPENDE DE FUMAR UMA VEZ QUE 60% DE CADA AMOSTRA (FUMADORES E NÃO FUMADORES), POSSUAM A DOENÇA.

Figura 4. Resolução da questão 1 pelo estudante E40 no pós-ensino.

Este estudante determinou e comparou as frequências relativas condicionadas em linha, enquanto outros determinaram e compararam as frequências relativas condicionadas em coluna e outros determinaram e compararam quatro frequências relativas condicionadas simultaneamente. Neste último caso, o estudante E23, no pós-ensino, apresentou o seguinte argumento:

Olhemos para os seguintes dados:

Def:  $P(F|B)$  a probabilidade de ser fumador dado que tem bronquite

$P(F|\bar{B})$  a probabilidade de não ser fumador dado que tem bronquite

Dados: 98 fumos

De entre estes podemos observar que:

$P(B|F) \rightarrow$  Probabilidade de ter na: ter bronquite após que é fumador é:

$P(B|F) = \frac{60}{100} = 60\%$

$P(\bar{B}|F) \rightarrow$  probabilidade de não ter bronquite dado que não é fumador

$P(\bar{B}|F) = \frac{40}{100} = 40\%$

$P(F|B) = \frac{98}{150} = 60\%$

$P(F|\bar{B}) = \frac{60}{150} = 40\%$

$P(F|B) = P(B|F) = 60\%$

$P(F|\bar{B}) = P(\bar{B}|F) = 40\%$

Esses resultados mostram que não existe nenhuma relação

Figura 5. Resolução da questão 1 pelo estudante E23 no pós-ensino.

O uso do critério de independência de dois eventos, correspondente à estratégia 1.4d), foi pouco utilizada (1 estudante no pré-ensino e 2 estudantes no pós-ensino) e consiste na comparação da probabilidade da interseção de dois eventos com o respetivo produto das probabilidades, isto é,  $P(B_j \cap F_i) = P(B_j) \times P(F_i)$ . Depois de ter afirmado a independência entre as duas variáveis, o estudante E34, no pós-ensino, apresentou o seguinte argumento.

Para responder esta questão há que analisar se os eventos "ter bronquite" e "fumar" são ou não independentes, nesse contexto para que os eventos sejam independentes, a probabilidade da interseção deve ser igual ao produto das probabilidades dos eventos!

Designando o evento ter "bronquite" por  $A$  e fumar por  $B$ , tem-se

$$P(A \cap B) = P(A) \cdot P(B)$$

baseando na tabela  $P(A \cap B) \rightarrow$  pessoas que fumam e têm bronquite

$$P(A \cap B) = \frac{90}{250} = \frac{9}{25} = 0,36$$

$$P(A) = \frac{150}{250} = \frac{15}{25} = 0,6$$

$$P(B) = \frac{150}{250} = \frac{15}{25} = 0,6$$

Substituído na relação

$$P(A \cap B) = P(A) \cdot P(B)$$

$$0,36 = 0,6 \cdot 0,6$$

$$0,36 = 0,36$$

Como verifica-se a relação, os eventos são independentes, quer dizer que "ter a doença de bronquite não depende de fumar".

Figura 6. Resolução da questão 1 pelo estudante E34 no pós-ensino.

A estratégia 1.4e), que consiste na comparação das diferenças  $a-b$  e  $c-d$ , foi adotada apenas pelo estudante E13 no pós-ensino. Este estudante apresentou a seguinte argumentação para justificar a independência entre as variáveis:

Porque pode-se observar que a diferença entre os fumadores com e sem bronquite é de 30 e no caso dos não fumadores a diferença é de 20, tendo em conta que o universo dos fumadores é 150 e dos não fumadores é 100.

Admitindo que no argumento do estudante está implícita a afirmação da igualdade

$\frac{a-b}{a+b} = \frac{c-d}{c+d}$ , e conseqüentemente que a diferença  $\frac{a-b}{a+b} - \frac{c-d}{c+d}$  é zero, chega-se à

conclusão que esta diferença é equivalente a  $\frac{2(ad-bc)}{(a+b)(c+d)}$ . Ora, como a diferença entre as

frequências relativas condicionadas é dada por  $\frac{a}{a+b} - \frac{c}{c+d} = \frac{ad-bc}{(a+b)(c+d)}$ , o que constitui

uma estratégia correta de avaliação da associação estatística em tabelas de contingência de 2x2 (Batanero et al., 1996), concluiu-se que a justificação apresentada pelo estudante é válida.

Também a estratégia 1.4f), que consiste em comparar uma frequência relativa condicionada com a correspondente frequência na amostra total, foi adotada por apenas 1 estudante no pré-ensino e por outro no pós-ensino. Neste caso, a igual proporção de incidência da bronquite na amostra total e no subconjunto dos fumadores, levou os estudantes a afirmarem

a independência entre fumar e ter bronquite, como aconteceu com o estudante E40, no pré-ensino.

$$P(\text{ter bronquite} / \text{fumar}) = \frac{P(\text{ter bronq.} \cap \text{fumar})}{P(\text{fumar})}$$

$$= \frac{\frac{90}{250}}{\frac{150}{250}} = 0,6 = 60\%$$

Agora: A doença de bronquite não dependerá de fumar se:  $P(\text{ter bronquite} / \text{fumar}) = P(\text{ter bronquite})$

$$0,6 = \frac{150}{250}$$

$$0,6 = 0,6$$

f: Como as probabilidades são iguais então a doença de bronquite não depende de fumar.

Figura 7. Resolução da questão 1 pelo estudante E40 no pré-ensino.

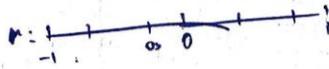
Por último, a estratégia 1.4g), que se baseou na determinação do valor do coeficiente de correlação de Pearson e na avaliação da sua grandeza, apenas foi adotada por 2 estudantes no pós-ensino. Embora se trate de uma estratégia incorreta, uma vez que o coeficiente de correlação de Pearson se aplica apenas a variáveis quantitativas contínuas, o facto de os estudantes terem obtido um valor próximo de zero ou de  $-1$  ou  $1$  para o seu valor levou-os a afirmar, respetivamente, a independência ou dependência das variáveis em jogo. Por exemplo, o estudante E50 apresentou o seguinte argumento:

$$d_2 r^2 = \frac{n \sum xy - \sum x \cdot \sum y}{\sqrt{(n \sum x^2 - (\sum x)^2) \cdot (n \sum y^2 - (\sum y)^2)}} = \frac{2 \cdot 2800 - 150 \cdot 100}{\sqrt{2 \cdot 11200 - 22500} \cdot \sqrt{2 \cdot 5800 - 10000}}$$

$$d_2 r^2 = \frac{36000}{(-19260)^2} = \frac{36000}{370452600} = 0,0009\%$$

$$d = 0,0009$$
  

$$r = 0,03 = 3\%$$



como  $r = 0,03$  que equivale a 3% de o coeficiente de determinação, infimo tudo mostra que ter bronquite não depende de fumar, exist outros fatores que influenciam na obtenção de doenças.

ou talvez em Pouca Palavras o fator fumar influencia na obtenção de doenças e a prova que não fumaras menos tivemos bronquite nos seguintes a tabela tudo mostra que nos o fumo de cigarro que dita a taxa de obtenção de bronquite.

Figura 8. Resolução da questão 1 pelo estudante E50 no pós-ensino.

A seleção pelos alunos do coeficiente de correlação de Pearson para avaliar a associação estatística entre as variáveis certamente terá sido influenciada pela ênfase que lhe foi dada no ensino da associação e correlação linear.

Analisando os resultados desta questão, de uma forma global, podemos concluir que os estudantes nem sempre utilizaram todos os dados relevantes do problema e que o desenvolvimento do conceito de associação não se revelou muito intuitivo. Antes da instrução formal do tema verificou-se que apenas 42,1% da totalidade dos estudantes apresentou a resposta correta, enquanto depois da instrução formal do tema essa percentagem aumentou ligeiramente, atingindo os 45,6%.

Já ao nível das estratégias corretas, que correspondem às estratégias 1.4b), 1.4c), 1.4d), 1.4e) e 1.4f), todas elas envolvendo o recurso à informação de quatro células, constata-se que foram adotadas por 21,1% da totalidade dos estudantes no pré-ensino e por 35,1% no pós-ensino. Neste caso, na globalidade, obtiveram-se percentagens inferiores às obtidas para as respostas corretas; contudo, no caso das estratégias corretas, verifica-se um aumento superior das percentagens do pré-ensino para o pós-ensino.

#### 4.1.2. Avaliação da associação estatística numa tabela de duas amostras

Neste contexto inserem-se as questões 2, 3 e 4, cujo objetivo é avaliar as estratégias e julgamentos de associação dos estudantes quando estão diante da comparação de uma variável numérica em duas amostras qualitativas dicotômicas ou quando estão diante da comparação de uma variável numérica em duas amostras relacionadas, ou seja, com dados emparelhados.

##### Questão 2

Depois de medir o nível de açúcar no sangue de crianças do sexo masculino e feminino de uma escola, obtiveram-se os dados do quadro seguinte.

	Crianças																			
	do sexo masculino									do sexo feminino										
Nível de açúcar no sangue	9	0	9	8	6	7	4	9	8	9	6	0	7	0	8	3	6	7	7	3

Usando a informação do quadro, achas que o nível de açúcar no sangue nesta amostra depende do sexo das crianças? Explica a tua resposta.

Nesta questão compara-se o nível de açúcar no sangue em crianças do sexo masculino e feminino. Depois de analisadas as respostas dos estudantes, antes e depois da leção do tema Correlação e Regressão Lineares, definiram-se as três seguintes estratégias: estratégias corretas, parcialmente corretas e incorretas, cujas frequências são apresentadas na tabela 6.

Tabela 7. Frequência absoluta (percentagem) de estudantes nas diferentes estratégias da questão 2

Respostas	Frequências (%)	
	Pré-ensino	Pós-ensino
<u>Corretas</u>		
2.1: comparação das médias	15 (26,3)	25 (43,9)
2.2: comparação dos totais	9 (15,8)	7 (12,3)
<u>Parcialmente corretas</u>		
2.3: comparação de alguns valores das distribuições	0 (0,0)	5 (8,8)
2.4: comparação global	23 (40,4)	15 (26,3)
<u>Incorretas</u>		
2.5: avaliação das coincidências nas duas distribuições	3 (5,3)	2 (3,5)
2.6: comparação dos valores mais elevados e mais baixos de ambas distribuições	4 (7,0)	2 (3,5)
2.7: comparação dos valores extremos das duas distribuições	3 (5,3)	1 (1,8)

Os resultados apresentados na tabela 6 mostram uma diversidade das estratégias adotadas pelos estudantes para analisarem a relação entre as variáveis em estudo, especificamente as variáveis sexo e nível de açúcar no sangue (Estepa et al., 1999).

Entre as diferentes estratégias, existem aquelas que classificamos como estratégias corretas, isto é, aquelas em que os estudantes optaram por comparar as médias amostrais para decidir sobre a existência de associação entre as variáveis (estratégia 2.1), sendo que consideravam haver dependência entre elas se existisse uma diferença nas respectivas médias amostrais. Por exemplo, um estudante afirmou que “O nível de açúcar no sangue depende do sexo, uma vez que olhando para os níveis de açúcar verifica-se que, em média, há maior nível nas crianças do sexo masculino (6,9) do que nas do sexo feminino (4,7)” (E10, pré-ensino).

Na mesma lógica, houve estudantes que compararam a soma dos níveis de açúcar nos dois grupos, que é uma estratégia igualmente correta (estratégia 2.2), como se verifica em respostas do tipo: “uma vez que o teor de açúcar no sangue das 10 crianças de sexo masculino é de 69 e nas 10 crianças do sexo feminino é de 47, então o nível de açúcar depende de sexo” (E38, pós-ensino). Deste modo, os estudantes que optaram por comparar as médias nas duas amostras conceberam a média como um valor representativo da amostra. O mesmo pode ser dito acerca da comparação dos totais, uma vez que os totais são valores representativos de um conjunto de dados.

No que diz respeito às respostas parcialmente corretas, houve estudantes que avaliaram a associação com base na comparação de alguns valores das duas distribuições (estratégia 2.3), isto é, compararam a frequência de cada um desses valores em ambas distribuições, partindo do pressuposto de que para existir dependência era necessário que tivessem frequências diferentes em ambas distribuições. Por exemplo: “O nível de açúcar depende do sexo da criança porque o nível máximo de açúcar no sangue foi observado nas crianças do sexo masculino 4 vezes e nenhuma vez nas crianças do sexo feminino” (E13, pós-ensino). Esta estratégia denota uma concepção local de associação, conforme foi referida anteriormente.

A estratégia 2.4 também levou os estudantes a darem uma resposta parcialmente correta, na medida em que fizeram uma comparação qualitativa global das duas amostras, sem especificaram que estatísticas foram usadas. Por exemplo: “O nível de açúcar no sangue depende do sexo porque o nível de açúcar nas crianças do sexo masculino é mais acentuado em relação ao feminino” (E3, pré-ensino). Esta estratégia, a par da de comparação das médias, foi a que mais se salientou nos estudantes, tanto no pré-ensino como no pós-ensino. Na estratégia de

comparação global observou-se uma ligeira diminuição da frequência do pré para o pós-ensino, ao contrário da estratégia de comparação das médias em que houve um aumento significativo de estudantes que optaram por esta estratégia do pré para o pós-ensino.

No que respeita às respostas incorretas, observaram-se três casos: 2.5: avaliação das coincidências nas duas distribuições; 2.6: comparação dos valores mais elevados e mais baixos de ambas distribuições e 2.7: comparação dos valores extremos das duas distribuições. Em todos estes casos, os estudantes utilizaram apenas uma parte dos dados para a tomada da decisão sobre a associação entre as variáveis.

Na estratégia 2.5 alguns estudantes afirmaram a independência por causa da coincidência de alguns casos, tal com ilustra a seguinte resposta: “Não depende pois existem crianças do sexo masculino com o mesmo nível [de açúcar no sangue] de crianças do sexo feminino. Por exemplo, crianças do nível 8 e 7” (E11, pré-ensino). Neste exemplo pode inferir-se que o estudante possui dois tipos de conceção, especificamente a conceção local e a conceção determinista, na medida em que, por um lado, utiliza parte dos dados do problema ao invés da totalidade dos dados e, por outro, o estudante não admite exceções para haver associação quando considera que a cada elemento de uma amostra devia corresponder um elemento diferente da outra amostra.

Em relação à estratégia 2.6, os estudantes justificaram os seus julgamentos com base na variabilidade dos valores de ambas distribuições, ou seja, pelo facto de existirem níveis de açúcar no sangue relativamente mais altos e mais baixos em crianças do sexo masculino do que em algumas crianças do sexo feminino e vice-versa. Por exemplo: “O nível de açúcar no sangue não depende do sexo porque, por exemplo, existem crianças do sexo masculino com níveis mais elevados e outras com níveis mais baixos” (E8, pré-ensino).

Por último, na estratégia 2.7 os estudantes optaram por comparar os valores extremos de ambas distribuições, tal como mostra a seguinte resposta: “O nível de açúcar no sangue depende do sexo porque tanto crianças do sexo masculino como do sexo feminino têm o nível mínimo 0, enquanto que o nível máximo 9 apenas existe nas do sexo masculino” (E34, pós-ensino).

Tanto na estratégia 2.6 como na estratégia 2.7 houve uma forte tendência de os estudantes ignorarem a totalidade dos dados a favor de parte destes. Particularmente na estratégia 2.7, apesar do estudante ter feito um julgamento correto, ele baseou-o numa estratégia incorreta.

### Questão 3

Mediu-se a pressão sanguínea num grupo de 10 mulheres, antes e depois de se aplicar um tratamento médico, tendo-se obtido os dados da tabela seguinte.

Mulher	A	B	C	D	E	F	G	H	I	J
Pressão sanguínea antes do tratamento	115	112	107	119	115	138	126	105	104	115
Pressão sanguínea depois do tratamento	128	115	106	128	122	145	132	109	102	117

Usando a informação da tabela, achas que a pressão sanguínea nesta amostra de mulheres depende do tempo em que foi medida (antes ou depois do tratamento)? Explica a tua resposta.

A questão central da pergunta é saber se as variáveis numéricas têm a mesma distribuição em ambas amostras, podendo para tal, por exemplo, ser calculada a diferença das médias ou construir a representação gráfica de ambas distribuições. As estratégias utilizadas pelos estudantes e as respetivas frequências e percentagens, antes e depois da leção do tema Correlação e Regressão Lineares, estão descritas na tabela 7.

Tabela 8. Frequência absoluta (percentagem) de estudantes nas diferentes estratégias da questão 3

Respostas	Frequências (%)	
	Pré-ensino	Pós-ensino
<u>Corretas</u>		
3.1: comparação das percentagens	5 (8,8)	9 (15,8)
3.2: análise da variação dos pontos no gráfico de dispersão	0 (0,0)	3 (5,3)
3.3: determinação e análise do coeficiente de correlação	0 (0,0)	10 (17,5)
3.4: comparação das médias	6 (10,5)	5 (8,8)
<u>Parcialmente corretas</u>		
3.5: análise de casos excecionais	8 (14,0)	6 (10,5)
3.6: comparação de alguns valores de ambas amostras	28 (49,1)	20 (35,1)
<u>Incorretas</u>		
3.7: comparação de todos valores de ambas amostras	6 (10,5)	2 (3,5)
3.8: outra	4 (7,0)	2 (3,5)

Os resultados apresentados na tabela 6 mostram existir três tipos de resposta dos estudantes: corretas, parcialmente corretas e incorretas. No rol das respostas corretas, definiram-se três categorias: 3.1: comparação das percentagens; 3.2: análise da variação dos

pontos no gráfico de dispersão; 3.3: determinação e análise do coeficiente de correlação e 3.4: comparação das médias.

Na estratégia 3.1: comparações das porcentagens alguns estudantes compararam os dois valores da pressão de sangue em cada mulher, correspondentes aos dois momentos, e basearam os seus julgamentos na proporção das porcentagens dos casos onde a pressão aumenta ou diminui. A resposta do estudante E20, no pré-ensino, exemplifica exatamente isso:

Nota-se que 8 das 10 mulheres (80% das mulheres) registaram um aumento das suas pressões sanguíneas depois do tratamento médico. As outras 2, das 10 (20% das mulheres), reduziram a sua pressão sanguínea depois do tratamento. Isto mostra que o tratamento faz aumentar a pressão sanguínea nos pacientes.

Como se pode observar pelas frequências e porcentagens de respostas, houve um ligeiro aumento do número de estudantes que responderam corretamente do pré para o pós-ensino, afirmando que existe dependência entre as variáveis pressão sanguínea e tempo em que foi medida.

Em relação às categorias 3.2: análise da variação dos pontos no gráfico de dispersão e 3.3: determinação e análise do coeficiente de correlação, observou-se que os estudantes chegaram às suas conclusões, no primeiro caso, a partir da representação dos pontos num diagrama de dispersão, ou seja, analisaram se os pontos tendiam ou não a ajustar-se a uma reta. Esta resposta foi obtida pelo estudante E29 no pós-ensino tal como ilustra a figura1:

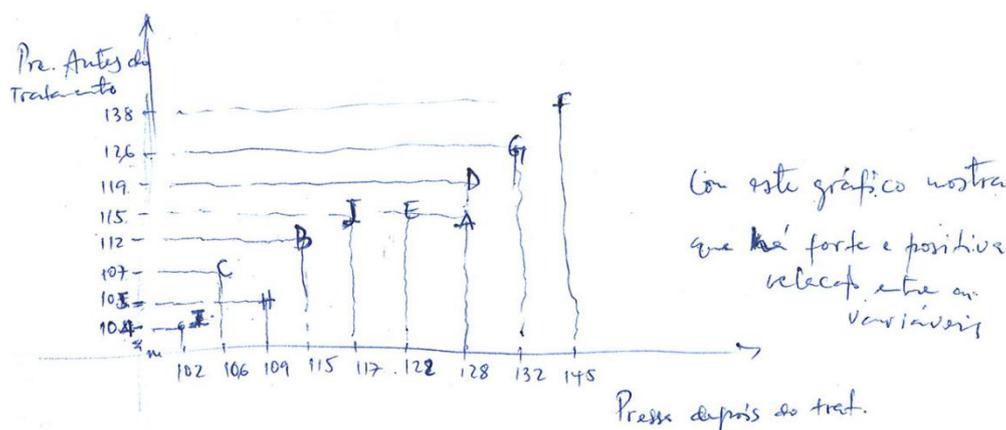


Figura 9. Resolução da questão 3 pelo estudante E29 no pós-ensino.

No segundo caso, determinaram o valor do coeficiente de correlação e avaliaram a sua intensidade, afirmando existir dependência se o coeficiente de correlação fosse diferente de zero, tal como mostra a resposta dada pelo estudante E29, no pós-ensino:

Depende do tempo em que foi medido, devido a existência de correlação muito forte

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \cdot \sqrt{n \sum y^2 - (\sum y)^2}} = \frac{10 \cdot 140354 - 1156 \cdot 1204}{\sqrt{10 \cdot 13459 - 1156^2} \cdot \sqrt{10 \cdot 146836 - 1204^2}}$$

$$r = \frac{1403540 - 1391824}{\sqrt{9564} \cdot \sqrt{15744}} \quad r = \frac{11716}{\sqrt{150575616}}$$

$$r = \frac{11716}{12270,92} = 0,95'$$

Figura 10. Resolução da questão 3 pelo estudante E29 no pós-ensino

Ambos os tipos de resposta, baseados nestas duas estratégias, foram observados apenas no pós-ensino.

No caso da estratégia 3.4: comparação das médias os estudantes afirmaram existir dependência entre as duas variáveis, justificando as suas afirmações com base na comparação das médias da pressão sanguínea em ambas amostras. Por exemplo, o estudante E53, no pós-ensino, respondeu: “Sim pois, em média, a pressão sanguínea aumentou com aplicação do tratamento médico. Concluiu-se, assim, que a pressão sanguínea destas mulheres depende do tempo do tratamento”. Aqui o estudante implicitamente usou a ideia correta de que a diferença das médias implica uma associação entre as variáveis.

Quanto às respostas parcialmente corretas, pela frequência de respostas, há que destacar a estratégia 3.6: comparação de alguns valores de ambas amostras, em que os estudantes admitiram haver dependência entre as variáveis a partir dos casos em que a pressão sanguínea subiu, tal como se constata na seguinte resposta: “A pressão sanguínea nesta amostra depende do tempo em que foi medida porque a pressão sanguínea subiu depois do tratamento na maior parte das mulheres” (E27, pós-ensino). Apesar do estudante ter dado uma resposta correta (existência de associação), ele seguiu um procedimento pouco fiável na medida em que avaliou a associação tendo por base apenas parte dos dados. Porém, para analisar a associação entre duas variáveis devem-se incluir todos os dados. Esta resposta foi a que mais se salientou tanto no pré ensino como no pós-ensino.

A outra estratégia que teve em conta a análise de apenas parte dos dados é a estratégia 3.5: análise de casos excepcionais. Nesta estratégia os estudantes responderam considerando os

casos em que a regra geral não era seguida, ou seja, os casos em que a pressão diminuía depois do tratamento. Considerámos esta estratégia parcialmente correta porque ela foi utilizada apenas para afirmar a associação, sem que os estudantes usassem percentagens, baseando os seus julgamentos em apenas alguns dados da distribuição, tal como se verifica na seguinte resposta:

Depende porque as frequências aumentaram depois do tratamento. Quase em todas mulheres A, B, C, D, E, F, G, H e J houve aumento, excepto na mulher I, que baixou a pressão do sangue depois do tratamento, de 104 para 102. (E19, pré-ensino)

A estratégia 3.7: comparação de todos valores de ambas amostras conduziu à resposta errada, uma vez que os estudantes compararam os valores correspondentes ao mesmo sujeito em ambas amostras, seguindo um raciocínio similar ao da estratégia 3.1, embora não tenham utilizado as percentagens. Os estudantes erraram nos seus julgamentos de associação por esperarem que todos os dados seguissem no mesmo sentido. A resposta dada pelo estudante E19, no pós-ensino, elucida claramente esse facto: “Não depende porque se dependesse do tempo ocorreria apenas aumento ou diminuição segundo uma proporcionalidade direta ou inversa”. Estepa et al. (1999) associam este tipo de estratégia à conceção determinista de associação porque os estudantes esperavam a existência de valores variando no mesmo sentido em todos os casos para cada amostra, ou seja, que crescesse ou diminuísse num mesmo sentido.

Para além das estratégias acima referidas, existiram também algumas respostas incluídas na categoria “Outra”. Nestas respostas, algumas vezes, as justificações eram incoerentes e não permitiram tirar qualquer ilação, enquanto noutras os estudantes deram respostas com base no conhecimento acerca do contexto da questão (condições para que haja aumento de pressão) ao invés de analisar os dados empíricos presentes no enunciado do problema, tal como mostra a seguinte resposta: “depende porque a pressão sanguínea envolve vários fatores, por exemplo o nervosismo pode ser um fator que influenciou na subida de pressão sanguínea” (E23, pré-ensino).

#### Questão 4

Na tabela seguinte indicam-se as taxas anuais de casamento e divórcio, por cada 1000 pessoas, em alguns países europeus.

País	Taxa de casamento	Taxa de divórcio
Áustria	5,6	2,0
Dinamarca	6,0	3,0
Finlândia	5,1	2,9
França	5,0	1,9
Alemanha	6,7	2,0
Hungria	6,3	2,4
Itália	5,4	0,4
Holanda	6,1	1,9
Noruega	4,9	2,2
Polónia	6,8	1,3
Suécia	5,2	2,2
Suíça	6,8	2,0
Reino Unido	6,1	2,9

Achas que existe ou não associação entre as taxas de casamento e de divórcio, isto é, que a taxa de divórcio dependa da taxa de casamento? Explica a tua resposta.

Esta questão, tal como a anterior, diz respeito à comparação de duas amostras emparelhadas. No entanto, ao contrário da questão anterior, em que se comparava duas variáveis de um mesmo contexto em duas etapas (pressão sanguínea antes e depois do tratamento), esta questão trata de comparar variáveis de contextos diferentes isto é, a taxa de divórcio com a taxa de casamento. As frequências e as percentagens de respostas corretas, parcialmente corretas e incorretas, no pré e pós-ensino, são apresentados na tabela 8.

Tabela 9. Frequência absoluta (percentagem) de estudantes nas diferentes estratégias da questão 4

Respostas	Frequências (%)	
	Pré-ensino	Pós-ensino
<u>Corretas</u>		
4.1: determinação e análise do coeficiente de correlação	0 (0,0)	29 (50,9)
<u>Parcialmente corretas</u>		
4.2: comparação de alguns valores de ambas as amostras	31 (54,4)	15 (26,3)
4.3: comparação das diferenças entre a taxa de casamento e de divórcio em todos países	8 (14,0)	0 (0,0)
<u>Incorretas</u>		

4.4: comparação de todos os valores de ambas as amostras	7 (12,3)	3 (5,3)
4.5: relação de causa e efeito	6 (10,5)	8 (14,0)
4.6: outra	5 (8,8)	2 (3,5)

Considerando como estratégia correta àquela em que se relaciona ambas as variáveis com base na determinação e análise do coeficiente de correlação, observa-se pela tabela que nenhum estudante utilizou esta estratégia no pré-ensino, tendo sido usada apenas no pós-ensino por 29 estudantes, o que corresponde a 50,9% da totalidade dos estudantes. Por exemplo, a resposta do estudante E29 no pós-ensino elucida este facto, na medida em que justificou a não associação como base na determinação e análise do coeficiente de correlação.

$$r = \frac{n \sum x_i y_i - \sum x_i \cdot \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \cdot \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

$$r = \frac{13 \cdot 158,11 - 76 \cdot 27,11}{\sqrt{13 \cdot 450,06 - 76^2} \cdot \sqrt{13 \cdot 62,33 - 27,1^2}}$$

$$r = \frac{-9,117}{75,3415} = r = -0,055 \approx -0,06$$

Como podemos ver pelo resultado do coeficiente de correlação, a relação é muito fraca ou quase inexistente pois o  $r$  está muito próximo de zero.

Figura 11. Resolução da questão 3 pelo estudante E29 no pós-ensino

Em relação às respostas parcialmente incorretas, constatámos dois tipos de estratégias, especificamente: 4.2: comparação de alguns valores de ambas as amostras e 4.3: comparação das diferenças entre a taxa de casamento e de divórcio em todos países. Quanto a primeira estratégia, alguns estudantes justificavam a não existência de dependência entre as duas variáveis pelo facto de em alguns países não se verificar uma variação conjunta das taxas no mesmo sentido, ou seja, à medida que a taxa de casamento aumenta, a taxa de divórcio devia aumentar também, e reciprocamente. Por exemplo, o estudante E43 no pré-ensino respondeu do seguinte modo:

Não, porque olhando para taxa do casamento na Áustria é de 5,6 e de divórcio é 2,0, mas na Finlândia a taxa de casamento desce em relação a Áustria para 5,1, enquanto a taxa de divórcio da Finlândia sobe em 2.9 em relação a Áustria.

Podemos usar o mesmo raciocínio para compararmos a Finlândia com a França, não se verificando tal fenómeno.

Esta estratégia foi a predominante no pré-ensino, com 31 estudantes a adotá-la, correspondente à 54,4% da totalidade dos estudantes. Já no pós-ensino registou-se um decréscimo 16 estudantes, correspondente a 28,1% do total de estudantes.

Quanto a segunda estratégia, os estudantes optaram por justificar as suas respostas socorrendo-se das diferenças entre a taxa de casamento e de divórcio registadas nos diferentes países, assumindo que não há dependência entre a taxa de divórcio e a taxa de casamento pelo facto de se registarem discrepâncias nas diferenças calculados. Esta estratégia, apesar de conduzir a uma resposta correta (independência), está acompanhada de certas incorreções pelo facto de assumirem que estas pequenas diferenças, por um lado, e as diferenças relativamente superiores, por outro, eram sinónimo da tradução de uma certa independência. Esta estratégia é exemplificada pela resposta do estudante E21, pré-ensino: “Eu penso que a taxa de divórcio não depende da taxa de casamento porque em alguns casos a diferença entre a taxa de casamento e de divórcio é bastante elevada e noutros não é”.

No que diz respeito às respostas incorretas, constatámos três estratégias: 4.4: comparação de todos os valores de ambas amostras; 4.5: Uso de teorias prévias e 4.6: outra. A primeira estratégia consistiu na comparação de cada valor da variável taxa de casamento com o correspondente da taxa de divórcio, para todos os elementos de ambas as amostras. Consideramos esta estratégia incorreta na medida em que nela se considera a associação estatística como uma relação entre os valores das variáveis que se verifica sempre, o que, no entanto, não garante tal associação estatística. Os estudantes que adotaram esta estratégia destacaram que em todos os países existia sempre uma taxa de casamento superior à taxa de divórcio, levando-os a concluir pela existência de associação estatística entre as variáveis. A resposta seguinte exemplifica o que acabamos de referir: “A taxa de divórcio depende da taxa de casamento porque em todos países a taxa de casamento é maior que a taxa de divórcio” (E21, no pré-ensino).

Em relação a estratégia 4.5: uso de teorias prévias, tal como a anterior, verificou tratar-se de uma estratégia pouco adequada para avaliar a associação estatística na medida em que os estudantes argumentaram que para haver divórcio era necessário que primeiro tivesse havido casamento, o que é óbvio. Portanto, neste caso, eles utilizaram suas próprias teorias sobre casamento e divórcio e não recorreram a qualquer dado do problema ou outro objeto

matemático. Por exemplo, o estudante E1, no pós-ensino, deu a seguinte resposta: “Sim existe uma associação entre as taxas de casamento e de divórcio porque o divórcio é consequência do casamento, ou seja, não há divórcio sem que haja casamento”. Nesta estratégia, do pré para o pós-ensino, houve um ligeiro aumento (de 2 estudantes), o que mostra alguma resistência na mudança da conceção de associação estatística.

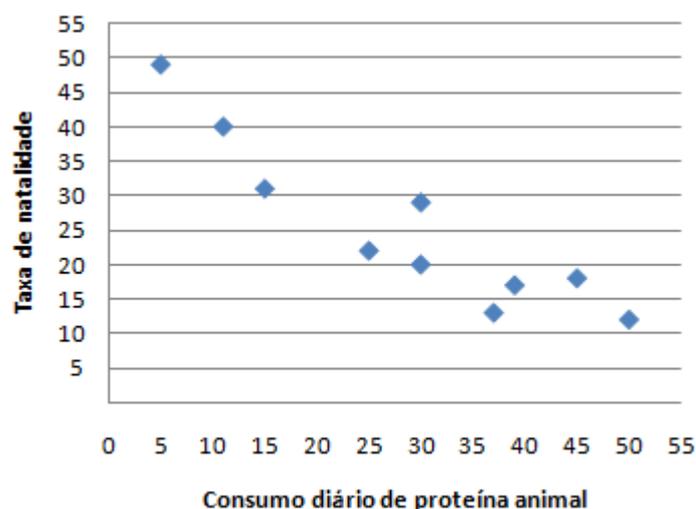
Para além das estratégias acima referidas, existiram também algumas respostas incluídas na categoria “Outra”. Nestas respostas, algumas vezes, as justificações eram incoerentes e não permitiram retirar delas o raciocínio que estava detrás das respostas e, outras vezes, não se associavam a quaisquer objetos matemáticos. Por exemplo, o estudante E8, no pré-ensino, respondeu do seguinte modo: “Não, porque a taxa depende de cada país, ou seja, da importância que as pessoas de um país dão ao casamento”.

Fazendo a análise global das respostas dos estudantes nesta questão, em termos da evolução das estratégias do pré para o pós-ensino, podemos inferir que houve alguma influência do ensino na mudança das conceções de associação estatística, como seria de esperar. No entanto, registou-se também alguma resistência em utilizar procedimentos mais institucionalizados por parte de outros estudantes, que mesmo depois de terem tido aulas sobre a determinação e interpretação da associação estatística entre duas variáveis não as utilizaram.

#### 4.1.3. Avaliação da associação estatística num diagrama de dispersão

##### Questão 5

Num estudo sociológico foram recolhidos dados relativos ao consumo diário de proteína animal e à taxa de natalidade em diferentes países. No gráfico seguinte estão representados os dados obtidos.



- a)** Em quantos países foram recolhidos dados relativos ao consumo diário de proteína animal e à taxa de natalidade?
- b)** Observando o gráfico, que tendência existe entre o consumo diário de proteína animal e a taxa de natalidade?
- c)** Achas que a taxa de natalidade nos diferentes países depende do consumo diário de proteína animal?  
Explica a tua resposta.

Na presente questão, na alínea a) estuda-se a capacidade de leitura e interpretação do gráfico de dispersão por parte dos estudantes, na alínea b) os estudantes são inquiridos sobre o reconhecimento da tendência de variação entre as variáveis e na alínea c) os estudantes devem avaliar a existência de associação entre as variáveis. Seguidamente apresenta-se a análise efetuada às respostas dos estudantes em cada uma das alíneas da questão.

- a)** Em quantos países foram recolhidos dados relativos ao consumo diário de proteína animal e à taxa de natalidade?

Na alínea a) pretendia-se saber a capacidade dos estudantes em fazer a leitura de um gráfico de dispersão, mais particularmente a interpretação dos pontos do gráfico de dispersão, ou seja, o que é que os pontos representam tendo em conta o problema em estudo. Os resultados mostram uma boa capacidade dos estudantes na leitura dos pontos do gráfico de dispersão, tendo-se obtido uma frequência (percentagem) de respostas corretas muito maiores do que de incorretas, tanto no pré-ensino 43 (75,4%) como no pós-ensino 52 (91,2%). Por outro lado, estes valores indicam também um aumento considerável das respostas corretas do pré-ensino para o pós-ensino.

Curcio (1989) definiu três níveis, de complexidade crescente, na leitura e interpretação de gráficos: primeiro, *ler os dados*; segundo, *ler entre os dados*; e terceiro, *ler além dos dados*. No nível *ler os dados* é necessário que o leitor faça uma leitura literal do gráfico, que se realiza através da leitura dos factos que nele estão representados; no nível *ler entre os dados* é necessário combinar e integrar a informação e identificar relações matemáticas através de algum conhecimento prévio sobre o assunto tratado no gráfico; e no nível *ler para além dos dados* pressupõe-se a capacidade de efetuar previsões a partir da informação do gráfico e um conhecimento prévio aprofundado sobre o assunto referente aos dados do gráfico. No caso da pergunta colocada aos estudantes, ela situa-se no nível *ler entre os dados* pois ela requer a determinação da dimensão da amostra em estudo, a qual não é diretamente observável a partir do diagrama de dispersão.

b) Observando o gráfico, que tendência existe entre o consumo diário de proteína animal e a taxa de natalidade?

A pergunta seguinte alínea b), tinha como objetivo reconhecer a tendência de variação existente entre as duas variáveis, ou seja, o sentido da variação de ambas variáveis (variação no mesmo sentido, variação em sentido contrário ou ausência de variação). Para o efeito, os estudantes podiam calcular o coeficiente de correlação, ou ainda de forma intuitiva ajustar uma reta ao gráfico de dispersão.

Em termos de resultados, verificou-se que a maior parte dos estudantes usou um raciocínio correto, na medida em que ajustaram uma reta à nuvem de pontos. Uns fizeram uma representação mais intuitiva, particularmente no pré-ensino, e outros mais concreta através da representação da reta no diagrama, isto no pós-ensino, tal como mostra a seguinte resposta do estudante E7:

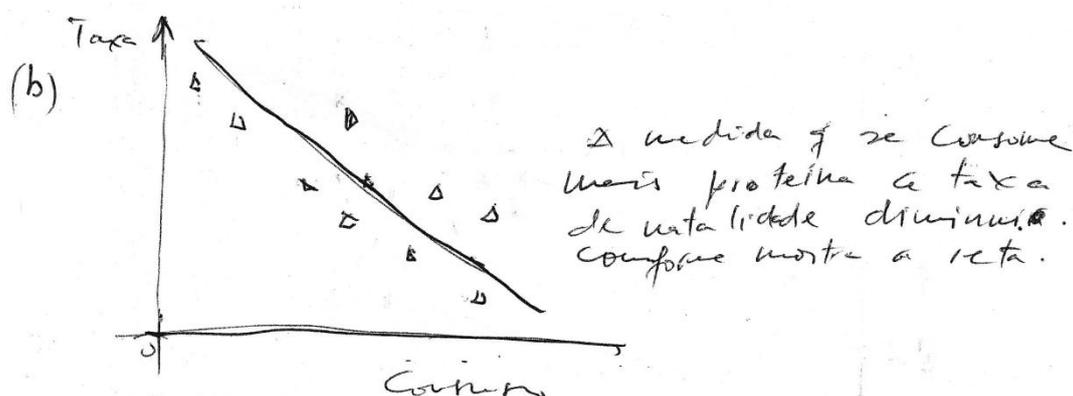


Figura 12. Resolução da questão 5b pelo estudante E7 no pós-ensino.

Este raciocínio resultou na resposta correta de que a tendência que existe entre as duas variáveis (consumo diário de proteína animal e taxa de natalidade) é decrescente, ou seja, à medida que o consumo de proteína animal aumenta, a taxa de natalidade diminui. Esta resposta foi verificada em 50 (87,7%) estudantes no pré-ensino e 57 (100%) no pós-ensino, que corresponde à grande maioria dos estudantes no pré-ensino e a todos no pós-ensino.

Quanto às respostas incorretas, verificadas apenas no pré-ensino, encontram-se aquelas em que os estudantes consideraram não haver uma tendência clara dos pontos de dispersão, uma vez que é possível encontrar um mesmo valor para países diferentes, tal como mostra a resposta do estudante E30 no pré-ensino: “Não existe nenhuma tendência notável porque alguns países possuem o mesmo consumo mas com taxas de natalidade diferentes, outros ainda têm maior consumo [mas] menor taxa de natalidade e vice-versa”. Com este exemplo, pode-se

depreender que o estudante tem uma perspectiva determinista de associação pois admite a não independência pelo facto de não haver uma correspondência unívoca entre valores de ambas as variáveis (Estepa & Batanero, 1996).

**c)** Achas que a taxa de natalidade nos diferentes países depende do consumo diário de proteína animal?

Com a alínea c) pretendia-se averiguar quais eram as estratégias dos estudantes quanto aos julgamentos de associação entre duas variáveis numéricas a partir de dados representados num diagrama de dispersão. É de referir que esta pergunta insere-se num dos objetivos centrais da presente pesquisa, isto é, estudar as conceções dos estudantes sobre a associação estatística e regressão lineares quando a distribuição das variáveis é representada através de um gráfico de dispersão, antes e depois da instrução. As frequências e percentagens das distintas estratégias adotadas pelos estudantes estão descritas na tabela 8.

Tabela 10. Frequência absoluta (percentagem) de estudantes nas diferentes estratégias da questão 5c)

Respostas	Frequências (%)	
	Pré-ensino	Pós-ensino
<u>Corretas</u>		
5c).1: tendência de variação dos pontos do gráfico de dispersão (variação no mesmo sentido, variação em sentido contrario ou ausência de variação)	27 (47,4)	30 (52,6)
<u>Parcialmente corretas</u>		
5c).2: ajustamento de uma reta à nuvem de pontos	0 (0,0)	22 (38,6)
<u>Incorretas</u>		
5c).3: utilização de pontos isolados do gráfico de dispersão	22 (38,6)	4 (7,0)
5c).4: causalidade	8 (14,0)	1 (1,8)

Como pode-se ver na tabela 9, a estratégia mais utilizada, tanto no pré-ensino como no pós-ensino, foi a estratégia 5c).1: tendência de variação dos pontos do gráfico de dispersão (crescente, decrescente e ausência), respetivamente com 27 (47,4%) estudantes no pré-ensino e 30 (52,6%) no pós-ensino. Para justificar o tipo de associação que existe entre as variáveis, alguns estudantes optaram por analisar a tendência dos pontos no gráfico de dispersão, isto é, a tendência de variação ou a ausência de qualquer tendência de variação dos pontos do diagrama de dispersão. Deste modo, os estudantes assumiram a existência de dependência entre o consumo de proteína animal e a taxa de natalidade pelo facto de ambas variáveis variarem

conjuntamente em sentidos diferentes, ou seja, à medida que o consumo diário de proteína animal aumenta, a taxa de natalidade diminuiu. A resposta do estudante E2, no pré-ensino, a seguir apresentada, elucida uma concepção correta de associação: “De acordo com o gráfico, a taxa de natalidade nos diferentes países depende do consumo diário de proteína animal porque quando a taxa de natalidade aumenta o consumo diário de proteína diminui”. Esta resposta mostra que o estudante possui uma concepção correta de associação, pois no caso de independência não haveria uma variação conjunta.

No caso da estratégia 5c).2: ajustamento de uma reta à nuvem de pontos, alguns estudantes afirmaram existir associação entre as variáveis com base no ajustamento de uma reta à nuvem de pontos do diagrama de dispersão, tal como mostra a resposta dada pelo estudante E54, no pós-ensino: “A taxa de natalidade depende do consumo diário de proteína porque os pontos tendem a uma reta, daí que há uma associação entre as duas variáveis”. Nesta estratégia, o estudante utilizou intuitivamente a ideia de que quando há um bom ajustamento do gráfico de uma reta à nuvem de pontos do diagrama de dispersão, então as duas variáveis estão correlacionadas. A utilização do ajustamento do gráfico da reta ao gráfico de dispersão garante a existência de correlação linear, mas já não adequada para avaliar a existência de uma correlação não linear. Neste último caso, o não ajustamento do gráfico de uma reta à nuvem de pontos não significa que não possa existir uma curva que produza esse ajustamento — por exemplo, o gráfico de uma função quadrática, o que evidenciaria a existência de uma correlação quadrática. Esta estratégia foi adotada apenas no pós-ensino por 22 estudantes, o que corresponde a 38,6%.

As estratégias 5c).3 e 5c).4, apresentadas na tabela 9, conduziram a respostas incorretas porque os estudantes, por um lado, utilizaram parte dos dados do gráfico de dispersão e, por outro, recorreram a outras variáveis causais que pudessem explicar a variável dependente, ao invés de utilizarem os dados disponíveis no gráfico.

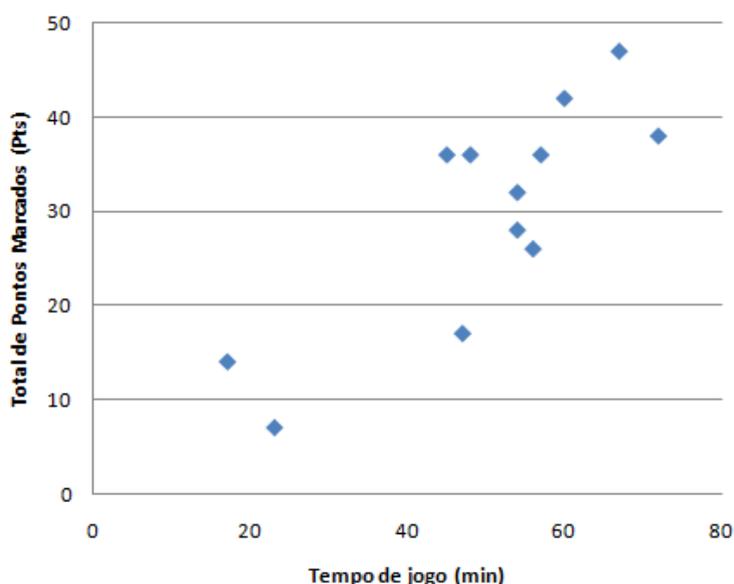
A estratégia 5c).3 consistiu em deduzir a associação entre as variáveis a partir de pontos isolados. Por exemplo, o estudante E33, no pré-ensino, respondeu que não existe dependência entre as variáveis pelo facto de existirem dois países com um mesmo valor de consumo de proteína animal: “A taxa de natalidade nos diferentes países não depende do consumo diário de proteína animal porque, por exemplo, o consumo diário de proteína animal de 30 tem uma taxa de natalidade de 20 e 30 ao mesmo tempo”. Esta resposta, para além de constituir um indicador de uma concepção local de associação, pode também ser um indicador de uma

conceção determinística de associação, ao considerar-se que para haver associação era necessário que a cada valor da variável consumo diário de proteína devia corresponder apenas um único valor da variável taxa de natalidade (Estepa & Batanero, 1996).

Em relação a estratégia 5c).4: causalidade, alguns estudantes afirmaram não existir associação entre as variáveis porque a variável independente não influencia diretamente a variável dependente, ou seja, admitiram existência de outras variáveis que podiam influenciar a variável dependente, tal como mostra a resposta do estudante E40, no pré-ensino: “Não, porque mesmo que não sejam consumidos alimentos que contenham proteína animal sempre existirá nascimentos, portanto a proteína animal não influencia na taxa de natalidade”. Esta resposta é um indicador de uma conceção causal de associação (Estepa & Batanero, 1996).

### Questão 6

No gráfico seguinte estão representados os pontos obtidos por cada jogador da seleção de basquetebol dos Estados Unidos da América nos Jogos Olímpicos de Inverno de 1992 e o seu tempo de jogo em minutos, no conjunto dos três jogos finais.



- a) Assinala no gráfico o ponto que representa o jogador que marcou mais pontos. Achas que esse jogador também jogou durante mais tempo?
- b) Achas que existe uma relação entre o tempo de jogo (em minutos) do jogador e o total de pontos marcados? Explica a tua resposta.
- c) Se um jogador joga durante 50 minutos, quantos pontos é de esperar que ele marque?

Tal como na questão 5, na questão 6 interessava-nos averiguar a capacidade dos estudantes, por lado, em fazer a leitura do gráfico de dispersão, com particular destaque dos eixos (alínea a), e, por outro, pretendia-se analisar o tipo de julgamento de associação que os estudantes fazem entre duas variáveis quantitativas apresentadas no gráfico de dispersão.

(alínea b). Além disso, na alínea c) é pedido aos estudantes que estimem o valor de uma das variáveis a partir do conhecimento do valor da outra variável. Seguidamente apresenta-se a análise efetuada às respostas dos estudantes em cada uma das alíneas da questão.

- a)** Assinala no gráfico o ponto que representa o jogador que marcou mais pontos. Achas que esse jogador também jogou durante mais tempo?

Na alínea a) pretendia-se avaliar a capacidade dos estudantes em fazer a leitura do gráfico de dispersão, mais especificamente a identificação do jogador que marcou mais pontos. Os resultados obtidos tanto no pré como no pós-ensino foram satisfatórios, na medida em que mais de 90 por cento dos estudantes (96% no pré-ensino e 98,2% no pós-ensino) conseguiram identificar o jogador que marcou mais pontos. Também na sua grande maioria (93% no pré-ensino e 98,2% no pós-ensino), os estudantes reconheceram também que este jogador, apesar de ter marcado mais pontos, não foi o jogador com mais tempo em campo.

Em termos dos diferentes níveis de leitura e interpretação de um gráfico, de Curcio (1989), qualquer destas duas perguntas se situa no nível de *ler os dados* pois a resposta pode ser obtida diretamente a partir da informação explícita no gráfico de dispersão.

- b)** Achas que existe uma relação entre o tempo de jogo (em minutos) do jogador e o total de pontos marcados? Explica a tua resposta.

Os resultados obtidos nesta questão demonstram uma variedade de respostas dos estudantes na identificação e justificação da associação entre as duas variáveis: o tempo de jogo (em minutos) e o número de pontos marcados. Na tabela 10 apresentam-se as frequências (percentagens) obtidas nos diferentes tipos de resposta (correta, parcialmente correta e incorreta), antes e depois da instrução.

Tabela 11. Frequência absoluta (percentagem) de estudantes nas diferentes respostas da questão 6b

Respostas	Frequências (%)	
	Pré-ensino	Pós-ensino
<u>Corretas</u>		
6b).1: tendência de variação dos pontos do gráfico de dispersão (variação no mesmo sentido, variação em sentido contrario ou ausência de variação).	18 (31,6)	11 (19,3)
6b).2: ajustamento de uma reta à nuvem de pontos.	0 (0,0)	10 (17,5)
<u>Parcialmente correta</u>		
6b).3: avaliação global da dispersão da nuvem de pontos.	0 (0,0)	23 (40,4)

#### Incorretas

6b).4: utilização de pontos isolados.	26 (45,6)	6 (10,5)
6b).5: atribuição da associação a outras variáveis distintas das explicitadas no problema.	9 (15,8)	4 (7,0)
6b).6: teorias prévias.	4 (7,0)	3 (5,3)

---

No conjunto das distintas estratégias apresentadas, constatamos que as duas seguintes estratégias conduziram à resposta correta: 1: tendência de variação dos pontos do gráfico de dispersão (variação no mesmo sentido, variação em sentido contrário e ausência de variação); e 2: ajustamento de uma reta à nuvem de pontos.

Em relação a estratégia 6b) 1, alguns estudantes julgaram existir ou não associação entre as duas variáveis através da análise da tendência de variação ou ausência de variação dos pontos do diagrama de dispersão. Assim, consideraram existir associação pelo facto de ambas as variáveis variarem conjuntamente no mesmo sentido, tal como mostra a resposta do estudante E2 no pré-ensino: “Sim, depende porque, de uma forma geral, o jogador com mais pontos marcados tende a ser aquele que jogou durante mais tempo”. Esta resposta baseia-se numa conceção correta de associação, e verificou-se uma diminuição de 7 estudantes do pré-ensino para o pós-ensino na adoção desta estratégia.

Na estratégia 6b) 2, para justificar a associação entre as variáveis, os estudantes avaliaram o ajustamento de uma reta à nuvem de pontos. Por exemplo, o estudante E29, no pós-ensino, respondeu o seguinte: “Sim porque a maneira como os dados se apresentam no gráfico mostram no fundo a ideia de existência de uma linha reta crescente”. Esta estratégia foi adotada por 10 (17,5%) estudantes, apenas no pós-ensino.

Quanto à estratégia 6b) 3, os estudantes partiram do pressuposto de que a maior ou a menor dispersão dos pontos implica uma fraca ou forte associação entre as variáveis, respetivamente. A resposta do estudante E8, no pós-ensino, exemplifica o que acabamos de referir: “Existe uma relação porque os pontos apresentam uma menor dispersão”. Esta justificação, adotada por 23 (40,4%) estudantes, apenas no pós-ensino, corresponde a uma resposta parcialmente correta uma vez que aludiram apenas a dispersão, sem um referente (uma curva) dessa dispersão.

As estratégias 6b).4, 6b).5 e 6b).6, apresentadas na tabela 10, produziram respostas incorretas na medida em que se basearam, por um lado, em parte dos dados para justificar a associação ou nas suas crenças subjetivas acerca da relação entre as variáveis em detrimento

dos dados do problema e, por outro lado, por considerarem a inexistência de associação pelo facto de existirem outras variáveis que pudessem influenciar a variável dependente.

A estratégia 6b).4 consistiu em deduzir a associação a partir de pontos isolados. Por exemplo, o estudante E3, no pré-ensino, considerou não existir associação entre o número de pontos marcados e o tempo de jogo pelo facto de se observar em alguns jogadores discrepâncias entre essas duas variáveis, ou seja, marcaram muitos pontos tendo jogado poucos minutos e vice-versa.

Não porque existem jogadores que jogaram muito mais tempo e marcaram poucos pontos e outros que jogaram poucos minutos e marcaram muitos pontos. Por exemplo, o jogador que jogou apenas 60 minutos marcou 43 pontos e o jogador que fez 70 minutos marcou aproximadamente 38 pontos. (E3, pré-ensino)

Nesta resposta o estudante mostra claramente que possui uma conceção local de associação, que provavelmente o terá conduzido a uma resposta incorreta de associação, neste caso de independência. Esta estratégia foi a mais utilizada pelos estudantes no pré-ensino, com 26 (45,6%) estudantes a adotá-la para responderem à questão 6b). Do pré-ensino para o pós-ensino registou-se uma diminuição assinalável do número de estudantes que adotaram esta estratégia, exatamente 20 estudantes, o que corresponde a uma diminuição de 35,1%.

Tal com na estratégia 5c).4: causalidade, a estratégia 6b).5: atribuição da associação a outras variáveis distintas das explicitadas no problema teve como base pressupostos causais, na medida em que os estudantes afirmaram não existir associação entre as variáveis porque a variável independente não influencia diretamente a variável dependente, ou seja, admitiram a existência de outras variáveis que podiam influenciar a variável dependente, tal como mostra a resposta do estudante E22, no pré-ensino: “Não há relação entre o tempo de jogo e o total de pontos marcados, isto porque a marcação de pontos depende da flexibilidade do jogador e da sua estratégia dentro do campo”. Esta estratégia, reveladora de um raciocínio causal por parte do estudante, foi adotada por um número menor de estudantes, tanto no pré-ensino como no pós-ensino, respetivamente 9 (15,8%) e 4 (7,0%) estudantes.

Finalmente, na estratégia 6b).6: teorias prévias os estudantes argumentaram as suas respostas com base nas suas crenças prévias acerca da relação entre o número de pontos marcados e o tempo de jogo, ao invés de analisar as evidências dos dados disponíveis no problema. A resposta do estudante E9, no pós-ensino, é reveladora disso mesmo: “É claro que sim, porque se um jogador permanecer muito tempo no campo irá marcar muitos pontos em relação ao jogador que permanece pouco tempo no campo”. O recurso a ideias prévias para

justificar a associação observou-se em apenas 4 (7,0%) estudantes no pré-ensino e 3 (5,3%) no pós-ensino. Chapman e Chapman (1969) designam este tipo de raciocínio por “correlação ilusória” porque as pessoas afirmam as suas crenças ao invés de ter em consideração as evidências dos dados

**c)** Se um jogador joga durante 50 minutos, quantos pontos é de esperar que ele marque?

Na alínea c) pretendia-se avaliar a capacidade dos estudantes em prever o valor de uma variável a partir do conhecimento do valor da outra variável. No caso concreto, pretendia-se estimar o número de pontos marcados a partir do conhecimento do tempo de jogo. Para o efeito, o estudante podia recorrer à reta de regressão e utilizar retas paralelas aos eixos coordenados para estimar o ponto correspondente a 50 minutos.

Constatámos que as respostas à alínea c), particularmente no pré-ensino, eram maioritariamente dadas na forma intervalar, ou seja, estimaram em termos médios o número de pontos marcados pelo jogador que tinha jogado 50 minutos. Assim, as respostas dos estudantes agruparam-se em três intervalos: i) intervalo de 10 a 20 pontos; ii) intervalo de 20 a 30 pontos e iii) intervalo de 30 a 40 pontos.

A maior parte dos 22 estudantes, o que corresponde a 38,6% de todos estudantes, respondeu que o jogador marcaria mais ou menos entre 20 e 30 pontos. O facto de nesta região passar a reta de ajustamento à nuvem de pontos pode indiciar que os estudantes, apesar de não terem dado uma resposta precisa, tiveram a capacidade de estimar uma região onde se encontrava o valor da previsão pedida. Recorrendo intuitivamente ao conceito de monotonia de uma reta, revela-se também que os estudantes ao estimarem a região tiveram como base a tendência de variação das variáveis, ou seja, se elas variavam no mesmo sentido, em sentidos contrários ou a ausência de um padrão de variação.

Já relativamente aos dois outros intervalos, verificou-se que para 12 (21,1%) estudantes o valor se situava no intervalo de 10 a 20 pontos e para 6 (10,5%) estudantes o valor se situava no intervalo de 30 a 40.

Contrariamente à resposta anterior, 17 (29,8%) estudantes responderam que era difícil determinar o valor pretendido porque as variáveis não estavam correlacionadas linearmente, tal como mostra a resposta do estudante E55 no pós-ensino: “Não havendo relação linear entre o tempo de jogo e o número de pontos marcados não se pode estimar o número de pontos que um jogador que teve 50 minutos em campo pode marcar”. Presume-se que as respostas dadas por estes estudantes foram muito condicionadas pela resposta dada pelos mesmos na alínea b),

pois a maior parte dos estudantes que deram esta resposta afirmaram na alínea b) que não existia relação entre as variáveis pelo facto de existirem outros fatores diferentes do tempo que podem influenciar a variável dependente (número de pontos marcados).

Em relação ao pós-ensino, contrariamente às nossas expectativas, constatou-se que apenas 14 (24,6%) estudantes responderam corretamente à questão. Estes estudantes, por observação do diagrama de dispersão, estimaram os valores das variáveis e definiram, de seguida, a reta de regressão para estimar o número de pontos marcados pelo jogador, sendo conhecido o seu tempo (em minutos) de jogo. O exemplo a seguir elucida o que acabamos de referir.

Tempo de jogo (x)	17	23	45	47	48	55	55	56	58	60	67	72
Pts. marcados (y)	14	7	37	17	37	28	32	26	37	43	47	38

$$\bar{y} = a + b \bar{x} \quad \bar{x} = 50,25 \quad b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$a = \bar{y} - b \bar{x} \quad \bar{y} = 30,25$$

$$a = 30,25 - 0,613 \cdot 50,25$$

$$a = 30,25 - 30,80$$

$$a = -0,55$$

$$b = \frac{12 \cdot 2006 - 603 \cdot 363}{12 \cdot 33179 - 363 \cdot 609}$$

$$b = \frac{24072 - 218899}{398148 - 363609}$$

$$b = \frac{2183}{34539} = 0,613$$

$$y^* = a + b x$$

$$y^* = -0,55 + 0,613 x$$

Para  $x = 50 \Rightarrow y^* = -0,55 + 0,613 \cdot 50 \Rightarrow y^* = 30$   
 R.: Espera-se que o jogador marque 30 pontos.

Figura 13. Resolução da questão 6c) pelo estudante E16 no pós-ensino.

Praticamente metade dos estudantes (27, isto é, 47,4%) responderam apresentando estimativas na forma intervalar. De entre estes estudantes, tal como no pré-ensino, 17 (29,8%) responderam que o jogador marcaria mais ao menos 20 a 30 pontos, pelas razões apontadas anteriormente, 1 (1,8%) estudante indicou que o jogador marcaria entre 10 a 20 pontos e 9 (15,8%) estudantes responderam que o jogador marcaria entre 30 a 40 pontos.

Os restantes 16 (28,1%) estudantes, em número semelhante ao pré-ensino, persistiram em afirmar que não era possível prever o valor pedido pelas razões apontadas anteriormente, isto é, porque as variáveis não estavam relacionadas.

Portanto, de acordo com os resultados acima apresentados, concluímos que apesar de os estudantes terem passado pelo ensino formal de conteúdos sobre associação estatística e regressão linear, verificamos que a maior parte deles mantiveram as suas estratégias e

concepções, baseadas fundamentalmente na leitura dos pontos do diagrama de dispersão e na identificação da relação de linearidade entre as variáveis ou da sua não existência. Do pré-ensino para o pós-ensino, a maior alteração nas respostas dos estudantes verificou-se no recurso à reta de regressão para estimar o valor pretendido, com a conseqüente diminuição do recurso à estimação do valor através de intervalos.

Tanto a questão 5 como a questão 6 envolvem problemas que dizem respeito à associação estatística e correlação linear em variáveis quantitativas representadas num diagrama de dispersão. Nestas situações, o estudo das estratégias dos estudantes permitiu-nos relacioná-las com o tipo de respostas (respostas corretas, parcialmente corretas e incorretas) e constituíram-se como indicadores das concepções subjacentes às avaliações de associação estatística. De entre essas estratégias, porque conduziram a respostas corretas e parcialmente corretas, há que destacar as seguintes:

- i) tendência de variação dos pontos do gráfico de dispersão;
- ii) ajustamento da reta de regressão ao gráfico de dispersão.

Na mesma linha de análise, constatámos as estratégias e argumentos incorretos, tais como a utilização de pontos isolados do gráfico de dispersão e a utilização de causalidade e/ou existência de outras variáveis que influenciam na variável dependente, para justificar a independência. Essas estratégias permitiram inferir algumas concepções incorretas de associação estatística por parte dos estudantes a saber: concepção determinística de associação; concepção local; concepção causal de associação e concepção baseada em teorias prévias, que alguns autores designam de correlação ilusória.

#### **4.1.4. Evolução dos significados sobre associação estatística com o ensino**

Da análise feita às estratégias de avaliação da associação entre duas variáveis dicotómicas, apresentadas na forma de tabela de contingência de  $2 \times 2$ , conclui-se que nem sempre os estudantes utilizaram todos os dados relevantes do problema para decidirem sobre a existência ou não de associação. A utilização de apenas uma única célula, geralmente a de maior frequência absoluta, e a comparação de frequências absolutas e frequências relativas entre si foram estratégias mais frequentemente referidas no pré-ensino do que no pós-ensino. A redução no recurso a apenas uma ou duas células no pós-ensino foi compensada por um aumento no recurso a quatro células.

Globalmente, do pré-ensino para o pós-ensino, praticamente mantém-se a percentagem de respostas corretas. Além disso, em termos de proporção de respostas corretas e incorretas, constata-se que quando o estudante usa informação de uma única célula o número de respostas corretas é semelhante ao número de respostas incorretas (6/8 no pré-ensino e 5/5 no pós-ensino), diminui quando o estudante usa na sua resposta informação de duas células (4/18 no pré-ensino e 1/13 no pós-ensino) e aumenta quando o estudante usa na sua resposta informação das quatro células (11/2 no pré-ensino e 20/5 no pós-ensino). Assim, podemos concluir que quando é usada a informação de uma única célula, a tendência é responder ao acaso; quando é usada a informação de duas células, a tendência é responder erradamente; e quando é usada a informação de quatro células, a tendência é responder corretamente.

Considerando que o envolvimento de um maior número de células nas respostas dos estudantes corresponde a uma maior sofisticação dos seus raciocínios, o que por sua vez se traduziu no aumento da proporção de respostas corretas, podemos concluir que o ensino por que passaram os estudantes produziu progressos limitados a esse nível.

Quanto às estratégias utilizadas pelos estudantes para decidirem sobre a existência de dependência na comparação de uma variável numérica em duas amostras qualitativas dicotómicas ou na comparação de uma variável numérica em duas amostras relacionadas, verificou-se que no pré-ensino houve uma maior proporção de estudantes que utilizaram estratégias parcialmente corretas e incorretas. Estas estratégias estavam associadas sobretudo à utilização parcial dos dados de ambas distribuições. Portanto, os estudantes não perceberam que para analisarem uma possível associação entre duas distribuições era necessário envolver todos os dados de ambas distribuições. Este erro foi verificado novamente no pós-ensino, porém com menor frequência.

As estratégias que conduziram a respostas corretas no pré-ensino resumiram-se às seguintes: i) comparação das médias; ii) comparação dos totais; e iii) comparação das percentagens de aumentos e diminuições de uma distribuição para a outra. A percentagem de respostas corretas teve um aumento significativo no pós-ensino, sobretudo pela inclusão das estratégias de análise da variação dos pontos no gráfico de dispersão e da determinação e análise do coeficiente de correlação, que obviamente foram estudadas durante o ensino implementado da associação estatística.

Por último, no que concerne ao estudo das estratégias de associação no diagrama de dispersão, identificámos estratégias que conduziram a respostas corretas e parcialmente, que

são indicadores de concepções corretas de associação estatística, e estratégias incorretas. Tal como nas situações anteriores, verificou-se no pré-ensino um maior número de respostas incorretas em resultado do recurso dos estudantes a *pontos isolados* para decidirem o tipo de correlação. Já no pós-ensino verificou-se uma ligeira diminuição da percentagem de respostas incorretas, acompanhada do correspondente aumento das respostas corretas, a partir da adoção das estratégias *avaliação da tendência de variação dos pontos do gráfico de dispersão (variação no mesmo sentido, variação em sentido contrário ou ausência de variação)* e *comparação do diagrama de dispersão com o gráfico de uma função conhecida*.

Em ambas as tarefas, aqui incluídas, realça-se ainda a utilização pelos estudantes das suas crenças e expectativas sobre a relação entre as variáveis para decidirem sobre a existência ou não de associação, em detrimento da análise empírica dos dados apresentados no problema. Na adesão a este fenómeno, identificado por Chapman e Chapman (1969) nos seus estudos, e que designaram por *correlação ilusória*, as pessoas mantêm as suas crenças sobre a associação entre variáveis apesar da evidência empírica suportar a independência das variáveis.

## **4.2. Aquisições conceituais da associação estatística com o ensino**

Nesta secção avaliam-se as aquisições conceituais dos estudantes sobre a associação estatística depois de terem passado pelo ensino do tema Correlação e Regressão Linear. Para tal, foi aplicado aos estudantes um teste depois de terem sido lecionadas as aulas sobre a associação estatística e regressão linear. São exatamente as respostas dos estudantes às questões desse teste que são analisadas nesta secção.

Por fim, apresente-se uma síntese dos aspetos mais relevantes subjacentes às respostas dos estudantes na globalidade do teste, enfatizando as percentagens de respostas corretas e incorretas ao longo das várias questões.

### **4.2.2. Análise das respostas dos estudantes às questões do teste**

A seguir iremos analisar as respostas dos estudantes a cada uma das questões do teste.

#### **Questão 1**

Sabendo-se que duas variáveis estão correlacionadas positivamente, indique se é verdadeira ou falsa cada uma das seguintes afirmações:

- a)** Quando uma aumenta, a outra também aumenta. \_\_\_\_\_
- b)** Quando uma diminui, a outra aumenta. \_\_\_\_\_
- c)** Quando uma diminui, a outra diminui. \_\_\_\_\_

d) A relação entre as duas variáveis é do tipo linear. \_\_\_\_\_

Nesta questão indagam-se os estudantes sobre consequências de duas variáveis estarem correlacionadas positivamente, ou seja, concretamente decidirem sobre o sentido da variação de ambas as variáveis e o tipo de relação entre elas.

A correlação linear permite estudar em que medida duas variáveis estão associadas, isto é, de que forma a variação de uma variável afeta a variação da outra. Ela pode variar quanto ao sentido e quanto a intensidade. Quanto ao sentido, a correlação pode ser positiva e negativa. Uma das características da correlação positiva é que as variáveis covariam na mesma direção, ou seja, quando uma variável aumenta, a outra também aumenta, ou quando uma variável diminui, a outra também diminui (Fernandes, 2004).

Os princípios acima descritos têm sido muitas vezes utilizados pelos estudantes para a determinação do sentido de correlação, tal como Cobo (1998, p.193) descreve: “Estes procedimentos têm tido muita influência na determinação do sentido da correlação pelos estudantes”.

De entre as diferentes afirmações, a) e c) são verdadeiras e b) e d) são falsas. As frequências absolutas (percentagens) de estudantes nas opções de resposta (verdadeira e falsa) das diferentes alíneas estão apresentadas na tabela 11.

Tabela 12. Frequência absoluta (percentagem) de estudantes nas respostas à questão 1

Sabendo que duas variáveis estão correlacionadas positivamente, indique se é verdadeira ou falsa cada uma das seguintes afirmações:	Frequências (%)	
	Verdadeira	Falsa
1a) Quando uma aumenta, a outra também aumenta	54 (94,7)	3 (5,3)
1b) Quando uma diminui, a outra aumenta	7 (12,3)	50 (87,7)
1c) Quando uma diminui, a outra diminui	40 (70,2)	17 (29,8)
1d) A relação entre as duas variáveis é do tipo linear	42 (73,7)	15 (26,3)

De acordo com os resultados apresentados na tabela 11, verifica-se que 54 (94,7%) estudantes assinalaram a afirmação como sendo verdadeira. Assim, eles afirmaram que em duas variáveis correlacionadas positivamente, quando uma aumenta, a outra também aumenta. Não obstante a opção 1c) (quando uma diminui, a outra diminui) ter o mesmo significado que a opção 1a), verifica-se nesta última que apenas 40 (70,2%) estudantes deram a resposta correta, o que corresponde uma diminuição de 14 (24,5%) sujeitos. Os estudantes 17 (29,8%) que consideraram a afirmação 1c) como falsa, presumiram que pelo facto de a correlação ser

positiva, a covariação das variáveis não poderia ser negativa, que é uma ideia errada acerca da correlação positiva.

Esta ideia já foi referida antes, no questionário, quando os estudantes foram interrogados sobre a existência de relação entre fumar e ter bronquite, numa tabela de contingência  $2 \times 2$  (questão 1). Nesta situação, a maior parte dos sujeitos utilizaram preferencialmente a célula C1, que representava a presença simultânea das duas características: fumar e ter bronquite. Embora seja um dos casos favoráveis à associação, os estudantes não perceberam que a célula C4, que representa a ausência simultânea de ambas características: não fumar e não ter bronquite, tem um papel semelhante à célula C1 na associação. Resultados semelhantes foram também obtidos por Inhelder e Piaget (1955) no seu estudo sobre associação estatística em tabelas de contingência.

Na alínea 1b) 7 (12,3%) estudantes deram uma resposta incorreta, afirmando que em duas variáveis correlacionadas positivamente, quando uma diminui, a outra aumenta.

Na última alínea desta questão, alínea d), avalia-se se os estudantes são conscientes sobre o facto de as variáveis estarem correlacionadas positivamente nem sempre implica uma relação do tipo linear. Os resultados deste item mostram que 42 (73,7%) estudantes declaram que quando as variáveis estão correlacionadas positivamente, a relação entre as duas variáveis é do tipo linear. Julgamos que estes resultados são a consequência de o ensino do tema Correlação e Regressão Lineares, muitas vezes, se centrar quase exclusivamente na análise do coeficiente de correlação linear, para saber o grau de relação ou de dependência entre duas variáveis, e na análise da regressão linear, com vista a encontrar uma reta que melhor se ajuste à nuvem de pontos e que permita explicar de forma mais conveniente a variação de uma variável a partir da outra. Consequentemente, os outros tipos de coeficientes de correlação e de ajuste, nomeadamente os não lineares, são tratados de forma superficial ou mesmo ignorados.

## Questão 2

Se  $r$  é o coeficiente de correlação entre duas variáveis  $X$  e  $Y$ , indique se é verdadeira ou falsa cada uma das seguintes afirmações:

- a)** Se  $r = 0$ , as variáveis  $X$  e  $Y$  são independentes. \_\_\_\_\_
- b)** Se  $r = 0,6$ , a correlação entre as variáveis  $X$  e  $Y$  é o dobro do que quando  $r = 0,3$ .  
\_\_\_\_\_
- c)** Uma relação funcional entre as variáveis  $X$  e  $Y$  corresponde a um valor de  $r$  de  $+1$  ou  $-1$ . \_\_\_\_\_
- d)** O coeficiente de correlação pode interpretar-se como uma percentagem da variância.  
\_\_\_\_\_

Na questão 2 trata-se da relação do valor do coeficiente de correlação com o tipo de associação. Basicamente, procura-se avaliar a percepção dos estudantes sobre as principais propriedades do coeficiente de correlação: na alínea 2a) pretende-se conhecer a interpretação dos estudantes do valor 0 (zero) do coeficiente de correlação; na alínea 2b) questiona-se a relação que os alunos estabelecem entre a correlação e a proporcionalidade; na alínea 2c) pretende-se verificar se os estudantes estabelecem uma relação correta entre a dependência funcional e os valores máximos do coeficiente de correlação; e na alínea 2d) pretende-se saber como é que os estudantes relacionavam o coeficiente de correlação com o coeficiente de determinação.

Na tabela 12 apresentam-se as frequências (percentagens) de estudantes nas opções de resposta (verdadeira e falsa) das diferentes alíneas.

Tabela 13. Frequência absoluta (percentagem) de estudantes nas respostas à questão 2

Se $r$ é o coeficiente de correlação entre duas variáveis $X$ e $Y$ , indique se é <u>verdadeira</u> ou <u>falsa</u> cada uma das seguintes afirmações:	Frequências (%)	
	Verdadeira	Falsa
2a) Se $r = 0$ , as variáveis $X$ e $Y$ são independentes	52 (91,2)	5 (8,8)
2b) Se $r = 0,6$ , a correlação entre as variáveis $X$ e $Y$ é o dobro do que quando $r = 0,3$	39 (68,4)	18 (31,6)
2c) Uma relação funcional entre as variáveis $X$ e $Y$ corresponde a um valor de $r$ de $+1$ ou $-1$	49 (86,0)	22 (14,0)
2d) O coeficiente de correlação pode interpretar-se como uma percentagem da variância	10 (17,5)	47 (82,3)

Na alínea a), verifica-se que 52 (91,2%) estudantes afirmaram que quando o coeficiente de correlação é igual a zero, as duas variáveis  $X$  e  $Y$  são independentes, que constitui a resposta correta.

Em relação à alínea 2b), verificou-se que 39 (68,4%) estudantes assinalaram a afirmação como verdadeira, ou seja, que a correlação entre duas variáveis duplica quando o coeficiente de correlação duplica. Provavelmente, os estudantes exibiram este raciocínio incorreto pelo facto de interpretarem o coeficiente de correlação como se tivesse um papel análogo à constante de proporcionalidade de uma função linear, pois acreditámos que uma vez aprendido em anos anteriores o critério da determinação da proporcionalidade direta, que define que se aumentamos ou diminuimos o valor de uma variável  $x \rightarrow kx$ , a outra aumenta ou diminui na mesma proporção, isto é,  $f(kx) \rightarrow kf(x)$ , aliado à ligação que, muitas vezes, os estudantes

fazem entre a correlação positiva ou negativa e a proporcionalidade direta ou inversa, respetivamente, pode ter sido um dos fatores que contribuiu para essas respostas incorretas.

No que diz respeito a alínea 2c), verificou-se que 49 (86,0%) estudantes souberam estabelecer uma relação correta entre os valores  $+1$  e  $-1$  do coeficiente de correlação com o tipo de dependência, neste caso uma dependência funcional entre as duas variáveis, afirmando que os valores do coeficiente de correlação  $+1$  e  $-1$  correspondem a relações perfeitas, respetivamente positiva e negativa. Daqui, presume-se que os estudantes fizeram esta associação por compreenderem que uma correlação perfeita é representada por uma função afim (relação funcional), ou seja, que todos os pontos do diagrama de dispersão estão sobre essa reta. Os restantes 8 (14,0%) estudantes responderam negativamente à afirmação, e julgamos que eles tenham cometido esse erro por consideraram que a afirmação estivesse relacionada com o domínio do coeficiente de correlação, que é  $[-1, +1]$ .

Por último, na alínea 2d), 47 (82,3%) estudantes consideraram falsa a afirmação, ou seja, que o coeficiente de correlação não pode ser interpretado como uma percentagem de variância. Para além disso, parece que alguns estudantes confundiram o coeficiente de correlação com o coeficiente de determinação, dado que 10 (17,5%) estudantes responderam que a afirmação é verdadeira.

### Questão 3

Sabendo-se que a covariância entre as variáveis  $X$  e  $Y$  é maior do que 0, indique se é verdadeira ou falsa cada uma das seguintes afirmações:

- a) A correlação entre  $X$  e  $Y$  é positiva. \_\_\_\_\_
- b) A correlação pode ser não linear. \_\_\_\_\_
- c)  $X$  e  $Y$  podem não estar correlacionadas. \_\_\_\_\_
- d) O declive da reta de regressão é positivo. \_\_\_\_\_
- e) O coeficiente de correlação entre  $X$  e  $Y$  é positivo. \_\_\_\_\_

Apesar do conceito de covariância ser importante no estudo da associação estatística e regressão linear, ele, muita das vezes, é relegado para um papel secundário, sendo o coeficiente de correlação considerado mais relevante para determinar a relação entre duas variáveis (Sánchez, 1996).

Na questão 3 faz-se alusão ao conceito de covariância, fundamentalmente a relação do sinal da covariância com o sinal do coeficiente de correlação  $r$  e a relação entre a covariância e o tipo de relação.

Os resultados sintetizados na tabela 13 mostram, em geral, uma boa prestação dos estudantes na relação entre a covariância e o coeficiente de correlação, assim como na relação entre a covariância e o tipo de relação.

Tabela 14. Frequência absoluta (percentagem) de estudantes nas respostas à questão 3

Sabendo-se que a covariância entre as variáveis $X$ e $Y$ é maior do que 0, indique se é <u>verdadeira</u> ou <u>falsa</u> cada uma das seguintes afirmações:	Frequências (%)	
	Verdadeira	Falsa
3a) A correlação entre $X$ e $Y$ é positiva	53 (93,0)	4 (7,0)
3b) A correlação pode ser não linear	18 (31,6)	39 (68,4)
3c) $X$ e $Y$ podem não estar correlacionadas	9 (15,8)	48 (84,2)
3d) O declive da reta de regressão é positivo	43 (75,4)	14 (24,6)
3e) O coeficiente de correlação entre $X$ e $Y$ é positivo	50 (87,7)	7 (12,3)

Na alínea 3a) interessava-nos saber se os estudantes associam o tipo de correlação ao sinal de covariância. Os resultados obtidos mostram que 53 (93,0%) estudantes associam o tipo de relação entre  $X$  e  $Y$  ao sinal de covariância entre ambas as variáveis. Isto é, os estudantes afirmaram que se a covariância entre  $X$  e  $Y$  é maior que 0, então a correlação entre elas é positiva. Contudo, no que respeita à relação entre a covariância e o coeficiente de correlação, alínea 3e), a percentagem das respostas corretas diminuiu em 5,3%. Assim, podemos inferir que os estudantes que deram uma resposta incorreta não terão tomado em consideração a fórmula de cálculo do coeficiente de correlação de Pearson  $\rho = \frac{\delta_{xy}}{\delta_x \delta_y}$ , da qual obviamente se deduz que de uma covariância maior que zero se obtêm uma correlação positiva e um coeficiente de correlação maior que zero.

Em relação a alínea 3b), interessava-nos saber se os estudantes são conscientes de que uma covariação positiva pode não resultar numa correlação não linear. Constatámos que houve um menor número de estudantes, 18 (31,6%), que deram uma resposta correta, ou seja, que afirmaram que se a covariância entre duas variáveis é maior que 0, então a correlação pode não ser linear. Pensamos que o reduzido número de respostas corretas resultou, tal como referimos anteriormente, do facto de o ensino da associação estatística e regressão linear dedicar maior atenção ao coeficiente de correlação linear e à reta de regressão linear, em detrimento de outros tipos de regressão e coeficientes de correlação.

Quanto a alínea 3c), constatámos que 48 (84,2%) estudantes declararam ser falsa a afirmação. Esta é uma resposta correta, na medida em que se analisarmos a fórmula de correlação de Pearson, referida anteriormente, podemos observar que quando a covariância decresce também decresce a intensidade da correlação e quando a covariância cresce a intensidade da correlação também cresce, que é o mesmo que dizer: quanto maior for a covariância entre duas variáveis, mais fortemente correlacionadas elas estão.

Por último, na alínea 3d), 43 (75,4%) estudantes deram a resposta correta. Assim, concluímos que os estudantes estabeleceram a relação entre o sinal da covariância e o sinal do declive da reta de regressão. Ademais, de acordo com os resultados da alínea 3a), os estudantes demonstraram saber relacionar a covariância e o sentido de correlação. Ora, uma vez que o declive determina o sentido da correlação, justifica-se que se tenha obtido uma maior percentagem de respostas corretas na alínea a).

#### **Questão 4**

Os números 0,5, -0,8, 0,2, -0,4 e 0 correspondem a diferentes valores do coeficiente de correlação entre duas variáveis. Escreva esses números segundo a ordem da maior para a menor correlação que definem entre essas variáveis.

\_\_\_\_\_ Maior valor de correlação  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_ Não existe correlação

O objetivo da questão 4 é saber se os estudantes são capazes de estabelecer a correspondência entre o valor do coeficiente de correlação e a intensidade da associação ou de dependência entre as variáveis. Nesta questão, verificou-se que 45 (78,9%) estudantes ordenaram corretamente os valores do coeficiente de correlação, ordenando-os por ordem decrescente dos seus valores absolutos, tal como mostra o seguinte exemplo: “-0,8; 0,5; -0,4; 0,2; 0” (E51).

Dos restantes, 3 (5,3%) estudantes responderam considerando o coeficiente de correlação como um número real, ignorando o seu significado estatístico. Estes estudantes ordenaram os valores do coeficiente de correlação segundo a ordem decrescente usual dos números reais. Por exemplo, o estudante E35 adotou esse procedimento, escrevendo: “0,5; 0,2; 0; -0,4; -0,8”.

Outro aspeto relevante que se constatou nesta questão foi o elevado número de estudantes (50, ou seja, 87,7%) que interpretaram o valor zero (0) como sendo indicando a não

existência de relação entre as duas variáveis. Esta constatação confirma o que se tinha afirmado antes na questão 2a): *Se  $r = 0$ , as variáveis  $X$  e  $Y$  são independentes*. Apesar de todos estes estudantes terem associado corretamente a independência ao valor 0 (zero) do coeficiente de correlação, 2 (3,5%) destes estudantes ordenaram os outros valores não nulos de forma incorreta. Especificamente, agruparam os coeficientes positivos do lado esquerdo e os coeficientes negativos do lado direito, ordenando os coeficientes de cada grupo seguindo a ordem numérica de que o número de maior valor absoluto precede o de menor valor absoluto. Por exemplo, o estudante E35 apresentou a resposta: “0,5 ; 0,2 ; -0,8 ; -0,4 ; 0”.

Outros 5 (8,8%) estudantes fizeram o mesmo agrupamento, porém não utilizaram ordem numérica usual, particularmente nos números negativos. Por exemplo, o estudante E28 agrupou os valores do coeficiente de correlação do seguinte modo: “0,5 ; 0,2 ; -0,4 ; -0,8 ; 0”. Nesta resposta podemos constatar que os estudantes ao compararem os valores negativos do coeficiente de correlação tendem a utilizar a ordenação usual em  $\mathbb{R}^-$ , atribuindo um nível de correlação superior ao valor de menor valor absoluto, já que ao ordenarmos estes valores negativos como números é maior aquele que tiver menor valor absoluto.

Finalmente, 2 (3,5%) estudantes agruparam de forma diferente, começando por agrupar os coeficientes negativos do lado esquerdo e depois os coeficientes positivos do lado direito. Na resposta do estudante E44: “-0,8 ; -0,4 ; 0,5 ; 0,2 ; 0” verifica-se que ele usou o sentido da reta numérica no caso dos valores negativos e o sentido oposto no caso dos valores positivos. A este respeito, González et al.(1990) referem que o conhecimento sobre a ordenação de números negativos constitui um obstáculo para ordenar a intensidade de correlação.

### **Questão 5**

Nesta questão pretende-se analisar as estratégias que os estudantes utilizam para estimar o valor do coeficiente de correlação de duas variáveis representadas através do diagrama de dispersão. Consideramos *a priori* que as estratégias aplicadas pelos estudantes abarcariam aspetos meramente intuitivos, passando pela avaliação do aspeto do diagrama de dispersão e pelo ajuste aproximado da nuvem de pontos através da reta de regressão.

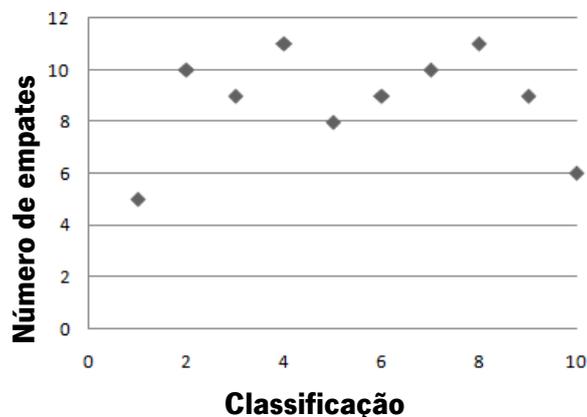
Para analisar esta questão, inicialmente registaram-se os valores dos coeficientes indicados pelos estudantes. Em seguida, determinaram-se as médias, os desvios padrão e amplitude total das distribuições dos valores dos coeficientes de correlação estimados pelos 57 estudantes e compararam-se com os valores dos coeficientes esperados (ver tabela 14).

Tabela 15. Valores dos coeficientes de correlação esperados e médias, desvios padrão e amplitudes totais dos valores dos coeficientes de correlação estimados pelos estudantes

	Questões			
	5a)	5b)	5c)	5d)
Coefficiente de correlação esperado	0,21	0,53	0,93	0,02
Média dos coeficientes de correlação estimados	0,24	0,55	0,95	0,10
Desvio padrão dos coeficientes de correlação estimados	0,32	0,11	0,03	0,21
Amplitude total dos coeficientes de correlação estimados	1	0,4	0,15	2

A seguir são analisados os argumentos apresentados pelos estudantes para justificarem os valores dos coeficientes de correlação por si estimados em cada uma das quatro alíneas da questão.

- a) Classificação das 10 primeiras equipas do campeonato nacional de futebol da última temporada e o número de jogos empatados.



$r =$  \_\_\_\_\_ Porquê?

Na alínea 5a) foi pedido aos estudantes para estimarem o valor do coeficiente de correlação, sem efetuar cálculos, a partir do diagrama de dispersão representando as variáveis *classificação* e *número de empates* de dez equipas do campeonato nacional de futebol. Foi-lhes também solicitado que justificassem a sua resposta.

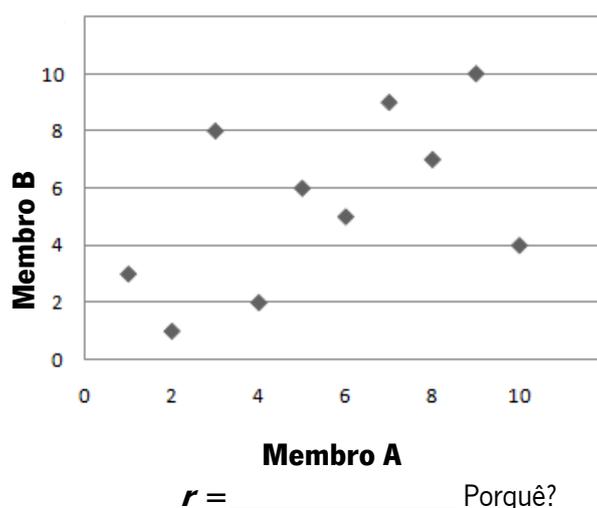
Constatámos, nesta alínea, que a amplitude dos valores estimados pelos estudantes foi de um (1) e a média foi de 0,24. Comparando o valor do coeficiente de correlação esperado (0,21) com o valor da média dos valores estimados pelos estudantes (0,24), observa-se um desvio de 0,03, que não constitui um valor muito elevado. Já em termos de variação, tanto o desvio padrão

como a amplitude total apresentam valores relativamente elevados, sobretudo no caso da amplitude total. Todavia esses valores devem-se a uma pequena percentagem de valores estimados pelos estudantes uma vez que se verificou que 48 (84,2%) estudantes associaram o diagrama de dispersão a um coeficiente de correlação positivo e com uma intensidade nula, baixa ou moderada, variando entre os valores 0 e 0,5 (inclusive).

As justificações que fundamentaram a determinação do sinal do coeficiente de correlação, assim como da sua intensidade, assentam em dois aspetos: primeiro, na análise da covariação das variáveis para explicar o sinal do coeficiente de correlação e, por conseguinte, o tipo de relação; segundo, na avaliação global da nuvem de pontos, isto é, na ideia de que uma maior dispersão dos pontos implica menor correlação, e na análise do ajuste da reta de regressão à nuvem de pontos. Estas duas últimas justificações foram utilizadas para atribuir uma certa intensidade à correlação, isto, é se era fraca, moderada e forte. Por exemplo, o estudante E7 respondeu o seguinte: “ $r = 0,3$  porque existem menos pontos no diagrama que formam uma linha reta, isto é, os pontos estão dispersos”.

Os outros 9 (15,8%) estudantes responderam simplesmente que não havia correlação, indicando para valor desse coeficiente o valor zero (0) pelo fato de os pontos do diagrama não mostrarem uma tendência linear crescente ou decrescente. Vejamos, por exemplo, a resposta do estudante E44: “ Não há correlação entre a classificação e o número de empates porque a dispersão dos pontos não mostra uma linha reta”.

**b)** Pontuações atribuídas pelos membros A e B de um júri a 10 projetos apresentados.



Na alínea 5b, pediu-se para os estudantes estimarem o coeficiente de correlação das pontuações atribuídas por dois membros de um júri, o *membro A* e o *membro B*, a partir da representação dos dados num diagrama de dispersão.

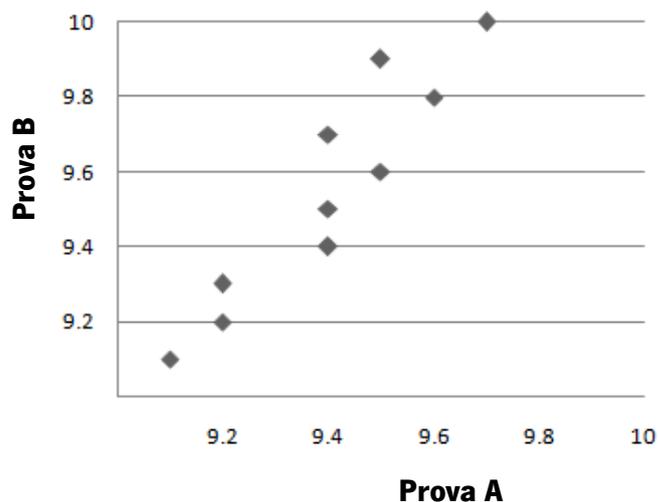
Em média, os coeficientes de correlação atribuídos pelos estudantes nesta alínea foi de 0,55 e a amplitude total desses valores foi de 0,4 (variando entre 0,4 e 0,8). Por outro lado, pelo valor do desvio padrão (0,11) podemos concluir que os valores estimados se encontram muito concentrados em torno da média. Portanto, o valor do coeficiente de correlação esperado (0,53) está muito próximo da média dos valores dos coeficientes estimados pelos estudantes (0,55), daí que a maior parte dos estudantes (49, que equivale a 86,0%) estimou um valor que estava próximo do coeficiente de correlação esperado, variando entre 0,4 e 0,6 (inclusive).

De uma forma geral, os raciocínios utilizados pelos estudantes para estimar os coeficientes de correlação de um diagrama de dispersão eram corretos. Tal como na questão anterior, a maior parte deles baseou as suas respostas na avaliação global da dispersão dos pontos do diagrama de dispersão ou no ajustamento de uma reta à nuvem de pontos para obter a intensidade de correlação, para além da análise da covariação das variáveis para determinar o sinal do coeficiente de correlação.

Os estudantes afirmaram que a correlação era positiva e moderada, justificando a sua resposta pelo facto de os pontos não estarem muito ajustados a uma reta. Por exemplo, o estudante E16 respondeu que “ $r = 0,7$  porque, pelo diagrama, os pontos aproximam-se mais ao menos de uma reta de regressão com sentido positivo. É uma correlação moderada positiva”. Para justificar o sinal da correlação, o estudante E27 refere: “ $r = 0,8$  porque a nuvem mostra a existência de uma relação entre as duas variáveis no sentido positivo, quer dizer que à medida que a pontuação do membro B aumenta a pontuação do membro A também aumenta”.

Dos restantes 8 (14,0%) estudantes, 5 indicaram o valor zero (0) para o coeficiente de correlação e justificaram que não havia correlação entre as pontuações dos dois júris. Os outros 3 estudantes responderam simplesmente que não havia correlação entre as pontuações dos dois júris e não apresentaram qualquer valor para o coeficiente de correlação.

- c)** Tempo gasto, em segundos, por 10 atletas olímpicos para correr 100 metros planos em duas provas, a prova A e a prova B.



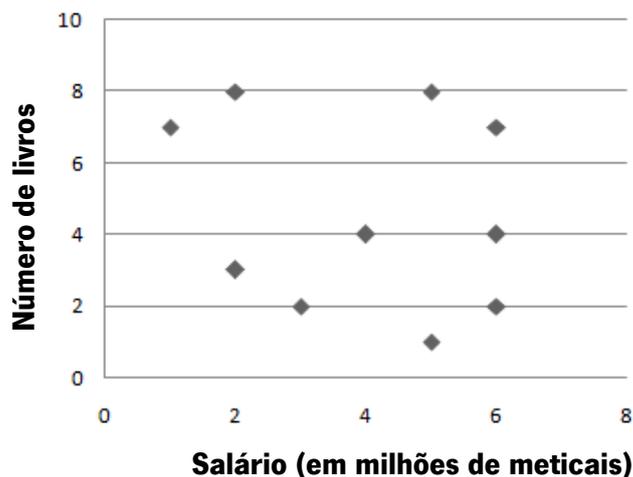
$r =$  \_\_\_\_\_ Porquê?

Na alínea 5c) pretendia-se que os estudantes estimassem o valor do coeficiente de correlação entre os tempos gastos pelos atletas na *prova A* e na *prova B*. Neste caso, o valor obtido para a média dos valores estimados pelos estudantes (0,95) é muito próximo do valor do coeficiente de correlação esperado (0,93), sendo também a amplitude total dos valores estimados relativamente pequena (0,15), correspondendo a uma variação entre 0,85 (valor mínimo) e 1 (valor máximo). Assim, tendo em conta os valores extremos e o desvio padrão (0,03), conclui-se que os valores estimados pelos estudantes se encontram concentrados em torno da média.

Em termos de justificações, a maior parte dos estudantes (55, o que equivale a 96,5%) respondeu que a correlação era positiva e muito forte porque não havia uma grande dispersão dos pontos e que estes se ajustavam a uma reta com declive positivo. Alguns salientaram que ambas variáveis variavam no mesmo sentido para justificarem o sinal do coeficiente de correlação. Por exemplo, o sujeito E49 deu a seguinte resposta: “ $r = 0,96$  porque os pontos estão ajustados a uma reta de regressão de sentido positivo”. Já o sujeito E15 respondeu que “ $r = 0,97$  porque quando o tempo obtido na prova A aumenta, o tempo obtido na prova B também aumenta”.

Os restantes dois estudantes apresentaram o valor 1 para o coeficiente de correlação, significando que a sua estimativa se refere a uma correlação perfeita.

- d)** Salário dos funcionários de uma empresa, em milhões de meticais, e o número de livros lidos ao longo de um ano.



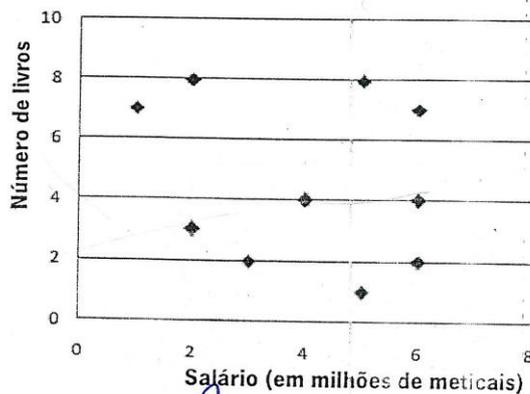
$r =$  \_\_\_\_\_ Porquê?

Por último, na alínea 5d) pedia-se aos estudantes para estimarem o valor do coeficiente de correlação das variáveis *salário* e *número de livros lidos* representadas através do diagrama de dispersão.

Nesta alínea, a média dos valores dos coeficientes de correlação estimados pelos estudantes foi de 0,02, o que constitui um valor consideravelmente afastado do coeficiente de correlação de esperado (0,10), indicando uma correlação positiva e muito fraca. Relativamente às alíneas anteriores, nesta alínea verificou-se uma maior variação dos valores estimados pelos estudantes, obtendo-se a amplitude total de 2, variando entre  $-1$  (valor mínimo) e  $1$  (valor máximo), e um desvio padrão de 0,21.

Do total dos 57 estudantes, 42 (73,7%) indicaram o valor zero (0) como estimativa do valor do coeficiente de correlação, e 13 (22,8%) indicaram valores para o coeficiente de correlação diferente de zero (0). Destes, 4 (7,0%) indicaram valores negativos, compreendidos entre  $-1$  e  $-0,01$  e 9 (15,8%) indicaram valores positivos de 0,1 e 1. Os restantes 2 (3,5%) não indicaram qualquer valor para o coeficiente de correlação, afirmando simplesmente que não havia correlação entre as variáveis.

Tanto os estudantes que indicaram o valor zero, valores diferentes de zero, como os que afirmaram a inexistência de correlação justificaram as suas respostas tendo em conta a análise global da dispersão dos pontos do diagrama. Vejamos a resposta do estudante E11:



$r = 0$  Porque? Porque as duas variáveis no gráfico estão representadas por pontos dispersos, logo não há correlação

Figura 14. Resolução da questão 5d) pelo estudante E11.

Os valores do coeficiente de correlação estimados pelos estudantes, assim como os seus argumentos para justificarem as suas escolhas, leva-nos a inferir que os estudantes têm um considerável conhecimento da relação existente entre o valor do coeficiente de correlação (seu sinal e intensidade) e o tipo de associação existente entre duas variáveis quantitativas, quando representadas através de um diagrama de dispersão.

Outro aspeto importante a reter, é que, na maior parte das justificações, os estudantes tiveram em conta a configuração global do diagrama em detrimento de conhecimentos prévios sobre as situações correlacionais entre as variáveis em estudo. Daqui, podemos inferir que a correlação ilusória, referida por Chapman e Chapman (1969), não esteve presente nos raciocínios dos estudantes em estudo.

### Questão 6

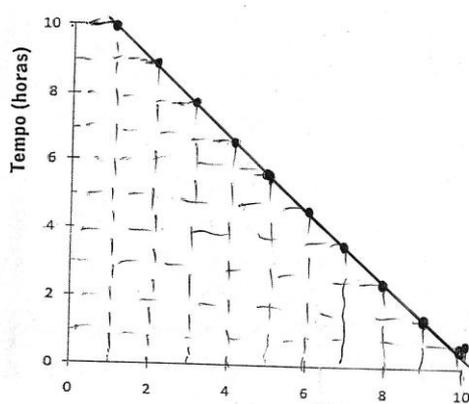
Na questão 6 pretende-se estudar a capacidade dos estudantes em construir um diagrama de dispersão que represente de forma coerente a relação entre pares de variáveis previamente definidas. Especificamente, pretendemos analisar quais são as estratégias que os estudantes utilizam para associar o diagrama de dispersão por ele desenhado ao tipo de correlação em questão. Nas três alíneas, que a seguir vamos analisar, solicitou-se aos estudantes que representassem 10 pontos num diagrama de dispersão, correspondentes a pares de variáveis previamente indicadas.

- a) Número de pintores pintando uma habitação e tempo em horas para acabar o trabalho

Na alínea 6a) pede-se aos estudantes para representarem o diagrama de dispersão com 10 pontos, representando a variação conjunta das variáveis *tempo* (em horas) e *número de operários* que estão envolvidos na pintura de uma habitação.

*A priori* sabe-se que existe uma relação de dependência entre as duas variáveis pois, em condições normais, quanto maior for o número de operários menor será o tempo de execução da pintura da habitação. Julgamos que o fato de a relação entre as duas variáveis ser clara, terá influenciado os estudantes a darem uma resposta com base em conhecimentos prévios acerca do contexto das variáveis. Assim, 45 (78,9%) sujeitos representaram corretamente o diagrama de dispersão sob ponto de vista do sentido de correlação. Eles tiveram como base umas das propriedades de covariação, que diz que se duas variáveis covariam em sentidos opostos, então a correlação entre elas é negativa.

De entre estes 45 (78,9%) estudantes, houve 25 (43,9%) estudantes que recorreram a noções sobre a dependência funcional. De entre estes, uns representaram apenas os pontos isolados sobre uma reta, descrevendo uma correlação perfeita negativa; enquanto os outros representaram os pontos e uniram-nos através de uma reta, mostrando um modelo funcional conhecido. No exemplo a seguir, o estudante E26, ao relacionar o tempo e o número de pintores, representou os pontos e uniu-os através de uma linha reta, o que mostra uma relação funcional, que não corresponde ao caso específico do problema em questão.



Explique a sua resposta: Numa empresa de pintura, quanto mais forem os operários menos será o tempo de trabalho, um operário pintando uma casa gasta mais o tempo que 2 ou 3, no normal.

Figura 15. Resolução da questão 6a) pelo estudante E26.

Tanto os estudantes que representaram os pontos isoladamente como os que uniram os pontos através de uma reta revelaram que possuem uma concepção determinística de associação estatística.

Os outros 20 (35,0%) estudantes construíram igualmente um diagrama que representava uma correlação negativa e distribuíram os pontos aleatoriamente em função do sinal de correlação (negativo). O raciocínio destes estudantes mostra que eles souberam distinguir uma relação funcional de uma relação aleatória (que é típica do estudo de associação estatística e regressão lineares). Por exemplo, o estudante E9 desenhou o diagrama do seguinte modo:

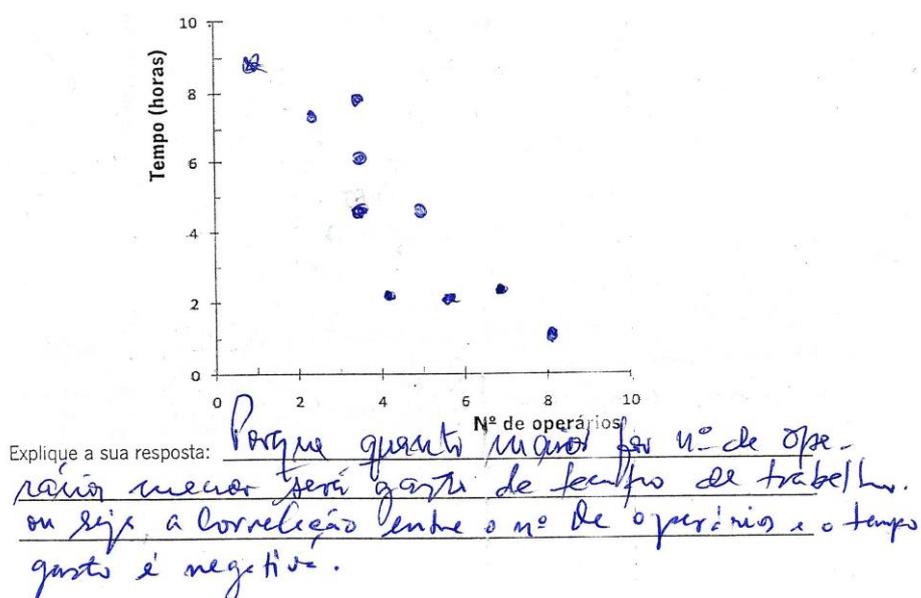


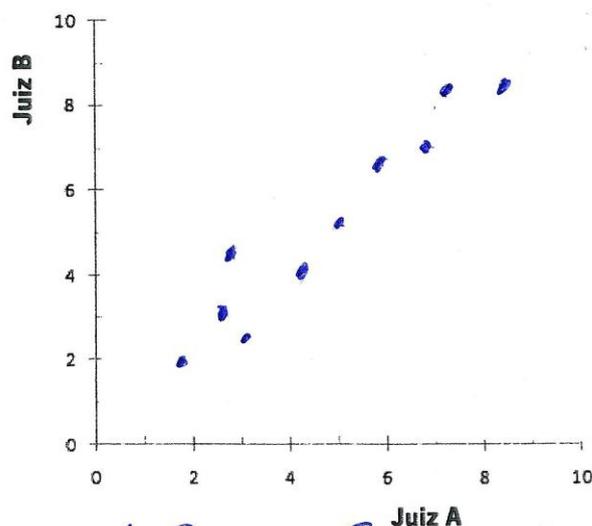
Figura 16. Resolução da questão 6a) pelo estudante E50.

Por último, 12 (21,0%) estudantes apresentaram respostas que considerámos incorretas, na medida em que construíram diagramas representando uma correlação positiva, que não era o caso da situação em estudo.

- b)** Classificações (0 até 10 pontos) atribuídas por dois juizes, A e B, a 10 concorrentes num concurso de beleza.

Contrariamente à alínea anterior, em que as respostas dos estudantes foram influenciadas pelos seus conhecimentos prévios da relação entre as variáveis implicadas, pelo fato delas denunciarem claramente uma relação de dependência entre as variáveis, na alínea 6b) não é muito clara a existência de uma relação do tipo causal. Trata-se de uma relação de concordância em que interessa saber se existe ou não uma consonância entre as classificações atribuídas por dois juizes independentes a um mesmo grupo de concorrentes.

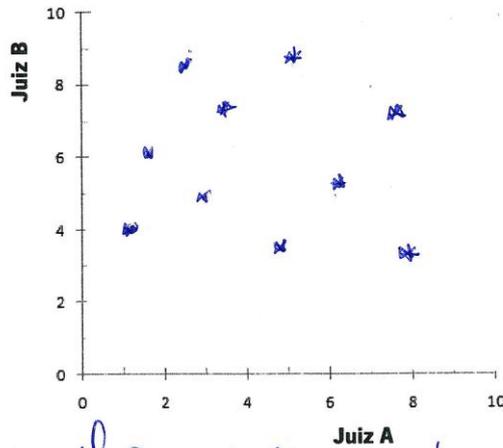
Nesta alínea existe uma correlação positiva entre as duas variáveis *classificação do juiz A* e *classificações do juiz B*. Na medida em que, sendo juizes de um concurso de beleza, era expectável que eles atribuissem as pontuações tendo em conta determinados critérios predefinidos, de modo a que as pontuações não fossem muito distintas. Assim, 39 (68,4%) estudantes desenharam um diagrama de dispersão representando uma correlação positiva forte, argumentando de que, sendo um grupo de júris de um mesmo concurso, os dois juizes não deviam apresentar pontuações muito diferentes, mas sim deviam apresentar pontuações muito próximas. Por exemplo, o estudante E39 apresentou o seguinte diagrama:



Explique a sua resposta: A Correlação é positiva porque sendo um concurso de beleza em que há critérios previamente definidos, quando a nota atribuída por um juiz é alta a do outro também será alta.

Figura 17. Resolução da questão 6b) pelo estudante E39.

Os restantes 18 (31,6%) estudantes desenharam um diagrama de dispersão onde não é evidente a existência de uma correlação positiva ou negativa. Neste caso, os estudantes justificaram as suas respostas argumentando que pelo fato de as variáveis (juiz A e juiz B) representarem sujeitos distintos, fazia sentido que cada um tivesse a sua opinião e atribuisse a pontuação que entendesse, independentemente da nota que outro juiz atribuiu. Vejamos, por exemplo, o diagrama e a justificação que o estudante E45 representou:

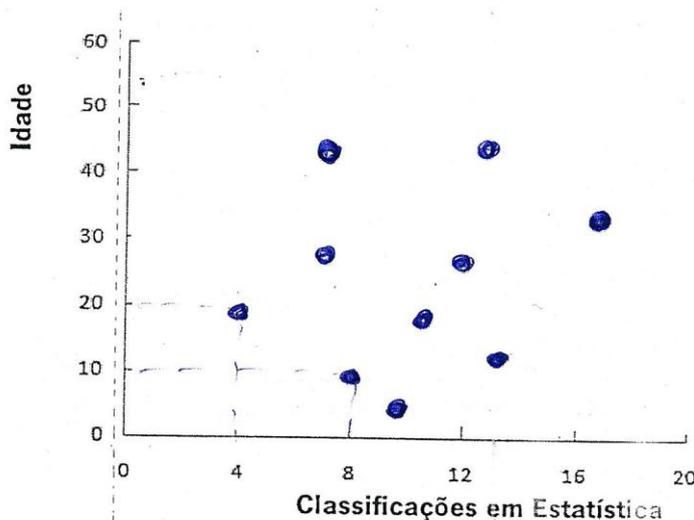


Explique a sua resposta: O grau de dispersão é maior pelo facto de a atribuição dos pontos ser individual, ou seja, cada juiz tem a sua própria opinião acerca do Concorrente que pode ser diferente da do outro juiz.

Figura 18. Resolução da questão 6b) pelo estudante E45.

**c) Classificações em Estatística e idade do aluno.**

Nesta questão pretendia-se que os estudantes representassem dez pontos num diagrama de dispersão que retratasse a covariação entre as variáveis *classificação em Estatística* e a *idade do aluno*. Nesta alínea 50 (87,7%) estudantes representaram os pontos de uma forma dispersa e associaram esta dispersão à independência entre as variáveis. Os estudantes em referência argumentaram que a classificação em Estatística não depende da idade. Por exemplo, o estudante E9 representou o diagrama da seguinte forma:



Explique a sua resposta: A idade e classificações em estatística não tem relação significativa que a relação é nula.

Figura 19. Resolução da questão 6c) pelo estudante E9.

Os restantes 7 (12,3%) estudantes desenharam um diagrama de dispersão representando uma correlação negativa muito forte. Neste caso, os estudantes referiram que quanto maior for a idade menor é a capacidade de aprender Estatística, e consequentemente obtêm menores classificações. Por exemplo, o estudante E55 desenhou o diagrama e justificou a sua resposta da seguinte forma:

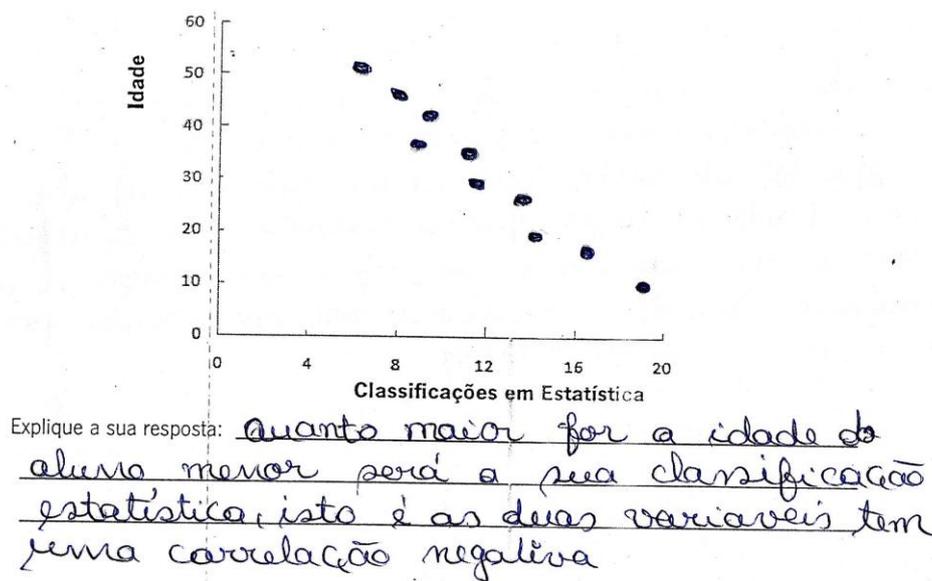


Figura 20. Resolução da questão 6c) pelo estudante E55.

Apesar das situações apresentadas nas questões 5 e 6 não serem muito habituais no ensino da associação estatística e regressão linear, os resultados obtidos demonstram que a maioria dos estudantes têm uma boa capacidade de estimar o valor do coeficiente de correlação e de avaliar a associação estatística a partir de um diagrama de dispersão, assim como são capazes de construir um diagrama de dispersão que expressa razoavelmente a covariação entre duas variáveis.

### Questão 7

Na tabela seguinte mostram-se as classificações obtidas por 10 alunos de uma escola nos exames de Matemática e Física.

Matemática	2	4	4	6	8	8	10	12	14	14
Física	6	14	12	4	4	14	8	10	6	10

- a)** Determine o valor do coeficiente de correlação  $r$ .
- b)** A partir do valor de  $r$ , calculado em **a)**, o que pode dizer-se acerca do sinal e da intensidade da correlação entre as classificações em Matemática e Física?

A finalidade desta questão era avaliar a capacidade dos estudantes em determinarem o valor coeficiente de correlação linear a partir de dados apresentados numa tabela e de o

interpretarem quanto ao sinal e à intensidade no contexto da situação apresentada, em que as variáveis envolvidas são as *classificações em Matemática* e as *classificações em Física* obtidas por 10 alunos. Seguidamente apresentam-se as respostas dos estudantes em cada uma das alíneas.

**a)** Determine o valor do coeficiente de correlação  $r$ .

Para determinação do coeficiente de correlação, os 57 (100%) estudantes recorreram a fórmula do coeficiente de correlação de Pearson  $r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$ . De

entre os 57 estudantes, 41 (71,9%) estudantes apresentaram o valor correto do coeficiente de correlação, ou seja, o valor de  $r = -0,09$ . O exemplo a seguir mostra os procedimentos utilizados pelo estudante E42 na determinação do valor do coeficiente de correlação.

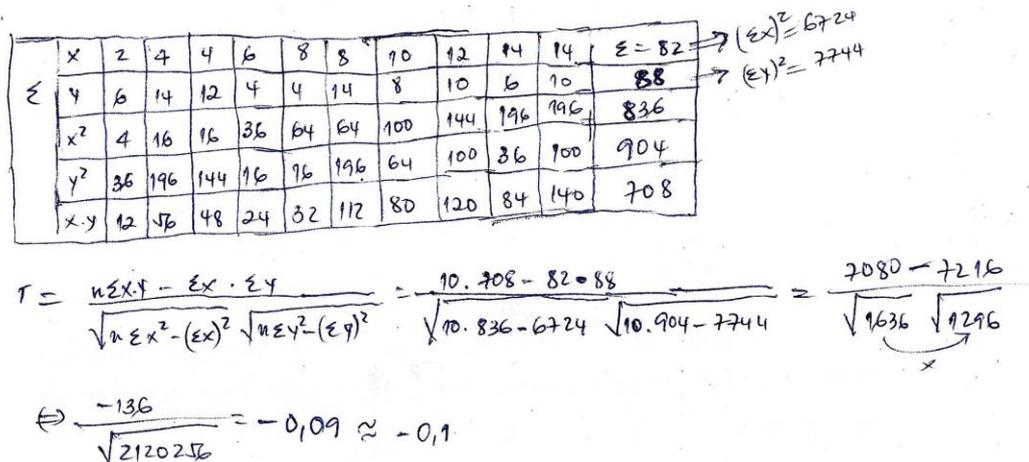


Figura 21. Resolução da questão 7a) pelo estudante E42.

Os restantes 16 (28%) estudantes, apesar de terem utilizado a fórmula correta para o cálculo do valor do coeficiente de correlação (fórmula de Pearson), deram uma resposta inadequada devido, fundamentalmente, a erros de cálculo. Entre as respostas erradas, importa referir que 5 (8,8%) estudantes atribuíam ao coeficiente de correlação um valor que não pertencia ao intervalo  $-1 \leq r \leq 1$ . Por exemplo, o estudante E18 apresentou o seguinte resultado do coeficiente de correlação: " $r = -141,31$ ", o que denota uma incompreensão do significado do coeficiente de correlação linear.

**b)** A partir do valor de  $r$ , calculado em **a)**, o que pode dizer-se acerca do sinal e da intensidade da correlação entre as classificações em Matemática e Física?

Em relação a esta alínea, constatámos que quase todos os estudantes (53, o que corresponde a 93%) interpretaram de forma parcialmente correta o sinal e a intensidade de correlação, na medida em que afirmaram simplesmente que o sinal da correlação era negativo. Consideramos que a resposta seria completa se, nas suas justificações, tivessem explicitado de forma clara a relação entre o sinal do coeficiente de correlação e o sentido de covariação, ou seja, explicassem as implicações que este sinal teria na covariação das duas variáveis (*classificações em Matemáticas e classificações em Física*). Acreditamos que esta constatação é resultado da forma como este tópico tem sido ensinado aos estudantes, no qual é dada muita relevância à determinação do resultado em detrimento da sua interpretação.

Apenas 4 (7,0%) estudantes deram uma resposta que consideramos completa, isto é, apresentaram o sinal e justificaram-no em termos da covariação das variáveis, tal como mostra a resposta do estudante E2.

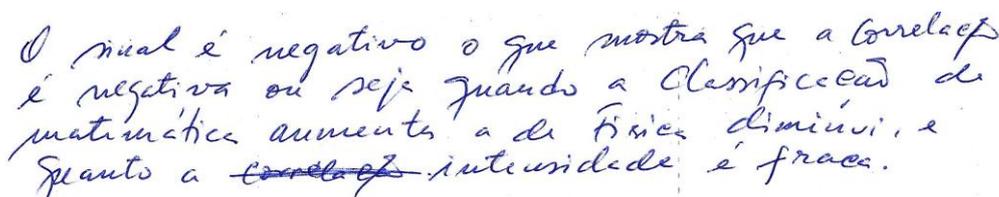


Figura 22. Resolução da questão 7b) pelo estudante E2.

### Questão 8

Na tabela seguinte estão registadas as idades de 20 casais à data dos seus casamentos.

Idade do Homem ( $x_i$ )	18	20	21	21	22	23	23	23	24	25	25	26	26	26	28	28	28	30	31	32
Idade da Mulher ( $y_i$ )	17	20	20	22	22	21	22	23	23	24	25	23	24	27	26	27	29	29	30	26

- Qual é a previsão da idade da mulher à data do casamento, sabendo que o homem tinha 29 anos?
- Qual é a previsão da idade do homem à data do casamento, sabendo-se que a mulher tinha 30 anos?
- Será seguro prever a idade da mulher à data do casamento, sabendo que o homem tinha 45 anos? Porquê?

Uma das finalidades do estudo de regressão é encontrar uma curva particular, precisamente aquela que melhor traduz a relação entre as variáveis, e através dela fazer predições do valor de uma das variáveis a partir do valor conhecido da outra variável. É assim que nas alíneas a) e b) procuramos saber, por um lado, os procedimentos usados pelos estudantes para a determinação da reta de regressão linear de  $y$  sobre  $x$  e de  $x$  sobre  $y$ ,

assim como a sua capacidade de prever o valor de uma variável a partir do valor conhecido da outra variável. Finalmente, na alínea c) questionam-se os estudantes sobre as condições em que é possível fazer previsões fiáveis a partir do modelo de regressão.

- a) Qual é a previsão da idade da mulher à data do casamento, sabendo que o homem tinha 29 anos?

Pedia-se nesta alínea que o estudante estimasse a idade da mulher ( $y$ ) à data do casamento, sabendo que a idade do homem ( $x$ ) na mesma data era de 29 anos.

Constatámos que a maior parte dos estudantes (41, ou seja, 71,9%) respondeu corretamente à questão, afirmando que a idade do homem à data do casamento era de 27 anos. Para chegarem a esta resposta, eles utilizaram a equação da reta de regressão linear  $y$  sobre  $x$  definida pela fórmula:  $y = a + bx$ , onde  $b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$ , e  $a = \bar{y} - b\bar{x}$ , tal como

mostra a resolução apresentada pelo estudante E27.

Handwritten solution showing the calculation of the regression line  $y = \alpha x + \beta$ . The slope  $\alpha$  is calculated as  $\frac{n \sum xy - \sum x \cdot \sum y}{n \sum x^2 - (\sum x)^2} = \frac{20 \cdot 12219 - 500 \cdot 480}{20 \cdot 12768 - 250000} = \frac{244380 - 240000}{255360 - 250000} = \frac{4380}{5360} \approx 0,8$ . The intercept  $\beta$  is calculated as  $\beta = \bar{y} - \alpha \bar{x} = 2 - 0,8 \cdot 25 = 29 - 20 = 9$ . The final prediction is  $\Rightarrow y = 0,8x + 9$ , para  $x = 29$  teremos  $y = 0,8 \cdot 29 + 9 = 27,2$ , então a previsão da idade da mulher à data do casamento é de 27 anos.

Figura 23. Resolução da questão 8a) pelo estudante E27.

Os restantes 16 (28,0%) estudantes deram respostas incorretas. Destes, 8 (14,0%) estudantes responderam sem efetuar cálculos, afirmando simplesmente que a mulher tinha 29 anos na altura do casamento.

Um aspeto importante, que também foi objeto de análise nesta questão, diz respeito ao significado do resultado obtido no contexto da situação apresentada. Neste âmbito, verificámos que, apesar de os restantes 8 (14,0%) estudantes terem utilizado um modelo adequado para determinação da reta de regressão, apresentaram uma resposta inadequada ao contexto da medida em que apresentaram como resposta, nuns casos, um valor na forma decimal e, noutros casos, um valor negativo, ao invés de responder com um número natural como requer o contexto (idade). Por exemplo, o estudante E38 apresentou como resultado o valor  $-25,64$  para idade da mulher à data do casamento, não mostrando ser capaz de questionar o resultado

obtido. Julgamos que os estudantes em referência prestaram mais atenção à determinação e manipulação do modelo estatístico do que propriamente ao significado do resultado obtido (Freudenthal, 1994).

**b)** Qual é a previsão da idade do homem à data do casamento, sabendo-se que a mulher tinha 30 anos?

Nesta alínea, para além do resultado da estimativa da idade do homem, pretendíamos essencialmente saber se ao determinarem essa estimativa, a partir da idade da mulher, os estudantes tinham em conta a expressão da reta de regressão  $x$  sobre  $y$ .

Uma vez que nesta alínea, ao contrário da alínea anterior, se pretendia estimar a idade do homem ( $x$ ) a partir da idade da mulher ( $y$ ), o procedimento correto seria determinar a reta de  $x$  sobre  $y$ . Assim, constatámos que 20 (35,1%) estudantes apresentaram um resultado incorreto, afirmando que a idade do homem à data do casamento era 32 anos, correspondente à idade 30 anos da mulher na mesma data. Estes estudantes recorreram a reta de regressão de  $y$  sobre  $x$ , substituíram a variável  $y$  pelo valor 30 e determinaram o correspondente valor de  $x$ . Portanto, considerámos este procedimento incorreto, dado que estes estudantes não discriminaram de forma adequada a variável independente (idade da mulher) da variável dependente (idade do homem). Por exemplo, o sujeito E16 respondeu da seguinte maneira:

$$y = ax + b = 30 = 0,82x + 3,5 \Rightarrow x = 32 \text{ anos}$$

R: A previsão da idade do homem é de 32 anos.

Figura 24. Resolução da questão 8b) pelo estudante E16.

Apenas 7 (12,3%) estudantes apresentaram o resultado correto de 31 anos. Neste caso, os estudantes souberam distinguir as duas retas de regressão linear, escolhendo o modelo adequado para o problema:  $x = a + by$ . Para encontrarem o resultado, estes estudantes substituíram na variável  $y$  o valor 30, para de seguida calcularam o desejado valor de  $x$ . Por exemplo, o estudante E48 apresentou a seguinte resolução:

$$\begin{aligned}
 X &= a + b\bar{y} \Leftrightarrow a = \bar{X} - b\bar{y} \Leftrightarrow a = 25 - 1,0046 \cdot 24 = 0,8896 \\
 b &= \frac{n \sum xy - \sum x \sum y}{n \sum y^2 - (\sum y)^2} = \frac{20 \cdot 12219 - 480 \cdot 500}{20 \cdot 11738 - 480^2} = \frac{4380}{4380} = 1,0046 \\
 X &= a + by \Leftrightarrow X = 0,8896 + 1,0046 y \\
 &X = 0,8896 + 1,0046 \cdot 30 \\
 &X = 31 \text{ anos}
 \end{aligned}$$

Figura 25. Resolução da questão 8b) pelo estudante E48.

Os restantes 27 (47,4%) estudantes afirmaram que a idade do homem à data do casamento era simplesmente 31 anos. Apesar destes estudantes terem apresentado o resultado correto, eles não utilizaram nenhuma das retas de regressão linear, limitando-se a consultar a tabela dada no enunciado para obterem o valor da variável dependente correspondente ao valor dado da variável independente (31 anos). Este tipo de procedimento, que consideramos incorreto, constitui um indicador de uma conceção funcional de associação, ao consideraram que para existir associação entre duas variáveis deve haver uma correspondência que associa a cada valor da variável independente apenas um valor da variável dependente (Estepa et al., 1999).

Tal como na alínea anterior, também nesta alínea houve 3 (5,3%) estudantes que apresentaram uma resposta um tanto desajustada, na medida em que apresentaram um valor da idade do homem muito elevado ou atribuíram à idade um valor totalmente descabido. Por exemplo, o estudante E28 respondeu: “estima-se que a idade do homem à data do casamento era de 308,8”.

- c)** Será seguro prever a idade da mulher à data do casamento, sabendo que o homem tinha 45 anos? Porquê?

O aspeto mais relevante que pretendíamos saber na alínea c) era se os estudantes eram capazes de avaliar os perigos de extrapolar os resultados da amostra para outras situações, afirmando algumas precauções na sua utilização.

Frequentemente os alunos não são conscientes dos cuidados que se deve ter quando se faz uma extrapolação, tal como avaliar se o valor para o qual pretendemos fazer previsões pertence ou não ao intervalo dos dados observados, dado que à medida que estes valores se afastam dos intervalos dos valores observados, menos fiáveis são as estimativas que podemos fazer de uma das variáveis a partir do conhecimento da outra (Truran, 1997; Fernandes, 2004).

Nesta situação, 22 (38,6%) estudantes, que afirmaram que era seguro prever a idade da mulher à data do casamento a partir da idade do homem, determinaram o valor da idade da mulher através da reta de regressão  $y$  sobre  $x$ , obtendo uma estimativa de 40 anos. Contudo, não fizeram qualquer observação acerca do valor estimado, tendo em conta que o valor dado (45 anos) está muito afastado do intervalo dos valores observados.

Os outros 35 (61,4%) estudantes, sem fazerem cálculos, responderam que não era fiável determinar uma estimativa para um valor que não estava dentro do intervalo observado. Por exemplo, o sujeito E36 afirmou: “Não é seguro prever a idade da mulher sabendo que a idade do homem à data do casamento era de 45 anos, pois estaríamos a estimar dados que não constam na tabela”.

Estes resultados indicam que, apesar de alguns estudantes se mostrarem indiferentes quanto aos perigos da extrapolação, a maioria dos sujeitos desta amostra foram capazes de concluir *a priori* que não era seguro extrapolar para um dado que estava muito distante do intervalo dos valores observados.

### **2.2.2. Síntese sobre as aquisições conceptuais dos estudantes**

Nesta secção faremos a síntese dos principais conhecimentos adquiridos pelos estudantes através das respostas dadas por eles nas 8 questões do teste, que foi amplamente analisado anteriormente. Para tal, a seguir, na Tabela 15, apresenta-se as frequências e percentagens de respostas corretas e incorretas obtidas nas diferentes questões propostas no teste.

No que diz respeito à relação estabelecida entre a covariância e a associação ou independência, concluímos que os estudantes perceberam que o sinal de covariância indica a direção da correlação existente entre duas variáveis estatísticas. Concretamente, 53 (93,0%) estudantes explicitaram a relação entre a covariância positiva e a associação positiva, reconhecendo, portanto, que o sinal da covariância é o indicador do sinal do coeficiente de correlação. Uma pequena percentagem de estudantes (9, ou seja, 15,8%) admitiu a possibilidades de duas variáveis não estarem relacionadas linearmente quando a covariância é positiva, o que quer dizer que a maior parte dos estudantes associaram a covariância positiva a apenas relações lineares, excluindo os outros tipos de relações possíveis.

Tabela 16. Frequências (percentagens) das respostas corretas e incorretas nas diferentes questões do teste

Questões	Frequências (%)	
	Corretas	Incorretas
Questão 1: relação entre o sinal de $r$ e o sentido de variação entre as variáveis	54 (94,7)	3 (5,3)
Questão 2: consequências do conhecimento do valor de $r$		
a) independência	52 (91,2)	5 (8,8)
b) proporcionalidade entre o valor de $r$ e a associação estatística	18 (31,6)	39 (68,4)
c) relação funcional ( $r = -1$ ou $r = 1$ )	49 (86,0)	8 (14,0)
d) variância explicada	47 (82,3)	10 (17,5)
Questão 3: covariância positiva		
a) correlação positiva	53 (93,0)	4 (7,0)
b) a correlação pode ser não linear	18 (31,6)	39 (68,4)
c) pode não existir correlação	48 (84,2)	9 (15,8)
d) declive da reta de regressão positivo	43 (75,4)	14 (24,6)
e) coeficiente de correlação positivo	50 (87,7)	7 (12,3)
Questão 4: relação entre o valor de $r$ e a intensidade da associação	45 (78,9%)	12 (21,0)
Questão 5: estimação do coeficiente de correlação a partir do diagrama de dispersão		
a) classificação e número de empates	48 (84,2%)	9 (16,0%)
b) pontuações dos membro A e B de um júri	49 (86,0%)	8 (14,0%)
c) tempo gasto em duas provas A e B	55 (96,5%)	2 (3,5%)
d) salário e número de livros lidos	42 (73,7%)	15 (26,3%)
Questão 6: construção do diagrama de dispersão		
a) Número de operários e tempo de pintura	45 (78,9%)	12 (21,0%)
b) Classificações dos juízes A e B	39 (68,4%)	18 (31,6%)
c) Classificação em estatística e idade do aluno	50 (87,7%)	7 (12,3%)
Questão 7		
a) determinação do valor de $r$	41 (71,9%)	16 (28%)
b) relação entre o valor de $r$ com o sinal a intensidade da associação	4 (7,0%)	53 (93%)
Questão 8: predição		
a) regressão linear $y$ sobre $x$	41 (71,9%)	16 (28,0%)
b) regressão linear $x$ sobre $y$	7 (12,3%)	50 (87,7%)
c) extrapolação	35 (61,4%)	22 (38,6%)

Verificámos também que quase todos os estudantes compreenderam a relação entre o sinal da correlação e o sentido de variação das variáveis, sobretudo quando o sentido da variação entre as variáveis é crescente. Um total de 54 (94,7%) estudantes respondeu que se a correlação é positiva, os valores da variável variam num mesmo sentido. Porém, apenas 17 (29,8%) estudantes perceberam que esta propriedade também é válida para os casos em que ambas variáveis diminuem simultaneamente.

Quando confrontados com a ordenação dos coeficientes de correlação segundo a intensidade da associação correspondente, os estudantes souberam estabelecer a respetiva correspondência entre o valor do coeficiente de correlação com a intensidade da correlação seguindo a ordem solicitada. Por exemplo, na questão 4), 45 (78,9%) estudantes ordenaram corretamente os coeficientes de correlação segundo a sua intensidade e atribuíram uma maior intensidade aos valores que estavam próximos de 1 ou de  $-1$ , nomeadamente  $-0,8$ , e menor intensidade ao valor  $0,2$ . Também relacionaram corretamente o coeficiente de correlação nulo com a inexistência de correlação, o que foi verificado tanto na alínea 2a) como na questão 4. Por último, um número significativo de estudantes 49 (86,0%) identificou uma correta relação entre a correlação perfeita ou dependência funcional com o valor do coeficiente de correlação correspondente aos valores  $-1$  e  $1$ .

Na questão seguinte, envolvendo quatro alíneas sobre a estimação do coeficiente de correlação correspondente à variação conjunta de duas variáveis representadas num gráfico de dispersão, verificámos de forma geral uma concordância entre o valor do coeficiente de correlação estimado pelos estudantes, o sentido de covariação das variáveis e a intensidade da relação. A maior parte deles justificaram a sua escolha tendo em conta a dispersão da nuvem de pontos, ou seja, do ajustamento ou não ajustamento a uma reta.

O mesmo se pode dizer em relação à construção do diagrama de dispersão a partir de pares de variáveis, assunto tratado nas três alíneas da questão 6. Como era de esperar, com base no conhecimento prévio da existência, ou não, de relação entre as variáveis, também nestas alíneas se constatou uma boa capacidade dos estudantes em representar coerentemente a covariação de duas variáveis num diagrama de dispersão, sobretudo em relação ao sentido da correlação. O mesmo não se pode dizer em relação à intensidade da correlação, verificando-se que quase metade dos estudantes que representaram corretamente o sentido da relação revelaram a propensão para desenhar um gráfico de correlação perfeita (que não era o caso) entre as variáveis, atribuindo-lhe uma natureza determinista, tal como aconteceu na alínea 6a),

envolvendo as variáveis *número de operários* e *tempo de pintura de uma habitação*. Já quando as variáveis eram interdependentes, como no caso concreto da alínea 6b), surgiu uma variedade de gráficos correspondentes às diferentes convicções dos estudantes. Todavia, verificámos que independentemente dessa variedade de respostas, elas estavam em consonância com o diagrama representado. Isto revela uma consistência na representação e interpretação da covariação entre duas variáveis, o que de certo modo confirma as constatações tiradas na questão anterior 5 sobre estimação do coeficiente de correlação com base na variação conjunta.

No que diz respeito as questões de cálculo e interpretação do coeficiente de correlação para dados apresentados numa tabela, a grande maioria dos estudantes (41, ou seja, 71,9%) demonstraram uma boa capacidade de cálculo e de interpretação do coeficiente de correlação.

Já na determinação das retas de regressão de  $Y$  sobre  $X$  e de  $X$  sobre  $Y$ , as maiores dificuldades que os estudantes enfrentaram resultaram do facto de utilizarem uma única reta de regressão, sobretudo a de  $Y$  sobre  $X$ , mesmo que não fosse esse o caso. Por exemplo, na questão 8b) a grande maioria dos estudantes (50, ou seja, 87,7%) recorreu à equação de  $Y$  sobre  $X$  para estimarem  $X$  a partir de  $Y$ , que é um procedimento incorreto e um indicador de uma conceção funcional da regressão linear. A esta dificuldade não terá sido alheia a confusão que os estudantes fizeram entre a variável dependente e a independente.

Quanto aos perigos de extrapolação, a maioria (35, ou seja, 61,4%) dos sujeitos desta amostra foram capazes de concluir *a priori* que não era seguro extrapolar para um dado que estava muito distante do intervalo dos valores observados.



## **CAPÍTULO V**

### **CONCLUSÃO, IMPLICAÇÕES E RECOMENDAÇÕES**

Este capítulo estrutura-se em cinco secções: na primeira secção faz-se uma breve síntese do estudo; seguidamente, na segunda e terceira secções, apresentam-se e discutem-se os principais resultados do estudo, tendo por referência as questões de investigação nele estabelecidas e a literatura que foi revista no estudo; na quarta secção referem-se algumas implicações do estudo para o ensino; e, por fim, na quinta secção apresentam-se limitações e sugestões de futuras investigações.

#### **5.1. Síntese do estudo**

O presente trabalho tem por objetivo determinar as ideias intuitivas de estudantes universitários que frequentam o curso de formação de professores de matemática sobre a associação estatística, antes e depois destes terem tido aulas sobre esta temática. A ideia central do estudo era compreender quais são as estratégias que os estudantes utilizam para avaliar a associação estatística entre duas variáveis representadas numa tabela de contingência de  $2 \times 2$ , num gráfico de dispersão e através da comparação de duas amostras dadas numa tabela.

Esta temática, nomeadamente o estudo das crenças, concepções e erros dos estudantes, tem vindo a ganhar relevância e atualidade, na medida em que é a partir delas que tem favorecido a transformação de algumas práticas de ensino pouco ajustadas a determinados conteúdos, implicando que, atualmente, as práticas de ensino tendem a ter em conta os erros exibidos pelos estudantes e as concepções por eles mais frequentemente partilhadas (Batanero, 2001). Nesse sentido, torna-se imperioso conhecê-las e torná-las disponíveis de modo a que possam ser utilizadas em prol da melhoria do ensino e aprendizagem da estatística, designadamente no caso concreto da associação e regressão linear. É assim que neste estudo se procura perceber a influência que o ensino teve na mudança das ideias e concepções dos estudantes.

Neste contexto, procurando-se determinar os significados que os estudantes universitários atribuem à associação estatística antes e depois do ensino desta temática, no presente estudo estabeleceram-se as duas seguintes questões de investigação:

- 1 – Que significados atribuem os estudantes à associação estatística? Esses significados alteram-se com o ensino do tema correlação e regressão linear?
- 2 – Quais as aquisições conceituais sobre a associação estatística entre duas variáveis que os estudantes adquirem com o ensino do tema correlação e regressão linear?

Para concretizar o estudo, considerou-se uma amostra de 57 estudantes do curso de Matemática de duas Delegações da Universidade Pedagógica de Moçambique, 20 estudantes da Delegação de Maputo e 37 da Delegação de Nampula. Em qualquer das delegações, os estudantes pertenciam a duas turmas do 2º ano do curso de Matemática, que ainda não tinham aprendido as noções de associação estatística e regressão linear na escola secundária nem na universidade.

O estudo foi conduzido, fundamentalmente, segundo uma abordagem de investigação qualitativa, de natureza interpretativa, centrada na identificação e descrição das estratégias adotadas pelos estudantes nas distintas tarefas sobre a temática em estudo. Para o efeito, utilizamos como instrumentos de recolha de dados um questionário, que foi aplicado antes e depois do ensino, e um teste que foi aplicado apenas no final do ensino.

Com o questionário visava-se saber as estratégias utilizadas pelos estudantes para avaliarem associação estatística existente entre duas variáveis representadas através de uma tabela de contingência de  $2 \times 2$ , da comparação de amostras dadas numa tabela e de gráficos de dispersão. Para além de analisarmos as estratégias, determinámos também as concepções a elas subjacentes, o que constituía uma das questões de pesquisa do nosso estudo. A classificação das concepções teve como base as classificações feitas em alguns estudos do mesmo âmbito, com particular destaque para os de Estepa e Batanero (1996), de Batanero et al. (1998) e de Batanero et al. (1996).

Relativamente ao teste, ele englobava questões que incluíam conceitos básicos desta temática tais como: a correlação, tipos correlação, coeficiente de correlação, intensidade do coeficiente de correlação, a covariação e reta de regressão linear. O objetivo destas questões era saber se os estudantes sabiam interpretar algumas propriedades da correlação, bem como relacioná-las entre si. Um outro aspeto incluído no teste, que requeria do estudante o conhecimento de distintas propriedades, foi a estimação do valor do coeficiente de correlação a

partir do gráfico de dispersão, assim como a construção de um diagrama de dispersão a partir de duas variáveis dadas. Por último, incluímos também questões meramente de cálculo, que incluíam o cálculo do valor do coeficiente de correlação a partir de dados representados numa tabela e a determinação da reta de regressão para efetuar previsões de valores de uma variável a partir do conhecimento dos valores da outra variável.

Após a aplicação dos instrumentos, seguiu-se a análise qualitativa e quantitativa das respostas dos estudantes ao questionário e ao teste. A análise qualitativa incidiu basicamente na análise descritiva dos argumentos e/ou estratégias utilizadas pelos estudantes nas distintas questões. Desta análise foi possível identificar e descrever as distintas respostas corretas e incorretas, e em alguns casos parcialmente corretas, e a partir das estratégias determinámos as conceções associadas a essas respostas. Finalmente, para avaliar se existiu alguma diferença nos significados atribuídos pelos estudantes à associação estatística antes e depois do ensino, fizemos uma comparação das frequências e percentagens de respostas corretas antes e depois do ensino.

A seguir, nas duas próximas secções, são apresentados e discutidas as principais conclusões do estudo, tendo por referência as duas questões de investigação formuladas no estudo.

## **5.2. Que significados atribuem os estudantes à associação estatística? Esses significados alteram-se com o ensino do tema correlação e regressão linear?**

O objetivo desta questão de investigação era realizar um estudo exploratório de avaliação da compreensão da associação entre as variáveis representadas numa tabela de contingência, num gráfico de dispersão e numa amostra, emparelhada ou não, em estudantes do 2º ano do curso de Ensino de Matemática. Tendo-se estudado as práticas matemáticas realizadas pelos estudantes na avaliação da associação, classificámos as distintas estratégias que, por sua vez, permitiram identificar as possíveis conceções corretas ou incorretas sobre a associação estatística. As respostas foram recolhidas em dois momentos do estudo (antes e depois do ensino), com o intuito de saber quais eram os significados de associação atribuídos pelos estudantes antes do ensino, e se os mesmos eram modificados em função do ensino do tema. Ou seja, procurámos perceber em que medida o ensino de associação estatística terá influenciado os significados que os estudantes apresentaram antes do ensino.

A análise feita às respostas dos estudantes ao questionário permitiu-nos tirar conclusões sobre os significados de associação atribuídos pelos estudantes e os conflitos a eles subjacentes. A seguir serão apresentados as conclusões das três categorias de análise anteriormente referidas.

No que diz respeito à avaliação da associação em tabelas de contingência  $2 \times 2$ , nos dois momentos do estudo, concluímos que as práticas realizadas pelos estudantes para decidirem acerca da associação tiveram como base diferentes objetos matemáticos, tais como: frequências absolutas, relativas, condicionadas e independência. Contudo, nem sempre estes objetos eram utilizados de forma adequada, motivada, em parte, pela confusão que os estudantes fizeram de alguns conceitos ou propriedades, ou mesmo da interpretação inadequada de algumas propriedades da associação estatística.

Na maior parte das situações, os estudantes não utilizaram toda a informação relevante do problema para decidirem sobre a existência ou não de associação, o que muitas vezes conduziu a respostas incorretas, ou seja, admitiram a existência de dependência entre as variáveis *fumar* e *ter bronquite*. Por exemplo, no pré-ensino, as estratégias que resultaram num maior número de respostas incorretas foram as que tiveram como base informação de duas células, especificamente quando os estudantes compararam as frequências absolutas ou relativas de duas células entre si (*estratégias 1.2a e 1.2b*), e de uma única célula, geralmente a de maior frequência absoluta (*estratégia 1.1a*). O elevado número de respostas com base apenas em informação de uma ou duas células demonstra o que já tinha sido constatado nos trabalhos de Batanero et al. (1996), que os alunos tendem a utilizar parte dos dados (*conceção local*), sobretudo quando estes confirmam as suas preconcepções ou *teorias prévias* sobre a relação entre as duas variáveis.

Por outro lado, julgamos que, no caso particular do recurso a uma única célula (célula de maior frequência), o facto de ela ser a primeira das frequências absolutas da tabela, que corresponde à presença simultânea de ambas as variáveis, pode ter causado um maior impacto na atenção que lhe foi dada em relação às outras células. Cañadas, Díaz, Batanero e Arteaga (2011) afirmam que em alguns trabalhos sobre julgamentos de associação estatística em tabelas de contingência de  $2 \times 2$ , desenvolvido por outros autores, um dos fatores que influenciou os julgamentos de associação foi o valor da célula [a] – “o valor da célula [a] parece ter maior impacto nas contingências estimadas” (p. 298). Para além deste fator, observou-se também nas respostas dos estudantes o fenómeno da *correlação ilusória* e o recurso a *teorias*

*prévias* sobre a relação existente entre *fumar* e *ter bronquite*, além da confusão que alguns estudantes fizeram entre *causalidade e associação*. Por exemplo, o estudante E22 considerou haver dependência com base no seu conhecimento sobre os efeitos negativos do tabaco para a saúde. Também Batanero et al. (1998) chegaram a resultados semelhantes nos seus estudos sobre o significado da associação estatística.

Notamos também na utilização da informação de uma única célula que houve igualmente estudantes que utilizaram a célula *c* (*estratégia 1.1b*) adotando uma perspectiva determinista de associação (Batanero et al., 1996), na medida em que eles esperavam que houvesse uma correspondência que atribuísse um único valor à variável *bronquite* para cada valor da variável *fumar*.

As estratégias mais frequentes nos dois momentos do estudo foram a comparação de duas frequências absolutas ou relativas entre si, sendo que destas estratégias resultaram quase sempre respostas incorretas. No primeiro caso, os estudantes compararam dois grupos distintos com base em frequências absolutas, considerando que quando as frequências absolutas são distintas existe associação entre duas variáveis. Este erro foi também descrito por Batanero et al. (1996.), argumentando-se que ele se deve a uma incompreensão do conceito de distribuição, pois os estudantes compararam distribuições utilizando frequências absolutas. No segundo caso, os estudantes calcularam frequências relativas de forma isolada e não relacionaram a frequência relativa condicional de uma variável com a da outra variável.

Comparativamente com as outras duas estratégias, na utilização das quatro células verificou-se uma maior proporção de respostas corretas, tendo a maior parte recorrido à comparação de frequências relativas (nível 5 de Pérez Echevarría (1990)). As estratégias mais utilizadas, tanto no pré-ensino como no pós-ensino, foram as de comparação de proporções em linha e/ou coluna e a de comparação de duas ou mais frequências relativas condicionadas.

De uma forma geral, concluímos que quando o estudante usava a informação de uma única célula, a tendência foi responder ao acaso; quando era usada a informação de duas células, a tendência foi responder erradamente; e quando era usada a informação de quatro células, a tendência foi responder corretamente.

No pós-ensino verificou-se uma ligeira diminuição no recurso a uma e duas células e um aumento da utilização de quatro células, constituindo esta última uma estratégia mais sofisticada que conduziu a respostas corretas, o que leva a concluir que o ensino da associação estatística produziu alterações limitadas nesta situação relativa à tabela de contingência.

Notou-se também nos estudantes uma evolução da forma de avaliar a associação, sobretudo na percepção de que a avaliação da associação estatística entre duas variáveis se devia fazer em termos das frequências relativas condicionadas e não através da comparação de frequências absolutas ou relativas entre si, como foi mais frequente no pré-ensino.

Da análise das estratégias utilizadas pelos estudantes na avaliação da associação na comparação de amostras com variáveis numéricas emparelhadas ou não emparelhadas, concluímos que alguns estudantes, sobretudo no pós-ensino, apresentaram estratégias normativas para comparar dois grupos, nomeadamente recorrendo à utilização da média, dos totais e das proporções, o que muitas vezes lhes permitiu obter a resposta correta sobre a associação entre as variáveis. Os estudantes que compararam as médias nas duas amostras independentes conceberam a média como um valor representativo dos dados, o que demonstra que eles possuem uma estrutura conceptual da média que lhes permite resolver um conjunto de problemas com ela relacionados. O mesmo pode ser dito para os estudantes que compararam os totais, onde igualmente consideraram os totais como valores representativos de conjuntos de dados.

Uma das estratégias, que foi muito utilizada tanto no pré-ensino como no pós-ensino, consistiu na comparação qualitativa global das distribuições sem serem especificadas as estatísticas usadas. Apesar dos estudantes não especificarem as estatísticas usadas para decidir sobre a associação, implicitamente notou-se que eles reconheciam a importância da utilização de todos os dados, que é um dos passos básicos para o julgamento de associação estatística (Estepa et al., 1999).

No caso das duas questões que envolviam amostras emparelhadas (questões 3 e 4) verificou-se, apenas no pós-ensino, que uma percentagem considerável de estudantes recorreu à determinação e interpretação do coeficiente de correlação de Pearson para decidir sobre a associação estatística entre as variáveis. A adoção desta estratégia, que teve um grande impacto nas percentagens de respostas corretas, era expectável por se tratar de um conteúdo que geralmente é muito enfatizado durante o ensino.

Para além das estratégias corretas, observaram-se algumas estratégias incorretas, designadamente consistindo na avaliação de coincidências das duas distribuições, na comparação de alguns valores de ambas as amostras e na comparação de valores mais elevados e mais baixos de ambas distribuições. Em todos estes casos, os estudantes utilizaram apenas parte dos dados, adotando assim uma conceção local de associação. Todavia, houve

também estudantes que manifestaram uma concepção determinista de associação ao não admitirem exceções para haver associação, o que também foi verificado no trabalho de Estepa et al. (1999).

A avaliar pelo número de respostas corretas no pós-ensino, onde se notou um aumento significativo dessas respostas, sobretudo pela adesão às estratégias de análise da variação dos pontos no gráfico de dispersão e da determinação e análise do coeficiente de correlação, podemos concluir que houve alguma influência do ensino na mudança das concepções de associação estatística, como seria de esperar. No entanto, registou-se ainda alguma resistência em utilizar procedimentos mais institucionalizados por parte de outros estudantes, que mesmo depois de terem tido aulas sobre a determinação e interpretação da associação estatística entre duas variáveis não as utilizaram.

Por último, no que concerne às estratégias de associação quando a situação era apresentada através de um diagrama de dispersão, identificámos estratégias que conduziram a respostas corretas e parcialmente corretas, que são indicadores de concepções corretas de associação estatística, e estratégias incorretas. Tal como nas situações anteriores, verificou-se no pré-ensino um maior número de respostas incorretas em resultado do recurso dos estudantes a *pontos isolados* para decidirem sobre o tipo de correlação. Já no pós-ensino verificou-se uma ligeira diminuição da percentagem de respostas incorretas, acompanhada do correspondente aumento das respostas corretas, a partir da adoção das estratégias: *avaliação da tendência de variação dos pontos do gráfico de dispersão (variação no mesmo sentido, variação em sentido contrário ou ausência de variação)* e *comparação do diagrama de dispersão com o gráfico de uma função conhecida*. As estratégias incorretas permitiram identificar a adesão, por parte dos estudantes, a diferentes concepções incorretas, especificamente a concepção determinista de associação, a concepção local de associação e a concepção causal, concepções que foram igualmente encontradas em estudantes do final do ensino secundário por Estepa e Batanero (1996).

Em ambas os tipos de problemas – avaliação da associação na comparação de duas amostras e avaliação da associação em gráficos de dispersão – os estudantes, ainda que em menor número, recorreram às suas crenças e expectativas sobre a relação entre as variáveis para decidirem sobre a existência ou não de associação, em detrimento da análise empírica dos dados apresentados no problema. Na adesão a este fenómeno, identificado por Chapman e Chapman (1969) nos seus estudos, e que designaram por *correlação ilusória*, as pessoas

mantêm as suas crenças sobre a associação entre variáveis apesar da evidência empírica suportar a independência das variáveis.

### **5.3. Quais as aquisições conceituais sobre a associação estatística entre duas variáveis que os estudantes adquirem com o ensino do tema correlação e regressão linear?**

Fazendo uma análise global das percentagens de respostas corretas (71,2%), concluímos que os significados adquiridos pelos estudantes sobre a correlação e regressão linear estão de acordo com os significados institucionais previstos na abordagem do tema.

Seguidamente, de modo a perceber as componentes essenciais do significado matemático das noções de correlação e regressão linear adquiridas pelos estudantes, fez-se uma análise exhaustiva às respostas dadas pelos estudantes às 8 questões inseridas um teste, centrada nos seguintes temas: (i) propriedades da covariação, do coeficiente de correlação, assim como da relação que os estudantes estabelecem entre estes conceitos; (ii) estimação da correlação a partir da representação gráfica, bem como a construção de um diagrama de dispersão a partir de pares de variáveis; e, por último, (iii) cálculo do coeficiente de correlação a partir de dados apresentados numa tabela e estimação do valor desconhecido de uma variável a partir do valor conhecido de outra variável através da reta de regressão. A seguir são apresentadas as principais conclusões tiradas na análise efetuada, organizadas segundo estes temas.

No que respeita às propriedades da covariação, do coeficiente de correlação e da relação entre estes conceitos, verificou-se que a grande maioria dos estudantes (76,1%) respondeu corretamente, o que permite concluir que os significados adquiridos pelos estudantes sobre a correlação e regressão linear estão de acordo com os significados institucionais previstos na abordagem do tema.

Nas situações apresentadas neste tema, a maioria dos estudantes mostrou compreender com facilidade a relação entre o sinal da correlação e o sentido da covariação e relacionar corretamente a correlação nula com a independência das variáveis e a correlação perfeita com a dependência funcional, envolvendo neste último caso um valor do coeficiente de correlação correspondente de 1 ou  $-1$ .

Contudo, houve algumas propriedades nas quais os estudantes tiveram maior dificuldade, destacando-se a relação entre a correlação e a proporcionalidade, onde 18 (31,6%) dos estudantes admitiram que a correlação entre duas variáveis duplica se duplicarmos o seu

coeficiente de correlação. Neste caso, os estudantes interpretaram o coeficiente de correlação como se tivesse um papel análogo ao da constante de proporcionalidade, o que nem sempre é válido. Outra dificuldade encontrada, agora revelada por mais de metade dos estudantes, diz respeito à identificação de uma correlação linear a partir do valor da covariação, parecendo que os estudantes não conhecem outros tipos de correlação, para além das lineares. Este facto releva a pouca atenção que tem sido dada às relações não lineares durante o estudo do tema correlação e regressão lineares.

No caso da estimação do valor do coeficiente de correlação a partir de um diagrama de dispersão e da construção de um diagrama de dispersão para representar razoavelmente a covariação entre duas variáveis, apesar dessas situações não serem muito habituais no ensino da associação estatística e regressão linear, os resultados obtidos demonstram que a maioria dos estudantes (82,2%) revelou uma boa capacidade na resolução destas tarefas.

Quanto às respostas dos estudantes na estimação do coeficiente de correlação a partir do diagrama de dispersão, verificámos existir na maioria das respostas concordância entre o valor do coeficiente de correlação estimado, o sentido de covariação entre as variáveis e a intensidade de correlação. A maior parte dos estudantes justificaram as suas estimativas tendo em conta a dispersão das nuvens de pontos dadas, ou seja, o ajustamento ou não ajustamento a uma reta. Dai, nestas situações, podemos inferir que a correlação ilusória, referida por Chapman e Chapman (1969), não esteve presente nos raciocínios dos estudantes em estudo.

Em relação à construção do diagrama de dispersão a partir de pares de variáveis dadas, verificou-se também uma boa capacidade dos estudantes em representar coerentemente a covariação de duas variáveis num diagrama de dispersão, sobretudo em relação ao sentido da correlação. As suas justificações foram quase sempre baseadas no conhecimento prévio da existência, ou não, de relação entre as variáveis. Todavia, observou-se que quase metade dos estudantes que representaram corretamente o sentido da relação revelaram a propensão para desenhar um gráfico de correlação perfeita (que não era o caso) entre as variáveis, atribuindo-lhe uma natureza determinista (Estepa & Batanero, 1996), tal como aconteceu na alínea 6a), envolvendo as variáveis *número de operários* e *tempo de pintura de uma habitação*.

Quando as variáveis eram interdependentes, observou-se que os estudantes representavam diferentes gráficos com base nas suas convicções. Todavia, verificámos que independentemente dessa variedade de respostas, elas estavam em consonância com o

diagrama representado, o que revela uma consistência na representação e interpretação da covariação entre duas variáveis.

No que diz respeito às questões do último tema, verificámos que elas se revelaram mais difíceis para os estudantes. Neste caso, menos de metade dos estudantes (44,9%) apresentou respostas corretas no cálculo e interpretação do coeficiente de correlação e na estimação do valor desconhecido de uma variável a partir do valor conhecido de outra variável através da reta de regressão.

No cálculo do coeficiente de correlação, a grande maioria dos estudantes fê-lo corretamente aplicando a respetiva fórmula, enquanto muito poucos foram capazes de relacionar o valor do coeficiente obtido com a intensidade e o sentido da associação. Já no que diz respeito à determinação de reta de regressão que melhor se ajusta a uma distribuição bidimensional, tendo em vista efetuar a predição de uma variável a partir da outra, as maiores dificuldades que os estudantes enfrentaram resultaram do facto de utilizarem uma única reta de regressão, sobretudo a de  $y$  sobre  $x$ , para efetuarem previsões de  $y$  a partir de  $x$  e de  $x$  a partir de  $y$ . Por exemplo, para estimarem  $x$  a partir de  $y$ , a grande maioria dos estudantes recorreu à equação de  $y$  sobre  $x$ , que é um procedimento incorreto e um indicador de uma conceção funcional de regressão linear (Estepa & Batanero, 1996). A esta dificuldade não terá sido certamente alheia a confusão que os estudantes fizeram entre os papéis desempenhados pelas variáveis dependente e independente no estabelecimento da equação de regressão linear.

Finalmente, quanto aos perigos de extrapolação de conclusões para além dos dados efetivamente observados, os sujeitos desta amostra foram capazes de concluir *a priori* que não era seguro extrapolar para um dado que estava muito distante do intervalo dos valores observados.

Considerando as aquisições conceituais globais reveladas pelos estudantes no teste, aqui relatadas, podemos concluir que elas foram satisfatórias, sobretudo quando comparadas com as aquisições mais intuitivas reveladas pelos estudantes no questionário, referidas na primeira questão de investigação. O reduzido impacto do ensino sobre o substrato intuitivo, por um lado, é também referido em outros estudos (e.g., Fernandes, 1990; Fischbein & Gazit, 1984) e, por outro lado, também se pode dever à pouca ênfase dada a certos conteúdos no ensino da correlação e regressão linear, como foi o caso da associação estatística em tabelas de contingência de  $2 \times 2$ .

#### 5.4. Implicações do estudo para o ensino

O presente estudo focalizou-se em determinar as concepções e estratégias dos estudantes sobre a avaliação associação estatística e regressão linear. Nesta secção faremos uma breve reflexão sobre as implicações dos resultados obtidos para o ensino, sobretudo quando contrariam os significados institucionais, e possíveis sugestões de investigação para a sua melhoria.

As dificuldades sentidas pelos estudantes na avaliação da associação estatística entre duas variáveis, representadas numa tabela de contingência, num gráfico de dispersão ou através da comparação de amostras dadas numa tabela, sugerem que se deve dar oportunidade aos estudantes de enfrentar situações problemáticas que lhes mostrem uma diversidade de tipos de covariação e tipos de dependência entre duas variáveis. Tal poderia contribuir para eliminar certas concepções erradas dos estudantes, que foram encontradas nesta investigação e nas investigações realizadas por Batanero e colaboradores.

Deve procurar, sempre que possível, efetuar-se uma aproximação do tema de correlação e de regressão linear de modo a clarificar a confusão que os estudantes normalmente fazem entre a associação e causalidade, pois nem sempre a associação forte entre duas variáveis implica uma relação de causa e efeito entre elas (Garfield, 1998).

Ainda que não se tenham verificado muitas dificuldades concernentes ao cálculo do coeficiente de correlação, é importante que se dê especial atenção à conexão deste conceito com outros conceitos estatísticos que jogam um papel destacado neste tema, tais como a covariância, o tipo de dependência (linear e não linear) e a dispersão de dados.

Dada a dificuldade que os sujeitos da amostra exibiram na hora de distinguir entre variável independente e dependente, sugerimos que se enfatize o ensino de modo a discriminar de forma conveniente a variável explicativa da variável explicada. Em consequência disto, pode esperar-se que os estudantes melhorem a sua capacidade de distinção entre ambas as retas de regressão linear, sendo mais conscientes da utilização da reta de regressão de  $X$  sobre  $Y$  ou de  $Y$  sobre  $X$ , consoante o modelo estatístico adequado à situação em estudo.

Por outro lado, também é importante que as tarefas estimulem a exploração de aspetos fundamentais da educação estatística, como a interpretação, a representação gráfica, a predição, a comparação, etc., pois considerando que “um bom raciocínio estatístico é tao necessário para exercer uma cidadania eficiente como a capacidade de ler e escrever”

(Campbell, 1981, p. 14), devem ser propostos aos estudantes problemas que sejam catalisadores deste tipo de raciocínio.

### **5.5. Limitações e recomendações de futuras investigações**

Em termos das limitações do estudo, salienta-se o recurso a uma amostra de estudantes de dimensão não muito elevada. A inexistência de mais estudantes, nas condições do estudo nas duas Delegações da Universidade Pedagógica a que tivemos acesso, impossibilitou a participação no estudo de um maior número de estudantes.

No que respeita a futuras investigações, considerando que os estudantes que participaram na presente investigação tinham estudado o tema da correlação e regressão linear apenas na universidade, recomenda-se a realização de outros estudos em que os estudantes tenham estudado esse tema mais cedo, ainda que a um nível mais intuitivo. Um tal estudo mais cedo do tema na escola pode contribuir para o desenvolvimento de intuições mais consonantes com o significado institucional da associação estatística, até porque alguns autores (e.g., Fischbein, 1975; Shaughnessy, 1992) consideram que a abordagem tardia na escola dos conceitos de probabilidades e estatística favorece o surgimento e a consolidação de conceções erróneas.

Na sequência do presente estudo, tendo em conta as dificuldades sentidas pelos estudantes e o impacto limitado do ensino regular do tema correlação e regressão linear nas estratégias e conceções de avaliação da associação estatística, justifica-se a realização de investigações focadas na experimentação de estratégias de ensino especialmente dirigidas à superação das dificuldades e erros revelados pelos estudantes.

## BIBLIOGRAFIA

- Abrantes (1986). *Porque ensinar Matemática: perspectivas e concepções de professores e futuros professores de matemática*. Lisboa: Associação de Professores de Matemática.
- Allyn & Bacon.
- Ausubel, D. P., Novak, J. D. & Hanesian, H. (1983). *Psicología educativa. Un punto de vista cognoscitivo*. México: Trilhas.
- Azevedo, A. (1993). *O computador no ensino da matemática: Uma contribuição para o estudo das concepções e práticas dos profesores*. Dissertação de mestrado, Universidade Nova de Lisboa. Lisboa: Associação de Professores de Matemática.
- Batanero, C. (2000). Dificultades de los estudiantes en los conceptos estadísticos elementares: El caso de las medidas de posición central. In C. Loureiro, O Oliveira, & L. Brunheira (Orgs.), *Ensino e aprendizagem de estatística* (pp.31-48). Lisboa: Sociedade Portuguesa de Estatística, Associação de Professores de Matemática e Departamento de Educação e de Estatística e Investigação Operacional da Faculdade de Ciências da Universidade de Lisboa.
- Batanero, C. (2001). *Didáctica de la estadística*. Granada: Universidade de Granada.
- Batanero, C., Estepa, A., Godino, J. D., & Green, D. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, 27(2), 151-169.
- Batanero, C., Godino J. D., Vallecillos, A., Green, D. R., & Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematics Education in Science and Technology*, 25(4), 527-547.
- Batanero, C., Godino, J., & Estepa, A. (1998). Building the meaning of statistical association through data analysis activities. In A Oliver., & K. Newslead (Eds.), *Proceedings of the 22nd Conference of the International Group for the Psychology of Mathematics Education*, (v. 1, pp. 221-236). University of Stellenbosch, South Africa.
- Beyth, R. (1982). Perception of correlation reexamined. *Memory and Cognition*, 10(6), 511-519.
- Bogdan, R. C., & Biklen, S. K. (1994). *Investigação qualitativa em educação: Uma introdução à teoria e aos métodos*. Porto: Porto Editora.
- Branco, J. (2000). Estatística no secundário: o ensino e seus problemas. In C. Loureiro, F. Oliveira, & Brunheira (Eds.), *Ensino e Aprendizagem da Estatística* (pp.11-30).Lisboa:

Sociedade Portuguesa de Estatística, Associação de Professores de Matemática e Departamento de Educação e de Estatística e Investigação Operacional da Faculdade de Ciências da Universidade de Lisboa.

- Campbel, S. K. (1981). *Equívocos y falacias en la interpretación de estadísticas*. Naucalpan de Juárez, México: Limusa.
- Cañadas, G., Díaz, C., Batanero, C., & Arteaga, J. P. (2011). Estimación por estudiantes de psicología. In M. Mari, G. Fernández, L. J. Blanco & M. M. Palarea (Eds.), *Investigación Matemática XV* (pp. 297-306). Ciudad Real: Sociedad Española de Investigación en Educación Matemática (SEIEM).
- Canavarro, A. (1993). *Concepções e práticas de professores de matemática: Três estudos de caso*. Universidade de Lisboa. Lisboa: Associação de Professores de Matemática.
- Chapman, L. J., & Chapman, J. P. (1969). Illusory correlations as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74*, 271-280.
- Cobo, F. T. S. (1998). *Significado de correlación y regresión para los estudiantes universitarios*. Tese de doutoramento, Universidade de Granada, Granada.
- Curcio, F.R. (1989). *Developing graph comprehension*. Reston, VA: National Council of Teachers of Mathematics.
- Estepa, A., & Batanero, C. (1996). Judgments of correlation in scatter plots: Students' intuitive strategies and preconceptions. *Hiroshima Journal of Mathematics Education, 4*, 25-41.
- Estepa, A., Batanero, C., & Sánchez, F. T. (1999). Students' intuitive strategies in judging association when comparing two sample. *Hiroshima Journal of Mathematics Education, 7*, 17-30.
- Falk, R., & Well, A. D. (1997). Many face of the correlation coefficient. *Journal of Statistics Education, 5*(3).
- Fernandes, J. A. (1990). *Concepções erradas na aprendizagem de conceitos probabilísticos*. Dissertação de mestrado, Universidade do Minho, Braga, Portugal.
- Fernandes, J. A. (2004). *Tecnologia no ensino da Matemática: Aplicação da calculadora gráfica no estudo das probabilidades*. Braga: Centro de Formação Prof. Agostinho Manuel da Silva.
- Fernandes, J. A. (2009). Ensino e aprendizagem da estatística: Realidades e desafios. In C. Costa, E. Mamede & F. Guimarães (Orgs.), *Actas do XIX Encontro de Investigação em*

*Educação Matemática*. Vila Real: Secção de Educação Matemática da Sociedade Portuguesa de Ciências da Educação.

- Fernandes, J. A., & Morais, P. C. (2011). Leitura e interpretação de gráficos estatísticos por alunos do 9º ano de escolaridade. *Educação Matemática Pesquisa*, 13(1), 95-115.
- Fernandes, J. A., Carvalho, C. F., & Correia, P. F. (2011). Contributos para a caracterização do ensino da Estatística nas escolas. *Boletim de Educação Matemática (BOLEMA)*, 24(39), 585-606.
- Fernandes, J. A., Carvalho, C., & Ribeiro, S. A. (2007). Caracterização e implementação de tarefas de Estatística: Um exemplo no 7.º ano de escolaridade. *Zetetiké*, 15(28), 27-61.
- Fernandes, J. A., Sousa, M. V., & Ribeiro, S. A. (2004). Ensino e aprendizagem de Probabilidades e Estatística. Braga: Centro de Investigação em Educação da Universidade do Minho.
- Fischbein, E. & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? *Educational Studies in Mathematics*, 15, 1-24.
- Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht: D. Reidel Publishing Company.
- Freudenthal, H. (1991). *Revisiting mathematics education*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Friel, S., Curcio, F., & Bright, G. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2), 124-155.
- Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational Research: An Introduction*. Boston:
- Garfield, J. B. (1998). Assessing statistical reasoning. *Statistics Education Research Journal*, 2(1), 22-38.
- Godino, J. D. (1996). Significado y comprensión de los objectos matemáticos. In L. Puig & A. Gutiérrez (Orgs.), *Proceedings of 20th PME Conference*, (v. 2, pp. 417- 424). Valencia.
- Godino, J. D., & Batanero, C. (1994). Significado institucional y personal de los objetos matemáticos. *Recherches en Didactique des Mathématiques*, 14(3), 325-355.
- González, J. L., Iriarte, M. D., Jimeno, M., Ortiz, A., Sanz, E., Vargas-Machuca, I (1990). *Números enteros*. Madrid: Editorial Síntesis.
- Guimarães, H. M. (1988). *Ensinar matemática: concepções e práticas*. Dissertação de mestrado, Universidade de Lisboa. Lisboa: Associação de Professores de Matemática.

- Guimarães, H. M. (2010). Concepções, crenças e conhecimento — afinidades e distinções essenciais. *Quadrante*, 19(2), 81-101.
- Inhelder, B., & Piaget, J. (1976). *Da lógica da criança à lógica do adolescente: Ensaio sobre a construção das estruturas operatórias formais*. São Paulo: Livraria Pioneira Editora.
- Jennings, D. L., Amabile, T. M., & Ross, L. (1982), Informal covariation assessment: Data-base versus theory-based judgments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgments under uncertainty: Heuristic and biases* (pp. 211-230). New York, NY: Cambridge University Press.
- Jolliffe, F (2007). The changing brave new world of statistics assessment. In *The proceedings of the ISI/IASE satellite on assessing student learning in statistics*. Voorburg: International Statistical Institute.
- Landwehr, J. & Watkins, A. E. (1986). *Exploring data*. Palo Alto, CA: Dale Seymour.
- Menezes, L. (1995). *Concepções e práticas de professores de matemática: Contributos para o estudo da pergunta* (Dissertação de mestrado, Universidade de Lisboa). Lisboa: Associação de Professores de Matemática.
- Ministério da Educação (2008). *Programa do ensino básico*. Maputo: MINED.
- Ministério da Educação (2010). *Programa de Matemática (8.ª- 12.ª Classe)*., Maputo: MINED.
- NCTM (1994). *Normas profissionais para o ensino da matemática*. Lisboa: Associação de Professores de Matemática/Instituto de Inovação Educacional.
- Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcoming of social judgments*. Englewood Cliffs, NJ: Prentice-Hall.
- Oliveira, H. (1998). *Actividades de investigação na aula de Matemática: Aspectos da prática do professor* (Dissertação de mestrado, Universidade de Lisboa). Lisboa: Associação de Professores de Matemática.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Journal Review of Educational Research*, 62(3), 307-332.
- Pérez Echevarria, M. P. (1990). *Psicología del razonamiento probabilístico*. Madrid: Universidad Autonoma de Madrid.
- Ponte, J. P. & Fonseca, H. (2000). A Estatística no currículo do Ensino Básico e Secundário. In C. Loureiro, F. Oliveira & Brunheira (Eds.), *Ensino e Aprendizagem da Estatística* (pp.179-194).Lisboa: Sociedade Portuguesa de Estatística, Associação de Professores de

Matemática e Departamento de Educação e de Estatística e Investigação Operacional da Faculdade de Ciências da Universidade de Lisboa.

- Pozo, J. I. (1987). *Aprendizaje de la ciencia y pensamiento causal*. Madrid: Visor.
- Rodrigues, E. F. (1993). *Perspectivas dos professores sobre o ensino da Matemática* (Dissertação de mestrado, Universidade de Lisboa). Lisboa: Associação de Professores de Matemática.
- Sánchez, F. S. (1996). *Análisis de la exposición teórica y de los ejercicios de correlación y regresión en los textos de bachillerato*. Memoria de Tercer Ciclo. Departamento de Didáctica de la Matemática. Universidade de Granada, Granada.
- Schild, M. (2006). Statistical literacy Survey analyses: Reading graphs and tables of rates and percentages. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. Cape Town: International Association for Statistical Education.
- Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (ed.), *Handbook of research on mathematics teaching and learning* (pp. 465-494). New York: Macmillan Publishing Company.
- Smedlund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology*, 4, 165-174.
- Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grows (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-146). Nova York, NY: MacMillan.
- Thompson (1982). *Teacher's beliefs and conceptions of Mathematics and mathematical teaching: Three case studies*. Tese de Doutoramento, Universidade de Geórgia, Geórgia.
- Truran, J.M. (1997). Understanding of association and regression by first year economics students from different countries as revealed in responses to the same examination questions. In J. B. Garfield & J. M. Truran (Eds.), *Research Papers on Stochastics from 1997*. Department Educational Psychology University of Minnesota.
- Tversky, A. & Kahneman, D. (1982). Causal schemas in judgments under uncertainty. In D. Kahneman, P. Slovic e A. Tversky (Eds), *Judgments under uncertainty: heuristics and biases* (pp. 117-128). Cambridge University Press.
- Vergnaud, G. (1982). Cognitive and developmental psychology and research in mathematics education: Some theoretical and methodological issues. *Learning of Mathematics*, 3(2), 31-41.



## **ANEXO**

Questionário e Teste



## Questionário

Caro/a Estudante

No âmbito da realização de um estudo de investigação sobre Correlação e Regressão Linear, venho pedir a tua colaboração para responder às seis questões que a seguir são apresentadas. Estas questões têm por finalidade principal identificar e caracterizar as ideias usadas pelos estudantes na resolução de problemas sobre Correlação e Regressão Linear.

Embora as tuas respostas não sejam usadas para te atribuir qualquer classificação à disciplina de Estatística Descritiva, é muito importante que leias cuidadosamente todas as questões e que respondas a todas as perguntas do questionário com sinceridade e empenho, explicando claramente a forma como pensaste. Nessas justificações podes utilizar números, cálculos, palavras, etc.

As tuas respostas às questões serão mantidas confidenciais e eu, enquanto única pessoa com acesso aos dados, comprometo-me a não divulgar as respostas a não ser para fins do estudo e sempre sob a forma de anonimato.

Muito obrigado pela colaboração.

Delson Mugabe

---

### I – DADOS PESSOAIS

Nome: \_\_\_\_\_

Sexo:  Masculino  Feminino

Em que subsistema terminaste o nível Médio?

Secundário  Técnico Profissional  Formação de professores

Outro. Qual? \_\_\_\_\_

Estudaste algumas noções de Estatística no ensino Médio?  Sim  Não

Se respondeste Sim, em que classe? \_\_\_\_\_

Classificação que obtiveste no exame de admissão à Universidade na disciplina de Matemática: \_\_\_\_\_

Qual o interesse da Estatística para tua formação?

Muito Pouco  Pouco  Algum  Muito

## II – QUESTÕES

1. Num centro médico, foram observadas 250 pessoas para determinar se o hábito de fumar tem alguma relação com a doença da bronquite, tendo-se obtido os resultados da tabela seguinte.

	Tem bronquite	Não tem bronquite	Total
Fuma	90	60	150
Não fuma	60	40	100
Total	150	100	250

Usando a informação da tabela, achas que nesta amostra de pessoas a doença da bronquite depende de fumar?

Explica a tua resposta:

2. Depois de medir o nível de açúcar no sangue de crianças do sexo masculino e feminino de uma escola, obtiveram-se os dados da tabela seguinte.

	Crianças																			
	do sexo masculino										do sexo feminino									
Nível de açúcar no sangue	9	0	9	8	6	7	4	9	8	9	6	0	7	0	8	3	6	7	7	3

Usando a informação do tabela, achas que o nível de açúcar no sangue nesta amostra depende do sexo das crianças?

Explica a tua resposta:

**3.** Mediu-se a pressão sanguínea num grupo de 10 mulheres, antes e depois de se aplicar um tratamento médico, tendo-se obtido os dados da tabela seguinte.

Mulher	A	B	C	D	E	F	G	H	I	J
Pressão sanguínea antes do tratamento	115	112	107	119	115	138	126	105	104	115
Pressão sanguínea depois do tratamento	128	115	106	128	122	145	132	109	102	117

Usando a informação da tabela, achas que a pressão sanguínea nesta amostra de mulheres depende do tempo em que foi medida (antes ou depois do tratamento)?

Explica a tua resposta:

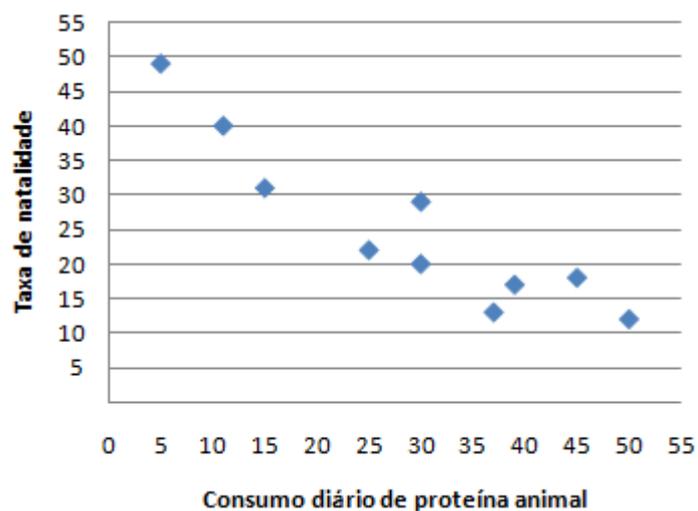
4. Na tabela seguinte indicam-se as taxas anuais de casamento e divórcio, por cada 1000 pessoas, em alguns países europeus.

País	Taxa de casamento	Taxa de divórcio
Áustria	5.6	2.0
Dinamarca	6.0	3.0
Finlândia	5.1	2.9
França	5.0	1.9
Alemanha	6.7	2.0
Hungria	6.3	2.4
Itália	5.4	0.4
Holanda	6.1	1.9
Noruega	4.9	2.2
Polónia	6.8	1.3
Suécia	5.2	2.2
Suíça	6.8	2.0
Reino Unido	6.1	2.9

Achas que existe ou não associação entre as taxas de casamento e de divórcio, isto é, que a taxa de divórcio dependa da taxa de casamento?

Explica a tua resposta:

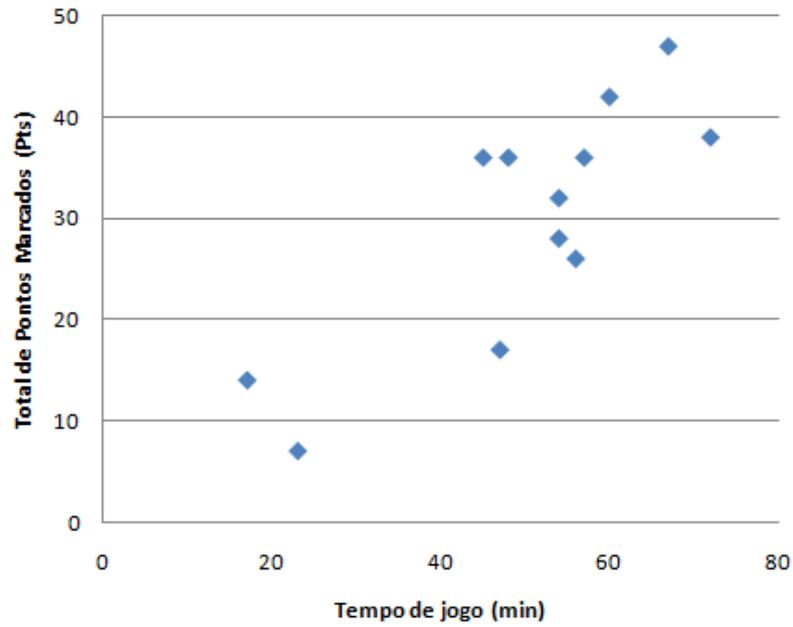
5. Num estudo sociológico foram recolhidos dados relativos ao consumo diário de proteína animal e à taxa de natalidade em diferentes países. No gráfico seguinte estão representados os dados obtidos.



- Em quantos países foram recolhidos dados relativos ao consumo diário de proteína animal e à taxa de natalidade?
- Observando o gráfico, que tendência existe entre o consumo diário de proteína animal e a taxa de natalidade?
- Achas que a taxa de natalidade nos diferentes países depende do consumo diário de proteína animal?

Explica a tua resposta:

6. No gráfico seguinte estão representados os pontos obtidos por cada jogador da selecção de basquetebol dos Estados Unidos da América nos Jogos Olímpicos de Inverno de 1992 e o seu tempo de jogo em minutos, no conjunto dos três jogos finais.



- a) Assinala no gráfico o ponto que representa o jogador que marcou mais pontos. Achas que esse jogador também jogou durante mais tempo?

- b)** Achas que existe uma relação entre o tempo de jogo (em minutos) do jogador e o total de pontos marcados?

Explica a tua resposta:

- c)** Se um jogador joga durante 50 minutos, quantos pontos é de esperar que ele marque?

---

FIM

## Teste

NOME: \_\_\_\_\_

UNIVERSIDADE: \_\_\_\_\_

Deve indicar todos os cálculos e raciocínios que realizou para obter as respostas apresentadas.

1. Sabendo-se que duas variáveis estão correlacionadas positivamente, indique se é verdadeira ou falsa cada uma das seguintes afirmações:
  - a) Quando uma aumenta, a outra também aumenta. \_\_\_\_\_
  - b) Quando uma diminui, a outra aumenta. \_\_\_\_\_
  - c) Quando uma diminui, a outra diminui. \_\_\_\_\_
  - d) A relação entre as duas variáveis é do tipo linear. \_\_\_\_\_
2. Se  $r$  é o coeficiente de correlação entre duas variáveis  $X$  e  $Y$ , indique se é verdadeira ou falsa cada uma das seguintes afirmações.
  - a) Se  $r = 0$ , as variáveis  $X$  e  $Y$  são independentes. \_\_\_\_\_
  - b) Se  $r = 0,6$ , a correlação entre as variáveis  $X$  e  $Y$  é o dobro do que quando  $r = 0,3$ .  
\_\_\_\_\_
  - c) Uma relação funcional entre as variáveis  $X$  e  $Y$  corresponde a um valor de  $r$  de  $+1$  ou  $-1$ . \_\_\_\_\_
  - d) O coeficiente de correlação pode interpretar-se como uma percentagem da variância.  
\_\_\_\_\_
3. Sabendo-se que a covariância entre as variáveis  $X$  e  $Y$  é maior do que  $0$ , indique se é verdadeira ou falsa cada uma das seguintes afirmações:
  - a) A correlação entre  $X$  e  $Y$  é positiva. \_\_\_\_\_
  - b) A correlação pode ser não linear. \_\_\_\_\_
  - c)  $X$  e  $Y$  podem não estar correlacionadas. \_\_\_\_\_
  - d) O declive da reta de regressão é positivo. \_\_\_\_\_
  - e) O coeficiente de correlação entre  $X$  e  $Y$  é positivo. \_\_\_\_\_
4. Os números  $0,5$ ,  $-0,8$ ,  $0,2$ ,  $-0,4$  e  $0$  correspondem a diferentes valores do coeficiente de correlação entre duas variáveis. Escreva esses números segundo a ordem da maior para a menor correlação que definem entre essas variáveis.

\_\_\_\_\_ Maior valor de correlação

\_\_\_\_\_

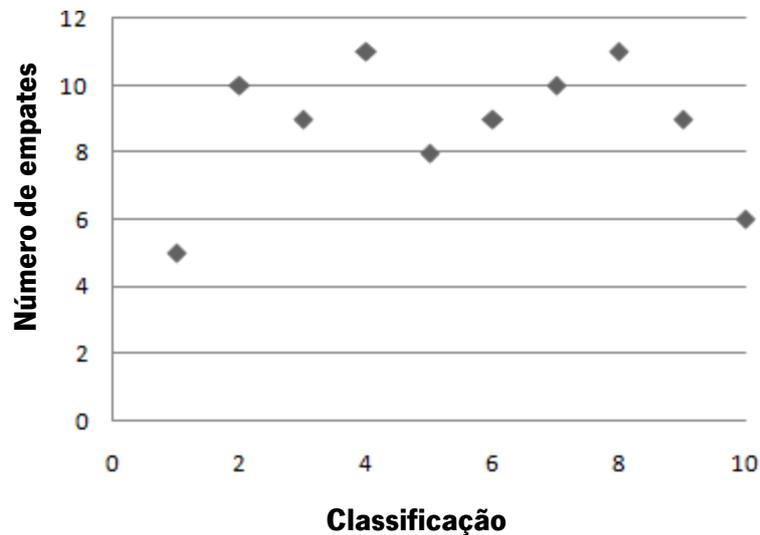
\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_ Não existe correlação

5. Dados os gráficos seguintes, que representam a variação conjunta de duas variáveis, estime o valor do seu coeficiente de correlação, sem efetuar cálculos, tendo em conta o tipo de relação – positiva ou negativa – e a sua intensidade – forte, moderada, fraca ou nula.

- a) Classificação das 10 primeiras equipas do campeonato nacional de futebol da última temporada e o número de jogos empatados.

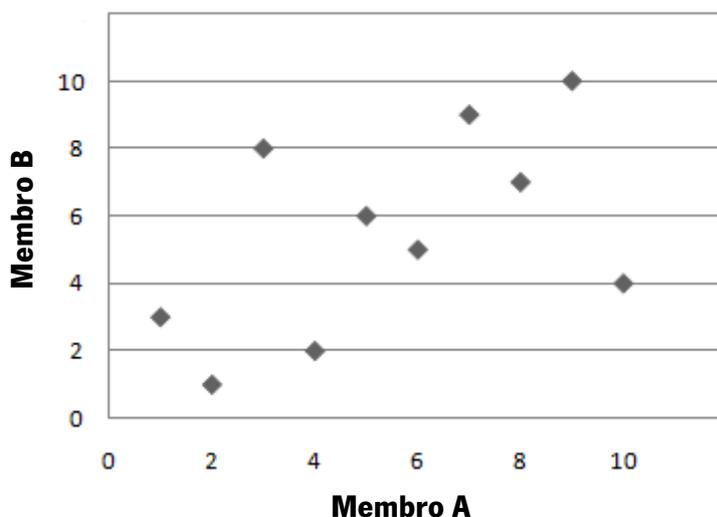


$r =$  \_\_\_\_\_ Porquê? \_\_\_\_\_

\_\_\_\_\_

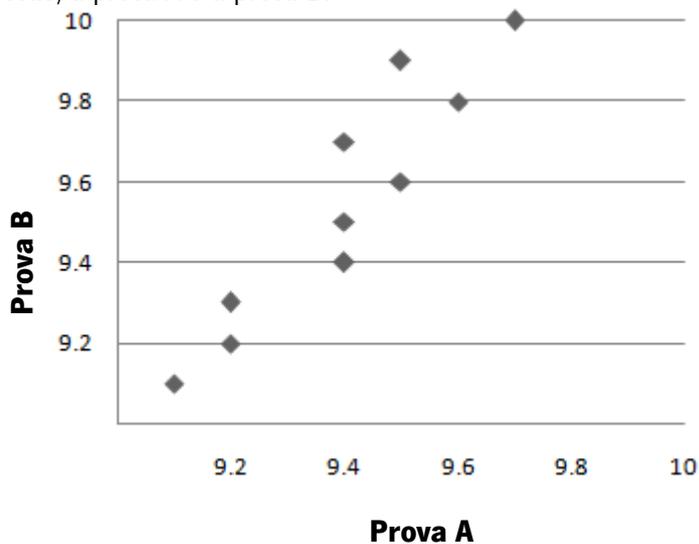
\_\_\_\_\_

**b)** Pontuações atribuídas pelos membros A e B de um júri a 10 projetos apresentados.



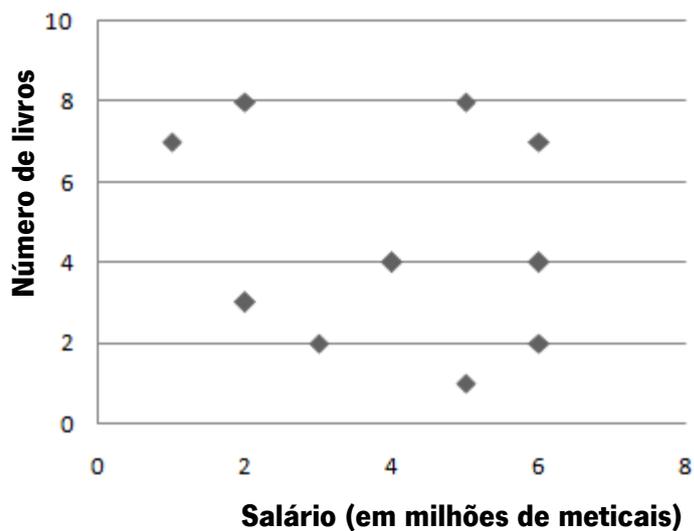
$r =$  \_\_\_\_\_ Porquê? \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_

**c)** Tempo gasto, em segundos, por 10 atletas olímpicos para correr 100 metros planos em duas provas, a prova A e a prova B.



$r =$  \_\_\_\_\_ Porquê? \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_

**d)** Salário dos funcionários de uma empresa, em milhões de meticais, e o número de livros lidos ao longo de um ano.



$r =$  \_\_\_\_\_ Porquê? \_\_\_\_\_

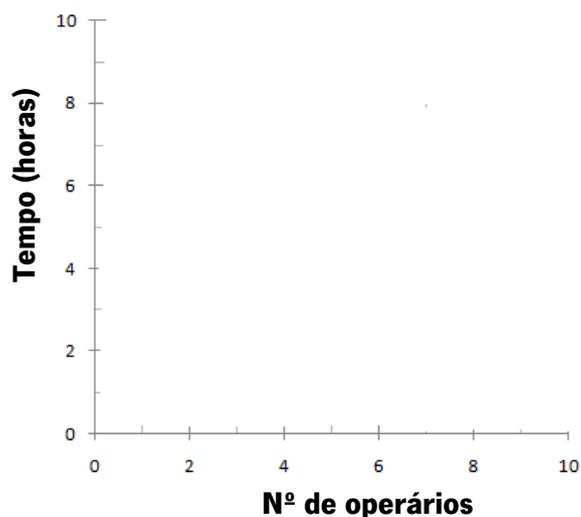
---



---

6. Dados os seguintes pares de variáveis, desenhe um diagrama de dispersão que contenha 10 pontos e que mostre razoavelmente a variação conjunta.

a) Número de operários pintando uma habitação e tempo em horas para efetuar o trabalho.



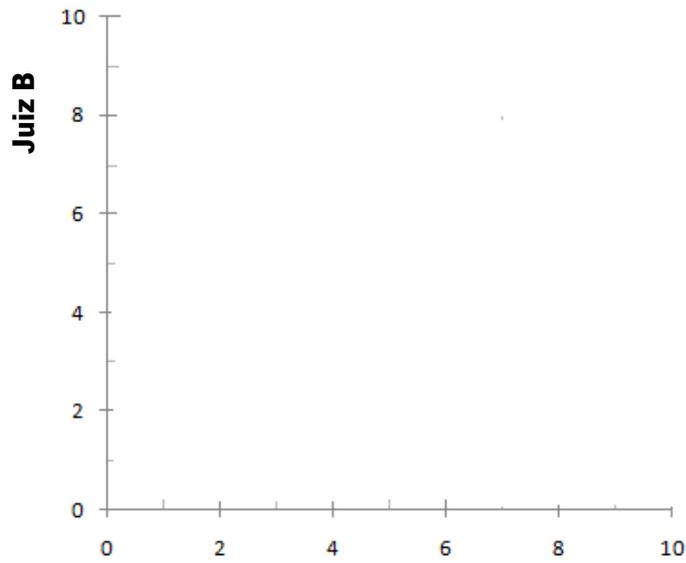
Explique a sua resposta: \_\_\_\_\_

---



---

b) Classificações (0 até 10 pontos) atribuídas por dois juizes, A e B, a 10 concorrentes num concurso de beleza.



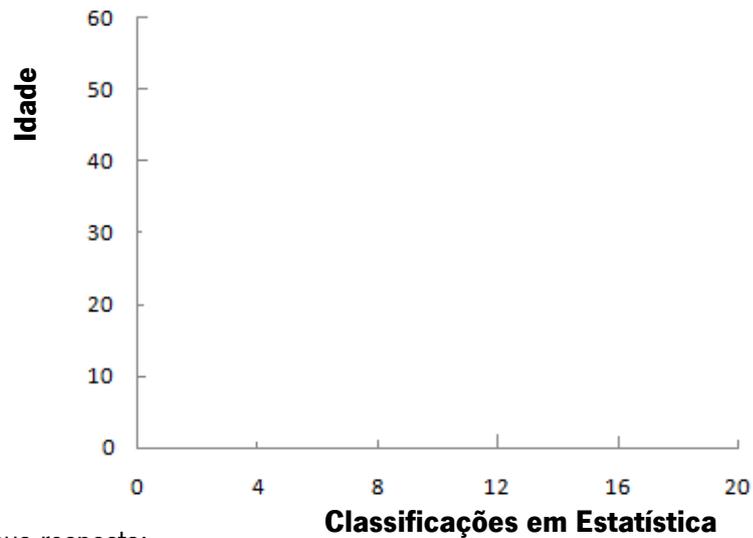
Explique a sua resposta: \_\_\_\_\_ **Juiz A** \_\_\_\_\_

---



---

**c)** Classificações em Estatística e idade do aluno.



Explique a sua resposta: \_\_\_\_\_

---



---

7. Na tabela seguinte mostram-se as classificações obtidas por 10 alunos de uma escola nos exames de Matemática e Física.

Matemática	2	4	4	6	8	8	10	12	14	14
Física	6	14	12	4	4	14	8	10	6	10

- a) Determine o valor do coeficiente de correlação  $r$ .

- b) A partir do valor de  $r$ , calculado em a), o que pode dizer-se acerca do sinal e da intensidade da correlação entre as classificações em Matemática e Física?

8. Na tabela seguinte estão registadas as idades de 20 casais à data dos seus casamentos.

Idade do Homem ( $x_i$ )	18	20	21	21	22	23	23	23	24	25	25	26	26	26	28	28	28	30	31	32
Idade da Mulher ( $y_i$ )	17	20	20	22	22	21	22	23	23	24	25	23	24	27	26	27	29	29	30	26

a) Qual é a previsão da idade da mulher à data do casamento, sabendo que o homem tinha 29 anos?

b) Qual é a previsão da idade do homem à data do casamento, sabendo-se que a mulher tinha 30 anos?

c) Será seguro prever a idade da mulher à data do casamento, sabendo que o homem tinha 45 anos? Porquê?

Muito obrigado pela colaboração  
Delson Mugabe

---

FIM