



Boletim de Educação Matemática

ISSN: 0103-636X

bolema@rc.unesp.br

Universidade Estadual Paulista Júlio de
Mesquita Filho
Brasil

Grings Sebastiani, Renate; Viali, Lori
Teste de Hipóteses: uma análise dos erros cometidos por alunos de engenharia
Boletim de Educação Matemática, vol. 24, núm. 40, diciembre, 2011, pp. 835-854
Universidade Estadual Paulista Júlio de Mesquita Filho
Rio Claro, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=291222113011>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica
Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto



Teste de Hipóteses: uma análise dos erros cometidos por alunos de engenharia

Hypothesis Test: an analysis of errors committed by engineering students

Renate Grings Sebastiani*

Lori Viali**

Resumo

O teste de hipóteses ou teoria da decisão faz parte dos conteúdos da maioria das disciplinas de Estatística Básica dos cursos universitários. Nas disciplinas dos cursos de Engenharia ele representa, aproximadamente, 20% do total de conteúdos abordados em uma disciplina padrão de quatro horas aulas semanais. Os testes são a parte mais visível da estatística em artigos científicos. Entendê-los é parte básica da literacia estatística para a compreensão da literatura acadêmica e científica. Este artigo faz um apanhado dos erros cometidos por alunos em provas e faz um comparativo com a literatura da área. Na literatura são encontrados alguns estudos sobre dificuldades e erros, mas o tema ainda carece de investigações. As questões analisadas referem-se aos testes de uma amostra para uma média (com desvio padrão desconhecido) e para uma proporção. Alguns tipos de erros são discutidos a partir das soluções apresentadas pelos alunos.

Palavras-chave: Ensino e Aprendizagem da Estatística. Ensino e Aprendizagem de Testes de Hipóteses. Erros Cometidos nos Testes de Hipóteses.

* Mestre em Educação em Ciências e Matemática pela Pontifícia Universidade Católica do Rio Grande do Sul (PUC-RS). Professora da Escola Estadual de Ensino Fundamental São José, Nova Petrópolis, RS, Brasil. Endereço para correspondência: Rua Vicente Prieto, s/nº, Pinhal Alto, CEP: 95150-000. Nova Petrópolis, RS, Brasil. *E-mail:* renete.sebastiani@acad.pucrs.br.

** Doutor. Professor titular da Faculdade de Matemática e permanente do Mestrado em Educação em Ciências e Matemática da Faculdade de Física da Pontifícia Universidade Católica do Rio Grande do Sul (PUC-RS), Porto Alegre, RS, Brasil. Professor adjunto do departamento de estatística da Universidade Federal do Rio Grande do Sul (UFRGS). Endereço para correspondência: Av. Ipiranga, 6681, CP 1429, CEP: 90619-900, Porto Alegre, RS, Brasil. *E-mail:* viali@pucrs.br.

Abstract

Hypothesis testing or decision theory is part of the contents of the majority of university-level basic statistics courses. In engineering, it represents approximately 20% of all content addressed in a standard four-hour course. These tests are the most visible part of statistics in the scientific literature. Understanding them is the basic part of statistical literacy for comprehending academic and scientific papers. Some studies on difficulties and mistakes are found in the literature, but the theme still lacks of more and deeper investigation. This article offers an overview of errors committed by students in tests and makes a comparison with some others found in the literature. The questions examined refer to tests for one sample for a mean (with an unknown standard deviation), and for a proportion. Some of the students' solutions are presented to discuss and exemplify the errors.

Keywords: Statistics Teaching and Learning. Teaching and Learning Hypothesis Tests. Hypothesis Tests Misunderstandings.

1 Introdução

A Estatística e a Probabilidade fazem parte do currículo dos cursos de Engenharia, em geral, por intermédio de uma única disciplina de 4 horas/aulas semanais. Os conteúdos propostos envolvem a totalidade da estatística básica, parte da probabilidade univariada e alguns conceitos de bivariada. Uma disciplina típica é formada por conteúdos que incluem: estatística descritiva, probabilidade, amostragem, estimação, testes de hipóteses, correlação, regressão e, eventualmente, análise de variância. É praxe que a disciplina seja iniciada com conteúdos de probabilidade ou estatística descritiva, a critério do professor ministrante. A inserção curricular da disciplina varia do terceiro ao nono semestre do curso. A disciplina, geralmente, é cursada por alunos de outros cursos além das engenharias, como os de Geologia, Química, Física e Matemática, sendo que esses três últimos incluem tanto os bacharéis quanto os licenciados. A exceção ocorre, em alguns casos, com a Engenharia de Produção, que pode ter uma disciplina específica ou mais de uma disciplina da área.

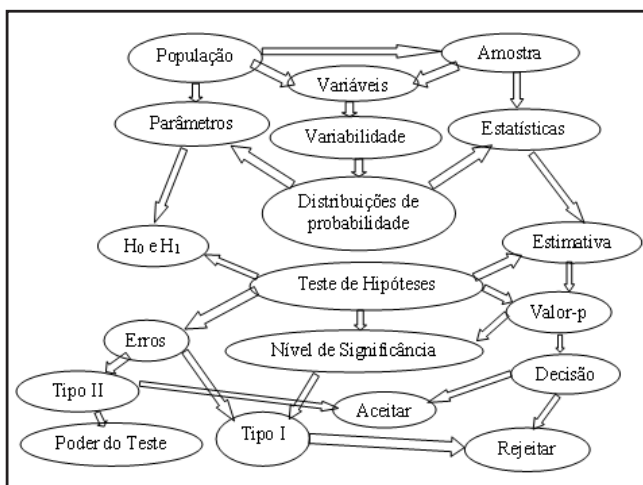


Figura 1 - Rede semântica com conceitos de testes de hipóteses.

Fonte: os autores.

Alguns obstáculos epistemológicos podem ser identificados. O primeiro deles surge no tratamento da dependência e independência na probabilidade, e já foi objeto de estudos de alguns trabalhos (DE TARI; DIBLASI, 2006; DÍAZ; BATANERO; CONTRERAS, 2010). O segundo surge na abordagem dos testes de hipóteses. Esse assunto ainda carece de trabalhos que tentem mapear quais são as dificuldades específicas enfrentadas pelos alunos no processo de ensino e aprendizagem. Conforme Haller e Krauss (2002), uma atenção particular deve ser dada às interpretações errôneas relacionadas aos testes de hipóteses, dado o uso generalizado dos testes estatísticos em todas as áreas de pesquisa.

Obviamente, a dificuldade básica desta disciplina, e de muitas outras de estatística, consiste na discrepância ou desalinhamento entre a quantidade de conteúdos propostos e o número de horas disponíveis para abordar esses conteúdos. Essa é, talvez, uma das prováveis causas de uma aprendizagem deficiente. A Estatística e a Probabilidade, a exemplo de outras disciplinas da área de Matemática, mantêm uma estrutura lógica, está baseada numa rede semântica (Figura 1) onde os nodos, geralmente, têm uma direção de sentido único. Assim, por exemplo, não é possível obter uma aprendizagem de conceitos de inferência pelo menos satisfatória ou significativa, sem que anteriormente tenha ocorrido uma compreensão de conceitos de descrição de dados amostrais e um mínimo de conceitos probabilísticos. Para se ter uma compreensão mínima de conceitos de testes de hipóteses, de que trata este trabalho, torna-se necessária,

ainda, uma boa compreensão de conceitos de amostragem. Para compreender amostragem, por outro lado, se faz necessário estabelecer conexões entre três contextos, dois próprios da Estatística (amostra e população) e um da Matemática (teoria da probabilidade). Colocando de outra forma, há necessidade de aliar o contexto prático, operacional dos dados amostrais, isto é, a realidade tangível de certo modo, com o conceito prático, mas não tangível de população ou universo. A ligação entre o contexto factível (amostra) e o idealizado (população) é moldada pela teoria, pela abstração por intermédio dos modelos probabilísticos que nos apresentam uma estimativa do quanto podemos estar errando ao tomarmos o recorte do universal pelo amostral. A Figura 2 apresenta um cenário desta estrutura.

Esses conteúdos, acima elencados, são pré-requisitos. Nas disciplinas da área de Exatas é comum exigir, ainda, um curso básico de Cálculo Diferencial e Integral para se poder cursar a disciplina de Estatística e Probabilidade.

Os testes de hipóteses apresentam muitas dificuldades de entendimento pela maioria dos alunos. Uma das possíveis razões é a abordagem meramente algorítmica que, muitas vezes, é praticada. Considerando-se que os professores seguem os textos disponíveis no mercado, pode-se perceber, em uma análise rápida, que a grande maioria pratica este tipo de abordagem. Nenhuma ou quase nenhuma atenção é dedicada aos conceitos básicos dos testes, tais como erros do Tipo I (nível de significância), Tipo II (poder de um teste), região crítica e valor-p. Isto ainda é agravado por abordagens que misturam a de Neyman-Pearson com a de Fisher uma vez que os textos disponíveis seguem a primeira, e os recursos computacionais (*softwares*) seguem a segunda.

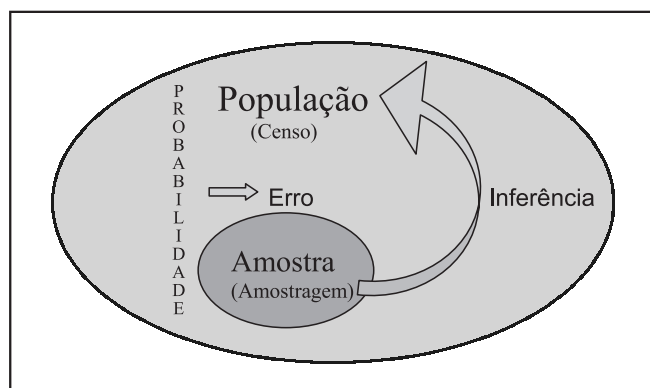


Figura 2 – Estrutura conceitual da amostragem.

Fonte: (VIALI, 2008, p. 5).

Considerando este cenário, o objetivo deste trabalho é avaliar os erros cometidos pelos alunos, em provas semestrais da disciplina de Estatística e Probabilidade dos cursos de Engenharia, em que os conteúdos avaliados são os testes de hipóteses. Por erros entende-se um conceito amplo, que pode englobar uma concepção errada, um pré-conceito equivocado, um juízo falso, uma má interpretação, um mal-entendido, um engano ou uma utilização imprópria, conforme Smith et al. (1993).

O interesse neste tópico em particular se deve ao fato de que ele é pouco presente na literatura, e a inferência estatística é um tópico relevante para o entendimento de boa parte da literatura técnica em várias áreas do conhecimento. Ainda, conforme Sotos et al. (2007) as ideias de inferência são especialmente sensíveis a interpretações equivocadas e os estudantes estão frequentemente inclinados a cometê-las, pois a inferência estatística requer o entendimento e a conexão de muitos conceitos abstratos tais como o de distribuições amostrais e nível de significância (vide a Figura 2).

2 Revisão da literatura

Dentre as pesquisas realizadas sobre as dificuldades na compreensão do teste de hipóteses apresentam-se os estudos de Vallecillos e Batanero (1997), Rodríguez (2006), Sotos et al. (2009) e Link (2002). No estudo de Vallecillos e Batanero (1997), os resultados obtidos foram analisados qualitativamente e foram agrupados em cinco categorias: interpretação das probabilidades condicionais, interpretação do nível de significância, conceito de hipótese nula e alternativa, lógica global do processo e parâmetro, estatística teste e sua distribuição. Em uma amostra de sete estudantes universitários do curso de Medicina, que responderam a um questionário, apenas dois alunos formularam corretamente as hipóteses conforme o contexto do segundo problema, o qual envolvia a média de uma variável com distribuição normal num contexto de controle de qualidade. O estudo evidenciou que os estudantes, embora tenham, teoricamente, conhecimento de que a hipótese nula deve ser formulada com o objetivo de ser rejeitada, na prática dificilmente conseguem enunciá-la de modo correto. Além disso, quase todos os alunos apresentaram indiferença em relação aos conceitos de estatística e parâmetro, como também, não reconheceram os diferentes tipos de média e suas distribuições de probabilidade.

O estudo de Rodríguez (2006) envolveu 96 alunos, sendo 29 estudantes do curso de Biologia, 22 do curso de Microbiologia e 45 estudantes de Agronomia. A pesquisadora aplicou um questionário composto de duas partes: uma para

identificar o conhecimento conceitual, com onze itens para assinalar verdadeiro ou falso, e a segunda com três problemas de aplicação. A principal concepção errônea identificada por Rodríguez (2006) foi em relação ao nível de significância. Os estudantes o consideram como sendo a probabilidade da hipótese nula ser verdadeira, dado que ela foi rejeitada. A maioria deles não diferencia probabilidade condicional, ou seja, para eles não há diferença entre $P(A|B)$ e $P(B|A)$. Grande parte dos participantes dessa investigação respondeu incorretamente as questões sobre a interpretação e definição do nível de significância e, também, sobre o erro do Tipo I.

No trabalho de Sotos et al. (2009) foram investigados 144 estudantes universitários de cursos introdutórios de Estatística, por meio de um questionário abordando três aspectos que, conforme a literatura, são de difícil compreensão num teste de hipóteses: definição de um teste de hipóteses, interpretação do valor-p e interpretação do nível de significância. Além de detectar os erros cometidos por esses estudantes, investigou-se, também, a confiança que eles têm em seus erros. Dentre os erros mais frequentes, os alunos acreditavam que um teste de hipótese é uma prova matemática da hipótese nula, ou então, que é uma prova probabilística por contradição. Em relação ao valor-p o erro mais comum foi considerá-lo como sendo a probabilidade de cometer um erro ao rejeitar a hipótese nula. O conceito de nível de significância foi o mais confuso para os estudantes. Os erros mais comuns vincularam-se à consideração sobre se o resultado do teste foi estatisticamente significativo para um nível de significância de 5%, a probabilidade de rejeitar a hipótese nula é igual a 95%, ou então, a probabilidade da hipótese nula ser verdadeira é igual a 5%.

Link (2002) analisou testes realizados por 295 estudantes de duas disciplinas do curso de Ciências Biológicas, em que uma tinha a disciplina de Cálculo como pré-requisito e a outra apenas Álgebra do Ensino Médio. Os erros encontrados foram na formulação das hipóteses, cálculo da estatística teste, determinação do valor-p e interpretação dos resultados.

A literatura existente sobre o assunto ainda é limitada e bastante localizada. Devido à diversidade de cursos universitários brasileiros que possuem uma disciplina de Estatística, os poucos estudos realizados não conseguiram mapear esse universo. Assim, a grande maioria dos cursos ainda não foi contemplada com uma investigação sobre as dificuldades dos alunos nestes conceitos. Particularmente, não foram encontrados trabalhos com alunos de Engenharia, portanto, o comparativo será realizado com os cursos encontrados (Agronomia, Biologia, Medicina e Microbiologia), mas o ideal seria fazer um confronto com outros estudos com alunos desta área.

3 Metodologia

A abordagem metodológica utilizada na investigação é predominantemente qualitativa. O objetivo principal não é quantificar os erros, mas sim analisar cada caso, categorizando conforme as semelhanças apresentadas. Os instrumentos de coleta de dados compreendem avaliações realizadas pelos alunos das turmas da disciplina de Probabilidade e Estatística para Engenharia, no segundo semestre de 2009, de três universidades do Rio Grande do Sul. As questões dessas avaliações foram elaboradas pelos professores das respectivas disciplinas.

Para analisar os dados coletados, foi utilizada a metodologia de análise de conteúdo. Cury (2007) coloca que, quando se analisam os erros dos alunos, está se analisando o conteúdo de uma produção, empregando-se uma metodologia de análise de conteúdo. Esta metodologia, conforme Bardin (2009, p. 44) é definida como:

Um conjunto de técnicas de análise das comunicações visando obter, por procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens, indicadores (quantitativos ou não) que permitam a inferência de conhecimentos relativos às condições de produção/recepção (variáveis inferidas) destas mensagens.

A mesma autora apresenta três fases para a análise de conteúdo: a pré-análise, a exploração do material e o tratamento dos resultados. Na primeira fase, em que são escolhidos os documentos, é feita uma leitura flutuante do material selecionado para constituir o *corpus*, isto é, o conjunto de documentos selecionados para a análise.

Nesta etapa, foi definido que seriam analisadas questões referentes a dois tipos de testes paramétricos: o teste para a média com o desvio padrão desconhecido e o teste para a proporção. Dentre as avaliações que foram analisadas, optou-se em utilizar as provas de sete turmas, pois nelas havia uma questão sobre cada tipo de teste escolhido, com exceção de uma turma, na qual foram analisados dois enunciados de cada tipo de teste. Assim, foi feita a análise de quinze enunciados diferentes, de provas aplicadas em sete turmas e que foram elaborados por seis professores.

Na fase da exploração do material é realizada uma leitura mais aprofundada do *corpus*. Segundo Bardin (2009), nesta etapa são definidas as unidades de registro e são formadas as categorias. O processo de agrupamento

pode ser definido previamente ou estabelecido durante a exploração do material. Bardin (2009) denomina de procedimento por *caixas* quando o sistema de categorias é fornecido previamente e as unidades de registro são organizadas à medida que vão sendo encontradas. O outro procedimento é denominado por *acervo*, neste caso, as categorias vão surgindo conforme o material é explorado. Nesta pesquisa foi adotado o segundo procedimento.

Inicialmente, foram criados códigos para identificar as universidades, os professores e as turmas. As universidades foram identificadas pelas letras X, Y e Z, os professores com números de 1 a 6 e as turmas com as letras A, B, C, D, E, F e G. A seguir, realizou-se um levantamento das questões respondidas corretamente, das que apresentavam algum tipo de erro, e, também, daquelas que não foram respondidas. Dentre as questões respondidas com erros, foram identificadas aquelas que seriam passíveis de análise. Destas foram identificados os tipos de erros cometidos, a frequência de ocorrência de cada um e foi feito um registro para, então, categorizar.

O levantamento quantitativo dos erros, bem como o seu registro, possibilitou uma maior impregnação no *corpus* e, a partir disso, foram construídas as unidades e formadas as categorias de erros. Definidas as categorias, prosseguiu-se com o tratamento dos resultados e as interpretações (BARDIN, 2009). Cury (2007) coloca que a descrição das categorias pode ser feita através da apresentação de tabelas com a distribuição de frequências das unidades, e também, aconselha a produção de textos-síntese.

4 Tipos de erros

O exame do material coletado possibilitou a identificação dos vários tipos de erros, relacionados às várias etapas de um teste de hipóteses. Conforme Bussab e Morettin (2005) e Viali (2009), estas etapas, geralmente, incluem seis passos que são:

- (i) definição das hipóteses;
- (ii) determinação da estatística teste e da distribuição amostral;
- (iii) cálculo da estatística teste;
- (iv) identificação e posicionamento da região crítica ou nível de significância do teste;
- (v) decisão e
- (vi) conclusão.

Inicialmente os erros foram apenas identificados e numerados e, quando era o caso, relacionados a uma das seis etapas do teste. Muitos erros não estão

diretamente relacionados a essas etapas, constituindo erros externos ao procedimento, como o erro de um cálculo, de uma fórmula, de um valor percentual etc. Cada tipo de erro encontrado foi agrupado em categorias, originando sete ao todo.

As categorias e os tipos de erros atribuídos a cada uma são os seguintes:

Categoria 1: envolve erros na formulação das hipóteses.

- 01: uso de teste bilateral ao invés do teste unilateral e vice-versa;
- 02: inversão do enunciado das hipóteses, isto é, o enunciado da hipótese nula refere-se ao que se deseja comprovar e a hipótese alternativa ao que se deseja testar;
- 03: uso de notação inadequada para enunciar as hipóteses;
- 04: não utilização do valor do parâmetro para enunciar a hipótese;
- 05: erro na forma de apresentação das hipóteses;
- 06: formulação de apenas uma hipótese;

Categoria 2: formada por erros cometidos na escolha da distribuição de probabilidade para a padronização da estatística teste.

- 07: uso de $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ para padronizar a estatística teste quando realiza

um teste para a média com desvio padrão desconhecido, isto é, utilizar z ao invés de t_{n-1} .

- 08: uso de $t_{n-1} = \frac{\bar{x} - \mu}{s / \sqrt{n}}$ para padronizar a estatística teste quando realiza

um teste para a proporção;

- 09: utilização do teste para a diferença entre duas proporções para padronizar a estatística teste quando realiza um teste para uma proporção;

Categoria 3: engloba os erros relacionados com a definição ou delimitação da região crítica.

- 10: utilização do valor crítico do teste bilateral quando está realizando um teste unilateral ou vice-versa;
- 11: erro na identificação do número de graus de liberdade;
- 12: utilização de um valor incorreto para o nível de significância;
- 13: uso da tabela errada (isto é, valor crítico errado);
- 14: erro não identificado na obtenção do valor crítico na tabela;
- 15: erro de sinal do valor crítico;

- 16: região crítica não é definida conforme o tipo de teste utilizado (unilateral à direita, unilateral à esquerda ou bilateral);
- 17: definição da região crítica sem ter formulado a hipótese alternativa;
- 18: erro de apresentação da região de rejeição (consequentemente de aceitação);

Categoria 4: envolve erros cometidos no cálculo da estatística teste.

- 19: uso de valores incorretos no cálculo da estatística teste por enganos;
- 20: uso de valores incorretos no cálculo da estatística teste por não diferenciar parâmetro de estatística;
- 21: erro na determinação de valores percentuais;
- 22: erro de digitação de valores;
- 23: erro na realização de operações aritméticas;
- 24: erro de transcrição do algoritmo (fórmula) do cálculo da estatística teste;

Categoria 5: engloba a utilização de notação inadequada para representar valores referentes à população e à amostra.

- 25: uso de σ (parâmetro) ao invés de s (estatística);
- 26: uso de μ (parâmetro) ao invés de \bar{x} (estatística);
- 27: uso de \bar{x} (estatística) ao invés de μ (parâmetro);
- 28: uso de N ao invés de n (tamanho da população quando deveria ser tamanho da amostra);
- 29: uso de s^2 (variância) ao invés de s (desvio padrão);
- 30: uso de π (proporção populacional) ao invés de μ (média populacional);
- 31: uso de x ao invés de μ (parâmetro);
- 32: uso de p (estatística) ao invés de π (parâmetro);
- 33: uso de π (parâmetro) ao invés de p (estatística);
- 34: uso de μ (média populacional) ao invés de π (proporção populacional);
- 35: uso de \bar{x} (média amostral) ao invés de p (proporção amostral);

Categoria 6: é formada por erros de tomada da decisão.

- 36: não rejeição da hipótese nula quando isto deveria ser feito;
- 37: rejeitar hipótese nula quando o valor calculado da estatística teste não indica que isto deva ser feito;
- 38: erro no posicionamento da estatística teste na reta real;
- 39: decisão sem determinação da região crítica, quando este é o caso;

Categoria 7: envolve os erros identificados na formulação da conclusão.

- 40: não faz referência ao nível de significância ao formular a conclusão;
- 41: fazer afirmações a respeito de H_0 , mesmo quando ela não é rejeitada;
- 42: conclui ao contrário do que os dados indicam.

Tabela 1 – Quantidade e percentual de erros em cada categoria.

Categorias	Erros	%
1	121	19,0
2	64	10,0
3	69	10,8
4	67	10,5
5	111	17,4
6	12	1,9
7	194	30,4
Total	638	100

A Tabela 1 fornece uma ideia global sobre a frequência de ocorrência dos erros em cada categoria. Ela indica que os estudantes estão apresentando maiores dificuldades em relação à formulação das hipóteses, à utilização da notação correta e à formulação da conclusão. Somando as categorias 1 (formulação das hipóteses) com a categoria 7 (elaboração da conclusão) temos praticamente 50% dos erros cometidos pelos alunos investigados. Note-se, ainda, que o menor número de erros, categoria 6, com cerca de 2% do total, ocorre na tomada da decisão, isto é, em aceitar ou rejeitar a hipótese nula com base nos resultados do algoritmo. Novamente, pode-se perceber aqui a excessiva operacionalização do ensino da Estatística, pois o aluno decide certo, mas, em seguida, interpreta errado, já que os erros de conclusão são os mais numerosos.

5 Análise dos dados

Nesta investigação foram selecionadas 357 questões de avaliações realizadas por sete turmas, sobre dois tipos de testes de hipóteses envolvendo uma amostra. A tabela 2 mostra o número e o percentual de questões constituintes do *corpus*, para cada tipo de teste.

Tabela 2 – Número e percentual de questões analisadas.

Univer- sidade	Professor/ Turma	Teste para a Média		Professor/ Turma	Teste para a Proporção	
		Nº de questões	%		Nº de questões	%
X	1/A	35	19,7	1/A	35	19,6
	2/B	35	19,7	2/B	35	19,6
	3/C	50	28,1	3/C	50	27,9
Y	4/D	44	24,7	4/D	44	24,6
	5/E	10	5,6	6/G	11	6,1
Z	6/F	4	2,2	6/F	4	2,2
Total		178	100		179	100

Na Tabela 3 é apresentado o resultado do levantamento do número de questões corretas, erradas e sem resposta. Dentre as 325 soluções erradas investigadas, em 23 o erro não pode ser identificado, totalizando, desta forma, 302 questões efetivamente analisadas.

A Tabela 3 mostra que são poucas as questões respondidas corretamente, sendo que na maioria delas é cometido pelo menos um tipo de erro. Alguns destes erros serão posteriormente exemplificados por meio das soluções apresentados pelos alunos. Pode-se considerar que os dados apresentados na Tabela 3 evidenciam que as dificuldades dos alunos, neste conteúdo, aparecem nas três universidades, como também em todas as turmas.

Após a exposição dos resultados, que teve o intuito de mostrar ao leitor os tipos de erros e a frequência com que foram detectados, apresenta-se a análise de algumas das soluções fornecidas pelos alunos, ilustrando os principais tipos de erros cometidos.

Tabela 3 – Tipos de testes e questões corretas, erradas e não respondidas.

Universidade	Professor/ Turma	Teste para a média			Teste para a proporção		
		Questões corretas	Questões com erros	Questões sem respostas	Questões corretas	Questões com erros	Questões sem respostas
X	1/A	5	30	0	4	29	2
	2/B	0	34	1	0	32	3
	3/C	5	45	0	3	42	5
Y	4/D	0	44	0	2	40	2
	5/E	0	10	0	0	11	0
Z	6/F	0	4	0	0	4	0
	Total	10	167	1	9	158	12

Na solução apresentada na Figura 3, a seguir, são identificados os erros do tipo 02, 07, 16 e 25. O erro do tipo 02 corresponde a 31,4% dos erros classificados na primeira categoria (que envolvem os erros na formulação das hipóteses). Dos 38 alunos que cometeram este erro, sete são da turma que respondeu a questão da Figura 3. A hipótese alternativa deve ser formulada de acordo com aquilo que o pesquisador deseja comprovar, enquanto que a hipótese nula expressa a intenção contrária. Portanto, a hipótese nula enuncia que este percentual seja igual, ou ainda, superior a 18%. Dessa forma, pode-se colocar que não existe a compreensão de uma ideia básica que é elaborar a hipótese nula com o objetivo de ser rejeitada.

Em relação ao erro tipo 07 da segunda categoria, que teve um percentual de ocorrência de 95,3%, o aluno optou em utilizar a distribuição de probabilidade normal, pois não diferencia o desvio padrão obtido na amostra do desvio padrão populacional, neste caso, desconhecido. Inclusive, comete o erro do tipo 25, ao

representar o desvio padrão da amostra com a letra grega σ , que é utilizada para representar o desvio padrão populacional. O aluno ainda escreve: *desvio padrão conhecido*; mostrando que não levou em consideração que este é um valor que pode ser obtido tanto da população, como também, da amostra. Isso pode indicar que para ele não há diferença entre esses dois valores. Este fato foi também observado na investigação de Vallecillos e Batanero (1997) e mostra que há uma dificuldade de compreensão em relação a esses dois universos distintos: o universo abstrato da população e o universo concreto da amostra.

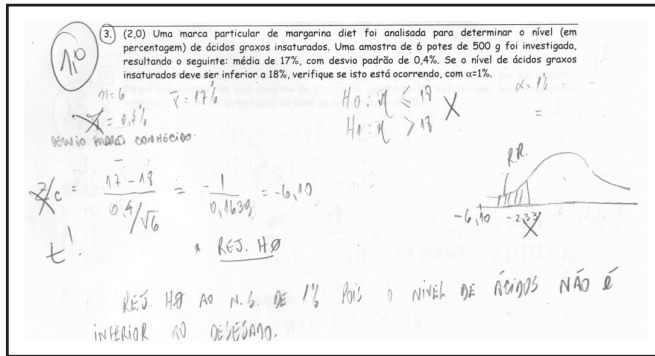


Figura 3 – Solução de um aluno a uma questão sobre o teste para uma média.

Ainda, em relação à Figura 3, observa-se que o aluno está utilizando um teste unilateral à direita e define a região crítica à esquerda (erro tipo 16). Um possível raciocínio feito por ele, para solucionar dessa forma, refere-se ao valor padronizado da estatística teste que resultou num número negativo. Na Figura 4 também se observa o mesmo tipo de erro.

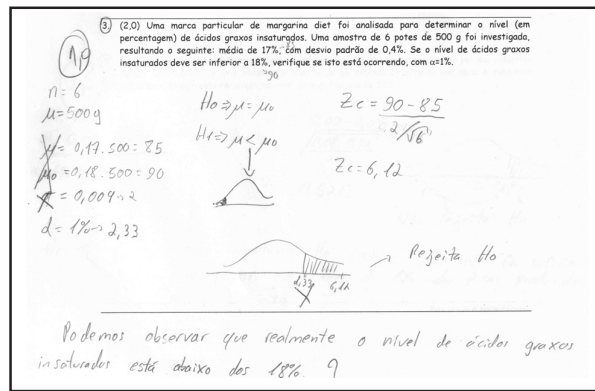


Figura 4 – Solução de um aluno da questão sobre o teste de uma média.

Na Figura 4 também foi detectado um erro do tipo 26, quando o aluno utiliza para a média da amostra a letra grega μ , que representa a média populacional. Além desse, na formulação da conclusão observa-se um erro do tipo 40. Esse tipo de erro ocorreu com muita frequência na categoria 7, sendo de 70,6% o seu percentual de ocorrência. Isto mostra que os estudantes não compreendem que a rejeição da hipótese nula não necessariamente implica na veracidade da hipótese alternativa, isto é, que ainda existe um erro (o nível de significância) associado a esta decisão. Uma possível causa dessa dificuldade pode estar associada à menor ênfase atribuída aos conceitos básicos e aos erros envolvidos, visto que praticamente todos os livros textos disponíveis priorizam os algoritmos, deixando de enfatizar a interpretação e análise, bem como o aspecto não determinístico do resultado. Além disso, poucos mencionam ou enfatizam o erro do tipo II, bem como o fato de ele ser desconhecido ou não controlado.

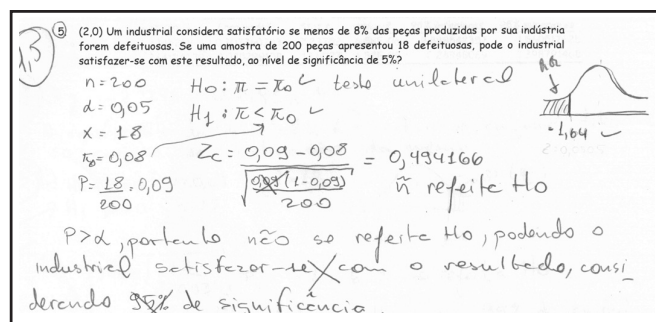


Figura 5 – Solução de um aluno a uma questão sobre o teste para uma proporção.

Na Figura 5 apresenta-se uma solução na qual o aluno calcula o erro padrão da proporção, utilizando o valor da proporção obtido na amostra. Novamente, percebe-se que há indiferença no tratamento dos valores provenientes de dados da população e os da amostra. Ainda, na formulação da conclusão, é feita uma afirmação quando não foi possível rejeitar a hipótese nula (erro tipo 41). Este erro evidencia, mais uma vez, a dificuldade de entender que a não rejeição da hipótese nula, a partir da evidência amostral, não nos permite fazer afirmações sobre ela, pois não se conhece o erro do tipo II envolvido. Nesta conclusão, o estudante também se mostra confuso entre o nível de significância e o nível de confiança, utilizado nos intervalos de confiança.

Para finalizar, apresenta-se na figura 6 uma solução na qual não foi utilizada a notação adequada para representar a média populacional no enunciado

das hipóteses (erro tipo 03). Também, houve dificuldade no cálculo de um percentual. Este erro operacional levou à tomada de uma decisão incorreta. Na formulação da conclusão pode-se discutir sobre dois erros comuns ocorridos na sua formulação. O aluno toma a decisão de não rejeitar a hipótese nula e na sua conclusão afirma o que foi enunciado na hipótese alternativa, assim como na conclusão apresentada na figura 5. Comete, também, o erro tipo 42, quando afirma o contrário daquilo que foi enunciado na hipótese nula.

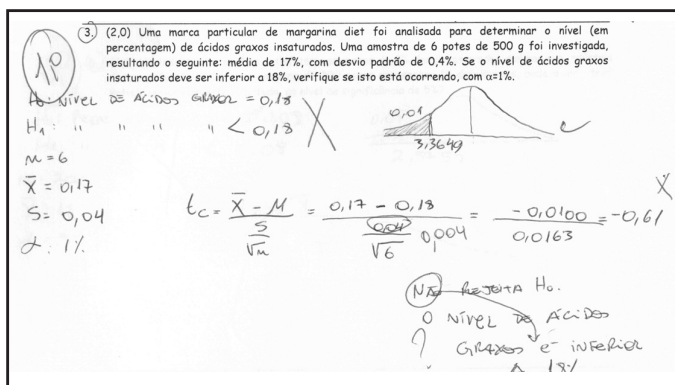


Figura 6 – Solução de um aluno a uma questão sobre o teste para uma média.

Nestas quatro soluções apresentadas, estão ilustrados alguns dos erros mais comuns detectados na análise das questões. A parte essencial na resolução de um problema desse conteúdo é a formulação das hipóteses e a conclusão, visto que, para escrever ou interpretar os resultados num artigo científico, por exemplo, a etapa que envolve os algoritmos é feita geralmente por meio de um *software*; no entanto, a interpretação dos resultados não é ainda possível de ser realizada com o uso da tecnologia.

A dificuldade dos alunos em compreender e, também, incorporar a visão não determinística na resolução desses problemas está relacionada com os erros que foram detectados nessas etapas. Quando um estudante, na sua conclusão coloca: *rejeita-se H₀, prova-se que este automóvel consome menos que 30 km/l*; está mostrando que não incorporou a ideia de que esse resultado foi obtido em uma amostra, e que na população não necessariamente isso é verdadeiro. O nível de significância, para esse aluno, parece não ter significado, sendo ignorado na formulação da conclusão, assim como, também, alguns professores não têm destacado este erro nas correções, considerando a questão como resolvida corretamente. Dessa forma, a análise das respostas dos alunos levou à

constatação de que muitos dos conceitos envolvidos num teste de hipóteses têm pouco ou nenhum significado para eles.

Os resultados dessa investigação reforçam a necessidade de o professor oportunizar situações nas quais o aluno vivencie os conceitos, possibilitando a emergência das conexões entre as diversas ideias envolvidas. A visão determinista, evidenciada pelos erros do tipo 40 (ao ignorar o nível de significância), pode ser superada, ou seja, abandonada, no momento em que são estabelecidas as relações entre os conceitos, principalmente, os de população, amostra e sua conexão com a probabilidade. Da mesma forma, os erros do tipo 20 e do tipo 07 podem resultar da prevalência de uma visão fragmentada dos conceitos, não havendo necessidade de diferenciar dados provindos de uma população ou de uma amostra.

6 Considerações finais

Esse trabalho teve como objetivo investigar e analisar os erros que os alunos cometem ao resolver questões sobre testes de hipóteses, classificá-los e comparar os resultados obtidos com os de investigações já realizadas sobre este tema. Foram analisadas as produções dos alunos ao responderem questões de avaliações elaboradas por professores das disciplinas de Estatística para cursos de Engenharia. Diversos erros foram detectados, alguns deles se repetiam com bastante frequência e, também, já foram identificados em estudos correlatos. Os erros ocorreram com maior frequência na formulação da conclusão, no uso da notação para representar parâmetros e estatísticas e na formulação das hipóteses. Os erros de conclusão representaram 30% de todos os observados. Isto representa 50% a mais do que os erros pertencentes à segunda categoria mais frequente, que foi a da formulação das hipóteses, com cerca de 20% do total de erros observados.

Uma possível explicação para que tantos erros ocorram nos procedimentos de formulação das hipóteses e na conclusão é que ainda não ocorreu a mudança preconizada por Ben-Zvi e Garfield (2005). Eles propõem uma nova abordagem para o ensino de Estatística, que deve mudar de uma perspectiva matemática, centrada em algoritmos e procedimentos, para uma ênfase nas ideias de raciocínio, pensamento e literacia estatística. Essa mudança, segundo eles, significa a transformação do ensino de Estatística com base na Probabilidade para um ensino centrado em dados, encorajando o uso, entre outros recursos, de dados reais e de tarefas em grupo como forma de melhorar as habilidades colaborativas e comunicativas por meio de discussões estatísticas.

A integração da tecnologia nas aulas de Estatística, em sinergia com uma postura orientada aos dados e uma pedagogia revisada, é, também, enfatizada por Dutton e Dutton (2005).

Nesta investigação, constatou-se que os estudantes têm ignorado o conceito de nível de significância em grande parte das conclusões por eles formuladas. Este fato evidencia a presença de uma concepção errônea a respeito do significado do resultado de um teste de hipóteses. Na conclusão apresentada na solução da Figura 4, o aluno coloca: *podemos observar que realmente o nível de ácidos graxos insaturados está abaixo dos 18%*; o que mostra que ele compreende que a rejeição da hipótese nula implica na veracidade da hipótese alternativa. Vallecillos e Batanero (1997) constataram que, para alguns alunos, as hipóteses referem-se a enunciados ou situações desconhecidas e se deseja determinar se são verdadeiras ou falsas.

O uso de notação inadequada para representar parâmetros e estatísticas foi um erro recorrente, evidenciando que os estudantes, em geral, não demonstram a necessidade de fazer uma diferenciação entre valores provenientes da população e da amostra. Albert (1995) e Link (2002) também encontraram que os estudantes apresentam dificuldades em reconhecer o parâmetro a ser testado na inferência estatística.

Em relação à etapa que envolve a formulação das hipóteses, foram detectadas algumas dificuldades. Assim como relatam Vallecillos e Batanero (1997) em sua investigação, também foi constatado que os alunos cometem erros quando enunciam as hipóteses a partir do contexto do problema. Percebe-se que alguns estudantes não compreendem que num teste de hipóteses são testados valores hipotéticos de parâmetros populacionais, como na solução apresentada na Figura 4, como, também, não formulam a hipótese nula com o objetivo de ser rejeitada (Figura 4). Aqui, temos uma dificuldade que também foi apontada em Bady (1979), que observou terem as pessoas uma forte tendência a procurar informações que verifiquem a hipótese ao invés de tentar falseá-la.

As dificuldades dos alunos, que resultam nos erros detectados nesta investigação, podem estar relacionadas com a metodologia adotada pelos professores, ministrantes das disciplinas de Estatística nos cursos superiores. O tempo disponível para trabalhar os conteúdos de Estatística Básica numa disciplina de 4 horas/aulas semanais, ou em alguns casos, apenas 2 horas/aulas semanais, pode ser um empecilho para o professor abordar com maior ênfase alguns conceitos. Mas, por outro lado, nas soluções apresentados pelos alunos, parece haver ênfase em praticar cálculos, ou seja, disponibilizar mais tempo para resolver algoritmos que poderiam ser resolvidos por um *software*.

Muitos conceitos, dentre eles os erros do tipo I (nível de significância), tipo II (poder de um teste), região crítica e valor-p, parecem não ser tão relevantes, quanto aos aspectos relacionados aos algoritmos e são, de certa forma, ignorados pelo professor. Os estudantes teriam uma melhor compreensão desses conceitos se lhes fossem oportunizados exercícios que os envolvam. Além disso, a forma como o conteúdo de testes de hipóteses é ensinado deve levar em conta a conexão entre os conteúdos de Estatística, assim como é apresentado na rede semântica da Figura 1, e, também, da ligação entre os dois contextos - o abstrato (população) e o concreto (amostra) - sugerida pela Figura 2.

Uma constatação deste estudo é que, na literatura, são relatados poucos erros operacionais, contudo constatou-se que eles representam uma parcela bem razoável. Se entendermos que as categorias 1 e 7 são de erros conceituais e as demais operacionais, então é possível constatar que os estudantes investigados se dividem de forma uniforme quanto aos tipos de erros cometidos, sendo aproximadamente 50% operacionais e os outros 50% conceituais.

Referências

ALBERT, J. Teaching inference about proportions using Bayes and discrete models. **Journal of Statistics Education**, Alexandria, VA, v. 3. n. 3, 1995. Disponível em: <<http://www.amstat.org/publications/jse/v3n3/albert.html>>. Acesso em: 15 maio 2010.

BADY, R. J. Students' understanding of the logic of hypothesis testing. **Journal of Research in Science Teaching**, Hoboken, NJ, v. 16, n. 1, p. 61-65, 1979.

BARDIN, L. **Análise de Conteúdo**. Lisboa: Edições 70, 2009.

BEN-ZVI, D; GARFIELD, J. Research on Statistical Literacy, Reasoning, and Thinking: Issues, Challenges, and Implications. In: BEN-ZVI, D; GARFIELD, J.(Eds.). **The Challenge of Developing Statistical Literacy, Reasoning and Thinking**. Netherlands: Springer, 2005. p. 397-409.

BUSSAB, W. O.; MORETTIN, P. A. **Estatística Básica**. São Paulo: Saraiva, 2005.

CURY, H. N. **Análise de Erros**: o que podemos aprender com as respostas dos alunos. Belo Horizonte: Autêntica, 2007.

DE TARI, A. D'A.; DIBLASI, A. Analysis of didactic situations suggested to distinguish disjunctive events and independent events. In: INTERNATIONAL CONFERENCE ON STATISTICS EDUCATION, 7th, Salvador, 2006. **Proceedings...** Salvador: IASE/ISI, 2006.

DÍAZ, C.; BATANERO, C.; CONTRERAS, J. M. Teaching Independence and Conditional Probability. **Boletín de Estadística e Investigación Operativa**, La Rioja, Espanha, v. 26, n. 2, p. 149-162, jun. 2010. Disponível em: <<http://www.emis.de/journals/BEIO/files/BEIOVol26Num2HyE-Web.pdf>>, Acesso em: 13 abr. 2010.

DUTTON, J.; DUTTON, M. Characteristics and Performance of Students in an Online Section of Business Statistics. **Journal of Statistics Education**, Alexandria, VA, v.13, n. 3, 2005. Disponível em: <<http://www.amstat.org/publications/jse/v13n3/dutton.html>>. Acesso em: 10 mar. 2010.

HALLER, Heiko, KRAUSS, Stefan. Misinterpretations of significance: a problem students share with their teachers? **Science Education**. Hoboken, NJ, v. 7, n. 1, p. 1-20, 2002. Disponível em: <http://www.dgps.de/fachgruppen/methoden/mpr-online/issue16/art1/haller.pdf>. Acesso em: 12 jan. 2010.

LINK, C. W. An Examination of Student Mistakes in Setting Up Hypothesis Testing Problems. **Proceedings of the Louisiana-Mississippi Section of the Mathematical Association of America**. Louisiana, Spring, 2002. Disponível em: <<http://sections.maa.org/lams/proceedings/spring2002/conway.link.pdf>>, Acesso em: 17 fev. 2010.

RODRÍGUEZ, M. I. Estudio Teórico y Experimental sobre Dificultades en la Comprensión del Contraste de Hipótesis en Estudiantes Universitarios. **Acta Latinoamericana de Matemática Educativa**, México, v. 19, s/n, p. 162-168, 2006. Disponível em: <<http://www.clame.org.mx/documentos/alme19.pdf>>. Acesso em: 06 abr. 2010.

SMITH, J. P. et al. Misconceptions reconceived: a constructivist analysis of knowledge in transition. **The Journal of the Learning Sciences**, London, England, v. 3, n. 2, p. 115-163, 1993. Disponível em: <<http://clint.sharedwing.net/research/misconceptions/disessa—recon.pdf>>. Acesso em: 29 nov. 2009.

SOTOS, A. E. C. et al. Student's misconceptions of statistical inference: a review of the empirical evidence from research on statistics education. **Educational Research Review**. Amsterdam, n. 2, p. 98-113, 2007. Disponível em: <<https://lirias.kuleuven.be/bitstream/123456789/136347/1/CastroSotos.pdf>>. Acesso em: 12 nov. 2009.

SOTOS, A. E. C. et al. How confident are students in their misconceptions about hypothesis tests? **Journal of Statistics Education**, Alexandria, VA, v. 17, n.2, s/p, 2009. Disponível em: <<http://www.amstat.org/publications/jse/v17n2/castrosotos.html>>. Acesso em: 19 fev. 2010.

VALLECILLOS, A.; BATANERO, C. Análisis del aprendizaje de conceptos clave en el contraste de hipótesis estadísticas mediante el estudio de casos. **Recherches en Didactique des Mathématiques**, Grenoble, France, v. 17, n. 1, p. 29-48, 1997. Disponível em: <<http://www.ugr.es/~batanero/publicaciones%20index.htm>>. Acesso em: 17 dez. 2009.

VIALI, L. **Material Didático**: Amostragem. 2008. Disponível em: <www.mat.ufrgs.br/~viali/>. Acesso em: 15 mar. 2010.

VIALI, L. **Material Didático**: Testes de Hipóteses. 2009. Disponível em: <www.pucrs.br/famat/viali/>. Acesso em: 27 jan. 2010.

Submetido em Junho de 2010.
Aprovado em Outubro de 2010.