

RENATE GRINGS SEBASTIANI

**ANÁLISE DE ERROS EM TESTES DE HIPÓTESES: UM ESTUDO  
COM ALUNOS DE ENGENHARIA**

Dissertação apresentada ao Programa de Pós-Graduação em Educação em Ciências e Matemática, da Pontifícia Universidade Católica do Rio Grande do Sul, como requisito parcial para a obtenção do grau de Mestre em Educação em Ciências e Matemática.

**Orientador: Prof. Dr. Lori Viali**

Porto Alegre, 2010

Dados Internacionais de Catalogação na Publicação ( CIP )

S443a Sebastiani, Renate Grings  
Análise de erros em testes de hipóteses: um estudo  
com alunos de engenharia / Renate Grings Sebastiani. –  
Porto Alegre, 2010.  
**82 f.**

Diss. (Mestrado) – Programa de Pós-Graduação em  
Ciências e Matemática, PUCRS  
Orientador: Prof. Dr. Lori Viali

1. Estatística – Ensino e Aprendizagem. 2. Teste de  
Hipóteses. 3. Análise de Erros. 4. Educação. I. Lori, Viali.  
II. Título.

**CDD 372.7  
510.7**

**Bibliotecária Responsável: Salete Maria Sartori, CRB 10/1363**

RENATE GRINGS SEBASTIANI

**ANÁLISE DE ERROS EM TESTES DE HIPÓTESES: UM ESTUDO COM  
ALUNOS DE ENGENHARIA**

Dissertação apresentada ao Programa de Pós-Graduação em Educação em Ciências e Matemática, da Pontifícia Universidade Católica do Rio Grande do Sul, como requisito parcial para a obtenção do grau de Mestre em Educação em Ciências e Matemática.

Aprovado em 21 de dezembro de 2010, pela Banca Examinadora.

BANCA EXAMINADORA:



---

Dr. Lorí Viali (Orientador - PUCRS)



---

Dra. Luciana Neves Nunes (UFRGS)



---

Dr. João Feliz Duarte de Moraes (PUCRS)

*Dedico este trabalho ao meu esposo Ivair, pela compreensão e pelo apoio em todos os momentos do  
Mestrado.*

## *AGRADECIMENTOS*

*A Deus pela vida e por permitir a conclusão de mais uma etapa da minha formação acadêmica.*

*Aos meus pais Valter e Therezinha pelos ensinamentos, incentivo e apoio em todos os momentos de minha vida.*

*Ao meu esposo Ivair pelo incentivo, apoio, compreensão, carinho e amor, principalmente nestes dois anos de estudos.*

*A minha irmã Ana Emília e ao meu cunhado Idivandro pelos favores prestados.*

*Aos meus avós, em especial ao meu vô Ermindo pelos momentos de descontração proporcionados durante os jogos de canastra.*

*A minha prima Letícia pelas correções e traduções.*

*Aos demais familiares pelo incentivo, apoio e carinho.*

*Ao professor e orientador Doutor Lori Viali pelos conhecimentos compartilhados, pela dedicação, amizade, competência e auxílio em todas as etapas do curso.*

*Ao professor Doutor João Feliz Duarte de Moraes e a professora Doutora Luciana Neves Nunes pelas valiosas sugestões dadas.*

*Aos professores e funcionários do Programa de Pós-Graduação em Educação em Ciências e Matemática.*

*Aos queridos colegas da turma de mestrado pela amizade.*

*Aos professores e alunos sujeitos da pesquisa.*

*A CAPES pelo apoio financeiro.*

*A todos que de uma forma ou de outra contribuíram para a realização deste trabalho.*

*Muito Obrigada!*

*O que aprendi, acima de tudo, é a seguir em frente, pois a grande ideia é a de que, como o acaso efetivamente participa de nosso destino, um dos importantes fatores que levam ao sucesso está sob nosso controle: o número de vezes que tentamos rebater a bola, o número de vezes que nos arriscamos, o número de oportunidades que aproveitamos. Pois até mesmo uma moeda viciada que tenda ao fracasso às vezes cairá do lado do sucesso.*

*(Leonard Mlodinow)*

## RESUMO

Esta investigação tem como objetivo analisar os erros cometidos por alunos dos cursos de Engenharia nas provas semestrais em que os conteúdos avaliados são os testes de hipóteses. Estes estão incluídos na ementa da disciplina de Probabilidade e Estatística e 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. Na literatura, são encontrados alguns estudos sobre dificuldades e erros, mas o tema ainda carece de investigações. Este trabalho, portanto, consiste em classificar os erros identificados nas soluções dos alunos, fazendo um comparativo com a literatura da área, além de analisar as respostas dadas a um questionário aplicado aos professores das turmas participantes, sobre as possíveis causas dos erros. Foram participantes desta pesquisa sete turmas de três universidades gaúchas e, também, seis professores que lecionam a disciplina. As respostas analisadas referem-se aos testes para uma amostra, totalizando 178 respostas sobre o teste para a média (com desvio padrão desconhecido) e 179 respostas sobre o teste para a proporção. Os erros foram classificados em sete categorias, e observou-se que as maiores dificuldades dos alunos estão relacionadas à formulação da conclusão, ou seja, à etapa da interpretação dos resultados. Os professores que responderam ao questionário consideram, em geral, que os alunos não compreendem a estrutura dos testes de hipóteses e, também, que lhes faltam pré-requisitos, como conhecimentos matemáticos elementares e de interpretação de textos. Ainda, apontam a falta de atenção, de estudo e de realização de exercícios como possíveis causas dos erros. Os resultados sugerem que, no ensino deste conteúdo, seja dada mais ênfase aos conceitos envolvidos do que aos procedimentos e cálculos, como forma de promover a aprendizagem significativa.

**Palavras-chave:** Ensino e aprendizagem da Estatística. Ensino e aprendizagem de testes de hipóteses. Erros cometidos em testes de hipóteses.

## ABSTRACT

This research aims to analyze the mistakes made by engineering students in semester exams in which the contents evaluated are the hypothesis tests. These contents are included in the subject syllabus of Probability and Statistics and are the most visible part of statistics in scientific papers. Understanding them is part of the basic statistical literacy for the comprehension of the scientific and scholarly literature. In literature, it is found few studies on the problems and mistakes, but the theme still needs investigation. This work, therefore, consists in classify the mistakes found in students' solutions, making a comparison with the literature, over an analysis of the responses to a questionnaire given to teachers of participating classes, on the possible causes of errors. Seven classes from three universities in southern Brazil and also six teachers who teach the discipline were the participants of this study. The answers analyzed are related to the testing for a sample, totaling 178 answers on the test for the mean (with unknown standard deviation) and 179 answers on the test for proportion. The errors were classified into seven categories, and it was observed that the greatest difficulties of the students are related to the formulation of conclusions, it means, the stage of the interpretation of results. Teachers who answered the questionnaire generally consider that students do not understand the structure of hypothesis testing, and also that they lack prerequisites such as basic mathematical skills and reading comprehension. Also, they indicate the lack of attention, study and exercises as the possible causes of errors. The results suggest that in teaching this content, more emphasis must be given to the concepts involved than to the procedures and calculations, in order to promote meaningful learning.

**Keywords:** Teaching and learning of Statistics. Teaching and learning hypothesis tests.

Mistakes made in hypothesis tests.

## LISTA DE FIGURAS

Figura 1 - Rede semântica com conceitos de testes de hipóteses .....	16
Figura 2 – Probabilidades de obter-se cara em 10 lançamentos de uma moeda equilibrada ...	22
Figura 3 – Estrutura conceitual da amostragem .....	26
Figura 4 – Percentual de acertos, erros e não realização das etapas dos testes de hipóteses ...	41
Figura 5 – Ilustração dos erros tipo 02, 07, 16 e 25 .....	47
Figura 6 – Ilustração do erro tipo 16 .....	49
Figura 7 – Ilustração dos erros tipo 03, 21, 41 e 42 .....	50
Figura 8 – Ilustração dos erros tipo 03, 07, 21 e 25 .....	51
Figura 9 – Ilustração dos erros tipo 02, 03, 07, 25 e 40 .....	51
Figura 10 – Ilustração dos erros tipo 02, 03, 20 e 40 .....	52
Figura 11 – Ilustração do erro tipo 40 .....	53
Figura 12 – Ilustração dos erros tipo 20 e 41 .....	53
Figura 13 – Ilustração dos erros tipo 02 e 41 .....	54
Figura 14 – Ilustração dos erros tipo 02 e 37 .....	54
Figura 15 – Ilustração dos erros tipo 01, 41 e 42 .....	55
Figura 16 – Ilustração de uma questão resolvida corretamente .....	56
Figura 17 – Ilustração da curva normal .....	58

## LISTA DE QUADROS

Quadro 1 – Participantes da pesquisa .....	34
Quadro 2 – Número e percentual de respostas analisadas .....	38
Quadro 3 – Tipos de testes e respostas corretas, erradas e em branco .....	39
Quadro 4 – Quantidade de acertos, erros e não realização das seis etapas dos testes de hipóteses .....	40
Quadro 5 – Frequência de ocorrência dos erros identificados.....	45
Quadro 6 – Categorias e tipos de erros.....	45
Quadro 7 – Quantidade e percentual de erros em cada categoria.....	46
Quadro 8 – Categorias relacionadas às causas dos erros segundo a opinião dos professores ..	59

## SUMÁRIO

1	INTRODUÇÃO .....	12
2	OBJETIVOS .....	18
2.1	Objetivo Geral .....	18
2.2	Objetivos Específicos .....	18
3	PROBLEMAS.....	19
3.1	Problema de Pesquisa .....	19
3.2	Questões de Pesquisa.....	19
4	FUNDAMENTAÇÃO TEÓRICA .....	20
4.1	Os Testes de Hipóteses .....	20
4.2	Aprendizagem Significativa .....	24
4.3	Análise de Erros .....	27
5	METODOLOGIA .....	33
5.1	Delineamento da Pesquisa .....	33
5.2	Participantes .....	34
5.3	Instrumentos de Coleta de Dados .....	34
5.4	Técnicas de Análise de Dados .....	35
6	RESULTADOS E DISCUSSÃO .....	38
6.1	Tipos e Categorias de Erros.....	38
6.2	Análise dos Erros.....	46
6.3	Análise das Respostas dos Professores.....	58
7	CONSIDERAÇÕES FINAIS.....	63
	REFERÊNCIAS .....	70
	ANEXO A – ENUNCIADOS DAS QUESTÕES ANALISADAS .....	74
	APÊNDICE A .....	77

# 1 INTRODUÇÃO

A Estatística é uma ciência que ocupa posição de destaque na atual sociedade tornando-se necessário que nos espaços onde ocorre a educação formal este conhecimento seja disseminado. De acordo com Batanero (2001), o desenvolvimento de um país está relacionado ao grau com o qual este produz estatísticas completas e confiáveis, porque esta informação é utilizada para tomar decisões econômicas, sociais e políticas. Por isso, a grande quantidade de informação disponível precisa ser transformada em conhecimento, para que seja possível tomar decisões sobre questões cotidianas.

Além disso, a informação já organizada, que é apresentada pela mídia, precisa ser avaliada, para que seja considerada se é ou não coerente. Sobre isso, Ponte, Brocardo e Oliveira (2006, p. 91) afirmam que “todo cidadão precisa saber quando um argumento estatístico está ou não a ser utilizado com propriedade”. Portanto, é imprescindível que os conteúdos de Estatística sejam abordados em todos os níveis de ensino, desde os primeiros anos do Ensino Básico, conforme estabelecem as diretrizes dos Parâmetros Curriculares Nacionais.

Nesse contexto, é necessário que a formação Estatística não se restrinja aos técnicos responsáveis pela produção das estatísticas, mas principalmente aos profissionais e cidadãos que devem interpretá-las e tomar decisões, como também aos que contribuem na obtenção dos dados (BATANERO, 2001). Praticamente, a maioria dos cursos de graduação apresenta pelo menos uma disciplina que trata dos conteúdos de Estatística. Porém, as reformulações que ocorreram nos currículos destes cursos nos últimos anos têm reduzido a carga horária destinada a estas disciplinas.

Conforme Viali (2007), nos cursos de Administração, Contábeis e Economia a carga horária foi reduzida pela metade. Em alguns cursos, como os de Engenharia, a disciplina sofreu menos redução, de 6 para 4 horas semanais. Além disso, outros cursos eliminaram totalmente a Estatística de seu currículo, como o curso de Pedagogia de algumas universidades.

O desalinhamento ou a discrepância entre a quantidade de conteúdos propostos e o número de horas disponíveis para abordá-los é, provavelmente, uma das causas da

aprendizagem deficiente. Este fato impõe que o professor adote estratégias de ensino que priorizem o desenvolvimento do raciocínio estatístico, dando menos ênfase aos aspectos mecânicos. Os extensos cálculos podem ser realizados utilizando-se o computador e, dessa forma, pode ser disponibilizado mais tempo para abordarem-se os conceitos e interpretarem-se os resultados.

O desenvolvimento tecnológico tem proporcionado o avanço das técnicas estatísticas, principalmente pelo uso do computador. Modernas técnicas de análise de dados foram desenvolvidas e, aliadas aos *softwares* estatísticos, tornaram muito mais simples e rápida a obtenção dos resultados. Porém, o que não teve a mesma evolução foi o ensino da estatística.

Conforme Batanero (2001) é deficiente a formação dos professores na área de estatística. No ensino básico, os professores licenciados em Matemática são os que ensinam a Estatística e, no ensino superior, há também bacharéis em Estatística que ministram a disciplina. O fato de profissionais da área da Matemática atuarem no ensino da Estatística contribui para que este tenha algumas características que não são próprias de sua natureza.

Sobre isso, Pamplona e Carvalho (2007) destacam que na formação dos professores de Matemática não é dada a devida importância ao desenvolvimento da “lógica da inferência”, priorizando-se o “raciocínio determinista” que dificulta a compreensão de questões de natureza não determinista. Como exemplo os autores mencionam a dificuldade de compreensão da diferença entre usar o termo “rejeitar” em vez de “não aceitar”, ou “não rejeitar” em vez de “aceitar”, utilizados em testes de hipóteses.

Aliada à falta de formação específica em educação estatística, Batanero (2001) também acrescenta que são poucas as investigações realizadas sobre a didática da estatística em comparação com as da área da matemática. Este fato não permite que as dificuldades dos alunos nos principais conceitos de estatística sejam conhecidas. Pela utilidade que a Estatística tem em praticamente todas as áreas de conhecimento, a autora destaca a necessidade de aprofundar os conhecimentos em relação às dificuldades no processo de ensino-aprendizagem, bem como os aspectos didáticos do tema e os erros cometidos pelos alunos após a aprendizagem.

Além disso, são encontradas poucas investigações sobre a inferência estatística na literatura e este tema é relevante para o entendimento de boa parte dos artigos técnicos em várias áreas do conhecimento. Em especial, os testes de hipóteses são um assunto que ainda carece de trabalhos que tentem mapear quais são as dificuldades específicas enfrentadas pelos

alunos no processo de ensino-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.

Por isso, pretende-se neste estudo investigar as dificuldades que se manifestam na forma de erros quando os alunos resolvem questões que envolvem os testes de hipóteses. As disciplinas que envolvem Estatística nos cursos de graduação tratam geralmente da Estatística Descritiva e da Probabilidade, que são também abordadas no Ensino Elementar e Médio, como ainda de Estatística Inferencial.

Especificamente, na disciplina de Probabilidade e Estatística do currículo dos cursos de Engenharia, os conteúdos propostos geralmente envolvem a totalidade da estatística básica, parte da probabilidade univariada e alguns conceitos de bivariada. A ementa de uma disciplina típica é formada por conteúdos que incluem: estatística descritiva, probabilidade, amostragem, estimação, testes de hipóteses, correlação, regressão e eventualmente análise de variância.

Com relação aos conteúdos ministrados na disciplina de Estatística, Lopes (2007) e Link (2002) afirmam que é nos testes de hipóteses que os alunos apresentam as maiores dificuldades de compreensão e que, para os professores, este é o mais difícil de ser ensinado. Os testes de hipóteses oferecem ferramentas que possibilitam fazer generalizações para uma população, tendo por base dados amostrais. Portanto, os testes de hipóteses são um procedimento que requer um raciocínio indutivo. Convém lembrar que o raciocínio indutivo está essencialmente associado a um fator de incerteza, isto é, as conclusões serão sempre probabilísticas. Com base em dados empíricos, obtidos a partir de um conjunto restrito de casos, denominado amostra, buscam-se evidências contra um conhecimento aceito como verdadeiro, generalizando-se os novos achados para um conjunto mais geral, chamado de população. De acordo com Fisher (1931, p. 16) “a hipótese nula nunca é provada ou estabelecida, mas é possivelmente refutada, no curso de uma experimentação. Cada experimento existe somente para oferecer aos fatos uma chance de refutar a hipótese nula”.

Apesar de serem criticados por alguns pesquisadores, os testes de hipóteses fazem parte dos conteúdos da maioria das disciplinas de Estatística Básica dos cursos universitários. Nas disciplinas dos cursos de Engenharia, representam aproximadamente 20% do total de conteúdos abordados em uma disciplina padrão de 4 horas/aula semanais.

Segundo Kirk (2001), as críticas ocorrem devido ao fato de os testes não se centrarem no significado prático do resultado e por não fornecerem o tamanho do efeito.

Ainda, acrescenta que os resultados dos testes de hipóteses deveriam ser acompanhados por outras medidas descritivas, além dos intervalos de confiança. Sobre isso, Daniel (1998) afirma que os testes têm sido importantes para o avanço das ciências sociais, no entanto, têm gerado controvérsias devido ao uso inadequado e à má interpretação dos resultados.

McLean e Ernest (1998) e Gliner, Leech e Morgan (2002) também apoiam os pesquisadores que recomendam que os resultados dos testes de hipóteses não somente apresentem a significância estatística, mas também que esta seja acompanhada do significado prático e da replicabilidade. Os autores ainda destacam a importância em usar-se o termo “estatisticamente significativo” no lugar da expressão “significativo”, porque nem sempre um resultado que tem significância estatística é significativo na prática, ou seja, no contexto do problema em estudo.

Segundo Batanero (2000), o ensino dos testes de hipóteses deve sofrer mudanças, que implicam a inclusão da replicação, a escolha de tamanhos de amostra ideais, a combinação de testes de hipóteses com intervalos de confiança ou estimativas do tamanho do efeito e especificação dos critérios antes do experimento. Ainda, acrescenta a necessidade de uma mudança na metodologia de ensino, no sentido de promover a participação ativa dos alunos, sendo uma das propostas, a inserção do computador nas aulas, oferecendo, assim, uma importante ferramenta de auxílio para os cálculos e representações gráficas e, também, para o tratamento de dados reais.

Neste contexto, salienta-se que o interesse em pesquisar sobre os erros cometidos pelos alunos nesse tema surgiu durante conversas com colegas professores da educação básica, licenciados em matemática e biologia. Pode-se perceber uma falta de compreensão dos conceitos envolvidos e, como consequência, os professores não conseguem enxergar aplicações dos testes em suas áreas de atuação. Um exemplo de fala de um colega, que diz “esse tal de 5%...”, mostra que conceitos fundamentais, como o nível de significância estatística, por não serem bem compreendidos, não têm significado.

Sobre isso, Sotos *et al.* (2007, p. 99) referem que as ideias de inferência são especialmente sensíveis a interpretações equivocadas e os estudantes estão frequentemente inclinados a cometê-las pelo fato da inferência requerer o entendimento e a conexão de muitos conceitos abstratos, tais como o de distribuições amostrais e o de nível de significância, entre outros (ver figura 1). Além disso, os autores (op. cit.) acrescentam que a estatística inferencial é um tópico relevante para o desenvolvimento das pesquisas nas ciências. Ainda, destacam que nos cursos de diversas áreas científicas, os testes de hipóteses e os intervalos de confiança

são ensinados aos alunos como métodos para avaliar hipóteses científicas. Assim, em função destas três realidades, os autores julgam interessante desenvolverem-se pesquisas sobre este assunto.



Figura 1 - Rede semântica com conceitos de testes de hipóteses

Nesta investigação, torna-se necessária uma diferenciação em relação à palavra erro, pois ela será empregada com o sentido de engano que o aluno comete na resolução de questões em avaliações e não, como erro em Estatística. Neste segundo uso, o termo é empregado quando se trata da imprecisão, pelo fato de trabalhar-se com dados provenientes de processos que envolvem medições ou de amostragens. O erro do aluno, por sua vez, entende-se como 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, diSessa e Roschelle (1993). Contudo, o contexto deixará claro ao leitor o sentido com o qual cada conceito está sendo empregado.

A análise de erros tem sido utilizada por diversos pesquisadores para propiciar ao aluno a reconstrução do seu conhecimento. Para Cury (2007), o erro que o aluno comete ao resolver determinado problema é um saber que ele construiu. Portanto, não deve ser visto como um elemento negativo, e sim aproveitado para elaborarem-se situações que propiciem ao aluno a reconstrução desse saber/conhecimento elaborado de forma inadequada.

O erro possibilita que o professor questione os alunos sobre suas respostas e convicções. A esse respeito, Pinto (2000) diz que, se o professor criar situações em que o aluno possa refletir sobre seu erro, ele estará contribuindo para que esse seja visto como algo positivo e não como motivo para punição. A autora considera “o erro como um elemento construtivo e de ajuda – não de sanção – para o aluno apropriar-se do conhecimento” (PINTO, 2000, p. 20).

A análise das respostas dos alunos é, portanto, uma ferramenta que o professor pode utilizar a seu favor. Por meio dela, podem-se detectar as possíveis causas dos erros e elaborarem-se atividades que possibilitem a reconstrução do conhecimento.

Nesse sentido, esta dissertação está dividida em sete capítulos. O primeiro apresenta uma introdução sobre o tema escolhido. A seguir, o segundo e o terceiro capítulo apresentam os objetivos e as questões da pesquisa, respectivamente. O quarto capítulo traz os pressupostos teóricos que embasaram a pesquisa. O quinto capítulo descreve a metodologia da pesquisa e, no sexto, são apresentados os resultados do estudo. No último capítulo são apresentadas as considerações finais.

## **2 OBJETIVOS**

### **2.1 Objetivo Geral**

O objetivo geral deste estudo é identificar quais são as dificuldades dos alunos dos cursos de Engenharia em relação aos testes de hipóteses e levantar suposições sobre suas causas.

### **2.2 Objetivos Específicos**

Como objetivos específicos para esta pesquisa, destacam-se:

- analisar e classificar os erros cometidos pelos alunos ao resolverem questões propostas em provas nas quais é avaliado o conteúdo de testes de hipóteses;
- comparar os erros identificados nas provas com a literatura da área;
- verificar os possíveis fatores que levam o aluno a cometer os erros identificados;
- investigar as opiniões dos professores que ministram a disciplina de Probabilidade e Estatística nos cursos de engenharia sobre as causas dos erros cometidos pelos alunos.

### **3 PROBLEMAS**

#### **3.1 Problema de Pesquisa**

Para a realização do presente estudo, tem-se como questão norteadora o seguinte problema de pesquisa:

**Quais são os erros mais frequentes que os alunos dos cursos de Engenharia cometem em provas nas quais são avaliadas questões que envolvem os testes de hipóteses?**

#### **3.2 Questões de Pesquisa**

A seguir, apresentam-se as questões de pesquisa que guiaram a realização deste estudo e que, ao longo do trabalho, tentaram ser respondidas.

- Quais são os erros que os alunos cometem ao resolverem questões propostas em provas nas quais é avaliado o conteúdo de testes de hipóteses?
- Os erros identificados nesta investigação já foram detectados em estudos correlatos?
- Quais são os possíveis fatores que levam o aluno a cometer os erros identificados?
- Quais as opiniões dos professores sobre as causas dos erros?

## 4 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, abordam-se os conceitos que embasaram teoricamente este estudo. Com o intuito de aproximar o leitor das ideias básicas sobre o tema desta investigação, em um primeiro momento expõem-se alguns tópicos sobre os conceitos envolvidos nos testes de hipóteses. A seguir, trata-se da teoria de aprendizagem significativa de David Ausubel, pois entende-se que a estatística e a probabilidade, a exemplo de outras disciplinas da área de matemática, mantêm uma estrutura lógica. Esta estrutura é baseada em uma rede semântica, cujos conteúdos só têm significado para o aluno quando ele consegue conectá-los ao conhecimento já existente em sua estrutura cognitiva. Além da aprendizagem significativa, aborda-se, também, a análise dos erros em Educação Matemática e faz-se uma revisão de literatura, apresentando-se estudos já realizados sobre as dificuldades dos alunos relacionadas ao conteúdo testes de hipóteses.

### 4.1 Os Testes de Hipóteses

Em muitas situações, cientistas e pesquisadores procuram explicações para os problemas que investigam, mas apenas têm acesso a uma pequena parcela de seus objetos de estudo. Por diversos motivos, é praticamente impossível investigar todos os elementos que constituem a população em estudo. Os principais obstáculos com os quais o pesquisador se depara são: o alto custo de uma pesquisa envolvendo um número elevado de casos a serem observados, o tempo necessário para realizá-la, o custo envolvido quando há necessidade de destruição de materiais (teste com lâmpadas e similares), entre outros.

Dessa forma, torna-se necessário que nesses estudos sejam utilizados procedimentos estatísticos. Dentre eles, os testes de hipóteses são ferramentas frequentemente empregadas, por terem como base geralmente pequenas amostras, por possibilitarem a inferência sobre a população, e também, por informarem o erro ao se tomar determinada decisão.

Os conceitos básicos envolvidos nos testes de hipóteses podem ser melhor compreendidos se for considerado um experimento que consiste em lançar uma moeda para

verificar se ela é honesta (ALVES-MAZZOTTI e GEWANDESZNAJDER, 2000; LOPES, 2007; VIALI, 2010). Entende-se por moeda honesta aquela em que ambas as faces têm a mesma probabilidade ( $\frac{1}{2}$ ) de caírem voltadas para cima ao ser lançada a moeda. Neste experimento, a hipótese que se deseja testar, denominada de hipótese nula ( $H_0$ ), é se a moeda é honesta. Ela é formulada com o objetivo de ser rejeitada. A hipótese que se deseja comprovar é chamada de hipótese alternativa ( $H_1$ ). Neste experimento as hipóteses são:

$H_0$ : a moeda lançada é honesta ( $\pi = 0,5$ ).

$H_1$ : a moeda lançada não é honesta ( $\pi \neq 0,5$ ).

Inicialmente, é preciso definir o número de lançamentos que serão feitos, isto é, o tamanho de amostra, pois, teoricamente, é possível lançar a moeda infinitas vezes (população infinita). Além disso, define-se a variável que será utilizada para verificar-se a moeda é ou não honesta. Como se pode perceber, nas duas frases anteriores estão expostos três conceitos tratados na Estatística Descritiva. Os conhecimentos sobre “população”, “amostra” e “variáveis” já devem estar presentes na estrutura cognitiva do aluno, para que ele possa associar ao novo conteúdo.

O tamanho da amostra será definido pelo número de vezes que a moeda for lançada. Dessa forma, se a variável escolhida for o número de caras, em 10 lançamentos, os possíveis resultados irão variar de 0 a 10 caras. A variável envolvida “número de caras” tem uma distribuição (modelo) de probabilidade, que está associado a cada uma das hipóteses:  $H_0$  e  $H_1$ . No experimento em questão, o modelo envolvido é o binomial. Este modelo está associado a fenômenos que se repetem um determinado número de vezes e as possibilidades de ocorrência de resultados são duas: sucesso e fracasso. No caso do experimento de lançar uma moeda, usa-se “cara” e “coroa”.

Conforme o que foi exposto no parágrafo anterior, o aluno também necessita de conhecimentos de probabilidade. Por isso, o conteúdo de testes de hipóteses exige que o aluno tenha conhecimentos prévios tanto sobre Estatística Descritiva quanto sobre Probabilidade.

Assumindo-se que a moeda utilizada no experimento é honesta, a probabilidade de obter-se 0 ou 10 caras é de apenas 0,2%. A figura 2 apresenta a distribuição de probabilidade da variável “número de caras”, obtida em 10 lançamentos de uma moeda equilibrada.

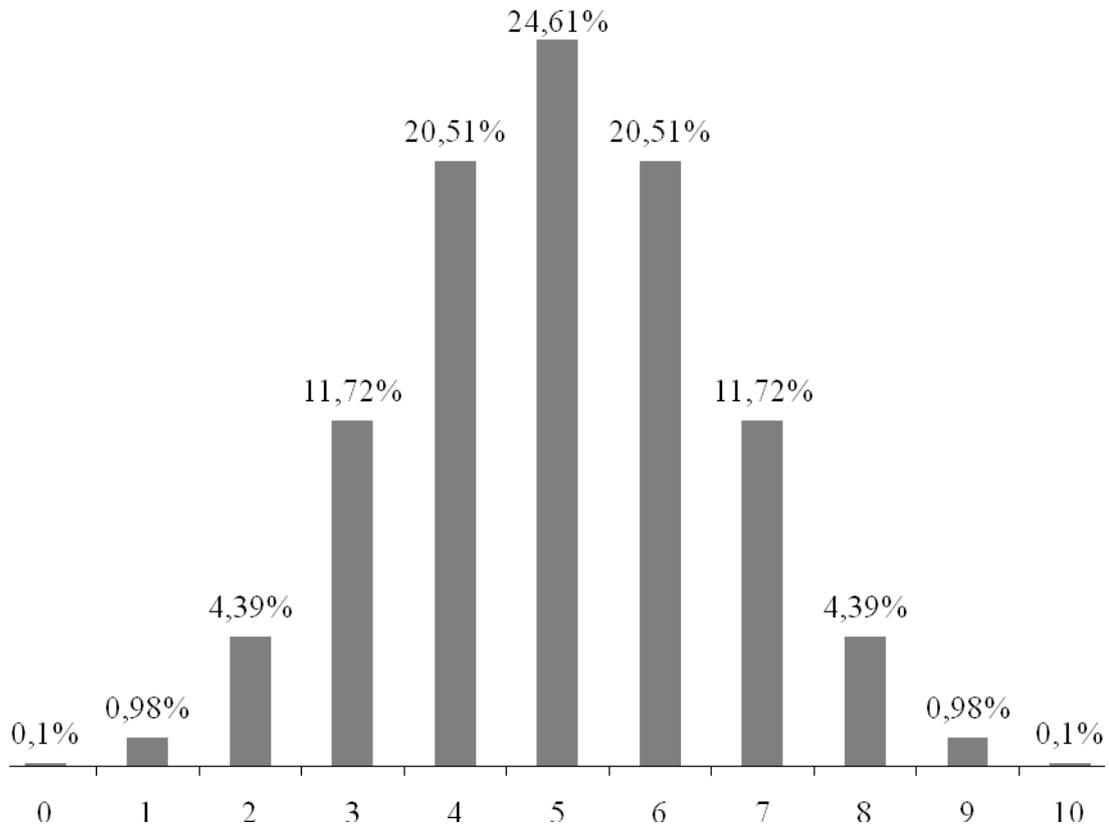


Figura 2 – Probabilidades de obter-se cara em 10 lançamentos de uma moeda equilibrada

Para tomar-se a decisão de rejeitar ou não a hipótese nula, isto é, a honestidade da moeda, é necessário estabelecer uma regra de decisão. Na situação em questão, poder-se-ia definir, por exemplo, que, se o número de caras nos dez lançamentos fosse igual a 0, 1, 2, 8, 9 e 10, a hipótese nula seria rejeitada. Ao rejeitá-la, é possível a ocorrência de um erro, pois, embora a probabilidade de ocorrer face cara 0, 1, 2, 8, 9 e 10 vezes seja pequena para uma moeda honesta, ela existe e não é zero. Este erro, que pode ocorrer ao rejeitar-se a hipótese nula quando ela é verdadeira recebe a denominação de erro do Tipo I, e a sua probabilidade de ocorrência é representada por  $\alpha$  (alfa).

Se a moeda for lançada dez vezes e não ocorrer nenhuma cara, pela regra de decisão estabelecida a hipótese nula seria rejeitada. Porém, há uma probabilidade de 0,001 de, ao lançar-se uma moeda honesta, obter-se zero caras. O estatístico Ronald Aylmer Fisher (1890 – 1962), entre 1920 e 1930, trouxe contribuições para o desenvolvimento dos testes de hipóteses (BATANERO, 2001). Para Fisher, a probabilidade 0,001 é denominada valor-p e os testes denominavam-se testes de significância. Um valor-p pequeno quer dizer que, se a

hipótese nula fosse verdadeira, os resultados observados teriam uma probabilidade baixa de ocorrer, isto é, seriam estatisticamente significativos.

Já os estatísticos Jerzy Neyman (1894 – 1981) e Egon Sharpe Pearson (1895 – 1990), deram uma interpretação frequencial a esta probabilidade. Um valor-p igual a 0,02 quer dizer que duas de cada 100 vezes que a hipótese nula é verdadeira, ela é, mesmo assim, rejeitada (BATANERO, 2001). Frequentemente, a hipótese nula é rejeitada se o valor-p encontrado for menor que 0,01, 0,05 ou, em algumas situações, 0,10. Estes valores são estabelecidos pelo pesquisador e correspondem à probabilidade do erro do Tipo I tolerada, também chamados de nível de significância do teste.

Pode ocorrer, também, um segundo tipo de erro, denominado erro do Tipo II, que consiste em não rejeitar a hipótese nula quando, na realidade, ela é falsa, isto é, quando ela deveria ser rejeitada. A probabilidade de isto ocorrer é representada por  $\beta$  (beta), ou seja,  $P(\text{Erro do Tipo II}) = \beta$ . Para determinar-se a probabilidade de ocorrência do erro do Tipo II, é necessário conhecer o valor de  $H_1$ , que, em geral, é desconhecido. Assim, para determinar-se essa probabilidade, é necessário fazer suposições sobre a realidade, isto é, sobre  $H_1$ . Desta forma, a probabilidade de cometer-se o erro do Tipo II é uma função que pode assumir um valor, vários valores ou uma infinidade (não) enumerável de valores.

A probabilidade complementar do erro do Tipo II denomina-se poder do teste. Quanto menor o erro do Tipo II, mais poderoso é o teste, isto é, maior a probabilidade de tomar-se a decisão correta de rejeitar a hipótese nula quando ela realmente é falsa. Em outras palavras, poder do teste é a “probabilidade de detectar aquela hipótese alternativa, se ela for verdadeira” (SALSBURG, 2009).

Além de depender da probabilidade de ocorrência do erro do tipo II e, conseqüentemente, da probabilidade de ocorrência do erro do Tipo I, o poder do teste também é influenciado pelo tamanho da amostra e pelo tamanho do efeito. O aumento do tamanho da amostra sempre acarreta um aumento no poder do teste. Conseqüentemente, amostras muito grandes tornam os testes de hipóteses muito sensíveis, isto é, capazes de detectar efeitos não significativos na prática. O tamanho do efeito, definido como a real magnitude do efeito na população, influencia no poder do teste pelo fato de um grande efeito ter maior probabilidade de ser detectado, elevando dessa forma o poder do teste.

Os testes de hipóteses, portanto, não permitem que se conheça, com certeza, se uma hipótese é verdadeira ou falsa. O que se determina nos testes de hipóteses é a probabilidade do erro cometido ao se tomar uma determinada decisão. Ao não rejeitar-se uma

hipótese, não se quer dizer que ela seja necessariamente verdadeira, assim como rejeitá-la não implica necessariamente em sua falsidade.

O ensino dos testes de hipóteses pode seguir dois enfoques: o clássico baseado nas ideias de Newman e Pearson e o computacional que segue a proposta de Fisher. Segundo o enfoque clássico, conforme Bussab e Morettin (2005) e Viali (2010), as etapas para resolver um problema geralmente incluem seis passos, sendo eles:

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

A decisão é tomada com base na região crítica previamente definida, isto é, conforme a regra de decisão estabelecida que fornece o ponto a partir do qual a hipótese nula é rejeitada. Este valor fixado antes da realização do teste é o nível de significância do teste.

Já na abordagem de Fisher, calcula-se a significância do resultado, isto é, o valor-p, ficando a critério do leitor a interpretação do resultado. Neste enfoque o valor-p é definido como a probabilidade de obter-se o mesmo valor que a estatística teste ou um valor mais extremo se a hipótese nula for verdadeira. Um valor-p pequeno significa que é provável que o valor do parâmetro seja diferente do que o valor enunciado na hipótese nula. Ou seja, é possível que o resultado não se deva ao acaso.

Pelo exposto nesta seção, faz-se necessária uma explanação sobre a teoria de aprendizagem proposta por David Ausubel. Trata-se, basicamente, da importância do conhecimento prévio do aluno, bem como da organização do ensino para facilitar a aprendizagem de conceitos.

## **4.2 Aprendizagem Significativa**

A ideia central da teoria de aprendizagem significativa de David Ausubel é a influência do conhecimento prévio do aluno na aprendizagem de novos conceitos. Ausubel

afirma que o fator mais influente para uma aprendizagem significativa é aquilo que o aprendiz já sabe. Por isso, o autor aconselha que os professores criem situações didáticas com a finalidade de descobrir esses conhecimentos, que se referem a um dos conceitos Ausubelianos mais conhecidos, denominado “organizadores prévios”.

Os organizadores prévios servem de âncora para a nova aprendizagem e desenvolvem conceitos subsunçores que facilitam a aprendizagem subsequente. A função deles “é servir de ponte entre o que o aprendiz já sabe e o que ele deve saber, a fim de que o material possa ser aprendido de forma significativa” (MOREIRA e MASINI, 1982, p.12). Segundo Coll *et al.* (1998, p. 61),

uma aprendizagem é tanto mais significativa quanto mais relações com sentido o aluno for capaz de estabelecer entre o que já conhece, seus conhecimentos prévios e o novo conteúdo que lhe é apresentado como objeto de aprendizagem.

Para Ausubel (apud MOREIRA e MASINI, 1982, p. 7), “aprendizagem significativa é um processo pelo qual uma nova informação se relaciona com um aspecto relevante da estrutura de conhecimento do indivíduo”. Em oposição à aprendizagem significativa, Ausubel (apud MOREIRA; MASINI, 1982, p. 8) define a “aprendizagem mecânica como sendo a aprendizagem de novas informações com pouca ou nenhuma associação com os conceitos relevantes existentes na estrutura cognitiva”. Na aprendizagem mecânica, portanto, não há interação entre a nova informação e a já armazenada.

Dentre os aspectos relevantes da teoria Ausubeliana, pode-se destacar a aquisição de conceitos, pois constitui fundamentação da aprendizagem significativa. Segundo Ausubel, Novak e Hanesian (1980), os conceitos são adquiridos através da formação e da assimilação. A formação de conceitos é característica da criança em idade pré-escolar. A criança adquire espontaneamente ideias genéricas por meio de experiências empírico-concretas. Pode-se dizer que é um tipo de aprendizagem por descoberta, pois consiste em um processo de abstração dos aspectos comuns característicos de uma classe de objetos que varia em um contexto.

Após a infância, no ambiente escolar, a aquisição de conceitos ocorre através da assimilação. Para Ausubel, Novak e Hanesian (1980), a assimilação é um processo que ocorre quando um conceito, potencialmente significativo, é assimilado sob uma ideia ou conceito mais inclusivo, já existente na estrutura cognitiva. A estrutura de conhecimento específica existente na estrutura cognitiva do indivíduo, que serve de âncora para a nova informação, é denominada por Ausubel de “subsunçores”. Na interação, tanto a nova informação como o conceito subsunçor sofrem modificações, resultando em um novo significado.

Para Ausubel, o armazenamento de informações no cérebro é altamente organizado, formado por uma hierarquia conceitual, na qual os elementos mais específicos de conhecimento são ligados e assimilados a conceitos mais gerais, mais inclusivos. Para ele, o desenvolvimento dos conceitos é facilitado quando os elementos mais gerais, mais inclusivos de um conceito são introduzidos em primeiro lugar, para então serem progressivamente diferenciados.

O conteúdo testes de hipóteses, desenvolvido na Estatística Inferencial (ver Figura 1) presta-se particularmente a essa hierarquização. Assim, os conteúdos de Estatística podem tirar proveito das ideias de Ausubel, pois são altamente organizados e estruturados. Essas estruturas podem ser vistas como uma rede de interconexões que se assemelham ao processo mental descrito.

Na figura 3, retirada de Viali (2008), apresenta-se a estrutura básica necessária para compreenderem-se as ideias abordadas na estatística inferencial.

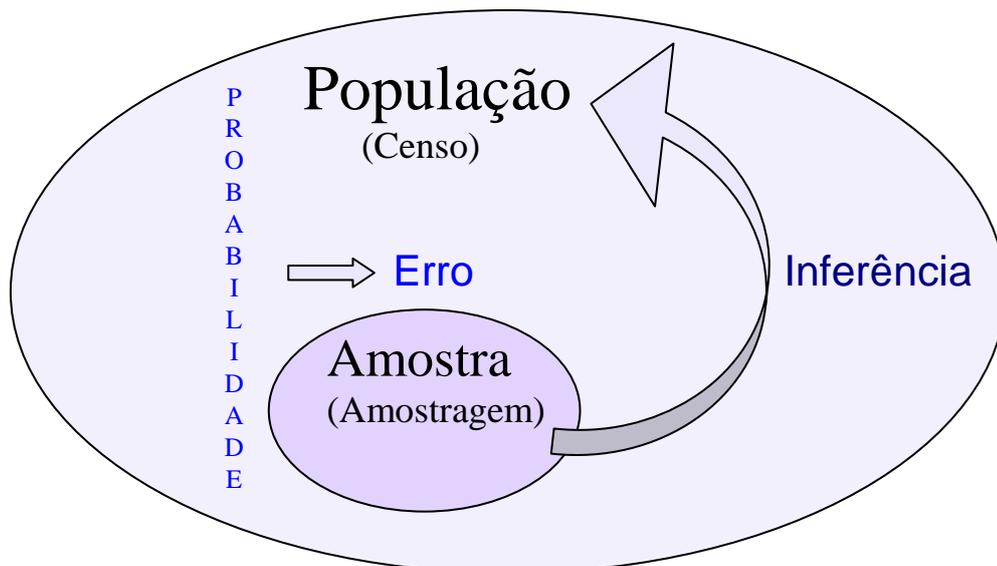


Figura 3 – Estrutura conceitual da amostragem

Delmas *et al.* (2007) enfatizam que já foram realizados estudos que indicam que os estudantes apresentam dificuldades em relação às distribuições amostrais e suas representações gráficas, e também aos conceitos relacionados à variação estatística, como as medidas de variabilidade e a variabilidade amostral. Os erros cometidos em conteúdos de estatística inferencial possivelmente possam estar associados a essas dificuldades identificadas. Assim, como exposto anteriormente, não é possível se obter uma aprendizagem de conceitos de inferência, ao menos satisfatória ou significativa, sem haver uma

compreensão prévia de conceitos de descrição de dados amostrais e um mínimo de conceitos probabilísticos, bem como de conceitos de amostragem.

### 4.3 Análise de Erros

Segundo Cury (2007, p. 17), a análise de erros “é uma tendência em Educação Matemática como área de conhecimento”. Batanero (2001, p. 66) afirma que grande parte das investigações da Didática da Matemática se interessa em explicar porque o aluno erra quando resolve determinadas tarefas. A autora acrescenta que muitas vezes certos erros se repetem e se produzem regularidades e associações com variáveis próprias das tarefas propostas. Pinto (2000) e Cury (2007) apresentam uma lista de trabalhos, estrangeiros e brasileiros, sobre a análise de erros em questões matemáticas.

A análise das produções dos alunos possibilita ao professor descobrir como eles apropriam-se de um saber e quais as influências das aprendizagens anteriores. Através da identificação dos erros, o professor pode fazer questionamentos, propiciando situações para o aluno construir seu conhecimento (CURY, 2007). Conforme Pinto (2000, p. 50), “trata-se de compreender o percurso do aluno nas tarefas propostas, sua representação sobre as tarefas e suas estratégias para resolvê-las. [...] A partir dessas informações, o professor reorganizará as atividades pedagógicas”.

O teste de hipóteses, por ser um conteúdo que envolve um grande número de conceitos e procedimentos, propicia também ao aluno cometer uma série de equívocos. Há diversos tipos de testes: teste para a diferença entre médias, análise de variância, testes não paramétricos, testes multivariados. Porém, todos eles têm em comum alguns conceitos básicos, como: hipótese nula e alternativa, estatística teste, nível de significância (erro do Tipo I), poder do teste (erro do Tipo II), dentre outros. Para aplicar corretamente os testes, o aluno precisa realizar uma série de escolhas, como, por exemplo, o tamanho da amostra, o nível de significância e a estatística apropriada ao problema (BATANERO, 2001).

Conforme a pesquisadora Carmen Batanero (2001, p. 106) as investigações sobre a compreensão dos métodos de inferência mostram que os estudantes têm concepções errôneas e as dificuldades apresentadas por eles são:

- (a) a determinação da hipótese nula ( $H_0$ ) e da hipótese alternativa ( $H_1$ );
- (b) a distinção entre os erros do tipo I e do tipo II;
- (c) a compreensão do propósito das curvas características de operação;

- (d) a compreensão da terminologia empregada para estabelecer a decisão; e
- (e) a decisão tomada e a interpretação em termos do problema sendo testado.

Dentre os testes de hipóteses, desde os mais simples, como os que são utilizados para verificar se um modelo é ou não adequado, até os mais complexos testes multivariados, exigem uma série de pré-conceitos que são comuns a todos. Desde a elaboração das hipóteses a serem testadas até a interpretação dos resultados, o aluno necessita de conhecimentos prévios, tais como: o conceito de população e amostra, aleatoriedade, variabilidade, distribuição de probabilidade da população e da amostra, distribuição de probabilidade das médias e da variância, variável aleatória, probabilidade condicional, entre outros. A presença destes conceitos na estrutura cognitiva do aluno possibilita que os conceitos relacionados aos testes de hipóteses ancorem-se neles e, assim, adquiram significado.

Sotos *et al.* (2007) identificaram os principais equívocos de estudantes universitários em relação aos testes de hipóteses por meio da análise de publicações sobre estudos que fornecem evidências empíricas sobre os mesmos. Segundo os autores, as abordagens para os testes de hipóteses, que muitas vezes misturam a de Neyman-Pearson com a de Fisher, podem causar confusão na interpretação dos resultados, pois nesta etapa as duas teorias diferem. Segundo a abordagem de Fisher, se a hipótese nula é verdadeira e o valor-p é pequeno, então o resultado é estatisticamente significativo, assumindo-se que os dados forneceram evidências contra a hipótese nula. Todavia, se a abordagem de Neyman-Pearson for seguida, então o valor-p é comparado com o nível de significância, previamente estabelecido, e é então tomada uma decisão para rejeitar-se ou não a hipótese nula. O nível de significância, de acordo com esta abordagem, é interpretado como um erro tolerável que se assume ao rejeitar-se a hipótese nula quando ela é verdadeira.

Além da influência dessa questão metodológica, os conceitos que não são compreendidos também são causadores de equívocos. Detectou-se que os alunos têm dificuldades em entender o significado da hipótese nula e alternativa, a natureza condicional do nível de significância, a interpretação do valor-p, a lógica dos testes de hipóteses e a avaliação da significância estatística. A seguir, apresentam-se alguns estudos nos quais estes equívocos foram evidenciados.

Dentre as pesquisas realizadas sobre as dificuldades na compreensão do teste de hipóteses, apresentam-se os estudos de Vallecillos e Batanero (1997), Haller e Krauss (2002), Link (2002), Rodríguez (2006), Delmas *et al.* (2007) e Sotos *et al.* (2009).

Vallecillos e Batanero (1997) apresentam parte de uma investigação realizada com uma amostra de sete estudantes universitários do curso de Medicina, que responderam a um questionário. A primeira questão era composta por três itens, que deveriam ser julgados como verdadeiros ou falsos: um sobre a compreensão do conceito de teste de hipóteses e dois sobre a interpretação do nível de significância. Na segunda questão, foi apresentado um problema típico sobre testes de hipóteses envolvendo a média de uma variável com distribuição normal em um contexto de controle de qualidade. Na última questão, foi apresentada uma tabela de contingência com dados sobre a capacidade diagnóstica de um teste para detectar câncer. Os resultados obtidos foram analisados qualitativamente e 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 relação à primeira categoria, apenas um aluno não soube interpretar corretamente as probabilidades condicionais na tabela de contingência, apesar de ter utilizado a notação correta para representá-las. No entanto, quando os alunos são solicitados a responderem as questões sobre a interpretação do nível de significância, apenas um aluno responde corretamente. Todos os demais interpretam o nível de significância como sendo a probabilidade da hipótese nula ser verdadeira dado que ela foi rejeitada. Sobre a formulação das hipóteses, três alunos relacionam a  $H_0$  (hipótese nula) com a região de aceitação e a  $H_1$  (hipótese alternativa) com a região de rejeição; três deles acreditam que a hipótese pode tanto referir-se a um parâmetro da população como também a uma estatística da amostra; e cinco alunos têm dificuldade em estabelecer as hipóteses de acordo com o contexto do problema. Em relação à lógica global do processo, cinco alunos acreditam que o procedimento permite calcular a probabilidade de uma hipótese e seis alunos não diferenciam a média da amostra e a média da população.

A investigação de Haller e Krauss (2002) procurou identificar as concepções de professores e alunos sobre alguns conceitos envolvidos nos testes de hipóteses. Foi elaborada uma questão com seis itens que deveriam ser julgados como verdadeiros ou falsos, aplicada a professores e alunos de seis universidades alemãs. Os participantes foram divididos em três grupos: os professores de estatística ( $n = 30$ ); os professores (psicólogos) não envolvidos no ensino da estatística ( $n = 39$ ); e os estudantes de psicologia ( $n = 44$ ).

Nesta investigação, destaca-se o resultado encontrado nos itens sobre o conceito de um teste de hipóteses. Dos 30 professores de estatística, 10% concordam com a seguinte

declaração: “a hipótese nula é absolutamente refutada”. Dentre os 39 professores que não ensinam estatística, 15% concordam com esta afirmação e, dos 44 alunos de psicologia, 34% estão de acordo. Os autores suspeitam que muitos equívocos possam estar relacionados aos livros de estatística utilizados pelos professores, além de os próprios professores terem concepções equivocadas dos conceitos.

Link (2002), por sua vez, analisou testes realizados por 295 estudantes de duas disciplinas do curso de Ciências Biológicas, sendo que uma tinha a disciplina de Cálculo como pré-requisito e a outra apenas a Álgebra do Ensino Médio. Foram avaliados os erros cometidos em uma série de etapas: na formulação das hipóteses, na determinação da estatística teste e do valor crítico, no cálculo da estatística teste, na comparação da estatística teste com o valor crítico, na decisão e na determinação do valor-p.

No estudo de Link (2002), os alunos tiveram maiores dificuldades em relação à determinação da estatística teste e do valor crítico, sendo o percentual de respostas incorretas igual a 47%. Os erros foram categorizados da seguinte forma: escolha da estatística teste correta, mas determinação do valor crítico incorreto (39,8%); escolha da estatística teste incorreta (4,1%); e resposta sem sentido (3,1%). Em segundo lugar, destacam-se as dificuldades na formulação das hipóteses, sendo de 31,6% o percentual de respostas incorretas. Em relação a esta etapa, foram identificadas quatro categorias de erros: não identificação do parâmetro correto, por exemplo,  $\mu$  ao invés de  $\pi$  (12,6%); declarações sobre a estatística ao invés do parâmetro (5,8%); valor do parâmetro incorreto, erro de sinal e de direção (12,6%); e declarações sem sentido (0,7%). Em seguida, destaca-se a etapa da determinação do valor-p, com 26,8% de respostas incorretas, e, também, a etapa do cálculo da estatística teste, na qual o percentual de resposta incorretas foi de 25,8%. Por último, o item que teve o menor percentual de erros foi a etapa da tomada da decisão, com apenas 16,3% de respostas incorretas.

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 por duas partes: uma para identificar o conhecimento conceitual, com onze itens para assinalar verdadeiro ou falso, e outra com três problemas de aplicação. A principal concepção errônea identificada por Rodríguez (2006) foi em relação ao nível de significância. Os estudantes o consideraram 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  $P(B|A)$ . Em torno de 65% dos

estudantes respondeu incorretamente a um item sobre a definição do nível de significância. Relacionado à interpretação frequencial do nível de significância, eles consideram que um nível de 0,05 quer dizer que, em média, cinco em cada 100 vezes que a hipótese nula é rejeitada, ela é verdadeira.

O estudo de Delmas *et al.* (2007) consiste na validação e na aplicação de um instrumento com 40 itens a estudantes que cursavam pela primeira vez uma disciplina de estatística no ensino superior. O instrumento foi aplicado no início e no final da disciplina e os resultados foram comparados. Dentre os itens, havia um que avaliava a compreensão sobre a lógica dos testes de hipóteses na situação em que a hipótese nula é rejeitada. Dos 716 estudantes que responderam ao item, um percentual de 26,7% no pré-teste (e 32,4% no pós-teste) acreditavam que “rejeitar a hipótese nula significa que a hipótese nula é definitivamente falsa” (p. 46). Outro item que merece ser destacado neste trabalho é o que se refere à interpretação do valor-p. No pré-teste, dos 712 estudantes que responderam ao item, 53,2% não interpretaram corretamente o valor-p, sendo que no pós-teste este percentual reduziu para 45,5%.

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 em um teste de hipóteses: a definição de um teste de hipóteses, a interpretação do valor-p e a interpretação do nível de significância. Além de detectarem-se os erros cometidos por esses estudantes, também, investigou-se sobre a confiança que eles têm em seus erros.

Dentre os erros mais frequentes, neste estudo, os alunos acreditavam que um teste de hipótese é uma prova matemática da hipótese nula, ou então que é uma prova probabilística por contradição. Com relação ao valor-p, o erro mais comum foi considerá-lo como sendo a probabilidade de cometer-se um erro ao rejeitar-se a hipótese nula. O conceito de nível de significância foi o mais confuso para os estudantes. Os erros mais comuns foram considerar, se o resultado do teste foi estatisticamente significativo para um nível de significância de 0,05, a probabilidade de rejeitar a hipótese nula é igual a 0,95 ou, então, a probabilidade da hipótese nula ser verdadeira é igual a 0,05.

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 ainda 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, assim, o comparativo dos resultados da presente pesquisa será realizado com os cursos encontrados (agronomia, biologia, medicina e microbiologia), embora saliente-se que o ideal seria fazer um confronto com outros estudos que abordassem alunos desta mesma área.

Acredita-se, portanto, que o mapeamento dos erros cometidos pelos alunos ao resolverem questões sobre testes de hipóteses possa contribuir para o ensino da Estatística, uma vez que são poucos os estudos existentes. Dessa forma, acredita-se ser possível mostrar aos professores onde concentram-se as principais dificuldades dos alunos.

## 5 METODOLOGIA

Neste capítulo, será apresentada a metodologia utilizada para a realização deste estudo, trazendo-se informações sobre os participantes que compõem o *corpus* da pesquisa, sobre a coleta dos dados e, também, sobre as técnicas de análise dos dados.

### 5.1 Delineamento da Pesquisa

Muitas vezes, o real sentido da palavra pesquisa não é entendido pelas pessoas que a usam. Lüdke e André (1986) destacam que a pesquisa tornou-se tão popular que o verdadeiro significado da palavra está comprometido. A mídia, principalmente, utiliza a pesquisa para fins políticos, ao apresentar as tendências eleitorais com objetivos, não raras vezes, de induzir uma determinada ideia aos eleitores. Na escola, os professores pedem aos alunos para pesquisarem sobre um tema, em uma enciclopédia ou em uma página da *web*. Os mesmos autores consideram que estes fatos levam as pessoas a acreditarem que a pesquisa reduz-se a estes aspectos somente. Entretanto, o seu objetivo central é responder questionamentos referentes a problemas que surgem no dia a dia. De acordo com Lüdke e André (1986, p. 1-2),

para se realizar uma pesquisa é preciso promover o confronto entre os dados, as evidências, as informações coletadas sobre determinado assunto e o conhecimento teórico acumulado a respeito dele. Em geral isso se faz a partir de um estudo de um problema, que ao mesmo tempo desperta o interesse do pesquisador e limita sua atividade de pesquisa a uma determinada porção do saber.

Uma pesquisa pode ter três abordagens: a quantitativa, a qualitativa e a mista. A diferença fundamental entre as duas primeiras refere-se aos objetivos, o que explica os métodos de seleção da amostra, as técnicas de coleta e a análise de dados de cada abordagem. O presente estudo terá uma abordagem predominantemente qualitativa. O objetivo principal, portanto, não é quantificar os erros, mas sim analisar cada caso, categorizando-os conforme as semelhanças apresentadas.

## 5.2 Participantes

A investigação realizada neste estudo contou com alunos matriculados na disciplina de Probabilidade e Estatística para Engenharia, no segundo semestre de 2009. As turmas escolhidas para participarem da pesquisa são de três universidades do Rio Grande do Sul, sendo uma delas pública e as demais privadas. O quadro 1 mostra as turmas de cada universidade e o número de alunos envolvidos na pesquisa.

Quadro 1 – Participantes da pesquisa

Universidades	Turmas	Número de alunos
X	A	35
	B	35
	C	50
Y	D	44
Z	E	10
	F	4
	G	11

Ainda, foram participantes da pesquisa os professores destas turmas, com os quais foi aplicado um questionário. No total, participaram seis professores, dos quais três eram da universidade X, um da universidade Y e dois da universidade Z.

## 5.3 Instrumentos de Coleta de Dados

Basicamente, os instrumentos de coleta de dados abordam avaliações realizadas pelos alunos e questionários respondidos pelos professores. As questões das avaliações foram elaboradas pelos professores das disciplinas, e os questionários foram construídos com base nos erros identificados nestas avaliações.

O questionário dos professores (ver Apêndice A), que contém questões abertas e fechadas, foi elaborado após a classificação dos erros detectados nas avaliações. Na primeira parte do questionário, foram apresentadas algumas soluções nas quais foram identificados os erros que apareceram com maior frequência. Solicitou-se aos professores que colocassem sua opinião sobre as causas destes erros. Na segunda parte do questionário, foram feitas algumas

perguntas sobre a abordagem do conteúdo. Para finalizar, foram formuladas duas questões pessoais.

## 5.4 Técnicas de Análise de Dados

Para analisar os dados coletados, foi utilizada a metodologia de análise de conteúdo. Conforme Cury (2007) 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.

Bardin (op. cit.) apresenta três fases para a análise de conteúdo: a pré-análise, a exploração do material e o tratamento dos resultados. Na primeira fase, escolhem-se os documentos, formulam-se as hipóteses e os objetivos da análise e elaboram-se indicadores para fundamentar a interpretação final. Neste momento, o pesquisador faz uma leitura flutuante do material selecionado para constituir o *corpus*, isto é, o conjunto de documentos selecionados para a análise.

Para a referente etapa definiu-se 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 analisadas, optou-se por 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 catorze enunciados diferentes (ver Anexo A), correspondentes a provas aplicadas em sete turmas, elaboradas 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. Uma unidade de registro é definida, segundo a autora, como uma “unidade de significação a codificar e corresponde ao segmento de conteúdo a considerar como unidade de base, visando à categorização e à contagem frequencial” (BARDIN, 2009, p.

130). As unidades de registro, conforme Bardin (2009), podem ser palavras, frases, temas ou documentos, entre outros.

Após, as unidades de registro são agrupadas em categorias, que “tem como primeiro objetivo [...] fornecer, por condensação, uma representação simplificada dos dados brutos” (BARDIN, 2004, p. 146-147). Cada grupo recebe um título e reúne unidades de registro com características comuns. 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. O título de cada grupo é definido ao final do processo, após a definição das categorias finais que são resultantes de um reagrupamento progressivo.

Nesse contexto, um conjunto de categorias deve apresentar algumas características que o qualificam. Segundo Bardin (2009), as qualidades são as seguintes: exclusão mútua, homogeneidade, pertinência, objetividade e fidelidade e produtividade. Uma mesma unidade de registro pode estar presente em apenas uma classe, por isso elas devem ser homogêneas, sendo constituídas de apenas um princípio de classificação. O conjunto de categorias, por sua vez, deve estar de acordo com os objetivos da investigação e, ao ser submetido a diversas análises, deve ser classificado da mesma maneira. Preferencialmente, o conjunto de tais grupos deve fornecer um resultado que ainda não foi encontrado em outras investigações.

Na fase da definição das unidades de registro e das categorias, inicialmente, foram criados códigos para identificarem-se 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 respostas corretas, das que apresentavam algum tipo de erro e também daquelas que não foram respondidas. Dentre as respostas com erros, foram identificadas aquelas que seriam passíveis de análise. Destas, foram destacados os tipos de erros cometidos e a frequência de ocorrência de cada um, fazendo-se um registro para, então, categorizá-los.

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, o pesquisador precisa fazer o tratamento dos resultados, bem como proceder às inferências e às interpretações (BARDIN, 2004). Cury (2007, p. 65) afirma que a descrição das classes pode ser feita através da apresentação de tabelas com a distribuição de frequências das unidades, além de aconselhar a produção de textos-síntese.

## 6 RESULTADOS E DISCUSSÃO

### 6.1 Tipos e Categorias de Erros

Nesta investigação foram seleccionadas 357 respostas (ou soluções) de avaliações realizadas por alunos de sete turmas, sobre dois tipos de testes de hipóteses envolvendo uma amostra. O quadro 2 mostra o número e o percentual de respostas constituintes do *corpus* de cada tipo de teste.

Quadro 2 – Número e percentual de respostas analisadas

Univer- sidade	Professor/ Turma	Teste para a Média		Professor/ Turma	Teste para a Proporção	
		Nº de respostas	%		Nº de respostas	%
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
Z	5/E	10	5,6	6/G	11	6,1
	6/F	4	2,2	6/F	4	2,2
Total		178	100		179	100

No quadro 3, a seguir, é apresentado o resultado do levantamento do número de respostas corretas, erradas e em branco. Dentre as 325 soluções erradas investigadas, em 23 delas foram dadas respostas sem sentido, totalizando, desta forma, 302 respostas efetivamente analisadas.

O quadro 3 mostra que são poucas as soluções resolvidas corretamente, sendo que, na maioria delas, é cometido pelo menos um tipo de erro. Alguns destes erros serão posteriormente exemplificados por meio das respostas apresentadas pelos alunos. Pode-se afirmar que os dados apresentados no quadro 3 evidenciam que as dificuldades dos alunos

neste conteúdo aparecem nas três universidades, como também em todas as turmas e professores.

Quadro 3 – Tipos de testes e respostas corretas, erradas e em branco

Universidade	Professor/ Turma	Teste para a média			Teste para a proporção		
		corretas	Respostas com erros	Respostas em branco	Respostas corretas	Respostas com erros	Respostas em branco
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
Z	5/E	0	10	0	0	11	0
	6/F	0	4	0	0	4	0
Total		10	167	1	9	158	12

A análise das respostas erradas possibilitou a identificação dos vários tipos de erros relacionados às diversas etapas de um teste de hipóteses. Segundo Link (2002), avaliar os erros cometidos nas etapas é uma forma de testar a compreensão do aluno sobre os testes de hipóteses, pois os erros em determinados passos indicam que a compreensão plena ainda não foi alcançada. Dessa forma, inicialmente, no quadro 4 apresenta-se a frequência com que as seis etapas que envolvem a resolução de um problema sobre testes de hipóteses paramétrico, citadas anteriormente, são respondidas corretamente (C), incorretamente (I) ou não são realizadas (NR).

Percebe-se, pelo quadro 4, que o maior número de respostas incorretas, assim como de casos em que o aluno não realizou a etapa, ocorreram na formulação da conclusão, enquanto na tomada da decisão a frequência de acertos é a maior. Nesse contexto, traz-se a figura 4 que mostra o percentual de acertos e de erros e, também, o percentual de soluções nas quais não foi respondida alguma das etapas seguidas na resolução dos problemas sobre testes de hipóteses.

Quadro 4 – Quantidade de acertos, erros e não realização das seis etapas dos testes de hipóteses

Etapas		Teste para a Média	Teste para a Proporção	Total
Definição das hipóteses	C	95	89	184
	I	54	38	92
	NR	12	14	26
Determinação da estatística teste e da distribuição amostral	C	100	134	234
	I	61	3	64
	NR	0	4	4
Identificação e posicionamento da região crítica ou nível de significância do teste	C	103	76	179
	I	32	35	67
	NR	26	30	56
Cálculo da estatística teste	C	132	95	227
	I	29	40	69
	NR	0	6	6
Decisão	C	139	118	257
	I	7	5	12
	NR	15	18	33
Conclusão	C	49	24	73
	I	76	90	166
	NR	36	27	63

Na figura 4, observa-se que, das 302 respostas analisadas, em apenas 24,2% delas a conclusão do problema é formulada adequadamente. Por outro lado, em 55,0% são cometidos erros e em 20,9% os alunos nem sequer a formulam. Em todas as demais etapas, o percentual de acertos é superior a 50%, indicando que é na formulação da conclusão que os equívocos ocorrem com maior frequência.

Na etapa da formulação das hipóteses, também se percebem muitos erros, como se verifica na figura 4, em que o percentual de respostas incorretas é de 30,5%. Na investigação de Link (2002), que analisou testes realizados por 295 estudantes do curso de Ciências Biológicas, o percentual de respostas incorretas na etapa da formulação das hipóteses foi de 31,6%, enquanto 67% estavam corretas e 1,4% não foram formuladas. Estes resultados estão próximos aos verificados nesta investigação, corroborando, assim, com este estudo.

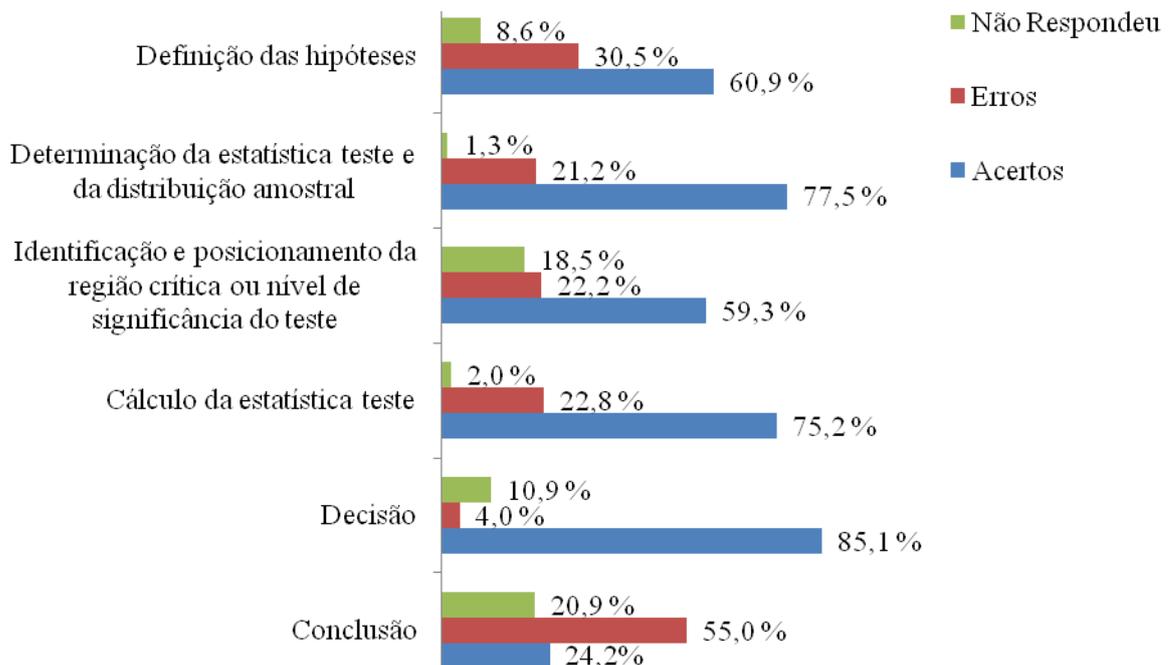


Figura 4 – Percentual de acertos, erros e não realização das etapas dos testes de hipóteses

Nota-se, também, que na determinação e no cálculo da estatística teste são observados os menores percentuais de não realização das etapas. Provavelmente, este fato indique que a ênfase é dedicada ao cálculo e não ao processo como um todo. Considerando-se que, hoje em dia, dificilmente na prática as análises quantitativas são realizadas sem o uso de um *software* estatístico, a prioridade deveria ser a compreensão dos conceitos envolvidos, para, então, saber-se utilizar os resultados corretamente.

O percentual de respostas incorretas que Link (2002) detectou na determinação da estatística teste foi de apenas 4,1%. Enquanto 86,7% das respostas foram corretas em seu estudo, 3,1% foram de respostas sem sentido e 6,1% foi o percentual de questões em que não foi realizada a etapa. No cálculo do valor da estatística teste, 65,1% foi o percentual de respostas corretas, 25,8% de respostas erradas, 0,3% de respostas sem sentido e, em 7,8% das questões, a etapa não foi realizada.

Sobre esse resultado, Link (2002) afirma que a determinação correta do valor da estatística teste não quer dizer que o aluno tenha a compreensão plena do processo dos testes de hipóteses. Considerando-se que o processo de substituir corretamente valores em uma fórmula é um procedimento que não envolve o raciocínio probabilístico, aqueles alunos que têm uma razoável base matemática conseguem resolver esta etapa sem maiores dificuldades.

Na etapa da identificação e posicionamento da região crítica, 18,5% dos alunos deixaram de fazer a sua representação, 59,3% a fizeram corretamente e 22,2% cometeram erros. Dentre os erros, os mais frequentes foram o posicionamento incorreto da região crítica conforme o tipo de teste utilizado, a identificação incorreta dos graus de liberdade e o uso do valor crítico do teste unilateral ao invés do bilateral, e vice-versa.

Na etapa referente à da decisão de rejeitar ou não a hipótese nula, observou-se o maior percentual de acertos, 85,1%. Por ser uma etapa que envolve apenas uma comparação do valor calculado da estatística teste com a região crítica estabelecida, a possibilidade de errar, neste caso, é pequena, pois o aluno, novamente, necessita somente do conhecimento matemático elementar. Link (2002) também encontrou em sua investigação um percentual elevado de respostas corretas (78,6%). Segundo o autor, como não era necessário realizar cálculos nem procurar algum valor em tabelas, provavelmente “chutes” ajudaram a aumentar o percentual de acertos, pois em muitos casos a resposta correta não estava relacionada aos valores observado e crítico da estatística teste.

Os erros verificados em cada uma das etapas, e também aqueles não diretamente relacionados a uma das seis etapas, constituindo erros externos ao procedimento, foram identificados e numerados. A seguir, são apresentados todos os tipos de erros encontrados, que serão referidos a partir de números:

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;

07: uso de  $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$  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;

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 (e, conseqüentemente, de aceitação);

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;

25: uso de  $\sigma$  (parâmetro) ao invés de  $s$  (estatística);

- 26: uso de  $\mu$  (parâmetro) ao invés de  $\bar{x}$  (estatística);
- 27: uso de  $\bar{x}$  (estatística) ao invés de  $\mu$  (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  $\pi$  (proporção populacional) ao invés de  $\mu$  (média populacional);
- 31: uso de  $x$  ao invés de  $\mu$  (parâmetro);
- 32: uso de  $p$  (estatística) ao invés de  $\pi$  (parâmetro);
- 33: uso de  $\pi$  (parâmetro) ao invés de  $p$  (estatística);
- 34: uso de  $\mu$  (média populacional) ao invés de  $\pi$  (proporção populacional);
- 35: uso de  $\bar{x}$  (média amostral) ao invés de  $p$  (proporção amostral);
- 36: não rejeição da hipótese nula quando isto deveria ser feito;
- 37: rejeição da 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;
- 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.

Quadro 5 – Frequência de ocorrência dos erros identificados

Tipos de erros	Frequência	%	Tipos de erros	Frequência	%	Tipos de erros	Frequência	%
<b>01</b>	<b>41</b>	<b>6,4</b>	15	1	0,2	29	1	0,2
<b>02</b>	<b>38</b>	<b>6,0</b>	<b>16</b>	<b>22</b>	<b>3,4</b>	30	1	0,2
03	16	2,5	17	2	0,3	31	1	0,2
04	11	1,7	18	2	0,3	32	12	1,9
05	9	1,4	<b>19</b>	<b>17</b>	<b>2,7</b>	33	7	1,1
06	6	0,9	20	13	2,0	34	7	1,1
<b>07</b>	<b>61</b>	<b>9,6</b>	<b>21</b>	<b>19</b>	<b>3,0</b>	35	2	0,3
08	1	0,2	22	3	0,5	36	7	1,1
09	2	0,3	23	5	0,8	37	2	0,3
10	11	1,7	24	10	1,6	38	1	0,2
11	13	2,0	<b>25</b>	<b>50</b>	<b>7,8</b>	39	2	0,3
12	7	1,1	26	16	2,5	<b>40</b>	<b>137</b>	<b>21,5</b>
13	6	0,9	27	9	1,4	<b>41</b>	<b>38</b>	<b>6,0</b>
14	5	0,8	28	5	0,8	<b>42</b>	<b>19</b>	<b>3,0</b>

Os dez tipos de erros que ocorreram com maior frequência nas respostas analisadas estão destacados no quadro 5, e correspondem a 69,3% dos erros identificados. Estes erros serão posteriormente apresentados e discutidos.

Após a identificação, 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 apresentados no quadro 6, a seguir.

Quadro 6 – Categorias e tipos de erros

Categorias	Tipos de erros
1: envolve erros na formulação das hipóteses.	01, 02, 03, 04, 05 e 06
2: formada por erros cometidos na escolha da distribuição de probabilidade para a padronização da estatística teste.	07, 08 e 09
3: engloba os erros relacionados com a definição ou delimitação da região crítica.	10, 11, 12, 13, 14, 15, 16, 17 e 18
4: envolve erros cometidos no cálculo da estatística teste.	19, 20, 21, 22, 23 e 24
5: engloba a utilização de notação inadequada para representar valores referentes à população e à amostra.	25, 26, 27, 28, 29, 30, 31, 32, 33, 34 e 35
6: formada por erros de tomada da decisão.	36, 37, 38 e 39
7: envolve os erros identificados na formulação da conclusão.	40, 41 e 42

O quadro 7, por sua vez, fornece uma ideia global sobre a frequência de ocorrência dos erros em cada categoria. Os dados verificados neste quadro indicam que os estudantes estão apresentando maiores dificuldades em relação à formulação das hipóteses, à utilização da notação correta e à formulação da conclusão.

Quadro 7 – Quantidade e percentual de erros em cada categoria

<b>Categorias</b>	<b>Erros</b>	<b>%</b>
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
<b>Total</b>	<b>638</b>	<b>100</b>

Somando-se as categorias 1 (formulação das hipóteses) e 7 (elaboração da conclusão), tem-se praticamente 50% dos erros cometidos pelos alunos investigados. Destaca-se, ainda, que o menor número de erros, referente à categoria 6, com 1,9% do total, ocorre na tomada da decisão, isto é, no momento de não rejeitar 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 de maneira correta, mas em seguida interpreta errado, já que os erros de conclusão são os mais numerosos.

Após a exposição dos resultados agregados, 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-se os principais tipos de erros cometidos.

## **6.2 Análise dos Erros**

A partir dos tipos de erros apresentados na seção anterior, parte-se para a análise desses erros, considerando-se as soluções fornecidas pelos alunos durante a realização das avaliações. As figuras 5, 6 e 7, a seguir, trazem as soluções referentes ao seguinte enunciado: *Uma marca particular de margarina diet foi analisada para determinar o nível (em percentagem) de ácidos graxos insaturados. Uma amostra de 6 potes de 500g foi investigada,*

resultando o seguinte: média de 17%, com desvio padrão de 0,4%. Se o nível de ácidos graxos insaturados deve ser inferior a 18%, verifique se isto está ocorrendo, com  $\alpha = 1\%$ . Nelas são identificados os erros do tipo 02, 07, 16 e 25 (figura 5), 16 (figura 6) e 03, 21, 41 e 42 (figura 7).

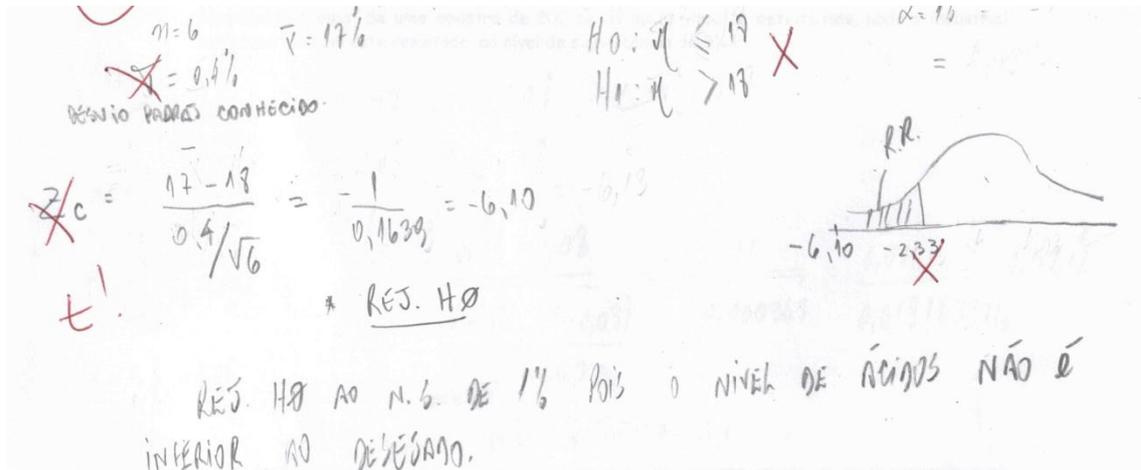


Figura 5 – Ilustração dos erros tipo 02, 07, 16 e 25

O erro do tipo 02 corresponde a 31,4% dos erros classificados na primeira categoria (que envolvem os erros na formulação das hipóteses). Na investigação de Link (2002), também foi verificado este tipo de erro na formulação das hipóteses. Segundo ele, foram identificados quatro tipos de erros na formulação das hipóteses: 12,6% dos alunos formularam as hipóteses utilizando o parâmetro populacional errado; 5,8% fizeram afirmações sobre a estatística da amostra; em 12,6% das vezes os alunos cometeram um erro no valor do parâmetro, do sinal ou da direção; e, 0,7% formularam hipóteses sem sentido.

A hipótese alternativa deve ser formulada de acordo com aquilo que o pesquisador deseja comprovar, enquanto a hipótese nula expressa a intenção contrária. O estudo de Vallecillos e Batanero (1997) também 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 corretamente. Segundo as pesquisadoras, a dificuldade dos estudantes para formular corretamente as hipóteses pode estar associada à falta de elementos que dão significado ao conceito de hipótese estatística.

Além da incompreensão da ideia básica que é elaborar a hipótese nula com o objetivo de ser rejeitada, que está diretamente ligado aos conceitos de erro do Tipo I e do Tipo II, destaca-se ainda, a falta de conhecimentos matemáticos elementares. Ainda, o aluno pode

ter equivocado-se ao usar os sinais de desigualdade, por não saber transpor a linguagem verbal para a linguagem matemática.

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  $\sigma$ , que é utilizada para representar o desvio padrão populacional. O aluno ainda escreve “desvio padrão conhecido”, mostrando que não está considerando que este é um valor que pode ser obtido tanto da população como também da amostra. Dessa forma, pode-se inferir que para ele não há diferença entre esses dois valores.

Esse fato foi também observado na investigação de Vallecillos e Batanero (1997), e mostra que há dificuldade de compreensão com relação a esses dois universos distintos: o universo abstrato da população e o universo concreto da amostra. As pesquisadoras dizem que os estudantes não apenas confundem a notação para representar parâmetros e estatísticas, mas em geral eles não são capazes de diferenciar estes conceitos.

Uma provável causa para esta dificuldade pode estar relacionada ao tipo de ensino que predomina nos cursos de Engenharia. Sobre isso, Ara (2006) destaca que as aulas de Estatística e Probabilidade nesses cursos seguem a linha tradicional, com aulas expositivas, que priorizam as técnicas ao invés da construção dos conceitos. No estudo de Ara (2006), é feita a análise de dez livros didáticos utilizados pelos professores em sala de aula, constatando-se que em apenas 30% dos livros analisados predomina a concepção aleatória da realidade. Assim, o autor conclui “que a maioria dos professores de Probabilidade e Estatística nos cursos de Engenharia adotam uma concepção determinística da realidade” (ARA, 2006, p. 79).

Ainda em relação à figura 5, 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 realizado por ele para definir a região crítica dessa forma refere-se ao valor padronizado da estatística teste que resultou em um número negativo. O aluno, de certa forma, não está tendo uma compreensão global do processo envolvido na resolução de um problema de testes de hipóteses. Inicialmente, para formular as hipóteses, o aluno utiliza um teste unilateral à direita. Após, ele realiza o cálculo da estatística teste e verifica que o resultado é um valor negativo, estabelecendo, provavelmente por causa desse fato, a região crítica à esquerda. No

entanto, parece que o aluno não se dá conta de verificar a contradição entre a hipótese que foi formulada e a região crítica.

Na figura 6, apresenta-se o recorte de uma solução em que também é identificado o erro tipo 16 e, ainda, um erro no cálculo da estatística teste. Provavelmente, este último erro pode ter ocorrido por causa da transcrição incorreta do algoritmo, pois o aluno faz a diferença entre o valor do parâmetro e a estatística, resultando em um valor positivo. Ainda, convém ressaltar que, apesar de o aluno acertar a formulação das hipóteses, ao definir a região crítica ele não observa que na definição da hipótese alternativa foi utilizado um teste unilateral à esquerda, definindo, então, a região crítica à direita.

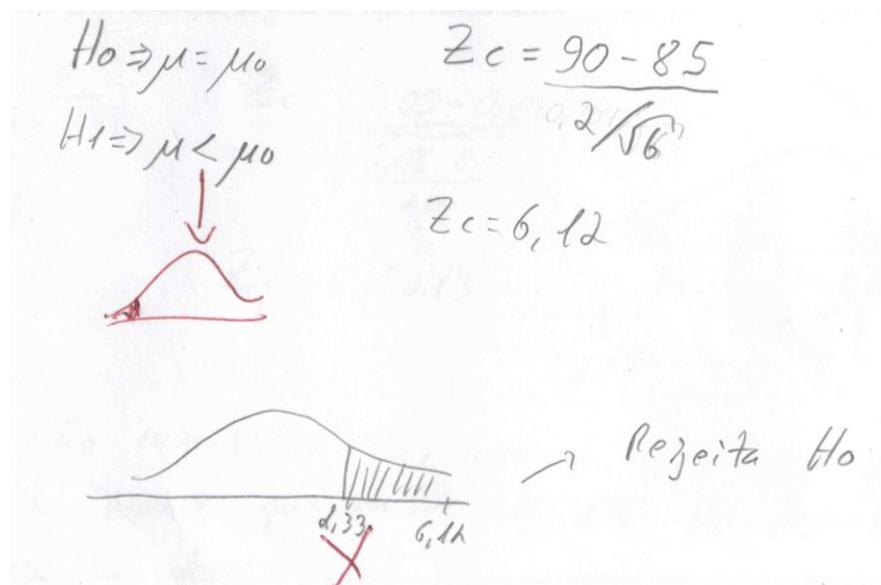


Figura 6 – Ilustração do erro tipo 16

Percebe-se que o aluno não estabelece conexão entre as etapas seguidas – provavelmente, ele associou a região crítica ao valor positivo da estatística teste, que não foi calculado corretamente. Assim, também na figura 5, parece que a região crítica foi definida de acordo com o valor da estatística teste e não conforme o tipo de teste utilizado.

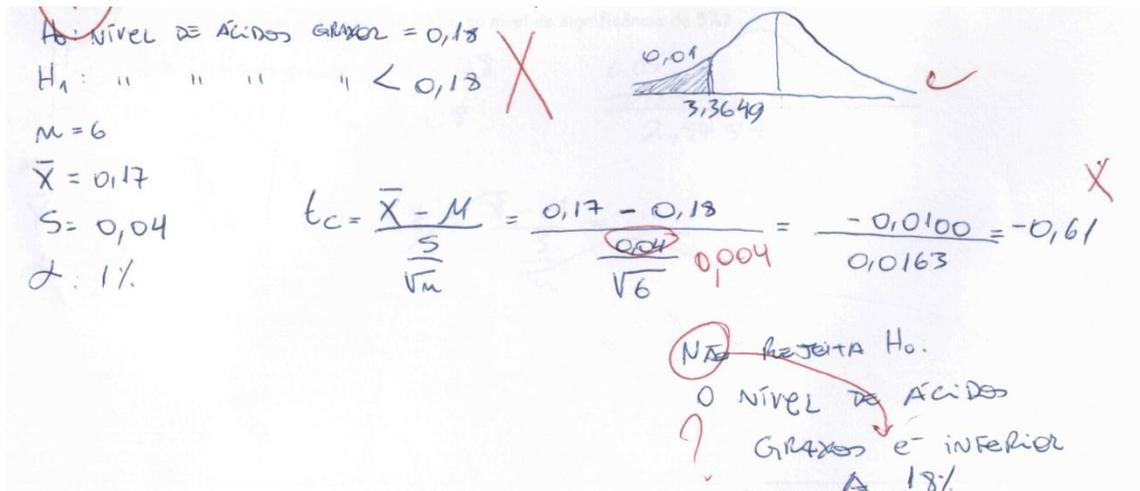


Figura 7 – Ilustração dos erros tipo 03, 21, 41 e 42

Na solução apresentada na figura 7, não foi utilizada a notação adequada para representar a média populacional no enunciado das hipóteses (erro tipo 03). Além disso, houve dificuldade no cálculo de um percentual (erro tipo 21). 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 (erro tipo 41). Ainda, comete o erro tipo 42, quando afirma o contrário daquilo que foi enunciado na hipótese nula.

Nota-se que o aluno não compreendeu a ideia de que não se podem fazer afirmações sobre a hipótese nula quando não foi possível rejeitá-la. Provavelmente, ele não tem compreensão dos erros envolvidos no processo (erro do Tipo I e erro do Tipo II) e, dificilmente, irá entender o porquê desta lógica. O erro tipo 42 mostra novamente que o aluno não consegue estabelecer as conexões entre as etapas, ou então, pode representar que ele tenha dificuldades em interpretar as hipóteses.

No recorte da solução apresentada na figura 8, o aluno utiliza a notação correta para representar a média da amostra, no entanto, para enunciar as hipóteses comete um erro. Ele utiliza o valor correto do parâmetro, mas não se dá conta que a notação para representá-lo é a da média amostral. Neste caso, parece não estar clara para este aluno a função de um teste de hipóteses, pois subentende-se que o estudante está testando a média da amostra que já é conhecida. Ainda, comete um erro ao representar o desvio padrão amostral com a letra grega  $\sigma$  e ao calcular o percentual, que o levou a encontrar um valor incorreto da estatística teste.

Figura 8 – Ilustração dos erros tipo 03, 07, 21 e 25

O próximo enunciado refere-se às soluções apresentadas nas figuras 9 e 10. O Ford Taurus é classificado como tendo uma eficiência média de combustível em rodovia de 30 quilômetros por litro (1995 Motor Trend New Car Buyer's Guide). Um grupo de consumidores realizou testes de quilometragem de automóveis, procurando por evidências estatísticas para mostrar que os fabricantes de automóveis exageraram as avaliações de quilômetros por litro para determinados modelos. Em uma amostra de 25 testes de quilometragem com o Ford Taurus, o grupo de interesse de consumidores encontrou uma avaliação às amostras de quilometragem média em rodovia de 25,5 quilômetros por litro e um desvio padrão de 1,8 quilômetros por litro. Que conclusão deve ser tirada a partir dos resultados obtidos? Use um nível de significância de 0,01.

Figura 9 – Ilustração dos erros tipo 02, 03, 07, 25 e 40

Nas figuras 9 e 10 (a seguir), pode-se observar um erro tipo 40 na formulação da conclusão. Esse tipo de erro ocorreu com muita frequência na categoria 7, sendo de 70,6% o 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. Vallecillos e Batanero (1997) também constataram que, para alguns alunos, as hipóteses referem-se a enunciados ou a situações desconhecidas, nas quais se deseja determinar se são absolutamente verdadeiras ou falsas.

Na literatura, são apontadas algumas concepções equivocadas que os alunos, em geral, possuem sobre a lógica de um teste de hipóteses. Nas investigações de Vallecillos (1999), Haller e Krauss (2002), Delmas *et al.* (2007) e Sotos *et al.* (2009), um número considerável de alunos entendem o teste de hipóteses como uma prova matemática. Uma

possível causa dessa dificuldade pode estar associada a pouca ênfase atribuída aos conceitos básicos e aos erros envolvidos, visto que praticamente todos os livros-texto disponíveis priorizam os algoritmos, deixando de enfatizar a interpretação e a análise, bem como o aspecto não determinístico do resultado. Além disso, poucos mencionam ou enfatizam o erro do Tipo II, além do fato de ele ser desconhecido ou não controlado.

A realização de atividades envolvendo simulações e experimentos para apresentar aos alunos os erros envolvidos nos testes de hipóteses, como o erro do Tipo I e do Tipo II, pode ajudar a superar a visão determinística que a maioria dos alunos expressou nas conclusões dos problemas. Ainda, é necessária, para a compreensão desses conceitos envolvidos nos testes de hipóteses, uma boa compreensão de conceitos de amostragem. Dessa forma, atividades que envolvam simulações para observar as distribuições amostrais dos estimadores possivelmente facilitarão o estabelecimento de conexões entre os três contextos, dois próprios da estatística (amostra e população) e um da matemática (teoria da probabilidade). Batanero (2000) defende a ideia de utilizar simulações, pois podem contribuir para o entendimento dos alunos sobre a variabilidade da amostra e das distribuições amostrais que são essenciais para a compreensão da lógica dos testes de hipóteses.

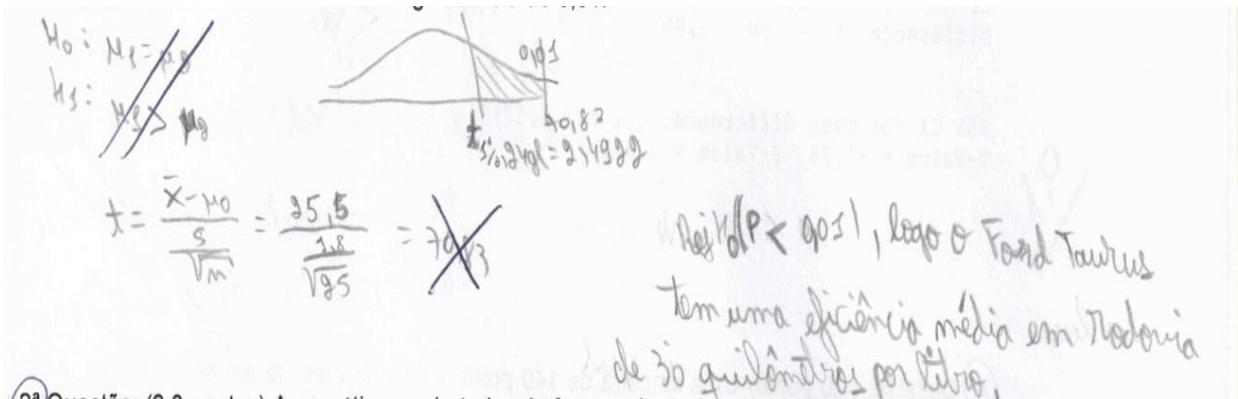


Figura 10 – Ilustração dos erros tipo 02, 03, 20 e 40

Muitas vezes, como se pode observar nas soluções das figuras 9 e 10, o professor que realizou a correção não apontou o erro tipo 40. No entanto, na figura 11, o professor destaca que o aluno não fez referência ao nível de significância. Assim, a visão determinista do professor pode ser uma das barreiras para o desenvolvimento do raciocínio probabilístico dos alunos.

Rejeitar  $H_0$ , ao nível... então o nível do ácido graxo está abaixo de 18.

Figura 11 – Ilustração do erro tipo 40

Outro tipo de erro identificado nas figuras 9 e 10 é o erro tipo 03, em que o aluno utiliza uma notação inadequada para enunciar as hipóteses. Nos casos observados nessas soluções, os alunos utilizam a notação  $\mu_1$  e  $\mu_2$  do teste para a diferença entre duas médias populacionais. Isso indica que, provavelmente, eles não distinguem a média da população (desconhecida) e a média da amostra.

Ainda, na figura 9, o aluno utiliza a letra grega  $\sigma$  para representar o desvio padrão da amostra, assim como, utiliza a estatística Z ao invés da distribuição t. E, na figura 10 é identificado um erro no cálculo, quando esquece de subtrair a média populacional. Convém destacar que nas soluções apresentadas nas figuras 9 e 10 os alunos utilizam a abordagem de Fisher, mencionando a significância do resultado, ou seja, o valor-p.

Nas figuras 12, 13 e 14 apresentam-se três soluções que se referem ao seguinte enunciado: *Um industrial considera satisfatório se menos de 8% das peças produzidas por sua indústria forem defeituosas. Se uma amostra de 200 peças apresentou 18 defeituosas, pode o industrial satisfazer-se com este resultado, ao nível de significância de 5%?*

$n = 200$   
 $\alpha = 0,05$   
 $x = 18$   
 $\pi_0 = 0,08$   
 $P = \frac{18}{200} = 0,09$

$H_0: \pi = \pi_0$  teste unilateral  
 $H_1: \pi < \pi_0$

$Z_c = \frac{0,09 - 0,08}{\sqrt{\frac{0,08(1-0,08)}{200}}} = 0,494166$

$\tilde{n}$  rejeita  $H_0$

$P > \alpha$ , portanto não se rejeita  $H_0$ , podendo o industrial satisfazer-se com o resultado, considerando 5% de significância

Figura 12 – Ilustração dos erros tipo 20 e 41

Na figura 12, apresenta-se uma solução na qual o aluno calcula o erro padrão da proporção utilizando o valor da proporção obtido na amostra (erro tipo 20). Novamente, percebe-se que há indiferença no tratamento dos valores provenientes de dados da população e 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 do estudante em entender que a não rejeição da hipótese nula, a partir da evidência amostral, não permite fazer afirmações sobre ela, pois não se conhece o erro do

Tipo II envolvido. Nesta conclusão, o estudante também apresenta certa confusão entre o nível de significância e o nível de confiança utilizado nos intervalos de confiança.

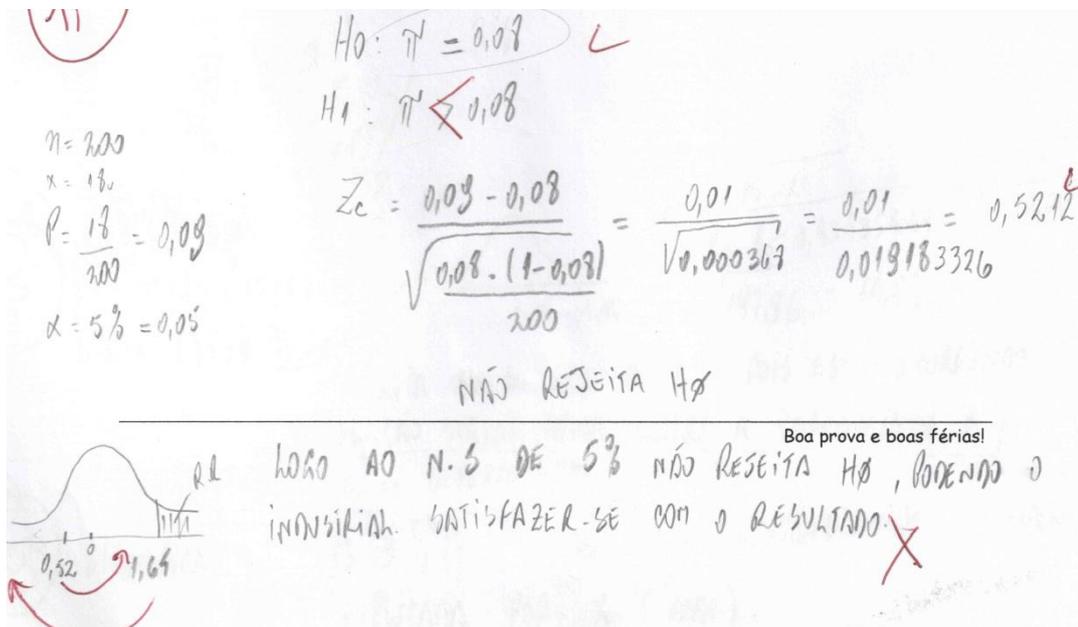


Figura 13 – Ilustração dos erros tipo 02 e 41

Nas soluções apresentadas nas figuras 13 e 14, foi cometido um erro na formulação das hipóteses (erro tipo 02). Os alunos tiveram dificuldade em interpretar corretamente o problema e, provavelmente, foram induzidos a cometer este erro pelo fato da proporção amostral ter sido maior que o parâmetro populacional.

Na figura 13, ainda que o aluno cometa um erro ao posicionar o valor da estatística teste no eixo, considerando-se as hipóteses formuladas, ele conclui corretamente, mas comete, também, o erro tipo 41 ao fazer afirmações sobre a hipótese nula.

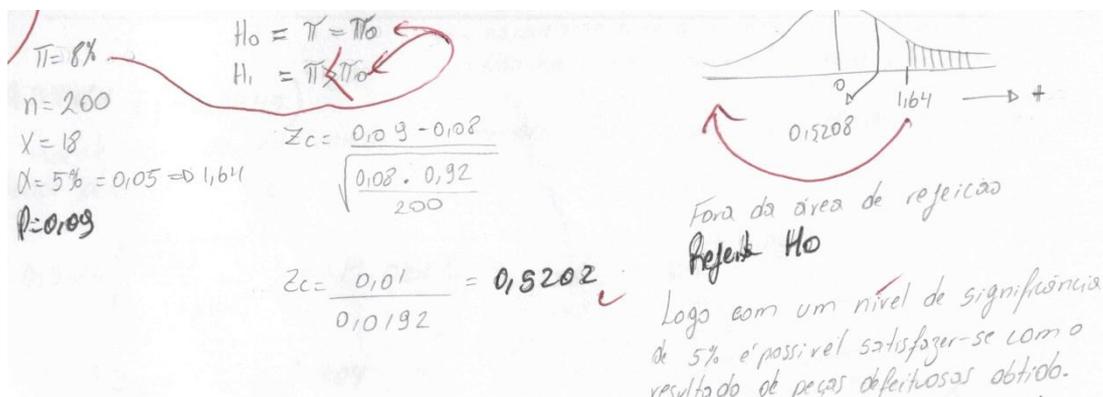


Figura 14 – Ilustração dos erros tipo 02 e 37

Na solução da figura 14, o aluno toma a decisão errada, mesmo escrevendo “fora da área de rejeição”. Possivelmente, isto tenha sido um lapso ou falta de atenção, pois a maioria dos alunos não teve dificuldades nesta etapa. Contudo, na conclusão o aluno expressa a ideia como se não tivesse rejeitado a hipótese nula.

O próximo enunciado refere-se à solução apresentada na figura 15: *Uma agência de viagens tem um tradicional plano de férias que é oferecido a seus clientes. O índice de respostas positivas é historicamente 20%. Este ano, uma amostra aleatória de 50 clientes mostrou que 15 adquiriram o plano. Teste ao nível de 5% de significância a hipótese de que o índice tenha aumentado.* Na figura 15, é identificado o erro tipo 01 (uso do teste bilateral ao invés do teste unilateral), que apresentou percentual de ocorrência de 6,4%. Ainda, aparecem erros na formulação da conclusão, mas que já foram analisados anteriormente.

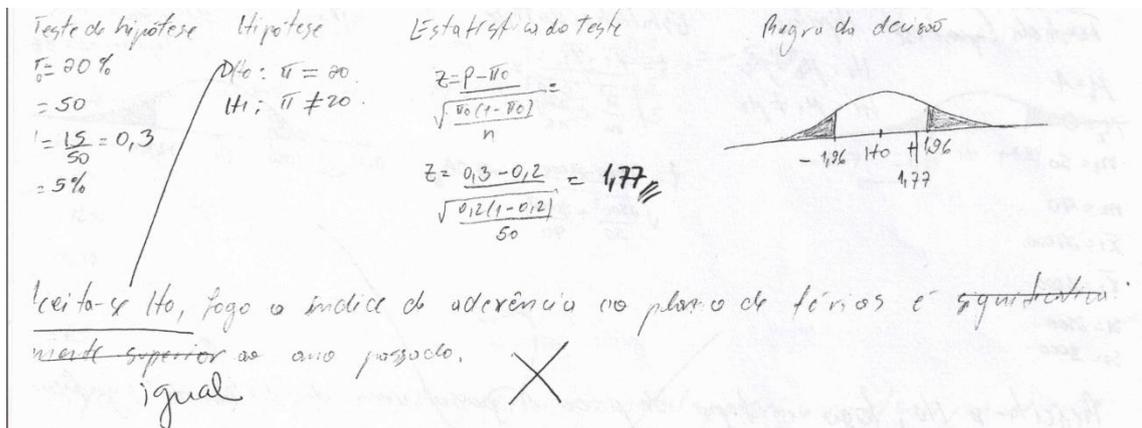


Figura 15 – Ilustração dos erros tipo 01, 41 e 42

Na solução da figura 15, o erro tipo 01 não foi destacado pelo professor que corrigiu a prova. Porém, seria mais coerente se o aluno tivesse usado um teste unilateral à direita, pois deseja-se testar a hipótese de que o índice de respostas positivas tenha aumentado. Nota-se novamente a presença da visão determinística, ao não se fazer referência ao fato de que se tivesse utilizado um teste unilateral à direita a hipótese nula seria rejeitada. Simplesmente, na conclusão o aluno afirma que aceita a hipótese nula, sem cogitar essa possibilidade de rejeitá-la, dependendo do tipo de teste utilizado. Pôde-se observar, diversas vezes, que o aluno tem a tendência de pensar que o resultado obtido é verdadeiro, sem considerá-lo probabilístico.

Já na solução apresentada na figura 16, a seguir, o aluno testa a hipótese utilizando um teste unilateral à esquerda e, ainda, um teste bilateral. Na conclusão por ele

apresentada, também, menciona que, com um nível de significância de 5%, rejeita a hipótese nula tanto com o teste bilateral à esquerda como com o teste bilateral.

2. Uma revista resolve testar a afirmação de uma fábrica de automóveis que seus veículos fazem 12 km por litro de combustível. A revista analisa 25 automóveis, obtendo nessa amostra uma média de 11,5 km por litro e um desvio padrão amostral de 1,2 km por litro.

a) Que tipo de teste deve ser efetuado? teste para a média

b) Apresente suas conclusões frente a este teste de hipóteses, usando 5% de significância.

Com uma significância de 5%, podemos dizer que a média não é de 12 km/litro. Calculando unilateralmente, podemos dizer ainda que (com 5% de significância) a média é menor que os 12 km/litro prometidos

---

2)  $\bar{x} = 11,5$   
 $H_0: \mu = 12$      $n = 25$      $t_{n-1; \alpha} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$      $t$  calculado  
 $H_1: \mu \neq 12$      $s = 1,2$      $\downarrow$   
 $\mu < 12$      $= \frac{11,5 - 12}{1,2 / \sqrt{25}} = -2,0833$   
 $t_{24; 0,1}$  SA unilateral     $\Rightarrow$   $1,7109$      $H_1: \mu \neq 12$   
 Se bilateral  $\Rightarrow 2,0639$

$|t_{\text{calculado}}| > |t_{\text{tabelado}}| \Rightarrow$  rejeita  $H_0$



Figura 16 – Ilustração de uma questão resolvida corretamente

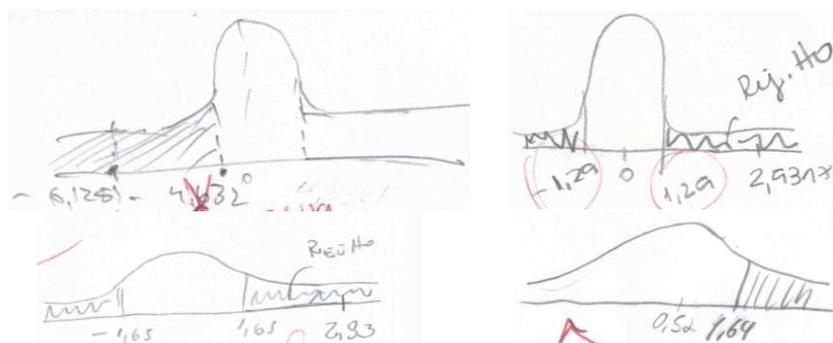
Nestas diferentes soluções apresentadas, estão ilustrados alguns dos erros detectados na análise das respostas. Salienta-se que as partes essenciais na resolução de um problema desse conteúdo são a formulação das hipóteses e a conclusão. Visto que, para escrever ou interpretar os resultados em um artigo científico, por exemplo, a etapa que envolve os algoritmos é feita, geralmente, por meio de um software, ainda assim a interpretação dos resultados não é 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, em sua conclusão, afirma “rejeita-se  $H_0$ , prova-se que este automóvel consome menos que 30 km/l”, mostra-se que este aluno não incorporou a ideia de que esse resultado foi obtido em uma amostra e que na população não necessariamente isso é verdadeiro. Nota-se que o aluno não assume que o valor calculado com base em uma amostra é um resultado estimado, ou seja, uma aproximação do verdadeiro valor que se obteria ao trabalhar-se com a população.

O nível de significância, para esse aluno, parece não ter significado, sendo ignorado na formulação da conclusão. Dessa mesma forma, também alguns professores não têm destacado este erro nas correções (figura 9 e 10), considerando a questão como resolvida corretamente. Dessa forma, a análise das respostas dos alunos levou à constatação de que muitos dos conceitos envolvidos em um teste de hipóteses têm pouco ou nenhum significado para eles. Isso indica que a aprendizagem não foi significativa, pois não foram estabelecidas as conexões entre os conceitos conforme propõe a teoria de Ausubel.

Para finalizar, destaca-se a pouca ou quase nenhuma atenção do aluno em relação a algumas ideias básicas sobre a distribuição normal, ao fazerem a representação gráfica da curva. Na figura 17, são apresentadas algumas representações dos alunos nas quais certas propriedades da curva normal não são observadas: a simetria em torno da média, a tendência da curva a se aproximar do eixo  $x$  quando o domínio tende ao infinito e o ponto de inflexão a um desvio padrão da média.

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. No mesmo sentido, os erros tipo 20 (uso de valores incorretos por não diferenciar parâmetro de estatística) e tipo 07 (uso da estatística  $Z$  ao invés da  $t_{n-1}$ ) podem resultar da prevalência de uma visão fragmentada dos conceitos, não havendo necessidade de diferenciar dados provindos de uma população ou de uma amostra.



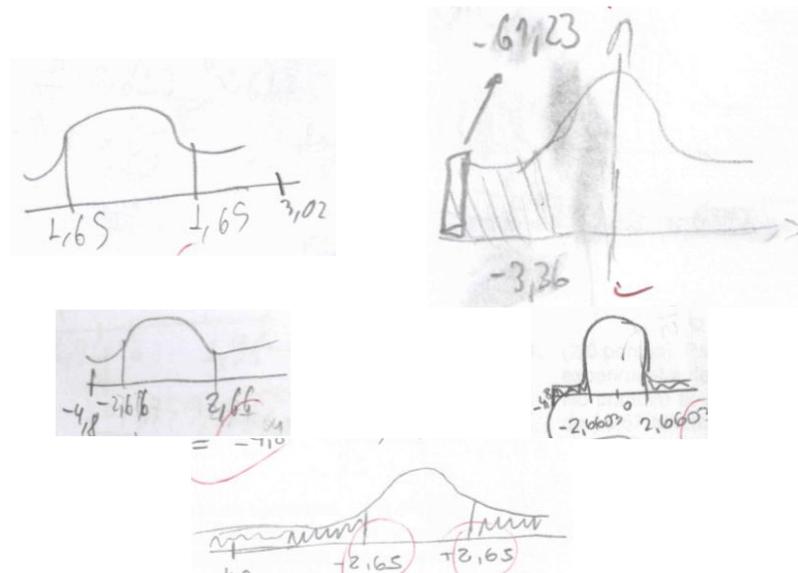


Figura 17 – Ilustração da curva normal

### 6.3 Análise das Respostas dos Professores

O questionário elaborado após a classificação dos erros foi aplicado a seis professores das turmas envolvidas na pesquisa. Com relação à formação, todos os professores são graduados em Estatística, sendo que todos possuem mestrado. Dois dos professores lecionam a disciplina de Probabilidade e Estatística há 20 anos, um deles há 22 anos, um há 3 anos, um há 2,5 anos e o outro há 2 anos, mas lecionou outras disciplinas por 13 anos.

No que se refere à abordagem dos testes de hipóteses em suas aulas, foram feitas três perguntas. A primeira delas refere-se ao tipo de abordagem que o professor costuma utilizar em suas aulas: região crítica, valor-p ou ambas. Três professores responderam que costumam utilizar ambas, dois responderam que utilizam a região crítica e um costuma utilizar a região crítica e, em algumas turmas, ambas.

A segunda pergunta versava sobre o momento em que os conceitos, tais como, erros do Tipo I (nível de significância), Tipo II (poder de um teste), CCO (Curva Característica de Operação), região crítica e valor-p, são introduzidos e quantas horas/aula são dedicadas à sua abordagem. Quatro professores responderam que abordam os conceitos antes da introdução dos tipos de testes, sendo que um deles dedica, em geral, meia hora/aula, outro em torno de uma hora, o terceiro duas horas/aula e, o último, quatro horas/aula. Um professor respondeu que aborda os tipos de erros e a região crítica antes da introdução dos tipos de

testes e o valor-p durante o estudo dos tipos de testes. Ainda, um dos professores respondeu que aborda os conceitos durante o estudo dos tipos de testes.

A terceira e última pergunta sobre a forma de abordagem do conteúdo indagava o professor se eram apresentados exemplos e oportunizados exercícios envolvendo o cálculo do erro do Tipo I, do erro do Tipo II e da curva característica de operação (CCO). Com exceção de um professor, todos os demais responderam que não realizam estas atividades com seus alunos. Este professor realiza exercícios em que o aluno determina o erro do Tipo I para situações em que a região crítica é dada ou o erro do Tipo II quando o valor do parâmetro sob a hipótese alternativa é fornecido. Também aplica um exercício no qual o aluno determina uma função para o poder, dependendo do valor do parâmetro sob a hipótese alternativa.

Quadro 8 – Categorias relacionadas às causas dos erros segundo a opinião dos professores

Categorias	Sub-categorias	Frequência
Não compreensão do conteúdo:	da estrutura geral	3
	das hipóteses formuladas	2
	da conexão entre as etapas	3
	dos objetivos do teste	2
Falta de pré requisitos:	conhecimentos de matemática	3
	conhecimentos de interpretação de textos	5
Falta de empenho do aluno:	falta de atenção	2
	falta de estudo	2
	falta de exercícios	3

Além do exposto, também foram apresentadas algumas soluções nas quais foram identificados os erros mais frequentes, para que os professores expressassem suas opiniões sobre as possíveis causas destes erros. Da análise das respostas dadas pelos professores, emergiram três categorias relacionadas às causas dos erros cometidos pelos alunos, expostas no quadro 8: não compreensão do conteúdo, falta de pré requisitos e falta de empenho do aluno.

Em relação à primeira categoria, foram evidenciados quatro aspectos relevantes, segundo as opiniões dos professores, sobre as prováveis causas dos erros identificados. Sobre o primeiro, dos seis professores que responderam ao questionário, três apontaram que o aluno erra porque não compreendeu a estrutura geral dos testes de hipóteses. O professor A coloca que “a causa, na minha opinião, parece ser uma falta de compreensão do todo, de entendimento do que está fazendo”. Já o professor E, diz que “os alunos não seguem

corretamente as etapas para a realização do teste, como  $p = 0,09$ , eles acharam que  $\pi > 0,08$ ”, referindo-se ao erro que o aluno comete quando utiliza um teste unilateral à direita ao invés de um teste unilateral à esquerda, pelo fato da proporção na amostra ter resultado em um valor superior à proporção na população. De fato, neste caso, o aluno não compreendeu que as hipóteses deveriam ser formuladas antes mesmo de saber-se o valor da estatística da amostra, com base no que o pesquisador deseja investigar.

Relacionado ao segundo aspecto, dois professores afirmaram que às vezes o aluno não compreende as hipóteses que formulou. Para o professor C, “um dos erros mais frequentes é a contradição na conclusão, decorrente do não entendimento do que representa a hipótese nula”. Este professor está se referindo ao erro que o aluno comete quando conclui o contrário do que os dados indicam. O professor D também afirma que “a confusão iniciada na formulação das hipóteses gera confusão semelhante na conclusão do teste, uma vez que o aluno não consegue entender as hipóteses que formulou”.

O terceiro aspecto abordado refere-se ao tratamento das etapas de forma independente, sem a devida conexão, citado por três professores. Para o professor B, muitos erros são decorrentes da “realização do exercício como uma série de etapas, sem perceber a concatenação entre elas”. O professor B complementa, colocando: “parece que decisão e conclusão do teste não se referem à mesma coisa, nem tem conexão com as hipóteses”.

Sobre o último aspecto, dois professores identificam que falta o entendimento dos objetivos dos testes de hipóteses. Para o professor C, “a identificação correta de qual teste estatístico deve ser utilizado e como são formuladas suas hipóteses faz parte do entendimento dos objetivos de cada ferramenta (para que serve, em que casos se aplica)”. Enquanto o professor D afirma “falta compreensão da lógica do teste”.

Conforme as opiniões expressas pelos professores, o aluno não entende ou não compreende os testes de hipóteses e por isso comete os erros. Isso permite fazer-se outro questionamento: por que, de fato, o aluno não entendeu este conteúdo? Um indício para a resposta desta pergunta pode estar associado à questão “Você apresenta exemplos e oportuniza a realização de exercícios envolvendo o cálculo do Erro do Tipo I, Erro do Tipo II, CCO?”, pois apenas um professor a respondeu positivamente. Dessa forma, pode-se colocar que os demais educadores não dão prioridade a estes conceitos fundamentais para se entender o resultado de um teste de hipóteses. Uma possível explicação para isso é que, conforme colocam Ben-Zvi e Garfield (2005), a literatura sobre a educação estatística, com os

resultados de pesquisa, muitas vezes não são conhecidos e, portanto, estes conhecimentos ainda não são utilizados pelos professores da área.

Não só conceitos específicos dos testes de hipóteses, mas também conceitos já abordados na Estatística Descritiva e na Probabilidade podem estar associados às causas dos erros apontadas pelos professores. Conforme apresentado no Capítulo 4, há muitos conceitos envolvidos nos testes de hipóteses (ver figura 2). Por exemplo, os conceitos de população e amostra e as implicações ou diferenças de trabalhar-se com dados provenientes de uma ou de outra devem ser bem explorados. Dessa forma, garante-se que o aluno, posteriormente, não tenha a ideia equivocada de que o resultado obtido ao se analisar dados amostrais é o mesmo daquele que seria obtido se toda a população fosse investigada. Segundo Ben-Zvi e Garfield (2005), é um desafio para o professor encontrar formas para ir além dos conceitos individuais, levando o aluno a uma compreensão de ideias mais gerais e de relações entre elas. Como exemplo, os autores citam que o aprendiz pode entender como calcular as medidas de posição central e de variabilidade sem, no entanto, compreendê-las e relacioná-las com ideias mais gerais, como a distribuição dos dados.

A segunda categoria é formada pelas possíveis causas dos erros, relacionadas a pré-requisitos necessários para resolverem-se os problemas, que podem ser a falta de conhecimentos matemáticos elementares ou as dificuldades em interpretação de textos. A metade dos professores considera que os alunos têm uma base matemática deficiente. É quase unânime a opinião dos educadores em relação às dificuldades dos alunos em interpretação de textos. Dos 6 professores, apenas um não fez esta afirmação. O professor C diz que os erros, “são decorrentes da dificuldade de interpretação do enunciado ou ainda da falta de uma base matemática mais forte”. Complementando, o professor D acrescenta: “há uma grande dificuldade na compreensão do enunciado do problema. Muitos confundem ‘pelo menos’ com ‘menos que’, ou não conseguem interpretar ‘no mínimo’. No meu entender, falta base, não só matemática, mas na interpretação de textos também”. Ainda, o professor F afirma: “em geral, os alunos conseguem efetuar as contas. As maiores dificuldades encontram-se na interpretação dos enunciados, apesar de fraca base matemática ser uma realidade”.

Para os professores, os três tipos de erros mais frequentes que foram identificados na etapa da formulação das hipóteses, ou seja, o uso de um teste bilateral ao invés do unilateral e vice-versa, a inversão do enunciado das hipóteses e o uso de notação inadequada para enunciar as hipóteses, são cometidos devido à dificuldade dos alunos em interpretarem o enunciado do problema.

Grande parte das dificuldades que o aluno apresenta ao realizar os cálculos é decorrente do ensino praticado no nível fundamental e médio. A pouca ênfase que geralmente é dada à utilização de números não inteiros pelos professores que lecionam nestes níveis de ensino, não permite aos alunos que aprendam a operá-los. Na estatística, é praticamente impossível trabalhar com números inteiros, uma vez que se trabalha com o real, no qual dificilmente um dado é expresso por um número exato. O professor F diz: “muitos alunos ficam surpresos em saber que 0,4% é igual a  $0,4/100$ , o que resulta em 0,004”; este fato, portanto, mostra que os estudantes possivelmente não estão acostumados a trabalhar com números decimais.

A terceira e última categoria envolve as causas relacionadas ao empenho do aluno para conseguir resolver as questões propostas nas avaliações. Os professores mencionaram três aspectos relevantes: a falta de atenção, a falta de estudo e a falta de realização de exercícios. O professor F diz que “grande parte dos alunos não fazem os exercícios propostos”. Da mesma forma, o professor B cita, além de outras causas, também a “falta de atenção ao fazer uma prova” e a “falta de exercícios”.

A esse respeito, na investigação realizada por Turik (2010), constatou-se que os alunos apresentam mais atitudes negativas que positivas e, também, sentem-se tensos durante as aulas de estatística. Segundo a autora, “o reconhecimento quanto à aplicabilidade do conteúdo e a sua autopercepção de desempenho na disciplina são fatores que se mostram relacionados com as atitudes em relação à Estatística” (TURIK, 2010, p. 57).

## 7 CONSIDERAÇÕES FINAIS

Esse trabalho teve como objetivo investigar e analisar os erros que os alunos cometem ao resolver questões sobre testes de hipóteses, além de classificá-los e compará-los com os resultados obtidos em investigações já realizadas sobre o tema. Também, procurou-se levantar algumas hipóteses sobre suas causas. Para isso, foram analisadas as produções dos alunos ao responderem questões de avaliações elaboradas por professores das disciplinas de Probabilidade e Estatística de cursos de Engenharia. Ainda, analisou-se a opinião desses professores sobre as causas dos erros identificados.

Para resolver problemas de testes de hipóteses e interpretar corretamente seus resultados, é necessário articular diversos conceitos. Uma boa parte deles é aprendido na Estatística Descritiva e na Probabilidade. Outros são conceitos mais específicos do conteúdo, como as hipóteses nula e alternativa, o nível de significância, o valor-p, o poder do teste e a região crítica. O fato é que, para a aprendizagem significativa proposta por Ausubel, não basta o aluno ter uma boa compreensão de cada conceito isoladamente: é também necessário estabelecer e articular ligações entre eles.

Diversos erros foram detectados, sendo que alguns deles ocorreram com grande frequência e notou-se, também, que isso já foi identificado em estudos correlatos. Nesse contexto, a seguir, comentam-se os principais erros evidenciados. Iniciou-se com a categoria que apresentou a maior frequência de erros.

Na etapa da formulação da conclusão, que corresponde à sétima categoria identificada nesse estudo, foi onde os erros ocorreram com maior frequência. Os erros de conclusão representaram 30% de todos os observados. Isto é 50% a mais do que os pertencentes à segunda categoria mais frequente, a da formulação das hipóteses, com cerca de 20% do total dos erros observados.

O erro que ocorreu mais frequentemente foi o de expressar na formulação da conclusão do problema, a ideia de que o resultado encontrado é absolutamente verdadeiro. 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 presente investigação, em uma boa parte das conclusões formuladas os alunos expressaram-se da seguinte forma: “Rejeita  $H_0$ , podemos observar que realmente o nível de ácidos graxos insaturados está abaixo dos 18%”, o que demonstra que eles acreditam que a hipótese nula foi plenamente provada ser falsa, evidenciando, dessa forma, 100% de certeza no resultado.

Ainda, relacionado à sétima categoria, o segundo erro mais frequente foi o de fazer afirmações a respeito da hipótese nula quando ela não era rejeitada. O terceiro e último erro identificado nesta classe foi o de concluir o contrário do que os dados indicavam.

Como segunda categoria com maior frequência de erros destaca-se a etapa da formulação das hipóteses, que corresponde à primeira categoria desse estudo. Percebe-se que alguns estudantes não distinguem quando devem realizar um teste unilateral ou então um teste bilateral. Ainda, não utilizam a notação adequada para enunciar as hipóteses, porque não conseguem identificar o parâmetro a ser testado e também não formulam a hipótese nula com o objetivo de ser rejeitada.

Estas duas etapas, que correspondem tanto à formulação das hipóteses quanto a da conclusão do problema, que são as etapas básicas de um teste de hipóteses, é onde os alunos apresentam as maiores dificuldades e, conseqüentemente, cometem a maioria dos erros. Estas são justamente as duas etapas que exigem o raciocínio e a interpretação do aluno e que não são realizadas com o recurso de *software*, isto é, não estão automatizadas. Portanto, um dos resultados desse estudo é a sugestão de se dedicar mais atenção na abordagem dos conceitos que são fundamentais para a compreensão da lógica do processo dos testes de hipóteses, evitando dessa forma as interpretações errôneas tanto da formulação das hipóteses quanto da conclusão do teste. Além de, principalmente, enfatizar a abordagem probabilística, fazendo com que não ocorram as conclusões determinísticas observadas.

Convém enfatizar que as etapas envolvendo algoritmos, as quais podem ser realizadas automaticamente com o auxílio de *softwares*, em geral, são mais fáceis para os alunos e parecem ser consideradas as mais importantes. Assim, das 302 respostas analisadas, em apenas 2% delas os alunos deixaram de realizar os cálculos. Entretanto, em 8,6% das soluções não foram enunciadas as hipóteses e, em 20,9% delas, não foi formulada a conclusão do problema.

O uso de notação inadequada para representar parâmetros e estatísticas também foi um erro recorrente. Os erros da quinta categoria corresponderam a 17,4% do total, sendo a terceira classe com maior frequência de equívocos. Usar notação inadequada evidencia que os

estudantes, em geral, não demonstram que seja necessário fazer uma diferenciação entre valores provenientes da população e da amostra, ficando evidente novamente a necessidade de um trabalho maior dos conceitos probabilísticos envolvidos.

Os enganos mais comuns referem-se à utilização de  $\sigma$  ao invés de  $s$  e de  $\mu$  no lugar de  $\bar{x}$  para representar o desvio padrão e a média da amostra, respectivamente. Nesse contexto, Albert (1995) destaca que em um curso introdutório de estatística devem ser enfatizados os conceitos de população e seus parâmetros e de amostra e suas respectivas estatísticas, ou seja, o caráter aleatório do trabalho com os testes de hipóteses.

A quarta categoria na qual os erros ocorreram com maior frequência é a que corresponde à etapa da delimitação da região crítica (categoria 3). O percentual de ocorrência de erros foi de 10,8% e o tipo que se observou com mais frequência foi o posicionamento incorreto da região crítica pelo fato de não associá-la as hipóteses formuladas. Em segundo lugar, destacou-se a identificação incorreta do número de graus de liberdade e como consequência da leitura errada do valor crítico. Ainda, foi identificado que alguns alunos utilizam o valor crítico do teste bilateral quando estão realizando um teste unilateral. Estes erros provavelmente se devem a pouca prática de exercícios que envolvem o uso da tabela, pois não é necessário nenhum conhecimento conceitual para a realização desta etapa. No estudo de Link (2002), os alunos também cometeram equívocos ao identificar o valor crítico da estatística teste.

Na etapa do cálculo da estatística teste, que corresponde à quarta categoria, ocorreram 10,5% dos erros, destacando-se em quinto lugar. Os três tipos mais frequentes foram a determinação incorreta de valores percentuais, a utilização de valores incorretos por enganos e a utilização de valores incorretos pela não diferenciação entre parâmetro e estatística (estimativa).

Como muitos alunos apresentaram dificuldades em diferenciar parâmetro e estatística, acabaram escolhendo a distribuição de probabilidade incorreta para padronizar a estatística teste. O erro mais comum dessa categoria, que se destacou em sexto lugar com 10% dos erros, foi a utilização da distribuição normal ( $Z$ ) ao invés da  $t$  de Student ( $t_{n-1}$ ).

Na etapa em que o aluno precisa decidir se rejeita ou não a hipótese nula, que corresponde à sexta categoria, a frequência dos erros é a menor. Os erros deste grupo provavelmente se devem à desatenção do aluno, pois basta comparar o valor da estatística teste com o valor crítico lido na tabela, ao contrário da categoria em que ocorreram a maior

quantidade de equívocos, que foi na formulação da conclusão, a qual exige raciocínio e interpretação por parte do aluno.

A análise das respostas dos professores ao questionário elaborado após a classificação dos erros identificados neste estudo permitiu levantar algumas hipóteses sobre suas causas, que serão apresentadas a seguir.

Segundo as opiniões dos professores entrevistados, os erros cometidos na formulação das hipóteses ocorrem porque os alunos têm dificuldades em interpretação de textos. Por outro lado, os erros também podem estar associados ao ensino que é praticado de forma excessivamente operacional, ou seja, a uma abordagem meramente algorítmica. De fato, são dedicadas poucas horas/aula para a abordagem conceitual, e praticamente a totalidade dos professores entrevistados não exemplifica e pratica exercícios envolvendo os conceitos básicos para possibilitar a sua compreensão.

O ensino baseado em fornecer ao aluno uma espécie de “passo a passo” ou “receita de bolo” a seguir para resolver com sucesso os problemas dificilmente o permitirá uma compreensão da lógica de cada processo. No caso dos testes de hipóteses, parece que os alunos apenas seguem as etapas, sem perceberem as conexões entre elas e, muito menos, o significado do que estão calculando. Este fato foi também destacado pelos professores entrevistados e isto é um indício de que a aprendizagem significativa não está acontecendo.

Dos seis professores, quatro tratam os conceitos antes da introdução dos tipos de testes e dedicam, em média, apenas duas horas/aula para trabalhar os conceitos básicos. Do total de conteúdo abordado em uma disciplina de Estatística Básica dos cursos de Engenharia, de 60 horas aula, aproximadamente 20% do conteúdo é representado pelos testes de hipóteses. Assim, o tempo dedicado para abordar a parte conceitual dos testes de hipóteses é em torno de 15% do tempo total dispensado para o conteúdo.

Dessa forma, uma possível explicação para que tantos erros ocorram nos procedimentos de formulação das hipóteses e na conclusão é a de que ainda não ocorreu a mudança preconizada por Ben-Zvi e Garfield (2005). Os autores propõem uma nova abordagem para o ensino de estatística, que deve mudar de uma perspectiva matemática centrada em algoritmos e procedimentos, para uma que enfatize o raciocínio, o pensamento e a literacia estatística. Essa mudança, segundo eles, significa a transformação do ensino de estatística com base na probabilidade (matemática), 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.

No entanto, não é de se esperar que tão somente uma abordagem diferenciada com a utilização de *softwares* estatísticos e planilhas venha a resolver os problemas no ensino e na aprendizagem dos testes de hipóteses. Delmas *et al.* (2007) afirmam que há evidências de que o ensino pode ter efeito positivo sobre a compreensão dos conceitos, mas muitos alunos podem ter dificuldades até mesmo se o professor empregar metodologias inovadoras e inserir recursos tecnológicos nas aulas.

Muitos erros identificados neste estudo evidenciam que também faltam pré-requisitos, tanto sobre Estatística Descritiva quanto sobre Probabilidade, para a diferenciação entre parâmetro e estatística e, assim, entender suas distribuições. Outra hipótese levantada pelos professores entrevistados é a de que falta atenção, por parte dos alunos, ao lerem o texto, pois conforme um deles coloca, “frequentemente o aluno lê ‘desvio padrão’ e conclui que é o da população, não verificando que o conjunto do exercício é amostral”.

Outra causa, que também foi levantada pelos professores entrevistados, é a da falta de estudo e de realização de exercícios. De fato, se o aluno não tem interesse e não se empenha em aprender o que lhe é proposto, dificilmente a aprendizagem será efetiva, independentemente de como o conteúdo é ensinado. Por outro lado, o estudante também necessita perceber a aplicabilidade do conteúdo, pois, normalmente, quando isso não acontece, surge o desinteresse.

Neste contexto, as dificuldades dos alunos, que resultam nos erros detectados nesta investigação, podem estar relacionadas também ao empenho dos mesmos em aprenderem o conteúdo. Por outro lado, o tempo disponível para trabalharem-se os conteúdos de Estatística Básica em uma disciplina é de 4 horas/aulas semanais ou, em alguns casos, apenas 2 horas/aulas semanais, pode ser um empecilho para o professor não abordar com maior ênfase alguns conceitos. Os estudantes teriam uma melhor compreensão de todo o processo se lhes fossem oportunizadas atividades que envolvessem os conceitos. Dentre eles os erros do Tipo I (nível de significância), do Tipo II (poder de um teste), a região crítica e o valor-p.

Para uma aprendizagem significativa, o aluno deve compreender esses conceitos mais específicos dos testes de hipóteses e estabelecer conexões com as ideias mais gerais, que já foram trabalhadas tanto em Estatística Descritiva como em Probabilidade. Um conhecimento anterior essencial para a compreensão dos testes de hipóteses é o da distribuição de probabilidade do estimador envolvido, que é denominada de distribuição amostral. No entanto, se os conhecimentos prévios necessários para a ancoragem do novo

conteúdo não estiverem presentes, dificilmente o aluno consegue dar um significado aos novos conceitos.

Uma possibilidade para a compreensão dos conceitos, como o nível de significância, o valor-p, o poder do teste e a região crítica, é realizar experimentos (ou simulações) envolvendo lançamentos de moedas ou dados. Essas situações concretas, nas quais é dada a oportunidade ao aluno de praticar e coletar dados para posteriores análises aproxima-o da realidade e pode facilitar o entendimento dos conceitos. Por meio do experimento, é possível mostrar os erros do Tipo I e do Tipo II e fazer as relações com o nível de significância e o poder do teste.

Salienta-se, ainda, que uma das constatações deste estudo é que na literatura são relatados diversos equívocos conceituais que dificultam a utilização correta dos testes de hipóteses e a interpretação adequada dos resultados. Contudo, nesta pesquisa evidenciou-se que os alunos também cometem erros operacionais, que podem levar a conclusões inadequadas. Estes erros merecem também ser destacados, pois muitos deles, possivelmente, estejam relacionados a dificuldades dos alunos em matemática elementar ou à interpretação de textos (leitura deficiente).

Sabe-se que, na escola básica, muitos alunos apresentam resistência e, também, má vontade em resolver questões que envolvem números não inteiros; conseqüentemente, apresentam dificuldades em realizar essas operações. Esta opinião é compartilhada por colegas que ministram aulas de Matemática ou outros componentes curriculares nos quais são realizados cálculos.

Por diversas vezes, alunos têm questionado, ao encontrarem um resultado não inteiro na resolução de um problema, se tal valor estaria correto. Parece que um valor não exato representa um resultado incorreto. Essas concepções são provavelmente reforçadas com a prática de um ensino baseado em atividades que não envolvem dados reais. Realizar experimentos nos quais são desempenhadas medições é uma atividade que pode desmistificar a ideia de que não é comum o resultado de um problema ser um valor não inteiro.

Muitas dessas ideias e concepções dos alunos persistem e isso pôde ser evidenciado por meio de alguns erros identificados nessa investigação. As dificuldades em cálculos percentuais e a localização de números na reta real são alguns exemplos.

Apesar das pesquisas em Educação Estatística terem iniciado recentemente, a partir da década de 1970 (CAZORLA, KATAOKA e SILVA, 2010), já existe um número considerável de publicações. Dentre as questões que são abordadas destaca-se a reforma no

ensino da Estatística. Esta prevê que seja dada mais ênfase às questões conceituais, à participação ativa do aluno, ao tratamento de dados reais e à utilização da tecnologia. Ainda, aponta-se a necessidade de fazer com que o aluno entenda a natureza dos problemas estatísticos.

Mesmo que a Estatística, no ensino básico, seja um tópico que deveria ser abordado no componente curricular de Matemática e no ensino superior, seja uma disciplina que utiliza conceitos e procedimentos matemáticos, a natureza dos seus problemas é distinta. Apesar desta ligação, os problemas em Estatística vão além da aplicação dos conhecimentos matemáticos. Ao contrário de um problema matemático, um problema estatístico pode ter várias soluções, sendo que estas podem estar sujeitas a erros, não sendo possível afirmar com certeza que elas são corretas ou incorretas.

Esta questão parece ser pouco ou não discutida com os alunos. Assim, eles apenas seguem os passos previamente estabelecidos pelo professor: aplicando o algoritmo chegam a um resultado que acreditam ser o verdadeiro. Não atentam ao fato de que estão trabalhando com dados provenientes de amostras e generalizam o resultado para a população, sem mencionarem a probabilidade de se estar cometendo um erro.

Para finalizar, este trabalho possibilitou conhecer quais são os principais erros cometidos pelos alunos e, também, levantar algumas hipóteses sobre suas causas. Porém, por envolver um número relativamente expressivo de alunos, não foi possível identificar-se o perfil dos alunos, como também entrevistá-los, para que se pudessem identificar as causas de cada tipo de erro detectado.

A partir das constatações deste estudo, deixa-se como sugestão para futuros trabalhos investigar, dentre as hipóteses levantadas nesta pesquisa, quais as causas dos erros classificados, por meio de um estudo comparativo que adote diferentes estratégias para abordar-se o conteúdo em diversas turmas. Da mesma forma, sugere-se realizar um estudo qualitativo, procurando identificar e compreender as concepções dos professores em relação aos conceitos envolvidos nos testes de hipóteses. Ainda, julga-se interessante que estas pesquisas envolvam outros cursos de graduação, pois esse ficou limitado aos erros cometidos na resolução de exercícios de alguns testes de hipóteses por alunos de Engenharia.

## REFERÊNCIAS

ALBERT, Jim. Teaching inference about proportions using Bayes and discrete models. **Journal of Statistics Education**. v. 3. n. 3, 1995.

ALVES-MAZZOTTI, Alda Judith; GEWANDSZNAJDER, Fernando. **O Método nas Ciências Naturais e Sociais: Pesquisa Quantitativa e Qualitativa**. São Paulo: Pioneira, 2000.

ARA, Amilton Braio. **O Ensino de Estatística e a busca do Equilíbrio entre os aspectos Determinísticos e Aleatórios da Realidade**. 2006. 113 f. Tese (Doutorado em Educação) – Faculdade de Educação, Universidade de São Paulo, São Paulo, 2006.

AUSUBEL, David P.; HANESIAN, Halen; NOVAK, Joseph Donald. **Psicologia Educacional**. 2 ed. Rio de Janeiro: Interamericana, 1980.

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

BATANERO, Carmen. Controversies around the Role of Statistical Tests in Experimental Research. **Mathematical Thinking and Learning**. v. 2, n. 1-2, p. 75-98, 2000.

\_\_\_\_\_. **Didáctica de la Estadística**. Granada: GEEUG, Departamento de Didáctica de la Matemática, Universidad de Granada, Espanha, 2001.

BEN-ZVI, Dani, GARFIELD, Joan. Research on Statistical Literacy, Reasoning, and Thinking: Issues, Challenges, and Implications. In: **The Challenge of Developing Statistical Literacy, Reasoning and Thinking**. Netherlands: Springer. p. 397-409, 2005.

BUSSAB, Wilton de O.; MORETTIN, Pedro A. **Estatística Básica**. São Paulo: Saraiva, 2005.

CAZORLA, Irene Mauricio; KATAOKA, Verônica Yumi; SILVA, Cláudia Borin da. Trajetória e perspectivas da Educação Estatística no Brasil: um Olhar a partir do GT12. In: LOPES, Celi Espasandin; COUTINHO, Cileda de Queiroz e Silva; ALMOULOUD, Saddo Ag. **Estudos e Reflexões em Educação Estatística**. Campinas: Mercado de Letras, 2010. p. 19-44.

COLL, C.; et al. **O Construtivismo na Sala de Aula**. 5. ed. São Paulo: Ática, 1998.

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

DANIEL, Larry G.. Statistical Significance Testing: A Historical Overview of Misuse and Misinterpretation with Implications for the Editorial Policies of Educational Journals. **Research in the Schools**. v. 5, n. 2, p. 23-32, 1998.

DELMAS, Robert *et al.* Assessing Students' Conceptual Understanding After a First Course in Statistics. **Statistics Education Research Journal**. v. 6, n. 2, p. 28-58, 2007.

FISHER, Sir Ronald A. **The Desing of Experiments**. New York: Hafner Press, 1931.

GLINER, Jeffrey A.; LEECH, Nancy L.; MORGAN, George A.. Problems With Null Hypothesis Significance Testing (NHST): What do the Textbooks Say? **The Journal of Experimental Education**. v. 71, n. 1. P. 83-92, 2002.

HALLER, Heiko; KRAUSS, Stefan. Misinterpretations of significance: a problem students share with their teachers? **Methods of Psychological Research**. v. 7, n. 1, p. 1-20, 2002.

KIRK, Roger E.. Promoting Good Statistical Practices: Some Suggestions. **Educational and Psychological Measurement**. v. 61, n. 2, p. 213-218, 2001.

LINK, Conway W. An Examination of Student Mistakes in Setting Up Hypothesis Testing Problems. **Proceedings of the Louisiana-Mississippi Section of the Mathematical Association of America**. Spring 2002.

LOPES, José Marcos. **Conceitos Básicos de Testes de Hipóteses através de Aulas Investigativas**. Encontro Nacional de Educação Matemática, IX, 2007, Belo Horizonte. Disponível em:  
[http://www.mat.feis.unesp.br/docentes2008/jose\\_marcos/ComunicacaoCientfica.pdf](http://www.mat.feis.unesp.br/docentes2008/jose_marcos/ComunicacaoCientfica.pdf)  
 Acesso em: 27 set. 2009.

LÜDKE, Menga; ANDRÉ, Marli E. D. A.. **Pesquisa em Educação**: abordagens qualitativas. São Paulo: EPU, 1986.

PAMPLONA, Admur Severino; CARVALHO, Dione Lucchesi de. **A Lógica da Inferência Estatística e seu Ensino na Licenciatura em Estatística**. Encontro Nacional de Educação Matemática, IX, 2007, Belo Horizonte. Disponível em:

[http://www.sbem.com.br/files/ix\\_enem/comunicacao\\_cientifica/trabalhos/cc54855837687t.do](http://www.sbem.com.br/files/ix_enem/comunicacao_cientifica/trabalhos/cc54855837687t.do)  
 Acesso em: 26 nov. 2009.

MCLEAN, James E.; ERNEST, James M.. The Role of Statistical Significance Testing in Educational Research. **Research in the Schools**. v. 5, n. 2, p. 15-22, 1998.

MOREIRA, Marco Antônio; MASINI, Elcie F. Salzano. **Aprendizagem Significativa: A teoria de David Ausubel**. São Paulo: Moraes, 1982.

PINTO, Neuza Bertoni. **O Erro como Estratégia Didática: estudo do erro no ensino da matemática elementar**. Campinas, SP: Papirus, 2000.

PONTE, João Pedro da; BROCARD, Joana; OLIVEIRA, Hélia. **Investigações Matemáticas na sala de Aula**. Belo Horizonte: Autêntica, 2006.

RODRÍGUEZ, María Inés. Estudio Teórico y Experimental sobre Dificultades en la Comprensión del Contraste de Hipótesis en Estudiantes Universitarios. **Acta Latinoamericana de Matemática Educativa**, México, v. 19, p. 162-168, 2006.

SALSBURG, David. **Uma Senhora Toma Chá...como a estatística revolucionou a ciência no século XX**. Rio de Janeiro: Zahar, 2009.

SMITH, John P.; DISESSA, Andrea A.; ROSCHELLE, Jeremy. Misconceptions reconceived: a constructivist analysis of knowledge in transition. **The journal of the Learning Sciences**. v. 3, n. 2, p. 115-163, 1993.

SOTOS, Ana Elisa Castro et al. Student's misconceptions of statistical inference: a review of the empirical evidence from research on statistics education. **Educational research review**. n. 2, p. 98-113, 2007.

\_\_\_\_\_. How confident are students in their misconceptions about hypothesis tests? **Journal of Statistics Education**. v. 17, n.2, 2009.

TURIK, Claudia. **Análise de atitudes de Alunos Universitários em relação à Estatística por meio da Teoria de Resposta ao Item (TRI)**. 2010. 79f. Dissertação (Mestrado em Educação em Ciências e Matemática) – Faculdade de Física, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2010.

VALLECILLOS, Angustias. **Some Empirical Evidences on Learning Difficulties about Testing Hypotheses**. 52nd Session of International Statistical Institute, 1999.

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**, v. 17, n. 1, p. 29-48, 1997.

VIALI, Lori. **Aprender Fazendo**: Como Tirar Proveito do Computador para Melhorar a Aprendizagem da Estatística. ENEM (Encontro Nacional de Educação Matemática). Belo Horizonte (MG.): de 18 a 21 de julho de 2007.

\_\_\_\_\_. **Material Didático: Amostragem**. Disponível em: [www.mat.ufrgs.br/~viali/](http://www.mat.ufrgs.br/~viali/). Acesso em: 15 de Março de 2008.

\_\_\_\_\_. **Material Didático: Testes de Hipóteses**. Disponível em: [www.pucrs.br/famat/viali/](http://www.pucrs.br/famat/viali/). Acesso em: 27 de Jan. 2010.

## ANEXO A – ENUNCIADOS DAS QUESTÕES ANALISADAS

- 1) Uma marca particular de margarina diet foi analisada para determinar o nível (em percentagem) de ácidos graxos insaturados. Uma amostra de 6 potes de 500g foi investigada, resultando o seguinte: média de 17%, com desvio padrão de 0,4%. Se o nível de ácidos graxos insaturados deve ser inferior a 18%, verifique se isto está ocorrendo, com  $\alpha = 1\%$ .
  
- 2) O Ford Taurus é classificado como tendo uma eficiência média de combustível em rodovia de 30 quilômetros por litro (1995 Motor Trend New Car Buyer's Guide). Um grupo de consumidores realizou testes de quilometragem de automóveis, procurando por evidências estatísticas para mostrar que os fabricantes de automóveis exageram as avaliações de quilômetros por litro para determinados modelos. Em uma amostra de 25 testes de quilometragem com o Ford Taurus, o grupo de interesse de consumidores encontrou uma avaliação as amostras de quilometragem média em rodovia de 25,5 quilômetros por litro e um desvio padrão de 1,8 quilômetros por litro. Que conclusão deve ser tirada a partir dos resultados obtidos? Use um nível de significância de 0,01.
  
- 3) Um fabricante de detergentes afirma que a média aritmética do peso de uma determinada caixa de detergente é 1,50 kg. Uma amostra aleatória de 64 caixas revelou uma média da amostra de 1,47 kg e um desvio padrão da amostra de 0,05 kg. Utilizando o nível de significância de 0,01 há evidências de que o peso médio das caixas seja diferente de 1,50 kg?
  
- 4) Uma revista resolve testar a afirmação de uma fábrica de automóveis que seus veículos fazem 12 km por litro de combustível. A revista analisa 25 automóveis, obtendo nessa amostra uma média de 11,5 km por litro e um desvio padrão amostral de 1,2 km por litro. Apresente suas conclusões frente a este teste de hipóteses, usando 5% de significância.
  
- 5) Uma amostra de 20 unidades de um produto foi pesquisada com o objetivo de verificar se o diâmetro médio é superior a 10 mm. Encontrou-se uma média de 10,5 mm, com desvio padrão de 0,3 mm. O que se pode concluir a 5% de significância?
  
- 6) As especificações de uma peça componente de um rádio exigem que seu peso não exceda 25 mg. O fabricante está examinando a inspeção de cada lote antes de sua liberação

para decidir entre entregá-lo ao consumo ou mantê-lo para reprocessamento. Em uma amostra de 30 unidades de um lote foi observado um peso médio de 24,3 mg com um desvio padrão de 3mg. Analise os dados e conclua se o lote deve ser entregue ao consumo ou não. Use 5% de significância.

7) A associação dos proprietários de indústrias metalúrgicas está preocupada com o tempo perdido com acidentes de trabalho, cuja média, nos últimos tempos, tem sido da ordem de 60 homens/hora por ano. Tentou-se um programa de prevenção de acidentes e, após o mesmo, tomou-se uma amostra aleatória de 36 indústrias e verificou-se que o tempo perdido baixou para 50 homens/hora ano com um desvio padrão de 15 homens/hora por ano. Você diria que, ao nível de 5% de significância, o programa surtiu efeito?

8) Um industrial considera satisfatório se menos de 8% das peças produzidas por sua indústria forem defeituosas. Se uma amostra de 200 peças apresentou 18 defeituosas, pode o industrial satisfazer-se com este resultado, ao nível de significância de 5%?

9) A garantia para baterias de fones móveis é estabelecida em 200 horas operacionais, seguindo os procedimentos adequados de recarga. Um estudo com 5000 baterias foi executado e 15 pararam de operar antes de 200 horas. Este experimento confirma a suspeita de que mais de 0,2% das baterias da companhia falhará durante o período de garantia, usando os procedimentos apropriados de recarga? Use o procedimento de teste de hipóteses, com  $\alpha = 0,01$ .

10) O gerente de marketing de uma fábrica de automóveis está interessado em determinar a proporção de novos proprietários de carros compactos que teriam adquirido um air-bag inflável para o lado do passageiro se o mesmo estivesse disponível a um custo adicional de \$300. Por informações anteriores o gerente acredita que a proporção é igual a 0,30. Suponha que seja feito um levantamento de 200 novos proprietários de carros compactos e 79 indicaram que teriam comprado os air-bags infláveis. Ao nível de significância de 0,10, há evidência de que a proporção é diferente de 0,30?

11) Uma agência de viagens tem um tradicional plano de férias que é oferecido a seus clientes. O índice de respostas positivas é historicamente 20%. Este ano, uma amostra aleatória de 50 clientes mostrou que 15 adquiriram o plano. Teste ao nível de 5% de significância a hipótese de que o índice tenha aumentado.

- 12) Uma amostra aleatória de 500 donas de casa mostrou que 300 delas preferiam conservas como milho e ervilha em embalagens Tetra Pak ao invés das latas de alumínio. Isso é suficiente para afirmar que mais da metade das donas de casa do mercado são favoráveis à utilização de embalagens Tetra Pak para conservas como milho e ervilhas em substituição às latas de alumínio? Analise os dados e conclua usando 5% de significância.
- 13) De acordo com a Polícia Rodoviária 80% das vítimas de desastres automobilísticos são homens. Um pesquisador resolveu testar tal informação, para isso observou no último ano uma amostra de 140 desastres automobilísticos e verificou que destes, 110 tiveram com vítimas homens. Analise os dados e conclua ao nível de significância de 5%.
- 14) Numa pesquisa de opinião para o segundo turno de uma eleição, 180 de 400 entrevistados declaram que votariam em “Honesto Quando Com-Vén”. Apresente suas conclusões frente a este teste de hipóteses, usando 5% de significância.

## APÊNDICE A

Caro professor,

Conto com a sua colaboração para responder este questionário que faz parte da pesquisa sobre análise de erros em Estatística Inferencial, que estou desenvolvendo no Mestrado em Educação em Ciências e Matemática da PUCRS. O objetivo da investigação é verificar, classificar e analisar os erros cometidos pelos alunos, procurando identificar as causas das dificuldades que se manifestam ao resolverem questões sobre os Testes de Hipóteses. Foram selecionadas 357 respostas sobre dois tipos de testes paramétricos: o teste para a média com o desvio padrão desconhecido e o teste para a proporção, dentre as quais, foi feita a classificação dos erros encontrados em 302 delas.

O questionário está dividido em três partes: inicialmente, apresentam-se algumas questões e soluções para a análise dos erros, após seguem perguntas referentes à abordagem do conteúdo e, por último, duas questões pessoais.

Não é necessário identificar-se.

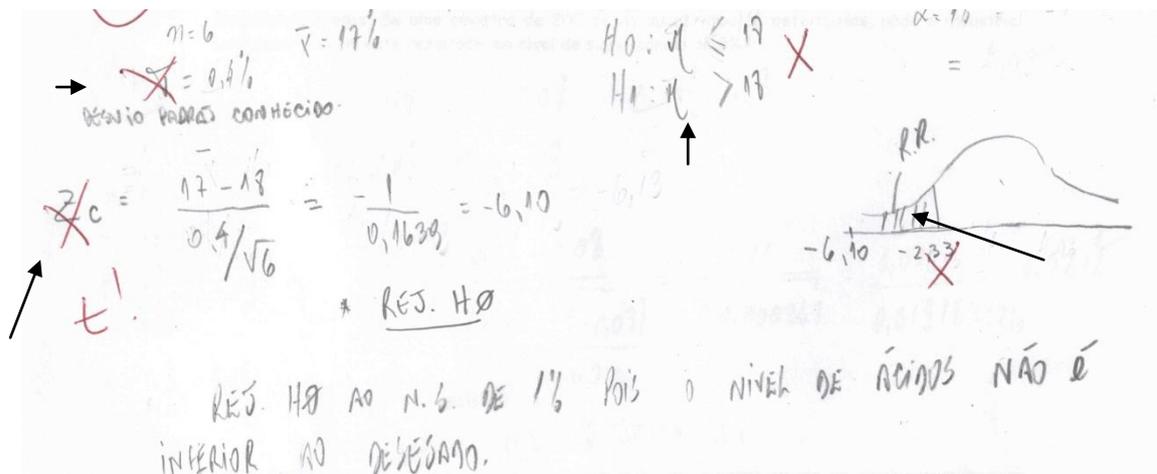
Obrigada!

Renate Grings Sebastiani

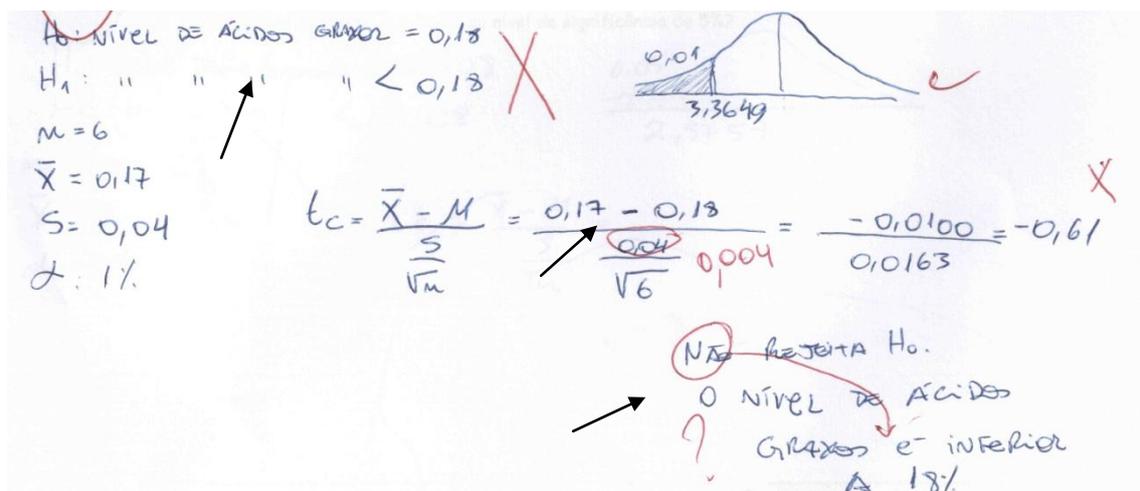
**A seguir, são apresentados dois enunciados de cada tipo de teste e, também, soluções de alunos nas quais foram identificados alguns erros. Analise estes erros e responda as questões, que seguem após cada figura, no espaço em branco abaixo de cada pergunta.**

*a) “Uma marca particular de margarina diet foi analisada para determinar o nível (em percentagem) de ácidos graxos insaturados. Uma amostra de 6 potes de 500g foi investigada, resultando o seguinte: média de 17%, com desvio padrão de 0,4%. Se o nível de ácidos graxos insaturados deve ser inferior a 18%, verifique se isto está ocorrendo, com  $\alpha = 1\%$ .”*

Solução 1:



Solução 2:



1) Os erros identificados nas soluções apresentadas acima ocorrem com frequência em suas aulas? Na sua opinião, qual é a causa destes erros?

---



---



---



---



---



---



---

b) "O Ford Taurus é classificado como tendo uma eficiência média de combustível em rodovia de 30 quilômetros por litro (1995 Motor Trend New Car Buyer's Guide). Um grupo

de consumidores realizou testes de quilometragem de automóveis, procurando por evidências estatísticas para mostrar que os fabricantes de automóveis exageram as avaliações de quilômetros por litro para determinados modelos. Em uma amostra de 25 testes de quilometragem com o Ford Taurus, o grupo de interesse de consumidores encontrou uma avaliação as amostras de quilometragem média em rodovia de 25,5 quilômetros por litro e um desvio padrão de 1,8 quilômetros por litro. Que conclusão deve ser tirada a partir dos resultados obtidos? Use um nível de significância de 0,01.”

Solução 1:

Handwritten solution for Solution 1:

- $\mu_1 = 30 \text{ km/l}$
- $n = 25$
- $\bar{x} = 25,5 \text{ km/l}$
- $\alpha = 0,01$
- $s = 1,8 \text{ km/l} \rightarrow \sigma^2 = 3,24 \text{ km}^2/\text{l}^2$
- $H_0: \mu_1 = \mu_2$
- $H_1: \mu_1 > \mu_2$
- $t = \frac{25,5 - 30}{1,8 / \sqrt{25}} = -12,5$
- $(P = 0,000 < \alpha = 0,01)$
- Rejeita-se  $H_0$ .
- Logo, os consumidores tem razão, a média é menor do que a dada pelos fabricantes.

Solução 2:

Handwritten solution for Solution 2:

- $H_0: \mu_1 = \mu_2$
- $H_1: \mu_1 > \mu_2$
- Normal distribution curve with a shaded area to the right of a critical value.
- $t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{25,5}{\frac{1,8}{\sqrt{25}}} = 70,83$
- Rejeita-se  $H_0$  ( $P < 0,01$ ), logo o Ford Taurus tem uma eficiência média em rodovia de 30 quilômetros por litro.

2) Os erros identificados nas soluções apresentadas acima ocorrem com frequência em suas aulas? Na sua opinião, qual é a causa destes erros?

---



---



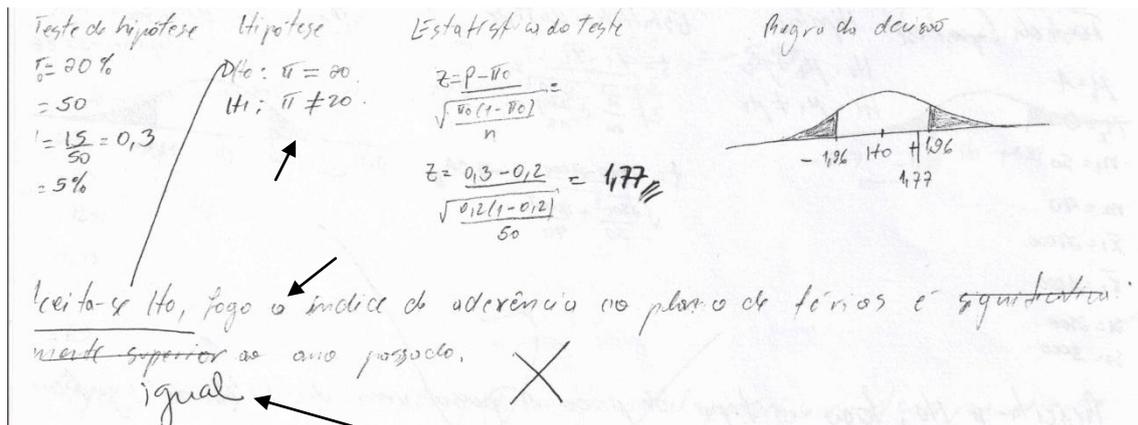
---



---

c) “Uma agência de viagens tem um tradicional plano de férias que é oferecido a seus clientes. O índice de respostas positivas é historicamente 20%. Este ano, uma amostra aleatória de 50 clientes mostrou que 15 adquiriram o plano. Teste ao nível de 5% de significância a hipótese de que o índice tenha aumentado.”

Solução 1:



3) Os erros identificados nas soluções apresentadas acima ocorrem com frequência em suas aulas? Na sua opinião, qual é a causa destes erros?

---



---



---



---



---



---



---

d) “Um industrial considera satisfatório se menos de 8% das peças produzidas por sua indústria forem defeituosas. Se uma amostra de 200 peças apresentou 18 defeituosas, pode o industrial satisfazer-se com este resultado, ao nível de significância de 5%?”

Solução 1:

$n = 200$   
 $x = 18$   
 $p = \frac{18}{200} = 0,09$   
 $\alpha = 5\% = 0,05$

$H_0: \pi = 0,08$   
 $H_1: \pi < 0,08$

$Z_c = \frac{0,08 - 0,08}{\sqrt{\frac{0,08 \cdot (1 - 0,08)}{200}}} = \frac{0,01}{\sqrt{0,000368}} = \frac{0,01}{0,019183326} = 0,5212$

NÃO REJEITA  $H_0$

Boa prova e boas férias!

RR Logo ao N.S DE 5% NÃO REJEITA  $H_0$ , PODENDO O INDIVÍDUO SATISFAZER-SE COM O RESULTADO.

Solução 2;

$\pi = 8\%$   
 $n = 200$   
 $x = 18$   
 $\alpha = 5\% = 0,05 \Rightarrow 1,64$   
 $p = 0,09$

$H_0 = \pi = 8\%$   
 $H_1 = \pi < 8\%$

$Z_c = \frac{0,08 - 0,08}{\sqrt{\frac{0,08 \cdot 0,92}{200}}}$

$Z_c = \frac{0,01}{0,0192} = 0,5202$

Fora da área de rejeição  
 Prefere  $H_0$

Logo com um nível de significância de 5% é possível satisfazer-se com o resultado de peças defeituosas obtido.

4) Os erros identificados nas soluções apresentadas acima ocorrem com frequência em suas aulas? Na sua opinião, qual é a causa destes erros?

---



---



---



---



---



---



---



---

**A seguir, seguem as perguntas referentes à abordagem do conteúdo.**

5) Qual é o tipo de abordagem que você costuma utilizar em suas aulas?

( ) Região crítica;

( ) valor-p;

( ) ou ambas.

6) Em relação aos conceitos, tais como, erros do Tipo I (nível de significância), Tipo II (poder de um teste), CCO (Curva Característica de Operação), região crítica e valor-p:

( ) são introduzidos antes do estudo dos tipos de testes;

( ) são tratados durante o estudo dos tipos de testes.

6.1) Caso estes conceitos sejam introduzidos antes da apresentação dos tipos de testes, quantas horas/aula são dedicadas para sua abordagem?

---

7) Você apresenta exemplos e oportuniza a realização de exercícios envolvendo o cálculo do Erro do Tipo I, Erro Tipo II, CCO?

( ) Sim;

( ) não.

7.1) Caso tenha respondido “Sim” na questão anterior, comente sobre os exemplos e exercícios utilizados para tratar estes conceitos.

---



---



---

**Para finalizar, gostaria de duas informações pessoais.**

8) Qual é a sua formação?

Graduação: \_\_\_\_\_

Pós-graduação: \_\_\_\_\_

9) Há quanto tempo você leciona a disciplina de Probabilidade e Estatística?

---

Muito obrigada, pela sua colaboração!