

# Desempenho de alunos de Engenharia em testes de hipóteses

Gabriela Gonçalves

Instituto Superior de Engenharia do Porto: Departamento de Matemática  
Instituto Politécnico do Porto  
Porto, Portugal  
gmc@isep.ipp.pt

José António Fernandes

Instituto da Educação  
Universidade do Minho  
Braga, Portugal  
jfernandes@ie.uminho.pt

Maria M. Nascimento

Universidade de Trás-os-Montes e Alto Douro - UTAD,  
www.utad.pt, Dep. de Matemática da ECT, Vila Real,  
Portugal, e LabDCT-UTAD/CIDTFF, UA  
mmsn@utad.pt

José Alexandre Martins

Instituto Politécnico da Guarda, Guarda, Portugal  
jasvm@ipg.pt

**Resumo** — Neste trabalho analisamos o desempenho de 223 alunos da Licenciatura de Engenharia Informática do Instituto Superior de Engenharia do Porto, no ano letivo 2012-2013, quando confrontados com várias questões sobre testes de hipóteses. Em termos de resultados, os alunos apresentaram muitas dificuldades no tema, a maior parte das quais similares às que são relatadas na literatura.

**Palavras-chave** — aprendizagem da Estatística; inferência estatística; testes de hipóteses; ensino superior

## I. INTRODUÇÃO

Nas últimas décadas, o ensino da estatística tem sido implementado, cada vez mais, nas escolas e nas universidades, não só pelo seu carácter instrumental, mas também pela importância do raciocínio estatístico na sociedade atual para lidar com a proliferação de informação e a tomada de decisões.

As tecnologias têm hoje uma utilização mais intensa nos campos da Engenharia, e o raciocínio e as metodologias da Estatística assumem-se, cada vez mais, como ferramentas de suporte para o trabalho e a investigação em engenharia. As revistas e publicações em Engenharia contêm várias informação estatística, tornando-se esta uma área imprescindível no currículo de um engenheiro, pois os seus fundamentos estão presentes noutras disciplinas e, conseqüentemente, no seu trabalho. Segundo Olivo [1], um engenheiro, na sua vida profissional, irá deparar-se com a análise de dados e a realização de inferências estatísticas.

Relativamente ao ensino dos testes de hipóteses, existe investigação a nível internacional [e.g. 2], mas em Portugal é uma área ainda pouco trabalhada em termos de investigação.

Na literatura aparecem poucos estudos nesta área, apesar de constituir um tópico relevante para a compreensão de boa parte da literatura científica e técnica em várias áreas do conhecimento, como a engenharia, as ciências, a matemática e as ciências sociais.

Por outro lado, apesar da sua relevância, é um tema em que os alunos apresentam muitas dificuldades na compreensão dos conceitos envolvidos, por exemplo, [2] [3] [4] [5].

Para Batanero [6], os testes de hipóteses, embora tenham um campo de aplicação muito amplo, eles constituem um tema pouco compreendido, envolvendo muitas confusões e mal-entendidos em estatística. Os vários conceitos implicados – hipótese nula e alternativa; erros tipo I e II; probabilidades destes erros; resultados significativos e não significativos; população e amostra; parâmetro e estimador e distribuição da população e da amostra – estão certamente na origem das interpretações erradas e confusões daqueles que recorrem aos testes de hipóteses. Ora, tais dificuldades justificam a realização de mais investigação nesta temática.

## II. INVESTIGAÇÕES PRÉVIAS SOBRE TESTES DE HIPÓTESES

As dificuldades que os alunos demonstram na compreensão dos testes de hipóteses têm sido objeto de diversos trabalhos de investigação. De entre as pesquisas realizadas sobre dificuldades e erros na compreensão dos testes de hipóteses, destacamos seguidamente algumas.

No seu estudo, Sotos, Vanhoof, Noortgate e Onghena [7] investigaram 144 alunos universitários de cursos introdutórios de estatística através de um questionário. O questionário teve como objetivo estudar três aspetos fundamentais que são de difícil compreensão num teste de hipóteses: definição de um teste de hipóteses; interpretação do valor de prova e interpretação do nível de significância. Os investigadores, além de quererem detetar os erros cometidos pelos alunos, também estudaram a confiança que tinham nas suas respostas erradas. Analisando os erros que os alunos cometeram, os investigadores concluíram que eles acreditavam que um teste de hipóteses é uma prova matemática da hipótese nula, ou que é uma prova probabilística por contradição. O erro mais comum em relação ao valor de prova foi considerá-lo como sendo a probabilidade de cometer um erro ao rejeitar a hipótese

nula. Os erros detetados com maior frequência estão relacionados com as seguintes afirmações: o resultado do teste foi estatisticamente significativo para um nível de significância de 5%; a probabilidade de rejeitar a hipótese nula é igual a 95%; e a probabilidade da hipótese nula ser verdadeira é igual a 5%.

Já Batanero, Vera e Díaz [8] realizaram um estudo com 224 alunos do 2.º ano da Licenciatura de Psicologia da Universidade de Huelva, que frequentavam a unidade curricular de Análise de Dados II, com o objetivo de avaliar as dificuldades dos alunos na compreensão dos testes de hipóteses. Mais especificamente, com base nas respostas dos alunos a um questionário curto (seis itens sobre testes de hipóteses), avaliaram a sua compreensão sobre diferentes conceitos relativos aos testes de hipóteses: diferença entre teste unilateral e bilateral; hipótese nula e alternativa; tipos de erros e suas probabilidades e tomada de decisão. A percentagem de alunos que deram respostas corretas em todos os itens foi superior a 50%, em que 84,8% souberam enunciar corretamente as hipóteses nula e alternativa. Relativamente ao erro tipo II e potência do teste, 50,9% dos alunos responderam corretamente, enquanto 64,7% discriminaram entre os erros tipo I e II. Quanto à relação entre o nível de significância e a região crítica, a percentagem de respostas corretas foi de 64,3% e na tomada de decisão foi de 58%. Face aos resultados obtidos, os investigadores reafirmam a recomendação de Harradine, Batanero e Rossman [9], de que o raciocínio inferencial não pode ser desenvolvido num curto espaço de tempo e que seria importante começar a introduzi-lo de forma informal desde o ensino secundário.

No estudo que Khrishnan e Idris [10] realizaram, com 150 alunos de duas universidades da região central da Malásia, pretendeu-se identificar os erros que os alunos cometem na realização dos testes de hipóteses. Para este estudo foram retirados sete itens abertos de um questionário original de 21 itens para a identificação dos erros. Esses itens consistiam em perguntas sobre a finalidade das hipóteses nula e alternativa, região crítica, tomada de decisão e valor de prova. Depois de analisados os itens, os investigadores identificaram quatro erros que os alunos cometeram acerca do propósito dos testes de hipóteses: o teste de hipóteses é realizado para estabelecer a estatística da amostra; o teste de hipóteses é realizado para decidir se a hipótese nula é verdadeira ou falsa; a hipótese nula deve ser rejeitada e a hipótese nula pode ser aceite. Nos itens relativos à inferência num dado contexto (eram três), em dois dos itens os alunos deram respostas insuficientes e no outro 43,3% deram a resposta correta. Nos itens relativos à comunicação e compreensão (eram dois), os alunos em geral responderam corretamente. Para os dois itens que avaliam a tomada de decisão, num deles 40,7% dos alunos deu uma resposta incompleta e no outro 47,3% dos alunos deu uma resposta estatisticamente incorreta (aceitar a hipótese nula). Perante estes dados, os autores concluíram que o principal resultado deste estudo é o de que os erros cometidos por estes alunos foram semelhantes aos identificados noutros estudos.

### III. METODOLOGIA

Neste texto estuda-se o desempenho de um grupo de alunos do curso de Engenharia Informática em conteúdos de testes de

hipóteses. Para tal, os alunos responderam a um questionário com 10 questões de escolha múltipla e dois problemas, sendo aqui tratadas apenas as 10 questões de escolha múltipla, as quais avaliam as seguintes dimensões: 1) Formulação das hipóteses/ Identificação do teste de hipóteses; 2) Interpretação dos erros tipo I e II/Interpretação do valor de prova. Na Figura 1 apresentamos as 10 questões de escolha múltipla propostas aos alunos.

1. Qual das seguintes hipóteses não é uma hipótese nula?  
 a)  $\mu_x = 10$ ; b)  $\sigma = 3$ ; c)  $\bar{x} = 35$ ; d)  $\mu_1 = \mu_2$ .

2. O candidato A, a Presidente de Junta de Freguesia afirma que vai ser eleito com 60% dos votos. O outro candidato B, concorrente de A, deseja contestar esta afirmação, e decidiu, para isso, efetuar uma sondagem a 150 eleitores onde obteve 105 votos favoráveis à sua candidatura. Qual das seguintes hipóteses elegeria como hipótese nula?  
 a)  $p_A = 0,7$  ( $105/150 = 0,7$ ); b)  $p_A = 0,6$ ; c)  $p_A < 0,6$ ; d)  $p_A > p_B$ .

3. Num teste de hipóteses, quando rejeitamos  $H_0$ , sendo  $H_1$  falsa,  
 a) Comete-se um Erro tipo I; b) Comete-se um Erro tipo II;  
 c) Comete-se um Erro tipo I e um erro tipo II; d) Tomou-se a decisão correta.

4. A estação de rádio FMP quer estimar o tempo médio que uma família dedica diariamente a ouvir esta estação. Para tal, selecionou-se uma amostra aleatória de 81 famílias, tendo sido calculadas uma média diária de audição de 2,4 horas com um desvio-padrão de 0,7 horas. Suponha que o diretor comercial da FMP tinha decidido o fecho da estação se a média diária de audiências fosse inferior a 2,5 horas. Que teste usaria para responder a esta questão?  
 a) Teste Bilateral para a média da população;  
 b) Teste Unilateral à esquerda para a diferença das médias das populações;  
 c) Teste Unilateral à direita para a média da população;  
 d) Teste Unilateral à esquerda para a média da população.

5. Afirmer que o nível de significância  $\alpha = 5\%$ , significa que:  
 a) Se a hipótese nula é verdadeira, a probabilidade de a rejeitar é igual a 0,05;  
 b) Se a hipótese nula é falsa, a probabilidade de a aceitar é 0,05;  
 c) Há uma probabilidade de 0,05 do resultado ser correto;  
 d) Se a hipótese nula for rejeitada, a probabilidade de erro é de 5%.

6. Consideremos os Testes 1 e 2 e as respetivas hipóteses apresentadas no quadro seguinte:

Teste 1	Teste 2
$H_0: \mu = 75$	$H_0: \mu = 75$
$H_1: \mu > 75$	$H_1: \mu < 75$

Numa grande amostra, selecionada aleatoriamente para uma variável que na população é normalmente distribuída e da qual se conhece a variância, sabe-se que  $Z_0 = 1.89$  e  $Z_c = 1.645$  ao nível de significância de 5%. Nestas condições podemos:

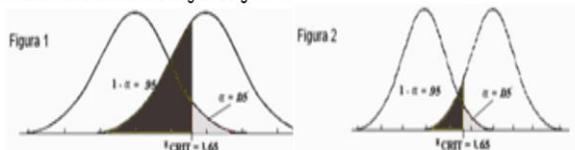
a) Não rejeitar  $H_0: \mu = 75$  no Teste 1, ou seja, o Teste 1 não é significativo;  
 b) Não rejeitar  $H_1: \mu > 75$  no Teste 1, ou seja, o Teste 1 é significativo;  
 c) Não rejeitar  $H_0: \mu = 75$  no Teste 2, ou seja, o Teste 2 não é significativo;  
 d) Não rejeitar  $H_1: \mu < 75$  no Teste 2, ou seja, o Teste 2 é significativo.

7. Suponha que leu a seguinte afirmação numa publicação científica: "Verificou-se uma diferença significativa (valor de prova  $p < 0,5$ ) entre as despesas médias dos alunos do ISEP e os da FEUP durante a semana académica". Este resultado significa que:  
 a) Os alunos do ISEP tendem, em média, a gastar mais do que os alunos da FEUP;  
 b) Os alunos do ISEP tendem, em média, a gastar menos do que os alunos da FEUP;  
 c) Com uma confiança de 95%, é possível afirmar que existe uma diferença entre os gastos médios;  
 d) Caso a despesa média dos alunos do ISEP seja igual à dos alunos da FEUP, então existe uma probabilidade menor ou igual a 5% de se obter uma diferença apenas devida ao acaso.

8. Se a hipótese  $H_0$  não for rejeitada a um nível de significância de 5%, o que podemos dizer acerca da probabilidade do Erro tipo II?  
 a) É igual a 95%; b) É menor que 5%; c) É igual a 0,05; d) Não pode ser determinada por falta de informação.

9. Considere as duas afirmações seguintes:  
 I — O nível de significância de um teste de hipóteses é a probabilidade de rejeitar a hipótese nula quando, na realidade, ela é falsa;  
 II — Comete-se um erro tipo II quando se aceita a hipótese nula quando, na realidade, ela é falsa.  
 a) Apenas a afirmação I é verdadeira; b) Apenas a afirmação II é verdadeira;  
 c) Ambas as afirmações I e II são verdadeiras; d) Ambas as afirmações I e II são falsas.

10. Observe atentamente as duas figuras seguintes:



Da figura 1 para a figura verifica-se uma diminuição dos erros de tipo I e tipo II. Isso deve-se a:  
 a) Um maior nível de confiança; b) Uma diminuição da potência do teste;  
 c) Um menor desvio-padrão; d) Uma menor média amostral.

Figura 1. Questões de escolha múltipla propostas aos alunos (as respostas assinaladas a vermelho são as corretas)

O questionário foi aplicado aos alunos do 1.º ano que frequentavam a disciplina de Matemática Computacional (MATCP), no ano letivo 2012-2013, do curso de Engenharia Informática do Instituto Superior de Engenharia do Porto. O questionário foi facultativo e não foi considerado na avaliação da unidade curricular. Dos 263 alunos que frequentavam a disciplina, responderam ao questionário 223 nas suas aulas teórico-práticas da disciplina de MATCP, na presença dos respetivos docentes, e os alunos dispuseram de 90 minutos para responder, revelando-se um tempo suficiente.

Os alunos responderam ao questionário na última aula do semestre (junho de 2013), por escrito e sem consulta, imediatamente depois da lecionação dos testes de hipóteses nas aulas. O ensino do tema realizou-se ao longo de uma aula teórica e duas aulas teórico-práticas, cada uma com a duração de duas horas, onde os alunos acompanharam o professor e tiveram a oportunidade de resolver exercícios e problemas para consolidação dos conceitos usando papel e lápis e, esporadicamente, o software R.

Na secção seguinte apresentam-se os diferentes tipos de respostas dadas pelos alunos nas diferentes questões.

#### IV. RESULTADOS

As respostas dadas pelos alunos foram categorizadas em corretas e erradas, sendo também considerada uma categoria para a ausência de resposta (não respostas). Na Tabela 1, apresentam-se, para cada questão, as frequências de respostas de cada categoria.

TABELA 1 – FREQUÊNCIAS DE RESPOSTAS CORRETAS E ERRADAS E DE NÃO RESPOSTAS EM CADA QUESTÃO

Questão	Corretas		Erradas		Não Respostas	
	N	%	N	%	N	%
1	25	11,2	186	83,4	12	5,4
2	110	49,3	109	48,9	4	1,8
3	166	74,4	51	22,9	6	2,7
4	146	65,5	73	32,7	4	1,8
5	84	37,7	128	57,4	11	4,9
6	73	32,7	123	55,2	27	12,1
7	64	28,7	146	65,5	13	5,8
8	85	38,1	124	55,6	14	6,3
9	147	65,9	65	29,1	11	4,9
10	84	37,7	119	53,4	20	9,0

Os resultados revelam a existência de uma percentagem baixa de respostas corretas na questão 1 (11,2%), enquanto na questão 2, de tipo semelhante, os resultados foram melhores (49,3%), podendo ser uma explicação para os maus resultados da questão 1 o facto de ela estar formulada numa forma diferente daquela a que alunos estão habituados. Estas questões avaliavam o estabelecimento/formulação das hipóteses.

Observamos que a percentagem de respostas corretas na questão 5 (37,7%), que se refere à interpretação do nível de significância, foi reduzida, sendo que uma possível explicação para o facto se prende com a confusão entre a  $P(\text{rejeitar } H_0 | H_0 \text{ verdadeira})$  e

$P(H_0 \text{ verdadeira} | \text{rejeitar } H_0)$ . Este erro, em geral, designado por falácia da condicional transposta, significa que o sujeito não discrimina entre a probabilidade condicional e a sua transposta.

A questão 6 recolheu apenas 32,7% de respostas corretas, indicando que a lógica e o processo de aplicação de um teste de hipóteses, que era o objetivo da questão, é um tema complexo para os alunos, levando-os a confundirem região de aceitação com região de rejeição.

A questão 7 apresenta uma percentagem de respostas corretas ainda mais baixa (28,7%), situação que pode ter ocorrido pelo facto de nas aulas teórico práticas o conceito de valor de prova não ter sido muito explorado. Já nas questões 3 e 9, relacionadas com os erros tipo I e tipo II, observaram-se percentagens elevadas de respostas corretas (74,4%, 65,9%, respetivamente), o que leva a concluir que os alunos conseguiram aplicar os conceitos em causa.

Na questão 8 podemos constatar uma percentagem baixa de respostas corretas, 38,1%, mostrando que os alunos não distinguiram claramente os conceitos de erro tipo I e II, ou entenderam-nos como acontecimentos complementares.

Finalmente, na questão 10 pretendia-se avaliar a definição e interpretação dos erros tipo I e tipo II. Esta questão revelou-se também de um nível de dificuldade bastante elevado para os alunos, uma vez que envolve a interpretação de figuras e o conhecimento e compreensão simultânea de muitos conceitos (erros tipo I e II, nível de confiança, desvio padrão, média amostral e potência do teste), necessários para a interpretação das figuras. Assim, essa exigência refletiu-se na baixa percentagem de alunos (37,7%) que respondeu corretamente à questão.

De acordo com as duas dimensões, antes estabelecidas, averiguamos se existe correlação entre as diversas questões do questionário, as relacionadas com formulação e interpretação (questões 1, 2, 4 e 6), que designámos por dimensão “formulação”, e as relacionadas com erros e nível de significância (questões 3, 5, 7, 8, 9 e 10), que designámos por dimensão “erro”.

Para analisar estas dimensões, para cada aluno, dividimos o número de respostas corretas pelo número de perguntas em cada dimensão, obtendo uma nova variável cujo valor se pode considerar a “taxa de respostas corretas para cada dimensão”. Foram calculados o valor médio e o desvio padrão desta nova variável e os resultados indicam que a dimensão “formulação e interpretação” tem menor valor médio (40%) e aproximadamente igual desvio padrão (22%) do que a dimensão “erro e n”, cujo valor médio é 47% e o desvio padrão 21%. Isto é, as questões relacionadas com problemas de formulação e interpretação, revelam-se mais difíceis, para os alunos do que as questões relacionadas com erro e nível de significância. A Tabela 2 mostra os resultados obtidos.

TABELA 2 – VALOR MÉDIO E DESVIO PADRÃO PARA O NÚMERO DE RESPOSTAS CORRETAS EM CADA DIMENSÃO

Dimensões	Média (%)	Desvio padrão (%)
Formulação e interpretação	40%	22%
Erro e nível de significância	47%	21%

Recorremos ao teste t-Student para amostras emparelhadas (para um nível de significância de 5%), com o intuito de averiguar se seria significativa a diferença observada em cada dimensão, na taxa de respostas corretas. Para tal, foi efetuado um teste bilateral, em que na hipótese nula se estabelece a igualdade das médias e, na hipótese alternativa, a diferença das médias,

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

O resultado do teste mostra que temos condições para rejeitar a hipótese nula ( $p = 0,000$ ), sendo de considerar que a taxa de respostas corretas das questões relacionadas com “formulação e interpretação” é significativamente diferente – em termos estatísticos – do que o relacionado com “erro e nível de significância”.

Deste modo, constatou-se, que os alunos apresentaram maior dificuldade na formulação e interpretação em problemas de testes de hipóteses, sendo necessário um maior investimento nos aspetos pedagógicos relacionados com esta vertente do tema.

## V. SÍNTESE

Os alunos mostraram ter muita dificuldade nos conteúdos do tema de testes de hipóteses, principalmente na interpretação das questões e na explicitação da forma como pensaram. Para além da dificuldade inerente ao tema, o facto de ter sido o último a ser lecionado repercutiu-se no pouco tempo de exploração nas aulas e no pouco tempo de estudo dos alunos.

Resumindo, estes alunos apresentaram muitas dificuldades na forma como raciocinaram em termos estatísticos. Em geral, podemos concluir que os erros cometidos por estes alunos foram semelhantes aos identificados em vários estudos, por exemplo, [7] [8] [10].

Perante esta situação, o trabalho aqui apresentado aponta a necessidade de se rever a forma de ensinar este tema. Para além da necessidade de motivar os alunos para o tema, é imprescindível ajudá-los. Tal passará por um ensino que implemente a exploração de dados, encoraje o uso de dados reais e a realização de tarefas em grupo como forma de melhorar as suas habilidades comunicativas através de

discussões estatísticas, como propõem Ben-Zvi e Garfield [11], utilizando uma metodologia trabalho de projeto com temas atuais e de interesse dos alunos.

Além das sugestões referidas, as dificuldades reveladas neste estudo podem ser usadas para que, no futuro, de delinear um ensino diferente nos testes de hipóteses, para além de motivar outros investigadores para a realização de outras investigações nesta temática.

## REFERÊNCIAS

- [1] E. Olivo, “Significado de los intervalos de confianza para los estudiantes de ingeniería en México”. Tese de doutoramento: Universidad de Granada: Granada, 2008.
- [2] I. Rodríguez, “Estudio teórico y experimental sobre dificultades en la comprensión del contraste de hipótesis en estudiantes universitarios”, *Acta Latinoamericana de Matemática Educativa*, pp.162–168, 2006.
- [3] A. Vallecillos, “Inferencia estadística y enseñanza: un análisis didáctico del contraste de hipótesis estadísticas”. Recife: Comares, 1996.
- [4] A. Vallecillos, C. Batanero e J. D. Godino, “Student’s understanding of the significance level on statistical tests”, In W. Geelling and K. Graham, Eds., *Proceedings of the XVII Conference on the Psychology of Mathematics Education*, vol.4, Universidad de Valencia: Spain, pp. 271-378, 1992.
- [5] O. Vera, C. Díaz e C. Batanero, “Dificultades en la formulación de hipótesis estadísticas por estudiantes de Psicología”, *UNIÓN*, vol. 27, pp. 41-61, 2011.
- [6] C. Batanero, “Didáctica de la Estadística”. Universidad de Granada: Granada, 2001.
- [7] C. Sotos, S. Vanhoof, W. Noortgate e P. Onghena, “How confident are students in their misconceptions about hypothesis test?”, *Journal of Statistics Education*, vol.17, 2009.
- [8] C. Batanero, O. Vera e C. Díaz, “Dificultades de estudiantes de Psicología en la comprensión del contraste de hipótesis”, *Números*, vol. 80, pp. 91-101, 2012.
- [9] A. Harradine, C. Batanero e A. Rossman, “Students and teachers knowledge of sampling and inference”, In C. Batanero, G. Burril and C. Reading, Eds., *Teaching statistics in school mathematics: Challenges for teaching and teacher education*. New York: Springer, pp. 235-246, 2011.
- [10] S. Krishnan e N. Idris, “Student’s Misconceptions about Hypothesis Test”. *REDIMAT*, vol.3, pp. 276-283, 2014.
- [11] D. Ben-Zvi e J. Garfield, “Research on Statistical Literacy, Reasoning, Thinking: Issues, Challenges and Implications”, In D. Ben-Zvi and J. Garfield, Eds., *The challenge of Developing Statistical Literacy, Reasoning and Thinking*. Netherlands: Springer, pp.397-409, 2005.