

Uso interativo de planilha eletrônica para o ensino de Estatística: O caso do valor de p .

Interactive use of spreadsheets to teach statistics: The case for the p value

FERNANDO FREI¹

Resumo

O objetivo deste artigo foi avaliar os efeitos do uso de planilha eletrônica interativa para simular eventos na aprendizagem do conceito estatístico inferencial do valor de p , que representa a probabilidade de o efeito observado entre os tratamentos ser devido a variação natural das amostras, e não aos fatores que estão sendo pesquisados.

O presente estudo possui uma abordagem metodológica quantitativa e qualitativa, baseada na descoberta guiada, em que os alunos recebem uma série de atividades que os levam a um objetivo predeterminado. Os resultados obtidos neste estudo demonstram que o procedimento adotado pode ser um aliado no ensino da inferência, e que a simulação torna as atividades mais ativas permitindo que os alunos descubram os próprios princípios, tornando o aprendizado mais efetivo.

Palavras-chave: Valor de p , Planilha Eletrônica, Inferência, Simulação.

Abstract

The aim of this article was to evaluate the effects of the use of interactive spreadsheets to simulate events for the learning of the inferential statistical concept to determine the p value, which represents the probability of the effect observed between the treatments being due to the natural variation of the samples, and not to the factors being investigated. The present study has developed a quantitative and qualitative methodological approach based on guided discovery in which students are offered a series of activities that leads them to a predetermined objective. The results obtained show that the procedure chosen can be helpful for teaching inference and also indicate that simulation makes activities more dynamic, allowing students to discover the principles themselves, which makes learning more effective.

Keywords: p value, Spreadsheets, Inference, Simulation.

¹ Doutor em Saúde Pública – USP e Bacharel em Estatística – UFSCar. Professor de Estatística - UNESP - FCLAssis, e-mail: fernando.frei@unesp.br

Introdução

O ensino e aprendizagem da estatística está incluído em todos os níveis de educação e ganhou reconhecimento em muitas áreas da ciência. Mesmo que visto de forma mais superficial no Ensino médio, a Estatística é disciplina obrigatória em diversos cursos de nível superior, não apenas na área de exatas. Em grande parte das disciplinas, a capacidade de entender, interpretar e criticar resultados de pesquisas está se tornando uma habilidade essencial (JAMIE, 2002).

Muitos estudantes universitários são obrigados a fazer cursos de estatística para sua formação e o Teste de Hipóteses em conjunto com os princípios inferenciais é frequentemente ensinado em cursos introdutórios, comumente como tópicos finais, mas de forma geral, tornam-se o objetivo principal desses cursos (AQUILONIUS; BRENNER, 2015; ANDERSON-COOK; DORAI-RAJ, 2003; ZIEFFLER et al. 2008). No Brasil, a Inferência Estatística é também desenvolvida em cursos introdutórios com carga horária de 60 horas aula concentradas em um semestre.

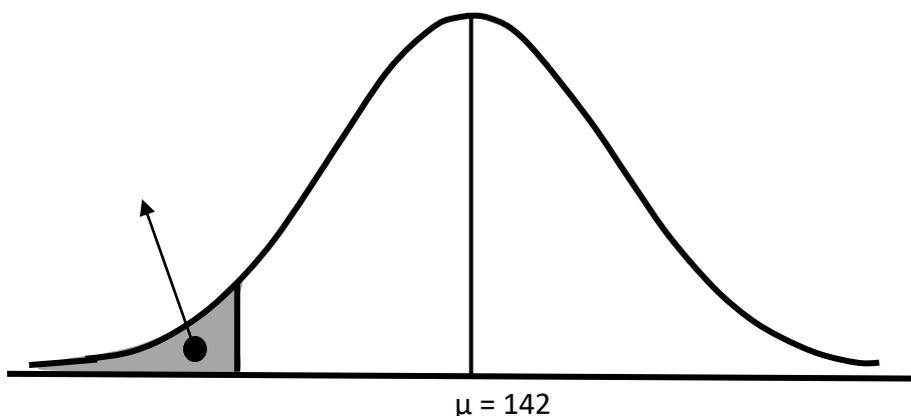
O valor de p

Diversas críticas têm sido feitas sobre o uso e as limitações do valor de p e os Teste de Significância da Hipótese Nula (NHST sigla em inglês para denominar o teste) (COHEN, 1994; LANG et al., 1998; GOODMAN, 2003; HUBBARD; LINDSAY, 2008; NUZZO, 2014; GREENLAND et al., 2016). No entanto, desde sua introdução na década de 1920, o valor p, usado nos testes de significância e a teoria dos testes de hipóteses - NHST, permeiam a pesquisa científica quase completamente. Essas teorias permitem que um pesquisador aborde hipóteses científicas por meio de hipóteses estatísticas. Nesses casos, os pesquisadores muitas vezes desejam refutar a hipótese nula, caracterizada pela afirmativa matemática de ausência de diferença entre parâmetros ou ausência de associação entre variáveis (BIAU et al., 2010).

A ideia de teste de significância foi proposta por Fisher, que relacionou o valor p como um índice a ser usado para medir a discrepância entre os dados amostrais e a hipótese nula denominada de H_0 (STERNE, 2002). Assim, o valor p é a probabilidade de obter um efeito igual ou mais extremo do que seria observado considerando a hipótese nula verdadeira (BIAU et al., 2010), ou de forma mais prática, o valor de p representa a probabilidade do efeito observado entre os tratamentos ser devido a variação natural das amostras, e não aos fatores que estão sendo pesquisados. Desta forma, poderíamos

avaliar, por exemplo, uma hipótese nula indicando que a média de uma dada população seria igual a 142 $\Rightarrow H_0: \mu = 142$. Para qualquer magnitude da média amostral obtida após um experimento, podemos calcular uma estatística, denominada de estatística do teste, que mede a diferença do que foi observado entre a média amostral e o valor indicado na Hipótese Nula, a suposta média populacional. Esta estatística do teste é convertida em probabilidade, ou seja, o valor p , usando a distribuição de probabilidade da estatística de teste sob a hipótese nula (BIAU et al., 2010). Para Fisher, o valor de uma probabilidade p muito pequena indica que um evento excepcionalmente raro ocorreu ou a teoria (H_0) não é verdadeira. Desta forma, o valor da probabilidade p representa um caminho para avaliar a plausibilidade da Hipótese Nula H_0 . De maneira geral, um valor muito pequeno de p pode indicar que H_0 não é verdadeira. Esses conceitos podem ser observados na figura 1 a seguir.

Figura 1. Gráfico da distribuição t-student; t_0 indica a estatística do teste; μ denota a média populacional e a área hachurada em cinza indica a probabilidade à esquerda de t_0 , ou seja, o valor de p .



Fonte: O autor

Por outro lado, a interpretação pouco objetiva a essa abordagem levou Neyman e Pearson a propor o que eles chamaram de "testes de hipóteses". Dessa forma, foi proposta uma regra de decisão, a qual delimita níveis para a rejeição ou não da hipótese H_0 baseada em valores de significância α , geralmente de 5% ($p < 0,05$) ou de 1% ($p < 0,01$). No caso da rejeição de H_0 , o pesquisador "aceitaria" uma segunda hipótese denominada de hipótese alternativa, denotada por H_1 . (CURRAN-EVERETT, 2008).

A teoria de Neyman-Pearson requer a especificação de uma hipótese alternativa precisa que é apresentada antecipadamente, juntamente com as taxas de erro tipo I – rejeitar H_0 .

quando ela é verdadeira (α) e tipo II – não rejeitar H_0 quando ela é falsa (β) (STERNE, 2002).

Ainda que gere críticas, é certo que o valor de p e os Teste de Significância da Hipótese Nula recebam atenção e sua aplicabilidade é inerente aos experimentos nas mais diversas áreas da ciência.

A aprendizagem do valor de p

Em artigo intitulado “*Eight common but false objections to the discontinuation of significance testing in analysis of research data*” Schmidt e Hunter (1997) fazem a afirmação de que: “*Professores de estatística sabem que é muito mais fácil para os alunos entenderem as estimativas por Intervalos de Confiança do que os testes de significância com sua lógica invertida*”. Este não é o único trabalho a tratar das dificuldades do ensino e aprendizagem do valor de p e NHST.

Diversos estudos apontam para as dificuldades na aprendizagem do conceito e aplicação do valor de p . Muitos estudantes confundem e realizam a troca dos conceitos do valor de p e do nível de significância α (FALK, 1986; HALLER; KRAUSS, 2002). No entanto, esse não é o único equívoco observado, existe uma grande variedade de interpretações do valor de p , e a maioria delas é incorreta (VALLECILLOS, 2002; BATANERO; DIAZ, 2006; SOTOS et al., 2007; GOODMAN, 2008; CUMMING, 2010; AQUILONIUS; BRENNER, 2015; BADENES-RIBERA et al., 2016).

O uso de planilhas eletrônicas no ensino de estatística

As décadas de 1980 e de 1990 foram marcadas pelo surgimento dos computadores pessoais no mercado de trabalho, lazer e educação. Esse advento acarretou um grande impacto no ensino e aprendizado da estatística com o desenvolvimento de programas computacionais de maneira a permitir que os estudantes realizem tarefas computacionais com mais rapidez e eficiência, liberando-os para se concentrarem mais em conceitos de estatística. Desta forma, o computador não apenas funciona como uma poderosa ferramenta computacional, mas também pode ajudar a reforçar conceitos específicos e fornecer condições para diversas aplicações (JAMIE, 2002).

As aplicações do uso de planilhas eletrônicas no ensino de Estatística são diversas. Hunt e Mashhoudy (2004) sugerem alguns princípios simples para a confecção de gráficos,

para a representação de dados univariados e bivariados. As planilhas também podem ser úteis no ensino de experimentos (SALEHI, 2016), testes envolvendo a distribuição Qui-Quadrado (CAÑADAS et al., 2012) e para explorar conceitos envolvendo intervalos de confiança e testes de hipóteses (QUINTELA-DEL-RÍO; FRANCISCO-FERNÁNDEZ, 2017). Duffy (2010) utiliza planilhas eletrônicas demonstrando a natureza e a frequência do erro de tipo I e Martin e College (2008) também usam planilhas eletrônicas para ilustrar como professores podem orientar seus alunos em tópicos relacionados ao teste F, aos testes t-Student e à multicolinearidade.

A simulação desenvolvida com o auxílio das planilhas eletrônicas ajuda a imitar a amostragem em situações reais, facilita a repetição de amostras de uma dada população, torna as atividades mais ativas, permitindo que os alunos descubram os próprios princípios interagindo com a simulação, enfim, o aprendizado é mais efetivo porque torna a teoria e conceitos mais concretos para o aluno (ERICKSON 2006; LANE E PERES, 2006; ZACHAROPOULOU 2006; CHANDRAKANTHA 2014).

A principal questão de pesquisa que norteou o estudo foi: Quais são os efeitos de usar a planilha eletrônica interativa na aprendizagem do conceito estatístico inferencial do valor de p ?

Métodos

O método empregado neste trabalho é fundamentado no aprendizado da descoberta. Esse estilo de aprendizagem é baseado na ideia de que o conhecimento que os estudantes encontram para si mesmos tem mais valor do que o conhecimento que foi meramente explicado (KLAHR; NIGAM, 2004). O objetivo do aprendizado da descoberta é envolver os alunos em atividades que os levam a “descobrir” um princípio. Verificou-se, no entanto, que os alunos nem sempre percebem o princípio que o instrutor pretendia que eles descobrissem. Isso ocorre dado que os alunos podem não perceber regularidades ou padrões nos dados. Além disso, os alunos têm dificuldades em mudar suas crenças pré-estabelecidas, mesmo se os dados contradizem essas crenças (DE JONG; VAN JOOLIGEN, 1998; MILLS, 2002). Tem sido sugerido, portanto, que tanto o aprendizado orientado quanto a descoberta, ou uma combinação de ambos são métodos de ensino necessários (DELMAS et. al., 1999; LANE; PERES, 2006). Com o aprendizado de

descoberta guiada, os alunos recebem uma série de atividades que os levam a um objetivo predeterminado (LANE; PERES, 2006).

Uma maneira de guiar o aprendizado da descoberta pode ser alcançada em classes de estatística com o uso de simulação de computador e outras estratégias didáticas. A simulação computacional permite o desenho de um grande número de amostras de uma população hipotética ou de um conjunto de dados rapidamente. Ademais, a simulação por computador pode, por exemplo, ser usada para modelar procedimentos de inferência.

Com o objetivo de combinar o aprendizado de descoberta por meio da planilha eletrônica interativa com o aprendizado guiado, foi disponibilizado um vídeo de dez minutos com os conceitos básicos envolvendo valor de p .

Sujeitos

A metodologia de simulação com imagens foi aplicada em um curso de estatística inferencial para 40 estudantes de graduação de Engenharia Biotecnológica da Faculdade de Ciências e Letras de Assis da Universidade Estadual Paulista – UNESP, São Paulo, Brasil. A carga horária para o curso foi de 60 horas durante um semestre.

Planilha eletrônica para o ensino do valor de p .

A atividade tem como objetivo realizar simulações para apresentar aos alunos os conceitos que envolvem o cálculo e uso do valor de p na inferência estatística.

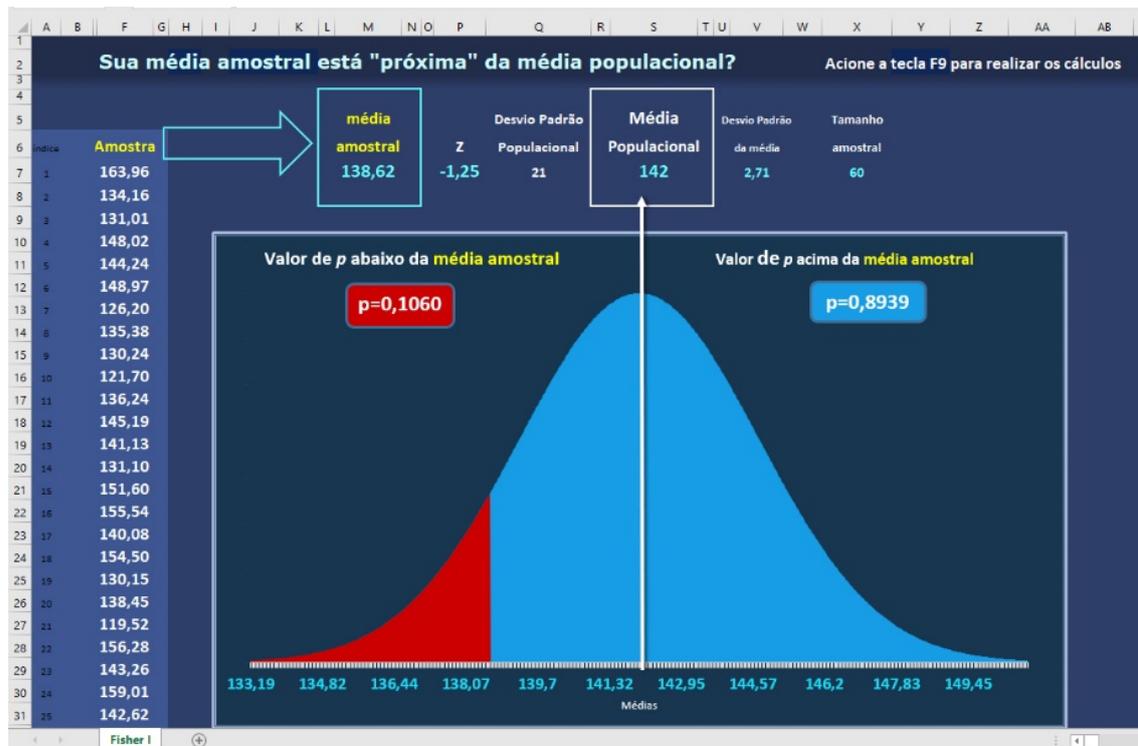
A principal questão apresentada aos estudantes foi: Sua média amostral está próxima da média populacional?

Para responder a esta questão, os alunos receberam uma planilha eletrônica Microsoft Excel® (figura 2) para desenvolver a atividade.

Inicialmente foi simulada uma população com distribuição normal com média 135 e desvio padrão 21 de tamanho 1000. Essa população é disposta na coluna B. A coluna A apresenta um índice, denominado de ponteiro, útil para identificar a posição a ser sorteada na população. Nas colunas C, D e E são dispostas funções para realizar o sorteio aleatório e na coluna F, nova função busca o índice de ordem e apresenta cada valor amostral. As colunas B, C, D e E são ocultas e o aluno visualizará somente a coluna F com amostra de tamanho 60. A coluna M apresenta o valor da média amostral e a coluna S contém o valor

da média populacional a ser posta à prova. Assim, a coluna S foi interpretada como sendo $H_0 = 142$. Esse método procura simular uma situação real em que o pesquisador não possui acesso à sua população, mas sim a uma amostra desta população. Esse conceito pode ser trabalhado pelo docente em aulas anteriores e/ou durante a atividade com a planilha eletrônica.

Figura 2. Planilha eletrônica interativa.



Fonte: O autor

Para realizar o cálculo da média amostral a coluna M recebe a função média e na coluna P é feito o cálculo do valor de Z_0 pela normal padronizada (1).

$$Z_0 = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \quad (1)$$

A curva da distribuição normal é apresentada com a respectiva área à esquerda do valor de Z_0 calculado para cada amostra. Os valores de p obtidos para cada área são ilustrados na figura (figura 2).

Todos os elementos apresentados na planilha, tratados de forma que o *layout* seja atraente, podem ser abordados pelo professor.

A simulação de situações com o uso da planilha pode ser ampliada de forma que outras populações possam ser confrontadas com o valor H_0 . Também é possível alterar o desvio padrão das populações da qual as amostras são extraídas, o que permite uma discussão sobre questões envolvendo a variabilidade. Outra possibilidade didática interessante seria ocultar o valor da média amostral para que o estudante avaliasse a área/valor de p .

A atividade consistiu em realizar 30 amostras observando a relação da média amostral, valor de Z_0 , área relativa a Z_0 e a probabilidade indicada pelo valor de p a esquerda de Z_0 . Para realizar uma nova amostragem o aluno acionou a tecla F9 o que permitiu efetuar automaticamente os diversos cálculos, bem como a visualização da área abaixo de Z_0 e o valor de p . Cada novo acionamento na tecla F9 permitiu a seleção de uma nova amostra e a repetição de todo o processo, com nova média, novo valor de Z_0 , nova área e novo valor de p .

Avaliação do conceito do valor de p

Após a atividade, cada aluno respondeu a quatro questões dissertativas sobre a temática envolvendo conceito do valor de p , proximidade da média amostral em relação à média populacional e Rejeição a H_0 . As respostas foram classificadas como respostas corretas, respostas corretas, mas incompletas e respostas incorretas.

Resultados e Discussão

Os resultados podem ser considerados satisfatórios visto que a percentagem de acertos superou 60% no geral. Os resultados de cada categoria de acerto são apresentados na tabela 1 a seguir. O maior percentual de erro está ligado ao conceito do valor de p . Pode-se observar que 22,5% das respostas foram incorretas quanto ao conceito desenvolvido por Fisher. Por outro lado, as três outras questões apresentam percentuais de 65% de acerto (tabela1).

Tabela 1. Avaliação das questões propostas aos estudantes.

Questões	Respostas (%)		
	Corretas	Corretas Incompletas	Incorretas
Qual o significado de rejeitar H_0 ?	65,0	25,0	10,0
Qual o significado de não rejeitar H_0	65,0	27,5	7,5
Você pode afirmar que suas amostras são de outra população com média menor que 142?	65,0	17,5	17,5
Para Fisher, o valor de uma probabilidade p muito pequeno indica...	47,5	30,0	22,5
Média	60,6	25,0	14,4

Fonte: O autor

Muitas das respostas incompletas sobre o conceito do valor de p estão relacionadas a falta do entendimento que um determinado valor de p pode indicar que a média amostral pode ter vindo da população, e que um evento raro teria ocorrido.

Algumas das respostas dos estudantes refletem o entendimento dos conceitos após a atividade com a planilha eletrônica interativa. Abaixo algumas respostas representativas de respostas corretas e incorretas:

Questão - Para Fisher, o valor de uma probabilidade p muito pequeno indica

Resposta: que há insuficiência de evidências para sustentar H_0 , ou seja, aquele dado amostral em questão possui um valor raro de ocorrer na população por variação natural, logo deve pertencer a outra população e aceita-se H_1

Questão - O que significa rejeitar H_0 ?

Resposta: Significa que a amostra coletada possui alta probabilidade de não ter sido retirada da população dada, sendo então de uma outra população diferente.

Questão - O que significa não rejeitar H_0 ?

Resposta: A não rejeição de H_0 indica que a amostra pode fazer parte de uma variação dentro da população, não se tratando de uma população diferente.

Resposta: Significa apenas que não se conseguiu, através dos dados disponíveis, demonstrar a sua falsidade, o que difere completamente de provar a sua veracidade.

Resposta: Significa que as médias amostrais obtidas fazem parte da variação natural.

Questão - Qual a interpretação que você faz do valor de p ?

Resposta: É uma medida de quanta evidência você tem contra a hipótese nula.

Questão - Depois de fazer todas as diversas análises para as várias amostras, como você responde à questão: Sua média amostral está "próxima" da média populacional? *

Resposta - ...como foi observado no gráfico podemos ver que os valores encontrados estão distantes do centro do gráfico que é a média populacional.

A resposta apresentada mostra que o aluno fez relação entre a conceitos de distância sem reportar a regra de que o valor de p seria menor que 5%, o que pode muitas das vezes confundir e trazer equívocas na tomada de decisão.

Por outro lado, erros na interpretação do valor de p são comuns (SEBASTIANI E VIALI, 2011) e neste aspecto, algumas respostas equivocadas dos alunos puderam ser observadas:

Questão - Qual a interpretação que você faz do valor de p ?

Resposta. valor p é o maior nível de significância com que se rejeitaria a hipótese H_0 .

A resposta é comum e mostra a troca de conceitos entre valor de p e o nível alfa estabelecido pela teoria de Newman-Pearson.

Questão - Você pode afirmar que suas amostras são de outra população com média menor que 142?

Resposta - Não é possível afirmar pois estamos trabalhando com probabilidade.

A resposta apresentada não foi única e evidencia que alguns alunos não tem a probabilidade como uma ferramenta para avaliar processos de decisão.

Por outro lado, as respostas corretas, mas incompletas ou mesmo as incorretas foram úteis para que o professor pudesse realizar uma intervenção individualizada – *feedback* - após a correção das questões. Assim, a utilização da planilha eletrônica interativa proporcionou, não só aos alunos a oportunidade de uma melhor aprendizagem sobre a

temática proposta, mas também a possibilidade de que o professor possa reduzir a distância entre a compreensão e o desempenho do aluno (HATTIE; TIMPERLAY, 2007). O resultado da avaliação dos alunos apresentou uma pontuação acima da média em uma escala de variação de zero a dez (tabela 2).

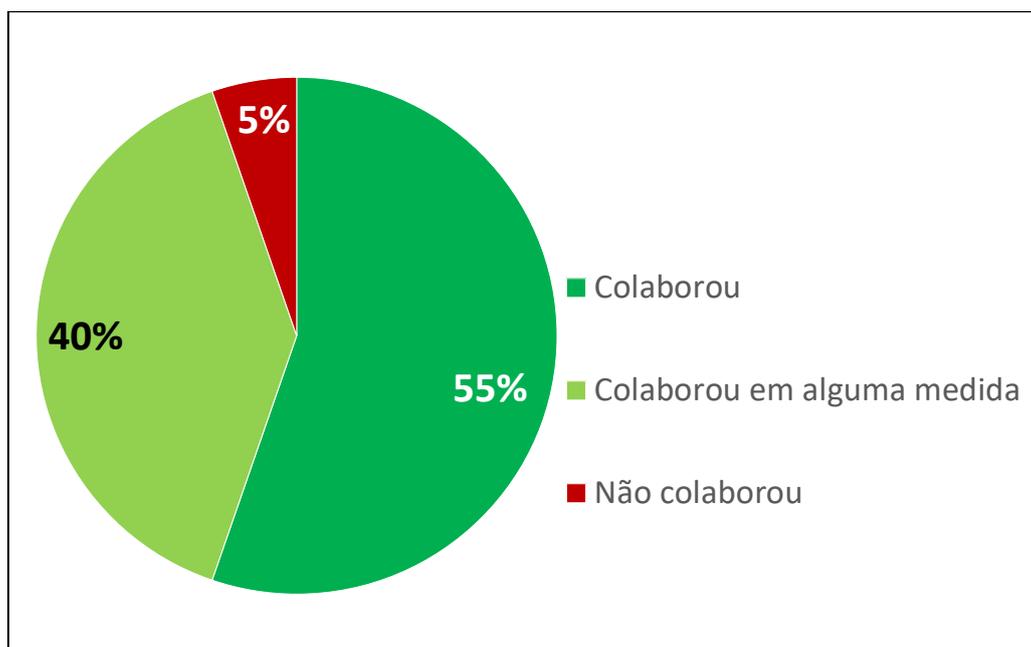
Tabela 2. Resultados gerais dos alunos (n = 40)

Medida	Resultado
Média	7,2
Desvio Padrão	2,0
Mediana	7,7
Coefficiente de Variação	27,7
Mínimo	2,7
Máximo	10,0

Fonte: O autor

Outro resultado promissor diz respeito a avaliação da atividade pelos próprios alunos. A figura 3 apresenta apenas 5% das respostas como negativas, ou seja, 2 alunos indicaram que a atividade não colaborou para o aprendizado dos conceitos relativos ao valor de p .

Figura 2. Avaliação da atividade pelos estudantes.



Fonte: O autor

Considerações Finais

Os resultados obtidos neste estudo com a utilização de planilha eletrônica interativa para simular eventos, demonstram que o procedimento adotado pode ser um aliado no ensino da inferência, mais especificamente nos conceitos que envolvem o valor de p empregado nos testes estatísticos.

Deve estar claro que o sucesso da aprendizagem não reside na utilização de uma única estratégia, mas sim da combinação da simulação e de outras atividades que possam enriquecer o ambiente de estudo e aproximar o estudante do contexto de pesquisa e da realidade social em que vive. Assim, como destaca Erickson 2006, o processo de simulação faz com que a atividade seja mais concreta para o aluno.

É importante ressaltar que a representação visual da distribuição Normal permitiu que os alunos fizessem conexões entre os conceitos subjacentes ao valor de p , o que evidencia a importância de implementar métodos de instruções mais envolventes e diversificados (LE, 2013). Isto é particularmente verdadeiro para os adolescentes, que vivenciam a era da tecnologia, onde a informação visual é a norma (MURPHY, 2009).

Este estudo apresentou uma alternativa ao ensino do conceito envolvendo o valor de p . Essa abordagem, utilizando planilhas eletrônicas interativas pode facilitar a aprendizagem dos conceitos de inferência estatística. Por outro lado, grande parte dos alunos tem acesso a planilhas eletrônicas, o que permite a disseminação de atividades, seja em sala de aula, ou em outros ambientes, sem que sejam necessários programas computacionais custosos.

Referências

ANDERSON-COOK, C. M.; DORAI-RAJ, S. Making the Concepts of Power and Sample Size Relevant and Accessible to Students in Introductory Statistics Courses using Applets. *Journal of Statistics Education* v. 11, n. 3, p. 1-12, 2003.

AQUILONIUS, B. C., BRENNER, M. E. Students' reasoning about p -values. *Statistics Education Research Journal*, v. 14, n. 2, p. 7-27, 2015.

BADENES-RIBERA, L.; FRIAS-NAVARRO, D.; IOTTI, B.; BONILLA-CAMPOS, A.; LONGOBARDI, C. Misconceptions of the p -value among Chilean and Italian Academic. *Psychologists*, v. 7, p. 1-7, 2016.

BATANERO, C.; DIAZ, C. Methodological and Didactical Controversies around Statistical Inference (2006). Proceedings of 38th Conference of the French Statistical Conference. Actes du 36ièmes Journées de la Societé Française de Statistique. Paris: Societé Française de Statistique. Disponível em <<http://www.ugr.es/~batanero/pages/ARTICULOS/SFE.pdf>>. Acesso em: 10 fevereiro de 2018.

BIAU, D. J.; JOLLES, B. M. PORCHER, R. P Value and the Theory of Hypothesis Testing: An Explanation for New Researchers, *Clin Orthop Relat Res*, v. 468, p. 885-892, 2010.

CANADAS, G. R.; BATANERO, C.; DIAZ, C.; ROA, R. Psychology students' understanding of the chi-squared test. *Statistique et Enseignement*, v. 3, n. 1, p. 3-8, 2012.

CHANDRAKANTHA, L. Simulation using excel data tables in teaching introductory statistics. *Journal of Computing Sciences in Colleges*, v. 29, p. 29-34, 2014.

COHEN, J. The earth is round ($p < 0.05$). *American Psychologist*, v. 49, n. 12, p. 997-1003, 1994.

CUMMING, G. (2010). Understanding, teaching and using p values. In A. Rossman & B. Chance (Eds.), ICOTS-10 Proceedings. Disponível em https://iase-web.org/documents/papers/icots8/ICOTS8_8J4_CUMMING.pdf. Acesso em: 10 fevereiro de 2018.

CURRAN-EVERETT, D. Explorations in statistics: hypothesis tests and P values. *Adv Physiol Educ*, v. 33, p. 81-86, 2009.

DE JONG, T.; VAN JOOLIGEN, W. Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, v. 68, n. 2, p. 179-201, 1998.

DELMAS, R.; GARFIELD, J.; CHANCE, B. A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education*, v. 7, n. 3, 1999. Disponível em: <<https://ww2.amstat.org/publications/jse/secure/v7n3/delmas.cfm>>. Acesso em: 10 de janeiro de 2018.

DUFFY, S. Random Numbers Demonstrate the Frequency of Type I Errors: Three Spreadsheets for Class Instruction. *Journal of Statistics Education*, v. 18, n. 2, p. 1-16, 2010.

ERICKSON, T. Using simulation to learn about inference. Proceedings of the Seventh International Conference on Teaching Statistics, ICOTS-7, 2006. Disponível em: https://www.researchgate.net/profile/Tim_Erickson2/publication/248282382_USING_SIMULATION_TO_LEARN_ABOUT_INFERENCE/links/5694828c08ae425c6896488d/USING-SIMULATION-TO-LEARN-ABOUT-INFERENCE.pdf. Acesso em: 18 de janeiro de 2018.

FALK, R. Misconceptions of statistical significance. *Journal of Structural Learning*, v. 9, p. 83-96, 1986.

GOODMAN S. A dirty dozen: twelve p-value misconceptions. *Semin Hematol*, v. 45, n. 3, p. 135-40, 2008.

GOODMAN, S. Commentary: The P-value, devalued. *International Journal of Epidemiology*, v. 32, p. 699-702, 2003.

GREENLAND, S.; SENN, S. J.; ROTHMAN, K. J.; CARLIN, J. B.; POOLE, C.; GOODMAN, S. N.; ALTMAN; D. G. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations, *Eur J Epidemiol*, v. 31, p. 337-350, 2016.

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

HATTIE, J.; TIMPERLAY, H. The power of feedback. *Review of Educational Research*, v. 77, n. 1, p. 81-112, 2007.

HUBBARD, R.; LINDSAY, R. M. Why P values are not a useful measure of evidence in statistical significance testing. *Theory Psychol*, v. 18, p. 69-88, 2008.

HUNT, N.; MASHHOUDY, H. Charts in Excel – A Series Matter. *Teaching Statistics*, v. 26, n. 2, p. 49-53, 2004.

JAMIE. D. M. Using Computer Simulation Methods to Teach Statistics: A Review of the Literature, *Journal of Statistics Education*, v. 10, n. 1, p. 1-20, 2002.

KLAHR, D.; NIGAM, M. The equivalence of learning paths in early science instruction. *Psychological Science*, v. 15, n. 10, p. 661-667, 2004.

LANE, D. M.; PERES, S. C. Interactive simulations in the teaching of statistics: Promise and Pitfalls. Proceedings of the Seventh Annual Meeting of the International Conference on the Teaching of Statistics, Salvador, Brazil, 2006. Disponível em: <<http://www.ruf.rice.edu/~lane/papers/interactive-simulations.pdf>>. Acesso em: 10 de fevereiro de 2018.

LANG, J. M.; ROTHMAN, K.; CANN, C. I. That confounded P-value. *Epidemiology*, v. 9, p. 7-8, 1998.

LE D. Bringing Data to Life into an Introductory Statistics Course with GAPMINDER. *Teaching Statistics*, v. 35, n. 3, p. 114-122, 2013.

MARTIN, D.; COLLEGE, D. A. Spreadsheet Tool for Learning the Multiple Regression F-test, t-tests, and Multicollinearity. *Journal of Statistics Education*, v. 16, n. 3, 2008. Disponível em: <<https://www.tandfonline.com/doi/pdf/10.1080/10691898.2008.11889573?needAccess=true>>. Acesso em: 15 de fevereiro de 2018.

MILLS, J. Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education*, v. 10, n. 1, p. 1-10, 2002.

MURPHY, S. J. (2009). The power of visual learning in secondary mathematics education. Pearson Education Inc. Disponível em <https://assets.pearsonschool.com/asset_mgr/legacy/200916/MatMon092291HS2011StuMur_LR_20702_1.pdf>. Acesso em: 09 de janeiro de 2018.

NUZZO, R. Scientific method: Statistical errors. *Nature*, v. 506, p. 150 –152, 2004.

QUINTELA-DEL-RÍO, A.; FRANCISCO-FERNÁNDEZ, M. Excel Templates: A Helpful Tool for Teaching Statistics. *The American Statistician*, V. 71, p. 317- 325, 2014.

SALEHI, M. Using MS Excel in Teaching Design of Experiment. *International Journal of Education and Learning Systems*, v. 1, p. 93 – 98, 2016.

SCHMIDT, F. L.; HUNTER, J. E. *What If There Were No Significance Tests? Eight common but false objections to the discontinuation of significance testing in analysis of research data.* HARLOW L., MULAİK S. e STEIGER J. (Eds.) New York. Routledge, p. 37-64, 1997.

SEBASTIANI, R. G.; VIALI, L. Teste de Hipóteses: uma análise dos erros cometidos por alunos de engenharia. *Bolema*, v. 24, n. 40, p. 835-854, 2011.

SOTOS, A. E. C.; VANHOOF, S.; NOORTGATE, W. V.; ONGHENA, P. Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, v. 2, p. 98-113, 2007.

STERNE, J. A. C. Teaching hypothesis tests – time for significant change? *Statistics in Medicine*, v. 21, p. 985-994, 2002.

VALLECILLOS, A. Empirical evidence about understanding of the level of significance concept in hypotheses testing by university students. *Themes in Education*, v. 3, n. 2, p. 183-198, 2002.

ZACHAROPOULOU, H. Two Learning Activities for a Large Introductory Statistics Class. *Journal of Statistics Education*, v. 14, n. 1, p. 1-10, 2006.

ZIEFFLER, A.; GARFIELD, J.; DELMAS, R.; READING, C. A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, v. 7, n. 2, p. 40-58, 2008.

Texto recebido: 13/06/2018
Texto aprovado: 01/05/2019