

ANÁLISE INFERENCIAL DE DADOS NA EDUCAÇÃO: PRIMEIROS PASSOS

Elizete Pereira Bahia Viana¹;

Universidade Estadual do Sudoeste da Bahia (PPGEEd/UESB), Vitória da Conquista, Bahia.
<https://orcid.org/0009-0000-3256-0799>

Marinalva Pereira de Souza²;

Universidade Estadual do Sudoeste da Bahia (PPGEEd/UESB), Vitória da Conquista, Bahia.
<https://orcid.org/0009-0009-8254-1962>

Luís Filipe de Oliveira Novais³;

Universidade Estadual do Sudoeste da Bahia (PPGEEd/UESB), Vitória da Conquista, Bahia.
<https://orcid.org/0009-0005-0850-2018>

Cláudio Pinto Nunes⁴;

Universidade Estadual do Sudoeste da Bahia (PPGEEd/UESB), Vitória da Conquista, Bahia.
<https://orcid.org/0000-0003-1514-6961>

Berta Leni Costa Cardoso⁵;

Universidade do Estado da Bahia (PPGEDuF/UNEB), Guanambi, Bahia.
Universidade Estadual do Sudoeste da Bahia (PPGEEd/UESB), Vitória da Conquista, Bahia.
<https://orcid.org/0000-0001-7697-0423>

Luiz Humberto Rodrigues Souza⁶.

Universidade do Estado da Bahia (UNEB), Guanambi, Bahia.
Universidade Estadual do Sudoeste da Bahia (PPGEEd/UESB), Vitória da Conquista, Bahia.
<https://orcid.org/0000-0001-9237-3928>

RESUMO: A análise inferencial de dados pode avaliar o efeito das intervenções educacionais e viabilizar a escolha de programas de tecnologias educacionais a serem empregadas nas diversas práticas educativas. Além disso, por meio de testes estatísticos, pode-se comparar desempenhos entre diferentes grupos de alunos, identificando padrões e tendências que podem influenciar práticas pedagógicas e fatores que estão estatisticamente associados ao sucesso acadêmico. Sendo assim, este estudo teve o objetivo de apresentar algumas pesquisas da educação que utilizaram a análise inferencial de dados para interpretar seus resultados. Para tanto, foi conduzida a pesquisa descritiva por meio de uma revisão bibliográfica de dois artigos científicos. O primeiro estudo utilizou o teste-t de *Student* e a

análise de correlação Pearson. Mais especificamente, o teste-t de *Student* foi usado para a comparar as médias de uma variável contínua (numérica) entre dois grupos independentes, enquanto a correlação de Pearson estimou a correlação entre a média de duas variáveis contínuas (numéricas). O segundo artigo apropriou do teste qui-quadrado para avaliar a associação entre as variáveis categóricas. Em síntese, nosso estudo sugeriu que a análise inferencial envolve um tratamento criterioso capaz de extrair conclusões importantes para a tomada de decisões educacionais. Por outro lado, é necessário considerar o amplo contexto educativo para não simplificar a complexidade inerente à essa área de conhecimento. Outra questão relevante é a presença de variáveis não observadas ou desconhecidas que podem influenciar nos resultados, exigindo uma abordagem cautelosa ao interpretar as conclusões.

PALAVRAS-CHAVE: Análise Inferencial. Educação. Pesquisa Quantitativa.

INFERENCEAL DATA ANALYSIS IN EDUCATION: FIRST STEPS

ABSTRACT: Inferential data analysis can assess the effect of educational interventions and enable the selection of educational technology programs to be used in different educational practices. In addition, statistical tests can be used to compare performances between different groups of students, identifying patterns and trends that may influence pedagogical practices and factors that are statistically associated with academic success. Therefore, this study aimed to present some educational research that used inferential data analysis to interpret its results. To this end, descriptive research was conducted through a literature review of two scientific articles. The first study used Student's t-test and Pearson correlation analysis. More specifically, Student's t-test was used to compare the means of a continuous (numerical) variable between two independent groups, while Pearson's correlation estimated the correlation between the means of two continuous (numerical) variables. The second article used the chi-square test to assess the association between categorical variables. In summary, our study suggested that inferential analysis involves careful treatment capable of extracting important conclusions for educational decision-making. On the other hand, it is necessary to consider the broad educational context so as not to simplify the complexity inherent in this area of knowledge. Another relevant issue is the presence of unobserved or unknown variables that may influence the results, requiring a cautious approach when interpreting the conclusions.

KEY-WORDS: Inferential Analysis. Education. Quantitative Research.

INTRODUÇÃO

A pesquisa qualitativa é reconhecida por usar dados não-numéricos, buscando compreender significados, contextos e padrões diversos. Utiliza-se a técnica interpretativa como a análise de conteúdo, proporcionando nesta abordagem uma compreensão mais

aprofundada e rica de fenômenos complexos, a fim de explorar as nuances e capturar a essência das experiências humanas, revelando determinados aspectos que podem escapar a métodos estritamente quantitativos. Por outro lado, a pesquisa quantitativa lida com dados mensuráveis e quantificáveis, utilizando métodos estatísticos para identificar padrões, relações e tendências (Thomas; Nelson; Silverman, 2012).

Previamente, Gatti (2007) afirmou que a abordagem quantitativa se pauta em pressupostos positivistas, na objetivação e generalização dos resultados, no distanciamento entre sujeito e objeto, e na neutralidade do pesquisador como elementos que asseguram e legitimam a cientificidade de uma pesquisa. Embora distintas, ambas não são mutuamente exclusivas, tendo em vista que muitos pesquisadores adotam abordagens mistas, combinando elementos das duas para obter uma compreensão mais abrangente e holística. Esta escolha entre uma abordagem e outra depende da natureza da pesquisa, dos objetivos do estudo e da complexidade do fenômeno em análise. Em suma, a análise de dados independentemente da abordagem é uma ferramenta poderosa para compreender a natureza dos dados coletados.

Outrossim, Gatti (2004) ressaltou que poucos estudos na educação empregam metodologias quantitativas e que há problemas educacionais que necessitam ser qualificados e contextualizados através de dados quantitativos. Para ela, compreender questões como o analfabetismo no Brasil e discutir políticas em relação a esse problema demandam compreensão quantitativa e, portanto, tal abordagem é muito importante para estabelecimento das relações que se quer entender a partir da análise dos dados quantificáveis.

Para entender tamanha importância do tratamento de dados quantitativos, Souza e Kerbauy (2017) apontaram que, a partir da década de 1930, o enfoque quantitativo subsidiou investigações realizadas na área, principalmente as vinculadas ao Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), intensificando nas décadas seguintes. No mesmo entendimento, Gatti (2004) ressaltou que estudos publicados no início dos anos 1970 nos mostraram que a pesquisa quantitativa em educação era muito escassa e mesmo incipiente, pois empregavam apenas a análise descritiva de tabelas de frequências.

Nesta dinâmica, a análise inferencial é entendida como um ramo da estatística que se concentra em realizar inferências ou generalizações sobre uma população com base em uma amostra representativa dos dados. Em vez de descrever características dos dados observados, ela busca tirar conclusões mais amplas e preditivas a partir da aplicação de testes estatísticos que avaliam se as diferenças ou relações observadas na amostra são significativas o suficiente para toda a população (Thomas; Nelson; Silverman, 2012).

A análise inferencial é de fundamental importância para a educação, pois ela avalia a eficácia das intervenções educacionais e viabiliza a escolha de programas de tutoria ou tecnologias educacionais a serem empregadas nas diversas práticas educativas. Além disso, por meio de testes estatísticos, pode-se comparar desempenhos entre diferentes

grupos de alunos, identificando tendências e fatores que estão estatisticamente associados ao sucesso acadêmico. Assim, este estudo teve o objetivo de apresentar algumas pesquisas da educação que utilizaram a análise inferencial de dados para interpretar seus resultados.

METODOLOGIA

Trata-se de uma pesquisa descritiva que utilizou a revisão bibliográfica de dois artigos que utilizaram a análise inferencial para interpretar seus dados. As pesquisas utilizadas para exemplificar o uso da análise inferencial foram extraídas do Google Acadêmico, a partir dos descritores “educação”, “análise inferencial” e “pesquisa quantitativa”, combinados com o operador booleano “AND”. A seleção dos estudos aconteceu a partir da leitura dos títulos e resumos, sob a responsabilidade de E.P.B.V, M.P.S e L.F.O.N, conferido por uma quarta pessoa (L.H.R.S).

RESULTADOS

A realização da análise inferencial de dados envolve a aplicação de materiais e métodos que visam extrair conclusões significativas e generalizáveis a partir de dados específicos. Este processo é fundamental para embasar decisões e práticas pedagógicas, contribuindo para um avanço na qualidade de ensino.

O estudo de Fialho, Cid e Coppi (2023) analisou as vantagens e as dificuldades encontradas por professores e alunos quanto ao uso de uma seleção de nove plataformas e tecnologias digitais (PTD) em atividades escolares. As análises estatísticas inferenciais utilizadas pelos autores foram o teste-t de *Student* e a análise de correlação. Mais especificamente, o teste-t de *Student* para amostras independentes foi utilizado para a comparar as médias de concordância das vantagens e das dificuldades na utilização das PTD por alunos e professores, enquanto a análise de correlação de Pearson foi realizada para estimar a correlação entre a média da frequência de uso geral das PTD e o nível de concordância das vantagens e das dificuldades na utilização das PTD por alunos e professores.

A análise dos dados a partir do teste-t de *Student* resultou na elaboração de doze figuras (seis quadros e seis gráficos) que sumarizaram os principais resultados oriundos do teste estatístico. Para fins didáticos, utilizamos o quadro 4 do estudo em questão como exemplo (Figura 1).

Figura 1: Resultados organizados a partir do teste-t de *Student*.

Quadro 4 – Diferença média dos problemas e das dificuldades do uso das plataformas e tecnologias digitais por alunos e professores.

| Problemas e dificuldades | Respondentes | N | M | DP | t(gl) |
|---|--------------|-----|------|------|----------------|
| Dificuldade com o uso das TI | Alunos | 104 | 2,35 | 1,05 | -1,08 (145) |
| | Professores | 43 | 2,56 | 1,14 | |
| Falta de conhecimentos como utilizador | Alunos | 105 | 2,27 | 1,00 | -2,03 (69,54)* |
| | Professores | 43 | 2,67 | 1,15 | |
| Dificuldades de funcionamento das plataformas | Alunos | 104 | 2,55 | 1,06 | -1,83 (145) |
| | Professores | 43 | 2,91 | 1,13 | |
| Parque informático obsoleto/antiquado | Alunos | 104 | 2,40 | 1,03 | -7,42 (145)*** |
| | Professores | 43 | 3,88 | 1,26 | |
| Insuficiência de equipamentos | Alunos | 104 | 2,60 | 1,14 | -6,83 (145)*** |
| | Professores | 43 | 3,98 | 1,06 | |
| Dificuldade no acesso à internet | Alunos | 103 | 3,42 | 1,22 | -2,13 (93,02)* |
| | Professores | 43 | 3,84 | 1,02 | |

TI: tecnologias da informação; N: número de respondentes; M: média; DP: desvio padrão; t: teste-t de Student; * p < 0,05; ***p < 0,001.

Fonte: Elaboração dos autores.

Fonte: Fialho; Cid; Coppi (2023).

A Figura 1 mostrou que os únicos problemas e dificuldades do uso das PTD que apresentaram diferença estatisticamente significativa entre os grupos (alunos e professores) foram a “falta de conhecimento como utilizador”, “obsolescência do parque informático”, “insuficiência de equipamentos” e “dificuldade no acesso à internet”. Os demais problemas e dificuldades (“dificuldade com o uso das TI” e “dificuldade com o funcionamento das plataformas”) apresentaram uma diferença numérica da média entre os grupos, porém essa diferença não foi considerada significativa pelo teste-t de *Student*.

Ainda no estudo de Fialho, Cid e Coppi (2023), a análise dos dados a partir da correlação de Pearson resultou na elaboração de uma figura (quadro 5) que apresentou o coeficiente de correlação entre a média da frequência de uso das PTD e o nível de concordância dos problemas e das dificuldades do uso das PTD de alunos e professores (Figura 2).

Figura 2: Resultados organizados a partir da correlação de Pearson.

Quadro 5 – Correlação entre a média da frequência de uso das plataformas e tecnologias digitais e o nível de concordância dos problemas e das dificuldades do uso das plataformas e tecnologias digitais de alunos e professores.

| Problemas e dificuldades | Média de uso das PTD por alunos | Média de uso das PTD por professores |
|---|---------------------------------|--------------------------------------|
| Dificuldade com o uso das TI | -0,094 | -0,333* |
| Falta de conhecimentos como utilizador | -0,118 | -0,314* |
| Dificuldades de funcionamento das plataformas | -0,133 | -0,274 |
| Parque informático obsoleto/antiquado | -0,203* | -0,090 |
| Insuficiência de equipamentos | -0,048 | 0,068 |
| Dificuldade no acesso à internet | -0,039 | 0,090 |

PTD: plataformas e tecnologias digitais; TI: tecnologias da informação; * a correlação é significativa no nível 0,05 (bilateral).
Fonte: Elaboração dos autores.

Fonte: Fialho; Cid; Coppi (2023).

A Figura 2 apresentou alguns coeficientes com o sinal negativo. Esse sinal matemático revelou que as variáveis em questão possuem grandezas inversamente proporcional. Dito isso, a interpretação dessa figura pode ser feita da seguinte forma: “os professores que em média fizeram mais uso das PTD, foram aqueles que apresentaram menor dificuldade com o uso das TI”. Usando o mesmo raciocínio, podemos dizer que “os professores que tiveram maior falta de conhecimento como utilizador, foram aqueles que obtiveram menor média de uso das PTD”. A interpretação para por aqui, pois a correlação de Pearson apenas testa a força do relacionamento entre duas variáveis, ou seja, o coeficiente de correlação é uma métrica estatística que mede o grau pelo qual duas variáveis tendem a mudar juntas (Field, 2009). Portanto, a correlação não é causalidade (no próximo capítulo, esse assunto será aprofundado).

Como explicar o coeficiente de correlação positivo? Neste caso, o sinal matemático indicou que as variáveis em questão possuem grandezas diretamente proporcional. Na Figura 2, embora tenhamos dois coeficientes positivos, não podemos fazer afirmações considerando o teste estatístico utilizado, pois a relação entre as variáveis não foi considerada significativa.

O estudo de Caporossi e Ferreira (2011) associou os hábitos vocais autorreferidos por professores do ensino fundamental e médio da rede pública à presença de rouquidão, cansaço ao falar, garganta seca e alteração de voz, também autorreferidos. A análise estatística inferencial utilizada pelas autoras foi o teste qui-quadrado, pois é capaz de avaliar a associação entre duas variáveis categóricas (Field, 2009).

A análise dos dados a partir do teste qui-quadrado resultou na elaboração de duas tabelas que sumarizaram os principais resultados oriundos do teste estatístico. Para fins didáticos, utilizamos a tabela 1 do estudo em questão como exemplo (Figura 3).

Figura 3: Resultados organizados a partir do teste qui-quadrado.

Tabela 1 – Distribuição numérica e percentual dos professores quanto ao sexo (M- Masculino; F- Feminino), idade (maior e menor à média do grupo), tempo de magistério (maior e menor à média do grupo), número de escolas que leciona (maior e menor à média do grupo), e carga horária semanal (maior e menor à média do grupo), segundo a autorreferência de alteração de voz, rouquidão, cansaço ao falar e garganta seca

| Variável | Categoria | Autorreferencia | | | | | | | | | | | |
|--|-----------|------------------|-----------------|--------|---------------|-----------------|--------|------------------|-----------------|--------|---------------|-----------------|--------|
| | | Alteração de voz | | | Rouquidão | | | Cansaço ao falar | | | Garganta seca | | |
| | | n (%) | Sim n (%) | P | n (%) | sim n (%) | P | n (%) | Sim n (%) | P | n (%) | sim n (%) | P |
| Sexo | M | 14 (53.85) | 12 (46.15) | 0.018* | 17 (68.00) | 8 (32.00) | 0.062# | 15 (62.50) | 9 (37.50) | 0.017* | 13 (54.17) | 11 (45.83) | 0.176 |
| | F | 17 (27.42) | 45 (72.58) | | 26 (45.61) | 31 (54.39) | | 20 (33.90) | 39 (66.10) | | 22 (37.93) | 36 (62.07) | |
| Idade (anos) | ≥ 42 | 12 (28.57) | 30 (71.43) | 0.390 | 19 (51.35) | 18 (48.65) | 0.818 | 11 (28.21) | 28 (71.79) | 0.012* | 11 (29.73) | 26 (70.27) | 0.033* |
| | < 41 | 15 (37.50) | 25 (62.50) | | 19 (48.72) | 20 (51.28) | | 22 (56.41) | 17 (43.59) | | 21 (53.85) | 18 (46.15) | |
| Tempo de magistério (anos) | ≥ 14 | 14 (29.17) | 34 (70.83) | 0.763 | 19 (42.22) | 26 (57.78) | 0.344 | 28 (60.87) | 18 (39.13) | 0.874 | 20 (44.44) | 25 (55.56) | 0.802 |
| | < 13 | 17 (43.59) | 22 (56.41) | | 19 (52.78) | 17 (47.22) | | 19 (52.78) | 17 (47.22) | | 15 (41.67) | 21 (58.33) | |
| Número de escolas que leciona | ≥ 2 | 20 (36.36) | 35 (63.64) | 0.773 | 24 (48.00) | 26 (52.00) | 0.314 | 20 (38.46) | 32 (61.54) | 0.376 | 22 (44.00) | 28 (56.00) | 0.763 |
| | 1 | 11 (33.33) | 22 (66.67) | | 19 (59.38) | 13 (40.63) | | 15 (48.39) | 16 (51.61) | | 13 (40.63) | 19 (59.38) | |
| Carga horária semanal (horas) | ≥ 40 | 10 (37.04) | 17 (62.96) | 0.777 | 13 (48.15) | 14 (51.85) | 0.473 | 12 (44.44) | 15 (55.56) | 0.874 | 13 (48.15) | 14 (51.85) | 0.571 |
| | <39 | 20 (33.90) | 39 (66.10) | | 30 (56.60) | 23 (43.40) | | 23 (42.59) | 31 (57.41) | | 22 (41.51) | 31 (58.49) | |

Teste de Chi²

*Significante

tendência

Fonte: Caporossi; Ferreira (2011).

Tomando como referência o valor-p (também será melhor discutido no próximo capítulo), usamos o ponto de corte de “0,05” para considerar a associação entre as variáveis significativa ou não. Assim, um valor-p menor que “0,05” indicou uma associação significativa, enquanto um valor-p maior que “0,05” sinalizou que não houve associação significativa entre as variáveis. Na Figura 3, o asterisco foi utilizado para sinalizar onde o teste qui-quadrado identificou uma associação significativa.

Assim, percebemos que a variável “sexo” apresentou uma associação significativa com a variável “alteração de voz”. Note que o valor-p foi de 0,018. Da mesma forma, a variável “sexo” também se associou significativamente com a variável “cansaço ao falar”(p = 0,017). Observe que a maioria das mulheres estudadas apresentou “alteração de voz” (n=45) e “cansaço ao falar” (n=39). Isso sugere que o sexo biológico feminino esteve mais propenso a apresentar alteração de voz e cansaço ao falar que o sexo masculino. Mas, para chegar a essa conclusão, é preciso outras análises (esses detalhes serão apresentados no próximo capítulo).

A variável “faixa etária” também apresentou associação significativa com alguns hábitos vocais autorreferidos pelos professores, de tal modo que aquelas pessoas com idade igual ou superior a 42 anos, estiveram associadas com “cansaço ao falar” (0,012) e “garganta seca” (0,033). Para finalizar, note que em nenhuma outra situação o teste qui-quadrado verificou associação significativa entre as variáveis apresentadas na Figura 3.

Em síntese, a análise inferencial envolve uma variedade de métodos e materiais, mas também um tratamento criterioso capaz de extrair conclusões importantes que possibilitam, por exemplo, o aprimoramento contínuo de um sistema de ensino.

Embora a utilização da análise inferencial nas pesquisas educacionais seja uma ferramenta valiosa, existem limitações que demandam cuidado. É necessário considerar o amplo contexto educativo para não simplificar a complexidade inerente à essa área de conhecimento. Outra questão relevante é a presença de variáveis não observadas ou desconhecidas que podem influenciar nos resultados, exigindo uma abordagem cautelosa ao interpretar as conclusões.

CONCLUSÃO

Este estudo explorou duas pesquisas na área da educação que usaram a análise inferencial de dados para apresentar seus resultados. No primeiro artigo, foi utilizado um teste de comparação entre dois grupos independentes (teste-t de *Student*) e um teste de associação (correlação de Pearson). Já no segundo artigo, foi utilizado o teste de associação para variáveis categóricas (teste qui-quadrado). Além disso, foram descritas as limitações e a importância da análise inferencial no contexto educacional.

DECLARAÇÃO DE INTERESSES

Nós, autores deste artigo, declaramos que não possuímos conflitos de interesses de ordem financeira, comercial, político, acadêmico e pessoal.

REFERÊNCIAS

CAPOROSSO, C.; FERREIRA, L. P. Sintomas vocais e fatores relativos ao estilo de vida em professores. **Revista CEFAC**, v. 13, p. 132-139, 2011.

FIALHO, I.; CID, M.; COPPI, M. Vantagens e dificuldades na utilização de plataformas e tecnologias digitais por professores e alunos. **Revista Brasileira de Educação**, v. 28, p. e280050, 2023.

FIELD, A. **Descobrendo a estatística usando o SPSS**. 2. ed. Porto Alegre: Artmed, 2009.

GATTI, B. A. **A construção da pesquisa em educação no Brasil**. Brasília: Líber Livro, 2007.

GATTI, B. A. Estudos quantitativos em educação. **Educação e Pesquisa**, v. 30, n. 1, p. 11-30, 2004.

SOUZA, K. R.; KERBAUY, M. T. M. Abordagem quanti-qualitativa: superação da dicotomia quantitativa-qualitativa na pesquisa em educação. **Educação e Filosofia**, v. 31, n. 61, p. 21–44, 2017.

THOMAS, J. R.; NELSON, J. K.; SILVERMAN, S. J. **Métodos de pesquisa em atividade física**. 6. ed. Porto Alegre: Artmed, 2012.