



TESTES DE NORMALIDADE EM ANÁLISES ESTATÍSTICAS: UMA ORIENTAÇÃO PARA PRATICANTES EM CIÊNCIAS DA SAÚDE E ATIVIDADE FÍSICA

Dahan da Cunha Nascimento
Ramires Alsamir Tibana
Gislane Ferreira de Melo
Jonato Prestes

Universidade Católica de Brasília – Brasil

Resumo: Aprender a aplicar as análises estatísticas em resultados de estudos experimentais leva tempo, maturidade e erros com o processo. Todavia, o entendimento da estatística para seus praticantes pode ser um processo duradouro e até mesmo amedrontador. Grande parte do temor associado envolve o entendimento de que a estatística está perdida em meio às fórmulas matemáticas. No entanto, o desentendimento sobre os propósitos e aplicações das análises estatísticas geram falhas que regularmente são encontradas nas análises de estudos publicados. Ademais, devemos considerar que previsões inadequadas devido ao uso inapropriado dessas análises comprometem negativamente a aplicabilidade clínica e prática dos resultados. Devemos indagar por que profissionais nas áreas das ciências da saúde e atividade física precisam saber estatística, o quanto sabemos, o quanto precisamos e como adquirimos conhecimentos estatísticos. Por isso, o objetivo deste estudo é de contribuir com os praticantes em estatística nas áreas de ciências da saúde e atividade física para a compreensão das limitações de técnicas estatísticas com ênfase nos testes de normalidade comumente utilizados. Além disso, entender suas desvantagens, quando utilizá-los e promover a reflexão sobre o uso cuidadoso, dever social, profissional e ético de praticantes em estatística envolvidos em análises de dados nas áreas de ciências da saúde e atividade física.

Palavras-chave: estatística; testes; ciências da saúde.

INTRODUÇÃO

Para grande parte dos estudantes em ciências da saúde e atividade física o entendimento da estatística é um processo duradouro e até mesmo amedrontador. Parte do temor associado envolve o entendimento de que a estatística está perdida em meio às fórmulas matemáticas, e o mais importante é que os estudantes não entendem a relevância dessa matéria e não apresentam interesse (ALTMAN; BLAND, 1991).

Percebemos que existe uma grande falta de entendimento sobre os propósitos e aplicações das análises estatísticas e, por consequência, falhas nas análises de estudos publicados. Em uma revisão de literatura elegante sobre erros estatísticos em pesquisas médicas realizada por Strasak et al. (2007), verificou-se que o uso errado de testes estatísticos e da inflação do erro tipo I (concluir que existe uma diferença quando na verdade não existe) são falhas e deficiências comumente encontradas. Estatísticos estipulam que cerca de

50% dos estudos apresentam erros estatísticos (ALTMAN; BLAND, 1991). Um dado alarmante, pois previsões inadequadas devido ao uso inapropriado de análises estatísticas acarretam negativamente a aplicabilidade clínica e prática dos resultados, e são essas pesquisas duvidosas que contribuem para a base de conhecimento que afetam a forma que o paciente/aluno/cliente é tratado e/ou diagnosticado.

Devemos indagar por que profissionais nas áreas das ciências da saúde e atividade física precisam saber estatística, o quanto sabemos, o quanto precisamos e como adquirimos conhecimentos estatísticos. Dentre as principais habilidades que um profissional da área da saúde deve ter é analisar criticamente as contribuições de estudos para suas respectivas áreas de atuação.

Assim, o propósito deste texto é de contribuir para que os praticantes em estatística nas áreas de ciências da saúde e atividade física tenham compreensão das limitações de técnicas estatísticas com ênfase nos testes de normalidade comumente utilizados, pois os mesmos testes nos indicam qual tipo de tratamento estatístico (testes paramétricos ou não paramétricos) será aplicado. Ademais, entender suas desvantagens, quando utilizá-las e, por último, criticar construtivamente o uso dessas análises.

SUPOSIÇÕES PARA O USO DE TESTES ESTATÍSTICOS

Muitos dos procedimentos estatísticos são baseados em suposições para sua aplicação (FIELD, 2009). Dentre as suposições, podemos citar a homogeneidade de variância: essa hipótese significa que as variâncias devem ser as mesmas para as diferentes populações consideradas. Os grupos devem proceder de uma população que segue a lei normal: essa hipótese significa que a população é distribuída normalmente. Dados em uma escala de intervalo ou de razão: isso significa que a distância entre os pontos de sua escala deve ser igual em todas as partes ao longo da escala e independência, em que o comportamento de um participante não influencia no comportamento de outro.

VERIFICAÇÃO DA NORMALIDADE DOS DADOS

A normalidade dos dados é uma das suposições frequentemente utilizada para determinar que tipo de teste estatístico será empregado e muitos dos procedimentos estatísticos são testes paramétricos, os quais requerem que os dados sejam retirados de uma população normalmente distribuída.

Se quisermos que nossas conclusões sejam precisas, o tipo de teste estatístico que utilizaremos deve ter uma boa representação para os dados que coletamos. Portanto, a aderência do modelo ou o grau com que o modelo estatístico representa os dados deve ser adequado.

De acordo com Field (2009), se você utiliza um teste paramétrico quando seus dados não são paramétricos (distribuição não normal), os resultados talvez não sejam apropriados. Dessa forma, as nossas conclusões quanto à aplicabilidade prática/clínica dos dados podem ser imprecisas ou até mesmo catastróficas.

TESTES DE NORMALIDADE

Dentre as opções para se verificar se uma distribuição é normal, o histograma, *box plot* e o gráfico de dispersão *normal Q-Q plot* são normalmente usados, no entanto, eles não informam se a distribuição está próxima suficiente da normalidade (FIELD, 2009; BARROS et al., 2012). Portanto, para dar suporte às análises gráficas, métodos objetivos são necessários para determinar se uma distribuição é ou não normal.

Dentre os testes comumente utilizados, os de *Shapiro-Wilk* (SW) e *Kolmogorov-Smirnov* (KS) estão disponíveis em programas estatísticos como o *Statistical Package for Social Sciences* (SPSS), apesar de existirem na literatura mais de 40 testes de normalidade (RAZALI; WAH, 2011).

Qual o número necessário de participantes para aplicar os testes de normalidade? Essa foi uma das principais perguntas que me fiz até este momento. Ouvia de professores experientes que a quantidade de participantes é o que determinava a aplicabilidade do teste e seu poder. Dentre as respostas dadas a mim, o ponto de corte de 50 era o número que eu apoiava acriticamente até o presente momento.

Dentre as recomendações que me foram apresentadas, o teste de SW deveria ser aplicado para uma amostra igual ou inferior a 50 e que acima disso o KS era a melhor opção.

Recentemente, entendi que a sugestão dos professores tinha uma fundamentação teórica. O ponto de corte ($N = 50$) foi citado recentemente no livro do professor Barros et al. (2012), mas sem a referência original dessa citação. No entanto, outros pontos de corte como 30 (sem referência) também encontrei (BISQUERRA; SARRIERA; MARTINEZ, 2004) e me foram sugeridos. Diante do exposto, fiquei com dúvidas quanto à aplicação desses testes já que o número de participantes influencia sua aplicabilidade. Para melhor entender o que será comentado posteriormente, preciso falar brevemente sobre hipóteses e erros.

Quando realizamos uma pesquisa, normalmente criamos hipóteses que podem ser confirmadas ou rejeitadas. A probabilidade que calculamos nos testes estatísticos é baseada na suposição de que não existe relacionamento ou diferenças entre os tratamentos ou variáveis. Essa suposição é a hipótese nula (H_0). Se a hipótese da pesquisa estabelece que haverá relacionamento ou diferenças entre os tratamentos (normalmente chamada de hipótese experimental ou alternativa), então a hipótese nula estabelece que não haverá relacionamento ou diferença (DANCEY; REIDY, 2011). Contudo, podemos ter cometido dois tipos de erros em nossas previsões. Para H_0 concluímos que não houve diferenças entre os tratamentos quando na verdade houve (erro tipo II) ou concluímos que houve diferenças entre os tratamentos quando na verdade não houve (erro tipo I). Agora sim posso comentar sobre outro termo importante que é poder ou potência de um teste.

O poder significa encontrar uma diferença quando na verdade ela existe. Extrapolando essa definição para os testes de normalidade, se utilizarmos o teste inadequado para uma quantidade de observações (exemplo: $N < 30$) insuficientes, concluiremos que os dados são normalmente distribuídos quando na verdade não são (erro tipo I).

Então, incorporo neste texto informações sobre o melhor teste de normalidade e qual o número de observações (N) necessárias para que os testes de normalidade apresentem poder ou potência suficientemente confiáveis para identificar se um dado é normalmente distribuído quando na verdade é realmente normalmente distribuído.

O teste de SW foi originalmente feito para amostras entre 3 a 50 (SHAPIRO; WILK, 1965), no entanto, eles não o estenderam para amostras acima de 50; estudos entre 2003 e 2011 encontraram resultados conflitantes e preocupantes.

O mais recente estudo, feito por Ahad et al. (2011), verificou a sensibilidade (capacidade do teste de identificar corretamente dados não normalmente ou normalmente distribuídos) do teste de SW, KS, *Anderson-Darling* (AD) e *Cramer-von* (CV) em dados propositalmente não normalmente distribuídos em diferentes tamanhos amostrais (39 a 201). O trabalho concluiu que o teste de SW é melhor que o KS, pois rejeita a hipótese nula de normalidade em amostras pequenas (N maior ou igual a 39), enquanto o teste de KS necessita de uma amostra maior (N maior ou igual a 77) para indicar que um dado não é normal. E recomendou aos praticantes em estatística o uso do SW por apresentar maior sensibilidade e demonstrou a deficiência do teste de KS entre outros analisados.

Razali e Wah (2011) corroboraram os resultados anteriormente citados. Nesse estudo, eles compararam os poderes de testes de normalidade SW, KS, AD e *Lilliefors* (LF). Os poderes dos testes analisados nesse estudo variaram de acordo com o tamanho da amostra (10 a 2000). Para uma amostra menor ou igual a 50 com um nível de significância de 5%, para todos os testes de normalidade analisados o poder ficou abaixo de 40%. Somente para uma amostra mínima de 50, o teste de SW quando comparado com os demais atingiu poder suficiente de indicar que um dado não é normal. Considerou-se então o teste de SW como o de maior

potência para a análise de normalidade quando comparado com os demais. Além disso, alertou aos praticantes de estatísticas para tomarem cuidado ao utilizarem esses testes de normalidade em amostras com número baixo de observações (N) por causa da falta de poder e inflação do erro tipo I. Outro estudo corrobora as mesmas conclusões de que o teste de SW tem maior potência na análise da normalidade dos dados e que sua utilização é fortemente recomendada (MENDES; PALA, 2003)

Com isso, através dos achados, sugiro que praticantes em estatística não dependam somente de gráficos para a análise da normalidade dos dados, estes devem ser combinados com testes objetivos de normalidade, como o de SW. Contudo peço atenção quando for aplicado esse tipo de teste em uma amostra pequena, pois o poder e inflação do erro tipo I diminui e aumenta, respectivamente.

CONSIDERAÇÕES FINAIS

Portanto, com base na leitura desses artigos. Convido a reflexão de praticantes em estatística e estudantes nas áreas de ciências da saúde e atividade física sobre os cuidados na aplicabilidade dos testes de normalidade em amostras pequenas.

O uso cuidadoso e correto da estatística nas áreas de ciências da saúde e da atividade física é de grande importância e deve ser repensado continuamente. Principalmente por considerar a realidade da maioria dos pesquisadores em que tanto a pesquisa e a análise dos dados são feitas por conta própria. Por isso, o conhecimento profundo ou adequado da estatística é desejável, especialmente se assistência estatística não está disponível (ALTMAN; BLAND, 1991).

É difícil demonstrar evidências de que o uso errado da estatística tem consequências adversas na prática clínica. No entanto, devemos lembrar que erros estatísticos podem acarretar conclusões incorretas na vida e no tratamento de alunos/pacientes/clientes. Com isso, praticantes de estatística têm a obrigação social de realizar seu trabalho de forma profissional e de maneira ética, e o trabalho conjunto de estatísticos e praticantes nas áreas de ciências da saúde e atividade física tem de ser continuamente reforçado.

Precisamos dos estatísticos, assim como precisamos estar cientes dos problemas relacionados à estatística.

NORMALITY TESTS FOR STATISTICAL ANALYSIS: A GUIDE FOR PRACTITIONERS IN HEALTH SCIENCES AND PHYSICAL ACTIVITY

Abstract: The understanding of statistics to its practitioners may take a life time and might be a frightening process. Much of the associated fear involves the understanding that the statistic is lost under mathematical formulas. However, the misunderstanding of the purposes and applications of statistical analyzes generate pitfalls that are regularly found in the analysis of published studies. Furthermore, we consider that incorrect statements due to inappropriate use of these analyzes imply a negative clinical and practical applicability of the results. Therefore, the aim of this study is to contribute to practitioners in statistical sciences in the areas of health and physical activity on the understanding of statistical techniques with emphasis on normality tests commonly used. Also, understand its disadvantages, when to use and encourage continuous reflection on the careful use, social, professional and ethical duty for the practitioners involved in statistical

data analysis in the areas of health sciences and physical activity.

Keywords: statistics; tests; health sciences.

REFERÊNCIAS

- AHAD, N.A. et al. Sensitivity of normality tests to non-normal data. **Sains Malaysiana**, Malaysia, v. 46, n. 6, p. 637-641, 2011.
- ALTMAN, D. G.; BLAND, J. Martin. Improving doctors' understanding of statistics. **Journal of the Royal Statistical Society**. Series A (Statistics in Society), New York, v. 154, n. 2, p. 223-267, 1991.
- BARROS, M.V. G. et al. **Análise de dados em saúde**. Londrina: Midiograf, 2012.
- BISQUERRA, R.; SARRIERA, J. C.; MARTÍNEZ, F. **Introdução à estatística**: enfoque informático com o pacote estatístico SPSS. Porto Alegre: Artmed, 2004.
- DANCEY, C. P.; REIDY, J. **Statistics without maths for psychology**. 5th ed. London: Pearson Education Limited, 2011.
- FIELD, A. **Descobrimo a estatística usando o SPSS**. São Paulo: Artmed, 2009.
- MENDES, M.; PALA, A. Type I error rate and power of three normality tests. **Pakistan Journal of Information Technology**, Karachi, v. 2, n. 2, p. 135-139, 2003.
- RAZALI, N. M.; WAH, Y. B. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. **Journal of Statistical Modeling and Analytics**, Malaysia, v. 2, n. 1, p. 21-23, 2011.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Bimetrika**, Oxford, Inglaterra, v. 52, n. 3-4, p. 591-611, 1965.
- STRASAK, A. M. et al. Statistical errors in medical research – a review of common pitfalls. **Swiss Med Wkly**, Zurich, v. 137, n. 3-4, p. 44-49, 2007.

Contato

Dahan da Cunha Nascimento
E-mail: dahanc@hotmail.com

Tramitação

Recebido em 21 de janeiro de 2014
Aceito em 22 de julho de 2014