

# TÓPICOS PARA UMA BOA ANÁLISE DE DADOS DE ESTUDOS EMPÍRICOS<sup>★</sup>

<http://dx.doi.org/10.1590/1984-0292/1136>

*Teresa Garcia-Marques*<sup>★★</sup>

*Instituto Superior de Psicologia Aplicada, Lisboa, Portugal*

## RESUMO

*Este artigo discute o processo de análise de dados em sete tópicos: a) modelos estatísticos, técnicas estatísticas e diferenciação de “receita estatística”; b) a associação da validade das conclusões estatísticas ao processo de delineamento e amostragem; c) a diferença e complementaridade da descrição de dados e do teste de hipóteses; d) como as hipóteses se relacionam com os testes estatísticos a realizar; e) a análise preliminar de dados e sua relevância; f) teste de hipótese na perspectiva de Fisher e de Neyman e Pearson; g) a apresentação dos dados como um processo comunicacional, e vantagens do uso de normas (APA).*

*Palavras-chave: análise de dados; estatística; teste de hipóteses; comunicação de dados*

## TOPIC FOR A GOOD ANALYSIS OF EMPIRICAL STUDIES' DATA

### ABSTRACT

*This paper discusses the data analysis process around seven critical topics: a) Distinguishing between a statistical model, a statistical technique and a statistical receipt; b) Sampling and Design's implications for statistical validity; c) Complementarities and distinctions between data description and hypothesis testing; d) How statistical hypothesis are framed by the statistical model and vice versa; e) relevance of preliminary analysis of data; f) Fisher's and Neyman-Pearson's perspectives of statistical inference; g) Data presentation as a communication matter.*

*Keywords: data analysis; statistic; hypothesis testing; data communication*

---

<sup>★</sup>Apoio e financiamento: FCT- Fundação para a Ciência e Tecnologia, Portugal

<sup>★★</sup>Morada: Rua Jardim do Tabaco, nº34. 1149-041 Lisboa – Portugal. E-mail: [gmarques@ispa.pt](mailto:gmarques@ispa.pt)

Todo o investigador de forma direta ou indireta já se colocou a si próprio a questão: O que é uma boa análise de dados?

Responder a esta questão é como responder à questão “o que é uma boa concepção de um dado prato de culinária?” Todos sabemos que existem muitas variações na forma como concebemos uma receita. E para além da variedade de receitas, existem diferentes cozinheiros a conseguir tirar melhor ou pior partido dos ingredientes que têm à sua disposição. Assim acontece com a análise de dados!

Neste artigo, em vez de fornecer uma boa receita para uma análise de dados, apresento o que são os ingredientes necessários à sua realização, a forma de maximizar a sua funcionalidade e, principalmente, aspectos a serem evitados no processo que a sustenta. Cabe a cada investigador tornar-se um “bom cozinheiro”. Este pode fazê-lo através de desenvolvimento profundo dos modelos estatísticos que utiliza na sua análise de dados e através da observação do modo como os peritos no campo analisam os seus dados (reportada em artigos científicos de revistas de qualidade). Apresento todos estes ingredientes em formato de tópicos de conhecimento, os quais deverão ser dominados por aquele que pretende analisar dados.

### **TÓPICO 1: DIFERENCIANDO MODELOS ESTATÍSTICOS, TÉCNICAS ESTATÍSTICAS E “RECEITAS ESTATÍSTICAS”**

A primeira ideia daquele que analisa dados é: qual é o instrumento a utilizar. O objetivo de uma abordagem estatística, desenvolvida pelos teóricos da área é desenvolver “modelos”. O estatístico vê o mundo que nos rodeia definido em “constantes” e “variáveis”. Apaixona-se por estas últimas, procurando perceber o padrão a que obedecem suas alterações. Assim, percebe que cada variável tem uma distribuição específica e caracteriza-a. Depois percebe que nem todas essas variações ocorrem de forma independente das características do seu contexto (outras variáveis). Assim, umas variáveis “ao variarem introduzem algum grau de variação nas outras”. Procura então perceber as regras subjacentes a estas covariações (se existir interdependência) ou a estes efeitos (se a dependência for unidirecional). Ao estudar a variabilidade e as suas regularidades, o estatístico apercebe-se da identidade de cada variável, e que características (parâmetros) das mesmas podem ser utilizadas para as descrever. Uma variável não tem cheiro nem cor, mas tem, por exemplo, um ponto que centra a sua variação, e um grau de dispersão dos valores relativamente a esse ponto, entre outras propriedades.

Kurt Lewin (1952, p. 169) proferiu certa vez : “*There is nothing more practical than a good theory*”.<sup>1</sup> Esta citação tem especial relevância para a abordagem estatística. As abordagens teóricas dos estatísticos são as que mais rapidamente são utilizadas como instrumentos nas nossas análises de dados. Usamos os parâmetros que os estatísticos detectaram para descrever uns aos outros a natureza das nossas variáveis. Utilizamos as estimativas que os estatísticos nos fornecem sobre o quanto a amostra que obtivemos da variável pode induzir em erro sobre o que são realmente as suas características (teoria de amostragem). Usamos os padrões identificados de covariação para ver se “esses modelos” se adequam aos dados. Postulamos atributos sobre os parâmetros e covariações das nossas variáveis, e es-

timamos a probabilidade com que acertamos essas previsões (abordagem de teste de hipóteses). Depois utilizamos esses dados como argumentos para referir a validade das afirmações que queremos fazer sobre os nossos dados (análise de dados).

O fato de fazermos do pensamento estatístico e seus conhecimentos e teorias instrumentos práticos, faz com que chamemos a alguns deles “técnicas”. Ora, como qualquer técnica, os modelos estatísticos existentes ajudam a alcançar os objetivos. Apenas há de se estar bem ciente que o uso da técnica por si não garante a validade das conclusões. Assim, ficará claro para todos os que lêem este texto que a facilitação do uso da técnica por meio de *softwares* estatísticos, pode levar a que se confunda o processo de “análise de dados” com o fazer correr um programa sobre os dados. Para recuperar a metáfora da receita, isso assemelha-se a colocar os ingredientes disponíveis numa batedeira, depois no forno, e pensar é garantia de um belo bolo. Basta a simples confusão que qualquer extraterrestre – como o são alguns dos utilizadores da estatística – pode fazer entre o pó (farinha) e o fermento para perceber que o resultado será desastroso.

Tentando evitar um mau uso da estatística e facilitar a vida de quem pretende analisar dados, alguns autores têm definido “receitas” ou “princípios de culinária” que se, seguidos à risca, garantem a validade das conclusões estatísticas. Na realidade, uma boa receita, e principalmente se colocada num robô de culinária (como são os *softwares* de estatística), previne erros graves e garante um resultado que não envergonha nenhuma mesa. O problema surge quando essas “receitas” passam a ser interpretadas como “o que se ‘deve’ fazer”. Se houvesse um modo único de cozinhar uma receita não haveria bons e maus cozinheiros. Muitos textos estatísticos e muitos especialistas defendem uma abordagem fundamentalista: “com estes dados a análise tem de ser esta”, ou “se o valor não atinge este patamar, então os dados ‘nada valem’”. Em boa verdade, onde há fundamentalistas, há também os ultraliberais, que indiscriminadamente fazem uso de qualquer instrumento e dados para sustentar qualquer argumento. Desde que seja um gráfico, ele representa os dados, o que é confundir o uso da técnica com a análise de dados.

Em suma, uma análise de dados deve perceber os modelos desenvolvidos que sustentam as técnicas a utilizar. Só esse conhecimento garante que se está a realizar uma verdadeira análise de dados, com vista a perceber a validade dos argumentos que irão sustentar as afirmações que se pretendem realizar sobre esses dados.

## **TÓPICO 2: A VALIDADE DAS CONCLUSÕES ESTATÍSTICAS PASSA PELO PROCESSO E DELINEAMENTO DE AMOSTRAGEM DOS DADOS**

Tal como a nossa ida ao supermercado determina as características de preparação de uma refeição, também a forma como se planeja a aquisição dos dados; determina as características da sua análise.

O processo de amostragem de dados é definido, em parte, pelo delineamento do estudo. Registam-se dados de forma aleatória de uma população ou restringindo (pré-definindo) a amostra de onde são retirados (amostragem estrati-

ficada). Esse processo de amostragem associado à natureza da variável mensurada; é passível de ser modelado estatisticamente (ver, por exemplo, a abordagem dos *General Linear Models*), e, com base nesse modelo colocar várias questões relativas aos padrões passíveis de serem detetados na variabilidade observada. Deste modo, a análise a ser efetuada é determinada, em parte, pelo delineamento do estudo, e noutra parte pela natureza distribucional das variáveis. Alguns exemplos: as Anovas estão associadas a delineamentos fatoriais; a regressão a delineamentos correlacionais; as tabelas de contingência e os modelos *log*-lineares a delineamentos com variáveis discretas, etc. Um bom cozinheiro planeja a ementa antes de se dirigir ao mercado. Caso não aconteça: falta algum dos ingredientes necessários e ele não pode confeccionar o prato almejado.

Concluindo: tal como um bom cozinheiro sabe o que pretende de cada um dos ingredientes à partida, também o bom investigador deve saber as questões empíricas que quer ver respondidas com a sua análise de dados antes de nela se envolver. Existe uma relação direta entre as hipóteses que se pretendem testar, o delineamento dos dados recolhidos e a sua análise estatística. Nota-se, todavia, que no processo de análise de dados há espaço para a criatividade e para novos questionamentos. Mas se fizermos disso regra, podemos simplesmente não ter uma refeição a servir aos nossos convidados.

### TÓPICO 3. DESCREVER DADOS E TESTAR HIPÓTESES

O propósito da análise de dados é obter respostas a questões. Se as questões são vagas não podemos obter respostas diretas, pelo que devemos apenas descrever cuidadosamente os dados com vista a obter informação sobre os mesmos e suas relações. Assim, descrevemos cada uma das nossas variáveis nas suas características distribucionais ( médias, desvios padrões, frequências etc.), nas suas relações binárias ( correlações) ou múltiplas ( correlações múltiplas) e nas suas distribuições condicionadas ( quando as características dessas relações são diversas em diferentes categorias). É como se tivéssemos acabado de cozinhar e apresentássemos todos os nossos produtos aos nossos convidados, com o objetivo de perceberem a natureza da refeição. Respondemos a qualquer questão relativa a cada confecção: se é doce, salgado, se tem pimenta, ou não etc. Até podemos sugerir que uns pratos são bons e outros nem tanto assim, que há mais vantagens em experimentar um do que o outro, por exemplo. Porém, o mais importante será provarmos cada sabor, de forma a podermos afirmar com segurança a qualidade destes dados. É aqui que entra a noção de “prova estatística”, de “teste estatístico”.

O teste estatístico apenas ocorre quando *a priori* o investigador prevê os seus resultados. Pelo menos, ele espera-os e coloca-os como hipótese. Ele quer afirmar com validade, “esta hipótese está correcta”, e dispõe-se a apresentar “prova” disso. Essa prova é um argumento construído em duas fases: a primeira é aquela em que perguntamos a todos os que nos rodeiam “se nesta condição eu obtiver estes dados, isso significa que eu tenho razão”? Consensualmente temos

de perceber o delineamento de um estudo como capaz de nos fornecer essa prova; a segunda fase é aquela em que afirmamos: “Eu obtive os dados que disse que ia obter e (aqui entra a análise de dados) eles não se devem ao acaso”.

Para poder referir que obtivemos os dados esperados pela nossa hipótese, e que tal não se deve a um acaso, recorremos a um “teste de hipóteses”. No entanto, é preciso perceber que nós não concluímos se os dados que observamos se adequam às nossas hipóteses ou não com base na estatística. São os nossos olhos que o afirmam ou negam ao observar os sumários estatísticos das nossas variáveis e suas relações. É uma média maior do que outra? Se o valor dela for maior do que a da outra, é claro que é! Se for menor, claro que não o é! O que fica por saber é se, o que nos aparenta aos olhos é verdade ou poderá ser fruto do acaso. Para afirmar que não é acaso, temos de obter informação que nos sirva como argumento para o sustentar. O argumento que é mais comumente usado, passa por estimar a probabilidade com que o efeito se esperaria por mero acaso. O argumento consiste em afirmar que essa probabilidade (valor  $p$ ) é muito pequena. “Bem, se é assim tão pequena, não deve ser acaso”, concluímos todos em nossas mentes.

Existem outros argumentos à nossa disposição para além do baseado no valor  $p$  e dos quais devemos fazer uso sempre que precisemos de sustentar a nossa crença na validade da observação que fazemos sobre os nossos dados (mesmo tendo sido alguns deles gerados, por vezes, como alternativas ao  $p$ , como é referido no ponto 5).

Um desses argumentos incide sobre as propriedades dos nossos, definidas no cálculo de intervalos de confiança dos parâmetros das variáveis e da relação entre elas. O significado de qualquer parâmetro como uma média ou uma proporção, pode diferir imensamente quando integrado no seu intervalo de confiança. É muito diferente dizer que a proporção de votos que se infere para a população, a partir de uma amostra é de 55%, quando esse valor se associa a um intervalo de ]48% e 62%[ ou a um intervalo de ]52% e 57%[. Pelo fato de o intervalo de confiança nos reportar para a variabilidade que atribuímos ao acaso – e por isso é expectável que ocorra independentemente de qualquer acontecimento – os intervalos de confiança para os efeitos concorrem com o valor  $p$  no quão bem nos informam sobre os resultados? serem apenas fruto do acaso (ver THOMPSON, 2002, para uma revisão).

Outro argumento é construído com base no índice de magnitude de efeito. Este informa-nos de quão significativo = relevante = importante = grande é o efeito observado (e, já visto, não resultar por mero acaso). Note-se que, esta informação é muitas vezes erradamente retirada do valor  $p$ , que apenas nos fala da significância estatística (não ser acaso) e não da significância prática. Existem vários índices que procuram comunicar essa grandeza e fazem-no em diferentes escalas. O índice específico de magnitude de efeito a usar, deve ter em conta: a) a tradição da revista a que se submete o artigo; b) a tradição do campo em que se realizam os estudos; c) a vantagem que as medidas estandardizadas têm relativamente às não estandardizadas, por permitir a comparação inter-estudos; e d)

a vantagem de usar medidas que traduzam diretamente a proporção de variabilidade explicada pelo efeito, por permitir igualmente as comparações com dados associados a diferente tipo de delineamentos.

Quando a análise do valor  $p$ , nos sugere a não existência de diferenças entre dois valores, ou significância de outro tipo de relação, podemos ser tentados a aceitar a “não diferença” como sendo suportada pelos nossos dados. Há, porém, um forte argumento contra essa conclusão: é que a aparente ausência de efeito pode estar associada à instabilidade dos nossos dados (ser então fruto do acaso) sendo o efeito obscurecido pela sua variabilidade. Essa variabilidade associada ao acaso tende a ser maior em pequenas amostras. E as amostras são sempre pequenas relativamente à população. Assim, há sempre a possibilidade de se confundir um efeito pequeno com “acaso”. Como contra-argumento pode apresentar-se a “potência” da análise estatística. O cálculo do “poder de teste” (ver COHEN, 1988), permite perceber a possibilidade de se cometer um erro de Tipo II, isto é rejeitar um efeito quando ele está presente”. Se a potência do teste é elevada, permitindo detetar um valor de uma grandeza  $x$ , pode-se dizer com segurança que a diferença observada não atinge esse valor, caso contrário teria sido detectada na análise (ver por exemplo, AZEVEDO; GARCIA-MARQUES, 1996).

Em suma: uma análise de dados pretende obter respostas a questões. Mesmo que algumas destas respostas possam ser alcançadas por uma mera descrição de dados, é importante perceber se a descrição do que se verifica na nossa amostra se generaliza à população. A estatística inferencial suporta os argumentos de generalização, por sugerir que o observado não é fruto do mero acaso. Para tal atendemos ao valor  $p$  das análises estatísticas. Porém, há ter em conta que este não informa, nem sobre a magnitude dos efeitos observados, nem sobre a confiança que podemos ter na ausência de efeitos. Estas duas informações devem ser alcançadas por índices de magnitude de efeito e de potência de teste.

#### **TÓPICO 4. TESTAR O MODELO VERSUS USAR O MODELO PARA TESTAR HIPÓTESES**

Algumas análises estatísticas parecem receitas culinárias que misturam todos os ingredientes, esperando que dessa mistura surja “uma boa refeição”. É feita a mesma análise estatística quando a hipótese em estudo é a de que “dos 5 grupos experimentais de homens apenas um vai diferir de todos os outros, e dos 5 grupos de mulheres” ou “todos os grupos experimentais diferem uns dos outros, tanto para homens como para mulheres”. É feita a mesma análise, porque esta é determinada pelo modelo estatístico associado ao delineamento do estudo e natureza da variável de medida. Assim, tendemos a simplesmente usar o modelo global e calcular todos os seus parâmetros e sua significância estatística “por defeito”. Assim, quando em presença de uma variável dependente com natureza contínua, e um delineamento fatorial  $2 \times 5$  (descrito acima), o investigador recorrer a uma Anova de dois fatores e testar todos os efeitos a ela associados. Ele, por defeito, testa o modelo completo, analisando dois efeitos principais e uma interação. Mas a verdade é que ao proceder a esta análise global, o investigador está a ignorar a

sua hipótese que tem uma natureza muito mais específica. No nosso primeiro exemplo a hipótese assume que todas as células do delineamento são iguais com exceção de uma. O teste direto desta hipótese deverá ser feito dentro do modelo Anova global mas em forma de “contraste”. O contraste para esta hipótese define a comparação de uma célula com todas as outras 9. Mas não devemos ficar pela realização deste contraste e teste da sua significância estatística. Ao concluir pela sua significância, devemos, a seguir, procurar perceber se essa diferença explica toda a variabilidade observada nos dados (ver por exemplo THOMPSON, 1994, sobre análise de contrastes planejados). Se o padrão de dados é o que postulamos, ele é a origem da variabilidade e por tal “explica toda essa variabilidade”. E esta seria a forma mais direta e indiscutível de testarmos a “nossa hipótese”.

A verdade é a de que tendemos a realizar análises globais aos dados mesmo quando delas não necessitamos. Estas análises globais têm, porém a vantagem de, quando publicadas, permitirem ao leitor testar as suas próprias hipóteses sobre os nossos dados. Esta pode ser uma das razões pelas quais os revisores insistem na sua presença. No entanto, convém ter o cuidado de nesse processo de comunicação científica não perdermos de vista os nossos objetivos. É que o efeito pretendido pode estar presente e não o detectarmos nessa análise global (ver GARCIA-MARQUES, 1997).

Outro cuidado a ter: não cair na tentação oposta de testar o papel do acaso em todas as comparações possíveis de fazer com os dados. De outro modo, não se devem testar todas as correlações entre todas as variáveis, só por mera curiosidade. E nem se deve comparar todas as médias entre si. A razão é simples: se um teste estatístico informa da probabilidade de se obter um efeito por mero acaso, um grande número de testes estatísticos aumenta a probabilidade de se obter um efeito por mero acaso. Pense só: se você estaria disposto a tentar a sua sorte com uma bala no cano de um revólver, disparando apenas uma vez ou disparando varias vezes. A probabilidade de se cometerem erros de Tipo I (rejeitar uma hipótese nula quando esta é verdadeira, usualmente definido pelo índice *alfa*), num teste entre vários realizados, aumenta numa base quase proporcional (para mais esclarecimentos sobre o processo de “inflação do *alfa*” (da probabilidade de cometer um erro Tipo I ( ver GARCIA-MARQUES; AZEVEDO, 1995). A regra de algibeira a usar nestas situações é procurar obter de uma resposta a partir dos dados com um número de testes que não ultrapasse o número de graus de liberdade associado ao delineamento do estudo (THOMPSON, 1994).

## **TÓPICO 5: A CORRETA UTILIZAÇÃO DOS MODELOS ESTATÍSTICOS ASSOCIA-SE A UM CLARO CONHECIMENTO DAS VARIÁVEIS A MODELAR**

Daniel Wright (2003, p. 134) refere “*Conducting data analysis is like drinking a fine wine. It is important to swirl and sniff the wine, to unpack the complex bouquet and to appreciate the experience. Gulping the wine doesn't work*”.<sup>2</sup>

Com uma analogia diferente da utilizada neste artigo, Daniel Wright chama a nossa atenção para o que Tukey definiu como “a análise preliminar dos dados”. A ideia é orientar o primeiro passo de uma análise de dados para o conhecimento

pessoal e detalhado de cada uma das variáveis em estudo. A principal relevância desta análise é perceber se os parâmetros com que descrevemos as variáveis são fidedignos, e não enviesados pela presença de *outliers* ( são dados que quebram o padrão observado em todos os dados, induzindo-nos a uma ideia errada do que é esse padrão). Apesar de não haver consenso no modo de lidar com *outliers* (BARNETT; LEWIS, 1994), existe consenso no fato de termos de lidar com eles, sendo esta uma medida preventiva dos enviesamentos na análise.

O segundo aspecto relevante nesta análise preliminar é perceber se é ou não possível modelarmos o padrão geral dos dados, com um dos modelos previamente estudados (*General Linear Models*, Regressão Logística etc). O que equivale a questionar sobre o modo como as variáveis seguem os pressupostos subjacentes a esses modelos. Se alguns autores referem que a maioria dos artigos ignora este aspecto (e.g. BRECKLER, 1990; KESELMAN et al., 1998; MICCERI, 1989) outros aconselham claramente que, quando uma variável não tem a esperada distribuição normal, se mude de análise ou se proceda a uma transformação com o objetivo dela se adequar a esse pressuposto (BEHRENS, 1997; FERKETICH; VERRAN, 1994).

## **TÓPICO 6: “UMA vs A” ABORDAGEM DE TESTE DE HIPÓTESES TEÓRICOS**

O utilizador dos modelos estatísticos com uma abordagem de teste de hipótese, muitas vezes ignora o fato de que os procedimentos que utiliza são eles próprios sujeitos a críticas de enviesamento das suas conclusões. Os próprios procedimentos estatísticos qualificam a validade dos argumentos que uma análise estatística nos oferece (ver GIGERENZER, 2004).

Claro que não cabe ao investigador que recorre à estatística para analisar os seus dados envolver-se em discussões sobre a validade de uma ou outra abordagem estatística. Regra geral o investigador acomoda-se à abordagem dominante. Podendo fazê-lo, será, no entanto, relevante perceber que “não há uma forma correcta de pensar a estatística inferencial e o teste de hipóteses”. O que fazemos ao seguir diretrizes é optar por uma ou outra abordagem como se fosse a “certa”.

A abordagem dominante na análise de dados tende a basear-se no cálculo de valores  $p$  e negligenciar a relevância da proposta Bayseana, que envolve o cálculo de probabilidades *a posteriori* associadas às nossas hipóteses e suas alternativas (HUBBARD; RYAN, 2000). Mas essa abordagem dominante é ela própria híbrida, sendo uma mistura de duas abordagens distintas: a de Fisher e a de Neyman e Pearson (N&P), que são oponentes (ver GIGERENZER, 1993 para um esclarecimento completo) e, se analisadas a fundo, nos induzem a seguir diferentes procedimentos. Tal como Fisher sugere, nós tendemos a referir o nível de significância como uma característica dos nossos dados, e congruentemente, segundo as normas da APA, reportamos o  $p = ^3$  a esse valor como representando a probabilidade de erros Tipo I (Fisher apenas contempla os erros de Tipo I e Ho). No entanto, utilizamos como regra *a priori*, valores de  $p = .05$  ou  $p = .001$ ,<sup>4</sup> usando-os como corte que definem a rejeição de uma hipótese  $H_0$  e aceitação da

hipótese alternativa H1, o que é uma abordagem de N&P. Além disso, não temos esses valores como a probabilidade de erros de Tipo I, como N&P recomendariam (e muito menos definimos *a priori* o valor dos erros Tipo II).

Não contemplar a existência destas duas abordagens e respectivas diferenças, explica porque razão alguns autores recomendam o cálculo cuidadoso do tamanho de uma amostra, enquanto outros não exibem tal preocupação. Com efeito, a adopção da visão de N&P exige tal cálculo, enquanto a concepção Fisheriana não faz disso uma exigência. A discrepância entre estas duas abordagens é igualmente a razão de ouvirmos, uns autores a referir uma decisão de aceitar ou rejeitar  $H_0$  com certos níveis de confiança (tal como sugerido por N&P), e outros, apenas a referir a aceitação de H1 quando  $p$  é de valor reduzido (como o recomenda Fisher).

As normas e recomendações das revistas, como as da APA, tendem a tentar agradar a gregos e troianos, requerendo uma abordagem híbrida (que alguns autores como GIGERENZER, 2004, referenciam como perigosa). Assim, para além de requererem valores exactos de “ $p$ ” (como Fisher recomenda) e lhes associar a noção de potência de teste (como N&P recomendam), tendem agora a requerer a apresentação de índices que se desenvolveram como abordagens alternativas ao simples teste de hipóteses com base nos valores de  $p$ : magnitude de efeitos (ver SCHMIDT, 1996) e intervalos de confiança (ver MAXWELL, 2004).

## **TÓPICO 7: A ARGUMENTAÇÃO ESTATÍSTICA É UM PROCESSO COMUNICACIONAL**

A apresentação dos dados é um processo comunicacional, que oferece um enquadramento ao processo de análise de dados. Existem regras gerais de Comunicação que devem ser mantidas neste processo, como a do “rigor” e a da “clareza”. Por rigor, entendemos a necessidade da apresentação de dados válidos, feita de modo a que não possa criar qualquer tipo de enviesamento na sua interpretação. A necessidade de clareza, visa facilitar a correcta compreensão das análises, garantindo que qualquer outro investigador as possa replicar. Pretende-se, deste modo, reduzir ao mínimo a existência de ambiguidades e maximizar a eficiência comunicacional.

Há duas outras direcções fundamentais para a apresentação dos nossos dados.

Para a primeira, retome-se a metáfora da culinária. Tal como num jantar se separam as entradas do prato principal, e se faz acompanhar um prato com uma salada, também na apresentação dos dados se deve seguir um guião previamente determinado. Os resultados são apresentados numa narrativa que pontua a razão pela qual se obteve um conjunto de dados, referindo a informação que dele se pretende retirar. Depois, explica-se que o padrão dos mesmos sugere uma resposta à questão de investigação, e acautela-se que tal não se deveu ao acaso. Assim, se submete essa resposta ao escrutínio do papel do acaso, com vista a concluir a sua significância estatística. A argumentação completa-se com a apresentação de índices de magnitude de efeito, análises *post-hoc*, contrastes específicos, etc. E arrematando-se informações obtidas por todas essas análises.

A segunda é a de que toda a argumentação estatística tem de estar ancorada nos valores numéricos, estatísticas, parâmetros, índices, que suportam os nossos argumentos. Todos estes dados visam permitir ao leitor verificar as nossas contas (para não haver algum engano). Se precisamos de apresentar as nossas variáveis aos leitores, definindo alguns dos seus parâmetros, devemos ser contidos e não sobrecarregar o leitor com informação que não é relevante para a compreensão do estudo e dos seus resultados. Instrumentos que podem aumentar a clareza são os sumários estatísticos em forma de tabelas e gráficos. Não podemos é cair na tentação de pensar que o seu uso nos garante a clareza. Existem boas e más tabelas, existem bons e maus gráficos. Há de se saber adequar o uso do gráfico aos dados e à informação que é por eles veiculada.

As Normas da APA são um bom guia ao iniciado que tem dúvidas sobre que dados que devem ser ou não ser apresentados com cada análise realizada (médias, desvios-padrão, correlações, estatísticas, graus de liberdade, valores de  $p$ , magnitude de efeito, MSEs, etc.)

### A TÍTULO DE CONCLUSÃO

Este texto procura enquadrar alguns dos principais tópicos subjacentes a uma análise de dados, esclarecendo-os de forma a evitarem-se alguns erros. O texto deixa bem claro que não há receitas de como proceder a uma boa análise de dados e que existe grande variedade na forma como diferentes autores encararam este processo. O paralelo com o papel do mestre de cozinha é fundamental para o leitor perceber que se existem livros de receitas, quem não os segue não incorre obrigatoriamente em erro. Mas se o leitor procura diretrizes mais concretas sobre todas estas questões, algo como um livro de Pantagruel, eles existem! Uma sugestão é a leitura do relatório da “*task force*” da *American Psychological Association* (APA) (WILKINSON; THE APA TASK FORCE ON STATISTICAL INFERENCE, 1999). Bom apetite!

### NOTAS

<sup>1</sup>“Não há nada mais prático que uma boa teoria”.

<sup>2</sup>“Analisar dados é como tomar um bom vinho. É importante cheirar e saborear o vinho, analisar o seu complexo *bouquet* e apreciar a experiência. Beber o vinho de uma só vez não funciona.”

<sup>3</sup>As normas da APA e de algumas outras publicações, recomendam a apresentação do valor  $p$  com três casas decimais. Apenas no caso em que essas casas referem o valor zero, é que é recomendado o uso de  $p < .001$ .

<sup>4</sup>Para uma discussão sobre o significado e relevância destes valores, ver o artigo de Rosnow e Rosenthal (1989, p. 1277 que refere que “*God loves 0.06 as much as 0.05*” – “Deus ama o valor 0.06 tanto quanto o valor 0.05”).

## REFERÊNCIAS

- AZEVEDO, M.; GARCIA-MARQUES, T. Que Confiança podemos ter nas conclusões estatísticas que apresentamos? *Psicologia*, Lisboa, v. 11, n. 1, p. 151-168, 1996.
- BARNETT, V.; LEWIS, T. *Outliers in Statistical Data*. 3rd ed. Chichester, England: John Wiley & Sons, 1994.
- BEHRENS, J. T. Principles and procedures of exploratory data analysis. *Psychological Methods*, [S.l.], v. 2, n. 2, p. 131-160, 1997.
- BRECKLER, S. J. Application of covariance structure modeling in psychology: cause for concern? *Psychological Bulletin*, [S.l.], v. 107, n. 2, p. 260-273, 1990.
- COHEN, J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum, 1988.
- FERKETICH, S.; VERRAN, J. An overview of data transformation. *Research in Nursing and Health*, [S.l.], v. 17, n. 5, p. 393-396, 1994.
- GARCIA-MARQUES, T. A HIPÓTESE de estudo determina a análise estatística: um exemplo com o modelo ANOVA. *Análise Psicológica*, [S.l.], v. 15 n. 1, p. 19-28, 1997.
- GARCIA-MARQUES, T.; AZEVEDO, M. A inferência estatística múltipla e os erros tipo I: um problema de inflação do alfa. *Psicologia*, [S.l.], n. 10, p. 1-11, 1995.
- GIGERENZER, G. The superego, the ego, and the id in statistical reasoning. In: KEREN, G.; LEWIS, C. (Ed.). *A handbook for data analysis in the behavioral sciences: methodological issues*. Hillsdale, NJ: Erlbaum, 1993. p. 311-339.
- GIGERENZER, G. Mindless statistics. *The Journal of Socio-Economics*, [S.l.], v. 33, p. 587-606, 2004
- HUBBARD, R.; RYAN, P. A. The historical growth of statistical significance testing in psychology and its future prospects. *Educational and Psychological Measurement*, v. 60, n. 5, p. 661-681, 2000.
- KESELMAN, H. J. et al. Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research*, [S.l.], v. 68, n. 3, p. 350-386, 1998.

LEWIN, K. *Field theory in social science: selected theoretical papers* by Kurt Lewin. London: Tavistock, 1952.

MAXWELL, S. E. The persistence of underpowered studies in psychological research: causes, consequences, and remedies. *Psychological Methods*, [S.l.], v. 9, n. 2, p. 147-163, 2004.

MICCERI, T. The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, [S.l.], v. 105, n. 1, p. 156-166, 1989.

ROSNOW, R. L.; ROSENTHAL, R. Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, v. 44, n. 10, p. 1276-1284, 1989.

SCHMIDT, F. L. Statistical significance testing and cumulative knowledge in psychology: implications for training of researchers. *Psychological Methods*, [S.l.], v. 1, n. 2, p. 115-129, 1996

THOMPSON, B. (Ed.). Planned versus unplanned and orthogonal versus nonorthogonal contrasts: the neo-classical perspective. In \_\_\_\_\_. *Advances in social science methodology*. Greenwich, CT: JAI Press. 1994. v. 3, p. 3-27.

THOMPSON, B. What future quantitative social science research could look like: confidence intervals for effect sizes. *Educational Researcher*, [S.l.], v. 31, n. 3, p. 25-32, Apr. 2002.

WILKINSON, L.; APA Task Force on Statistical Inference. Statistical methods in psychology journals: guidelines and explanations. *American Psychologist*, [S.l.], v. 54, n. 8, p. 594-604, 1999. Disponível em: <<http://www.apastyle.org/manual/related/wilkinson-1999.pdf>>. Acesso em: 20 abr. 2013.

WRIGHT, D. B. Making friends with your data: improving how statistics are conducted and reported. *British Journal of Educational Psychology*, [S.l.], v. 73, n. 1, p.123-136, 2003.

Recebido em: 01 de julho de 2013

Aceito em: 26 de agosto de 2014