



EDITORIAL

PREPARAÇÃO DE DADOS E BOAS PRÁTICAS EM PESQUISAS QUANTITATIVAS

 **Flavio Santino Bizarrias**

Pós-Doutor
Universidade Nove de Julho – UNINOVE.
São Paulo, SP – Brasil.
flavioxsp@hotmail.com

 **Luciano Ferreira da Silva**

Pós-Doutor
Universidade Nove de Julho – UNINOVE.
São Paulo, SP – Brasil.
lf_silvabr@yahoo.com.br

 **Renato Penha**

Pós-Doutor
Universidade Nove de Julho – UNINOVE.
São Paulo, SP – Brasil.
renato.penha.12@gmail.com

Resumo

Estudos quantitativos são considerados aqueles que recorrem a representações numéricas dos aspectos relacionados ao fenômeno de estudo, de maneira descritiva, ou pode-se ainda se considerar como estudos quantitativos aqueles em que o pesquisador estabelece hipóteses que serão testadas por meio de técnicas estatísticas inferenciais. Em ambos os casos, antes mesmo de se realizar análises sobre os dados coletados, o pesquisador deve inspecionar os dados obtidos de maneira a prepará-los para que as inferências sejam feitas sobre uma base de dados confiável, o mais livre possível de causas de viés que venham a distorcer as análises realizadas. Este comentário editorial tem por objetivo apontar alguns cuidados importantes que os pesquisadores devem ter nos seus estudos quantitativos, antes mesmo de aplicar qualquer teste estatístico. Estas recomendações se iniciam desde a etapa de planejamento da pesquisa, passando pela coleta, alcançando a etapa de preparação dos dados e análises prévias que se recomenda que sejam feitas para a apresentação dos resultados. Portanto, são abordadas recomendações diversas, como por exemplo, o tratamento de vieses do método comum, dados faltantes, multicolinearidade, tratamento de *outliers*, normalidade dos dados, preparação de variáveis, pressupostos dos testes, entre outras.

Palavras-chave: Pesquisa quantitativa. Preparação de dados. Análise de dados. Pressupostos de técnicas estatísticas.

Cite como

American Psychological Association (APA)

Bizarrias, F. S, Silva. L. F., Penha, R., (2023, jan./abr.). Preparação de dados e boas práticas em pesquisas quantitativas. Editorial. *Revista de Gestão e Projetos (GeP)*, 14(1), 1-10.
<https://doi.org/10.5585/gep.v14i1.22024>.

Muitos pesquisadores recorrem a estudos quantitativos para estudar fenômenos de interesse em projetos. Toda técnica de pesquisa, independente do paradigma adotado pelo pesquisador, demanda que sejam seguidos alguns procedimentos metodológicos para que o estudo possa ser confiável, robusto, baseado em critérios científicos, bem como ser passível de ser reproduzido por outros pesquisadores. Deste modo, o método científico representa um conjunto de etapas que estão relacionadas ao objetivo de pesquisa (Whitley & Kite, 2013; Bèrnie & Fernandez, 2012). Porém, pode ser observado que em estudos quantitativos a exploração dos dados, antes da realização de inferências, ou testes estatísticos e testes de hipótese, tem negligenciado a observação de adequação dos dados à inferência estatística. Esta prática vem se tornando comum em função de muitos fatores. Desde à maior popularização do emprego de testes estatísticos, antes de acesso mais restrito, até a maior amplitude de pesquisadores adotando técnicas quantitativas em seus estudos, o que promove a maior complexidade das análises multivariadas. Vale salientar que este não é um aspecto novo no ambiente de pesquisa.

Este comentário editorial tem por objetivo sinalizar aos pesquisadores que optem por empregar técnicas quantitativas, procedimentos que se recomendam que sejam realizados desde o planejamento da pesquisa, passando pela etapa de coleta de dados, até a exploração dos dados, antes da realização de inferências ou testes de hipóteses, ou qualquer outra análise inferencial para apresentação de resultados.

Na etapa de exploração dos dados, a inspeção prévia dos dados irá preparar os dados para que análises mais robustas e conclusivas para o estudo sejam realizadas. Isto evita retrabalho do pesquisador, críticas de avaliadores, rejeição de estudos ou, o mais grave, que resultados sejam tidos como os esperados pela teoria, quando não o são. Estes procedimentos de análise prévia dos dados, e até mesmo antes disto, devem ser reportados, quando pertinentes, antes dos resultados das análises relacionadas às hipóteses do estudo, ou demais análises.

O pesquisador deve sempre se pautar pela teoria que guia seus estudos, para que não se caia na armadilha da mera exploração de relações entre variáveis, e aplicação de técnicas diversas, desconectadas dos objetivos do estudo, com a inferência inapropriada dos resultados alcançados. A exploração prévia dos dados demanda tempo, esforço e conhecimento do pesquisador, e deve ser vista como um “investimento” do pesquisador na qualidade do seu estudo (Hair *et al.*, 2019).

Desde o planejamento do estudo, o pesquisador já pode adotar procedimentos que deverão ser reportados e trarão robustez à pesquisa. É recomendável a criação de um protocolo de pesquisa, ou seja, um documento em que o pesquisador possa registrar e acessar procedimentos realizados no seu estudo, registrar análises, testes, codificações adotadas, passo-a-passo de análises feitas. Isto requer organização do pesquisador, o que irá o ajudar quando seu estudo passar pelo escrutínio dos avaliadores e editores de revistas. Neste sentido, a documentação do processo de pesquisa viabiliza, caso seja necessária, a prestação de informações na revisão do estudo, que podem constar das análises feitas, mas que não foram reportadas.

Estudos quantitativos frequentemente se utilizam de escalas de mensuração de variáveis representativas dos construtos pertinentes ao estudo. Os procedimentos de tradução reversa adotados no estudo (quando ocorrer) deverão constar no método, ao se tratar das medidas utilizadas (Klotz, Swider, & Kwon, 2022). Isto traz confiabilidade e rigor ao estudo. Além da tradução dos itens, podem ser utilizadas outras técnicas de validação de face do instrumento de pesquisa, como a avaliação de juízes, pesquisadores especializados no tema, para a construção do instrumento de pesquisa. Caso estes procedimentos ocorram, na sua totalidade, ou parcialmente, deverão ser reportados. A observação obtida sobre as escalas utilizadas pode sinalizar a presença de itens reversos (assertivas que vão em sentido contrário às demais, para fins de captura de atenção do respondente). Itens reversos são apontados em estudos geralmente com a presença de um “r” minúsculo à frente do item, ou relacionado a uma nota em uma tabela (DeVellis & Thorpe 2021). O pesquisador deve se atentar a reverter a resposta de todos os respondentes dos itens reversos antes de se realizar qualquer análise dos dados.

A coleta dos dados deve seguir a busca de uma amostra representativa do universo de pesquisa, tanto no aspecto qualitativo desta representatividade, com participantes relacionados ao fenômeno de estudo, quanto com relação ao tamanho de amostra minimamente necessário para a condução do estudo. Outro aspecto bastante relevante, comumente negligenciado por pesquisadores, é o pré-teste do instrumento de pesquisa, onde a avaliação semântica, de tempo de preenchimento, bem como o *feedback* de representantes do universo de pesquisa, possibilita a avaliação da adequação do instrumento de pesquisa pretendido para a coleta principal. Neste momento, não se buscam validações estatísticas, mas tão somente avaliações da adequação da pesquisa ao estudo. Outro aspecto também regularmente não reportado em estudos

quantitativos é o emprego de técnicas de randomização dos itens, para se mitigar o viés do método comum, que será novamente abordado mais adiante.

Com os dados coletados, o trabalho de inspeção prévia dos dados se inicia. O primeiro aspecto a ser observado é a existência de dados faltantes, sejam estes dados primários ou secundários. Com o uso de plataformas eletrônicas de coleta de dados, a existência de dados faltantes não deve ocorrer. Entretanto, dependendo da forma como os dados foram coletados é possível que ocorra a ausência de dados (*missing values*) e o pesquisador deverá optar por uma forma de tratar esta questão, e reportar quando pertinente. Muitos softwares possuem nos seus algoritmos opções para tratar os dados faltantes, como o complemento dos dados faltantes pela média de respostas, ou a eliminação do respondente.

De qualquer forma, o pesquisador deve avaliar isto previamente, e reportar se este aspecto foi identificado no estudo, e como tratou a mitigação do seu impacto. Hair *et al.* (2019) indicam que é aceitável se adotar alguma técnica de complemento dos dados até o limite de 10% do total dos casos ou observações coletados. Isto é um recurso comum, e existem diversas formas de tratar dados faltantes baseadas no modelo a ser testado, que cada vez mais são incorporados aos softwares de análise. Para um maior detalhamento das opções disponíveis ver Hair *et al.* (2019, p. 59). De maneira prática, se a amostra obtida é grande o suficiente, o pesquisador pode adotar a eliminação do respondente com dados faltantes, sem prejudicar a exigência de tamanho de amostra.

Uma segunda etapa necessária na exploração de dados se refere à análise de *outliers*. *Outliers* são casos, ou respondentes, cujo padrão de respostas foge aos dos demais casos e respondentes da amostra. A ocorrência de *outliers* pode se dever a diversos fatores, como a própria divergência de avaliação do respondente, a diversidade de tipos de dados coletados, o erro de preenchimento, a dificuldade na compreensão do que é questionado, entre outros fatores. Quando a discrepância dos dados ocorre em uma única variável, dizemos que há *outliers* univariados. De maneira mais crítica, quando a discrepância dos dados ocorre em mais de uma variável, chamamos isto de *outliers* multivariados. Aqui vale um ponto de atenção, dependendo da técnica de análise de dados empregada no estudo, um *outlier* pode ser compreendido de maneira distinta. Seja por conta da característica distinta do dado em relação aos demais, ou em relação à discrepância do valor previsto. Este último caso, por exemplo, ocorre quando se utiliza a regressão, em que *outlier* é mais bem compreendido pelos resíduos, ou a diferença entre o

valor coletado e o previsto pelo modelo proposto. Um *outlier* não pode ser caracterizado sempre como algo ruim, ou bom.

Se por um lado um *outlier* representa uma discrepância de apontamento de um dado, e isto pode afetar os resultados, por outro lado, um *outlier* pode representar uma parcela de casos em que há características distintas entre os indivíduos estudados que não são passíveis de identificação à primeira vista. Para o tratamento de *outliers* o pesquisador deve antes identificá-los, avaliar seu peso em relação à quantidade de dados obtidos e seu impacto sobre o estudo, e então decidir como tratar estes respondentes, considerando inclusive a sua eliminação da base total. A distância de Mahalanobis (D^2) é comumente aceita como técnica para se detectar os *outliers* multivariados. A recomendação é que após avaliar se o *outlier* é um registro diferente do padrão, ou uma representação de um grupo de casos com perfil distinto, o pesquisador deve optar por manter ou retirar o caso *outlier*. Grupos de perfil distintos devem ser mantidos nos dados, enquanto *outliers* puramente discrepantes, sem significação de aspectos claros na amostra, devem ser retirados.

O uso de técnicas multivariadas é geralmente sensível à presença de multicolinearidade, ou a existência de uma excessiva correlação entre os itens do estudo. A presença de multicolinearidade está associada à instabilidade de um modelo, prejudicando a análise de dados, bem como resultando em super ou subestimação dos resultados. O dado comumente associado à representação de multicolinearidade é o VIF (*variance inflation factor*, ou fator de inflação da variância). O VIF está associado ao item, não ao respondente. O pesquisador deve estipular previamente o valor aceito. Estudos sugerem ser aceitável valor de VIF abaixo de 5 para retenção do item no estudo.

Um próximo passo antes da análise inferencial dos dados deve ser a avaliação de vieses comuns ao método (*commom method bias*). O viés comum ao método é um tipo de viés de mensuração que ocorre quando a variação nos dados se deve a aspectos do método de mensuração, e não aos construtos mensurados no estudo. É um erro sistemático que surge do uso de um único método de coleta de dados, levando a uma superestimação ou subestimação da verdadeira relação entre as variáveis. Desta forma, o viés comum ao método pode ser mitigado por procedimentos, ou avaliado por técnicas estatísticas. Podsakoff *et al.* (2003) apresentam uma série de recomendações que permitem mitigar os efeitos do viés do método comum. Não se trata de eliminar este aspecto, mas lidar com ele de maneira a se ter os menores impactos possíveis. Assim, podem ser adotadas medidas de procedimento de pesquisa e testes

estatísticos. Ao realizar um estudo quantitativo, se busca compreender a variância dos construtos de interesse. Quando esta variância se deve ao método de mensuração, e não aos construtos do estudo, temos potencialmente um viés de método comum, ou relacionado aos métodos de mensuração, desde a coleta dos dados. Este viés de método comum pode afetar os resultados alcançados em um estudo, desta forma, o pesquisador deve tratá-los. Diversas fontes podem gerar viés do método comum, como a mesma pessoa responder as variáveis do estudo, o esforço dos respondentes em manter consistência cognitiva nas suas respostas, correlações ilusórias e teorias implícitas dos próprios respondentes sobre aquilo que lhe é perguntado, desejabilidade social, leniência com o pesquisador, tendência à concordância, estados afetivos, entre outras (ver Podsakoff *et al.*, 2003, p. 882, Tabela 2).

Possíveis tratamentos para prevenir estes vieses são abordados por Podsakoff *et al.* (2003), assim recomendamos que ao menos um procedimento deva ser adotado nos estudos quantitativos. Estas abordagens se dividem basicamente em duas fontes, relacionadas aos procedimentos e relacionadas a testes estatísticos. Dentro das possibilidades de procedimento que podem ser adotadas, destacamos a coleta das variáveis do estudo de diferentes fontes, naturalmente se for possível e adequado aos objetivos do estudo. Pode-se ainda separar temporalmente a coleta das variáveis preditiva e de critério (dependente), pode-se garantir ao respondente o sigilo das informações, randomizar os itens do instrumento de pesquisa, além de uma atenção especial quanto ao significado dos itens utilizados nas escalas do estudo.

Ao utilizarmos em nossos estudos escalas de outros contextos culturais, é natural que os instrumentos apresentem diferenças de significado que podem gerar erros de mensuração, que deveriam ser invariantes em grupos distintos (Henseler, Ringle, & Sarstedt, 2016). Desta forma, o pesquisador deve investir tempo na observância das escalas referentes aos construtos que irá utilizar. Um erro frequente na preparação dos dados é não inverter os itens que foram desenvolvidos de maneira reversa, e analisar os dados sem esta alteração, produzindo resultados não esperados, e incorretos.

Com relação às técnicas estatísticas para mitigação do viés comum ao método, Podsakoff *et al.* (2003) apresentam algumas possibilidades. O mais comum destes testes é o fator único de Harman, em que todos os itens do estudo são analisados com um único fator em uma análise fatorial exploratória, e avaliar a estrutura fatorial não rotacionada, que deverá produzir fatores dispersos, ou se emergir um único fator, em que a variância total explicada seja reduzida (abaixo de 50%). Outra técnica comumente utilizada é avaliar o efeito de uma variável

não relacionada ao estudo, o que produz correlações parciais distintas. Aqui é recomendado avaliar se a desajustabilidade social tem este efeito no teste.

Busca-se evidenciar que uma medida não relacionada com o estudo não produz efeito sobre a correlação entre as variáveis, e desta forma, não produz um viés. Estas técnicas se ampliam com suas variações, em diversas outras abordagens de testes estatísticos para avaliar o viés do método comum, como por meio de variáveis latentes e variáveis de desajustabilidade social ou de afeto, carregarem no mesmo fator, ou atuarem como variáveis de controle na correlação parcial. Recomendamos que os estudos que são baseados em *surveys* de intenção comportamental devem possuir ao menos uma técnica de controle estatístico do viés comum ao método reportado. Em gestão de projetos, estudos que envolvem avaliações atitudinais ou comportamentais são comuns, pois existem diversos *stakeholders* nos projetos, ou envolvidos com o projeto de alguma forma, representando oportunidades de estudos com pessoas.

Como recomendação final, os estudos quantitativos que apresentem técnicas de análise multivariada devem ter especial atenção à descrição dos pressupostos de cada técnica para sua utilização no estudo. Cada técnica estatística possui um conjunto de pressupostos para a sua aplicação e, desta forma, o pesquisador deve sinalizar domínio da relação objetivo-técnica ao demonstrar a adequação da aplicação ao seu contexto de pesquisa, respeitando as condições necessárias de cada técnica. Isto contribui não apenas para a robustez metodológica do estudo, bem como para a reprodutibilidade do estudo por outros pesquisadores.

É usual que estudos quantitativos que envolvam dados primários busquem apresentar após a exploração dos dados, os indicadores de ajuste do modelo proposto, baseados em validade convergente, discriminante e confiabilidade das medidas (Hair *et al.*, 2019). Desta forma, o pesquisador sinaliza aos interessados na sua pesquisa que o desenho do estudo proposto, além de representar a abordagem teórica apropriada, também possui os índices de ajuste adequados para o teste de hipóteses e demais inferências que vier a realizar.

Estudos mais robustos vêm adotando técnicas diferentes, abordagens mistas de investigação, que possam apresentar abordagens novas aos fenômenos de interesse (Kline, 2019). O pesquisador deve ainda trazer explicação das técnicas de análise que adotou e, principalmente, justificativa de adequação da técnica empregada aos objetivos do estudo. Com relação aos resultados, sempre que possível, o pesquisador deve buscar reportar o tamanho do efeito e intervalos de confiança dos seus resultados, buscando procedimentos baseados em reamostragem (*bootstrapping*), procurando trazer mais robustez aos achados do trabalho

(Kline, 2013). Sempre que possível, é importante para a diversidade de audiência dos estudos que se traga uma apresentação gráfica do principal resultado do estudo, por meio de modelos conceituais, gráficos, curvas de interação, ou alguma apresentação do comportamento dos dados.

Em suma, na medida em que a maior capacidade computacional, e disseminação de técnicas evoluem, os estudos quantitativos ganham mais alcance e variedade de possibilidades de aplicação. O maior acesso a métodos quantitativos diversos aumenta a popularização do seu uso, mas isto não deve vir acompanhado do menor cuidado com o rigor de sua aplicação. A Tabela 1 apresenta uma síntese de recomendações e boas práticas para estudos quantitativos.

Tabela 1.

Recomendações e Boas Práticas em Pesquisa Quantitativa

Etapa	Recomendação	Explicação	Objetivo	Referência
Planejamento da pesquisa	Análise detida dos itens das escalas de interesse.	Observar o significado dos itens das escalas, <i>versus</i> os construtos que representam. Avaliar se há itens revertidos. Avaliar a necessidade de adaptação do instrumento.	Avaliar a pertinência da escala aos objetivos do estudo e ao universo de pesquisa, se evitando possíveis erros de mensuração.	Hair <i>et al.</i> (2019); DeVellis & Thorpe (2021)
	Tradução reversa*	Realizar a tradução ou adaptação dos itens das escalas do instrumento de pesquisa, por meio de procedimentos adequados.	Obter confiabilidade do instrumento de pesquisa.	Klotz, Swider, & Kwon (2022)
	Protocolo de pesquisa	Ter o registro e controle dos pressupostos do estudo, como escalas, procedimentos de tradução, codificações, sintaxes e construção de hipóteses.	Traz controle, segurança e robustez ao estudo.	Whitley & Kite (2013); Bèrnie & Fernandez (2012)
Coleta de dados	Pré-teste*	Testar o instrumento de pesquisa em uma amostra semelhante à da coleta final.	Avaliar a adequação dos itens ao universo de pesquisa, avaliar e corrigir erros semânticos, ou entendimento do instrumento de pesquisa, bem como avaliar o tempo de resposta.	Whitley & Kite (2013); Hair <i>et al.</i> (2019)
	Abordagem da amostra*	Descrever de maneira clara como a amostra do estudo foi contactada e os respondentes engajados no estudo.	Demonstra a pertinência da amostra em relação ao universo de pesquisa.	Whitley & Kite (2013); Hair <i>et al.</i> (2019)

Etapa	Recomendação	Explicação	Objetivo	Referência
	Aleatorização da coleta*	Conduzir coleta de dados com alternância de itens, de fontes de dados e de formas de coleta.	Reduz o viés do método comum.	Podsakoff <i>et al.</i> (2003)
Exploração dos dados	Análise de dados faltantes*	Observar na amostra a existência de dados faltantes, seu padrão de ocorrência, alcance dentro dos construtos, e dentro do estudo como um todo. Escolher um método de tratamento.	Construir uma base adequada para análises inferenciais.	Hair <i>et al.</i> (2019)
	Análise de outliers*	Avaliar a presença e impacto de outliers uni e multivariados.	Preparar a amostra para análises inferenciais.	Hair <i>et al.</i> (2019)
	Análise de multicolinearidade*	Avaliar o grau de explicação de uma variável por meio das demais.	Preparar a amostra para análises inferenciais.	Hair <i>et al.</i> (2019)
	Análise de viés comum ao método*	Busca avaliar a existência de indícios de erros de mensuração.	Mitigar possíveis diferenças na variância devido a métodos de mensuração.	Podsakoff <i>et al.</i> (2003)
	Pressupostos das técnicas*	Avaliar os pressupostos para a aplicação da técnica de análise dos dados.	Utilizar adequadamente cada técnica de análise, respeitando seus pré-requisitos.	Hair <i>et al.</i> (2019); Whitley & Kite (2013); Marôco (2018)
Apresentação de resultados	Tamanho do efeito, intervalo de confiança e apresentação gráfica*	Avaliar o resultado dos dados por meio de indicadores robustos	Trazer robustez para o estudo.	Kline (2013); Hair <i>et al.</i> (2019)

Nota: * recomenda-se que seja reportado como foi tratado no estudo,

Fonte: Os autores, 2023.

Estas indicações de práticas não pretendem ser um guia *sine qua non* para as pesquisas quantitativas, pois cada estudo possui uma série de particularidades que devem ser observadas para sua construção, mas sim, uma forma de apresentar aos pesquisadores boas práticas para publicações de impacto. A forma de cálculo de muitas das abordagens aqui descritas se encontra distribuída em diversos estudos, e muitas vezes são abordados em cursos e seminários. A implementação de técnicas de análise depende ainda do *software* empregado, para que se possa seguir a parametrização dos dados nos pacotes estatísticos disponíveis. Para informações com relação a pressupostos e aplicação destas técnicas recomenda-se o trabalho de Marôco (2018), com aplicações por meio do software SPSS, com equivalentes em outras plataformas.

Tomadas em conjunto, a combinação parcial de ações aqui propostas, devem ser valorizadas nos estudos quantitativos, como esforço do pesquisador para a elaboração de estudos robustos metodologicamente, desde o planejamento da pesquisa até a apresentação dos resultados.

Referências

- Bèrnie, D. A.; Fernandez, B. P. (2012). (Orgs). *Métodos e Técnicas de Pesquisa - Modelando As Ciências Empresariais*, 1a edição. Saraiva.
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage publications.
- Hair Jr., J.F.; William, B.; Babin, B.; and Anderson, R.E (2019). *Multivariate data analysis*. 8th.ed. Hampshire: Cengage Learning EMEA.
- Kline, R. B. (2013). *Beyond significance testing: Statistics reform in the behavioral sciences*. American Psychological Association.
- Kline, R. B. (2019). *Becoming a behavioral science researcher: A guide to producing research that matters*. Guilford Press, Second ed.
- Klotz, A. C., Swider, B. W., & Kwon, S. H. (2022). Back-translation practices in organizational research: Avoiding loss in translation. *Journal of Applied Psychology*.
- Marôco, J. (2018). *Análise Estatística com o SPSS Statistics*.: 7ª edição. Report Number, Lda.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology*, 88(5), 879.
- Whitley Jr, B. E., & Kite, M. E. (2013). *Principles of research in behavioral science*. Routledge.