

**MÉTODOS ESTATÍSTICOS ADOTADOS EM TESES DE DOUTORADO EM
CIÊNCIAS DO MOVIMENTO HUMANO: UM ESTUDO DESCRITIVO**

**STATISTICAL METHODS ADOPTED IN PHD THESES IN HUMAN MOVEMENT
SCIENCES: A DESCRIPTIVE STUDY**

**MÉTODOS ESTADÍSTICOS ADOPTADOS EN TESIS DOCTORALES EN
CIENCIAS DEL MOVIMIENTO HUMANO: UN ESTUDIO DESCRIPTIVO**

ROCHELLE ROCHA COSTA

Doutorado em Ciências do Movimento Humano pela Universidade Federal do Rio Grande do Sul (UFRGS). Professora da Faculdade SOGIPA, Porto Alegre/RS, e pesquisadora do Programa de Pós-Graduação em Ciências do Movimento Humano da UFRGS.

rochelle.costa@ufrgs.br

FLÁVIO ANTÔNIO DE SOUZA CASTRO

Doutorado em Ciências do Movimento Humano pela Universidade Federal do Rio Grande do Sul (UFRGS), Professor Associado da Escola de Educação Física, Fisioterapia e Dança (ESEFID) da UFRGS e professor permanente do Programa de Pós-Graduação em Ciências do Movimento Humano da UFRGS.

souza.castro@ufrgs.br

Resumo

Introdução: O conhecimento acerca dos principais métodos estatísticos utilizados em estudos de uma determinada área do conhecimento permite conhecer o desenho metodológico predominante, bem como os principais objetivos a serem testados, e verificar se há limitações no uso de métodos mais elaborados e atualizados. **Objetivo:** descrever os principais métodos estatísticos adotados em teses de doutorado desenvolvidas em um programa de pós-graduação na área de Educação Física e descrever os métodos estatísticos utilizados nos artigos publicados com dados oriundos dessas teses. **Metodologia:** o presente estudo analisou teses de caráter quantitativo, defendidas entre 2003 e 2013 e seus correspondentes artigos publicados. Os seguintes dados foram extraídos das teses e dos artigos: quantidade de artigos publicados com dados oriundos de cada tese, dados relativos à análise descritiva, testes de normalidade, se foram ou não realizados testes estatísticos inferenciais e, em caso afirmativo, quais foram adotados. **Resultados:** um total de 46 teses e 88 artigos foram incluídos nas análises do presente estudo. A mediana de publicações por tese foi de dois artigos, sendo que 10 teses não apresentaram nenhum artigo publicado e três teses apresentaram cinco estudos publicados. Verificou-se a predominância de estudos transversais, a preferência pelo uso da estatística inferencial, com predominância de testes paramétricos. Ainda, evidenciou-se elevada utilização de métodos de comparações entre médias, e uso de correlações. **Conclusões:** de forma geral, os métodos estatísticos adotados nas teses e nos artigos publicados são adequados e coerentes com os objetivos propostos nos estudos.

Palavras-Chave: Estatística. Monografia. Artigo de revista.

Abstract

Introduction: The knowledge about the main statistical methods used in studies of a certain area of knowledge allows to know the predominant methodological design, as well as the main objectives to be tested, and to check if there are limitations in the use of more elaborated and updated methods. **Objective:** to describe the main statistical methods adopted in doctoral theses developed in a postgraduate program in the area of Physical Education and to describe the statistical methods used in published articles with data from these theses. **Methodology:** the present study analyzed quantitative theses, defended between 2003 and 2013 and their corresponding published articles. The following data were extracted from the theses and articles: number of articles published with data from each thesis, data related to descriptive analysis, normality tests, whether inferential statistical tests were carried out and, if so, which ones were adopted. **Results:** a total of 46 theses and 88 articles were included in the analyzes of the present study. The median of publications per thesis was two articles, with 10 theses presenting no published articles and three theses presenting five published studies. There was a predominance of cross-sectional studies, a preference for the use of inferential statistics, with a predominance of parametric tests. Furthermore, there was a high use of methods of comparisons between means, and the use of correlations. **Conclusions:** in general, the statistical methods adopted in the theses and published articles are adequate and consistent with the objectives proposed in the studies.

Keywords: Statistic. Monography. Journal article.

Resumen

Introducción: El conocimiento sobre los principales métodos estadísticos utilizados en los estudios de un área de conocimiento determinada permite conocer el diseño metodológico predominante, así como los objetivos principales a ser probados, y verificar si existen limitaciones en el uso de métodos más elaborados y actualizados. **Objetivo:** describir los principales métodos estadísticos adoptados en las tesis doctorales desarrolladas en un programa de posgrado en el área de Educación Física y describir los métodos estadísticos utilizados en los artículos publicados con datos de estas tesis. **Metodología:** el presente estudio analizó tesis cuantitativas, defendidas entre 2003 y 2013 y sus correspondientes artículos publicados. Los siguientes datos se extrajeron de las tesis y artículos: número de artículos

publicados con datos de cada tesis, datos relacionados con el análisis descriptivo, pruebas de normalidad, si se realizaron pruebas estadísticas inferenciales y, en caso afirmativo, cuáles se adoptaron. Resultados: se incluyeron un total de 46 tesis y 88 artículos en los análisis del presente estudio. La mediana de publicaciones por tesis fue de dos artículos, con 10 tesis que no presentan artículos publicados y tres tesis que presentan cinco estudios publicados. Hubo un predominio de los estudios transversales, una preferencia por el uso de estadísticas inferenciales, con predominio de las pruebas paramétricas. Aún así, hubo un alto uso de métodos de comparación entre medias y el uso de correlaciones. Conclusiones: en general, los métodos estadísticos adoptados en las tesis y artículos publicados son adecuados y consistentes con los objetivos propuestos en los estudios.

Palabras Clave: Estadística. Monografía. Artículo de revista.

1. INTRODUÇÃO

A utilização dos métodos estatísticos adequados para o contexto da pesquisa é um tema que tem se tornado o foco de diversas investigações nas ciências do exercício e do esporte (HOPKINS et al., 2009; TEIXEIRA et al., 2015). Consiste em um assunto relevante à medida que a adoção de técnicas inapropriadas pode conduzir a conclusões equivocadas acerca dos resultados, ou dificultar a correta interpretação dos achados da pesquisa. A definição do método estatístico pertinente tem início na formulação do problema de pesquisa e na consequente definição das hipóteses. Para decidir se uma determinada hipótese é confirmada por um conjunto de dados, é necessário que haja um planejamento de procedimentos objetivos que permitam aceitar ou rejeitar a hipótese nula traçada (SIEGEL; CASTELLAN, 2006). Neste sentido, os métodos estatísticos são utilizados para o planejamento e condução de um estudo, descrição dos seus dados e tomada de decisões finais.

O conhecimento acerca dos principais métodos estatísticos utilizados em estudos de uma determinada área do conhecimento permite conhecer o desenho metodológico predominante, bem como os principais objetivos a serem testados, e verificar se há limitações no uso de métodos mais elaborados e atualizados. Quando essa análise se dá, de forma mais específica, dentro de um programa de pós-graduação, com curso de doutorado, esses benefícios são ampliados. A completa análise e descrição dos procedimentos estatísticos utilizados em um curso de doutorado possibilita conhecer o perfil dos trabalhos desenvolvidos, verificar se há predominância de determinados métodos e objetivos de pesquisa, verificar se os procedimentos estatísticos adotados são adequados aos objetivos dos estudos, identificar possíveis limitações da aplicação do conhecimento adquirido durante a formação em estatística no curso de

doutorado, bem como identificar falhas e fraquezas no ensino de determinados conteúdos nas disciplinas oferecidas pelo curso.

A detecção desse conjunto de informações pode auxiliar na melhoria da qualidade dos estudos desenvolvidos no curso de doutorado, bem como melhorar os índices de publicação em revistas de melhores fatores de impacto, além de melhorar a qualidade do ensino dos métodos estatísticos nos cursos. Sendo assim, os objetivos do presente estudo foram: (i) descrever os principais métodos estatísticos adotados em teses de doutorado desenvolvidas em um programa de pós-graduação na área de Educação Física e (ii) descrever os métodos estatísticos utilizados nos artigos publicados com dados oriundos dessas teses.

Hipotetiza-se que será encontrada uma predominância de estudos transversais, do uso de estatística inferencial com testes paramétricos, e de métodos de comparações entre médias/medianas, tanto nas teses quanto nos artigos publicados, em razão dos objetivos da maioria dos estudos da área.

2. MÉTODO

2.1. DELINEAMENTO DO ESTUDO

Este trabalho caracteriza-se como um estudo transversal, descritivo, de abordagem quantitativa.

2.2. SELEÇÃO DAS TESES

As teses utilizadas no presente estudo são oriundas do Programa de Pós-Graduação em Ciências do Movimento Humano (PPGCMH) da Universidade Federal do Rio Grande do Sul. Todas as teses foram obtidas via repositório digital, sendo de acesso público e, por isso, não necessitando de avaliação anterior por Comitê de Ética em Pesquisa. Foram incluídos trabalhos monográficos de doutorado apresentados no período de janeiro de 2003 a dezembro de 2013, tendo em vista que este estudo faz parte de um projeto aprovado pela Comissão de Pesquisa da unidade acadêmica da Universidade Federal do Rio Grande do Sul, onde o estudo foi realizado, e pelo comitê do órgão de fomento que o financiou (CAPES), com coletas para este período. Estes trabalhos foram, primeiramente, catalogados e inseridos no software gerenciador de referências EndNote®(<https://endnote.com/>).

Como critérios de inclusão foram adotados incluir apenas estudos de caráter quantitativo (aqueles que determinaram método formal, objetivo e sistemático para geração de

dados numéricos que foram usados para estabelecer relações entre as variáveis dependentes e independentes do estudo, adotando métodos estatísticos já padronizados da literatura científica da área de estudo) e defendidos perante banca examinadora durante o período de janeiro de 2003 a dezembro de 2013. Também foram incluídos estudos mistos, com abordagens quantitativa e qualitativa simultaneamente (sendo adotadas, para a análise do presente estudo, as abordagens quantitativas desses trabalhos). Em contrapartida, foram excluídos os estudos de caráter exclusivamente qualitativo, além daqueles indisponíveis no repositório digital.

2.3. SELEÇÃO E OBTENÇÃO DOS ARTIGOS

Após a identificação das teses, os artigos publicados com dados coletados para as respectivas teses foram identificados. Nesta fase, a identificação foi realizada consultando-se o currículo Lattes (de acesso público) na plataforma Lattes (<http://lattes.cnpq.br/>), a plataforma acadêmica *Research Gate* e via contato direto com os autores por e-mail, quando disponível. A seguir foram realizadas buscas pelos artigos completos usando os motores de busca *PubMed*, *Scopus*, *Scielo*, Biblioteca Virtual em Saúde e Google Acadêmico, além dos sites dos jornais científicos nos quais foram publicados.

2.4. COLETA DOS DADOS

Os dados foram coletados usando um formulário padronizado, por meio do qual foram extraídas as informações: (i) quantidade de artigos publicados com dados oriundos de cada tese, e (ii) informações relativas à análise descritiva dos dados, presença de testes de normalidade dos dados, realização de técnicas estatísticas inferenciais e de testes complementares (*post hoc*) e, em caso afirmativo, quais foram adotados. Estes dados foram extraídos das teses e dos respectivos artigos, sendo tabulados em planilhas digitais do software Microsoft Excel® 2018 versão 16.18 (<https://www.microsoft.com/pt-br/>).

2.5. ANÁLISE DOS DADOS

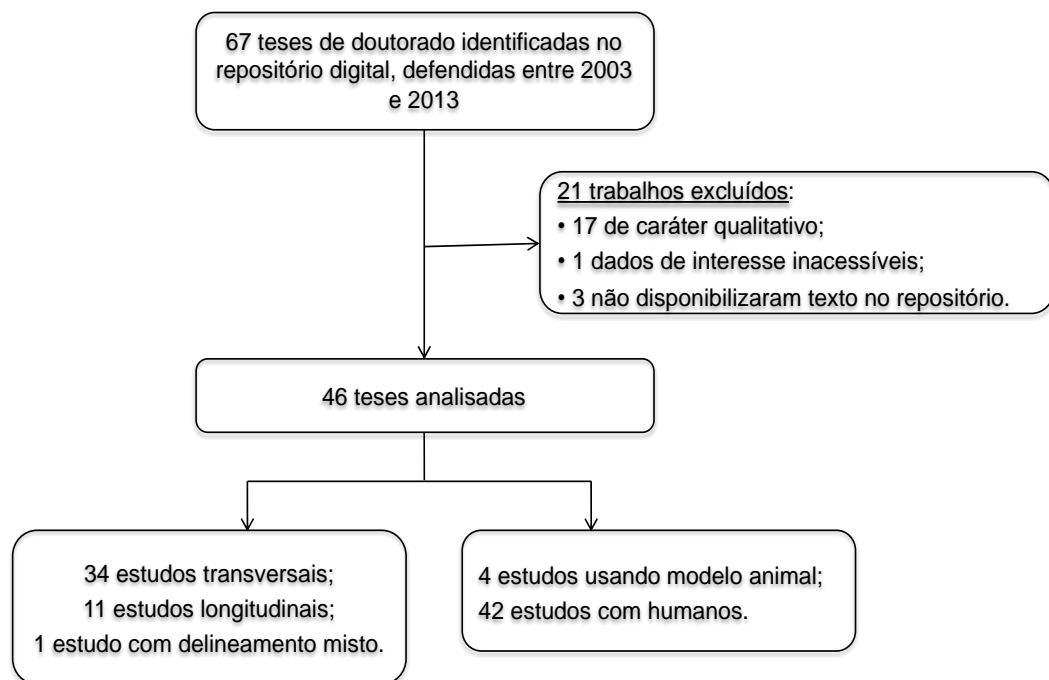
As características dos estudos estão apresentadas de forma descritiva, com valores relatados a partir de frequências absolutas (n) e relativas (percentual em relação ao total de estudos quantitativos incluídos ou ao total de estudos na subanálise em questão) e medidas de tendência central (média e mediana). Os estudos foram agrupados quanto aos tipos de testes

estatísticos adotados: (i) para dados categóricos, (ii) paramétricos ou (iii) não-paramétricos; e (iv) quanto às diferentes classificações de finalidade (comparações de médias, testes de associação, análises de regressão, etc), quando possível. Tais análises foram realizadas no software Microsoft Excel® 2018 versão 16.18 (<https://www.microsoft.com/pt-br/>).

3. RESULTADOS

Foram identificadas 67 teses de doutorado no repositório digital defendidas no período de janeiro de 2003 a dezembro de 2013. Destas, 46 foram incluídas na análise do presente estudo, por atenderem aos critérios de inclusão. A Figura 1 apresenta o fluxograma das etapas de identificação, seleção e análise das teses.

Figura 1 - Fluxograma das teses analisadas.

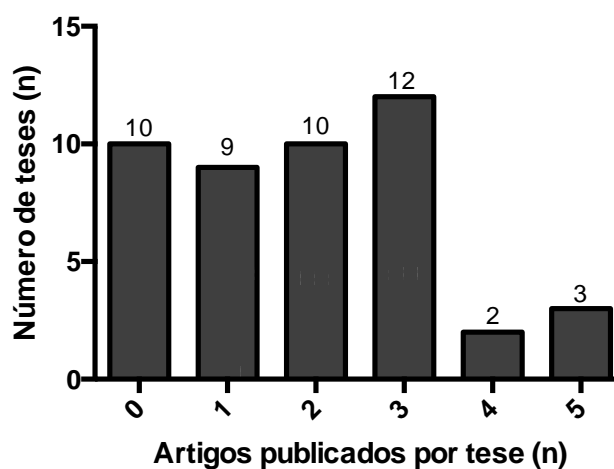


Fonte: Dados da pesquisa.

Correspondentes as 46 teses, foram identificados 88 artigos publicados em revistas científicas, sendo 50 publicados em revistas internacionais (57%) e 15 artigos apresentando o delineamento de revisão bibliográfica (17%). A mediana de publicações por tese foi de dois

artigos, sendo que 10 teses não apresentaram nenhum artigo publicado e três teses apresentaram cinco estudos publicados (Figura 2).

Figura 2 - Frequências absolutas do número de artigos publicados por tese.



Fonte: Dados da pesquisa.

No que se refere às análises apresentadas, duas teses (4%) citam ter realizado apenas análise descritiva dos dados, sem adoção de testes inferenciais. Ambas adotam como forma de apresentação dos resultados a média e o desvio-padrão de suas variáveis. Além disso, uma delas apresenta dados percentuais (frequências relativas) e a outra o coeficiente de variação das variáveis de interesse dentro da amostra. Em contrapartida, seis artigos (7%) apresentam exclusivamente análise descritiva, sendo apenas um deles correspondente a uma das teses que

também aplica este método de forma isolada. Os demais artigos apresentam seus dados de formas variadas, sendo um deles com frequências absolutas e relativas, três com média, desvio-padrão e frequências relativas (percentuais), e um com valores absolutos das respostas individuais e seu coeficiente de variação.

Vinte e cinco teses (54%) citam ter realizado testes estatísticos no intuito de verificar a distribuição dos dados, enquanto 37 artigos (42%) relatam ter adotado essa análise. Tanto nas teses quanto nos artigos, o teste mais adotado para esta finalidade foi o Shapiro-Wilk, estando presente em 16 das 25 teses e em 31 dos 37 artigos. Do total de 46 teses analisadas, 24 (52%) adotaram apenas testes estatísticos paramétricos em suas análises, enquanto 45 artigos apresentam análise exclusivamente paramétrica, representando 51% do total. Em contrapartida, apenas duas teses (4%) e três artigos (3%) optaram por utilizar exclusivamente testes não-paramétricos em suas análises. Adotando tanto testes paramétricos quanto não-paramétricos, foram encontradas sete teses defendidas (15%) e nove artigos publicados (10%) com a análise combinada.

Adicionalmente, uma tese (2%) adotando uma análise mista, composta por testes não-paramétricos somada à análise de dados categóricos, foi encontrada. Contudo, nenhum artigo que tenha sido publicado com essa metodologia foi identificado dentre os 88 oriundos das teses. Também de forma mista, três das 46 teses (7%) e três dos 88 artigos (3%) adotaram tanto testes paramétricos, quanto análises para dados categóricos em seus tratamentos estatísticos. Três outras teses (4%) adotaram uma análise mista mais completa, composta por testes paramétricos, não-paramétricos e também análise de dados categóricos. Dentre os 88 artigos identificados, nenhum foi publicado utilizando esse desenho.

Em uma análise descritiva quanto aos tipos de testes estatísticos utilizados nas teses, nota-se que predominam os métodos de comparações entre médias (ou medianas/postos), totalizando 38 das 46 teses (83%). Neste sentido, 24 delas adotaram testes t de Student (63%), 31 teses utilizaram análises de variância ou de covariância (82%), 14 teses realizaram os testes de Wilcoxon ou U de Mann Whitney (37%), e cinco teses optaram por realizar os testes de Kruskal-Wallis ou de Friedman (13%). De forma similar, os métodos de comparações entre médias são os que predominam entre os artigos publicados, estando presentes em 58 dos 88 artigos (66%). Dentre estes, 31 artigos apresentam análises com testes t de Student (53%), 47 adotaram análises de variância ou de covariância (81%), 10 usaram os testes de Wilcoxon ou

U de Mann Whitney (17%), e apenas 2 adotaram os testes de Kruskal-Wallis ou de Friedeman (3%).

A homogeneidade das variâncias foi testada em sete das 46 teses (15%), tendo sido utilizado o teste de Levene em todas elas. Estas sete teses originaram, em conjunto, 13 dos 88 artigos publicados (15%), todos constando a verificação da homogeneidade das variâncias por meio do teste de Levene. De forma a complementar às análises de variância, 14 teses (30%) e 25 artigos (28%) adotaram testes complementares de comparações múltiplas entre médias (*post hoc*). Neste contexto, o *post hoc* mais utilizado foi o de Bonferroni, estando presente em nove teses (64%) e em 15 artigos (60%), seguido do teste de Tuckey, adotado em três teses (21%) e em sete artigos (28%). Ainda, o teste complementar *Least Significant Difference* foi utilizado em três teses (21%) e em quatro artigos (16%), o teste de Student-Newman-Keuls foi adotado como *post hoc* em uma tese (7%) e em um artigo (4%). Por fim, o teste de Duncan foi realizado em uma tese (7%) e nenhum artigo publicado relata tê-lo utilizado.

O segundo método estatístico mais adotado corresponde às correlações, estando presente em 18 das 46 teses (39%) e em 24 dos 88 artigos publicados (27%). Estratificando pelos testes de correlação específicos, foram encontradas 13 teses (72%) e 15 artigos (62%) adotando o teste de Coeficiente de Correlação Linear Produto-Momento de Pearson, três teses (17%) e dois artigos (8%) usando a Coeficiente de Correlação de Postos de Spearman, além de cinco teses (28%) e sete artigos (29%) utilizando o Índice de Correlação Intra-classe (ICC). De uma forma geral, nos estudos analisados, os testes Coeficiente de Correlação Linear Produto-Momento de Pearson e de Postos de Spearman foram usados no intuito de verificar associações entre variáveis escalares, enquanto o ICC foi adotado quando se buscou testar a reprodutibilidade de medidas. Com utilização mais restrita, as regressões lineares foram utilizadas em três das 46 teses (7%) e em sete dos 88 artigos publicados (8%).

A determinação do tamanho de efeito foi realizada em apenas uma tese das 46 analisadas (2%), e em sete dos 88 artigos encontrados (8%). A tese que apresenta este resultado não cita qual método foi adotado, bem como não originou artigos publicados. Assim, os sete artigos que apresentam a determinação do tamanho de efeito são provenientes de teses que não realizaram esses cálculos. Como método adotado para a determinação do tamanho de efeito, três artigos (43%) utilizaram Eta Quadrado (η^2) e os demais quatro estudos (57%) realizaram o cálculo do *d* de Cohen.

Com o objetivo primário de validação de instrumentos de pesquisa, seis teses foram realizadas (13%) e dois artigos (2%) foram publicados. Dentre elas, o método estatístico mais utilizado foi o modelo de consistência interna Alpha de Cronbach, estando presente em quatro teses (67%) e nos dois artigos (100%). A análise fatorial foi adotada em três teses (50%) e em apenas um dos artigos (50%).

Em relação aos testes utilizados para a análise de dados categóricos, 10 das 46 teses (22%) e seis dos 88 artigos (7%) utilizaram o teste qui-quadrado em suas análises. O teste de McNemar para frequências pareadas foi adotado em duas teses (4%) e em um único artigo (1%). Objetivando a proposição de pontos de corte, a curva ROC (*Receiver Operating Characteristic*) foi utilizada em duas teses (4%) e em três artigos (3%). Ainda, o Índice de Concordância Kappa foi o método adotado em uma tese (2%) e em dois artigos (2%) para descrever a concordância inter-observador. Utilizando a regressão de Poisson foi encontrada uma tese dentre as 46 analisadas (2%), e dois artigos dentre os 88 analisados (2%). Um destes artigos é oriundo da tese que adota a regressão de Poisson, enquanto o outro é originário da única tese que realizou regressão logística binária dentre as 46 verificadas (2%). Não foram encontrados artigos utilizando a regressão logística binária.

4. DISCUSSÃO

Este estudo objetivou descrever os métodos estatísticos adotados em teses de doutorado de um programa de pós-graduação da área de educação física, e em seus artigos correspondentes publicados. Neste sentido, observou-se que grande parte dos estudos adota estatística inferencial na análise dos dados, com predominância de testes paramétricos. Evidenciou-se, também, a predominância de métodos de comparações entre médias/medanas, tanto nas teses quanto nos artigos publicados, como era hipotetizado, em razão dos objetivos da maioria dos estudos da área.

Como resultado da análise dos estudos incluídos neste trabalho, evidenciou-se que duas teses e seis artigos foram desenvolvidos usando isoladamente análise descritiva dos seus dados, sem realização de testes inferenciais. De acordo com Callegari-Jacques (CALLEGARI-JACQUES, 2003), a descrição detalhada das variáveis em estudo é imprescindível para a adequada interpretação dos resultados de uma pesquisa quantitativa, e a metodologia empregada faz parte da estatística descritiva. A utilização da estatística descritiva envolve o uso de medidas de tendência central (ou de postos/posição), e de dispersão (ou de variabilidade).

Essas são valores calculados com o objetivo de representar os dados de uma forma condensada, por meio de valores únicos. Conforme esperado pela abordagem quantitativa dos estudos incluídos, a medida de tendência central mais comumente utilizada nas Ciências do Exercício e do Esporte consiste na média aritmética que, além de ser fácil de calcular, apresenta uma interpretação familiar e propriedades estatísticas que a tornam útil na comparação entre populações e outras situações que envolvam inferências estatísticas. Tal aplicabilidade pode justificar o fato de todas as teses, e a maioria dos artigos, que adotam de forma exclusiva a análise descritiva de seus dados, terem escolhido a média como forma de agrupamento dos seus conjuntos de informações.

Neste sentido, as medidas de dispersão servem como complemento à informação agrupada das medidas de tendência central, informando sobre a variabilidade dos dados. Dentre os estudos incluídos, o desvio-padrão da média constitui a medida de dispersão mais adotada nas teses e artigos analisados. O desvio-padrão reflete a variabilidade das observações em relação à média, e não varia com o aumento do tamanho amostral (LUNET; SEVERO; BARROS, 2006). Este consiste na raiz quadrada positiva da variância e possui a vantagem de ser apresentado na unidade original da variável em questão (CALLEGARI-JACQUES, 2003; ALTMAN; BLAND, 2005).

Por estas características, o uso isolado da análise descritiva dos dados pode ser considerado adequado e suficiente quando o estudo em questão objetiva realizar uma descrição detalhada de determinado fenômeno em uma amostra, sem realizar generalizações para a população em geral (VETTER, 2017). Este foi o caso das duas teses e dos seis artigos incluídos no presente estudo, todos com desenho transversal, com objetivo exclusivamente descritivo. Nestes casos não é necessário que sejam realizados procedimentos estatísticos inferenciais, que se justificam na generalização dos achados para a população.

Embora 38 teses e 77 artigos tenham adotado testes paramétricos em suas análises, que pressupõem a distribuição normal dos dados, apenas 25 teses e 37 artigos relatam ter realizado procedimentos estatísticos para testar a distribuição das suas observações. De acordo com Vetter (2017), a distribuição normal dos dados consiste em um pressuposto importante em muitos testes inferenciais, sendo sua testagem fundamental para a correta utilização destas técnicas estatísticas e, conseqüente obtenção de resultados confiáveis. Neste contexto, o teste de Shapiro-Wilk tem sido um dos mais amplamente utilizados em análises quantitativas (VETTER, 2017). De fato, nas análises do presente estudo, este foi o método mais adotado com

o objetivo de verificar a normalidade dos dados coletados, tanto nas teses, quanto nos artigos publicados. Embora outros testes também tenham sua utilização difundida, o teste de Shapiro-Wilk fornece elevado poder estatístico, sendo mais confiável e robusto do que outros como, por exemplo, o teste de Kolmogorov-Smirnov (VETTER, 2017), segundo mais utilizado nas análises do presente estudo.

Observou-se, dentre os estudos incluídos, predominância de uso de testes paramétricos, seja de forma isolada, seja em análises mistas com testes não-paramétricos associados. Testes paramétricos são aqueles que estimam parâmetros de distribuição para os dados (em escala intervalar), e assumem que, de alguma forma, os dados seguem distribuição normal e variância uniforme entre os grupos ou ao longo do acompanhamento do estudo (no caso de estudos longitudinais com medidas repetidas no tempo). Esses testes apresentam maior poder estatístico quando comparados aos testes não-paramétricos (desde que seus pressupostos estejam atendidos). Dessa forma, a probabilidade de detectar um efeito significativo, quando ele realmente existir, é maior quando se usam técnicas paramétricas comparadas às não-paramétricas.

Em contrapartida, os testes não-paramétricos utilizam, para o cálculo de sua estatística, postos atribuídos aos dados ordenados e são livres da distribuição de probabilidades dos dados estudados. Como características, é possível ressaltar que (i) são menos exigentes do que os testes paramétricos, não possuindo tantos pressupostos para sua utilização; (ii) são técnicas apropriadas quando não se conhece a distribuição dos dados ou esta é assimétrica; (iii) são testes úteis quando as variâncias dos dados são heterogêneas. Contudo, uma desvantagem dos testes não-paramétricos consiste no fato de que, mesmo sendo úteis para testar hipóteses, não fornecem estimativas úteis, como por exemplo, a diferença média entre dois grupos ou entre dois momentos de análise. Além disto, dificultam as generalizações para situações mais complexas, como quando se deseja utilizar métodos de regressão com ajuste de diversos fatores (ALTMAN; BLAND, 2009). Além disso, são testes que extraem menos informações do experimento por serem técnicas empregadas em dados mensurados em escalas não quantitativas, em sua maioria, ou em dados quantitativos reduzidos a uma escala qualitativa ordenável. Dessa forma, perdem-se informações relativas à variabilidade da variável e impossibilita a testagem de interações entre as variáveis, o que restringe sua aplicação a algumas pesquisas. Essas características podem justificar a ampla utilização dos testes paramétricos nos estudos da área,

considerando que fornecem estimativas, intervalos de confiança e permitem generalizações para métodos mais complexos.

Em uma análise mais detalhada acerca dos tipos de testes adotados, evidenciou-se que as abordagens mais utilizadas, tanto nas teses quanto nos artigos, foram as comparações entre médias (ou medianas). Estes achados são consistentes com outros já publicados que encontraram elevado uso de testes t de Student, análises de variância, e seus testes não-paramétricos correspondentes (STRASAKA et al., 2007; TEIXEIRA et al., 2015; TRESSOLDI et al., 2013). O teste t de Student aparece como o mais utilizado método de comparação entre médias. Isso se justifica pelo desenho predominante dos estudos desenvolvidos no curso de doutorado, cujos objetivos principais consistem em comparar grupos de indivíduos quanto à determinadas características (por exemplo: treinados *versus* sedentários; saudáveis *versus* população clínica), ou verificar a eficácia de intervenções com análises pré e pós-intervenção.

De forma surpreendente, menos da metade das teses e dos artigos que realizaram análises de variância, citam ter testado a homogeneidade das variâncias, um pressuposto importante para sua utilização. Callegari-Jacques (CALLEGARI-JACQUES, 2003) sugere que, para que os resultados de análises de variâncias sejam válidos, é necessário que as variâncias amostrais sejam semelhantes nos diferentes conjuntos de dados. A realização de testes estatísticos cujos pressupostos básicos (como a homogeneidade das variâncias, por exemplo) não foram atendidos, poderá conduzir a interpretações equivocadas dos resultados (GREENLAND et al., 2016). Keselman et al. (KESELMAN; HUBERTY; LIX, 1998) relatam que a checagem de pressupostos para utilização de testes estatísticos é reportada em apenas 8 a 25% dos estudos. De fato, a testagem do pressuposto da homogeneidade nos estudos analisados, foi relatada em apenas 15% das teses e dos artigos publicados. Por outro lado, a verificação da distribuição dos dados foi relatada em 54% das teses incluídas e em 42% dos artigos publicados.

No mesmo sentido, menos da metade das teses e dos artigos que adotaram análises de variância, relatam ter realizado testes complementares de comparações múltiplas entre médias. Neste contexto, um resultado significativo na análise de variância não indica quais médias diferem entre si do ponto de vista estatístico, ela apenas demonstra que existe ao menos uma diferença entre as médias estudadas. Assim sendo, a utilização de testes complementares torna-se fundamental para “localizar” tais diferenças, considerando que a indicação de quais médias diferem entre si (em uma comparação duas a duas) é dada por este tipo de teste. Como característica em comum, os testes *post hoc* controlam o nível de significância ao levar em

conta o número de comparações realizadas (CALLEGARI-JACQUES, 2003). Dentre os estudos analisados, o Teste de Bonferroni foi mais utilizado para esta finalidade. Este teste possui como vantagem não requerer que as várias comparações sejam independentes, ampliando, dessa forma, sua aplicabilidade em estudos com medidas repetidas em amostras pareadas, desenho frequente nos estudos das Ciências do Exercício e dos Esportes. Por outro lado, produz um teste mais conservador quando comparado a outros *post hoc*, como por exemplo, o teste de Tuckey, o que pode dificultar a obtenção de significância estatística, sendo, por isso considerado rígido e rigoroso.

Na sequência dos testes de comparações entre médias, as correlações foram o segundo método mais adotado nas teses e nos artigos. Este achado era esperado, tendo em vista que a verificação de possíveis associações entre variáveis quantitativas é prática comum nas Ciências do Exercício e do Esporte. Igualmente, é ampla a utilização de coeficiente de correlação intraclassa (ICC) para verificação da reprodutibilidade das medidas realizadas nos experimentos desta área (HOPKINS et al., 2010). Dentre as teses e artigos analisados notou-se a predominância do uso do Coeficiente de Correlação Linear Produto-Momento de Pearson. De acordo com Callegari-Jacques (CALLEGARI-JACQUES, 2003), o Coeficiente de Correlação Produto-Momento de Pearson consiste no instrumento mais adequado para medir a associação entre os dados de duas variáveis quantitativas, com distribuição normal. O resultado deste teste não informa, no entanto, se há relação de causa-efeito, apenas indica se as variáveis em questão variam juntas. É sempre aconselhável que se considere a possibilidade de que haja outros fatores influenciando a variação das variáveis em questão. Nestes casos, o uso dos modelos de regressão podem ser úteis e constituem os mais apropriados. Contudo, a utilização de modelos de regressão foi pequena entre os estudos analisados, tanto nas teses, quanto nos artigos, o que indica baixa prevalência de trabalhos preditivos na área de estudo das Ciências do Exercício e do Esporte.

Com um baixo índice de utilização nas teses e nos artigos analisados, o tamanho de efeito vem sendo amplamente solicitado pelas revistas científicas e tido como essencial na apresentação dos resultados das variáveis principais dos estudos (ESPÍRITO-SANTO; DANIEL, 2015; NAKAGAWA; CUTHILL, 2007). De acordo com a última edição do Manual da Associação Americana de Psicologia (APA, 2010), deve-se incluir a apresentação do tamanho de efeito e os seus intervalos de confiança mesmo quando se aplicam estatísticas inferenciais na análise. O tamanho do efeito se refere à estimativa da magnitude da relação entre

variáveis, ou do efeito de uma variável sobre a outra, ou da diferença entre duas amostras (HEDGES, 1981). Dentre os estudos analisados, o método mais utilizado para determinação do tamanho de efeito foi o d de Cohen. Este método é apropriado para utilização tanto com amostras independentes, quanto com amostras pareadas (COHEN, 1992), tendo sido, desta forma, adequadamente adotado pelos quatro artigos analisados. Sua grande aplicabilidade justifica a alta prevalência de utilização como método preferencial para determinação do tamanho do efeito nos estudos da área das Ciências do Exercício e dos Esportes.

Poucos estudos objetivando validar instrumentos de pesquisa foram encontrados na análise das teses e dos artigos. Para esta finalidade, o modelo de consistência interna Alpha de Cronbach foi adequadamente utilizado nos estudos analisados, considerando que Hopkins et al. (HOPKINS et al., 2009) recomendam a utilização deste modelo estatístico quando se objetiva validar instrumentos na área das Ciências do Exercício e da Medicina Esportiva. Este método consiste na determinação de um limite superior para a correlação de validade e confiabilidade do instrumento em questão (FIELD, 2009; HOPKINS et al., 2009). A baixa prevalência de estudos de validação de instrumentos, dentre os trabalhos incluídos, pode ser justificada pela elevada quantidade de instrumentos já validados disponíveis para utilização nas avaliações comumente realizadas na área das ciências do exercício e dos esportes.

De forma similar, foram poucos os estudos que realizaram análise inferencial de dados categóricos. Neste contexto, o teste qui-quadrado foi o mais utilizado quando se buscou comparar as frequências absolutas de observações de duas variáveis categóricas. Este teste é apropriado para verificar a existência de associação entre duas variáveis categóricas (FIELD, 2009), ou comparar as proporções entre duas ou mais populações quanto a uma variável categórica de interesse (CALLEGARI-JACQUES, 2003), tendo sido, desta forma, corretamente adotado nas teses e nos artigos do presente estudo. Sua ampla utilização pode ser justificada por suas características principais, como: (i) aplicabilidade a dados categóricos dicotômicos (com duas opções de categorias de respostas); (ii) aplicabilidade a dados categóricos politômicos (com três ou mais categorias de respostas); (iii) possível utilização tanto para dados independentes, quanto para dados pareados; (iv) fácil execução dos seus cálculos, mesmo quando realizados em procedimentos manuais.

A obtenção desta análise descritiva dos principais métodos utilizados permitiu conhecer o perfil dos trabalhos desenvolvidos no doutorado deste programa de pós-graduação e confirmar a hipótese da predominância de (i) estudos transversais, (ii) uso de estatística

inferencial, (iii) testes paramétricos, (iv) métodos de comparações entre médias e medianas. Além disso, os achados do presente estudo demonstraram algumas fraquezas metodológicas dos trabalhos, como (i) a falta de informações acerca da realização ou não de testes para verificar a normalidade dos dados e a homogeneidade das variâncias, quando estes eram pressupostos essenciais para os testes principais; (ii) a ausência de cálculos do tamanho de efeito em quase totalidade dos estudos; (iii) a baixa incidência de estudos preditivos com uso de modelos de regressão; (iv) a ausência de testes mais robustos na análise de medidas repetidas (equações de estimativas generalizadas, modelos mistos, dentre outros). Essas limitações podem ser reflexo (i) do ensino falho da bioestatística nas disciplinas do curso de doutorado, havendo insuficiente abordagem das questões citadas, ou (ii) da pouca motivação dos pesquisadores em realizar as análises da forma mais completa e robusta o possível, ou ainda (iii) do descuido ou da falta de interesse dos pesquisadores em relatar de forma completa os procedimentos realizados em seus experimentos.

Por outro lado, evidenciou-se pontos fortes na análise dos estudos. Em geral, observou-se que os métodos estatísticos adotados nas teses e nos artigos estão de acordo com os objetivos propostos nos estudos, e foram adequadamente planejados para os desenhos dos trabalhos, respondendo as suas questões de pesquisa. Além disso, o uso de análises mistas, com testes paramétricos somados a testes não-paramétricos, no intuito de melhor confirmar ou refutar todas as hipóteses do estudo, adotando os procedimentos adequados de acordo com a distribuição dos dados, deve ser ressaltado.

Ainda, é possível destacar algumas limitações do presente estudo, como a análise de trabalhos oriundos de um único curso de doutorado da área das ciências do exercício e do esporte, unicamente coletados no estado do Rio Grande do Sul, a inclusão de teses defendidas durante apenas 10 anos, a ausência de análise de dissertações de mestrado, bem como a falta de análise das teses inexistentes no repositório digital, e, por fim, a falta de análise de alguns artigos científicos publicados aos quais não foi possível obter acesso ao texto na íntegra.

5. CONCLUSÃO

Os resultados do presente estudo permitiram conhecer o perfil metodológico dos trabalhos desenvolvidos no curso de doutorado, no que concerne aos procedimentos estatísticos adotados. Dentre os principais achados, verificou-se a predominância de estudos com desenho transversal, a preferência pelo uso da estatística inferencial, com predominância de testes

paramétricos. Ainda, evidenciou-se a utilização preponderante de métodos de comparações entre médias e medianas, e trabalhos de associações entre variáveis, com uso de correlações. De forma geral, os métodos estatísticos adotados nas teses e nos artigos publicados são adequados e coerentes com os objetivos propostos nos estudos. Contudo, foram identificadas algumas limitações importantes nos estudos analisados, como a falta de uma descrição adequada da realização das testagens de pressupostos aos testes das análises principais (normalidade e homogeneidade), a baixa utilização de medidas de tamanho de efeito e de estudos preditivos com uso de modelos de regressão, além da ausência de testes mais robustos na análise de medidas repetidas.

REFERÊNCIAS

ALTMAN, D. & BLAND, M. Parametric v non-parametric methods for data analysis. **British Medical Journal**, v. 338, p. a3167, 2009.

_____. Standard deviations and standard errors. **British Medical Journal**, v. 331, p. 903, 2005.

AMERICAN PSYCHOLOGICAL ASSOCIATION (APA). **Publication of manual of the American Psychological Association**. 6. ed. Washington DC: APA; 2010.

CALLEGARI-JACQUES, S. **Bioestatística: princípios e aplicações**. Porto Alegre: Artmed, 2003.

COHEN, J. A power primer. **Psychological Bulletin**, v. 112, n. 1, p. 155-159, 1992.

ESPÍRITO-SANTO, H & DANIEL, F. Calcular e apresentar tamanhos do efeito em trabalhos científicos (1): as limitações do $p < 0,05$ na análise de diferenças de médias de dois grupos. **Revista Portuguesa de Investigação Comportamental e Social**, v. 1, n. 1, p. 3-16, 2015.

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

GREENLAND, S. et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. **European Journal of Epidemiology**, v. 31, p. 337-350, 2016.

HEDGES, L. Distribution theory for glass's estimator of effect size and related estimators. **Journal of Educational and Behavioral Statistics**, v. 6, n. 2, p. 107-128, 1981.

HOPKINS, W. et al. Progressive statistics for studies in sports. **Medicine and Science in Sports and Exercise**, v. 41, n. 1, p. 3-12, 2009.

KESELMAN, H. et al. Statistical practices of educational researchers: an analysis of their Anova, Manova, and Ancova analyses. **Review of Educational Research**, v. 68, p. 350-86, 1998.

LUNET, N.; SEVERO, M.; BARROS, H. Desvio padrão ou erro padrão. **Arquivos de Medicina**, v. 20, n. 1, p. 55-59, 2006.

NAKAGAWA, S.; CUTHILL, I. Effect size, confidence interval and statistical significance: a practical guide for biologists. **Biological Reviews of the Cambridge Philosophical Society**, v. 82, p. 591-605, 2007.

SIEGEL, S.; CASTELLAN J. R. **Estatística não-paramétrica para ciências do comportamento**. 2. ed. Porto Alegre: Artmed; 2006.

STRASAKA, A. et al. The use of statistics in medical research: a comparison of the New England Journal of Medicine and Nature Medicine. **The American Statistician**, v. 61, n. 1, p. 47-55, 2007.

TEIXEIRA, I. P. et al. Uso da estatística na educação física: análise das publicações nacionais entre os anos de 2009 e 2011. **Revista Brasileira de Educação Física e Esporte**, v. 29, p. 139-47, 2015.

TRESSOLDI, P. et al. High impact = high statistical standards? Not necessarily so. **PLoS One**, v. 8, n. 2, p. e56180, 2013.

VETTER, T. Fundamentals of research data and variables: the devil is in the details. **Anesthesia & Analgesia**, v.1 p. 1-6, 2017.