

# BASES METODOLÓGICAS PARA PESQUISA EM FISIOTERAPIA: Descrição e Sumarização da Informação

## *Methodological Basis for Researches in Physical Therapy: Description and Summarization of Information*

Lidiane Isabel Filippin<sup>1</sup>, Mario B. Wagner<sup>2</sup>

### Resumo

Ter conhecimento dos processos de descrição e sumarização de dados é fundamental para poder acompanhar a literatura científica e desenvolver uma avaliação crítica da informação que se recebe. Este artigo busca, de forma sucinta, esclarecer os aspectos da descrição e sumarização da informação na área biomédica.

### Palavras-chave:

descrição, sumarização

### Abstract

Understanding the process of data summarization is essential for health care professionals interested in critically appraising health information. This paper aims to present, in a condensed format, key points regarding the description and summarization of health information.

### Key words:

description, summarization

## INTRODUÇÃO

A Bioestatística é a aplicação da estatística na área biomédica. Ela é essencial ao planejamento, coleta, avaliação e interpretação dos dados obtidos em pesquisa e é fundamental à epidemiologia e à prática baseada em evidência.

Para pesquisadores é redundante falar na relevância da estatística: ela é importante tanto no momento do planejamento experimental, quanto na hora da análise dos dados. No entanto, muitos profissionais, especialmente aqueles mais próximos da prática clínica e distantes da pesquisa, tendem a crer que têm pouco a ver com a estatística.

Dois erros, no mínimo, se relacionam com esta idéia. Primeiro, esse profissional não poderá acompanhar corretamente a evolução da pesquisa em sua área de interesse, que faz uso permanente da estatística. Segundo, algumas noções de métodos estatísticos podem permitir ao profissional não somente tomar consciência da variabilidade dos dados com os quais lida, como também favorecer um melhor entendimento e controle desta variabilidade.

Este artigo busca esclarecer alguns conceitos básicos de Bioestatística a fim de facilitar a leitura de artigos científicos.

### 1. Variáveis e Dados

As pesquisas biomédicas podem ser consideradas estudos de relações entre variáveis. Comumente são estudadas as diferenças ou associações entre variáveis observa-

das em um paciente ou em uma população (CALLEGARI-JACQUES, 2004). Portanto, *variável* é toda característica que, quando observada, é passível de modificar-se.

Chame-se *dado* todo o valor resultante da mensuração de uma variável (KIRKWOOD E STERNE, 2003; CALLEGARI-JACQUES, 2004). A principal divisão ocorre entre variáveis quantitativas e qualitativas (categóricas).

### 2. Tipos de variáveis e seus níveis de medida

As variáveis podem ser classificadas como *dependente* (desfecho de interesse) e *exposição* (fator em estudo) a qual pode ser associada ou não com o desfecho (FLETCHER E FLETCHER, 2006). Por exemplo, se um pesquisador desejasse avaliar a associação entre câncer de esôfago e o consumo de chimarrão, o consumo de chimarrão seria a exposição (fator em estudo) e o câncer de esôfago seria o desfecho.

Tanto a variável dependente como a exposição podem assumir diferentes valores de acordo com as características dos dados gerados. Dessa forma, as variáveis podem ser classificadas segundo o nível de medida, para facilitar sua compreensão e tratamento estatístico.

### Variáveis Qualitativas ou Categóricas

*i. Variável Nominal:* neste nível de medida diferencia-se uma categoria da outra simplesmente por meio de uma denominação. Essas variáveis podem ser binomiais, binárias ou dicotômicas, quando composta de duas cate-

1. Fisioterapeuta Doutoranda Pós-Graduação em Ciências Médicas Faculdade de Medicina, Universidade Federal do Rio Grande do Sul.

2. Doutor em Epidemiologia (Universidade de Londres) Professor Adjunto, Departamento de Medicina Social, Faculdade de Medicina, Universidade Federal do Rio Grande do Sul.

Recebido: 07/01/2010

Aceito: 19/01/2010

Autor para correspondência: Lidiane Isabel Filippin

E-mail: l.filippin@terra.com.br

gorias (sim/não; Rh+/Rh-) e polinomiais ou politômicas, quando apresentam mais de duas categorias (sistema sanguíneo ABO).

ii. *Variável Ordinal*: existe uma ordem entre as várias categorias. Exemplos de variáveis ordinais são escores (sensação dolorosa – Escala Analógica de Dor, onde zero indica nenhuma dor e 10, uma dor insuportável).

### Variáveis Quantitativas

i. *Variáveis quantitativas discretas*: são variáveis onde os dados somente podem apresentar valores inteiros. Exemplos deste tipo de variável são o número de filhos e baixas hospitalares por paciente.

ii. *Variáveis quantitativas contínuas*: são aquelas cujos dados podem apresentar qualquer valor dentro de um intervalo possível de variação. Exemplos desta variável são peso, altura, tempo, dosagem de colesterol, goniometria.

## 3. Principais sumarizadores numéricos

### 3.1. Medidas de tendência central

A estatística descritiva objetiva sumarizar um conjunto de dados. A média, a mediana e a moda são medidas de tendência central para uma distribuição. Cada uma delas possui suas particularidades e são usadas em situações distintas.

A *média* possui o maior poder matemático e é a medida descritiva indicada em dados quantitativos. Seu uso é recomendado quando os dados possuem uma distribuição simétrica, pois é afetada por valores extremos (KIRKWOOD E STERNE, 2003; CALLEGARI-JACQUES, 2004).

A *mediana* é uma medida de posicionamento que representa o valor que ocupa o meio da série, ou seja, 50% dos valores que estão abaixo e 50% acima da mediana. Por este fato, é a medida de tendência central recomendada para distribuições assimétricas.

A *moda*, pouco utilizada, somente descreve o valor mais frequente (KIRKWOOD E STERNE, 2003; CALLEGARI-JACQUES, 2004).

### 3.2. Medidas de Dispersão

As medidas de dispersão informam a variabilidade da amostra em relação à medida de tendência central, ou seja, o quanto os dados de uma série estão agrupados em torno da tendência central.

As medidas de dispersão mais comuns são amplitude, amplitude interquartil, variância e desvio padrão.

A *amplitude* é considerada uma medida simples e pouco informativa, pois se refere apenas aos valores, máximo e mínimo, dos dados. Por exemplo, na série de dados 22, 39, 19, 30 a amplitude é  $39 - 19 = 20$ . Apesar de ser um único número, frequentemente a amplitude não é calculada e são apresentados os valores mínimo e máximo observados na série (CALLEGARI-JACQUES, 2004).

A *amplitude interquartil (AIQ)*. Derivada da expressão inglesa *interquartile range*, indica a dispersão (amplitude) da metade (50%) da distribuição (KIRKWOOD E STERNE, 2003). Deve-se lembrar que a mediana divide a série em duas partes iguais: o quartil inferior também é conhecido como percentil 25 (P25) e é a mediana da metade inferior;

o quartil superior, ou percentil 75 (P75), é a medida da metade superior. Assim, a AIQ é a diferença entre P25 e P75 e envolve os dados compreendidos nos 50% centrais da série.

*Variância e desvio padrão*. A variância e o desvio padrão são considerados as medidas mais representativas da dispersão ou variabilidade dos dados, visto que levam em conta todos os elementos da série (KIRKWOOD E STERNE, 2003). Como a variância é expressa em unidades ao quadrado seu uso é bem mais restrito. Já o desvio padrão (DP), em palavras simples, representa qual a “oscilação média” para mais ou para menos que os valores de uma série apresentam ao redor da média. Geralmente em um conjunto de dados quantitativos é possível afirmar que entre a  $média \pm 1DP$  encontram-se 68% dos valores e que entre a  $média \pm 2DP$  temos 95% dos valores.

## Considerações Finais

Não existe uma técnica estatística aplicável a todas as situações, por isso a necessidade do leitor ter noções básicas sobre a bioestatística. No que se refere à descrição e sumarização de informações, pode-se dizer que, de forma geral, quando os dados forem simétricos a média é a melhor medida de tendência central acompanhada do desvio padrão como medida de dispersão. Nesta situação, os dados são frequentemente apresentados como  $média \pm DP$ . Uma boa indicação de simetria é quando o valor do desvio padrão não ultrapassa a metade do valor da média. No caso de assimetria dos dados, utiliza-se a mediana, uma vez que ela não é influenciada por dados discrepantes. Adicionalmente utiliza-se como medida de dispersão a AIQ ou P25 e P75.

Por fim, pode-se dizer que os procedimentos paramétricos são baseados na média e no desvio padrão e são geralmente utilizados em variáveis quantitativas. Por outro lado, quando as variáveis são categóricas empregam-se procedimentos e testes não-paramétricos.

## REFERÊNCIAS

- Callegari-Jacques, S. M. Bioestatística Princípios e Aplicações. Porto Alegre: ARTMED Editora. 2004
- Fletcher, R. W. e S. E. Fletcher. Epidemiologia Clínica - Elementos Essenciais. London: Artmed. 2006
- Kirkwood, B. R. e J. A. C. Sterne. Essential Medical Statistics. 2003.