DBR
DEPARTAMENTO DE BIOFISICA E RADIOBIOLOGIA

V BIOFISICA
ENCONTRO ANUAL 2018

ESTATÍSTICA BÁSICA EM ARTIGOS PUBLICADOS NAS REVISTAS BIOFÍSICAS: INCOSISTÊNCIAS E ANÁLISE HISTÓRICA

Otacilio Antunes Santana[1,2]*, Mayara Lopes de Freitas Lima[1], Marcela Karolinny da Silva Costa[1], Raquel Bernardo de Melo[2], Clodoaldo de Lima[2], Carla Valéria de Miranda Costa Duarte[2], Susana Carvalho de Souza[2,3], Paulo Euzébio Cabral Filho[1,2]

[1]Departamento de Biofísica e Radiobiologia, Centro de Biociências, UFPE
[2]Mestrado Profissional em Rede Nacional para Ensino das Ciências Ambientais - PROFCIAMB, UFPE
[3]Departamento de Micologia, Centro de Biociências, UFPE
*otacilio.santana@ufpe.br

INTRODUÇÃO

Antes da década de 60 (séc. XX), todas as análises estatísticas efetuadas para publicações técnicas e científicas eram realizadas de forma manual, com auxílio eventual de calculadoras mecânicas, inventadas em 1623, e de calculadoras eletrônicas, inventadas em 1957 (IFRAH et al., 2000). O primeiro software lançado com alcance à comunidade científica foi em 1968 o *Statistical Package for the Social Sciences* (SPSS), Pacote Estatístico para as Ciências Sociais, desenvolvido para usuários com noções básicas em matemá tica, probabilidade e estatística e para encurtar a complexidade entre a entrada (*input*) e saída (*output*) dos dados (NIE; BENT; HULL, 1970). A partir de 1975, com a criação jurídica da SPSS Inc., esse pacote estatístico se disseminou, por conta da versão projetada para computadores da *International Business Machines* (IBM) e a *Israel Chemicals* Ltd. (ICL) (NORUSIS, 1993).

A popularização e a acessibilidade dos programas estatísticos registrados nos trabalhos acadêmicos começaram principalmente a partir da criação das Planilhas Eletrônicas e da disseminação dos microcomputadores – *desktop* e *notebook* (década de 90; KVANLI, 1988). As planilhas eletrônicas continham funções como: medida de tendência central, análise entre grupos amostrais e análises entre variáveis (Microsoft Office Excel criado em 1987; LEVINE; BERENSON; STEPHAN, 1999). Outro marco importante, anterior aos pacotes estatísticos, está relacionado às revistas científicas das á reas de biológicas e saúde, solicitarem na estrutura dos artigos: i) o delineamento experimental que seja conduzido por uma hipótese estatística de significância, de proporcionalidade ou de agrupamento; e, ii) a análise estatística que sigam os princípios da experimentação (repetição, controle local e casualização).

As revistas acadêmicas, a partir da década de 90, já começavam a se preocupar com a produção de artigos no qual os autores conduziam suas pesquisas sem delineamentos experimentais pré-definidos, sem seguir os princípios da experimentação e, até mesmo, sem ter suficiência amostral (MARINO, 1995; SITAR; CHEANG, 1996; GOODMAN, 1999). Mais recentemente, os editores demonstraram preocupação com a utilização estatística sem cumprir as premissas básicas de cada teste e análise, gerando conclusões aceitas que em outro momento ou análise poderiam ser refutadas (GALE; HOCHHAUS; ZHANG, 2016; WASSERSTEIN; LAZAR, 2016). Muitas dessas revistas com escopo principal na Biofísica, uma área peculiar por sua interface entre as ciências exatas e as ciências biológicas (JOHNSON, 2013).

Com base nessas afirmações, a hipótese desse trabalho foi que a partir da popularização dos programas estatísticos para microcomputadores (década de 90) há uma crescente produção de inferências estatísticas em trabalhos acadêmicos que não seguem as premissas básicas de cada teste e análise estabelecida. Sendo assim, tem-se como objetivos deste trabalho: i) fazer um levantamento das revistas com maiores impactos com escopo na Biofísica; ii) analisar se os artigos dessa revista estão seguindo as premissas estatísticas em seus testes e análises; iii) verificar se há diferenças estatísticas entre a frequências de não cumprimento das premissas antes e depois da popularização dos programas estatí sticos.

MATERIAIS E MÉTODOS

A primeira etapa metodológica foi o levantamento das revistas com principal escopo na área do conhecimento da Biofísica (Biofísica Molecular, Biofísica Celular, Biofísica de Processos e Sistemas, e, Radiologia e Fotobiologia), revistas que estavam indexadas e que continham Fator de Impacto relatado pela *Journal Citation Reports* – JCR 2016 (Thomson Reuters, 2017). Fator de Impacto de um perió dico é calculado como o número médio de citações dos artigos que foram publicados durante o biênio anterior (GARFIELD, 2006). Duzentos artigos para cada revista escolhida foram avaliados antes e depois do ano de 90 (n = 7.200 artigos, 18 revistas: 200 antes e 200 depois), recuperados de forma aleatória em cada período. Depois de 90, os artigos recuperados foram até a data de 30 de novembro de 2017.

A etapa seguinte foi avaliar nos artigos publicados se haviam ou nã o o cumprimento das premissas estatísticas, a se basear no livro *Biostatistical Analysis* de Jerrold H. Zar (1999), o livro mais citado da história em trabalhos científicos: 86.974 citações (Google Acadê mico, 2017). Situações e testes (*e.g.* inconsistência na significâ ncia) foram classificados e registrados suas frequências. Depois dessa avaliação, uma análise de Qui-quadrado foi executada para as frequências antes e depois do ano de 90.

RESULTADOS

O número de revistas encontradas com escopo principal na área de conhecimento da Biofísica que estavam indexadas e que continham fator de impacto relatado pela JCR – 2016 foi de 18 (Tabela 1). A revista mais antiga amostrada foi a *Archives of Biochemistry and*

ANAIS: ENCONTRO ANUAL DA BIOFÍSICA (2018): 77-80
BIOFÍSICA E RADIOBIOLOGIA, BIOCIÊNCIAS, UFPE
05 E 06 DE FEVEREIRO, RECIFE, PERNAMBUCO, BRASIL
ISSN: 2526-6071
DOI: 10.5151/biofisica2018-027

Biophysics, fundada em 1951, e a de maior fator de impacto registrado em 2016 foi a *Annual Review of Biophysics* (10,676).

Tabela 1. Revistas da área da Biofísica com ISSN (*International Standard Serial Number*), maiores fatores de impacto (JCR, 2016) e ano de início de suas publicações.

Revista	ISSN	Fator de Impacto JCR	Ano de Início
Annual Review of Biophysics	1936-122X	10.676	1972
Quarterly Reviews of Biophysics	0033-5835	5.267	1968
Biochimica et Biophysica Acta - Bioenergetics	0005-2728	4.932	1967
Biophysical Journal	0006-3495	3.656	1960
Progress in Biophysics & Molecular Biology	0079-6107	3.227	1963
Archives of Biochemistry and Biophysics	0003-9861	3.165	1951
Biochemical and Biophysical Research Communications	0006-291X	2.466	1959
Biophysical Chemistry	0301-4622	2.402	1973
Radiation and Environmental Biophysics	0301-634X	2.398	1963
BMC Biophysics	2046-1682	2.292	2011
Acta Biochimica et Biophysica Sinica	1672-9145	2.200	2004
Food Biophysics	1557-1858	1.704	2006
European Biophysics Journal with Biophysics Letters	0175-7571	1.472	1974
Cell Biochemistry and Biophysics	1085-9195	1.320	1979
General Physiology and Biophysics	0231-5882	1.170	1982
Indian Journal of Biochemistry & Biophysics	0301-1208	0.827	1964
Doklady Biochemistry and Biophysics	1607-6729	0.471	2000
Progress in Biochemistry and Biophysics	1000-3282	0.341	1974

Os artigos recuperados e examinados apresentaram inconsistências em aproximadamente nove premissas estatísticas nos dois períodos avaliados (Figura 1): i) na delimitação da hipótese; ii) na suficiência amostral (S.A.) e no consequentemente desdobramento na premissa da distribuição normal dos dados; iii) no estabelecimento da medida de tendência central (MTC) que representa a população ou grupo amostral; iv) na importância da variância dos dados (desvio padrão e coeficiente de variação); v) na interpretação da significação do *p* em relação ao *n* amostral e em relação ao efeito causado pelo tratamento; vi) na relação de causa e efeito entre as variáveis correlacionadas; vii) na acumulação de erros em ajustes dos dados e suas interpretações; viii) na linearidade de relação das variáveis biológicas sem uma visível saturação; e ix) na discussão que foge do delineamento experimental proposto. Outras inconsistências apareceram, principalmente a partir dos artigos avaliados no período acima da década de 90.

Entre os dois períodos avaliados, o número de artigos com inconsistência estatística foi maior ($p < 0,001$) depois de 1990 (1.710) do que antes (144). Isso pode estar relacionado principalmente ao fato de que com a utilização dos pacotes estatísticos e a simplicidade na execução das análises, os usuários não se aprofundaram em construir um conhecimento sobre as premissas estatísticas. Os autores na escolha da análise estatística a ser aplicada pelo seu conjunto de dados preferem utilizar a estatística mais frequentemente empregada por outros autores de trabalhos já publicados na área, muitas vezes até reproduzindo e replicando erros e inconsistências. Outro destaque é que a maioria dos trabalhos acadêmicos só definem seus critérios estatísticos pós-coleta de dados, o que faz, na maioria dos casos, uma inversão conceitual, os dados que se adaptaram aos testes e não o contrário (*e.g.* a busca para aceitar uma hipótese alternativa). Outra possível causa, que não está relacionada com a delimitação do teste, é o *n* amostral nos trabalhos científicos que acabam sendo limitados por protocolos estabelecidos, comitês de ética ou falta de recurso para replicar experimentos.

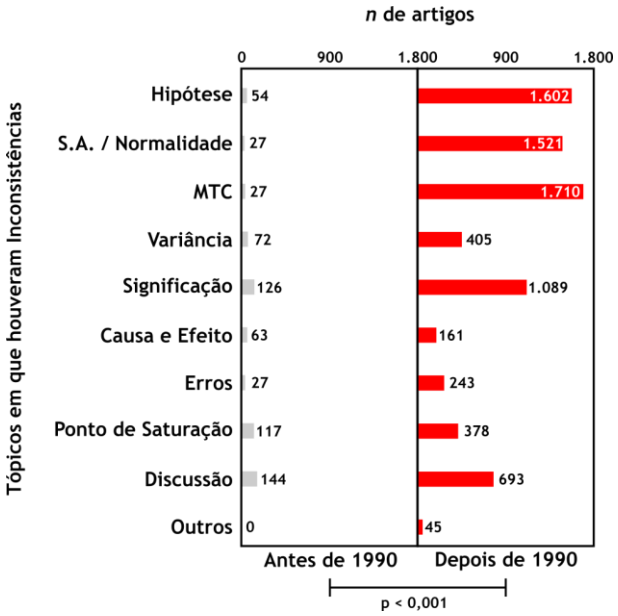


Figura 1. Tópicos em que houveram inconsistências nas premissas estatísticas aplicadas, antes e depois do ano de 1990 - p = resultado do Teste de Qui-Quadrado, S.A. = suficiência amostral e MTC = medida de tendência central.

A maioria dos artigos avaliados não descreveram explicitamente hipóteses ou delineamentos experimentais. Isso principalmente depois de 90 (1.602/1.800 artigos), os autores se preocuparam em querer uma aceitação do que era proposto (*e.g.* efeito do medicamento) do que em uma interpretação do efeito nos *n* amostrais. Porém, no conjunto dos dados, quando a maioria se comporta de maneira previsível ou casualizada, o importante são as exceções (ZAR, 1999). A delimitação do delineamento experimental também pode ajudar aos autores e aos leitores entenderem os possíveis vieses de condução da experimentação e de toda casualização, controle local e repetição da amostragem. Talvez essas lacunas possam demonstrar a falta de domínio dos autores nos fundamentos estatísticos.

Outra carência nos artigos é de uma discussão sobre a suficiência amostral dos dados (1.512/1.800 artigos depois de 90). O que define aquele *n* amostral utilizado pela maioria dos artigos? Ningué m se sabe, acabou sendo um padrão literário. E isso cria uma dú vida nos leitores se a partir de um conjunto de dados que não se sabe se satisfaz uma suficiência amostral, a distribuição da frequê ncia dos dados sobre as classes de valores encontradas se comporta de maneira normal, pré-requisito fundamental para as variáveis biológicas e biométricas. Sem uma distribuição normal dos dados, ou até mesmo uma distribuição dos dados sem se saber a suficiê ncia amostral não se tem como definir critérios de parametria ou intervalo de confiança.

Os autores por conta própria definem se seus dados são paramé tricos ou não, e escolhem qual a medida de tendência central irá representar seus dados. Geralmente, os autores escolhem a média aritmética dos dados com sua representante, mas como escolher a média se por vezes a maior frequência dos dados ou nenhuma ocorrência está sob ela no eixo cartesiano. Na Figura 2, se vê duas situações, uma em que a média não representa o conjunto dos dados (Figura 2A), e uma em que a média representa o conjunto dos dados (Figura 2B). E isso é importante no momento da decisão da parametria e do uso de um teste de média ou não (ver Equação 1: Teste paramétrico t - Teste de média e Equação 2: Teste não paramétrico U - Teste de mediana), o que as próprias funções dos testes (Equação 1) já nos indicam e que os *inputs* e *outputs* dos

pacotes estatísticos não revelam explicitamente. Essa foi a inconsistência mais registrada quando da utilização das premissas estatísticas nos artigos avaliados (1.710/1.800 artigos depois de 90).

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$ Equação 1

Teste t

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - \mathrm{T}$$ Equação 2

Teste Mann-Whitney

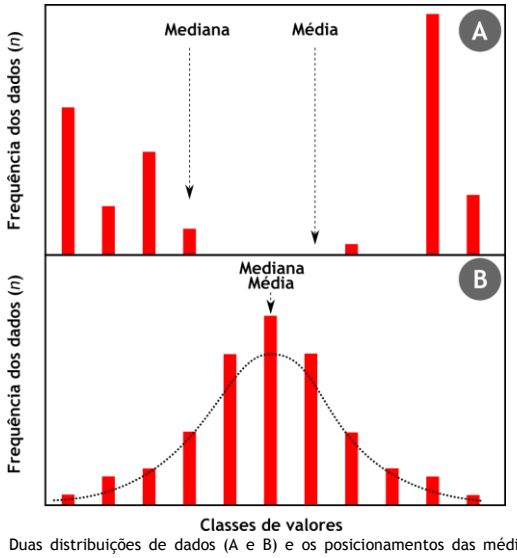


Figura 2. Duas distribuições de dados (A e B) e os posicionamentos das médias e das medianas no eixo cartesiano para cada uma delas.

Em conexão com essas inconsistências encontradas nos artigos, outra análise que gera dúvida, principalmente na revelação dos resultados e na discussão dos artigos, foi que com a média e seu desvio poderia se estabelecer alguma significância ou diferença entre grupos amostrais. Dois grupos amostrais podem ter a mesma média ou os mesmos desvios e possuírem uma distribuição dos dados de maneiras distintas, ou vice-versa. Na Figura 3, as situaçõ es A e B podem, sem uma análise de distribuição, levar a interpretações equivocadas de semelhança ou significação nas hipó teses estatísticas. E isso foi nos artigos, principalmente a partir de 90 (405/1.800 artigos).

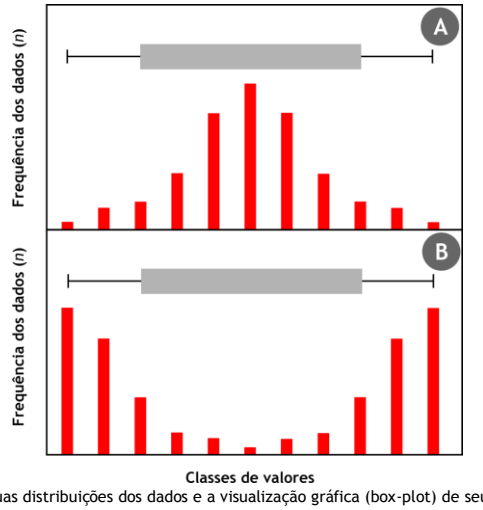


Figura 3. Duas distribuições dos dados e a visualização gráfica (box-plot) de seu desvio.

A variância dos dados biológicos é uma característica natural da variável estudada, porém alguns vieses são notados, como por exemplo, a posição do dado na matriz ou na posição de coleta (Figura 4). Se uma placa de Petri está próxima à porta da estufa (Situação 1), ela terá algum resultado diferente da placa de Petri no fundo da estufa (Situação 2)? Essa avaliação da variância, do desvio padrão e do coeficiente de variação é fundamental para eliminações de efeitos que não seja os do tratamento. A variância também pode decidir o resultado da significância de uma análise estatística. Médias distintas entre dois grupos amostrais podem resultar em diferenças significativas (Figura 5A) e não significativas (Figura 5B). Porém, dois grupos amostrais sob a mesma média, podem resultar em diferenças significativas, levando em consideraç ão a variância dos dados de um grupo.

Figura 4. Duas posições na matriz de cálculo estatístico.

A inserção de cada dado (*n* acumulado) na sua coleta, faz com que o valor de p em uma relação intergrupos se comportem ou para uma significância (aceitação da hipótese alternativa – 1ª Situação – Figura 5) ou para uma não significância (aceitação da hipótese nula alternativa – 3ª Situação – Figura 5).

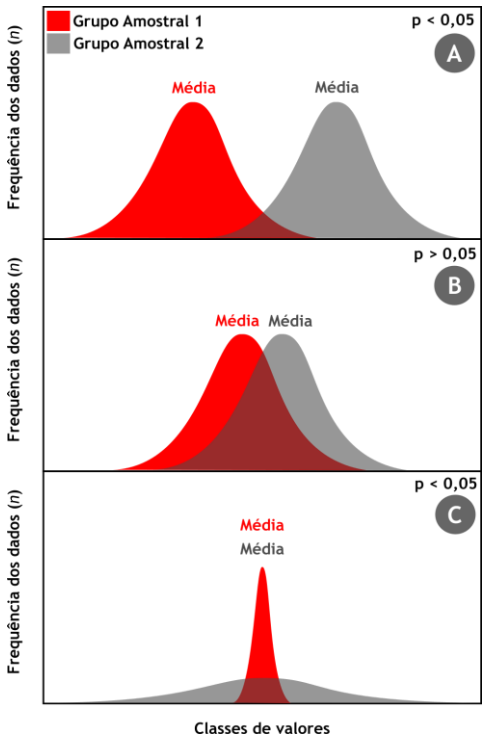


Figura 5. Três situações de significância estatística em relação a média.

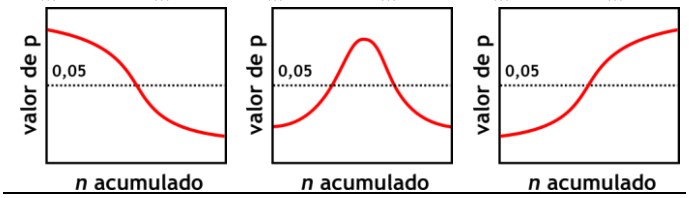
E há um outro caso em que com o passar da coleta dos dados, com pouco *n* amostral não há significância, a partir de um *n* amostral, há significância, e depois se torna na significativa a relação intergrupos (2ª Situação – Figura 5). Essas situações exemplificadas na Tabela 2 mostram que se deve prestar atenção de como a partir do acréscimo de dados na matriz isso influenciará no valor de p. Geralmente, os autores dos artigos registram apenas o valor do p

do seu conjunto dos dados, não registrando se em uma replicação da pesquisa resultará com um *n* maior ou menor terá a mesma forç a de aceitação ou refutação da hipótese. Com essa variação do valor de p, se pode ter uma ideia, por exemplo, da proximidade do valor $F_{calculado}$ em relação ao $F_{tabelado}$ e o pesquisador referendar sua significância não por uma eventualidade, mas por um forte probabilismo. 1.089/1.800 artigos avaliados, depois de 90, tiveram inconsistências quanto a análise de significância da hipótese (valor de p).

Na relação entre variáveis, cinco inconsistências foram destaques nos artigos avaliados: i) a falta de uma relação de causa e efeito entre as variáveis (*e.g.* colocar o tempo como uma variável independente em algumas relações) (161/1.800 artigos depois de 90); ii) a multicolinearidade entre duas variáveis, ou seja, variáveis que explicam a mesma coisa; iii) a falta de um ponto de saturação da relação (*e.g.* relações biométricas lineares ou que seguem ao infinito de forma linear) (378/1.800 artigos depois de 90) (Figura 6A); iv) a eliminação do erro nas equações ou na discussão pós ajuste dos dados (243/1.800 artigos depois de 90); e, v) a discussão de uma sub ou super estimação do ajuste dos dados (Figura 6B).

Tabela 2. Situações do comportamento do valor do p com a acumulação do dado.

1ª Situação		2ª Situação		3ª Situação	
n	p	n	p	n	p
5	0,0778	5	0,0454	5	0,0119
6	0,0767	6	0,0460	6	0,0147
7	0,0724	7	0,0474	7	0,0164
8	0,0678	8	0,0482	8	0,0177
9	0,0603	9	0,0523	9	0,0191
10	0,0585	10	0,0537	10	0,0197
11	0,0564	11	0,0634	11	0,0212
12	0,0558	12	0,0674	12	0,0268
13	0,0532	13	0,0701	13	0,0274
14	0,0505	14	0,0670	14	0,0315
15	0,0487	15	0,0583	15	0,0374
16	0,0465	16	0,0524	16	0,0415
17	0,0464	17	0,0491	17	0,0424
18	0,0452	18	0,0479	18	0,0444
19	0,0330	19	0,0464	19	0,0510
20	0,0327	20	0,0433	20	0,0536
21	0,0300	21	0,0421	21	0,0576
22	0,0298	22	0,0402	22	0,0725
23	0,0287	23	0,0399	23	0,0757
24	0,0248	24	0,0295	24	0,0832
...	...	...	...	...	...



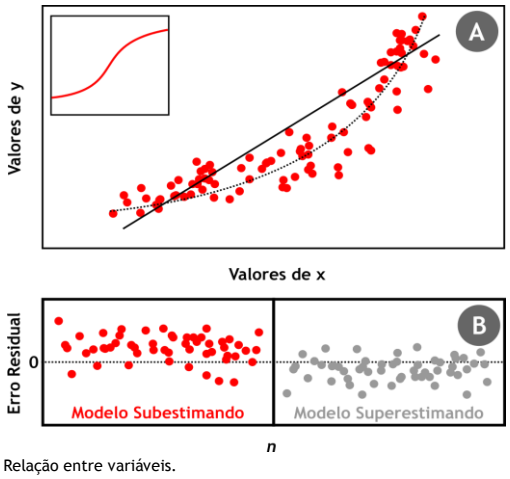


Figura 6. Relação entre variáveis.

Por fim, a inconsistência de se construir uma discussão que não seja pela hipótese estatística e pelo delineamento experimental estabelecidos abre margem a um enviesamento daquilo que o autor quer que seja aceito. Os autores esquecem o método e suas limitaç ões, esquecem dos probabilismos e discutem deterministicamente. Esse tópico foi o mais frequente dos artigos publicados antes do ano 90 (144/1.800 artigos) e teve uma representação dos artigos publicados depois de 1990 (643/1.800 artigos). Outras inconsistê ncias foram registradas depois de 90, como a plotagem de gráficos que não representavam a estatística (*e.g.* gráficos de linhas em que os grupos eram independentes) e interpretação errada do parâ metro (*e.g.* p > 0,05 = diferença significativa).

CONSIDERAÇÕES FINAIS

A hipótese trabalho foi aceita que a partir da popularização dos programas estatísticos para microcomputadores (década de 90) há uma crescente produção de inferências estatísticas em trabalhos acadêmicos que não seguem as premissas básicas de cada teste e análise estabelecida. Nove inconsistências estatísticas foram destaques nos artigos publicados nas revistas biofísicas apresentadas sendo a principal a escolha da medida de tendência central que represente o conjunto dos dados e seu desdobramento no uso e interpretação (suficiência amostral e normalidade). Os pacotes estatísticos ajudaram os trabalhos acadêmicos, porém se vê uma perda do conhecimento do fundamento estatístico certificado pelas inconsistências latentes nos artigos publicados mais recentemente.

REFERÊNCIAS

IFRAH, G.; HARDING, E. F.; BELLOS, D.; WOOD, S. **The universal history of computing**: From the abacus to quantum computing. Hoboken: John Wiley & Sons. 2000.

GALE, R. P.; HOCHHAUS, A.; ZHANG, M. Z. What is the (p-) value of the P-value? **Leukemia**, v. 30, p. 1965–1967, 2016. doi:10.1038/leu.2016.193

GARFIELD, E. The history and meaning of the journal impact factor. **Jama**, v. 295, n. 1, p. 90-93, 2006. doi:10.1001/jama.295.1.90

GOODMAN, S. N. Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy. **Annals of Internal Medicine**, v. 130, p. 995–1004, 1999. doi: 10.7326/0003-4819-130-12-199906150-00008

GOOGLE ACADÊMICO. **Busca.** Disponível em: https://scholar.google.com.br/. Acesso em 13 de nov. de 2017.

JOHNSON, V. E. Revised standards for statistical evidence. PNAS, v. 110, n. 48, p. 19313–19317, 2013. doi: 10.1073/pnas.1313476110

KVANLI, A. H. **Statistics**: a computer integrated approach. Eagan: West Publishing Co. 1988.

LEVINE, D. M., BERENSON, M. L., & STEPHAN, D. **Statistics for managers using Microsoft Excel**. Upper Saddle River: Prentice Hall. 1999.

MARINO, P. Survival curves is the P-value enough? **Lung Cancer**, v. 12, n. 1–2, p. 87-89, 1995. doi: 10.1016/0169-5002(94)00411-F.

NIE, N. H.; BENT, D. H.; HULL, C. H. **SPSS**: Statistical package for the social sciences (No. HA29 S6). New York: McGraw-Hill. 1970.

NORUSIS, M. J. **SPSS for windows**: advanced statistics, release 6.0. Chicago: SPSS inc. 1993.

SITAR, D. S.; CHEANG, M. Error in statistics program. **Journal of Clinical Epidemiology**, v. 49, n. 5, p. 605-625, 1996. doi: 10.1016/0895-4356(95)00035-6.

WASSERSTEIN, R. L.; LAZAR, N. A. The ASA's Statement on p-Values: Context, Process, and Purpose. **The American Statistician**, v. 70, n. 2, p. 129-133, 2016. doi: 10.1080/00031305.2016.1154108

ZAR, Jerrold H. et al. **Biostatistical analysis**. Pearson Education India, 1999.