

Utilização do Software R em pesquisas na terapia ocupacional

Maysa Marinho Antunes Ramos^a , Pedro Luiz Ramos^b , Francisco Louzada Neto^b ,
Patrícia Carla de Souza Della Barba^c 

^a Universidade Federal de São Carlos – UFSCar, São Carlos, SP, Brasil.

^b Universidade de São Paulo – USP, São Carlos, SP, Brasil.

^c Departamento de Terapia Ocupacional, Universidade Federal de São Carlos – UFSCar, São Carlos, SP, Brasil.

Resumo: O presente artigo tem como objetivo fornecer subsídios para que pesquisadores em terapia ocupacional possam realizar procedimentos de estatística básica de maneira mais flexível e independente, a partir do sistema R, um software livre e gratuito, cuja popularidade vem aumentando consideravelmente no âmbito acadêmico. Apresentar-se-á, assim, o passo a passo de como instalar e utilizar o programa na realização de leitura e sumarização de dados, bem como no cálculo de estatísticas básicas, nas representações gráficas de dados quantitativos e qualitativos, no cálculo do tamanho amostral e na efetuação de testes de hipóteses e de teste de correlação linear. O banco de dados utilizado nas demonstrações apresentadas é oriundo de uma pesquisa em terapia ocupacional e as análises aqui realizadas resultam de uma demanda compreendida no decorrer de uma disciplina obrigatória do Programa de Pós-Graduação em Terapia Ocupacional da Universidade Federal de São Carlos (UFSCar), por meio da qual foi possível levantar os principais procedimentos estatísticos utilizados pelos pesquisadores em suas pesquisas.

Palavras-chave: *Análise Estatística, Software R, Terapia Ocupacional.*

Using Software R in research in occupational therapy

Abstract: In this paper, it is presented a simple guide for researchers in occupational therapy to perform basic statistical analysis in a flexible and independent way, using the R software, that is a free open source software which its popularity has been increased considerably in many fields. We have presented a step-by-step guide about how to install such software, it is also discussed the necessary steps to include the data set and perform basic statistical analysis, such as the calculation of sample size, basic statistics, graphical presentation, hypothesis tests and the linear correlation test. The dataset considered in this study comes from a research in Occupational Therapy and the topics considered are result of the common statistical procedures that were face in the course of the Post Graduation Program in Occupational Therapy at the Federal University of São Carlos (UFSCar), in which were possible to find the principal statistical procedures used by the researchers in applications.

Keywords: *Statistical Analysis, Software R, Occupational Therapy.*

1 Introdução

A essencialidade da pesquisa na terapia ocupacional se dá, em parte, pela responsabilidade dos profissionais em basearem suas práticas em evidências científicas, visto que a escolha por ferramentas e estratégias de intervenção depende de precedentes confiáveis capazes de garantir uma validade maior do trabalho a ser realizado (KIELHOFNER, 2006).

Todavia, o caráter sistemático das pesquisas científicas implica uma gama de conhecimentos que ultrapassa os saberes comuns a cada área, demandando a apropriação de recursos externos para a realização e a compreensão de estudos suficientemente consistentes (SAMPAIO; MANCINI; FONSECA, 2002).

Dentre eles, fazem-se presentes os recursos estatísticos, fortemente demandados e, muitas vezes, evitados por pesquisadores de outras áreas.

Um artigo de Ottenbacher e Petersen (1985), publicado na *The American Journal of Occupational Therapy*, ao discutir as implicações do uso crescente de procedimentos quantitativos na literatura de terapia ocupacional, revelou que

[...] a expansão de uma literatura de pesquisa na profissão tem sido acompanhada por uma sofisticação emergente na utilização de modelos de pesquisa e análises estatísticas (OTTENBACHER; PETERSEN, 1985, p. 240).

Segundo Sampaio, Mancini e Fonseca (2002), terapeutas ocupacionais precisam ser produtores e consumidores críticos de informação. Entretanto, existe o fato de se mostrarem, muitas vezes, desencorajados quando o assunto é estatística, visto que muitos revelam ignorar a seção de análise estatística ao lerem artigos científicos (KIELHOFNER, 2006). Assim, é fundamental se engajar, sempre que possível, a aprender a lidar com esses recursos, visto que eles estão disponíveis para ajudar os pesquisadores e profissionais da prática nessa caminhada árdua, porém essencial para a consolidação da profissão.

Atualmente, existem softwares capazes de gerar estatísticas de forma rápida e acessível (KIELHOFNER, 2006). Além disso, mais do que praticidade, pode-se ter liberdade e gratuidade na realização de tais procedimentos, visto que a barreira tecnológica advinda com os softwares privados também impede um contato maior com essa etapa essencial a muitas pesquisas.

Recentemente, a utilização de softwares livres tem sido intensificada, inclusive pelo constante incentivo governamental. Além de focalizar a redução de custos, o aumento da concorrência

e a geração de empregos, o governo vislumbra, sobretudo, maior independência e colaboratividade na produção e difusão de conhecimentos necessários para o desenvolvimento tecnológico do país. Dados oriundos do Serviço Federal de Processamento de dados (Serpro) apontam uma economia do Governo Federal de, aproximadamente, R\$370 milhões com o uso de softwares livres nos últimos anos e esse número torna-se expressivo quando se têm inúmeras outras demandas a serem atendidas (COSTA, 2009).

Todavia, embora haja um movimento em prol desses softwares, ainda se despende muito dinheiro em licenças que, além de serem caras, possuem prazo de expiração. No âmbito acadêmico existe uma forte dependência dos pesquisadores em relação aos softwares pagos para a realização de análises estatísticas. Além do alto custo atribuído a eles, seu uso fica restrito a poucos computadores e aos espaços dos laboratórios que, muitas vezes, não podem ser acessados rotineiramente por todos.

Dentre as opções que vêm sendo disseminadas na comunidade científica para substituir softwares pagos, destaca-se aqui o R (R CORE TEAM, 2018), um software livre, gratuito, multiplataforma e expansível, que vem ganhando popularidade no âmbito acadêmico, podendo ultrapassar, nos próximos anos, o uso de softwares pagos como, por exemplo, o SAS, SPSS, Statistica, Minitab, entre outros. No entanto, vale ressaltar que nada impede uma instituição ou pesquisador de utilizar um software pago se assim o desejar, mas que essa, ou esse, possam ter à sua disposição outras possibilidades com vantagens mais evidentes e escolher aquela que mais se ajusta às suas necessidades.

Posto que não há literatura disponível que aborde o uso do R no âmbito da terapia ocupacional, o presente artigo tem como objetivo instruir os pesquisadores da área para o uso desse software na obtenção de estatísticas básicas, conferindo-lhes maior independência e flexibilidade científica.

O banco de dados utilizado nas demonstrações aqui apresentadas é oriundo de uma pesquisa em terapia ocupacional intitulada “*O Ages and Stages Questionnaires Brasil (ASQ-BR) como instrumento de triagem do desenvolvimento no contexto da educação infantil*” de Della Barba (2014), cujo principal objetivo foi analisar o desempenho de crianças que frequentam a educação infantil de um município do interior do estado de São Paulo em um instrumento americano de triagem do desenvolvimento. Já as análises aqui realizadas (leitura e sumarização de dados, cálculo do tamanho amostral, testes de hipóteses e teste de correlação linear) resultam de uma demanda compreendida no decorrer de uma

disciplina do Programa de Pós-Graduação em Terapia Ocupacional da Universidade Federal de São Carlos (UFSCar).

Para tanto, o artigo encontra-se organizado da seguinte forma:

No Capítulo 1 serão apresentados o Software *R*, os passos para a sua instalação e o seu layout. No Capítulo 2 será demonstrado como realizar a leitura de dados no *R* e adicionar demais informações. No Capítulo 3 será tratada a sumarização dos dados, desde o cálculo de estatísticas descritivas (variância e desvio padrão) até o desenvolvimento de representações gráficas tanto de dados quantitativos quanto qualitativos. No Capítulo 4 será explicado como fazer o cálculo do tamanho amostral de amostras aleatórias simples. No Capítulo 5 serão apresentados os comandos para realizar um teste de hipótese em duas populações. E, por fim, no Capítulo 6 será demonstrado como executar cálculos para verificar correlações entre variáveis.

1.1 Software R

Desenvolvido a priori por Ross Ihaka e Robert Gentleman e, posteriormente, aditado por colaboradores de outras partes do mundo, o *R* é um programa computacional direcionado a operações estatísticas e gráficas amplamente demandadas para o tratamento, a sistematização e a divulgação de dados informativos (R CORE TEAM, 2018).

Posto que existam outros programas com a mesma finalidade, faz-se necessário elencar as vantagens atribuídas ao uso do *R* que o tornam uma opção diferenciada dos demais concorrentes.

Primeiro, trata-se de um software livre, o que permite ao pesquisador propor novas sub-rotinas e implementar novos métodos de análise conforme a sua necessidade. Segundo, ele é gratuito e, portanto, não tem prazo de expiração, podendo ser utilizado com mais flexibilidade. Terceiro, por ser multiplataforma, pode ser executado pelo Windows, Macintosh e Unix/Linux. E quarto, é expansível, visto que oferece inúmeros serviços desde os mais básicos até os mais complexos, por exemplo, novas técnicas estatísticas que são publicadas em periódicos vêm, em geral, acompanhadas de pacotes com funções implementadas em *R*, possibilitando dessa forma que o pesquisador tenha acesso a tais metodologias e as aplique facilmente.

Diante disto, justifica-se a crescente popularidade do *R* em relação aos demais programas. Percebe-se, na figura 1, que concomitantemente ao declínio do SPSS está a ascensão do *R* no âmbito acadêmico.

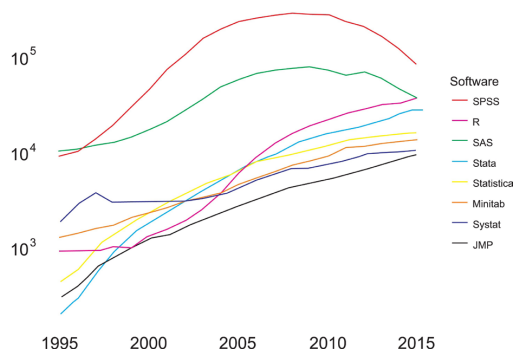


Figura 1. Número de acessos a diferentes softwares estatísticos no google scholar.

1.1.1 Instalação

Passo 1 - Acesse o link disponível em FIOCRUZ (2019).

Passo 2 - Escolha a plataforma na qual será executado o *R*.

Passo 3 - Clique em “*install R for the first time*”.

Passo 4 - Clique em “*Download R 3.3.3 for*”.

Executado os procedimentos para a instalação o software estará pronto para uso (Figura 2). É importante ressaltar que ao realizar o passo 4 uma versão atualizada poderá estar disponível. O software é atualizado constantemente para acomodar novas tecnologias, no entanto os procedimentos discutidos aqui não se alteram para qualquer versão.

1.1.2 Layout

O layout do *R* compreende uma janela intitulada “*console*”, espaço no qual o usuário irá inserir, alterar ou salvar os dados e os códigos das análises a serem efetuadas (Figura 3).

No intuito de facilitar o processo operacional, sugere-se a abertura de uma janela complementar intitulada “*new script*” para que os dados, bem como os comandos, possam ser organizados e transportados ao *console* imediatamente, sem erros de digitação, por meio da combinação *Ctrl R*. Para abrir uma janela complementar basta ir em “*file*” e clicar na opção “*new script*”.

2 Leitura de Dados

Aqui será apresentado como realizar leitura de dados no *R* (Tabela 1).



CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

Thank R

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages. **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.



CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

Subdirectories:

[base](#)

[contrib](#)

[old contrib](#)

[Rtools](#)

Binaries for base distribution (managed by Duncan Murdoch). This is what you want to [install R for the first time](#).

Binaries of contributed CRAN packages (for R \geq 2.11.x; managed by Uwe Ligges). There is also information on [third party software](#) available for CRAN Windows services and corresponding environment and make variables.

Binaries of contributed CRAN packages for outdated versions of R (for R $<$ 2.11.x; managed by Uwe Ligges). Tools to build R and R packages (managed by Duncan Murdoch). This is what you want to build your own packages on Windows, or to build R itself.

R for Windows



CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)

R-3.3.3 for Windows (32/64 bit)

[Download R 3.3.3 for Windows](#) (71 megabytes, 32/64 bit)

[Installation and other instructions](#)
[New features in this version](#)

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the [md5sum](#) of the .exe with the [fingerprint](#) on the master server. You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

Figura 2. Instalação do R.

```
R R Console

R version 3.3.3 (2017-03-06) -- "Another Canoe"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R é um software livre e vem sem GARANTIA ALGUMA.
Você pode redistribuí-lo sob certas circunstâncias.
Digite 'license()' ou 'licence()' para detalhes de distribuição.

R é um projeto colaborativo com muitos contribuidores.
Digite 'contributors()' para obter mais informações e
'citation()' para saber como citar o R ou pacotes do R em publicações.

Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda,
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digite 'q()' para sair do R.

[Área de trabalho anterior carregada]

> |
```

Figura 3. Console do R.

Onde o X representa os grupos do sexo masculino, o Y os do sexo feminino e os adendos 1, 2, 3, 4 e 5 as categorias: Comunicação, Coordenação motora ampla, Coordenação motora fina, Resolução de problemas e Pessoal/Social, respectivamente.

No Brasil é padrão utilizar a vírgula como separação de casas decimais, enquanto o padrão internacional é dado por ponto. Nesse caso, o R utiliza o ponto para definir casas decimais e a vírgula para distinguir os elementos.

Para realizar a leitura dos dados é preciso, inicialmente, inseri-los, “manualmente”, na página complementar:

```
x1=c(25,60,50,50,40,45,55,60,50,40,60)
x2=c(50,55,50,60,50,35,60,60,60,50,55)
x3=c(20,30,50,40,30,40,55,50,55,45,50)
x4=c(35,40,45,60,50,45,60,50,60,30,60)
x5=c(30,60,40,45,50,25,40,40,45,60,60)
y1=c(25,60,55,55,60,60,55,60,55,60)
y2=c(60,60,55,55,60,60,60,15,60,55)
y3=c(55,40,30,30,55,60,45,60,60,50)
y4=c(50,35,40,55,55,50,50,50,60,60)
y5=c(50,50,50,60,60,55,50,60,60,45)
```

Para inserir informações adicionais no comando basta digitar # e a informação. As informações adicionais têm como objetivo sinalizar, complementar ou diferenciar os dados apresentados, devendo ser utilizadas sempre que houver a necessidade de especificar algo, seja, por exemplo, por meio de um título ou um subtítulo.

```
#Grupo Meninos Comunicação 16-27 meses
#Grupo Meninas Comunicação 16-27 meses
```

O R também possibilita a leitura direta de softwares como, por exemplo, Excel, Minitab, SPSS, entre outros. Por exemplo, suponha que o interesse seja fazer a leitura dos dados que estão inseridos no Excel (veja Figura 4 painel esquerdo). Uma forma simples de inserir os dados sem a necessidade de instalar novos pacotes é salvar os dados em formato .csv.

Tabela 1. Escores em diferentes áreas de desempenho obtidos por crianças de 16 a 27 meses em um município do interior de São Paulo.

Sexo	Categoria	Escore										
M	X1	25	60	55	55	60	60	55	60	55	60	
	X2	60	60	55	55	60	60	60	15	60	55	
	X3	55	40	30	30	55	60	45	60	60	50	
	X4	50	35	40	55	55	50	50	50	60	60	
	X5	50	50	50	60	60	55	50	60	60	45	
F	Y1	25	60	50	50	40	45	55	60	50	40	60
	Y2	50	55	50	60	50	35	60	60	60	50	55
	Y3	20	30	50	40	30	40	55	50	55	45	50
	Y4	35	40	45	60	50	45	60	50	60	30	60
	Y5	30	60	40	45	50	25	40	40	45	60	60

	A	B	C	D
1	Comunicacao	CM Ampla	CM Fina	Sexo
2	25	50	20	1
3	60	55	30	1
4	50	50	50	1
5	50	60	40	1
6	40	50	30	1
7	45	35	40	1
8	55	60	55	1
9	60	60	50	1
10	50	60	55	1
11	40	50	45	1
12	60	55	50	1
13	25	60	55	0
14	60	60	40	0
15	55	55	30	0
16	55	55	30	0
17	60	60	55	0
18	60	60	60	0
19	55	60	45	0
20	60	15	60	0
21	55	60	60	0
22	60	55	50	0

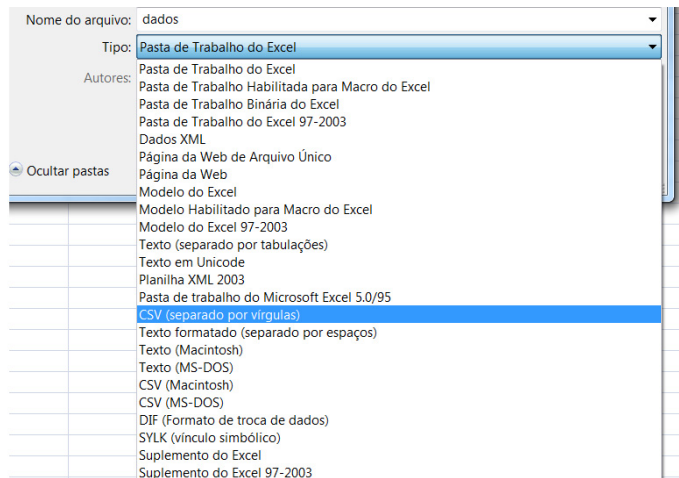


Figura 4. Inserção de dados no Excel.

O Excel possibilita salvar os dados nesse formato (veja Figura 4 painel direito).

Por fim, ao salvar o arquivo dentro da uma pasta destino, por exemplo, D:/, a leitura é feita da seguinte forma:

```
dados=read.csv("D:/dados.csv",header = TRUE,
sep=";",")
> dados
Comunicacao CM.Ampla CM.Fina Sexo
1 25 50 20 1
2 60 55 30 1
....
21 60 55 50 0
t1=dados$Comunicacao
> t1
[1] 25 60 50 50 40 45 55 60 50 40 60 25 60 55 55
60 60 55 60 55 60
x1=dados$Comunicacao[dados$Sexo==1]
y1=dados$Comunicacao[dados$Sexo==0]
> x1
[1] 25 60 50 50 40 45 55 60 50 40 60
> y1
[1] 25 60 55 55 60 60 55 60 55 60
```

O comando *read.csv* é utilizado para realizar a leitura do arquivo, “D:/dados.csv” é a pasta e o nome do arquivo salvo, *header = TRUE* refere-se ao caso em que a tabela possui cabeçalho, caso não haja troque para *header = FALSE, sep=";* apresenta a forma que separa as informações do arquivo .csv (o Excel salva dessa forma). Caso o interesse seja declarar as informações separadamente como descrito no exemplo anterior, o comando é *x1=dados\$NOME*, onde *NOME* é o nome da variável definida no cabeçalho.

É importante ressaltar que aqui as variáveis não foram inicialmente separadas entre masculino e feminino e que a coluna *Sexo* é utilizada para discriminar o gênero. Caso o interesse seja trabalhar com *x1* e *y1* separadamente, podemos utilizar o termo *dados\$NOME[dados\$Sexo==1]* para selecionar apenas o sexo masculino e trocar *1* por *0* para selecionar apenas o feminino.

3. Sumarização de Dados

O processo de sumarização e descrição dos resultados é importante, pois é a partir dele que os dados serão organizados e apresentados ao leitor.

Pensar sobre essa etapa implica lançar mão de recursos cada vez mais compreensíveis, visto que a comunicação no âmbito científico é fundamental para a divulgação de informações e para a construção de novos conhecimentos (KIELHOFNER, 2006).

3.1 Medidas de tendência central e de variabilidade

Estatísticas básicas como *média*, *variância* e *desvio padrão* são as mais utilizadas, pois permitem que tenhamos uma *medida de tendência central* e uma *medida de dispersão*. Ressalta-se que os comandos estão ligados ao termo em inglês como, por exemplo, “*var*” que está ligado ao termo em inglês “*variance*”, e “*sd*” que está ligado ao termo “*standard deviation*”.

Comandos para calcular tais estatísticas:

```
mean(x1)
[1] 48.63636
var(x1)
[1] 115.4545
sd(x1)
[1] 10.74498
```

Além destas, pode-se obter também o *mínimo*, *1º quartil*, *mediana*, *3º quartil* e o *máximo*:

```
summary(x1)
Min. 1st Qu. Median Mean 3rd Qu. Max.
25.00 42.50 50.00 48.64 57.50 60.00
```

3.2 Representação gráfica

Obter gráficos no R é muito simples, sejam eles gerados a partir de dados quantitativos ou qualitativos. Entretanto, vale ressaltar que existem gráficos apropriados para cada caso, porém serão apresentados aqui apenas aqueles mais comuns.

3.2.1 Dados quantitativos

Tratando-se de dados quantitativos, considere o *histograma* e o *boxplot*.

Comando para gerar o *histograma*:

```
#Figura Esquerda
hist(x1)
#Figura Direita
hist(x1,main="Grupo Comunicação 16-27 meses",
xlab="Score", ylab="Frequencia",
col="deepskyblue4")
```

No primeiro caso, utilizando apenas o comando *hist(x1)*, obtemos o gráfico de interesse. No entanto, personalizações são possíveis adicionando informações como título (*main*), informação no eixo x (*xlab*) ou

a frequência no eixo y (*ylab*) e diferentes cores (*col*). Para modificar a cor dos gráficos basta trocar o nome no comando. O nome das cores disponíveis pode ser acessado em Columbia University (2019).

Para o *boxplot*, no eixo horizontal temos o(s) *fator(es) de interesse* e no eixo vertical a *variável* a ser analisada. O *boxplot* é um gráfico muito informativo, pois possibilita localizar a distribuição dos dados, a variabilidade, a simetria ou assimetria, fornece também um critério para identificação de valores extremos e, por fim, possibilita a comparação desses resultados para diferentes grupos. A primeira linha da caixa representa o *primeiro quartil*, a segunda a *mediana* e a terceira o *terceiro quartil*. A reta vertical conectada ao topo representa o *máximo* e a reta inferior o *mínimo*. Assim, o *boxplot* pode ser utilizado para visualizar se determinados conjuntos de dados possuem ou não equivalência.

Comando para gerar o *boxplot*:

```
#Figura Esquerda
boxplot(x3,y3)
# Figura Direita
boxplot(x3,y3,col=c("deepskyblue4","gray70"),
main="Boxplot para
Coordenação Motora Fina", names=c("Masculino",
"Feminino"), ylab="Score",
horizontal=FALSE)
```

Para gerar a figura 5 do *boxplot* é necessário utilizar o comando *boxplot()*. Caso o interesse seja visualizar apenas a variável *x3*, pode-se utilizar o *boxplot(x3)*, já para duas variáveis utiliza-se *boxplot(x3,y3)*. Cores podem ser personalizadas por meio do *col*, enquanto o *names* discrimina quais são as variáveis em estudo. Por fim, o argumento *horizontal* é utilizado para informar se desejamos que o *boxplot* seja apresentado verticalmente (*horizontal=FALSE*) ou de forma horizontal (*horizontal=TRUE*).

3.2.2 Dados qualitativos

Tratando-se de variáveis qualitativas, considere o gráfico de *pizza* e o de *barras*.

Comando para gerar o gráfico de *pizza*:

```
pie(c(length(x1),length(y1)))
pie(c(length(x1),length(y1)),
col=c("deepskyblue4","gray70"),
labels=c("Masculino", "Feminino"), main="Sexo")
```

Neste caso, têm-se os seguintes comandos: gerar gráfico de *pizza* (*pie*), selecionar as cores (*col*) e definir as legendas (*labels*) e o título da figura (*main*). Utiliza-se o comando *length* para calcular a frequência de ocorrência da variável, no entanto podemos colocar os valores manualmente

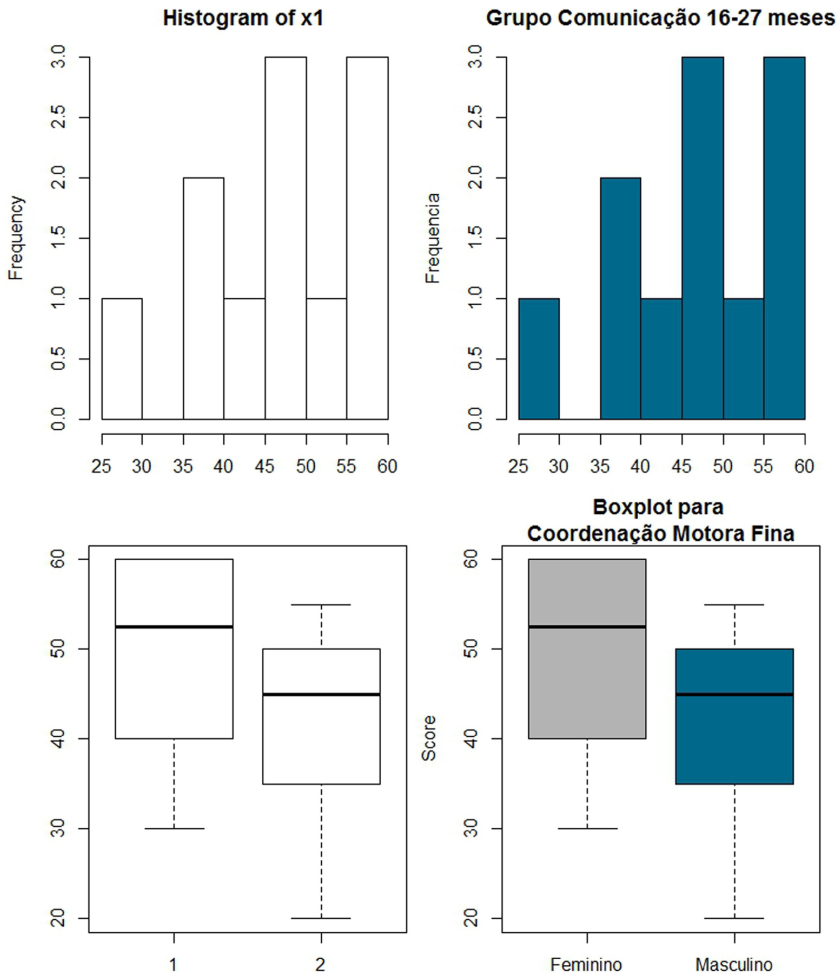


Figura 5. Histograma e Boxplot do escore Coordenação motora fina para os grupos de meninos e meninas com 16-17 meses.

caso haja o interesse. Por fim, aqui utilizou-se apenas duas categorias, mas pode-se estender para quantas forem necessárias, incluindo uma vírgula depois do último item de cada argumento seguido de suas informações.

Caso os dados fossem inseridos diretamente pelo Excel e o interesse fosse trabalhar diretamente com *dados\$Sexo*, o gráfico poderia ser obtido da seguinte forma:

```
pie(c(sum(dados$Sexo),length(dados$Sexo)-
sum(dados$Sexo)))
pie (c(sum(dados$Sexo),length(dados$Sexo)-
sum(dados$Sexo)), col=c("deepskyblue4",
"gray70"),labels=c("Masculino", "Feminino"),
main="Sexo")
```

Em que *sum()* calcula a soma de todos elementos. Como definiu-se o sexo masculino como 1 e o feminino como 0, a soma de todos os valores

retornará o número de meninos na amostra enquanto o complementar trará o total de meninas.

Comando para gerar o gráfico de *barras*:

```
barplot(c(length(x1),length(y1)),
main="Número de alunos dividido por sexo",
col=c("deepskyblue4","gray70"),
names.arg=c("Masculino", "Feminino"))
```

Assim como no exemplo anterior, o gráfico de *barras* é gerado através do *barplot*, enquanto as categorias são nomeadas através do *names.arg* (Figura 6).

4 Cálculo do Tamanho Amostral

Em qualquer estudo a margem de erro e o nível de confiança estão intimamente ligados ao tamanho da amostra (BUSSAB; MORETTIN, 2010). Assim, quando não há a possibilidade de se trabalhar com

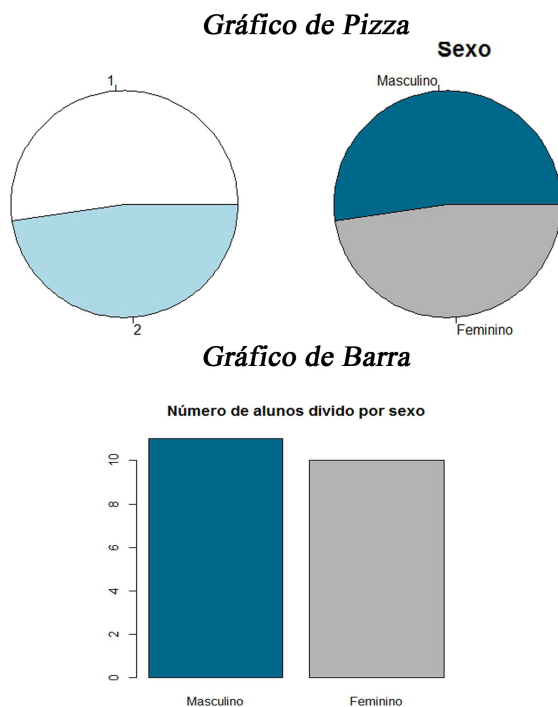


Figura 6. Distribuição de crianças por sexo.

toda a população de interesse, seja pelo tempo limitado, custos elevados, questões éticas, entre outras limitações, faz-se necessário lançar mão de técnicas de amostragem.

O cálculo do tamanho amostral trata-se de uma etapa importante da pesquisa. Problemas no decorrer dessa etapa podem comprometer a análise e a interpretação dos resultados (MIOT, 2011).

Será apresentado aqui apenas o cálculo do tamanho amostral de amostras aleatórias simples. Primeiro deve-se definir o *nível de significância* (α) e a *margem de erro* (ϵ). O *nível de significância* é utilizado para a construção dos intervalos de confiança, que nos dizem qual percentual de todas as amostras possíveis satisfazem a *margem de erro*, enquanto a *margem de erro* revela o quão perto da amostra está do parâmetro da população.

Della Barba (2014), ao tomar como verdadeiro parâmetro a proporção de crianças de 4 a 60 meses que apresentam desenvolvimento típico e que frequentam creches e pré-escolas de um município do interior de São Paulo, definiu que a *margem de erro* seria de dez pontos percentuais e que o intervalo de confiança seria de 95%. Isso significa que se na amostra 80% das crianças apresentarem um desenvolvimento típico, na população esse grupo deveria ter entre 70% e 90% de crianças com desenvolvimento típico. Além disso, se o intervalo de confiança é de 95%, é possível que cinco em cada 100 pesquisas

realizadas, respeitando os mesmos procedimentos metodológicos, apresentem um resultado fora do intervalo.

4.1 Tamanho amostral para uma população

O cálculo do tamanho amostral pode ser realizado em quatro situações. Considerando variáveis qualitativas:

- 1) Variável qualitativa com população menor de 10000.
- 2) Variável qualitativa com população maior ou igual a 10000.

A seguir propõe-se uma rotina (*tal*) a partir da qual será possível calcular o tamanho amostral para dados qualitativos, ou seja, quando o interesse é estimar a proporção de alguma característica na população. Os argumentos da função são: *alpha*, que é o nível de significância, *E*, que é o erro assumido (valores entre 0 e 1), *N*, que é o tamanho da população (se *N* for maior que 10000 não precisa inseri-lo), e *p*, que é a proporção obtida anteriormente em um pré-teste ou trabalhos anteriores. Os últimos dois elementos só deverão ser inseridos quando houver

esse tipo de informação (BUSSAB; BOLFARINE, 2005).

```
ta1<-function(alpha,E,N=NA,p=0.5){
z<-qnorm(1-alpha/2,0,1)
if(is.na(N)) { n<-(z*sqrt(p*(1-p))/E)^2 } else {
n<-((N*p*(1-p)*(z^2))/(((N-1)*(E^2))+(p*(1-
p)*(z^2)))) }
return(round(n))}
```

Ao inserir o comando acima, gera-se o comando para o cálculo do tamanho amostral que dependerá da inclusão das informações sobre o *alpha*, *E*, *N* e *p*:

```
ta1(alpha,E,N,p)
```

De acordo com Miot (2011), é de suma importância realizar um pré-teste com 30 a 40 indivíduos e considerar os resultados como estimativa populacional para o cálculo do tamanho amostral. Para variáveis quantitativas o pré-teste é necessário, já para as variáveis qualitativas ele é opcional. Neste caso, pode-se utilizar o fato de que a proporção que levará ao maior tamanho da amostra dentre todas as estimativas de proporção é de $p = 0.5$. Este resultado é o mais conservador e é comumente aplicado na ausência de um pré-teste. Embora um pré-teste envolva uma etapa adicional no trabalho, ele pode diminuir em mais de quatro vezes o tamanho amostral (em relação ao intervalo conservativo), possibilitando corte de custos e maior agilidade na tabulação e na obtenção dos resultados.

Exemplo 1: Supondo que o interesse seja encontrar a proporção de crianças que se encontram na zona de risco de desenvolvimento.

- a) Calcule o tamanho amostral necessário para estimar tal proporção com nível de significância de 0.05 (95% de confiança) e uma margem de erro de 0.03.

```
ta1(alpha=0.05,E=0.03)
[1] 1067
```

Neste caso, seria necessária uma amostra de 1067 pessoas para estimar tal proporção, considerando 95% de confiança e uma margem de erro de 3 pontos percentuais.

- b) Supondo que um pré-teste foi realizado e, por meio dele, constatou-se que 6% das crianças se encontravam na zona de risco. Qual seria o tamanho amostral necessário sob as mesmas suposições do exemplo anterior?

```
ta1(0.05,0.03,0.06)
[1] 241
```

Percebe-se que a realização do pré-teste possibilitou uma redução significativa do tamanho amostral.

c) Supondo que o interesse seja estimar a proporção de crianças que se encontram na zona de risco em uma cidade pequena. Supondo que o número da população é de 1000 crianças, calcule o tamanho amostral sob as mesmas suposições.

Sem dados de pré-teste:

```
ta1(0.05,0.03,1000)
[1] 516
```

Com dados de pré-teste:

```
ta1(0.05,0.03,1000,0.06)
[1] 194
```

Sem um pré-teste seriam necessárias 516 crianças para realização do experimento. Já com um pré-teste indicando 6% da população na zona de risco, o tamanho amostral seria de 194 crianças.

Em casos relacionados a variáveis quantitativas, ou seja, quando o interesse é estimar a média de alguma característica na população, temos as seguintes possibilidades:

- 3) Variável quantitativa com população menor de 10000.
- 4) Variável quantitativa com população maior ou igual a 10000.

A rotina a seguir deve ser inserida no *R* antes da realização dos cálculos. Ela utiliza os valores de *alpha*, o nível de significância considerado, e *E* que é a margem de erro da estimativa, o *sigma*, o desvio-padrão das variáveis de interesse e o *N* que é o tamanho da população (BUSSAB; BOLFARINE, 2005).

```
#Função para dados quantitativos
ta2<-function(alpha,E,sigma,N=NA){
z<-qnorm(1-alpha/2,0,1)
if(is.na(N)) { n<-(z*sigma/(100*E))^2 } else {
n<-((N*(sigma^2)*(z^2))/
(((N-1)*((100*E)^2)+(sigma^2*(z^2)))) }
return(round(n))}
```

A função a ser chamada no *R* será então:

```
ta2(alpha,E,sigma,N=NA)
```

Diferentemente do caso anterior, não há escolha conservadora para *sigma* e tal resultado precisa ser obtido por meio de um pré-teste ou pesquisas relacionadas. O valor de *N* deve ser inserido para pequenas populações ($N < 10000$), caso não seja inserido o programa irá considerá-lo seu complementar ($N \geq 10000$).

Exemplo 2: Caso o objetivo seja descrever o escore da coordenação motora fina de crianças de 16-27 meses de uma cidade, pode-se utilizar como pré-teste os dados apresentados na Tabela 1. Neste caso, tem-se que σ é de 11.77. Assumindo um nível de significância de 0.05 e um erro de 0.02:

- a) Calcule o tamanho amostral necessário, assumindo que a população de crianças nessa faixa etária é de 4000.

```
ta2(0.05,0.02,11.77,4000)
[1] 129
```

Constata-se que seriam necessárias 129 crianças para encontrar uma estimativa do escore populacional com o nível de significância, o desvio-padrão e a margem de erro assumidos. É importante ressaltar que conforme aumenta a variabilidade do conjunto de dados (σ), maior será o tamanho amostral necessário para encontrar a estimativa populacional.

- b) Calcule o tamanho amostral necessário assumindo que a população de crianças é de 30000.

```
ta2(0.05,0.02, 11.77,30000)
[1] 132
```

Neste caso, como $N > 10000$, não devemos preencher o último item.

```
ta2(0.05,0.02, 11.77)
[1] 133
```

Percebe-se que, tanto inserindo a informação quanto não inserindo, obtêm-se resultados similares. No entanto, o segundo caso é o correto. Os métodos aqui apresentados para cálculo de tamanho de amostra são usados para objetivos básicos de estimação de uma proporção ou de uma média. Para outros objetivos como, por exemplo, comparação de proporções ou médias de vários grupos, delineamentos com repetição ao longo do tempo, entre outros, é importante consultar um profissional da estatística, pois cada método de análise possui fórmulas diferentes para calcular tamanhos amostrais.

Também é importante ressaltar que cálculos de tamanho amostral assumem que o processo de seleção dos indivíduos é aleatório, o que na prática nem sempre é o caso. Sendo assim, sempre é importante discutir e ponderar cada caso.

5. Teste de Hipóteses em Duas Populações

Teste de hipóteses trata-se de um método de inferência estatística utilizado para avaliar parâmetros populacionais desconhecidos, por meio de evidências

que uma amostra fornece (MIGON; GAMERMAN; LOUZADA, 2014).

Este método é amplamente empregado para validar pesquisas clínicas quando o principal interesse é, por exemplo, verificar se houve diferença entre um tratamento padrão e um tratamento alternativo. Tal método permite também responder diversas perguntas sobre os parâmetros de interesse, como a média, a proporção, a variância e etc. Desta forma, assumimos duas hipóteses, a *nula* (H_0) e a *alternativa* (H_A), e tomamos uma decisão de aceitar ou rejeitar a *hipótese nula*. Embora seja mais comum definir a *hipótese alternativa* como aquela proposta pelo pesquisador e a *hipótese nula* como o complemento da *hipótese alternativa*, no R não é necessário definir as *hipóteses nula* e *alternativa*, o programa em si já faz isso, é necessário apenas verificar quais foram as hipóteses assumidas pelo R.

5.1 Hipótese de normalidade

Diversos testes estatísticos utilizam como base a suposição de que os dados provêm de uma distribuição normal, o que nos proporciona ótimas ferramentas para a realização de testes de hipótese.

Há na literatura um grande número de testes para verificar a suposição de normalidade. Yap e Sim (2011), ao compararem vários testes, concluíram que, em geral, o teste de *Shapiro-Wilk* é mais sensível para detecção ou não de normalidade. Ele pode ser realizado através do comando:

```
shapiro.test()
```

As hipóteses a serem testadas serão:

$$\begin{cases} H_0 & \text{A amostra provém de uma distribuição normal} \\ H_a & \text{A amostra não provém de uma distribuição normal} \end{cases}$$

Ao tomar uma decisão em relação a uma hipótese, pode-se cometer os seguintes erros: o erro do tipo I refere-se a rejeitar a H_0 quando na verdade a H_0 é verdadeira. Por outro lado, pode-se cometer o erro do tipo II, que é não rejeitar a H_0 quando na verdade a H_0 é falsa. A probabilidade de ocorrer o erro do tipo I é denotada por α , conhecida como nível de significância. Tal valor é determinado, em geral, pelo pesquisador antes da coleta de dados. Em diversas aplicações o nível de significância assumido é de 0.05. Uma regra simples a ser seguida para a tomada de decisão é considerar o valor-p. O valor-p pode ser visto como a probabilidade de se obter uma estatística de teste igual ou mais extrema que a observada, por meio de uma amostra da população considerando como verdadeira a H_0 . Desta forma, após definir o nível de significância, rejeita-se a H_0 se o valor-p

for menor que α ou não se rejeita a H_0 se o valor-p for maior que α .

Assim, para o conjunto de dados x_3 , aplicando a função *shapiro.test(x3)*, têm-se os seguintes resultados:

```
>shapiro.test(x3)
Shapiro-Wilk normality test
data: x3
W = 0.905, p-value = 0.2125
```

Tal comando tem como saída a estatística do teste (W) e o seu valor-p (p-value) associado. Como definiu-se o $\alpha = 0.05$, a regra de decisão será se *valor-p* < α , rejeita-se a H_0 , ou seja, os dados não provêm de uma distribuição normal. Se *valor-p* > α , não se rejeita a H_0 , isto é, os dados provêm de uma distribuição normal. Como $0.2125 > 0.05$, então, com um nível de significância de 0.05, não há evidências de que os dados não sigam essa distribuição. É preciso também verificar a suposição de normalidade para a outra variável y_3 a ser comparada.

```
>shapiro.test(y3)
Shapiro-Wilk normality test
data: y3
W = 0.85874, p-value = 0.07374
```

Análoga a x_3 , como $0.07374 > 0.05$, então, com um nível de significância de 0.05, os dados provêm de uma distribuição normal.

5.2 Hipótese de normalidade aceita

Verificada as suposições de normalidade para ambas variáveis, realiza-se o teste *t* para comparar as médias entre os dois grupos (x_3 e y_3), possibilitando afirmar (a um nível de significância pré-definido) se as médias diferem ou não estatisticamente. No entanto, os testes serão diferentes caso as variâncias dos grupos sejam iguais ou diferentes. Portanto, antes de realizar tal procedimento, deve-se verificar a igualdade ou diferença das variâncias. Para tanto, utiliza-se outro teste de hipótese conhecido como teste *F*. Nesse caso tem-se as seguintes suposições:

- { H_0 As amostras possuem variâncias iguais
- { H_a As amostras possuem variâncias diferentes

A partir daí, utiliza-se a mesma regra de decisão adotada na seção anterior. Assumindo um nível de significância de $\alpha = 0.05$, a regra de decisão será se *valor-p* < α rejeitamos H_0 , ou seja, as amostras possuem variâncias diferentes. Se *valor-p* > α não se rejeita a H_0 , isto é, as amostras possuem variâncias iguais. Comando para o teste *F*:

```
>var.test(x3,y3,conf.level = 0.95)
F test to compare two variances
data: x3 and y3
F = 0.9472, numdf = 10, denomdf = 9, p-value =
0.926
alternative hypothesis: true ratio of variances is not
equal to 1
```

Como *p-value* > 0.05, não se rejeita a H_0 , isto é, o teste de hipóteses apresenta evidências de que não há diferença significativa entre as variâncias.

Verificada a suposição de igualdade ou diferença entre as variâncias de x_3 e y_3 , pode-se considerar o teste *t* para a comparação das médias:

```
t.test(x, y, paired = FALSE, var.equal=FALSE,
conf.level = 0.95)
```

Sendo x o primeiro grupo, y o segundo grupo, *var.equal* se as variâncias são iguais ou não (*TRUE* ou *FALSE*) e *conf.level* intervalo de confiança ($1-\alpha$). Se o nível de significância for 0.05, então o *conf.level* = $1-0.05=0.95$. Caso não seja inserida nenhuma informação, o *R* irá considerar como padrão o valor 0.95. O argumento *paired* refere-se ao caso em os dados são pareados. Uma amostra pareada significa que cada observação da primeira amostra está relacionada com a respectiva observação da segunda amostra, um exemplo seria considerar uma medida que é observada pré e pós teste em um mesmo indivíduo. Caso os dados sejam pareados deve-se definir que *paired* = *TRUE*.

Exemplo1: Faça um teste de hipótese para verificar se há diferença entre a média dos meninos e das meninas de 16-27 meses para a variável coordenação motora fina a um nível de significância de 0.05.

```
t.test(x3,y3, paired = FALSE, var.equal = TRUE,
conf.level = 0.95)
Two Sample t-test
data: x3 and y3
t = -1.2236, df = 18.696, p-value = 0.2363
alternative hypothesis: true difference in means is
not equal to 0
95 percent confidence interval: -16.890851 4.436305
```

A hipótese a ser testada aqui é *igualdade versus diferença* entre os grupos, onde a *hipótese nula* é de *igualdade*. Portanto, utilizando a regra de decisão já definida $0.236 > 0.050$, com um nível de significância de 0.05, não se rejeita a H_0 , isto é, não há diferença entre as médias. Note que, embora a média dos meninos seja 42.27 e a das meninas de 48.5, ao realizar o teste de hipótese não se pode afirmar que ambas diferem a um nível de significância de 0.05.

Exemplo 2: Supondo que o interesse seja realizar um teste de hipótese para verificar se há diferença entre a média dos escores dos meninos e das meninas em relação à variável pessoal/social a um nível de significância de 0.05.

```
t.test(x5,y5, paired = FALSE, var.equal = TRUE,
conf.level = 0.95)
Two Sample t-test
data: x5 and y5
t = -2.2536, df = 14.66, p-value = 0.03999
alternative hypothesis: true difference in means is
not equal to 0
```

Neste caso, utilizando a mesma regra de decisão, tem-se que $0.03999 < 0.05$, com um nível de significância de 0.05, logo rejeita-se a H_0 , isto é, há diferença entre as médias. No entanto, tais resultados não são válidos, veja que aqui não se testou a normalidade de y_5 .

```
shapiro.test(y5)
Shapiro-Wilk normality test
data: y5
W = 0.82495, p-value = 0.02909
```

Como $0.02909 < 0.05$, então, com um nível de significância de 0.05, não se pode afirmar que os dados provêm de uma distribuição normal (lembre-se que se pode aceitar hipótese de normalidade se $p\text{-valor} > 0.05$). Neste caso, deve-se lançar mão de outro tipo de teste estatístico que será discutido a seguir.

5.3 Hipótese de normalidade rejeitada

Quando os dados não possuem uma distribuição normal, os testes anteriores não devem ser aplicados. Uma alternativa é considerar o teste de *Wilcoxon-Mann-Whitney* (*WMW*) para comparar as distribuições dos dois grupos. Nesse caso, a comparação não é feita diretamente por meio da média entre os grupos, mas por meio da distribuição dos dados.

Diferente do procedimento anterior o teste *WMW* não necessita da suposição de igualdade ou de diferença entre as variâncias.

Comando para o teste *WMW*:

```
wilcox.test(x, y, paired = FALSE, conf.level =
0.95)
```

Exemplo 3: Considere o exemplo anterior (onde y_5 não provêm de uma distribuição normal) e verifique se há diferença na média dos escores dos meninos e das meninas em

relação a variável pessoal/social a um nível de significância de 0.05.

```
wilcox.test(x5,y5,paired = FALSE, exact=FALSE,
conf.level = 0.95)
Wilcoxon rank sum test with continuity correction
data: x5 and y5
W = 28, p-value = 0.05497
alternative hypothesis: true location shift is not
equal to 0
```

A hipótese a ser testada aqui é *igualdade versus diferença* entre os grupos onde a *hipótese nula* é de *igualdade*. Portanto, utilizando-se a regra de decisão já definida $p\text{-value} > 0.05$, com um nível de significância de 0.05, não se rejeita a H_0 , isto é, não há diferença na média dos escores dos meninos e das meninas em relação à variável pessoal/social, a um nível de significância de 0.05.

É importante ressaltar que, quando se utiliza o teste *t* de forma errônea o mesmo leva a rejeitar a hipótese de igualdade, enquanto o teste *WMW* não levou à rejeição da hipótese de igualdade. Desta forma, as suposições de normalidade e igualdade de variância sempre devem ser checadas.

6. Correlação Linear

Em diversas pesquisas há o interesse em estabelecer relações entre variáveis, porém, mais que isso, é preciso que essas sejam, de fato, relevantes e não uma mera coincidência. O site disponível em Vigen (2019) apresenta algumas correlações estranhas (Figura 7).

Desta forma, conclui-se que uma solução para diminuir a taxa de divórcios em Maine seria eliminar o consumo de margarina per capita. Sabe-se que não há nenhuma relação entre essas variáveis e essa alta correlação é apenas uma mera coincidência. Logo, durante a construção das hipóteses a serem verificadas, o pesquisador deve se valer do bom senso.

6.1 Cálculo do coeficiente de correlação linear de Pearson

A saber, existe uma *variável explicativa* (x) e uma *variável resposta* (y) a ser encontrada. A *variável explicativa* é aquela que busca explicar a variável resposta, enquanto a *variável resposta* é a resposta propriamente dita. Para saber o quão elas estão relacionadas, utilizar-se-á, aqui, o *coeficiente de regressão linear de Pearson*.

O coeficiente de regressão linear (*cor*) sempre se encontrará entre -1 e 1. Assim, se $cor < 0$ significa correlação linear negativa, se $cor > 0$ significa correlação linear positiva e se $cor = 0$ significa ausência de

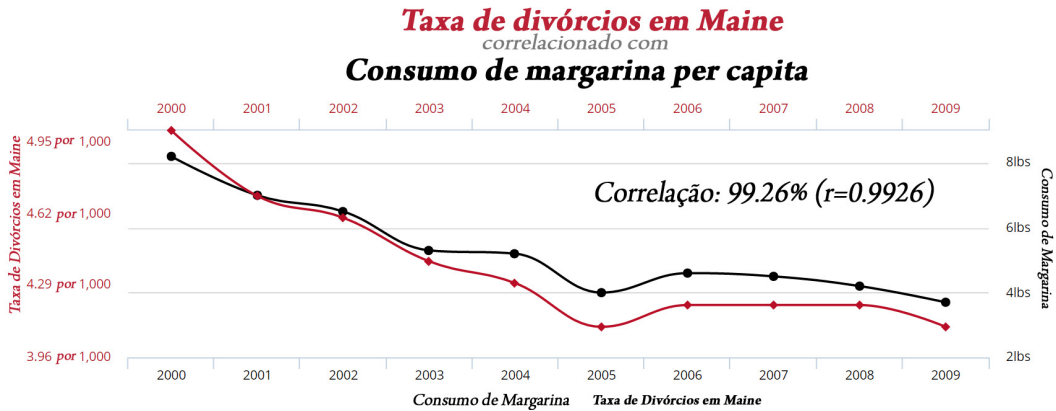


Figura 7. Correlação entre a taxa de divórcio sem Maine e o consumo de margarina per capita. Fonte: Vigen (2019).

correlação linear (BUSSAB; MORETTIN, 2010). No R, ele é dado por meio da expressão:

`cor(x,y)`

Exemplo 1: Verifique se há relação linear entre as variáveis comunicação e coordenação motora ampla, assim como entre comunicação e coordenação motora fina para meninos, assumindo um nível de confiança de 95%.

```
#Relação entre as variáveis comunicação e
coordenação motora ampla para meninos
>cor(x1,x2)
[1] 0.462047
#Relação em as variáveis comunicação e
coordenação motora fina para meninos
>cor(x1,x3)
[1] 0.6153194
```

No primeiro caso a correlação é de 0.46 , o que indica uma *correlação linear fraca*. Enquanto no segundo caso a correlação é de 0.61 , o que indica uma *correlação linear positiva*. No entanto, para saber se a correlação é, de fato, significativa, realiza-se um teste de hipótese.

6.2 Teste de hipótese

Comando para gerar o nível de significância da correlação:

```
cor.test(x,y,conf.level = 0.95)
```

Onde x é o primeiro grupo, y é o segundo grupo e *conf.level* é o intervalo de confiança. Caso nenhuma informação seja inserida, o programa considera como padrão o nível de confiança de 95% (relacionado a 5% de significância).

```
cor.test(x2,x3)
```

Pearson's product-moment correlation

data: x1 and x2

t = 1.563, df = 9, p-value = 0.1525

alternative hypothesis: true correlation is not equal to 0

sample estimates: cor 0.462047

A hipótese a ser testada aqui é $H_0 : r = 0$ versus $H_0 : r \neq 0$, assumindo um nível de significância de 0.05 . Não se rejeita a H_0 se o *p-valor* > 0.05 , o que indicará falta de uma relação linear entre as variáveis.

Considerando o exemplo anterior, percebe-se que, embora a correlação obtida seja de 0.46 , ao realizar um teste de hipótese obteve-se $0.1525 > 0.05$, o que significa que a mesma não é significativa. Desta forma, não se pode afirmar que há uma correlação linear entre as variáveis. Por outro lado para outras variáveis tem-se o seguinte resultado:

```
cor.test(x1,x3)
```

Pearson's product-moment correlation

data: x1 and x3

t = 2.3418, df = 9, p-value = 0.0439

alternative hypothesis: true correlation is not equal to 0

Sample estimates: cor 0.6153194

Tem-se que $0.0439 < 0.05$. Logo, com um nível de significância de 0.05 , rejeita-se a H_0 , isto é, a correlação linear entre as variáveis é diferente de zero. É importante salientar que, dependendo da área de pesquisa e do tamanho da amostra, uma correlação de, por exemplo, $0,5$ pode ser diferente de zero, mas não ser considerada uma alta correlação linear.

7 Considerações Finais

É evidente o atual movimento mundial pró *software livre* que, no Brasil, vem se estabelecendo desde a década de 1990. Esse movimento, composto

por representantes dos diferentes âmbitos sociais, surge como uma maneira de se alcançar a liberdade necessária para engrenar o desenvolvimento tecnológico nacional, descentralizando os conhecimentos acumulados e potencializando a construção de novas possibilidades.

Em consonância com este movimento, o presente artigo teve como objetivo fornecer subsídios para que pesquisadores em terapia ocupacional possam realizar procedimentos estatísticos básicos de maneira mais flexível e independente, a partir do R, um software livre e gratuito, cuja popularidade vem aumentando consideravelmente no âmbito acadêmico.

O atributo “livre”, associado a menores ou a nenhum custo, ao nosso ver, é essencial no âmbito das universidades brasileiras, visto que, muitas pesquisas são desenvolvidas com base em financiamentos extremamente modestos que necessitam ser rateados e, muitas vezes, compensados com recursos próprios. Pensando nisso, vimos no R a possibilidade de não apenas alimentar esse movimento, mas, também, de oportunizar o uso de uma ferramenta diferenciada e o remanejamento de verbas para outros fins.

Embora o R ofereça tantas outras possibilidades estatísticas, nós buscamos demonstrar, de uma maneira mais prática, apenas aquelas análises que nos foram percebidas no âmbito de um grupo de pesquisadores em terapia ocupacional que, muitas vezes, depara-se com a necessidade de analisar seus dados, com recursos e tempo limitados. Vale ressaltar que, em muitos casos, será necessária a aplicação de métodos mais complexos que exigirão o suporte de um profissional com expertise em estatística.

Referências

BUSSAB, W. O.; BOLFARINE, H. *Elementos de amostragem*. São Paulo: Ed. Edgard Blucher, 2005.

BUSSAB, W. O.; MORETTIN, P. A. *Estatística básica*. São Paulo: Saraiva, 2010.

COLUMBIA UNIVERSITY. DEPARTMENT OF STATISTICS. *Colors in R*. 2019. Disponível em: <www.stat.columbia.edu/~tzheng/files/Rcolor.pdf>. Acesso em: 23 maio 2018.

COSTA, G. Governo economiza R\$ 370 milhões com sistemas operacionais de computador. *Serpro Sede*, Brasília, 5 abr. 2009. Disponível em: <<http://www.serpro.gov.br/menu/noticias/noticias-antigas/governo-economiza-r-370-milhoes-com-sistemas-operacionais-de-computador>>. Acesso em: 23 maio 2018.

DELLA BARBA, P. C. S. *O empoderamento de pais para o conhecimento sobre o desenvolvimento de seus filhos: o Ages and Stages Questionnaire - ASQ-BR*. 2014. Relatório (Pós-Doutorado em Estudos da Criança) – Universidade do Minho, Braga, 2014.

FUNDAÇÃO OSWALDO CRUZ – FIOCRUZ. COMPREHENSIVE R ARCHIVE NETWORK – CRAN. *Apresenta o link para download do software R*. Vienna, 2019. Disponível em: <<https://cran.fiocruz.br/>>. Acesso em: 1 fev. 2019.

KIELHOFNER, G. *Research in occupational therapy: Methods of inquiry for enhancing practice*. Philadelphia: FA Davis, 2006.

MIGON, H. S.; GAMERMAN, D.; LOUZADA, F. *Statistical inference: na integrated approach*. Boca Raton: Taylor & Francis, 2014.

MIOT, H. A. Tamanho da amostra em estudos clínicos e experimentais. *Jornal Vascular Brasileiro*, Botucatu, v. 10, n. 4, p. 275-278, 2011.

OTTENBACHER, K.; PETERSEN, P. Quantitative trends in occupational therapy research: Implications for practice and education. *American Journal of Occupational Therapy*, Bethesda, v. 39, n. 4, p. 240-246, 1985.

R CORE TEAM. *R. A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, 2018.

SAMPAIO, R. F.; MANCINI, M. C.; FONSECA, S. T. Produção científica e atuação profissional: aspectos que limitam essa integração na fisioterapia e na terapia ocupacional. *Revista Brasileira de Fisioterapia*, São Carlos, v. 6, n. 3, p. 113-118, 2002.

VIGEN, T. *Spurious Correlations*. New York: Hachette Books, 2019. Disponível em: <<http://www.tylervigen.com/spurious-correlations>>. Acesso em: 01 fev. 2019.

YAP, B. W.; SIM, C. H. Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, Blacksburg, v. 81, n. 12, p. 2141-2155, 2011.

Contribuição dos Autores

Maysa Marinho Antunes Ramos foi responsável pela concepção do texto; organização de fontes e/ou análises, redação do texto. Pedro Luiz Ramos foi responsável pela concepção do texto; organização de fontes e/ou análises, redação do texto. Francisco Louzada Neto foi responsável pela concepção do texto; Revisão do texto. Patrícia Carla de Souza Della Barba foi responsável pela concepção do texto; revisão do texto. Todos os autores aprovaram a versão final do texto.