

CADERNO N.º 2

ELTON NUNES BRITTO
DAMIÃO PEDRO MEIRA FILHO
MARIA RAIMUNDA SILVA REGO
GEISIANE DA ROCHA SARMENTO

ESTATÍSTICA APLICADA À BIOLOGIA

USANDO O 

MANIPULAÇÃO DE TABELAS, TESTES ESTATÍSTICOS
E ABORDAGENS DE ANOVA PARA A ANÁLISE DE
DADOS BIOLÓGICOS

ESTATÍSTICA APLICADA À BIOLOGIA

USANDO O 

MANIPULAÇÃO DE TABELAS, TESTES ESTATÍSTICOS
E ABORDAGENS DE ANOVA PARA A ANÁLISE DE
DADOS BIOLÓGICOS

2025 by Atena Editora

Copyright © 2025 Atena Editora

Copyright do texto © 2025, o autor

Copyright da edição © 2025, Atena Editora

Os direitos desta edição foram cedidos à Atena Editora pelo autor.

Open access publication by Atena Editora

Editora chefe

Profª Drª Antonella Carvalho de Oliveira

Editora executiva

Natalia Oliveira Scheffer

Imagens da capa

iStock

Edição de arte

Yago Raphael Massuqueto Rocha



Todo o conteúdo deste livro está licenciado sob a Licença Creative Commons Atribuição 4.0 Internacional (CC BY 4.0).

A Atena Editora mantém um compromisso firme com a integridade editorial em todas as etapas do processo de publicação, assegurando que os padrões éticos e acadêmicos sejam rigorosamente cumpridos. Adota políticas para prevenir e combater práticas como plágio, manipulação ou falsificação de dados e resultados, bem como quaisquer interferências indevidas de interesses financeiros ou institucionais.

Qualquer suspeita de má conduta científica é tratada com máxima seriedade e será investigada de acordo com os mais elevados padrões de rigor acadêmico, transparência e ética.

O conteúdo da obra e seus dados, em sua forma, correção e confiabilidade, são de responsabilidade exclusiva do autor, não representando necessariamente a posição oficial da Atena Editora. O download, compartilhamento, adaptação e reutilização desta obra são permitidos para quaisquer fins, desde que seja atribuída a devida autoria e referência à editora, conforme os termos da Licença Creative Commons Atribuição 4.0 Internacional (CC BY 4.0).

Os trabalhos nacionais foram submetidos à avaliação cega por pares, realizada pelos membros do Conselho Editorial da editora, enquanto os internacionais passaram por avaliação de pareceristas externos. Todos foram aprovados para publicação com base em critérios de neutralidade e imparcialidade acadêmica.

Caderno no. 2 Estatística Aplicada à Biologia usando o R: Manipulação de Tabelas, Testes Estatísticos e Abordagens de ANOVA para a Análise de Dados Biológicos

| Autores:

Elton Nunes Britto
Damião Pedro Meira Filho
Maria Raimunda Silva Rego
Geisiane da Rocha Sarmento

| Revisão:

Os autores

| Diagramação:

Jeniffer Paula dos Santos

Dados Internacionais de Catalogação na Publicação (CIP)

C122 Caderno nº. 2 - Estatística aplicada à biologia usando o R: manipulação de tabelas, testes estatísticos e abordagens de ANOVA para análise de dados biológicos / Elton Nunes Britto, Damião Pedro Meira Filho, Maria Raimunda Silva Rego, et al. – Ponta Grossa - PR: Atena, 2025.

Outra autora
Geisiane da Rocha Sarmento

Formato: PDF
Requisitos de sistema: Adobe Acrobat Reader
Modo de acesso: World Wide Web
Inclui bibliografia
ISBN 978-65-258-3765-9
DOI: <https://doi.org/10.22533/at.ed.659250110>

1. Estatística biológica. I. Britto, Elton Nunes. II. Meira Filho, Damião Pedro. III. Rego, Maria Raimunda Silva. IV. Título.

CDD 570.1

Elaborado por Bibliotecária Janaina Ramos – CRB-8/9166

Atena Editora

☎ +55 (42) 3323-5493

☎ +55 (42) 99955-2866

🌐 www.atenaeditora.com.br

✉ contato@atenaeditora.com.br

CONSELHO EDITORIAL

CONSELHO EDITORIAL

Prof. Dr. Alexandre Igor Azevedo Pereira – Instituto Federal Goiano
Profª Drª Amanda Vasconcelos Guimarães – Universidade Federal de Lavras
Prof. Dr. Antonio Pasqualetto – Pontifícia Universidade Católica de Goiás
Profª Drª Ariadna Faria Vieira – Universidade Estadual do Piauí
Prof. Dr. Arinaldo Pereira da Silva – Universidade Federal do Sul e Sudeste do Pará
Prof. Dr. Benedito Rodrigues da Silva Neto – Universidade Federal de Goiás
Prof. Dr. Cirênio de Almeida Barbosa – Universidade Federal de Ouro Preto
Prof. Dr. Cláudio José de Souza – Universidade Federal Fluminense
Profª Drª Daniela Reis Joaquim de Freitas – Universidade Federal do Piauí
Profª Drª. Dayane de Melo Barros – Universidade Federal de Pernambuco
Prof. Dr. Eloi Rufato Junior – Universidade Tecnológica Federal do Paraná
Profª Drª Érica de Melo Azevedo – Instituto Federal do Rio de Janeiro
Prof. Dr. Fabrício Menezes Ramos – Instituto Federal do Pará
Prof. Dr. Fabrício Moraes de Almeida – Universidade Federal de Rondônia
Profª Drª Glécilla Colombelli de Souza Nunes – Universidade Estadual de Maringá
Prof. Dr. Humberto Costa – Universidade Federal do Paraná
Prof. Dr. Joachin de Melo Azevedo Sobrinho Neto – Universidade de Pernambuco
Prof. Dr. João Paulo Roberti Junior – Universidade Federal de Santa Catarina
Profª Drª Juliana Abonizio – Universidade Federal de Mato Grosso
Prof. Dr. Julio Candido de Meirelles Junior – Universidade Federal Fluminense
Profª Drª Keyla Christina Almeida Portela – Instituto Federal de Educação, Ciência e Tecnologia do Paraná
Profª Drª Miranilde Oliveira Neves – Instituto de Educação, Ciência e Tecnologia do Pará
Prof. Dr. Sérgio Nunes de Jesus – Instituto Federal de Educação Ciência e Tecnologia
Profª Drª Talita de Santos Matos – Universidade Federal Rural do Rio de Janeiro
Prof. Dr. Tiago da Silva Teófilo – Universidade Federal Rural do Semi-Árido
Prof. Dr. Valdemar Antonio Paffaro Junior – Universidade Federal de Alfenas

APRESENTAÇÃO

APRESENTAÇÃO

Este segundo livro intitulado “Caderno no. 2 Estatística Aplicada a biologia usando o R: manipulação de tabelas, testes estatísticos e abordagens de ANOVA para a análise de dados biológicos” faz parte da continuidade das atividades semestrais do projeto de extensão coordenado pelo Professor EBTT Elton Nunes Britto no Instituto Federal de Educação Ciência e Tecnologia do Pará, IFPA – Campus Santarém intitulado “Bioestatística Aplicada com software R para Biólogos e áreas afins”. A ideia principal do projeto é e continua sendo a de compartilhar os conhecimentos de R com o usuário de R para a aplicação da estatística descritiva e inferencial. Cada tópico deste segundo livro compõe-se de passos básicos para a construção de tabelas, realização de testes estatísticos, sendo a última parte voltada para a aplicação dos tipos de somatórios da Análise de variância - ANOVA quando ocorre a mudança de posição das variáveis preditivas a partir da função de ANOVA dois e três fatores no R. Quanto a seção das tabelas, o intuito não foi de se aprofundar no assunto, e sim de passar para o usuário de R as formas de se construir novas tabelas utilizando-se diferentes funções encontradas em diferentes pacotes, a partir de um banco de dados tomando-se o cuidado de passar as linhas básicas de comando, a fim de facilitar as operações para o usuário de R. Quanto ao tópico dos testes estatísticos, a ideia principal foi de levar ao conhecimento dos usuários de R como é útil trabalhar com o pacote “Rcmdr” mostrando-se que é possível utilizar o R de forma semelhante aos softwares tradicionais de estatística, onde clicam-se em ícones, realizando-se de forma rápida e prática a estatística descritiva e inferencial. Quanto ao último tópico sobre os somatórios utilizados na ANOVA, com destaque, nos testes de ANOVA tipo II e III, o enfoque principal foi de mostrar como variam os resultados destes testes com a simples mudança de posição de entrada das variáveis preditivas na função da ANOVA para os tipos II e III, uma vez que não há uma ordem de qual variável preditiva irá entrar primeiro na função da ANOVA para esses tipos. Destaca-se que a simples modificação da ordem das variáveis preditivas pode alterar o resultado final e causar confusão na mente do usuário de R. Este tópico então, visa esclarecer os procedimentos que devem ser tomados quando isto ocorre indicando o que fazer para obter os resultados corretos da ANOVA tipo II e III. Desta forma, a continuidade desse projeto de extensão dá-se não somente pela realização de cursos de R para discentes do IFPA e de faculdades e universidades de Santarém, e sim também pela produção de materiais didáticos que servirão de material de apoio ficando disponíveis na forma de e-book e na forma impressa para os usuários de R, sendo de grande valia para o registro das atividades do projeto e também como fonte de consulta para o público em geral.

SUMÁRIO

SUMÁRIO

BANCO DE DADOS E TABELAS	1
Manipulação de tabelas no R	1
Exemplo da utilização das funções dos pacotes supracitados no R para manipular tabelas no R.....	2
USANDO O COMANDO SELECT() DO PACOTE “DPLYR”	3
FILTRANDO LINHAS COM O COMANDO FILTER() DO PACOTE “DPLYR”	4
ADICIONANDO UMA COLUNA CONTENDO 3 LINHAS UTILIZANDO-SE O PACOTE “TIBBLE”	5
ADICIONANDO UMA COLUNA ANTES DA COLUNA 1 QUE É A RENDA...	6
ADICIONANDO UMA COLUNA DEPOIS DA COLUNA 2 QUE É A ESCOLARIDADE	7
ADICIONANDO NOMES NAS COLUNAS CRIADAS (ADICIONADAS) DE UMA ÚNICA VEZ.....	8
USANDO A FUNÇÃO MUTATE() PARA CALCULAR O ÍNDICE IMC= PESO/ALTURA²	9
MOSTRANDO UMA TABELA COM APENAS UMA COLUNA COM A FUNÇÃO TRANSMUTATE()	11
CRIANDO UM VETOR OU UM CONJUNTO NUMÉRICO COM A FUNÇÃO PULL() A PARTIR DE UMA TABELA.....	12
ALTERANDO A ORDEM DOS VALORES DO IMC PARA ORDEM CRESCENTE COM A FUNÇÃO ARRANGE()	13
CONSTRUINDO NOVAS TABELAS COMBINANDO AS FUNÇÕES SELECT(), ARRANGE() E FILTER().....	14
UNINDO TABELAS DIFERENTES CONTENDO VARIÁVEIS EM COMUM COM A FUNÇÃO BIND_COLS()	15
UNINDO TABELAS FORMADAS COM VARIÁVEIS DIFERENTES COM A FUNÇÃO BIND_COLS().....	16
FILTRANDO DUAS VARIÁVEIS COM A FUNÇÃO FILTER().....	17

SUMÁRIO

SUMÁRIO

USANDO A FUNÇÃO `BIND_ROWS()` PARA ADICIONAR LINHAS EM TABELAS COM O MESMO NÚMERO DE LINHAS OU REPETIÇÕES..... 19

Primeiro passo: criando o vetor idade.....	19
Vetor idade	19
Verificando o “n” amostral do vetor idade.....	19
Segundo passo: adicionando o vetor idade na tabela rr com o pacote “tibble”	19
Divisão da nova tabela rr11.....	20
Primeiro passo: criando a tabelax, sendo representada pelas 5 primeiras linhas da tabela rr11, da 1ª até 5ª linha e n=5	20
Segundo passo: criando a tabelay, sendo representada pelas 5 últimas linhas da tabela rr11, da 6ª até a 10ª linha com n=5.....	21
Unindo as tabelas x e y com a função <code>bind_rows()</code> retornando a tabela rr11 com n=10 linhas	21

USANDO A FUNÇÃO `BIND_ROWS()` PARA ADICIONAR LINHAS EM TABELAS COM NÚMERO DIFERENTE DE LINHAS OU REPETIÇÕES..... 22

Primeiro passo: criando a tabelax1 formada com as quatro primeiras linhas da tabela rr11, da 1ª a 4ª linha e n=4.....	22
Segundo passo: criando a tabelay1 formada com as seis últimas linhas da tabela rr11, da 6ª a 10ª linha e n=6	22
Unindo as tabelasx1 e tabelay1 com a função <code>bind_rows()</code> retornando a tabela rr11 com n=10.....	23

CONSTRUINDO GRÁFICOS COM O PACOTE “RCMDR”24

Abrindo a interface do pacote “Rcmdr”	24
---------------------------------------------	----

FAZENDO GRÁFICOS COM O PACOTE “RCMDR”: GRÁFICO DE DISPERSÃO DE PONTOS.....25

Caminho para fazer o gráfico de dispersão de pontos na interface do “Rcmdr”	25
Conjunto de passos para gerar o gráfico de dispersão de pontos	25

SUMÁRIO

SUMÁRIO

Acessando a pasta no computador com a linha de comando “Importar arquivo de dados”	25
Resultado do gráfico de diagrama de dispersão de pontos no console do “Rcmdr”	26
Linha de comando gerada para o gráfico de dispersão de pontos na interface do “Rcmdr”	27
Opções de salvamento do gráfico gerado em PDF	27
Opção “ver conjunto de dados”	28
Opção “Editar conjunto de dados”	29
Mudando o valor 147 da coluna “baltura” para 150	29
FAZENDO GRÁFICO DE LINHAS NA INTERFACE DO “RCMDR”	31
Caminho para fazer o gráfico de linhas na interface do “Rcmdr”	31
Linha de comando gerada para o gráfico de linhas na interface do “Rcmdr”	31
FAZENDO GRÁFICO DE DISPERSÃO DE PONTOS XY (CONDICIONADO) NA INTERFACE DO “RCMDR”	32
Caminho para fazer o gráfico de dispersão xy (condicionado) na interface do “Rcmdr”	32
Linha de comando gerada para o gráfico XY (dispersão) condicionado na interface do “Rcmdr”	33
Adicionando título e nome dos eixos x e y no gráfico XY (dispersão) condicionado	33
FAZENDO GRÁFICO DE HISTOGRAMA NA INTERFACE DO “RCMDR” ..	35
Caminho para fazer o gráfico de histograma na interface do “Rcmdr”	35
Linha de comando gerada para o gráfico de histograma na interface do “Rcmdr”	36
Adicionando “Legendas” ao histograma	36

SUMÁRIO

SUMÁRIO

FAZENDO GRÁFICO DE BOX-PLOT NA INTERFACE DO PACOTE “RCMDR”	37
Conjunto de passos para gerar o gráfico de Boxplot simples	37
Resultado do gráfico Boxplot no console do “Rcmdr”	37
Linha de comando gerada para o gráfico Boxplot na interface do “Rcmdr”	38
Conjunto de passos para gerar o gráfico de Boxplot da idade por sexo	38
Escolhendo a variável independente considerando o grupo por sexo.....	38
Linha de comando gerada para o gráfico Boxplot da idade versus o tipo de sexo na interface do “Rcmdr”	39
FAZENDO GRÁFICO DE BARRAS NA INTERFACE DO PACOTE “RCMDR”	40
Conjunto de passos para gerar o gráfico de barras por sexo.....	40
Escolhendo a variável independente considerando o grupo por sexo.....	40
Linha de comando gerada para o gráfico barras para o sexo na interface do “Rcmdr”	41
Conjunto de passos para gerar o gráfico de barras por sexo e escolaridade	41
Linha de comando gerada para o gráfico de barras por tipo de sexo e tipo de escolaridade na interface do “Rcmdr”	42
FAZENDO GRÁFICO DE PIZZA NA INTERFACE DO PACOTE “RCMDR”	43
Conjunto de passos para gerar o gráfico de pizza por sexo	43
Escolhendo a variável independente sexo para fazer o gráfico de pizza.....	43
Linha de comando gerada para o gráfico de Pizza por tipo de sexo na interface do “Rcmdr” com a resposta expressada em porcentagem	44
Escolhendo a variável independente sexo para fazer o gráfico de pizza com a resposta por frequência absoluta	44
Linha de comando gerada para o gráfico de Pizza por tipo de sexo na interface do “Rcmdr” e resposta expressada em frequência absoluta.....	45

SUMÁRIO

SUMÁRIO

TESTE T DE STUDENT E TEST U DE MANN-WHITNEY NO “RCMDR” ... 46

Teste T de Student: comparação da idade média x sexo	46
Separando as idades de homens e de mulheres pelo uso dos colchetes	47
Amostra de idade de homem, repetições de idade e “n” amostral	47
Amostra de idade de mulher, repetições de idade e “n” amostral	47
Caminho para fazer teste T de Student no Rcommander	48
Resultado do teste T de Student no pacote “Rcmdr”	49
Linha de comando do resultado do teste T de Student no pacote “Rcmdr”	49
Extraindo a média da idade da amostra de homem e de mulher pela função <code>tapply()</code>	49
Extraindo a média de idade da amostra de homem e de mulher pela função <code>describeBY</code> do pacote “psych”	50

CALCULANDO-SE AS PREMISSAS PARA FAZER O TESTE T DE STUDENT51

Caminho para fazer o teste de Normalidade no Rcommander	51
Linha de comando gerada para o teste de normalidade na interface do “Rcmdr”	51

HIPÓTESES DO TESTE DE NORMALIDADE52

Teste de Hipótese segundo Crawley (2013)	52
Resultado do teste de normalidade no console do “Rcmdr”	52
Caminho para se fazer o teste da Normalidade por grupos	52
Resultado do teste de Normalidade por grupos no pacote “Rcmdr”	53

CAMINHO PARA FAZER O TESTE DE HOMOGENEIDADE DAS VARIÂNCIAS NO “RCMDR”55

Linha de comando gerada para o teste da homogeneidade de variâncias por sexo na interface do “Rcmdr”	56
---------------------------------------------------------------------------------------------------------------	----

SUMÁRIO

SUMÁRIO

HIPÓTESES DO TESTE DE HOMOGENEIDADE DAS VARIÂNCIAS	57
Teste de Hipótese segundo Crawley (2013)	57
Hipóteses das variâncias:	57
Resultado do teste da homogeneidade de variâncias no console do “Rcmdr”	57
TESTE U DE MANN-WHITNEY NO “RCMDR”	59
Caminho para fazer o teste U de Mann-Whitney no “Rcmdr”	59
HIPÓTESES DO TESTE U DE MANN-WHITNEY	61
Linha de comando gerada para o teste U de Mann-Whitney idade por sexo na interface do “Rcmdr”	61
Resultado do teste U de Mann-Whitney no console do “Rcmdr”	61
FAZENDO O TESTE DE ANOVA ONE WAY NO “RCMDR”	62
Caminho para fazer teste de ANOVA one way no Rcommander	62
Linha de comando gerada para o teste de ANOVA one way no “Rcmdr” com a saída mostrando a média e o desvio padrão	62
Linha de comando gerada para o teste de ANOVA one way no “Rcmdr” com a saída mostrando a média e o erro padrão.	63
HIPÓTESES DO TESTE DE ANOVA ONE WAY	65
Teste de Hipótese segundo Zar (1999)	65
FAZENDO O TESTE DE TUKEY NO “RCMDR”	66
Linha de comando gerada para o teste de Tukey no “Rcmdr”	66
Resultado do teste de Tukey no console do Rstudio	67
PREMISSAS PARA FAZER O TESTE DE ANOVA ONE WAY NO “RCMDR”	71
Linha de comando gerada para o teste de normalidade no “Rcmdr” da variável peso	71

SUMÁRIO

SUMÁRIO

HIPÓTESES DO TESTE DE NORMALIDADE	72
Teste de Hipótese segundo Crawley (2013).....	72
Resultado do teste de Normalidade no Rstudio ou R raiz da variável peso	72
Caminho para fazer o teste de Normalidade no “Rcmdr” para a variável resposta peso considerando os níveis de religião.....	72
Linha de comando gerada para o testes de normalidade no “Rcmdr” por grupos	73
HIPÓTESES DO TESTE DE NORMALIDADE.....	74
Teste de Hipótese segundo Crawley (2013).....	74
Resultado do teste de normalidade no console do Rstudio	74
FAZENDO TESTE DE HOMOGENEIDADE DE VARIÂNCIAS NO “RCMDR”	75
Caminho para fazer o teste de homogeneidade de variâncias no “Rcmdr”	75
Linha de comando gerada para o teste de homogeneidade de variâncias de Bartlett	75
PREMISSAS DO TESTE DE HOMOGENEIDADE DE VARIÂNCIAS	76
Teste de Hipótese segundo Crawley (2013).....	76
Resultado do teste de homogeneidade de variâncias no console do Rstudio.....	76
FAZENDO TESTE DE CORRELAÇÃO NO “RCMDR”	77
Caminho para fazer o teste de correlação no “Rcmdr”	77
Linha de comando gerada para o teste correlação no “Rcmdr”	78
Gráfico de dispersão no “Rcmdr” para avaliar a correlação entre peso x altura	78
Caminho para fazer um gráfico de dispersão no “Rcmdr”	78
Resultado do gráfico de dispersão do “Rcmdr”	78
Resultado do teste de correlação no console do Rstudio.....	79

SUMÁRIO

SUMÁRIO

ESCOLHA DO TIPO DE SOMATÓRIO TIPO I, II E III PARA CALCULAR ANOVA TWO WAY E THREE WAY NO R	80
HIPÓTESES RELATIVAS AO FATOR RELIGIÃO	81
Teste de Hipótese segundo Zar (1999).....	81
HIPÓTESES RELATIVAS AO FATOR SEXO	82
Teste de Hipótese segundo Zar (1999).....	82
HIPÓTESES RELATIVAS À INTERAÇÃO ENTRE OS FATORES RELIGIÃO E SEXO	83
Teste de Hipótese segundo Zar (1999).....	83
TESTE DE ANOVA TWO WAY COM O SOMATÓRIO DOS QUADRADOS DO TIPO I NO R.....	84
Separação das amostras de peso em função do sexo e de peso em função da religião	84
Aplicando-se o teste de ANOVA two way como somatório do tipo I entrando primeiro com a variável preditora religião.....	84
Aplicando-se o teste de ANOVA two way como somatório do tipo I entrando primeiro com a variável preditora sexo.....	85
Aplicando-se o teste de ANOVA two way como somatório do tipo II entrando primeiro com a variável preditora religião.....	85
Aplicando-se o teste de ANOVA two way como somatório do tipo II entrando primeiro com a variável preditora sexo.....	86
Aplicando-se o teste de ANOVA two way como somatório do tipo III entrando primeiro com a variável preditora religião.....	87
Aplicando-se o teste de ANOVA two way como somatório do tipo III entrando primeiro com a variável preditora sexo.....	87
REFERÊNCIAS.....	89
SOBRE OS AUTORES	91



C A P Í T U L O 1

BANCO DE DADOS E TABELAS

Este tópico não tem a pretensão de mostrar como se trabalha com banco de dados no R, e sim das principais ferramentas usadas para a manipulação de tabelas no R. Inicialmente é interessante destacar que segundo Carvalho (2020) um banco de dados é formado campos ou atributos, linhas ou tuplas onde são formadas as tabelas. Os campos ou atributos são as colunas; as tuplas registram a informação das colunas e as tabelas são formadas por colunas e linhas. Segundo este autor o banco de dados é um lugar que armazena informações para consultas ou utilização. Um dos tipos de dados armazenados no banco de dados é o do tipo relacional formado por variáveis quantitativas e qualitativas. E o relacionamento das tabelas do banco de dados ocorre por uma chave de ligação primária ou primary key – PK, e chave estrangeira ou foreign key – FK. Se pensarmos que um banco de dados relaciona várias tabelas entre si, então uma tabela representa uma parte de um banco de dados, e que de forma simples pode ser consultada e utilizada para extrair informações à luz da estatística.

Manipulação de tabelas no R

Prosseguindo neste tópico vamos apontar pacotes usados para a manipulação de tabelas no R, tais como: “tidyverse” por Wickham et al., (2019), “dplyr” por Wickham et al., (2023) e “magrittr” por Bache e Wickham (2022) e “tibble” por Muller e Wickham (2023). Observação: na linguagem R o nome das variáveis e das tabelas deve ser o mais simples possível, não sendo recomendado acentuação gráfica, letras maiúsculas misturados com minúsculas e palavras compostas separadas com hífen ou espaços.

Exemplo da utilização das funções dos pacotes supracitados no R para manipular tabelas no R

Com a tabela fictícia chamada de `renda.xlsx` iremos começar a aplicar as ferramentas do `para` para a manipulação de tabelas segundo Alcoforado (2021) no R ou Rstudio. Para importar o comando desta tabela é necessário instalar e carregar o pacote do Excel chamado de “`readxl`” por Wickham e Bryan (2025). Esta tabela será copiada e salva no objeto `rr` (tabela `rr`) sendo convertida em uma `data.frame` pelo comando: `nome da tabela<- as.data.frame(nome da tabela)`, no caso `rr<-as.data.frame(rr)`.

Tabela 1. Tabela `rr` para a manipulação de linhas e colunas no R

n	peso	altura	sexo	est civil	locomocao	renda	escolaridade	estuda
1	75	185	0	casado	carro	2	graduado	ufopa
2	65	198	1	casado	onibus	6	ensino medio	ufopa
3	40	178	1	solteiro	carro	5	graduado	uepa
4	75	152	1	casado	carro	5	posgraduado	ifpa
5	54	187	0	solteiro	onibus	8	posgraduado	uepa
6	56	188	0	solteiro	onibus	5	graduado	ifpa
7	65	189	0	solteiro	carro	9	ensino medio	ufopa
8	48	167	1	casado	carro	7	graduado	uepa
9	55	187	0	solteiro	onibus	8	posgraduado	uepa
10	56	188	1	solteiro	onibus	5	graduado	ifpa

`estcivil`=estado civil; `ufopa`= Universidade Federal do Oeste do Pará;
`ifpa`=Instituto Federal do Pará; e `uepa`=Universidade Estadual do Pará.

A estrutura desta tabela apresentou: as variáveis `peso` (kg), `altura` (cm), `renda` (número de salários) como numéricas e `sexo` (0=feminino e 1=masculino), `estcivil` (casado e solteiro), `locomocao` (carro e onibus), `escolaridade` (ensinomedio, graduado e posgraduado) e `estuda` (`ufopa`, `uepa` e `ifpa`) como caractere. Posteriormente as variáveis do tipo caractere foram convertidas em factor pelo comando `rr$nome da variavel<- as.factor(rr$nome da variável)`, como exemplo para a variável `sexo`, `rr$sexo<- as.factor(rr$sexo)`.



CAPÍTULO 2

USANDO O COMANDO SELECT() DO PACOTE “DPLYR”

Começaremos a usar o pacote “dplyr” e a função `select()` para fazer a seleção das variáveis “renda”, “escolaridade” e “estuda” e utilizaremos o comando: nome da tabela seguido do operador pipe identificado pelo símbolo `%>%` e `select(renda, escolaridade, estuda)` e `control + enter`. Nesta seleção foram extraídas da tabela `rr` as variáveis supracitadas e reunidas formando uma nova tabela denominada de `tab1`.

Tabela 2. A tabela `tab1` originada pelo comando: `tab1<- rr %>% select(renda, escolaridade, estuda)`

n	renda	escolaridade	estuda
1	2	graduado	ufopa
2	6	ensinomedio	ufopa
3	5	graduado	uepa
4	5	posgraduado	ifpa
5	8	posgraduado	uepa
6	5	graduado	ifpa
7	9	ensinomedio	ufopa
8	7	graduado	uepa
9	8	posgraduado	uepa
10	5	graduado	ifpa

Fonte. Autores (2025)



CAPÍTULO 3

FILTRANDO LINHAS COM O COMANDO FILTER() DO PACOTE “DPLYR”

Continuando com a aplicação das funções de manipulação de tabelas segundo Alcoforado (2021) agora iremos utilizar a função `filter()` para separar informações mais específicas das variáveis. Lembrando que usamos até aqui usamos a `tab1<-rr %>% select(renda, escolaridade, estuda)`. Agora com a função `filter()` escolheremos somente quem estuda na ufopa copiando e salvando na nova tabela, a tabela `tab2` ficando com a linha de comando `tab2<-tab1 %>% filter(estuda=="ufopa")`. Com este comando a `tab2` mostra o que foi filtrado da tabela `tab1` contendo quem estuda na ufopa, assim como as colunas e linhas onde contém quem estuda na ufopa (ver tabela 3).

Tabela 3. Tabela `tab2` com a linha de comando `tab2<-tab1 %>% filter(estuda=="ufopa")` para fazer uma separação mais específica de somente quem estuda na ufopa.

n	renda	escolaridade	estuda
1	2	graduado	ufopa
2	6	ensinomedio	ufopa
7	9	ensinomedio	ufopa

Fonte. Autores (2025)

Observação: Na tabela 3 o número 7 corresponde ao estudante número 7 conforme visualizado na tabela 2.



CAPÍTULO 4

ADICIONANDO UMA COLUNA CONTENDO 3 LINHAS UTILIZANDO- SE O PACOTE “TIBBLE”

Dando continuidade as alterações nas tabelas originadas pela aplicação das funções que manipulam as tabelas segundo Alcoforado (2021) vamos adicionar a tabela `tab2` uma coluna que contenha até três linhas formando outra nova tabela, a `tab3`. A `tab3` terá a linha de comando `tab3<-tab2`. Lembrando que a `tab2` é `tab2<-tab1 %>% filter(estuda=="ufopa")`. Então a `tab3` será `tab3<-tab2 %>% tibble::add_column(adicionada=1:3)`. Aqui os dois pontos duplos indicam que estamos utilizando a função `add_column` que é específica do pacote “tibble”, com o argumento “`adicionada=1:3`” de uma até três linhas, adicionando a nova coluna após a última coluna. Esta adição é feita depois da última coluna como um default da função (ver tabela 4).

Tabela 4. A `tab3 <- tab2 %>% tibble::add_column(adicionada=1:3)` para adicionar uma nova coluna com até 3 linhas a tabela `tab2` que já contém um filtro de somente quem estuda na ufopa.

n	renda	escolaridade	estuda	adicionada
1	2	graduado	ufopa	1
2	6	ensinomedio	ufopa	2
7	9	ensinomedio	ufopa	3

Fonte. Autores (2025)



CAPÍTULO 5

ADICIONANDO UMA COLUNA ANTES DA COLUNA 1 QUE É A RENDA

Agora vamos adicionar uma nova coluna antes da variável “renda” formando uma nova tabela, a tab4. A tab4 será `tab4<- tab3 %>% tibble::add_column(adicionada.1=1:3, .before=”renda”)`. Uma particularidade é que para diferenciar a nova coluna adicionada temos que chamar a nova coluna de adicionada.1.

Tabela 4.1. A tab4 com adição da coluna adicionada.1 antes da variável “renda” sob a linha de comando `tab4<-tab3 %>% tibble::add_column(adicionada.1=1:3,.before=”renda”)`.

n	adicionada.1	renda	escolaridade	estuda	adicionada
1	1	2	graduado	ufopa	1
2	2	6	ensinomedio	ufopa	2
7	3	9	ensinomedio	ufopa	3

Fonte. Autores (2025)



CAPÍTULO 6

ADICIONANDO UMA COLUNA DEPOIS DA COLUNA 2 QUE É A ESCOLARIDADE

Prosseguindo com a manipulação de tabelas segundo Alcoforado (2021) agora iremos adicionar uma nova coluna que chamaremos de adicionada.2 depois da variável “escolaridade” para diferenciar das outras colunas adicionadas anteriormente formando outra nova tabela, a tab5. Novamente para facilitar esta operação iremos copiar e salvar no objeto tab5. A tab5 será `tab5<- tab4 %>% tibble::add_column(adicionada.2=1:3,after=”escolaridade”)`. Para que possamos fazer esta adição iremos acrescentar o argumento `after` que significa depois em português (ver tabela 5).

Tabela5. A `tab5<-tab4%>%tibble::add_column(adicionada.2=1:3,after=”escolaridade”)` com a adição da coluna `adicionada.2` após a variável `escolaridade`.

n	adicio- nada1	renda	escolaridade	adicionada2	estuda	adicio- nada
1	1	2	graduado	1	ufopa	1
2	2	6	ensino medio	2	ufopa	2
7	3	9	ensino medio	3	ufopa	3

Fonte. Autores (2025)



CAPÍTULO 7

ADICIONANDO NOMES NAS COLUNAS CRIADAS (ADICIONADAS) DE UMA ÚNICA VEZ

Prosseguindo na aplicação das funções manipuladoras de tabela no R segundo Alcoforado (2021), iremos acrescentar nomes em todas as colunas ao mesmo tempo, numa única operação copiando e salvando numa nova tabela a `tab6` que tem a linha de comando `tab6<-tab5`. É importante destacar que a tabela `tab5`, sendo `tab5<- tab4 %>% tibble::add_column(adicionada.2=1:3,after="escolaridade")` vem acumulando todas as tabelas anteriores. As mudanças de nomes nas operações permite nos organizarmos melhor, apesar de que todas estas operações acumuladas poderiam ser feitas com o nome de um único objeto, porém iria ficar bem confuso para o usuário de R entender, contrariando o objetivo deste livro, que é facilitar a vida do usuário de R.

Desta forma, para acrescentarmos o nome a todas as colunas, ao mesmo tempo, usaremos a linha de comando `tab6 %>% rename(n=adicionada.1, horasestudo=adicionada.2, horasestudo1=adicionada)` a `adicionada.1` passou a se chamar "n", a `adicionada.2` de `horasestudo` e `adicionada` de `Horasestudo1`. Assim estas alterações feitas com a `tab6` serão copiadas e salvas em outra nova tabela, a `tab7`, ficando a com a linha de comando: `tab7<-tab6 %>% rename(n=adicionada.1,horasestudo=adicionada.2, horasestudo1=adicionada)` (ver tabela 6).

Tabela 6. A `tab7 <- tab6 %>% rename(n=adicionada.1, horasestudo=adicionada.2, horasestudo1=adicionada)`.

n	renda	escolaridade	horasestudo	estuda	horasestudo1
1	2	graduado	1	ufopa	1
2	6	ensinomedio	2	ufopa	2
7	9	ensinomedio	3	ufopa	3

Fonte. Autores (2025)



CAPÍTULO 8

USANDO A FUNÇÃO MUTATE() PARA CALCULAR O ÍNDICE IMC= PESO/ALTURA²

Continuando com as funções do R para a manipulação de tabelas segundo Alcoforado (2021) usando as funções do pacote “dplyr” segundo Wickham et al., (2023). Usaremos a função mutate() para calcular o índice de massa corporal - imc com as variáveis peso e altura da tabela 7 que tem como fórmula $imc = peso / altura^2$. Utilizaremos o objeto rr ou tabela rr que inicialmente foi a nossa tabela mãe e que copiaremos e salvaremos o objeto rr1 ou tabela rr1, que conterá apenas as cinco primeiras linhas do objeto rr, usando-se o `tabela rr1 <- rr [-c(6:9),]`. Com esta nova tabela rr1 nós iremos usar a função mutate(), para isto usaremos a linha de comando `rr1 %>% mutate(imc=peso/altura^2)` e esta nova linha de comando será copiada e salva na nova tabela rr2, sendo a nova tabela `rr2 <- rr1 %>% mutate(imc=peso/altura^2)`. Ver a tabela rr na tabela 7 e a tabela rr1 na tabela 8 e a tabela rr2 na tabela 9. Para o cálculo do imc a variável altura foi convertida para centímetros.

Tabela7. A tabela rr é tabela tabela mãe contendo a altura em centímetros.

n	peso	altura	sexo	estcivil	locomocao	renda	escolaridade	estuda
1	75	185	0	casado	carro	2	graduado	ufopa
2	65	198	1	casado	onibus	6	ensinomedio	ufopa
3	40	178	1	solteiro	carro	5	graduado	uepa
4	75	152	1	casado	carro	5	posgraduado	ifpa
5	54	187	0	solteiro	onibus	8	posgraduado	uepa
6	56	188	0	solteiro	onibus	5	graduado	ifpa
7	65	189	0	solteiro	carro	9	ensinomedio	ufopa
8	48	167	1	casado	carro	7	graduado	uepa
9	55	187	0	solteiro	onibus	8	posgraduado	uepa
10	56	188	1	solteiro	onibus	5	graduado	ifpa

estcivil=estado civil; ufopa= universidade federal do oeste do pará;
ifpa=instituto federal do pará; e uepa=universidade estadual do pará.

Fonte. Autores (2025)

Tabela 8. Tabela rr1 contendo as primeiras cinco linhas com o comando `rr1<-rr[-c(6:9),]` o peso em (Kg) e a altura em (m) para formar a nova tabela rr2.

n	peso	altura	sexo	estcivil
1	75	1.85	0	casado
2	65	1.98	1	casado
3	40	1.78	1	solteiro
4	75	1.52	1	casado
5	54	1.87	0	solteiro

Fonte. Autores (2025)

A tabela rr1 foi copiada e salva na nova tabela rr2 pelo comando `rr2<- rr1 %>% mutate(imc=peso/altura^2)` (ver tabela 9).

Tabela 9. Tabela rr2 contendo o cálculo do Índice de massa corporal

n	peso	altura	sexo	estcivil	imc
1	75	1.85	0	casado	21.91381
2	65	1.98	1	casado	16.57994
3	40	1.78	1	solteiro	12.62467
4	75	1.52	1	casado	32.46191
5	54	1.87	0	solteiro	15.44225

Fonte. Autores (2025)



CAPÍTULO 9

MOSTRANDO UMA TABELA COM APENAS UMA COLUNA COM A FUNÇÃO TRANSMUTATE()

De acordo com Alcoforado (2021) e usando-se a função `transmutate()` criaremos uma tabela contendo somente a coluna da variável `imc`. A nova tabela chamada de `rr3` (ver tabela 10) terá a linha de comando: `rr3<-rr2 %>% transmutate(imc=peso/altura^2)`.

Tabela 10. Tabela `rr3` contendo somente a coluna do `imc`.

Imc
21.91381
16.57994
12.62467
32.46191
15.44225

Fonte. Autores (2025)



C A P Í T U L O 10

CRIANDO UM VETOR OU UM CONJUNTO NUMÉRICO COM A FUNÇÃO `PULL()` A PARTIR DE UMA TABELA

Dando prosseguimento as funções de manipulação de tabelas no R segundo Alcoforado (2021) usaremos a função `pull()` para transformarmos uma tabela em um vetor ou conjunto numérico. Faremos isto com os valores de `imc` da tabela `rr3` (ver tabela 10). Então a tabela `rr3` terá a linha de comando `rr3 %>% pull(imc)` sendo convertida no conjunto numérico abaixo:

```
> [1] 21.91381 16.57994 12.62467 32.46191 15.44225
```



CAPÍTULO 11

ALTERANDO A ORDEM DOS VALORES DO IMC PARA ORDEM CRESCENTE COM A FUNÇÃO ARRANGE()

Continuando com os valores do imc da tabela 9 vamos ordená-los em ordem crescente segundo Alcoforado (2021) usando a função `arrange()`. Utilizando a tabela `rr2` sob a linha de comando `rr2 %>% arrange(imc)` obteremos a nova tabela a `rr4` sob a linha de comando `rr4<- rr2 %>% arrange(imc)` (ver tabela 11).

Tabela 11. Tabela `rr4` com os valores de imc ordenados em ordem crescente

n	peso	altura	sexo	estcivil	imc
1	40	1.78	1	solteiro	12.62467
2	54	1.87	0	solteiro	15.44225
3	65	1.98	1	casado	16.57994
4	75	1.85	0	casado	21.91381
5	75	1.52	1	casado	32.46191

Fonte. Autores (2025)

Avançando na manipulação de tabelas no R segundo Alcoforado (2021) e usando a função `arrange()` na mesma tabela 9 converteremos os valores do imc para ordem decrescente. Para fazermos isto basta adicionarmos um sinal negativo antes da variável que esta como argumento da função `arrange(- nome da variável)`. Desta forma, com comando `rr2 %>% arrange(- imc)` os valores de imc da tabela 9 ficarão em ordem decrescente ficando a tab `rr4<- rr2 %>% arrange(- imc)` com os valores do imc em ordem decrescente (ver tabela 12).

Tabela 12. Tabela `rr4` com os valores do imc em ordem decrescentes.

n	peso	altura	sexo	estcivil	imc
5	75	1.52	1	casado	32.46191
4	75	1.85	0	casado	21.91381
3	65	1.98	1	casado	16.57994
2	54	1.87	0	solteiro	15.44225
1	40	1.78	1	solteiro	12.62467

Fonte. Autores (2025)



CAPÍTULO 12

CONSTRUINDO NOVAS TABELAS COMBINANDO AS FUNÇÕES SELECT(), ARRANGE() E FILTER()

Dando prosseguimento a manipulação de tabelas no `r` apoiado em Alcoforado (2021) vamos misturar várias funções usando o operador `%>%` (pipe). Vamos juntar as funções `select()`, `arrange()` e `filter()` para formar uma nova tabela que chamaremos de `rr5`. A nova tabela `rr5` terá as variáveis (`sexo`, `renda`, `altura` e `peso`) arranjadas por (`sexo` e `renda`) e filtrando os pesos maiores que 45Kg. A nova tabela terá a linha e comando `rr5<- rr %>% select(sexo,renda,altura,peso) %>% arrange(sexo,renda) %>% filter(peso>45)`. Na tabela `rr5` os valores estão ordenados em ordem crescente pelos valores da variável `sexo`, visto ser utilizada a função `arrange()` tendo como o primeiro argumento a variável `sexo` (ver tabela 13).

Tabela 13. A tabela `rr5` formada pela seleção das variáveis `sexo`, `renda`, `altura` e `peso` e o arranjo em ordem crescente da variável `sexo` em primeiro lugar contendo pesos maiores que 45Kg.

sexo	renda	altura	peso
0	2	1.85	75
0	8	1.87	54
1	5	1.52	75
1	6	1.98	65

Fonte. Autores (2025)



CAPÍTULO 13

UNINDO TABELAS DIFERENTES CONTENDO VARIÁVEIS EM COMUM COM A FUNÇÃO BIND_COLS()

Continuando com a manipulação de tabelas no R em conformidade com Alcoforado (2021) uniremos diferentes tabelas que contenham variáveis em comum usando a função `bind_cols()`. Trabalhando com a tabela `rr1` `%>% mutate(imc=peso/altura^2)` vamos construir a tabela A uma nova tabela, a tabela A será formada pelas variáveis `peso` e `altura` e a tabela B pelas variáveis `peso` e `imc`. A nova tabela A foi formada com o comando `tabelaA<- rr2 %>% select(peso,altura)` e a nova `tabelaB<- rr2 %>% select(peso, imc)`. A variável em comum é `peso` e a nova tabela será formada pela linha de comando `bind_cols(tabelaA,tabelaB)` sendo denominada de tabela `rr6` (ver tabela 14) sob a linha de comando `rr6<- bind_cols(tabelaA,tabelaB)`. Neste caso a variável `peso...1` é porque está na coluna 1 e `peso...3` na coluna 3. Um detalhe peculiar do R é que os nomes das tabelas podem ser escritos pelo usuário como: `tabela 1` ou `tabela rr` ou objeto `rr` ou `rr1` e etc, sendo feitos conforme a conveniência dos autores.

Tabela 14. Nova tabela `rr6` foi gerada pela função `bind_cols(tabelaA,tabelaB)`.

peso...1	altura	peso...3	imc
75	1.85	75	21.91381
65	1.98	65	16.57994
40	1.78	40	12.62467
75	1.52	75	32.46191
54	1.87	54	15.44225

Fonte. Autores (2025)



CAPÍTULO 14

UNINDO TABELAS FORMADAS COM VARIÁVEIS DIFERENTES COM A FUNÇÃO BIND_COLS()

Prosseguindo com a manipulação de tabelas no R de acordo com Alcoforado (2021), agora usando distintas variáveis e a mesma função `bind_cols()` é possível unir estas novas tabelas formando outra tabela que chamaremos de `rr7`. A nova tabela `rr7` formada pela tabela `rr2` e pela união de duas tabelas: a `tabelaA` e `tabelaB`. A tabela `A` com a linha de comando `tabelaA<- rr2 %>% select(renda,altura)` e `tabelaB<- rr2 %>% select(imc,peso)`. A linha de comando da `rr7<-rr2 bind_cols(tabelaA,tabelaB)` ver (tabela 15).

Tabela 15. A tabela `rr7<-rr2 bind_cols(tabelaA,tabelaB)`.

renda	altura	Imc	peso
2	1.85	21.91381	75
6	1.98	16.57994	65
5	1.78	12.62467	40
5	1.52	32.46191	75
8	1.87	15.44225	54

Fonte. Autores (2025)



CAPÍTULO 15

FILTRANDO DUAS VARIÁVEIS COM A FUNÇÃO FILTER()

Prosseguindo com a manipulação de tabelas no R segundo Alcoforado (2021) vamos utilizar uma tabela criada para aplicar a função `filter()` chamada de tabela `rr8` (ver tabela 16). Com a tabela `rr8` vamos usar a função `filter()` para selecionar indivíduos com alturas maiores que 1.78m e renda maior que 5 salários. Copiaremos e salvaremos esta função na nova tabela, a tabela `rr9` `rr2 %>% filter(altura > "1.78 & renda > "5")` (ver tabela 17). Observação: para produzir a tabela `rr9` copia-se a tabela `rr8` criada para um arquivo do bloco de notas e importa-se para o R com a função `read.table("nome salvo da tabela na extensão. txt, T)` e depois prossegue-se a criação da tabela `rr9` com a função `filter()`.

Tabela 16. A nova Tabela `rr8` criada

n	peso	altura	sexo	estcivil	locomocao	renda	escolaridade	estuda	imc
1	75	1.85	0	casado	carro	2	graduado	ufopa	21.91381
2	65	1.98	1	casado	onibus	6	ensinomedio	ufopa	16.57994
3	40	1.78	1	solteiro	carro	5	graduado	uepa	12.62467
4	75	1.52	1	casado	carro	5	posgraduado	ifpa	32.46191
5	54	1.87	0	solteiro	ônibus	8	posgraduado	uepa	15.44225

Fonte. Autores (2025)

Tabela 17. Tabela `rr9` `rr2 %>% filter(altura > "1.78 & renda > "5")`.

n	peso	altura	sexo	estcivil	locomocao	renda	escolaridade	estuda	imc
2	65	1.98	1	casado	onibus	6	ensinomedio	ufopa	16.57994
5	54	1.87	0	solteiro	onibus	8	posgraduado	uepa	15.44225

Fonte. Autores (2025)

Continuando com a aplicação da função filter() vamos filtrar pessoas com o peso maior que 40Kg copiando e salvando na nova tabela rr10 com a linha de comando `rr10<-rr8 %>% filter(peso>"40" & altura<"1.98")` (ver tabela 18).

Tabela 18. Tabela rr10 <-rr8 %>% filter(peso>"40" & altura<"1.98").

n	peso	altura	sexo	estcivil	locomocao	renda	escolaridade	estuda	imc
1	75	1.85	0	casado	carro	2	graduado	ufopa	21.91381
4	75	1.52	1	casado	carro	5	posgraduado	ifpa	32.46191
5	54	1.87	0	solteiro	onibus	8	posgraduado	uepa	15.44225

Fonte. Autores (2025)



C A P Í T U L O 16

USANDO A FUNÇÃO `BIND_ROWS()` PARA ADICIONAR LINHAS EM TABELAS COM O MESMO NÚMERO DE LINHAS OU REPETIÇÕES

Continuando com a manipulação de tabelas no `r` segundo Alcoforado (2021) vamos usar a função `bind_rows()` para adicionar linhas em tabelas com o mesmo número de repetições. Para fazer isto criaremos vetores com o mesmo número de linhas da tabela onde serão adicionadas as linhas. Usaremos a tabela `rr` a nossa tabela mãe inicialmente e adicionaremos a variável `idade`.

Primeiro passo: criando o vetor `idade`

O vetor `idade` será criado com a linha de comando `idade<-c(as repetições de idade separadas por vírgula)`. Depois confirmaremos o número de repetições ou o `n` amostral com a função `length(idade)` do objeto `idade`.

Vetor `idade`

```
idade<-c(25,27,36,42,65,54,52,42,63,47)
```

```
[1] 25 27 36 42 65 54 52 42 63 47
```

Verificando o “`n`” amostral do vetor `idade`

```
length(idade)
```

```
[1] 10
```

Segundo passo: adicionando o vetor `idade` na tabela `rr` com o pacote “`tibble`”

Depois de criado o vetor `idade`, ele será adicionado a tabela `rr` pela linha de comando `rr %>% tibble::add_column(idade)`. A nova tabela copiaremos e salvaremos na nova tabela `rr1`, sendo `rr1<- rr %>% tibble::add_column(idade)` (ver a tabela 19).

Tabela 19. Tabela rr11 < rr %>% tibble::add_column(idade)

n	peso	altura	sexo	estcivil	locomocao	renda	escolaridade	estuda	idade
1	75	185	0	casado	carro	2	graduado	ufopa	25
2	65	198	1	casado	onibus	6	ensinomedio	ufopa	27
3	40	178	1	solteiro	carro	5	graduado	uepa	36
4	75	152	1	casado	carro	5	posgraduado	ifpa	42
5	54	187	0	solteiro	onibus	8	posgraduado	uepa	65
6	56	188	0	solteiro	onibus	5	graduado	ifpa	54
7	65	189	0	solteiro	carro	9	ensinomedio	ufopa	52
8	48	167	1	casado	carro	7	graduado	uepa	42
9	55	187	0	solteiro	onibus	8	posgraduado	uepa	63
10	56	188	1	solteiro	onibus	5	graduado	ifpa	47

Fonte. Autores (2025)

Divisão da nova tabela rr11

A nova tabela rr11 será dividida em duas tabelas, a tabela x e a tabela y, onde `tabelax<-rr11[-c(6:10),]` será formada com as cinco primeiras linhas da tabela rr11. E a `tabelay<-rr11[-c(1:5),]` com as cinco últimas linhas da tabela rr11. Agora com a função `bind_rows()` uniremos as tabelas x e y. A linha de comando `bind_rows(tabelax,tabelay)` retorna a tabela rr11.

Primeiro passo: criando a tabelax, sendo representada pelas 5 primeiras linhas da tabela rr11, da 1ª até 5ª linha e n=5

De acordo com Britto *et al.* (2021) utilizando-se os colchetes exclui-se as cinco últimas linhas com a linha de comando `rr11[-c(6:10),]` permanecendo uma tabela com as cinco primeiras colunas copiada e salva na `tabelax<- rr11[-c(6:10),]` (ver tabela 20).

Tabela 20. Tabelax

n	peso	altura	sexo	estcivil	locomocao	renda	escolaridade	estuda	idade
1	75	185	0	casado	carro	2	graduado	ufopa	25
2	65	198	1	casado	onibus	6	ensinomedio	ufopa	27
3	40	178	1	solteiro	carro	5	graduado	uepa	36
4	75	152	1	casado	carro	5	posgraduado	ifpa	42
5	54	187	0	solteiro	onibus	8	posgraduado	uepa	65

Fonte. Autores (2025)

Segundo passo: criando a tabelay, sendo representada pelas 5 últimas linhas da tabela rr11, da 6ª até a 10ª linha com n=5.

Conforme Britto *et al.* (2021) na tabelay<-rr11[-c(1:5),] foram retiradas as cinco primeiras linhas, permanecendo da sexta até a décima linha (ver tabela 21).

Tabela 21. Tabelay

n	peso	altura	sexo	estcivil	locomocao	renda	escolaridade	estuda	idade
6	56	188	0	solteiro	onibus	5	graduado	ifpa	54
7	65	189	0	solteiro	carro	9	ensinomedio	ufopa	52
8	48	167	1	casado	carro	7	graduado	uepa	42
9	55	187	0	solteiro	onibus	8	posgraduado	uepa	63
10	56	188	1	solteiro	onibus	5	graduado	ifpa	47

Fonte. Autores (2025)

Unindo as tabelas x e y com a funcao bind_rows() retornando a tabela rr11 com n=10 linhas

Continuando com a manipulação de tabelas no R segundo Alcoforado (2021) e usando a função `bind_rows(tabelax,tabelay)` retornando a tabela original rr11 (ver tabela 22).

Tabela 22. Tabela rr11 formada da união das tabelasx e tabelay

n	peso	altura	sexo	estcivil	locomocao	renda	escolaridade	estuda	idade
1	75	185	0	casado	carro	2	graduado	ufopa	25
2	65	198	1	casado	onibus	6	ensinomedio	ufopa	27
3	40	178	1	solteiro	carro	5	graduado	uepa	36
4	75	152	1	casado	carro	5	posgraduado	ifpa	42
5	54	187	0	solteiro	onibus	8	posgraduado	uepa	65
6	56	188	0	solteiro	onibus	5	graduado	ifpa	54
7	65	189	0	solteiro	carro	9	ensinomedio	ufopa	52
8	48	167	1	casado	carro	7	graduado	uepa	42
9	55	187	0	solteiro	onibus	8	posgraduado	uepa	63
10	56	188	1	solteiro	onibus	5	graduado	ifpa	47

Fonte. Autores (2025)



CAPÍTULO 17

USANDO A FUNÇÃO `BIND_ROWS()` PARA ADICIONAR LINHAS EM TABELAS COM NÚMERO DIFERENTE DE LINHAS OU REPETIÇÕES

Continuando a manipulação de tabelas no `r` segundo Alcoforado (2021) a partir da tabela `rr11` iremos dividi-la em uma `tabelax1` contendo as quatro primeiras linhas e a `tabelay1` contendo as seis últimas linhas para depois com a função `bind_rows()` formar a mesma tabela `rr11`.

Primeiro passo: criando a `tabelax1` formada com as quatro primeiras linhas da tabela `rr11`, da 1ª a 4ª linha e `n=4`

Usando a linha de comando `rr11[-c(5:10),]` foram retiradas as seis últimas linhas ficando uma tabela com as quatro primeiras linhas (ver tabela 23).

Tabela 23. `Tabelax1`

n	peso	altura	sexo	estcivil	locomocao	renda	escolaridade	estuda	idade
1	75	185	0	casado	carro	2	graduado	ufopa	25
2	65	198	1	casado	onibus	6	ensinomedio	ufopa	27
3	40	178	1	solteiro	carro	5	graduado	uepa	36
4	75	152	1	casado	carro	5	posgraduado	ifpa	42

Fonte. Autores (2025)

Segundo passo: criando a `tabelay1` formada com as seis últimas linhas da tabela `rr11`, da 6ª a 10ª linha e `n=6`

Usando a linha de comando `rr11[-c(1:4),]` foi feita a retirada das quatro primeiras linhas ficando uma `tabelay1` com as seis últimas linhas (ver tabela 24).

Tabela 24. Tabelay1

n	peso	altura	sexo	estcivil	locomocao	renda	escolaridade	estuda	idade
5	54	187	0	solteiro	onibus	8	posgraduado	uepa	65
6	56	188	0	solteiro	onibus	5	graduado	ifpa	54
7	65	189	0	solteiro	carro	9	ensinomedio	ufopa	52
8	48	167	1	casado	carro	7	graduado	uepa	42
9	55	187	0	solteiro	onibus	8	posgraduado	uepa	63
10	56	188	1	solteiro	onibus	5	graduado	ifpa	47

Fonte. Autores (2025)

Unindo as tabelasx1 e tabelay1 com a função bind_rows() retornando a tabela rr11 com n=10.

Prosseguindo com a manipulação de tabelas com r segundo Alcoforado (2021) e a linha de comando `bind_rows(tabelax1,tabelay1)` retornaremos a tabela original rr11 (ver tabela 25).

Tabela 25. Tabela formada pela união das tabelasx1 e tabelay1

n	peso	altura	sexo	estcivil	locomocao	renda	escolaridade	estuda	idade
1	75	185	0	casado	carro	2	graduado	ufopa	25
2	65	198	1	casado	onibus	6	ensinomedio	ufopa	27
3	40	178	1	solteiro	carro	5	graduado	uepa	36
4	75	152	1	casado	carro	5	posgraduado	ifpa	42
5	54	187	0	solteiro	onibus	8	posgraduado	uepa	65
6	56	188	0	solteiro	onibus	5	graduado	ifpa	54
7	65	189	0	solteiro	carro	9	ensinomedio	ufopa	52
8	48	167	1	casado	carro	7	graduado	uepa	42
9	55	187	0	solteiro	onibus	8	posgraduado	uepa	63
10	56	188	1	solteiro	onibus	5	graduado	ifpa	47

Fonte. Autores (2025)



CAPÍTULO 18

CONSTRUINDO GRÁFICOS COM O PACOTE “RCMDR”

Abrindo a interface do pacote “Rcmdr”

Abre-se um script no R raiz ou Rstudio , instala-se e carrega-se o pacote “Rcmdr”. Feito isto irá aparecer uma interface deste pacote denominada R Commander (ver figura 1)

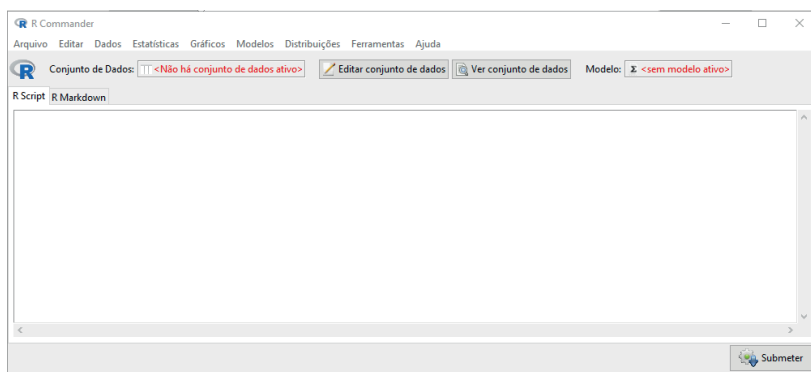


Figura 1. Interface do pacote “Rcmdr”



C A P Í T U L O 19

FAZENDO GRÁFICOS COM O PACOTE “RCMDR”: GRÁFICO DE DISPERSÃO DE PONTOS

Com a interface do “R Commander” aberta utilizou-se o arquivo em Excel `alturapeso.xls` para fazer todos os gráficos deste tópico. O pacote “Rcmdr” segundo Fox e Bouchet-Valat (2024) exhibe a interface chamada de “R Commander” onde aparecem os ícones na barra de tarefas, tais como: Arquivo, Editar, Dados, Estatísticas, Gráficos, Modelos, Distribuições, Ferramentas, Ajuda (ver figura 1).

Caminho para fazer o gráfico de dispersão de pontos na interface do “Rcmdr”

Conjunto de passos para gerar o gráfico de dispersão de pontos

Dados - Importar arquivo de dados – Importar o conjunto em Excel - Defina um nome do conjunto de dados – nomeie o arquivo em Excel- Acesse o diretório onde está o arquivo `alturapeso.xls` – selecione o arquivo em Excel `alturapeso.xls` – Gráficos – Diagrama de dispersão... - Dados – nome da variável x – nome da variável y – ok ou aplicar e é gerado o gráfico de dispersão de pontos.

Acessando a pasta no computador com a linha de comando “Importar arquivo de dados”

Após clicar em dados e Importar arquivo do tipo Excel selecionando “do arquivo Excel” do conjunto de dados ativo Dataset teremos no final desta operação acesso ao diretório do computador onde está o arquivo `alturapeso.xls` e no Rscript do “Rcmdr” aparecerá este caminho percorrido (ver figura 2).

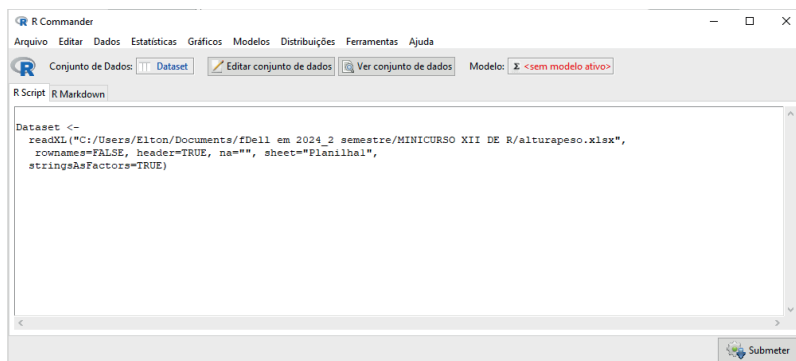


Figura 2. Caminho que foi percorrido mostrando o diretório de onde foi obtido o arquivo alturapeso.xls agora chamado de Dataset.

Observação: quando o conjunto de dados não está ativo no “R Commander” aparece ao lado da caixa “Conjunto de dados” a frase “não há conjunto de dados ativo” escrito em cor vermelha (ver figura 1). E quando o conjunto de dados está ativo aparece ao lado da caixa “Conjunto de dados” o nome do arquivo definido pelo usuário escrito em cor azul (ver figura 2). Desta forma pode-se beneficiar das funções do pacote “Rcmdr” e fazer estatísticas e gráficos.

Resultado do gráfico de diagrama de dispersão de pontos no console do “Rcmdr”

O resultado mostra um gráfico de dispersão de pontos na janela gráfica. O gráfico mostrado pode ser salvo nos formatos: PDF, PNG, e JPEG nos níveis de qualidade: 50%, 75% e 100%. Existem outros formatos, mas citaremos os mais importantes para o nosso uso. Este gráfico pode ser salvo na pasta de escolha do usuário (ver figura 3).

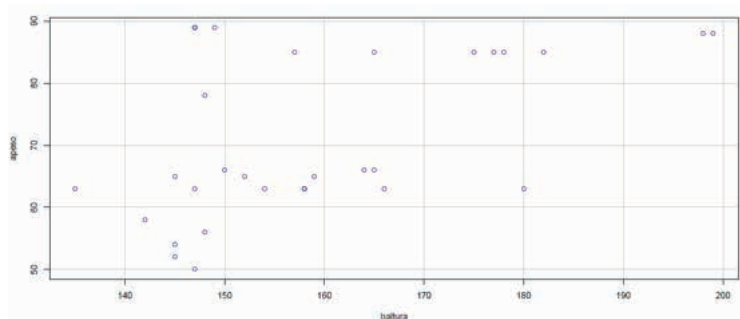


Figura 3. Gráfico de dispersão de pontos do pacote “Rcmdr”

Linha de comando gerada para o gráfico de dispersão de pontos na interface do “Rcmdr”

> scatterplot(apeso~baltura, regLine=FALSE, smooth=FALSE, boxplots=FALSE, data=pontos) (ver figura 4).

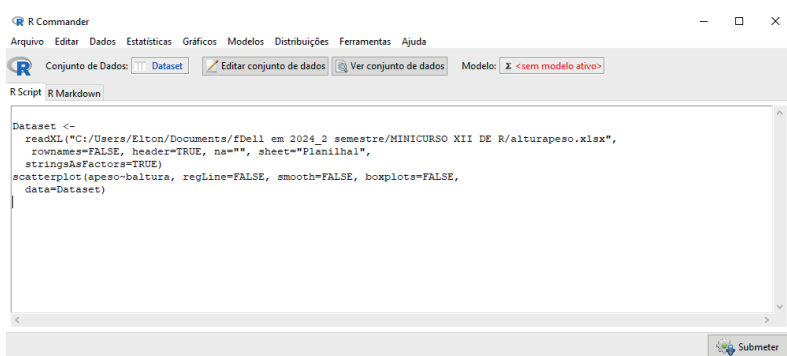


Figura 4. Caminho do scatterplot indicando a linha de comando.

Opções de salvamento do gráfico gerado em PDF

Para salvar o gráfico em PDF percorre-se o caminho: Arquivo – Salvar como – PDF. Isto gerará um arquivo em PDF que poderá ser salvo na pasta da escolha do usuário. Existem outras opções interessantes, tais como: Arquivo – copiar para a área de transferência – como um Bitmap que é o comando ctrl + c, nesta opção você copia e salva no texto em Word, por exemplo. Usando a opção ctrl + c gerou-se o gráfico copiado para esta página (ver figura 5). Há a opção Arquivo – Imprimir gerando uma impressão do gráfico.

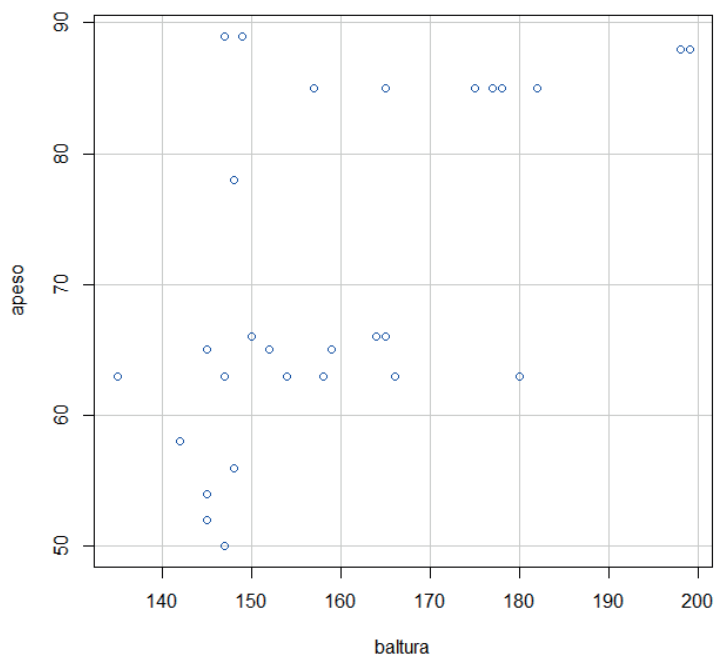


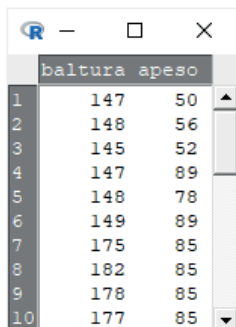
Figura 5. Gráfico copiado para esta página com o comando Arquivo – copiar para a área de transferência – como um Bitmap – ctrl + c.

Opção “ver conjunto de dados”

Nesta opção podemos ver a tabela alturapeso.xls salva com o nome de Dataset

Percorremos o caminho – “Ver conjunto de dados” (ver figura 1) onde aparece o ícone

“ver conjunto de dados” e (figura 6) visualizando o conjunto de dados.



	baltura	apeso
1	147	50
2	148	56
3	145	52
4	147	89
5	148	78
6	149	89
7	175	85
8	182	85
9	178	85
10	177	85

Figura 6. Após clicar em “Ver conjunto de dados” temos a nossa tabela com os dados de altura e peso.

Opção “Editar conjunto de dados”

Nesta opção podemos editar o conjunto de dados de qualquer coluna da tabela altura e peso. Percorremos o caminho – “Editar conjunto de dados” (ver figuras 1 e 7). Nesta opção é possível adicionar tanto linhas quanto colunas e modificar o valor das variáveis como podemos verificar na posição do cursor no valor da linha 1 e coluna baltura registrado como 147 seguido do cursor posicionado a direita do número.

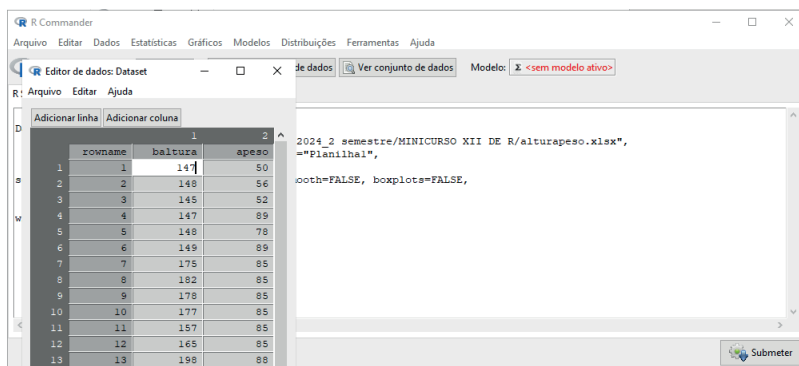


Figura 7. Opção “Editar o conjunto de dados”

Mudando o valor 147 da coluna “baltura” para 150

Depois de mudar o valor de 147 para 150 ou fazer qualquer alteração deve-se clicar em Arquivo – sair e salvar ficando registrada a alteração (ver figura 8). Posteriormente alteramos novamente o valor para 147 para prosseguir com os demais gráficos.

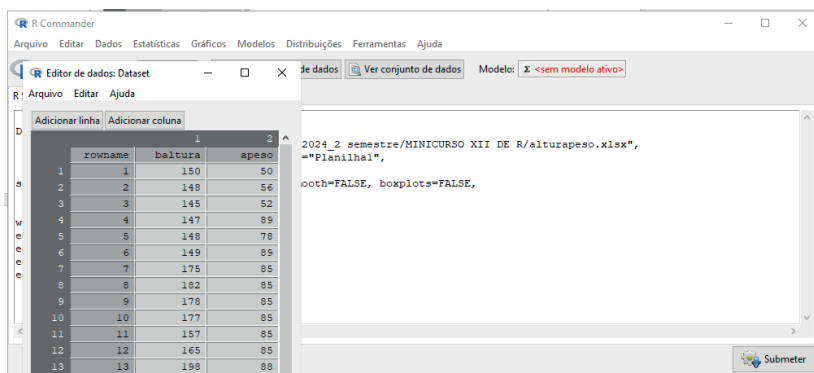


Figura 8. Mudança do valor 147 para 150



CAPÍTULO 20

FAZENDO GRÁFICO DE LINHAS NA INTERFACE DO “RCMDR”

Caminho para fazer o gráfico de linhas na interface do “Rcmdr”

Dados - Importar arquivo de dados – Importar o conjunto em Excel - Defina um nome do conjunto de dados – nomeie o arquivo em Excel- Acesse o diretório onde está o arquivo alturapeso.xls – selecione o arquivo em Excel – Gráficos - Gráfico de Linha - Dados – nome da variável x – nome da variável y – ok ou aplicar e é gerado o gráfico de Linhas. (ver figura 9).

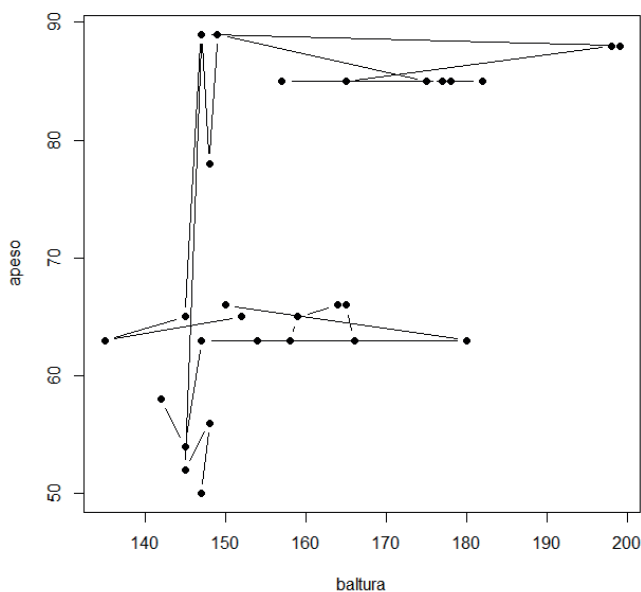


Figura 9. Gráfico de linhas

Linha de comando gerada para o gráfico de linhas na interface do “Rcmdr”

```
> with(Dataset, lineplot(baltura, apeso))
```



CAPÍTULO 21

FAZENDO GRÁFICO DE DISPERSÃO DE PONTOS XY (CONDICIONADO) NA INTERFACE DO “RCMDR”

Caminho para fazer o gráfico de dispersão xy (condicionado) na interface do “Rcmdr”

Dados - Importar arquivo de dados – Importar o conjunto em Excel - Defina um nome do conjunto de dados – nomeie o arquivo em Excel- Acesse o diretório onde está o arquivo alturapeso.xls – selecione o arquivo em Excel – Gráficos – gráfico XY (dispersão) condicionado - Dados – nome da variável x ou explicativa – nome da variável y ou resposta – ok ou aplicar (ver figura 10). Foi gerado um gráfico de pontos, no argumento type=”p” e formato do ponto pelo argumento pch=”16” indicado para o formato de losango.

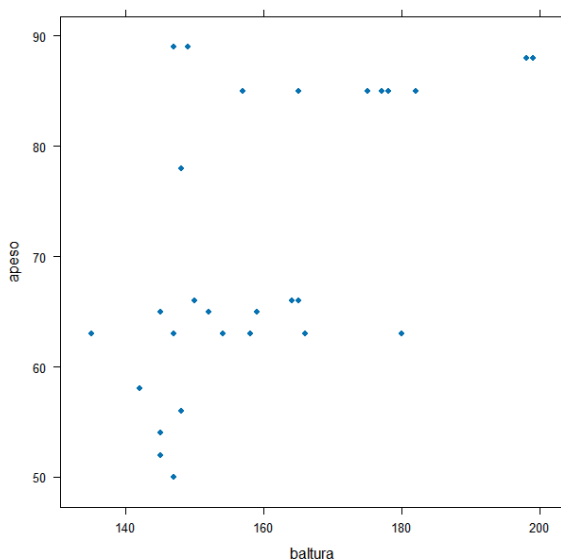


Figura 10. Gráfico XY (dispersão) condicionado

Linha de comando gerada para o gráfico XY (dispersão) condicionado na interface do “Rcmdr”

> xyplot(apeso ~ baltura, type="p", pch=16, auto.key=list(border=TRUE), par.settings=simpleTheme(pch=16),

Adicionando título e nome dos eixos x e y no gráfico XY (dispersão) condicionado

Para esta alteração percorremos o caminho: Gráficos – gráfico XY (dispersão) condicionado – Opções – Legendas adicionando-se os nomes nos: rótulo do eixo-x, rótulo do eixo-y e Título do gráfico. O default destas opções é automático. (ver figuras 10). Na figura 11 verificamos no gráfico que a legenda dos eixos foi alterada: no eixo do y, Altura foi adicionada a unidade de medida metros e no eixo do x, Peso em quilogramas, bem como o título nomeado de Gráfico de pontos. E na figura 12 verificamos o gráfico XY (dispersão) condicionado com as legendas dos eixos x e y e títulos alterados.

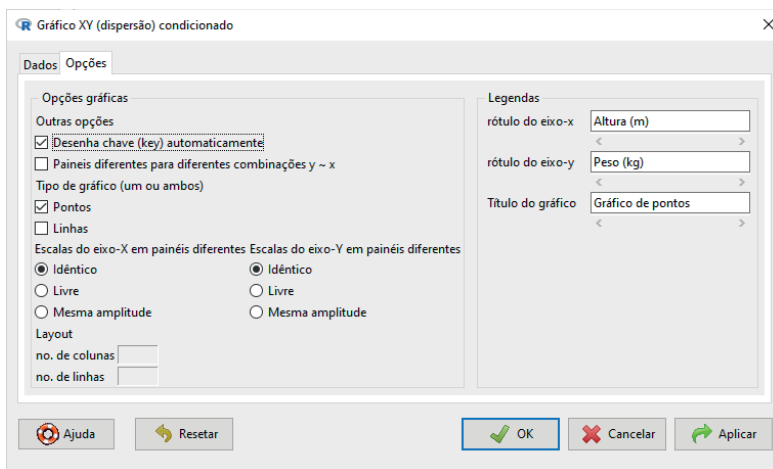


Figura 11. Adicionando-se as Legendas ao gráfico XY (dispersão) condicionado.

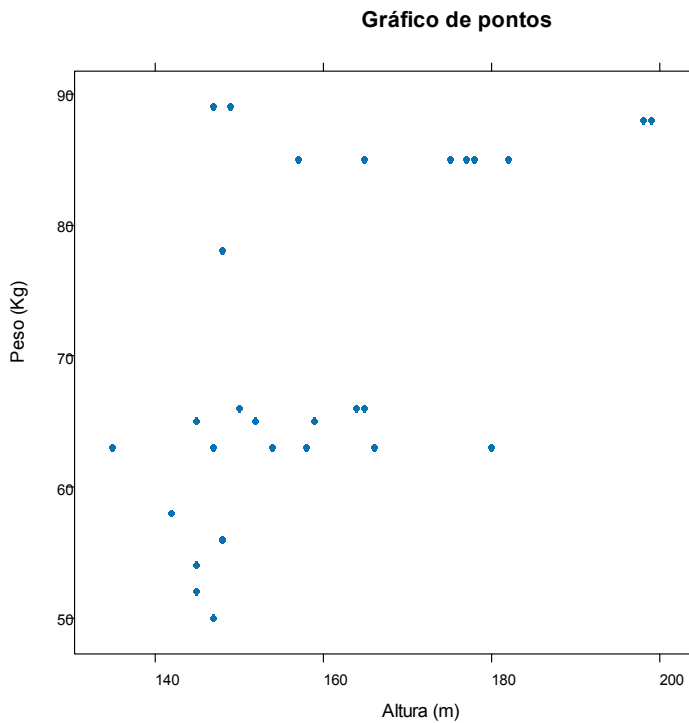


Figura 12. Gráfico XY (dispersão) condicionado com legenda dos eixos x e y e do título.



CAPÍTULO 22

FAZENDO GRÁFICO DE HISTOGRAMA NA INTERFACE DO “RCMDR”

Caminho para fazer o gráfico de histograma na interface do “Rcmdr”

Dados - Importar arquivo de dados – Importar o conjunto em Excel - Defina um nome do conjunto de dados – nomeie o arquivo em Excel- Acesse o diretório onde está o arquivo `alturapeso.xls` – selecione o arquivo em Excel – Gráficos – Histograma - Dados – nome da variável (selecione uma) – ok ou aplicar. Este comando gerará um gráfico de histograma simples. O gráfico de histograma é feito para se observar a distribuição de frequência da variável resposta, que no caso agora será a variável “`apeso`”. Quando é gerado o gráfico histograma a nossa variável resposta passa ser a “Frequência simples” de ocorrência dos valores de `apeso`, onde os valores de peso passam a ser agora nossa variável independente. Neste gráfico a nossa variável resposta agora é “`frequency`” a frequência absoluta e variável independente são os dados de “`apeso`” agrupados em classes de peso (ver figura 13) para um histograma simples.

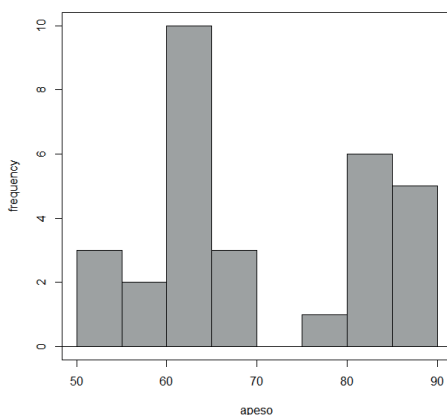


Figura 13. Histograma simples com a variável resposta a “`frequency`” ou a frequência simples e a variável independente a variável “`apeso`”.

Linha de comando gerada para o gráfico de histograma na interface do "Rcmdr"

```
> with(Dataset, Hist(apeso, scale="frequency", breaks="Sturges", col="darkgray"))
```

Adicionando "Legendas" ao histograma

Podemos adicionar nome nos rótulos acessando o caminho: Gráficos – Histograma – Opções – Legendas e modificando o nome dos rótulos dos eixos x e y e do título do gráfico. Também é possível na opção "Escala do eixo" mudar de Contagem de frequência para Porcentagens e Densidades (ver figura 14). Aqui não iremos colocar os nomes dos eixos x e y e do título porque esta operação é a mesma que gerou a figura 11 e 12.

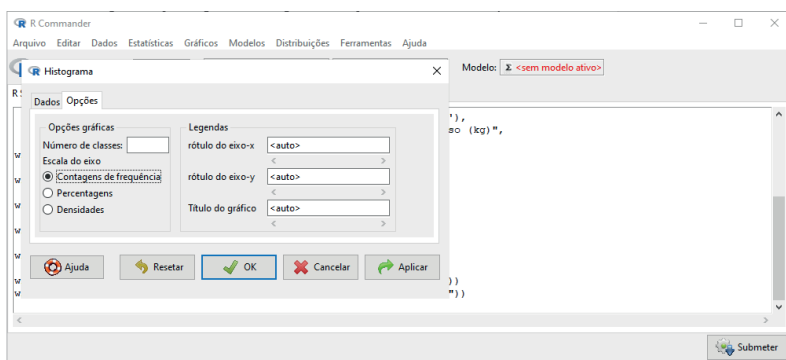


Figura 14. Alteração do nome dos rótulos do eixo x e y e título e tipo de escalas do histograma feito com a variável apeso.



CAPÍTULO 23

FAZENDO GRÁFICO DE BOX-PLOT NA INTERFACE DO PACOTE “RCMDR”

Conjunto de passos para gerar o gráfico de Boxplot simples

Dados - Importar arquivo de dados - Importar o conjunto em Excel - Defina um nome do conjunto de dados - nomeie o arquivo em Excel- Acesse o diretório onde está o arquivo estado civil_idade_peso_altura.xlsx - selecione o arquivo em Excel - Gráficos - Boxplot - Dados - nome da variável resposta, a idade - ok ou aplicar e será gerado um gráfico de Boxplot (ver figura 15).

Resultado do gráfico Boxplot no console do “Rcmdr”

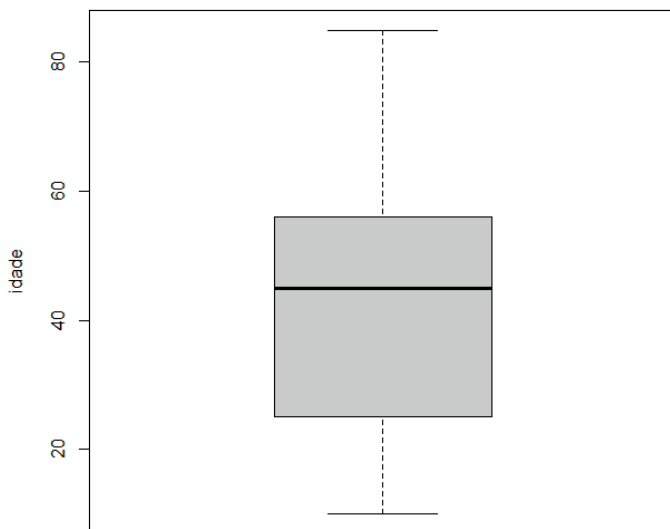


Figura 15. Gráfico Boxplot da idade no RComander.

Linha de comando gerada para o gráfico Boxplot na interface do “Rcmdr”

```
>Boxplot( ~ idade, data=Dataset, id=list(method="y"))
```

Conjunto de passos para gerar o gráfico de Boxplot da idade por sexo

Dados - Importar arquivo de dados – Importar o conjunto em Excel - Defina um nome do conjunto de dados – nomeie o arquivo em Excel- Acesse o diretório onde está o arquivo estado civil_idade_peso_altura.xlsx – selecione o arquivo em Excel – Gráficos - Boxplot - Dados – nome da variável independente, o sexo e o nome da variável resposta, a idade – ok ou aplicar, e será gerado um gráfico de Boxplot por grupos (ver figura 16 e 17 e 18).

Escolhendo a variável independente considerando o grupo por sexo

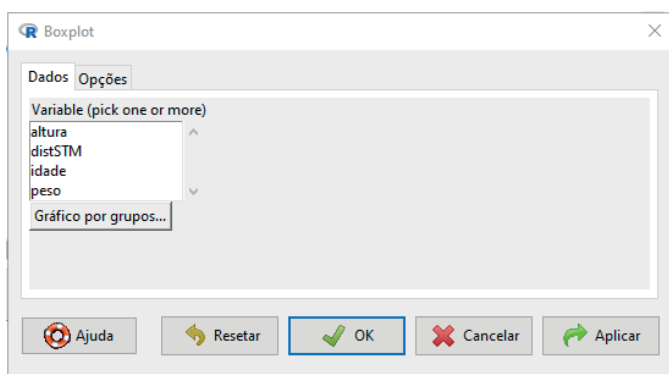


Figura 16. Escolhendo a idade por grupos

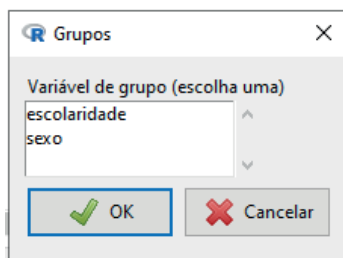


Figura 17. Escolhendo a idade por sexo

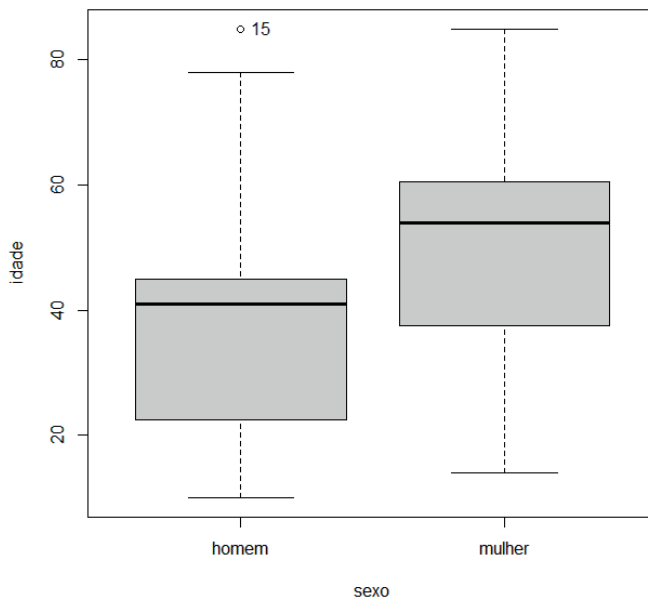


Figura 18. Gráfico Boxplot da idade por sexo nos níveis: homem e mulher.

Linha de comando gerada para o gráfico Boxplot da idade versus o tipo de sexo na interface do “Rcmdr”

```
> Boxplot(idade ~ sexo, data=Dataset, id=list(method="y"))
```



C A P Í T U L O 24

FAZENDO GRÁFICO DE BARRAS NA INTERFACE DO PACOTE “RCMDR”

Conjunto de passos para gerar o gráfico de barras por sexo

Dados - Importar arquivo de dados – Importar o conjunto em Excel - Defina um nome do conjunto de dados – nomeie o arquivo em Excel- Acesse o diretório onde está o arquivo estado_civil_idade_peso_altura.xlsx – selecione o arquivo em Excel – Gráficos – Gráfico de Barras - Dados – nome da variável independente, o sexo – ok ou aplicar. Neste caso, a variável resposta passa a ser a frequência e será gerado um gráfico de barras por grupos (ver figura 19 e 20).

Escolhendo a variável independente considerando o grupo por sexo

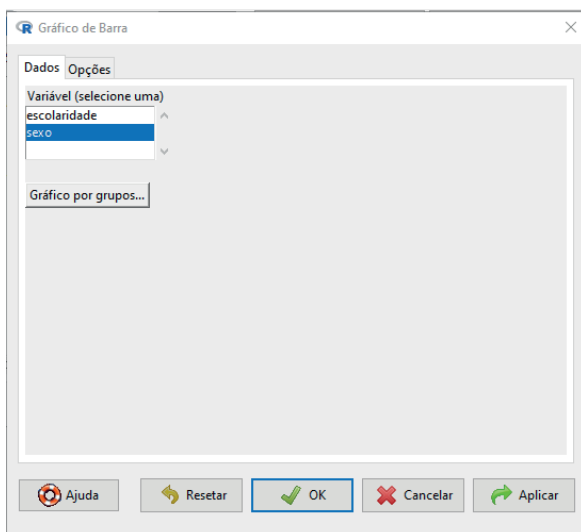


Figura 19. Selecionando a variável sexo para fazer o gráfico de barras.

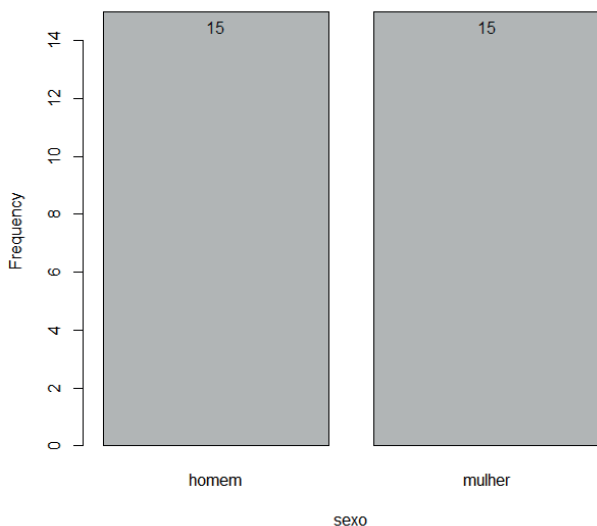


Figura 20. Gráfico de barras por sexo

Linha de comando gerada para o gráfico barras para o sexo na interface do “Rcmdr”

```
>with(Dataset, Barplot(sexo, xlab="sexo", ylab="Frequency", label.bars=TRUE))
```

Conjunto de passos para gerar o gráfico de barras por sexo e escolaridade

Dados - Importar arquivo de dados – Importar o conjunto em Excel - Defina um nome do conjunto de dados – nomeie o arquivo em Excel- Acesse o diretório onde está o arquivo estado_civil_idade_peso_altura.xlsx – selecione o arquivo em Excel – Gráficos – Gráfico de barras - Dados – nome da variável independente, o sexo – ok ou aplicar, e depois e clica-se em por grupos selecionando-se a variável escolaridade, e em ok ou aplicar (ver figura 21). E para ver o gráfico (figura 22).

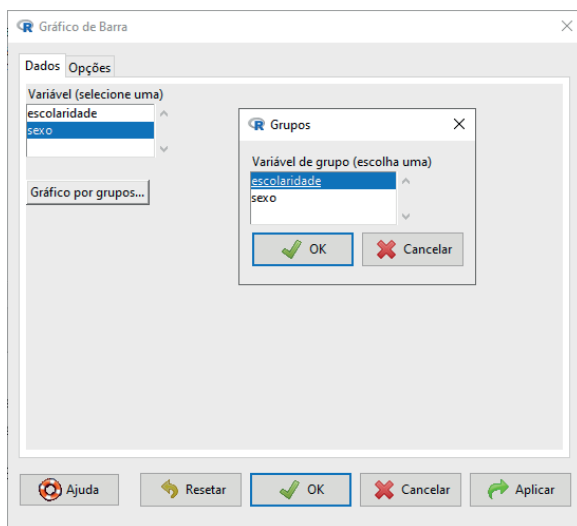


Figura 21. Escolhendo o grupo escolaridade para fazer o gráfico de barras por sexo e escolaridade.

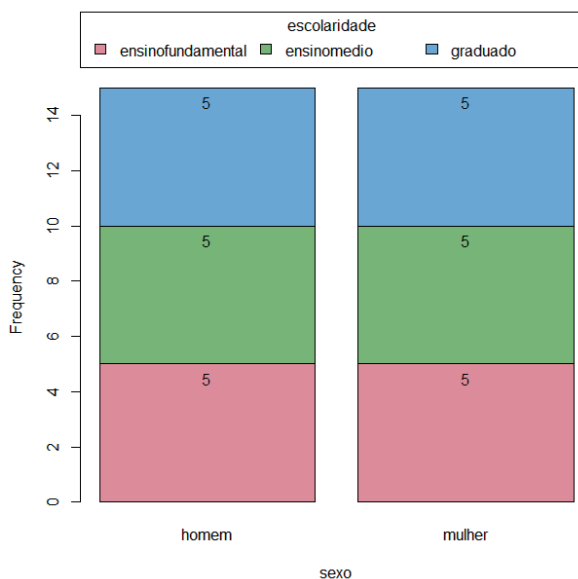


Figura 22. Gráfico de barras por sexo e escolaridade.

Linha de comando gerada para o gráfico de barras por tipo de sexo e tipo de escolaridade na interface do "Rcmdr"

```
>with(Dataset, Barplot(sexo, by=escolaridade, style="divided", legend.  
pos="above", xlab="sexo", ylab="Frequency", label.bars=TRUE))
```




CAPÍTULO 25

FAZENDO GRÁFICO DE PIZZA NA INTERFACE DO PACOTE “RCMDR”

Conjunto de passos para gerar o gráfico de pizza por sexo

Dados - Importar arquivo de dados – Importar o conjunto em Excel - Defina um nome do conjunto de dados – nomeie o arquivo em Excel- Acesse o diretório onde está o arquivo estado_civil_idade_peso_altura.xlsx – selecione o arquivo em Excel – Gráficos – Gráfico de Pizza - Dados – nome da variável independente, o sexo – ok ou aplicar. Neste caso, a variável resposta passa a ser a frequência em porcentagem (ver figura 23 e 24).

Escolhendo a variável independente sexo para fazer o gráfico de pizza

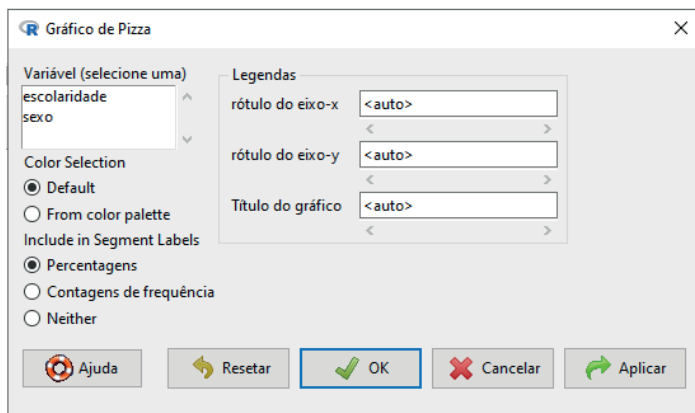


Figura 23. Escolhendo a variável sexo para o gráfico de pizza.

Neste caso da figura 24 escolhe-se a variável sexo e o gráfico tem o default para cor e com resposta em porcentagem. Para gerar o gráfico clica-se em ok ou aplicar.

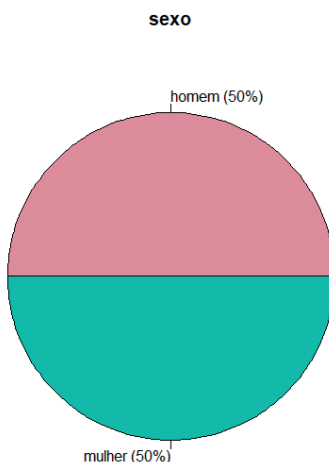


Figura 24. Gráfico de pizza por sexo com o resultado mostrado em frequência relativa ou porcentagem.

Linha de comando gerada para o gráfico de Pizza por tipo de sexo na interface do "Rcmdr" com a resposta expressada em porcentagem

```
>with(Dataset, piechart(sexo, xlab="", ylab="", main="sexo", col=rainbow_hcl(2), scale="percent"))
```

Escolhendo a variável independente sexo para fazer o gráfico de pizza com a resposta por frequência absoluta

Para fazer o gráfico de Pizza por sexo e a variável resposta frequência absoluta devemos selecionar em "Include in Segment Labels" marcar Contagens de frequência (ver figura 25). Com isto é gerado um gráfico de Pizza com a resposta em frequência absoluta conforme a figura 26.

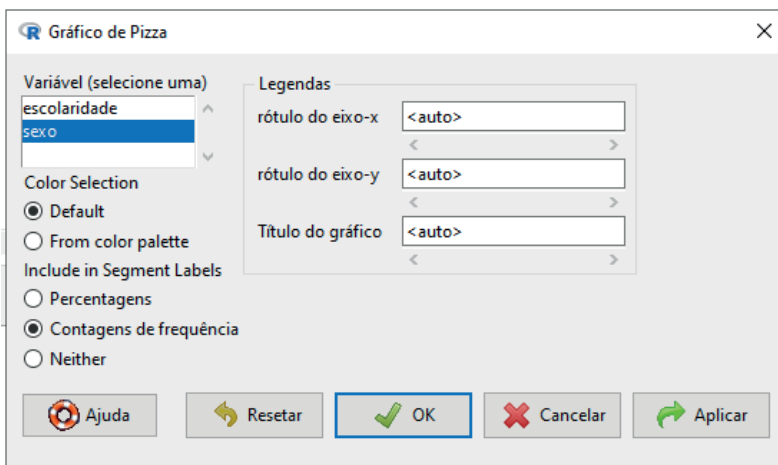


Figura 25. Selecionando a variável resposta para frequência absoluta.

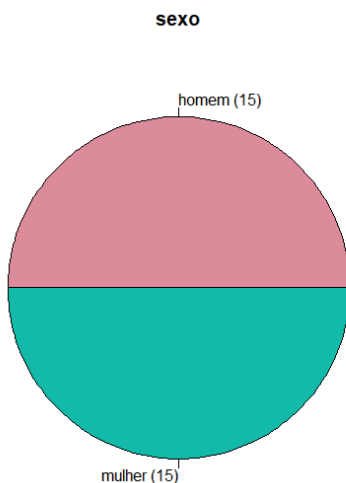


Figura 26. Gráfico de Pizza com a variável resposta expressada em frequência absoluta.

Linha de comando gerada para o gráfico de Pizza por tipo de sexo na interface do "Rcmdr" e resposta expressada em frequência absoluta

```
>with(Dataset, piechart(sexo, xlab="", ylab="", main="sexo", col=rainbow_hcl(2),
scale="frequency"))
```



CAPÍTULO 26

TESTE T DE STUDENT E TEST U DE MANN-WHITNEY NO “RCMDR”

Teste T de Student: comparação da idade média x sexo

É importante citar que os resultados dos testes feitos na interface do pacote “Rcmdr” sairão no R console do RStudio ou do R raiz de onde foi instalado e carregado inicialmente. Antes do teste iremos separar com o auxílio dos colchetes segundo Britto *et al.* (2021) as amostras de idade de homem e de mulher do arquivo em Excel chamado de `estcivl_idade_peso_altura1.xlsx`. Utilizando-se o ícone “ver o conjunto de dados” do pacote “Rcmdr” iremos verificar os dados da tabela usada para a realização do teste T ou U (ver figura 27). Conforme a figura 27 verificamos que o nível da variável sexo que entra primeiro é “homem” e depois o da “mulher”.

	escolaridade	sexo	distSTM	idade	peso	altura	
1	ensinomedio	home	homem	5	21	50	147
2	ensinomedio	homem	5	24	56	148	
3	ensinomedio	homem	5	25	52	145	
4	ensinomedio	homem	5	63	89	147	
5	ensinomedio	homem	5	36	78	148	
6	graduado	homem	5	45	89	149	
7	graduado	homem	5	41	85	175	
8	graduado	homem	5	44	85	182	
9	graduado	homem	5	45	85	178	
10	graduado	homem	5	45	85	177	
11	ensinofundamental	homem	5	16	85	157	
12	ensinofundamental	homem	5	14	85	165	
13	ensinofundamental	homem	5	10	88	198	
14	ensinofundamental	homem	10	78	88	199	
15	ensinofundamental	homem	10	85	89	147	
16	ensinomedio	mulher	10	14	65	145	
17	ensinomedio	mulher	10	45	63	135	
18	ensinomedio	mulher	10	42	65	152	
19	ensinomedio	mulher	10	47	66	150	
20	ensinomedio	mulher	10	85	63	180	
21	graduado	mulher	10	65	63	158	
22	graduado	mulher	22	69	63	154	
23	graduado	mulher	22	85	63	158	
24	graduado	mulher	22	23	65	159	
25	graduado	mulher	22	33	66	164	
26	ensinofundamental	mulher	22	54	66	165	
27	ensinofundamental	mulher	22	54	63	166	
28	ensinofundamental	mulher	22	56	63	147	
29	ensinofundamental	mulher	22	26	54	145	
30	ensinofundamental	mulher	22	56	58	142	

Figura 27. Visualização dos dados para separação da amostra de idade de homem e de idade de mulher para fazer o teste T de Student no pacote “Rcmdr”

Separando as idades de homens e de mulheres pelo uso dos colchetes

Amostra de idade de homem, repetições de idade e “n” amostral

Para selecionarmos a amostra de idade de homem criamos o objeto `ih` que contém o vetor da idade de homem que se inicia em 21 e termina em 85 (ver figura 27). Para termos um vetor ou conjunto numérico usamos a linha de comando: `nome do vetor<- nome do vetor$variável resposta[nome do vetor$variável independente=="nível da variável independente"]`.

```
> ih<-i$idade[i$sexo=="homem"]
> ih
[1] 21 24 25 63 36 45 41 44 45 45 16 14 10 78 85
> length(ih)
[1] 15
```

Amostra de idade de mulher, repetições de idade e “n” amostral

Para selecionarmos a amostra de idade mulher criamos o objeto `im` que contém o vetor idade do nível mulher que se inicia em 14 e termina em 56 (ver figura 27).

```
> im<-i$idade[i$sexo=="mulher"]
> im
[1] 14 45 42 47 85 65 69 85 23 33 54 54 56 26 56
> length(im)
[1] 15
```

Média de idade de homem e de mulher com a função `mean(nome do objeto)`

```
> mean(ih)
[1] 39.46667
> mean(im)
[1] 50.26667
```

Caminho para fazer teste T de Student no Rcommander

Para aplicar o teste T de Student no Rcommander assumimos que as idades em cada nível do fator sexo tenham distribuição normal e variâncias homogêneas. O caminho para fazer o teste T é: dados - importar arquivo de dados - do arquivo Excel - abre-se a pasta no computador, seleciona-se o arquivo que foi `estcivl_idade_peso_altura1.xlsx` - dá-se um nome a este arquivo - estatísticas - médias - teste t para amostras independentes - seleciona-se a variável independente, o sexo, e depois a variável dependente (ver figura 28). Antes de realizarmos o teste clicamos em "opções" e selecionamos o teste para variâncias iguais (ver figura 29).

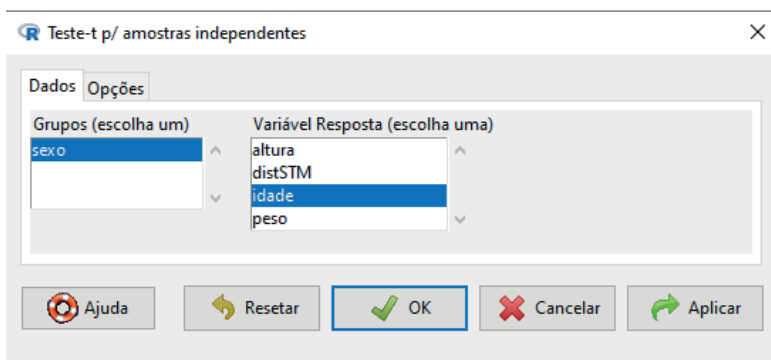


Figura 28. Seleção da variável independente, sexo e variável resposta idade

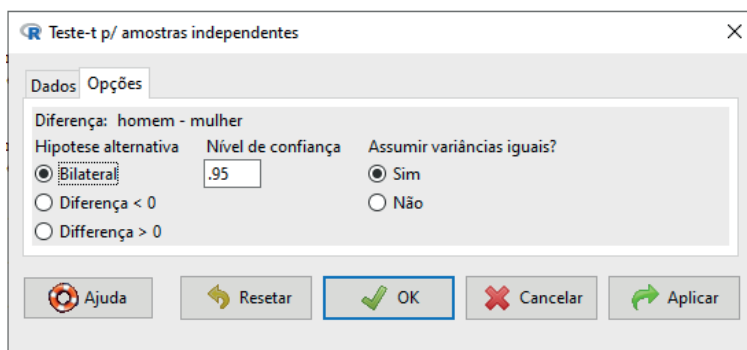


Figura 29. Seleção da opção assumir variâncias iguais, clicando em Sim.

Obs: Como já calculamos as idades médias devemos observar se as médias são referentes as idades de homens e de mulheres após o teste T de Student no pacote "Rcmdr".

Resultado do teste T de Student no pacote “Rcmdr”

```
Rcmdr> t.test(idade~sexo, alternative='two.sided', conf.level=.95, var.equal=TRUE,
```

```
Rcmdr+ data=Dataset)
```

Two Sample t-test

data: idade by sexo

t = -1.361, df = 28, p-value = 0.1844

alternative hypothesis: true difference in means between group homem and group mulher is not equal to 0

95 percent confidence interval:

-27.054221 5.454221

sample estimates:

mean in group homem mean in group mulher

39.46667 50.26667

Previamente a aplicação do teste T de Student no pacote “Rcmdr” foi selecionado trabalhar-se com o nível de significância ou $\alpha=5\%$, e com nível de confiança de 0.95 para ter-se resultado com o intervalo de confiança de 95%.

Linha de comando do resultado do teste T de Student no pacote “Rcmdr”

```
t.test(idade~sexo, alternative='two.sided', conf.level=.95, var.equal=TRUE,
```

```
data=Dataset)
```

Observamos conforme o resultado do teste T de Student pelo pacote “Rcmdr” que não houve diferença estatística significativa na idade média de homem e de mulher pelo p-valor: 0.1844 para o $\alpha=5\%$.

Extraindo a média da idade da amostra de homem e de mulher pela função tapply()

Uma função interessante para confirmar os resultados com a função mean() no RStudio ou R raiz é usar função tapply(). Enquanto que com a função mean() obtém-se a média de uma amostra, com a função tapply(objeto\$variável resposta, objeto\$variável independente, função mean) obtemos a média de mais de uma amostra por vez.

```
> tapply(i$idade,i$sexo,mean)

homem  mulher
39.46667 50.26667
```

Extraindo a média de idade da amostra de homem e de mulher pela função describeBY do pacote “psych”

Depois de instalar e carregar o pacote “psych” segundo Revelle (2020) no R Studio ou R raíz calculamos a estatística descritiva para o grupo homem e para o grupo mulher. Usaremos a linha de comando describeBY(i\$idade,i\$sexo).

Descriptive statistics by group

group: homem

vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	15	39.47	22.51	41	38.23	25.2	10	85	75	0.57	-0.79	5.81

group: mulher

vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	15	50.27	20.92	54		50.38	17.79	14	85	71	0.04	-0.98

A razão de calcular média da idade pelas funções mean(), tapply() e describeBy() foi para verificar a resposta da média depois do teste T de Student feito no pacote “Rcmdr” onde consta no final, primeiro o resultado da amostra de idade de homem e depois de idade de mulher. Estes resultados tem a ver com a ordem dos níveis que constam na tabela original dos dados em Excel de onde são extraídas as variáveis para o cálculo do teste T de Student e que fica salva na pasta criada pelo usuário no seu notebook ou computador pessoal.



CAPÍTULO 27

CALCULANDO-SE AS PREMISSAS PARA FAZER O TESTE T DE STUDENT

Caminho para fazer o teste de Normalidade no Rcommander

Este caminho serve para aplicar o teste de normalidade no Rcommander considerando todas as repetições da variável resposta de ambos os grupos que se deseja comparar. O caminho para fazer o teste de normalidade é: Estatísticas – Resumo – Test of normality...- selecione a variável, no caso a resposta é a idade – Normality test – Shapiro-Wilk – ok

Linha de comando gerada para o teste de normalidade na interface do "Rcmdr"

```
normalityTest(~idade, test="shapiro.test", data=Dataset)
```



C A P Í T U L O 28

HIPÓTESES DO TESTE DE NORMALIDADE

Teste de Hipótese segundo Crawley (2013)

Ho: Há distribuição normal para a idade

H1: não há distribuição normal para a idade

Resultado do teste de normalidade no console do “Rcmdr”

```
Rcmdr> normalityTest(~idade, test="shapiro.test", data=Dataset)
```

Shapiro-Wilk normality test

data: idade

W = 0.95146, p-value = 0.185

O teste indica que há normalidade para a variável resposta idade para o valor do alfa escolhido, mesmo sendo o $\alpha=1\%$ ou $\alpha=5\%$ porque o p-valor do teste é maior que o α escolhido. Lembrando que a escolha do $\alpha=5\%$ deve ser feita antes de se fazer o teste estatístico.

Caminho para se fazer o teste da Normalidade por grupos

Este caminho serve para aplicar o teste de normalidade no Rcommander da variável resposta idade para cada amostra de idade por sexo. O caminho para fazer o teste de normalidade é: Estatísticas – Resumo – Test of normality...- selecione a variável, no caso a resposta – Normality test – Shapiro-Wilk. Depois selecione o ícone test by groups... e em sexo e clica-se em ok e tem-se o resultado da normalidade por grupos (ver figuras 30 e 31).

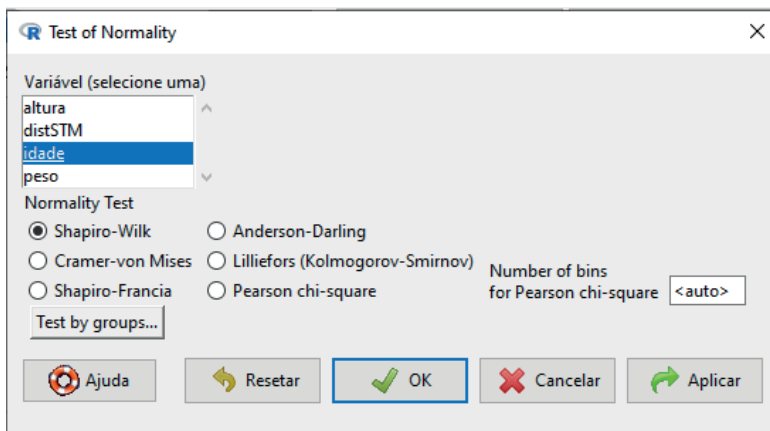


Figura 30. Seleção da idade para o teste de normalidade.

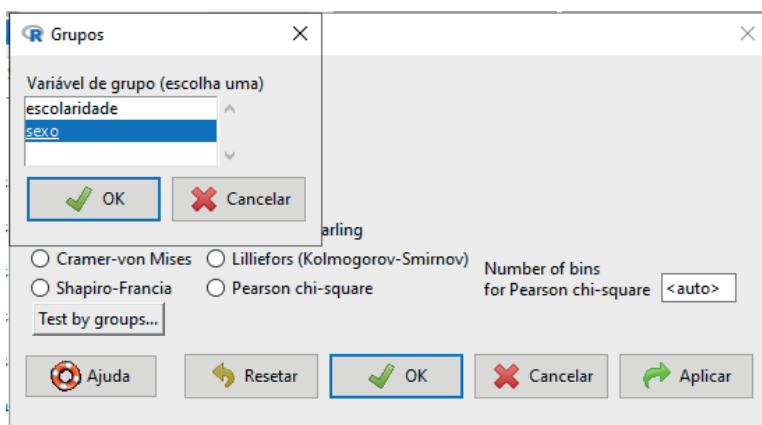


Figura 31. Seleção do grupo sexo para o teste de normalidade por grupos.

Resultado do teste de Normalidade por grupos no pacote "Rcmdr"

```
Rcmdr> normalityTest(idade ~ sexo, test="shapiro.test", data=Dataset)
```

```
-----
```

```
sexo = homem
```

```
Shapiro-Wilk normality test
```

```
data: idade
```

W = 0.92391, p-value = 0.2209

sexo = mulher

Shapiro-Wilk normality test

data: idade

W = 0.96688, p-value = 0.8095

Pelo resultado dos p-valor da amostra de idade de homem de $p=0.2209$ e idade de mulher de $p=0.8095$ verificamos que ambas as amostras de idade têm distribuição normal e podemos aplicar o teste T de Student



CAPÍTULO 29

CAMINHO PARA FAZER O TESTE DE HOMOGENEIDADE DAS VARIÂNCIAS NO “RCMDR”

Para aplicar o teste da homogeneidade das variâncias no Rcommander utilizamos a idade separada por grupos. O caminho para se fazer o teste da homogeneidade das variâncias é: Estatísticas – Resumo – Variâncias- Teste F p/ duas variâncias – selecionar a variável independente em Grupos, no caso a variável sexo na caixa da esquerda – e a variável resposta na caixa da direita, idade e clicamos em ok ou aplicar (ver figura 32 e 33). O default do R no teste de variâncias é fazer o teste bilateral.

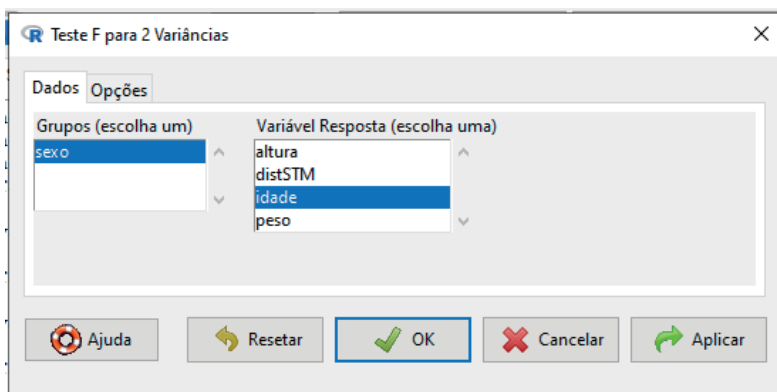


Figura 32. Teste F para duas variâncias na interface do pacote “Rcmdr”

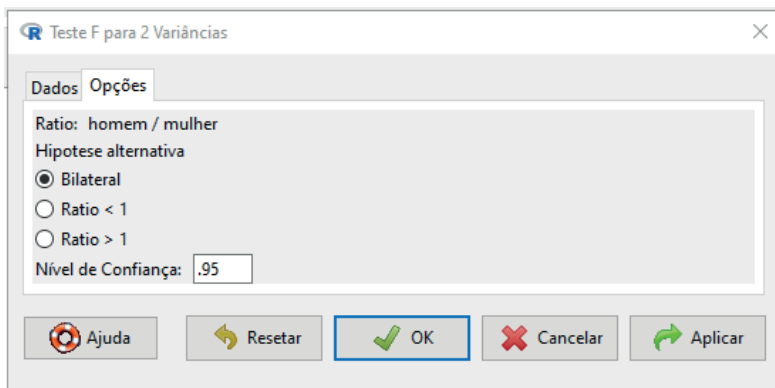


Figura 33. Teste F para duas variâncias mostrando o default para teste bilateral no “Rcmdr”.

Linha de comando gerada para o teste da homogeneidade de variâncias por sexo na interface do “Rcmdr”

```
Tapply(idade ~ sexo, var, na.action=na.omit, data=Dataset)
```

```
# variances by group
```

```
var.test(idade ~ sexo, alternative='two.sided', conf.level=.95,
```

```
data=Dataset)
```



C A P Í T U L O 30

HIPÓTESES DO TESTE DE HOMOGENEIDADE DAS VARIÂNCIAS

Teste de Hipótese segundo Crawley (2013)

Hipóteses das variâncias:

Ho: as variâncias são homogêneas ou não há diferença entre as variâncias da idade de homens e mulheres

H1: as variâncias são heterogêneas ou há diferença entre as variâncias da idade de homens e mulheres

Resultado do teste da homogeneidade de variâncias no console do “Rcmdr”

Pelo resultado do p-valor do teste de homogeneidade, as variâncias são homogêneas porque o p-valor do teste é maior que o $\alpha=5\%$.

```
Rcmdr> Tapply(idade ~ sexo, var, na.action=na.omit, data=Dataset)
```

```
Rcmdr+ # variances by group
```

```
homem  mulher
```

```
506.8381 437.6381
```

```
Rcmdr> var.test(idade ~ sexo, alternative='two.sided', conf.level=.95,
```

```
Rcmdr+ data=Dataset)
```

```
F test to compare two variances
```

```
data: idade by sexo
```

```
F = 1.1581, num df = 14, denom df = 14, p-value = 0.7874
```

```
alternative hypothesis: true ratio of variances is not equal to 1
```

95 percent confidence interval:

0.3888157 3.4495663

sample estimates:

ratio of variances

1.158122



CAPÍTULO 31

TESTE U DE MANN-WHITNEY NO “RCMDR”

Assumindo não existir ou normalidade ou homogeneidade das variâncias ou os dois, e mesmo se depois de transformar a variável resposta em cada nível do fator sexo estas condições não forem satisfeitas, fazemos então o teste U de Mann-Whitney.

Caminho para fazer o teste U de Mann-Whitney no “Rcmdr”

O caminho para se fazer o teste U de Mann-Whitney no R Commander é: Dados - selecionar arquivo de dados - arquivo em Excel - abre-se a pasta - seleciona-se e nomeia-se o arquivo - Estatísticas - Testes Não-paramétricos - Teste de Wilcoxon (2 amostras) – seleciona-se a variável independente em Grupos, o sexo – a variável resposta, a idade – ok. O default do teste de Wilcoxon para 2 amostras é fazer um teste bilateral (ver figuras 34 e 35). No R Commander o teste U de Mann-Whitney é feito como teste de Wilcoxon para duas amostras e seu resultado é expressado como Wilcoxon rank sum test no R raíz ou R Studio.

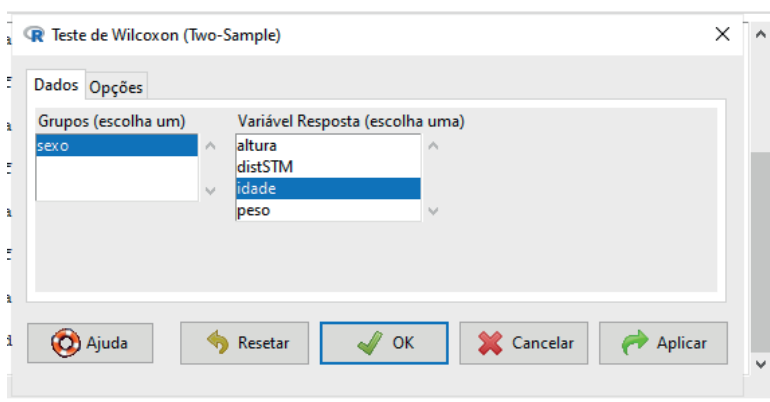


Figura 34. Teste de Wilcoxon da idade por sexo

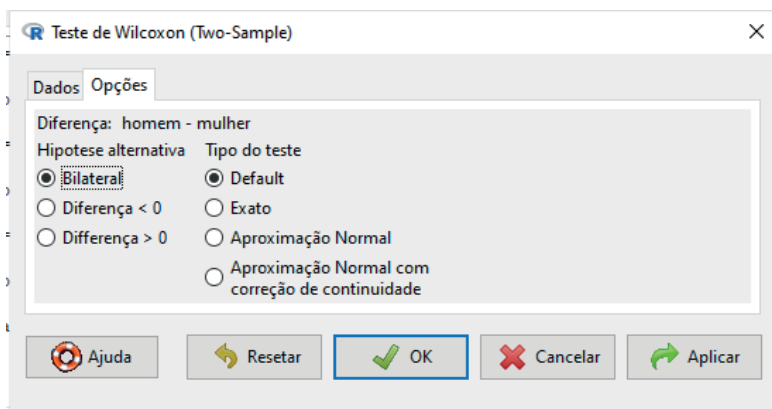


Figura 35. Teste de Wilcoxon para duas amostras da idade por sexo com default de teste bilateral



C A P Í T U L O 32

HIPÓTESES DO TESTE U DE MANN-WHITNEY

Ho: a mediana dos postos de idade de homens e mulheres não difere

H1: a mediana dos postos de idade de homens e mulheres difere

Linha de comando gerada para o teste U de Mann-Whitney idade por sexo na interface do "Rcmdr"

```
Tapply(idade ~ sexo, median, na.action=na.omit, data=Dataset)
```

```
# medians by group
```

```
wilcox.test(idade ~ sexo, alternative="two.sided", data=Dataset)
```

Resultado do teste U de Mann-Whitney no console do "Rcmdr"

```
Rcmdr> Tapply(idade ~ sexo, median, na.action=na.omit, data=Dataset)
```

```
Rcmdr+ # medians by group
```

```
homem mulher
```

```
41 54
```

```
Rcmdr> wilcox.test(idade ~ sexo, alternative="two.sided", data=Dataset)
```

```
Wilcoxon rank sum test with continuity correction
```

```
data: idade by sexo
```

```
W = 74, p-value = 0.1143
```

No Rcommander o teste U de Mann-Whitney é gerado com correção de continuidade.

Conforme o p-valor: 0.1143 do teste U de Mann-Whitney, a mediana dos postos de idade não difere para o α escolhido.

FAZENDO O TESTE DE ANOVA ONE WAY NO “RCMDR”

Caminho para fazer teste de ANOVA one way no Rcommander

Depois de carregar o arquivo em Excel e nomeá-lo - abre-se Estatísticas - Médias – ANOVA para um fator (one way)...- seleciona-se a variável independente em Grupos, a religião e a variável resposta, o peso e clica-se em ok. Para fazermos o teste de ANOVA 1 fator assumimos que os dados de peso entre os níveis de religião tinham distribuição normal e variâncias homogêneas (ver figura 36).

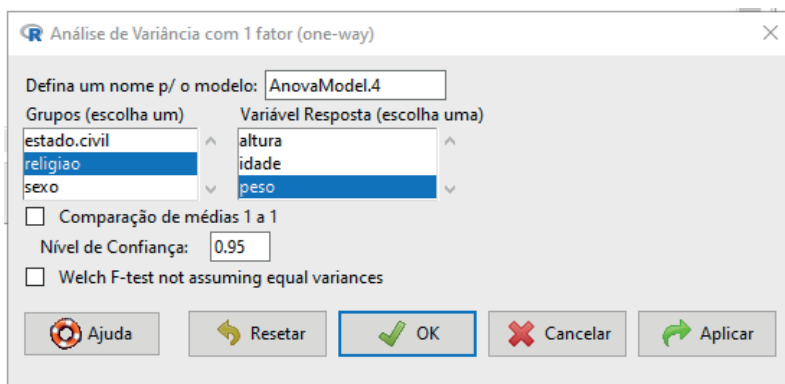


Figura 36. Teste de ANOVA one way no “Rcmdr”.

Linha de comando gerada para o teste de ANOVA one way no “Rcmdr” com a saída mostrando a média e o desvio padrão

```
AnovaModel.4 <- aov(peso ~ religiao, data=Dataset)
summary(AnovaModel.4)
with(Dataset, numSummary(peso, groups=religiao, statistics=c("mean", "sd")))
```

O default do R Commander irá gerar resultados com média e desvio padrão, sendo “mean” e “sd” respectivamente.

Linha de comando gerada para o teste de ANOVA one way no "Rcmdr" com a saída mostrando a média e o erro padrão.

Se quisermos o resultado do teste de ANOVA one way e a comparação de média e erro padrão devemos substituir no rscript do "Rcmdr" o resultado da ANOVA one way de "sd" para "se" e clicar em submeter. O resultado poderá ser visto no R console do RStudio ou Rraíz.

De

```
AnovaModel.4 <- aov(peso ~ religiao, data=Dataset)
summary(AnovaModel.4)
with(Dataset, numSummary(peso, groups=religiao, statistics=c("mean", "sd")))
```

Para

```
AnovaModel.4 <- aov(peso ~ religiao, data=Dataset)
summary(AnovaModel.4)
with(Dataset, numSummary(peso, groups=religiao, statistics=c("mean", "se")))
Rcmdr> AnovaModel.4 <- aov(peso ~ religiao, data=Dataset)
```

Ficando:

```
Rcmdr> summary(AnovaModel.4)
      Df Sum Sq Mean Sq F value Pr(>F)
religiao  2  1289   644.5   4.832 0.0161 *
Residuals 27  3601   133.4
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
Rcmdr> with(Dataset, numSummary(peso, groups=religiao, statistics=c("mean",
"se")))
```

mean se(mean) data:n

```
catolico  77.84615 4.066508  13
evangelico 70.25000 4.003347   8
protestante 62.33333 1.312335   9
```

O resultado do erro padrão mostrado foi feito pela fórmula: $ep = sd / \text{raiz quadrada } (n)$ do grupo que se quer, por exemplo, para a amostra de peso de católicos com $n=13$. O resultado do desvio padrão foi 14.662004 primeiramente gerado pelo default

quando se calcula a ANOVA one way no "Rcmdr". Com a mudança de "sd" para "se" o erro padrão ficou $ep=4.066508$, ou seja, foi aplicada a fórmula: $14.662004/\text{raiz quadrada}(13)$. No entanto, para mostrar este resultado na tabela comparativa de média da Análise de variância e gerar as letras para diferenciar as amostras de peso quanto a religião e trabalhando-se com um $\alpha=5\%$ ou com 2 desvios padrões ou $z=1.96$, o erro padrão seria $ep=7.9704$, fruto da multiplicação do $ep=4.066508 \times 1.96$.



C A P Í T U L O 34

HIPÓTESES DO TESTE DE ANOVA ONE WAY

Teste de Hipótese segundo Zar (1999)

$H_0: \text{pesomédio1} = \text{pesomédio2} = \text{pesomédio3}$

$H_1: \text{pesomédio1} \neq \text{pesomédio2} \neq \text{pesomédio3}$

Pelo teste de ANOVA one way e o p-valor de 0.0161, há diferença em pelo menos um peso médio de uma religião segundo o $\alpha=5\%$ escolhido.

FAZENDO O TESTE DE TUKEY NO “RCMDR”

Para fazer o teste de Tukey no “Rcmdr” depois de percorrer o caminho para fazer ANOVA one way, marca-se a caixa “Comparação de média 1 a 1” como na figura 37.

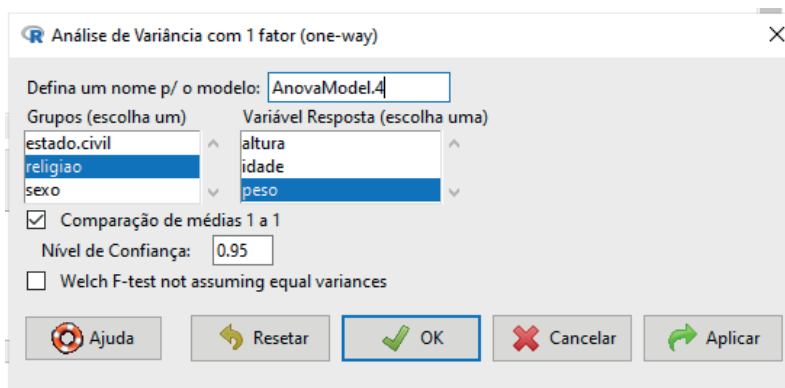


Figura 37. Caminho para fazer o teste de Tukey no “Rcmdr” marcando a caixa Comparação de médias 1 a 1.

Segundo a figura 37 foram selecionadas as variáveis, a independente em Grupos, selecionando-se a religião e a variável resposta, o peso. Com estes comandos tem-se os resultados da ANOVA one way e clicando-se na caixa “Comparação das médias 1 a 1, o comando para ser feito o teste de Tukey.

Linha de comando gerada para o teste de Tukey no “Rcmdr”

```
AnovaModel.4 <- aov(peso ~ religiao, data=Dataset)
summary>AnovaModel.4)
with(Dataset, numSummary(peso, groups=religiao, statistics=c("mean", "sd")))
local{
```



```
.Pairs <- glht(AnovaModel.4, linfct = mcp(religiao = "Tukey"))
print(summary(.Pairs)) # pairwise tests
print(confint(.Pairs, level=0.95)) # confidence intervals
print(cld(.Pairs, level=0.05)) # compact letter display
old.oma <- par(oma=c(0, 5, 0, 0))
plot(confint(.Pairs))
par(old.oma)
})
```

Resultado do teste de Tukey no console do Rstudio

```
Rcmdr> AnovaModel.4 <- aov(peso ~ religiao, data=Dataset)
```

```
Rcmdr> summary(AnovaModel.4)
```

```
      Df Sum Sq Mean Sq F value Pr(>F)
religiao  2  1289   644.5   4.832 0.0161 *
Residuals 27  3601   133.4
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Rcmdr> with(Dataset, numSummary(peso, groups=religiao, statistics=c("mean",
"sd")))
```

```
      mean    sd data:n
catolico  77.84615 14.662004   13
evangelico 70.25000 11.323175    8
protestante 62.33333  3.937004    9
```

```
Rcmdr> local({
```

```
Rcmdr+ .Pairs <- glht(AnovaModel.4, linfct = mcp(religiao = "Tukey"))
```

```
Rcmdr+ print(summary(.Pairs)) # pairwise tests
```

```
Rcmdr+ print(confint(.Pairs, level=0.95)) # confidence intervals
```

```
Rcmdr+ print(cld(.Pairs, level=0.05)) # compact letter display
Rcmdr+ old.oma <- par(oma=c(0, 5, 0, 0))
Rcmdr+ plot(confint(.Pairs))
Rcmdr+ par(old.oma)
Rcmdr+ })
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = peso ~ religiao, data = Dataset)

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t)
evangelico - catolico == 0	-7.596	5.190	-1.464	0.3231
protestante - catolico == 0	-15.513	5.008	-3.098	0.0121 *
protestante - evangelico == 0	-7.917	5.612	-1.411	0.3488

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Adjusted p values reported -- single-step method)

Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = peso ~ religiao, data = Dataset)

Quantile = 2.4768

95% family-wise confidence level

Linear Hypotheses:

	Estimate	lwr	upr
evangelico - catolico == 0	-7.5962	-20.4496	5.2573

protestante - catolico == 0 -15.5128 -27.9163 -3.1093

protestante - evangelico == 0 -7.9167 -21.8157 5.9824

catolico evangelico protestante

"a" "ab" "b"

Paralelamente quando selecionamos o caminho para o teste de ANOVA one way e marcamos o caminho para fazer o teste de Tukey é gerado um gráfico comparativo das diferenças do peso médio entre os níveis de religião conforme a figura 38.

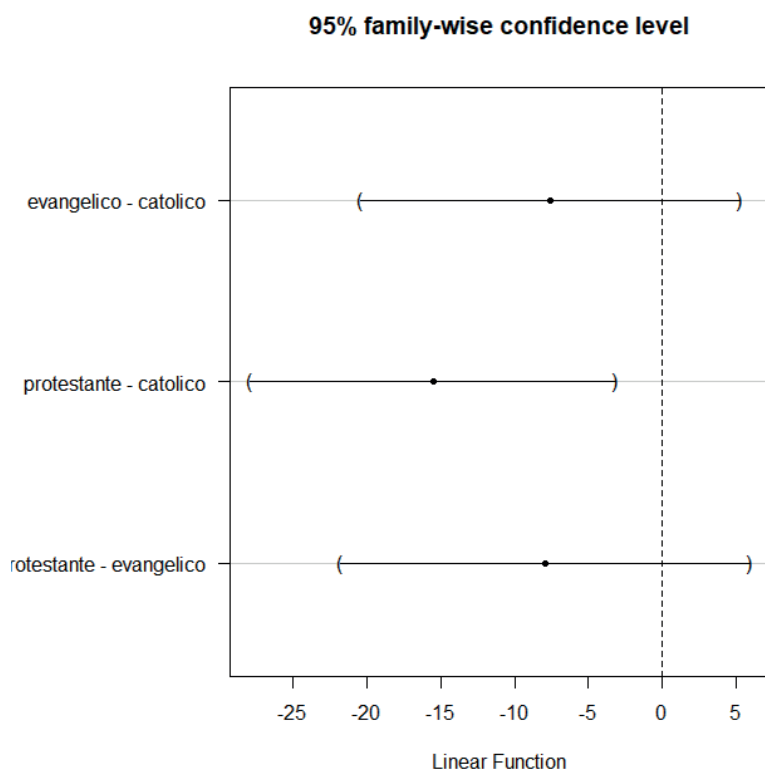


Figura 38. Gráfico comparativo do teste de Tukey para a diferença entre os pesos médios dos níveis do fator religião.

Conforme a figura 38 os grupos evangélico-católico e protestante-evangélico coincidem com a linha vertical onde as diferenças nas médias de pesos são iguais a zero ou não existem diferença estatística significativa para estes grupos para o intervalo de confiança de 95%. Os grupos protestante-católico apresentam diferenças estatísticas significativa entre os pesos médios.

Os resultados da figura 38 podem ser confirmados pelos resultados do p-valor gerados no R console do Rstudio, onde os comparativos entre os pesos médios de evangélico-católico têm o p-valor de 0.3231; os pesos médios entre protestante-católico têm o p-valor de 0.0121; e entre protestante-evangélico o p-valor de 0.3488. Resultado com este tipo de gráfico pode ser visualizado e confirmado no trabalho de Cravo *et al.* (2019) onde eles compararam diferenças na pressão intraocular média de grupos de pacientes com HIV em um estudo de caso (ver nas referências).

No final dos resultados do teste de Tukey no Rconsole do Rstudio são mostradas as letras, onde letras iguais não indicam diferença estatística significativa ao $\alpha=5\%$, e letras diferentes indicam diferença estatística significativa ao $\alpha=5\%$ conforme os resultados abaixo:

catolico evangelico protestante

"a" "ab" "b"

Comparando-se católico-evangélico verificamos que ambos têm letra "a" indicando não existir diferença significativa; e católico com letra "a" e protestante com letra "b" indicam que há diferença estatística significativa conforme o $\alpha=5\%$. E evangélico-protestante ambos possuindo a letra "b", estes resultados indicam não haver diferença estatística significativa.

PREMISSAS PARA FAZER O TESTE DE ANOVA ONE WAY NO “RCMDR”

Caminho para fazer o teste de Normalidade no “Rcmdr” para a variável resposta peso independente dos níveis de religião

O caminho para fazer o teste de normalidade é: Estatísticas – Resumo – Test of normality...- selecione a variável, no caso a resposta – Normality test – seleciona-se a variável depende, o peso e clica-se ok! Obtem-se o resultado da normalidade para os valores da variável peso independente do nível. (ver figura 39)

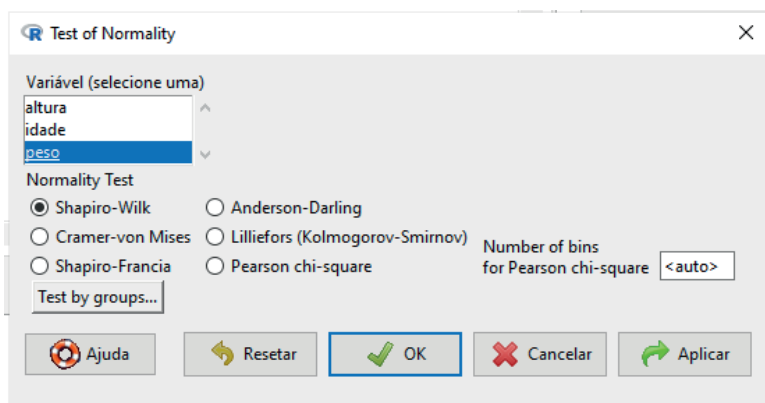


Figura 39. Teste de Normalidade Shapiro-Wilk da variável peso independente dos níveis.

Linha de comando gerada para o teste de normalidade no “Rcmdr” da variável peso

```
normalityTest(~peso, test="shapiro.test", data=Dataset)
```



C A P Í T U L O 37

HIPÓTESES DO TESTE DE NORMALIDADE

Teste de Hipótese segundo Crawley (2013)

Ho: os pesos têm distribuição normal

H1: os pesos não têm distribuição normal

Resultado do teste de Normalidade no Rstudio ou R raíz da variável peso

```
Rcmdr> normalityTest(~peso, test="shapiro.test", data=Dataset)
```

Shapiro-Wilk normality test

data: peso

W = 0.8593, p-value = 0.0009769

Pelo resultado do p-valor não há distribuição normal para a variável peso.

Caminho para fazer o teste de Normalidade no "Rcmdr" para a variável resposta peso considerando os níveis de religião

Para obter a normalidade por grupos clicando-se na caixa Test by groups abre-se Grupos e seleciona-se religião e oK (ver figura 40) verificando-se no RStudio ou Rraíz o resultado do teste da normalidade do peso por níveis de religião.

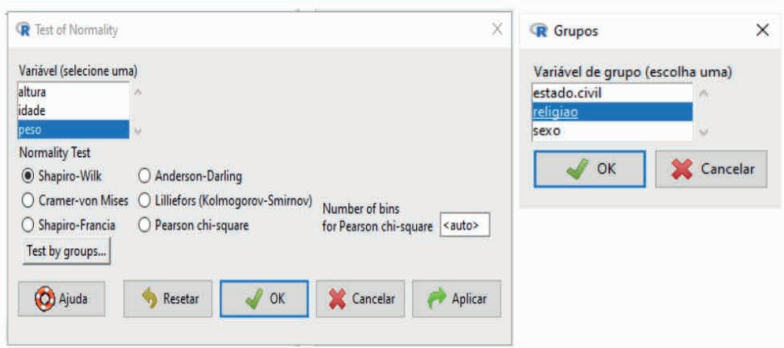


Figura 40. Clicando-se na caixa Test by groups para selecionar a variável independente religiao.

Linha de comando gerada para o testes de normalidade no “Rcmdr” por grupos

normalityTest(peso ~ religiao, test="shapiro.test", data=Dataset)



C A P Í T U L O 38

HIPÓTESES DO TESTE DE NORMALIDADE

Teste de Hipótese segundo Crawley (2013)

Ho: os pesos para os níveis de religião têm distribuição normal

H1: os pesos para os níveis de religião não têm distribuição normal

Resultado do teste de normalidade no console do Rstudio

```
normalityTest(peso ~ religiao, test="shapiro.test", data=Dataset)
```

```
-----
```

```
religiao = catolico
```

```
Shapiro-Wilk normality test
```

```
data: peso
```

```
W = 0.68732, p-value = 0.0004156
```

```
-----
```

```
religiao = evangelico
```

```
Shapiro-Wilk normality test
```

```
data: peso
```

```
W = 0.65061, p-value = 0.000615
```

```
-----
```

```
religiao = protestante
```

```
Shapiro-Wilk normality test
```

```
data: peso
```

```
W = 0.81528, p-value = 0.03047
```

```
-----
```

São gerados os testes de normalidade dos pesos médios por nível de religião, e pelos p-valores não há distribuição normal, pois, valores menores que o $\alpha=5\%$ selecionado.

FAZENDO TESTE DE HOMOGENEIDADE DE VARIÂNCIAS NO “RCMDR”

Caminho para fazer o teste de homogeneidade de variâncias no “Rcmdr”

O caminho para fazer o teste de homogeneidade de variâncias é: Estatísticas – Variâncias – Teste de Bartlett...- selecione a variável independente em Fatores, no caso religião e a variável resposta, o peso e ok (ver figura 41).

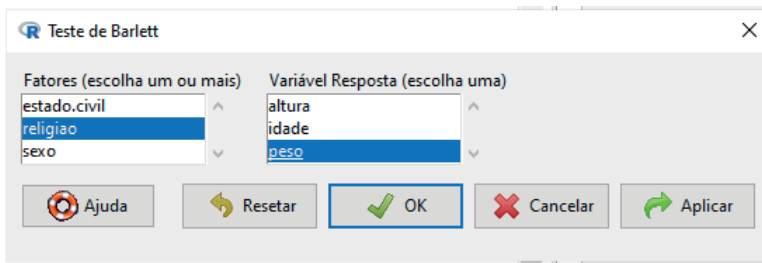


Figura 41. Teste de homogeneidade de variâncias do peso por níveis de religião.

Linha de comando gerada para o teste de homogeneidade de variâncias de Bartlett

```
Tapply(peso ~ religiao, var, na.action=na.omit, data=Dataset)
```

```
# variances by group
```

```
bartlett.test(peso ~ religiao, data=Dataset)
```



C A P Í T U L O 4 0

PREMISSAS DO TESTE DE HOMOGENEIDADE DE VARIÂNCIAS

Teste de Hipótese segundo Crawley (2013)

Ho: as variâncias dos pesoscatolicos=pesosevangelicos=pesoprotestantes

H1: as variâncias dos pesoscatolicos≠pesosevangelicos≠pesoprotestantes

Resultado do teste de homogeneidade de variâncias no console do Rstudio

```
Rcmdr> Tapply(peso ~ religiao, var, na.action=na.omit, data=Dataset)
```

```
Rcmdr+ # variances by group
```

```
catolico evangelico protestante
```

```
214.9744 128.2143 15.5000
```

```
Rcmdr> bartlett.test(peso ~ religiao, data=Dataset)
```

```
Bartlett test of homogeneity of variances
```

```
data: peso by religiao
```

```
Bartlett's K-squared = 11.182, df = 2, p-value = 0.003732
```

Pela resposta do teste de homogeneidade de variâncias dos pesos por níveis de religião, as variâncias não são homogêneas segundo o p-valor 0.003732 ser menor que o p-valor do $\alpha=5\%$ selecionado.



CAPÍTULO 41

FAZENDO TESTE DE CORRELAÇÃO NO “RCMDR”

Vamos fazer a correlação com as variáveis peso x altura obtida de um arquivo em Excel e assumimos que as variáveis altura e peso tem distribuição normal e variâncias homogêneas.

Caminho para fazer o teste de correlação no “Rcmdr”

Depois de carregar o arquivo em Excel e nomeá-lo – clicamos em: Estatísticas - Resumos - Teste de correlação - selecionar variáveis: a seleção das variáveis no “Rcmdr” ocorre em ordem alfabética de cima para baixo na caixa de seleção das variáveis na correlação, e a primeira variável é a variável independente, sendo a altura e a segunda variável é a variável dependente, o peso. Depois clica-se em ok! (ver figura 42) o resultado é visto no RStudio ou Rraíz.

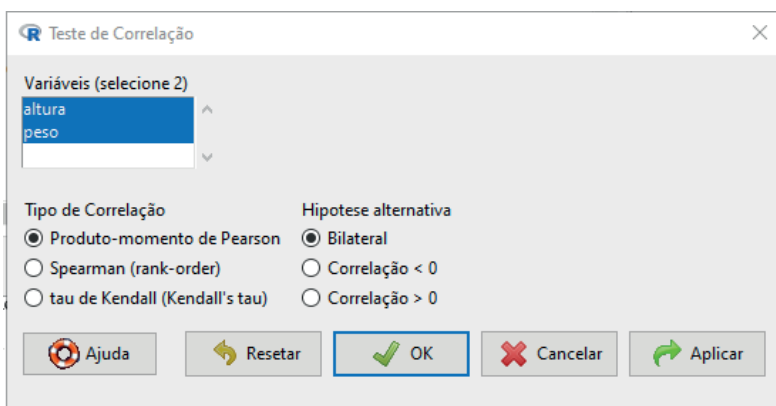


Figura 42. Teste de correlação no “Rcmdr”

Linha de comando gerada para o teste correlação no "Rcmdr"

```
with(Dataset, cor.test(altura, peso, alternative="two.sided",  
method="pearson"))
```

Gráfico de dispersão no "Rcmdr" para avaliar a correlação entre peso x altura

Caminho para fazer um gráfico de dispersão no "Rcmdr"

Dados - Importar arquivo de dados - Importar o conjunto em Excel - Defina um nome do conjunto de dados - nomeie o arquivo em Excel - Acesse o diretório onde está o arquivo alturapeso.xls - selecione o arquivo em Excel alturapeso.xls - Gráficos - Diagrama de dispersão... - Dados - nome da variável x, altura - nome da variável y, peso - ok! (ver figura 43).

Resultado do gráfico de dispersão do "Rcmdr"

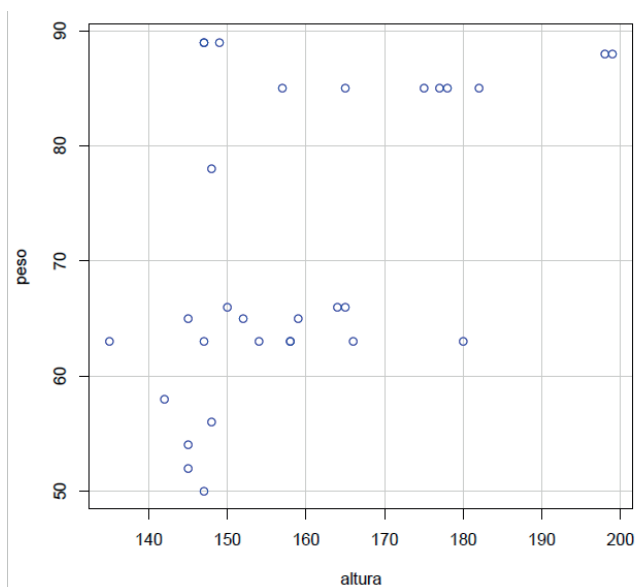


Figura 43. Gráfico de dispersão de pontos da variável resposta peso x a variável independente altura.

Resultado do teste de correlação no console do Rstudio

```
Rcmdr> with(Dataset, cor.test(altura, peso, alternative="two.sided",  
Rcmdr+ method="pearson"))
```

Pearson's product-moment correlation

data: altura and peso

$t = 3.1637$, $df = 28$, $p\text{-value} = 0.003732$

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.1875736 0.7371513

sample estimates:

cor

0.5131647

Pelo resultado do gráfico da figura 43 percebemos visualmente uma média dispersão dos valores de peso em função da altura, e isto refletiu no teste de correlação gerado no Rconsole do Rstudio por apresentar um valor do $r=0.5113$ que segundo Levin (1985) indica uma correlação variando entre moderada a forte. Verificamos que o modelo é significativo, porém com este coeficiente de correlação paramos por aqui, visto que o coeficiente de explicação dado pelo r^2 é de 26,14%, mostrando que a associação do peso e altura é explicada por 26,14% destes dados apenas, e 73,85% estaria relacionado a outros fatores desconhecidos. Desta forma levaria a equação de regressão com um r^2 baixo e inviável para serem feitas estimações de valores de peso em função de valores de altura não coletados.



C A P Í T U L O 42

ESCOLHA DO TIPO DE SOMATÓRIO TIPO I, II E III PARA CALCULAR ANOVA TWO WAY E THREE WAY NO R

Para fazer o teste de ANOVA two way assumiu-se que as amostras de peso em função do sexo e peso em função da religião tinham distribuição normal e variâncias homogêneas sendo os valores independentes. No exemplo a tabela tem como variável resposta o peso e as variáveis preditoras, a religião com três níveis (católico, evangélico e protestante), e o sexo, com dois níveis (homem e mulher). O α escolhido para a tomada de decisão foi $\alpha=5\%$. Neste exemplo assumiu-se que os dados tinham distribuição normal, variâncias homogêneas e que as repetições eram independentes. Estes testes foram feitos no Rstudio.



C A P Í T U L O 43

HIPÓTESES RELATIVAS AO FATOR RELIGIÃO

Teste de Hipótese segundo Zar (1999)

Ho: não há diferença entre o peso médio dentre os níveis católicos, evangélicos e protestantes

H1: há diferença entre o peso médio dentre os níveis católicos, evangélico e protestantes



C A P Í T U L O 44

HIPÓTESES RELATIVAS AO FATOR SEXO

Teste de Hipótese segundo Zar (1999)

Ho: não há diferença entre o peso médio dentre os níveis homem e mulher

H1: há diferença entre o peso médio dentre os níveis homem e mulher



C A P Í T U L O 4 5

HIPÓTESES RELATIVAS À INTERAÇÃO ENTRE OS FATORES RELIGIÃO E SEXO

Teste de Hipótese segundo Zar (1999)

Ho: não há interação entre o peso médio dentre os níveis de religião e sexo

H1: há interação entre o peso médio dentre os níveis de religião e sexo

Utilizou para este estudo o objeto `ecipaa<-read_excel("ecipaa18052025.xlsx")` que tinha como variável numérica o peso, e os fatores religião e sexo, que foram convertidos para factor segundo a linha de comando `ecipaa$religiao<-as.factor(ecipaa$religiao)` e `ecipaa$sexo<-as.factor(ecipaa$sexo)`. Posteriormente o objeto `ecipaa` foi convertido em `data.frame` sob o comando `ecipaa<-as.data.frame(ecipaa)` e armazenado no objeto `ee` sob o comando `ee<-ecipaa`.



CAPÍTULO 46

TESTE DE ANOVA TWO WAY COM O SOMATÓRIO DOS QUADRADOS DO TIPO I NO R

Separação das amostras de peso em função do sexo e de peso em função da religião

Para fazer a ANOVA two-way separou-se o peso por grupo em função dos fatores sexo e religião, nas denominações: `ee$ph<-ee$peso[ee$sexo=="homem"]`; e `ee$pm<-ee$peso[ee$sexo=="mulher"]`. Verificou-se o (n) amostral para cada amostra do peso x sexo que tinham 15 repetições cada. Para as amostras de peso x religião, nas denominações: `ee$pc<- ee$peso[ee$religiao=="catolico"]`; `ee$pe<-ee$peso[ee$religiao=="evangelico"]`; e `ee$pp<-ee$peso[ee$religiao=="protestante"]`. E tinham (n) amostral variando de n=13; n=8 e n=9 respectivamente, indicando um desenho desbalanceado.

Aplicando-se o teste de ANOVA two way como somatório do tipo I entrando primeiro com a variável preditora religião

Para saber o efeito das variáveis sobre o peso, fez-se o mesmo teste usando o somatório tipo I variando a ordem de entrada das variáveis preditoras. No primeiro teste entrou primeiro no modelo a variável religião e depois a variável sexo segundo o modelo nome do objeto `anova2<-aov(ee$peso ~ ee$religiao* ee$sexo,ee)`. O resultado foi armazenado na função `summary(nome do objeto do teste de anova)`, como: `summary(anova2)`. Utilizou-se neste teste o sinal asterisco entre as variáveis preditoras ou fatores que avalia o efeito principal de cada fator mais o efeito da interação entre os fatores, no caso entre religião e sexo.

```
> anova2<-aov(ee$peso ~ ee$religiao*ee$sexo,ee)
```

```
> summary(anova2)
```

```
Df Sum Sq Mean Sq F value Pr(>F)
```

```
ee$religiao 2 1289.0 644.5 6.176 0.00639 **
```

```
ee$sexo 1 888.2 888.2 8.512 0.00718 **
```

```
Residuals 26 2713.0 104.3
```

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Segundo este resultado não houve interação, apenas o efeito dos fatores principais, tendo-se ambos os fatores com efeitos significativos sobre a variável peso conforme o p-valor menor que 5%.

Aplicando-se o teste de ANOVA two way como somatório do tipo I entrando primeiro com a variável preditora sexo

Agora com a linha de comando: `anova3<-aov(ee$peso ~ ee$sexo*ee$religiao,ee)` visualizando-se o resultado com a função `summary(anova3)`.

```
> anova3<-aov(ee$peso ~ ee$sexo*ee$religiao,ee)
```

```
> summary(anova3)
```

```
      Df Sum Sq Mean Sq F value Pr(>F)
```

```
ee$sexo  1 1968.3  1968.3 18.863 0.00019 ***
```

```
ee$religiao 2  208.8   104.4  1.001 0.38134
```

```
Residuals 26 2713.0  104.3
```

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Percebe-se que a mudança da ordem de entrada da variável preditora no modelo de ANOVA two way influenciou no resultado final, sendo significativo apenas para o fator sexo. A explicação para isto é que segundo Field *et al.* (2012) quando se usa o somatório tipo I, a ordem de entrada das variáveis preditoras influencia na resposta final, isto é, variando segundo a ordem de entrada.

Aplicando-se o teste de ANOVA two way como somatório do tipo II entrando primeiro com a variável preditora religião

Fazendo-se o mesmo teste, porém agora com somatório dos quadrados do tipo II segundo a linha de comando: `anova4<-aov(ee$peso ~ ee$religiao+ee$sexo,ee)`. Como não havia interação fez-se o teste apenas comparando-se os efeitos principais, por isto utilizou-se o sinal de + entre as variáveis preditoras. O sinal de mais segundo Field *et al.*, (2012) indica o cálculo direcionado para os efeitos principais apenas. Os resultados foram visualizados pela função `Anova(nome do objeto do teste da ANOVA, type="II")` do pacote "car". O argumento `type="II"` indica os resultados expressados no somatório tipo II.

```
anova4<-aov(ee$peso ~ ee$religiao+ee$sexo,ee)
```

```
> Anova(anova4, type="II")
```

Anova Table (Type II tests)

Response: ee\$peso

	Sum Sq	Df	F value	Pr(>F)
--	--------	----	---------	--------

ee\$religiao	208.84	2	1.0007	0.381343
--------------	--------	---	--------	----------

ee\$sexo	888.17	1	8.5117	0.007184 **
----------	--------	---	--------	-------------

Residuals	2713.03	26		
-----------	---------	----	--	--

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Percebe-se que se usando o somatório do tipo II, apenas o sexo teve efeito significativo sobre o peso.

Aplicando-se o teste de ANOVA two way como somatório do tipo II entrando primeiro com a variável preditora sexo

Agora armazenou-se o resultado no modelo de ANOVA: `anova5<-aov(ee$peso ~ ee$sexo+ee$religiao,ee)` verificando-se o resultado em: `Anova(anova5,type="II")`

```
> Anova(anova5,type="II")
```

Anova Table (Type II tests)

Response: ee\$peso

	Sum Sq	Df	F value	Pr(>F)
--	--------	----	---------	--------

ee\$sexo	888.17	1	8.5117	0.007184 **
----------	--------	---	--------	-------------

ee\$religiao	208.84	2	1.0007	0.381343
--------------	--------	---	--------	----------

Residuals	2713.03	26		
-----------	---------	----	--	--

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Percebe-se que se invertendo a ordem de entrada da variável preditora, ainda o fator sexo é o que apresenta efeito significativo sobre o peso e o fator religião mantendo-se um efeito não significativo sobre o peso, tendo-se os mesmos resultados. Segundo Field et al. (2012) a ANOVA com somatório tipo II fornece um teste mais

poderoso para se avaliar os efeitos principais, não sendo indicado quando se tem interações, sendo vantajoso para se avaliar os efeitos principais. Outra vantagem, segundo os mesmos autores, é que este teste funciona para avaliar os efeitos principais independente da ordem de entrada das variáveis preditoras e do tipo de contraste utilizado. Também se fez o teste priorizando-se os efeitos principais, porque já no teste com o somatório tipo I não havia interação.

Aplicando-se o teste de ANOVA two way como somatório do tipo III entrando primeiro com a variável preditora religião

Utilizou-se a linha de comando: `anova6<-aov(ee$peso ~ ee$religiao+ee$sexo,ee)` e a visualização com: `Anova(anova6,type="III")`

```
Anova(anova6,type="III")
```

```
Anova Table (Type III tests)
```

```
Response: ee$peso
```

```
Sum Sq Df F value Pr(>F)
```

```
(Intercept) 78780 1 754.9829 < 2.2e-16 ***
```

```
ee$religiao 209 2 1.0007 0.381343
```

```
ee$sexo 888 1 8.5117 0.007184 **
```

```
Residuals 2713 26
```

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Percebe-se que ao utilizar-se o teste de ANOVA two way com o somatório dos quadrados do tipo III que os resultados estão de acordo com os do somatório do tipo II, novamente sendo o sexo o fator significativo sobre o peso.

Aplicando-se o teste de ANOVA two way como somatório do tipo III entrando primeiro com a variável preditora sexo

Utilizou-se a linha de comando: `anova7<-aov(ee$peso ~ ee$sexo+ee$religiao,ee)` e a visualização: `Anova(anova7, type="III")`

```
> Anova(anova7,type="III")
```

```
Anova Table (Type III tests)
```

Response: ee\$peso

Sum Sq Df F value Pr(>F)

(Intercept) 78780 1 754.9829 < 2.2e-16 ***

ee\$sexo 888 1 8.5117 0.007184 **

ee\$religiao 209 2 1.0007 0.381343

Residuals 2713 26

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Percebe-se que se invertendo a ordem de entrada da variável preditora, o resultado da ANOVA two way manteve-se de acordo com a ANOVA two way como somatório tipo II, mantendo-se o sexo com efeito significativo sobre o peso. Segundo Field et al. (2012) o cálculo de ANOVA two way com somatório do tipo III é mais indicado quando se tem efeito significativo sobre a interação, não se tendo a interação é indicado o somatório tipo II por ser mais poderoso para avaliar os efeitos principais. Também o cálculo com o somatório tipo III é mais utilizado para modelos desbalanceados, onde a interação é significativa e onde apenas os preditores são codificados e com constrastes ortogonais. Ainda segundo os mesmos autores, o somatório tipo III é usado como um pacote fechado de muitos softwares e a mudança da ordem das variáveis preditoras não afeta nos resultados. Neste exemplo os dados eram desbalanceados quanto a religião, não havia interação, portanto, conforme Field et al. (2012) seria mais indicado usar os resultados da ANOVA com somatório tipo II. Desta forma conforme Field et al., (2012) ao se fazer o teste de ANOVA two-way e também ANOVA threeway no software R é importante verificar as hipóteses; o tipo de contraste; e os tipos de somatórios que se adequam a escolha dos efeitos principais e das interações.

REFERÊNCIAS

ALCOFORADO, L. F. **Utilizando a Linguagem R: Conceitos, Manipulação, Visualização, Modelagem e Elaboração de Relatórios**. Rio de Janeiro: Alta Books, 2021. 384p.

BACHE S, WICKHAM, H. (2022). `_magrittr`: A Forward-Pipe Operator for R_. R package version 2.0.3, <<https://CRAN.R-project.org/package=magrittr>>.

BRITTO, E. N.; FILHO, D. P. M.; REGO, M. R. S. **Estatística Aplicada à Biologia usando o R: estatística descritiva e inferencial**. Paraná: Atena, 2021.

CARVALHO, V. **MySQL: Comece com o principal banco de dados open source do mercado**. 1ª ed. Casa do código, ALURA, 2020, 165p.

CRAWLEY, M. **The R book**. 2ª ed. John Wiley & Sons, Ltd. 2013, 975p.

CRAVO LMS, BIANCARDI AL, RONCONI TB, XAVIER VL, MORAES H. JR. Pressão intraocular em pacientes com vírus da imunodeficiência humana, correlação com TCD4, carga viral e características demográficas: estudo caso-controle. Clin Ophthalmol (em inglês). 2019 Aug 5;13:1477-1483. Doi: 10.2147/OPHTH.S216759. PMID: 31496638; PMCID: PMC6689137.

FIELD, A., JEREMY, M., ZOE, F. 2012. **Discovering Statistics Using R**. London, England: SAGE Publications, 2012.

FOX, J., AND BOUCHET-VALAT, M. (2024). `Rcmdr`: R Commander. R package version. 2.9-5.

LEVIN, J. **Estatística Aplicada a Ciências Humanas**. São Paulo: Harbra, 1985.

MÜLLER, K., WICKHAM, H. (2023). `_tibble`: Simple Data Frames_. R package version 3.2.1, <<https://CRAN.R-project.org/package=tibble>>.

REVELLE, W. 2020. `Psych`: procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, <https://CRAN.R-project.org/package=psych>=2.0.9, R Development Core Team R: **A Language and Environment for Statistical Computing**, 2008. R

WICKHAM, H., AVERICK, M., BRYAN, J., CHANG, W., MCGOWAN, LD., FRANÇOIS, R., GROLE-MUND, G., HAYES, A., HENRY, L., HESTER, J., KUHN, M., PEDERSEN, TL., MILLER, E., BACHE, SM., MÜLLER, K., OOMS, J., ROBINSON, D., SEIDEL, DP., SPINU, V., TAKAHASHI, K., VAUGHAN, D., WILKE, C., WOO, K., YUTANI, H. (2019). "Welcome to the tidyverse." *_Journal of Open Source Software_*, *4*(43), 1686. doi:10.21105/joss.01686 <<https://doi.org/10.21105/joss.01686>>.

WICKHAM, H., FRANÇOIS, R., HENRY, L., MÜLLER, K., VAUGHAN, D. (2023). *_dplyr: A Grammar of Data Manipulation_*. R package version 1.1.4, <<https://CRAN.R-project.org/package=dplyr>>.

WICKHAM, H., BRYAN, J. (2025). *_readxl: Read Excel Files_*. R package version 1.4.5, <<https://CRAN.R-project.org/package=readxl>>.

ZAR, J. H. **Biostatistical analysis**. 4ªed. New Jersey, Prentice-Hall, Inc, 1999. 663p.

SOBRE OS AUTORES

ELTON NUNES BRITTO - graduou-se em Oceanologia pela Fundação Universidade do Rio Grande – FURG/RS na cidade de Rio Grande/RS em 1998 e em Licenciatura em Ciências Biológicas pelo Centro Universitário Leonardo Da Vinci - UNIASSELVI – Indaial/SC em Manaus/AM em 2016. Concluiu o Mestrado em 2003 e o Doutorado em 2008, ambos em Biologia de Água Doce e Pesca Interior – BADPI pelo Instituto Nacional de Pesquisas da Amazônia – INPA/AM em Manaus. Durante o mestrado e doutorado cursou disciplinas de estatística aplicada trabalhando com os softwares estatísticos: STATISTICA, SYSTAT e BIOESTAT. Neste período do INPA participou também como Monitor da disciplina Fundamentos de Estatística para discentes dos cursos de Pós-graduação em Entomologia e Biotecnologia do convênio INPA/UFAM em 2006, e ministrou o curso de Estatística Aplicada a Microbiologia em 2007. Depois do doutorado entre 2008 à 2015 lecionou estatística em diversas faculdades de Manaus para discentes das áreas: Biológica, Saúde, e de Ciências Sociais. Desde de 2016 faz parte do quadro dos servidores efetivos do Instituto Federal de Educação, Ciência e Tecnologia do Pará - IFPA - Campus Santarém - PA como professor do Curso Técnico em Nível Médio em Aquicultura Subsequente ensinando dentre outras disciplinas estatística básica desde 2017. Em 2016 no IFPA também criou o Projeto de Extensão intitulado “Bioestatística Aplicada com o software BIOESTAT para biólogos e áreas afins”, ministrando cinco cursos de extensão de bioestatística aplicada, beneficiando os discentes do ensino médio do IFPA e finalistas de cursos de Graduação das Instituições de Ensino Públicas de Santarém, como a Universidade Federal do Oeste do Pará – UFOPA dos cursos de Engenharia de Pesca, Agronomia e Biologia; da Universidade Estadual do Pará – UEPA dos cursos de Biologia e Enfermagem; das Instituições Privadas como a Universidade Luterana do Brasil – ULBRA dos cursos de Biologia e Agronomia. De 2019 a 2021 cursou seis cursos de Estatística à distância sobre o software R. Em 2020 inseriu o software R ao projeto de Extensão supracitado, e em 2021 ministrou o curso intitulado “Curso de Bioestatística com Rstudio EAD” para servidores do Laboratório Federal de Defesa Agropecuária do Ministério da Agricultura Pecuária e Abastecimento – MAPA, Campinas- SP. Em 2021 publicou o livro Caderno no. 1 Estatística Aplicada à Biologia usando o R: Estatística Descritiva e Inferencial. Em 2023 concluiu a Especialização, MBA em Data Science e Statistics da EASEBRASIL- Faculdade de Negócios, EASEBRASIL, Brasil com carga-horária de 420h. <http://lattes.cnpq.br/7630462808118280>.

Damião Pedro Meira Filho - graduou-se em Física pela Universidade Federal do Pará– UFPA em 2003 e em Matemática pela Universidade do Estado do Pará – UEPA em 2002. Concluiu o mestrado em Física Teórica pela Universidade Federal do Pará em 2006, e durante o mestrado estagiou no Instituto de Física Teórica da Universidade Estadual Paulista – IFT-UNESP. Concluiu o Doutorado em Física-Matemática pelo Instituto de Física da Universidade de São Paulo–IFUSP em 2010. Desde 2009 faz parte do quadro de servidores efetivos como professor do Instituto Federal de Educação, Ciência e Tecnologia do Pará - IFPA - Campus Santarém, atualmente atua como docente do IFPA Campus Belém. Em 2024, alcançou o nível de Professor Titular. Desde 2010, desenvolve projeto de pesquisa sobre “mecânica quântica e métodos matemáticos” bem como desenvolve projeto de extensão intitulado “Minicursos sobre Tópicos de Física e Matemática” e lecionou as disciplinas de Estatística Básica e Física Estatística, para cursos de graduação, especialização e mestrado. Frequentou o curso de extensão de Estatística Aplicada “Curso de Estatística Aplicada utilizando o software BIOESTAT”. Atualmente é membro docente do projeto de extensão de Bioestatística Aplicada com o software BIOESTAT e R para biólogos e áreas afins, onde dedicou-se ativamente na confecção desse caderno, contribuindo nas análises e aplicações do software R. <http://lattes.cnpq.br/2267580178817640>.

MARIA RAIMUNDA SILVA REGO - concluiu o ensino médio em 2001 e desde 2003 faz parte do quadro de servidores públicos efetivos da Prefeitura Municipal de Santarém como Chefe do Setor de Recursos Humanos. Desde 2019 é discente do Curso Técnico Subsequente em Aquicultura do Instituto Federal de Educação Ciência e Tecnologia do Pará – IFPA, Campus Santarém e neste período do IFPA participou do III Congresso de Tecnologia e Desenvolvimento da Amazônia e da VIII Jornada Científica e Semana Integrada do IFPA. cursou a disciplina de Estatística Básica e atualmente é membro discente do projeto de extensão de Bioestatística Aplicada com o software BIOESTAT e R para biólogos e áreas afins, onde dedicou-se ativamente na confecção desse caderno, contribuindo nas análises e aplicações do software R. <http://lattes.cnpq.br/7618720047491204>.

GEISIANE DA ROCHA SARMENTO -Graduou-se no Bacharelado Interdisciplinar em Ciências da Terra pela Universidade Federal do Oeste do Pará (UFOPA) e, atualmente, é graduanda em Ciências Atmosféricas pela mesma instituição. Atuou como bolsista do Programa Institucional de Bolsas de Extensão (PIBEX), entre 2022 e 2023, desenvolvendo o projeto Clima e Turismo: Influências Socioeconômicas em Alter do Chão, Santarém, PA. Também participou, entre 2023 e 2024, do Programa Institucional de Bolsas de Iniciação Científica (PIBIC), com financiamento da Fundação Amazônia de Amparo a Estudos e Pesquisas (FAPESPA), integrando o projeto de pesquisa Influência da Turbulência nas Reações de Compostos Químicos Reativos na Atmosfera, em Diferentes Condições de Sazonalidade e de Superfície na Região Amazônica. Desde 2024, integra o projeto de extensão Bioestatística Aplicada com o Programa R para Biólogos e Áreas Afins, colaborando na elaboração de material didático e na condução de análises estatísticas aplicadas com os softwares BIOESTAT e R. Atualmente, é bolsista PIBIC do Instituto Nacional de Pesquisas Espaciais (INPE), onde desenvolve atividades de pesquisa no projeto Variabilidade Sazonal do Ozônio Estratosférico e sua Relação com a Precipitação de Partículas sobre a Anomalia Magnética da América do Sul. Currículo Lattes: <https://lattes.cnpq.br/7036145550370433>.

ESTATÍSTICA APLICADA À BIOLOGIA



MANIPULAÇÃO DE TABELAS, TESTES ESTATÍSTICOS
E ABORDAGENS DE ANOVA PARA A ANÁLISE DE
DADOS BIOLÓGICOS



www.atenaeditora.com.br



contato@atenaeditora.com.br



[@atenaeditora](https://www.instagram.com/atenaeditora)



www.facebook.com/atenaeditora.com.br

ESTATÍSTICA APLICADA À BIOLOGIA

USANDO O 

MANIPULAÇÃO DE TABELAS, TESTES ESTATÍSTICOS
E ABORDAGENS DE ANOVA PARA A ANÁLISE DE
DADOS BIOLÓGICOS



www.atenaeditora.com.br



contato@atenaeditora.com.br



[@atenaeditora](https://www.instagram.com/atenaeditora)



www.facebook.com/atenaeditora.com.br