Denilson Junio Marques Soares
Talita Emidio Andrade Soares
Wagner dos Santos

ANÁLISE ESTATÍSTICA E SEU USO NA PESQUISA EDUCACIONAL

com práticas no *software* R

ANÁLISE ESTATÍSTICA E SEU USO NA PESQUISA EDUCACIONAL

com práticas no *software* R

Denilson Junio Marques Soares
Talita Emidio Andrade Soares
Wagner dos Santos

ANÁLISE ESTATÍSTICA E SEU USO NA PESQUISA EDUCACIONAL: COM PRÁTICAS NO *SOFTWARE* R

Edição 1

Belém-PA

Rfb Editora

2021

https://doi.org/10.46898/rfb.9786558891093

Catalogação na publicação
Elaborada por Bibliotecária Janaina Ramos – CRB-8/9166

M152

Soares, Denilson Junio Marques

Análise estatística e seu uso na pesquisa educacional: com práticas no software R / Denilson Junio Marques Soares, Talita Emidio Andrade Soares, Wagner dos Santos – Belém: RFB, 2021.

Livro em PDF

88 p., il.

ISBN: 978-65-5889-109-3
DOI: 10.46898/rfb.9786558891093

1. Estatística educacional. I. Soares, Denilson Junio Marques. II. Soares, Talita Emidio Andrade. III. Santos, Wagner dos. IV. Título.

CDD 370.21

Índice para catálogo sistemático

I. Estatística educacional

Diagramação:

Laiane Borges.

Arte da capa:

Pryscila Rosy Borges de Souza.

Imagens da capa:

https://www.canva.com/

Revisão de texto:

Os autores.

Bibliotecária

Janaina Karina Alves Trigo Ramos

Assistente editorial

Manoel Souza.

Home Page: www.rfbeditora.com.
E-mail: adm@rfbeditora.com.
Telefone: (91)3085-8403/98885-7730.
CNPJ: 39.242.488/0001-07.
Barão de Igarapé Miri, sn, 66075-971, Belém-PA.

SUMÁRIO

APRESENTAÇÃO

Este material foi desenvolvido com a finalidade de ser o guia didático do curso *Análise Estatística e seu uso na Pesquisa Educacional*, ministrado nos dias 17 e 18 de junho de 2019 no Laboratório de Pesquisa do Instituto de Pesquisa em Educação e Educação Física (PROTEORIA) da Universidade Federal do Espírito Santo.

O objetivo principal deste curso é oferecer um embasamento teórico para a aplicação das técnicas estatísticas em pesquisas que envolvem as áreas de Educação e Educação Física, ressaltando a importância de cada análise para efetivação da interpretação dos resultados obtidos ao longo dessas pesquisas.

O livro está estruturado em quatro capítulos. O capítulo 1 apresenta conceitos básicos de amostragem, estatística descritiva e suas representações em gráficos e tabelas. O capítulo 2 traz alguns dos testes de hipótese mais utilizados em pesquisas das áreas de ciências humanas e sociais aplicadas, como testes de comparação de médias e de associação entre variáveis. O capítulo 3 é destinado ao estudo de correlação e regressão e o capítulo 4 apresenta uma introdução à análise de avaliações, pautada na teoria clássica dos testes.

Em todos os capítulos são apresentados exemplos práticos para serem executados no software estatístico R, adotado como instrumento facilitador na aprendizagem, ilustrando aspectos básicos com ênfase na compreensão da estrutura do software e na forma de operar seus comandos.

O software R pode ser obtido gratuitamente em `http://www.R-project.org`, em que são apresentadas versões para os principais sistemas operacionais: Linux, MacOS X e Windows. Durante o processo de instalação são criados atalhos na área de trabalho que podem ser acessados para rodar o programa. Caso o leitor precise cita-lo em suas publicações, basta digitar na janela aberta pelo software, conhecida como *R Console*, o comando *citation()*.

Gostaríamos de agradecer ao Instituto de Pesquisa em Educação e Educação Física (PROTEORIA) e ao Núcleo de Estudos e Pesquisas em Políticas Educacionais(NEPE) da Universidade Federal do Espírito Santo (UFES) pela oportunidade de ofertar este curso.

CAPÍTULO 1

CONCEITOS INICIAIS

1 Conceitos Iniciais

A Estatística é uma ciência que estuda, de forma sistemática, técnicas para coletar, organizar, descrever, analisar e interpretar dados oriundos de estudos ou experimentos realizados em qualquer área do conhecimento. Trata-se de uma ferramenta capaz de extrair informações acerca de uma população, pautada em conceitos inferenciais probabilísticos.

Em seu estudo, é fundamental compreender conceitos como o de amostragem, tipos de variáveis, medidas-resumo e suas representações.

1.1 Amostragem

Podemos definir uma **população** como sendo um conjunto de elementos que detém alguma característica comum, factível de ser estudada. Qualquer subconjunto dessa população, constituem uma **amostra**. Entretanto, para garantir que uma amostra seja representativa para uma população, mantendo suas características essenciais, é preciso muito cuidado no processo seleção dos seus elementos, conhecido como processo de amostragem. A Figura 1 ilustra este processo.

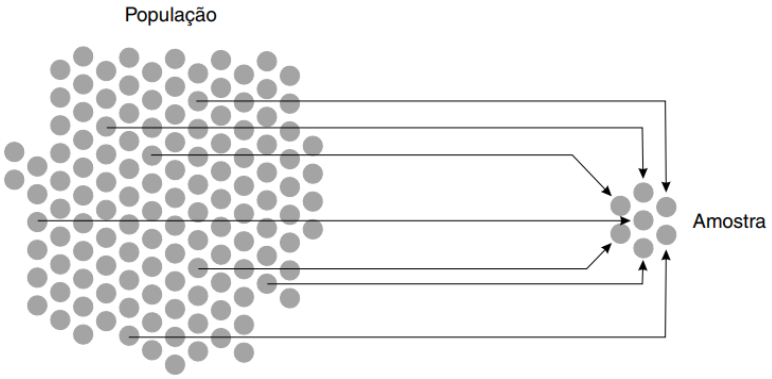

Figura 1: População e Amostra.
Fonte: Santos (2007, p. 17)

É muito comum trabalharmos com amostras, ao invés de populações, devido as dificuldades operacionais (e financeiras) que teríamos, caso contrário.

1.1.1 Cálculo do tamanho de uma amostra

Existem algumas fórmulas para o cálculo do tamanho de uma amostra. A que traremos aqui é uma que acreditamos ser de fácil compreensão por não envolver cálculos complexos e nem necessitar de valores tabelados de apoio. Primeiramente, precisamos adotar um valor para o erro amostral tolerável (E_0), para calcularmos uma primeira aproximação para o tamanho da amostra (n_0). Este valor pode ser considerado quando não se conhece o tamanho da população.

$$n_0 = \frac{1}{E_0^2} \tag{1}$$

Caso se conheça o tamanho da população (N), pode-se calcular o tamanho da amostra da seguinte maneira:

$$n = \frac{N \times n_0}{N + n_0} \tag{2}$$

Como exemplo, suponha que deseja-se selecionar uma amostra de uma população de tamanho 200 mil, considerando que os erros amostrais não ultrapassem 2%. Assim, das Equações 1 e 2, tem-se:

$$n_0 = \frac{1}{0,02^2} = 2500 \qquad \Longrightarrow \qquad n = \frac{200000 \times 2500}{200000 + 2500} \approx 2469,136.$$

Portanto, devem ser amostradas 2470 pessoas para se obter uma amostra 98% confiável.

1.1.2 Processos de Amostragem

São quatro os principais processos de amostragem: aleatório simples, sistemático, estratificado e por conglomerados. Na **amostragem aleatória simples**, cada elemento do conjunto tem a mesma probabilidade de ser selecionado. Esta seleção muitas vezes se dá por geradores de números aleatórios ou por qualquer outro tipo de processo randômico.

Na **amostragem sistemática**, primeiramente ordena-se os elementos e em seguida seleciona-se (aleatoriamente) um elemento inicial para compor a amostra. Os demais elementos são selecionados em intervalos constantes de unidades, cujo tamanho é determinado pela razão entre o tamanho da população e o da amostra a ser selecionada.

Por exemplo, suponha que um pesquisador da área de educação física acredite que a seleção de 139 dos 2363 atletas de Atletismo das Olimpíadas 2016 é suficiente para realizar sua pesquisa. Ele opta pelo processo de amostragem sistemático, considerando que os atletas estão organizados alfabeticamente, e sorteia o número 15 para ser o elemento inicial da sua amostra. Em seguida, ele calcula o salto amostral obtendo $\frac{2363}{139} = 17$. Assim, os elementos amostrados serão os de ordem $15; 15 + 17 = 32; 15 + 2 \times 17 = 49; 15 + 3 \times 17 = 66; \cdots ; 15 + 138 \times 17 = 2361$.

Na **amostragem por conglomerados**, primeiramente a população é dividida em grupos. Em seguida, um ou mais grupos são selecionados de maneira aleatória e todos os elementos desses grupos são amostrados. Por exemplo, considere que o diretor de uma escola pretende entrevistar os alunos. Para isto, primeiramente ele seleciona, aleatoriamente, 3 salas de aula e, em seguida, entrevista todos os alunos pertencentes a elas. Este tipo de amostragem é comum em pesquisas de satisfação realizadas por companhias aéreas, em que geralmente seleciona-se um número de aeronaves cujos passageiros são todos entrevistados.

Na **amostragem estratificada**, os elementos são primeiramente dispostos em grupos com

alguma similaridade, chamados de estratos. Em seguida, proporcionalmente ao tamanho de cada grupo, é realizada uma seleção aleatória dos elementos desses grupos para compor a amostra. Os grandes institutos de pesquisa, como o Ibope e o Datafolha, geralmente utilizam a amostragem aleatória estratificada em que as divisões ocorrem por renda, idade, escolaridade, gênero, entre outras variáveis.

1.1.3 Exemplo no R

Para gerar números aleatórios no R, podemos usar o comando *sample()*. Por exemplo, para gerar 5 números aleatórios no intervalo de 1 a 100, fazemos:

```
> sample(1:100, 5)

[1] 62 52 17  1 39
```

Cada vez que se executa o comando, o R retornará novos valores. É possível definir uma semente aleatória, através do comando *set.seed()*, para garantir que seja utilizado o mesmo número aleatório gerado, sempre que for necessário repeti-lo. A mesma simulação acima é realizada, considerando a semente número 125:

```
> set.seed(125)
> sample(1:100, 5)

[1] 30 63 72 24 51
```

1.2 Tipos de variáveis

Para termos uma boa amostragem dos dados, precisamos nos preocupar com as variáveis que interferem nos resultados. Por exemplo, um candidato A pode ser o preferido para os eleitores de baixa renda, enquanto que um candidato B é o mais querido para os eleitores de melhor condição financeira. Dessa forma, a renda da população é uma variável que precisa ser levada em consideração em uma pesquisa eleitoral. Outras variáveis que devem ser analisadas neste caso são o sexo, a idade, o grau de escolaridade, entre muitas outras.

Quando uma variável expressa uma qualidade ou preferência de um entrevistado ela é chamada de **qualitativa**. Se a variável expressa valores numéricos, ela é chamada de **quantitativa**. Assim, sexo e grau de escolaridade são variáveis qualitativas enquanto que a renda familiar e a idade do entrevistado são variáveis quantitativas.

As variáveis qualitativas podem ser classificadas em **ordinais**, quando podem ser ordenadas, ou **nominais**, caso contrário. Assim, o grau de escolaridade é uma variável qualitativa ordinal e a variável sexo é uma variável qualitativa nominal.

As variáveis quantitativas podem ser classificadas em **discreta**, quando é expressa por um número inteiro, ou **contínua**, quando expressa por um número real não-inteiro. Assim, o número

de acesso a uma plataforma online pode ser considerada como uma variável quantitativa discreta e a altura de um grupo de adolescentes como uma variável quantitativa contínua. A Figura 2 resume estas classificações.

Figura 2: Classificação de uma variável.

Considere que um pesquisador esteja interessado em fazer um levantamento sobre algumas características dos professores de Educação Física da rede pública de ensino da cidade de Serra-ES. Para isto, selecionou uma amostra com os dados de 30 desses profissionais, conforme ilustra a Tabela 1. Os dados em questão são hipotéticos.

Tabela 1: Informações sobre sexo, estado civil, idade, escolaridade, tempo de serviço e renda de 30 professores de Educação Física da rede pública de ensino da cidade de Serra-ES.

Número	Sexo	Estado Civil	Idade	Escolaridade	Tempo de serviço (em anos)	Renda (× sal. mín.)
1	masculino	solteiro	32	graduação	7	3,15
2	feminino	solteiro	23	graduação	1	2,20
3	feminino	casado	30	mestrado	5	4,04
4	masculino	casado	25	graduação	2	2,20
5	feminino	casado	27	mestrado	2	3,00
6	masculino	casado	45	graduação	20	3,50
7	feminino	casado	52	graduação	28	3,80
8	feminino	solteiro	30	doutorado	5	5,20
9	masculino	casado	30	mestrado	7	3,25
10	feminino	solteiro	35	graduação	10	3,00
11	masculino	solteiro	25	mestrado	2	3,00
12	feminino	casado	26	mestrado	2	3,00
13	masculino	casado	36	graduação	11	3,00
14	masculino	casado	44	graduação	18	3,50
15	masculino	solteiro	28	mestrado	6	3,25
16	feminino	casado	51	doutorado	25	6,10
17	masculino	casado	30	mestrado	6	4,60
18	masculino	casado	38	graduação	15	3,40
19	feminino	casado	28	mestrado	1	3,00
20	feminino	casado	55	graduação	30	3,80
21	feminino	casado	26	mestrado	1	3,00
22	masculino	solteiro	30	graduação	4	2,50
23	feminino	casado	34	doutorado	10	5,90
24	feminino	solteiro	25	mestrado	1	3,00
25	masculino	solteiro	60	graduação	35	4,00
26	masculino	casado	35	graduação	12	2,80
27	masculino	casado	30	mestrado	4	3,20
28	feminino	casado	30	graduação	6	2,60
29	feminino	casado	35	mestrado	10	3,50
30	masculino	solteiro	26	graduação	2	2,20

As variáveis qualitativas nominais presentes na Tabela 1 são o sexo e o estado civil. Estas variáveis podem ser dicotomizadas, auxiliando em alguns processos de análises estatísticas. Observe que é possível ordenar a variável escolaridade, logo trata-se de uma variável qualitativa do tipo ordinal. As variáveis idade e tempo de serviço (em anos) são variáveis quantitativas discretas e a variável renda é uma variável quantitativa contínua.

1.2.1 Representação global do conjunto de dados

Muitas vezes, o pesquisador está interessado em uma representação global do comportamento de uma variável. Para uma leitura mais objetiva dos dados, podemos construir tabelas de frequên-

cia e gráficos, como mostra a Tabela 2 que apresenta a distribuição de frequências da variável escolaridade da Tabela 1.

Tabela 2: Distribuição de frequências da variável escolaridade.

Escolaridade	**Frequência**	**Proporção**	**Porcentagem**
Graduação	15	$\frac{15}{30} = 0,50$	50%
Mestrado	12	$\frac{12}{30} = 0,40$	40%
Doutorado	3	$\frac{3}{30} = 0,10$	10%
Total	30	1,00	100%

Para representar variáveis quantitativas contínuas é comum a dispor os dados em classes com a mesma amplitude, que deve ser determinada de acordo com a familiaridade do pesquisador com os dados. Entretanto, é importante ressaltar que um número pequeno de classes causa perda de informações. Bussab e Morettin (2010) sugerem o uso de 5 a 15 classes para representar uma variável. Na Tabela 3 há informações acerca da variável salário.

Tabela 3: Distribuição de frequências da variável renda.

Salários	**Frequência**	**Proporção**	**Porcentagem**
$[2, 3)$	6	$\frac{6}{30} = 0,20$	20%
$[3, 4)$	18	$\frac{18}{30} = 0,60$	60%
$[4, 5)$	3	$\frac{3}{30} = 0,10$	10%
$[5, 6)$	2	$\frac{2}{30} \approx 0,0667$	6,67%
$[6, 7)$	1	$\frac{1}{30} \approx 0,0333$	3,33%
Total	30	1,00	100%

Observe que foi utilizada a notação matemática $[x, y)$ para designar os intervalos que contém o extremo x mas que não contém o extremo y. A representação em gráficos ocorre de maneiras distintas para variáveis qualitativas e quantitativas. Para o primeiro grupo, destacam-se a representação em gráficos de barras e setores. Já para o segundo, destacam-se a representação em gráficos de dispersão, histogramas, boxplot e ramo-e-folhas.

1.2.2 Exemplos no R

Para uma melhor leitura dos dados no software R, dispomos as respostas dos 30 professores entrevistados em uma planilha do Excel, conforme ilustra a Figura 3.

Número	Sexo	Estado Civil	Idade	Escolaridade	Tempo de Serviço	Renda
1	masculino	solteiro	32	graduação	7	3.15
2	feminino	solteiro	23	graduação	1	2.20
3	feminino	casado	30	mestrado	5	4.04
4	masculino	casado	25	graduação	2	2.20
5	feminino	casado	27	mestrado	2	3.00
6	masculino	casado	45	graduação	20	3.50
7	feminino	casado	52	graduação	28	3.80
8	feminino	solteiro	30	doutorado	5	5.20
9	masculino	casado	30	mestrado	7	3.25
...	...	...	...	...	...	...
30	masculino	solteiro	26	graduação	2	2.20

Figura 3: Dados dispostos em uma planilha do Excel

Para a leitura dos dados provenientes da planilha Excel no R, o arquivo foi previamente salvo como *Texto (separado por tabulações)(*.txt)* e, em seguida, utilizou-se o comando *read.table()*:

```
> dados=read.table("C:/Users/Usuario/Desktop/Curso no R/dados1.txt",head=T)
```

Para facilitar, pode-se clicar com o botão direito do *mouse* no arquivo salvo, clicar em propriedades e copiar o local em que o trabalho foi salvo. Entretanto, deve-se lembrar de inverter o sentido das barras que separam as respectivas pastas. Em tempo, a função *head=T* ou *head=TRUE* indica que há um cabeçalho no arquivo analisado. O comando *head()* nos fornece um resumo da planilha analisada.

```
> head(dados)
```

```
  Numero      Sexo EstadoCivil Idade Escolaridade TempodeServico Renda
1      1 masculino    solteiro    32    graduacao              7  3.15
2      2  feminino    solteiro    23    graduacao              1  2.20
3      3  feminino      casado    30     mestrado              5  4.04
4      4 masculino      casado    25    graduacao              2  2.20
5      5  feminino      casado    27     mestrado              2  3.00
6      6 masculino      casado    45    graduacao             20  3.50
```

É possível também selecionar apenas uma das variáveis do conjunto de dados. Para isto, acrescentamos o símbolo $ junto ao nome da variável que queremos selecionar. Por exemplo, para selecionar apenas a variável estado civil, fazemos:

```
> dados$EstadoCivil
```

```
 [1] solteiro solteiro casado   casado   casado   casado   casado   solteiro
 [9] casado   solteiro solteiro casado   casado   casado   solteiro casado
[17] casado   casado   casado   casado   casado   solteiro casado   solteiro
[25] solteiro casado   casado   casado   casado   solteiro
Levels: casado solteiro
```

Como visto na seção anterior, é comum trabalharmos com amostras para representar uma população. Suponhamos que seja de interesse selecionar uma amostra de 5 elementos pertencentes ao conjunto de dados representados na Tabela 1. Para isto, podemos utilizar o comando *sample* e a semente 14302.

```
> set.seed(14302)
> dados[sample(nrow(dados),5),]
```

```
   Numero      Sexo EstadoCivil Idade Escolaridade TempodeServico Renda
16     16  feminino       casado    51    doutorado             25  6.10
17     17 masculino       casado    30     mestrado              6  4.60
26     26 masculino       casado    35    graduacao             12  2.80
9       9 masculino       casado    30     mestrado              7  3.25
30     30 masculino     solteiro    26    graduacao              2  2.20
```

Para a construção de qualquer tipo de gráfico no R, há algumas opções comuns que usaremos no decorrer deste tópico. São elas:

xlim: contém os limites do eixo x;
ylim: contém os limites do eixo y;
xlab: nomeia o eixo x;
ylab: nomeia o eixo y;
main: título do gráfico;
col: cor de preenchimento do gráfico;
pch: formato dos pontos do gráfico;
type: formato do segmento que une os pontos do gráfico;
text: insere um texto nas coordenadas definidas.

Vejamos alguns exemplos. Na Figura 4, temos o gráfico de dispersão para as variáveis Idade e Tempo de Serviço. Observe que estas variáveis parecem estar altamente correlacionadas, o que veremos nos próximos capítulos.

```
> plot(dados$Idade, dados$TempodeServico, pch=1,
+ xlab="Idade", ylab="Tempo de Serviço",
+ main="Gráfico de dispersão para Idade versus Tempo de Serviço")
```

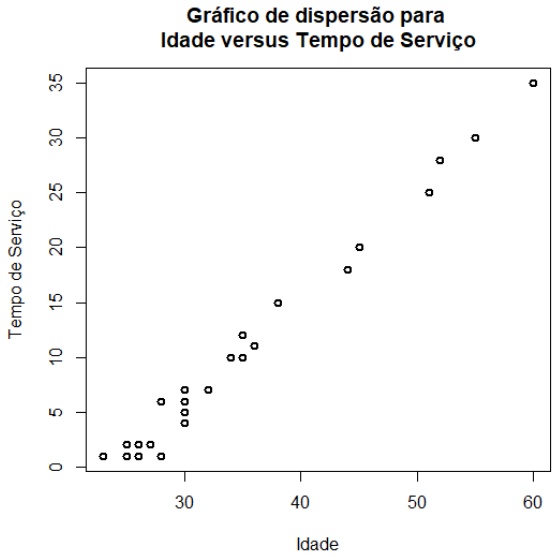

Figura 4: Gráfico de dispersão para as variáveis Idade e Tempo de Serviço

Na Figura 5 temos a representação em gráfico de barras vertical da variável Estado Civil. Para construir o gráfico de barras horizontal, representado na Figura 6, basta inverter os eixos coordenados e acrescentar o comando *horiz=TRUE* no script abaixo.

```
> barplot(table(dados$EstadoCivil), col=c("green","blue"),ylim=c(0,25),
+ main="Dados referentes à variável Estado Civil",
+ xlab="Sexo", ylab="Número de Entrevistados")
```

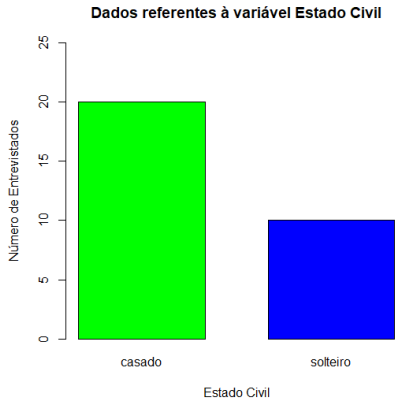

Figura 5: Dados referentes à variável Estado Civil

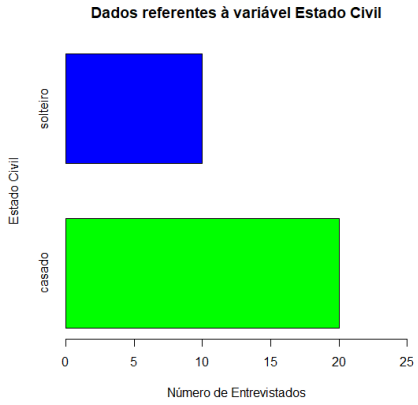

Figura 6: Dados referentes à variável Estado Civil

Na Figura 7 temos a representação em gráfico de setores da variável Escolaridade. Como não

colocamos o comando para definir as cores, no script, o software retornará com as cores definidas como padrão.

```
> pie(table(dados$Escolaridade),
+ main="Dados referentes à Escolaridade")
```

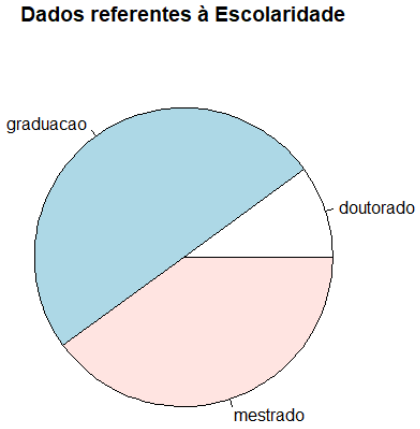

Figura 7: Dados referentes à Escolaridade

Na Figura 8 temos um histograma que representa a variável Renda. A definição das classes também ocorreu de forma padrão no software.

```
> hist(dados$Renda, main="Histograma para a variável Renda", prob=F,
+ xlab="Renda (em salários mínimos)", ylab="Frequência de Entrevistados",
+ col=c("orange"), xlim=c(2,7), ylim=c(0,10))
```

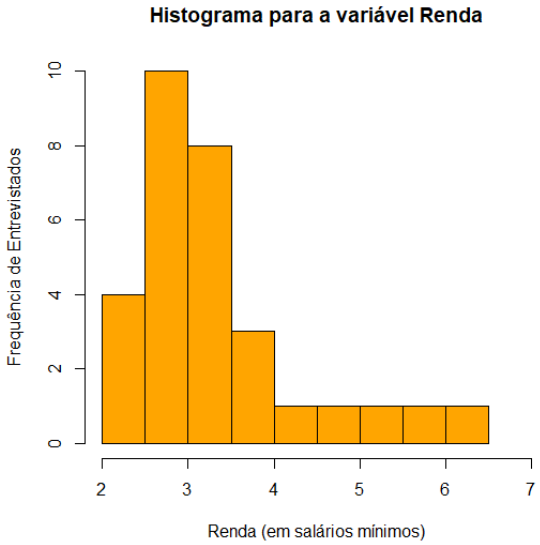

Figura 8: Dados referentes à Renda

Na Figura 9 temos um Boxplot representando a variável Idade. Trata-se de um modelo de gráfico bastante útil que contém informações essenciais para uma análise mais robusta dos dados. O traço horizontal representa a mediana dos dados, ou seja, o valor que divide o conjunto de dados em partes iguais. Os limites inferior e superior da caixa representam o primeiro e o terceiro quartis, respectivamente. As linhas pontilhadas indicam aproximadamente o valor de dois desvios-padrão e são conhecidas como *whiskers*. Caso algum dado for plotado fora dessas linhas, ele é considerado um *outlier*, ou seja, um valor atípico, que apresenta um grande afastamento dos demais valores do conjunto de dados. Veremos, no próximo tópico, estas medidas de dispersão mais detalhadamente.

```
> boxplot(dados$Idade, main="Boxplot para a variável Idade",
+ ylab="Idade", col=("pink"))
```

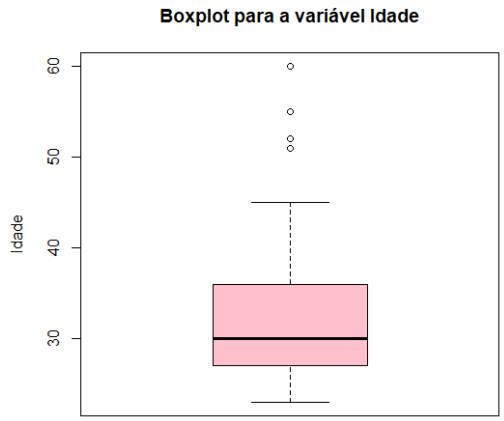

Figura 9: Dados referentes à Renda

Podemos também relacionar duas variáveis e representá-las através do Boxplot. A Figura 10 ilustra esta situação, relacionando as variáveis Sexo e Renda.

```
> boxplot(dados$Renda ~ dados$Sexo,
+ main="Boxplot para as variáveis Sexo e Renda",
+ xlab="Sexo", ylab="Renda", col=c("yellow","orange"))
```

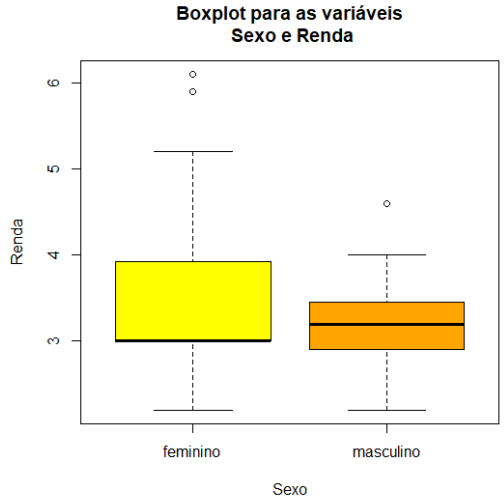

Figura 10: Dados referentes à Renda e o Sexo

Um tipo de gráfico bastante útil para resumir um conjunto de dados, quando se tem interesse na forma de sua distribuição (simétrica ou assimétrica) é o gráfico ramo-e-folhas. Nele, também é possível analisar a frequência de observações e a presença de *outliers*. Em sua construção, as

observações devem ser divididas em duas partes: o ramo, colocado à direita de uma linha vertical, e as folhas, colocadas à esquerda. O comando para gerá-lo no R é o *stem()*.

```
> stem(dados$Idade)

  The decimal point is 1 digit(s) to the right of the |

  2 | 3
  2 | 555666788
  3 | 000000024
  3 | 55568
  4 | 4
  4 | 5
  5 | 12
  5 | 5
  6 | 0
```

Neste exemplo, o ramo representa a casa das dezenas das idades e a folha representa a casa das unidades. Observe que os dados estão distribuídos assimetricamente à esquerda.

1.3 Medidas-Resumo

Como vimos, a representação dos dados por meio de gráficos e tabelas resume e fornece informações globais sobre o comportamento das variáveis. Entretanto, as vezes precisamos resumir ainda mais estes dados, apresentando um ou outro valor que represente o conjunto de dados como um todo. Esses valores são denominados "Medidas de Posição, Medidas de Tendência Central ou de Centralidade".

Podemos também estar interessados na variabilidade de um conjunto de dados, ou seja, no quanto seus elementos são dispersos entre eles. Estas informações são omitidas pelas medidas de posição, mas contém informações importantes que devem ser levadas em consideração na análise de um conjunto de dados.

Para ilustração, observe a Tabela 4 que apresenta as notas de dois candidatos que disputam um cargo em uma empresa de telecomunicações.

Tabela 4: Notas de dois candidatos em um processo seletivo.

Disciplina	**Candidato 1**	**Candidato 2**
Português	7,3	6,0
Matemática	7,5	8,5
Legislação	7,7	9,0
Informática	7,5	6,5

Na Tabela 4, embora as médias e medianas sejam iguais, o candidato 1 apresentou notas mais

regulares do que o candidato 2. Em outras palavras, há menor variabilidade nas notas obtidas pelo candidato 1. Dessa forma, se a empresa tiver interessada em um candidato cujos conhecimentos são mais homogêneos, seria interessante contratá-lo. Observe que as medidas de posição não trazem essa informação.

1.3.1 Medidas de Posição

As principais medidas de posição são a média aritmética, a mediana, a moda, os quartis e os percentis. Para encontrarmos a média aritmética de um conjunto de dados basta somar os valores das observações e dividir o resultado pelo número delas. Assim, a média aritimética é dada por

$$\overline{x} = \frac{x_1 + \ldots + x_n}{n} \tag{3}$$

A moda (Mo) é o valor mais frequente do conjunto de dados e a mediana (Md) é o valor observado que divide o este conjunto (depois de ordenado) em duas partes com a mesma quantidade. Matematicamente,

$$Md = X_{(\frac{n+1}{2})}, \text{ se } n \text{ é impar ou } Md = \frac{X_{(\frac{n}{2})} + X_{(\frac{n+1}{2})}}{2}, \text{ se } n \text{ é par.} \tag{4}$$

em que $X_{(n)}$ representa o elemento de ordem n.

Os quartis são os valores que dividem uma amostra de dados em quatro partes iguais. Abaixo do 1º e acima do 3º quartil encontram-se 25% dos dados, o 2º quartil coincide com a mediana, ou seja, representa o elemento central dos dados ordenados. A Figura 11 ilustra a situação descrita.

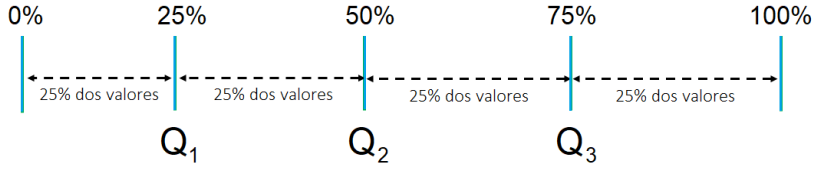

Figura 11: Representação para quartis.

Semelhante ao que acontece com os quartis, os percentis são medidas que dividem a amostra (por ordem crescente dos dados) em 100 partes. Dessa forma, o 1º percentil determina os 1% menores dos dados e o 65º percentil determina os 65% menores dos dados.

1.3.2 Medidas de Dispersão

As principais medidas de dispersão são a amplitude, variância, desvio-padrão e coeficiente de variação. Para o cálculo da amplitude, basta subtrair o menor do maior elemento de uma amostra. O cálculo da variância é um pouco mais elaborado: faz-se a razão entre a soma dos quadrados dos

desvios dos dados (encontrados ao subtrairmos o valor observado pela média aritmética) pelo total de dados observados. Matematicamente:

$$Var(X) = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n} \tag{5}$$

em que X representa o conjunto formado pelos n dados x_i e $\bar{x}$ representa a média aritmética desses dados. Há uma correção para o cálculo da variância de uma amostra que consiste na divisão por $n-1$ ao invés de n na Equação 5.

O desvio-padrão de uma amostra é obtido extraindo a raiz quadrada da variância e o coeficiente de variação é obtido pela razão entre o desvio-padrão e a média aritmética de uma amostra. O coeficiente de variação (CV) é utilizado quando precisamos comparar variáveis que apresentam médias diferentes, sendo o mais homogêneo aquele que apresentar menor CV.

Calculemos o CV das notas dos candidatos 1 e 2 da Tabela 4 para exemplificar estes conceitos. Para o candidato 1, temos:

$$\bar{X} = \frac{7,3 + 7,5 + 7,7 + 7,5}{4} = 7,5 \quad \text{e}$$

$$Var(X) = \frac{(7,3 - 7,5)^2 + (7,5 - 7,5)^2 + (7,7 - 7,5)^2 + (7,5 - 7,5)^2}{4} = 0,02$$

$$DP(X) = \sqrt{0,02} \approx 0,141 \quad \Longrightarrow \quad CV(X) = \frac{0,141}{7,5} = 0,0188.$$

Para o candidato 2, temos:

$$\bar{Y} = \frac{6 + 8,5 + 9 + 6,5}{4} = 7,5;$$

$$Var(Y) = \frac{(6 - 7,5)^2 + (8,5 - 7,5)^2 + (9 - 7,5)^2 + (6,5 - 7,5)^2}{4} = 1,625;$$

$$DP(Y) = \sqrt{1,625} \approx 1,275 \quad \Longrightarrow \quad CV(X) = \frac{1,275}{7,5} = 0,17.$$

Portanto, como anteriormente observado, as notas obtidas pelo candidato 2 são mais regulares.

1.3.3 Exemplos no R

Os comandos do R para as principais medidas-resumo, estão representados na Tabela 5.

Tabela 5: Comandos para as principais Medidas-Resumo

Função	Significado
mean()	Média
median()	Mediana
table()	Moda
max()	Máximo
min()	Mínimo
quantile()	Quantis
sd()	Desvio-padrão
var()	Variância

Um resumo para as medidas de posição pode ser obtido através da função *summary()*. Como exemplo, consideremos a variável Idade da Tabela 1.

```
> summary(dados$Idade)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  23.00   27.25   30.00   34.03   35.75   60.00
```

Observe que as idades variam entre 23 e 60 anos e possuem média de $34,03$ e mediana 30. O primeiro e o terceiro quartis foram $27,25$ e $35,75$, respectivamente.

CAPÍTULO 2

ALGUNS TESTES ESTATÍSTICOS

2 Alguns Testes Estatísticos

Em análises realizadas em três revistas de engenharia do Brasil entre 1999 e 2004, Mello; Alencar e Peternelli (2004) encontraram a presença de análises estatísticas em cerca de 63% dos mais de 360 artigos publicados. Em educação Física, Teixeira et al. (2015) analisaram todos os artigos publicados no triênio 2009-2011, em periódicos com estratificação $B2$ ou superior no QUALIS CAPES, e perceberam a presença de análises estatísticas em pelo menos 56,7% deles.

Entretanto, segundo Inácio; Encinas e Santana (2012), as análises e testes estatísticos são poucos utilizados nos artigos e trabalhos científicos na área das ciências humanas, principalmente na área de educação. Neste capítulo, discutiremos alguns dos principais testes de hipóteses que podem ser utilizados nas mais diversas áreas do conhecimento.

Um teste de hipóteses é um procedimento estatístico que permite aceitar ou rejeitar uma hipótese (hipótese de nulidade), utilizando os dados amostrados. A hipótese de nulidade é simplificada como H_0 e caso o pesquisador a rejeite, a decisão é pela hipótese alternativa H_1 (ou H_a). Passos para a realização de um teste de hipóteses:

1. Enunciar a hipótese H_0 a ser testada (hipótese de nulidade) e a hipótese H_1 alternativa;
2. Especificar o nível de significância (α) a ser adotado no teste, que é, em termos práticos, a probabilidade de se rejeitar incorretamente a hipótese nula quando ela é verdadeira, ou seja, de se cometer um erro estatístico conhecido como erro do tipo I;
3. Identificar a estatística do teste a ser realizado;
4. Determinar a região crítica do teste e a região de não rejeição da hipótese de nulidade, em função do nível de significância adotado e através das tabelas estatísticas;
5. Calcular, para a amostra selecionada, o valor da estatística do teste (valor calculado).
6. Concluir pela rejeição ou não de H_0, caso o valor calculado pertença ou não à região crítica do teste, respectivamente.

Em geral, os softwares de análises estatísticas retornam um valor para a probabilidade de se obter uma estatística de teste igual ou mais extrema que a observada em uma amostra, sob a hipótese nula, conhecido como valor-p. Em termos práticos, um valor-p pequeno significa que obter um valor da estatística do teste como o observado é muito improvável, levando assim à rejeição da hipótese nula. A análise pelo valor-p pode auxiliar os pesquisadores a concluírem se suas hipóteses estão ou não corretas.

$$\text{valor-p} \leq \alpha \Rightarrow \text{Rejeita-se } H_0$$

$$\text{valor-p} \geq \alpha \Rightarrow \text{Não Rejeita-se } H_0$$

No decorrer deste texto, consideraremos $\alpha = 5\%$ como nível de significância.

2.1 Testes Paramétricos vs Não-paramétricos

Os testes de hipóteses se dividem em testes paramétricos e testes não-paramétricos. Segundo Reis e Junior (2007), os paramétricos são aqueles que utilizam os parâmetros da distribuição (ou uma estimativa deles) para o cálculo de sua estatística. Já os não-paramétricos utilizam postos atribuídos aos dados ordenados. Normalmente, os testes paramétricos são mais rigorosos e possuem mais pressuposições para sua validação.

Segundo Favero e Favero (2015), para a aplicação dos testes paramétricos, as observações devem ser independentes, oriundas de populações com uma determinada distribuição (geralmente a normal) e as variáveis em estudo devem ser mensuráveis. Para testes de comparação de duas médias populacionais emparelhadas ou mais de duas médias populacionais, essas populações devem ter mesma variância.

Nas ciências sociais aplicadas é mais comum o uso de testes não-paramétricos por não exigirem hipóteses sobre a distribuição de probabilidade da população e por permitirem trabalhar com amostras menores. Entretanto, testes paramétricos são mais poderosos e devem ser escolhidos, sempre que possível.

2.2 Testes para normalidade e homogeneidade de variâncias

Existem alguns testes para verificarmos a hipótese de que os dados são normalmente distribuídos, sendo que os mais utilizados são os de Kolmogorov-Smirnov (K-S) e Shapiro-Wilk, mais adequado para amostras pequenas. Na Tabela 6, temos alguns testes estatísticos para normalidade dos dados e seu respectivo comando para análise no software R. Estes testes pertencem ao pacote *nortest* (Gross; Ligges e Ligges (2015)).

Tabela 6: Comandos para verificação da normalidade dos dados.

Teste	**Comando no R**
Kolmogorov-Smirnov	ks.test(dados, “pnorm”, mean(dados), sd(dados))
Lilliefors	lillie.test(dados)
Cramér-von Mises	cvm.test(dados)
Shapiro-Wilk	shapiro.test(dados)
Shapiro-Francia	sf.test(dados)
Anderson-Darling	ad.test(dados)

Também podemos construir o gráfico de dispersão dos quantis amostrais *versus* os quantis teóricos (da distribuição normal), para verificar se os dados seguem uma distribuição normal de probabilidades, o que será verdade quando os pontos plotados se dispuserem em torno de uma reta, conforme ilustra a Figura 12. No R, usamos a função *qqnorm()*, para a construção deste tipo de gráfico.

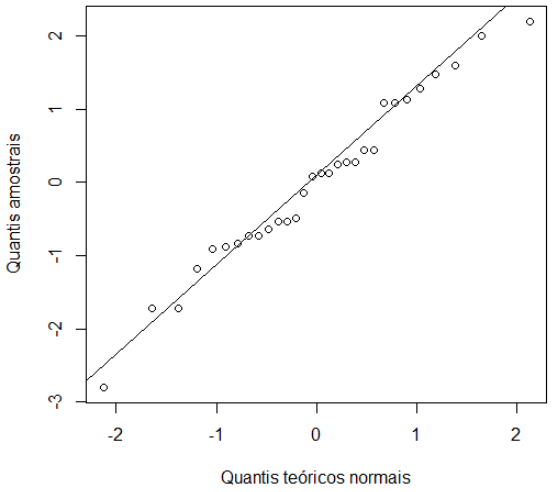

Figura 12: Gráfico Quantil-Quantil para análise da normalidade dos dados.

Para verificar se as variâncias são homogêneas, podemos utilizar os testes expostos na Tabela 7. Neles, a hipótese nula é a de que as variâncias dos grupos testados são iguais (homocedasticidade). Entretanto, a análise gráfica de Boxplots, embora mais subjetiva, também pode indicar esta característica.

Tabela 7: Comandos para verificação da homocedasticidade dos dados.

Teste	**Comando no R**
F de Fisher	var.test(x, y)
Bartlett	bartlett.test(x,y)
Levene	leveneTest(x,y)

2.3 Testes para comparação de duas médias

Muitos problemas aparecem quando se deseja testar hipóteses sobre médias de populações diferentes. Por exemplo, podemos querer comparar a médias das notas de duas turmas de uma determinada disciplina, uma com e outra sem aulas de monitoria, a fim de verificar a significância da mesma.

No caso paramétrico, para a comparação de médias de duas amostras independentes, sugere-se o uso do teste t. Entretanto, para seu uso, é preciso verificar, inicialmente, se as pressuposições de normalidade e homocedasticidade para as populações foram satisfeitas. O comando no R para este teste é *t.test(x,y)* e a hipótese de nulidade é a de que as médias populacionais são iguais para ambos os grupos (x e y).

Em algumas situações, não há independência entre os grupos. Por exemplo, pode-se comparar o

desempenho da mesma turma em dois momentos distintos: antes e depois da aplicação do método inovador. Nestes casos, dizemos que os grupos a serem comparados são pareados e acrescentamos o termo *paired=t* no comando do R, que passa a ser *t.test(x,y,paired=t)*.

No caso não-paramétrico, para amostras independentes, sugere-se o uso do teste de Wilcoxon-Mann-Whitney, cujo comando no R é *wilcox.test(x,y)* e para amostras dependentes, sugere-se o uso do teste de Wilcoxon, cujo comando no R é *wilcox.test(x,y,paired=T))*.

2.3.1 Exemplo no R

Para verificar se a adição de aulas de monitoria é significativa para a melhoria das notas de uma turma de fisiologia, foram amostradas as notas de 25 estudantes de duas turmas distintas: uma com (x) e outra sem (y) essas aulas. Os dados estão disponíveis na tabela 8.

Tabela 8: Notas de duas turmas de Fisiologia.

Notas sem a monitoria	20, 25, 75, 50, 63, 60, 65, 11, 75, 23, 28, 37, 86, 54, 22, 62, 44, 35, 44, 46, 50, 60, 52, 26, 73.
Notas com a monitoria	55, 60, 30, 55, 75, 85, 76, 62, 25, 54, 86, 77, 74, 28, 49, 57, 63, 66, 84, 58, 60, 60, 75, 45, 40

Primeiramente, precisamos verificar se trata-se de um teste paramétrico ou não. Para isto, utilizaremos o teste de Shapiro-Wilk para normalidade.

```
> x=c(20, 25, 75, 50, 63, 60, 65, 11, 75, 23, 28, 37,
+ 86, 54, 22, 62, 44, 35, 44, 46, 50, 60, 52, 26, 73)
> y=c(55, 60, 30, 55, 75, 85, 76, 62, 25, 54, 86, 77,
+ 74, 28, 49, 57, 63, 66, 84, 58, 60, 60, 75, 45, 40)
> shapiro.test(x)

        Shapiro-Wilk normality test

data:  x
W = 0.97061, p-value = 0.6607

> shapiro.test(y)

        Shapiro-Wilk normality test

data:  y
W = 0.9493, p-value = 0.2418
```

Observe que ambos possuem uma distribuição normal. Para verificar a homocedasticidade, utilizaremos o teste F.

```
> var.test(x,y)

        F test to compare two variances

data:  x and y
F = 1.374, num df = 24, denom df = 24, p-value = 0.4422
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6054606 3.1178915
sample estimates:
ratio of variances
          1.373958
```

Observe que as variâncias das populações são homogêneas. Assim, como os pressupostos foram satisfeitos, iremos comparar os grupos x e y através do teste t.

```
> t.test(x,y)

        Welch Two Sample t-test

data:  x and y
t = -2.3604, df = 46.838, p-value = 0.02247
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -23.191775  -1.848225
sample estimates:
mean of x mean of y
    47.44     59.96
```

Como o valor-p calculado foi menor do que o nível de significância adotado (5%), concluímos pela rejeição da hipótese H_0 que indicava que as médias das notas eram as mesmas para ambos os grupos. Assim, como a média de y é maior do que a média de x, em termos absolutos, concluímos que a adição de aulas de monitoria é significativa para uma melhoria das notas dos estudantes.

2.4 Testes para a comparação entre três ou mais grupos

Agora imagine o caso em que temos mais do que duas categorias para a avaliar. No caso em que todas as categorias analisadas são independentes, temos a ANOVA de Welch, cuja função no R é *anova()*, no caso paramétrico, e o teste de Kruskal-Wallis, função *kruskal.test()*, no caso não-paramétrico.

Se as categorias analisados forem dependentes, usamos a ANOVA para medidas repetidas, cuja função no R é *aov()*, no caso paramétrico, e o teste de Friedman, função *friedman.test()*, no caso não-paramétrico.

A Hipótese de nulidade para estes testes é a de que todas as médias das categorias analisadas são iguais. Caso a rejeitamos, concluímos que pelo menos uma dessas médias é estatisticamente diferente das demais, mas não sabemos qual (is). Para identificá-la (s), precisamos realizar um teste de comparações múltiplas. No caso paramétrico, o mais utilizado é o teste de Tukey, cuja função no R é *TukeyHSD()*. No caso não-paramétrico, destaca-se o teste de Nemenyi, função *posthoc.kruskal.nemenyi.test()*.

2.4.1 Exemplo no R

Imagine agora que o professor da disciplina Fisiologia queira comparar o rendimento de três grupos de estudantes de sua turma: Nutrição, Educação Física e Fisioterapia. Para isto, o professor selecionou uma amostra de 25 estudantes de cada curso, representada pela Tabela 9 e considerou-se que todos os pressupostos para a realização dos testes paramétricos foram satisfeitos.

Tabela 9: Notas de três turmas de Fisiologia.

Nutrição	20, 25, 65, 50, 63, 60, 65, 11, 65, 23, 28, 37, 86, 54, 22, 62, 44, 35, 44, 46, 50, 60, 52, 26, 73.
Educação Física	55, 60, 30, 55, 75, 85, 76, 62, 25, 54, 86, 77, 74, 28, 49, 57, 63, 66, 84, 58, 60, 60, 75, 45, 40
Fisioterapia	10, 13, 25, 42, 65, 44, 30, 62, 25, 54, 24, 77, 76, 28, 49, 37, 63, 44, 40, 15, 22, 75, 60, 45, 40.

Vamos construir uma ANOVA para verificarmos se há algum grupo diferente dos demais. Para isto, usaremos a função *aov*. Observe que, para criar o conjunto de dados e organizá-los em vetor, usamos os comandos *data.frame()* e *stack*.

```
> nut=c(20, 25, 65, 50, 63, 60, 65, 11, 65, 23, 28, 37,
+ 86, 54, 22, 62, 44, 35, 44, 46, 50, 60, 52, 26, 73)
> edfis=c(55, 60, 30, 55, 75, 85, 76, 62, 25, 54, 86, 77,
+ 74, 28, 49, 57, 63, 66, 84, 58, 60, 60, 75, 45, 40)
> fisio=c(10, 13, 25, 42, 65, 44, 30, 62, 25, 54, 24, 77,
+  76, 28, 49, 37, 63, 44, 40, 15, 22, 75, 60, 45, 40)
> dados2<-data.frame(nut,edfis,fisio)
> dat<-stack(dados2)
> anova=aov(dat$values~dat$ind)
> summary(anova)

            Df Sum Sq Mean Sq F value  Pr(>F)
dat$ind      2   4126  2063.0   5.803 0.00461 **
Residuals   72  25597   355.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Como rejeitamos a hipótese de que todos os grupos apresentam a mesma média (p-valor $< 5\%$), precisamos agora identificar qual(is) grupo (s) apresenta (m) média (s) diferente (s) dos demais. Para isto, vamos utilizar o teste de Tukey, considerando a análise paramétrica.

```
> TukeyHSD(anova)

  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = dat$values ~ dat$ind)

$`dat$ind`
               diff         lwr      upr     p adj
edfis-nut     13.32   0.5574904 26.08251 0.0388156
fisio-nut     -4.04 -16.8025096  8.72251 0.7300593
fisio-edfis -17.36 -30.1225096 -4.59749 0.0048622
```

Observe que não há diferenças entre as médias dos grupos nutrição e fisioterapia. Entretanto, o grupo educação física apresenta média diferente dos demais. Uma análise através do gráfico Boxplot, conforme Figura 13, também auxilia nas interpretações.

```
>  boxplot(dat$values~dat$ind,
+ xlab="Curso", ylab="Notas", col=c("yellow","orange", "pink"))
```

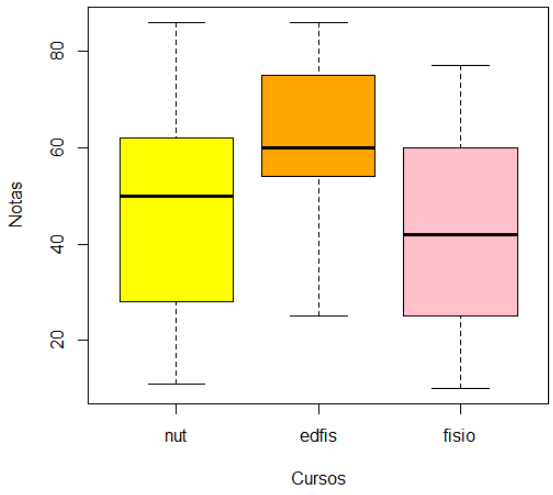

Figura 13: Boxplot para comparação de notas de Fisiologia.

Dessa forma, a análise descritiva, a análise pela ANOVA e a análise pelo teste de Tukey, nos

permitem concluir que os estudantes do curso de Educação Física possuem, em média, notas estatisticamente maiores na disciplina de Fisiologia.

A Figura 14 apresenta um resumo dos principais testes de comparação de médias.

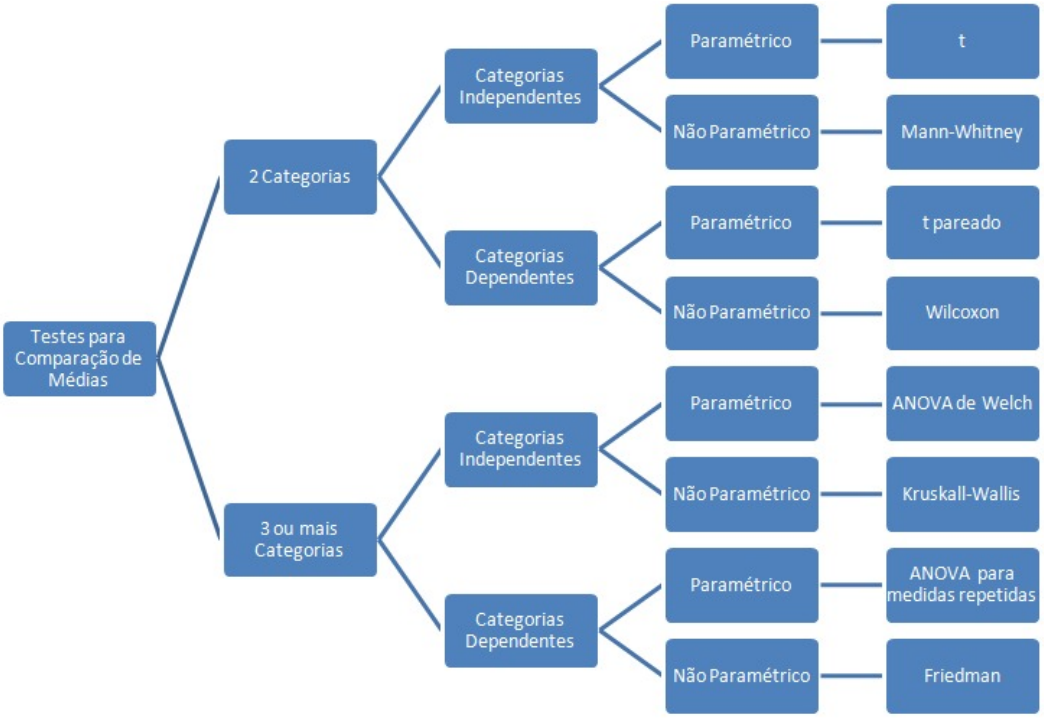

Figura 14: Esquema para testes de comparação de médias.

2.5 Testes de Associação

Em muitas pesquisas relacionadas às ciências sociais aplicadas é comum querer verificar se há dependência ou não entre duas variáveis qualitativas. Para verificar, por exemplo, a significância estatística da aparente associação entre sexo e alguma modalidade de esporte, escolhida para participação em aulas de educação física, podemos construir uma tabela de contingência, conforme Tabela 10. Observe que a maioria das meninas optam por fazer aulas de Vôlei. Já os meninos tem preferência para as aulas de futebol.

Tabela 10: Tabela de valores observados.

		Esporte			
		Futebol	Vôlei	Handbol	Total
Sexo	Feminino	20	45	35	100
	Masculino	60	25	15	100
	Total	80	70	50	200

Dos testes para associação, os mais utilizados são o teste Qui-Quadrado e o teste exato de

Fisher. A hipótese nula em ambos é de que não existe associação entre as variáveis.

A estatística do teste Qui-Quadrado é definida por:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{6}$$

em que n é o tamanho da amostra, O : frequência observada e E : frequência esperada. A estatística χ^2 é conhecida como estatística qui-quadrado.

Para testar a significância do coeficiente precisamos verificar o valor crítico da estatística χ^2. Em suma, um valor grande de χ^2 indica associação entre as variáveis.

Vamos calcular o valor da estatística χ^2 para o exemplo acima citado. A Tabela 11 representa os valores esperados para cada categoria.

Tabela 11: Tabela de valores esperados.

		Esporte			
		Futebol	Vôlei	Handbol	Total
Sexo	Feminino	$\frac{80\times100}{200} = 40$	$\frac{70\times100}{200} = 35$	$\frac{50\times100}{200} = 25$	100
	Masculino	$\frac{80\times100}{200} = 40$	$\frac{70\times100}{200} = 35$	$\frac{50\times100}{200} = 25$	100
	Total	80	70	50	200

Calculemos agora o valor da estatística qui-quadrado:

$$\chi^2 = \frac{(20-40)^2}{40} + \frac{(45-35)^2}{35} + \frac{(35-25)^2}{25} + \frac{(60-40)^2}{40} + \frac{(25-35)^2}{35} + \frac{(15-25)^2}{25} = 33,7143$$

Quando os valores esperados são menores do que 5 ou quando as amostras são pequenas, o teste χ^2 não apresenta resultados confiáveis. Nos casos em que conseguimos dispor os dados em uma tabela $2x2$, conforme Tabela 12, é preferível usar o teste exato de Fisher.

Tabela 12: Modelo de tabela para o teste exato de Fisher.

	Grupo 1	**Grupo 2**	**Total**
Grupo X	a	b	$a+b$
Grupo Y	c	d	$a+d$
Total	$a+c$	$c+d$	n

A probabilidade de ocorrência das frequências observadas, se faz com o uso da distribuição hipergeométrica:

$$P_a = P[X = a] = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}} = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!n!}$$

O teste exato de Fisher consiste na determinação das probabilidades de tabelas com as mesmas margens e com menores valores na entrada cujo valor, conforme ilustra a Figura 15.

a-1	b+1	a+b
c+1	d-1	c+d
a+c	b+d	n

a-2	b+2	a+b
c+2	d-2	c+d
a+c	b+d	n

▪ ▪ ▪

0	a+b	a+b
c+d	0	c+d
c+d	a+b	n

Figura 15: Esquema para teste exato de Fisher.

Rejeitamos a hipótese de independência entre as variáveis, se a soma $P_a + P_{a-1} + ... + P_0$ for inferior ao nível de significância α, adotado.

2.5.1 Exemplos no R

Vamos refazer o exemplo que verifica a relação de dependência entre as variáveis sexo e modalidade de esporte escolhida para participação em aulas de educação física, no R. Para isto, utilizaremos a função *chisq.test()*.

```
> M<-as.table(rbind(c(20,45,35),c(60,25,15)))
> chisq.test(M)

        Pearson's Chi-squared test

data:  M
X-squared = 33.714, df = 2, p-value = 4.776e-08
```

Como a hipótese de independência foi rejeitada, verificamos que as variáveis analisadas estão associadas.

Imagine agora que os dados da Tabela 8 fossem reduzidos e que só considerássemos como modalidade de esporte o futebol e o vôlei. A Tabela 13 retrata esta nova situação.

Tabela 13: Tabela reduzida de valores observados.

		Esporte		
		Futebol	Vôlei	Total
Sexo	Feminino	2	5	7
	Masculino	8	5	13
	Total	10	10	20

Neste caso, como há valores esperados menores do que 5 e a amostra é pequena, não seria prudente analisar pelo teste χ^2. Procederemos à analise pelo teste exato de Fisher, cuja função no R é *fisher.test()*.

```
> M<-as.table(rbind(c(2,5),c(8,5)))
> fisher.test(M)

        Fisher's Exact Test for Count Data

data:  M
p-value = 0.3498
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.01858382 2.47022220
sample estimates:
odds ratio
 0.2689102
```

Observe que, para a nova amostra, não há indícios para rejeitar a hipótese de independência entre as variáveis.

CAPÍTULO 3

CORRELAÇÃO E REGRESSÃO

3 Correlação e Regressão

Existem situações nas quais há interesse em estudar a relação entre duas ou mais variáveis. Tanto a correlação, quanto a regressão são técnicas que visam estimar esta possível relação, sendo que a primeira preocupa-se em quantificá-la e a segunda em matematizá-la.

O comportamento conjunto de duas variáveis quantitativas pode ser observado por meio do gráfico de dispersão, como o da Figura 4, em que é possível verificar uma relação linear entre as variáveis idade e tempo de serviço. Na Figura 16 traçamos uma reta para melhor visualização desta relação.

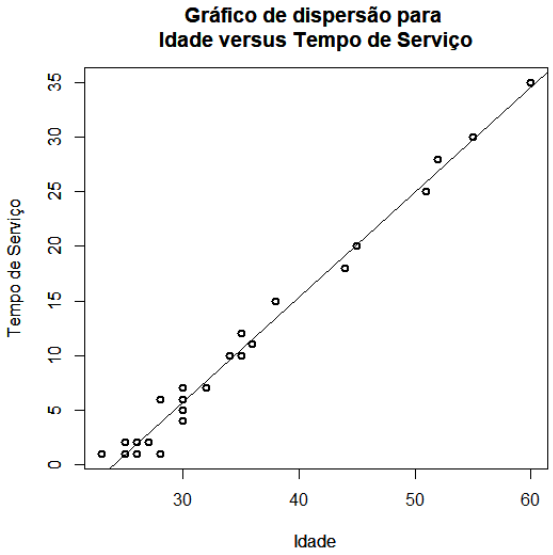

Figura 16: Análise da relação entre as variáveis Idade e Tempo de Serviço

O gráfico mostra que quanto maior a idade dos entrevistados maior o tempo de serviço. Em suma, quanto mais próximos de uma reta, maior a relação linear entre as variáveis.

3.1 Coeficientes de Correlação

Os coeficientes de correlação tem como objetivo a mensuração da intensidade e a direção da relação (linear ou não) entre duas variáveis. Existem vários coeficientes que quantificam esta intensidade. Nesta seção, apresentaremos os mais citados na literatura.

3.1.1 Coeficientes de Correlação para variáveis quantitativas

O coeficiente de correlação para variáveis quantitativas mais utilizado é o de Pearson, que pode ser calculado através da fórmula:

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{7}$$

em que $x_1, x_2, \cdots, x_n$ e $y_1, y_2, \cdots, y_n$ são os valores medidos de ambas as variáveis e $\bar{x}$ e $\bar{y}$ são as médias aritméticas dessas observações.

O coeficente de correlação de Pearson varia entre -1 e 1, sendo que, quanto mais perto dos extremos, maior é a correlação entre as variáveis. O sinal indica a direção, se a correlação é positiva ou negativa. Na Figura 17 temos um exemplo para correlação positiva (a), negativa (b) e nula (c).

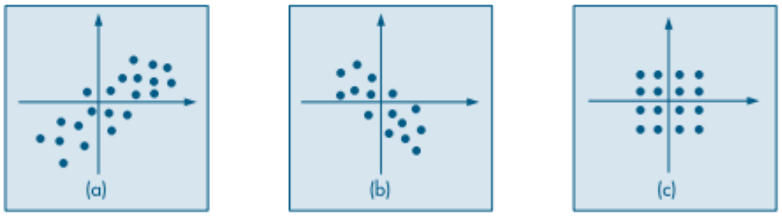

Figura 17: Exemplo de correlações entre duas variáveis
Fonte: Bussab & Morettin (2010, p. 83)

Para avaliar o grau de intensidade da correlação entre duas variáveis, Mello et al. (2011) apresentam a classificação apresentada na Tabela 14.

Tabela 14: Classificação para o Coeficiente de Correlação de Pearson.

Coeficiente de Correlação de Pearson	**Classificação**
0,00 a 0,19	Correlação bem fraca
0,20 a 0,39	Correlação fraca
0,40 a 0,69	Correlação moderada
0,70 a 0,89	Correlação forte
0,90 a 1,00	Correlação muito forte

Para se utilizar o coeficiente de correlação de Pearson, o relacionamento entre as variáveis deve ser linear, as variáveis envolvidas devem ser aleatórias e possuírem uma distribuição normal. Por conta desta última hipótese, este coeficiente de correlação é dito paramétrico, como visto no último capítulo. Para conjunto de dados não-paramétricos, ou que não possuam uma relação linear, mas monótona, isto é, se uma aumenta (ou diminuiu) a outra tende a aumentar (ou diminuir), podemos utilizar o coeficiente de correlação de Spearman (para amostras maiores) ou Kendall (para amostras menores).

Para o caso de variáveis dicotômicas, há algumas generalizações da correlação de Pearson. A correlação ponto-bisserial é recomendada quando uma das variáveis é genuinamente dicotômica, a correlação bisserial para quando uma das variáveis é artificialmente dicotomizada, a correlação phi quando as duas variáveis são genuinamente dicotômicas e a correlação tetracórica quando as duas variáveis forem artificialmente dicotomizadas. A Tabela 15 apresenta um resumo de quando usar cada tipo de correlação.

Tabela 15: Coeficientes de correlação e quando usá-los.

Coeficiente	**Variável X**	**Variável Y**
Pearson	Contínua	Contínua
Spearman ou Kendall	Ordinal	Contínua
Spearman ou Kendall	Ordinal	Ordinal
Bisserial	Artificialmente dicotômica	Contínua
Ponto-Bisserial	Genuinamente dicotômica	Contínua
Phi	Genuinamente dicotômica	Genuinamente dicotômica
Tetracórica	Artificialmente dicotômica	Artificialmente dicotômica

3.1.2 Coeficiente de Contingência para variáveis qualitativas

Entre as medidas que quantificam a associação entre variáveis qualitativas, apresentaremos o chamado coeficiente de contingência, introduzido por Pearson. A fórmula para o cálculo deste coeficiente é:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}, \tag{8}$$

em que n é o tamanho da amostra e χ^2 é a estatística qui-quadrado.

Como exemplo, calculemos o coeficiente de contingência para as variáveis sexo e modalidade de esporte, escolhida para a prática nas aulas de Educação Física, conforme Tabela 11. Neste caso, o coeficiente de contingência pode ser calculado da seguinte maneira:

$$C = \sqrt{\frac{33,7143}{33,7143 + 200}} \approx 0,1443$$

Para testar a significância do coeficiente podemos realizar o teste χ^2, como visto no capítulo anterior.

3.1.3 Exemplos no R

Para os coeficientes de correlação de Pearson, Spearman ou Kendall, o comando no R é o mesmo, mudando apenas nas opções extras do comando:

```
> cor.test(variável1, variável2, method = "pearson")
> cor.test(variável1, variável2, method = "spearman")
```

```
> cor.test(variável1, variável2, method = "kendal")
```

Estes comandos também retornam um teste de significância para o coeficiente de correlação. Caso o valor-p retornado seja maior do que o nível de significância adotado (usualmente 5%), então não há correlação significativa entre as variáveis.

O pacote *psych* (Revelle (2014)) do R oferece comandos para o cálculo dos coeficientes de correlação Bisserial, Phi e Tetracório. São eles, respectivamente:

```
> biserial(variável1,variável2)
> phi(matriz)
> tetrachoric(matriz)
```

Já a correlação Ponto-Bisserial pode ser obtida pelo pacote *ltm* (Rizopoulos (2006)), cujo comando é:

```
> biserial.cor(variável1,variável2)
```

Para ilustração, imaginemos que os pressupostos para o cálculo do coeficiente de correlação de Pearson foram atendidos e calculemos a correlação entre as variáveis Idade e Renda, da Tabela 1.

```
> cor.test(dados$Idade,dados$Renda,method="pearson")

        Pearson's product-moment correlation

data:  dados$Idade and dados$Renda
t = 2.7265, df = 28, p-value = 0.01092
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1170822 0.7023950
sample estimates:
      cor
0.4580311
```

Este resultado aponta para uma correlação moderada e significativa (a 5%) entre as variáveis.

3.2 Modelos de Regressão Linear

Na Figura 16 traçamos uma reta para visualizar o comportamento linear entre as variáveis Idade e Tempo de Serviço, da Tabela 1. Podemos encontrar a equação dessa reta e, através dela, fazer previsões acerca do comportamento dessas variáveis. Este processo é chamado de regressão linear. Para o exemplo dado, a variável Tempo de Serviço é dependente da variável independente Idade.

3.2.1 Regressão Linear Simples e Múltipla

Uma regressão é dita linear quando a curva ajustada é uma reta. Caso contrário, a regressão é dita não-linear. Caso haja a presença de apenas uma variável independente, a regressão linear é dita simples. Entretanto, na maioria dos problemas, para explicar uma variável dependente é necessário mais do que uma variável independente. Nestes casos, a regressão linear é dita múltipla.

O modelo de regressão linear simples é:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \text{para} \;\; i = 1, \ldots, n \tag{9}$$

em que Y_i representa os valores da variável dependente, x_i representa os valores da variável independente (também chamado de variável explicativa ou regressora), ϵ_i representa o erro experimental aleatório, β_0 e β_1 representam os coeficientes do modelo, que serão estimados, e que definem a reta de regressão e n é o tamanho da amostra.

O modelo estimado para representar a relação entre as variáveis idade e tempo de serviço, representado na Figura 16 é:

$$\widehat{y}_i = -23,1103 + 0,9611 \; x_{1i}, \qquad i = 1, \cdots 30$$

Com esse modelo podemos prever, por exemplo, o tempo médio de serviço de um (a) professor (a) com 35 anos de idade, que será indicado por $\widehat{y}(35)$:

$$\widehat{y}(35) = -23,1103 + 0,9611 \times 35 = 10,5282$$

A vantagem de matematizar esta relação está na possibilidade de fazer estimativas para dados não observáveis.

Para os modelos de regressão linear múltipla, temos a inclusão de variáveis independentes $x_{2i}, x_{3i}, \cdots x_{ki}$. Dessa forma, temos o modelo:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} \varepsilon_i, \quad \text{para} \;\; i = 1, \ldots, n \tag{10}$$

em que $\beta_2, \beta_3 \cdots \beta_k$ são os coeficientes das variáveis independentes $x_{2i}, x_{3i}, \cdots x_{ki}$, respectivamente, e os demais termos são igualmente definidos na Equação 9.

Americo e Lacruz (2017) descrevem a relação entre o "contexto" e o "desempenho" escolar das escolas estaduais do Espírito Santo, considerando as notas obtidas na Prova Brasil em 2013, por meio de regressão linear múltipla. Neste trabalho, os autores concluem que a permanência do docente em uma mesma escola tem impacto positivo nas notas, e que um aumento na taxa de abandono produziria um efeito negativo. Frente a estas conclusões, os autores discutem a construção de políticas públicas educacionais para solucionar os problemas evidenciados.

A Figura 18 mostra os coeficientes do modelo obtidos no processo de regressão que considerou a nota na Prova Brasil (NPB) como variável dependente e o Índice de Regularidade Docente (IRD), Indicador de Esforço Docente (IED) e Taxa de Abandono (TA) como variáveis independentes no

modelo estimado.

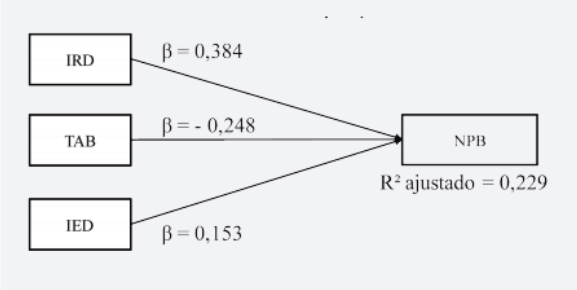

Figura 18: Relação entre"contexto" e "desempenho" escolar em escolas estaduais do Espírito Santo.
Fonte: Americo & Lacruz (2017, p. 868)

O coeficiente R^2 exibido na Figura 18 expressa o coeficiente de determinação, que mensura o quanto o modelo consegue explicar os valores observados. Em suma, trata-se do quadrado do coeficiente de correlação de Pearson e varia entre 0 e 1, sendo que, quanto maior, melhor o ajustamento do modelo aos dados.

3.2.2 Análise dos Resíduos

Para verificar a adequabilidade de um modelo de regressão linear simples ou múltipla, é preciso validar as suposições do modelo ajustado para que os resultados obtidos sejam confiáveis. Para isto, podemos nos basear nos resíduos gerados e realizar o que chamamos de Análise dos Resíduos.

A análise de resíduos consiste em um conjunto de técnicas para verificar os pressupostos de normalidade, homocedasticidade e independência dos resíduos. Também deve-se verificar, para a significância do modelo, a linearidade e a ausência de pontos atípicos influentes, conhecidos como *outliers*.

No caso da Regressão Linear Múltipla, além desses pressupostos, precisamos também verificar a multicolinearidade, que ocorre no caso em que as variáveis são altamente correlacionadas, provocando efeitos nas estimativas dos coeficientes do modelo de regressão e, consequentemente, na aplicabilidade geral do modelo estimado.

Na próxima seção apresentaremos alguns meios de verificação destes pressupostos. Caso o leitor deseje aprofundar-se em seus estudos, poderá consultar obras clássicas como Gujarati e Porter (2011), Hair Júnior et. al (2010) e Bussab e Morettin (2010).

3.2.3 Exemplos no R

O R apresenta um agrupamento dos principais pacotes relacionados a Regressão. Alguns deles, como o *Econometrics*, possuem funções para estimação e análises dos modelos da regressão. Para instalá-los simultaneamente basta digitar no *R Console*:

```
> install.packages("ctv")
> library(ctv)
> install.views("Econometrics")
```

Vamos verificar se existe uma relação linear, considerando o Tempo de Serviço como variável dependente (y) e a Idade (x_1) e o Salário (x_2) como variáveis independentes, na Tabela 1. Para isto, usaremos a função *lm()* do software R, cujo formato é ($y \sim x_1 + x_2$). A Figura **??** representa o retorno do R.

```
> lm(dados$TempodeServico~dados$Idade+dados$Renda)

Call:
lm(formula = dados$TempodeServico ~ dados$Idade + dados$Renda)

Coefficients:
(Intercept)  dados$Idade  dados$Renda
   -22.7653       0.9694      -0.1829
```

O R estima o valor dos coeficientes $\widehat{\beta_0}$ (intercepto), $\widehat{\beta_1}$ (da variável Idade) e $\widehat{\beta_2}$) da variável Renda, através do Método de Mínimos Quadrados, método estatístico que consiste na minimização dos erros do modelo estimado. Para o exemplo em questão, a equação da reta ajustada é dada por $\widehat{y} = -22,7653 + 0,9694\ x_1 - 0,1829\ x_2$. Para encontrar medidas descritivas para analisar o ajuste dos dados, podemos utilizar a função *summary*:

```
> summary(lm(dados$TempodeServico~dados$Idade+dados$Renda))

Call:
lm(formula = dados$TempodeServico ~ dados$Idade + dados$Renda)

Residuals:
     Min       1Q   Median       3Q      Max
-2.82876 -0.60677  0.02154  0.92114  2.21696

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -22.76526    0.93955 -24.230   <2e-16 ***
dados$Idade   0.96938    0.02553  37.972   <2e-16 ***
dados$Renda  -0.18289    0.25901  -0.706    0.486
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.194 on 27 degrees of freedom
```

```
Multiple R-squared:  0.9852,        Adjusted R-squared:  0.9841
F-statistic: 897.1 on 2 and 27 DF,  p-value: < 2.2e-16
```

Para verificar a significância da regressão e dos coeficientes individualmente, podemos utilizar os testes F e t, respectivamente, exibidos na Figura **??**, verificando os valores-p exibidos. Observe que o coeficiente da variável salário não foi significativo no modelo (valor-p > nível de significância), portanto, pode ser retirado da equação ajustada. Dessa forma, caímos em um processo de regressão linear simples, em que a variável dependente é o Tempo de Serviço (y) e a independente é a Idade (x):

```
> lm(dados$TempodeServico~dados$Idade)

Call:
lm(formula = dados$TempodeServico ~ dados$Idade)

Coefficients:
(Intercept)  dados$Idade
   -23.1103       0.9611
```

Conforme expressa pela Figura **??**, a equação da reta ajustada é dada por $\widehat{y} = -23,1103 + 0,9611\ x$. Verifiquemos, agora, a significância do modelo e de seus coeficientes:

```
> summary(lm(dados$TempodeServico~dados$Idade))

Call:
lm(formula = dados$TempodeServico ~ dados$Idade)

Residuals:
    Min      1Q  Median      3Q     Max
-2.8012 -0.7235  0.1016  0.9223  2.1988

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -23.11028    0.79528  -29.06   <2e-16 ***
dados$Idade   0.96112    0.02249   42.74   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.183 on 28 degrees of freedom
Multiple R-squared:  0.9849,        Adjusted R-squared:  0.9844
F-statistic:  1826 on 1 and 28 DF,  p-value: < 2.2e-16
```

A Figura **??** confirma esta significância e exibe, assim como a Figura **??**, os erros-padrão das

estimativas dos coeficientes de regressão e o coeficiente de determinação (R^2), que para este caso foi de $0,9844$, o que indica um bom ajustamento dos dados ao modelo.

Vamos, agora, verificar se as pressuposições do modelo de regressão foram atendidas. Para avaliar a homocedasticidade (variância constante) dos resíduos, podemos construir os gráficos para os valores ajustados da variável dependente e variável independente em função dos resíduos. Para isto, usaremos os comandos *fitted()* e *residuals()*, conforme ilustram as linhas de comando abaixo e as Figuras 19 e 20, geradas pelo software.

```
> dadosajustados=lm(dados$TempodeServico~dados$Idade)
> plot(fitted(dadosajustados),residuals(dadosajustados))
> abline(h=0)
> plot(dados$Idade,residuals(dadosajustados))
> abline(h=0)
```

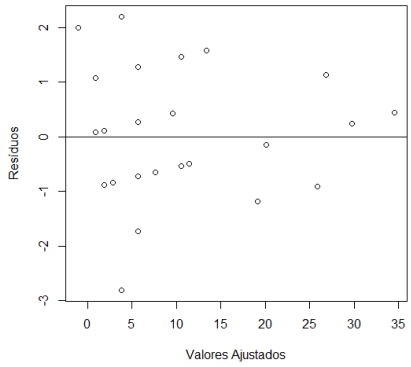

Figura 19: Resíduos versus valores ajustados.

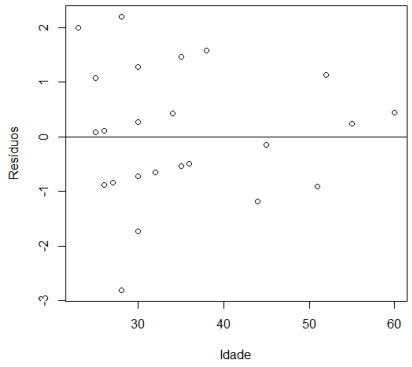

Figura 20: Resíduos versus variável independente.

Note que, para melhor visualização dos comandos, nomeamos o modelo como "dadosajustados" e plotamos a reta $y = 0$. Os gráficos plotados pelo software, não apresentam nenhum comportamento ou tendência. Assim, temos indícios de que o pressuposto da homogeneidade da variância dos resíduos foi atendido (Gráfico 19), assim como da independência (Gráfico 20).

Para verificar o pressuposto da normalidade dos resíduos, podemos construir o gráfico de probabilidade normal dos resíduos, cujo comando no R é *qqnorm()*:

```
> qqnorm(residuals(dadosajustados),
+ ylab="Resíduos",xlab="Quantis teóricos normais")
> qqline(residuals(dadosajustados))
```

A Figura 21 exibe o gráfico gerado.

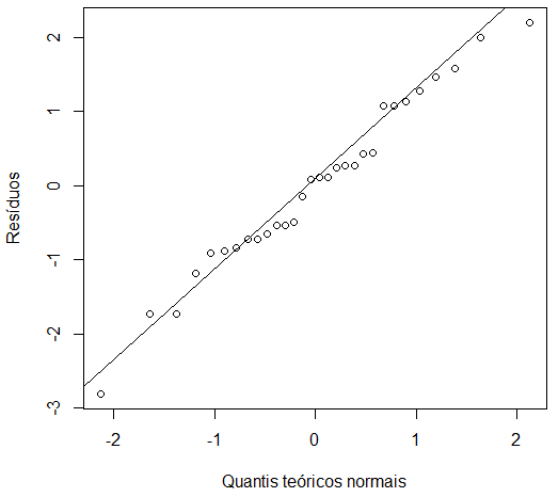

Figura 21: Análise gráfica da normalidade dos resíduos.

Observe que os pontos do se dispõem em torno de uma linha, o que nos permite considerar que possuem uma distribuição normal. Uma outra forma de verificar o pressuposto da normalidade dos resíduos é através do teste de Shapiro-Wilk, cujo comando no R é *shapiro.test()*. A Figura **??** mostra o retorno do software para este comando:

```
> shapiro.test(residuals(dadosajustados))

        Shapiro-Wilk normality test

data:  residuals(dadosajustados)
W = 0.97966, p-value = 0.8165
```

Neste caso, como o valor-p retornado é maior que o nível de significância adotado (5%), logo, aceita-se a hipótese de normalidade dos resíduos.

3.3 Regressão Logística Binária

Assim como na regressão linear, os modelos de regressão logística tratam de técnicas que permitem explicar a relação entre uma variável dependente, e um conjunto de variáveis independentes. O que difere uma da outra é que, no caso da regressão logística, a variável dependente é binária (dicotômica), comumente classificada como sucesso ($y = 1$) ou fracasso ($y = 0$) e segue uma distribuição binomial de probabilidade.

Nos modelos de regressão logística binária, as variáveis independentes podem ser métricas ou categóricas. Neles, é possível verificar a probabilidade de ocorrência de um evento e o quanto cada variável do modelo influencia no resultado da análise.

3.3.1 Modelo logístico

A equação da regressão logística pode ser escrita da seguinte forma:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1\ x_1 + \beta_2\ x_2 + \cdots + \beta_n\ x_n \tag{11}$$

em que p representa a probabilidade de êxito quando a variável preditiva é x e $\beta_0, \beta_1, \cdots \beta_n$ são os coeficientes de regressão, estimados com métodos de máxima verossimilhança.

Com algumas manipulações algébricas, a Equação 10 pode ser escrita da seguinte forma:

$$p = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_n\ x_n}}{1+e^{\beta_0+\beta_1 x_1+\cdots+\ \beta_n x_n}} = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots+\ \beta_n x_n)}} \tag{12}$$

Em suma, utiliza-se este modelo para encontrar a probabilidade de estar em uma categoria, baseado na combinação de variáveis independentes. Na Figura 22, temos a representação gráfica da regressão logística, cujo formato é de uma curva sigmoide.

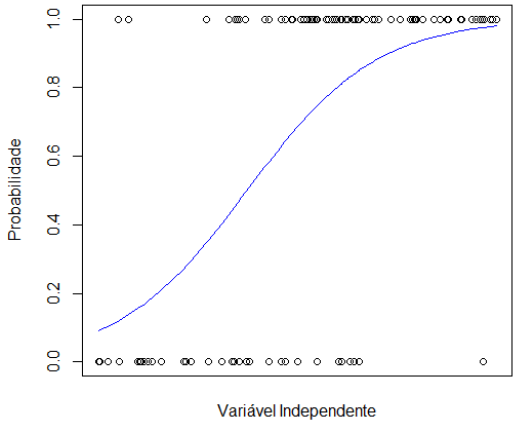

Figura 22: Curva sigmoide para representar um exemplo de regressão logística binária.

Um conceito bastante utilizado na regressão logística é o de razão de chances (*odds ratio*). A chance é definida como a razão entre a probabilidade de ocorrência de um evento pela sua não ocorrência. A razão de chances é definida como a razão entre a chance de um evento ocorrer em um determinado grupo e a chance de ocorrer em outro grupo.

Em termos práticos, podemos interpretar a razão de chances como sendo o aumento estimado na probabilidade de sucesso associado a mudança em uma unidade no valor da variável independente. Quando se altera em d unidades a variável independente, a razão de chances tem um aumento estimado em $e^{d\beta}$.

3.3.2 Teste de significância dos coeficientes e de qualidade de ajuste

Para verificar a significância do modelo estimado, podemos realizar o teste de razão de verossimilhanças cuja hipótese de nulidade é a que todos os coeficientes do modelo são nulos. Para isto, este teste compara a diferença entre o logaritmo da função de verossimilhança do modelo completo, com o logaritmo da verossimilhança do modelo sem a variável analisada, conhecido como estatística G:

$$G = -\ln\left[\frac{\text{verossimilhança do modelo sem a variável}}{\text{verossimilhança do modelo com a variável}}\right] \tag{13}$$

Para testar a significância de cada coeficiente do modelo, separadamente, podemos realizar o teste de Wald. Trata-se de uma generalização do teste t de Student, cuja estatística W é dada por:

$$W = \frac{\widehat{\beta}_i}{\widehat{EP}(\widehat{\beta}_i)} \tag{14}$$

em que β_i é o coeficiente testado e EP é o seu respectivo erro-padrão.

Para verificar se o modelo está bem ajustado, podemos utilizar o Teste de Hosmer & Lemeshow, cuja hipótese de nulidade é a de que o modelo se ajusta bem aos dados. Caso rejeitada essa hipótese, o modelo não é capaz de produzir estimativas e classificações confiáveis.

Uma outra medida para qualidade de ajuste é através do pseudo-R^2, que é similar ao coeficiente de determinação obtido nos modelos de regressão linear e cujos valores também estão entre 0 e 1, sendo que quanto mais perto de 1 melhor o ajuste do modelo.

3.3.3 Exemplo no R

Como exemplo, iremos refazer as análises presentes no artigo *"Um modelo estatístico para gestão de programas de pós-graduação"*, de autoria de Mesquita e Nogueira (2015), cujo objetivo foi o de estimar a probabilidade de obtenção de melhores conceitos CAPES em programas de pós-graduação, bem como indicar as variáveis mais relevantes para esta melhoria, baseados em um modelo de regressão logística binária.

Para isto, foram amostrados o desempenho de 540 programas de pós-graduação na avaliação trienal de 2013, cujos conceitos conceitos Muito Bom (MB), Bom (B), Regular (R), Fraco (F) ou Deficiente (D) são atribuídos aos itens e respectivos quesitos da ficha de avaliação organizados na Tabela 16, utilizada para atribuir uma nota de 3 a 7 para o programa.

Tabela 16: Quesitos e critérios de avaliação da CAPES.

	I- Proposta do Programa
x_1	Coerência, consistência, abrangência e atualização das áreas de concentração, linhas de pesquisa, projetos e proposta curricular.
x_2	Planejamento do programa com vistas a seu desenvolvimento futuro, contemplando os desafios internacionais da área na produção do conhecimento.
x_3	Infraestrutura para ensino, pesquisa e, se for o caso, extensão.
	II - Corpo docente
x_4	Perfil do corpo docente, consideradas titulação, diversificação na origem da formação, aprimoramento e experiência, e sua compatibilidade e adequação à proposta do programa.
x_5	Adequação e dedicação dos docentes permanentes em relação às atividades de pesquisa e de formação do programa.
x_6	Distribuição das atividades de pesquisa e de formação entre os docentes do programa.
x_7	Contribuição dos docentes para atividades de ensino e/ou de pesquisa na graduação, com atenção tanto à repercussão que este item pode ter na formação de futuros ingressantes na pós-graduação, quanto na formação de profissionais mais capacitados no plano da graduação.
	III - Corpo discente, teses e dissertações
x_8	Quantidade de teses e dissertações defendidas no período de avaliação, em relação ao corpo docente permanente e à dimensão do corpo docente.
x_9	Distribuição das orientações das teses e dissertações defendidas no período de avaliação em relação aos docentes do programa.
x_{10}	Qualidade das teses e dissertações e da produção de discentes autores da pós-graduação e da graduação na produção científica do programa, aferida por publicações e outros indicadores pertinentes à área.
x_{11}	Eficiência do programa na formação de mestres e doutores bolsistas: Tempo de formação de mestres e doutores e percentual de bolsistas titulados.
	IV - Produção intelectual
x_{12}	Publicações qualificadas do programa por docente permanente.
x_{13}	Distribuição de publicações qualificadas em relação ao corpo docente permanente do programa.
x_{14}	Produção técnica, patentes e outras publicações consideradas relevantes.
	V - Inserção social
x_{15}	Inserção e impacto regional e (ou) nacional do programa.
x_{16}	Integração e cooperação com outros programas e centros de pesquisa e desenvolvimento profissional, relacionados à área de conhecimento do programa, com vistas ao desenvolvimento da pesquisa e da pós graduação.
x_{17}	Visibilidade ou transparência dada pelo programa à sua atuação.

Fonte: Mesquita e Nogueira (2015)

O desempenho dos programas de pós-graduação amostrados, oriundos da base de dados da CAPES, foram tabulados em uma planilha do Excel, conforme ilustra a Figura 23.

Amostra	IES	PROGRAMA DE POS GRADUAÇÃO	Nota	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17
1	UFES	CIÊNCIAS SOCIAIS	3	B	R	B	B	MB	B	MB	F	B	B	B	F	R	B	F	B	B
2	UFJF	ENGENHARIA ELÉTRICA	4	MB	MB	MB	MB	MB	B	B	R	MB	MB	MB	MB	MB	MB	MB	MB	MB
3	UEM	ENGENHARIA MECÂNICA	3	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
4	UNB	GEOCIÊNCIAS APLICADAS	4	B	B	B	MB	R	B	B	B	B	R	MB	B	R	MB	MB	B	R
5	UNIPAR	CIÊNCIA ANIMAL	3	MB	B	MB	MB	MB	B	MB	R	MB	R	B	R	B	MB	B	B	MB
6	UEL	SERVIÇO SOCIAL E POLÍTICA SOCIAL	4	MB	MB	MB	B	MB	B	MB	MB	MB	B	MB	F	B	MB	MB	MB	B
7	UFRA	AGRONOMIA	4	MB	MB	MB	MB	MB	MB	MB	MB	MB	B	MB	B	MB	R	MB	MB	MB
8	UFSC	BIOLOGIA CELULAR E DO DESENVOLVIMENTO	4	MB	B	MB	B	B	B	MB	MB	R	R	MB	B	R	R	B	F	MB
9	UFPE	ENGENHARIA DE PRODUÇÃO	6	MB	MB	B	MB	MB	MB	MB	MB	MB	MB	MB	MB	MB	MB	MB	MB	MB
10	USP/SC	ENGENHARIA MECÂNICA	6	MB	MB	B	MB	MB	MB	MB	B	MB	MB	MB	MB	MB	MB	MB	MB	MB
11	UFPR	CIÊNCIAS BIOLÓGICAS (ENTOMOLOGIA)	6	MB	MB	MB	MB	MB	MB	MB	MB	MB	MB	MB	MB	MB	MB	MB	MB	MB
12	PUC/MG	LETRAS	5	B	B	MB	R	B	MB	MB	MB	B	B	MB	B	MB	MB	MB	MB	MB

Figura 23: Desempenho dos programas de pós-graduação - Fragmento do Excel.

Para construção do modelo proposto, os dados exibidos na Figura 23, foram codificados conforme a Tabela 17.

Tabela 17: Critérios para codificação das variáveis

	Antes	**Depois**
Nota da avaliação 2013 (y_i)	≥ 4	1
	≤ 4	0
Item avaliado (x_i)	Muito Bom "MB"	5
	Bom "B"	4
	Regular "R"	3
	Fraco "F"	2
	Deficiente "D"	1

Fonte: Mesquita e Nogueira (2015)

Para a leitura dos dados codificados no Excel, utilizou-se o comando *read.table()*.

```
> dados2=read.table("C:/Users/Usuario/Desktop/UFES/dados2.txt",head=T)
```

O comando *head()* exibe o cabeçalho dos dados codificados lidos pelo software.

```
> head(dados2)
```

```
  Nota x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13 x14 x15 x16 x17
1    0  4  3  4  4  5  4  5  2  4   4   4   2   3   4   2   4   4
2    1  5  5  5  5  5  4  4  3  5   5   5   5   5   5   5   5   5
3    0  3  3  3  3  3  3  3  3  3   3   3   3   3   3   3   3   3
4    1  4  4  4  5  3  4  4  4  4   3   5   4   3   5   5   4   3
5    0  5  4  5  5  5  4  5  3  5   3   4   3   4   5   4   4   5
6    1  5  5  5  4  5  4  5  5  5   4   5   2   4   5   5   5   4
```

Para compor o modelo, considerando todas as variáveis, usaremos a função *glm*. Para resumir os dados e fazer uma análise inicial, usaremos a função *summary()*, conforme ilustra a Figura **??**.

```
> modelo1=glm(Nota~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11+
+ x12+x13+x14+x15+x16+x17,family=binomial(link="logit"))
> summary(modelo1)

Call:
glm(formula = Nota ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 +
    x9 + x10 + x11 + x12 + x13 + x14 + x15 + x16 + x17, family = binomial(link = "logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.6175   0.0249   0.0476   0.1964   2.3325

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -25.81148    3.39473  -7.603 2.88e-14 ***
x1            0.31849    0.35114   0.907  0.36440
x2            0.14835    0.39617   0.374  0.70806
x3            0.15694    0.37051   0.424  0.67186
x4            0.72507    0.31793   2.281  0.02257 *
x5           -0.13541    0.31480  -0.430  0.66709
x6           -0.37612    0.27345  -1.375  0.16899
x7            0.02768    0.25109   0.110  0.91223
x8            1.11846    0.26734   4.184 2.87e-05 ***
x9            0.25011    0.26280   0.952  0.34124
x10           1.42358    0.24890   5.719 1.07e-08 ***
x11          -0.13289    0.35237  -0.377  0.70609
x12           1.58393    0.30176   5.249 1.53e-07 ***
x13           0.73632    0.23153   3.180  0.00147 **
x14           0.21002    0.21999   0.955  0.33974
x15           0.63219    0.38371   1.648  0.09943 .
x16           0.05273    0.33953   0.155  0.87658
x17           0.02089    0.28279   0.074  0.94111
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 574.58  on 539  degrees of freedom
Residual deviance: 180.91  on 522  degrees of freedom
AIC: 216.91
```

```
Number of Fisher Scoring iterations: 7
```

As variáveis significativas para compor o modelo, considerando 5% de significância, estão acompanhadas por pelo menos um asterisco na Figura **??**. Entretanto, nem todas as variáveis entram neste grupo. Para uma melhor seleção dessas variáveis, utilizaremos o método *stepwise*, através da função *step* do pacote stats. A Figura **??** ilustra parte do retorno do R para esta função.

```
> step(modelo1, data=dados2)

Start:  AIC=216.91
Nota ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11 +
    x12 + x13 + x14 + x15 + x16 + x17

       Df Deviance    AIC
- x17   1   180.91 214.91
- x7    1   180.92 214.92
- x16   1   180.93 214.93
- x2    1   181.05 215.05
- x11   1   181.05 215.05
- x3    1   181.09 215.09
- x5    1   181.09 215.09
- x1    1   181.74 215.74
- x14   1   181.81 215.81
- x9    1   181.83 215.83
- x6    1   182.81 216.81
<none>      180.91 216.91
- x15   1   183.70 217.70
- x4    1   186.30 220.30
- x13   1   191.68 225.68
- x8    1   201.05 235.05
- x12   1   216.51 250.51
- x10   1   222.93 256.93

Step:  AIC=214.92
Nota ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11 +
    x12 + x13 + x14 + x15 + x16

       Df Deviance    AIC
- x7    1   180.93 212.93
- x16   1   180.94 212.94
- x2    1   181.05 213.05
- x11   1   181.06 213.06
```

```
- x3    1   181.09 213.09
- x5    1   181.10 213.10
- x1    1   181.81 213.81
- x14   1   181.82 213.82
- x9    1   181.83 213.83
- x6    1   182.82 214.82
<none>      180.91 214.91
- x15   1   183.81 215.81
- x4    1   186.30 218.30
- x13   1   191.70 223.70
- x8    1   201.24 233.24
- x12   1   216.55 248.55
- x10   1   222.93 254.93

Step:  AIC=212.93
Nota ~ x1 + x2 + x3 + x4 + x5 + x6 + x8 + x9 + x10 + x11 + x12 +
    x13 + x14 + x15 + x16

       Df Deviance    AIC
- x16   1   180.96 210.96
- x11   1   181.06 211.06
- x2    1   181.07 211.07
- x5    1   181.10 211.10
- x3    1   181.12 211.12
- x9    1   181.84 211.84
- x1    1   181.87 211.87
- x14   1   181.87 211.87
- x6    1   182.88 212.88
<none>      180.93 212.93
- x15   1   183.83 213.83
- x4    1   186.37 216.37
- x13   1   191.77 221.77
- x8    1   201.27 231.27
- x12   1   216.89 246.89
- x10   1   223.38 253.38

Step:  AIC=210.96
Nota ~ x1 + x2 + x3 + x4 + x5 + x6 + x8 + x9 + x10 + x11 + x12 +
    x13 + x14 + x15

       Df Deviance    AIC
```

```
- x11   1   181.09 209.09
- x2    1   181.11 209.11
- x5    1   181.12 209.12
- x3    1   181.17 209.17
- x1    1   181.87 209.87
- x9    1   181.88 209.88
- x14   1   181.90 209.90
- x6    1   182.88 210.88
<none>      180.96 210.96
- x15   1   184.85 212.85
- x4    1   186.47 214.47
- x13   1   191.84 219.84
- x8    1   201.30 229.30
- x12   1   216.89 244.89
- x10   1   224.70 252.70

Step:  AIC=209.09
Nota ~ x1 + x2 + x3 + x4 + x5 + x6 + x8 + x9 + x10 + x12 + x13 +
    x14 + x15

       Df Deviance    AIC
- x2    1   181.21 207.21
- x5    1   181.25 207.25
- x3    1   181.29 207.29
- x1    1   182.03 208.03
- x9    1   182.06 208.06
- x14   1   182.06 208.06
- x6    1   183.03 209.03
<none>      181.09 209.09
- x15   1   185.17 211.17
- x4    1   186.56 212.56
- x13   1   191.88 217.88
- x8    1   202.40 228.40
- x12   1   217.38 243.38
- x10   1   224.70 250.70

Step:  AIC=207.21
Nota ~ x1 + x3 + x4 + x5 + x6 + x8 + x9 + x10 + x12 + x13 + x14 +
    x15

       Df Deviance    AIC
```

```
- x5    1   181.38 205.38
- x3    1   181.52 205.52
- x9    1   182.22 206.22
- x14   1   182.25 206.25
- x1    1   182.67 206.67
- x6    1   183.08 207.08
<none>      181.21 207.21
- x15   1   185.74 209.74
- x4    1   186.77 210.77
- x13   1   192.34 216.34
- x8    1   203.45 227.45
- x12   1   217.42 241.42
- x10   1   225.34 249.34

Step:  AIC=205.38
Nota ~ x1 + x3 + x4 + x6 + x8 + x9 + x10 + x12 + x13 + x14 +
    x15

       Df Deviance    AIC
- x3    1   181.65 203.65
- x9    1   182.28 204.28
- x14   1   182.45 204.45
- x1    1   182.77 204.77
- x6    1   183.22 205.22
<none>      181.38 205.38
- x15   1   185.75 207.75
- x4    1   186.81 208.81
- x13   1   192.35 214.35
- x8    1   204.87 226.87
- x12   1   217.81 239.81
- x10   1   225.36 247.36

Step:  AIC=203.65
Nota ~ x1 + x4 + x6 + x8 + x9 + x10 + x12 + x13 + x14 + x15

       Df Deviance    AIC
- x9    1   182.41 202.41
- x14   1   182.90 202.90
- x6    1   183.33 203.33
- x1    1   183.50 203.50
<none>      181.65 203.65
```

```
- x15   1   187.44 207.44
- x4    1   187.61 207.61
- x13   1   192.76 212.76
- x8    1   206.96 226.96
- x12   1   218.73 238.73
- x10   1   225.96 245.96

Step:  AIC=202.4
Nota ~ x1 + x4 + x6 + x8 + x10 + x12 + x13 + x14 + x15

       Df Deviance    AIC
- x6    1   183.56 201.56
- x14   1   183.91 201.91
<none>      182.41 202.41
- x1    1   184.63 202.63
- x15   1   187.83 205.83
- x4    1   188.61 206.61
- x13   1   194.23 212.23
- x8    1   211.77 229.77
- x12   1   218.73 236.73
- x10   1   226.91 244.91

Step:  AIC=201.56
Nota ~ x1 + x4 + x8 + x10 + x12 + x13 + x14 + x15

       Df Deviance    AIC
- x14   1   184.97 200.97
- x1    1   185.39 201.39
<none>      183.56 201.56
- x15   1   188.56 204.56
- x4    1   189.25 205.25
- x13   1   194.53 210.53
- x8    1   212.11 228.11
- x12   1   218.75 234.75
- x10   1   227.20 243.20

Step:  AIC=200.97
Nota ~ x1 + x4 + x8 + x10 + x12 + x13 + x15

       Df Deviance    AIC
<none>      184.97 200.97
```

```
- x1    1   187.20 201.20
- x4    1   190.50 204.50
- x15   1   191.81 205.81
- x13   1   197.37 211.37
- x8    1   213.49 227.49
- x12   1   220.03 234.03
- x10   1   227.69 241.69

Call:  glm(formula = Nota ~ x1 + x4 + x8 + x10 + x12 + x13 + x15, family = binomial(link =

Coefficients:
(Intercept)           x1           x4           x8          x10          x12
   -24.9949       0.4284       0.6740       1.1130       1.3689       1.4476
        x13          x15
     0.7602       0.7557

Degrees of Freedom: 539 Total (i.e. Null);  532 Residual
Null Deviance:            574.6
Residual Deviance: 185         AIC: 201
```

Observe que o função selecionou as variáveis $x_1, x_4, x_8, x_{10}, x_{12}, x_{13}$ e x_{15} para compor o modelo logístico. Vamos agora realizar a mesma análise inicialmente realizada, através da função *summary*, para verificar a significância dos coeficientes dessas variáveis. A Figura **??** retrata o retorno do R.

```
> modelo2=glm(Nota~x1+x4+x8+x10+x12+x13+x15,family=binomial(link="logit"))
> summary(modelo2)

Call:
glm(formula = Nota ~ x1 + x4 + x8 + x10 + x12 + x13 + x15, family = binomial(link = "logit"

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.5491   0.0294   0.0532   0.1908   2.3588

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -24.9949     3.0010  -8.329  < 2e-16 ***
x1            0.4284     0.2909   1.472 0.140898
x4            0.6740     0.2950   2.285 0.022340 *
x8            1.1130     0.2334   4.768 1.86e-06 ***
x10           1.3689     0.2368   5.781 7.42e-09 ***
x12           1.4476     0.2731   5.300 1.16e-07 ***
```

```
x13           0.7602     0.2237   3.399 0.000676 ***
x15           0.7557     0.2932   2.578 0.009950 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 574.58  on 539  degrees of freedom
Residual deviance: 184.97  on 532  degrees of freedom
AIC: 200.97

Number of Fisher Scoring iterations: 7
```

Observe que, quando analisada junta com as demais variáveis selecionadas, a variável x_1 não foi significativa para o modelo, considerando o teste de Wald, cuja estatística é representada por *zvalue*, na Figura **??**. Retirando-a e procedendo a mesma análise, temos o resultado exposto na Figura **??**.

```
> modelo3=glm(Nota~x4+x8+x10+x12+x13+x15,family=binomial(link="logit"))
> summary(modelo3)

Call:
glm(formula = Nota ~ x4 + x8 + x10 + x12 + x13 + x15, family = binomial(link = "logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.5082   0.0330   0.0583   0.1923   2.2792

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -23.6859     2.7427  -8.636  < 2e-16 ***
x4            0.7612     0.2856   2.665 0.007693 **
x8            1.1370     0.2326   4.889 1.01e-06 ***
x10           1.3509     0.2326   5.808 6.34e-09 ***
x12           1.3628     0.2609   5.223 1.76e-07 ***
x13           0.8071     0.2225   3.628 0.000286 ***
x15           0.8210     0.2862   2.869 0.004121 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
```

```
    Null deviance: 574.58  on 539  degrees of freedom
Residual deviance: 187.20  on 533  degrees of freedom
AIC: 201.2

Number of Fisher Scoring iterations: 7
```

Portanto, o modelo estimado é:

$$P(Y=1) = P(Nota \geq 4) = \frac{1}{1+e^{23,6859-0,7612x_4-1,1370x_8-1,3509x_{10}-1,3628x_{12}-0,8071x_{13}-0,8210x_{15}}}$$

Através do modelo estimado, podemos concluir que um caso um programa de pós-graduação tenha obtido conceito bom nas variáveis consideradas significativas do modelo, a probabilidade de obter um conceito maior ou igual a 4 será:

$$P(Y=1) = P(Nota \geq 4) = \frac{1}{1+e^{23,6859-0,7612\times4-1,1370\times4-1,3509\times4-1,3628\times4-0,8071\times4-0,8210\times4}} \approx 78,14\%.$$

Verifiquemos, agora, a qualidade de ajuste pelo teste de Hosmer & Lemeshow:

```
> library(ResourceSelection)
> hoslem.test(Nota,fitted(modelo3))

        Hosmer and Lemeshow goodness of fit (GOF) test

data:  Nota, fitted(modelo3)
X-squared = 6.4298, df = 8, p-value = 0.5992
```

Como o valor-p calculado para o teste foi maior do que o nível de significância adotado (5%), não rejeitamos a hipótese de nulidade e concluímos que o modelo se ajusta bem aos dados.

Vamos calcular agora, a razão de chances (OR), para as variáveis do modelo. Para isto, progamamos o software para calcular a exponencial dos coeficientes do modelo. O comando *round()* foi utilizado para arredondamento, considerando 3 casas decimais.

```
> OR=exp(modelo3$coefficients)
> round((cbind(OR)),3)

               OR
(Intercept) 0.000
x4          2.141
x8          3.118
x10         3.861
```

```
x12         3.907
x13         2.242
x15         2.273
```

Observe, através da Figura **??**, que para cada mudança de 1 unidade no conceito da variável x_4 - Perfil do corpo docente, consideradas titulação, diversificação na origem da formação, aprimoramento e experiência, e sua compatibilidade e adequação à proposta do programa - são 2,141 maiores as chances de se obter uma nota maior ou igual a 4, já para a variável x_8 - Quantidade de teses e dissertações defendidas no período de avaliação, em relação ao corpo docente permanente e à dimensão do corpo docente - essas chances são 3,118 maiores e, assim, sucessivamente.

Como discussão, observe também que das seis variáveis do modelo estimado, as três com maior efeito no conceito CAPES são relacionadas com a produção científica do corpo docente. Assim, conclui-se, em conformidade com Mesquita e Nogueira (2015), que os programas de pós-graduação devem investir na quantidade e qualidade das publicações de seus docentes e discentes.

CAPÍTULO 4

ANÁLISE CLÁSSICA DE AVALIACÕES NO SOFTWARE R

4 Análise Clássica de Avaliações no Software R

A necessidade cada vez maior de se produzir avaliações consistentes, com itens capazes de estimar com precisão o grau de conhecimento em determinada área e selecionar talentos, fez surgir, no campo da Psicometria, uma teoria de análise de testes, conhecida como Teoria Clássica dos Testes (TCT).

A TCT faz uso de algoritmos estatísticos com o intuito de avaliar diversos aspectos dos itens que compõe o teste. Existem várias informações que podem ser usadas para determinar se um item é útil como instrumento do que se propõe a medir e sobre o como ele funciona em relação aos outros itens de um teste. Neste capítulo definiremos algumas delas.

O softwate R apresenta um agrupamento dos principais pacotes relacionados a Psicometria, chamado *Psychometrics*, devido ao extenso número de pacotes que realizam análises psicométricas. Para instalar todos estes pacotes de uma vez basta digitar no *R Console*:

```
> install.packages("ctv")
> library(ctv)
> install.views("Psychometrics")
```

Os principais pacotes utilizados neste texto são *ltm* (Rizopoulos (2006)), *mirt* (Chalmers (2012)) e *psych* (Revelle (2014)). Entretanto, outros pacotes como *irtoys* (Partchev (2009)) e *CTT* (Willse e Shu (2014)), também podem ser utilizados na análise de itens.

4.1 Banco de dados

Analisaremos um banco de dados referente a resposta de 477 indivíduos à 50 itens de um teste. Os gabaritos individuais encontram-se digitalizados em uma planilha do Excel, conforme ilustra a Figura 24.

	A	B	C	D	E	F		AY
1		**Questão 1**	**Questão 2**	**Questão 3**	**Questão 4**	**Questão 5**	**...**	**Questão 50**
2	**Gabarito**	D	D	A	A	B	...	D
3	**Respondente 1**	A	D	A	A	B	...	D
4	**Respondente 2**	B	E	A	C	B	...	C
5	**Respondente 3**	D	D	A	A	B	...	B
6	**Respondente 4**	D	B	A	A	B	...	C
7	**Respondente 5**	A	D	E	A	E	...	B
8	**Respondente 6**	A	C	A	A	E	...	D
9	**Respondente 7**	C	A	A	A	B	...	B
10	**Respondente 8**	D	D	A	A	B	...	D
11	**Respondente 9**	C	D	A	C	B	...	D
12	**Respondente 10**	A	A	D	A	C	...	B
	...	B	D	C	A	C	...	B
479	**Respondente 477**	C	A	B	A	B	...	B

Figura 24: Exemplo de banco de dados.

Para a leitura dos dados, utilizou-se o comando *read.table()*:

```
> dados3=read.table("C:/Users/Usuario/Desktop/UFES/dados3.txt")
```

É possível também selecionar apenas uma linha, ou coluna, da planilha. Para isto para acrescentar ao lado do nome dado à planilha o número da linha ou coluna (nesta ordem) desejado. Por exemplo, se quisermos apenas o gabarito, que encontra-se na segunda linha da planilha, fazemos:

```
> gabarito=as.character(as.matrix(dados3[2,]))
```

Da mesma forma, podemos excluir algumas linhas ou colunas indesejadas:

```
> gabarito=as.character(as.matrix(dados3[2,-1]))
```

Observe que o comando acima excluirá a primeira coluna do gabarito, restando apenas as opções corretas. No modelo analisado, pode obter apenas as opções marcadas pelos respondentes através do seguinte comando:

```
> respostas=as.matrix(dados3[-2:-1,-1])
```

Existe uma função no R para dicotomizar os dados, de acordo com o gabarito, transformando-os em 0 para respostas incoerentes e 1 para respostas coerentes. Trata-se da função *mult.choice()*, Além disso, o comando *dim()* fornece a dimensão dos dados analisados.

```
> prova.dicotomizada=mult.choice(respostas,gabarito)
```

```
> dim(prova.dicotomizada)
```

```
[1] 477  50
```

Dessa forma, pode-se somar as linhas da planilha para obter o número de acertos de cada respondente e assim, proceder a análises clássicas, a serem vistas na seção 4

```
> notas=rowSums(prova.dicotomizada)
```

4.2 Estatísticas descritivas

É comum começar uma análise de um conjunto de dados pelas estatísticas descritivas, como as medidas de posição e de dispersão. O mesmo ocorre com os itens de um teste. Geralmente, quanto maior a variabilidade do item e quanto mais a média do item estiver no ponto central da distribuição, melhor será o item (KLINE, 2005).

Um resumo das estatísticas clássicas pode ser obtido através da função *summary()*:

```
> summary(notas)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   9.00   21.00   27.00   27.32   34.00   49.00
```

4.3 Índice de dificuldade

Conforme Borgatto e Andrade (2012), a dificuldade D_i de um determinado item i é definida em termos da proporção $\hat{p}$ de respondentes que acertam tal item.

Para classificação do nível de dificuldade dos itens, pode-se tomar como base Tabela a 18, proposta por Condé (2001).

Tabela 18: Critério para avaliação do índice de dificuldade de um determinado item

Classificação	Índice de dificuldade
Fácil	$D_i > 0,70$
Moderado	$0,30 < D_i \leq 0,70$
Difícil	$D_i \leq 0,30$

O índice de dificuldade dos itens do teste analisado pode ser obtido pelo comando *reliability* do pacote *CTT*:

```
> library(CTT)
> reliability(prova.dicotomizada)$itemMean
```

```
       V1         V2         V3         V4         V5         V6         V7
0.85953878 0.80712788 0.53039832 0.29559748 0.65408805 0.91823899 0.54507338
       V8         V9        V10        V11        V12        V13        V14
0.53878407 0.90356394 0.09224319 0.53878407 0.87211740 0.74842767 0.87211740
       V15        V16        V17        V18        V19        V20        V21
0.76100629 0.42767296 0.54297694 0.77358491 0.22222222 0.32494759 0.37526205
       V22        V23        V24        V25        V26        V27        V28
0.31656184 0.71069182 0.70020964 0.65199161 0.51991614 0.40461216 0.77987421
       V29        V30        V31        V32        V33        V34        V35
0.45073375 0.35639413 0.57442348 0.50314465 0.75681342 0.65618449 0.41509434
       V36        V37        V38        V39        V40        V41        V42
0.98322851 0.27882600 0.58700210 0.37945493 0.73794549 0.24947589 0.59958071
       V43        V44        V45        V46        V47        V48        V49
0.22431866 0.60167715 0.38155136 0.41090147 0.28721174 0.37735849 0.24318658
       V50
0.58071279
```

4.4 Discriminação do item

A discriminação do item avalia a capacidade de diferenciar indivíduos com bom desempenho daqueles indivíduos com baixo rendimento, no mesmo teste. Para isto, podemos analisar estatisticamente este índice através da criação de grupos-critério, de uma análise gráfica ou da correlação de cada item com o escore total observado.

4.4.1 Índice de discriminação

Pasquali (2003) sugere uma estatística para análise do poder discriminativo de um item baseado na criação de grupos-critérios: superior (formado pelos indivíduos com maior rendimento no teste) e inferior (formado pelos indivíduos com menor rendimento no teste). Para isto, baseado na distribuição sugerida por Kelley (1939), em que a porcentagem referente a cada grupo deve ser de 27% do total de indivíduos, define-se o índice de discriminação como o valor absoluto da diferença entre o índice de dificuldade calculado para cada um desses grupos. Para ilustrar esta situação, observe a Figura 25.

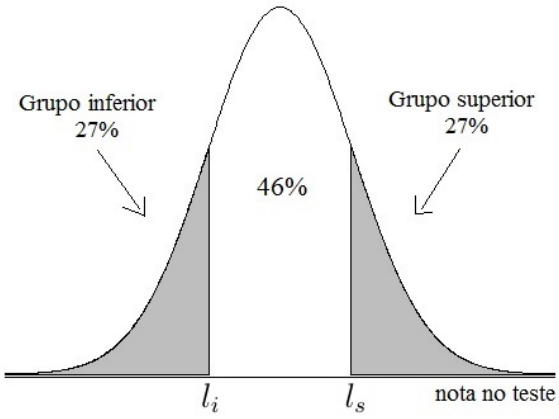

Figura 25: lustração do método de Kelley.

Como não foi encontrado nenhum comando no R que retornava esta estatística foi preciso construí-la manualmente. Uma das grandes vantagens de trabalhar com este software é a fácil manipulação e compreensão de seus comandos.

Primeiramente, encontremos as notas que correspondem aos limites inferior e superior:

```
> quantile(notas, c(0.27, 0.73))
```

```
27% 73%
 21  33
```

Encontrados os valores de 21 e 33, respectivamente, temos que 27% dos estudantes encontram-se com nota menor ou igual a 21 e, analogamente, temos o mesmo percentual para estudantes com média maior ou igual a 33. Assim, podemos determinar os grupos-critérios, através da criação de subconjuntos do conjunto de dados:

```
> dadosx=cbind(prova.dicotomizada,notas)
> grupoinferior=subset(dadosx, notas<=21)
> gruposuperior=subset(dadosx, notas>=33)
```

Finalmente, podemos determinar o índice de discriminação (ID), como o valor absoluto da diferença entre o índice de dificuldade encontrado para os dois grupos.

```
> ID=reliability(gruposuperior)$itemMean-reliability(grupoinferior)$itemMean
```

```
        V1         V2         V3         V4         V5         V6
0.16305165 0.27815956 0.58455922 0.45817674 0.60909481 0.08525717
        V7         V8         V9        V10        V11        V12
0.02067003 0.56297648 0.12815419 0.12847632 0.59894771 0.21448513
       V13        V14        V15        V16        V17        V18
0.48759798 0.29469559 0.40846129 0.63889187 0.69354666 0.42338666
       V19        V20        V21        V22        V23        V24
0.17727907 0.29109846 0.47659186 0.39369698 0.63975089 0.47186728
       V25        V26        V27        V28        V29        V30
0.73354451 0.63626114 0.45420380 0.42311822 0.38414045 0.38118759
       V31        V32        V33        V34        V35        V36
0.69435198 0.51820037 0.29174273 0.51637496 0.46993450 0.04316547
       V37        V38        V39        V40        V41        V42
0.37023516 0.45984108 0.44754644 0.28347471 0.31880168 0.64506604
       V43        V44        V45        V46        V47        V48
0.39181789 0.65145496 0.33002255 0.47100827 0.56292280 0.63916031
       V49        V50      notas
0.31133899 0.54885644 21.20664662
```

Segundo Arias; Lloreda e Lloreda (2006), uma boa referência para a classificação da qualidade discriminativa de um item é a descrita na Tabela 19.

Tabela 19: Classificação do item do teste, de acordo com o índice de discriminação

Índice de discriminação	Classificação do item
$ID \leq 0,20$	Ineficiente. Sugere-se eliminá-lo ou revisá-lo totalmente
$0,20 < ID \leq 0,30$	Necessita ser revisado
$0,30 < ID \leq 0,40$	Aceitável, não sendo necessária uma revisão.
$ID > 0,40$	Satisfatório. Deve permanecer no teste.

4.4.2 Análise gráfica

Outra maneira de medir o poder de discriminação de um item é através da análise gráfica do escore total *versus* proporção de acerto dos itens do teste. Este método de análise foi desenvolvido por Rasch (1960), que traçou os escores totais de um teste em função das taxas de aprovação em itens cognitivos.

A Figura 26 traz a análise de três itens do teste. Observe que o item 3 tem grau de dificuldade muito baixo, visto que independente do escore do respondente a proporção de acerto está sempre próxima de 1. Em contrapartida, o item 1 apresenta um grau de dificuldade elevado, pois para obter uma boa proporção de acerto, é preciso que o respondente tenha obtido um escore alto no teste. Estes itens não apresentam um bom poder discriminativo, pois para diferentes grupos-critérios, a proporção de acerto não difere muito. Por outro lado, o item 2 se destaca, quanto

a discriminação, por apresentar diferentes proporções de acerto para respondentes com diferentes escores.

```
> plot(descript(prova.dicotomizada),items=c(10,11,36),type="b")
```

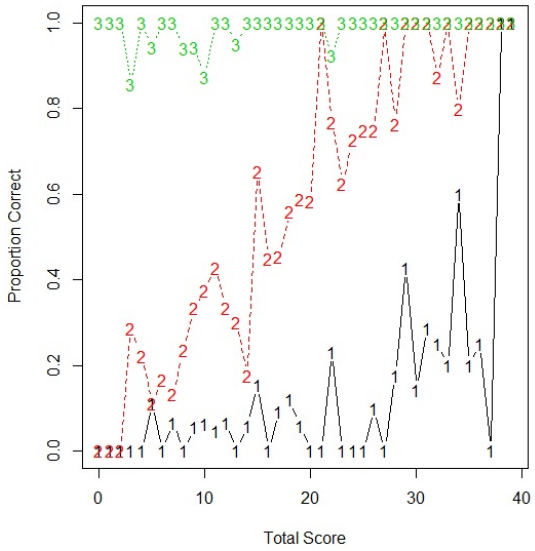

Figura 26: Escore total *versus* proporção de acerto para três itens do teste.

Embora visual, esta análise pode ser bastante subjetiva.

4.4.3 Coeficientes de correlação

Segundo Pasquali (2003), se desejamos estudar a correlação entre uma variável contínua e uma variação dicotômica, devemos usar os coeficientes de correlação bisserial, (ρ_b) ou ponto-bisserial (ρ_{pb}). Ambas tratam de uma estimativa derivada do coeficiente de correlação de Pearson.

Em nossa análise, a variável analisada é naturalmente dicotômica, ou seja, sua classificação já é determinada de forma natural: acerto *versus* erro. Neste caso, utiliza-se o coeficiente de correlação ponto bisserial. Entretanto, existirão casos em que uma variável numérica é artificialmente dicotomizada. Por exemplo, caso queiramos codificar a idade de um grupo de pessoas, podemos fazê-la comparando com um valor específico: se maior que este valor, recebe 1, caso contrário 0. Nestes casos, segundo Pasquali (2003), utiliza-se o coeficiente de correlação bisserial. Como pressuposto, para o cálculo do ρ_b é necessário que a variável contínua a ser dicotomizada siga uma distribuição normal. Para obter estas estimativas, pode-se utilizar, respectivamente, comandos abaixo. Aqui, os coeficientes foram calculados excluindo os respectivos itens.

```
> reliability(prova.dicotomizada)$pBis

 [1]  0.16586934  0.23521025  0.43059858  0.37162144  0.48226086  0.08320766
 [7] -0.03027969  0.42955587  0.15051627  0.20490619  0.46700528  0.23117620
[13]  0.39759758  0.34330411  0.33142067  0.45081533  0.50575102  0.35405813
[19]  0.11108886  0.18652716  0.35301238  0.30232517  0.51041532  0.36564821
[25]  0.54416168  0.42011930  0.33818037  0.36713635  0.28946029  0.27880944
[31]  0.50429454  0.38946130  0.25354209  0.37950519  0.33660198  0.09586229
[37]  0.34049016  0.33515950  0.33916906  0.19794198  0.28975971  0.49270573
[43]  0.38485001  0.44964539  0.23990932  0.37358864  0.46163060  0.49467733
[49]  0.28148164  0.40653983

> reliability(prova.dicotomizada)$bis

 [1]  0.27608473  0.35623844  0.53084833  0.46739622  0.63010852  0.15970077
 [7] -0.03803848  0.53322851  0.27237941  0.32477415  0.57371056  0.39826970
[13]  0.57029091  0.62592156  0.48437082  0.54549108  0.61871558  0.52715830
[19]  0.15181538  0.23606209  0.43404702  0.37936073  0.70179856  0.49973257
[25]  0.71511836  0.51635440  0.41516445  0.54572739  0.35749162  0.34360002
[31]  0.62320509  0.47955758  0.36183309  0.50680731  0.41277344  0.35589076
[37]  0.42904958  0.42042140  0.41532340  0.27190546  0.36413868  0.62820524
[43]  0.48880082  0.56826500  0.29701527  0.45642958  0.56580208  0.59477389
[49]  0.35437591  0.51250470
```

De acordo com Tôrres (2015), de maneira geral, espera-se que o coeficiente de correlação ponto-bisserial assuma valores positivos e superiores a $0,30$ para que sejam considerados de boa discriminação.

Em suma, espera-se de uma resposta a um item discriminativo que os estudantes que saem-se bem na prova como um todo, acertem-no, e por sua vez, aqueles que não vão bem, errem-no. Quanto maior forem os coeficientes de correlação bisserial e ponto-bisserial, maior a capacidade do item de discriminar grupos de indivíduos que construíram determinada competência e habilidade, daqueles que não as construíram. Além disso, os itens com coeficiente de correlação baixo não diferenciam o indivíduo que construiu determinada competência e habilidade, daquele que não a construiu (SANTOS, 2008).

4.5 Coeficiente alpha de Cronbach

A precisão de um teste está associada ao erro de medida, ou seja, a diferença entre os escores observado e verdadeiro em um teste. A precisão de um teste pode ser usada para estimar o erro padrão dos escores e para estabelecer intervalos de confiança em torno dos valores observados. Uma estimativa usual da precisão é dada pelo estimador alpha de Cronbach.

O coeficiente alpha de Cronbach mede a correlação entre respostas em um teste através da análise das respostas dadas pelos respondentes, apresentando uma correlação média entre as perguntas. Dessa forma, ele pode ser utilizado para medir a consistência interna do instrumento de medida. No R, pode-se determiná-lo da seguinte maneira:

```
> reliability(prova.dicotomizada)$alpha

[1] 0.8832461
```

A função *cronbach.alpha()* do pacote *ltm* também retorna o coeficiente:

```
> library(ltm)
> cronbach.alpha(prova.dicotomizada)

Cronbach's alpha for the 'prova.dicotomizada' data-set

Items: 50
Sample units: 477
alpha: 0.883
```

Para determinar o coeficiente alpha de Cronbach estimado para cada item, excluindo-o, pode-se fazer:

```
> reliability(prova.dicotomizada)$alphaIfDeleted

 [1] 0.8832190 0.8825171 0.8797144 0.8806843 0.8789553 0.8837806 0.8870200
 [8] 0.8797324 0.8832372 0.8827225 0.8791212 0.8824837 0.8803430 0.8812799
[15] 0.8812750 0.8793947 0.8784876 0.8809716 0.8842070 0.8834612 0.8809635
[22] 0.8817200 0.8786303 0.8807716 0.8779810 0.8798845 0.8812030 0.8808014
[29] 0.8819929 0.8821074 0.8785270 0.8803833 0.8823495 0.8805519 0.8812304
[36] 0.8834183 0.8811437 0.8812529 0.8811814 0.8831570 0.8818549 0.8787329
[43] 0.8805570 0.8794264 0.8827324 0.8806401 0.8793603 0.8787226 0.8819647
[50] 0.8801114
```

Esse coeficiente varia de zero a um, sendo o teste mais consistente a medida que se aproxima de um. De acordo com Hair Júnior et. al (2010), valores acima de $0,7$ são considerados satisfatórios.

4.6 A função *descript()*

A função *descript()* apresenta um resumo das principais análises descritivas a serem realizadas, como a frequência com que cada escore é obtido, o percentual de erros e acertos de cada item, o logit da proporção para as respostas, o coeficiente alpha de Cronbach para todos os itens e também para os itens individuais, excluindo-os, o coeficiente de correlação bisserial de cada item

com a pontuação total incluindo e excluindo no cálculo da pontuação total e uma análise do grau de associação entre pares de itens através de um teste Qui-quadrado, realizado através da construção de tabelas de contingência para todos os possíveis pares de itens.

```
> descript(prova.dicotomizada)

Descriptive statistics for the 'prova.dicotomizada' data-set

Sample:
 50 items and 477 sample units; 0 missing values

Proportions for each level of response:
         0      1   logit
V1  0.1405 0.8595  1.8115
V2  0.1929 0.8071  1.4315
V3  0.4696 0.5304  0.1217
V4  0.7044 0.2956 -0.8684
V5  0.3459 0.6541  0.6371
V6  0.0818 0.9182  2.4187
V7  0.4549 0.5451  0.1808
V8  0.4612 0.5388  0.1554
V9  0.0964 0.9036  2.2375
V10 0.9078 0.0922 -2.2865
V11 0.4612 0.5388  0.1554
V12 0.1279 0.8721  1.9198
V13 0.2516 0.7484  1.0902
V14 0.1279 0.8721  1.9198
V15 0.2390 0.7610  1.1582
V16 0.5723 0.4277 -0.2914
V17 0.4570 0.5430  0.1723
V18 0.2264 0.7736  1.2287
V19 0.7778 0.2222 -1.2528
V20 0.6751 0.3249 -0.7311
V21 0.6247 0.3753 -0.5097
V22 0.6834 0.3166 -0.7696
V23 0.2893 0.7107  0.8987
V24 0.2998 0.7002  0.8483
V25 0.3480 0.6520  0.6278
V26 0.4801 0.5199  0.0797
V27 0.5954 0.4046 -0.3863
V28 0.2201 0.7799  1.2649
V29 0.5493 0.4507 -0.1977
```

```
V30 0.6436 0.3564 -0.5910
V31 0.4256 0.5744  0.2999
V32 0.4969 0.5031  0.0126
V33 0.2432 0.7568  1.1353
V34 0.3438 0.6562  0.6463
V35 0.5849 0.4151 -0.3429
V36 0.0168 0.9832  4.0712
V37 0.7212 0.2788 -0.9503
V38 0.4130 0.5870  0.3516
V39 0.6205 0.3795 -0.4919
V40 0.2621 0.7379  1.0353
V41 0.7505 0.2495 -1.1014
V42 0.4004 0.5996  0.4037
V43 0.7757 0.2243 -1.2407
V44 0.3983 0.6017  0.4125
V45 0.6184 0.3816 -0.4830
V46 0.5891 0.4109 -0.3602
V47 0.7128 0.2872 -0.9090
V48 0.6226 0.3774 -0.5008
V49 0.7568 0.2432 -1.1353
V50 0.4193 0.5807  0.3257


Frequencies of total scores:
     0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
Freq 0 0 0 0 0 0 0 0 0 1  0  2  3  7  9 18 12 15 17 18 16 21 15 20 17 26 20 22
     28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
Freq 25 17 12 17 13  8 11 13 19 13 17  7  7  7  8  5  5  5  4  1  2  2  0


Point Biserial correlation with Total Score:
    Included Excluded
V1    0.2048   0.1659
V2    0.2781   0.2352
V3    0.4770   0.4306
V4    0.4166   0.3716
V5    0.5239   0.4823
V6    0.1146   0.0832
V7    0.0273  -0.0303
V8    0.4760   0.4296
V9    0.1838   0.1505
```

V10	0.2368	0.2049
V11	0.5115	0.4670
V12	0.2676	0.2312
V13	0.4394	0.3976
V14	0.3771	0.3433
V15	0.3749	0.3314
V16	0.4958	0.4508
V17	0.5480	0.5058
V18	0.3960	0.3541
V19	0.1584	0.1111
V20	0.2385	0.1865
V21	0.4015	0.3530
V22	0.3508	0.3023
V23	0.5487	0.5104
V24	0.4111	0.3656
V25	0.5823	0.5442
V26	0.4671	0.4201
V27	0.3879	0.3382
V28	0.4082	0.3671
V29	0.3417	0.2895
V30	0.3295	0.2788
V31	0.5463	0.5043
V32	0.4379	0.3895
V33	0.2997	0.2535
V34	0.4260	0.3795
V35	0.3866	0.3366
V36	0.1106	0.0959
V37	0.3859	0.3405
V38	0.3852	0.3352
V39	0.3883	0.3392
V40	0.2465	0.1979
V41	0.3352	0.2898
V42	0.5350	0.4927
V43	0.4255	0.3849
V44	0.4942	0.4496
V45	0.2925	0.2399
V46	0.4220	0.3736
V47	0.5023	0.4616
V48	0.5364	0.4947
V49	0.3268	0.2815
V50	0.4536	0.4065

```
Cronbach's alpha:
               value
All Items     0.8832
Excluding V1  0.8832
Excluding V2  0.8825
Excluding V3  0.8797
Excluding V4  0.8807
Excluding V5  0.8790
Excluding V6  0.8838
Excluding V7  0.8870
Excluding V8  0.8797
Excluding V9  0.8832
Excluding V10 0.8827
Excluding V11 0.8791
Excluding V12 0.8825
Excluding V13 0.8803
Excluding V14 0.8813
Excluding V15 0.8813
Excluding V16 0.8794
Excluding V17 0.8785
Excluding V18 0.8810
Excluding V19 0.8842
Excluding V20 0.8835
Excluding V21 0.8810
Excluding V22 0.8817
Excluding V23 0.8786
Excluding V24 0.8808
Excluding V25 0.8780
Excluding V26 0.8799
Excluding V27 0.8812
Excluding V28 0.8808
Excluding V29 0.8820
Excluding V30 0.8821
Excluding V31 0.8785
Excluding V32 0.8804
Excluding V33 0.8823
Excluding V34 0.8806
Excluding V35 0.8812
Excluding V36 0.8834
```

```
Excluding V37 0.8811
Excluding V38 0.8813
Excluding V39 0.8812
Excluding V40 0.8832
Excluding V41 0.8819
Excluding V42 0.8787
Excluding V43 0.8806
Excluding V44 0.8794
Excluding V45 0.8827
Excluding V46 0.8806
Excluding V47 0.8794
Excluding V48 0.8787
Excluding V49 0.8820
Excluding V50 0.8801


Pairwise Associations:
   Item i Item j p.value
1       2     36   1.000
2       6     40   1.000
3       7     10   1.000
4       7     26   1.000
5       7     42   1.000
6      12     36   1.000
7      20     36   1.000
8      26     36   1.000
9      28     36   1.000
10     36     37   1.000
```

CONSIDERAÇÕES FINAIS

5 Considerações finais

Com a popularização do uso de computadores e o desenvolvimento de softwares específicos, as análises estatísticas se tornaram acessíveis e fáceis de serem executadas por pesquisadores das mais diversas áreas do conhecimento. Segundo Gatti (2004):

> "Estas análises [...] trazem subsídios concretos para a compreensão de fenômenos educacionais indo além dos casuísmos e contribuindo para a produção/enfrentamento de políticas educacionais, para planejamento, administração/ gestão da educação, podendo ainda orientar ações pedagógicas de cunho mais geral ou específico." (GATTI, 2004, p. 26)

Apesar de introdutório, espera-se que este curso sirva para a difusão dos procedimentos de análises estatísticas em trabalhos acadêmicos, especialmente nas áreas de ciências humanas e sociais aplicadas. Para isto, sugere-se o uso do software R, por se tratar de um software livre, de código aberto de fácil manipulação e com uma extensão de funções imensuráveis.

REFERÊNCIAS

AMERICO, B. L.; LACRUZ, A. J. Contexto e desempenho escolar: análise das notas na Prova Brasil das escolas capixabas por meio de regressão linear múltipla, **Rev. Adm. Pública** [online]. 2017, vol.51, n.5, pp.854-878.

ARIAS, M. R. M.; LLOREDA, M. V. H.; LLOREDA, M. J. H. **Psicometría**. [S.1.]: Alianza Editorial, 2006. 488 p.

BORGATTO, A. F.; ANDRADE, D. F. Análise clássica de teste com diferentes graus de dificuldade. **Estudos em Avaliação Educacional**, v. 23, n. 52, p. 146-156, 2012.

BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica**. 6 ed. São Paulo: Saraiva, 2010. 540 p.

CHALMERS, R. P. mirt: A multidimensional item response theory package for the R environment. **Journal of Statistical Software**, v. 48, n. 6, p. 1-29, 2012.

CONDÉ, F. N. **Análise empírica de itens**. Technical report, Instituto Nacional de Estudos e Pesquisas Educacionais-DAEB/INEP/MEC, Brasília, 2001. 193 p.

FÁVERO, L. P.; BELFIORE, P.; SILVA, F. L.; CHAN, B. L. **Análise de dados: modelagem multivariada para tomada de decisões.** Rio de Janeiro: Campus/Elsevier, 2009. 646 p.

FÁVERO, L. P.; FÁVERO, P. **Estatística aplicada: Para cursos de Administração, Contabilidade e Economia com Excel e SPSS.** Rio de Janeiro: Elsevier Brasil, 2015. 480 p.

GATTI, B. A. Estudos quantitativos em educação. **Educação e pesquisa**, v. 30, n. 1, p. 11-30, São Paulo, jan./abr. 2004.

GROSS, J.; LIGGES, U.; LIGGES, M. U., I. **Nortest: Tests for Normality**. Five omnibus tests for testing the composite hypothesis of normality. R package version 1.0-3. Publicado em 26/02/2015 Disponível em: http://CRAN.R-project.org/package=nortest. Acesso em 23/05/2018.

GUJARATI, D. N.; PORTER, D. C. **Econometria Básica.** 5 ed. Porto Alegre: McGraw Hill, 2011, 924 p.

HAIR JÚNIOR, J.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. **Multivariate data analysis**. 7th ed. Upper Saddle River: Prentice Hall, 2010. 785 p.

INÁCIO, E. S. B.; ENCINAS, J. I.; SANTANA, O. A. Testes estatísticos utilizado em trabalhos científicos apresentados nos congressos internacionais de educação à distância da ABED (2001 a 2011). In: Congresso Internacional de Educação a Distância, 18., 2012, Recife. **Anais...** Recife: Associação Brasileira de Educação a Distância, 2012. 100 p.

KELLEY, T. L. The selection of upper and lower groups for the validation of test items. **Journal of educational psychology**, Warwick & york, v. 30, n. 1, p. 17-24, 1939.

KLINE, T. **Psychologicl testing: A practical approach to design and evaluation.** Thousand Oaks, CA: Sage, 2005. 363p.

MELLO, M. P.; ALENCAR, E. R.; PETERNELLI, L. A. Ocorrências de análises estatísticas em revistas científicas de engenharia. In: Simpódio de Iniciação científica da Universidade Federal de Viçosa, 14., 2004, Viçosa, MG. **Resumos**... Viçosa, MG: UFV, 2004. (n. 0775). CD-Room.

MELLO, C. B. et al. Versão abreviada do WISC-III: correlação entre QI estimado e QI total em crianças brasileiras. **Psicologia: Teoria e Pesquisa**. 2011, vol.27, n.2, p. 149-155, 2011.

MESQUITA, P. S. B; NOGUEIRA, R. T. . Um modelo estatístico para gestão de programas de pós-graduação. **Revista GEPROS**, v. 10, n. 2, p. 173-186, 2015.

PARTCHEV, I. **irtoys: Simple Interface to the Estimation and Plotting of IRT Models**. R package version 0.1.3, v. 2, 2009. Disponível em: http://CRAN.R-project.org/package=irtoys. Acesso em 04/12/2017.

PASQUALI, L. **Psicometria: teoria dos testes na psicologia e na educação.** Petrópolis: Vozes; 2003. 397 p.

R DEVELOPMENT CORE TEAM. **R:** A Language and Environment for Statistical Computing. Vienna: R Foundation on Statistical Computing, 2019. Disponível em: https://www.r-project.org. Acesso: 24/05/2019.

RASCH, G. **Probabilistic models for some intelligence and achievement tests**. Copenhagen: Danish Institute for Education Research, 1960. 18 4p.

REIS, G. M.; JUNIOR, J. I. R. Comparação de testes paramétricos e não-paramétricos aplicados em delineamentos experimentais. In: Simpósio Acadêmico de Engenharia de Produção, v. 3, 2007.**Anais...**, Viçosa: UFV, 2007.

REVELLE, W. **psych: Procedures for personality and psychological research**. Northwestern University, Evanston. Illinois, USA, 2014. Disponível em: https://cran.r-project.org/web/packages/psych/index.html. Acesso em 04/12/2017.

RIZOPOULOS, D. ltm: An R package for latent variable modeling and item response theory analyses. **Journal of statistical software**, v. 17, n. 5, p. 1-25, 2006.

SANTOS, L. M. **Desempenho escolar em Pernambuco: análise dos itens e das habilidades usando teoria clássica e tri**. 2008. 104 p. Dissertação (Mestrado em Estatística) - Universidade Federal de Pernambuco, Recife. 2008.

SANTOS, C. **Estatística descritiva** - Manual de auto-aprendizagem. 1.ed. Lisboa: Edições Silabo, 2007. 264p.

TEIXEIRA, I. P. et al. Uso da estatística na Educação Física: análise das publicações nacionais entre os anos de 2009 e 2011. **Revista Brasileira de Educação Física e Esporte**, v. 29, n. 1, p. 139-147, 2015.

TÔRRES, F. C. **Uma aplicação da teoria de resposta ao item em um simulado de matemática no modelo enem.** 2015. 116p. Dissertação (Mestrado em Matemática) - Programa de Mestrado Profissional em Matemática em Rede Nacional, Universidade de Brasília, Brasília, 2015.

WILLSE, J. T.; SHU, Z. **CTT: Classical test theory functions**. R package version, v. 2, 2014. Disponível em: http://CRAN.R-project.org/package=CTT. Acesso em: 04/12/2017.

SOBRE OS AUTORES

Denilson Junio Marques Soares

Doutorando em Educação pela Universidade Federal do Espírito Santo (UFES); Mestre em Estatística Aplicada e Biometria e Licenciado em Matemática pela Universidade Federal de Viçosa (UFV). É professor EBTT do Instituto Federal de Minas Gerais (IFMG) - Campus Piumhi.

Talita Emidio Andrade Soares

Mestranda em Educação pela Universidade Federal do Espírito Santo (UFES). Especialista em Ensino de Matemática pela Faculdade Única e Licenciada em Matemática pela Universidade Federal de Viçosa (UFV).

Wagner dos Santos

Doutor em Educação pela Universidade Federal do Espírito Santo (UFES), onde atua como professor dos Programas de Pós-Graduação em Educação e em Educação Física. Líder do Instituto de Pesquisa em Educação e Educação Física (Proteoria). Bolsista de Produtividade em Pesquisa do CNPq - Nível 2.

ANÁLISE ESTATÍSTICA E SEU USO NA PESQUISA EDUCACIONAL

com práticas no *software* R

ANÁLISE ESTATÍSTICA E SEU USO NA PESQUISA EDUCACIONAL

com práticas no *software* R

ISBN 978-655889109-3

Rfb Editora