

**CURSO DE PÓS-GRADUAÇÃO “LATO SENSU”
(ESPECIALIZAÇÃO) A DISTÂNCIA
GESTÃO DE EMPRESAS COM ÊNFASE EM QUALIDADE**

INTRODUÇÃO À ESTATÍSTICA BÁSICA COM R

**Eric Batista Ferreira
Marcelo Silva de Oliveira**

**Universidade Federal de Lavras - UFLA
Fundação de Apoio ao Ensino, Pesquisa e Extensão - FAEPE
Lavras - MG
2008**

Parceria

Universidade Federal de Lavras - UFLA
Fundação de Apoio ao Ensino, Pesquisa e Extensão - FAEPE

Reitor

Antônio Nazareno Guimarães Mendes

Vice-Reitor

Elias Tadeu Fialho

Diretor da Editora

Marco Antônio Rezende Alvarenga

Pró-Reitor de Pós-Graduação

Joel Augusto Muniz

Pró-Reitor Adjunto de Pós-Graduação “Lato Sensu”

Marcelo Silva de Oliveira

Coordenador do Curso

Daniel Carvalho de Rezende

Presidente do Conselho Deliberativo da FAEPE

Luiz Antônio Lima

Editoração

Centro de Editoração/FAEPE

Impressão

Gráfica Universitária/UFLA

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Ferreira, Eric Batista; Oliveira, Marcelo Silva de
Introdução à Estatística Básica com R / Eric Batista Ferreira.
Marcelo Silva de Oliveira. — Lavras : UFLA/FAEPE, 2008, 1ª Edição.
124 p. : il. - Curso de Pós-Graduação “Lato Sensu” (Especialização)
a Distância - Gestão de Empresas com Ênfase em Qualidade.

Bibliografia.

1. Qualidade. 2. Gestão. 3. Método estatístico. i. Ferreira, E. B.;
Oliveira, M. S. de. ii. Universidade Federal de Lavras. iii. Fundação de
Apoio ao Ensino, Pesquisa e Extensão. iv. Título.

CDD-658

Nenhuma parte dessa publicação pode ser reproduzida, por qualquer meio ou forma, sem a
prévia autorização da FAEPE.

LISTA DE TABELAS	vi
LISTA DE FIGURAS	viii
1 INTRODUÇÃO	1
1.1 Introdução ao R	1
1.1.1 Baixando e instalando	1
1.1.2 Iniciando o R	2
1.1.3 Objetos	2
1.1.4 Lendo arquivos	8
1.1.5 Funções básicas	8
1.1.6 Pedindo ajuda	9
1.2 Alguns conceitos importantes	10
1.3 Principais aplicações da Estatística	12
1.4 Bibliografia consultada e recomendada	14
2 TÉCNICAS DE SOMATÓRIO	15
3 ESTATÍSTICA DESCRITIVA	19
3.1 Introdução	19
3.2 Variáveis qualitativas	20
3.3 Variáveis quantitativas discretas	23
3.4 Variáveis quantitativas contínuas	25
3.5 Medidas de posição	30
3.5.1 Média (Me)	30
3.5.2 Mediana (Md)	32
3.5.3 Moda (Mo)	34
3.6 Medidas de dispersão	36
3.6.1 Amplitude (A)	36
3.6.2 Variância e Desvio Padrão	37
3.6.3 Coeficiente de Variação (CV)	40
4 PROBABILIDADE	43
4.1 Distribuições de probabilidade discretas	47
4.1.1 Distribuição Binomial	48
4.1.2 Distribuição Poisson	49

4.2	Distribuições de probabilidades contínuas	50
4.2.1	Distribuição Normal de probabilidades	51
4.2.2	Outras distribuições contínuas	57
5	AMOSTRAGEM	59
5.1	Amostragens não-aleatórias	59
5.2	Amostragens aleatórias	60
5.2.1	Amostragem aleatória simples (AAS)	60
5.2.2	Amostragem aleatória estratificada (AAE)	60
5.2.3	Amostragem aleatória por conglomerado (AAC)	61
5.2.4	Amostragem aleatória sistemática (AS)	61
6	INFERÊNCIA ESTATÍSTICA	63
6.1	Distribuição de amostragem	63
6.1.1	Distribuição de funções da média amostral (populações normais)	63
6.2	Estimação	66
6.2.1	Estimação por ponto	66
6.2.2	Estimação por intervalo	66
6.3	Testes de hipóteses	70
6.3.1	Teste de homogeneidade de Variâncias (teste F):	70
6.3.2	Teste sobre μ (populações infinitas)	72
6.3.3	Teste sobre proporções (p)	74
7	CORRELAÇÃO E REGRESSÃO	75
7.1	Correlação	75
7.1.1	Coeficiente de correlação linear (r ou ρ)	75
7.1.2	Coeficiente de determinação (r^2 ou ρ^2)	75
7.2	Regressão	77
7.2.1	Regressão Linear Simples	77
7.2.2	Regressão Múltipla	84
8	APÊNDICE A: EXERCÍCIOS PROPOSTOS	85
9	APÊNDICE B: TABELAS	101

Lista de Tabelas

3.1	Distribuição de frequências absolutas (fa), relativa (fr) e percentual (fp) da atividade em propriedades de uma região.	20
3.2	Distribuição de frequências absolutas (fa), relativa (fr) e percentual (fp) do número de filhos por casal de uma cidade.	23
3.3	Distribuição de frequências absoluta (fa), relativa (fr) e percentual (fp) do peso observado em potinhos de canela em pó.	27
3.4	Frequências absolutas acumuladas abaixo ($Fa\downarrow$) e acima de ($Fa\uparrow$).	29
5.1	Diferenças básicas entre a amostragem aleatória estratificada (AAE) e a amostragem aleatória por conglomerado (AAC).	61
6.1	Representação tabular dos resultados possíveis em um teste de hipóteses e os erros e acertos que eles acarretam.	72
7.1	Volume de polpa (cm^3), volume de água (cm^3) e teor de cálcio ($mg/100ml$) em 20 cocos verdes.	80
7.2	Tabela auxiliar para cálculo dos coeficientes do modelo linear.	82
8.1	Distribuição de frequência de rendas familiares de 100 entrevistados, em Lavras, MG.	87
8.2	Distribuição de frequências das médias diárias de produção de leite no período de lactação de 201 vacas da raça holandesa, de um rebanho pertencente ao Núcleo de Criadores de Gado Holandês do Sul de Minas Gerais. Lavras, 1992.	88
9.1	Probabilidades (α) da distribuição normal padrão $N(0, 1)$ para valores do quantil Z_t padronizado de acordo com o seguinte evento: $P(0 < Z < Z_t) = \alpha$	102
9.2	Probabilidades (α) da distribuição normal padrão $N(0, 1)$ para valores do quantil Z_t padronizado de acordo com o seguinte evento: $P(Z > Z_t) = \alpha$	103
9.3	Quantis superiores da distribuição de qui-quadrado (χ^2_α) com ν graus de liberdade e para diferentes valores da probabilidade (α) de acordo com o seguinte evento: $P(\chi^2 > \chi^2_\alpha) = \alpha$	104
9.4	Quantis superiores da distribuição de qui-quadrado (χ^2_α) com ν graus de liberdade e para diferentes valores da probabilidade (α) de acordo com o seguinte evento: $P(\chi^2 > \chi^2_\alpha) = \alpha$	105
9.5	Quantis superiores da distribuição de F ($F_{0,10}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador, valor da probabilidade (α) de 10% de acordo com o seguinte evento: $P(F > F_{0,10}) = 0,10$	106

9.6	Quantis superiores da distribuição de F ($F_{0,10}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador, valor da probabilidade (α) de 10% de acordo com o seguinte evento: $P(F > F_{0,10}) = 0,10$	107
9.7	Quantis superiores da distribuição de F ($F_{0,05}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador valor da probabilidade (α) de 5% de acordo com o seguinte evento: $P(F > F_{0,05}) = 0,05$	108
9.8	Quantis superiores da distribuição de F ($F_{0,05}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador, valor da probabilidade (α) de 5% de acordo com o seguinte evento: $P(F > F_{0,05}) = 0,05$	109
9.9	Quantis superiores da distribuição de F ($F_{0,025}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador valor da probabilidade (α) de 2,5% de acordo com o seguinte evento: $P(F > F_{0,025}) = 0,025$	110
9.10	Quantis superiores da distribuição de F ($F_{0,025}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador valor da probabilidade (α) de 2,5% de acordo com o seguinte evento: $P(F > F_{0,025}) = 0,025$	111
9.11	Quantis superiores da distribuição de F ($F_{0,01}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador valor da probabilidade (α) de 1% de acordo com o seguinte evento: $P(F > F_{0,01}) = 0,01$	112
9.12	Quantis superiores da distribuição de F ($F_{0,01}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador valor da probabilidade (α) de 1% de acordo com o seguinte evento: $P(F > F_{0,01}) = 0,01$	113
9.13	Quantis superiores da distribuição de F ($F_{0,005}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador valor da probabilidade (α) de 0,5% de acordo com o seguinte evento: $P(F > F_{0,005}) = 0,005$	114
9.14	Quantis superiores da distribuição de F ($F_{0,005}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador valor da probabilidade (α) de 0,5% de acordo com o seguinte evento: $P(F > F_{0,005}) = 0,005$	115
9.15	Quantis superiores da distribuição t de Student (t_α) com ν graus de liberdade e para diferentes valores da probabilidade (α) de acordo com o seguinte evento: $P(t > t_\alpha) = \alpha$	116

Lista de Figuras

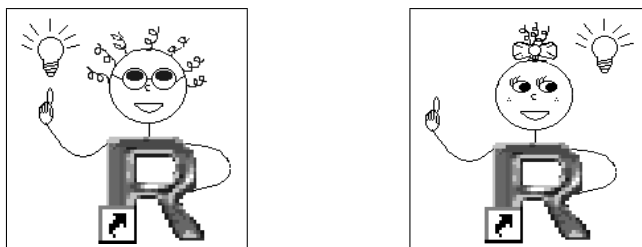
1.1	Tela inicial do R em ambiente Windows.	3
1.2	Mensagem ao tentar sair do R, em ambiente Windows.	3
1.3	Esquema de um objeto.	4
1.4	Gráfico ilustrativo do comportamento do peso de suínos (kg) em função de sua ingestão diária de ração ao longo do período de engorda.	13
2.1	Esquema ilustrativo dos elementos típicos de um somatório.	15
3.1	(a) Gráfico de colunas das principais atividades em propriedades rurais. (b) Gráfico de barras da mesma situação.	21
3.2	Setograma ou gráfico de pizza.	22
3.3	(a) gráfico de linhas da variável número de filhos por casal. (b) gráfico de colunas da mesma variável.	25
3.4	(a) Histograma do peso de potinhos de canela em pó em uma linha de produção. (b) O mesmo histograma com polígono de frequência.	28
3.5	Ogivas representando as frequências absolutas acumuladas <i>acima de</i> e <i>abaixo de</i> e seu respectivo código em R.	30
3.6	Histograma ilustrando geometricamente método de Czuber.	34
4.1	Simulação do lançamento de uma moeda honesta 500 vezes, comportamento de sua frequência relativa.	44
4.2	Esquema da integral de uma função densidade de probabilidade.	46
4.3	Esquema da generalização teórica de histogramas para funções densidade de probabilidade, quando $n \rightarrow \infty$	50
4.4	Esquema destacando a area acima de $100km/h$ em uma distribuição de media $60km/h$ e variância $400(km/h)^2$	52
4.5	Esquema destacando a area entre 40 e $100km/h$ em uma distribuição de media $60km/h$ e variância $400(km/h)^2$	54
4.6	Esquema destacando o intervalo que contem 90% dos veículos em uma distribuição de media $60km/h$ e variância $400(km/h)^2$	55
6.1	Demonstração da curva Normal (linha cheia) padrão e da curva t com 5 (linha tracejada) e 30 (linha pontilhada) graus de liberdade.	65
6.2	Distribuição F de probabilidade ressaltando a região de aceitação de um teste de homogeneidade de variâncias (de 1 até um F_c qualquer).	70

6.3	(a) Ilustração do teste unilateral superior. (b) Ilustração do teste bilateral. A area hachurada representa a região de rejeição do teste.	73
7.1	Representação de possíveis retas (pontilhadas) e aquela estimada por quadrados mínimos (linha cheia) em uma massa de dados fictícia.	79
7.2	Diagrama de dispersão entre a variável independente volume de água de coco e a variável dependente volume de polpa de coco.	81
7.3	Reta de regressão estimada e pontos amostrais.	83

1.1 Introdução ao R

R é uma linguagem e ambiente para computação estatística e gráfica. É um projeto GNU similar à linguagem e ambiente S desenvolvida no *Bell Laboratories* por John Chambers e colaboradores.

O software disponibiliza uma grande variedade de métodos estatísticos (modelagem linear e não-linear, testes estatísticos clássicos, séries temporais, classificação, métodos multivariados, etc) e técnicas gráficas. Um dos pontos fortes do R é a facilidade com que gráficos bem delineados e de alta qualidade para impressão podem ser produzidos com possibilidade de inclusão de fórmulas e símbolos matemáticos quando necessário.



Ele se presta a diversas funções, desde uma calculadora científica, passando pela integração e derivação de funções matemáticas, até a realização de complexas análises estatísticas.

Além disso, o R também apresenta uma série de recursos gráficos que permitem a descrição detalhada de todos os aspectos que se pode querer personalizar em um gráfico, como cor, tipo e tamanho de letra, símbolos, títulos e sub-títulos, pontos, linhas, legendas, planos de fundo e muito mais.

1.1.1 Baixando e instalando

O R é disponibilizado sob os termos da GNU General Public License da Free Software Foundation na forma de código aberto. Ele pode ser compilado e roda em um grande número de plataformas UNIX e similares (incluindo FreeBSD e Linux). Também pode ser compilado e roda em Windows 9x/NT/2000 e MacOS.

O download do R é gratuito de qualquer espelho do site www.r-project.org. Após entrar nesse site, clique em *CRAN*, logo abaixo da palavra *Download*. Em seguida, escolha um espelho perto de

você, por exemplo, o espelho da Universidade Federal do Paraná: <http://cran.br.r-project.org/>. Agora escolha seu sistema operacional, por exemplo, *Windows (95 and later)*. Aqui você opta entre baixar o conjunto de pacotes básicos ou os pacotes contribuídos. Supondo que você está baixando o R pela primeira vez, escolha a opção *base*. Nesta página o CRAN lhe apresenta uma série de opções como o *readme*, *changes*, etc. Escolha o arquivo executável, por exemplo, *R-2.7.0-win32.exe*. Pronto! É só baixar e executar.

A instalação do R é muito fácil e autodirecionada. Nas versões mais recentes é possível, inclusive, selecionar o idioma *Português (Brasil)* para as barras de menus e mensagens de erro. Porém, vale notar que, pelo menos até a versão 2.7.0, os nomes das funções, dos atributos e os *helps*, continuam em Inglês.

A cada ano, duas versões do R são disponibilizadas no CRAN, pelo menos, uma a cada semestre. Além disso, sempre existem duas versões disponíveis concomitantemente, uma versão alfa (revisada) e uma versão beta (não revisada, mas mais recente).

1.1.2 Iniciando o R

Uma vez que você chama o R, em ambiente Windows, sua tela é aberta com uma barra de menu, algumas mensagens básicas e um *prompt* vermelho (Figura 1.1).

As informações básicas se referem ao registro do R, às suas regras de distribuição, seus colaboradores, como citar o R, como pedir uma demonstração, como pedir ajuda e como sair do R.

A barra de menu traz diversos botões de atalho para a manipulação de arquivos, pacotes, ajudas, etc. O que se vê na Figura 1.1, à frente do sinal de "maior que" em vermelho, é o *prompt*. O *prompt* é herdado de linguagens como o MS-DOS e indica o ponto onde se deve inserir as linhas de comando. Lembre-se: tudo o que você disser ao R ficará impresso na tela na cor vermelha e tudo que o R lhe responder ficará impresso em azul.

Ao tentar sair do R, pela barra de menu ou pelo comando *q()*, é mostrada uma mensagem perguntando se o usuário deseja salvar a *área de trabalho*, ou seja, se os objetos atribuídos devem permanecer com os mesmo valores ou se tudo que foi feito deve ser ignorado (Figura 1.4). Quando se inicia novamente o R, após ter salvo a área de trabalho, os objetos anteriormente criados são carregados automaticamente. Aconselha-se que toda a informação desejada seja gravada em outro tipo de arquivo e a área de trabalho seja raramente salva. Isso evitará confusões quanto ao valor de objetos ao se fazer uma conta.

1.1.3 Objetos

Mais que um software que realiza análises estatísticas, R é um ambiente e uma linguagem de programação orientada a objeto. Nele, números, vetores, matrizes, *arrays*, *data frames* e listas podem ficar armazenados em objetos (Figura 1.3). Pode-se entender objeto como uma caixinha onde você pode guardar o que quiser. A partir daí todas as operações matemáticas podem ser feitas usando esses objetos. Isso torna as coisas mais simples.

Para criar um objeto é só atribuir um valor a um nome, ou seja, quando se coloca um valor dentro de um objeto, este passa a existir automaticamente. Uma atribuição pode ser feita, basicamente, de duas maneiras, usando o sinal de = ou usando uma seta formada pela junção dos

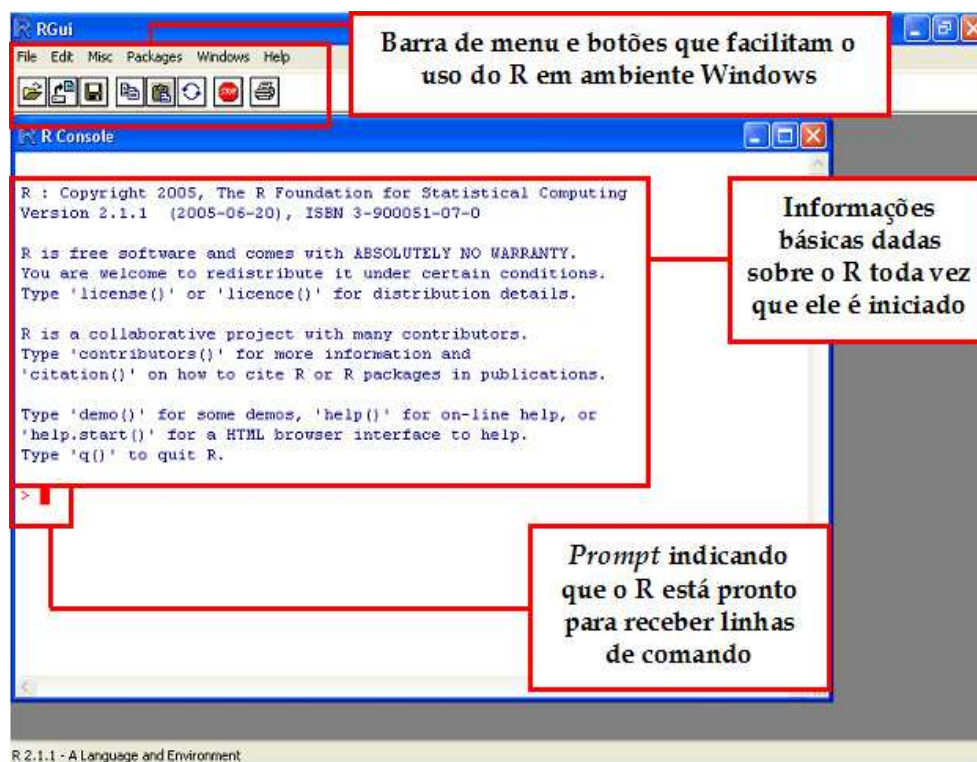


Figura 1.1: Tela inicial do R em ambiente Windows.

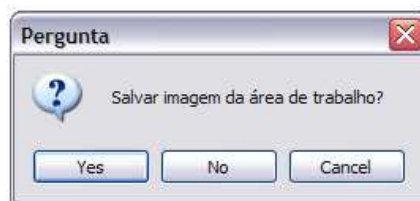


Figura 1.2: Mensagem ao tentar sair do R, em ambiente Windows.

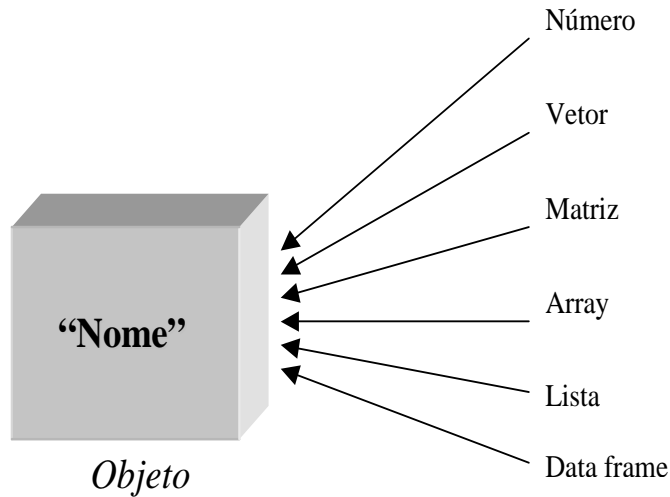


Figura 1.3: Esquema de um objeto.

siniais de *menor que* e *menos* $< -$. Note que essa seta sempre deve levar o valor ao objeto, ou seja, sempre apontar para o objeto. Portanto, é possível usar a setinha em ambas as direções ($< -$ ou $- >$).

Outro sinal muito útil na linguagem R é o sinal de comentário $\#$, ou seja, sinal a partir do qual o que for escrito não será interpretado como comando.

- (a) Número: É possível atribuir apenas um número a um objeto. Por exemplo, atribua o número 2 ao objeto *a* e o número 5 ao objeto *x*.

```
| > a<-2          #a recebe 2
| > x<-5          #x recebe 5
```

Para verificar quanto vale cada objeto, apenas digite seu nome tecle *enter*. O um entre colchetes se refere à primeira posição do vetor, ou seja, um número é entendido com o vetor de uma posição.

```
| > a
| [1] 2
```

Uma vez criados, os objetos podem ser usados em contas, equações, funções, sistemas etc.

```
| > a+x           #soma
| [1] 7
| > a-x           #subtração
| [1] -3
| > a*x           #produtos de escales
| [1] 10
| > a/x           #divisão
```

```
| [1] 0.4
| > a^x                #potenciação
| [1] 32
```

O resultado de uma conta, por sua vez, pode ser guardado dentro de um terceiro objeto.

```
| > c<-2*a + 300/x
| > c
| [1] 64
```

- (b) Vetor: O vetor da linguagem R tem um significado um pouco diferente que o vetor da Matemática. Para o R, um vetor é qualquer conjunto unidimensional de valores. Esses valores podem ser números, *strings* (palavras) ou valores lógicos (F para falso e T para verdadeiro). Em outras palavras, para o R, o vetor tem um significado mais amplo que para a Matemática.

Para se atribuir um conjunto de valores a um objeto pode-se usar o comando *c()*, onde os valores vêm separados por vírgulas, dentro dos parênteses.

```
| > d<-c(5,8,12,3.5,9,1)          #d recebe um vetor
| > d
| [1] 5.0 8.0 12.0 3.5 9.0 1.0
```

É possível se referir especificamente a uma posição do vetor. Imagine que se deseje saber qual o valor que ocupa a quarta posição do vetor *d*. Essa referência é feita enter colchetes, após o nome do objeto. | > d[4] #4a posicao do vetor d

```
| [1] 3.5
```

- (c) Matriz: Uma matriz é atribuída a um objeto pela função *matrix()*. Essa função tem como argumentos o conjunto de dados, o número de linhas e o número de colunas da matriz, nessa ordem. Note que o conjunto de dados deve ser escrito na ordem das colunas, ou seja, como se as colunas estivessem enfileiradas, umas sobre as outras. Observe o exemplo.

```
| > e<-matrix(c(5,8,12,3.5,9,1),2,3) #e recebe uma matriz
| > e
|      [,1] [,2] [,3]
| [1,] 5 12.0 9
| [2,] 8 3.5 1
```

Nas matrizes também é possível referenciar uma linha, uma coluna ou um elemento. Novamente deve-se usar números entre colchetes, porém respeitando a ordem: primeira posição se refere a linha e segunda posição se refere a coluna.

```
| > e[2,] #linha 2 da matriz e
| [1,] 8 3.5 1
| > e[,3] #coluna 3 da matriz e
```

```
| [1] 9 1
| > e[1,3] #elemento da linha 1, coluna 3
| [1] 9
```

- (d) Array: Esse termo em Inglês não possui tradução adequada. Ele representa uma hipermatriz, ou seja, um conjunto de números arranjados em mais de 2 dimensões. Quando tem 3 dimensões, um array pode ser entendido como um conjunto de matrizes de mesma dimensão.

Aqui, a referência a linhas e colunas é a mesma das matrizes, e a terceira posição dos colchetes se refere ao valor de interesse na terceira dimensão.

O comando usado é o `array()`. Uma forma de atribuir um array a um objeto é inserir um array de zeros (nas dimensões desejadas) e depois preenche-lo com os valores adequados. Outra opção é fazer um vetor que respeite a ordem: por coluna, por matriz; e usa-lo já na construção do array.

A primeira posição da função `array()` se refere aos argumentos das matrizes e a segunda posição se refere às dimensões do mesmo.

```
| > f<-array(0,c(2,2,2)) #array de zeros
| > f[,1]<-matrix(c(1,2,3,4),2,2) #primeira matriz
| > f[,2]<-matrix(c(5,6,7,8),2,2) #segunda matriz
```

ou

```
| > f<-array(c(1,2,3,4,5,6),c(2,2,2)) #de uma só vez
| > f
| , , 1
|
|      [,1] [,2]
| [1,]    1    3
| [2,]    2    4
|
| , , 2
|
|      [,1] [,2]
| [1,]    5    7
| [2,]    6    8
```

Analogamente, pode-se perguntar qual é o valor que ocupa a primeira linha, da segunda coluna, da segunda matriz, do objeto `f`.

```
| > f[1,2,2]
| [1] 7
```

- (e) Data frame: Essa estrutura de dados é uma espécie de tabela, de estrutura bidimensional de
-

dados. Podem fazer parte de um mesmo data frame números e *strings*. Além disso, podem ser dados nomes às colunas. Sua função é *data.frame()*. Veja o exemplo.

```
| > g<-data.frame('Marca'=c('Wolks','Fiat','Ford'),'Preço'=  
| + c(32000,28000,29500))  
| > g  
|   Marca Preço  
| 1 Wolks 32000  
| 2  Fiat 28000  
| 3  Ford 29500
```

- (f) Lista: Uma lista é um conjunto de objetos de tamanhos e naturezas diferentes. Ela é regida pela função *list()*. Essa é a estrutura mais geral da linguagem R. Suas posições são designadas por números entre dois parênteses `[[]]`. Considere o exemplo de lista que contém um número na primeira posição, uma matriz na segunda, uma palavra na terceira e um vetor na quarta.

```
| > h<-list(3,matrix(c(1,2,3,4),2,2),'lista',c(5,6,7,8))  
| > h  
| [[1]]  
| [1] 3  
|  
| [[2]]  
|      [,1] [,2]  
| [1,]    1    3  
| [2,]    2    4  
|  
| [[3]]  
| [1] "lista"  
|  
| [[4]]  
| [1] 5 6 7 8
```

Suponha que se deseje saber o terceiro elemento do vetor que está alocado na posição quarta posição da lista *h*.

```
| > h[[4]][3]  
| [1] 7
```

1.1.4 Lendo arquivos

A maneira mais fácil de inserir dados em objetos no R é a leitura de arquivos. O R pode ler arquivos de estruturas simples como as extensões *.txt* e *.r*. Também é possível importar outros tipos de arquivos mais complexos, como os *.xls*, mas os procedimentos de importação não são triviais, e não serão tratados aqui. O que se aconselha, quando se tem um arquivo *.xls*, é salva-lo como *.txt* e fazer a leitura normalmente.

Vale lembrar que, quando se salva uma área de trabalho, apenas os valores dos objetos são guardados. Todos os comandos dados e todos os resultados não armazenados em objetos são perdidos.

Por esses motivos, é fortemente recomendado que se trabalhe ao R em associação a um editor de texto da sua preferência. Alguns editores de texto muito úteis são: o *script* do R, o Bloco de notas do Windows, o Tinn-R, o WinEdt e o Emacs. Esses editores são usados tanto para elaborar os arquivos de dados que serão lidos pelo R, quanto para armazenar rotinas (conjuntos de linhas de comando) para a repetição futura da análise.

Para ler um arquivo no R, a função mais usada é a *read.table()*. Essa função lê o arquivo e o armazena (se desejado) na forma de data frame em um objeto. O primeiro argumento dessa função se refere ao nome do arquivo a ser lido. Esse argumento deve vir entre aspas. Entretanto, o endereço desse arquivo também deve ser passado ao R. Para isso, tem-se duas opções: (1) Na barra de menu, botão Arquivo, mudar diretório para o lugar onde se encontra o arquivo; (2) Escrever todo o endereço do arquivo dentro do primeiro argumento da função *read.table()*. O segundo argumento dessa função se refere ao cabeçalho (nome) das colunas de dados contidas no arquivo. Se as colunas tiverem cabeçalho (header), então deve-se digitar *h=TRUE*, caso contrário, *h=FALSE*.

Exemplos de comando de leitura de arquivo quando se muda o diretório de leitura para o lugar onde o arquivo está armazenado, e quando o endereço é informado na função.

```
| > read.table('nome.txt',h=TRUE)
| > read.table('C:\\Meus Documentos\\nome.txt',h=TRUE)
```

1.1.5 Funções básicas

Aqui são apresentadas algumas funções de uso constante, pertencentes aos pacotes básicos do R:

sum(x): soma todos os elementos de um objeto x.

length(x): retorna o comprimento de um objeto x.

rep(x,n): repete o número x, n vezes.

seq(a,b,by=c): gera uma seqüência de números contidos entre *a* e *b*, distantes *c* unidades um do outro.

table(x): retorna uma tabela com as freqüências absolutas de ocorrência da cada elemento de x.

1.1.6 Pedindo ajuda

O jeito mais fácil de se aprender a usar R é consultando constantemente seus tópicos de ajuda. Existem basicamente quatro tipos de ajuda no R:

- (a) `help('função()')`: Essa ajuda deve ser solicitada quando se sabe da existência de uma função (sabe-se seu nome exato), mas existe dúvidas em como usá-la. Se o pacote que contém essa função estiver instalado e carregado, será aberta a documentação da mesma para esclarecimentos.
- (b) `help.search(' ')`: Quando se deseja investigar a existência de uma função, essa ajuda recebe uma palavra-chave (em Inglês) e retorna todas aquelas funções que contém aquela palavra em sua documentação. A busca é feita nos pacotes existentes no computador em questão, ou seja, se uma busca não retornar nenhum resultado adequado, não significa que a função não exista. Significa que ela não existe, pelo menos, em seu computador.
- (c) Ajuda Html: Essa ajuda pode ser chamada pela barra de menu, no botão Ajuda (Help). Quando acionada, ela abre um documento em html que contém diversas informações sobre o R, sua linguagem, suas funções básicas, seus pacotes, seus autores, sua licença, perguntas mais freqüentes etc.
- (d) `RSiteSearch(' ')`: Quando conectado à internet, essa ajuda faz a busca de uma palavra-chave em todas as páginas da internet relacionadas com o R, principalmente aquelas páginas publicadas com as perguntas e respostas das listas de discussões do R. Existem diversos tipos de listas de discussões que podem ser encontradas na página do R. Nelas são tiradas dúvidas mais grave, são dadas sugestões para as novas versões do R, são desvendados e descoberto pequenos erros de programação etc. Elas colocam os usuários do R em contato com os estatísticos que fazem e mantêm o R.

Quando se deseja saber informações acerca de uma dada função existente deve-se digitar `help("função")` ou, simplesmente, `?função()`. Caso se deseje saber se um tópico possui função no R, o comando deve ser: `help.search("tópico")`.

1.2 Alguns conceitos importantes

Estatística é conjunto de técnicas para a coleta, organização, análise e interpretação de dados, para a descrição de populações.

Dado em Estatística, dado é o valor assumido por uma variável aleatória em um dado experimento.

População é o todo que se quer descrever. Conjunto de elementos com características em comum.

Ex.: Deseja-se saber o teor médio de açúcar (graus Brix) de uma determinada variedade de laranja.

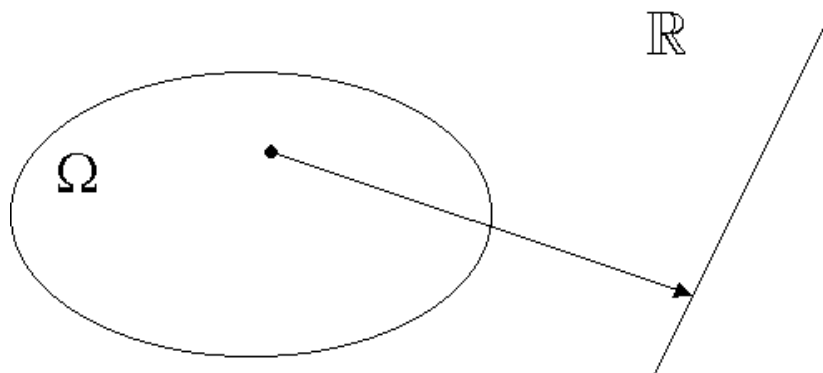
CLASSIFICAÇÕES DE UMA POPULAÇÃO


- Finita (ou real): fixa no tempo.
Ex: Conjunto de árvores em um talhão.
- Infinita (ou conceitual): engloba elementos não existentes.
Ex: Plantas da cultivar *carioca* que (existiram, existem ou virão a existir).

Espaço amostral é o conjunto de todos os resultados possíveis de serem observados num dado fenômeno, ou seja, todos os resultados possíveis de um experimento. Representado simbolicamente pela letra grega Omega maiúscula Ω .

Ex: $\Omega = \{\text{cara, coroa}\}$.

Variável aleatória é a função que associa um valor na reta real a cada ponto do espaço amostral.



Ex.: No lançamento de um dado honesto, o espaço amostral Ω é composto por suas seis diferentes faces. A cada uma delas está associado um número natural de 1 a 6. Suponha que a variável aleatória X descreve o resultado desse lançamento e o dado cai com face  virada para cima. Nesse caso associamos um valor a este evento, ou seja, $x = 4$.

Amostra é o subconjunto com n elementos da população.

Censo é a observação exaustiva de todos os N elementos da população.

Evento é cada possível resultado em um experimento. Ex: A face “cara” cair voltada para cima no lançamento de uma moeda honesta.

Estatística experimental tem por objetivo comparar mais de duas populações simultaneamente (tratamentos).

TIPOS DE VARIÁVEIS QUALITATIVAS

São aquelas variáveis que indicam qualidades, atributos, características não numéricas de forma geral.

(a) Qualitativas nominais:

São aquelas que não permitem uma ordenação natural.

Ex: O conjunto de espécies: Cedro, Cassia e Ipê.

(b) Qualitativas ordinais:

Por sua vez, são aquelas que admitem uma ordenação natural.

Ex: O ciclo de uma cultura: precoce, médio e tardio.

QUANTITATIVAS

Resumem-se a medidas, pesagens ou contagens.

(a) Quantitativas discretas:

São representadas pelas contagens.

Ex: N° de espigas por planta de milho.

(b) Quantitativas contínuas:

São representadas pelas medições ou pesagens (R)

Ex: Produtividade (t/ha).

1.3 Principais aplicações da Estatística

(a) Pesquisa científica:

A Estatística, em muitos momentos, confunde-se com o próprio conceito de fazer ciência¹. Ela tem o objetivo de administrar a incerteza acerca de um fenômeno, permitindo uma melhor compreensão do ambiente em que vivemos e permitindo a tomada de decisões e a realização de previsões.

(b) Processos produtivos:

A Estatística é usada diariamente no controle de processos produtivos empresariais por meio do Controle de Qualidade.

	Determinísticos	Principalmente usados na Física, Química e Matemática. As relações são exatas, ou seja, possíveis variações casuais são desprezadas. Ex.: $6CO_2 + 6H_2O \rightarrow C_6H_{12}O_6 + 6O_2$
Modelos		
	Probabilísticos	As variações naturais não são desprezadas e são descritas por meio de um componente probabilístico. Ex.: Peso de suínos em função da quantidade de ração ingerida ao longo de sua engorda (Figura 1.4).

(c) Levantamentos em geral:

Um dos usos mais populares da Estatística se dá em censos demográficos, pesquisas eleitorais, porém há outros levantamentos importantes para a sociedade como as pesquisas de mercado, os inventários florestais, etc.



Este gráfico foi feito no R. Quer saber como? Dê uma olha na rotina.

```
x<-seq(0,30)
y<--x^2+50*x+5
plot(x,y,'l',ylab='Peso (kg)',xlab='Ingesta de ração (kg/dia)')
e<-rnorm(length(y),0,25)
yo<-y+e
points(x,yo,pch=19)
```

¹**Ciência** é o conjunto metódico de conhecimentos obtidos mediante a observação e a experiência. Observação de evidências para refutar (ou afirmar) hipóteses sobre fenômenos naturais.

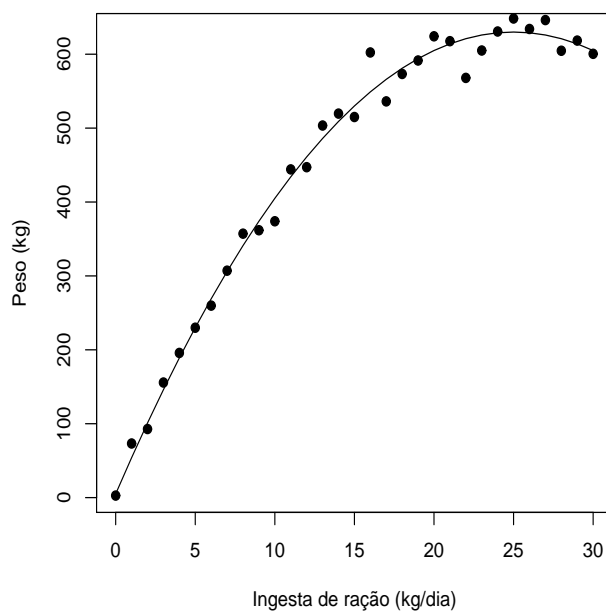
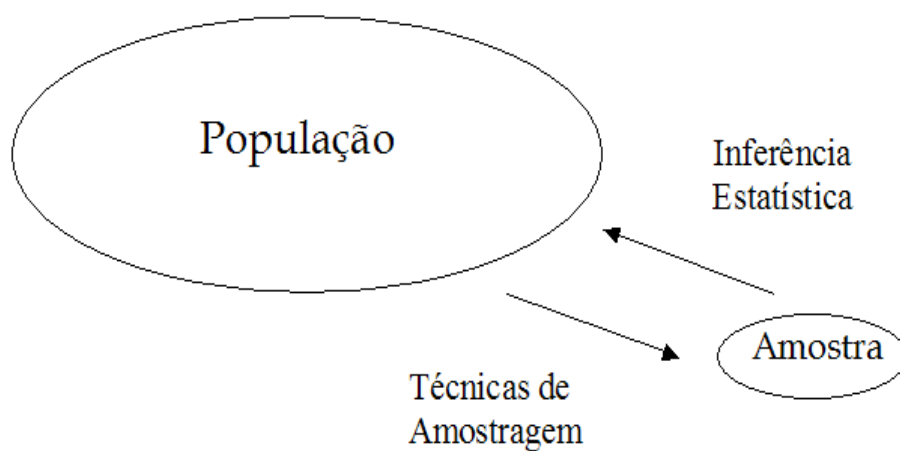


Figura 1.4: Gráfico ilustrativo do comportamento do peso de suínos (kg) em função de sua ingestão diária de ração ao longo do período de engorda.

DIAGRAMA DA ESTATÍSTICA



1.4 Bibliografia consultada e recomendada

- Bussab, W.O.; Morettin, P.A. Estatística básica. 1987. Atual.
 - Cochran, W.G. Técnicas de amostragem. 1965. Fundo de Cultura.
 - Ferreira, D. F. Estatística básica. Editora UFLA, Lavras, 2005. 676p.
 - Meyer, P. Probabilidade: aplicações à estatística. 1983. LTC.
 - Oliveira, M.S.; Bearzoti, E.; Vilas Boas, F.L.; Nogueira, D.A.; Nicolau, L.A. Introdução à Estatística. DEX/UFLA, Lavras, 2005. 329p.
 - Stevenson, W.J. Estatística aplicada à Administração. 1981. Harper & Row.
 - Spiegel, M.R. Estatística. 1993. McGraw Hill.
 - Werkema, M.C.C. Série Ferramentas de Qualidade Total. Vários volumes. Fundação Cristiano Otoni. UFMG.
 - <http://www.est.ufpr.br/~paulojus/>
 - <http://wikipedia.org/>
-

Notação: Esta é a letra grega sigma maiúscula: Σ .

Objetivo: O somatório (ou a somatória) tem por objetivo simplificar a notação de uma soma de termos, ou seja, de um polinômio.

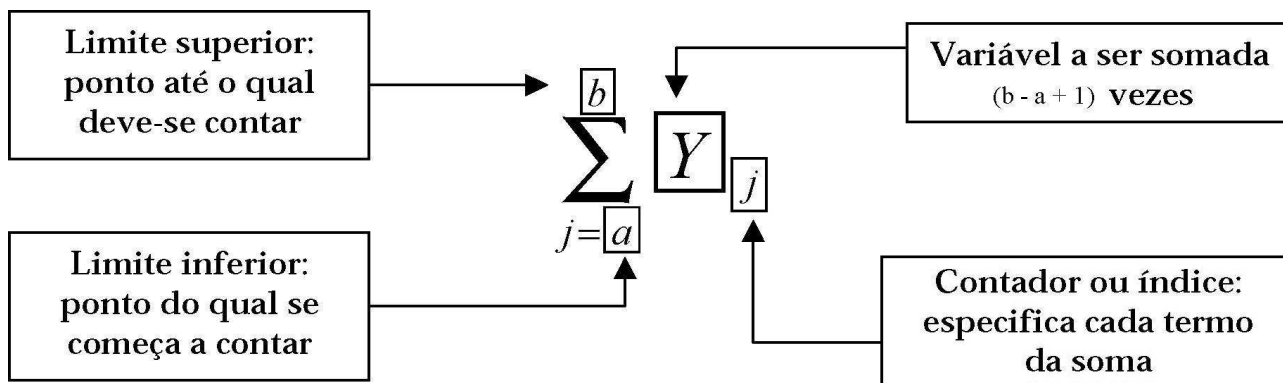


Figura 2.1: Esquema ilustrativo dos elementos típicos de um somatório.

$$\text{Ex}_1: 4 + 4 + 4 = \sum_{i=1}^3 4.$$

$$\text{Ex}_2: y_1 + y_2 + y_3 + y_4 = \sum_{i=1}^4 y_i.$$

$$\text{Ex}_3: (x_1 + y_1)^2 + (x_2 + y_2)^2 + \dots + (x_k + y_k)^2 = \sum_{i=1}^k (x_i + y_i)^2.$$

Ex₄: Seja $A = \{a_1, a_2, a_3, a_4, a_5\}$ um conjunto de dados. Então o somatório dos elementos de A pode ser escrito como: $a_1 + a_2 + a_3 + a_4 + a_5 = \sum_{i=1}^5 a_i$.

PROPRIEDADES DOS SOMATÓRIOS

1) Seja a uma constante e X uma variável aleatória, então

$$\sum_{i=1}^n ax_i = a \sum_{i=1}^n x_i. \quad (2.1)$$

Demonstração:

$$\sum_{i=1}^n ax_i = ax_1 + ax_2 + \dots + ax_n = a(x_1 + x_2 + \dots + x_n) = a \sum_{i=1}^n x_i.$$

$$\text{Ex}_5: \sum_{i=1}^3 2y_i = 2y_1 + 2y_2 + 2y_3 = 2(y_1 + y_2 + y_3) = 2 \sum_{i=1}^3 y_i.$$

2) Sejam X e Y variáveis aleatórias, então

$$\sum_{i=1}^n x_i y_i \neq \left(\sum_{i=1}^n x_i \right) \times \left(\sum_{i=1}^n y_i \right). \quad (2.2)$$

Demonstração:

$$\begin{aligned} \sum_{i=1}^n x_i y_i &\neq \left(\sum_{i=1}^n x_i \right) \times \left(\sum_{i=1}^n y_i \right) \\ x_1 y_1 + x_2 y_2 + \dots + x_n y_n &\neq (x_1 + x_2 + \dots + x_n) \times (y_1 + y_2 + \dots + y_n) \\ x_1 y_1 + x_2 y_2 + \dots + x_n y_n &\neq x_1 y_1 + x_1 y_2 + \dots + x_1 y_n + \dots + x_2 y_1 + x_2 y_2 + \dots \\ &\quad + \dots + x_2 y_n + \dots + x_n y_1 + x_n y_2 + \dots + x_n y_n \end{aligned}$$

Ex₆:

$$\begin{aligned} \sum_{i=1}^2 2x_i 3y_i &\neq \left(\sum_{i=1}^2 2x_i \right) \left(\sum_{i=1}^2 3y_i \right) \\ 2x_1 3y_1 + 2x_2 3y_2 &\neq (2x_1 + 2x_2) (3y_1 + 3y_2) \\ 2x_1 3y_1 + 2x_2 3y_2 &\neq 2x_1 3y_1 + 2x_2 3y_1 + 2x_1 3y_2 + 2x_2 3y_2. \end{aligned}$$

3) Sejam a e b constantes e X e Y variáveis aleatórias, então

$$\sum_{i=1}^n ax_i \pm by_i = a \sum_{i=1}^n x_i \pm b \sum_{i=1}^n y_i. \quad (2.3)$$

Demonstração:

$$\begin{aligned} \sum_{i=1}^n ax_i + by_i &= ax_1 + ax_2 + \dots + ax_n + by_1 + by_2 + \dots + by_n \\ &= a(x_1 + x_2 + \dots + x_n) + b(y_1 + y_2 + \dots + y_n) \\ &= a \sum_{i=1}^n x_i + b \sum_{i=1}^n y_i. \end{aligned}$$

Ex₇:

$$\begin{aligned} \sum_{i=1}^2 3x_i + 4y_i &= 3x_1 + 3x_2 + 4y_1 + 4y_2 \\ &= 3(x_1 + x_2) + 4(y_1 + y_2) \\ &= 3 \sum_{i=1}^2 x_i + 4 \sum_{i=1}^2 y_i. \end{aligned}$$

4) Seja k uma constante, então

$$\sum_{i=1}^n k = nk \quad (2.4)$$

Demonstração:

$$\begin{aligned} \sum_{i=1}^n k &= \underbrace{k + k + \dots + k}_n = nk. & \text{Ou então,} \\ \sum_{i=1}^n k &= k \sum_{i=1}^n 1 \\ &= k \underbrace{(1 + 1 + \dots + 1)}_n = nk. \end{aligned}$$

$$\text{Ex}_7: \sum_{i=1}^4 5 = \underbrace{5 + 5 + \dots + 5}_4 = 4 \times 5 = 20.$$



O R possui uma função básica chamada *sum()*. Essa função faz a soma de todos os elementos de um objeto que for seu argumento: *sum(objeto)*. Por exemplo:

```
x<-c(4,3,6,2,1,3,2,4,5)
sum(x)
```

Se você desejar somar apenas alguns valores de seu objeto (por exemplo, os valores da posição a até a posição b, em um vetor), é só indicar o intervalo desejado da seguinte maneira: *sum(objeto[a:b])*.

```
sum(x[2:5])
```

3.1 Introdução

Um bom trabalho de coleta de dados experimentais pode render uma massa de dados confiável, porém desordenados, isto é, brutos. Na sua forma bruta os dados não querem dizer muita coisa, isto é, não são considerados *informação*. Por isso, o objetivo da Estatística Descritiva é apresentar uma série de técnicas de descrição de dados válidas para censos e amostras.

ALGUMAS DEFINIÇÕES PERTINENTES

Frequência de forma geral, indica com que frequência determinado valor (ou intervalo de valores) ocorre na massa de dados.

Distribuição de frequência é a função que associa valores da variável com suas frequências de ocorrência.

Tipos de frequência

- (a) Frequência absoluta (fa): representa o número de vezes que um valor (ou intervalo) ocorre nos dados.
- (b) Frequência relativa (fr): representa, em forma decimal, a proporção de ocorrências de um valor em relação ao tamanho da massa de dados,

$$fr = \frac{fa}{n}. \quad (3.1)$$

- (c) Frequência percentual (fp): representa, em forma percentual, a proporção de ocorrências de um valor em relação ao tamanho da massa de dados,

$$fp = fr \times 100. \quad (3.2)$$

3.2 Variáveis qualitativas

Experimentos ou pesquisas que possuem como foco variáveis qualitativas podem ser descritos (resumidos) por meio de distribuições de frequência e suas representações gráficas. A seguir, um exemplo ilustra o procedimento.

Ex.: Um engenheiro agrônomo faz um levantamento das principais atividades agrícolas em uma amostra contendo 20 propriedades de certa região. O croquis a seguir representa esquematicamente o resultado da pesquisa.

C	L	L	C	S	LA	C	C	L	M
C	M	So	M	L	C	C	M	C	L

- Massa de dados: amostra.
- População: finita: conjunto de todas as propriedades rurais desta região que atualmente apresentam atividades agrícolas.
- Variável aleatória: qualitativa nominal: atividade agrícola.
- Valores assumidos pela variável aleatória na pesquisa: café (C), leite (L), silvicultura (S), milho (M), soja (So), laranja (LA).
- Distribuição de frequência:

Tabela 3.1: Distribuição de frequências absolutas (fa), relativa (fr) e percentual (fp) da atividade em propriedades de uma região.

Atividade	fa	fr	fp (%)
Café	8	0,40	40,00
Leite	5	0,25	25,00
Milho	4	0,20	20,00
Outras	3	0,15	15,00
Total	20	1,00	100,00

Fonte: Dados fictícios.

Nota: classes pouco frequentes podem ser agrupadas em uma categoria “outras”, em último lugar.



Pode-se facilmente fazer uma distribuição de frequências no R compondo-se um objeto (*df*) de forma conveniente. No exemplo:

```
at<-c('C','L','L','M','C','M','So','L','L','C','M','C','S','L','C','LA','C','M',
      'C','C')
tab.at<-table(at)
df<-matrix(0,5,3)
colnames(df)<-c("fa","fr","fp")
rownames(df)<-c("Café","Leite","Milho","Outras","Total")
df[1,1]<-tab.at["C"]
df[2,1]<-tab.at["L"]
df[3,1]<-tab.at["M"]
df[4,1]<-sum(tab.at["So"], tab.at["S"], tab.at["LA"])
df[5,1]<-length(at)
for(i in 1:5) {df[i,2]<-df[i,1]/length(at)}
for(i in 1:5) {df[i,3]<-df[i,2]*100}
```

REPRESENTAÇÕES GRÁFICAS DA DISTRIBUIÇÃO DE FREQUÊNCIA

GRÁFICO DE BARRAS E COLUNAS

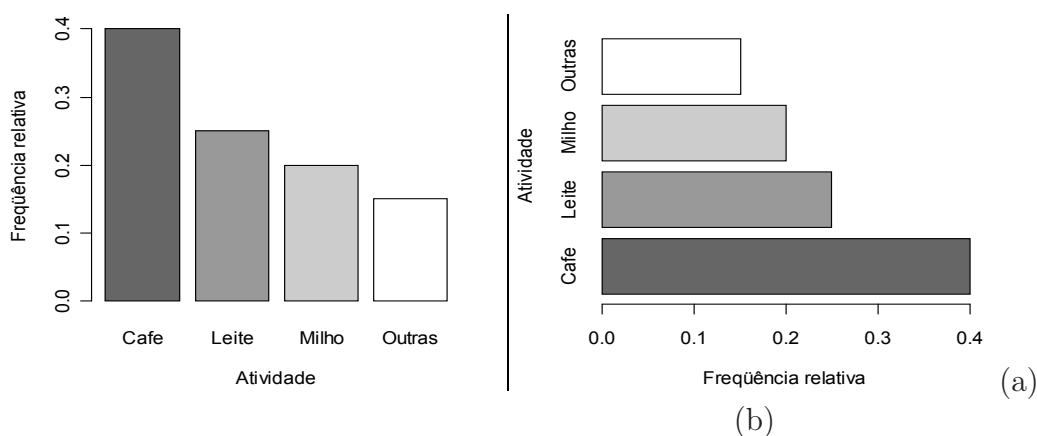


Figura 3.1: (a) Gráfico de colunas das principais atividades em propriedades rurais. (b) Gráfico de barras da mesma situação.



Reproduza em seu computador a rotina do gráfico de barras...

```
gc<-barplot(df[1:4,2],xlab="Atividade",ylab="Frequência relativa",  
            col = gray(seq(0.4,1.0,length=4)))
```

...e do gráfico de colunas

```
gc<-barplot(df[1:4,2],horiz=TRUE,ylab="Atividade",xlab="Frequência relativa",  
            col = gray(seq(0.4,1.0,length=4)))
```

Note que os gráficos de barras e colunas são feitos com a mesma função (*barplot*), a única diferença é o argumento *horiz* que deve ser verdadeiro no caso das barras. Mas lembre-se, troque o nome dos eixos ao inverter o gráfico ou eles ficarão trocados.

GRÁFICO DE PIZZA OU SETOGRAMA

O gráfico de pizza, torta ou setograma é um círculo com setores de área proporcional às frequências de ocorrência de cada valor da variável aleatória.

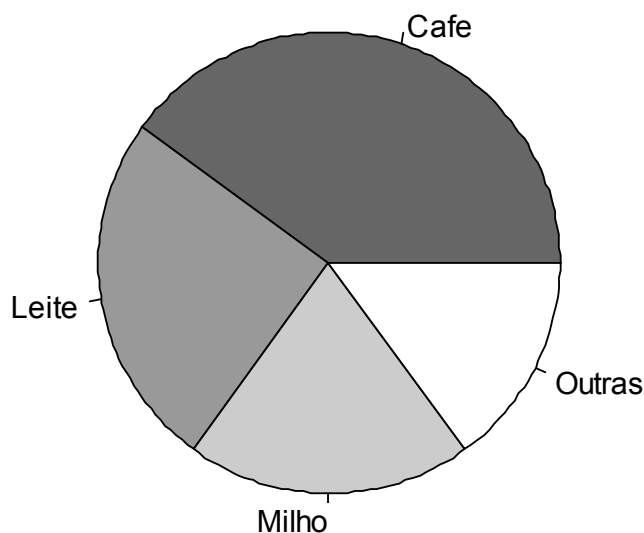


Figura 3.2: Setograma ou gráfico de pizza.



Confira como fazer um setograma no R:

```
pie(df[1:4,2], col = gray(seq(0.4,1.0, length=4)), radius = 1.05)
```

A função *pie* exige como argumento um objeto contendo números decimais que somem 1, ou seja, frequências relativas.

3.3 Variáveis quantitativas discretas

Variáveis quantitativas discretas podem ser vistas como casos particulares de variáveis quantitativas contínuas. Pode-se tratar uma massa de dados de variáveis quantitativas discretas como se fosse de variáveis qualitativas, ou seja, cada valor assumido pela variável pode ser visto como uma classe. Porém, quando a variável, apesar de assumir valores discretos, puder assumir uma quantidade muito grande de valores, ela pode ser tratada como uma variável quantitativa contínua, ou seja, construindo-se classes. Os procedimentos indicados para a manipulação de variáveis quantitativas contínuas serão apresentados a seguir.

A representação gráfica das variáveis quantitativas discretas se dá de forma semelhante a das qualitativas ordinais.

Veja o seguinte exemplo real: uma pesquisa da Secretaria de Saúde Pública de um município investigou o número de filhos por casal. A seguir está apresentada uma parte dos resultados obtidos:

3	4	3	1	3	2	1	1	2	2	4	4	1	3	2	2	4	4	3	3
1	0	2	1	3	2	2	4	2	1	1	4	1	0	1	3	3	0	3	3

A Tabela 3.2 apresenta a distribuição de frequência do número de filhos por casal em um determinado município.

Tabela 3.2: Distribuição de frequências absolutas (fa), relativa (fr) e percentual (fp) do número de filhos por casal de uma cidade.

Classes	fa	fr	fp (%)
0	3	0,075	7,50
1	10	0,250	25,00
2	9	0,225	22,50
3	11	0,275	27,50
4	7	0,175	17,50
Total	40	1,000	100,00

Fonte: Dados fictícios.



De forma semelhante, pode-se fazer uma distribuição de frequências no R compondo-se um objeto (*df*). No exemplo:

```
filhos<-c(3,4,3,1,3,2,1,1,2,2,4,4,1,3,2,2,4,4,3,3,1,0,2,1,3,2,2,4,2,1,1,4,1,0,
1,3,3,0,3,3)
tab.filhos<-table(filhos)
df<-matrix(0,6,3)
colnames(df)<-c("fa","fr","fp")
rownames(df)<-c(0,1,2,3,4,"Total")
df[1,1]<-tab.filhos["0"]
df[2,1]<-tab.filhos["1"]
df[3,1]<-tab.filhos["2"]
df[4,1]<-tab.filhos["3"]
df[5,1]<-tab.filhos["4"]
df[6,1]<-length(filhos)
for(i in 1:6) {df[i,2]<-df[i,1]/length(filhos)}
for(i in 1:6) {df[i,3]<-df[i,2]*100}
```

GRÁFICO DE LINHAS

Uma das formas de representar graficamente a distribuição de frequências de variáveis quantitativas discretas é o gráfico de linhas. Ao contrário do que se costuma chamar de gráfico de linhas, esse é um gráfico que representa as alturas de cada ocorrência da variável por meio de linhas (Figura 3.3). Esse é o limite do gráfico de colunas quando a largura da coluna tende a zero. Isso faz sentido já que, nesse caso, a classe se resume a um ponto, ou seja, a amplitude de classe (c) é zero.

Apesar de não fazer muito sentido, o gráfico de colunas tem sido muito utilizado para representar variáveis quantitativas discretas devido a seu apelo visual. A Figura 3.3 também apresenta o gráfico de colunas para esse exemplo.



Representação da distribuição de frequências de uma variável qualitativa:

```
gl<-plot(df[1:5,2], type="h", xlab="Número de filhos",ylab="fr") #linhas
gb<-barplot(df[1:5,2],col=gray(seq(0.4,1.0,length=5)),xlab="Número de filhos",
ylab="fr") #colunas
```

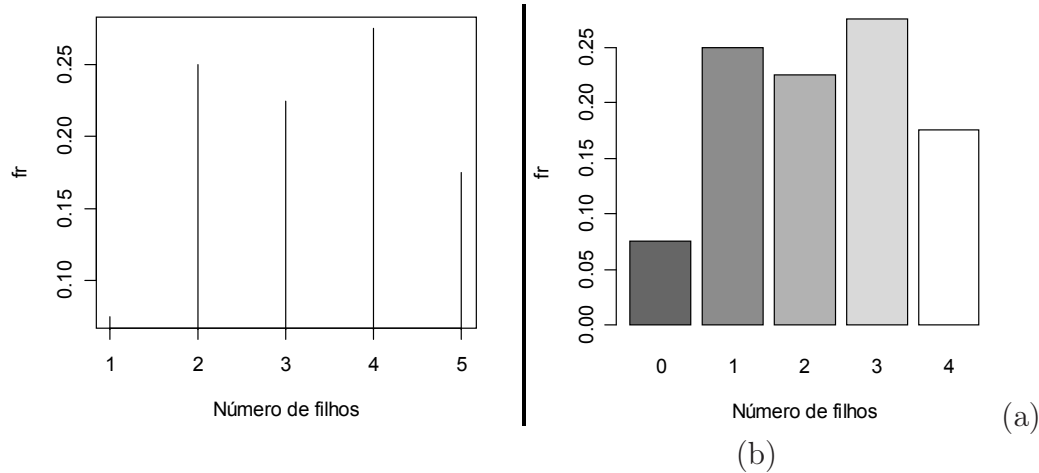



Figura 3.3: (a) gráfico de linhas da variável número de filhos por casal. (b) gráfico de colunas da mesma variável.

3.4 Variáveis quantitativas contínuas

Aqui se descreve uma sequência de passos indicados para a construção de uma distribuição de frequências para variáveis quantitativas contínuas. No entanto, é importante ressaltar que essa é apenas uma de uma infinidade de maneira que se poderia construir uma distribuição de frequência eficiente e compreensível. Portanto, na literatura especializada facilmente pode-se encontrar diferentes sugestões de procedimento.

- (1) Determinar o número de classes (k):

<p>Critério empírico</p> $k \simeq \begin{cases} \sqrt{n}, & \text{se } n < 100 \\ 5 \log n, & \text{se } n > 100 \end{cases}$	<p>Critério de Scott (1979)</p> $k \simeq 1 + \frac{An^{1/3}}{3,495}$
---	--

n : número de elementos da amostra.

- (2) Cálculo da Amplitude Total (A):

$$A = MVO - mvo, \quad (3.3)$$

em que MVO é o maior valor observado; e mvo é o menor valor observado.

- (3) Cálculo da amplitude de classe (c):

Sejam as seguintes fórmulas, se a massa de dados em questão se tratar de censos

$$c = \frac{A}{k} \quad (3.4)$$

ou amostras

$$c = \frac{A}{k-1}. \quad (3.5)$$

(4) Limite inferior da primeira classe (LI_1):

Em censos:

$$LI_1 = mvo \quad (3.6)$$

ou amostras

$$LI_1 = mvo - \frac{c}{2}. \quad (3.7)$$

(5) Demais limites:

$$LS_i = LI_i + c \quad (3.8)$$

e

$$LS_i = LI_{i+1}, \quad (3.9)$$

para todo $i = 1, \dots, k$.

Para ilustrar a seqüência de passos descrita, considere o seguinte exemplo: Em uma linha de envasamento de potinhos de canela em pó, a especificação é enchê-los com 50g do produto. Se a envasadora colocar mais que o especificado, a empresa estará sendo lesada. Caso contrário, o consumidor será enganado. Por isso, é conveniente fazer o acompanhamento dos potinhos envasados. Coletou-se uma amostra de 50 potinhos dessa linha de produção, que aqui são dispostos em ordem crescente, em g.

45,2	45,3	45,4	45,7	45,9	46,1	46,1	46,2	46,5	46,6	46,9	47,9	48,1	48,1	48,3
48,5	48,8	48,8	49,1	49,2	49,3	49,7	49,8	49,9	50,1	50,2	50,3	50,4	50,5	50,5
50,5	50,6	50,8	51,0	51,1	51,4	51,6	51,6	51,7	51,9	52,5	52,7	52,8	53,0	54,9
55,0	55,2	55,3	55,7	55,7										

Portanto,

$$(1) n < 100 \Rightarrow k = \sqrt{n} = \sqrt{50} = 7,07 \sim 7 \text{ classes.}$$

$$(2) A = 55,7 - 45,2 = 10,5g$$

$$(3) c = \frac{A}{k-1} = \frac{10,5}{7-1} = 1,75g$$

$$(4) LI_1 = 45,2 - \frac{1,75}{2} \sim 44,33g$$

Tabela 3.3: Distribuição de frequências absoluta (fa), relativa (fr) e percentual (fp) do peso observado em potinhos de canela em pó.

Classes	fa	fr	fp (%)
[44, 33; 46, 08)	5	0,10	10,0
[46, 08; 47, 83)	6	0,12	12,00
[47, 83; 49, 58)	10	0,20	20,00
[49, 58; 51, 33)	14	0,28	28,00
[51, 33; 53, 08)	9	0,18	18,00
[53, 08; 54, 83)	0	0,00	00,00
[54, 83; 56, 58)	6	0,12	12,00
Total	50	1,000	100,00

Fonte: Dados fictícios.



Veja como construir uma distribuição de frequências de uma variável quantitativa contínua:

```
canela<-c(45.2,45.3,45.4,45.7,45.9,46.1,46.1,46.2,46.5,46.6,46.9,47.9,48.1,48.1,
48.3,48.5,48.8,48.8,49.1,49.2,49.3,49.7,49.8,49.9,50.1,50.2,50.3,50.4,50.5,
50.5,50.5,50.6,50.8,51.0,51.1,51.4,51.4,51.6,51.7,51.9,52.5,52.7,52.8,53.0,
54.9,55.0,55.2,55.3,55.7,55.7)
df<-matrix(0,8,3)
colnames(df)<-c("fa","fr","fp")
rownames(df)<-c('[44,33;46,08)', '[46,08;47,83)', '[47,83;49,58)', '[49,58;51,33)',
'[51,33;53,08)', '[53,08;54,83)', '[54,83;56,58)', 'Total')1
tab.canela<-table(cut(canela,breaks=c(44.33,46.08,47.83,49.58,51.33,53.08,
54.83,56.58)))
df[1:7,1]<-tab.canela
df[8,1]<-length(canela)
for(i in 1:8) {df[i,2]<-df[i,1]/length(canela)}
for(i in 1:8) {df[i,3]<-df[i,2]*100}
```

HISTOGRAMA

A representação gráfica mais usada para representar variáveis quantitativas contínuas é o histograma. Histogramas são gráficos de barras verticais justapostas em um eixo contínuo. Neles, o eixo x recebe a variável em estudo, ou seja, abriga as classes. A largura das colunas representa a amplitude das classes. O eixo y recebe as frequências (absolutas, relativas, percentuais ou densidades de frequência).

Densidades de frequência são razões entre as frequências de ocorrência e as amplitudes de classe. Elas *traduzem* o que realmente acontece nas classes quando elas possuem amplitudes diferentes,

$$dfi_i = \frac{f_i}{c_i}, \quad (3.10)$$

com i variando da classe 1 à classe k .

Podem ser calculadas densidades de frequência absolutas, relativas ou percentuais, de acordo com o interesse, dividindo-se as respectivas frequências pelas amplitudes. Contudo, aconselha-se o uso das densidades de frequência relativas (dfr), pois no histograma, quando a altura das colunas representa a dfr, a área corresponde à frequência relativa ou probabilidade.

Nota: Aconselha-se evitar construir classes vazias, pois elas são pouco informativas. No exemplo da canela em pó, as classes 6 e 7 podem ser fundidas em uma só classe. Dessa forma, a frequência relativa passa a valer 0,12 e a amplitude de classe vale 3,5g. Então, por exemplo, calculando-se a dfr, tem-se

$$dfr_6 = \frac{0,00 + 0,12}{3,5} = 0,0343.$$

Dessa maneira pode-se construir o histograma referente ao exemplo (Figura 3.4).

Outro dispositivo visual comumente usado é o *polígono de frequência*. O polígono de frequência nada mais é do que a união, por meio de segmentos de reta, dos pontos médios das classes (Figura 3.4).

Podem ser úteis também as frequências absolutas acumuladas *para cima* (ou *acima de*) e *para baixo* (ou *abaixo de*). Podendo informar, por exemplo, quantos potinhos de canela em pó contêm menos de 48g. Uma tabela pode ser construída para demonstrar essas frequências explicitando-se os limites das classes e quantos elementos da amostra se encontram abaixo ou acima daquele valor (Tabela 3.4). Os dispositivos gráficos usados para representá-las chamam-se *ogivas* (Figura 3.5).

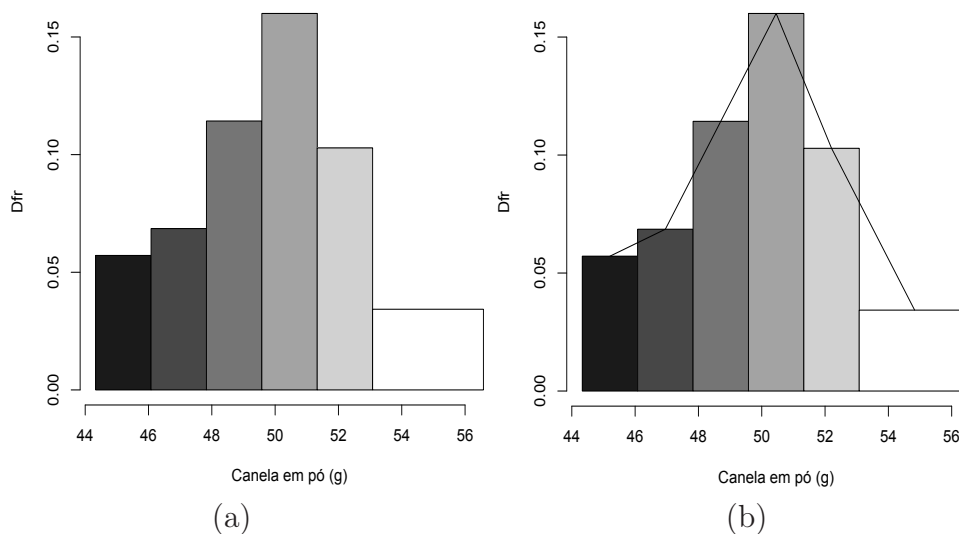


Figura 3.4: (a) Histograma do peso de potinhos de canela em pó em uma linha de produção. (b) O mesmo histograma com polígono de frequência.

Tabela 3.4: Freqüências absolutas acumuladas abaixo ($Fa\downarrow$) e acima de ($Fa\uparrow$).

Limite de classe (g)	$Fa\downarrow$	$Fa\uparrow$
44,33	0	50
46,08	5	45
47,83	11	39
49,58	21	29
51,33	35	15
53,08	44	6
54,83	44	6
56,58	50	0

Fonte: Dados fictícios.



A rotina para construir histogramas usa a função `hist()` do R. Nela atributos como a densidade de freqüência relativa e cores das colunas pode ser facilmente modificados. Figura 3.4(a):

```
h<-hist(canela,breaks=c(44.33,46.08,47.83,49.58,51.33,53.08,56.58),freq=FALSE,
        ylab="Dfr",xlab="Canela em pó (g)",main="",col=gray(seq(0.1,1.0,length=6)))
```

Figura 3.4(b) (além da rotina para a Figura 3.4(a)):

```
points(h$mids,h$density,"l")
```



A seguir, a rotina usada para construir a Figura 3.5.

```
lim<-c(44.33,46.08,47.83,49.58,51.33,53.08,54.83,56.58)
ab<-c(0, 5, 11, 21, 35, 44, 44, 50)
ac<-c(50, 45, 39, 29, 15, 6, 6, 0)
plot(lim,ab,'l',ylab='Freqüência acumulada',xlab='Canela em pó (g)')
points(lim,ac,"l")
```

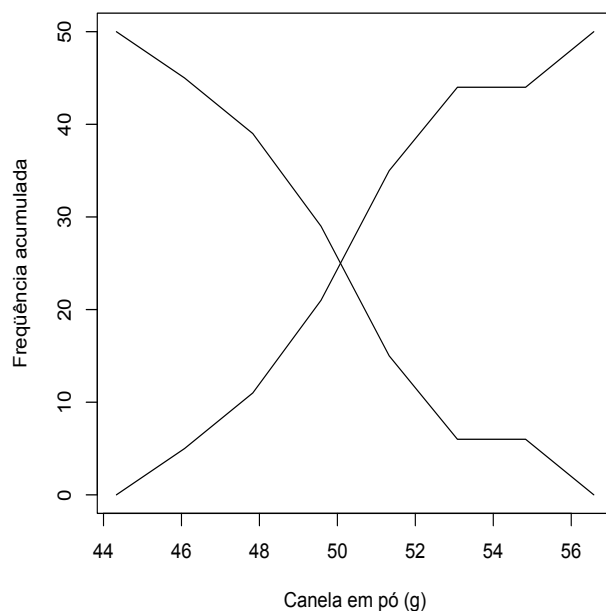


Figura 3.5: Ogivas representando as frequências absolutas acumuladas *acima de* e *abaixo de* e seu respectivo código em R.

3.5 Medidas de posição

Quando se tratam de variáveis quantitativas, os dados podem ser resumidos sob a forma de *distribuições de frequência* ou por *medidas descritivas*. Medidas descritivas são formas de, em um único número, tentar expressar a informação trazida pelos dados.

As duas categorias de medidas descritivas são: *medidas de posição* e *medidas de dispersão*.

As medidas de posição indicam a posição global dos dados na escala de valores possíveis.

3.5.1 Média (Me)

Outras notações que você pode encontrar são: μ , μ_r e \bar{Y} .

Dados não agrupados Dados agrupados

$$\bar{Y} = \sum_{i=1}^n \frac{Y_i}{n} \qquad \bar{Y} = \sum_{i=1}^k fr_i m_i$$

sendo m_i o ponto central da classe i ,

$$m_i = \frac{LS_i + LI_i}{2}. \quad (3.11)$$

PROPRIEDADES DA MÉDIA

Sejam X e Y variáveis aleatórias e k uma constante.

(1) Se $X = Y + k$, então $\bar{X} = \bar{Y} + k$.

Demonstração:

$$\begin{aligned}\bar{X} &= \frac{\sum_{i=1}^n X_i}{n} \\ &= \frac{x_1 + x_2 + \dots + x_n}{n} \\ &= \frac{(y_1 + k) + (y_2 + k) + \dots + (y_n + k)}{n} \\ &= \frac{(y_1 + y_2 + \dots + y_n) + (k + \dots + k)}{n} \\ &= \frac{(y_1 + y_2 + \dots + y_n) + nk}{n} \\ &= \frac{\sum_{i=1}^n Y_i}{n} + k \\ &= \bar{Y} + k\end{aligned}$$

(2) Se $X = Y \times k$, então $\bar{X} = \bar{Y} \times k$

Demonstração:

$$\begin{aligned}\bar{X} &= \frac{\sum_{i=1}^n X_i}{n} \\ &= \frac{x_1 + x_2 + \dots + x_n}{n} \\ &= \frac{(y_1 k) + (y_2 k) + \dots + (y_n k)}{n} \\ &= \frac{y_1 + y_2 + \dots + y_n}{n} \times k \\ &= \frac{\sum_{i=1}^n Y_i}{n} \times k \\ &= \bar{Y} \times k\end{aligned}$$

(3) Seja $e_i = y_i - \bar{y}$ o i -ésimo desvio, então, para $i = 1, \dots, n$, $\sum_{i=1}^n e_i = 0$.

Demonstração:

$$\begin{aligned}
 \sum_{i=1}^n e_i &= \sum_{i=1}^n (y_i - \bar{y}) \\
 &= \sum_{i=1}^n y_i - \sum_{i=1}^n \bar{y} \\
 &= \sum_{i=1}^n y_i - \sum_{i=1}^n \left(\sum_{i=1}^n \frac{y_i}{n} \right) \\
 &= \sum_{i=1}^n y_i - n \sum_{i=1}^n \frac{y_i}{n} \\
 &= \sum_{i=1}^n y_i - \sum_{i=1}^n y_i = 0
 \end{aligned}$$

3.5.2 Mediana (Md)

É aquele elemento que ocupa a posição central, ou seja, divide a massa de dados em duas partes iguais.

Dados não agrupados (porém, ordenados) **Dados agrupados**

$$Md(Y) = \begin{cases} y_{\frac{n+1}{2}}, & \text{se } n \text{ ímpar} \\ \frac{y_{\frac{n}{2}} + y_{\frac{n}{2}+1}}{2}, & \text{se } n \text{ par} \end{cases} \quad \begin{array}{l} \text{É o valor que separa a área do gráfico em} \\ \text{duas partes iguais.}^2 \end{array}$$

Ex.: $Y = \{3, 5, 6, 8, 9\} \rightarrow Md(Y) = 6$ e $\bar{y} = 6,2$

PROPRIEDADES DA MEDIANA

(1) Se $X = Y + k$, então, $Md(X) = Md(Y) + k$.

Demonstração:

• Se n é ímpar:

$$\begin{aligned}
 Md(X) &= x_{\frac{n+1}{2}} \\
 &= y_{\frac{n+1}{2}} + k \\
 &= Md(Y) + k
 \end{aligned}$$

- Se n é par:

$$\begin{aligned} Md(X) &= \frac{y_{\frac{n}{2}} + y_{\frac{n}{2}+1}}{2} \\ &= \frac{x_{\frac{n}{2}} + k + x_{\frac{n}{2}+1} + k}{2} \\ &= \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1} + 2k}{2} \\ &= \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} + k \\ &= Md(Y) + k \end{aligned}$$

(2) Se $X = Y \times k$, então, $Md(X) = Md(Y) \times k$.

Demonstração:

- Se n é ímpar:

$$\begin{aligned} Md(X) &= x_{\frac{n+1}{2}} \\ &= y_{\frac{n+1}{2}} \times k \\ &= Md(Y) \times k \end{aligned}$$

- Se n é par:

$$\begin{aligned} Md(X) &= \frac{y_{\frac{n}{2}} + y_{\frac{n}{2}+1}}{2} \\ &= \frac{(x_{\frac{n}{2}} \times k) + (x_{\frac{n}{2}+1} \times k)}{2} \\ &= \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} \times k \\ &= Md(Y) \times k \end{aligned}$$

3.5.3 Moda (Mo)

É valor mais freqüente, aquele que mais se repete.

Variáveis discretas

Verifica-se o valor que mais se repete.

Variáveis contínuas

Aconselha-se trabalhar com dados agrupados, pelo método de Czuber.

MÉTODO DE CZUBER

O método de Czuber permite encontrar-se a moda em dados agrupados. Como era de se esperar, a moda estará contida na classe mais freqüente ou, no histograma, a coluna mais alta. Essa classe recebe o nome de classe modal. Dentro da classe modal a moda se situará mais próximo àquela classe adjacente que for mais consecutivamente mais alta. Analise a fórmula e entenda sua lógica no histograma ilustrativo da Figura 3.6.

$$Mo(Y) = LI_M + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times c_M, \quad (3.12)$$

sendo $\Delta_1 = Dfr_M - Dfr_{M-1}$ e $\Delta_2 = Dfr_M - Dfr_{M+1}$.

Em que LI_M é o limite inferior da classe modal; Dfr_M é a densidade de freqüência relativa da classe modal; Dfr_{M-1} é a densidade de freqüência relativa da classe anterior à modal; Dfr_{M+1} é a densidade de freqüência relativa da classe posterior à modal; c_M : amplitude da classe modal.

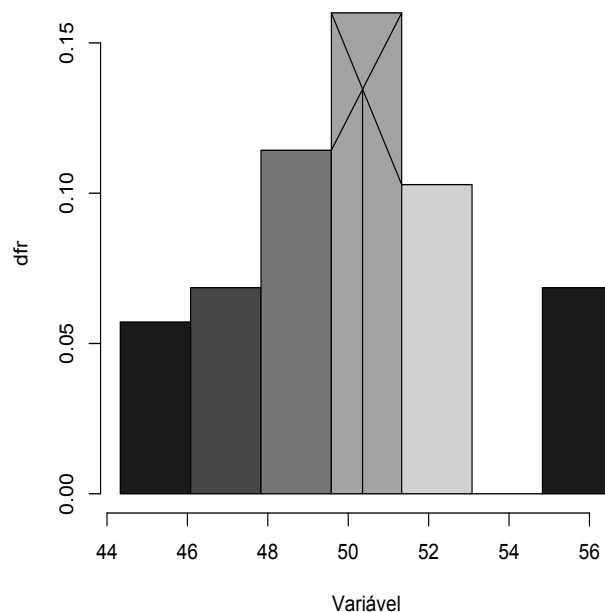


Figura 3.6: Histograma ilustrando geometricamente método de Czuber.



Esta é a rotina usada para fazer o gráfico da Figura 3.6:

```
h<-hist(canela,breaks=c(44.33,46.08,47.83,49.58,51.33,53.08,54.83,56.58),
main='',freq=FALSE,ylab='dfr',xlab='Variável',col=gray(seq(0.1,1.0,length=6)))
points(c(49.58,51.33),c(h$density[3],h$density[4]),"l")
points(c(49.58,51.33),c(h$density[4], h$density[5]),"l")
points(c(50.357,50.357), c(0.134,0),"l")
```

PROPRIEDADES DA MODA

(1) Se $X = Y + k$, então, $Mo(X) = Mo(Y) + k$.

Demonstração:

$$\begin{aligned} Mo(X) &= x \text{ mais freqüente.} \\ &= y \text{ mais freqüente} + k \\ &= Mo(Y) + k \end{aligned}$$

(2) Se $X = Y \times k$, então, $Mo(X) = Mo(Y) \times k$.

Demonstração:

$$\begin{aligned} Mo(X) &= x \text{ mais freqüente.} \\ &= y \text{ mais freqüente} \times k \\ &= Mo(Y) \times k \end{aligned}$$



Das medidas de posição apresentadas, apenas a moda não se encontra implementada nos comandos básicos do R.

```
ex<-c(3,5,6,8,9)
mean(ex)          #média
median(ex)        #mediana
```

INFLUÊNCIA SOFRIDA POR DADOS EXTREMOS:

$$Me > Md > Mo$$

3.6 Medidas de dispersão

As medidas de dispersão indicam quanto os dados variam.

Considere o seguinte exemplo: Imagine três situações distintas em que certa variável X é medida em quatro observações (quatro elementos em cada amostra). Os três conjuntos de dados resultantes estão explicitados na tabela a seguir. Caso você aplique nesses conjuntos de dados as medidas de posição conhecidas até o momento, o resultado será exatamente o mesmo, sugerindo que as três massas de dados são iguais. Mas isso claramente não é verdade! O que fazer?

Observação	I	II	III
1	100	80	10
2	100	100	100
3	100	100	100
4	100	120	190
\bar{x}	100	100	100
$Md(x)$	100	100	100
$Mo(x)$	100	100	100

Fica claro que as medidas de posição, por si só, não são suficientes para descrever um conjunto de dados. No exemplo, os três conjuntos de dados diferem quanto à *variabilidade*. Por exemplo, o conjunto III varia muito mais que os outros dois. Aí está evidenciada a importância das medidas de dispersão ou variabilidade.

3.6.1 Amplitude (A)

Amplitude total (ou simplesmente Amplitude), como já mencionando na construção de histogramas, é o intervalo total de variação dos dados.

$$A = MVO - mvo$$

No exemplo,

	I	II	III
Amplitude	$100 - 100 = 0$	$120 - 80 = 40$	$190 - 10 = 180$

Os conjuntos já começam a mostrar suas diferenças. Mas há uma desvantagem: *amplitudes só podem ser comparadas se os conjuntos tiverem o mesmo número de dados*. É intuitivo que se dois conjuntos apresentam números de elementos diferentes, o conjunto maior tem mais chance de ter uma amplitude também maior. Nesse caso, a diferença entre as amplitudes dos conjuntos refletiria a diferença no número de elementos e não a variabilidade dos dados.

Além disso, essa é uma medida de dispersão limitada, pois só leva em conta valores os extremos.

Considere agora os conjuntos de dados:

I	5	15	15	15	40	$A_I = 35$
II	5	10	20	30	40	$A_{II} = 35$

Os conjuntos são diferentes, apresentam variabilidades diferentes, porém a amplitude não conseguiu detectar esse fato.

PROPRIEDADES DA AMPLITUDE

(1) Se $X = Y + k$, então, $A(X) = A(Y)$.

Demonstração:

$$\begin{aligned} A(X) &= MVO_x - mvo_x \\ &= (MVO_y + k) - (mvo_y + k) \\ &= MVO_y - mvo_y = A(Y) \end{aligned}$$

(2) Se $X = Y \times k$, então, $A(X) = A(Y) \times k$.

Demonstração:

$$\begin{aligned} A(X) &= MVO_x - mvo_x \\ &= (MVO_y \times k) - (mvo_y \times k) \\ &= (MVO_y - mvo_y) \times k \\ &= A(Y) \times k \end{aligned}$$

3.6.2 Variância e Desvio Padrão

A variância e o desvio padrão são as duas medidas de dispersão mais usadas. Elas são grandezas proporcionais, por isso serão tratadas em um mesmo tópico.

Ambas se valem de todas as observações para calcular suas quantidades e se baseiam no desvio em relação à média

$$e_i = y_i - \bar{y}.$$

VARIÂNCIA

Notação: normalmente, a variância da população é designada pela letra grega sigma minúsculo ao quadrado (σ^2); e a variância da amostra, pela letra S^2 (quando se tratar da variável aleatória) e s^2 (quando se tratar de uma estimativa de S^2).

Se X é uma variável aleatória $V(X)$ também denota a Variância de X .

Censos	Amostras	Dados agrupados
$\sigma^2 = \frac{\sum_{i=1}^n e_i^2}{N}$	$S^2 = \frac{\sum_{i=1}^n e_i^2}{n - 1}$	$\sigma^2 = \sum_{i=1}^k fr_i (m_i - \bar{y})^2$

Sendo m_i o ponto médio da classe i ($i = 1, 2, \dots, k$).

No exemplo,

	I	II	III
Variância	0	266,67	5400

Nesse caso, a variância identificou a diferença na variabilidade dos conjuntos, porém seu valor absoluto não é interpretável praticamente porque ela se expressa no quadrado da unidade dos dados. Por exemplo, o peso de um grupo de bovinos alimentados com certa ração varia $266,67kg^2$. (!!??)

PROPRIEDADES DA VARIÂNCIA

(1) Se $X = Y + k$, então, $V(X) = V(Y)$.

Demonstração:

$$\begin{aligned}
 V(X) &= \frac{\sum_{i=1}^n e_i^2}{n-1} \\
 &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \\
 &= \frac{\sum_{i=1}^n (y_i + k - (\bar{y} + k))^2}{n-1} \\
 &= \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} = V(Y)
 \end{aligned}$$

(2) Se $X = Y \times k$, então, $V(X) = V(Y) \times k^2$.

Demonstração:

$$\begin{aligned}
 V(X) &= \frac{\sum_{i=1}^n e_i^2}{n-1} \\
 &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \\
 &= \frac{\sum_{i=1}^n (y_i \times k - (\bar{y} \times k))^2}{n-1} \\
 &= \frac{\sum_{i=1}^n ((y_i - \bar{y}) \times k)^2}{n-1} \\
 &= \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} \times k^2 \\
 &= V(Y) \times k^2
 \end{aligned}$$

DESVIO PADRÃO

Notação: O desvio padrão recebe a mesma notação que a variância, porém sem o quadrado, ou seja, o desvio padrão da população é designado por σ ; e o desvio padrão da amostra, por S (variável aleatória) ou s (estimativa de S).

Se X é uma variável aleatória $DP(X)$ também denota o Desvio Padrão de X .

$$\begin{array}{cc} \text{Censos} & \text{Amostras} \\ \sigma = \sqrt{\sigma^2} & S = \sqrt{S^2} \end{array}$$

No exemplo,

	I	II	III
Desvio Padrão	0	16,33	73,48

Além de utilizar todos os dados para computar sua medida de variabilidade, o desvio padrão ainda retorna um valor expresso na unidade dos dados, o que o torna mais facilmente interpretável. Por exemplo, quando se diz que o peso de bovinos alimentados com certa ração costuma variar 16,33kg ao redor da sua média de peso, o leitor consegue ter uma idéia prática da variação.

PROPRIEDADES DO DESVIO PADRÃO

(1) Se $X = Y + k$, então, $DP(X) = DP(Y)$.

Demonstração:

$$\begin{aligned} DP(X) &= \sqrt{V(X)} \\ &= \sqrt{V(Y)} \\ &= DP(Y) \end{aligned}$$

(2) Se $X = Y \times k$, então, $DP(X) = DP(Y) \times k$.

Demonstração:

$$\begin{aligned} DP(X) &= \sqrt{V(X)} \\ &= \sqrt{V(Y) \times k^2} \\ &= DP(Y) \times k \end{aligned}$$

3.6.3 Coeficiente de Variação (CV)

O Coeficiente de variação é uma medida de variabilidade padronizada, ou seja, expressa percentualmente a variação dos dados em relação à média.

$$\begin{array}{cc} \text{Censos} & \text{Amostras} \\ CV(\%) = \frac{\sigma}{\mu} & CV(\%) = \frac{S}{\bar{X}} \end{array}$$

Sua grande vantagem é permitir a comparação de grandezas diferentes, que estão em unidades diferentes (por exemplo: o que é mais variável, o ganho de peso de suínos ou a altura de plantas de milho?).

Por outro lado, ele possui sérias restrições de uso e inspira cuidados. Primeiro, quando a média da variável aleatória em questão tende a zero, o CV tende ao infinito (o que não faz sentido prático). Segundo, de acordo com as propriedades da média e do desvio padrão, a adição de uma constante às observações altera a média da nova variável aleatória, mas não altera seu desvio padrão, ou seja, por meio de algumas transformações de variáveis o CV pode ser criminosamente manipulado.

No exemplo,

	I	II	III
CV(%)	0	16,33	73,48

Nesse caso a interpretação se torna ainda mais imediata, porém não podemos nos esquecer das ressalvas feitas anteriormente.

PROPRIEDADES DO COEFICIENTE DE VARIAÇÃO

$$(1) \text{ Se } X = Y + k, \text{ então, } \begin{cases} CV(X) < CV(Y), & \text{se } k > 0 \\ CV(X) > CV(Y), & \text{se } k < 0 \end{cases}$$

Demonstração:

$$\begin{aligned} CV(X) &= \frac{DP(X)}{\bar{X}} \\ &= \frac{DP(Y + k)}{\bar{Y} + k} \\ &= \frac{DP(Y)}{\bar{Y} + k} \end{aligned}$$

$$\text{Se } k > 0, \text{ então } \frac{DP(Y)}{\bar{Y} + k} < \frac{DP(Y)}{\bar{Y}}$$

$$\text{Se } k < 0, \text{ então } \frac{DP(Y)}{\bar{Y} + k} > \frac{DP(Y)}{\bar{Y}}$$

(2) Se $X = Y \times k$, então, $CV(X) = CV(Y)$.

Demonstração:

$$\begin{aligned} CV(X) &= \frac{DP(X)}{\bar{X}} \\ &= \frac{DP(Y \times k)}{\bar{Y} \times k} \\ &= \frac{DP(Y) \times k}{\bar{Y} \times k} \\ &= \frac{DP(Y)}{\bar{Y}} = CV(Y) \end{aligned}$$



As medidas de dispersão encontradas de forma direta no R são a variância e o desvio padrão, porém a amplitude e o CV podem ser implementados facilmente de acordo com o seguinte código:

```
c1<-c(100,100,100,100)
c2<-c(80,100,100,120)
c3<-c(10,100,100,190)
Ac1<-range(c1)[2]-range(c1)[1]  #Amplitude
var(c2)                         #Variância
sd(c2)                          #Desvio padrão
CVc3<-sd(c3)/mean(c3)*100      #Coeficiente de variação
```


Para todos nós, noções de probabilidade ou respostas intuitivas a questões de probabilidade são comuns desde a mais tenra idade. Qualquer pessoa, por menos conhecimento estatístico que tenha é capaz de responder à pergunta: Qual a probabilidade de se retirar uma carta de ouros de um baralho honesto?

Fazemos intuitivamente: “Se um baralho honesto tem 13 cartas do naipe ouros e o total de cartas do baralho é 52, então a chance de uma carta de ouros ser tirada ao acaso é $\frac{13}{52} = \frac{1}{4} = 1 : 4$ ou 25%”.

Formalizando o que foi feito intuitivamente, temos

$$P(A) = \frac{|B|}{|C|} \quad (4.1)$$

em que $|\cdot|$ é a cardinalidade de um conjunto;¹ A é um evento e; B e C são conjuntos. Mais especificamente, B é o conjunto que contém todos os eventos de interesse e C é o espaço amostral (Ω). No exemplo, $B = \{x|x \text{ é uma carta do naipe de ouros}\}$ e $C = \{x|x \text{ é uma carta do baralho}\}$ ².

LEI DOS GRANDES NÚMEROS

A lei dos grandes números é um conceito fundamental em Estatística e Probabilidade que descreve como a média de uma amostra, suficientemente grande e selecionada aleatoriamente, se torna provável de estar perto da média da população. Uma definição é trazida pelo site Wikipédia (2006): "Se um evento de probabilidade p for observado repetidamente ao longo de realizações independentes, a relação da frequência observada desse evento ao número total das repetições converge para p enquanto o número das repetições se torna arbitrariamente grande." Dizendo com outras palavras e colocando no contexto da construção de histogramas, pode-se entender que quando $n \rightarrow \infty$, as frequências das classes tendem a se estabilizar.

Considere o seguinte exemplo: no lançamento de uma moeda honesta qual a probabilidade de sair cara? Resp.: 50%. Isto é intuitivo devido à lei dos grandes números.

Figura 4.1 traz a simulação de 500 lançamentos de uma moeda honesta, a contagem do número de caras obtidas em cada lançamento e a plotagem da frequência relativa de caras (número de caras/números de lançamentos). Note que, quanto mais n (número de lançamentos) aumenta,

¹Cardinalidade: número de elementos de um conjunto.

²Para uma perfeita compreensão deste capítulo é necessária uma revisão de Teoria dos Conjuntos.

mais a frequência relativa tende a se estabilizar em 50%, o que corrobora a afirmação intuitiva anterior.

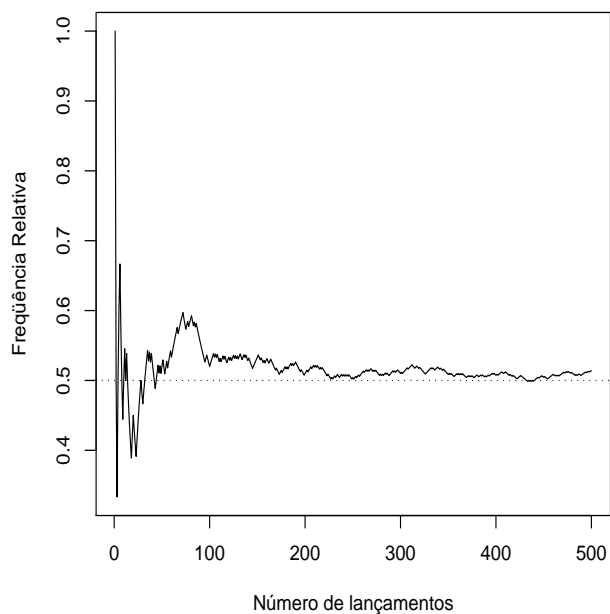


Figura 4.1: Simulação do lançamento de uma moeda honesta 500 vezes, comportamento de sua frequência relativa.



Rotina R da simulação do lançamento de uma moeda honesta 500 vezes.

```
cara<-0
fr<-vector("numeric",500)
for (i in 1:500) {
  moeda<-runif(1,0,1)
  if (moeda>0.5) {cara<-cara+1}
  fr[i]<-cara/i
}
x<-seq(1:500)
plot(x,fr,"l",xlab="Número de lançamentos",ylab="Frequência Relativa")
abline(h=.5,lty=3)
```

REGRA DO “E” E REGRA DO “OU”

Aqui serão apresentadas as regras do “e” e do “ou” apenas para eventos independentes. Quando os eventos são dependentes, são necessários alguns cuidados que não são objetos desse curso introdutório.

- Regra do “e”: A probabilidade de ocorrerem dois eventos A e B simultaneamente é

$$P(A \text{ e } B) = P(A) \times P(B) = P(A \cap B). \quad (4.2)$$

Exemplo: no lançamento de 2 dados honestos, qual a probabilidade de se tirar 3 e 5?

$$P(3) \times P(5) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} = 0,0278 = 2,78\%$$

- Regra do “ou”: A probabilidade de ocorrer o evento A ou o evento B em um dado experimento é

$$P(A \text{ ou } B) = P(A) + P(B) = P(A \cup B). \quad (4.3)$$

Exemplo: no lançamento de um dado honesto qual a probabilidade de se tirar 3 ou 5?

$$P(3) + P(5) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = 0,3333 = 33,33\%$$

ALGUMAS DEFINIÇÕES ÚTEIS

Probabilidade é a frequência relativa associada a uma variável descritora de uma população infinita.

Distribuição de probabilidade é a distribuição de frequência relativa em uma população infinita.

Parâmetro de uma distribuição é a constante que determina (estabelece) a forma³ da distribuição.

AXIOMAS DA PROBABILIDADE

- (i) Se A é um evento pertencente a Ω , então $P(A) > 0$.
- (ii) $P(\Omega) = 1$.
- (iii) Sejam A_1, A_2, A_3, \dots eventos disjuntos pertencentes a Ω (intersecção nula), então $P(A_1 \cup A_2 \cup A_3 \cup \dots) = P(A_1) + P(A_2) + P(A_3) + \dots$

³No sentido mais amplo da palavra

PROPRIEDADES DERIVADAS DOS AXIOMAS

- (i) $P(\bar{A}) = 1 - P(A)$.
- (ii) $P(\emptyset) = 0$.
- (iii) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- (iv) $0 \leq P(A) \leq 1$.

FUNÇÃO DENSIDADE DE PROBABILIDADE

Função densidade de probabilidade (fdp) é a função que descreve a probabilidade de uma variável aleatória associada a uma população infinita, onde a área representa a probabilidade e a altura, a densidade de probabilidade.

PROPRIEDADES DA FDP

- (i) A área total abaixo da curva é igual a 1,

$$\int_{-\infty}^{+\infty} f(x)dx = 1.$$

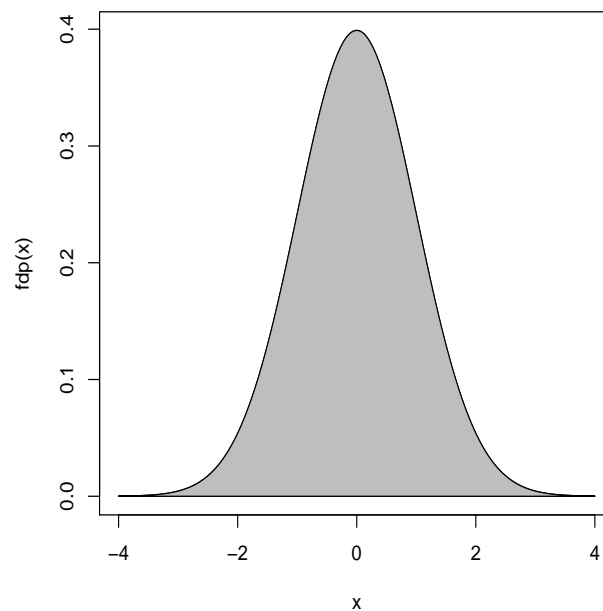


Figura 4.2: Esquema da integral de uma função densidade de probabilidade.



A rotina a seguir mostra como foi feita a Figura 4.2.

```
x <- seq(-4,4,by=.01)
y<-dnorm(x, mean=0,sd=1, log = FALSE)
rx<-rev(x)
ry<-vector("numeric", length(rx))
x<-c(x,rx)
y<-c(y,ry)
plot(x,y,"l",xlab="x",ylab="fdp(x)")
polygon(x, y, col = "gray")
```

(ii) Não existe probabilidade negativa.

$$f(x_0) \geq 0 \quad \forall x_0 \in D(X)$$

4.1 Distribuições de probabilidade discretas

Se caracteriza pela função $f(X)$, em que X é uma variável aleatória discreta. Análogo à variável aleatória contínua, se X assume k valores, então:

$$\sum_{i=1}^k P[X = x_i] = 1.$$

DEFINIÇÕES ÚTEIS

Esperança matemática de X é o valor médio esperado para infinitas realizações da variável aleatória discreta X .

$$E[X] = Me(X) = \sum_{i=1}^k x_i P[X = x_i]$$

Variância de X é uma medida da variabilidade das infinitas realizações da variável aleatória discreta X .

$$\sigma_X^2 = \sum_{i=1}^k [(X - E[X])^2 \times P[X = x_i]]$$

4.1.1 Distribuição Binomial

A distribuição Binomial de probabilidades é uma distribuição discreta que se caracteriza pela seguinte função densidade:

$$P[X = x] = C_{n,x} p^x q^{n-x}, \quad (4.4)$$

em que,

$$C_{n,x} = \frac{n!}{x!(n-x)!}. \quad (4.5)$$

Os parâmetros da Binomial são: n (número de eventos) e p (probabilidade de sucesso). q não é considerado um parâmetro porque ele é função de p , $q = 1 - p$.

A Binomial possui quatro características marcantes:

- (a) Ser uma soma de ensaios de Bernoulli, ou seja, de ensaios que possuem apenas dois resultados possíveis (sucesso ou fracasso).
- (b) As realizações desses ensaios são eventos independentes.
- (c) A probabilidade de sucesso (p) é constante ao longo dos ensaios.
- (d) O número de ensaios é finito.

A esperança matemática (média) e a variância da distribuição Binomial são funções de seus dois parâmetros, n e p , a saber:

$$E[X] = np; \quad (4.6)$$

$$\sigma_X^2 = V(X) = npq. \quad (4.7)$$

Exemplo: Em uma ninhada de aves, nove ovos foram chocados. Qual a probabilidade de nascerem sete machos?

Sabe-se que a probabilidade de um filhote ser macho é 50% ($p = 0,5$); o número de ovos nessa ninhada é 9 ($n = 9$); e a variável aleatória (X), desse experimento, representa o número de filhotes machos. Além disso, queremos descobrir a probabilidade dessa variável assumir o valor sete, ou seja, nascerem sete machos nessa ninhada ($x = 7$); então

$$P[X = 7] = C_{9,7} p^7 q^{9-7} = 36 \times 0,5^7 \times 0,5^2 \cong 7\%.$$

4.1.2 Distribuição Poisson

Distribuição discreta de probabilidades caracterizada pela seguinte função densidade de probabilidade:

$$P[X = x] = \frac{e^{-\lambda} \lambda^x}{x!}. \quad (4.8)$$

Neste curso, a distribuição de Poisson será usada apenas como uma aproximação a distribuição Binomial, quando n é muito grande e p é muito pequeno. Portanto, vale ressaltar que, além das características da distribuição Binomial, a Poisson apresenta as seguintes particularidades:

- (a) Pode descrever eventos raros.
- (b) A variável discreta assume apenas valores inteiros positivos ($X = 0, 1, \dots$).
- (c) $n > 50$ e $p < 0,1$.

O parâmetro da distribuição Poisson é o λ . Note, pelas fórmulas a seguir que λ também é a média dessa distribuição.

$$E[X] = V[X] = \lambda; \quad (4.9)$$

$$\lambda = np. \quad (4.10)$$

Exemplo: Supondo que a ocorrência média de chuvas acima de $50mm/h$, em uma região, seja de 1,5 por ano, qual a probabilidade de, em um dado ano, não chover mais de $50mm/h$? E ter apenas uma chuva intensa? E duas? Faça a distribuição de probabilidade dessa variável até 6 chuvas intensas.

$$\text{fdp: } P[X = 0] = \frac{e^{-1,5} 1,5^0}{0!} = 0,2231.$$

$$\text{fdp: } P[X = 1] = \frac{e^{-1,5} 1,5^1}{1!} = 0,3347.$$

$$\vdots$$

$$\text{fdp: } P[X = 5] = \frac{e^{-1,5} 1,5^5}{5!} = 0,0141.$$

$$\text{fdp: } P[X = 6] = \frac{e^{-1,5} 1,5^6}{6!} = 0,0035.$$

Portanto, temos a seguinte distribuição de frequências:

X	0	1	2	3	4	5	6
$P[X = x]$	0,2231	0,3347	0,2510	0,1255	0,0421	0,0142	0,0035

4.2 Distribuições de probabilidades contínuas

Pode-se entender distribuições contínuas de probabilidade como generalizações de histogramas construídos para grandes tamanhos amostrais ($n \rightarrow \infty$).

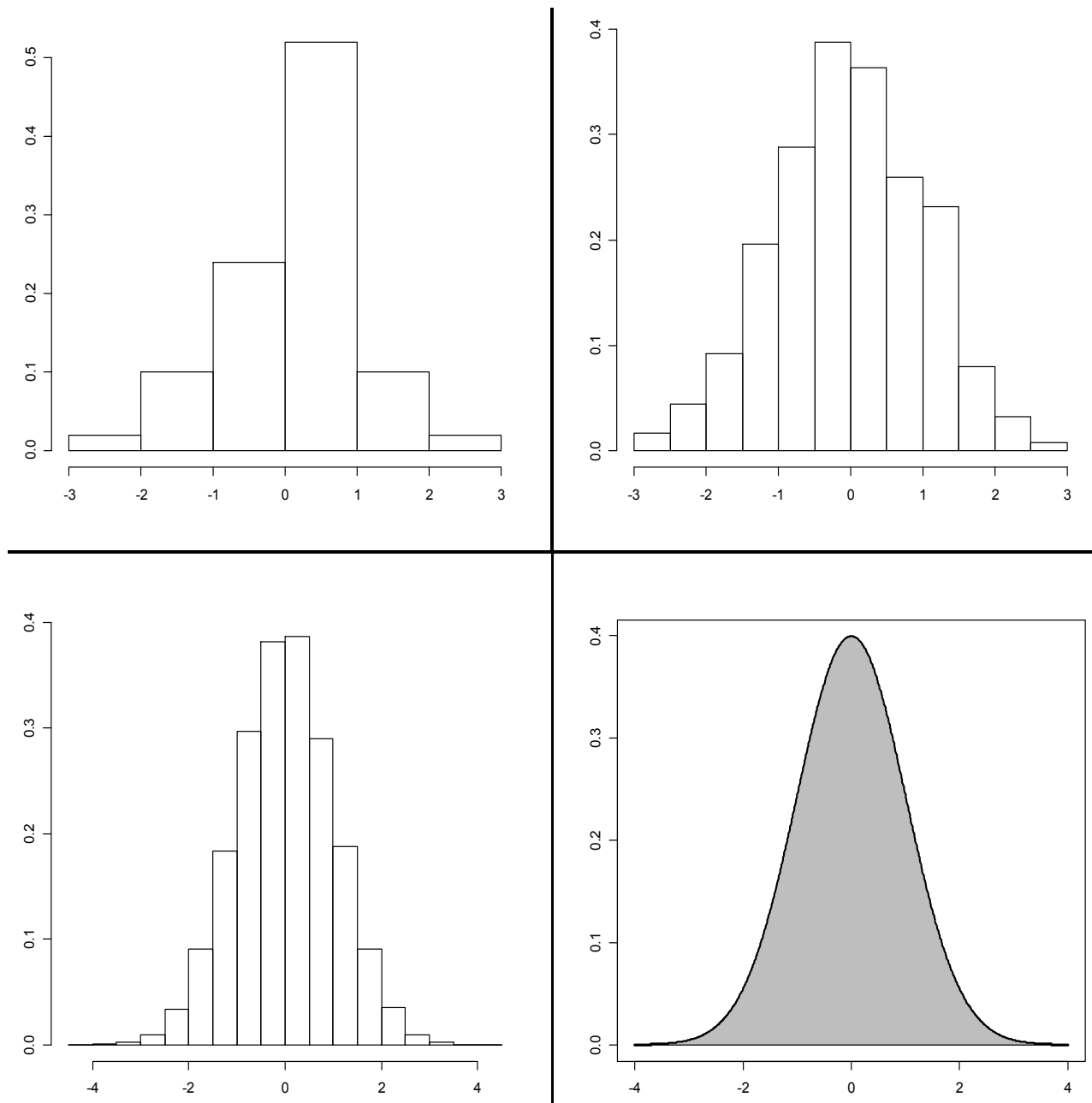


Figura 4.3: Esquema da generalização teórica de histogramas para funções densidade de probabilidade, quando $n \rightarrow \infty$.



Veja a rotina utilizada para a construção dos gráficos da Figura 4.3.

```
x1 <- rnorm(50,0,1)
hist(x1,freq=FALSE,xlab="",ylab="",main="")
x2<- rnorm(500,0,1)
hist(x2,freq=FALSE,xlab="",ylab="",main="")
x3 <- rnorm(50000,0,1)
hist(x3,freq=FALSE,xlab="",ylab="",main="")
x4<-seq(-4,4,by=.01)
y<-dnorm(x4, mean=0, sd=1, log = FALSE)
plot(x4,y,"l",lwd=3, xlab="",ylab="",main="")
polygon(x4, y, col = "gray")
```

Vamos entender o que acontece quando o tamanho da amostra cresce. Sejam as seguintes regras de determinação da amplitude e número de classes de um histograma convencional:

$$k = 5 \log n \quad c = \frac{A}{k-1}.$$

4.2.1 Distribuição Normal de probabilidades

Definição: uma variável aleatória contínua tem distribuição Normal se ela segue a seguinte função densidade de probabilidade:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2}. \quad (4.11)$$

Notação:⁴

$$N(\mu, \sigma^2).$$

Portanto, a média e a variância são os dois parâmetros da distribuição Normal, sendo μ o parâmetro de locação e σ^2 o parâmetro de forma.

$$Me(X) = \mu; \quad (4.12)$$

$$V(X) = \sigma^2. \quad (4.13)$$

⁴Lê-se: distribuição Normal com média μ e variância σ^2 .

PROPRIEDADES DA DISTRIBUIÇÃO NORMAL

- (1) É simétrica.
- (2) Tem forma de sino.
- (3) Está definida de $-\infty$ a $+\infty$, com caudas assintóticas ao eixo X.
- (4) Apresenta a particularidade: $Me = Mo = Md$.
- (5) Dois parâmetros: De forma: σ^2 e de locação μ .

Exemplo: A velocidade de veículos em uma rodovia segue uma distribuição Normal com média 60km/h e variância $400(\text{km/h})^2$.

(a) Qual a probabilidade de um veículo ser flagrado a mais de 100km/h ?

Solução:

$$\int_{100}^{\infty} f(x)dx = \int_{100}^{\infty} \frac{1}{20\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-60}{20} \right)^2} dx.$$

Essa integral não é trivial!

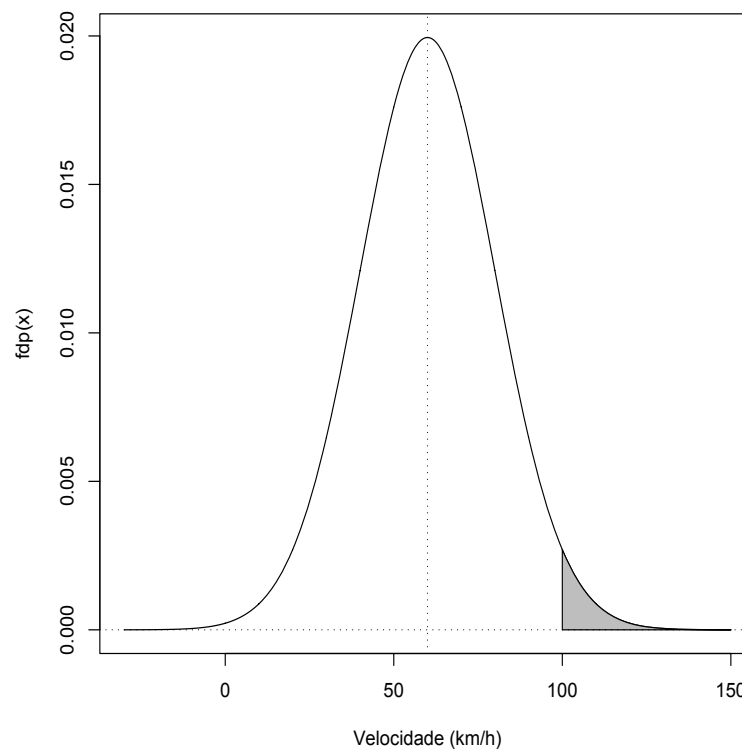


Figura 4.4: Esquema destacando a area acima de 100km/h em uma distribuição de media 60km/h e variância $400(\text{km/h})^2$.



Veja a rotina utilizada para fazer a Figura 4.4.

```
x<-seq(-30,150,by=.01)
y<-dnorm(x, mean=60,sd=20, log = FALSE)
rx<-seq(100,150,by=.1)
ry<-numeric(2*length(rx))
ry[1:length(rx)]<-dnorm(rx,mean=60,sd=20,log=FALSE)
rx<-c(rx,rev(rx))
plot(x,y,'l',xlab='Velocidade (km/h)',ylab='fdp(x)')
polygon(rx, ry, col = "gray")
abline(v=60,h=0,lty=3)
```

DISTRIBUIÇÃO NORMAL PADRONIZADA (REDUZIDA OU PADRÃO)

Uma solução simples e eficiente para o problema apresentado acima é definição da distribuição Normal padronizada. Essa distribuição apresenta, como principais características, a média igual a zero e a variância (e o desvio padrão) igual a 1.

Mediante as seguintes propriedades, uma distribuição Normal qualquer (média e variância quaisquer) pode ser temporariamente transformada em uma Normal padrão, por conveniência. Devido a essa correspondência, resultados de integrais calculadas (por processos numéricos) para atender à Normal padrão, servem para gerar resultados de quaisquer integrais que se deseje em outras Normais.

Seja X uma variável aleatória que segue uma distribuição $N(\mu, \sigma^2)$. As seguintes afirmações se verificam:

- (I) $(X - \mu) \sim N(0, \sigma^2)$;
- (II) $\frac{X}{\sigma} \sim N\left(\frac{\mu}{\sigma}, 1\right)$; e, portanto,
- (III) $\left(\frac{X - \mu}{\sigma}\right) \sim N(0, 1)$

Devido a definição de seus parâmetros, ou seja, sabendo que $\mu = 0$ e $\sigma^2 = 1$, a função densidade de probabilidade da Normal Padronizada se reduz a

$$f(X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{X^2}{2}}. \quad (4.14)$$

Então, retomando o problema, podemos definir uma variável aleatória Z , tal que

$$Z = \frac{X - \mu}{\sigma}.$$

Daí, $z = \frac{100 - 60}{20} = 2$ e utilizando a Tabela 9.2 do Apêndice ??, temos que $P(X > 100) = P(Z > 2) = 0,5 - 0,4772 = 0,0228 = 2,28\%$.

(b) E qual a chance de um automóvel estar trafegando entre 40 e 70km/h?

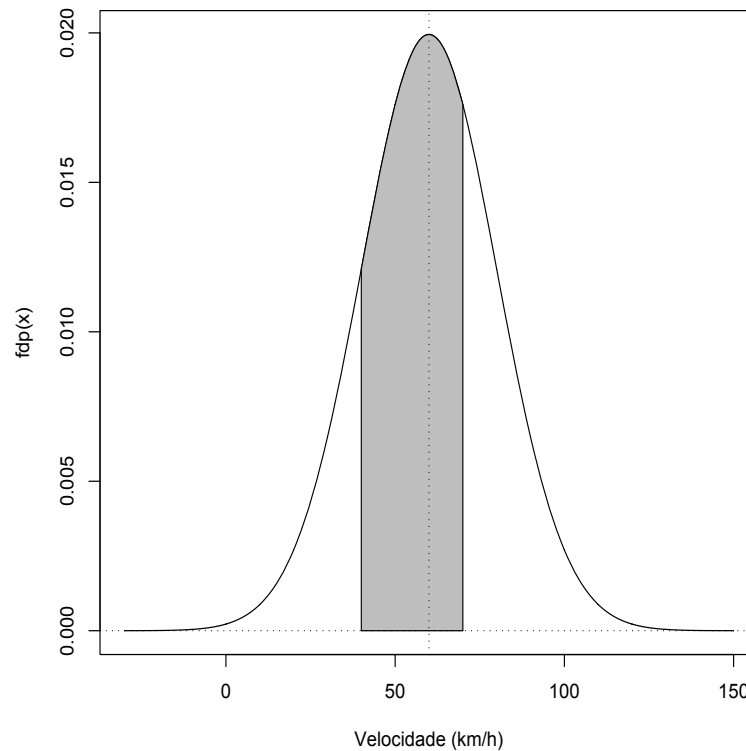


Figura 4.5: Esquema destacando a area entre 40 e 100km/h em uma distribuição de media 60km/h e variância $400(km/h)^2$.



Veja a rotina utilizada para fazer a Figura 4.5.

```
x<-seq(-30,150,by = .01)
y<-dnorm(x, mean=60,sd=20, log = FALSE)
rx<-seq(40,70,by =.1)
ry<-numeric(2*length(rx))
ry[1:length(rx)]<-dnorm(rx,mean=60,sd=20, log = FALSE)
rx<-c(rx,rev(rx))
plot(x,y,'l',xlab='Velocidade (km/h)',ylab='fdp(x)')
polygon(rx, ry, col = 'gray')
abline(v=60,h=0,lty=3)
```

Aqui, devemos transformar dois pontos: $z_1 = \frac{40 - 60}{20} = -1$ e $z_2 = \frac{70 - 60}{20} = 0,5$. E olhando na tabela: $P(40 < X < 70) = P(-1 < z < 0,5) = 0,3413 + 0,1915 = 53,28\%$.

(c) Qual intervalo contém 90% dos veículos?

O processo aqui é o inverso. Eu tenho a probabilidade (ou porcentagem de veículos) e desejo saber os pontos, ou seja, as velocidades que delimitam essa área.

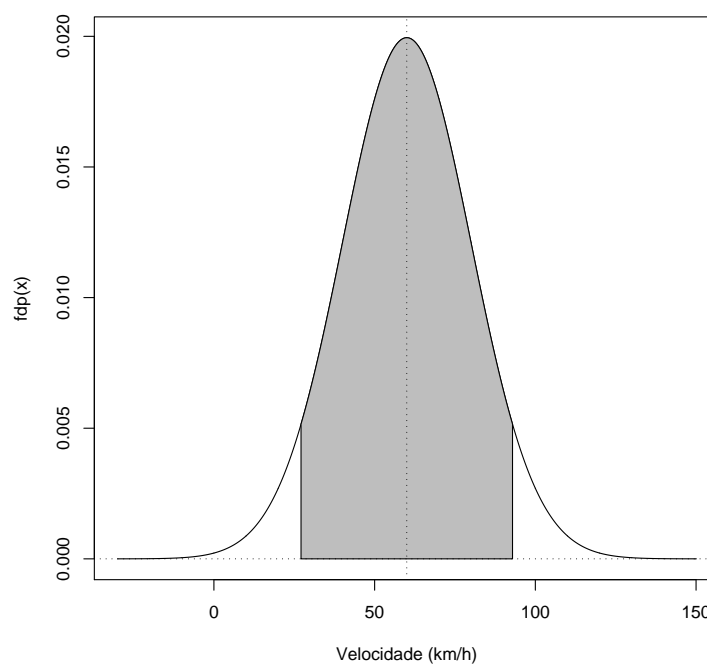


Figura 4.6: Esquema destacando o intervalo que contém 90% dos veículos em uma distribuição de média 60km/h e variância $400(\text{km/h})^2$.



Veja a rotina utilizada para fazer a Figura 4.6.

```
x<-seq(-30,150,by = .01)
y<-dnorm(x, mean=60,sd=20, log = FALSE)
rx<-seq(27.1,92.9,by=.1)
ry<-numeric(2*length(rx))
ry[1:length(rx)]<-dnorm(rx,mean=60,sd=20, log = FALSE)
rx<-c(rx,rev(rx))
plot(x,y,'l',xlab='Velocidade (km/h)',ylab='fdp(x)')
polygon(rx, ry, col = 'gray')
abline(v=60,h=0,lty=3)
```

$$1,645 = \frac{X_2 - 60}{20} \implies X_2 = 92,9 \text{ km/h}$$

$$-1,645 = \frac{X_1 - 60}{20} \implies X_1 = 27,1 \text{ km/h}$$

$$P(27,1 < X < 92,9) = 90\%$$

APROXIMAÇÃO DA BINOMIAL À NORMAL

Quando nos deparamos com uma situação em que uma variável aleatória é Binomial, mas n e p são muito grande grandes,

$$np > 5 \text{ e } npq > 5$$

as contas usuais da distribuição Binomial se tornam mais difíceis de executar. Em tais situações, podemos utilizar uma aproximação pela distribuição Normal para fazer os cálculos de interesse. Devemos seguir as seguintes relações de transformação:

$$\mu = np \text{ e } \sigma^2 = npq.$$

Exemplo: Um Eng. Agrônomo faz um teste de germinação com $n = 500$ sementes, sabendo que seu poder nominal de germinação é de $p = 83\%$. Se seu poder de germinação estiver correto, qual é a probabilidade de que, pelo menos, 430 sementes germinem?

$$\mu = 500 \times 0,83 = 415 \text{ sementes.}$$

$$\sigma^2 = 500 \times 0,83 \times 0,17 = 70,55 \implies \sigma = 8,4 \text{ sementes.}$$

$$z = \frac{430 - 415}{8,4} \approx 1,79.$$

$$P(X > 430) = P(Z > 1,79) = 0,0367 = 3,67\%.$$

APROXIMAÇÃO DA POISSON À NORMAL

Em situação semelhante, porém, dessa vez se tratando de uma variável Poisson, uma média muito grande,

$$\lambda > 15,$$

também podemos aproximar à uma distribuição Normal seguindo as transformações:

$$\mu = \lambda \text{ e } \sigma^2 = \lambda.$$

Exemplo: Suponha que a média de chuvas fracas por ano em certa região seja de $\lambda = 30$. Qual a probabilidade de ocorrerem mais de 45 chuvas desse tipo no próximo ano?

$$\lambda = \mu = \sigma^2 = 30$$

$$\sigma = 5,48 \text{ chuvas fracas}$$

$$z = \frac{45 - 30}{5,8} = 2,74$$

$$P(X < 45) = P(Z < 2,74) = 0,9969 = 99,69\%$$

4.2.2 Outras distribuições contínuas

Existe uma infinidade de outras distribuições contínuas muito usadas em Estatística Aplicada que, infelizmente, não serão tratadas aqui, a saber: distribuição Uniforme, Exponencial, Weibull, Gama, etc.

Na grande maioria das situações práticas, o exame exaustivo de todos os elementos de uma população (censo) não é possível por ser cara, demorada ou mesmo impossível (populações infinitas).

Nestes casos é imperativo o uso de técnicas de amostragem. O uso de amostras traz diversas vantagens como:

- menor custo;
- maior rapidez;
- boa acurácia (se coletada corretamente);
- torna viável o exame em análises destrutivas, etc.

As amostras devem ser representativas, ou seja, guardar semelhanças com a população. Para isso é necessário, sempre que possível, a aleatorização, casualização ou sorteio.

5.1 Amostragens não-aleatórias

São utilizadas quando o sorteio não é possível, mas se faz um esforço para garantir a representatividade. Algumas delas:

- (a) Dificil acesso: Exemplo: Amostragem de minério em vagões. Embora seja praticamente impossível tomarmos minérios em todas as partes do vagão, essa tende a ser uma situação homogênea e a coleta em pontos da superfície pode ser suficiente.
 - (b) Coleta a esmo: Exemplo: Amostragem de solo. Nesse caso costuma-se fazer um caminharmento em zigue-zague, que é um esforço de aleatorização.
 - (c) Por conveniência: Exemplo: Pesquisa onde se fazem perguntas a pessoas próximas ao pesquisador. Em casos em que se pode considerar que as pessoas que estão passando, naquele momento, ao lado do pesquisador estão ali por puro acaso.
 - (d) Intencional: Exemplo: Escolha de cada indivíduo. O pesquisador determina exatamente que será observado, mas se esforça garantir a representatividade de sua escolha.
 - (e) Auto-escolha: Exemplo: Voluntários para uma pesquisa médica. A área médica muitas frequentemente se enquadra nesta amostragem, onde não se pode escolher aquelas pessoas que ficam doentes, por exemplo. Neste caso, pode-se considerar que as pessoas adoecem aleatoriamente.
-

5.2 Amostragens aleatórias

5.2.1 Amostragem aleatória simples (AAS)

- Deve ser realizada em populações estritamente homogêneas.
- Ela se caracteriza pelo sorteio de n elementos de uma população.
- Pode ser feita com ou sem reposição dos elementos amostrados à população.

SORTEIO

- (a) Tabelas de números aleatórios: 0 a 99.999;
- (b) Calculadoras: Tecla RAN \times N (tamanho da população);
- (c) Softwares estatísticos.

INCONVENIENTES

- Populações estritamente homogêneas são pouco comuns.
- É muito trabalhosa em populações grandes, porque todos os elementos devem ser numerados (ou, pelo menos, identificados), e impossível em populações infinitas.

MODELO ESTATÍSTICO

$$Y_i = \mu + e_i, \quad (5.1)$$

em que Y_i é a observação do indivíduo i da amostra; μ é a média da população; e e_i é o desvio aleatório referente ao indivíduo i .

5.2.2 Amostragem aleatória estratificada (AAE)

- É utilizada em populações heterogêneas.
- Deve-se dividir a população em estratos homogêneos (dentro), mas diferentes entre si. Daí sorteiam-se elementos de cada estrato proporcionalmente a seu tamanho.

MODELO ESTATÍSTICO

$$Y_{ij} = \mu + t_i + e_{ij} \quad (5.2)$$

em que Y_{ij} é o valor do indivíduo j do estrato i ; μ é a média populacional; t_i é o efeito do estrato i ; e_{ij} é o efeito aleatório (desvio) do indivíduo j do estrato i .

Obs.: Média do estrato i : $\mu_i = \mu + t_i$.

5.2.3 Amostragem aleatória por conglomerado (AAC)

Conglomerado subdivisão da população objetivando economia de recursos, pois somente alguns serão sorteados.

- Principal objetivo: economia de tempo e recursos.
- Há homogeneidade *entre* conglomerados e espera-se que a variabilidade de população esteja representada *dentro* de cada um deles.

Exemplo: Em uma pesquisa dentre os domicílios de Lavras, se uma AAS fosse feita, provavelmente os sorteados ficariam muito espalhados, sendo difícil de serem observados. Daí sorteia-se 7 bairros e 10 domicílios por bairro para facilitar o processo.

MODELO ESTATÍSTICO

$$Y_{ij} = \mu + c_i + e_{ij}, \quad (5.3)$$

em que Y_{ij} é o valor do indivíduo j do conglomerado i ; μ é a média populacional; t_i é o efeito (aleatório) do conglomerado i ; e_i é o efeito aleatório (desvio) do indivíduo j do conglomerado i .

Tabela 5.1: Diferenças básicas entre a amostragem aleatória estratificada (AAE) e a amostragem aleatória por conglomerado (AAC).

AAE	AAC
Todo estado é observado	Alguns conglomerados são sorteados
Efeito fixo	Efeito aleatório
Estratos são diferentes entre si	Conglomerados são semelhantes entre si
Objetivo: > representatividade	Objetivo: < custo

5.2.4 Amostragem aleatória sistemática (AS)

- É utilizada em situações em que elementos da população estão dispostos em série.
- Apenas o 1º elemento é sorteado, os demais são tomados sistematicamente, e se distanciam distantes um *passo de amostragem* (k).
- Objetivo: facilitar o processo.

POPULAÇÕES FINITAS

- (a) Defini-se o passo de amostragem: $k = \frac{N}{n}$.
- (b) Sorteia-se o primeiro elemento dentre os k primeiros.
- (c) Tornam-se os demais de k em k .

POPULAÇÕES MUITO GRANDES OU INFINITAS

- (a) Toma-se o 1 elemento a esmo.
- (b) Tomam-se os demais elementos de maneira igualmente espaçada.

MODELO ESTATÍSTICO¹

$$Y_i = \mu + u_i, \tag{5.4}$$

em que

$$u_i = \rho u_{ij} + e_i. \tag{5.5}$$

Geralmente admi-se ρ (parâmetro de auto correlação²) igual a zero, daí,

$$Y_i = \mu + e_i. \tag{5.6}$$

¹Um dos modelos possíveis.

²Note que, quando $\rho = 0$, o modelo estatístico é igual ao da amostragem aleatória simples.

Inferência é o conjunto de técnicas que generalizam informações amostrais para toda a população.

Grandes áreas Estimação (Intervalos de confiança) e Decisão (Testes de hipóteses).

TEOREMA CENTRAL DO LIMITE

Seja uma população qualquer com média μ e variância σ^2 . Se infinitas amostras de tamanho n são coletadas dessa população, então as médias (\bar{X}) das amostras terão distribuição aproximadamente Normal com média μ e variância $\frac{\sigma^2}{n}$, à medida que n tende ao infinito.

6.1 Distribuição de amostragem

Definição: É a distribuição de probabilidade de estimadores $\hat{\theta}$ ou $f(\hat{\theta})$ ao longo de infinitas amostras aleatórias.

6.1.1 Distribuição de funções da média amostral (populações normais)

Sabemos que se X tem distribuição Normal, $Z = \frac{X - \mu}{\sigma}$ tem distribuição Normal padronizada, $N(0, 1)$. Agora, como saber qual a distribuição da média \bar{X} ?

Se $X \sim N(\mu, \sigma^2)$, então as seguintes afirmações se verificam:

$$(I) \sum X \sim N(n\mu, n\sigma); \text{ e}$$

$$(II) \frac{\sum X}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right); \text{ portanto}$$

$$(III) \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Se padronizarmos a média, ou seja, subtraírmos da média populacional e dividirmos por seu desvio padrão, conseguimos encontrar a quantidade que segue uma Normal padrão:

$$f(\bar{X}) = Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1). \quad (6.1)$$

Mas como σ é geralmente desconhecido, tem-se que utilizar uma quantidade que mais se assemelhe. Essa quantidade é o seu estimador, S . Porém, quando substituímos σ por S , devemos pagar um preço. A nova quantidade não segue mais uma distribuição Normal padrão (Z), mas uma aproximação da Normal padrão, a distribuição *t de Student*, com $n - 1$ *graus de liberdade*, t_{n-1} .

$$f(\bar{X}) = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1} \quad (6.2)$$

Graus de liberdade: O número de graus de liberdade para um conjunto de dados corresponde ao número de valores que podem variar após terem sido impostas certas restrições a todos os valores.

A distribuição *t de Student* é uma distribuição de probabilidades muito parecida com a distribuição Normal padrão. Ela é simétrica, tem forma de sino, é centrada em zero, mas é mais “larga” e sua forma varia em função do tamanho da amostra (n). Quanto maior a amostra mais a distribuição *t* se assemelha a uma distribuição *z*, ou seja,

$$n \rightarrow \infty, t_{n-1} \rightarrow N(0, 1).$$

PROPRIEDADES IMPORTANTES DA DISTRIBUIÇÃO T DE STUDENT¹

- (1) A distribuição *t de Student* é diferente, conforme o tamanho da amostra.
- (2) Ela tem a mesma forma geral simétrica (forma de sino) que a distribuição Normal, mas reflete a maior variabilidade (com distribuições mais amplas) que é esperada em pequenas amostras.
- (3) Tem média $t = 0$.
- (4) O desvio padrão da distribuição *t de Student* varia com o tamanho da amostra, mas é superior a 1.
- (5) Na medida em que aumenta o tamanho n da amostra, a distribuição *t de Student* se aproxima mais e mais da distribuição Normal padronizada. Para valores $n > 30$, as diferenças são tão pequenas que podemos utilizar os valores críticos Z em lugar de valores críticos t .

¹Adaptado de Triola, 1999.

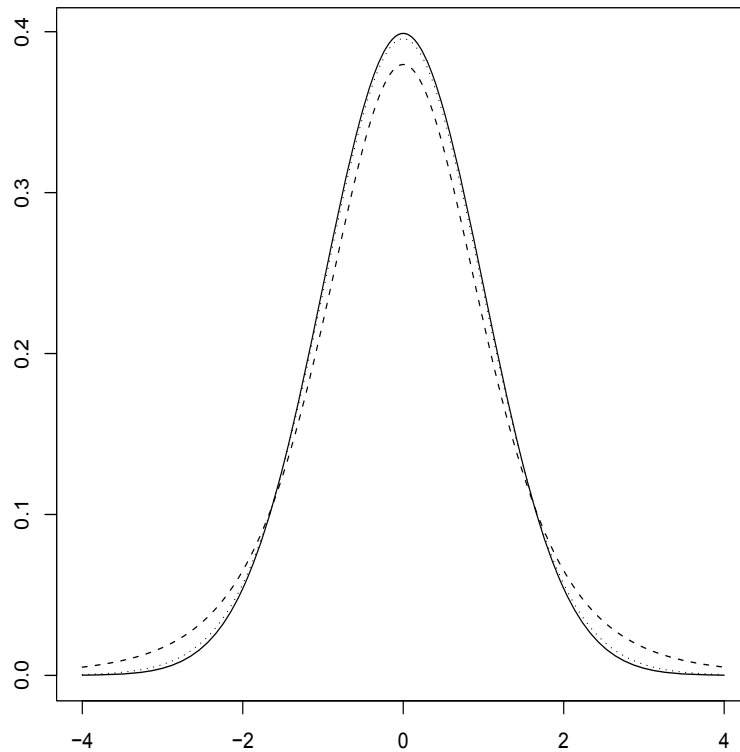


Figura 6.1: Demonstração da curva Normal (linha cheia) padrão e da curva t com 5 (linha tracejada) e 30 (linha pontilhada) graus de liberdade.



Veja a rotina utilizada para fazer a Figura 6.1.

```
x<-seq(-4,4,by=.01)
yn<-dnorm(x,mean=0,sd=1)
yt1<-dt(x,df=5)
yt2<-dt(x,df=30)
plot(x,yn,'l',lty=1,xlab='',ylab='')
points(x,yt1, 'l', lty=2)
points(x, yt2, 'l', lty=3)
```

6.2 Estimação

6.2.1 Estimação por ponto

É a obtenção de uma estimativa do valor paramétrico com base em informações vindas da amostra.

Estimadores mais comuns:

$$\begin{aligned}\hat{\mu} &= \bar{X} = \frac{\sum_i X_i}{n} \\ \hat{\sigma}^2 &= \frac{\sum_i (X_i - \bar{X})^2}{n - 1} \\ \hat{p} &= \frac{x}{n}, \quad x = \text{número de sucessos} \\ \widehat{\mu_1 - \mu_2} &= \hat{\mu}_1 - \hat{\mu}_2 = \bar{X}_1 - \bar{X}_2 \\ &\dots\end{aligned}$$

6.2.2 Estimação por intervalo

Intervalo de confiança (IC): é o intervalo que contém o parâmetro real com $1 - \alpha$ de confiança.

Confiança: $\gamma = 1 - \alpha$

Significância: α

A maioria dos intervalos de confiança segue a seguinte forma geral:

$$IC_\gamma(\theta) = \hat{\theta} \pm q_{\alpha/2} EP_\theta \quad (6.3)$$

em que γ é a confiança associada ao intervalo; θ é uma função qualquer do parâmetro de interesse; $\hat{\theta}$ é o estimador desse parâmetro; $q_{\alpha/2}$ é um quantil de uma distribuição associada ao estimador e; EP_θ é o erro padrão do estimador.

IC para a media (μ)

σ^2 CONHECIDO

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1) \quad (6.4)$$

Então, podemos deduzir o intervalo de confiança para a média:

$$\begin{aligned} P[-z < Z < z] &= \gamma \\ P\left[-z_{\alpha/2} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2}\right] &= \gamma \\ P\left[-z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right] &= \gamma \\ P\left[-\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right] &= \gamma \\ P\left[\bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} > -\mu > \bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right] &= \gamma \\ P\left[\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < -\mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right] &= \gamma \\ IC_{\gamma}(\mu) &= \left[\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right] \end{aligned}$$

σ^2 DESCONHECIDO

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1} \quad (6.5)$$

Analogamente, consegue-se o intervalo

$$IC_{\gamma}(\mu) = \left[\bar{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}\right]$$

FATOR DE CORREÇÃO PARA POPULAÇÕES FINITAS

Em populações finitas o erro (e) do IC deve ser corrigido pelo fator de correção:

$$\sqrt{\frac{N-n}{N-1}}. \quad (6.6)$$

Ou seja, o intervalo de confiança geral passa a ser

$$IC_{\gamma}(\theta) = \left[\hat{\theta} \pm e \sqrt{\frac{N-n}{N-1}}\right] \quad (6.7)$$

IC para proporção (p)

Existem diversas aproximações mais precisas, porém muito mais trabalhosas. Por isso podemos utilizar a aproximação binomial à normal.

Este procedimento **não** é recomendado quando

$$n\hat{p} \leq 5 \text{ ou } n(1 - \hat{p}) \leq 5.$$

Se X é binomial e n é grande

$$X \sim N(np, npq)$$

$$\frac{X}{n} \sim N\left(p, \frac{pq}{n}\right)$$

$$\hat{p} \sim N\left(p, \frac{pq}{n}\right)$$

A partir daí fica fácil enxergar que

$$IC_{\gamma}(p) = \left[\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} \right]$$

IC para a diferença entre duas medias ($\mu_1 - \mu_2$)**VARIÂNCIAS CONHECIDAS**

Se uma variável aleatória

$$X_1 \sim N(\mu_1, \sigma_1^2), \text{ então } \bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right).$$

E se outra variável

$$X_2 \sim N(\mu_2, \sigma_2^2), \text{ então } \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right).$$

Daí,

$$(\bar{X}_1 - \bar{X}_2) \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right).$$

Contudo, as variâncias consideradas conhecidas podem ser iguais ou diferentes.

$$\underline{\sigma_1^2 \neq \sigma_2^2}$$

$$IC_{\gamma}(\mu_1 - \mu_2) = \left[(\bar{X}_1 - \bar{X}_2) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right].$$

$$\underline{\sigma_1^2 = \sigma_2^2 = \sigma^2}$$

$$IC_\gamma(\mu_1 - \mu_2) = \left[(\bar{X}_1 - \bar{X}_2) \pm z_{\alpha/2} \sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \right].$$

VARIÂNCIAS DESCONHECIDAS

No caso em que não são conhecidas as variâncias populacionais, a tendência natural é utilizar seus estimadores. S^2 é o estimador de σ^2 , porém a distribuição de $\bar{X}_1 - \bar{X}_2$ não é mais normal, mas sim uma t .

E mesmo quando as variâncias populacionais são desconhecidas, deve-se decidir por considerá-las iguais ou diferentes.

$$\underline{\sigma_1^2 \neq \sigma_2^2}$$

$$IC_\gamma(\mu_1 - \mu_2) = \left[(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right].$$

Em que os graus de liberdade de t são dados por

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{\left(\frac{S_1^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2} \right)^2}{n_2 - 1}}.$$

$$\underline{\sigma_1^2 = \sigma_2^2 = \sigma^2}$$

$$IC_\gamma(\mu_1 - \mu_2) = \left[(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2} \sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \right].$$

Em que

$$S_p = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

e os graus de liberdade são dados por

$$\nu = n_1 + n_2 - 2.$$

6.3 Testes de hipóteses

Teste de hipótese: é uma ferramenta que permite testar se um valor é pertinente de ser o valor real de parâmetro em questão.

6.3.1 Teste de homogeneidade de Variâncias (teste F):

Estatística de teste: $F_c = \frac{S_{maior}^2}{S_{menor}^2} \geq 1$, em que F_c é uma variável aleatória que segue uma distribuição F de probabilidades com graus de liberdade $\nu_1 = n_1 - 1$ e $\nu_2 = n_2 - 1$, ou seja,

$$F_c \sim F(\nu_1, \nu_2) = F[(n_1 - 1), (n_2 - 2)].$$

DISTRIBUIÇÃO F DE PROBABILIDADES

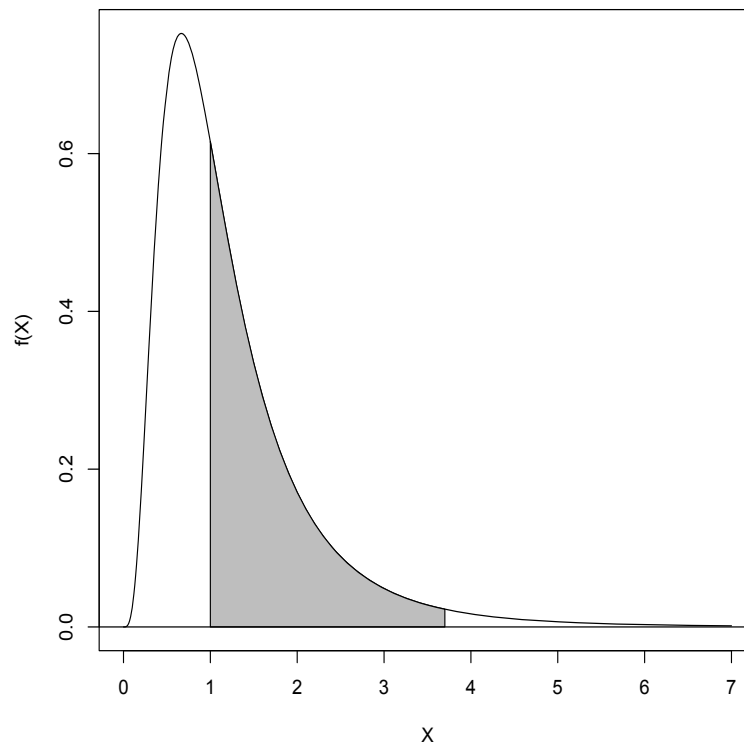


Figura 6.2: Distribuição F de probabilidade ressaltando a região de aceitação de um teste de homogeneidade de variâncias (de 1 até um F_c qualquer).



Veja a rotina utilizada para fazer a Figura 6.2.

```
x<-seq(0,7,by=.01)
y<-df(x,10,10)
plot(x, y, "l", xlab= 'X', ylab='f(X)')
rx<-seq(1,3.7,by =.1)
ry<-vector('numeric',2*length(rx))
ry[1:length(rx)]<-df(rx,10,10)
rx<-c(rx,rev(rx))
plot(x,y,'l',xlab='X',ylab='f(X)')
polygon(rx, ry, col = "gray")
abline(h=0)
```

HIPÓTESES

Todo teste de hipóteses considera duas hipóteses: a *Hipótese Nula* ou *Hipótese de nulidade* (H_0); e a *Hipótese Alternativa* (H_1).

A hipótese H_0 é aquela que descreve aquilo que suspeita, ou seja, aquela afirmação que se quer testar efetivamente. Já a hipótese H_1 traz aquela afirmação que sera considerada verdade se a hipótese H_0 for considerada não plausível (se H_0 for rejeitada).

Vamos voltar a considerar o caso particular do teste F. Considerando as definições das hipóteses H_0 e H_1 , o teste F deseja testar

$$\left\{ \begin{array}{l} H_0 : \frac{\sigma_{maior}^2}{\sigma_{menor}^2} = 1 \\ H_1 : \frac{\sigma_{maior}^2}{\sigma_{menor}^2} \geq 1 \end{array} \right.,$$

ou ainda,

$$\left\{ \begin{array}{l} H_0 : \sigma_{maior}^2 = \sigma_{menor}^2 \\ H_1 : \sigma_{maior}^2 \geq \sigma_{menor}^2 \end{array} \right.$$

RESULTADOS E TIPOS DE ERROS POSSÍVEIS

Erro tipo I: é o erro que se comete ao rejeitar H_0 , se ela é verdadeira.

Erro tipo II: é o erro que se comete ao aceitar H_0 , se ela é falsa.

Nível de significância do teste: é o valor da probabilidade de se cometer o erro tipo I.

Poder do teste: é a probabilidade de se rejeitar H_0 , quando ela é realmente falsa.

Tabela 6.1: Representação tabular dos resultados possíveis em um teste de hipóteses e os erros e acertos que eles acarretam.

		Verdade	
		H_0 é verdadeira	H_0 é falsa
Decisão	Aceita-se H_0	Ok!	Erro tipo II
	Rejeita-se H_0	Erro tipo I	Ok!

6.3.2 Teste sobre μ (populações infinitas)

σ^2 DESCONHECIDO

Estatística de teste: $t_c = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$.

As hipóteses consideradas nesse teste são:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases},$$

ou ainda,

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \quad \text{ou} \quad \mu < \mu_0 \end{cases}.$$

Teste bilateral:

$$\begin{cases} -t_{\alpha/2} < t_c < t_{\alpha/2}, & \text{Aceita-se } H_0 \\ \text{Caso contrário,} & \text{Rejeita-se } H_0 \end{cases}.$$

Teste unilateral superior

$$\begin{cases} t_c < t_{\alpha}, & \text{Aceita-se } H_0 \\ \text{Caso contrário,} & \text{Rejeita-se } H_0 \end{cases},$$

e teste unilateral inferior

$$\begin{cases} t_c > -t_{\alpha}, & \text{Aceita-se } H_0 \\ \text{Caso contrário,} & \text{Rejeita-se } H_0 \end{cases}$$

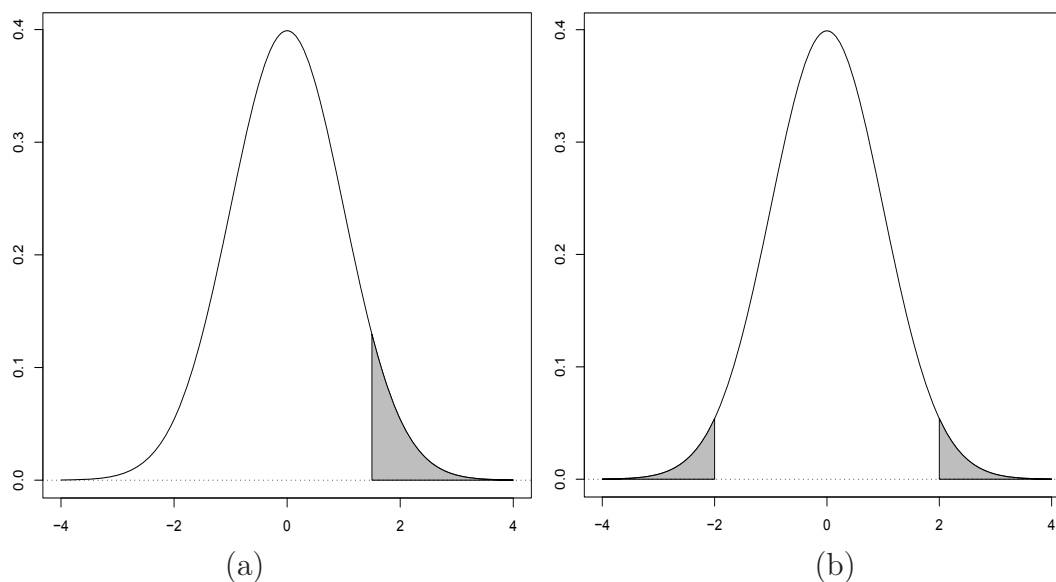


Figura 6.3: (a) Ilustração do teste unilateral superior. (b) Ilustração do teste bilateral. A área hachurada representa a região de rejeição do teste.



Veja como fazer as Figura 6.3 (a)...

```
x<-seq(-4,4,by =.01)
y<-dnorm(x, mean=0,sd=1, log = FALSE)
rx<-seq(1.5,4,by =.1)
ry<-numeric(2*length(rx))
ry[1:length(rx)]<-dnorm(rx, mean=0,sd=1, log = FALSE)
rx<-c(rx,rev(rx))
plot(x,y,'l',xlab='',ylab='')
polygon(rx, ry, col = "gray")
abline(h=0,lty=3)

... e (b)
rx<-seq(2,4,by =.1)
ry<-numeric(2*length(rx))
ry[1:length(rx)]<-dnorm(rx, mean=0,sd=1, log = FALSE)
rx<-c(rx,rev(rx))
rx1<- seq(-4,-2,by =.1)
ry1<-numeric(2*length(rx1))
ry1[1:length(rx1)]<-dnorm(rx1, mean=0,sd=1, log = FALSE)
rx1<-c(rx1,rev(rx1))
plot(x,y,'l',xlab='',ylab='')
polygon(rx, ry, col = "gray")
polygon(rx1, ry1, col = "gray")
abline(h=0,lty=3)
```

6.3.3 Teste sobre proporções (p)

Estatística de teste: $z_c = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}\hat{q}}{n}}}$

As hipóteses consideradas nesse teste são:

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases},$$

ou ainda,

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p > p_0 \text{ ou } p < p_0 \end{cases}.$$

Teste bilateral:

$$\begin{cases} -z_{\alpha/2} < z_c < z_{\alpha/2}, & \text{Aceita-se } H_0 \\ \text{Caso contrário,} & \text{Rejeita-se } H_0 \end{cases}.$$

Teste unilateral superior

$$\begin{cases} z_c < z_{\alpha}, & \text{Aceita-se } H_0 \\ \text{Caso contrário,} & \text{Rejeita-se } H_0 \end{cases},$$

e teste unilateral inferior

$$\begin{cases} z_c > -z_{\alpha}, & \text{Aceita-se } H_0 \\ \text{Caso contrário,} & \text{Rejeita-se } H_0 \end{cases}.$$

Correlação: mede o grau de relacionamento entre duas ou mais variáveis.

Regressão: é o estudo que busca ajustar uma equação a um conjunto de dados de forma que a relação entre variáveis possa ser descrita matematicamente.

7.1 Correlação

7.1.1 Coeficiente de correlação linear (r ou ρ)

Mede a correlação de duas variáveis.

$$\rho = \frac{\sum_{i=1}^n X_i Y_i - \frac{\left(\sum_{i=1}^n X_i\right) \left(\sum_{i=1}^n Y_i\right)}{n}}{\sqrt{\left(\sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i\right)^2}{n}\right) \left(\sum_{i=1}^n Y_i^2 - \frac{\left(\sum_{i=1}^n Y_i\right)^2}{n}\right)}} \quad (7.1)$$

7.1.2 Coeficiente de determinação (r^2 ou ρ^2)

Indica, percentualmente, quanto da variação da variável dependente (Y) que é explicada pelo modelo de regressão. Note que, no caso da Regressão Linear Simples, o modelo em questão é a reta.

$$\rho^2 = \frac{\text{Variação explicada pelo modelo}}{\text{Variação total}} \quad (7.2)$$

Vamos nos ater ao caso em que nosso modelo é uma reta, ou seja, estamos lidando com Regressão Linear. Nesse caso particular o coeficiente de determinação se torna

$$\rho^2 = \frac{\text{Variação explicada pela reta}}{\text{Variação total}} = \frac{SQRL}{SQT}, \quad (7.3)$$

em que $SQRL$ significa Soma de Quadros de Regressão Linear, SQT significa Soma de Quadrados Total e essas quantidades são calculadas pelas expressões

$$SQRL = \frac{\sum_{i=1}^n X_i Y_i - \frac{\left(\sum_{i=1}^n X_i\right) \left(\sum_{i=1}^n Y_i\right)}{n}}{\sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i\right)^2}{n}} \quad (7.4)$$

$$SQT = \frac{\left(\sum_{i=1}^n Y_i\right)^2}{n} \quad (7.5)$$

A diferença entre a variação total e a variação explicada pela reta de regressão é chamada de *desvio*, e pode ser calculada pela Soma de Quadrados de Desvios (SQD)

$$SQD = SQT - SQRL. \quad (7.6)$$

Além de estimar os coeficientes da reta de regressão é possível testar se eles são significativos. Por exemplo, pode-se testar se o coeficiente β_1 é estatisticamente igual a zero ou não. Se for considerado igual a zero, ou seja, se aceita-se H_0 nesse teste, isso significa que apenas a constante β_0 seria suficiente para explicar os dados, Y não varia com a variação de X . Por outro lado, se β_1 for considerado significativo, esse é um bom indicativo a favor do modelo estimado.

O seguinte teste pode ser feito o seguinte teste F para o ajuste de uma regressão linear.

HIPÓTESES

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 > \text{ ou } < \end{cases}$$

ESTATÍSTICA DE TESTE

$$F_c = \frac{SQRL}{S^2},$$

em que

$$S^2 = \frac{SQD}{n-2}.$$

TESTE

$$\begin{cases} F_c < F(\alpha, \nu_1, \nu_2), & \text{Aceita-se } H_0 \\ F_c > F(\alpha, \nu_1, \nu_2), & \text{Rejeita-se } H_0 \end{cases},$$

em que $F(\alpha, \nu_1, \nu_2)$ é o valor tabelado para a distribuição F com ν_1 e ν_2 graus de liberdade, e $100\alpha\%$ de probabilidade.

7.2 Regressão

Um estudo de regressão busca essencialmente associar uma variável Y (denominada variável-resposta) a um conjunto de outras p variáveis X_1, X_2, \dots, X_p (denominadas covariáveis ou variáveis explicadoras). Esta associação é segundo uma forma funcional do tipo

$$Y = f(X_1, X_2, \dots, X_p),$$

onde a função f pode ser, à princípio, qualquer uma. Quando f assume a forma funcional linear (isto é, f é uma combinação linear das covariáveis

$$f(X_1, X_2, \dots, X_p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p,$$

em que os coeficientes β_i ($i = 1, 2, \dots, p$) são números fixos chamados parâmetros), a regressão é chamada linear, caso contrário, é uma regressão não-linear. Numa regressão linear, quando $p = 1$ denominamos o estudo como regressão linear simples, caso contrário, denominamos regressão linear múltipla.

7.2.1 Regressão Linear Simples

Explica, por meio de uma reta, a relação entre duas variáveis.

MODELO ESTATÍSTICO

$$Y_i = \beta_0 + \beta_1 X_{1i} + e_i, \tag{7.7}$$

em que Y_i é o valor da variável Y para o indivíduo i ; β_0 é o intercepto; β_1 é o coeficiente linear; X_{1i} é o valor da (co)variável X para o indivíduo i ; e_i é o erro aleatório associado ao indivíduo i .

ESPERANÇA MATEMÁTICA

$$E[Y_i] = \beta_0 + \beta_1 X_{1i} \quad (7.8)$$

ESTIMAÇÃO DE PARÂMETROS POR MEIO DO MÉTODO DOS QUADRADOS MÍNIMOS

Existem diversos métodos que permitem estimar os parâmetros de interesse no contexto de Regressão Linear Simples (β_0 e β_1), mas aqui trataremos apenas do Método dos Quadrados Mínimos.

O Método dos Quadrados Mínimos fornece aqueles valores de β_0 e β_1 que minimizam a soma de quadrados dos resíduos, ou seja, que minimizam a distancia entre os valores observados e os estimados pelo modelo (reta), ao longo de todas as observações ao mesmo tempo.

Por exemplo, considere a Figura 7.1. Ela traz a representação de uma massa de dados fictícia. Duas variáveis quaisquer X e Y se relacionam de maneira diretamente proporcional, ou seja, são *positivamente correlacionadas*. Esse tipo de relacionamento sugere que uma reta pode ser o modelo ideal para descrever o comportamento, por exemplo, de Y , de acordo com o comportamento de X . A Figura 7.1 destaca que é possível infinitas retas passando por entre os pontos, porém, apenas uma delas possui a propriedade de estar a menor distância quadrática de todos os pontos ao mesmo tempo. Os coeficientes dessa reta podem ser estimados pelo Método dos Quadrados Mínimos por meio dos estimadores 7.2.1 e 7.2.1.



A seguir, a rotina para você reproduzir a Figura 7.1.

```
x<-seq(0,100)
y<-2*x+35
y1<-y+rnorm(101,0,50)
reg<-lm(y1~x)
a<-reg$coefficients[1]
b<-reg$coefficients[2]
y2<-a + b*x
y3<-(y2[51]-50*(b-1))+(b-1)*x
y4<-(y2[51]-50*(b+1))+(b+1)*x
y5<-(y2[51]-50*(b+2))+(b+2)*x
plot(x,y1,pch=19,xlab='X',ylab='Y')
lines(x,y2,lwd=2)
lines(x,y3,lty=3)
lines(x,y4,lty=3)
lines(x,y5,lty=3)
```

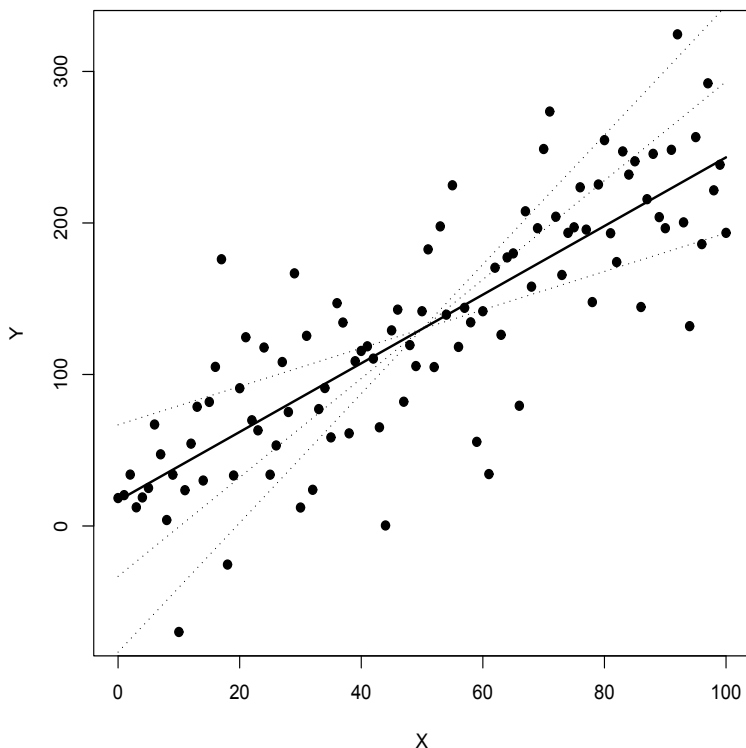


Figura 7.1: Representação de possíveis retas (pontilhadas) e aquela estimada por quadrados mínimos (linha cheia) em uma massa de dados fictícia.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n X_i Y_i - \frac{\left(\sum_{i=1}^n X_i\right) \left(\sum_{i=1}^n Y_i\right)}{n}}{\sum_{i=1}^n X_i^2 - \frac{\sum_{i=1}^n (X_i)^2}{n}} \quad (7.9)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = \frac{\sum_{i=1}^n Y_i}{n} - \hat{\beta}_1 \left(\frac{\sum_{i=1}^n X_i}{n} \right) \quad (7.10)$$

Exemplo: As famosas *cocadeiras* baianas costumam produzir suas próprias quitandas. Para isso, elas cumprem a difícil tarefa de quebrar dezenas de cocos por dia. Querendo evitar trabalho desnecessário, elas desejam quebrar apenas aqueles frutos que contiverem uma grande quantidade de polpa. Portanto, procedem da seguinte maneira: furam o coco, medem sua quantidade de água e, com base em sua experiência, decidem se vale a pena quebrar o mesmo. Como esse procedimento é impreciso, as baianas desejam a nossa ajuda. Considerando os dados apresentados na Tabela 7.1, vamos estimar um modelo adequado para estimar o volume de polpa de frutos de coco (Y) a

partir de sua quantidade de água (X).

O primeiro passo para a resolução desse problem é a *análise exploratória* dos dados, por exemplo, por meio de um *diagrama de dispersão* (Figura 7.2).

Tabela 7.1: Volume de polpa (cm^3), volume de água (cm^3) e teor de cálcio ($mg/100ml$) em 20 cocos verdes.

Fruto	Polpa	Água	Cálcio
1	9,02	17,87	7,52
2	13,10	13,75	24,77
3	14,76	12,72	30,74
4	21,54	6,98	7,58
5	15,62	11,01	23,33
6	18,34	10,48	20,49
7	20,23	10,19	14,84
8	8,88	19,11	5,51
9	14,06	12,72	31,03
10	23,59	0,45	4,33
11	16,62	10,67	21,75
12	21,93	1,59	4,92
13	10,56	14,91	19,50
14	12,28	14,14	20,16
15	20,68	9,40	12,20
16	9,53	16,23	11,54
17	13,73	12,74	29,39
18	5,73	20,64	3,79
19	15,08	12,34	23,40
20	21,57	6,44	7,47

Fonte: Dados fictícios.



A seguir, a rotina para você reproduzir a Figura 7.2

```
polpa<-c(9.02,13.10,14.76,21.54,15.62,18.34,20.23,8.88,14.06,23.59,16.62,21.93,
10.56,12.28,20.68,9.53,13.73,5.73,15.08,21.57)
agua<-c(17.87,13.75,12.72,6.98,11.01,10.48,10.19,19.11,12.72,0.45,10.67,1.59,
14.91,14.14,9.40,16.23,12.74,20.64,12.34,6.44)
plot(agua,polpa,xlab=expression(paste('Água (', cm^3,')'),),
ylab=expression(paste('Polpa (', cm^3 ,')'),))
```

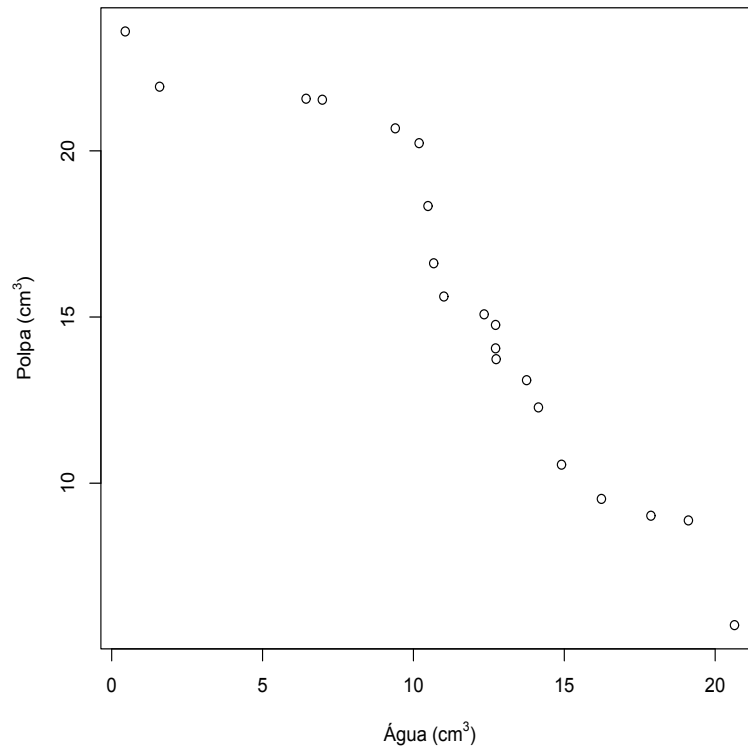



Figura 7.2: Diagrama de dispersão entre a variável independente volume de água de coco e a variável dependente volume de polpa de coco.

Os dados amostrais sugerem que, quanto mais água tem um coco, menos polpa ele possui. Além dessa informação, o diagrama de dispersão nos faz suspeitar que as duas variáveis possuem uma relação linear, ou seja, uma reta pode ser um bom modelo para este caso. Vamos estimar os coeficientes dessa reta de duas formas.

(a) Usando uma calculadora: Uma calculadora das mais simples pode nos ajudar a estimar os coeficientes β_0 e β_1 facilmente. Primeiramente, façamos uma tabela auxiliar com todas as quantidades de interesse (Tabela 7.2).

De posse dessa tabela só precisamos substituir as quantidades nas fórmulas dos coeficientes.

$$\hat{\beta}_1 = \frac{3119,430 - \frac{(234,38)(306,85)}{20}}{3253,462 - \frac{(234,38)^2}{20}} = \frac{3119,430 - 3595,975}{3253,462 - 2746,7} \approx -0,94$$

$$\hat{\beta}_0 = \frac{306,85}{20} + 0,94 \left(\frac{234,38}{20} \right) \approx 26,36$$

Tabela 7.2: Tabela auxiliar para cálculo dos coeficientes do modelo linear.

X	Y	X^2	XY
17,87	9,02	319,3369	161,1874
13,75	13,10	189,0625	180,1250
12,72	14,76	161,7984	187,7472
6,98	21,54	48,7204	150,3492
11,01	15,62	121,2201	171,9762
10,48	18,34	109,8304	192,2032
10,19	20,23	103,8361	206,1437
19,11	8,88	365,1921	169,6968
12,72	14,06	161,7984	178,8432
0,45	23,59	0,2025	10,6155
10,67	16,62	113,8489	177,3354
1,59	21,93	2,5281	34,8687
14,91	10,56	222,3081	157,4496
14,14	12,28	199,9396	173,6392
9,40	20,68	88,3600	194,3920
16,23	9,53	263,4129	154,6719
12,74	13,73	162,3076	174,9202
20,64	5,73	426,0096	118,2672
12,34	15,08	152,2756	186,0872
6,44	21,57	41,4736	138,9108
234,38	306,85	3253,462	3119,430

(b) Usando o R: É ainda mais fácil resolver tal problema usando o R. Primeiro devemos inserir a massa de dados, por exemplo, em vetores. Em seguida, podemos usar a função *lm()*, que ajusta modelos lineares. Para especificarmos que desejamos uma reta, basta dizer que os dados do vetor *polpa* são função dos dados do vetor *água*. Da seguinte maneira

```
polpa<-c(9.02, 13.10, 14.76, 21.54, 15.62, 18.34, 20.23, 8.88,
14.06, 23.59, 16.62, 21.93, 10.56, 12.28, 20.68, 9.53, 13.73,
5.73, 15.08, 21.57)
agua<-c(17.87, 13.75, 12.72, 6.98, 11.01, 10.48, 10.19, 19.11,
12.72, 0.45, 10.67, 1.59, 14.91, 14.14, 9.40, 16.23, 12.74,
20.64, 12.34, 6.44)
lm(polpa~agua)
```

O resultado obtido por esse procedimento é exatamente o mesmo. As únicas diferenças se devem a arredondamentos feito no item anterior. Isso mostra que ambos procedimentos são equivalentes.

Finalmente, podemos informar para as baianas que o modelo estimado é a reta $Polpa = 26,36 - 0,94 * Água$, e ilustrar o resultado com a reta estimada plotada junto aos pontos amostrais (Figura 7.3).

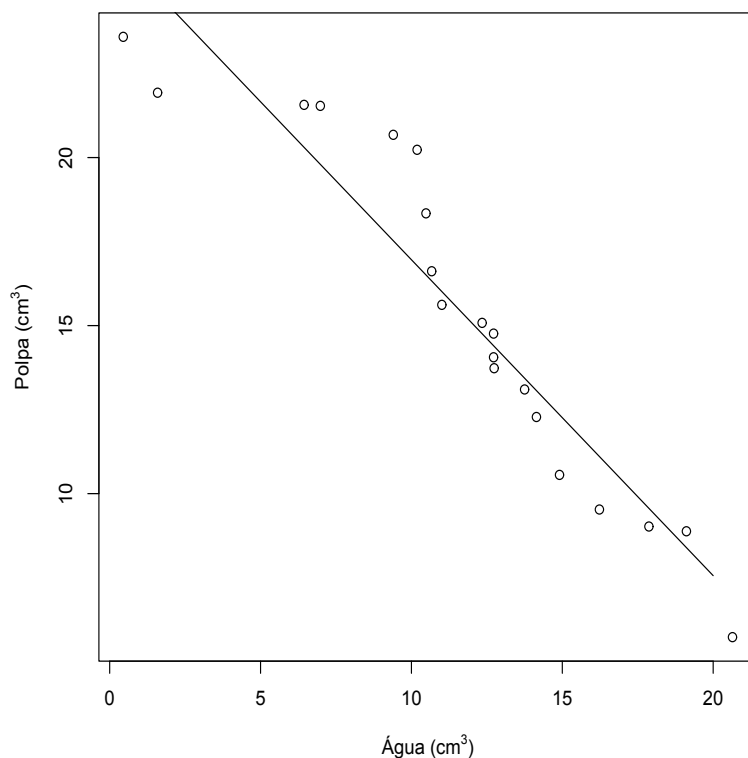


Figura 7.3: Reta de regressão estimada e pontos amostrais.



Para se reproduzir a Figura 7.3 deve-se repetir o procedimento do diagrama de dispersão e acrescentar os seguintes comandos:

```
x<-seq(1:20)
reta<-26.36-0.94*x
lines(x,reta)
```

7.2.2 Regressão Múltipla

Sistema de Equações Normais (SEN):

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \dots + \beta_k x_{k1} + e_1 \\ y_2 &= \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \dots + \beta_k x_{k2} + e_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \dots + \beta_k x_{kn} + e_n \end{aligned}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

Expressando esse sistema em notação matricial,

$$Y = X\Theta + \varepsilon \quad (7.11)$$

Diferenciando-se essa expressão com respeito a Θ e igualando-se a zero, tem-se que o estimador de quadrados mínimos do vetor Θ é

$$\begin{aligned} X'X\Theta &= X'Y \\ \hat{\Theta} &= (X'X)^{-1}X'Y. \end{aligned} \quad (7.12)$$

APÊNDICE A: EXERCÍCIOS PROPOSTOS

LISTA 1: Técnicas de Somatório

1. Sejam os conjuntos $X = \{2, 4, 4, 3, 2\}$ e $Y = \{1, 2, 3, 6, 7\}$. Obtenha:

$$1.1. \sum_{i=1}^4 X_i$$

$$1.2. \sum_{i=1}^5 Y_i$$

$$1.3. \sum_{i=1}^4 4X_i^2$$

$$1.4. \sum_{i=1}^5 X_i Y_i$$

$$1.5. \sum_{i=1}^5 (3X_i + 2Y_i)$$

$$1.6. \sum_{i=2}^4 X_i Y_i + \sum_{i=1}^5 Y_i^2.$$

2. Duas importantes estatísticas são a média e a variância amostral, respectivamente, \bar{X} e S^2 . Essas duas quantidades dependem de somatórios. Portanto, dada a amostra $X = \{2, 4, 5, 6, 1, 8\}$, calcule a sua média e variância:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}, S^2 = \frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n} \right].$$

3. Demonstre numérica e algebricamente que $\sum_{i=1}^n (X_i - \bar{X}) = 0$. Use os dados do exercício anterior para demonstrar numericamente.

4. Descubra qual é o valor de A para que a função $Q(A) = \frac{\sum_{i=1}^n (X_i - A)^2}{n-1}$ atinja seu mínimo. (Dica: sendo $Q(A)$ uma função quadrática em A , seu mínimo pode ser encontrado usando sua derivada.)
-

LISTA 2: Distribuição de freqüências e histograma

1. Para cada um dos cenários a seguir, responda as perguntas de (a) a (e):

I. Uma cooperativa agrícola deseja realizar uma pesquisa com o objetivo de caracterizar as propriedades rurais de seus cooperados. Como forma de obter estas informações foram distribuídos questionários para todos eles, por meio dos quais procurou-se avaliar o nível tecnológico adotado (baixo, médio ou alto), a atividade predominante na fazenda (café, milho, leite etc.) e o número de empregados da propriedade.

II. Um pesquisador, para obter informações a respeito de uma cultura atualmente plantada no sul de Minas Gerais, visita 50 propriedades e faz uma avaliação referente ao tamanho da área plantada com a cultura (ha), a produção obtida (Kg) e as principais pragas e doenças.

- (a) Qual é a população em estudo?
- (b) Classifique essa população (finita ou infinita).
- (c) Utilizou-se uma amostra para realizar o estudo? Por quê?
- (d) Quais foram as variáveis estudadas?
- (e) Classifique essas variáveis quanto à sua natureza.

2. Foi feito um levantamento dos pesos (Kg) de 48 crianças de 8 a 12 anos:

19,1	20,0	23,3	24,2	28,2	28,2	30,5	30,6	31,0	31,1	32,0	32,0
34,0	34,2	35,2	35,3	36,3	36,5	36,7	37,2	37,2	38,6	38,7	39,0
39,0	39,1	39,2	39,4	39,4	41,0	42,1	42,2	43,0	43,0	44,3	44,4
44,5	45,2	47,5	48,3	49,2	50,1	56,2	57,1	57,2	58,2	60,2	60,3

- (a) Construa uma distribuição de freqüências de ocorrência.
- (b) Construa o histograma e o polígono de freqüência.
- (c) Obtenha as distribuições de freqüências acumuladas “acima de” e “abaixo de” e construa as ogivas.

3. Nos cenários a seguir identifique: (i) a população; (ii) a classificação da população; (iii) a variável de interesse; (iv) a classificação da variável.

- (a) Estudo da distribuição da renda familiar dos moradores da cidade de Lavras (MG), utilizando os dados do Censo 2000 do IBGE.
 - (b) Estudo do nível escolar de todos os funcionários de uma cooperativa.
 - (c) Estudo da identificação das espécies florestais que ocorrem na região a ser alagada pela Hidrelétrica do Funil. Serão demarcadas regiões representativas para este levantamento.
 - (d) Estudo da área (em ha) das propriedades rurais do sul de Minas Gerais. Como há muitas propriedades, apenas algumas delas serão visitadas.
-

4. Foi contado o número de lagartas rosca (*Agrotis ipisilon*) em todos os 50 canteiros de mudas de eucalipto da Fazenda Experimental da UFLA, encontrando-se o seguinte resultado:

5	1	5	3	1	2	2	1	1	0	1	1	3	2	3	1	2	3	4	3	2
4	2	0	4	4	4	4	3	3	2	4	0	2	0	5	2	3	4	3	4	0

- Estes dados são de uma população ou de uma amostra?
- Classifique a variável em questão.
- Construa a distribuição de frequências (fa) absolutas e acumuladas ($F \downarrow$, $F \uparrow$).
- Construa um gráfico adequado para representar essas observações.

5. Uma pesquisa sobre renda familiar foi feita com 100 pessoas entrevistadas ao acaso no centro de Lavras (Tabela 8.1).

Tabela 8.1: Distribuição de frequência de rendas familiares de 100 entrevistados, em Lavras, MG.

Classes (R\$)	\bar{x}_i	fa	fr	dfr
[0, 00; 300, 00)	150,00	70	0,70	$0,0023333 = 23.333 \times 10^7$
[300, 00; 1.000, 00)	650,00	25	0,25	$0,0003571 = 3.571 \times 10^7$
[1.000, 00; 5.000, 00)	3.000,00	4	0,04	$0,0000100 = 100 \times 10^7$
> 5.000, 00		1	0,01	$0,0000004 = 4 \times 10^7$
Total	-	100	1,00	-

Fonte: Dados fictícios.

- Faça o histograma com a frequência relativa (fr) de ocorrência e outro com a densidade de frequência relativa (dfr). Compare-os. Em que eles se diferem? Qual dos dois é correto? Por quê?
- Qual é o limite superior da última classe utilizado pelo autor para calcular suas frequências?

LISTA 3: Medidas de posição

1. Calcule a média, mediana e a moda dos exercícios 2 e 4 da segunda aula prática. Interprete cada medida.
2. Considere a seguinte distribuição de frequências.

Tabela 8.2: Distribuição de frequências das médias diárias de produção de leite no período de lactação de 201 vacas da raça holandesa, de um rebanho pertencente ao Núcleo de Criadores de Gado Holandês do Sul de Minas Gerais. Lavras, 1992.

Produção de Leite	fa	fr	fp
[5, 25; 8, 15)	2	0,0100	1,00
[8, 15; 11, 05)	5	0,0249	2,49
[11, 05; 13, 95)	23	0,1144	11,44
[13, 95; 16, 85)	38	0,1891	18,91
[16, 85; 19, 75)	48	0,2388	23,88
[19, 75; 22, 65)	37	0,1841	18,41
[22, 65; 25, 55)	29	0,1443	14,43
[25, 55; 28, 45)	13	0,0646	6,46
[28, 45; 31, 35)	3	0,0149	1,49
[31, 35; 34, 25]	3	0,0149	1,49
Totais	201	1,0000	100,00

Fonte: Trabalho prático de estudantes de Estatística do ano de 1992.

- (a) Calcule as medidas de posição usando os dados elaborados.
- (b) Calcule os percentis de 1%, 5%, 25%, 50%, 75%, 95% e 99%, usando a fórmula:

$$P_i = L_{P_i} + \left[\frac{\frac{in}{100} - \sum fa}{f_{p_i}} \right] c,$$

em que L_{P_i} é o limite inferior da classe do percentil; $\sum fa$ é o somatório das frequências das classes anteriores; f_{p_i} é a frequência da classe do percentil.

LISTA 4: Medidas de dispersão

1. Demonstre a igualdade algébrica a seguir e utilize-a para estabelecer uma fórmula menos trabalhosa para o desvio-padrão:

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i\right)^2}{n}.$$

Nota: Por menos trabalhosa entenda um número menor de operações aritméticas, dado um conjunto de dados de tamanho n . Verifique que, para a fórmula do desvio padrão que utiliza o primeiro membro como numerador, são necessárias $4n + 2$ operações, enquanto que, ao utilizar a identidade algébrica acima serão necessárias apenas $3n + 4$ operações.

2. Calcule o desvio-padrão, a variância e o coeficiente de variação dos exercícios 2 e 4 da segunda aula prática. Interprete cada medida.
 3. Considere a tabela 8.2 da lista de exercícios número 3. Calcule as medidas de variabilidade usando os dados elaborados.
-

LISTA 5: Probabilidade

- A probabilidade que um homem esteja vivo daqui a 30 anos é $\frac{2}{5}$ e a da sua mulher é $\frac{2}{3}$. Determinar a probabilidade que daqui a 30 anos:
 - Ambos estejam vivos;
 - Somente o homem esteja vivo;
 - Somente a mulher esteja viva;
 - Nenhum esteja vivo;
 - Pelo menos um esteja vivo.
- Uma companhia que fura poços artesianos trabalha numa região escolhendo aleatoriamente o ponto de furo. Não encontrando água nessa tentativa, sorteia outro local e, caso também não obtenha sucesso, faz uma terceira e última tentativa. Admita probabilidade 0,7 de encontrar água em qualquer ponto dessa região. Determine o espaço amostral e calcule a probabilidade de:
 - Encontrar água na segunda tentativa.
 - Encontrar água em até duas tentativas.
 - Encontrar água.
- A tabela a seguir apresenta dados dos 1000 ingressantes de uma universidade, com informações sobre área de estudo e classe sócio econômica.

Área	Classe		
	Alta	Média	Baixa
Exatas	120	156	68
Humanas	72	85	112
Biológicas	169	145	73

Se um aluno ingressante é escolhido ao acaso, determine a probabilidade de:

- Ser da classe econômica mais alta.
 - Estudar na área de Exatas.
 - Estudar na área de Humanas, sendo de classe média.
 - Ser da classe baixa, dado que estuda na área de Biológicas.
- Estatísticas dos últimos anos do departamento estadual de estradas são apresentadas na tabela a seguir, contendo o número de acidentes com vítimas, fatais ou não, e as condições do principal motorista envolvido, sóbrio ou alcoolizado.

Motoristas	Vítimas	
	Não fatais	Fatais
Sóbrio	1228	275
Alcoolizado	2393	762

Você diria que o fato do motorista estar ou não alcoolizado interfere na ocorrência de vítimas fatais?

LISTA 6: Distribuição Binomial/Poisson

1. Sabe-se que 5% de um rebanho bovino está com febre aftosa. Qual a probabilidade de que num lote de seis animais retirados desse rebanho tenha-se:

- (a) nenhum animal com febre aftosa?
- (b) dois animais com febre aftosa?
- (c) mais de um animal com febre aftosa?

2. Numa leitegada de cinco leitões qual a probabilidade de:

- (a) não haver fêmeas?
- (b) haver duas fêmeas?
- (c) haver pelo menos duas fêmeas?

3. Numa criação de coelhos 40% são machos. Num dia em que nasçam vinte coelhos, qual a probabilidade de nascer:

- (a) cinco coelhos machos?
- (b) pelo menos dois coelhos machos?

4. Suponha que a percentagem de germinação de uma semente de feijoeiro seja de 60%. Serão semeadas três sementes por cova em um canteiro com vinte e quatro covas.

- (a) Qual a probabilidade de obter-se pelo menos uma cova falhada no canteiro?
- (b) Qual será o número esperado de covas falhadas no canteiro?

5. Numa lâmina verificou-se que existem em média 5 bactérias/ cm^2 . A lâmina foi subdividida em 300 quadros de $1cm^2$. Em quantos destes quadros em média você espera encontrar no máximo 6 bactérias? Qual é a probabilidade de se encontrar mais de 4 bactérias por centímetro quadrado?

6. Um pesquisador da área de zootecnia conseguiu uma série de dados dos últimos 120 anos, com o registro do número de ocorrências de uma doença rara em eqüinos da localidade em que trabalhava. Os dados obtidos foram:

Ocorrências	0	1	2	3	4	5
Anos	55	40	17	5	2	1

- (a) Estime o número médio de doenças/ano.
- (b) Calcule para cada valor de X , as probabilidades associadas. Suponha que X possua distribuição de Poisson.
- (c) Calcule a frequência esperada (em anos) para cada valor de X .
- (d) Compare os resultados esperados com os observados. Com base nesta comparação, você pode afirmar que a distribuição de Poisson é adequada para explicar a ocorrência desta doença na região de estudo? Justifique.

7. A granja acima havia comprado um lote de vacinas com deficiência nominal de imunização rotulada como “1 animal vacinado não imunizado em cada 2500 animais vacinados, em média”. Para um teste, a vacina foi aplicada em um lote de 5000 animais. Depois de decorrido algum tempo, constatou-se que 4 animais manifestaram a doença. Como você analisaria estatisticamente este resultado?

LISTA 7: Distribuição Normal

1. Considerando que os pesos dos coelhos Norfolk ao abate aos 90 dias obedeça uma distribuição Normal de frequências, com média de $2,70kg$ e variância de $0,04kg^2$. Responda:
 - (a) Qual a frequência de coelhos com peso acima de $2,90 kg$?
 - (b) E entre $2,90$ e $3,00 kg$?
 - (c) Qual o peso que é superado por apenas 1% dos coelhos?
 2. O diâmetro X das esferas de rolamentos fabricadas por certa indústria tem distribuição Normal, com média $0,6140cm$ e variância $0,00252cm^2$. O lucro T de cada esfera depende de seu diâmetro segundo a função abaixo: $T = R\$10,00$ se a esfera é boa ($0,6080 < X < 0,6200$) ou $T = R\$1,10$ se a esfera é defeituosa ($X < 0,6080$ ou $X > 0,6200$) Calcule:
 - (a) as frequências de ocorrência de esferas boas.
 - (b) as frequências de ocorrência de esferas defeituosas.
 - (c) a média de T .
 3. Uma máquina de empacotar café produz pacotes de café com pesos segundo uma distribuição Normal de frequências, com média de $500g$ e desvio-padrão de $3g$.
 - (a) Num lote de 10.000 pacotes, em quantos você espera encontrar menos de $490g$ de café?
 - (b) Considere que é possível ajustar a média com que os pacotes são cheios. Qual deve ser o ajuste da média para que 99% dos pacotes não tenham peso inferior a $500g$?
 4. Em uma granja de frangos, um estudo estatístico cuidadoso determinou que a lei de frequências de ocorrência de peso final (em kg) desses frangos, dentro de um período pré-especificado de tempo, é Normal, com $1.881g$ de média e desvio-padrão igual a $210g$.
 - (a) Você julga possível estes animais serem submetidos a condições de manejo, alimentação, ambiente e outros fatores não-homogêneos?
 - (b) Se um frango é escolhido aleatoriamente, qual seria a medida de sua expectativa de que tal animal teria peso superior a $2.000g$?
-

LISTA 8: Amostragem

1. Uma pessoa retirou três maçãs da superfície de cada uma das caixas de maçãs que estavam em uma quitanda, para verificar a sua qualidade. Isto é uma amostra aleatória? Existe algum problema com este método de amostragem? Comente.

2. Deseja-se testar durante um mês um novo tipo de ração alimentícia em vacas leiteiras. O objetivo é conhecer o incremento médio da produção de leite por vaca, quando é utilizada a nova ração. Para isto, planejou-se determinar a diferença entre a produção do leite do mês em que foi fornecida a nova ração e a produção do mês anterior de cada vaca. Sabe-se que, em qualquer caso, antes e depois da ração, a produção de leite de vacas jovens é superior (ou pelo menos diferente) da produção de vacas adultas, sendo esta diferença significativa. A granja conta atualmente com 653 vacas leiteiras, sendo que, após uma análise estatística e de custos, determinou-se aplicar a ração em 36 animais.

- (a) Qual é a população em estudo?
- (b) Qual é amostra?
- (c) Qual é o tamanho da população e qual é o tamanho da amostra?
- (d) A população é finita ou infinita? Por quê?
- (e) Qual seria o parâmetro que se deseja conhecer? É possível conhecer o valor exato desse parâmetro?
- (f) Qual é o estimador que você utilizaria para estimar o parâmetro do item (e)?
- (g) Para esse tipo de estudo, qual tipo de amostragem você recomendaria utilizar? Por quê?

3. Planeje uma amostragem aleatória sistemática para amostrar 20 hastes de amortecedores da linha de produção da Magnetti Marelli Cofap de Lavras, durante um turno de produção de 6300 hastes aproximadamente. Sorteie as hastes que deverão ser selecionadas.

4. Uma empresa cafeeira do Sul de Minas dispõe de 3.200 funcionários distribuídos nas diversas atividades conforme o quadro abaixo:

Atividade	Empregados
Campo	1.600
Armazém	720
Indústria	480
Administração	240
Gerência	160

(a) Na sua opinião seria razoável levantar as informações desejadas através de uma amostragem aleatória simples de $n = 160$ funcionários? Justifique.

(b) Planeje uma amostragem estratificada de $n = 160$, determinando o tamanho da amostra para cada atividade.

(c) Usando a calculadora, ou um software, ou a tabela de números aleatórios, sorteie os componentes da amostra para os empregados que trabalham na gerência.

5. Foi tomada uma amostra aleatória simples de 17 bovinos no rebanho de uma grande fazenda de gado de corte, fornecendo os seguintes valores para o peso destes animais:

222,5	250,0	268,7	267,3	277,0	206,1	241,4	265,4	233,1
215,0	212,2	227,9	231,6	207,3	235,3	240,2	225,9	

6. Quantos animais adicionais você recomenda amostrar para que o erro de estimação caia 50%, em relação ao erro calculado com o tamanho da amostra inicial?

7. Quantos animais devem ser amostrados se não se admitir errar mais do que 1% da média para mais ou para menos?

LISTA 9: Distribuição de amostragem da média e intervalo de confiança

1. Numa Universidade foi tomada uma amostra de 40 estudantes, anotando-se as suas alturas em cm. Os resultados forneceram:

$$\sum_{i=1}^{40} x_i = 6950 \quad \sum_{i=1}^{40} x_i^2 = 1213463$$

- Encontre as estimativas por ponto de m e s .
 - Construa o intervalo de confiança de 95% para a média da população. Interprete.
2. Determine o intervalo de confiança para média com 90%, na seguinte situação:

$$\bar{x} = 15 \quad S_x = 2 \quad n = 16 \quad N = 200$$

3. Um pecuarista se entusiasmou por uma nova ração amplamente divulgada pelos meios de comunicação. Para verificar a eficiência da ração, ele selecionou uma AAS de 49 bois de seu rebanho e os alimentou por 30 dias obtendo um ganho de peso médio de $31,7kg$ com um desvio padrão de $2,6kg$.

- Construa o intervalo de confiança de 95% para a média e interprete.
- Qual deverá ser o tamanho da amostra para que o erro não seja superior a $0,7kg$ com probabilidade de 95%.

4. Foi feito uma amostra aleatória simples de tamanho $n = 30$ do rebanho $N = 201$ do Núcleo dos Criadores de Gado Holandês do Sul de Minas com o objetivo de descrever a produção de leite. Os dados obtidos na amostra foram:

17,7	20,7	19,3	19,3	18,0	16,9	19,7	20,1	21,0	21,2
23,3	15,3	23,7	18,8	25,2	18,0	22,8	21,1	18,8	25,9
19,3	19,6	26,6	14,3	19,7	32,7	14,1	16,8	19,7	19,3

- Estime a média e a variância da população
- Estime a proporção dos animais que produzem menos que $20kg$ de leite.
- Construa o intervalo de confiança para a produção total do rebanho com coeficiente de confiança de 95% e considere $s = 4,91kg$.

5. Um levantamento amostral sobre aspectos de higiene e saúde envolvendo bairros periféricos de Lavras, mostrou, entre outros fatos, a seguinte resposta a pergunta: "você lava sua caixa d'água?"

Resposta	Número de residências
Nunca	13
3 em 3 meses	11
6 em 6 meses	4
Anualmente	22
Poucas vezes	18
Total	68

Nota: Considere 5.000 residências na periferia de Lavras, quando esta pesquisa foi feita (1992).

Construa um IC para a proporção de todas as residências na periferia de Lavras que nunca lavam, ou lavam poucas vezes. Construa um IC para o número total de residências na periferia de Lavras que nunca lavam ou poucas vezes lavam.

LISTA 10: Intervalo de Confiança e Testes de Hipótese

1. Para comparar 2 sistemas do plantio de arroz, foram cultivadas 16 parcelas de $10m^2$ ($2m \times 5m$), anotando-se os rendimentos (kg). Os resultados foram:

Arroz de sequeiro	1,1	1,0	1,4	1,3	1,5	0,9				
Arroz irrigado	2,0	2,5	2,4	2,1	1,8	1,9	1,9	2,3	2,5	2,6

Calcule um IC para a verdadeira diferença de produtividade dos 2 sistemas. Considere variâncias iguais.

2. Um administrador colecionou dados sobre aumento na produtividade no último ano para uma amostra de empresas que produzem equipamentos para mecanização agrícola. As firmas foram classificadas de acordo com seus investimentos em pesquisa e desenvolvimento nos últimos 3 anos. Os resultados do estudo seguem abaixo, onde o aumento na produtividade é medido numa escala de 0 à 10.

Investimento	Firmas					
	1	2	3	4	5	6
Baixo	7,6	8,2	6,8	5,8	6,9	6,6
Alto	8,5	9,7	10,1	7,8	9,6	9,5

Considere as variâncias populacionais diferentes, e construa um IC para a verdadeira diferença entre as médias de produtividade.

3. Acredita-se que peixes do gênero *Astymax* (Lambari) da represa de Camargos possuem seus pesos distribuídos segundo uma Normal, com média $13,36g$ e um desvio padrão de $4,79g$. Um Lambari com $25,2g$ foi pescado recentemente. Esta observação coloca em dúvida o que se acredita? Embase estatisticamente sua resposta.

4. Os produtores de uma certa semente de milho afirmam que seu poder de germinação é de 92%. Um fazendeiro tomou 10 sementes e plantou-as para testar se é verdadeira tal afirmação. Oito sementes germinaram. Este fato pode substituir cientificamente a acusação de que as sementes não têm 92% de germinação?

5. Indique claramente quais as hipóteses de nulidade e alternativa que seriam convenientes para serem testadas nas seguintes situações:

(a) Um laticínio afirma que o teor de gordura do seu creme de leite é de 25%, mas alguns consumidores têm dúvidas a esse respeito.

(b) Um produtor rural calcula que a viabilidade econômica de sua infra-estrutura é atingida com uma produtividade de $8,2t/ha$, e deseja saber se corresponde à produtividade atual.

(c) Um pecuarista utiliza medicamentos diferentes em dois rebanhos e deseja saber se eles determinam condições sanitárias diferentes.

(d) O limite de tolerância para comercialização de uma certa virose em campos de produção de batata-semente é de 10%; deseja-se saber se determinado campo deve ser condenado.

6. Um gerente de cooperativa suspeita que um produtor tenha fraudado o leite.

- (a) Quais os tipos de erro para a hipótese H_0 : “leite normal”.
- (b) Qual o tipo de erro que é menos grave?

7. Uma montadora de automóveis anuncia que seus carros consomem, em média, 11 litros a cada 100Km, com desvio padrão de 0,8 litros (considere que o consumo possua uma distribuição Normal). Você compra um automóvel dessa marca, obtendo 11,3 litros por 100Km, como consumo médio. O que você pode concluir sobre o anúncio da montadora a um nível de significância de 5%.

LISTA 11: Regressão e correlação

1. Use os valores dados abaixo para estimar a equação de regressão e plote a reta de regressão:

$$\sum_{i=1}^{20} x_i = 200 \quad \sum_{i=1}^{20} y_i = 300 \quad \sum_{i=1}^{20} x_i y_i = 6200 \quad \sum_{i=1}^{20} x_i^2 = 3600 \quad n = 20$$

2. Foi feito um estudo sobre adição de sulfato de cálcio (CaSO_4) à uréia e seu efeito no ganho de peso de novinhos tratados com cana. Os resultados obtidos foram os seguintes:

X	0	5	10	15	20
Y	495	560	590	620	615

Em que X é a concentração de CaSO_4 e Y o ganho de peso diário em gramas.

- Faça um diagrama de dispersões para os dados.
- Estime a equação de regressão que se ajusta aos dados
- Plote a equação estimada.
- Interprete praticamente a equação de regressão.
- Calcule e interprete as três variações admitidas no modelo
- Teste o ajuste de regressão linear aos dados através do Teste F ($\alpha = 5\%$)
- Calcule o coeficiente de correlação e intérprete
- Encontre o coeficiente de determinação (r^2).

3. Procurou-se realizar um estudo com o objetivo de saber o efeito na produção de leite de um grupo de vacas tratadas com diferentes níveis de proteína. Foram obtidos os seguintes dados:

X	10	12	14	16	18	20	22
Y	11,8	10,2	12,1	13,2	12,1	15,4	15,6

Pede-se o mesmo que no exercício anterior.

Tabela 9.2: Probabilidades (α) da distribuição normal padrão $N(0, 1)$ para valores do quantil Z_t padronizado de acordo com o seguinte evento: $P(Z > Z_t) = \alpha$.

Z_t	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,000

Tabela 9.3: Quantis superiores da distribuição de qui-quadrado (χ^2_α) com ν graus de liberdade e para diferentes valores da probabilidade (α) de acordo com o seguinte evento: $P(\chi^2 > \chi^2_\alpha) = \alpha$.

ν	0,995	0,990	0,975	0,950	0,900	0,750	0,500
1	0,000039	0,000157	0,000982	0,003932	0,015791	0,101532	0,455
2	0,010025	0,020101	0,050636	0,102587	0,210721	0,575364	1,386
3	0,071721	0,114831	0,215793	0,351843	0,584369	1,213	2,366
4	0,206989	0,297109	0,484418	0,710723	1,064	1,923	3,357
5	0,411742	0,554298	0,831212	1,145	1,610	2,675	4,351
6	0,675727	0,872090	1,237	1,635	2,204	3,455	5,348
7	0,989256	1,239	1,690	2,167	2,833	4,255	6,346
8	1,344	1,646	2,180	2,733	3,490	5,071	7,344
9	1,735	2,088	2,700	3,325	4,168	5,899	8,343
10	2,156	2,558	3,247	3,940	4,865	6,737	9,342
11	2,603	3,053	3,816	4,575	5,578	7,584	10,341
12	3,074	3,571	4,404	5,226	6,304	8,438	11,340
13	3,565	4,107	5,009	5,892	7,042	9,299	12,340
14	4,075	4,660	5,629	6,571	7,790	10,165	13,339
15	4,601	5,229	6,262	7,261	8,547	11,037	14,339
16	5,142	5,812	6,908	7,962	9,312	11,912	15,338
17	5,697	6,408	7,564	8,672	10,085	12,792	16,338
18	6,265	7,015	8,231	9,390	10,865	13,675	17,338
19	6,844	7,633	8,907	10,117	11,651	14,562	18,338
20	7,434	8,260	9,591	10,851	12,443	15,452	19,337
21	8,034	8,897	10,283	11,591	13,240	16,344	20,337
22	8,643	9,542	10,982	12,338	14,041	17,240	21,337
23	9,260	10,196	11,689	13,091	14,848	18,137	22,337
24	9,886	10,856	12,401	13,848	15,659	19,037	23,337
25	10,520	11,524	13,120	14,611	16,473	19,939	24,337
26	11,160	12,198	13,844	15,379	17,292	20,843	25,336
27	11,808	12,879	14,573	16,151	18,114	21,749	26,336
28	12,461	13,565	15,308	16,928	18,939	22,657	27,336
29	13,121	14,256	16,047	17,708	19,768	23,567	28,336
30	13,787	14,953	16,791	18,493	20,599	24,478	29,336
40	20,707	22,164	24,433	26,509	29,051	33,660	39,335
50	27,991	29,707	32,357	34,764	37,689	42,942	49,335
60	35,534	37,485	40,482	43,188	46,459	52,294	59,335
120	83,852	86,923	91,573	95,705	100,624	109,220	119,334
240	187,324	191,990	198,984	205,135	212,386	224,882	239,334
480	403,949	410,874	421,189	430,198	440,745	458,754	479,334
960	850,891	861,015	876,028	889,081	904,291	930,093	959,333

Tabela 9.4: Quantis superiores da distribuição de qui-quadrado (χ^2_α) com ν graus de liberdade e para diferentes valores da probabilidade (α) de acordo com o seguinte evento: $P(\chi^2 > \chi^2_\alpha) = \alpha$.

ν	0,500	0,250	0,100	0,050	0,025	0,010	0,005
1	0,454940	1,323	2,706	3,841	5,024	6,635	7,879
2	1,386	2,773	4,605	5,991	7,378	9,210	10,597
3	2,366	4,108	6,251	7,815	9,348	11,345	12,838
4	3,357	5,385	7,779	9,488	11,143	13,277	14,860
5	4,351	6,626	9,236	11,070	12,833	15,086	16,750
6	5,348	7,841	10,645	12,592	14,449	16,812	18,548
7	6,346	9,037	12,017	14,067	16,013	18,475	20,278
8	7,344	10,219	13,362	15,507	17,535	20,090	21,955
9	8,343	11,389	14,684	16,919	19,023	21,666	23,589
10	9,342	12,549	15,987	18,307	20,483	23,209	25,188
11	10,341	13,701	17,275	19,675	21,920	24,725	26,757
12	11,340	14,845	18,549	21,026	23,337	26,217	28,300
13	12,340	15,984	19,812	22,362	24,736	27,688	29,819
14	13,339	17,117	21,064	23,685	26,119	29,141	31,319
15	14,339	18,245	22,307	24,996	27,488	30,578	32,801
16	15,338	19,369	23,542	26,296	28,845	32,000	34,267
17	16,338	20,489	24,769	27,587	30,191	33,409	35,718
18	17,338	21,605	25,989	28,869	31,526	34,805	37,156
19	18,338	22,718	27,204	30,144	32,852	36,191	38,582
20	19,337	23,828	28,412	31,410	34,170	37,566	39,997
21	20,337	24,935	29,615	32,671	35,479	38,932	41,401
22	21,337	26,039	30,813	33,924	36,781	40,289	42,796
23	22,337	27,141	32,007	35,172	38,076	41,638	44,181
24	23,337	28,241	33,196	36,415	39,364	42,980	45,559
25	24,337	29,339	34,382	37,652	40,646	44,314	46,928
26	25,336	30,435	35,563	38,885	41,923	45,642	48,290
27	26,336	31,528	36,741	40,113	43,195	46,963	49,645
28	27,336	32,620	37,916	41,337	44,461	48,278	50,993
29	28,336	33,711	39,087	42,557	45,722	49,588	52,336
30	29,336	34,800	40,256	43,773	46,979	50,892	53,672
40	39,335	45,616	51,805	55,758	59,342	63,691	66,766
50	49,335	56,334	63,167	67,505	71,420	76,154	79,490
60	59,335	66,981	74,397	79,082	83,298	88,379	91,952
120	119,334	130,055	140,233	146,567	152,211	158,950	163,648
240	239,334	254,392	268,471	277,138	284,802	293,888	300,182
480	479,334	500,519	520,111	532,075	542,599	555,006	563,561
960	959,333	989,180	1016,566	1033,193	1047,760	1064,867	1076,621

Tabela 9.5: Quantis superiores da distribuição de F ($F_{0,10}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador, valor da probabilidade (α) de 10% de acordo com o seguinte evento: $P(F > F_{0,10}) = 0,10$.

	ν_1										
ν_2	1	2	3	4	5	6	7	8	9	10	11
1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,20	60,47
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,40
3	5,54	5,46	5,39	5,34	5,31	5,28	5,26	5,25	5,24	5,23	5,22
4	4,55	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,91
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30	3,28
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	2,92
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70	2,68
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	2,52
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,40
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,30
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25	2,23
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19	2,17
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14	2,12
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10	2,07
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06	2,04
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03	2,01
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00	1,98
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98	1,95
19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96	1,93
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94	1,91
21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	1,92	1,90
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90	1,88
23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	1,89	1,87
24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88	1,85
25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87	1,84
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86	1,83
27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87	1,85	1,82
28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84	1,81
29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83	1,80
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82	1,79
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76	1,74
50	2,81	2,41	2,20	2,06	1,97	1,90	1,84	1,80	1,76	1,73	1,70
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71	1,68
120	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68	1,65	1,63
240	2,73	2,32	2,11	1,97	1,87	1,80	1,74	1,70	1,66	1,63	1,60
480	2,72	2,31	2,10	1,96	1,86	1,79	1,73	1,68	1,64	1,61	1,58
960	2,71	2,31	2,09	1,95	1,85	1,78	1,72	1,68	1,64	1,61	1,58
∞	2,71	2,30	2,08	1,94	1,85	1,77	1,72	1,67	1,63	1,60	1,57

Tabela 9.6: Quantis superiores da distribuição de F ($F_{0,10}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador, valor da probabilidade (α) de 10% de acordo com o seguinte evento: $P(F > F_{0,10}) = 0,10$.

	ν_1										
ν_2	12	13	14	15	20	30	40	60	120	240	∞
1	60,71	60,90	61,07	61,22	61,74	62,26	62,53	62,79	63,06	63,19	63,33
2	9,41	9,41	9,42	9,42	9,44	9,46	9,47	9,47	9,48	9,49	9,49
3	5,21	5,21	5,20	5,20	5,18	5,16	5,15	5,14	5,13	5,11	5,13
4	3,90	3,89	3,88	3,87	3,84	3,82	3,80	3,79	3,78	3,77	3,76
5	3,27	3,26	3,25	3,24	3,21	3,17	3,16	3,14	3,12	3,12	3,10
6	2,90	2,89	2,88	2,87	2,84	2,80	2,78	2,76	2,74	2,73	2,72
7	2,67	2,65	2,64	2,63	2,59	2,56	2,54	2,51	2,49	2,48	2,47
8	2,50	2,49	2,48	2,46	2,42	2,38	2,36	2,34	2,32	2,30	2,29
9	2,38	2,36	2,35	2,34	2,30	2,25	2,23	2,21	2,18	2,17	2,16
10	2,28	2,27	2,26	2,24	2,20	2,16	2,13	2,11	2,08	2,07	2,06
11	2,21	2,19	2,18	2,17	2,12	2,08	2,05	2,03	2,00	1,99	1,97
12	2,15	2,13	2,12	2,10	2,06	2,01	1,99	1,96	1,93	1,92	1,90
13	2,10	2,08	2,07	2,05	2,01	1,96	1,93	1,90	1,88	1,86	1,85
14	2,05	2,04	2,02	2,01	1,96	1,91	1,89	1,86	1,83	1,81	1,80
15	2,02	2,00	1,99	1,97	1,92	1,87	1,85	1,82	1,79	1,77	1,76
16	1,99	1,97	1,95	1,94	1,89	1,84	1,81	1,78	1,75	1,73	1,72
17	1,96	1,94	1,93	1,91	1,86	1,81	1,78	1,75	1,72	1,70	1,69
18	1,93	1,92	1,90	1,89	1,84	1,78	1,75	1,72	1,69	1,67	1,66
19	1,91	1,89	1,88	1,86	1,81	1,76	1,73	1,70	1,67	1,65	1,63
20	1,89	1,87	1,86	1,84	1,79	1,74	1,71	1,68	1,64	1,63	1,61
21	1,87	1,86	1,84	1,83	1,78	1,72	1,69	1,66	1,62	1,60	1,59
22	1,86	1,84	1,83	1,81	1,76	1,70	1,67	1,64	1,60	1,59	1,57
23	1,84	1,83	1,81	1,80	1,74	1,69	1,66	1,62	1,59	1,57	1,55
24	1,83	1,81	1,80	1,78	1,73	1,67	1,64	1,61	1,57	1,55	1,53
25	1,82	1,80	1,79	1,77	1,72	1,66	1,63	1,59	1,56	1,54	1,52
26	1,81	1,79	1,77	1,76	1,71	1,65	1,61	1,58	1,54	1,52	1,50
27	1,80	1,78	1,76	1,75	1,70	1,64	1,60	1,57	1,53	1,51	1,49
28	1,79	1,77	1,75	1,74	1,69	1,63	1,59	1,56	1,52	1,50	1,48
29	1,78	1,76	1,75	1,73	1,68	1,62	1,58	1,55	1,51	1,49	1,47
30	1,77	1,75	1,74	1,72	1,67	1,61	1,57	1,54	1,50	1,48	1,46
40	1,71	1,70	1,68	1,66	1,61	1,54	1,51	1,47	1,42	1,40	1,38
50	1,68	1,66	1,64	1,63	1,57	1,50	1,46	1,42	1,38	1,35	1,33
60	1,66	1,64	1,62	1,60	1,54	1,48	1,44	1,40	1,35	1,32	1,29
120	1,60	1,58	1,56	1,55	1,48	1,41	1,37	1,32	1,26	1,23	1,19
240	1,57	1,55	1,53	1,52	1,45	1,38	1,33	1,28	1,22	1,18	1,13
480	1,56	1,54	1,52	1,50	1,44	1,36	1,31	1,26	1,19	1,15	1,09
960	1,55	1,53	1,51	1,49	1,43	1,35	1,30	1,25	1,18	1,14	1,06
∞	1,55	1,52	1,50	1,49	1,42	1,34	1,30	1,24	1,17	1,12	1,00

Tabela 9.7: Quantis superiores da distribuição de F ($F_{0,05}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador valor da probabilidade (α) de 5% de acordo com o seguinte evento: $P(F > F_{0,05}) = 0,05$.

	ν_1										
ν_2	1	2	3	4	5	6	7	8	9	10	11
1	161,45	199,50	215,70	224,58	230,16	234,0	236,8	238,9	240,5	241,9	242,98
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,40
3	10,13	9,55	9,27	9,11	9,01	8,94	8,88	8,84	8,81	8,78	8,76
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,24
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,22
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20
26	4,23	3,37	2,97	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,17
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,14
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,87
240	3,88	3,03	2,64	2,41	2,25	2,14	2,05	1,98	1,92	1,87	1,83
480	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,81
960	3,85	3,01	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84	1,80
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,79

Tabela 9.8: Quantis superiores da distribuição de F ($F_{0,05}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador, valor da probabilidade (α) de 5% de acordo com o seguinte evento: $P(F > F_{0,05}) = 0,05$.

ν_2	ν_1										
	12	13	14	15	20	30	40	60	120	240	∞
1	243,91	244,69	245,36	245,95	248,0	250,1	251,1	252,2	253,3	253,8	254,31
2	19,41	19,42	19,42	19,43	19,45	19,46	19,47	19,48	19,49	19,49	19,50
3	8,74	8,72	8,71	8,69	8,65	8,60	8,57	8,54	8,49	8,42	8,53
4	5,91	5,89	5,87	5,86	5,80	5,75	5,72	5,69	5,66	5,64	5,63
5	4,68	4,66	4,64	4,62	4,56	4,50	4,46	4,43	4,40	4,39	4,36
6	4,00	3,98	3,96	3,94	3,87	3,81	3,77	3,74	3,70	3,69	3,67
7	3,57	3,55	3,53	3,51	3,44	3,38	3,34	3,30	3,27	3,25	3,23
8	3,28	3,26	3,24	3,22	3,15	3,08	3,04	3,01	2,97	2,95	2,93
9	3,07	3,05	3,03	3,01	2,94	2,86	2,83	2,79	2,75	2,73	2,71
10	2,91	2,89	2,86	2,85	2,77	2,70	2,66	2,62	2,58	2,56	2,54
11	2,79	2,76	2,74	2,72	2,65	2,57	2,53	2,49	2,45	2,43	2,40
12	2,69	2,66	2,64	2,62	2,54	2,47	2,43	2,38	2,34	2,32	2,30
13	2,60	2,58	2,55	2,53	2,46	2,38	2,34	2,30	2,25	2,23	2,21
14	2,53	2,51	2,48	2,46	2,39	2,31	2,27	2,22	2,18	2,15	2,13
15	2,48	2,45	2,42	2,40	2,33	2,25	2,20	2,16	2,11	2,09	2,07
16	2,42	2,40	2,37	2,35	2,28	2,19	2,15	2,11	2,06	2,03	2,01
17	2,38	2,35	2,33	2,31	2,23	2,15	2,10	2,06	2,01	1,99	1,96
18	2,34	2,31	2,29	2,27	2,19	2,11	2,06	2,02	1,97	1,94	1,92
19	2,31	2,28	2,26	2,23	2,16	2,07	2,03	1,98	1,93	1,90	1,88
20	2,28	2,25	2,22	2,20	2,12	2,04	1,99	1,95	1,90	1,87	1,84
21	2,25	2,22	2,20	2,18	2,10	2,01	1,96	1,92	1,87	1,84	1,81
22	2,23	2,20	2,17	2,15	2,07	1,98	1,94	1,89	1,84	1,81	1,78
23	2,20	2,18	2,15	2,13	2,05	1,96	1,91	1,86	1,81	1,79	1,76
24	2,18	2,15	2,13	2,11	2,03	1,94	1,89	1,84	1,79	1,76	1,73
25	2,16	2,14	2,11	2,09	2,01	1,92	1,87	1,82	1,77	1,74	1,71
26	2,15	2,12	2,09	2,07	1,99	1,90	1,85	1,80	1,75	1,72	1,69
27	2,13	2,10	2,08	2,06	1,97	1,88	1,84	1,79	1,73	1,70	1,67
28	2,12	2,09	2,06	2,04	1,96	1,87	1,82	1,77	1,71	1,68	1,65
29	2,10	2,08	2,05	2,03	1,94	1,85	1,81	1,75	1,70	1,67	1,64
30	2,09	2,06	2,04	2,01	1,93	1,84	1,79	1,74	1,68	1,65	1,62
40	2,00	1,97	1,95	1,92	1,84	1,74	1,69	1,64	1,58	1,54	1,51
50	1,95	1,92	1,89	1,87	1,78	1,69	1,63	1,58	1,51	1,48	1,44
60	1,92	1,89	1,86	1,84	1,75	1,65	1,59	1,53	1,47	1,43	1,39
120	1,83	1,80	1,78	1,75	1,66	1,55	1,50	1,43	1,35	1,31	1,25
240	1,79	1,76	1,73	1,71	1,61	1,51	1,44	1,37	1,29	1,24	1,17
480	1,77	1,74	1,71	1,69	1,59	1,48	1,42	1,35	1,26	1,20	1,12
960	1,76	1,73	1,70	1,68	1,58	1,47	1,41	1,33	1,24	1,18	1,08
∞	1,75	1,72	1,69	1,67	1,57	1,46	1,39	1,32	1,22	1,15	1,00

Tabela 9.9: Quantis superiores da distribuição de F ($F_{0,025}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador valor da probabilidade (α) de 2,5% de acordo com o seguinte evento: $P(F > F_{0,025}) = 0,025$.

	ν_1										
ν_2	1	2	3	4	5	6	7	8	9	10	11
1	647,79	799,50	864,14	899,58	921,85	937,1	948,2	956,7	963,3	968,6	973,03
2	38,51	39,00	39,15	39,25	39,30	39,33	39,36	39,37	39,39	39,40	39,41
3	17,44	16,03	15,42	15,08	14,87	14,71	14,60	14,51	14,44	14,39	14,34
4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84	8,79
5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,57
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46	5,41
7	8,07	6,54	5,89	5,52	5,29	5,12	5,00	4,90	4,82	4,76	4,71
8	7,57	6,06	5,41	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,24
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,91
10	6,94	5,46	4,82	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,66
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53	3,47
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37	3,32
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25	3,20
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21	3,15	3,09
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	3,01
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	2,99	2,93
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	2,92	2,87
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93	2,87	2,81
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	2,82	2,76
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,72
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,80	2,73	2,68
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76	2,70	2,65
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73	2,67	2,62
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64	2,59
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	2,61	2,56
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,65	2,59	2,54
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,63	2,57	2,51
28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,61	2,55	2,49
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,59	2,53	2,48
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,46
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,33
50	5,34	3,97	3,39	3,05	2,83	2,67	2,55	2,46	2,38	2,32	2,26
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27	2,22
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22	2,16	2,10
240	5,09	3,75	3,17	2,84	2,62	2,46	2,34	2,25	2,17	2,10	2,05
480	5,06	3,72	3,14	2,81	2,59	2,43	2,31	2,22	2,14	2,08	2,02
960	5,04	3,70	3,13	2,80	2,58	2,42	2,30	2,21	2,13	2,06	2,01
∞	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05	1,99

Tabela 9.10: Quantis superiores da distribuição de F ($F_{0,025}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador valor da probabilidade (α) de 2,5% de acordo com o seguinte evento: $P(F > F_{0,025}) = 0,025$.

	ν_1										
ν_2	12	13	14	15	20	30	40	60	120	240	∞
1	976,71	979,84	982,53	984,87	993,1	1001	1006	1010	1014	1016	1018
2	39,41	39,42	39,43	39,43	39,45	39,46	39,47	39,48	39,49	39,49	39,50
3	14,30	14,27	14,24	14,21	14,11	14,00	13,94	13,85	13,68	13,43	13,90
4	8,75	8,71	8,68	8,66	8,56	8,46	8,41	8,36	8,31	8,28	8,26
5	6,53	6,49	6,45	6,43	6,33	6,22	6,17	6,12	6,06	6,03	6,02
6	5,37	5,33	5,30	5,27	5,17	5,07	5,01	4,96	4,90	4,88	4,85
7	4,67	4,63	4,60	4,57	4,47	4,36	4,31	4,26	4,20	4,17	4,14
8	4,20	4,16	4,13	4,10	4,00	3,89	3,84	3,78	3,73	3,70	3,67
9	3,87	3,83	3,80	3,77	3,67	3,56	3,51	3,45	3,39	3,36	3,33
10	3,62	3,58	3,55	3,52	3,42	3,31	3,26	3,20	3,14	3,11	3,08
11	3,43	3,39	3,36	3,33	3,23	3,12	3,06	3,00	2,94	2,91	2,88
12	3,28	3,24	3,21	3,18	3,07	2,96	2,91	2,85	2,79	2,76	2,72
13	3,15	3,12	3,08	3,05	2,95	2,84	2,78	2,72	2,66	2,63	2,60
14	3,05	3,01	2,98	2,95	2,84	2,73	2,67	2,61	2,55	2,52	2,49
15	2,96	2,92	2,89	2,86	2,76	2,64	2,59	2,52	2,46	2,43	2,40
16	2,89	2,85	2,82	2,79	2,68	2,57	2,51	2,45	2,38	2,35	2,32
17	2,82	2,79	2,75	2,72	2,62	2,50	2,44	2,38	2,32	2,28	2,25
18	2,77	2,73	2,70	2,67	2,56	2,44	2,38	2,32	2,26	2,22	2,19
19	2,72	2,68	2,65	2,62	2,51	2,39	2,33	2,27	2,20	2,17	2,13
20	2,68	2,64	2,60	2,57	2,46	2,35	2,29	2,22	2,16	2,12	2,09
21	2,64	2,60	2,56	2,53	2,42	2,31	2,25	2,18	2,11	2,08	2,04
22	2,60	2,56	2,53	2,50	2,39	2,27	2,21	2,14	2,08	2,04	2,00
23	2,57	2,53	2,50	2,47	2,36	2,24	2,18	2,11	2,04	2,01	1,97
24	2,54	2,50	2,47	2,44	2,33	2,21	2,15	2,08	2,01	1,97	1,94
25	2,51	2,48	2,44	2,41	2,30	2,18	2,12	2,05	1,98	1,94	1,91
26	2,49	2,45	2,42	2,39	2,28	2,16	2,09	2,03	1,95	1,92	1,88
27	2,47	2,43	2,39	2,36	2,25	2,13	2,07	2,00	1,93	1,89	1,85
28	2,45	2,41	2,37	2,34	2,23	2,11	2,05	1,98	1,91	1,87	1,83
29	2,43	2,39	2,36	2,32	2,21	2,09	2,03	1,96	1,89	1,85	1,81
30	2,41	2,37	2,34	2,31	2,20	2,07	2,01	1,94	1,87	1,83	1,79
40	2,29	2,25	2,21	2,18	2,07	1,94	1,88	1,80	1,72	1,68	1,64
50	2,22	2,18	2,14	2,11	1,99	1,87	1,80	1,72	1,64	1,59	1,55
60	2,17	2,13	2,09	2,06	1,94	1,82	1,74	1,67	1,58	1,53	1,48
120	2,05	2,01	1,98	1,94	1,82	1,69	1,61	1,53	1,43	1,38	1,31
240	2,00	1,96	1,92	1,89	1,77	1,63	1,55	1,46	1,35	1,29	1,21
480	1,97	1,93	1,89	1,86	1,74	1,60	1,52	1,42	1,31	1,24	1,14
960	1,96	1,92	1,88	1,85	1,72	1,58	1,50	1,41	1,29	1,21	1,10
∞	1,94	1,90	1,87	1,83	1,71	1,57	1,48	1,39	1,27	1,19	1,00

Tabela 9.11: Quantis superiores da distribuição de F ($F_{0,01}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador valor da probabilidade (α) de 1% de acordo com o seguinte evento: $P(F > F_{0,01}) = 0,01$.

	ν_1										
ν_2	1	2	3	4	5	6	7	8	9	10	11
1	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056	6083
2	98,50	99,01	99,05	99,24	99,30	99,33	99,36	99,37	99,39	99,40	99,41
3	34,12	30,74	29,34	28,60	28,11	27,77	27,52	27,32	27,16	27,03	26,92
4	21,20	18,00	16,68	15,98	15,52	15,21	14,97	14,80	14,66	14,55	14,45
5	16,26	13,28	12,05	11,39	10,96	10,67	10,45	10,29	10,15	10,05	9,96
6	13,75	10,92	9,77	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,77
11	9,65	7,21	6,21	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,62
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,43
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,29
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,24
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,09
25	7,77	5,57	4,67	4,18	3,85	3,63	3,46	3,32	3,22	3,13	3,06
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	3,02
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,99
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,96
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,93
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,91
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,73
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70	2,63
60	7,08	4,98	4,12	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,40
240	6,74	4,69	3,86	3,40	3,09	2,88	2,71	2,59	2,48	2,40	2,32
480	6,69	4,65	3,82	3,36	3,06	2,84	2,68	2,55	2,44	2,36	2,28
960	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34	2,27
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,25

Tabela 9.12: Quantis superiores da distribuição de F ($F_{0,01}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador valor da probabilidade (α) de 1% de acordo com o seguinte evento: $P(F > F_{0,01}) = 0,01$.

	ν_1										
ν_2	12	13	14	15	20	30	40	60	120	240	∞
1	6106	6126	6143	6157	6209	6261	6287	6313	6339	6353	6366
2	99,42	99,42	99,43	99,43	99,45	99,47	99,47	99,48	99,49	99,49	99,50
3	26,82	26,74	26,67	26,60	26,34	26,02	25,79	25,43	24,62	23,38	26,13
4	14,37	14,31	14,25	14,20	14,02	13,84	13,74	13,65	13,56	13,51	13,46
5	9,88	9,82	9,77	9,72	9,55	9,37	9,28	9,19	9,08	9,01	9,02
6	7,72	7,66	7,60	7,56	7,40	7,23	7,14	7,06	6,97	6,92	6,88
7	6,47	6,41	6,36	6,31	6,16	5,99	5,91	5,83	5,74	5,70	5,65
8	5,67	5,61	5,56	5,52	5,36	5,20	5,12	5,03	4,95	4,90	4,86
9	5,11	5,05	5,01	4,96	4,81	4,65	4,57	4,48	4,40	4,36	4,31
10	4,71	4,65	4,60	4,56	4,41	4,25	4,17	4,08	4,00	3,95	3,91
11	4,40	4,34	4,29	4,25	4,10	3,94	3,86	3,78	3,69	3,65	3,60
12	4,16	4,10	4,05	4,01	3,86	3,70	3,62	3,54	3,45	3,41	3,36
13	3,96	3,91	3,86	3,82	3,66	3,51	3,43	3,34	3,25	3,21	3,17
14	3,80	3,75	3,70	3,66	3,51	3,35	3,27	3,18	3,09	3,05	3,00
15	3,67	3,61	3,56	3,52	3,37	3,21	3,13	3,05	2,96	2,91	2,87
16	3,55	3,50	3,45	3,41	3,26	3,10	3,02	2,93	2,84	2,80	2,75
17	3,46	3,40	3,35	3,31	3,16	3,00	2,92	2,83	2,75	2,70	2,65
18	3,37	3,32	3,27	3,23	3,08	2,92	2,84	2,75	2,66	2,61	2,57
19	3,30	3,24	3,19	3,15	3,00	2,84	2,76	2,67	2,58	2,54	2,49
20	3,23	3,18	3,13	3,09	2,94	2,78	2,69	2,61	2,52	2,47	2,42
21	3,17	3,12	3,07	3,03	2,88	2,72	2,64	2,55	2,46	2,41	2,36
22	3,12	3,07	3,02	2,98	2,83	2,67	2,58	2,50	2,40	2,35	2,31
23	3,07	3,02	2,97	2,93	2,78	2,62	2,54	2,45	2,35	2,31	2,26
24	3,03	2,98	2,93	2,89	2,74	2,58	2,49	2,40	2,31	2,26	2,21
25	2,99	2,94	2,89	2,85	2,70	2,54	2,45	2,36	2,27	2,22	2,17
26	2,96	2,90	2,86	2,81	2,66	2,50	2,42	2,33	2,23	2,18	2,13
27	2,93	2,87	2,82	2,78	2,63	2,47	2,38	2,29	2,20	2,15	2,10
28	2,90	2,84	2,79	2,75	2,60	2,44	2,35	2,26	2,17	2,12	2,06
29	2,87	2,81	2,77	2,73	2,57	2,41	2,33	2,23	2,14	2,09	2,03
30	2,84	2,79	2,74	2,70	2,55	2,39	2,30	2,21	2,11	2,06	2,01
40	2,66	2,61	2,56	2,52	2,37	2,20	2,11	2,02	1,92	1,86	1,80
50	2,56	2,51	2,46	2,42	2,27	2,10	2,01	1,91	1,80	1,74	1,68
60	2,50	2,44	2,39	2,35	2,20	2,03	1,94	1,84	1,73	1,67	1,60
120	2,34	2,28	2,23	2,19	2,03	1,86	1,76	1,66	1,53	1,46	1,38
240	2,26	2,20	2,16	2,11	1,96	1,78	1,68	1,57	1,43	1,35	1,25
480	2,22	2,17	2,12	2,08	1,92	1,74	1,63	1,52	1,38	1,29	1,17
960	2,20	2,15	2,10	2,06	1,90	1,72	1,61	1,50	1,35	1,26	1,11
∞	2,18	2,13	2,08	2,04	1,88	1,70	1,59	1,47	1,32	1,22	1,00

Tabela 9.13: Quantis superiores da distribuição de F ($F_{0,005}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador valor da probabilidade (α) de 0,5% de acordo com o seguinte evento: $P(F > F_{0,005}) = 0,005$.

	ν_1										
ν_2	1	2	3	4	5	6	7	8	9	10	11
1	16211	20000	21614	22500	23056	23437	23715	23925	24091	24224	24334
2	198,50	199,04	198,70	199,21	199,29	199,3	199,4	199,4	199,4	199,4	199,41
3	55,55	49,49	47,03	45,76	44,90	44,29	43,83	43,46	43,16	42,91	42,68
4	31,33	26,28	24,23	23,15	22,45	21,97	21,62	21,35	21,14	20,96	20,82
5	22,78	18,31	16,51	15,55	14,93	14,51	14,19	13,95	13,76	13,61	13,48
6	18,64	14,54	12,90	12,03	11,46	11,07	10,79	10,57	10,39	10,25	10,13
7	16,24	12,40	10,87	10,05	9,52	9,16	8,89	8,68	8,51	8,38	8,27
8	14,69	11,04	9,59	8,80	8,30	7,95	7,69	7,50	7,34	7,21	7,10
9	13,61	10,11	8,71	7,96	7,47	7,13	6,89	6,69	6,54	6,42	6,31
10	12,83	9,43	8,08	7,34	6,87	6,54	6,30	6,12	5,97	5,85	5,75
11	12,23	8,91	7,60	6,88	6,42	6,10	5,86	5,68	5,54	5,42	5,32
12	11,75	8,51	7,22	6,52	6,07	5,76	5,52	5,35	5,20	5,09	4,99
13	11,37	8,19	6,92	6,23	5,79	5,48	5,25	5,08	4,94	4,82	4,72
14	11,06	7,92	6,68	6,00	5,56	5,26	5,03	4,86	4,72	4,60	4,51
15	10,80	7,70	6,47	5,80	5,37	5,07	4,85	4,67	4,54	4,42	4,33
16	10,58	7,51	6,30	5,64	5,21	4,91	4,69	4,52	4,38	4,27	4,18
17	10,38	7,35	6,15	5,50	5,07	4,78	4,56	4,39	4,25	4,14	4,05
18	10,22	7,21	6,02	5,37	4,96	4,66	4,44	4,28	4,14	4,03	3,94
19	10,07	7,09	5,91	5,27	4,85	4,56	4,34	4,18	4,04	3,93	3,84
20	9,94	6,99	5,81	5,17	4,76	4,47	4,26	4,09	3,96	3,85	3,76
21	9,83	6,89	5,73	5,09	4,68	4,39	4,18	4,01	3,88	3,77	3,68
22	9,73	6,81	5,65	5,02	4,61	4,32	4,11	3,94	3,81	3,70	3,61
23	9,63	6,73	5,58	4,95	4,54	4,26	4,05	3,88	3,75	3,64	3,55
24	9,55	6,66	5,52	4,89	4,49	4,20	3,99	3,83	3,69	3,59	3,50
25	9,48	6,60	5,46	4,83	4,43	4,15	3,94	3,78	3,64	3,54	3,45
26	9,41	6,54	5,41	4,79	4,38	4,10	3,89	3,73	3,60	3,49	3,40
27	9,34	6,49	5,36	4,74	4,34	4,06	3,85	3,69	3,56	3,45	3,36
28	9,28	6,44	5,31	4,70	4,30	4,02	3,81	3,65	3,52	3,41	3,32
29	9,23	6,40	5,27	4,66	4,26	3,98	3,77	3,61	3,48	3,38	3,29
30	9,18	6,35	5,24	4,62	4,23	3,95	3,74	3,58	3,45	3,34	3,25
40	8,83	6,07	4,97	4,37	3,99	3,71	3,51	3,35	3,22	3,12	3,03
50	8,63	5,90	4,82	4,23	3,85	3,58	3,38	3,22	3,09	2,99	2,90
60	8,49	5,79	4,73	4,14	3,76	3,49	3,29	3,13	3,01	2,90	2,82
120	8,18	5,54	4,50	3,92	3,55	3,28	3,09	2,93	2,81	2,71	2,62
240	8,03	5,42	4,38	3,82	3,45	3,19	2,99	2,84	2,71	2,61	2,52
480	7,95	5,36	4,33	3,77	3,40	3,14	2,94	2,79	2,67	2,56	2,48
960	7,92	5,33	4,30	3,74	3,37	3,11	2,92	2,77	2,64	2,54	2,46
∞	7,88	5,30	4,28	3,72	3,35	3,09	2,90	2,74	2,62	2,52	2,43

Tabela 9.14: Quantis superiores da distribuição de F ($F_{0,005}$) com ν_1 graus de liberdade do numerador e ν_2 graus de liberdade do denominador valor da probabilidade (α) de 0,5% de acordo com o seguinte evento: $P(F > F_{0,005}) = 0,005$.

	ν_1										
ν_2	12	13	14	15	20	30	40	60	120	240	∞
1	24426	24505	24572	24630	24836	25044	25148	25253	25359	25411	25464
2	199,42	199,42	199,43	199,43	199,4	199,5	199,5	199,5	199,5	199,5	199,50
3	42,49	42,32	42,16	42,02	41,43	40,59	39,93	38,85	36,37	32,90	41,83
4	20,70	20,60	20,51	20,44	20,16	19,89	19,75	19,61	19,47	19,39	19,32
5	13,37	13,28	13,20	13,13	12,89	12,64	12,50	12,36	12,20	12,07	12,14
6	10,03	9,95	9,88	9,81	9,59	9,36	9,24	9,12	9,00	8,94	8,88
7	8,18	8,10	8,03	7,97	7,76	7,54	7,42	7,31	7,20	7,15	7,08
8	7,01	6,94	6,87	6,81	6,61	6,40	6,29	6,18	6,06	6,01	5,95
9	6,23	6,15	6,09	6,03	5,83	5,63	5,52	5,41	5,30	5,25	5,19
10	5,66	5,59	5,53	5,47	5,27	5,07	4,97	4,86	4,75	4,69	4,64
11	5,24	5,16	5,10	5,05	4,86	4,65	4,55	4,45	4,34	4,28	4,23
12	4,91	4,84	4,77	4,72	4,53	4,33	4,23	4,12	4,01	3,96	3,90
13	4,64	4,57	4,51	4,46	4,27	4,07	3,97	3,87	3,76	3,70	3,65
14	4,43	4,36	4,30	4,25	4,06	3,86	3,76	3,66	3,55	3,49	3,44
15	4,25	4,18	4,12	4,07	3,88	3,69	3,59	3,48	3,37	3,32	3,26
16	4,10	4,03	3,97	3,92	3,73	3,54	3,44	3,33	3,22	3,17	3,11
17	3,97	3,90	3,84	3,79	3,61	3,41	3,31	3,21	3,10	3,04	2,98
18	3,86	3,79	3,73	3,68	3,50	3,30	3,20	3,10	2,99	2,93	2,87
19	3,76	3,70	3,64	3,59	3,40	3,21	3,11	3,00	2,89	2,83	2,78
20	3,68	3,61	3,55	3,50	3,32	3,12	3,02	2,92	2,81	2,75	2,69
21	3,60	3,54	3,48	3,43	3,24	3,05	2,95	2,84	2,73	2,67	2,61
22	3,54	3,47	3,41	3,36	3,18	2,98	2,88	2,77	2,66	2,60	2,55
23	3,47	3,41	3,35	3,30	3,12	2,92	2,82	2,71	2,60	2,54	2,48
24	3,42	3,35	3,30	3,25	3,06	2,87	2,77	2,66	2,55	2,49	2,43
25	3,37	3,30	3,25	3,20	3,01	2,82	2,72	2,61	2,50	2,44	2,38
26	3,33	3,26	3,20	3,15	2,97	2,77	2,67	2,56	2,45	2,39	2,33
27	3,28	3,22	3,16	3,11	2,93	2,73	2,63	2,52	2,41	2,35	2,29
28	3,25	3,18	3,12	3,07	2,89	2,69	2,59	2,48	2,37	2,31	2,25
29	3,21	3,15	3,09	3,04	2,86	2,66	2,56	2,45	2,33	2,27	2,21
30	3,18	3,11	3,06	3,01	2,82	2,63	2,52	2,42	2,30	2,24	2,18
40	2,95	2,89	2,83	2,78	2,60	2,40	2,30	2,18	2,06	2,00	1,93
50	2,82	2,76	2,70	2,65	2,47	2,27	2,16	2,05	1,93	1,86	1,79
60	2,74	2,68	2,62	2,57	2,39	2,19	2,08	1,96	1,83	1,76	1,69
120	2,54	2,48	2,42	2,37	2,19	1,98	1,87	1,75	1,61	1,52	1,43
240	2,45	2,39	2,33	2,28	2,09	1,89	1,77	1,64	1,49	1,40	1,28
480	2,40	2,34	2,28	2,23	2,05	1,84	1,72	1,59	1,43	1,33	1,19
960	2,38	2,32	2,26	2,21	2,02	1,81	1,69	1,56	1,40	1,29	1,13
∞	2,36	2,29	2,24	2,19	2,00	1,79	1,67	1,53	1,36	1,25	1,00

Tabela 9.15: Quantis superiores da distribuição t de Student (t_α) com ν graus de liberdade e para diferentes valores da probabilidade (α) de acordo com o seguinte evento: $P(t > t_\alpha) = \alpha$.

ν	0,250	0,200	0,150	0,100	0,050	0,025	0,010	0,005	0,001
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,599
3	0,765	0,979	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,500	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,768
24	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660	3,460
120	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
240	0,676	0,843	1,039	1,285	1,651	1,970	2,342	2,596	3,332
480	0,675	0,842	1,038	1,283	1,648	1,965	2,334	2,586	3,311
960	0,675	0,842	1,037	1,282	1,646	1,962	2,330	2,581	3,301
1920	0,675	0,842	1,037	1,282	1,646	1,961	2,328	2,578	3,296
3840	0,675	0,842	1,037	1,282	1,645	1,961	2,327	2,577	3,293
∞	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291