

# Estatística & Probabilidade aplicado às Engenharias e Ciências

Ben Dêivide de Oliveira Batista

2021-05-07



# Contents

<b>1</b>	<b>Definições Gerais da Estatística e técnicas de somatórios</b>	<b>15</b>
1.1	Introdução . . . . .	15
1.2	Definições gerais da Estatística . . . . .	16
1.2.1	Estatística na pesquisa científica . . . . .	21
1.3	Definições básicas . . . . .	21
1.4	Técnicas de somatório . . . . .	21
1.4.1	Normas de arredondamento . . . . .	21
<b>2</b>	<b>Coleta, organização e apresentação dos dados</b>	<b>23</b>
<b>3</b>	<b>Medidas de posição</b>	<b>25</b>
3.1	Introdução . . . . .	25
3.2	Média . . . . .	25
3.3	Mediana . . . . .	25
3.4	Moda . . . . .	25
<b>4</b>	<b>Medidas de dispersão</b>	<b>27</b>
4.1	Introdução . . . . .	27
4.2	Amplitude total . . . . .	27
4.3	Variância . . . . .	27
4.4	Desvio padrão . . . . .	27
4.5	Coefficiente de Variação . . . . .	27
4.6	Erro padrão da média . . . . .	27

<b>5</b>	<b>Probabilidades</b>	<b>29</b>
5.1	Conceitos básicos . . . . .	29
5.1.1	Leis básicas da teoria de conjuntos . . . . .	29
5.1.2	Teoremas clássicos da teoria de conjuntos . . . . .	29
5.2	Definições de probabilidades . . . . .	29
5.3	Propriedades . . . . .	29
5.3.1	Regra de adição de probabilidades . . . . .	29
5.3.2	Probabilidade de um evento complementar . . . . .	29
5.4	Eventos independentes e probabilidade condicional . . . . .	29
5.5	Teorema de Bayes . . . . .	29
5.6	Variável Aleatória . . . . .	29
5.7	Distribuição de probabilidade discreta . . . . .	29
5.8	Distribuição de probabilidade contínua . . . . .	30
5.9	Função de distribuição acumulada (FDA) . . . . .	30
5.9.1	FDA para $X$ discreta . . . . .	30
5.9.2	FDA para $X$ contínua . . . . .	30
5.10	Esperança matemática e variância . . . . .	30
5.10.1	Esperança matemática . . . . .	30
5.10.2	Mediana . . . . .	30
5.10.3	Moda . . . . .	30
5.10.4	Variância e desvio padrão . . . . .	30
<b>6</b>	<b>Distribuições de probabilidade</b>	<b>31</b>
6.1	Introdução . . . . .	31
6.2	Distribuições discretas de probabilidades . . . . .	31
6.2.1	Distribuição Bernoulli . . . . .	31
6.2.2	Distribuição Binomial . . . . .	31
6.2.3	Distribuição Poisson . . . . .	31
6.2.4	Relação entre a distribuição Binomial e Poisson . . . . .	31
6.3	Distribuições contínuas de probabilidades . . . . .	31
6.3.1	Distribuição normal . . . . .	31

<b>7</b>	<b>Amostragem</b>	<b>33</b>
7.1	Introdução . . . . .	33
7.2	Amostragem não-probabilística e probabilística . . . . .	33
7.3	Técnicas de amostragem probabilística . . . . .	33
<b>8</b>	<b>Distribuição de amostragem</b>	<b>35</b>
8.1	Introdução . . . . .	35
8.2	Distribuição de amostragem da média . . . . .	35
8.3	Distribuição de amostragem de proporções . . . . .	35
8.4	Distribuição de amostragem de diferença entre médias . . . . .	35
8.5	Distribuições amostrais para variância . . . . .	35
<b>9</b>	<b>Teoria de estimação</b>	<b>37</b>
9.1	Introdução . . . . .	37
9.2	Conceitos básicos . . . . .	37
9.3	Tipos de estimador . . . . .	37
9.4	Propriedades de um estimador . . . . .	37
9.5	Estimação por ponto . . . . .	37
9.6	Estimação por intervalo . . . . .	37
9.6.1	Intervalo de confiança para a média . . . . .	37
9.6.2	Intervalo de confiança para a variância . . . . .	37
9.6.3	Intervalo de confiança de confiança para a diferença entre médias . . . . .	37
9.7	Dimensionamento de amostras . . . . .	37
<b>10</b>	<b>Teoria da decisão</b>	<b>39</b>
10.1	Introdução . . . . .	39
10.2	Teste de hipóteses . . . . .	39
10.3	Erros tipo I e II . . . . .	39
10.4	Teste unilateral e bilateral . . . . .	39
10.5	Passos para a construção de um teste de hipótese . . . . .	39
10.6	Testes de hipóteses para a média . . . . .	39
10.7	Testes de hipóteses para a proporção . . . . .	39

10.8 Testes de hipóteses para a variância . . . . .	39
10.9 Testes de hipóteses para a diferença entre duas médias . . . . .	39
<b>11 Correlação e regressão linear simples</b>	<b>41</b>
11.1 Introdução . . . . .	41
11.2 Correlação linear . . . . .	41
11.2.1 Coeficiente de correlação linear . . . . .	41
11.2.2 Teste de hipóteses acerca do coeficiente de correlação linear	41
11.3 Regressão linear simples . . . . .	41
11.3.1 Modelo . . . . .	41
11.3.2 Estimação dos parâmetros do modelo . . . . .	41
11.3.3 Teste de hipóteses para o modelo de regressão . . . . .	41
11.3.4 Medidas de adequação do modelo . . . . .	41

# List of Tables

1.1	Volume, em MMbbl/d, da Produção e consumo de petróleo do Brasil nos períodos de 2007 a 2017. . . . .	19
-----	--	----





# List of Figures

1.1	Produção e consumo de petróleo do Brasil nos períodos de 2007 a 2017. . . . .	18
1.2	[Ilustração/Vídeo animado sobre as disposições gerais da Estatística.]( <a href="https://youtu.be/mASLUwyaC5Q">https://youtu.be/mASLUwyaC5Q</a> ) . . . . .	21

<! -



# Bem-vindo

Esse é um *livro digital* da 1ª edição do “**Estatística & Probabilidade aplicado às Engenharias e Ciências**”, um livro com o selo **Democratizando Conhecimento (DC)**. O Livro é voltado para quem deseja iniciar no estudo sobre a estatística. Daremos as bases e fundamentos, de modo aplicado a problemas na área das Engenharias e Ciências, de assuntos desde o que é uma população, amostra, até estudos sobre a teoria de decisão, estudo de regressão, dentre outros assuntos básicos, para que assim, a partir desse material, o leitor tenha base para ler, assuntos mais aprofundados na área da estatística.

## Licença

Este trabalho está licenciado com uma Licença Creative Commons - Atribuição-NãoComercial 4.0 Internacional.



**Democratizando  
Conhecimento**



# Epígrafe

*O pulsar de minha existência está intensa e totalmente no presente da vida* (Ben Dêivide)



## Chapter 1

# Definições Gerais da Estatística e técnicas de somatórios

### 1.1 Introdução

Em pleno século XXI, passamos por um processo de transformação na era digital. Uma grande massa de informações surge instantaneamente a cada momento sobre os mais diversos temas possíveis. Por exemplo, nas redes sociais quando percebemos o número de curtidas de uma determinada declaração, número de downloads de um determinado vídeo, a repercussão que determinada declaração proporciona, o número de propagandas, etc, tudo isso cria um grande banco de dados sobre os usuários, que hoje se torna mais valiosa do que a própria moeda local. Isso é a nova revolução chamada “Big Data”. Por meio de grande banco de dados, podemos por exemplo, traçar um perfil dos usuários, como eles se comportam, quais as suas preferências, escolhas, diversão, etc. Contudo, o entendimento dessas informações podem não ser tão claras, ou devido a complexidade do problema, ou pela quantidade de informações recebidas rapidamente, ou outros fatores. Diversos outros exemplos poderiam ser citados, tudo isso para mostrar a necessidade de entender o que está por trás desses dados, cuja compreensão é o grande objetivo nessa era global.

Nesse enfoque, temos a Estatística como Ciência que fornece métodos para coleta, organização, descrição, análise e interpretação de dados (observacionais ou experimentais) e para a utilização dos mesmos na tomada de decisões. Os dados são informações retiradas de um conjunto de elementos de interesse. Podemos estar interessados na Produção anual de Gás Natural Não associado com o petróleo (GASN) de um determinado país, e ao longo dos anos, coletarmos

informações para ao final, por exemplo, termos informações que nos indique o potencial energético desse recurso natural nesse local ou consequências dessa fonte energética na economia do país.

Assim, por meio da utilização de técnicas estatísticas, tentamos entender as informações contidas nos dados. Devido a complexidade dessas informações em algumas situações, o estudo sobre essas técnicas têm aumentado, fazendo parte do nosso cotidiano. Nessa era digital, a grande quantidade de informações é gigantesca e valiosa, e as empresas tentam entender o que está por trás desses dados, ou você acha que o Facebook foi criado simplesmente para gerar entretenimento às pessoas? Ou você também acha que o Google criou uma plataforma de pesquisa simplesmente para facilitar a vida das pessoas? Algo muito nobre está por trás de tudo isso, os dados.

No passado, tratar uma grande massa de números era uma tarefa custosa e cansativa, que exigia horas de trabalho tedioso. Porém, hoje, esse volume de informações pode ser analisado rapidamente por meio de um computador pessoal e programas adequados. O computador contribui positivamente na difusão e uso dos métodos estatísticos. Você já se perguntou como é que lojas virtuais lhe oferta produtos sendo que nunca acessou aquele site antes? Já percebeu que o Netflix quando lhe oferece uma série a tela de entrada as vezes se altera? Tudo isso é fruto das técnicas de máquinas de aprendizagem (do inglês, *Machine Learning*), uma área da inteligência artificial. Juntamente com a estatística, essas ferramentas estão presentes em várias das tecnologias que utilizamos hoje.

## 1.2 Definições gerais da Estatística

Inicialmente, podemos dividir a Estatística em três ramos:

- Estatística Descritiva;
- Probabilidade;
- Estatística Inferencial.

Definimos cada uma dessas áreas a seguir.

Um conjunto de técnicas estatísticas destinadas a coleta, descrição e síntese dos dados, a fim de podermos entender características de interesse da população<sup>1</sup>, é chamada de Estatística Descritiva ou Estatística Dedutiva.

As técnicas mencionadas na Definição 1.2 são: coleta, organização, tabulação, representação gráfica, bem como medidas que sintetizam todas as informações contidas nos dados.

As quatro primeiras técnicas serão abordadas no Capítulo 2. As medidas serão estudadas nos Capítulos 3 e 4. Essa fase é de grande relevância, pois com base

---

<sup>1</sup>Ver Definição 1.6



na Estatística Descritiva podemos sintetizar as informações contidas nos dados, e torná-las mais compreensíveis que de outro modo seriam complexas de serem entendidas.

No Capítulo 7, daremos uma maior ênfase sobre a definição de uma população. De modo simples, podemos definir como um conjunto de elementos com uma característica em comum. Uma vez definida a característica que delimita essa população (Ver Subseção 1.2.1), e também a característica de interesse (chamada de variável) da pesquisa, faremos a coleta dos valores observados da variável em cada elemento da população ou de um subconjunto (amostra). Os valores observados são chamados de dados.

A característica de interesse observada em cada elemento da população é definida como valor observado ou dado.

Todas as técnicas mencionadas anteriormente auxiliam na descrição dos dados, uma não necessariamente sobrepõe a outra. Vejamos, a [British Petroleum Company \(2018\)](#) lançou o relatório técnico de 2018 sobre os diversos tipos de produção de energias dos países, e na Figura 1.1 é mostrado um gráfico que sintetiza a produção e o consumo de petróleo do Brasil, em milhões de barris por dia (MMbbl/d), nos períodos de 2007 a 2017.

```
# BP Statistical Review 2018

# Producao e consumo de petroleo no brasil (Milhares de barris por dia)
ano <- as.factor(c(2007:2017, 2007:2017))
prodcons <- c(1831, 1897, 2029, 2137, 2179, 2145, 2110, 2341, 2525, 2608, 2734, 2308, 2481, 2498,
id <- c(rep("Produção", 11), rep("Consumo", 11))

# Objeto que armazena as informações
dados <- data.frame(ano, Legenda = id, prodcons)

# Funcao para criacao do grafico de barras
ggplot(dados, aes(x=ano, y=prodcons, fill=Legenda)) +
  geom_bar(position="dodge", stat="identity") +
  xlab("Ano") + ylab("Petróleo (Milhões de barris por dia)")
```

O gráfico nos revela que o Brasil produz petróleo abaixo do que necessitaria para o consumo, de tal modo que a produção é 27,71% a menos do que o consumo. Isso explica o porquê do Brasil como grande produtor de petróleo, ainda assim, necessita importar essa fonte de energia. Contudo, o gráfico não apresenta um resumo perfeito. Por exemplo, mesmo a produção de petróleo sendo mais baixa do que o consumo, o coeficiente de variação (assunto abordado no Capítulo 4 dessas dos dados dessas variáveis, são 12,99% e 10,93%, respectivamente, calculados de acordo com a Tabela 1.1. Isso implica, que as variações da produção de petróleo são maiores do que as do consumo, no Brasil. Observe que essas

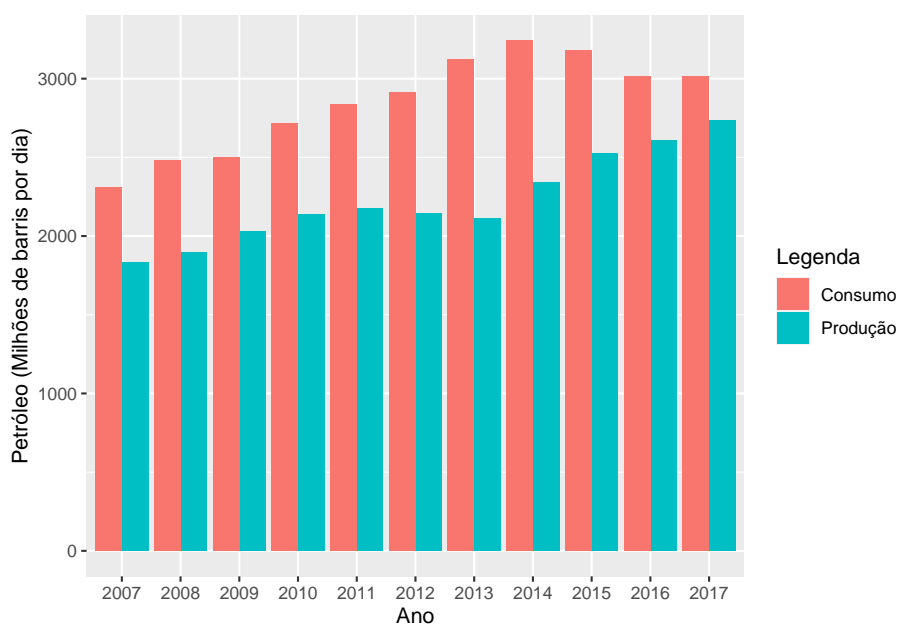


Figure 1.1: Produção e consumo de petróleo do Brasil nos períodos de 2007 a 2017.

últimas informações não podem ser vistas facilmente na Figura 1.1, mas juntamente com o auxílio das medidas numéricas (medidas de posição e dispersão) e as medidas gráficas, podemos complementar as informações, e assim, obter uma melhor descrição sobre essas informações.

Quando precisamos estender as informações contidas em um subconjunto (amostra) de dados para todo o conjunto (população), necessitamos de técnicas específicas dentro da estatística para garantir que estas informações sejam o mais verossímil possível. Técnicas estas são chamadas de Estatística Inferencial, definida a seguir.

O estudo de técnicas que visam estender (extrapolar) a informação contida na amostra à população, chamamos de Estatística Inferencial ou Estatística Indutiva.

As técnicas abordadas na Estatística Inferencial estão relacionadas a determinar características (parâmetros) populacionais desconhecidas, ou até mesmo fazer afirmações sobre esses parâmetros.

A determinação de parâmetros por meio de características amostrais (estimadores) que chamamos de Estimação será abordado no Capítulo 9. As afirmações realizadas sobre estes parâmetros, chamadas de hipóteses, serão estudadas no Capítulo 10.

Table 1.1: Volume, em MMbbl/d, da Produção e consumo de petróleo do Brasil nos períodos de 2007 a 2017.

Ano	Tipo	Volume
2007	Produção	1831
2008	Produção	1897
2009	Produção	2029
2010	Produção	2137
2011	Produção	2179
2012	Produção	2145
2013	Produção	2110
2014	Produção	2341
2015	Produção	2525
2016	Produção	2608
2017	Produção	2734
2007	Consumo	2308
2008	Consumo	2481
2009	Consumo	2498
2010	Consumo	2716
2011	Consumo	2839
2012	Consumo	2915
2013	Consumo	3124
2014	Consumo	3242
2015	Consumo	3181
2016	Consumo	3013
2017	Consumo	3017

A Definição 1.3 nos mostra que por meio da Estatística, podemos tomar decisões sobre uma população através da amostra. Isso se faz necessário muitas vezes em uma pesquisa, devido a duas coisas preciosas: tempo e dinheiro. Muito embora, se tivermos acesso a todos os elementos de uma população, não se faz necessário o uso de técnicas da inferência estatística, e daí realizamos o que chamamos de Censo.

A forma de como se obter uma amostra é um dos passos mais importante em todo o processo da análise, uma vez que não adianta está com todo o aparato técnico se as informações contidas nessa amostra não são representativas da população. Para isso, temos uma área na estatística chamada Amostragem, que será responsável pelo desenvolvimento de métodos de como selecionar os elementos populacionais para compor a amostra de modo que as principais características contidas na população sejam preservadas na amostra. Esse assunto será estudado no Capítulo 7.

Contudo, sabemos que entender uma população por um subconjunto desta, gera-se uma incerteza ou erro. A estatística tenta minimizar esse erro o máximo possível, isto é, reduzir as incertezas das informações contidas na amostra e extrapolar essas informações para a população. Para isso, usamos a probabilidade como suporte, assunto estudado nos Capítulos 5, 6 e 8.

A teoria matemática que estuda a incerteza de fenômenos aleatórios é chamada de probabilidade.

Os fenômenos aleatórios estão relacionados a situações que dificilmente saberemos com certeza do que pode acontecer. Por exemplo, se arremessarmos um dado de seis faces de tamanhos iguais e desejarmos saber a face superior desse dado antes do arremesso, não temos como afirmar com certeza qual o valor, se considerarmos as faces numerados de 1 a 6. Observe que, mesmo sabendo quais os valores das faces, não podemos afirmar com exatidão qual o valor da face superior antes do arremesso. Mas, por meio da probabilidade, podemos minimizar essa incerteza e dizer que há aproximadamente 17% de chance de um número escolhido ocorrer.

Em nosso cotidiano, a probabilidade auxilia na decisão de um fabricante de cola de empreender uma grande publicidade no seu produto visando aumentar sua participação no mercado, ou na decisão de parar de imunizar pessoas com menos de vinte anos contra determinada doença, ou ainda na decisão de arriscar-se a atravessar uma rua no meio do quarteirão. Esses pequenos exemplos mostram a relação que a probabilidade tem com a inferência estatística, pois ela nos auxiliará a tomar decisões em procedimento inferências tentando traduzir para a nossa linguagem do dia-a-dia.

Ao final dessas definições gerais, podemos mostrar uma ilustração, Figura 1.2, que facilitará a compreensão do que abordamos nessa seção. Por fim, um último assunto estudado nesse livro, Capítulo 11, será o estudo da correlação e regressão linear, quando estamos interessado em estudar a forma e o grau relação entre duas ou mais variáveis.

Todos esses assuntos estudaremos nos capítulos seguintes com um certo detalhamento, dando ênfase a exemplos práticos estudados em nosso campo de trabalho. Alguns Capítulos poderão conter uma seção chamada *Aprofundamento*, com o intuito de apresentar uma maior profundidade sobre o tema estudado.

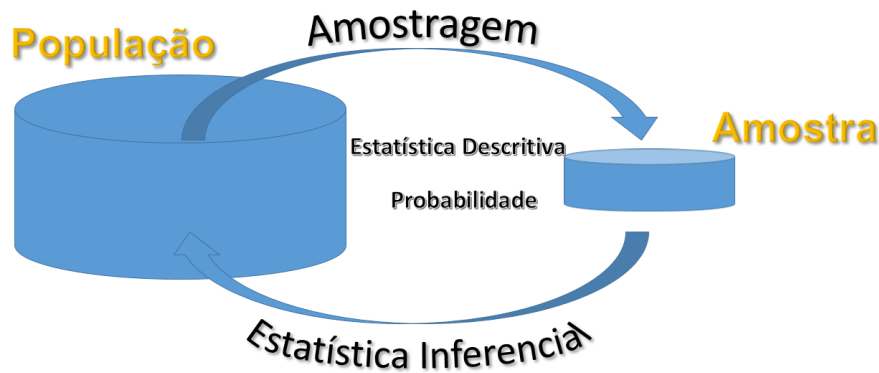


Figure 1.2: [Ilustração/Vídeo animado sobre as disposições gerais da Estatística.](<https://youtu.be/mASLUwyaC5Q>)

### 1.2.1 Estatística na pesquisa científica

O trabalho estatístico é parte integrante do método científico. Segundo ao Gilberto Corrêa da Silva (2007), definimos,

Um conjunto de regras e procedimentos para a obtenção de resultados, isto é, uma conclusão, sobre um determinado problema de uma pesquisa científica, é denominado de Método científico.

A pesquisa científica por sua vez, desenvolve conhecimentos para um saber mínimo de um determinado fenômeno estudado. A pesquisa científica se inicia a partir de um problema dentro da população em estudo. Por meio desse problema surgem diversas indagações.

## 1.3 Definições básicas

### 1.4 Técnicas de somatório

#### 1.4.1 Normas de arredondamento

## **Exercícios propostos**

## Chapter 2

# Coleta, organização e apresentação dos dados





## Chapter 3

# Medidas de posição

### 3.1 Introdução

### 3.2 Média

### 3.3 Mediana

### 3.4 Moda



## Chapter 4

# Medidas de dispersão

4.1 Introdução

4.2 Amplitude total

4.3 Variância

4.4 Desvio padrão

4.5 Coeficiente de Variação

4.6 Erro padrão da média



## Chapter 5

# Probabilidades

### 5.1 Conceitos básicos

#### 5.1.1 Leis básicas da teoria de conjuntos

#### 5.1.2 Teoremas clássicos da teoria de conjuntos

### 5.2 Definições de probabilidades

### 5.3 Propriedades

#### 5.3.1 Regra de adição de probabilidades

#### 5.3.2 Probabilidade de um evento complementar

### 5.4 Eventos independentes e probabilidade condicional

### 5.5 Teorema de Bayes

### 5.6 Variável Aleatória

### 5.7 Distribuição de probabilidade discreta

## 5.8 Distribuição de probabilidade contínua

## 5.9 Função de distribuição acumulada (FDA)

### 5.9.1 FDA para $X$ discreta

### 5.9.2 FDA para $X$ contínua

## 5.10 Esperança matemática e variância

### 5.10.1 Esperança matemática

### 5.10.2 Mediana

### 5.10.3 Moda

### 5.10.4 Variância e desvio padrão

## Exercícios propostos

## Chapter 6

# Distribuições de probabilidade

### 6.1 Introdução

### 6.2 Distribuições discretas de probabilidades

#### 6.2.1 Distribuição Bernoulli

#### 6.2.2 Distribuição Binomial

#### 6.2.3 Distribuição Poisson

#### 6.2.4 Relação entre a distribuição Binomial e Poisson

### 6.3 Distribuições contínuas de probabilidades

#### 6.3.1 Distribuição normal

##### 6.3.1.1 Distribuição normal padrão





## Chapter 7

# Amostragem

### 7.1 Introdução

### 7.2 Amostragem não-probabilística e probabilística

### 7.3 Técnicas de amostragem probabilística



## Chapter 8

# Distribuição de amostragem

### 8.1 Introdução

### 8.2 Distribuição de amostragem da média

### 8.3 Distribuição de amostragem de proporções

### 8.4 Distribuição de amostragem de diferença entre médias

### 8.5 Distribuições amostrais para variância



## Chapter 9

# Teoria de estimação

### 9.1 Introdução

### 9.2 Conceitos básicos

### 9.3 Tipos de estimador

### 9.4 Propriedades de um estimador

### 9.5 Estimação por ponto

### 9.6 Estimação por intervalo

#### 9.6.1 Intervalo de confiança para a média

#### 9.6.2 Intervalo de confiança para a variância

#### 9.6.3 Intervalo de confiança de confiança para a diferença entre médias

### 9.7 Dimensionamento de amostras



## Chapter 10

# Teoria da decisão

10.1 Introdução

10.2 Teste de hipóteses

10.3 Erros tipo I e II

10.4 Teste unilateral e bilateral

10.5 Passos para a construção de um teste de hipótese

10.6 Testes de hipóteses para a média

10.7 Testes de hipóteses para a proporção

10.8 Testes de hipóteses para a variância

10.9 Testes de hipóteses para a diferença entre duas médias





## Chapter 11

# Correlação e regressão linear simples

### 11.1 Introdução

### 11.2 Correlação linear

#### 11.2.1 Coeficiente de correlação linear

#### 11.2.2 Teste de hipóteses acerca do coeficiente de correlação linear

### 11.3 Regressão linear simples

#### 11.3.1 Modelo

#### 11.3.2 Estimação dos parâmetros do modelo

#### 11.3.3 Teste de hipóteses para o modelo de regressão

#### 11.3.4 Medidas de adequação do modelo



# Apêndice A - Introdução ao R



# Bibliography

ao Gilberto Corrêa da Silva, J. (2007). *Planejamento da Pesquisa Experimental*. Instituto de Física e Matemática, Pelotas.

British Petroleum Company (2018). Bp statistical review of world energy. Technical report, British Petroleum Company, London.