

# Laboratório de Fonologia



## Estatística para Linguística

**Prof. Dr. Adelino Pinheiro Silva**

# Sumário

Introdução

Introdução

Estatística Descritiva

Probabilidades

Estimação de Parâmetros

Teste de Significância

Comparação de dois grupos

Associação de Variáveis Categóricas

Regressão Linear e Correlação

Relação Multivariável

Regressão Múltipla e Correlação

Análise de Variância - ANOVA

Preditores Quantitativos e Categóricos

Modelos com Regressão Múltipla

Regressão Logística

Introdução a métodos aprofundados

Encerramento

Referências

Introdução

**Introdução**

Estatística Descritiva

Probabilidades

Estimação de Parâmetros

Teste de Significância

Comparação de dois grupos

Associação de Variáveis Categóricas

Regressão Linear e Correlação

Relação Multivariável

Regressão Múltipla e Correlação

Análise de Variância - ANOVA

Preditores Quantitativos e Categóricos

Modelos com Regressão Múltipla

Regressão Logística

Introdução a métodos aprofundados

Encerramento

Referências

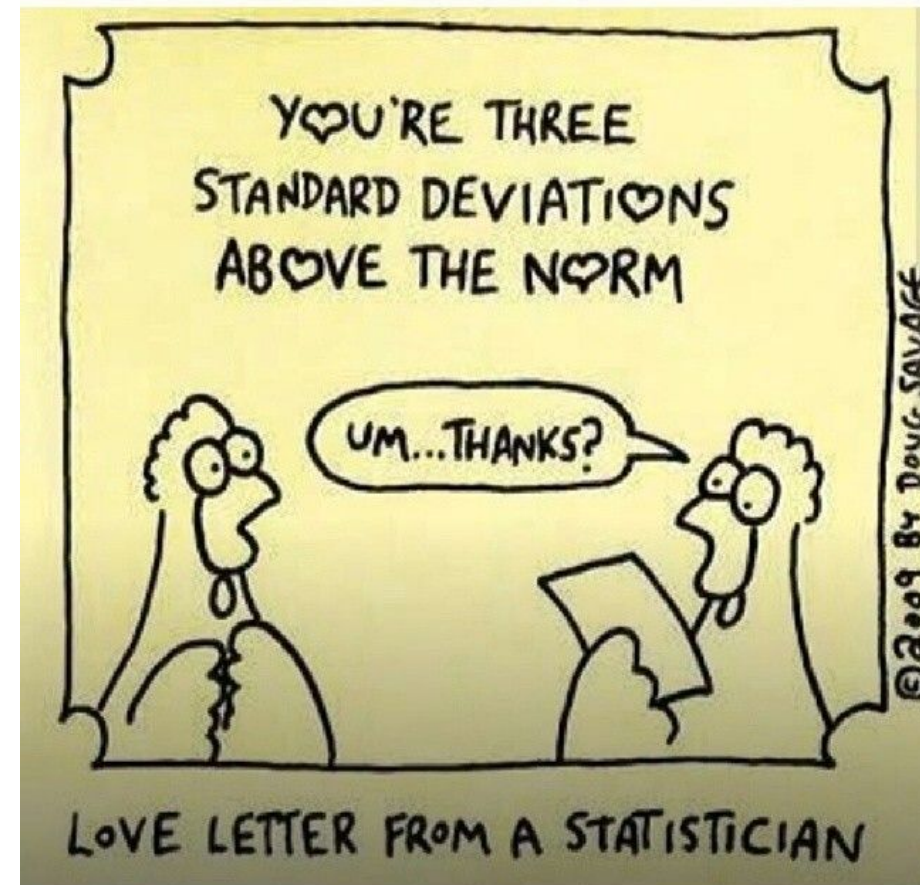
# In a hole in the ground there lived a...

Por que estudar estatística?

- Compreender **fatores** que afetam um resultado.
- Julgar de forma crítica as informações recebidas.
- Argumentar estatisticamente.

O que é estatística (Agresti, 2018)?

- Conjunto de métodos para se **obter** e **analisar** dados.
- Metodologia baseada na **ocorrência** para realizar **previsão**.



# ► In a hole in the ground there lived a...

*“Acho que somos forçados a concluir que a gramática é autônoma e independente do significado, e que os modelos probabilísticos não fornecem nenhum entendimento particular dentro de alguns problemas básicos da estrutura sintática. (tradução minha)” (Chomsky, 2009, p.-17) citado em (Levshina, 2015, p. 2)*

O que é estatística não pode fazer (Levshina, 2015)

- O *software* estatístico não pode fazer a pesquisa por você.
- As estatísticas não respondem o “por quê”.
- A causalidade é sempre imposta pelo pesquisador com base em suas considerações teóricas, dados empíricos e senso comum.

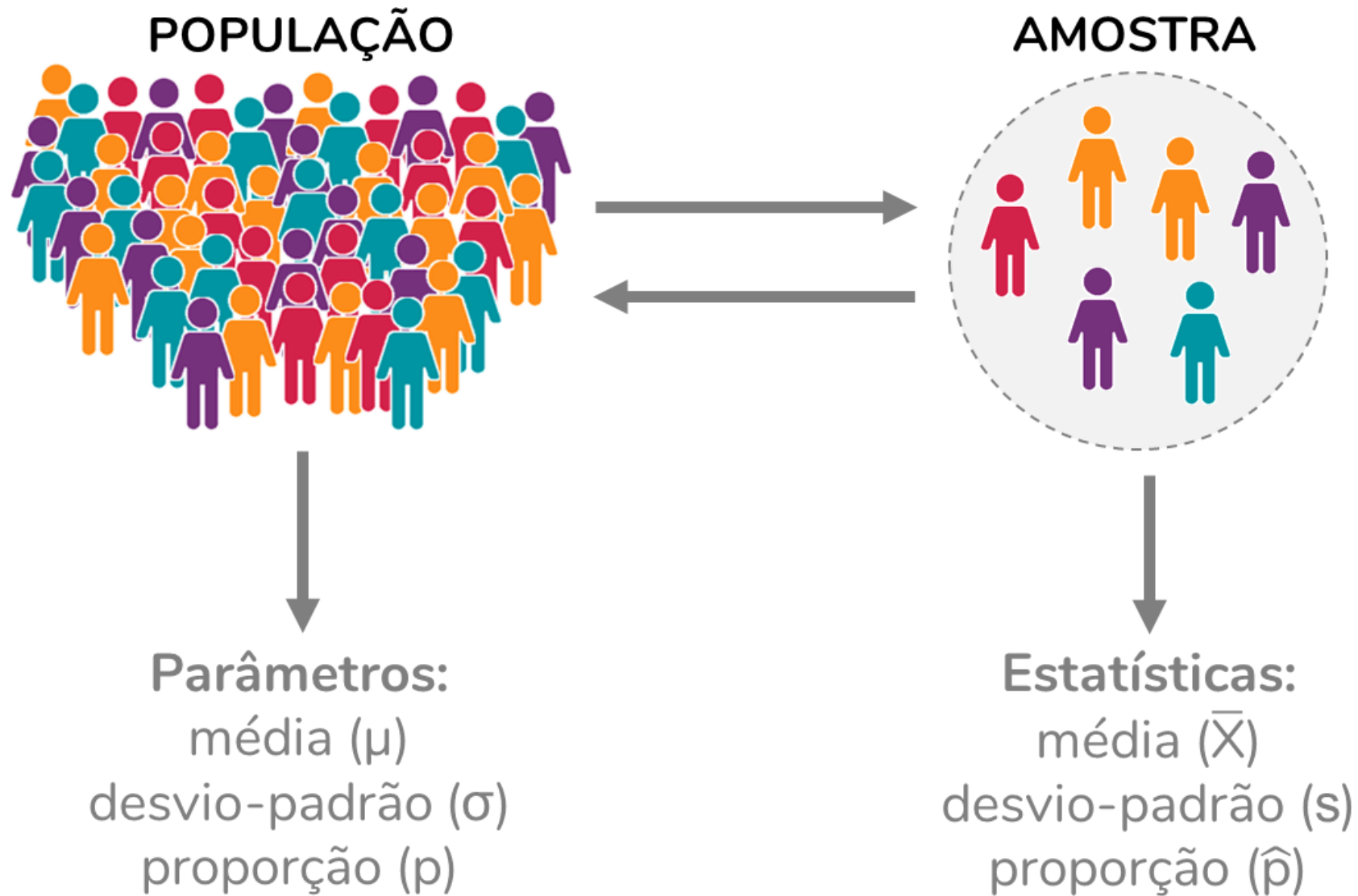
# I Have the High Ground

Alguns termos para começar

- **Dado:** Observação obtida sobre o objeto de interesse.
- **Observação:** Medida, ou informação coletada (sujeita a ruído e erros).
- **Base de dados:** Conjunto de dados, e.g., *general social survey*.
- **População:** Conjunto total dos elementos (desconhecido, inacessível).
- **Amostra:** subconjunto da população, dados (medidas) coletados.
- **Parâmetro:** Fator (resumo) numérico da população (dica: letras gregas).
- **Estatística:** Valor obtido da amostra !!!!!
- **Ferramental:** R-studio



# I Have the High Ground



# Medida e amostra

Maneiras de extrair informações de interesse.

- **Variável aleatória:** Característica que pode variar com os elementos da população ou amostra.
- **Escala de medição:** Extensão onde a variável aleatória pode ser medida. Exemplos:
  - Categóricas: (cara, coroa), (derrota, empate, vitória); ou
  - Quantitativas:  $\{x \in \mathbb{R} | 0 \leq x \leq 1\}$ ,  $[0, 1]$

Se caracteriza a variável aleatória como um resultado de uma experiência aleatória, que pode ser classificada como:

- **Categóricas:** valores aceitos dentro de um limite de categorias (qualitativos?).
- **Quantitativas:** valores numéricos de qualquer conjunto, e.g.,  $\mathbb{N}$ ,  $\mathbb{R}$ ,  $\mathbb{C}$



# Medida e amostra

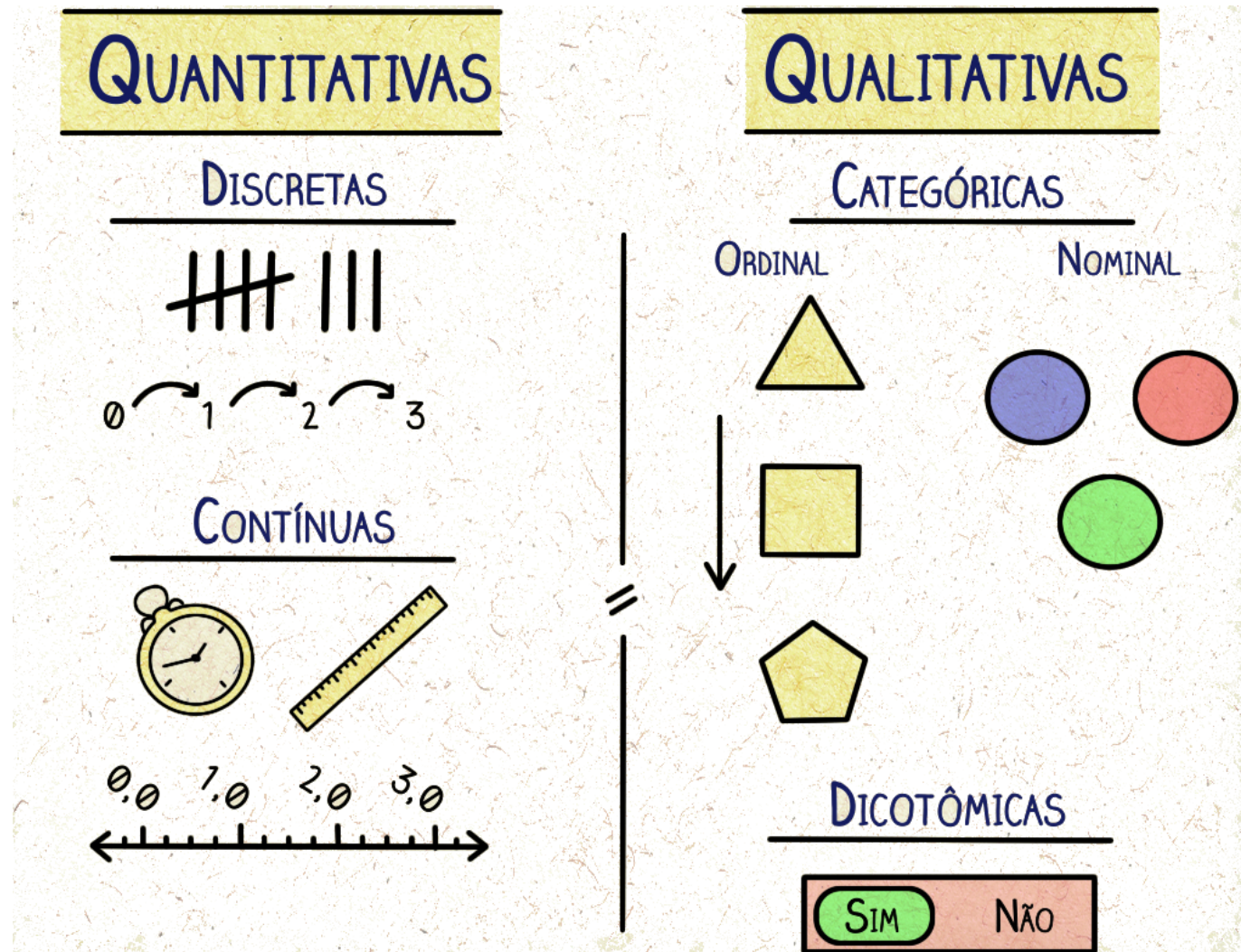
Escalas:

- **Intervalar:** delimitação numérica.
- **Nominal:** Nomes/categorias “não ordenáveis”, e.g., preferência de cores;
- **Ordenáveis:** Nomes/categorias que podem ser ordenadas em níveis, e.g., expectativa do curso (baixa, sem expectativa, alta).

Detalhe: Em escalas categóricas é muito difícil garantir uma homogeneidade dos intervalos, i.e., se os intervalos das categorias possuem escalas de mesmo tamanho.

# Variáveis estatísticas

- **Amostra aleatória simples:** todas amostras de mesmo tamanho possuem a mesma “chance”. Seria um retrato da população(?).
- **Métodos de amostragem, *sample survey*:** Sistemática, estratificada, grupo (*cluster*), multiestágios.
- **Amostra enviesada:** alunos de uma sala de aula (?).



# Estudo experimental

**Experimento:** Controlar variáveis independentes e observar a variação de variáveis dependentes para dar suporte ou refutar uma hipótese.

- Compara “tratamentos”.
- Unidades de testes.
- Grupos, pelo menos, “controle” e “tratamento”.
- Variáveis estranhas (predatórias).

## Problemas experimentais

- Variação do instrumento (ou pessoa que conduz parte dele).
- Regressão analítica.
- Viés de seleção.
- Perda de unidade



# Estudo experimental

Efeitos do teste: principal e interativo

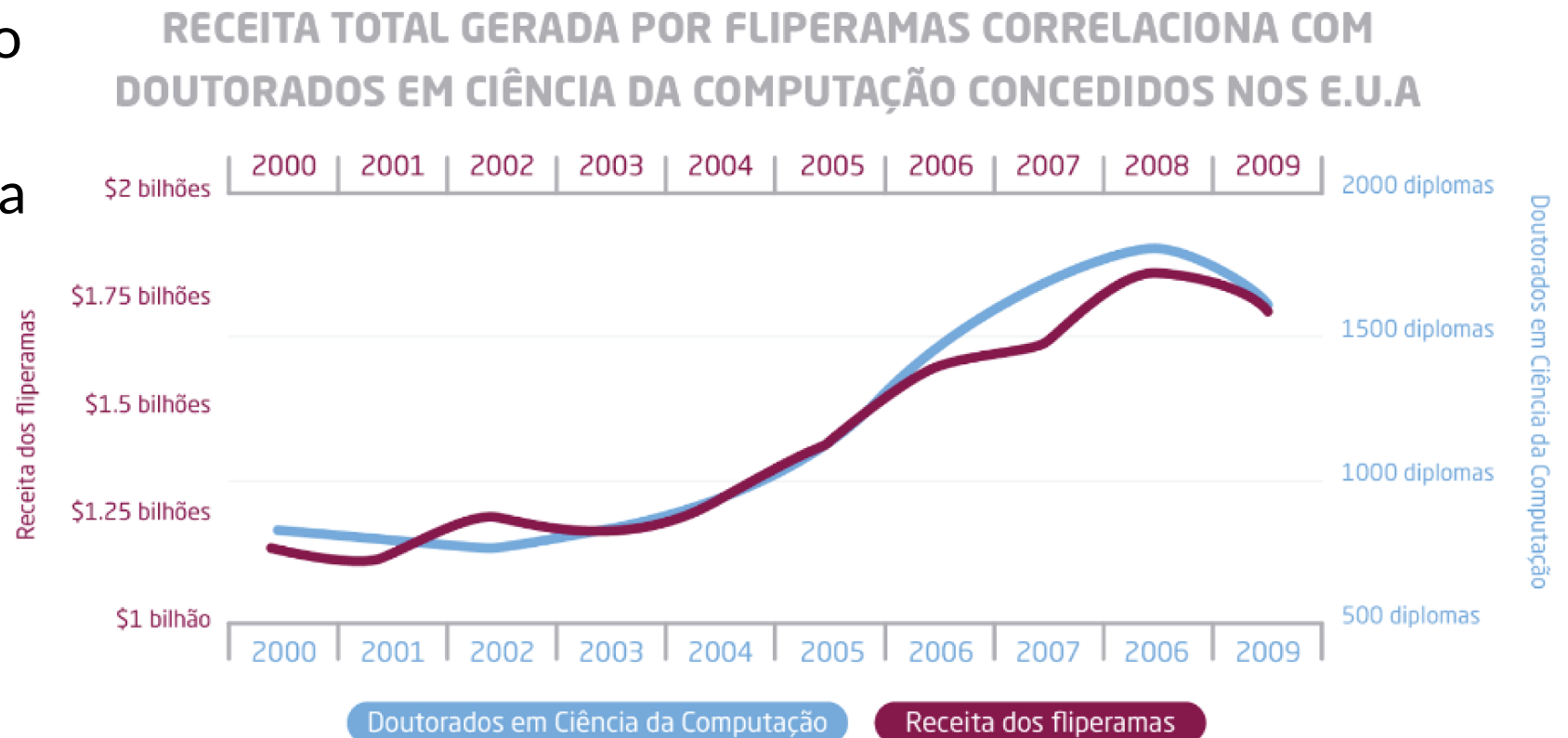
## Soluções para experimentos:

- Aleatorização.
- Emparelhamento.
- controle estatístico.
- Planejamento.
- Medições *a posteriori*.



# Estudo de Observação

- Sem manipulação do objeto de estudo.
- Grupos desbalanceados, difícil de realizar uma comparação adequada.
- **Não permite estabelecer causa e efeito.**
- Pode indicar uma relação entre variáveis.
- Uma variável não medida pode ser responsável pelo padrão observado.



# Variabilidade amostral e viés

**Erro de amostragem:** erro ocorrido ao utilizar uma estatística da amostra para prever um parâmetro da população. Exemplo: Erro da pesquisa eleitoral com  $n = 1000$  de  $\pm 3\%$ .

**Viés:** erro quando a amostra é enviesada, e.g., voluntários ou respostas de carta.

- **Viés de resposta** ocorre quando a pergunta é confusa, e.g., referendo do desarmamento;
- **viés de falha de dados** apenas uma fatia da amostra responde.

# Fim da introdução - Dever de casa

---

## **Exercícios do livro Agresti (2018):**

- Capítulo 1: 1.1, 1.3, 1.5-1.8, 1.14, 1.16;
- Capítulo 2: 2.2-2.10, 2.27, 2.35-2.37, 2.39

## **Preparação do terreno**

- Instalar o R-studio.



Introdução

Introdução

**Estatística Descritiva**

Probabilidades

Estimação de Parâmetros

Teste de Significância

Comparação de dois grupos

Associação de Variáveis Categóricas

Regressão Linear e Correlação

Relação Multivariável

Regressão Múltipla e Correlação

Análise de Variância - ANOVA

Preditores Quantitativos e Categóricos

Modelos com Regressão Múltipla

Regressão Logística

Introdução a métodos aprofundados

Encerramento

Referências



# Estatística descritiva

Primeiro passo para entender os dados coletados

Facilitar a assimilação de informação

Medidas de:

- tendência central (média), variabilidade, associação, etc...

Análise e regressão: predizer uma variável a partir de outras.

Um pouco de código R para tratar com dados

```
data_lemas <- read.table("../Dados/lexporbr_alfa_lemas_txt.txt", header = TRUE,
  sep = "\t", dec = ",", quote = "\"")
head(data_lemas)
dim(data_lemas)
summary(data_lemas)
```

Dados de Corpus Léxico do português

# Tabelas e gráficos

## Extraindo o cabeçalho dos dados

```
> head(data_lemas)
```

## Gera a saída:

	id	ortografia	cat_gram	inf_gram	freq_orto	freq_orto.M	log10_freq_orto	zipf_escala	nb_letras
1	1	o	gram	det	4364416	139093.06	6.6399	8.1433	1
2	2	de	gram	prp	2553292	81372.90	6.4071	7.9105	2
3	3	,	gram	pu	2133025	67979.08	6.3290	7.8324	1
4	4	.	gram	pu	1603184	51093.15	6.2050	7.7084	1
5	5	em	gram	prp	1044260	33280.36	6.0188	7.5222	2
6	6	e	gram	kc	667736	21280.61	5.8246	7.3280	1

A dimensionalidade dos dados, onde cada linha indica uma medição com as colunas indicando as informações

```
> dim(data_lemas)
```

Que é um total de 169.606 linhas com 9 colunas

```
[1] 169606      9
```

# Tabelas de contingência

## Construindo uma tabela

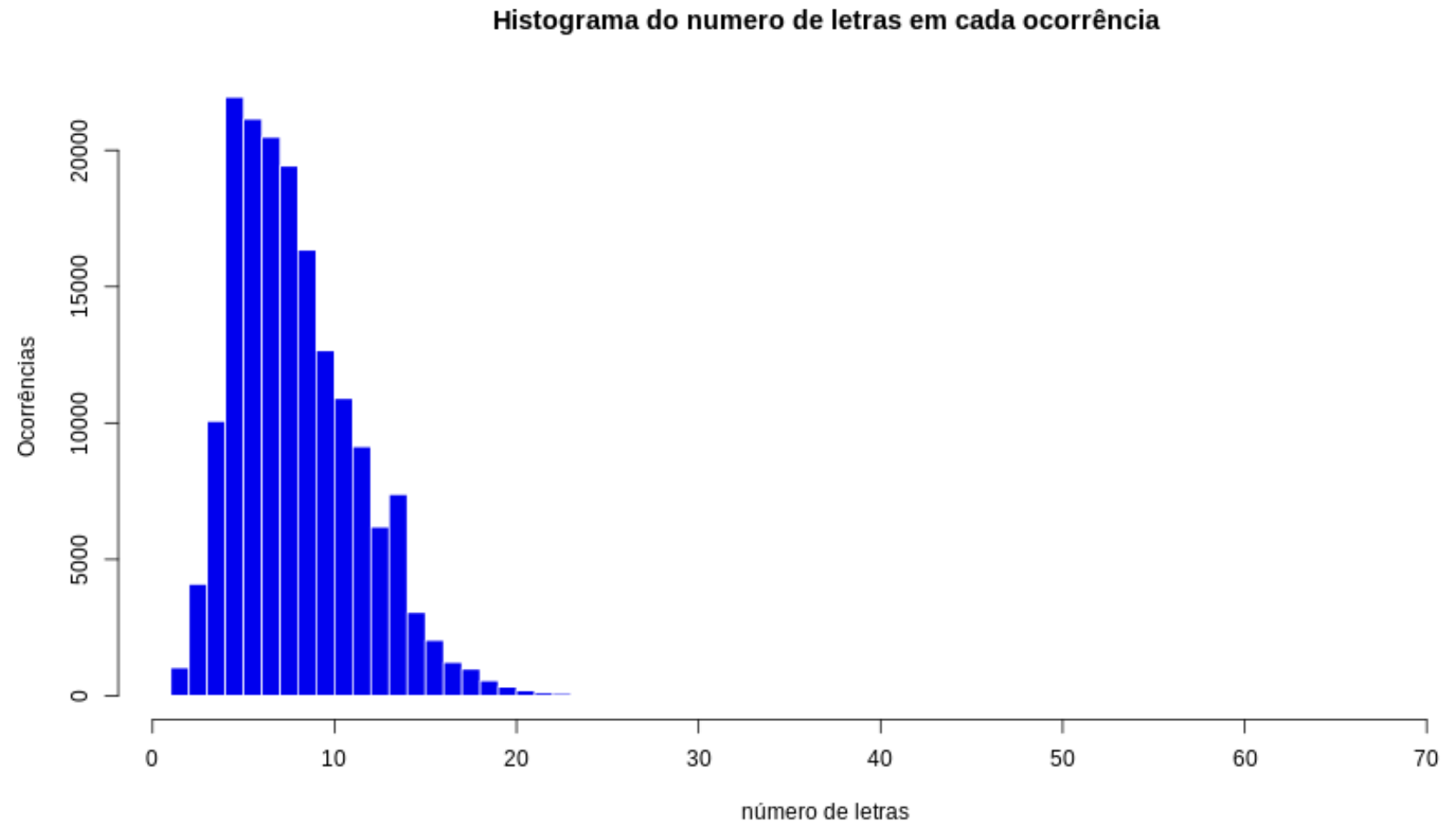
```
> tab <- table(data_lemas$cat_gram, data_lemas$nb_letras)
```

	1	2	3	4	5	6	7	8	9	10	11	12	13
adj	3	41	355	699	1061	1490	2320	2817	3117	3106	2672	2099	1575
adv	2	15	31	57	80	96	109	115	148	235	255	313	330
gram	56	60	96	75	75	47	22	12	14	8	1	2	0
nom	57	500	1517	3375	4992	5924	7396	7504	7397	6719	5912	4278	3151
num	11	275	2045	5678	14704	11792	8534	6535	3553	913	941	1630	657
ver	1	11	51	175	1037	1796	2099	2442	2115	1678	1123	815	475

# Histograma

## Histograma em uma figura PNG...

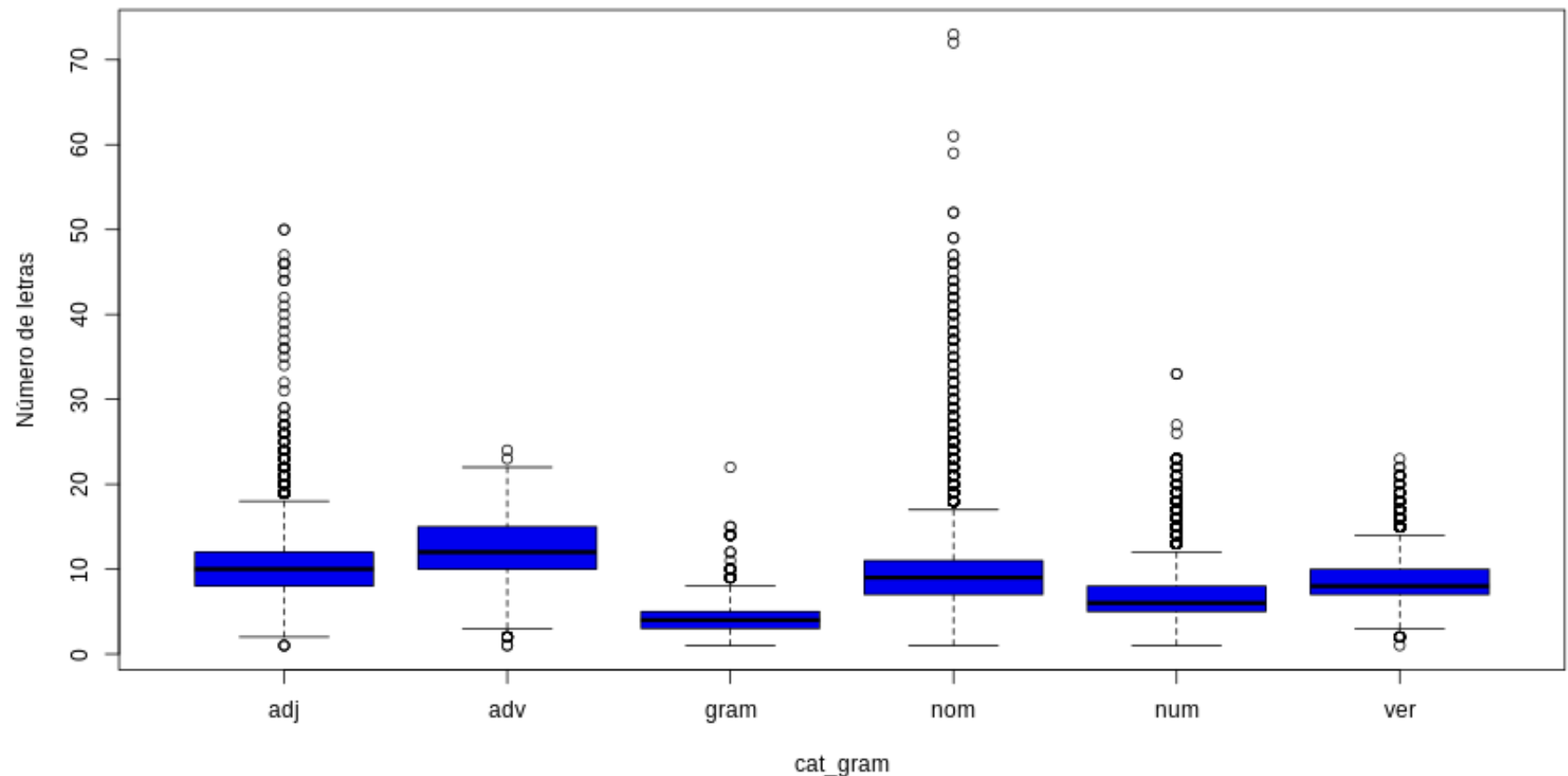
```
png(file = "../Imagens/histograma.png",width = 864, height = 486, units = "px")
hist(data_lemas$nb_letras,main="Histograma do numero de letras em cada ocorrencia",
breaks=40,xlab = "numero
border="white")
dev.off()
```



# Diagrama de caixa

Diagrama de caixa (*boxplot*) em uma figura PNG...

```
png(file = "../Imagens/Box_plot.png", width = 864, height = 486, units = "px")
boxplot(data_lemas$nb_letras ~ cat_gram, data = data_lemas, ylab = "Número de
  letras",
col = "blue2", borde
dev.off()
```

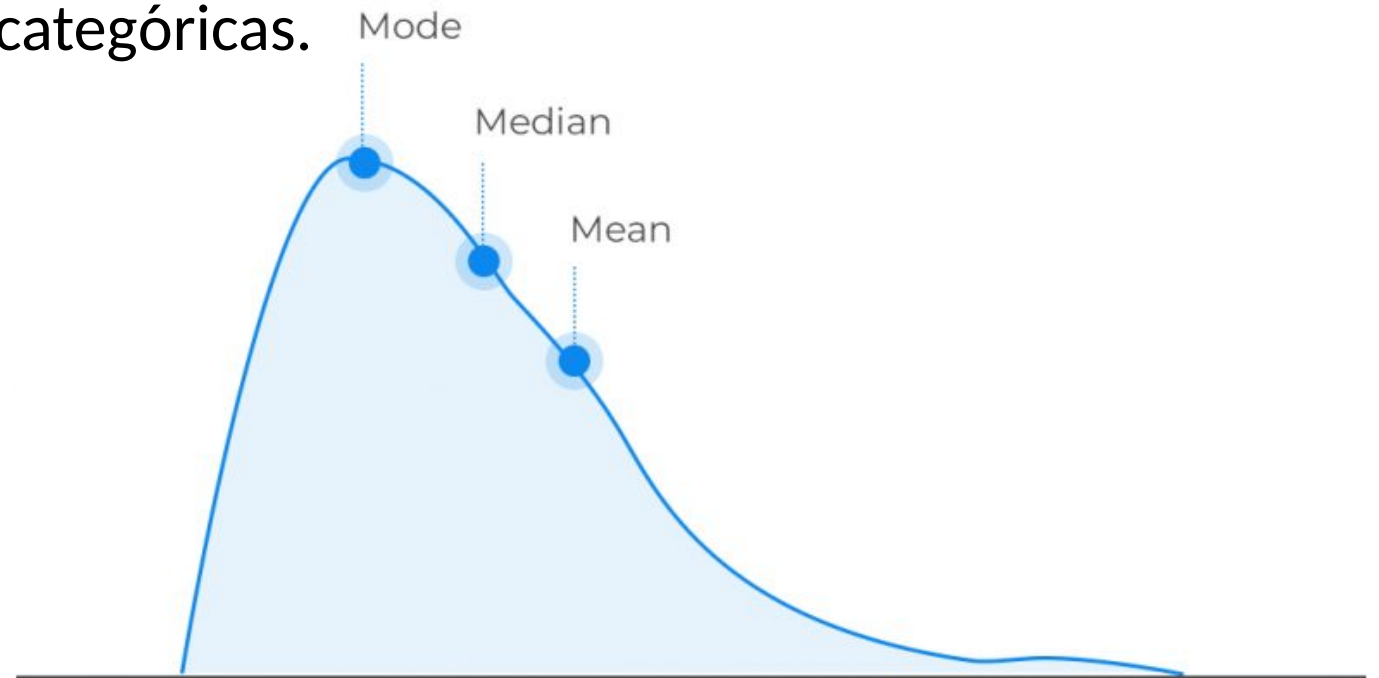


```
> tab <- stem(data_lemas$nb_letras[1:300])
```

[illegible]

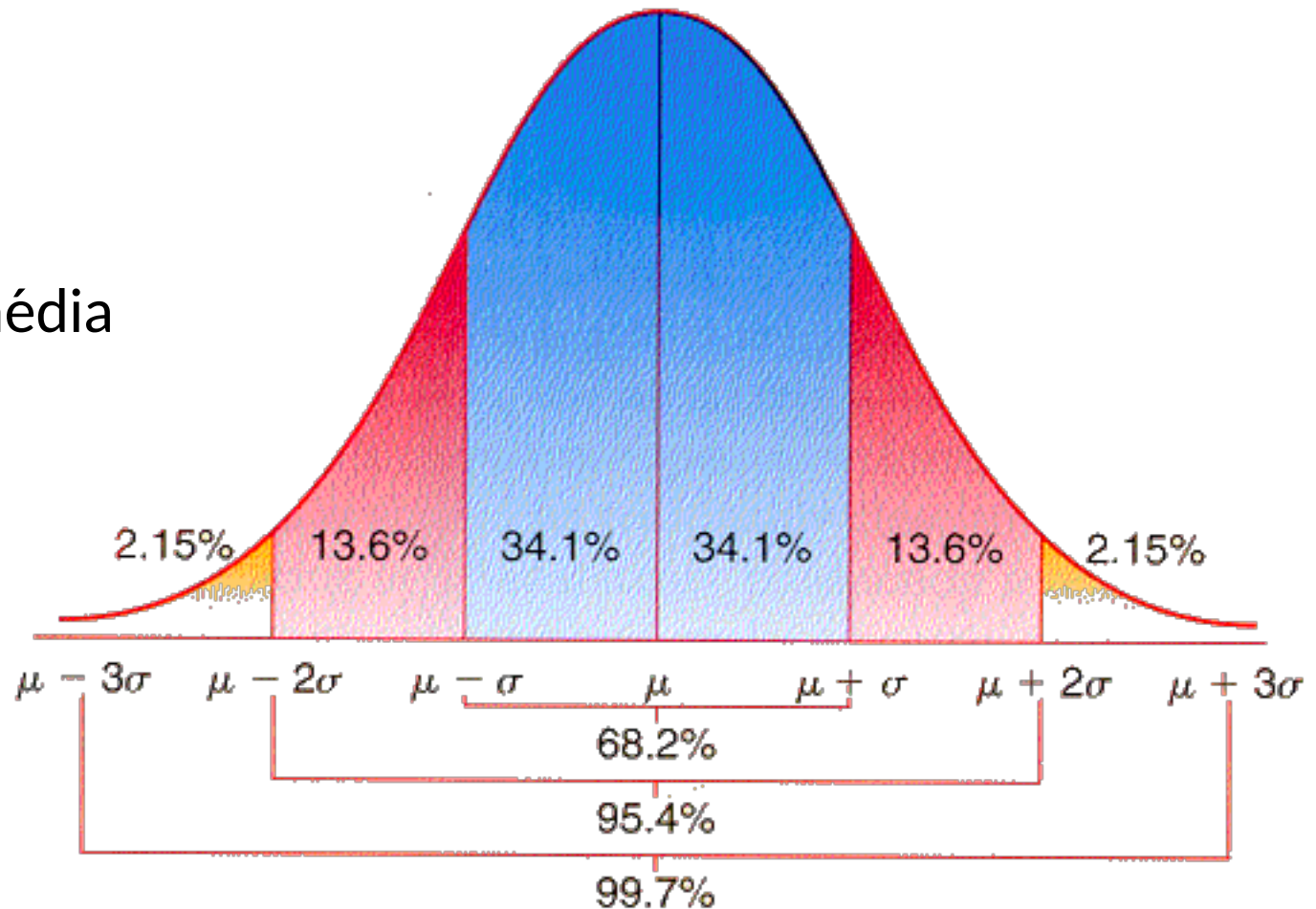
# Medidas de tendência central

- Média
  - Aritmética: problema que *outliers* podem alavancar.
  - Truncada (*winsorized*)
  - Ponderada
- Mediana: menos problemas com *outliers*.
- Moda: bem indicada para variáveis categóricas.
- Tri-média: utiliza quartis.



# Medidas de dispersão

- *Range*, alcance, diferença entre mínimo e máximo.
- desvios
  - variância e desvio padrão.
  - Soma dos desvios quadrados.
- Distância inter-percentis.
- **Erro padrão**: desvio padrão da média
- Regra do  $\sigma$ :
  - $0,67\sigma \rightarrow 50\%$
  - $1\sigma \rightarrow 68,3\%$
  - $1,96\sigma \rightarrow 95\%$
  - $2\sigma \rightarrow 95,4\%$
  - $3\sigma \rightarrow 99,7\%$

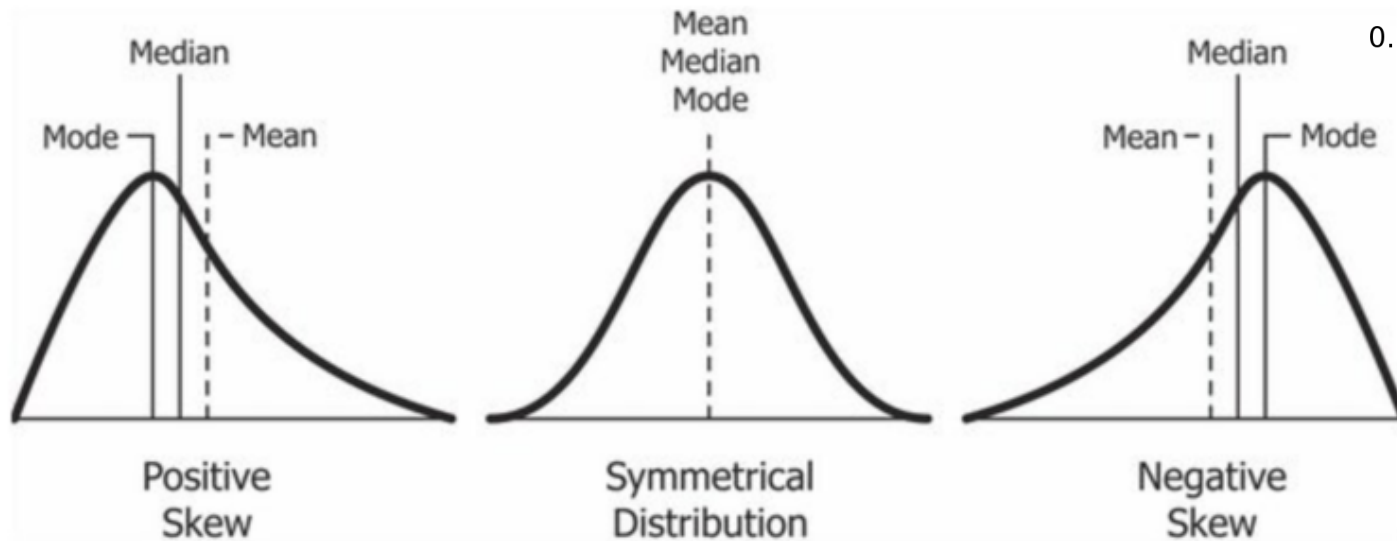
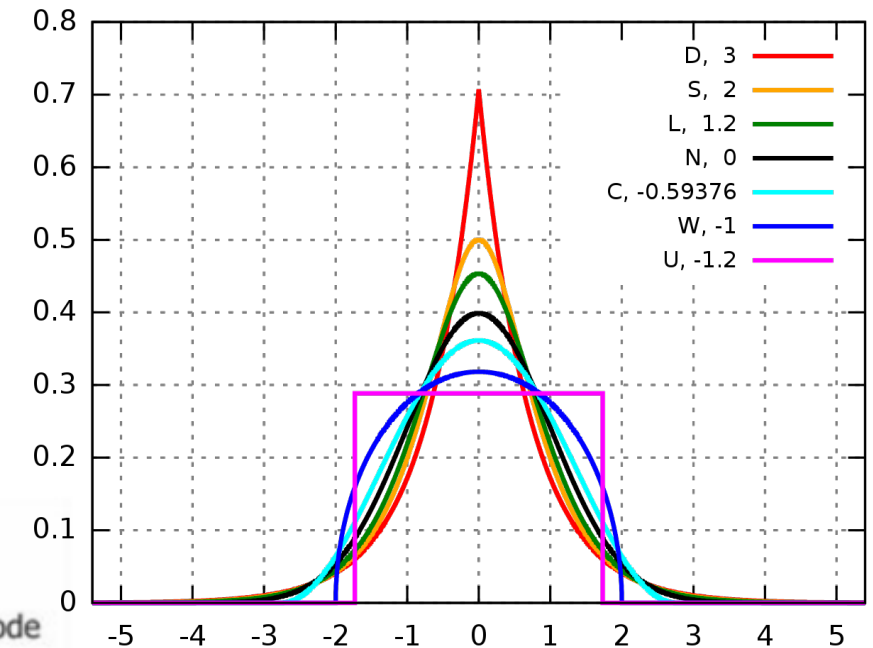




# Achatamento(curtose) e (As)simetria

## Mais medidas de caracterização dos dados

```
library(moments)
curt_data <- kurtosis(vec_n_letras_sel)
assi_data <- skewness(vec_n_letras_sel)
```

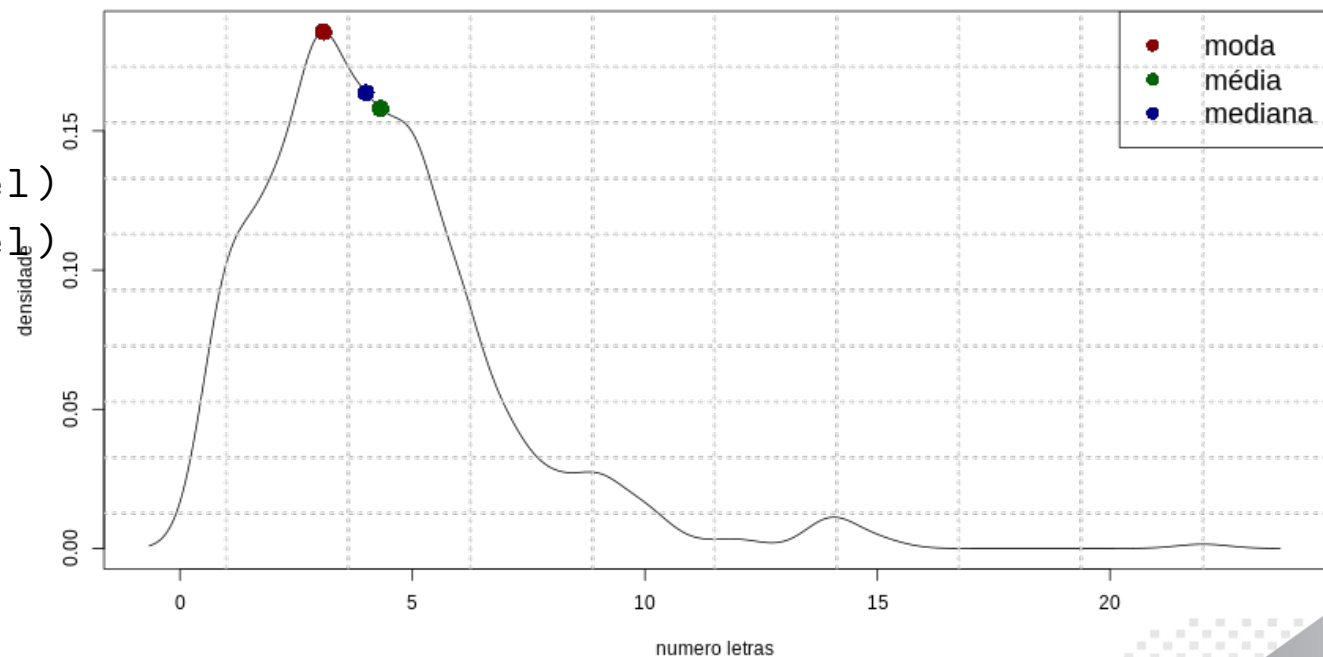


# Exemplos de Estatísticas

Extraindo algumas estatísticas dos dados:

```
library(moments)
vec_n_letras_sel <- data_lemas[data_lemas$cat_gram
  %in% 'gram',]$nb_letras
data_density <- density(vec_n_letras_sel,n=4096,
  bw=1.2*bw.nrd(vec_n_letras_sel))
idx_max <- which.max(data_density$y)
moda_data <- data_density$y[idx_max]
mean_data <- mean(vec_n_letras_sel)
medi_data <- median(vec_n_letras_sel)
stdv_data <- sd(vec_n_letras_sel)
curt_data <- kurtosis(vec_n_letras_sel)
assi_data <- skewness(vec_n_letras_sel)
```

```
3.088693
4.313808
4
2.719188
8.536706
1.746169
```

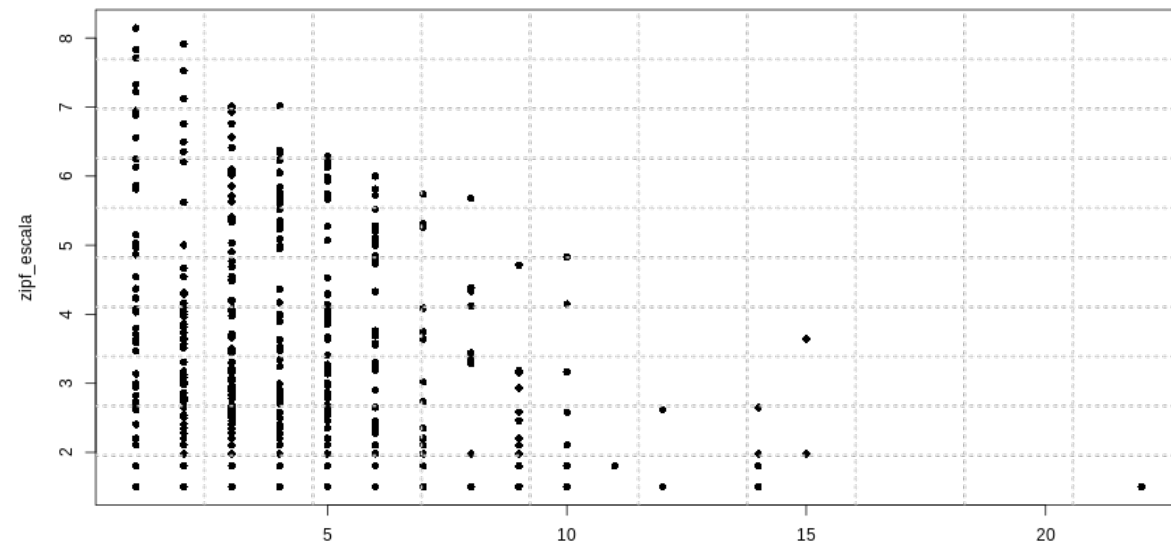


# Representações bivariadas

- Tabelas de contingência
- Gráficos de dispersão
- Correlação
  - Pearson, Kendall, Spearman
- Informação mútua...

Extraindo dados de duas variáveis:

```
vec_X_sel <- data_lemas[data_lemas$cat_gram %in% 'gram',]$nb_letras
vec_Y_sel <- data_lemas[data_lemas$cat_gram %in% 'gram',]$zipf_escala
png(file = "../Imagens/scatter_plot_02.png", width = 864, height = 486, bg = "
transparent")
plot(x=vec_X_sel, y=vec_Y_sel, type='p', pch=16, xlab="log10_freq_orto", ylab="zipf_
escala")
grid(10, lwd = 2)
dev.off()
cor(vec_X_sel, vec_Y_sel)
```



# Fim da Estatística Descritiva - Dever de casa

## Exercícios do livro Agresti (2018):

- Capítulo 3: 1.1, 1.3, 1.5-1.8, 1.14, 1.16;

## Preparação do terreno

- Reproduzir os exemplos no R-Studio.

Lembrete:

**Parâmetros** de populações geralmente são representados por letras gregas, e.g.,  $\mu$  (média),  $\sigma^2$  (variância),  $\pi$  (proporção), etc...

**Estatísticas** são extraídas das amostras e representadas por letras latinas, com ou sem complemento, e.g.,  $m$ ,  $s^2$ ,  $p$ .

# Assunto

Introdução

Introdução

Estatística Descritiva

**Probabilidades**

Estimação de Parâmetros

Teste de Significância

Comparação de dois grupos

Associação de Variáveis Categóricas

Regressão Linear e Correlação

Relação Multivariável

Regressão Múltipla e Correlação

Análise de Variância - ANOVA

Preditores Quantitativos e Categóricos

Modelos com Regressão Múltipla

Regressão Logística

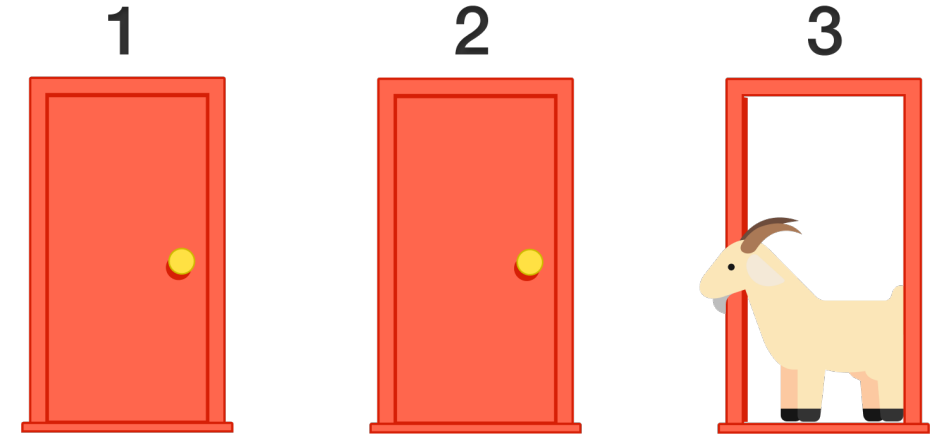
Introdução a métodos aprofundados

Encerramento

Referências

# Algunas definições

- Valor que indica o quão suscetível um evento está de ocorrer.
- Proporção de um evento em particular dada uma longa sequência de observações
- Bases para o cálculo de probabilidades:
  - Axiomas de Kolmogorov
  - Teorema do limite central (ou central do limite?).
  - $\sigma$ -álgebra
  - lei dos grandes números
- Exemplo do problema de Monty Hall.



**Parâmetros:** de populações geralmente são representados por letras gregas, e.g.,  $\mu$  (média),  $\sigma^2$  (variância),  $\pi$  (proporção), etc...

**Estatísticas** são extraídas das amostras e representadas por letras latinas, com ou sem complemento, e.g.,  $m$ ,  $s^2$ ,  $p$ .

# Notação e regras básicas I, mais em (Halperin et al., 1965)

A variável aleatória  $x \in X$  significa que um resultado particular (amostra)  $x$  pertence  $\in$  a variável aleatória/conjunto (população)  $X$ .

Se a variável tem seus parâmetros conhecidos, por exemplo, vem de uma distribuição normal (Gaussiana  $\mathcal{N}(\mu, \sigma)$ ) com média igual a 1,7 e desvio padrão de 0,4 podemos escrever  $x \in \mathcal{N}(1, 7, 0, 4)$ , ou  $X \sim \mathcal{N}(1, 7, 0, 4)$ .

A normal padrão possui média igual a zero e desvio unitário  $\mathcal{N}(0, 1)$

Uma probabilidade de um evento  $A$  é definida como  $P(\omega : X(\omega) \in A)$  ou simplesmente  $P(A)$  (vide nota de rodapé).

## Notação e regras básicas II, mais em (Halperin et al., 1965)

Costuma-se fazer referência ao espaço amostral como  $\Omega$ , assim  $P(\Omega) = 1$  Se

um evento  $A$  tem probabilidade  $P(A)$  de ocorrer.

$P(\bar{A}) = 1 - P(A)$  é a probabilidade do evento não ocorrer.

Dados dois eventos mutualmente independentes  $A$  e  $B$  (e.g., rodadas diferentes de um lançamento de moeda) e suas probabilidades  $P(A)$  e  $P(B)$ , a probabilidade de ocorrerem:

- $P(A)$  ou  $P(B)$  é:  $P(A \cup B) = P(A) + P(B)$  (um ou o outro ou os dois)
- $P(A)$  e  $P(B)$  é:  $P(A \cap B) = P(A) \times P(B)$  (concomitantemente)

**Independente:** Dois valores de uma mesma característica categórica, e.g., frequência fundamental grave ou aguda.



## Notação e regras básicas III, mais em (Halperin et al., 1965)

Considere dois eventos **não** mutualmente independentes  $A$  e  $B$ , como valores de características diferentes (e.g., frequência fundamental grave e presença de frênulo lingual) e suas probabilidades  $P(A)$  e  $P(B)$ .

A probabilidade condicional, de ocorrer uma condição dada outra é:

$P(B|A)$  lê-se  $P(B)$  dado  $A$

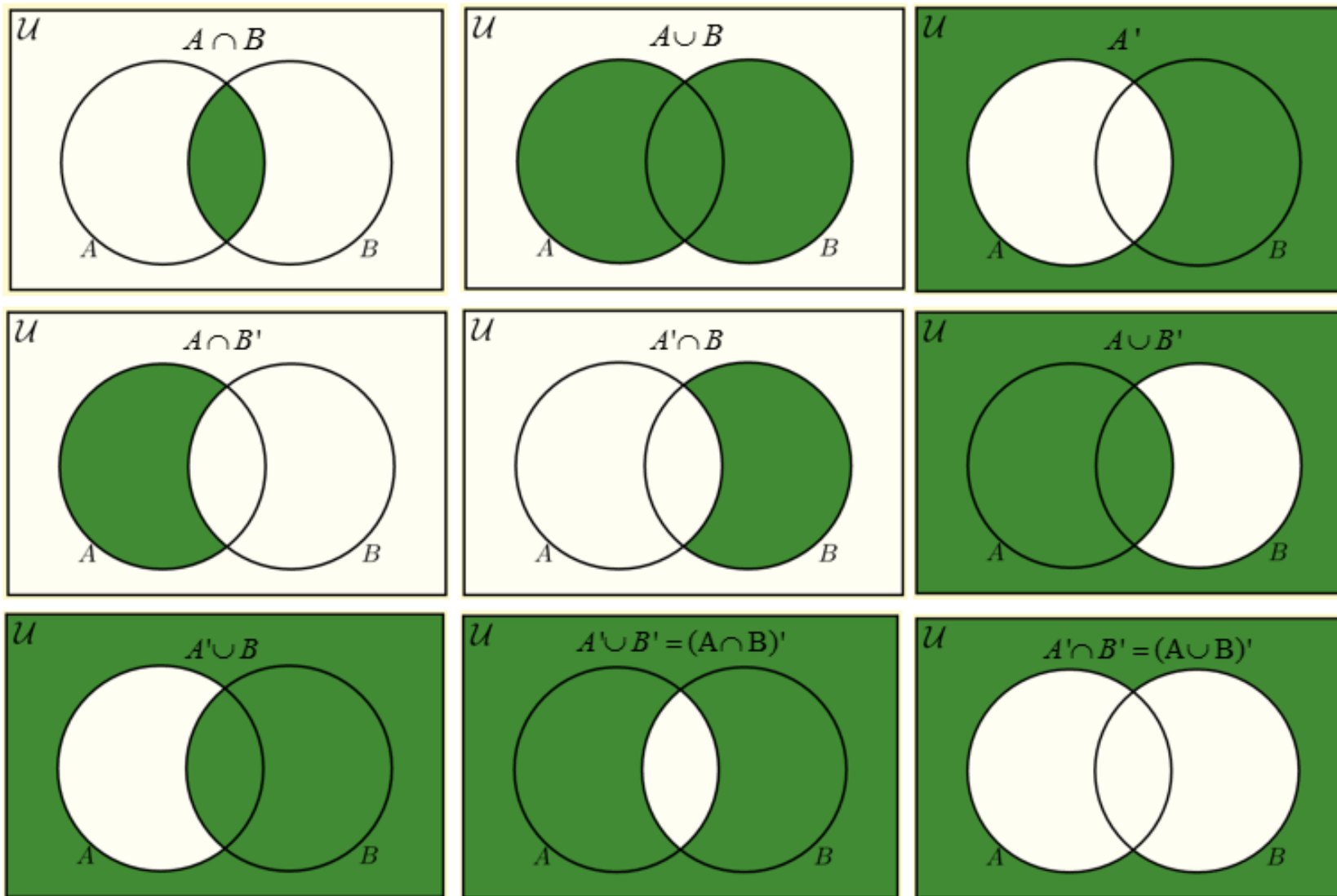
Neste caso:

$$P(A \cap B) = P(A) \times P(B|A).$$

Teorema de Bayes:

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

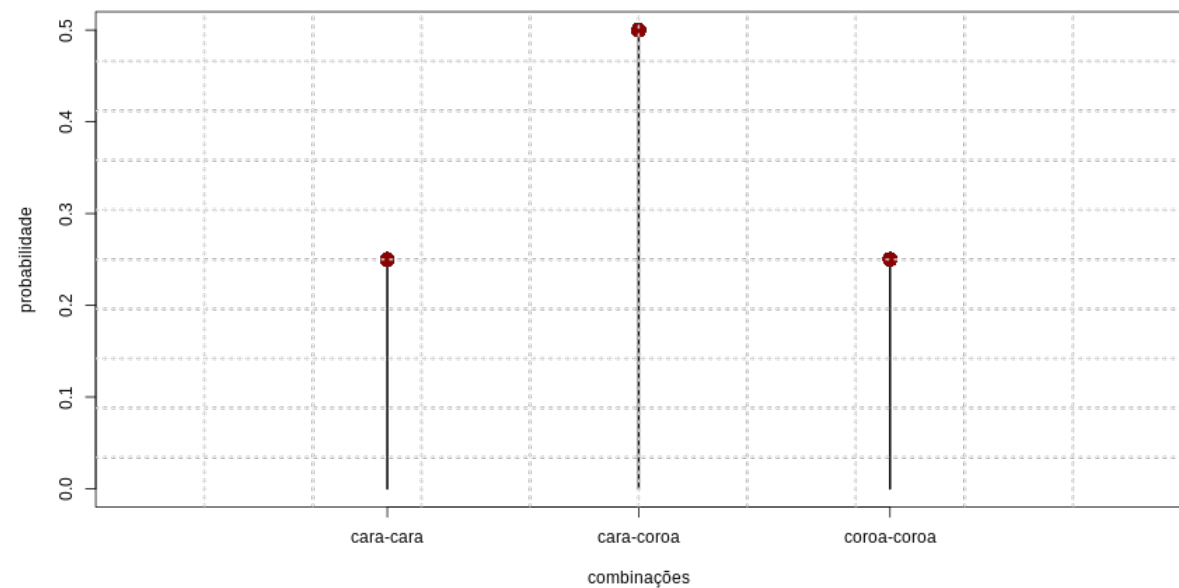
# Probabilidades em Diagramas de Venn



# Distribuições

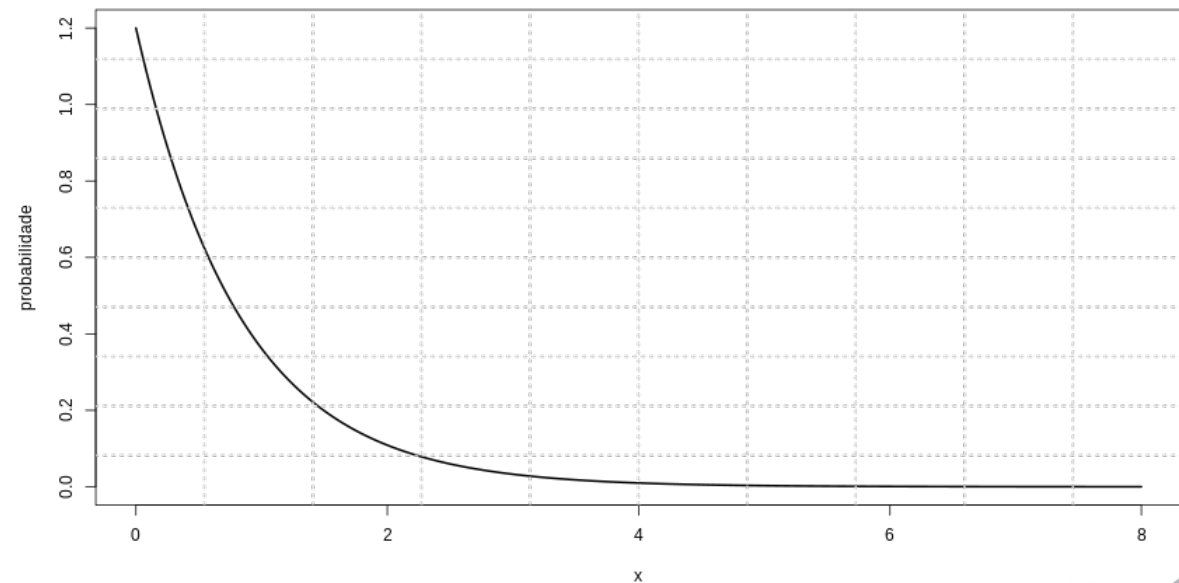
## Distribuição de uma variável discreta

- $0 \leq P(x) \leq 1$ .
- $\sum P(x) = 1$ .
- Função massa de probabilidade.
- Probabilidade está diretamente em  $P(x)$ .



## Distribuição de uma variável contínua

- $0 \leq P(x) \leq 1$ .
- $\int_{-\infty}^{\infty} p(x)dx = 1$ .
- Função densidade de probabilidade  $p(x)$ .
- $P(1 \leq x \leq 3) = \int_1^3 p(x)dx$ .



# ▶ Parâmetros - vide Casella and Berger (2011)

Valor esperado

$$E[x] = \mu = \sum_{x \in X} xP(x) \quad \text{ou} \quad \mu = \int_{x \in X} xp(x)dx$$

Variância

$$E[(x - \mu)]^2 = \sigma^2 = \sum_{x \in X} (x - \mu)^2 P(x) \quad \text{ou} \quad \sigma = \int_{x \in X} (x - \mu)^2 p(x)dx$$

Momentos estatísticos

$$E[x]^n = \sum_{x \in X} x^n P(x) \quad \text{ou} \quad \int_{x \in X} x^n p(x)dx$$

Momentos centrais

$$E[(x - \mu)]^n = \sum_{x \in X} (x - \mu)^n P(x) \quad \text{ou} \quad \int_{x \in X} (x - \mu)^n p(x)dx$$

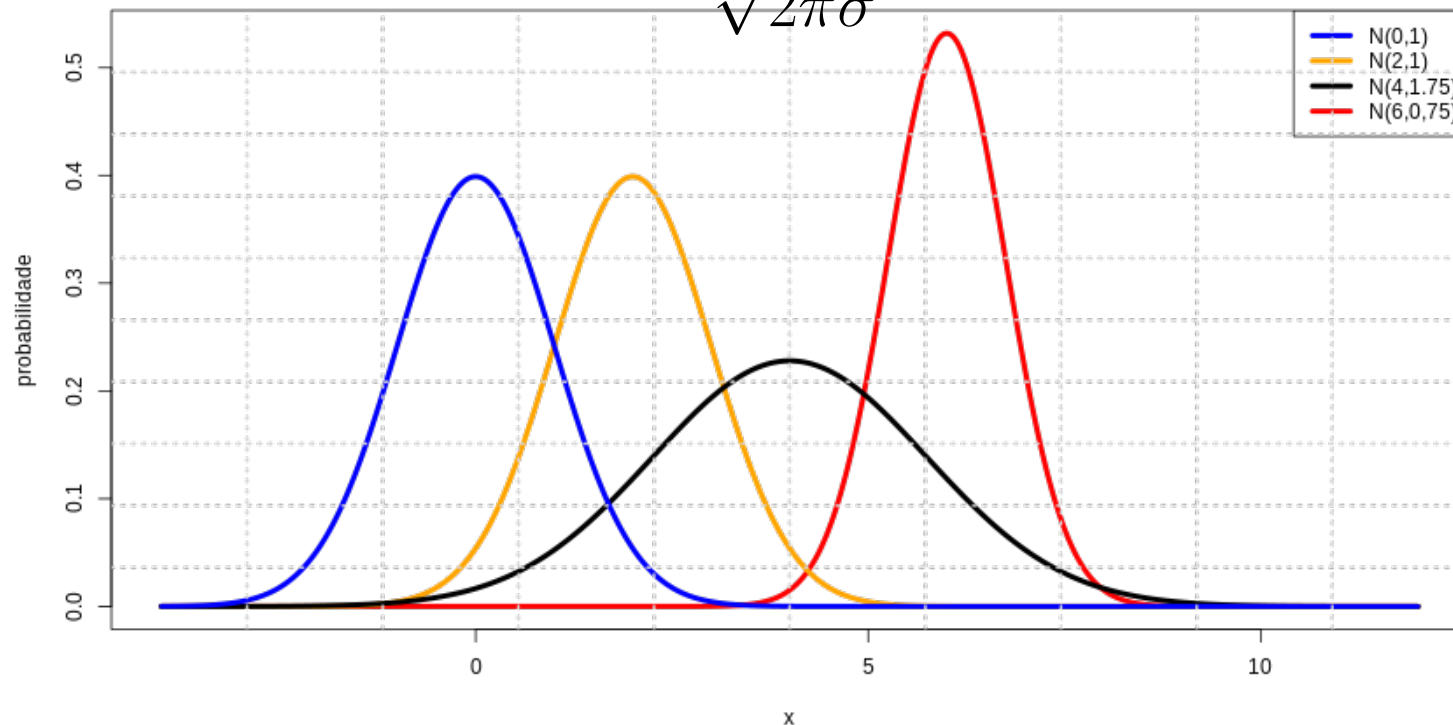
# Função Normal

Padrão com  $\mu = 0$  e  $\sigma = 1$

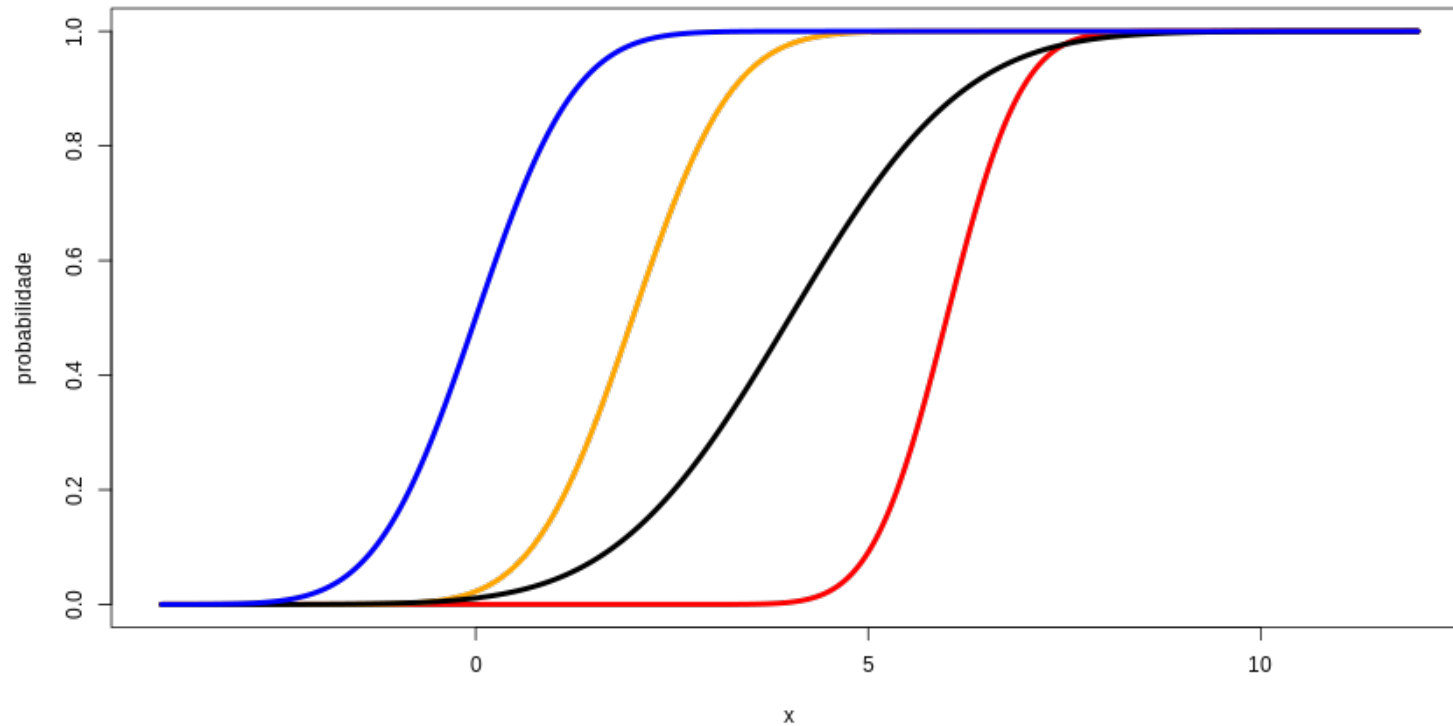
$$\mathcal{N}(x|0, 1) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{x^2}{2}}$$

Geral, lembrar da regra do  $\sigma$

$$\mathcal{N}(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



# Probabilidade acumulada



Covariância

$$\text{cov}(x, y) = E[(x - \mu_x)(y - \mu_y)]$$

Correlação

$$\text{cov}(x, y) = E\left[\frac{(x - \mu_x)}{\sigma_x} \frac{(y - \mu_y)}{\sigma_y}\right]$$

# Erro padrão

O erro padrão de uma **estatística** (na maioria das vezes a estimativa de um **parâmetro**) é o desvio padrão da distribuição amostral de uma estatística.

Erro padrão da média:

$$\sigma_{\bar{x}} = \frac{\sigma}{n}$$

Em geral, a distribuição amostral da média  $\bar{x}$  tende a uma normal independente da distribuição de  $X$ .

# Fim da Probabilidades - Dever de casa

---

## **Exercícios do livro Agresti (2018):**

- Capítulo 4: 4.1 - 4.7, 4.18 - 4.20, 4.26 - 4.32.





# Assunto

Introdução

Introdução

Estatística Descritiva

Probabilidades

**Estimação de Parâmetros**

Teste de Significância

Comparação de dois grupos

Associação de Variáveis Categóricas

Regressão Linear e Correlação

Relação Multivariável

Regressão Múltipla e Correlação

Análise de Variância - ANOVA

Preditores Quantitativos e Categóricos

Modelos com Regressão Múltipla

Regressão Logística

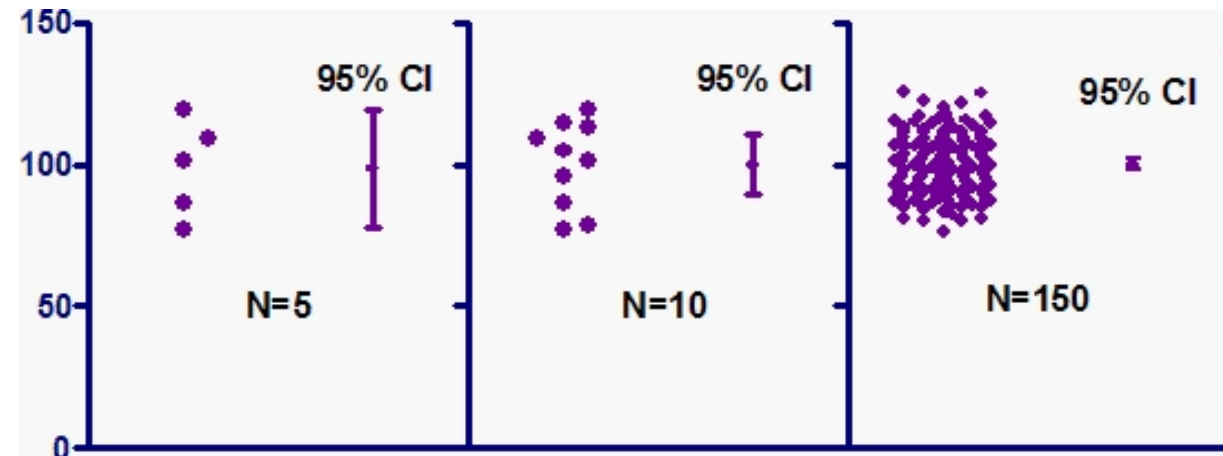
Introdução a métodos aprofundados

Encerramento

Referências

# Definições

- **Estimador pontual:** é uma melhor sugestão para um parâmetro.
  - **Estimador intervalar:** um intervalo ao redor da estimativa pontual em que acredita-se conter o parâmetro.
  - **Estimador não enviesado:** é centrado no parâmetro e possui a menor dispersão possível
- \* Eficiente → se fechado ao redor do parâmetro.



# Proporção

Uma proporção  $\hat{p}$  calculada de  $N$  é uma estimativa não enviesada do parâmetro da proporção da população  $\pi$ . Na proporção:

$$P(0) = 1 - \hat{p} \quad \text{e} \quad P(1) = \hat{p}.$$

$$\hat{m} = 0 \cdot (1 - \hat{p}) + 1 \cdot \hat{p} = \hat{p}$$

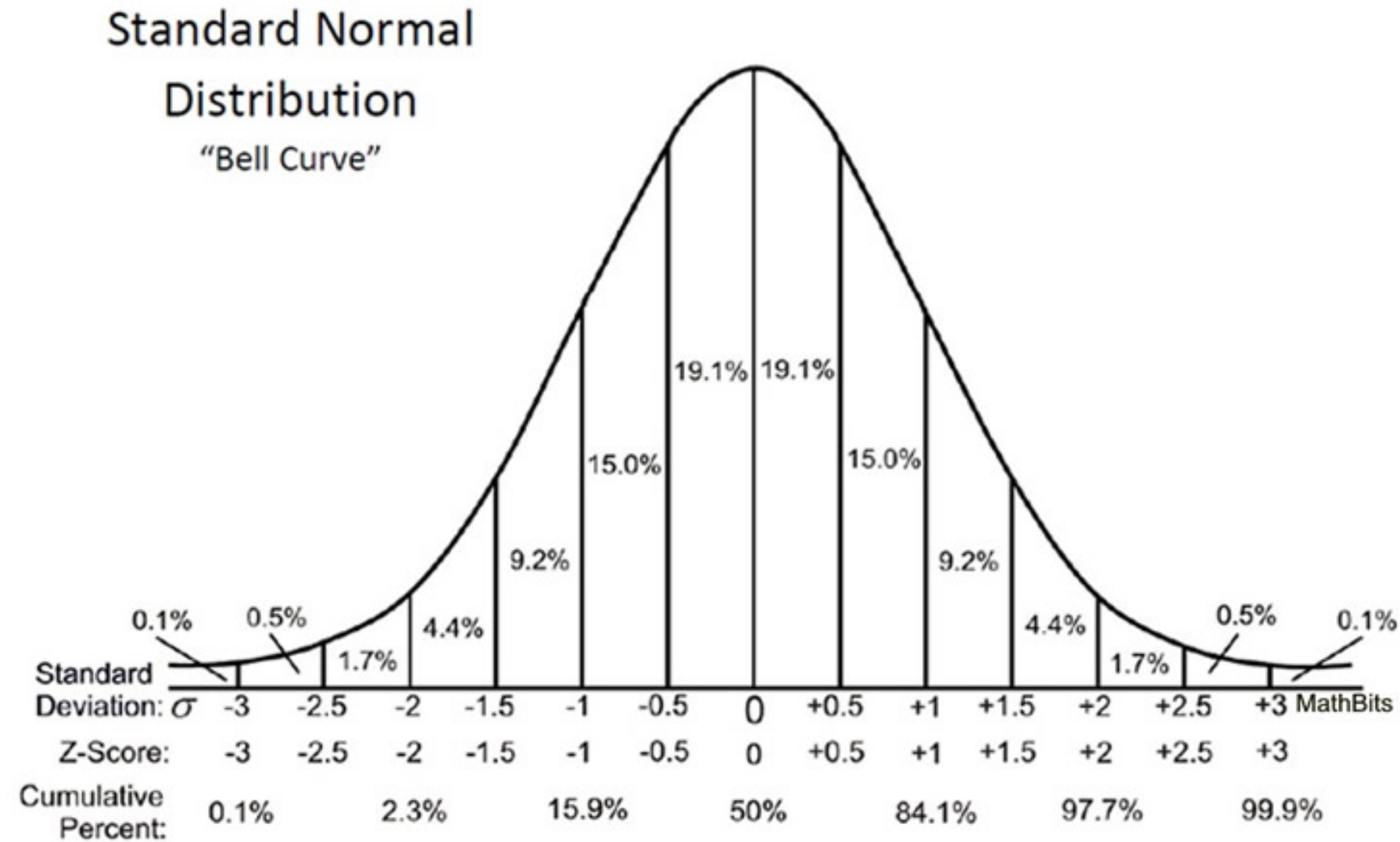
$$\hat{s}^2 = (0 - \hat{p})^2 \cdot (1 - \hat{p}) + (1 - \hat{p})^2 \cdot \hat{p} = (1 - \hat{p}) \cdot [\hat{p} \cdot (1 - \hat{p}) + \hat{p}^2] = \hat{p} \cdot (1 - \hat{p})$$

$$\hat{s}_{\mu}^2 = \frac{\hat{s}^2}{N} = \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{N}}$$

$$N = \hat{p} \cdot (1 - \hat{p}) \left( \frac{Z_{\alpha}}{M} \right)^2$$

$$Z_{\frac{\alpha}{2}} \cdot \hat{s}_{\mu}^2 \leq \pi \leq Z_{1-\frac{\alpha}{2}} \cdot \hat{s}_{\mu}^2$$

# Z-score



A média aritmética  $\bar{m}$  é um estimador não enviesado da média populacional  $\mu$  e pode ser calculada de N amostras:

$$\bar{m} = \frac{1}{N} \sum_N x_i$$

$$\hat{s}^2 = \frac{1}{N-1} \sum_N (x_i - \bar{m})^2$$

$$\hat{s}_{\mu}^2 = \frac{\hat{s}^2}{N}$$

$$N = \hat{s}^2 \left( \frac{t_{(\frac{\alpha}{2}, N-1)}}{\delta^*} \right)^2$$

$$t_{(\frac{\alpha}{2}, N-1)} \cdot \hat{s}_{\mu}^2 \leq \mu \leq t_{(1-\frac{\alpha}{2}, N-1)} \cdot \hat{s}_{\mu}^2$$

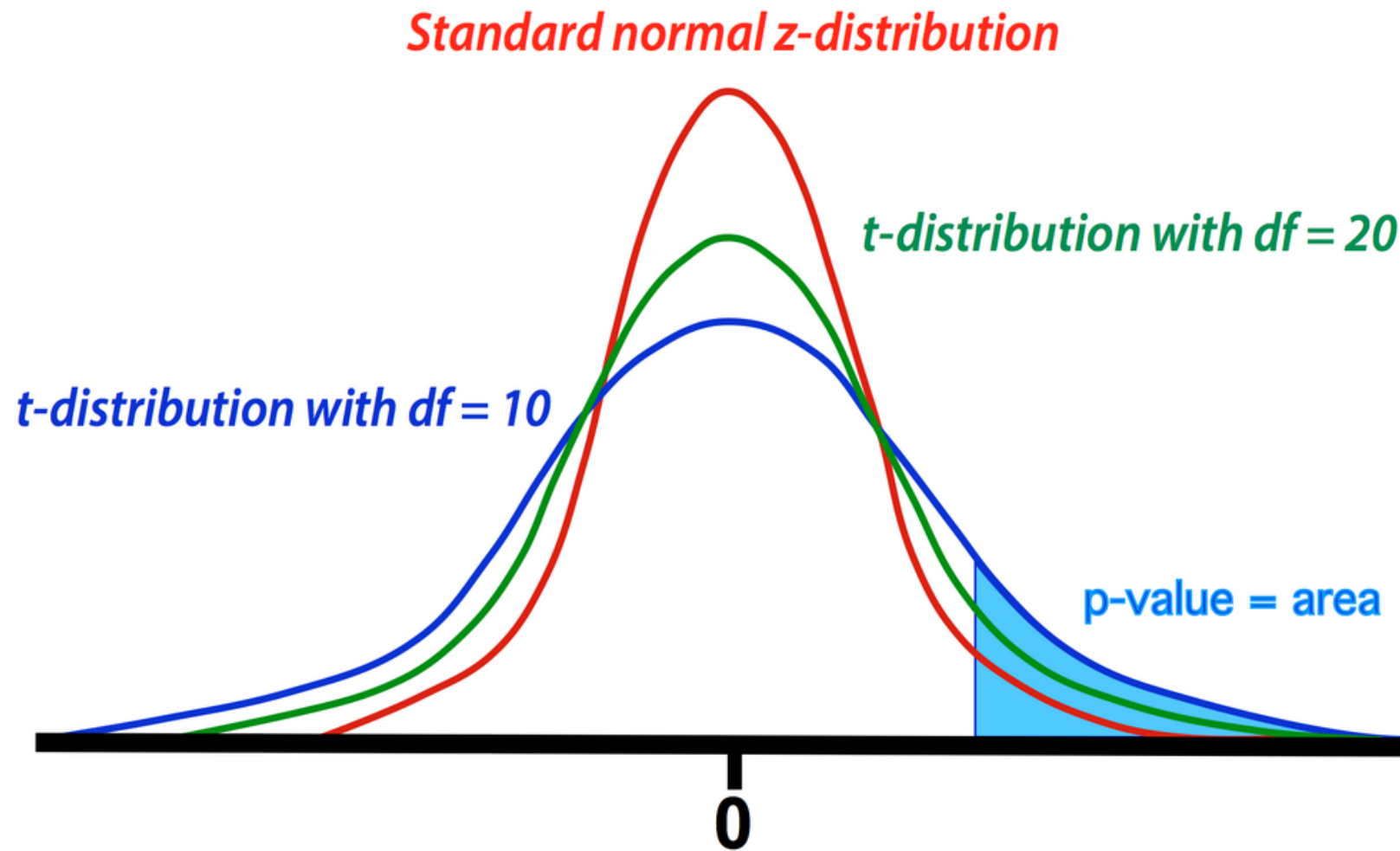
$\alpha$ : proporção da FDP ou significância.

A confiança  $\gamma = 1 - \alpha$ .

$M$  é proporção de erro aceitável (e.g. se erro 3%  $M = 0.003$ ).

$\delta^*$  é o mínimo efeito de interesse, na dimensão da variável (e.g. 5 gramas em medidas de massa).

# Distribuição t-Student



# Desvio padrão e testes

O intervalo de estimativa do desvio padrão

$$\frac{N-1}{\chi^2_{(\frac{\alpha}{2}, N-1)}} \cdot \hat{s}^2 \leq \sigma^2 \leq \frac{N-1}{\chi^2_{(1-\frac{\alpha}{2}, N-1)}} \cdot \hat{s}^2$$

Onde  $\chi^2_{(\alpha, N-1)}$  é a distribuição qui-quadrado na proporção  $\alpha$  com N-1 graus de liberdade.

Métodos de estimativa baseado em subamostragens

- *bootstrap*
- *jack knife*

# Fim de Estimação de parâmetros - Dever de casa

---

## **Exercícios do livro Agresti (2018):**

- Capítulo 5: 5.1, 5.9, 5.24, 5.25, 5.33, 5.34, 5.36, 5.40.





# Assunto

Introdução

Introdução

Estatística Descritiva

Probabilidades

Estimação de Parâmetros

**Teste de Significância**

Comparação de dois grupos

Associação de Variáveis Categóricas

Regressão Linear e Correlação

Relação Multivariável

Regressão Múltipla e Correlação

Análise de Variância - ANOVA

Preditores Quantitativos e Categóricos

Modelos com Regressão Múltipla

Regressão Logística

Introdução a métodos aprofundados

Encerramento

Referências

# Assunto

Introdução

Introdução

Estatística Descritiva

Probabilidades

Estimação de Parâmetros

Teste de Significância

**Comparação de dois grupos**

Associação de Variáveis Categóricas

Regressão Linear e Correlação

Relação Multivariável

Regressão Múltipla e Correlação

Análise de Variância - ANOVA

Preditores Quantitativos e Categóricos

Modelos com Regressão Múltipla

Regressão Logística

Introdução a métodos aprofundados

Encerramento

Referências



# ▶ Dever de casa

---



# Assunto

Introdução

Introdução

Estatística Descritiva

Probabilidades

Estimação de Parâmetros

Teste de Significância

Comparação de dois grupos

**Associação de Variáveis Categóricas**

Regressão Linear e Correlação

Relação Multivariável

Regressão Múltipla e Correlação

Análise de Variância - ANOVA

Preditores Quantitativos e Categóricos

Modelos com Regressão Múltipla

Regressão Logística

Introdução a métodos aprofundados

Encerramento

Referências





Introdução

Introdução

Estatística Descritiva

Probabilidades

Estimação de Parâmetros

Teste de Significância

Comparação de dois grupos

Associação de Variáveis Categóricas

**Regressão Linear e Correlação**

Relação Multivariável

Regressão Múltipla e Correlação

Análise de Variância - ANOVA

Preditores Quantitativos e Categóricos

Modelos com Regressão Múltipla

Regressão Logística

Introdução a métodos aprofundados

Encerramento

Referências





# ▶ Dever de casa

---



# Assunto

Introdução

Introdução

Estatística Descritiva

Probabilidades

Estimação de Parâmetros

Teste de Significância

Comparação de dois grupos

Associação de Variáveis Categóricas

Regressão Linear e Correlação

## Relação Multivariável

Regressão Múltipla e Correlação

Análise de Variância - ANOVA

Preditores Quantitativos e Categóricos

Modelos com Regressão Múltipla

Regressão Logística

Introdução a métodos aprofundados

Encerramento

Referências

# ▶ Dever de casa

---





Introdução

Introdução

Estatística Descritiva

Probabilidades

Estimação de Parâmetros

Teste de Significância

Comparação de dois grupos

Associação de Variáveis Categóricas

Regressão Linear e Correlação

Relação Multivariável

**Regressão Múltipla e Correlação**

Análise de Variância - ANOVA

Preditores Quantitativos e Categóricos

Modelos com Regressão Múltipla

Regressão Logística

Introdução a métodos aprofundados

Encerramento

Referências

# ▶ Dever de casa

---







Introdução

Introdução

Estatística Descritiva

Probabilidades

Estimação de Parâmetros

Teste de Significância

Comparação de dois grupos

Associação de Variáveis Categóricas

Regressão Linear e Correlação

Relação Multivariável

Regressão Múltipla e Correlação

**Análise de Variância - ANOVA**

Preditores Quantitativos e Categóricos

Modelos com Regressão Múltipla

Regressão Logística

Introdução a métodos aprofundados

Encerramento

Referências

# ▶ Dever de casa

---



# ▶ Dever de casa

---



# Assunto

Introdução

Introdução

Estatística Descritiva

Probabilidades

Estimação de Parâmetros

Teste de Significância

Comparação de dois grupos

Associação de Variáveis Categóricas

Regressão Linear e Correlação

Relação Multivariável

Regressão Múltipla e Correlação

Análise de Variância - ANOVA

**Preditores Quantitativos e Categóricos**

Modelos com Regressão Múltipla

Regressão Logística

Introdução a métodos aprofundados

Encerramento

Referências

# ▶ Dever de casa

---





# Assunto

Introdução

Introdução

Estatística Descritiva

Probabilidades

Estimação de Parâmetros

Teste de Significância

Comparação de dois grupos

Associação de Variáveis Categóricas

Regressão Linear e Correlação

Relação Multivariável

Regressão Múltipla e Correlação

Análise de Variância - ANOVA

Preditores Quantitativos e Categóricos

**Modelos com Regressão Múltipla**

Regressão Logística

Introdução a métodos aprofundados

Encerramento

Referências







# Assunto

Introdução

Introdução

Estatística Descritiva

Probabilidades

Estimação de Parâmetros

Teste de Significância

Comparação de dois grupos

Associação de Variáveis Categóricas

Regressão Linear e Correlação

Relação Multivariável

Regressão Múltipla e Correlação

Análise de Variância - ANOVA

Preditores Quantitativos e Categóricos

Modelos com Regressão Múltipla

**Regressão Logística**

Introdução a métodos aprofundados

Encerramento

Referências





# Assunto

Introdução

Introdução

Estatística Descritiva

Probabilidades

Estimação de Parâmetros

Teste de Significância

Comparação de dois grupos

Associação de Variáveis Categóricas

Regressão Linear e Correlação

Relação Multivariável

Regressão Múltipla e Correlação

Análise de Variância - ANOVA

Preditores Quantitativos e Categóricos

Modelos com Regressão Múltipla

Regressão Logística

**Introdução a métodos aprofundados**

Encerramento

Referências







# Assunto

Introdução

Introdução

Estatística Descritiva

Probabilidades

Estimação de Parâmetros

Teste de Significância

Comparação de dois grupos

Associação de Variáveis Categóricas

Regressão Linear e Correlação

Relação Multivariável

Regressão Múltipla e Correlação

Análise de Variância - ANOVA

Preditores Quantitativos e Categóricos

Modelos com Regressão Múltipla

Regressão Logística

Introdução a métodos aprofundados

**Encerramento**

Referências





# Sobre este material

Esta obra está licenciada sob a licença *Creative Commons* CC BY-NC-SA 4.0 (mais detalhes neste *link*)

Favor fazer referência a este trabalho como:

Silva, A. P. (2022), *Notas de Aulas de Estatística para Linguística*. Online:  
<https://github.com/adelinocpp/estatistica-para-linguistica>

```
@Misc{Silva2022,  
  title={Notas de Aulas de Notas de Aulas de Estatística para Linguística},  
  author={Adelino Pinheiro Silva},  
  howPublished={\url{https://github.com/adelinocpp/estatistica-para-linguistica}},  
  year={2022},  
  note={Version 1.0; Creative Commons BY-NC-SA 4.0.},  
}
```

# Assunto

Introdução

Introdução

Estatística Descritiva

Probabilidades

Estimação de Parâmetros

Teste de Significância

Comparação de dois grupos

Associação de Variáveis Categóricas

Regressão Linear e Correlação

Relação Multivariável

Regressão Múltipla e Correlação

Análise de Variância - ANOVA

Preditores Quantitativos e Categóricos

Modelos com Regressão Múltipla

Regressão Logística

Introdução a métodos aprofundados

Encerramento

**Referências**

# Referências I

- Agresti, A. (2018). *Statistical methods for the social sciences*. Number 300.72 A3. Pearson.
- Casella, G. and Berger, R. L. (2011). Inferência estatística-tradução da 2a edição norte-americana. *Centage Learning*, page 259.
- Chomsky, N. (2009). *Syntactic structures*. De Gruyter Mouton.
- Halperin, M., Hartley, H. O., and Hoel, P. G. (1965). Recommended standards for statistical symbols and notation: Copss committee on symbols and notation. *The American Statistician*, 19(3):12–14.
- Levshina, N. (2015). *How to do linguistics with R*. Université catholique de Louvain.
- Wikipédia (2017). Notação em probabilidade e estatística. <https://pt.wikipedia.org/wiki/Nota> Acessado em 19/03/2022.