

Introdução à Estatística

Bernardo dos Santos, MSc.



18 de abril de 2024

Roteiro

- 1 Introdução
- 2 Planejamento
- 3 Análise Exploratória
- 4 Inferência Estatística

Roteiro

- 1 **Introdução**
- 2 Planejamento
- 3 Análise Exploratória
- 4 Inferência Estatística

Introdução

Estatística é uma coleção de métodos para planejar experimentos, obter dados e tirar conclusões sobre eles.

Introdução

Estatística é uma coleção de métodos para planejar experimentos, obter dados e tirar conclusões sobre eles.

- Planejamento (tipo de estudo, tamanho da amostra)

Introdução

Estatística é uma coleção de métodos para planejar experimentos, obter dados e tirar conclusões sobre eles.

- Planejamento (tipo de estudo, tamanho da amostra)
- Coleta de dados (amostragem)

Introdução

Estatística é uma coleção de métodos para planejar experimentos, obter dados e tirar conclusões sobre eles.

- Planejamento (tipo de estudo, tamanho da amostra)
- Coleta de dados (amostragem)
- Análise

Introdução

Estatística é uma coleção de métodos para planejar experimentos, obter dados e tirar conclusões sobre eles.

- Planejamento (tipo de estudo, tamanho da amostra)
- Coleta de dados (amostragem)
- Análise
- Interpretação

População e amostra

População e amostra

População Todos os indivíduos que podem fazer parte da pesquisa e sobre quem se quer tirar conclusões.

População e amostra

População Todos os indivíduos que podem fazer parte da pesquisa e sobre quem se quer tirar conclusões.

Amostra Uma parte da população que de fato faz parte da pesquisa.

População e amostra

- População** Todos os indivíduos que podem fazer parte da pesquisa e sobre quem se quer tirar conclusões.
- Amostra** Uma parte da população que de fato faz parte da pesquisa.
- Variável** Qualquer característica da amostra/população.

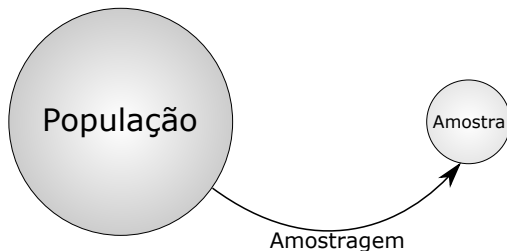
População e amostra

- População** Todos os indivíduos que podem fazer parte da pesquisa e sobre quem se quer tirar conclusões.
- Amostra** Uma parte da população que de fato faz parte da pesquisa.
- Variável** Qualquer característica da amostra/população.



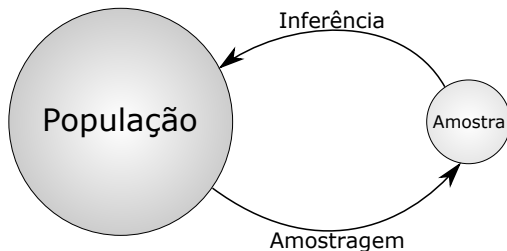
População e amostra

- População** Todos os indivíduos que podem fazer parte da pesquisa e sobre quem se quer tirar conclusões.
- Amostra** Uma parte da população que de fato faz parte da pesquisa.
- Variável** Qualquer característica da amostra/população.



População e amostra

- População** Todos os indivíduos que podem fazer parte da pesquisa e sobre quem se quer tirar conclusões.
- Amostra** Uma parte da população que de fato faz parte da pesquisa.
- Variável** Qualquer característica da amostra/população.



Tipos de variáveis

Segundo sua natureza

Tipos de variáveis

Segundo sua natureza

Variáveis categóricas separam as unidades amostrais em grupos

Tipos de variáveis

Segundo sua natureza

Variáveis categóricas separam as unidades amostrais em grupos

Nominal As categorias não têm ordem entre si
(*e.g., gênero, tratamento, estado civil*)

Tipos de variáveis

Segundo sua natureza

Variáveis categóricas separam as unidades amostrais em grupos

Nominal As categorias não têm ordem entre si
(*e.g., gênero, tratamento, estado civil*)

Se tiver duas categorias, também pode ser chamada de binária ou dicotômica

Tipos de variáveis

Segundo sua natureza

Variáveis categóricas separam as unidades amostrais em grupos

Nominal As categorias não têm ordem entre si
(*e.g., gênero, tratamento, estado civil*)

Se tiver duas categorias, também pode ser chamada de binária ou dicotômica

Ordinais As categorias formam uma ordem
(*e.g., escolaridade, gravidade da doença*)

Tipos de variáveis

Segundo sua natureza

Variáveis categóricas separam as unidades amostrais em grupos

Nominal As categorias não têm ordem entre si
(*e.g., gênero, tratamento, estado civil*)

Se tiver duas categorias, também pode ser chamada de binária ou dicotômica

Ordinais As categorias formam uma ordem
(*e.g., escolaridade, gravidade da doença*)

Variáveis numéricas quantificam uma característica da unidade amostral

Tipos de variáveis

Segundo sua natureza

Variáveis categóricas separam as unidades amostrais em grupos

Nominal As categorias não têm ordem entre si
(*e.g., gênero, tratamento, estado civil*)

Se tiver duas categorias, também pode ser chamada de binária ou dicotômica

Ordinais As categorias formam uma ordem
(*e.g., escolaridade, gravidade da doença*)

Variáveis numéricas quantificam uma característica da unidade amostral

Discretas Há um conjunto pré determinado de valores possíveis
(*e.g., quantidade de filhos, renda*)

Tipos de variáveis

Segundo sua natureza

Variáveis categóricas separam as unidades amostrais em grupos

Nominal As categorias não têm ordem entre si
(*e.g., gênero, tratamento, estado civil*)

Se tiver duas categorias, também pode ser chamada de binária ou dicotômica

Ordinais As categorias formam uma ordem
(*e.g., escolaridade, gravidade da doença*)

Variáveis numéricas quantificam uma característica da unidade amostral

Discretas Há um conjunto pré determinado de valores possíveis
(*e.g., quantidade de filhos, renda*)

Contínuas Qualquer número real dentro de um intervalo
(*e.g., IMC, tempo de doença*)

Tipos de variáveis

Segundo sua função

Tipos de variáveis

Segundo sua função

Variável resposta

Também chamada de variável dependente ou desfecho. É aquela sobre a qual se faz inferências e é influenciada pelas variáveis explicativas.

Tipos de variáveis

Segundo sua função

Variável resposta

Também chamada de variável dependente ou desfecho. É aquela sobre a qual se faz inferências e é influenciada pelas variáveis explicativas.

Variável explicativa

Também chamada de variável independente ou preditor. São as variáveis que podem modificar a resposta. Podem ou não serem controladas a priori.

Distribuição de probabilidades

Distribuição de probabilidades

Definição

É uma função que descreve a probabilidade de ocorrência de cada possível resultado do experimento.

Distribuição de probabilidades

Definição

É uma função que descreve a probabilidade de ocorrência de cada possível resultado do experimento.

Para variáveis discretas, define as probabilidades de cada evento individualmente.

Distribuição de probabilidades

Definição

É uma função que descreve a probabilidade de ocorrência de cada possível resultado do experimento.

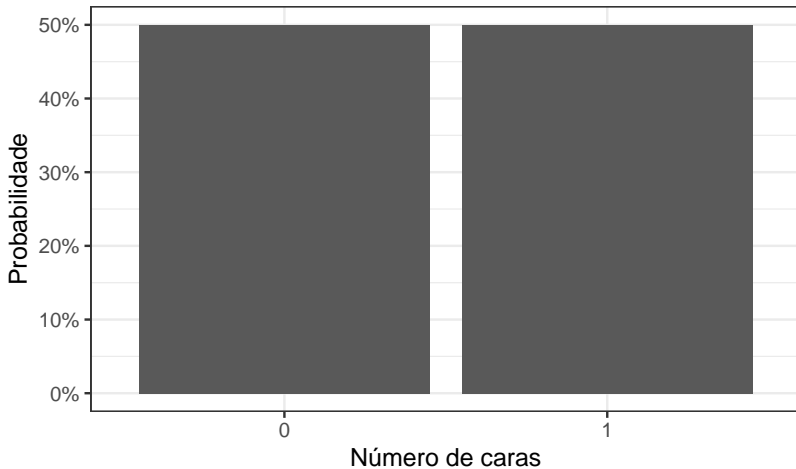
Para variáveis discretas, define as probabilidades de cada evento individualmente.

Para variáveis contínuas, a probabilidade é descrita em intervalos.

Distribuição de probabilidades

Binomial

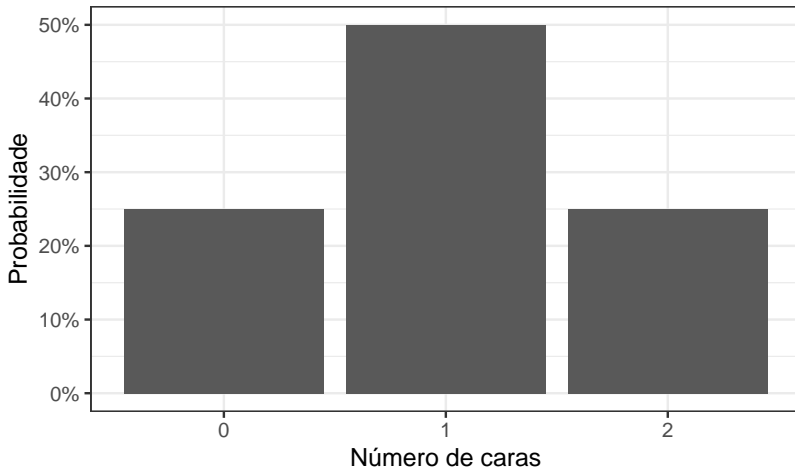
Lançamento de uma moeda



Distribuição de probabilidades

Binomial

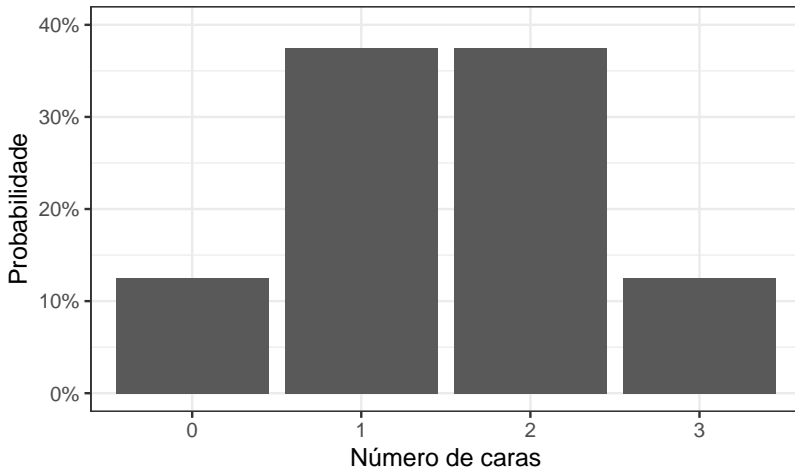
Lançamento de duas moedas



Distribuição de probabilidades

Binomial

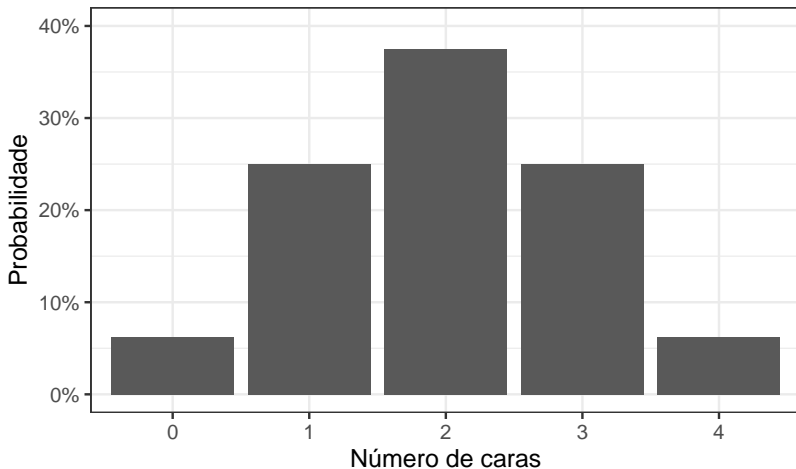
Lançamento de três moedas



Distribuição de probabilidades

Binomial

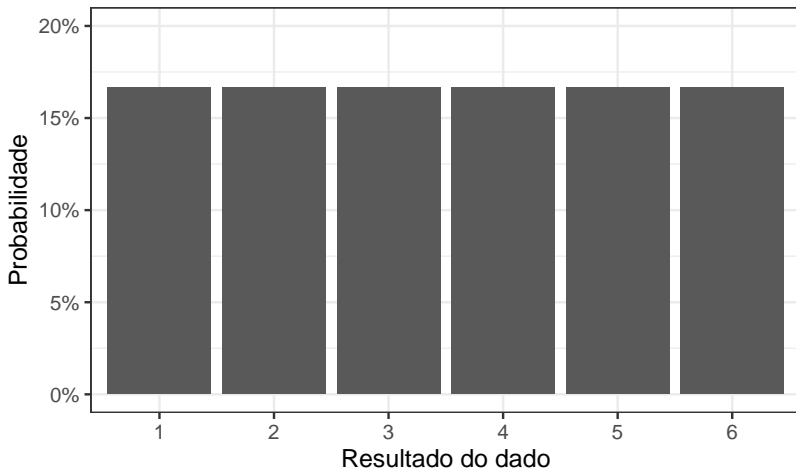
Lançamento de quatro moedas



Distribuição de probabilidades

Uniforme discreta

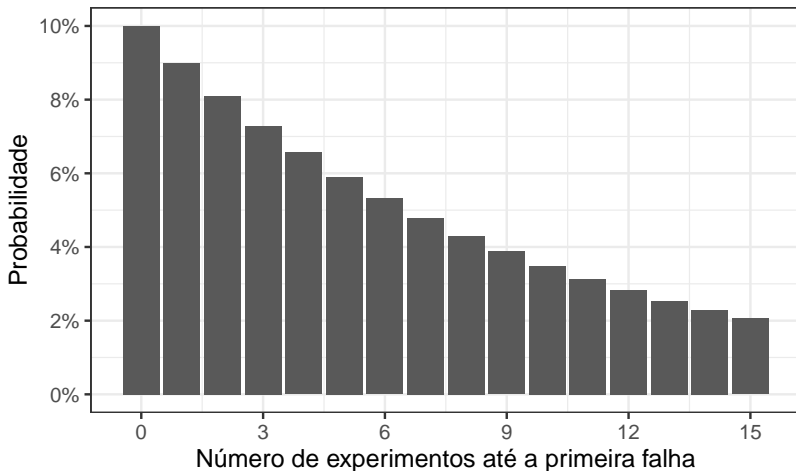
Lançamento de um dado



Distribuição de probabilidades

Geométrica

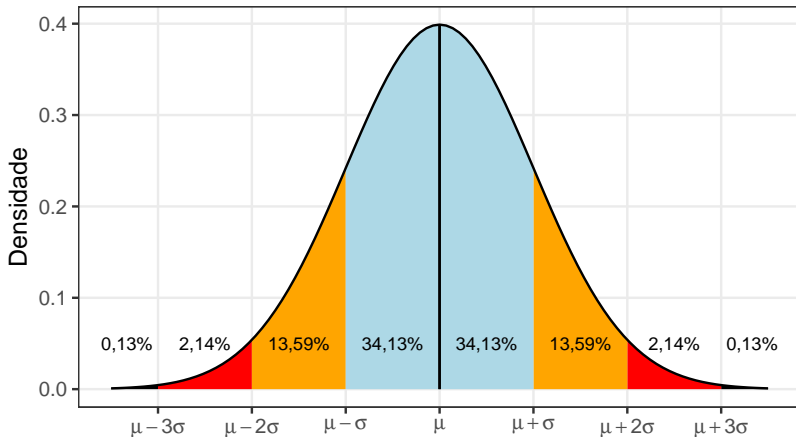
Experimento com taxa de falha 10%



Distribuição de probabilidades

Normal (gaussiana)

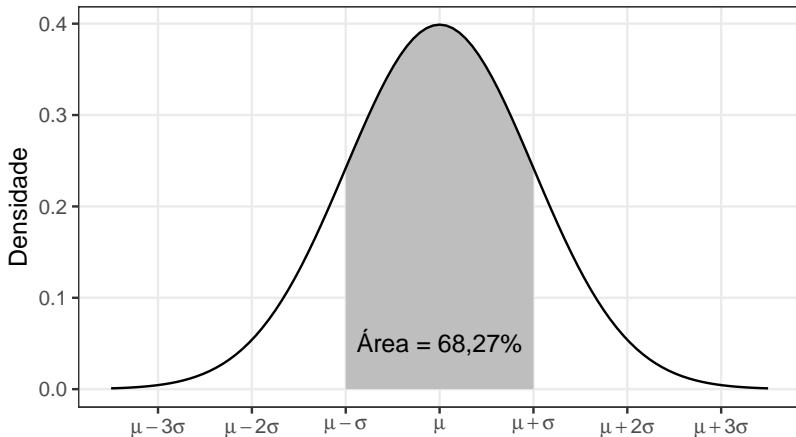
Distribuição normal com média μ e desvio-padrão σ



Distribuição de probabilidades

Normal (gaussiana)

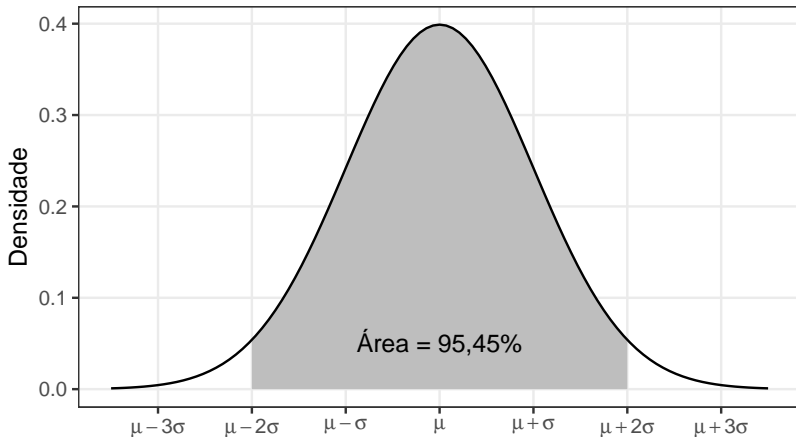
Distribuição normal com média μ e desvio-padrão σ



Distribuição de probabilidades

Normal (gaussiana)

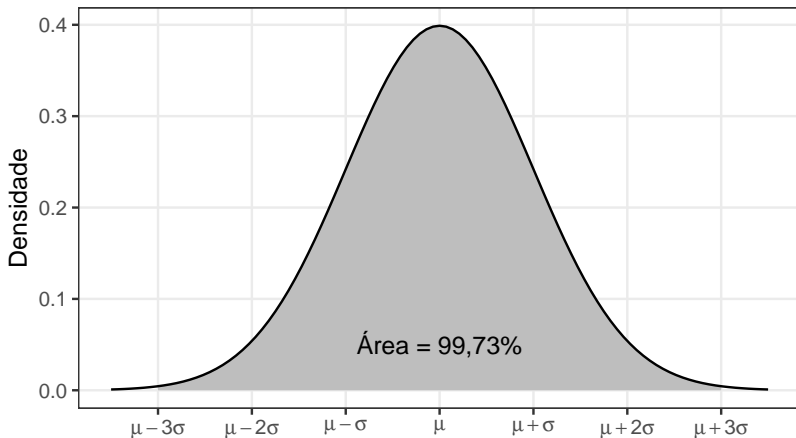
Distribuição normal com média μ e desvio-padrão σ



Distribuição de probabilidades

Normal (gaussiana)

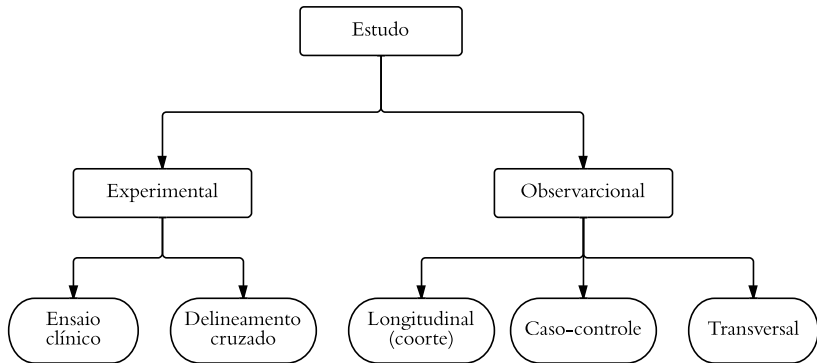
Distribuição normal com média μ e desvio-padrão σ



Roteiro

- 1 Introdução
- 2 Planejamento**
- 3 Análise Exploratória
- 4 Inferência Estatística

Tipos de estudo



Dimensionamento amostral

Dimensionamento amostral

Depende de:

- Menor efeito a ser detectado

Dimensionamento amostral

Depende de:

- Menor efeito a ser detectado
 - Teste a ser utilizado

Dimensionamento amostral

Depende de:

- Menor efeito a ser detectado
 - Teste a ser utilizado
 - Tipo de estudo

Dimensionamento amostral

Depende de:

- Menor efeito a ser detectado
 - Teste a ser utilizado
 - Tipo de estudo
 - Tipo do desfecho principal

Dimensionamento amostral

Depende de:

- Menor efeito a ser detectado
 - Teste a ser utilizado
 - Tipo de estudo
 - Tipo do desfecho principal
- Erros tipo I e II desejados

Dimensionamento amostral

Depende de:

- Menor efeito a ser detectado
 - Teste a ser utilizado
 - Tipo de estudo
 - Tipo do desfecho principal
- Erros tipo I e II desejados

Dimensionamento amostral

Depende de:

- Menor efeito a ser detectado
 - Teste a ser utilizado
 - Tipo de estudo
 - Tipo do desfecho principal
- Erros tipo I e II desejados

Voltaremos nisso!

Amostragem

Amostragem

Amostra aleatória simples

A amostra é coletada ao acaso entre todos os possíveis participantes. Cada indivíduo da população tem a mesma probabilidade de ser selecionado.

Amostragem

Amostra aleatória simples

A amostra é coletada ao acaso entre todos os possíveis participantes. Cada indivíduo da população tem a mesma probabilidade de ser selecionado.

Amostra estratificada

Grupos de indivíduos semelhantes (estratos) são formados e a amostra é coletada dentro de cada estrato. A alocação da amostra pode ser uniforme (mesma amostra para cada estrato), proporcional (amostra maior para estratos maiores) ou otimizada (amostras maiores para estratos mais heterogêneos). Quanto mais homogêneos são os estratos, mais confiável é o resultado final.

Amostragem

Amostra por conglomerados (*cluster*)

Os indivíduos são divididos em grupos heterogêneos (conglomerados) que representam a população toda. A amostra pode ser um conglomerado inteiro (um estágio) ou ainda uma amostra dentro do conglomerado (dois estágios). Quanto mais diversificado forem os conglomerados, mais confiável é o resultado final.

Amostragem

Amostra por conglomerados (*cluster*)

Os indivíduos são divididos em grupos heterogêneos (conglomerados) que representam a população toda. A amostra pode ser um conglomerado inteiro (um estágio) ou ainda uma amostra dentro do conglomerado (dois estágios). Quanto mais diversificados forem os conglomerados, mais confiável é o resultado final.

Amostra de conveniência

A amostra é composta por todos os indivíduos disponíveis da população alvo.

Roteiro

- 1 Introdução
- 2 Planejamento
- 3 Análise Exploratória**
- 4 Inferência Estatística

Medidas resumo

Variáveis categóricas

Medidas resumo

Variáveis categóricas

Tabela de frequências

Sexo	Frequência absoluta	Frequência relativa
Masculino	28	56,0%
Feminino	22	44,0%

Medidas resumo

Variáveis categóricas

Tabela de frequências

Sexo	Frequência absoluta	Frequência relativa
Masculino	28	56,0%
Feminino	22	44,0%

Tabela de contingência

	Masculino		Feminino		Total	
	N	%	N	%	N	%
Controle	13	52,0	12	48,0	25	100,0
Experimental	15	60,0	10	40,0	25	100,0
Total	28	56,0	22	44,0	50	100,0

Medidas resumo

Variáveis categóricas

Tabela de frequências

Sexo	Frequência absoluta	Frequência relativa
Masculino	28	56,0%
Feminino	22	44,0%

Tabela de contingência

	Controle		Experimental		Total	
	N	%	N	%	N	%
Masculino	13	46,4	15	53,6	28	100,0
Feminino	12	54,6	10	45,4	22	100,0
Total	25	50,0	25	50,0	50	100,0

Medidas resumo

Variáveis categóricas

Tabela de frequências

Sexo	Frequência absoluta	Frequência relativa
Masculino	28	56,0%
Feminino	22	44,0%

Tabela de contingência

	Controle		Experimental		Total	
	N	%	N	%	N	%
Masculino	13	52,0	15	60,0	28	50,0
Feminino	12	48,0	10	40,0	22	50,0
Total	25	100,0	25	100,0	50	100,0

Gráfico de barras

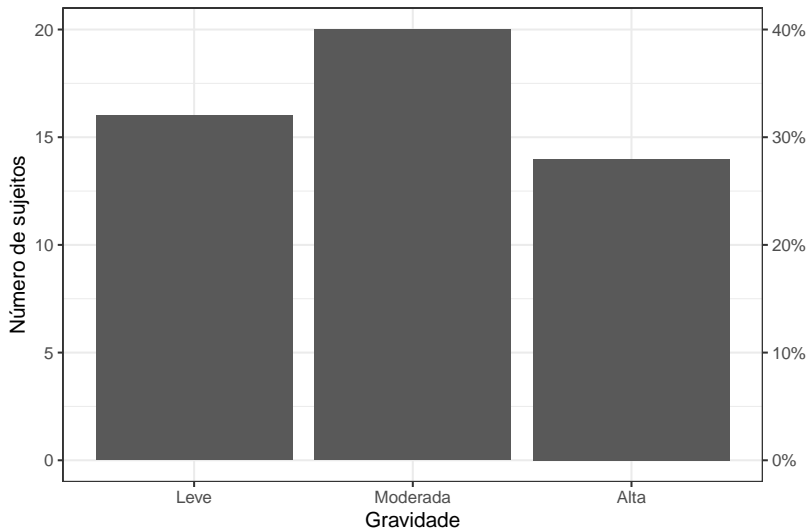


Gráfico de barras

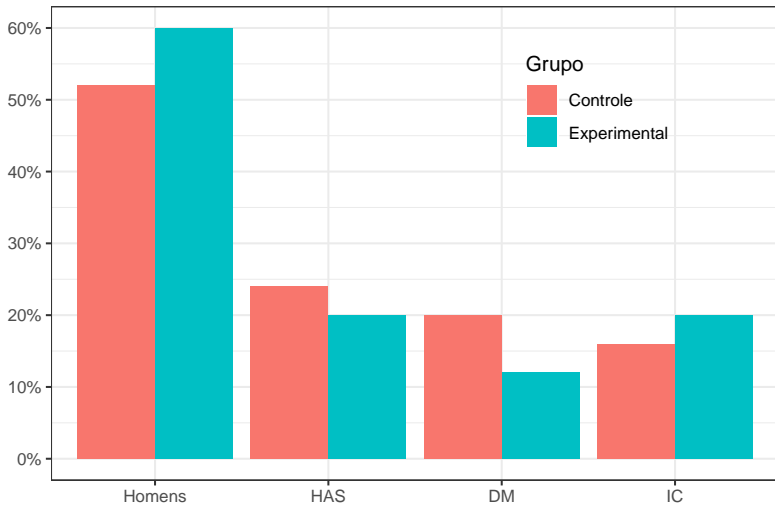


Gráfico de barras empilhadas

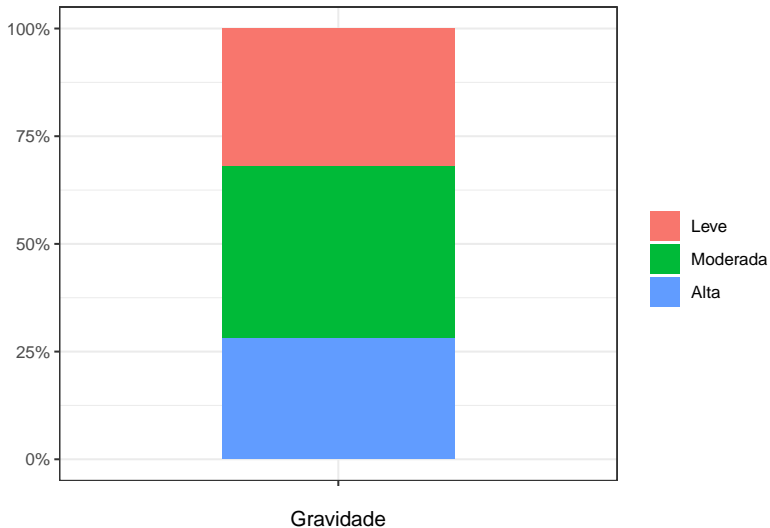


Gráfico de barras empilhadas

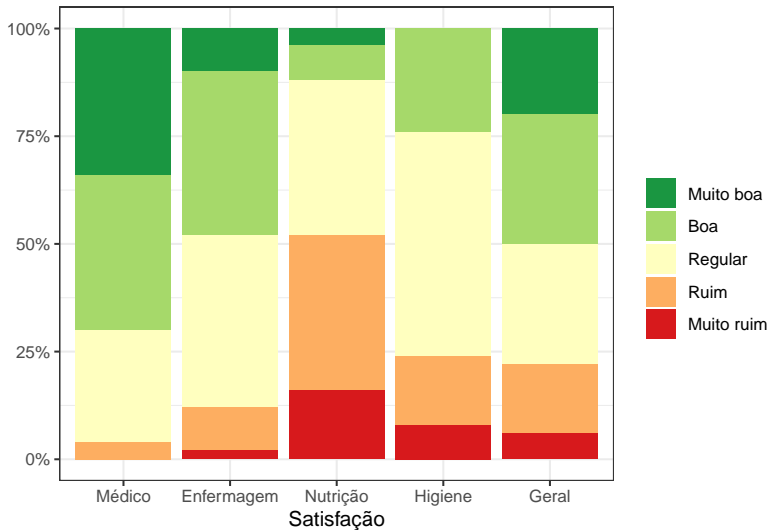


Gráfico de barras empilhadas

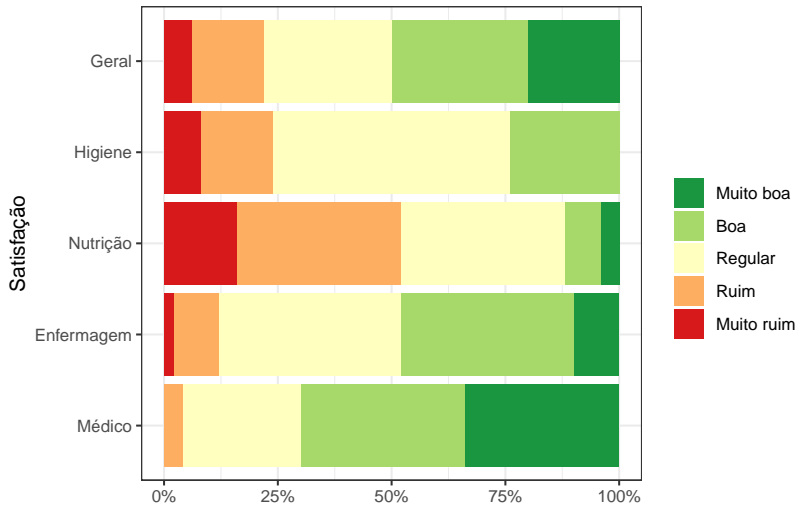


Gráfico de setores

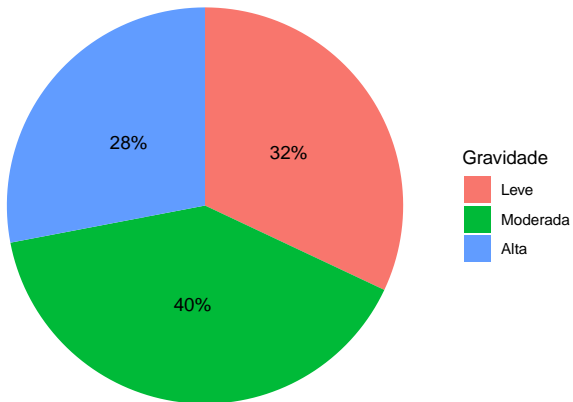


Gráfico de setores

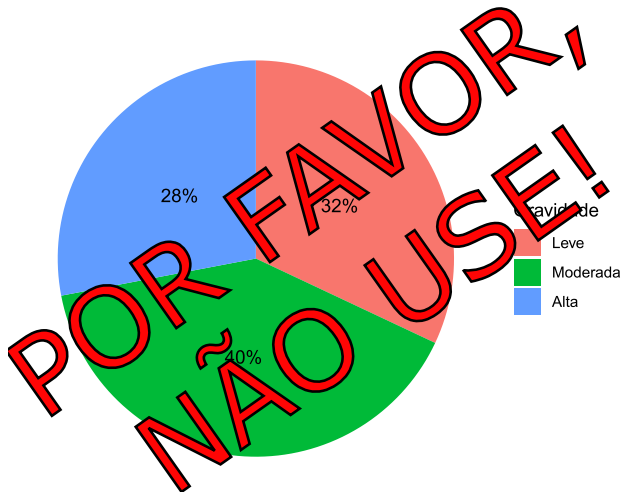


Gráfico de setores

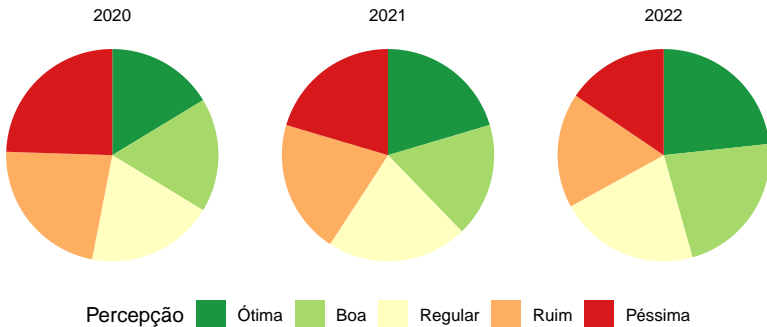


Gráfico de setores – fazendo melhor

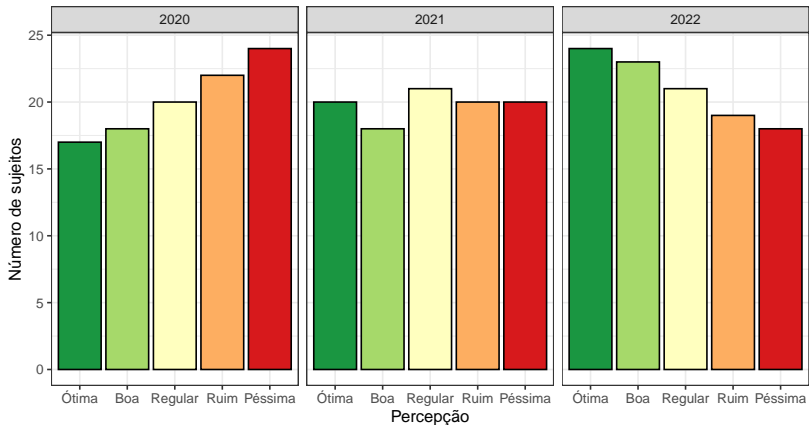
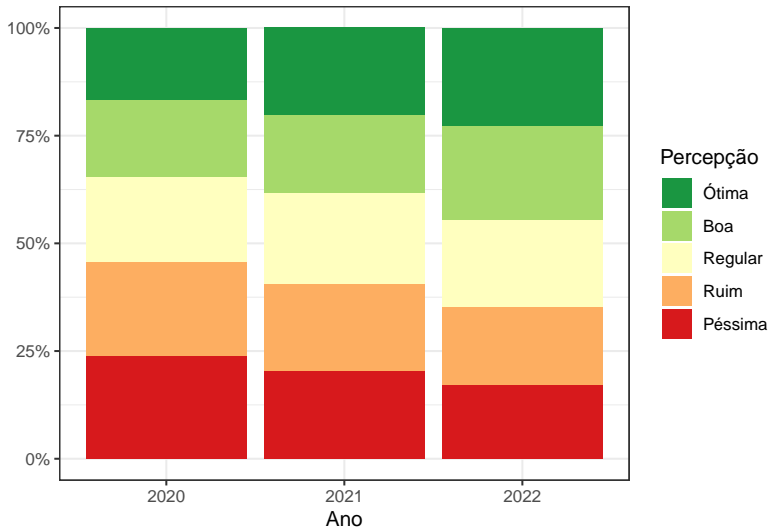


Gráfico de setores – fazendo melhor



Medidas resumo

Variáveis numéricas

Medidas resumo

Variáveis numéricas

$$\text{Média } \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Medidas resumo

Variáveis numéricas

$$\text{Média } \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\text{Desvio-padrão amostral } s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Medidas resumo

Variáveis numéricas

Média $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

Desvio-padrão amostral $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$

Percentil α Valor no qual $\alpha\%$ da amostra seja menor do que ele

Medidas resumo

Variáveis numéricas

$$\text{Média } \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\text{Desvio-padrão amostral } s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Percentil α Valor no qual $\alpha\%$ da amostra seja menor do que ele

$\alpha = 0\%$ Mínimo

Medidas resumo

Variáveis numéricas

$$\text{Média } \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\text{Desvio-padrão amostral } s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Percentil α Valor no qual $\alpha\%$ da amostra seja menor do que ele

$\alpha = 0\%$ Mínimo

$\alpha = 25\%$ Primeiro quartil

Medidas resumo

Variáveis numéricas

$$\text{Média } \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\text{Desvio-padrão amostral } s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Percentil α Valor no qual $\alpha\%$ da amostra seja menor do que ele

$\alpha = 0\%$ Mínimo

$\alpha = 25\%$ Primeiro quartil

$\alpha = 50\%$ Mediana

Medidas resumo

Variáveis numéricas

$$\text{Média } \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\text{Desvio-padrão amostral } s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Percentil α Valor no qual $\alpha\%$ da amostra seja menor do que ele

$\alpha = 0\%$ Mínimo

$\alpha = 25\%$ Primeiro quartil

$\alpha = 50\%$ Mediana

$\alpha = 75\%$ Terceiro quartil

Medidas resumo

Variáveis numéricas

$$\text{Média } \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\text{Desvio-padrão amostral } s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Percentil α Valor no qual $\alpha\%$ da amostra seja menor do que ele

$\alpha = 0\%$ Mínimo

$\alpha = 25\%$ Primeiro quartil

$\alpha = 50\%$ Mediana

$\alpha = 75\%$ Terceiro quartil

$\alpha = 100\%$ Máximo

Medidas resumo

Variáveis numéricas

$$\text{Média } \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\text{Desvio-padrão amostral } s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Percentil α Valor no qual $\alpha\%$ da amostra seja menor do que ele

$\alpha = 0\%$ Mínimo

$\alpha = 25\%$ Primeiro quartil

$\alpha = 50\%$ Mediana

$\alpha = 75\%$ Terceiro quartil

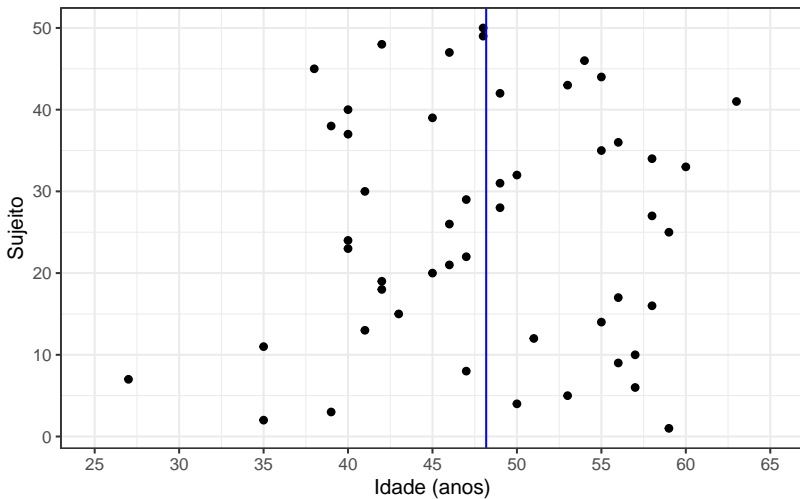
$\alpha = 100\%$ Máximo

	Média	DP	Mín.	Q1	Mediana	Q3	Máx.
Idade	48,18	7,92	27	42	48	55	63

Medidas resumo

Variáveis numéricas

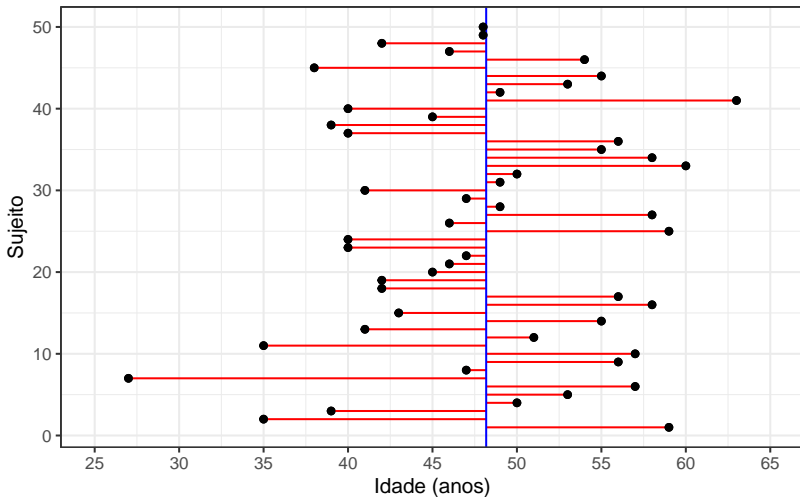
Média = 48,18



Medidas resumo

Variáveis numéricas

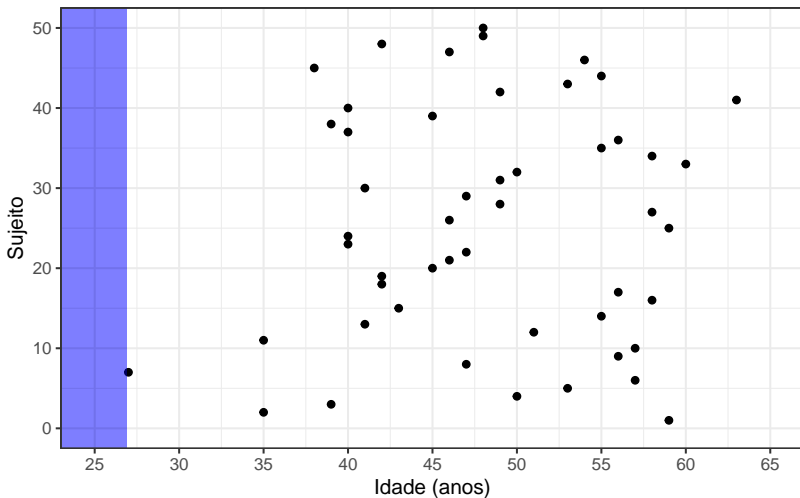
DP = 7,92



Medidas resumo

Variáveis numéricas

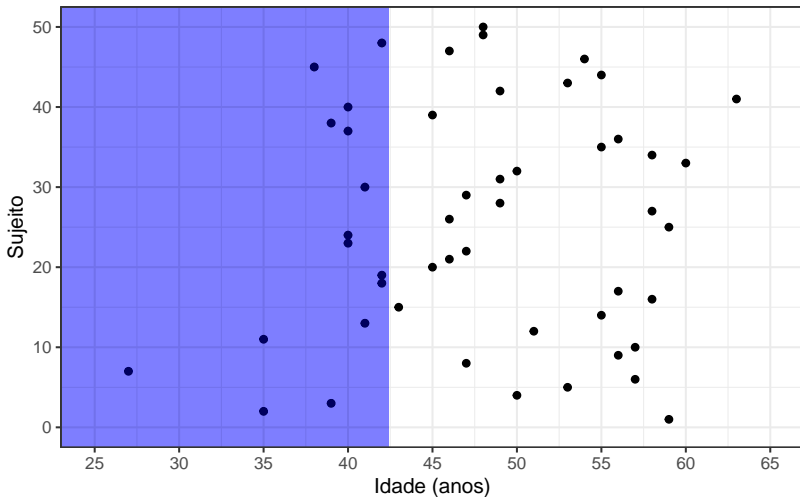
Mínimo = Percentil 0% = 27



Medidas resumo

Variáveis numéricas

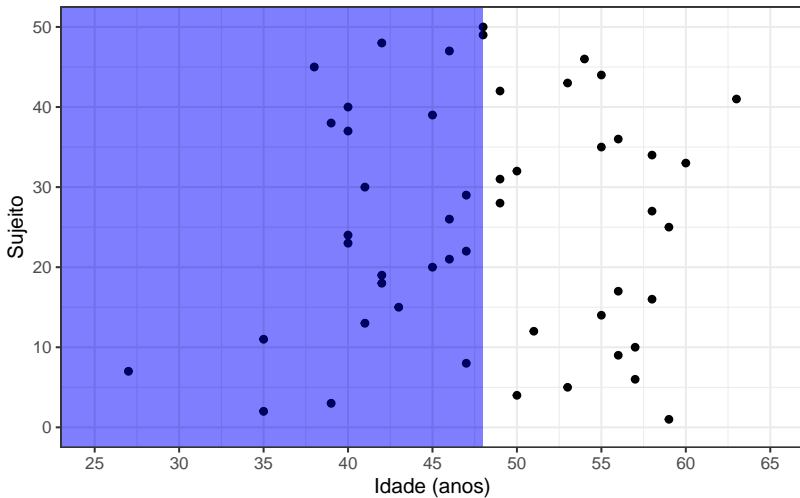
Q1 = Percentil 25% = 42



Medidas resumo

Variáveis numéricas

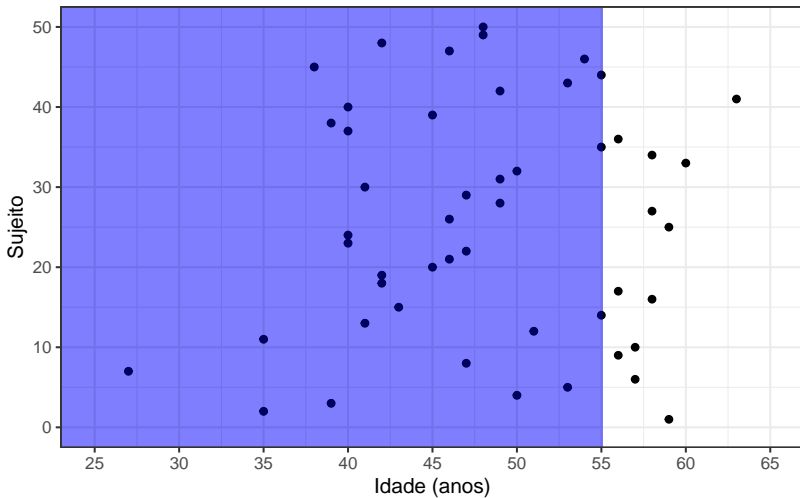
Mediana = Percentil 50% = 48



Medidas resumo

Variáveis numéricas

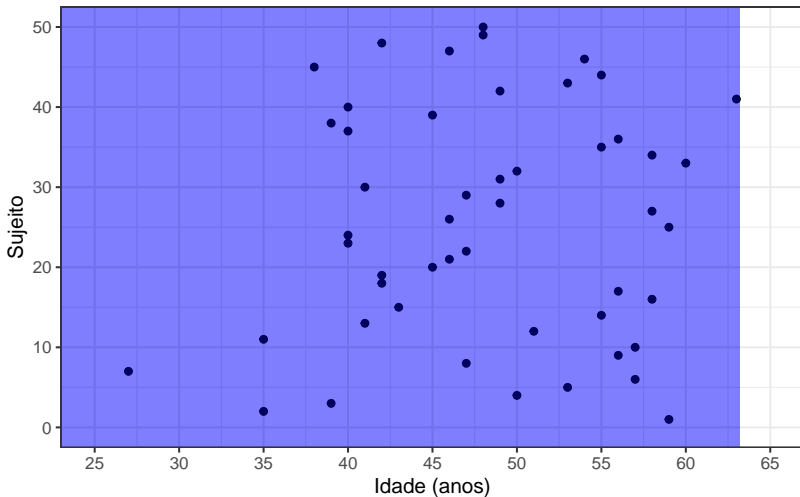
Q3 = Percentil 75% = 55



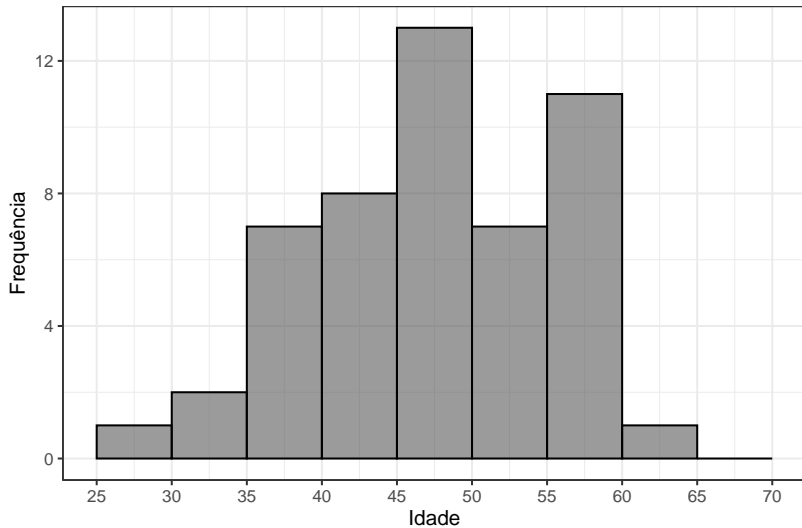
Medidas resumo

Variáveis numéricas

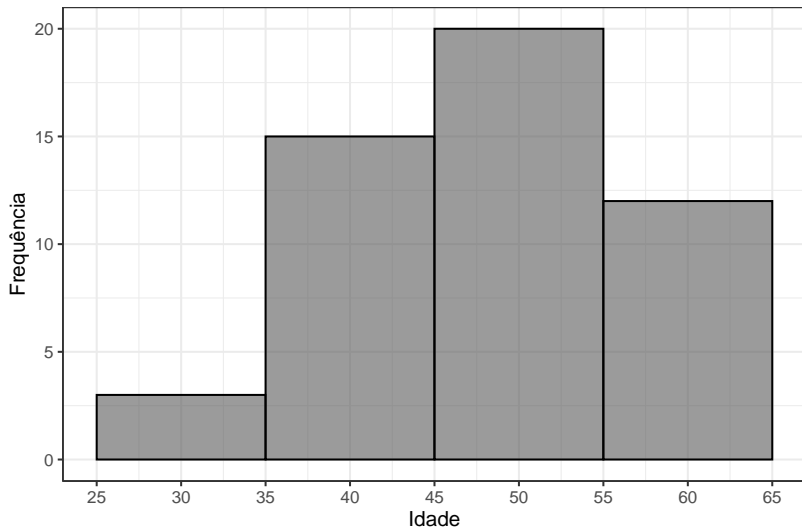
Máximo = Percentil 100% = 63



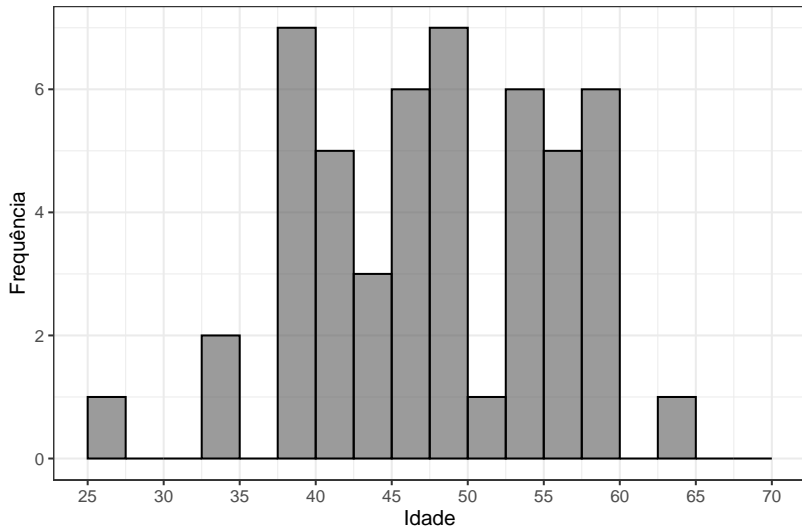
Histograma



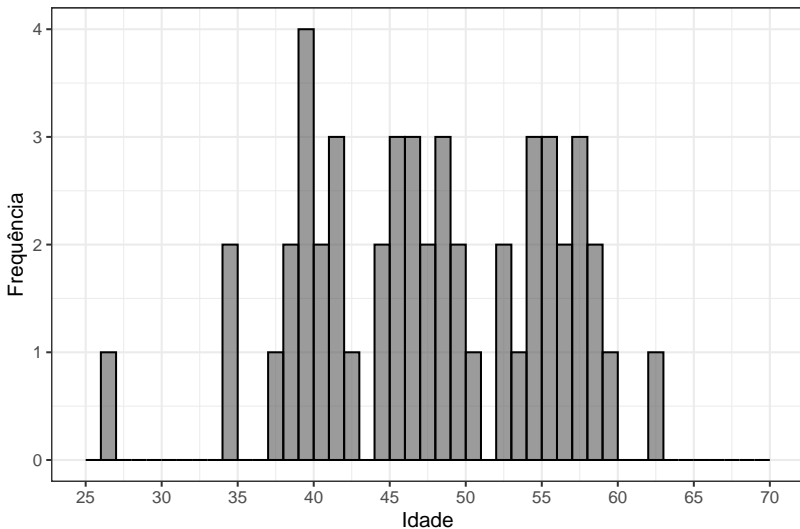
Histograma



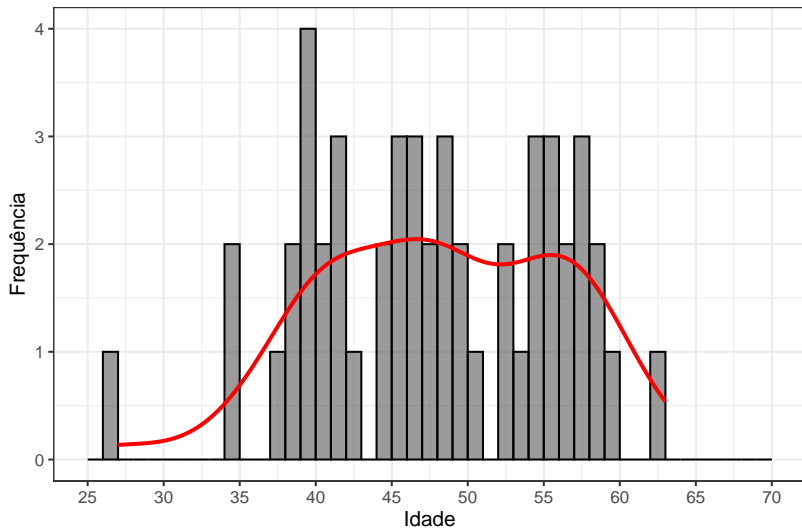
Histograma



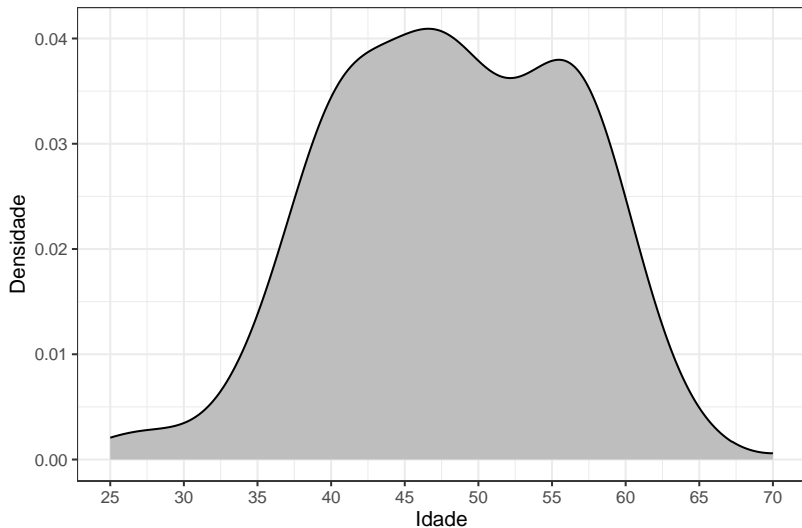
Histograma



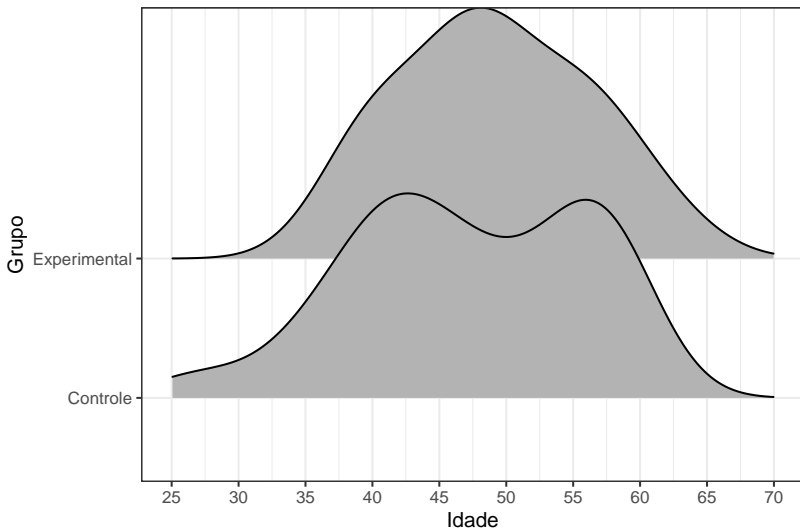
Estimativa de densidade kernel



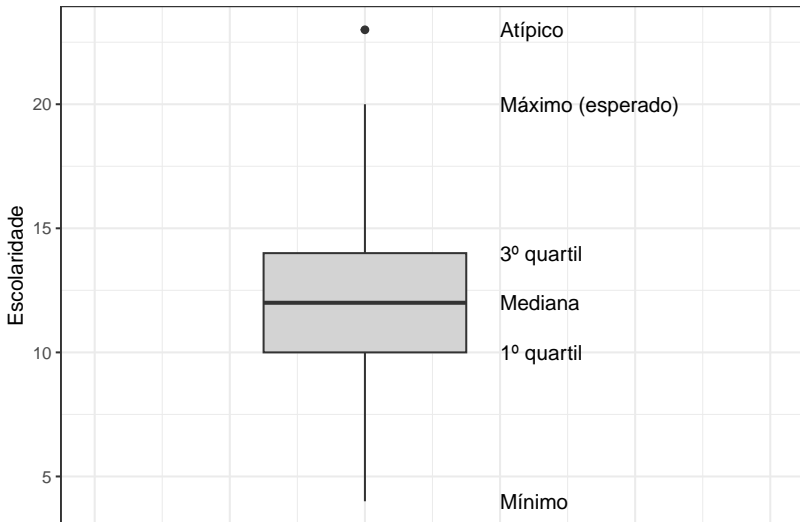
Estimativa de densidade kernel



Estimativa de densidade kernel



Boxplot



Boxplot

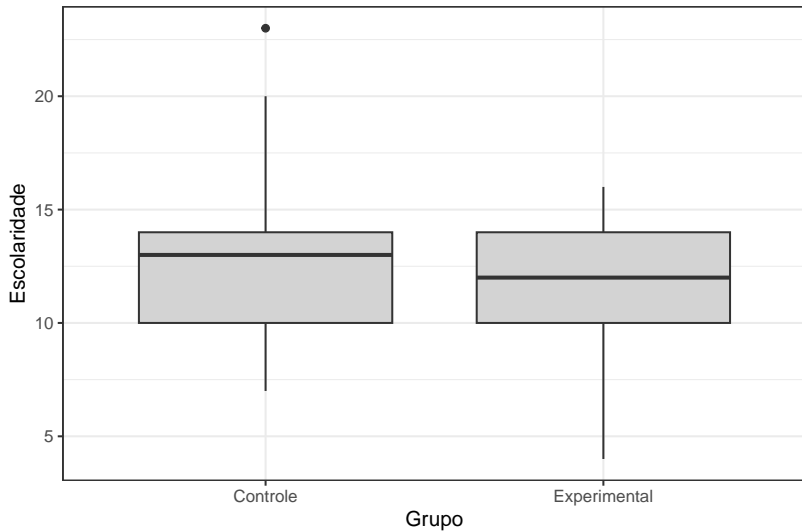
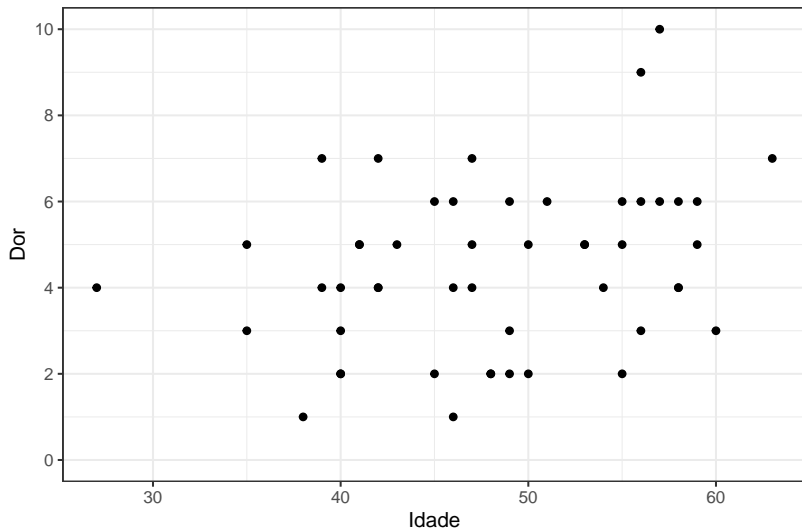


Gráfico de dispersão



Roteiro

- 1 Introdução
- 2 Planejamento
- 3 Análise Exploratória
- 4 Inferência Estatística**

Estimadores

Objetivo

Encontrar o melhor valor possível para um parâmetro populacional baseado em uma amostra.

Estimadores

Objetivo

Encontrar o melhor valor possível para um parâmetro populacional baseado em uma amostra.

Média populacional μ (valor desconhecido)

Estimadores

Objetivo

Encontrar o melhor valor possível para um parâmetro populacional baseado em uma amostra.

Média populacional μ (valor desconhecido)

Estimador pontual média amostral \bar{x}

Estimadores

Objetivo

Encontrar o melhor valor possível para um parâmetro populacional baseado em uma amostra.

Média populacional μ (valor desconhecido)

Estimador pontual média amostral \bar{x}

Erro-padrão medida de imprecisão de um estimador

Estimadores

Objetivo

Encontrar o melhor valor possível para um parâmetro populacional baseado em uma amostra.

Média populacional μ (valor desconhecido)

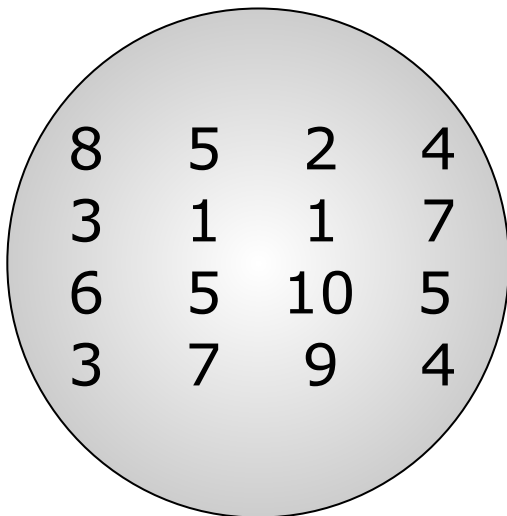
Estimador pontual média amostral \bar{x}

Erro-padrão medida de imprecisão de um estimador

Estimador intervalar intervalo de confiança

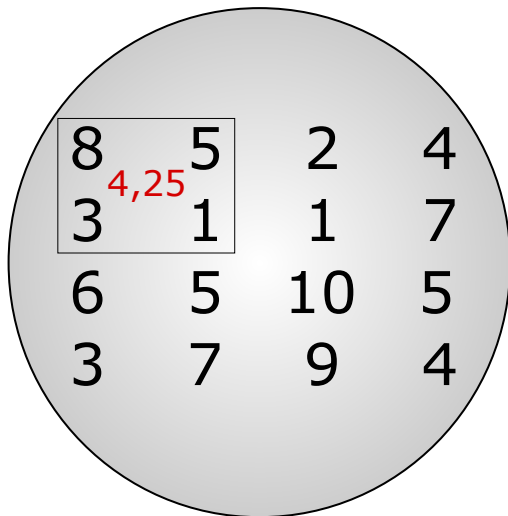
Média populacional e estimadores

Exemplo: notas de uma prova



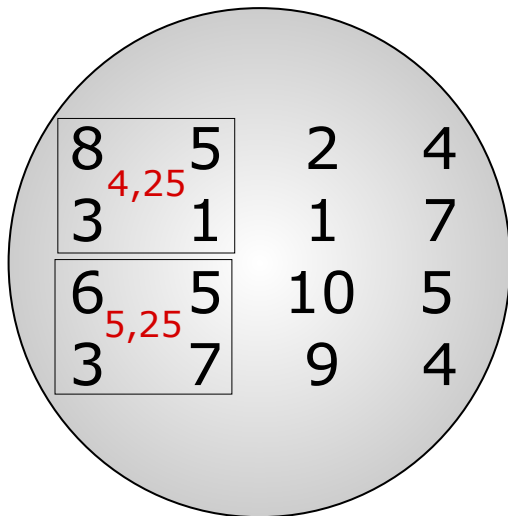
Média populacional e estimadores

Exemplo: notas de uma prova



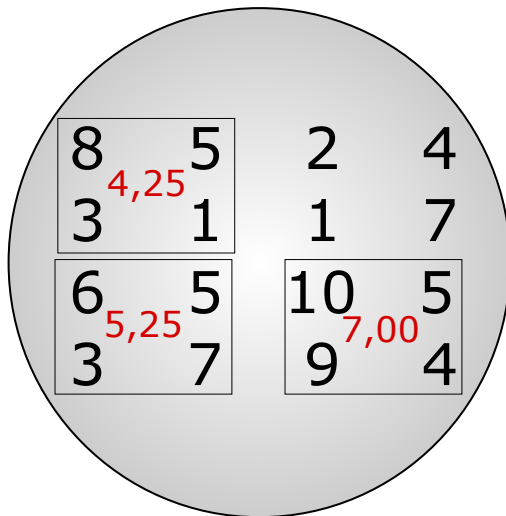
Média populacional e estimadores

Exemplo: notas de uma prova



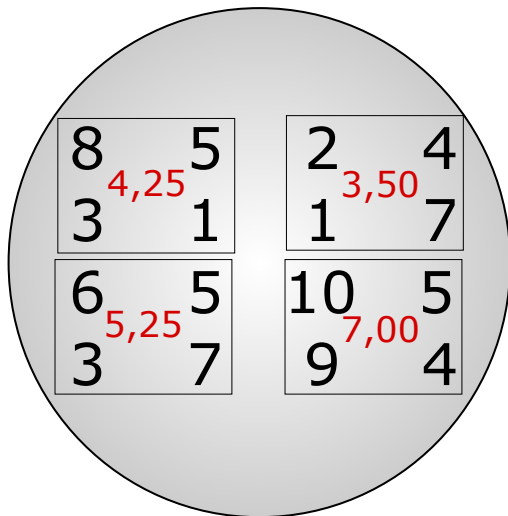
Média populacional e estimadores

Exemplo: notas de uma prova



Média populacional e estimadores

Exemplo: notas de uma prova



Parâmetros populacionais e seus estimadores

Média populacional: $\mu = 5$

Desvio-padrão populacional: $\sigma = 2,622$

Número de amostras possíveis de tamanho 4: 1820

Desvio-padrão de estimadores de tamanho 4: $EP = 1,173$

Parâmetros populacionais e seus estimadores

Média populacional: $\mu = 5$

Desvio-padrão populacional: $\sigma = 2,622$

Número de amostras possíveis de tamanho 4: 1820

Desvio-padrão de estimadores de tamanho 4: EP = 1,173

Estimativas de amostras de tamanho 4

Amostra	Média	Desvio-padrão	Erro-padrão
1	4,25	2,986	1,335
2	5,25	1,708	0,764
3	7,00	2,944	1,317
4	3,50	2,646	1,183

Parâmetros populacionais e seus estimadores

Média populacional: $\mu = 5$

Desvio-padrão populacional: $\sigma = 2,622$

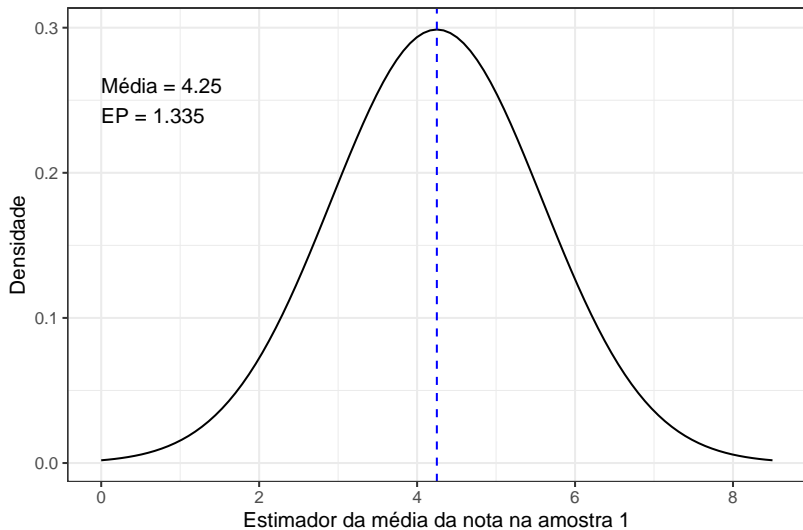
Número de amostras possíveis de tamanho 4: 1820

Desvio-padrão de estimadores de tamanho 4: EP = 1,173

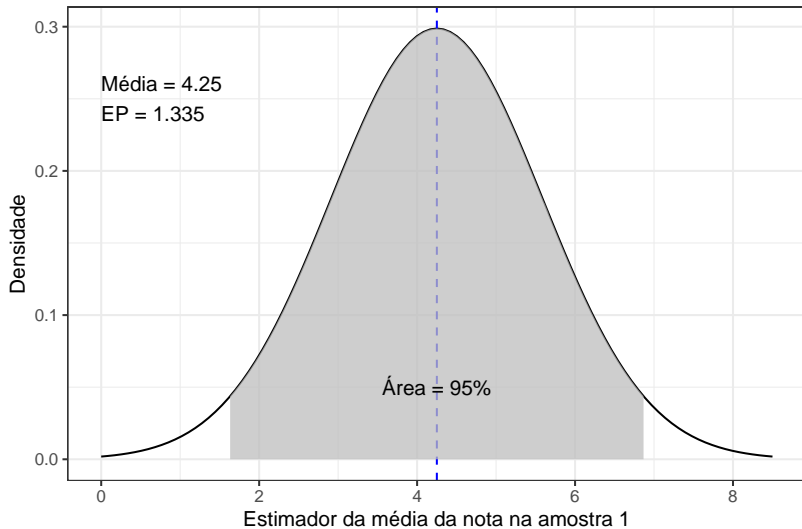
Estimativas de amostras aleatórias

N	Média	Desvio-padrão	Erro-padrão
6	4,50	2,345	0,782
8	5,25	2,659	0,687
10	5,20	3,084	0,617
12	5,50	2,780	0,414

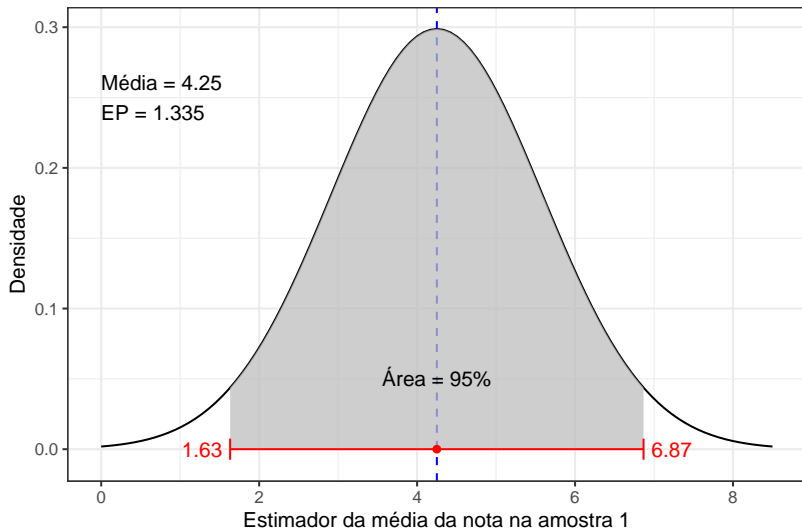
Intervalo de confiança



Intervalo de confiança



Intervalo de confiança

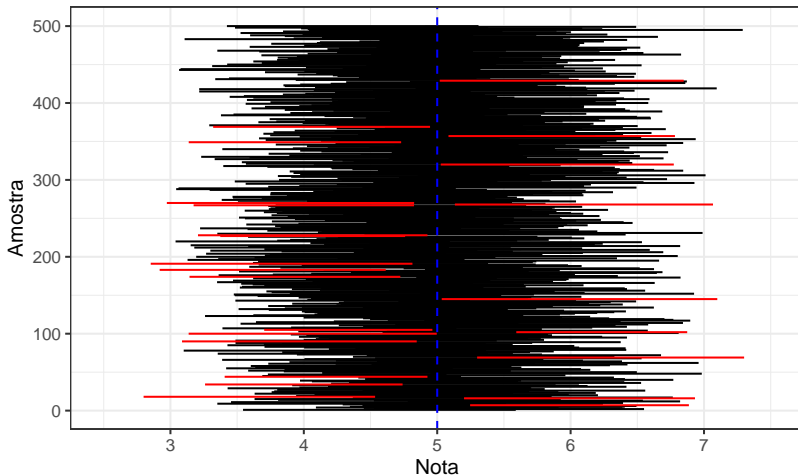


Intervalo de confiança

Amostras de tamanho 30

Intervalos contendo verdadeira média: 95.2%

Tamanho médio do intervalo: 1.95

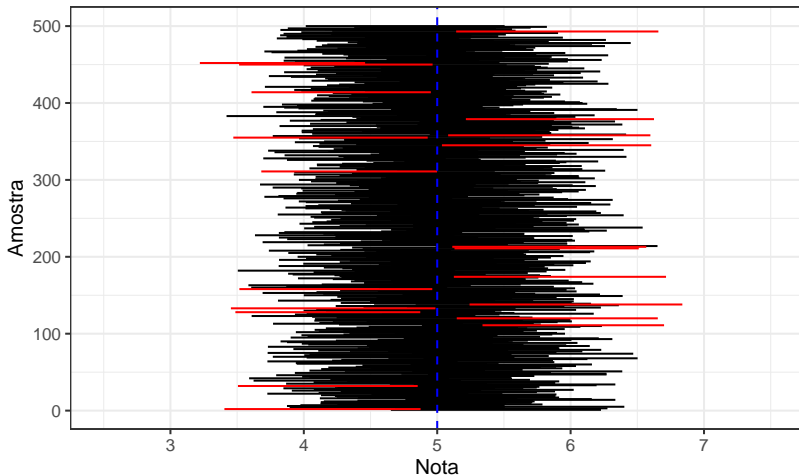


Intervalo de confiança

Amostras de tamanho 50

Intervalos contendo verdadeira média: 96.0%

Tamanho médio do intervalo: 1.48

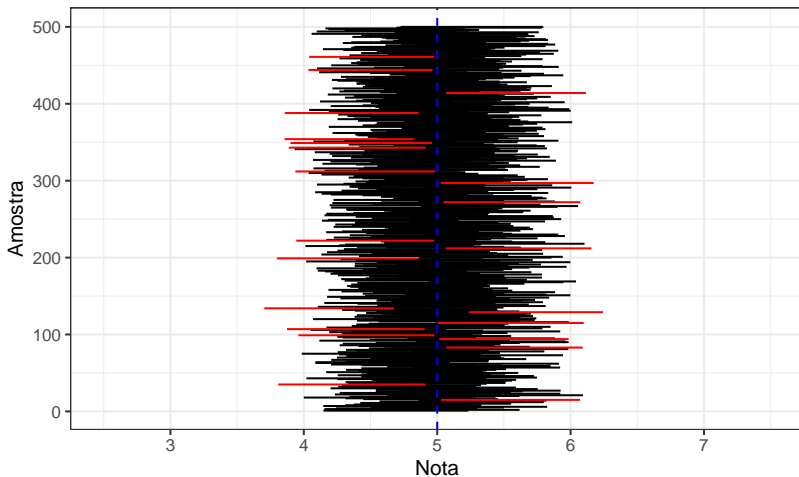


Intervalo de confiança

Amostras de tamanho 100

Intervalos contendo verdadeira média: 95.6%

Tamanho médio do intervalo: 1.04



Testes de hipóteses

Definição e conceitos

Testes de hipóteses

Definição e conceitos

Definição

Uma série de procedimentos para avaliar se uma amostra é consistente com uma hipótese feita *a priori* sobre algum parâmetro da população dada uma certa margem de erro.

Testes de hipóteses

Definição e conceitos

Definição

Uma série de procedimentos para avaliar se uma amostra é consistente com uma hipótese feita *a priori* sobre algum parâmetro da população dada uma certa margem de erro.

Hipóteses nula (H_0) ausência de relação entre dois fenômenos (usualmente)

Testes de hipóteses

Definição e conceitos

Definição

Uma série de procedimentos para avaliar se uma amostra é consistente com uma hipótese feita *a priori* sobre algum parâmetro da população dada uma certa margem de erro.

Hipóteses nula (H_0) ausência de relação entre dois fenômenos (usualmente)

Hipótese alternativa (H_a) complementar de H_0

Testes de hipóteses

Definição e conceitos

Definição

Uma série de procedimentos para avaliar se uma amostra é consistente com uma hipótese feita *a priori* sobre algum parâmetro da população dada uma certa margem de erro.

Hipóteses nula (H_0) ausência de relação entre dois fenômenos (usualmente)

Hipótese alternativa (H_a) complementar de H_0

Estatística de teste alguma função da amostra que mede o quão longe ela está de H_0

Testes de hipóteses

Definição e conceitos

Definição

Uma série de procedimentos para avaliar se uma amostra é consistente com uma hipótese feita *a priori* sobre algum parâmetro da população dada uma certa margem de erro.

Hipóteses nula (H_0) ausência de relação entre dois fenômenos (usualmente)

Hipótese alternativa (H_a) complementar de H_0

Estatística de teste alguma função da amostra que mede o quão longe ela está de H_0

Nível descritivo (valor-p) probabilidade de se obter uma amostra com estatística de teste maior ou igual à obtida assumindo que H_0 é verdadeira

Testes de hipóteses

Definição e conceitos

População	Resultado do teste	
	Não rejeitar H_0	Rejeitar H_0
H_0 verdadeira	✓	Erro tipo I
H_a verdadeira	Erro tipo II	✓

Testes de hipóteses

Definição e conceitos

População	Resultado do teste	
	Não rejeitar H_0	Rejeitar H_0
H_0 verdadeira	✓	Erro tipo I
H_a verdadeira	Erro tipo II	✓

Erro tipo I (α) probabilidade de rejeitar H_0 quando ela é verdadeira

Testes de hipóteses

Definição e conceitos

População	Resultado do teste	
	Não rejeitar H_0	Rejeitar H_0
H_0 verdadeira	✓	Erro tipo I
H_a verdadeira	Erro tipo II	✓

Erro tipo I (α) probabilidade de rejeitar H_0 quando ela é verdadeira

Nível de confiança ($1 - \alpha$) probabilidade de não de rejeitar H_0 quando ela é verdadeira

Testes de hipóteses

Definição e conceitos

População	Resultado do teste	
	Não rejeitar H_0	Rejeitar H_0
H_0 verdadeira	✓	Erro tipo I
H_a verdadeira	Erro tipo II	✓

Erro tipo I (α) probabilidade de rejeitar H_0 quando ela é verdadeira

Nível de confiança ($1 - \alpha$) probabilidade de não rejeitar H_0 quando ela é verdadeira

Erro tipo II (β) probabilidade de não rejeitar H_0 quando H_a é verdadeira

Testes de hipóteses

Definição e conceitos

População	Resultado do teste	
	Não rejeitar H_0	Rejeitar H_0
H_0 verdadeira	✓	Erro tipo I
H_a verdadeira	Erro tipo II	✓

Erro tipo I (α) probabilidade de rejeitar H_0 quando ela é verdadeira

Nível de confiança ($1 - \alpha$) probabilidade de não rejeitar H_0 quando ela é verdadeira

Erro tipo II (β) probabilidade de não rejeitar H_0 quando H_a é verdadeira

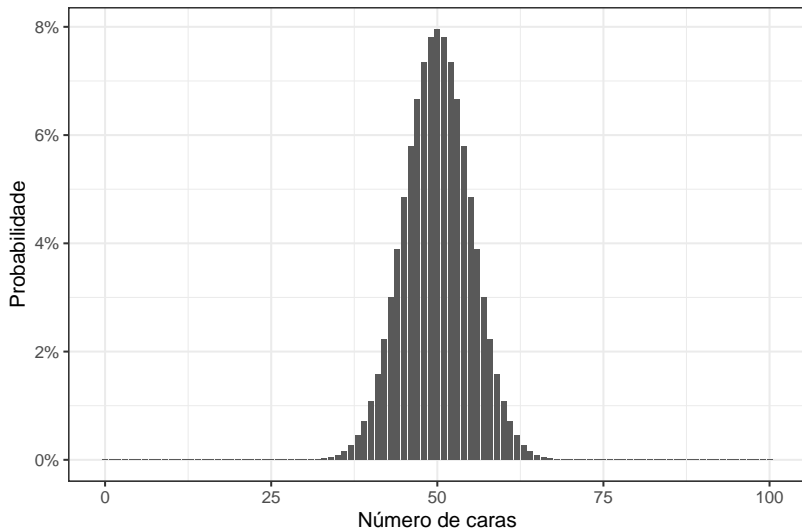
Poder de teste ($1 - \beta$) probabilidade de rejeitar H_0 quando H_a é verdadeira

Testes de hipóteses

Experimento: lançar uma moeda

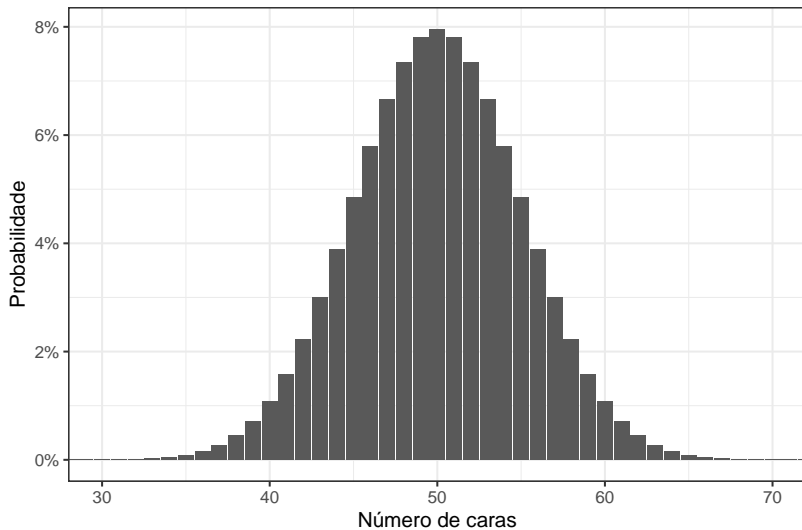
Testes de hipóteses

Experimento: lançar uma moeda



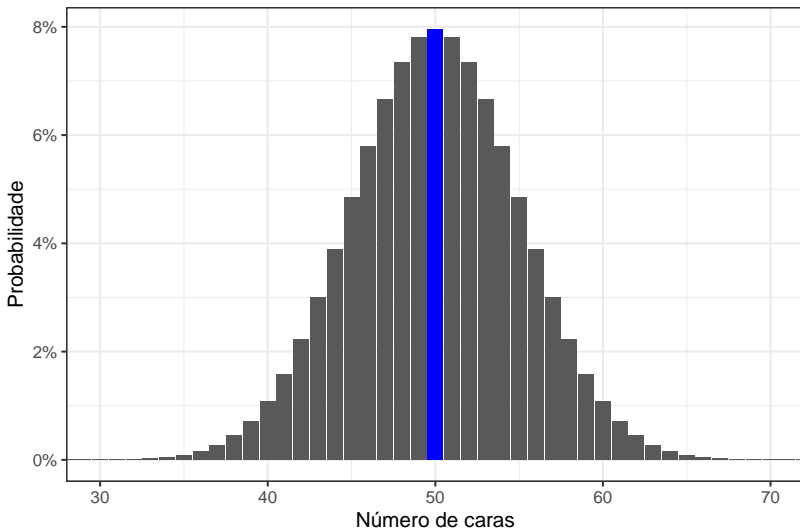
Testes de hipóteses

Experimento: lançar uma moeda



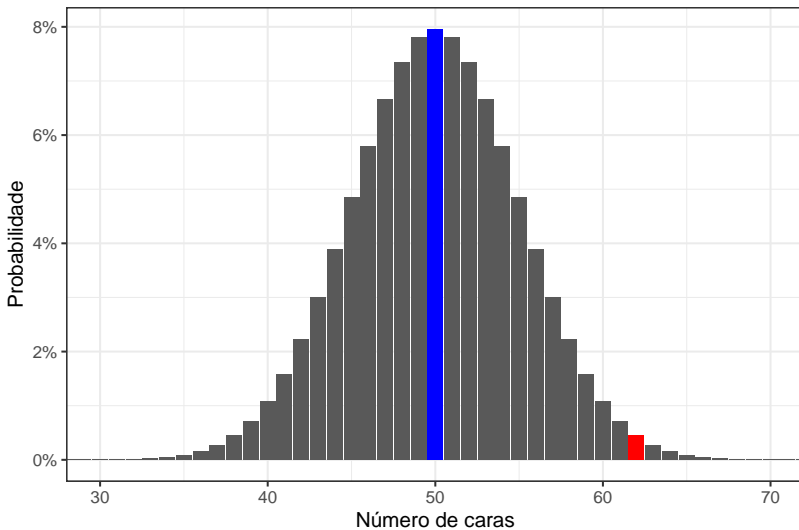
Testes de hipóteses

Experimento: lançar uma moeda



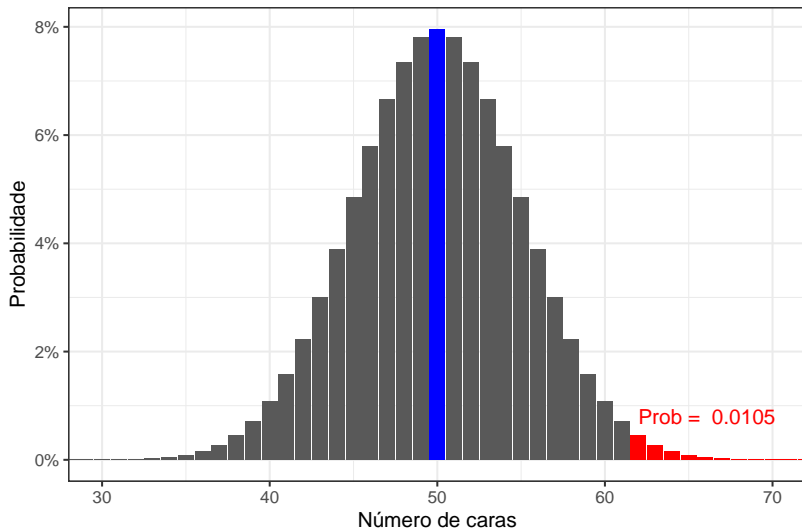
Testes de hipóteses

Experimento: lançar uma moeda



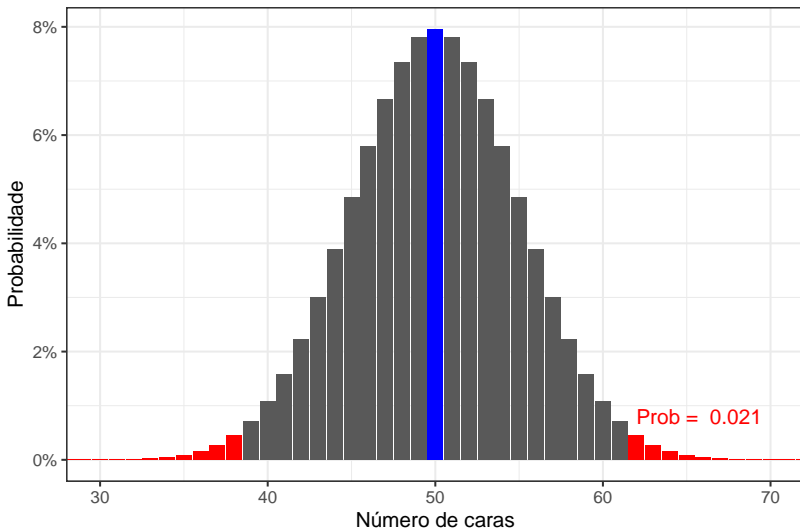
Testes de hipóteses

Experimento: lançar uma moeda



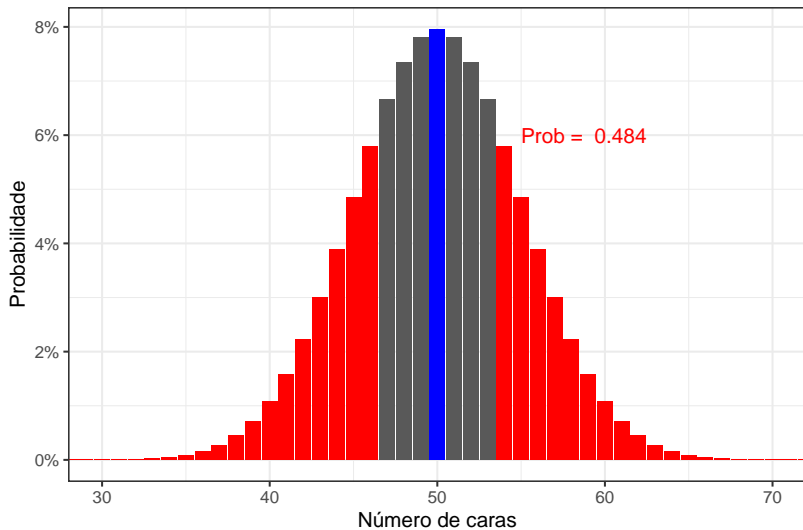
Testes de hipóteses

Experimento: lançar uma moeda



Testes de hipóteses

Experimento: lançar uma moeda



Testes de hipóteses

Algoritmo

- 1 Definir hipóteses nula e alternativa;

Testes de hipóteses

Algoritmo

- 1 Definir hipóteses nula e alternativa;
- 2 Extrair informação da amostra contra a hipótese nula;

Testes de hipóteses

Algoritmo

- 1 Definir hipóteses nula e alternativa;
- 2 Extrair informação da amostra contra a hipótese nula;
- 3 Calcular probabilidade de retirar essa amostra assumindo que a hipótese nula é verdadeira;

Testes de hipóteses

Algoritmo

- 1 Definir hipóteses nula e alternativa;
- 2 Extrair informação da amostra contra a hipótese nula;
- 3 Calcular probabilidade de retirar essa amostra assumindo que a hipótese nula é verdadeira;
- 4 Agregar evidência estatística e conhecimento sobre o assunto para tomar uma decisão.

Testes de hipóteses

Tipos de testes

Testes de hipóteses

Tipos de testes

Teste paramétrico

Faz suposições acerca da distribuição dos dados ou resíduos e estima seus parâmetros. Essas suposições costumam ser condições fortes sobre os dados e, por consequência, os testes são mais poderosos.

Testes de hipóteses

Tipos de testes

Teste paramétrico

Faz suposições acerca da distribuição dos dados ou resíduos e estima seus parâmetros. Essas suposições costumam ser condições fortes sobre os dados e, por consequência, os testes são mais poderosos.

Teste não paramétrico

Não estima nenhum parâmetro de distribuição no procedimento. Usualmente são casos mais gerais e menos poderosos do que testes paramétricos.

Testes de hipóteses

Tipos de testes

Teste paramétrico

Faz suposições acerca da distribuição dos dados ou resíduos e estima seus parâmetros. Essas suposições costumam ser condições fortes sobre os dados e, por consequência, os testes são mais poderosos.

Teste independente de distribuição

Faz suposições gerais sobre os dados de forma a não depender de distribuição, mas manipula a amostra para encaixá-la em alguma distribuição conhecida.

Teste não paramétrico

Não estima nenhum parâmetro de distribuição no procedimento. Usualmente são casos mais gerais e menos poderosos do que testes paramétricos.

Testes e modelos mais comuns

Testes e modelos mais comuns

Teste qui-quadrado associação de duas
Teste exato de Fisher variáveis categóricas

Testes e modelos mais comuns

Teste qui-quadrado associação de duas
Teste exato de Fisher variáveis categóricas

Teste t de Student/Welch comparação de variável
Teste de Wilcoxon-Mann-Whitney numérica entre dois grupos
Teste de Brunner-Munzel independentes

Testes e modelos mais comuns

Teste qui-quadrado associação de duas
Teste exato de Fisher variáveis categóricas

Teste t de Student/Welch comparação de variável
Teste de Wilcoxon-Mann-Whitney numérica entre dois grupos
Teste de Brunner-Munzel independentes

Correlação de Pearson/Kendall associação entre duas variáveis
numéricas ou ordinais

Associação entre variáveis categóricas

Associação entre variáveis categóricas

H_0 : independência entre as variáveis

H_a : as variáveis são dependentes

Associação entre variáveis categóricas

H_0 : independência entre as variáveis

H_a : as variáveis são dependentes

	Doente		Saudável	
	N	%	N	%
Controle	11	44,0	14	56,0
Experimental	2	8,0	23	92,0
Total	13	26,0	37	74,0

Associação entre variáveis categóricas

H_0 : independência entre as variáveis

H_a : as variáveis são dependentes

	Doente		Saudável	
	N	%	N	%
Controle	11	44,0	14	56,0
Experimental	2	8,0	23	92,0
Total	13	26,0	37	74,0

Teste qui-quadrado (χ^2) de E. Pearson

$$\chi^2 = 8,252, 1 \text{ g.l.}, p = 0,0041$$

Associação entre variáveis categóricas

H_0 : independência entre as variáveis

H_a : as variáveis são dependentes

	Doente		Saudável	
	N	%	N	%
Controle	11	44,0	14	56,0
Experimental	2	8,0	23	92,0
Total	13	26,0	37	74,0

Teste exato de Fisher

$$p = 0,0083$$

Tamanho de efeito: risco relativo e razão de chances

	Risco relativo (RR)	Razão de chances (OR)
Outro nome	Razão de prevalências	<i>Odds ratio</i>

Tamanho de efeito: risco relativo e razão de chances

	Risco relativo (RR)	Razão de chances (OR)
Outro nome	Razão de prevalências	<i>Odds ratio</i>
Definição	Compara a probabilidade de ocorrência de um evento entre dois grupos	Compara a chance de ocorrência de um evento entre dois grupos

Tamanho de efeito: risco relativo e razão de chances

	Risco relativo (RR)	Razão de chances (OR)
Outro nome	Razão de prevalências	<i>Odds ratio</i>
Definição	Compara a probabilidade de ocorrência de um evento entre dois grupos	Compara a chance de ocorrência de um evento entre dois grupos
Controle	Probabilidade de cura $\frac{14}{25} = 0,56 = 56\%$	Chance de cura $\frac{0,56}{1 - 0,56} = 1,273$
Experimental	$\frac{23}{25} = 0,92 = 92\%$	$\frac{0,92}{1 - 0,92} = 11,50$

Tamanho de efeito: risco relativo e razão de chances

	Risco relativo (RR)	Razão de chances (OR)
Outro nome	Razão de prevalências	<i>Odds ratio</i>
Definição	Compara a probabilidade de ocorrência de um evento entre dois grupos	Compara a chance de ocorrência de um evento entre dois grupos
Controle	Probabilidade de cura $\frac{14}{25} = 0,56 = 56\%$	Chance de cura $\frac{0,56}{1 - 0,56} = 1,273$
Experimental	$\frac{23}{25} = 0,92 = 92\%$	$\frac{0,92}{1 - 0,92} = 11,50$
Cálculo	$RR = \frac{0,92}{0,56} = 1,643$	$OR = \frac{11,50}{1,273} = 9,036$

Tamanho de efeito: risco relativo e razão de chances

	Risco relativo (RR)	Razão de chances (OR)
Outro nome	Razão de prevalências	<i>Odds ratio</i>
Definição	Compara a probabilidade de ocorrência de um evento entre dois grupos	Compara a chance de ocorrência de um evento entre dois grupos
Controle	Probabilidade de cura $\frac{14}{25} = 0,56 = 56\%$	Chance de cura $\frac{0,56}{1 - 0,56} = 1,273$
Experimental	$\frac{23}{25} = 0,92 = 92\%$	$\frac{0,92}{1 - 0,92} = 11,50$
Estimativa	RR = 1,643 IC95% = (1,139; 2,369)	OR = 9,036 IC95% = (1,741; 46,890)

Tamanho de efeito: risco relativo e razão de chances

	Risco relativo (RR)	Razão de chances (OR)
Outro nome	Razão de prevalências	<i>Odds ratio</i>
Definição	Compara a probabilidade de ocorrência de um evento entre dois grupos	Compara a chance de ocorrência de um evento entre dois grupos
Estimativa	RR = 1,643 IC95% = (1,139; 2,369)	OR = 9,036 IC95% = (1,741; 46,890)
Interpretação	Sujeitos do grupo experimental têm probabilidade de cura 64,3% maior do que os do grupo controle	Sujeitos do grupo experimental têm chance de cura 9 vezes maior do que os do grupo controle

Tamanho de efeito: risco relativo e razão de chances

	Risco relativo (RR)	Razão de chances (OR)
Outro nome	Razão de prevalências	<i>Odds ratio</i>
Definição	Compara a probabilidade de ocorrência de um evento entre dois grupos	Compara a chance de ocorrência de um evento entre dois grupos
Estimativa	RR = 0,609 IC95% = (0,422; 0,878)	OR = 0,111 IC95% = (0,021; 0,574)
Interpretação	Sujeitos do grupo controle têm probabilidade de cura 39,1% menor do que os do grupo experimental	Sujeitos do grupo controle têm 88,9% menor chance de cura do que os do grupo experimental

Associação entre variáveis categóricas

	Diabetes			
	Não		Sim	
	N	%	N	%
Controle	20	80,0	5	20,0
Experimental	22	88,0	3	12,0
Total	42	84,0	8	16,0

Teste exato de Fisher $p = 0,7019$

Razão de chances OR = 0,545

IC95% = (0,115; 2,581)

Associação entre variáveis categóricas

	Gravidade da doença					
	Leve		Moderada		Grave	
	N	%	N	%	N	%
Controle	8	32,0	6	24,0	11	44,0
Experimental	8	32,0	14	56,0	3	12,0
Total	16	32,0	20	40,0	14	28,0

Teste qui-quadrado $\chi^2 = 7,616$, 2 g.l., $p = 0,0222$

Teste exato de Fisher $p = 0,0210$

Associação entre variáveis categóricas

Suposições

- Teste χ^2 de K. Pearson
 - Observações independentes
 - Células com valor esperado maior ou igual a 5
 - Nenhuma célula com valor observado igual a zero
- Teste χ^2 de E. Pearson
 - Observações independentes
 - Células com valor esperado maior ou igual a 1
 - Nenhuma célula com valor observado igual a zero
- Teste exato de Fisher
 - Observações independentes
 - As marginais são fixas

Comparação de variável numérica entre dois grupos

Comparação de variável numérica entre dois grupos

H_0 : a variável numérica tem a mesma distribuição nos dois grupos

H_a : a variável numérica tem distribuição diferente nos dois grupos

Comparação de variável numérica entre dois grupos

H_0 : a variável numérica tem a mesma distribuição nos dois grupos

H_a : a variável numérica tem distribuição diferente nos dois grupos

Escolaridade por grupo

Grupo	Média	DP	Mediana	IC95%	
				Inferior	Superior
Controle	12,6	4,19	13	10,87	14,33
Experimental	11,2	3,14	12	9,91	12,49

Comparação de variável numérica entre dois grupos

H_0 : a variável numérica tem a mesma distribuição nos dois grupos

H_a : a variável numérica tem distribuição diferente nos dois grupos

Escolaridade por grupo

Grupo	Média	DP	Mediana	IC95%	
				Inferior	Superior
Controle	12,6	4,19	13	10,87	14,33
Experimental	11,2	3,14	12	9,91	12,49

Teste	Estatística	g.l.	valor-p
t de Student	1,3369	48	0,1876
t de Welch	1,3370	44,45	0,1881
Wilcoxon-Mann-Whitney	351		0,4526
Brunner-Munzel	0,7443	44,85	0,4606

Tamanho de efeito: d de Cohen

Escolaridade por grupo			
Grupo	N	Média	DP
Controle	25	12,6	4,19
Experimental	25	11,2	3,14

Definição $d = \frac{\bar{x}_1 - \bar{x}_2}{s_p}$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Tamanho de efeito: d de Cohen

Escolaridade por grupo			
Grupo	N	Média	DP
Controle	25	12,6	4,19
Experimental	25	11,2	3,14

$$s_p = \sqrt{\frac{(25 - 1) \times 4,19^2 + (25 - 1) \times 3,14^2}{25 + 25 - 2}} = 3,70$$

$$d = \frac{12,6 - 11,2}{3,70} = 0,378$$

$$IC95\% = (-0,193; 0,949)$$

Tamanho de efeito: d de Cohen

Escolaridade por grupo			
Grupo	N	Média	DP
Controle	25	12,6	4,19
Experimental	25	11,2	3,14

$$d = 0,378$$

$$IC95\% = (-0,193; 0,949)$$

Interpretação: sujeitos do grupo controle têm escolaridade 0,378 desvios-padrão maior do que os do grupo experimental.

Comparação de variável numérica entre dois grupos

Dor por grupo

	Média	DP	Mediana	IC95%	
				Inferior	Superior
Controle	5,52	1,71	5	5,0	6,0
Experimental	3,44	1,61	3	2,5	4,0

Teste	Estatística	g.l.	valor-p
t de Student	4,428	48	< 0,001
t de Welch	4,428	47,82	< 0,001
Wilcoxon-Mann-Whitney	508,5		< 0,001
Brunner-Munzel	5,190	45,61	< 0,001

$$d = 1,252$$

$$\text{IC95\%} = (0,631; 1,873)$$

Comparação de variável numérica entre dois grupos

Suposições

- **Teste t de Student**
 - Independência
 - Homoscedasticidade
 - Normalidade residual
- **Teste t de Welch**
 - Independência
 - Normalidade residual
- **Teste de Wilcoxon-Mann-Whitney**
 - Independência
 - Distribuições idênticas exceto por parâmetro de locação
- **Teste de Brunner-Munzel**
 - Independência

Correlação entre duas variáveis numéricas ou ordinais

Correlação entre duas variáveis numéricas ou ordinais

Conceito

Correlação mede a força de associação entre duas variáveis numéricas. O coeficiente ρ varia entre -1 e 1, sendo 0 a total falta de associação. A correlação em si já é um tamanho de efeito.

Correlação entre duas variáveis numéricas ou ordinais

Conceito

Correlação mede a força de associação entre duas variáveis numéricas. O coeficiente ρ varia entre -1 e 1, sendo 0 a total falta de associação. A correlação em si já é um tamanho de efeito.

$\rho > 0$ quando uma variável aumenta, a outra também aumenta

$\rho < 0$ quando uma variável aumenta, a outra diminui

Correlação entre duas variáveis numéricas ou ordinais

Conceito

Correlação mede a força de associação entre duas variáveis numéricas. O coeficiente ρ varia entre -1 e 1, sendo 0 a total falta de associação. A correlação em si já é um tamanho de efeito.

$\rho > 0$ quando uma variável aumenta, a outra também aumenta

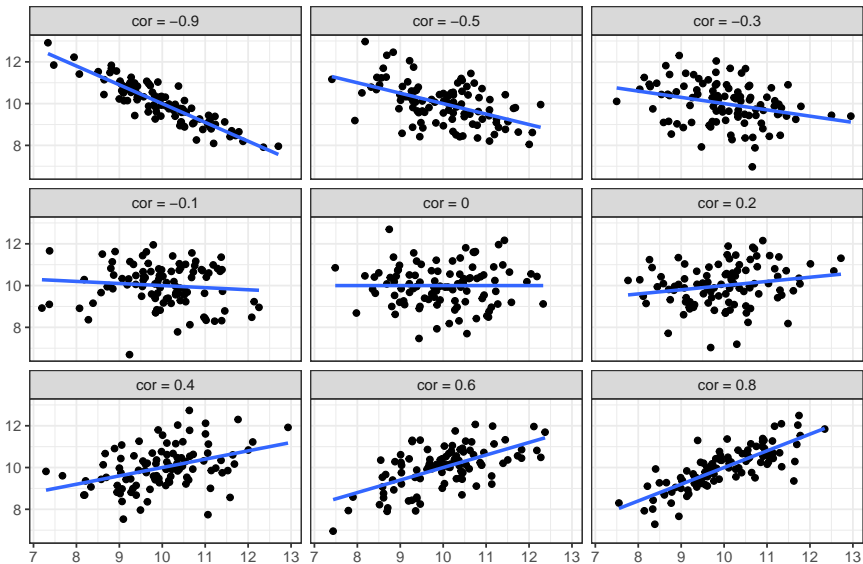
$\rho < 0$ quando uma variável aumenta, a outra diminui

Tipos de coeficiente

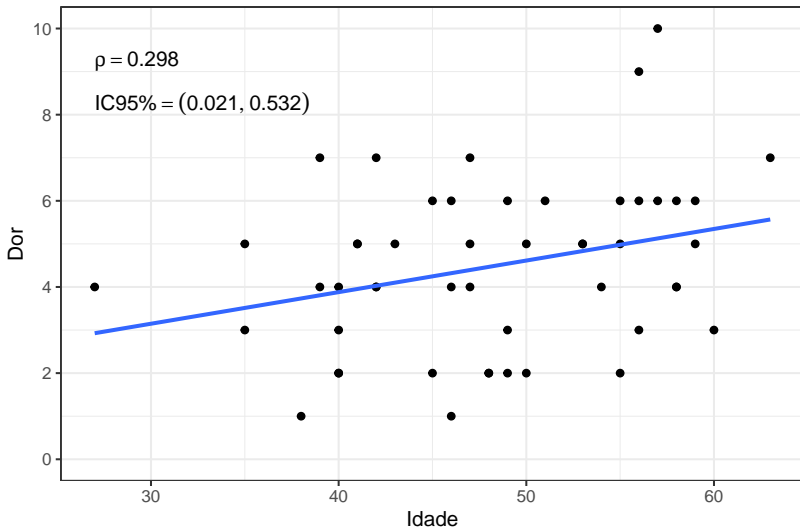
Pearson mede associação linear de duas variáveis numéricas

Kendall mede associação monotônica entre duas variáveis numéricas ou ordinais

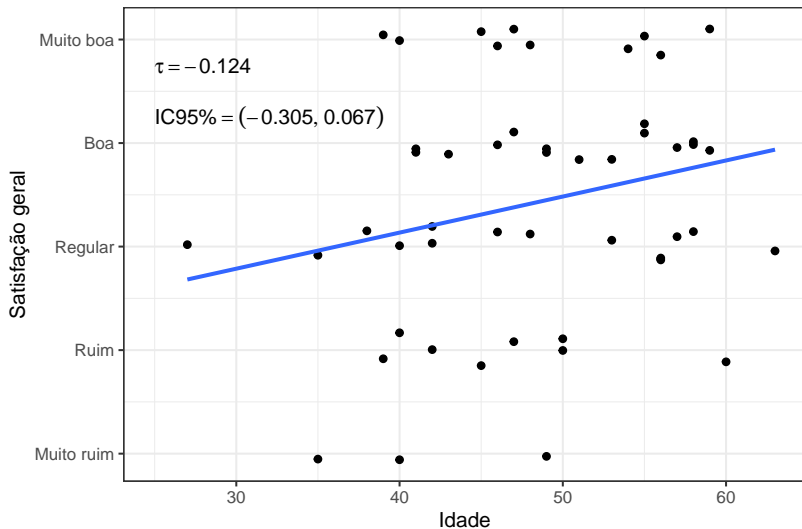
Vários exemplos de força de correlação



Correlação entre duas variáveis numéricas ou ordinais



Correlação entre duas variáveis numéricas ou ordinais



Regressão linear

Regressão linear

Conceito

Regressão é um modelo para estimar a relação entre uma variável resposta (desfecho ou dependente) com uma ou mais variáveis explicativas (preditores ou independentes). O objetivo pode ser tanto entender como os preditores explicam o desfecho como prever novos dados.

Regressão linear

Conceito

Regressão é um modelo para estimar a relação entre uma variável resposta (desfecho ou dependente) com uma ou mais variáveis explicativas (preditores ou independentes). O objetivo pode ser tanto entender como os preditores explicam o desfecho como prever novos dados.

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

$$\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

Regressão linear

Conceito

Regressão é um modelo para estimar a relação entre uma variável resposta (desfecho ou dependente) com uma ou mais variáveis explicativas (preditores ou independentes). O objetivo pode ser tanto entender como os preditores explicam o desfecho como prever novos dados.

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

$$\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

Preço =

Regressão linear

Conceito

Regressão é um modelo para estimar a relação entre uma variável resposta (desfecho ou dependente) com uma ou mais variáveis explicativas (preditores ou independentes). O objetivo pode ser tanto entender como os preditores explicam o desfecho como prever novos dados.

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

$$\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

Preço =

km rodado

Regressão linear

Conceito

Regressão é um modelo para estimar a relação entre uma variável resposta (desfecho ou dependente) com uma ou mais variáveis explicativas (preditores ou independentes). O objetivo pode ser tanto entender como os preditores explicam o desfecho como prever novos dados.

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

$$\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

Preço = Bandeirada

km rodado

Regressão linear

Conceito

Regressão é um modelo para estimar a relação entre uma variável resposta (desfecho ou dependente) com uma ou mais variáveis explicativas (preditores ou independentes). O objetivo pode ser tanto entender como os preditores explicam o desfecho como prever novos dados.

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

$$\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

Preço = Bandeirada + Preço por km × km rodado

Regressão linear

Conceito

Regressão é um modelo para estimar a relação entre uma variável resposta (desfecho ou dependente) com uma ou mais variáveis explicativas (preditores ou independentes). O objetivo pode ser tanto entender como os preditores explicam o desfecho como prever novos dados.

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

$$\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

Preço = Bandeirada + Preço por km × km rodado
+ Preço por minuto × Tempo da corrida

Regressão linear

Conceito

Regressão é um modelo para estimar a relação entre uma variável resposta (desfecho ou dependente) com uma ou mais variáveis explicativas (preditores ou independentes). O objetivo pode ser tanto entender como os preditores explicam o desfecho como prever novos dados.

$$y_i = \alpha + \beta x_i + \varepsilon_i$$
$$\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

Preço = Bandeirada + Preço por km × km rodado
+ Preço por minuto × Tempo da corrida
+ arredondamento

Regressão linear

Conceito

Regressão é um modelo para estimar a relação entre uma variável resposta (desfecho ou dependente) com uma ou mais variáveis explicativas (preditores ou independentes). O objetivo pode ser tanto entender como os preditores explicam o desfecho como prever novos dados.

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

$$\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

Preço = Bandeirada + Preço por km × km rodado
+ Preço por minuto × Tempo da corrida
+ arredondamento

Regressão linear

Conceito

Regressão é um modelo para estimar a relação entre uma variável resposta (desfecho ou dependente) com uma ou mais variáveis explicativas (preditores ou independentes). O objetivo pode ser tanto entender como os preditores explicam o desfecho como prever novos dados.

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

$$\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

$$H_0: \beta = 0$$

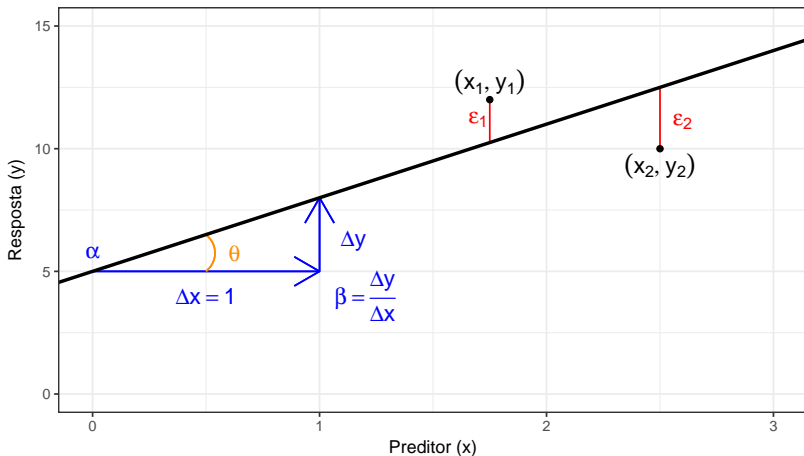
$$H_a: \beta \neq 0$$

Preço = Bandeirada + Preço por km × km rodado
+ Preço por minuto × Tempo da corrida
+ arredondamento

Regressão linear

Interpretação geométrica

$$y_i = \alpha + \beta x_i + \varepsilon_i = 5 + 3x_i + \varepsilon_i$$



Regressão linear múltipla

Modelo que associa uma variável numérica com uma combinação linear de preditores:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

Regressão linear múltipla

Modelo que associa uma variável numérica com uma combinação linear de preditores:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$
$$\varepsilon \stackrel{iid}{\sim} N(0, \sigma^2)$$

Regressão linear múltipla

Modelo que associa uma variável numérica com uma combinação linear de preditores:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$
$$\varepsilon \stackrel{iid}{\sim} N(0, \sigma^2)$$

- Variável resposta numérica
- Preditores não multicolineares
- Resíduos normais, independentes e homoscedásticos
- Modelo exato independente do tamanho da amostra

Regressão linear múltipla

Modelo que associa uma variável numérica com uma combinação linear de preditores:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$
$$\varepsilon \stackrel{iid}{\sim} N(0, \sigma^2)$$

Interpretação

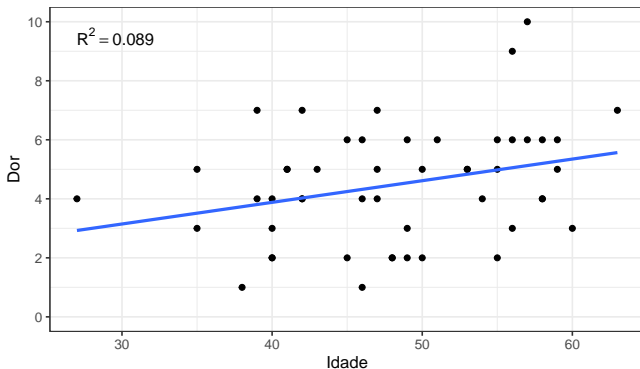
Preditor numérico Aumentar o valor da variável em uma unidade muda o desfecho em β unidades em média

Preditor categórico Observações da categoria observada têm média β unidades diferente da categoria de referência ou da média geral

Regressão linear simples

Exemplo

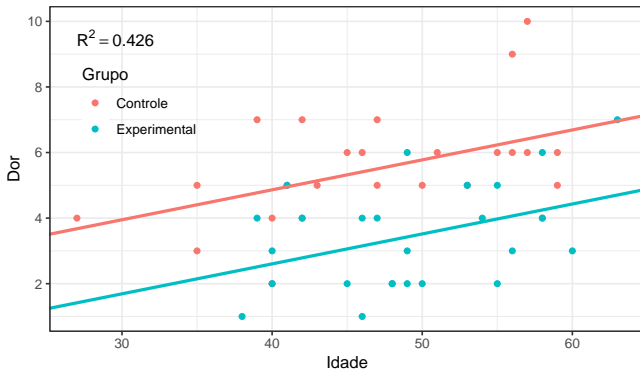
Variável	Coeficiente	IC95%		Valor-p
		Inferior	Superior	
Intercepto	0,947	-2,382	4,276	0,570
Idade	0,073	0,005	0,142	0,036



Regressão linear múltipla

Exemplo

Variável	Coeficiente	IC95%		Valor-p
		Inferior	Superior	
Intercepto	1,210	-1,465	3,884	0,366
Idade	0,091	0,036	0,147	0,002
Grupo experimental	-2,259	-3,125	-1,393	< 0,001



Dimensionamento amostral (finalmente!)

Dimensionamento amostral (finalmente!)

Depende de:

- Tamanho de efeito

Dimensionamento amostral (finalmente!)

Depende de:

- Tamanho de efeito
- Erro tipo I

Dimensionamento amostral (finalmente!)

Depende de:

- Tamanho de efeito
- Erro tipo I
- Poder de teste (oposto do erro tipo II)

Dimensionamento amostral (finalmente!)

Depende de:

- Tamanho de efeito
- Erro tipo I
- Poder de teste (oposto do erro tipo II)
- Soluções não são exatas, apenas aproximações!

Dimensionamento amostral (finalmente!)

Depende de:

- Tamanho de efeito
- Erro tipo I
- Poder de teste (oposto do erro tipo II)
- **Soluções não são exatas, apenas aproximações!**

Valores de referência (segundo Cohen)

Efeito	d de Cohen	Correlação (ρ)	V de Cramér
Pequeno	0,2	0,1	0,1
Médio	0,5	0,3	0,3
Grande	0,8	0,5	0,5

Dimensionamento amostral

Um exemplo

Uma pesquisa está sendo planejada para comparar o tempo que pessoas demoram para cumprir uma certa tarefa de atenção concentrada em duas salas com condições diferentes.

Dimensionamento amostral

Um exemplo

Uma pesquisa está sendo planejada para **comparar** o tempo que pessoas demoram para cumprir uma certa tarefa de atenção concentrada em **duas** salas com **condições diferentes**.

Dimensionamento amostral

Um exemplo

Uma pesquisa está sendo planejada para comparar o tempo que pessoas demoram para cumprir uma certa tarefa de atenção concentrada em duas salas com condições diferentes. Uma amostra piloto sugere que o tempo médio é de 12 minutos na sala com menos distrações e 15 minutos na sala com mais distrações, ambas com desvio-padrão de 5 minutos.

Dimensionamento amostral

Um exemplo

Uma pesquisa está sendo planejada para comparar o tempo que pessoas demoram para cumprir uma certa tarefa de atenção concentrada em duas salas com condições diferentes. Uma **amostra piloto** sugere que o **tempo médio** é de 12 minutos na sala com menos distrações e 15 minutos na sala com mais distrações, ambas com **desvio-padrão** de 5 minutos.

Dimensionamento amostral

Um exemplo

Uma pesquisa está sendo planejada para comparar o tempo que pessoas demoram para cumprir uma certa tarefa de atenção concentrada em duas salas com condições diferentes. Uma amostra piloto sugere que o tempo médio é de 12 minutos na sala com menos distrações e 15 minutos na sala com mais distrações, ambas com desvio-padrão de 5 minutos. O tamanho de efeito nesse caso é $d = \frac{\bar{x}_1 - \bar{x}_2}{s} = \frac{15 - 12}{5} = 0,6$, ou seja, a sala mais barulhenta faz os sujeitos levarem 0,6 desvios-padrão a mais para terminar a tarefa do que a outra.

Dimensionamento amostral

Um exemplo

Uma pesquisa está sendo planejada para comparar o tempo que pessoas demoram para cumprir uma certa tarefa de atenção concentrada em duas salas com condições diferentes. Uma amostra piloto sugere que o tempo médio é de 12 minutos na sala com menos distrações e 15 minutos na sala com mais distrações, ambas com desvio-padrão de 5 minutos. O **tamanho de efeito** nesse caso é $d = \frac{\bar{x}_1 - \bar{x}_2}{s} = \frac{15 - 12}{5} = 0,6$, ou seja, a sala mais barulhenta faz os sujeitos levarem 0,6 desvios-padrão a mais para terminar a tarefa do que a outra.

Dimensionamento amostral

Um exemplo

Uma pesquisa está sendo planejada para comparar o tempo que pessoas demoram para cumprir uma certa tarefa de atenção concentrada em duas salas com condições diferentes. Uma amostra piloto sugere que o tempo médio é de 12 minutos na sala com menos distrações e 15 minutos na sala com mais distrações, ambas com desvio-padrão de 5 minutos. O tamanho de efeito nesse caso é $d = \frac{\bar{x}_1 - \bar{x}_2}{s} = \frac{15 - 12}{5} = 0,6$, ou seja, a sala mais barulhenta faz os sujeitos levarem 0,6 desvios-padrão a mais para terminar a tarefa do que a outra. Para que essa diferença seja significativa num teste t de Student com poder de 95% ($\beta = 0,05$) e erro tipo I de 5% ($\alpha = 0,05$) são necessários ao menos 74 sujeitos.

Dimensionamento amostral

Um exemplo

Uma pesquisa está sendo planejada para comparar o tempo que pessoas demoram para cumprir uma certa tarefa de atenção concentrada em duas salas com condições diferentes. Uma amostra piloto sugere que o tempo médio é de 12 minutos na sala com menos distrações e 15 minutos na sala com mais distrações, ambas com desvio-padrão de 5 minutos. O tamanho de efeito nesse caso é $d = \frac{\bar{x}_1 - \bar{x}_2}{s} = \frac{15 - 12}{5} = 0,6$, ou seja, a sala mais barulhenta faz os sujeitos levarem 0,6 desvios-padrão a mais para terminar a tarefa do que a outra. Para que essa diferença seja significativa num teste **t de Student** com **poder de 95%** ($\beta = 0,05$) e **erro tipo I de 5%** ($\alpha = 0,05$) são necessários ao menos 74 sujeitos.

Dimensionamento amostral

Tipos de estudo e tamanhos de efeito

Objetivo	Elementos necessários	Tamanho de efeito
Comparar dois ou mais grupos	Tamanho das amostras, médias e DP dos grupos	d de Cohen η^2 parcial
Comparar duas proporções	Proporção base e razão de chances ou risco relativo	Razão de chances Risco relativo
Comparar várias proporções	Estatística χ^2	V de Cramér
Correlacionar duas variáveis	Correlação	ρ ou τ
Regressão	Coeficiente de determinação	R^2
Estudo longitudinal	Médias e DP de todos os grupos e condições	η^2 parcial

Dimensionamento amostral

Estimação de prevalência

Dimensionamento amostral

Estimação de prevalência

- Estimativa da prevalência
- Margem de erro
- Nível de confiança

Dimensionamento amostral

Estimação de prevalência

$$n \geq \frac{p(1-p)z_{1-\alpha/2}^2}{\epsilon^2}$$

p estimativa da prevalência a ser estudada

ϵ margem de erro a ser adotada na estimativa

$1 - \alpha$ nível de confiança adotado

$z_{1-\alpha/2}$ quantil $1 - \alpha/2$ da distribuição normal padrão. É o que garante a cobertura do nível de confiança adotado

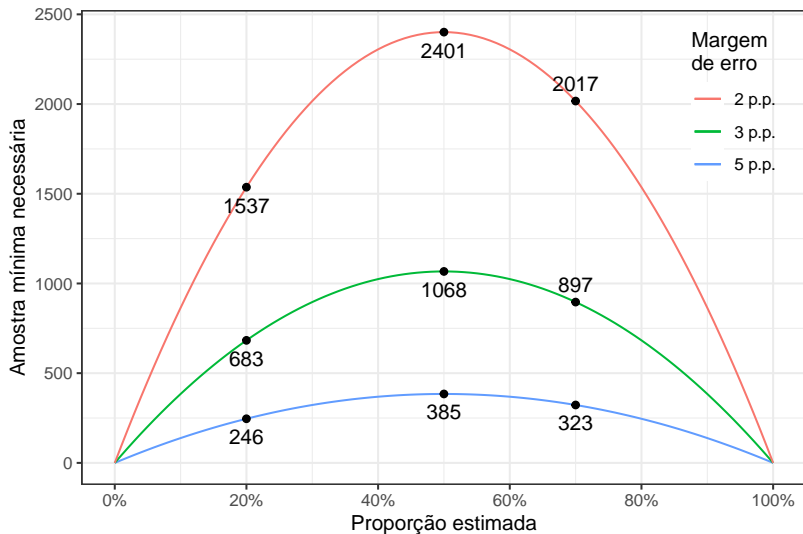
Dimensionamento amostral

Estimação de prevalência

Tamanho da amostra (com confiança de 95%)				
Prevalência	Margem de erro			
	1 p.p.	2 p.p.	5 p.p.	10 p.p.
5% ou 95%	1825	457	73	19
10% ou 90%	3458	865	139	35
20% ou 80%	6147	1537	246	62
25% ou 75%	7203	1801	289	73
33% ou 66%	8537	2135	342	86
50%	9604	2401	385	97

Dimensionamento amostral

Estimação de prevalência



Referências

- [1] Pedro Alberto Morettin e Wilton Oliveira Bussab. *Estatística básica*. Saraiva, 2017.
- [2] George Casella e Roger L Berger. *Statistical inference*. 2ª ed. Duxbury, 2002.
- [3] Bernardo dos Santos. *Conceitos de estatística aplicada*. 2018.
URL: <https://abrir.link/livroestatistica>.

Email: bernardo.santos@usp.br

Essa apresentação: <https://abre.ai/aulaestatistica>