

Arthur de Sá Ferreira

Ciência com R

Perguntas e respostas para pesquisadores e analistas de dados

Arthur de Sá Ferreira

Ciência com R

Perguntas e respostas para pesquisadores e analistas de dados

2026

Copyright © 2023–2026 Arthur de Sá Ferreira

Publicado em Niterói, Brasil
Edição do autor / Publicação independente

Todos os direitos reservados. Nenhuma parte deste livro pode ser reproduzida ou transmitida por qualquer meio eletrônico, mecânico, fotocópia, gravação ou outro, sem a permissão prévia por escrito do autor, exceto nos casos previstos pela lei de direitos autorais.

Nenhuma garantia é dada em relação ao conteúdo desta obra, incluindo adequação a finalidades específicas. O uso é de responsabilidade exclusiva do leitor.

As opiniões expressas nesta obra são de responsabilidade exclusiva do autor.

Para solicitar permissões, entre em contato: cienciaomr@gmail.com

1^a edição — 2026
Capa dura: ISBN
Brochura: ISBN
E-book: ISBN

Depósito legal: Biblioteca Nacional, Brasil

Arthur de Sá Ferreira
Ciência com R. – Rio de Janeiro: Editora, 2026.
429 p. : il. (alguma cor).
ISBN XXX-XX-XXXX-XXX-X
1. Estatística aplicada. 2. Metodologia científica. 3. Análise de dados. 4. R (Linguagem de programação). 5. Pesquisa científica. 6. Modelagem estatística. 7. Reprodutibilidade científica. 8. Boas práticas em pesquisa.
I. Título. II. Educação. III. Tecnologia.

Sumário

Sumário	v
Lista de Figuras	xv
Lista de Tabelas	xxi
Dedicatória	xxv
Agradecimentos	xxvii
Apresentação	xxix
Sobre o autor	xxxi

PARTE 1: PENSAMENTO CIENTÍFICO	1
1 Pensamento probabilístico	3
1.1 Experimento	3
1.2 Espaço amostral e eventos discretos	3
1.3 Espaço amostral e eventos contínuos	4
1.4 Probabilidade	4
1.5 Independência e probabilidade	6
1.6 Leis dos números anômalos	7
1.7 Leis dos pequenos números	7
1.8 Leis dos grandes números	9
1.9 Teorema central do limite	9
1.10 Regressão para a média	13
2 Pensamento estatístico	17
2.1 Unidade de análise	17
2.2 População	17
2.3 Amostra	18
2.4 Amostragem	18
2.5 Reamostragem	19
2.6 Subamostragem	22
2.7 Superamostragem	22
3 Pensamento metodológico	25
3.1 Metodologia da pesquisa	25
3.2 Relação Estatística-Metodologia	25

3.3	Pesquisa quantitativa vs. qualitativa	25
3.4	Pesquisa de métodos mistos	27
3.5	Pesquisa exploratória vs. confirmatória	27
3.6	Pesquisa translacional	28
3.7	Pré-registro	28
3.8	Reprodutibilidade	28
3.9	Robustez	29
3.10	Replicabilidade	29
3.11	Generalização	29
4	Pensamento computacional	31
4.1	R	31
4.2	RStudio	32
4.3	Scripts	32
4.4	Pacotes	35
4.5	Aplicativos	36
4.6	Manuscritos reproduzíveis	36
4.7	Compartilhamento	38
5	Letramento estatístico	41
5.1	Introdução ao letramento estatístico	41
5.2	Elementos centrais do letramento estatístico	42
5.3	Hierarquia de letramento estatístico	42
5.4	Habilidades de letramento estatístico baseadas no pensamento crítico	43
<hr/>		
<i>PARTE 2: AMEAÇAS À QUALIDADE DA EVIDÊNCIA CIENTÍFICA</i>		45
6	Vieses metodológicos	47
6.1	Vieses metodológicos	47
6.2	Tipos de vieses metodológicos	47
6.3	Efeitos relacionados aos vieses metodológicos	47
6.4	Diretrizes para redação	48
7	Falácia estatísticas	49
7.1	Falácia	49
8	Paradoxos estatísticos	51
8.1	Paradoxos	51
9	Práticas questionáveis em pesquisa	55
9.1	Práticas Questionáveis em Pesquisa	55
9.2	Prática não intencional e má conduta	55
9.3	Prevenindo práticas questionáveis em pesquisa	58
9.4	Reações éticas e institucionais práticas questionáveis em pesquisa	58
<hr/>		
<i>PARTE 3: DO MUNDO REAL À TABELA</i>		61
10	Variáveis e fatores	63
10.1	Variáveis	63
10.2	Transformação de variáveis	64

10.3 Centralização de variáveis (<i>centering</i>)	67
10.4 Padronização de variáveis	67
10.5 Categorização de variáveis contínuas	69
10.6 Dicotomização de variáveis contínuas	69
10.7 Representação de variáveis categóricas	71
10.8 Fatores	72
11 Dados e metadados	75
11.1 Dados	75
11.2 Metadados	76
12 Medidas e instrumentos	79
12.1 Escalas	79
12.2 Medição e Medidas	81
12.3 Erro de medida	84
12.4 Instrumentos	84
12.5 Acurácia e precisão	86
12.6 Viés e variabilidade	86
13 Tabulação de dados	89
13.1 Planilhas eletrônicas	89
14 Dados perdidos e imputados	93
14.1 Dados perdidos	93
14.2 Mecanismos geradores de dados perdidos	94
14.3 Estratégias para lidar com dados perdidos	97
14.4 Dados imputados	98
15 Dados anonimizados e sintéticos	101
15.1 Dados anonimizados	101
15.2 Dados sintéticos	102
<hr/>	
PARTE 4: ANÁLISES DESCRIPTIVAS E EXPLORATÓRIAS	103
16 Distribuições e parâmetros	105
16.1 Distribuições de probabilidade	105
16.2 Distribuições multivariadas	117
16.3 Parâmetros	118
16.4 Erro	119
16.5 Tendência central	120
16.6 Dispersão	121
16.7 Proporção	124
16.8 Extremos	125
16.9 Distribuição	125
16.10 Robustez em medidas de localização	127
16.11 Parâmetros robustos	129
17 Análise inicial de dados	131
17.1 Análise inicial de dados	131
18 Análise exploratória de dados	133
18.1 Análise exploratória de dados	133

18.2 Ingredientes da análise exploratória de dados	134
19 Análise descritiva	137
19.1 Análise descritiva	137
19.2 Apresentação de resultados numéricos	137
19.3 Tabelas	139
19.4 Tabela 1	140
19.5 Tabela 2	141
19.6 Visualização efetiva de dados	142
19.7 Gráficos	144
19.8 Tipos de gráficos	147
19.9 Gráficos dinâmicos	163
20 Análise robusta	165
20.1 Raciocínio inferencial robusto	165
20.2 Valores discrepantes	166
20.3 Valores influentes	169
20.4 Métodos robustos de tratamento de <i>outliers</i>	170
21 Análise preditiva	173
21.1 Predição	173
21.2 Interpretação e aplicação	173
21.3 Análise de curva de decisão	174
22 Análise causal	175
22.1 Causalidade	175
22.2 Abordagens filosóficas e estatísticas da causalidade	176
22.3 Ilusões de causalidade	176
22.4 Inferência causal em estudos observacionais	177
22.5 Critérios de Hill para inferência causal	177
22.6 Críticas contemporâneas aos critérios de Hill	178
22.7 Visão atual sobre os critérios de Hill	178
22.8 Linguagem causal em estudos observacionais	179
22.9 Efeitos diretos e indiretos	179
22.10 O papel do tempo e a causalidade dinâmica	179
22.11 Diagrama acíclico direcionado (DAG)	180
23 Análise qualitativa	183
23.1 Análise qualitativa	183
23.2 Representação de texto	183
PARTE 5: ANÁLISES INFERENCIAIS	185
24 Suposições inferenciais	187
24.1 Suposições gerais em análises inferenciais	187
24.2 Suposições implícitas e explícitas nos testes	187
24.3 Suposições causais que conectam dados observados a efeitos causais	188
24.4 Diagnóstico e verificação	188
25 Seleção de testes	191
25.1 Multiverso de análises estatísticas	191
25.2 Escolha de testes para análise inferencial	191

26 Análise inferencial	199
26.1 Raciocínio inferencial	199
26.2 Hipóteses científicas	199
26.3 Hipóteses estatísticas	200
26.4 Testes de hipóteses	200
26.5 Intervalos de confiança e raciocínio de longo prazo	201
26.6 Comparações múltiplas	203
26.7 Inferência visual	207
26.8 Interpretação de análise inferencial	207
26.9 Erros de inferência I, II, S e M	208
27 Tamanho do efeito e P-valor	215
27.1 Tamanho do efeito	215
27.2 Tipos de tamanho do efeito	215
27.3 Conversão entre tamanhos do efeito	218
27.4 Efeitos bruto e padronizado	218
27.5 P-valor	218
27.6 P-valor de 2 ^a geração	219
27.7 Distribuição de confiança	222
27.8 Boas práticas	222
28 Testes estatísticos	225
28.1 Variáveis categóricas	225
28.2 Variáveis contínuas	225
29 Descrição	231
29.1 Análise de descrição	231
29.2 Estimação	232
30 Comparação	233
30.1 Análise inferencial de comparação	233
30.2 F-teste	233
31 Associação	235
31.1 Análise inferencial de associação	235
31.2 Associação bivariada	235
31.3 Associação multivariada	236
32 Correlação	237
32.1 Análise inferencial de correlação	237
32.2 Coeficientes de correlação	241
32.3 Colinearidade	244
32.4 Correlação entre conjuntos de variáveis	244
33 Regressão	249
33.1 Análise de regressão	249
33.2 Estruturas de análise de regressão	250
33.3 Tipos e famílias de regressão	250
33.4 Preparação de variáveis	257
33.5 Multicolinearidade	259
33.6 Redução de dimensionalidade	261
33.7 Seleção de variáveis em regressão	262
34 Redes	265

34.1 Análise de redes	265
<hr/>	
PARTE 6: MODELAGEM	267
35 Modelos	269
35.1 Modelos	269
35.2 Modelos estocásticos	269
35.3 Efeito fixo	270
35.4 Efeito aleatório	270
35.5 Efeito misto	272
35.6 Efeito principal	272
35.7 Efeito de interação	273
35.8 Efeito de mediação	275
35.9 Efeito de modificação	275
35.10 Preditores e desfechos	275
35.11 Desempenho e estabilidade de modelos	276
35.12 Comparação de modelos	276
35.13 Avaliação de modelos	276
35.14 Validação de modelos	278
35.15 Calibração de modelos	278
36 Modelagem temporal	279
36.1 Modelos temporais	279
37 Modelagem espacial	281
37.1 Modelos espaciais	281
38 Modelagem de sobrevida	283
38.1 Sobrevida	283
38.2 Análise de sobrevida	283
<hr/>	
PARTE 7: REPRESENTAÇÃO, APRENDIZADO DE MÁQUINA E INTELIGÊNCIA ARTIFICIAL	285
39 Representações	287
39.1 Representações de dados e extração de atributos	287
40 Inteligência artificial	289
40.1 Inteligência artificial	289
40.2 IA generativa e grandes modelos de linguagem	289
41 Aprendizado de máquina	293
41.1 Aprendizado de máquina	293
41.2 Tipos de aprendizado	293
41.3 Principais algoritmos	295
41.4 Regressão logística	296
41.5 Máquina de vetores de suporte	296
41.6 <i>K-nearest neighbours</i>	296
41.7 <i>K-means Clustering</i>	296
41.8 Árvores de decisão	296
41.9 Análise de componentes principais	299

41.10 <i>Random forests</i>	299
41.11 <i>Ensemble</i>	299
41.12 Desbalanceamento de classes	299
42 Redes neurais	301
42.1 Neurônios artificiais	301
42.2 Rede neural artificial	302
42.3 Funções de ativação	302
42.4 Espaço de decisão	304
42.5 Redes neurais multicamadas	304
<hr/>	
PARTE 8: PLANEJAMENTO DE ESTUDOS	307
43 Poder estatístico	309
43.1 Poder do teste	309
44 Tamanho da amostra	313
44.1 Tamanho da amostra	313
44.2 Saturação em pesquisas qualitativas	314
44.3 “Fome de dados”	317
44.4 Eventos por variável (EPV) em modelos preditivos	317
44.5 Cálculo do tamanho da amostra	318
44.6 Perdas de amostra	320
44.7 Ajustes no tamanho da amostra	320
44.8 Justificativa do tamanho da amostra	320
45 Plano de análise	323
45.1 Plano de análise estatística	323
45.2 Diretrizes para redação	323
<hr/>	
PARTE 9: DELINEAMENTOS E SÍNTESE DE EVIDÊNCIAS	325
46 Delineamento de estudos	327
46.1 Critérios de delineamento	327
46.2 Alocação	327
46.3 Cegamento	327
46.4 Pareamento	327
46.5 Aleatorização	328
46.6 Taxonomia de estudos	328
47 Simulação computacional	331
47.1 Simulações computacionais	331
47.2 Características	332
47.3 Métodos de simulação	332
47.4 Diretrizes para redação	336
48 Estudos observacionais	337
48.1 Características	337
48.2 Diretrizes para redação	337

49 Propriedades psicométricas	339
49.1 Características	339
49.2 Análise fatorial exploratória	339
49.3 Análise fatorial confirmatória	339
49.4 Validade de conteúdo	340
49.5 Validade de face	340
49.6 Validade do construto	340
49.7 Validade fatorial	340
49.8 Validade convergente	340
49.9 Validade discriminante	341
49.10 Validade de critério	341
49.11 Validade concorrente	341
49.12 Responsividade	341
49.13 Concordância	341
49.14 Confiabilidade	344
49.15 Diretrizes para redação	345
50 Desempenho diagnóstico	347
50.1 Características	347
50.2 Tabelas 2x2	347
50.3 Tabelas 2x3	351
50.4 Curvas ROC	352
50.5 Gráficos crosshair	354
50.6 Interpretação da validade de um teste	356
50.7 Diretrizes para redação	356
51 Ensaios quase-experimentais	357
51.1 Características	357
51.2 Diretrizes para redação	357
52 Ensaios experimentais	359
52.1 Ensaio clínico aleatorizado	359
52.2 Modelos de análise de comparação	360
52.3 Comparação na linha de base	361
52.4 Comparação intragrupos	362
52.5 Comparação entre grupos	362
52.6 Comparação de subgrupos	363
52.7 Efeito de interação	363
52.8 Ajuste de covariáveis	363
52.9 Imputação de dados perdidos	364
52.10 Diretrizes para redação	365
53 Ensaios cruzados	367
53.1 Características	367
53.2 Diretrizes para redação	367
54 N de 1	369
54.1 Ensaio N-de-1	369
54.2 Aspectos metodológicos	369
54.3 Limitações e cuidados	370
55 Revisão sistemática	371
55.1 Tipologia de revisões	371

55.2 Revisão sistemática de literatura	371
55.3 Tipos de revisão sistemática	372
55.4 Diretrizes para redação	373
56 Meta-análise	375
56.1 Características	375
56.2 Modelos de meta-análise	375
56.3 Conversão de Medidas em Meta-análises	376
56.4 Interpretação de efeitos em meta-análise	377
56.5 Interpretação do <i>forest plot</i>	378
56.6 Viés de publicação em meta-análises	381
56.7 Diretrizes para redação	383
57 Revisão guarda-chuva	385
57.1 Revisão guarda-chuva	385
57.2 Diretrizes para redação	385
58 Pesquisa qualitativa	387
58.1 Pesquisa qualitativa	387
58.2 Diretrizes para redação	387
<hr/>	
PARTE 10: COMUNICAÇÃO E RELATO CIENTÍFICO	389
59 Redação de resultados	391
59.1 Resultados da análise estatística	391
59.2 Diretrizes e Listas	391
60 Diretrizes e Listas	393
60.1 Diretrizes	393
60.2 Listas de verificação	394
61 Fontes externas	395
61.1 Fontes de informação externas	395
<hr/>	
REFERÊNCIAS	397
Referências	399

RASCUNHO

Lista de Figuras

1.1	Exemplos de espaço amostral discreto. Superior: Todas as faces de uma moeda. Inferior: Todas as faces de um dado.	4
1.2	Exemplos de evento de experimento. Superior: 1 lançamento de 1 moeda. Inferior: 1 lançamento de 1 dado.	5
1.3	Espaço de eventos: União dos eventos face = 3 e face = 4 de um dado.	5
1.4	Superior: Eventos independentes. Inferior: Eventos dependentes.	7
1.5	Esquerda: Evento (face = 4). Direita: Experimentos de 1 lançamento de 1 dado (superior), 3 lançamentos de 1 dado (central), 10 lançamentos de 1 dado (inferior).	8
1.6	Esquerda: Histogramas de uma variável aleatória com distribuição uniforme ($N = 100$). Direita: Histogramas da média de 100 amostras de tamanhos 5, 50, 500 e 5000 tomadas da população com reposição e igual probabilidade.	10
1.7	Esquerda: Histogramas de lançamento de 1 dado com distribuição uniforme ($N = 100$). Direita: Histogramas da média de 100 amostras de tamanhos 5, 50, 500 e 5000 tomadas da população com reposição e igual probabilidade.	11
1.8	Esquerda: Histogramas de lançamento de 1 moeda com distribuição uniforme ($N = 100$). Direita: Histogramas da média de 100 amostras de tamanhos 5, 50, 500 e 5000 tomadas da população com reposição e igual probabilidade.	12
1.9	Representação gráfica da regressão para a média em medidas repetidas. A segunda medida (dado 2) é mais próxima da média (valor real) do que a primeira medida (dado 1).	14
2.1	Representação esquemática da amostragem: seleção de uma população para a amostra.	19
2.2	Representação esquemática do erro de amostragem: seleção de várias amostras independentes de uma população.	20
2.3	Representação esquemática da amostragem de uma população para a amostra.	21
2.4	Representação esquemática da reamostragem de uma amostra.	22
2.5	Representação esquemática da subamostragem de uma amostra.	23
2.6	Representação esquemática da superamostragem de uma população.	23
3.1	Mapa mental da relação entre o pensamento estatístico e o pensamento metodológico.	26
3.2	Representação esquemática da generalização de uma amostra para a população.	30
4.1	Interface do RStudio. Fonte: https://docs.posit.co/ide/user/	33
8.1	Paradoxo de Simpson representado com dados simulados. Os pontos no gráfico representam observações individuais e as linhas de tendência representam as regressões lineares ajustadas para os dados desagregados da população e agregados por subpopulação.	52
10.1	Transformações de variáveis com assimetria à direita (Original, Raiz quadrada, Log natural, Log10, Inversa).	65
10.2	Transformações de variáveis com assimetria à esquerda (Original, Reflexão + Raiz quadrada, Reflexão + Log natural, Reflexão + Log10, Reflexão + Inversa).	66

10.3 Comparação entre variáveis originais e padronizadas (Z-score e Min-Max).	68
12.1 Exemplo de instrumento com 3 itens tipo Likert com 5 categorias cada.	80
12.2 Ilustração do conceito de erro de medida em um modelo simples com erro normal. A linha tracejada indica o valor verdadeiro (desconhecido na prática). A área sombreada representa a probabilidade de cair na faixa $ X - \theta \leq \varepsilon$, que é > 0 . A probabilidade de 'acertar no ponto' $X = \theta$ é 0 (linha tracejada).	85
12.3 Acurácia e precisão como propriedades de uma medida.	86
12.4 Viés e variabilidade de uma medida.	87
14.1 Representação gráfica de dados perdidos completamente ao acaso (MCAR) em um estudo randomizado controlado (RCT).	94
14.2 Representação gráfica de dados perdidos ao acaso (MAR) em um estudo randomizado controlado (RCT).	95
14.3 Representação gráfica de dados perdidos não ao acaso (MNAR) em um estudo randomizado controlado (RCT).	96
14.4 Impacto de métodos de imputação na distribuição de uma variável contínua com dados perdidos.	99
16.1 Histogramas com diferentes métodos de binning.: Sturges, Scott e Freedman-Diaconis.	106
16.2 Distribuições discretas e suas funções de probabilidade.	108
16.3 Distribuições contínuas básicas e suas funções de densidade.	109
16.4 Distribuições contínuas aproximadas e suas funções de densidade.	110
16.5 Distribuições contínuas aproximadas e suas funções de densidade.	111
16.6 Distribuições contínuas para inferência e suas funções de densidade.	112
16.7 Distribuições contínuas para dados específicos e suas funções de densidade.	113
16.8 Distribuições contínuas para probabilidades e proporções e suas funções de densidade.	114
16.9 Distribuições contínuas com caudas pesadas e suas funções de densidade.	115
16.10 Distribuições e funções de probabilidade.	116
16.11 Distribuição normal e métodos de visualização e testes de normalidade.	117
16.12 Distribuição normal bivariada e amostra simulada com histogramas marginais.	118
16.13 Distribuições unimodal, bimodal e multimodal.	121
16.14 Parâmetros de tendência central em distribuições assimétricas e normais.	122
16.15 Parâmetros de dispersão em distribuições normais.	123
16.16 Regressão linear com valores extremos.	126
16.17 Parâmetros de distribuição: Assimetria e Curtose.	127
16.18 Parâmetros de distribuição: Curtose em distribuições simétricas (normal vs. uniforme).	128
18.1 Séries temporais e autocorrelação de duas séries simuladas com fraca e forte autocorrelação.	136
19.1 Exemplo de gráfico que mostra os dados brutos junto com um resumo estatístico (média e dispersão).	143
19.2 Exemplos de gráficos com barras de erro e dados brutos.	145
19.3 Exemplos de gráficos com barras de erro e dados brutos em diferentes cenários.	146
19.4 Gráfico de barras simples representando frequências por categoria.	147
19.5 Gráfico de barras empilhadas representando frequências por categoria.	148
19.6 Gráficos de barras representando médias, barras de erro e dados individuais.	149
19.7 Histograma da variável 'valor'.	150
19.8 Gráfico de densidade da variável 'valor'.	151
19.9 Boxplot por grupo.	152
19.10 Violin plot por grupo.	153
19.11 Gráfico de pontos da variável 'valor'.	154
19.12 Gráfico de dispersão representando a relação entre duas variáveis.	155
19.13 Gráfico de bolhas representando a relação entre três variáveis.	156
19.14 Sankey plot representando fluxos entre categorias.	156

19.15 Gráfico de categorias paralelas (parcats) representando transições entre categorias ao longo do tempo.	157
19.16 Gráfico de pares representando correlações entre múltiplas variáveis.	157
19.17 Gráfico spaghetti representando dados longitudinais.	158
19.18 Gráfico de linha representando uma série temporal.	159
19.19 Gráfico de correlação entre duas variáveis com linha de tendência.	160
19.20 Matriz de dispersão representando relações entre múltiplas variáveis.	161
19.21 Mapa de calor da correlação entre variáveis.	162
19.22 Gráfico radar representando múltiplas variáveis.	162
19.23 Exemplo de gráfico interativo com Plotly.	164
20.1 Regressão linear com valores discrepantes	166
20.2 Regressão linear com valores influentes.	169
20.3 Alavancagem vs Resíduos Padronizados com distância de Cook para análise da influência de pontos.	170
20.4 Boxplots comparando dados originais e dados Winsorizados.	171
22.1 Padrões causais básicos: independência, cadeia, garfo e colisor.	180
24.1 Diagnóstico de regressão para avaliar suposições do modelo: linearidade, normalidade dos resíduos, homocedasticidade e alavancagem.	189
26.1 Simulação ilustrativa de intervalos de confiança (IC) em 100 experimentos independentes, cada um com 1.000 observações amostradas de uma população normal padrão (média = 0, desvio-padrão = 1). Os ICs são construídos no nível de 95%. O gráfico superior mostra os ICs individuais para cada experimento, indicando se o IC cobre ou não a média verdadeira ($\mu = 0$). O gráfico inferior apresenta a distribuição das médias amostrais obtidas nos experimentos, juntamente com o IC teórico para a média populacional. Observe que aproximadamente 95% dos ICs individuais cobrem a média verdadeira, ilustrando o conceito de cobertura no longo prazo associado aos intervalos de confiança.	202
26.2 Representação gráfica de um teste de hipótese unicaudal à direita, aplicado quando se busca evidência de efeitos positivos (valores significativamente maiores que o esperado sob H_0).	204
26.3 Representação gráfica de um teste de hipótese unicaudal à esquerda, aplicado quando se busca evidência de efeitos negativos (valores significativamente menores que o esperado sob H_0).	205
26.4 Representação gráfica de um teste de hipótese bicaudal, aplicado quando se busca evidência de efeitos positivos ou negativos (valores significativamente diferentes do esperado sob H_0).	206
26.5 Representação gráfica dos erros tipo I e tipo II em um teste de hipótese (bicaudal).	209
26.6 Erro tipo I: Distribuição dos p-valores em 100 testes de hipótese de amostras aleatórias de tamanho 30. A linha vermelha pontilhada indica o nível de significância estatística de 0,05.	210
26.7 Erro tipo II: Distribuição dos p-valores em 100 testes de hipótese de amostras aleatórias de tamanho 10. A linha vermelha pontilhada indica o nível de significância estatística de 0,05.	211
26.8 Representação gráfica do erro tipo S (sinal) em um teste de hipótese (bicaudal).	212
26.9 Representação gráfica do erro tipo M (magnitude) em um teste de hipótese (bicaudal).	213
27.1 Visualização espacial de $p < 0,05$ (5 quadrados aleatórios em 100).	220
27.2 Distribuição de confiança para o tamanho do efeito estimado.	223
32.1 Exemplo de diferentes forças e direção de correlação entre duas variáveis X e Y.	238
32.2 Gráfico de dispersão do Quarteto de Anscombe para representação gráfica de conjuntos de dados bivariados com parâmetros quase idênticos e relações muito distintas.	240
32.3 Exemplo de análise de correlação canônica (CCA) entre dois conjuntos de variáveis.	247
33.1 Regressão linear.	251
33.2 Regressão polinomial.	252
33.3 Regressão não-linear.	253

33.4 Regressão logística.	254
33.5 Regressão multinomial	255
33.6 Regressão de Poisson.	256
33.7 Regressão ridge.	258
33.8 Multicolinearidade entre variáveis candidatas em modelos de regressão multivariável.	260
 35.1 Cadeia de Markov com 3 estados (a, b, c) e suas probabilidades de transição.	270
35.2 Trajetória de estados e proporção acumulada por estado em uma cadeia de Markov com 3 estados (a, b, c).	271
35.3 Efeitos fixos, aleatórios e mistos em dados simulados com paradoxo de Simpson. As linhas vermelhas representam os efeitos dentro dos grupos, enquanto as linhas cinza e preta representam os efeitos globais (naive e fixo, respectivamente). O modelo misto (linhas coloridas) captura os efeitos dentro dos grupos sem extrapolar além dos dados observados.	272
35.4 Análise de efeito de interação (direta) entre grupos e tempo. Retas paralelas sugerem ausência de efeito de interação.	273
35.5 Análise de efeito de interação (inversa) entre grupos e tempo. Retas paralelas sugerem ausência de efeito de interação.	274
35.6 Exemplos de ajuste de modelos de regressão linear simples ($y \times x$) com diferentes níveis de ruído (R^2). Cada painel mostra a reta ajustada (cinza) e os valores observados (pontos). Os valores anotados indicam o coeficiente angular simulado (β), o coeficiente angular estimado ($\hat{\beta}$) e o R^2 observado.	277
 38.1 Curvas de Kaplan–Meier simuladas para dois grupos (controle e tratamento).	284
40.1 Representação esquemática de um modelo de linguagem grande (LLM)	290
41.1 Mapa mental de algoritmos de aprendizado de máquina.	294
41.2 Exemplo de árvore de decisão para predizer depressão a partir de idade, tabagismo e sintomas. .	297
41.3 Comparação entre modelos de regressão logística e árvore de decisão.	298
 42.1 Representação esquemática de um neurônio computacional.	301
42.2 Representação esquemática de uma rede neural simples com camada de entrada e saída. .	302
42.3 Gráficos das funções de ativação mais comuns.	303
42.4 Espaço de decisão de um perceptron (regressão logística).	304
42.5 Comparação do espaço de decisão entre um modelo linear (regressão logística) e um modelo não linear (MLP).	305
42.6 Representação esquemática de uma rede neural multicamadas com 2 camadas escondidas além das camadas de entrada e saída.	306
 44.1 Curvas de poder para testes t (quantitativo). Linhas sólidas: $\alpha=0,05$ tracejadas: $\alpha=0,01$ linhas horizontais em 80% e 90% de poder.	315
44.2 Curvas de saturação para estudos qualitativos de descoberta de temas.	316
 46.1 Alocação 1:1 entre dois grupos de participantes	328
47.1 Dados simulados a partir de diferentes distribuições: Normal(0,1), Binomial(1,0.4), Poisson(2) e Exponencial(1).	333
47.2 Convergência do histograma para a PDF teórica da Normal(0,1) com o aumento do tamanho amostral ($n = 10, 100, 1000, 10000$).	334
47.3 Convergência da média e do desvio-padrão amostral para os valores teóricos (0 e 1, respectivamente) com o aumento do tamanho amostral ($n = 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000$).	335

50.1 Árvore de frequência do desempenho diagnóstico de uma tabela de confusão 2x2 representando um método novo (dicotômico) comparado ao método padrão-ouro ou referência (dicotômico)	348
50.2 Trade-off entre sensibilidade e especificidade em função do limiar de probabilidade (t) para um modelo de classificação	349
50.3 Curva ROC (Receiver Operating Characteristic) para um modelos de classificação com diferentes desempenhos diagnósticos	354
50.4 Gráfico *crosshair* em espaço ROC (Receiver Operating Characteristic) para 15 estudos simulados de desempenho diagnóstico	355
56.1 Comparação entre modelos de efeito fixo e aleatório com 10 ensaios clínicos simulados	376
56.2 Forest plot de uma meta-análise de efeito fixo com 10 ensaios clínicos simulados	378
56.3 Forest plot de uma meta-análise de efeito aleatório com 10 ensaios clínicos simulados	379
56.4 Forest plots ilustrativos para faixas usuais de I^2	380
56.5 Gráficos de funil simulados com baixa e alta heterogeneidade	381

RASCUNHO

Lista de Tabelas

9.1	Classificação das práticas questionáveis em pesquisa segundo sua intencionalidade.	56
10.1	Tabela de variáveis indicadoras (dummy variables) criadas a partir de variáveis categóricas Sexo e Grupo.	71
12.1	Descrição dos itens tipo Likert do instrumento.	79
12.2	Dados brutos com medidas únicas.	81
12.3	Dados brutos com medidas repetidas.	82
12.4	Dados brutos com medidas repetidas agregadas.	82
12.5	Dados brutos com medidas seriadas não agregadas.	83
12.6	Dados brutos com medidas seriadas não agregadas.	83
12.7	Dados brutos com medidas múltiplas.	84
13.1	Estrutura básica de uma tabela de dados.	89
13.2	Formato longo para medidas repetidas (múltiplas linhas por sujeito; colunas = variáveis)	90
13.3	Formato largo para medidas repetidas (1 variável; colunas = tempos)	90
13.4	Formatação recomendada para tabela de dados.	91
13.5	Formatação não recomendada para tabela de dados.	91
14.1	Simulação de uma amostra (n=10) de um ensaio clínico aleatorizado (dados com perdas aleatórias).	93
19.1	Quantidade de casas decimais e dígitos significativos.	138
19.2	Valores originais, arredondamentos e erros de arredondamento por casas decimais.	138
19.3	Características da amostra por grupo.	141
19.4	Exemplo clássico de apresentação suscetível à Falácia da ‘Tabela 2’.	142
19.5	Exposição (OR; 95% IC) com e sem ajuste.	142
25.1	Tabela de seleção de testes estatísticos: Descrição.	192
25.2	Tabela de seleção de testes estatísticos: Comparação.	192
25.3	Tabela de seleção de testes estatísticos: Associação.	193
25.4	Tabela de seleção de testes estatísticos: Predição.	194
25.5	Tabela de seleção de testes estatísticos: Contagem.	194
25.6	Tabela de seleção de testes estatísticos: Sobrevida.	195
25.7	Tabela de seleção de testes estatísticos: Concordância.	195
25.8	Tabela de seleção de testes estatísticos: Ajuste multivariado.	195
25.9	Tabela de seleção de testes estatísticos: Desempenho diagnóstico.	196
25.10	Tabela de seleção de testes estatísticos: Diagnóstico.	196
25.11	Tabela de seleção de testes estatísticos: Transição de estados.	197
26.1	Tabela de erros tipos I e II de inferência estatística.	208
26.2	Tabela de erro tipo S de inferência estatística.	209

26.3 Tabela de erro tipo M de inferência estatística.	213
27.1 Comparação entre p-valor (bicaudal, inferido do IC 95%) e SGPV (p_δ) nos cenários simulados.	221
28.1 Teste Qui-quadrado (com correção de Yates)	226
28.2 Teste Qui-quadrado (sem correção de Yates)	226
28.3 Teste exato de Fisher	226
28.4 Teste de McNemar	227
28.5 Teste Q de Cochran	227
28.6 Teste de Cochran–Armitage	227
28.7 Medidas de associação	228
28.8 Teste t de Student	228
28.9 Teste t de Welch	228
28.10 Teste de Mann–Whitney (Wilcoxon rank-sum)	228
28.11 Teste de Wilcoxon (signed-rank)	229
28.12 Análise de variância de um fator	229
28.13 Análise de variância de Welch	229
28.14 Teste de Kruskal–Wallis	230
32.1 Quarteto de Anscombe.	239
32.2 Análise descritiva do Quarteto de Anscombe demonstrando os conjuntos de dados bivariados com parâmetros quase idênticos.	239
35.1 Métricas de desempenho do modelo de regressão linear.	278
49.1 Tabela de confusão sobre propriedades psicométricas de instrumentos.	339
49.2 Tabela de confusão 2x2 para análise de concordância de testes e variáveis dicotômicas.	341
49.3 Tabela de confusão 3x3 para análise de concordância de testes e variáveis dicotômicas.	343
50.1 Tabela de confusão 2x2 para análise de desempenho diagnóstico de testes e variáveis dicotômicas.	347
50.2 Probabilidades calculados a partir da tabela de confusão 2x2 para análise de desempenho diagnóstico de testes e variáveis dicotômicas.	350
50.3 Tabela de confusão 3-vias (2×3) com totais: referência vs decisão (3WD).	351

RASCUNHO

RASCUNHO

Dedicatória

Esta obra é dedicada a todos que, em princípio, buscam conhecimento para melhorar a qualidade da pesquisa científica — seja a sua própria, a de colegas ou a de desconhecidos — mas, em última análise, desejam mesmo prover melhores condições de saúde e desenvolvimento da sociedade.

Dedico também ao leitor eventual que chegou aqui por acaso.

RASCUNHO

Agradecimentos

Este trabalho não seria possível sem o apoio e suporte da minha esposa Daniele, minha irmã Mônica, meu pai José Victorino, minha mãe Angela (*in memoriam*) e meus filhos Giovanna, Victor e Lucas.

RASCUNHO

Apresentação

No âmbito da análise estatística de dados, os processos envolvidos são marcados por uma série de escolhas críticas. Estas decisões abrangem considerações metodológicas e ações operacionais que moldam toda a jornada analítica. Deve-se selecionar, cuidadosamente, um delineamento de estudo para enfrentar os desafios únicos colocados por um projeto de pesquisa. Além disso, a escolha de métodos estatísticos adequados para lidar com os dados gerados pelo delineamento escolhido tem um peso importante. Estas decisões necessitam de uma base construída sobre as evidências mais convincentes da literatura existente e na adesão a práticas sólidas de investigação.

Interpretar os resultados destas análises não é uma tarefa simples. Confiar apenas na formação educacional convencional, no bom senso e na intuição para decifrar tabelas e gráficos pode revelar-se inadequado. Interpretações errôneas podem gerar consequências indesejáveis, incluindo a utilização de testes diagnósticos imprecisos ou o endosso de tratamentos ineficazes.

Este livro emerge do reconhecimento desses desafios.

A proposta gira em torno da organização de um compêndio abrangente de métodos e técnicas de ponta, para análise estatística de dados em pesquisa científica, apresentados em formato de perguntas e respostas. Esse formato promove um diálogo direto e objetivo com o leitor, respondendo a dúvidas comumente colocadas por alunos de graduação, pós-graduação lato sensu, pós-graduação stricto sensu (mestrado e doutorado), bem como por pesquisadores.

O objetivo geral de cada capítulo é elucidar as questões metodológicas fundamentais: “*O que é?*”, “*Por que usar?*”, “*Quando usar?*”, “*Quando não usar?*” e “*Como fazer?*”. Em cada capítulo, diversas questões específicas são propostas e respondidas sistematicamente, permitindo ao leitor uma melhor elaboração do conteúdo e resultado do seu trabalho. Todos eles com citações de fontes confiáveis referências, que podem ser consultadas para aprofundamento e verificação das informações apresentadas — um total de 511 referências foram incluídas para apoiar as informações e recomendações apresentadas.

Os capítulos foram organizados para seguir uma progressão de conceitos e aplicações. Embora sejam fragmentados para maior clareza instrucional, as referências cruzadas ajudam a mitigar a fragmentação do conteúdo e reforçar a interconexão dos tópicos.

O público-alvo comprehende pesquisadores, professores, analistas de dados, profissionais e estudantes que regularmente lidam com a tomada de decisões em pesquisa. Os estudantes de pós-graduação encontrarão aqui uma obra repleta de exemplos para adaptar na análise dos dados de seus projetos de pesquisa. Professores de graduação e pós-graduação terão acesso a uma obra didática de referência, direcionada para seus alunos. Pesquisadores e analistas de dados iniciantes descobrirão um valioso acervo de informações e referências para a construção de projetos e manuscritos. Pesquisadores e os cientistas mais experientes podem recorrer às referências e esclarecimentos mais atuais sobre vieses, paradoxos, mitos e mal práticas em pesquisa. E mesmo os leitores não familiarizados ainda com as técnicas de análise de dados em pesquisa terão a oportunidade de apreciar o papel fundamental de colocar e responder suas perguntas na busca do conhecimento científico.

Arthur de Sá Ferreira

RASCUNHO

Sobre o autor



Arthur de Sá Ferreira

Obtive minha Graduação em Fisioterapia pela Universidade Federal do Rio de Janeiro (1999), Formação em Acupuntura pela Academia Brasileira de Arte e Ciência Oriental (2001), Mestrado em Engenharia Biomédica pela Universidade Federal do Rio de Janeiro (2002) e Doutorado em Engenharia Biomédica pela Universidade Federal do Rio de Janeiro (2006).

Tenho experiência em docência no ensino superior, atuei com professor da graduação em cursos de Fisioterapia, Enfermagem e Odontologia, entre outros (2001-2018); pós-graduação *lato sensu* em Fisioterapia (2001-atual) e *stricto sensu* níveis mestrado e doutorado (2010-atual).

Como pesquisador, sou Professor Adjunto do Centro Universitário Augusto Motta (UNISUAM), atuando nos Programas de Pós-graduação em Ciências da Reabilitação (PPGCR; 2009-atual) e Desenvolvimento Local (PPGDL; 2018-atual). Também sou pesquisador do Instituto D'Or de Pesquisa e Ensino (IDOR; 2024-atual). Fundei o Laboratório de Simulação Computacional e Modelagem em Reabilitação (LSCMR) em 2012, onde desenvolvo projetos de pesquisa principalmente nos seguintes temas: Bioestatística, Modelagem e simulação computacional, Processamento de sinais biomédicos, Movimento funcional humano, Medicina tradicional (chinesa), Distúrbios musculoesqueléticos, Doenças cardiovasculares e Doenças respiratórias.

Dentre os editais públicos que obtive financiamento, destaco os Programas Jovem Cientista do Nossa Estado (JCNE; 2012-2015; 2015-2017) e Cientista do Nossa Estado (2021-atual) da Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ; e Bolsista Produtividade em Pesquisa pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; 2021-atual).

Como gestor, estou na Coordenação do Programa de Pós-Graduação *stricto sensu* em Ciências da Reabilitação (PPGCR; 2016-atual). Coordeno o Curso Superior de Tecnologia em Radiologia da Faculdade IDOR de Ciências Médicas (IDOR; 2024-atual). Atuei como coordenador do Comitê de Ética em Pesquisa (CEP) do Centro Universitário Augusto Motta (UNISUAM; 2020-2024) e como Coordenador do Curso de Graduação em Fisioterapia da Universidade Salgado de Oliveira (UNIVERSO; 2004-2009).

Sou membro da Associação Brasileira de Pesquisa e Pós-Graduação em Fisioterapia (ABRAPG-FT) (2007-atual), Consórcio Acadêmico Brasileiro de Saúde Integrativa (CABSIN) (2019-atual), e Royal Statistical Society (RSS) (2021-atual). Fui membro do Committee on Publication Ethics (COPE) (2018-2024).

Componho o corpo editorial e de revisores de periódicos nacionais e internacionais como Scientific Reports, Frontiers in Rehabilitation Sciences, The Journal of Clinical Hypertension, Chinese Journal of Integrative Medicine, Journal of Integrative Medicine, Brazilian Journal of Physical Therapy, Fisioterapia e Pesquisa.

RASCUNHO

PARTE 1: PENSAMENTO CIENTÍFICO

Conceitos essenciais para pensar cientificamente e evitar armadilhas comuns

RASCUNHO

Capítulo 1

Pensamento probabilístico

1.1 Experimento

1.1.1 O que é um experimento?

- Um experimento é um processo de simulação ou medição cujo resultado é chamado de desfecho.¹
- Tentativa se refere a uma repetição de um experimento.¹

1.1.2 O que é um experimento aleatório?

- Em um experimento aleatório, o desfecho de cada tentativa é imprevisível.¹

1.2 Espaço amostral e eventos discretos

1.2.1 O que é espaço amostral discreto?

- O espaço amostral S de um experimento aleatório é definido como o conjunto de todos os desfechos possíveis de um experimento.¹
- Em probabilidade discreta, o espaço amostral S pode ser enumerado e contado.¹

1.2.2 O que é evento discreto?

- Um evento E é um único desfecho ou uma coleção de desfechos.¹
- Um evento E é um subconjunto do espaço amostral S de um experimento.¹

1.2.3 O que é espaço de eventos discretos?

- Um espaço de eventos E_s também é um subconjunto do espaço amostral S de um experimento.¹
- A união de dois eventos $E_1 \cup E_2$ é o conjunto de todos os desfechos que estão em E_1 , em E_2 ou em ambos.¹
- A intersecção de dois eventos $E_1 \cap E_2$, ou evento conjunto, é o conjunto de todos os desfechos que estão em ambos os eventos.¹
- O complemento de um evento E^C consiste em todos os desfechos que não estão incluídos no evento E .¹

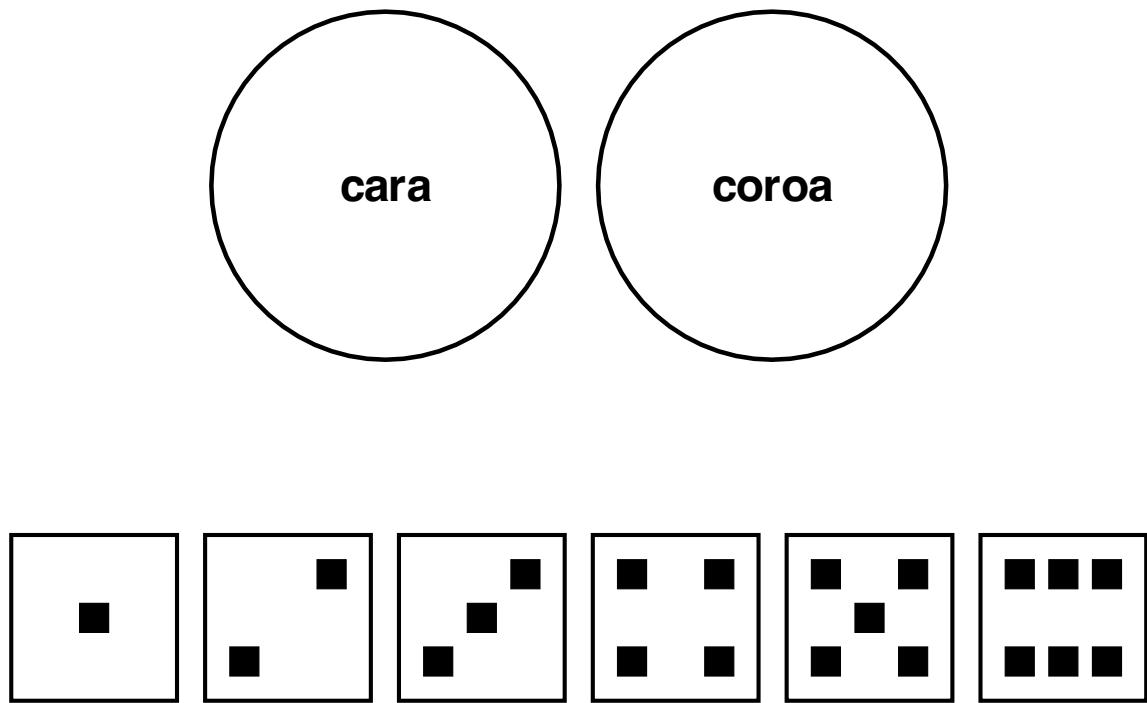


Figura 1.1: Exemplos de espaço amostral discreto. Superior: Todas as faces de uma moeda. Inferior: Todas as faces de um dado.

1.3 Espaço amostral e eventos contínuos

1.3.1 O que é espaço amostral contínuo?

- ?

1.3.2 O que é evento contínuo?

- ?

1.3.3 O que é espaço de eventos contínuo?

- ?

1.4 Probabilidade

1.4.1 O que é probabilidade?

- Com um espaço amostral S finito e não vazio de desfechos igualmente prováveis, a probabilidade P de um evento E é a razão entre o número de desfechos no evento E e o número de desfechos no espaço amostral S (1.1).¹

$$P(E) = \frac{\text{número de desfechos em } E}{\text{número de desfechos em } S} \quad (1.1)$$

- Um evento E impossível não contém um desfecho e, portanto, nunca ocorre: $P(E) = 0$.^{1,2}

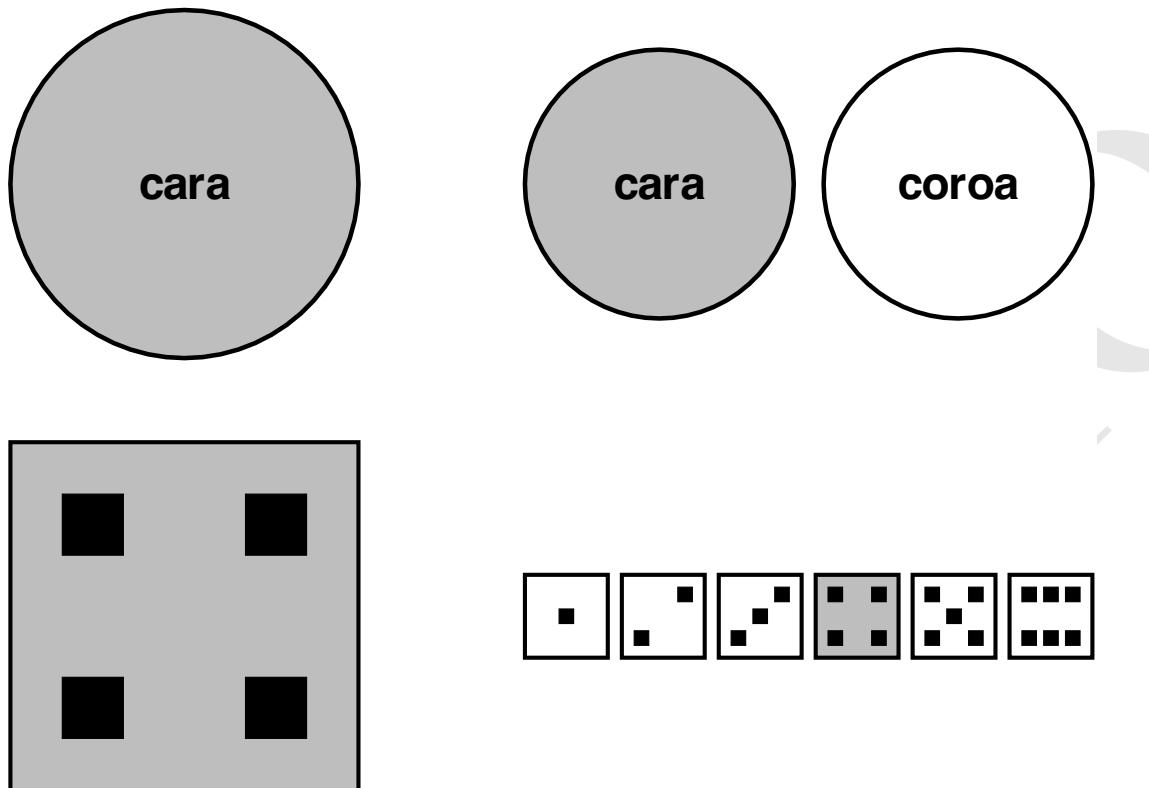


Figura 1.2: Exemplos de evento de experimento. Superior: 1 lançamento de 1 moeda. Inferior: 1 lançamento de 1 dado.

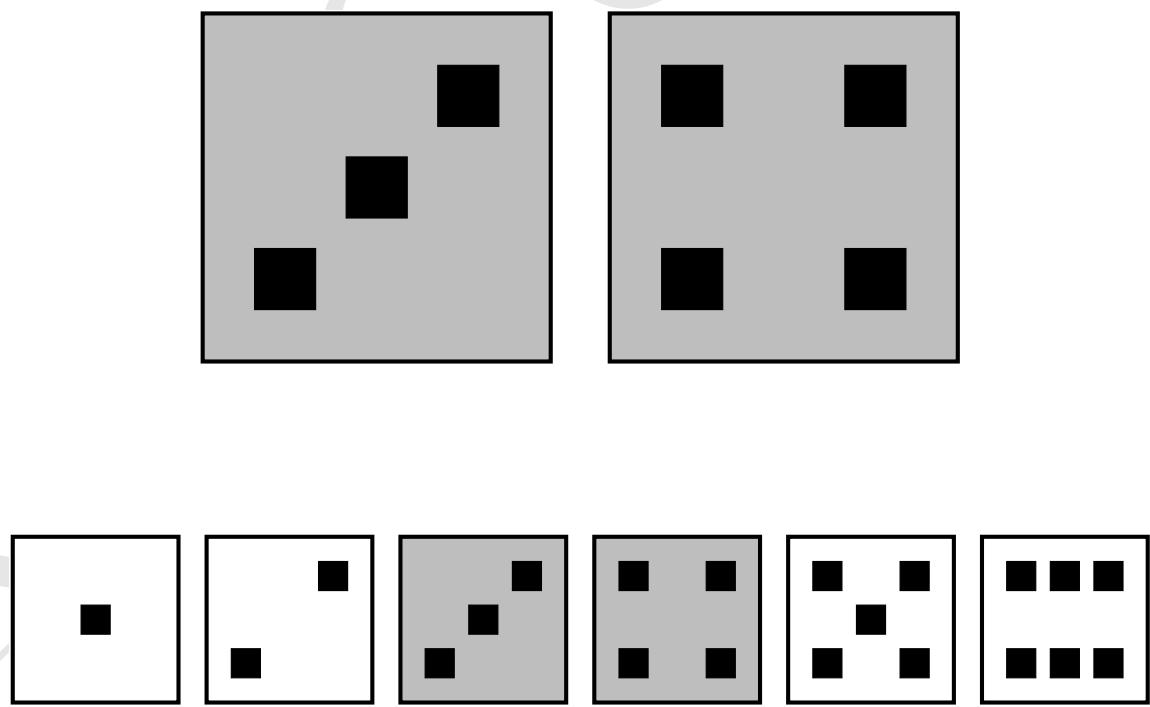


Figura 1.3: Espaço de eventos: União dos eventos face = 3 e face = 4 de um dado.

- Um evento E é certo consiste em qualquer um dos desfechos possíveis e, portanto, sempre ocorre: $P(E) = 1$.¹

1.4.2 Quais são os axiomas da probabilidade?

- A probabilidade de um evento é um número real que satisfaz os seguintes axiomas descritos por Andrei Nikolaevich Kolmogorov em 1950:^{1,2}
 - Axioma I. Probabilidades de um evento E são números não-negativos: $P(E) \geq 0$.
 - Axioma II. Probabilidade de todos os eventos do espaço amostral A ocorrerem é 100%: $P(S) = 1$.
 - Axioma III. A probabilidade de um conjunto k de eventos mutuamente exclusivos é igual a soma da probabilidade de cada evento: $P(E_1 \cup E_2 \cup \dots \cup E_k) = P(E_1) + P(E_2) + \dots + P(E_k)$.
- Os axiomas possuem as seguintes consequências:¹
 - A soma da probabilidade de dois eventos que dividem o espaço amostral é 100%: $P(E) + P(E)^C = 1$.
 - O valor máximo de probabilidade de um evento é 100%: $P(S) \leq 1$.
 - A probabilidade é uma função não decrescente do número de desfechos de um evento.

1.5 Independência e probabilidade

1.5.1 O que é independência em estatística?

- Em experimentos aleatórios, é comum assumir que os eventos de tentativas separadas são independentes devido a independência física de eventos e experimentos.¹
- Se a ocorrência do evento E_1 não tiver efeito na ocorrência do evento E_2 , os eventos E_1 e E_2 são considerados estatisticamente independentes.¹
- Eventos são mutuamente exclusivos, ou disjuntos, se a ocorrência de um exclui a ocorrência dos outros.¹
- Se dois eventos E_1 e E_2 são mutuamente exclusivos, então os eventos E_1 e E_2 não podem ocorrer ao mesmo tempo e, portanto, são eventos independentes.¹
- Em experimentos independentes, o desfecho de uma tentativa é independente dos desfechos de outras tentativas, passadas e/ou futuras. Uma tentativa em um experimento aleatório é independente se a probabilidade de cada desfecho possível não mudar de tentativa para tentativa.¹

1.5.2 O que é probabilidade marginal?

- Probabilidade marginal é a probabilidade de ocorrência de um evento E independentemente da(s) probabilidade(s) de outro(s) evento(s).¹

1.5.3 O que é probabilidade conjunta?

- Probabilidade conjunta é a probabilidade de ocorrência de dois ou mais eventos independentes E_1, E_2, \dots, E_k , independentemente da(s) probabilidade(s) de outro(s) evento(s).¹
- Se a probabilidade conjunta dos eventos é nula ($E_1 \cup E_2 = 0$), esses dois eventos E_1 e E_2 são mutuamente exclusivos ou disjuntos.¹

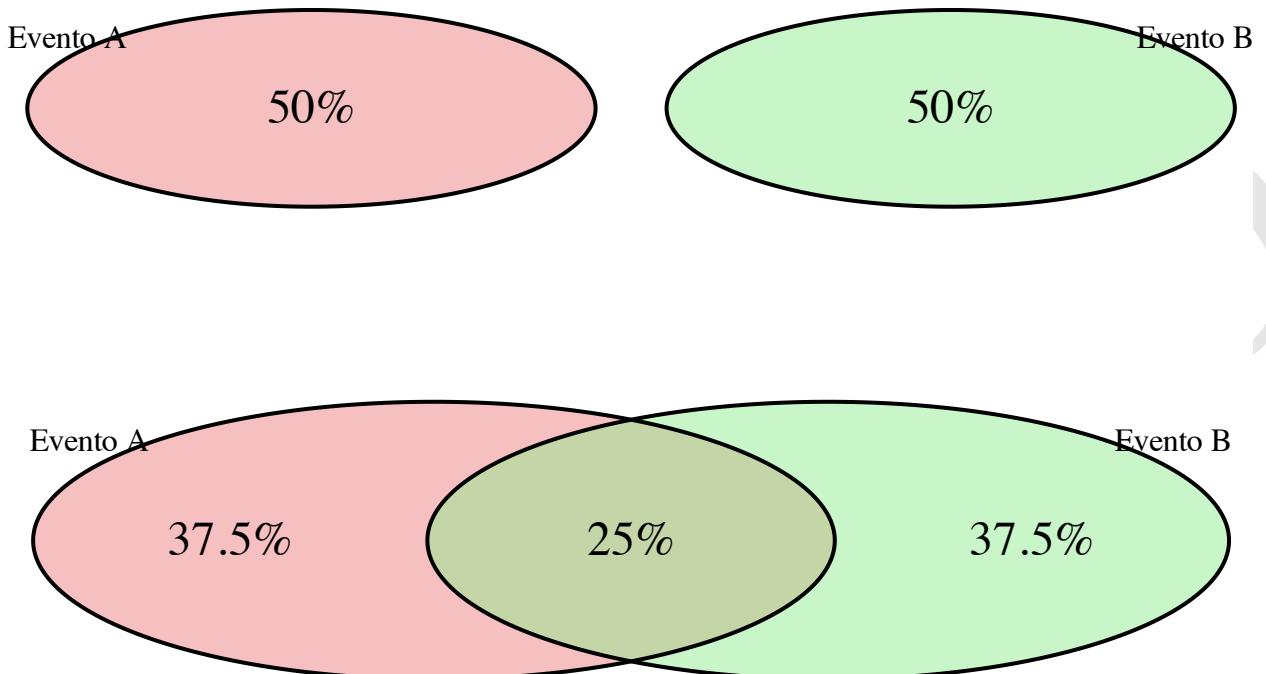


Figura 1.4: Superior: Eventos independentes. Inferior: Eventos dependentes.

1.5.4 O que é probabilidade condicional?

- Probabilidade condicional é a probabilidade de ocorrência do evento E_2 quando se sabe que o evento E_1 já ocorreu $P(E_2|E_1)$.¹
- A probabilidade condicional $P(E_2|E_1)$ representa que a ocorrência do evento E_1 fornece informação sobre a ocorrência do evento E_2 .¹
- Se a ocorrência do evento E_1 tiver alguma influência na ocorrência do evento E_2 , então a probabilidade condicional do evento E_2 dado o evento E_1 pode ser maior ou menor do que a probabilidade marginal.¹

1.6 Leis dos números anômalos

1.6.1 O que é a lei dos números anômalos?

- A lei dos números anômalos — lei de Benford — é uma distribuição de probabilidade que descreve a frequência de ocorrência do primeiro dígito em muitos conjuntos de dados do mundo real.³

1.7 Leis dos pequenos números

1.7.1 O que é a lei dos pequenos números?

- A crença exagerada na probabilidade de replicar com sucesso os achados de um estudo, pela tendência de se considerar uma amostra como representativa da população.⁴
- A crença na lei dos pequenos números se refere à tendência de superestimar a estabilidade das estimativas provenientes de estudos com amostras pequenas.⁵
- Quando se percebe um padrão, pode não ser possível identificar se tal padrão é real.⁶

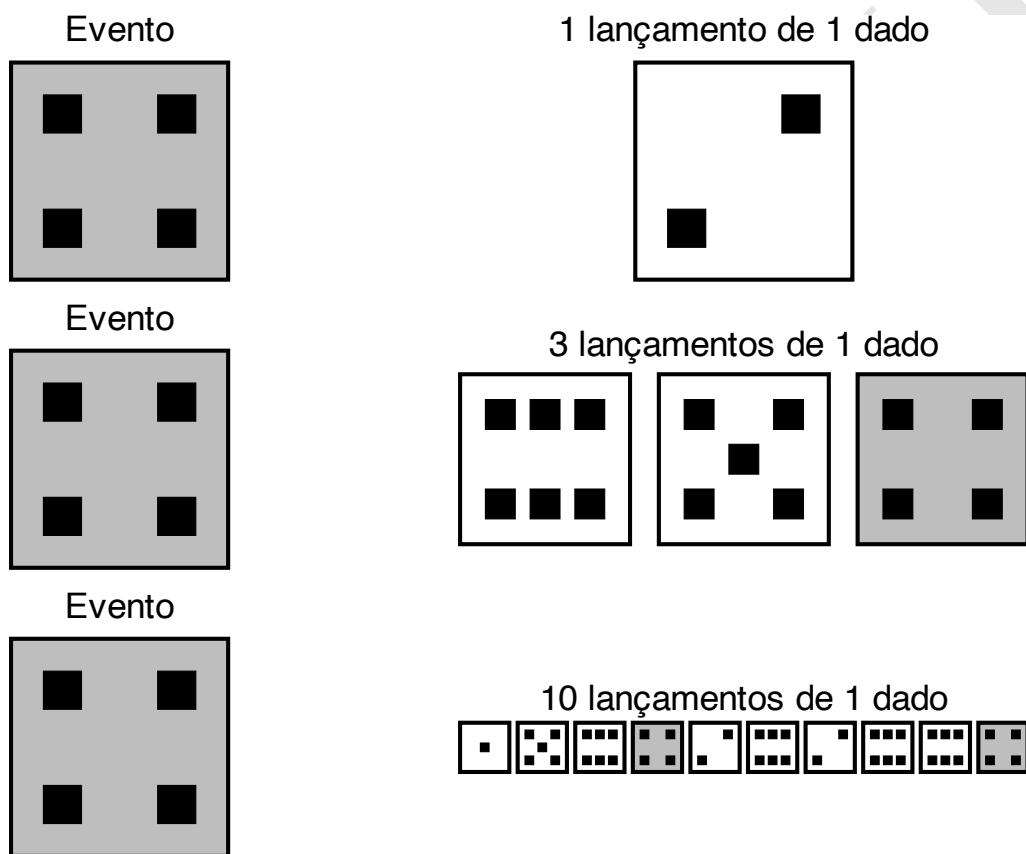


Figura 1.5: Esquerda: Evento (face = 4). Direita: Experimentos de 1 lançamento de 1 dado (superior), 3 lançamentos de 1 dado (central), 10 lançamentos de 1 dado (inferior).

1.7.2 Quais são as versões da lei dos pequenos números?

- 1a Lei Forte dos Pequenos Números: “Não há pequenos números suficientes para atender às muitas demandas que lhes são feitas”.⁶
- 2a Lei Forte dos Pequenos Números: “Quando dois números parecem iguais, não são necessariamente assim”.⁷

1.8 Leis dos grandes números

1.8.1 O que é a lei dos grandes números?

- A lei dos grandes números descreve que, ao realizar o mesmo experimento E um grande número de vezes (n), a média μ dos resultados obtidos tende a se aproximar do valor esperado $E[\bar{X}]$ à medida que mais experimentos forem realizados ($n \rightarrow \infty$).⁸
- De acordo com a lei dos grandes números, a média amostral converge para a média populacional à medida que o tamanho da amostra aumenta.⁹

1.8.2 Quais são as versões da lei dos grandes números?

- Lei Fraca dos Grandes Números (de Poisson): ““?“”.
- Lei Fraca dos Grandes Números (de Bernoulli): ““?“”.
- Lei Forte dos Grandes Números: ““?“”.

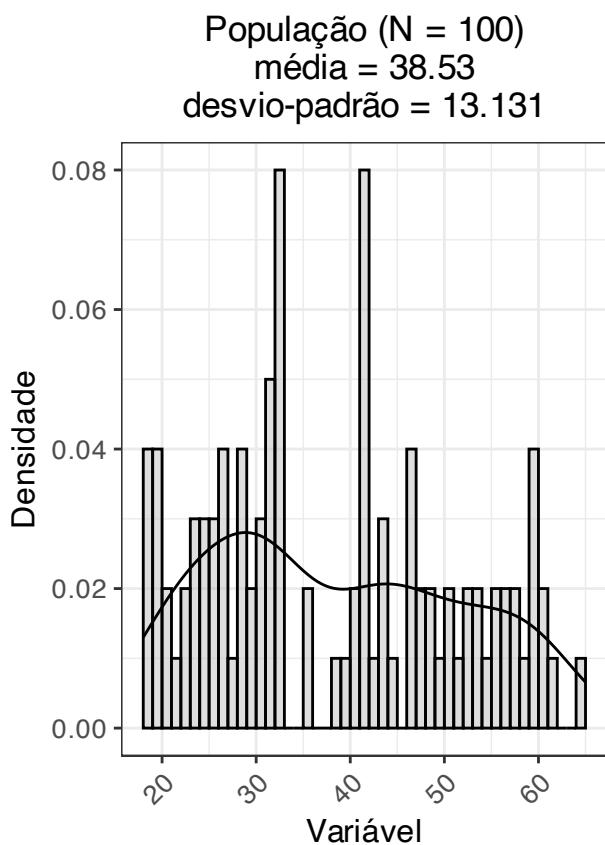
1.9 Teorema central do limite

1.9.1 O que é teorema central do limite?

- O teorema central do limite (1.2) afirma que, para uma amostra aleatória de tamanho n de uma população com valor esperado igual à média $E[\bar{X}_i] = \mu$ e variância $Var[\bar{X}_i]$ igual a σ^2 , a distribuição amostral da média de uma variável \bar{X} se aproxima de uma distribuição normal N com média μ e variância σ^2/n à medida que n aumenta ($n \rightarrow \infty$):⁸

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{n \rightarrow \infty} N(0, \sigma^2) \quad (1.2)$$

- O teorema central do limite demonstra que se o tamanho da amostra n for suficientemente grande, a distribuição amostral das médias obtidas utilizando reamostragem com substituição será aproximadamente normal, com média μ e variância σ^2/n , independentemente da distribuição da população.⁸
- No exemplo abaixo, uma variável aleatória numérica com distribuição uniforme no espaço amostral $S = [18; 65]$ tem média $\mu = 38.53$ e variância $\sigma^2 = 172.433$. A distribuição amostral da média de 100 amostras de tamanho 5, 50, 500 e 5000 tomadas da população com reposição e igual probabilidade se aproxima de uma distribuição normal com média $\mu = 38.493$ e variância $\sigma^2 = 0.038$, independentemente da distribuição da população:
- Em outro exemplo, o lançamento de um dado com distribuição uniforme no espaço amostral $S = \{1, 2, 3, 4, 5, 6\}$ tem média $\mu = 3.77$ e variância $\sigma^2 = 3.169$. A distribuição amostral da média de 100 amostras de tamanho 5, 50, 500 e 5000 tomadas da população com reposição e igual probabilidade se aproxima de uma distribuição normal com média $\mu = 3.767$ e variância $\sigma^2 = 0.001$, independentemente da distribuição da população:



Histogramas representando as média de 100 amostras de tamanhos diferentes tomadas da população com reposição e igual probabilidade

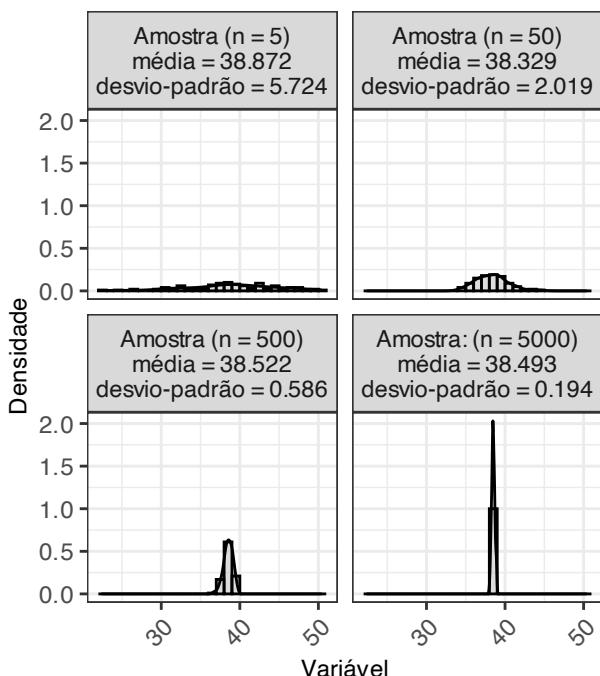


Figura 1.6: Esquerda: Histogramas de uma variável aleatória com distribuição uniforme ($N = 100$). Direita: Histogramas da média de 100 amostras de tamanhos 5, 50, 500 e 5000 tomadas da população com reposição e igual probabilidade.

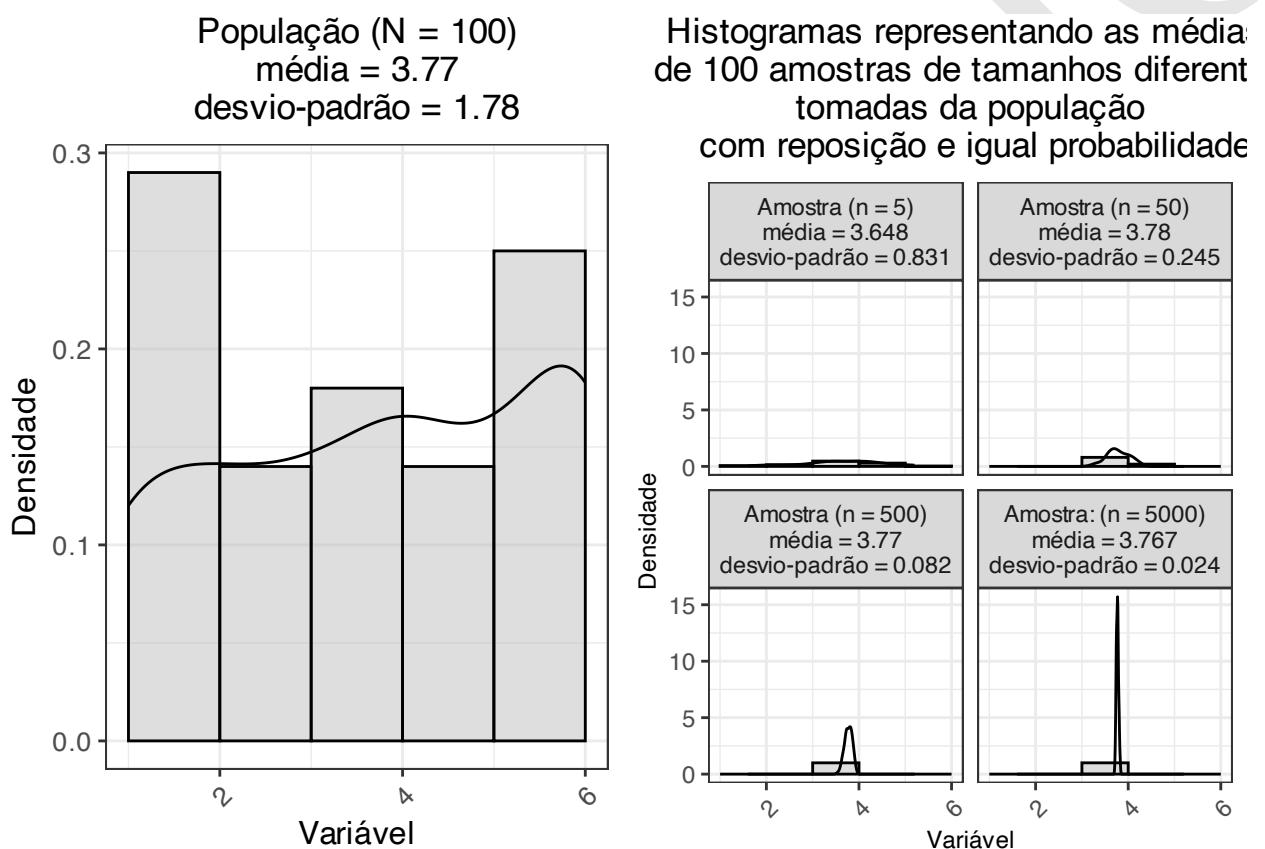


Figura 1.7: Esquerda: Histogramas de lançamento de 1 dado com distribuição uniforme ($N = 100$). Direita: Histogramas da média de 100 amostras de tamanhos 5, 50, 500 e 5000 tomadas da população com reposição e igual probabilidade.

- Mais um exemplo, o lançamento de uma moeda com distribuição uniforme no espaço amostral $S = \{0, 1\}$ — codificado para $succeso = 1$ e $insucesso = 0$ — tem média $\mu = 0.53$ e variância $\sigma^2 = 0.252$. A distribuição amostral da média de 100 amostras de tamanho 5, 50, 500 e 5000 tomadas da população com reposição e igual probabilidade se aproxima de uma distribuição normal com média $\mu = 0.531$ e variância $\sigma^2 = 0$, independentemente da distribuição da população:

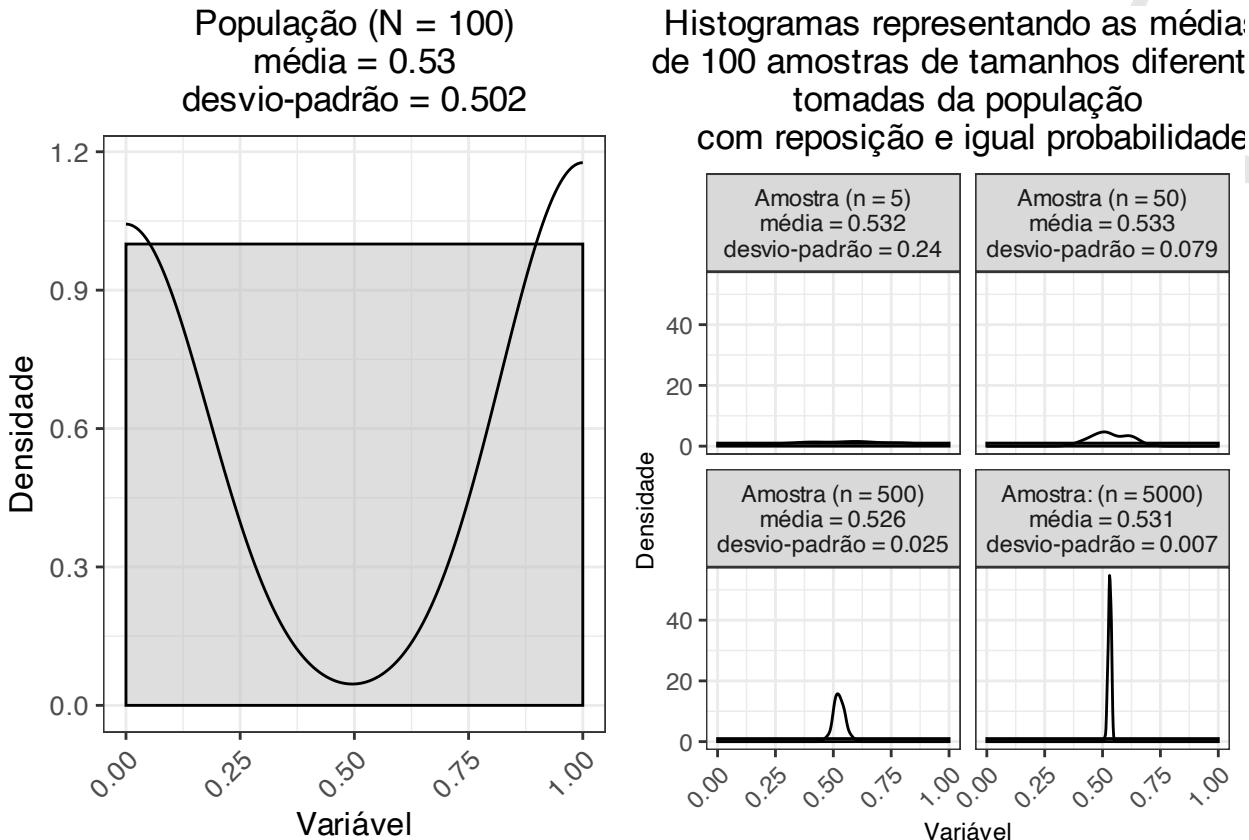


Figura 1.8: Esquerda: Histogramas de lançamento de 1 moeda com distribuição uniforme ($N = 100$). Direita: Histogramas da média de 100 amostras de tamanhos 5, 50, 500 e 5000 tomadas da população com reposição e igual probabilidade.

1.9.2 Quais as condições de validade do teorema central do limite?

- As condições de validade do teorema central do limite são:⁸
 - As variáveis aleatórias devem ser independentes e identicamente distribuídas (*independent and identically distributed* ou i.i.d.);
 - As variáveis aleatórias devem ter média μ e variância σ^2 finitas;
 - O tamanho da amostra deve ser suficientemente grande (geralmente, $n \geq 30$).

1.9.3 Qual a relação entre a lei dos grandes números e o teorema central do limite?

- A lei dos grandes números é um precursor do teorema central do limite, pois estabelece que a média da amostra se torna cada vez mais próxima da média populacional (isto é, mais representativa) à medida que o tamanho da amostra aumenta, e o teorema central do limite demonstra que a distribuição da soma das

variáveis aleatórias se aproxima de uma distribuição normal também à medida que o tamanho da amostra aumenta.⁸

1.9.4 Qual a relevância do teorema central do limite para a análise estatística?

- O teorema central do limite explica porque os testes paramétricos têm maior poder estatístico do que os testes não paramétricos, os quais não requerem suposições de distribuição de probabilidade.⁸
- O teorema central do limite implica que os métodos estatísticos que se aplicam a distribuições normais podem ser aplicados a outras distribuições quando suas suposições são satisfeitas.⁸
- Como o teorema central do limite determina a distribuição amostral Z (1.3) das médias com tamanho amostral suficientemente grande, a média pode ser padronizada para uma distribuição normal com média 0 e variância 1, $N(0, 1)$.⁸

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad (1.3)$$

- Para amostras com $n \geq 30$, a distribuição amostral Student- t se aproxima da distribuição normal padrão Z e, portanto, as suposições sobre a distribuição populacional não são mais necessárias de acordo com o teorema central do limite. Neste cenário, a suposição de distribuição normal pode ser usada para a distribuição de probabilidade.⁸

1.10 Regressão para a média

1.10.1 O que é regressão para a média?

- Regressão para a média⁹ é um fenômeno estatístico que ocorre quando uma variável aleatória X é medida na mesma unidade de análise em dois ou mais momentos diferentes, X_1, X_2, \dots, X_t e X_t é mais próximo da média populacional do que X_1 , ou seja, $E(X_t)$ é mais próxima de $E(X)$ do que $E(X_1)$ é de $E(X)$.¹⁰
- O valor real — sem erros aleatório ou sistemático — em geral não é conhecido, mas pode ser estimado pela média de várias observações.¹⁰
- Regressão para a média pode ocorrer em qualquer pesquisa cujo delineamento envolva medidas repetidas.¹¹
- Em medidas repetidas, a média de várias observações é mais próxima da média verdadeira do que qualquer observação individual, pois o erro aleatório é reduzido pela média.¹⁰
- Valores extremos — em direção ao mínimo ou máximo — em uma medição inicial tendem a ser seguidos por valores mais próximos da média (valor real) na medição subsequente.¹⁰
- No exemplo abaixo, a 2a medida (dado 2 = 121) é mais próxima da média (valor real = 120) do que a 1a medida (dado 1 = 118):

1.10.2 Qual a causa da regressão para a média?

- A regressão para a média pode ser atribuída ao erro aleatório, que é a variação não sistemática nos valores observados em torno de uma média verdadeira (por exemplo, erro de medição aleatório ou variações aleatórias em um participante).¹⁰
- Regressão para a média é uma consequência da observação de que dados extremos não se repetem com frequência.¹¹
- Deve-se assumir que a regressão para a média ocorreu até que os dados mostrem o contrário.¹⁰

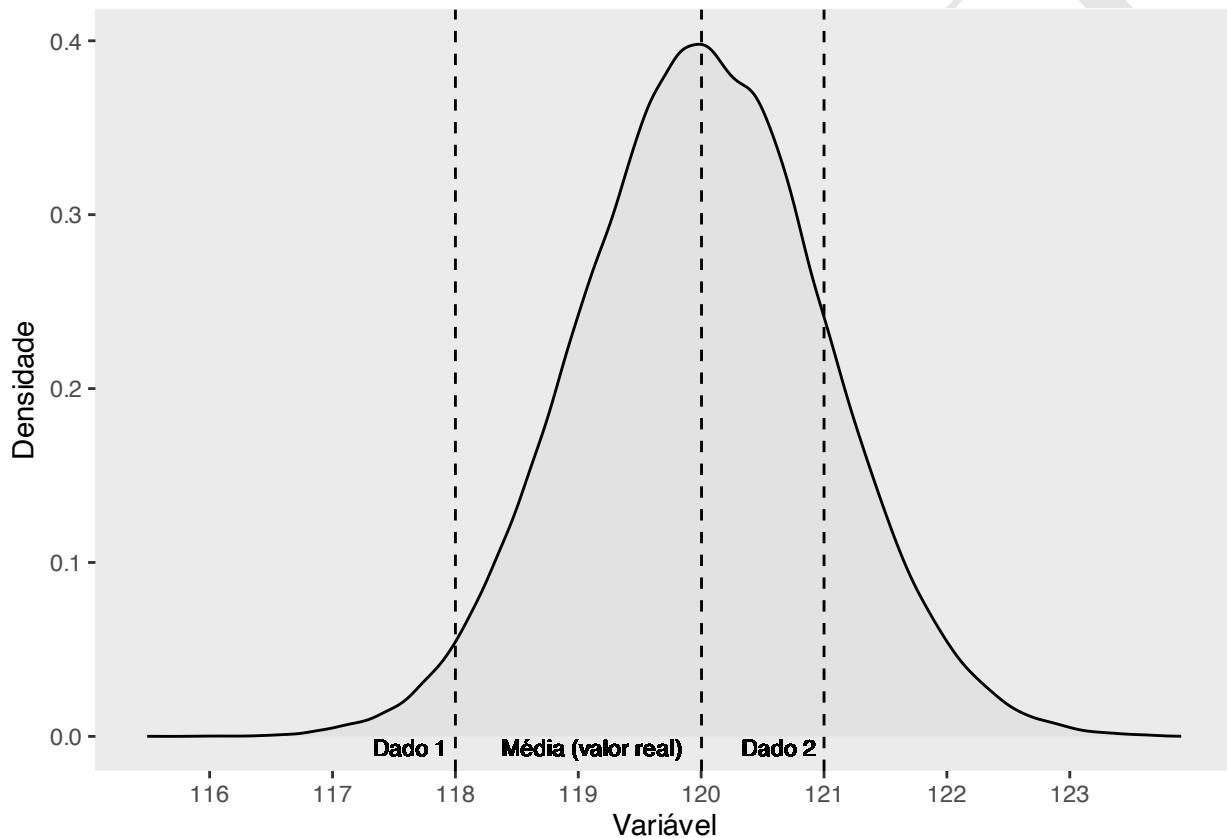


Figura 1.9: Representação gráfica da regressão para a média em medidas repetidas. A segunda medida (dado 2) é mais próxima da média (valor real) do que a primeira medida (dado 1).

1.10.3 Por que detectar o fenômeno de regressão para a média?

- A regressão para a média pode levar a conclusões errôneas sobre a eficácia de uma intervenção, pois a mudança observada pode ser devida ao erro aleatório e não ao tratamento.¹¹

1.10.4 Como detectar o fenômeno de regressão para a média?

- O fenômeno de regressão para a média pode ser detectado por meio de gráfico de dispersão da diferença (estudos transversais) ou mudança (estudos longitudinais) versus os valores da 1a medida.¹⁰



O pacote *regtomean*¹² fornece as funções *cordata*^a para calcular a correlação entre medidas tipo antes-e-depois e *meechua_reg*^b para ajustar modelos lineares de regressão.

^a<https://www.rdocumentation.org/packages/regtomean/versions/1.1/topics/cordata>

^bhttps://www.rdocumentation.org/packages/regtomean/versions/1.1/topics/meechua_reg

1.10.5 Como o fenômeno de regressão para a média pode ser evitado?

- Aloque os participantes de modo aleatório nos grupos de tratamento e controle para reduzir o fenômeno de regressão para a média.¹⁰
- Selecione participantes com base em medidas repetidas ao invés de medidas únicas.¹⁰

RASCUNHO

Capítulo 2

Pensamento estatístico

2.1 Unidade de análise

2.1.1 O que é unidade de análise?

- A unidade de análise (ou unidade experimental) de pesquisas na área de saúde geralmente é o indivíduo.¹³
- A unidade de análise também pode ser a instituição em estudos multicêntricos (ex.: hospitais, clínicas) ou um estudo publicado em meta-análise (ex.: ensaios clínicos).¹³

2.1.2 Por que identificar a unidade de análise de um estudo?

- É fundamental identificar corretamente a unidade de análise para evitar inflação do tamanho da amostra (ex.: medidas bilaterais resultando em o dobro de participantes), violações de suposições dos testes de hipótese (ex.: independência entre medidas e/ou unidade de análise) e resultados espúrios em testes de hipótese (ex.: P-valores menores que aqueles observados se a amostra não estivesse inflada).^{13,14}

2.1.3 Que medidas podem ser obtidas da unidade de análise de um estudo?

- Da unidade de análise podem ser coletadas informações em medidas únicas, repetidas, seriadas ou múltiplas.

2.2 População

2.2.1 O que é população?

- População — ou população-alvo — refere-se ao conjunto completo sobre o qual se pretende obter informações.^{15,16}
- População é metodologicamente delimitada pelos critérios de inclusão e exclusão do estudo.¹⁵
- Em estudos observacionais, inicialmente as características geográficas e/ou demográficas, por exemplo, definem a população a ser estudada.¹⁵
- Em estudos analíticos, a população é inicialmente definida pelos objetivos da pesquisa e, posteriormente, as observações são realizadas na amostra.¹⁵

2.2.2 O que é representatividade e por que ela importa?

- Representatividade refere-se ao grau em que uma amostra reflete com fidelidade as características da população de referência.¹⁶
- Quando a amostra contém menos indivíduos do que o número mínimo necessário, mas mantém a representatividade, a inferência estatística ainda é possível, embora possa haver redução da precisão e/ou do poder estatístico para detectar os efeitos.¹⁶
- Amostras não representativas comprometem a validade da inferência estatística, mesmo quando o tamanho da amostra atende aos requisitos de poder da análise.¹⁶

2.3 Amostra

2.3.1 O que é amostra?

- Amostra é uma parte finita da população do estudo.^{15,16}
- Em pesquisa científica, utilizam-se dados de uma amostra de participantes (ou outras unidades de análise) para realizar inferências sobre a população.¹⁷

2.3.2 Por que usar dados de amostras?

- Estudos com amostras, em vez de censos, são preferíveis por diversas razões, dentre elas: questões éticas; limitações orçamentárias; desafios logísticos; restrição de tempo; e tamanho populacional desconhecido.¹⁶
- Dados de uma amostra de tamanho suficiente e características representativas podem ser utilizados para inferência sobre uma população.⁸
- Em geral, amostras de tamanhos maiores possuem médias mais próximas da média populacional e menores variâncias.⁸

2.4 Amostragem

2.4.1 O que é amostragem?

- Amostragem é o processo pelo qual se seleciona uma parte de uma população para constituir a amostra que será efetivamente estudada.¹⁶

2.4.2 Quais métodos de amostragem são usados para obter uma amostra da população?

- O método de amostragem é geralmente definido pelas condições de viabilidade do estudo, no que diz respeito a acesso aos participantes, ao tempo de execução e aos custos envolvidos, entre outras.¹⁵
- Não-probabilísticas ou intencionais:^{15,16}
 - Bola de neve.
 - Conveniência.
 - Participantes encaminhados.
 - Proposital.
- Probabilísticas:^{15,16}
 - Simples.
 - Sistemática.

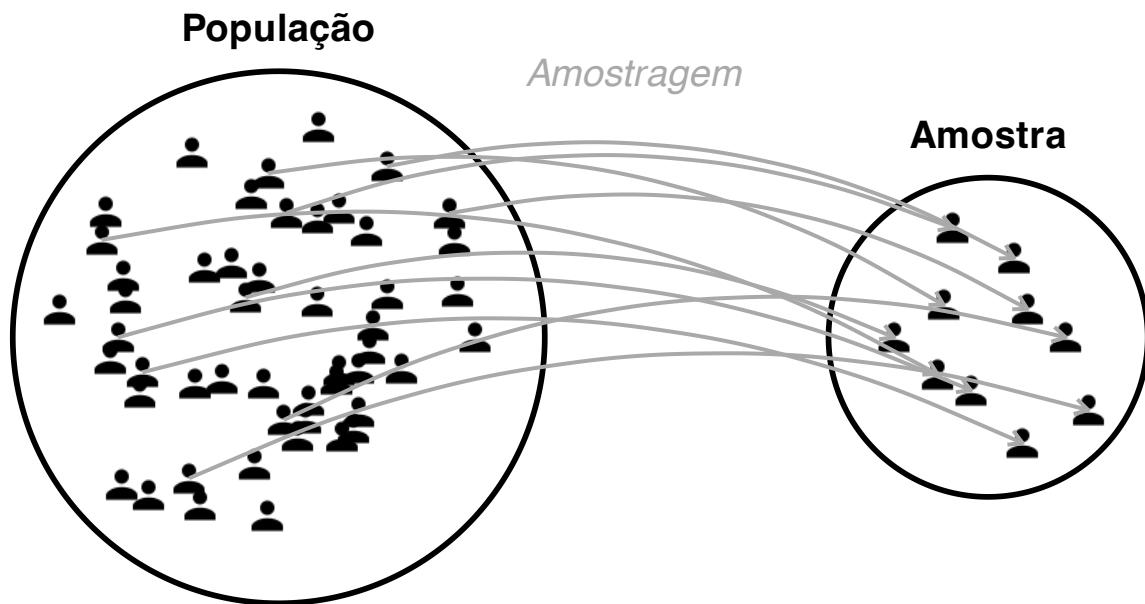


Figura 2.1: Representação esquemática da amostragem: seleção de uma população para a amostra.

- Multiestágio.
- Estratificada.
- Agregada.

2.4.3 O que é erro de amostragem?

- Erro de amostragem é a variação natural entre os resultados obtidos a partir de uma amostra e os resultados que seriam obtidos caso toda a população fosse examinada. Reflete o grau de incerteza inerente à generalização de uma amostra para a população.¹⁶

2.5 Reamostragem

2.5.1 O que é reamostragem?

- Reamostragem é um procedimento que cria vários conjuntos de dados sorteados a partir de um conjunto de dados real - a amostra da população - sem a necessidade de fazer suposições sobre os dados e suas distribuições.¹⁷
- O procedimento é repetido várias vezes para usar a variabilidade dos resultados para obter um intervalo de confiança do parâmetro no nível de significância α pré-estabelecido.¹⁷

2.5.2 Por que utilizar reamostragem?

- Quando se dispõe de dados de apenas 1 amostra, as diversas suposições que são feitas podem não ser atingidas.¹⁷
- Procedimentos de reamostragem produzem um conjunto de observações escolhidas aleatoriamente da amostra, igualmente representativo da população original.¹⁷
- Procedimentos de reamostragem permitem estimar o erro-padrão e intervalos de confiança sem a necessidade de tais suposições, sendo, portanto, um conjunto de procedimentos não-paramétricos.¹⁷

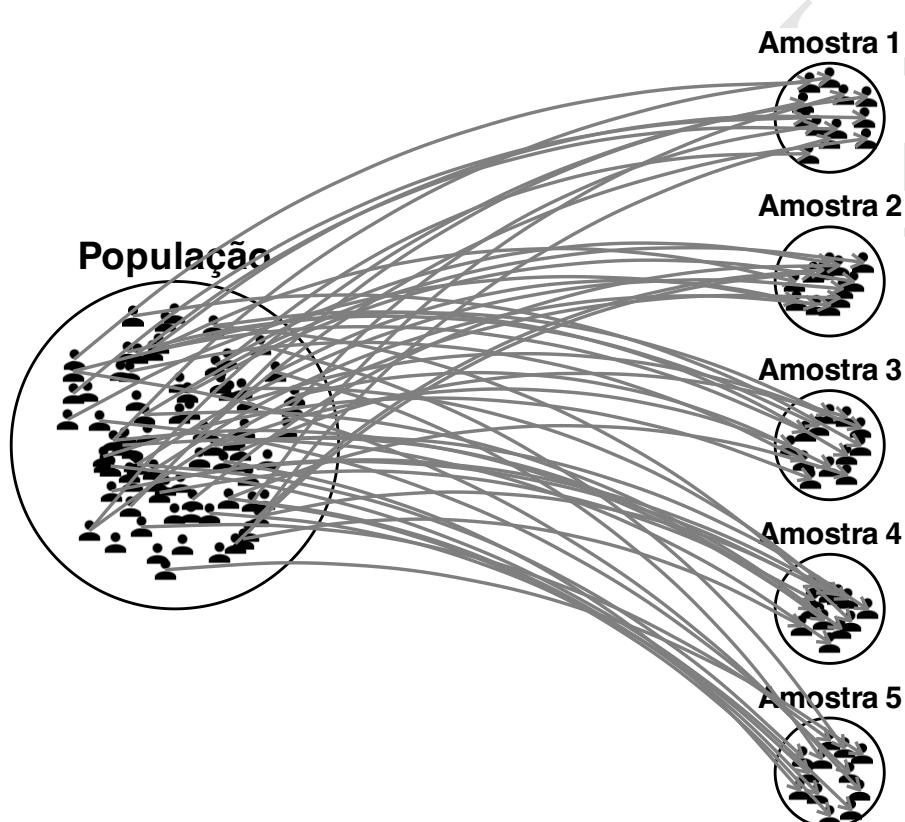


Figura 2.2: Representação esquemática do erro de amostragem: seleção de várias amostras independentes de uma população.

Mesma população → amostras diferentes
Cada amostra tem sua própria média

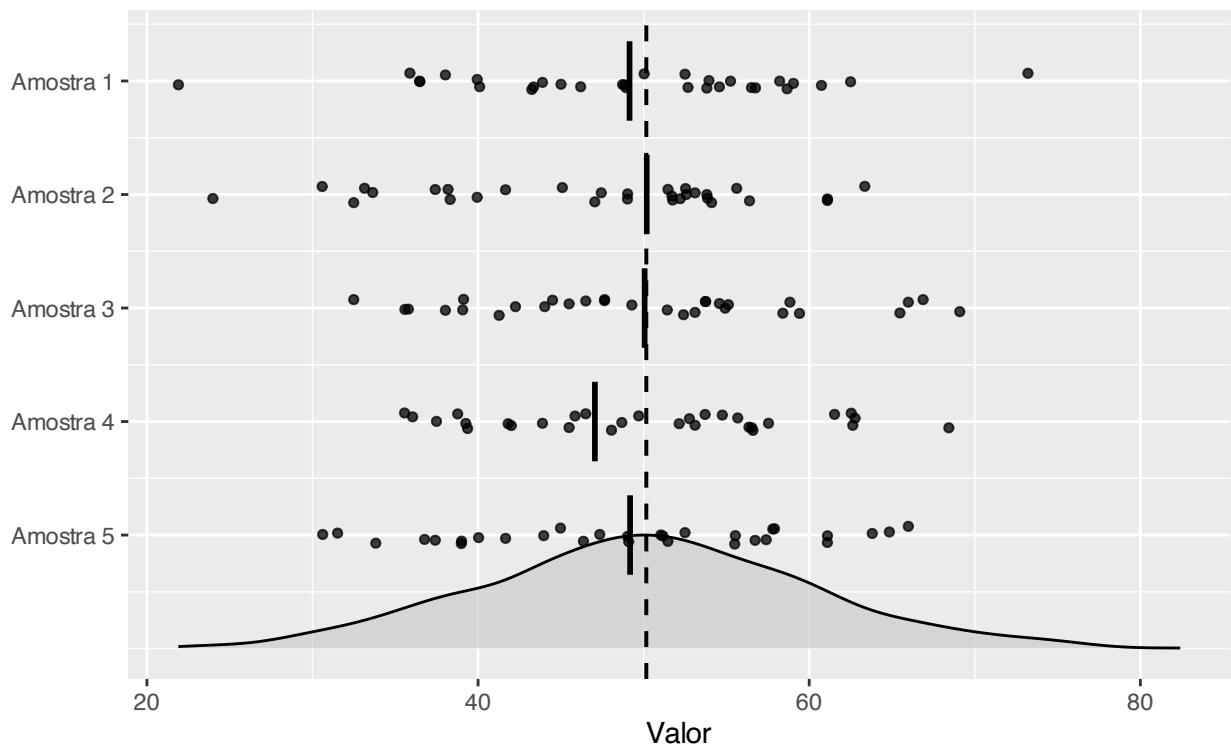


Figura 2.3: Representação esquemática da amostragem de uma população para a amostra.

2.5.3 Quais procedimentos de reamostragem podem ser realizados?

- *Bootstrap*: Cada iteração gera uma amostra *bootstrap* do mesmo tamanho do conjunto de dados original escolhendo aleatoriamente observações reais, uma de cada vez. Cada observação tem chances iguais de ser escolhida a cada vez, portanto, algumas observações serão escolhidas mais de uma vez e outras nem serão escolhidas.¹⁷

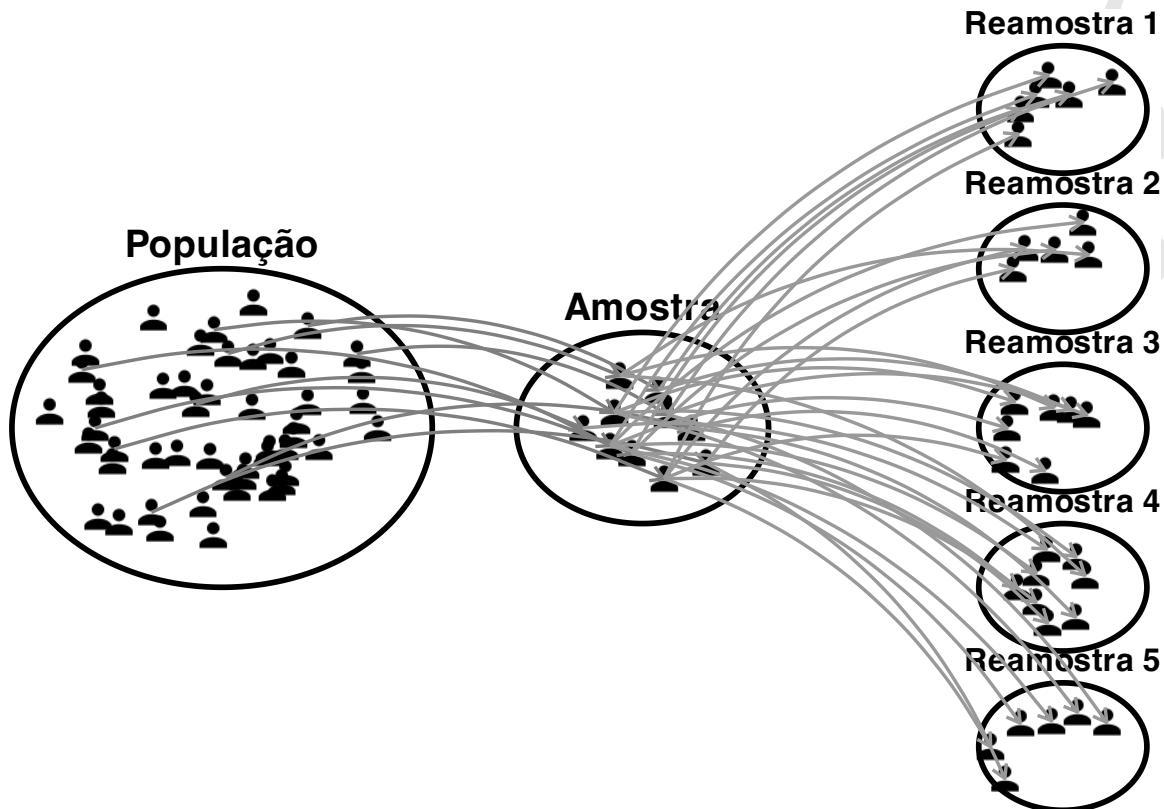


Figura 2.4: Representação esquemática da reamostragem de uma amostra.

2.6 Subamostragem

2.6.1 O que é subamostragem?

- ?

2.7 Superamostragem

2.7.1 O que é superamostragem?

- ?

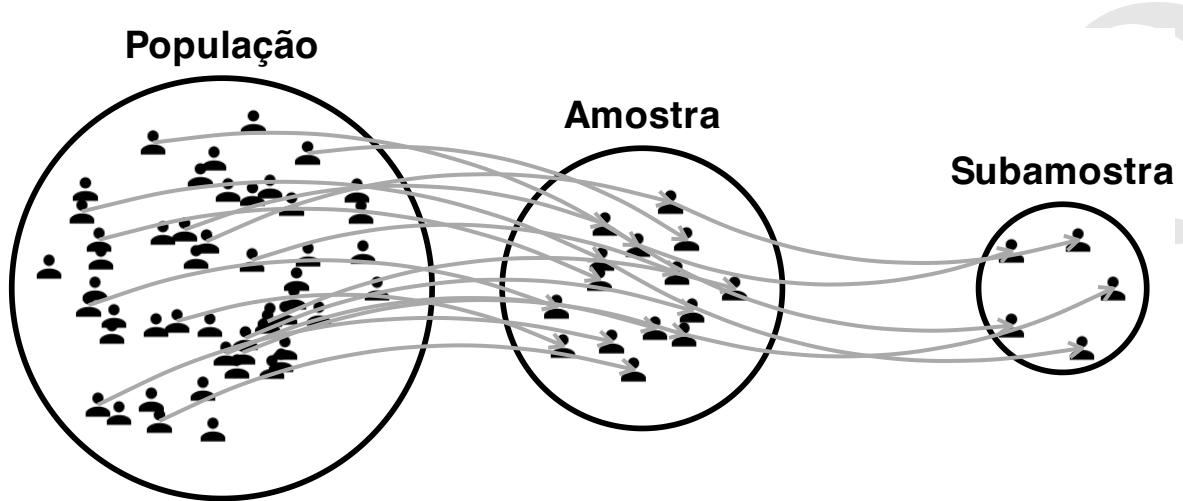


Figura 2.5: Representação esquemática da subamostragem de uma amostra.

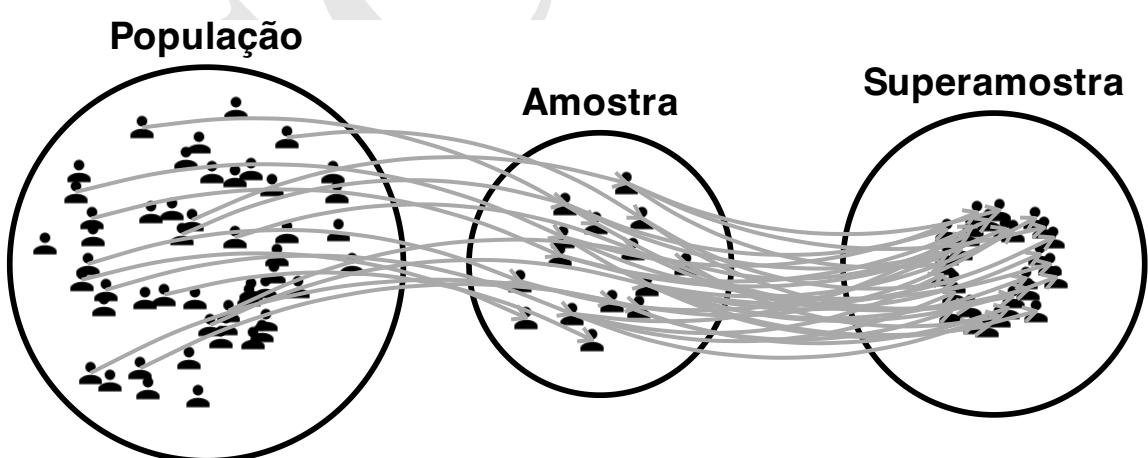


Figura 2.6: Representação esquemática da superamostragem de uma população.

RASCUNHO

Capítulo 3

Pensamento metodológico

3.1 Metodologia da pesquisa

3.1.1 O que é metodologia da pesquisa?

- A utilização de um vocabulário próprio — incluindo termos frequentemente usados em metodologia, epidemiologia e estatística — facilita a discussão na comunidade científica e melhora a compreensão das publicações.^{18,19}

3.2 Relação Estatística-Metodologia

3.2.1 Qual a relação entre estatística e metodologia da pesquisa?

- ²⁰

3.3 Pesquisa quantitativa vs. qualitativa

3.3.1 O que significa a distinção entre pesquisa qualitativa e quantitativa?

- A divisão entre quantitativo e qualitativo é amplamente usada, mas é considerada por muitos autores como superficial ou imprecisa. Em geral, associa-se o qualitativo à exploração detalhada de casos e significados, e o quantitativo ao uso de estatística e amostras maiores.²¹
- Tais associações ocultam múltiplas dimensões — por exemplo, análise estatística vs. não estatística e teste de hipóteses vs. indução — que não coincidem perfeitamente.²¹

3.3.2 Por que essa dicotomia pode ser problemática?

- Ao assumir apenas duas categorias, deixamos de lado possibilidades metodológicas úteis, como indução estatística (uso de estatística para construir teorias a partir dos dados) e teste de hipóteses não estatístico (avaliação de hipóteses em estudos de caso ou comparações conceituais).²¹
- A consequência é restringir artificialmente a variedade de métodos possíveis e criar mal-entendidos sobre o que cada termo implica.²¹

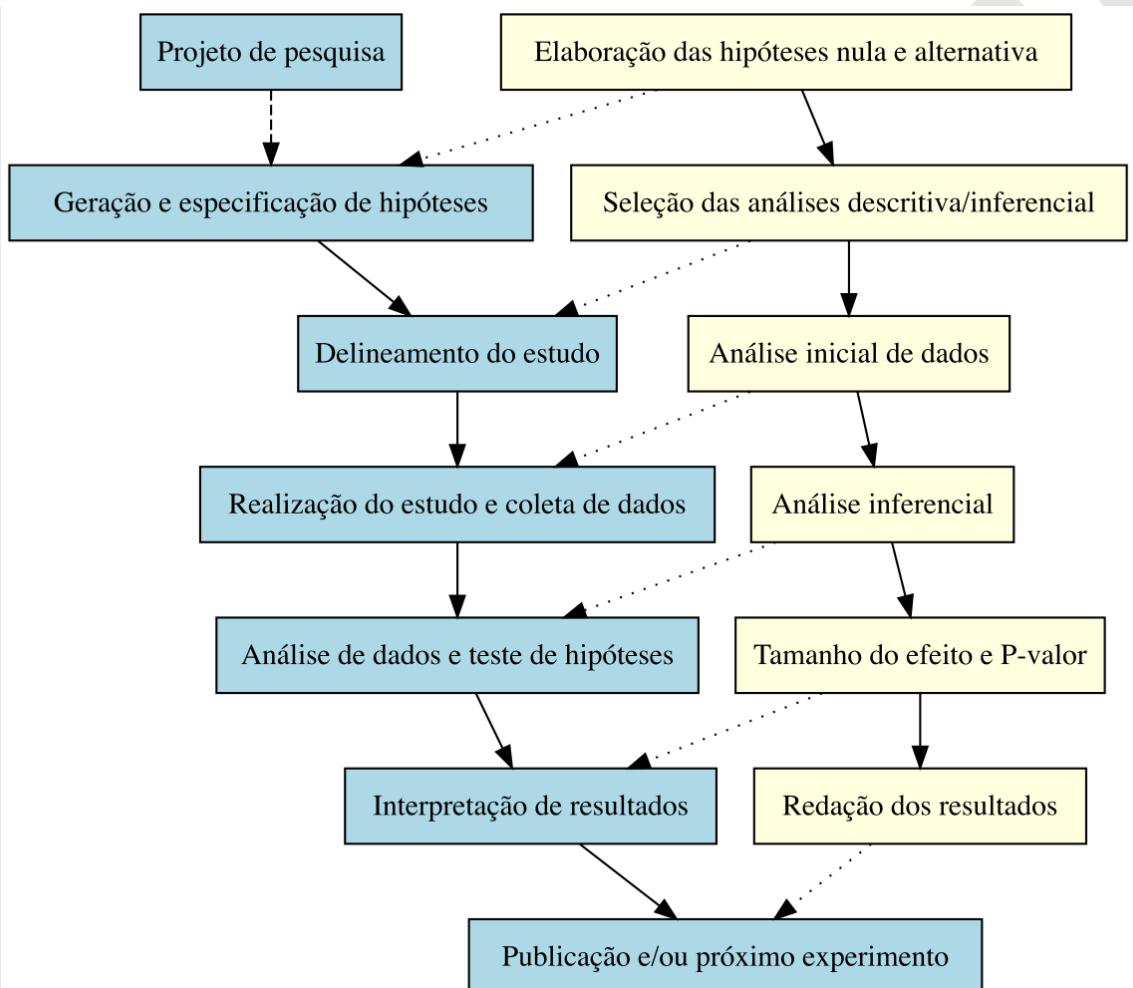


Figura 3.1: Mapa mental da relação entre o pensamento estatístico e o pensamento metodológico.

3.3.3 Qual é uma alternativa para pensar o debate?

- Usar termos mais específicos como “dados ricos” (*rich data*), “abordagem estatística”, “ilustração de possibilidades”, “teste de hipóteses”, “seguimento de paradigma”.²¹
- Descrever com clareza como os dados foram coletados, analisados e interpretados, sem recorrer a rótulos amplos que podem confundir ou carregar preconceitos metodológicos.²¹

3.4 Pesquisa de métodos mistos

3.4.1 O que é pesquisa de métodos mistos?

- Método misto é uma metodologia que integra de forma sistemática abordagens quantitativas e qualitativas em um único estudo, com o objetivo de responder a perguntas de pesquisa de maneira mais completa.²²
- Essa integração não é apenas a justaposição de duas técnicas; trata-se de um processo intencional de “mistura” de dados e interpretações em etapas como coleta, análise e interpretação, criando uma compreensão mais robusta.²²

3.4.2 Quais são as principais dimensões do desenho de métodos mistos?

- O desenho de pesquisa em métodos mistos deve considerar dimensões como propósito do estudo, orientação teórica, tempo (simultâneo ou sequencial), pontos de integração entre componentes, complexidade e se o desenho é planejado ou emergente.²³
- Entre as razões clássicas para combinar métodos estão: triangulação, complementaridade, desenvolvimento (um método orienta o outro), iniciação (explorar contradições) e expansão (ampliar o alcance da pesquisa).²³

3.4.3 Quais são os delineamentos centrais em pesquisa de métodos mistos?

- Três delineamentos principais são descritos como centrais: convergente, sequencial explanatório e sequencial exploratório.²²
- Convergente: coleta e análise de dados quantitativos e qualitativos em paralelo, com integração na interpretação.²²
- Sequencial explanatório: inicia com dados quantitativos, seguidos por qualitativos para explicar ou expandir os achados.²²
- Sequencial exploratório: inicia com dados qualitativos, seguidos por quantitativos que testam ou generalizam os resultados iniciais.²²
- Tipologias adicionais incluem delineamentos incorporados (*embedded*), transformativos (inspirados em perspectivas críticas, feministas ou de justiça social) e multifásicos, que combinam várias fases ao longo do tempo.²³

3.5 Pesquisa exploratória vs. confirmatória

3.5.1 O que são pesquisas exploratórias e confirmatórias?

- Confirmatória: teste planejado *a priori* de hipóteses com plano analítico predefinido (variáveis, modelos, critérios de exclusão, correções para múltiplos testes). Favorece controle de erro tipo I e interpretações diretas.²⁴
- Exploratória: testes pós-hoc motivados pelos dados, voltados a descoberta de padrões, geração/refinamento de hipóteses e checagens de plausibilidade. Pode revelar relações não antecipadas e orientar estudos futuros.²⁴

3.5.2 Por que a dicotomia é limitada?

- Na prática, há um *continuum* entre exploração e confirmação; muitos estudos combinam elementos de ambos em momentos distintos (p.ex., análises principais confirmatórias + análises de sensibilidade/descoberta).²⁴
- Análises exploratórias não são inherentemente inferiores: quando bem justificadas e comparando explicações alternativas, podem aumentar a rigorosidade do teste e produzir inferências informativas.²⁴
- A discussão contemporânea sobre a distinção entre estudos confirmatórios e exploratórios também enfatiza a necessidade de justificar o tamanho de efeito mínimo de interesse e aplicar testes severos conforme o arcabouço de falsificacionismo metodológico.²⁵

3.5.3 Quais são as boas práticas de transparência?

- Rotular claramente quais análises são confirmatórias e quais são exploratórias.²⁴
- Pré-registrar hipóteses e plano confirmatório; documentar desvios e justificá-los.²⁴
- Relatar análises de sensibilidade (modelos alternativos, decisões analíticas razoáveis) para avaliar robustez.²⁴
- Disponibilizar dados e código sempre que eticamente possível, distinguindo scripts confirmatórios de scripts exploratórios.²⁴

3.6 Pesquisa translacional

3.6.1 O que é pesquisa translacional?

- ?

3.7 Pré-registro

3.7.1 O que é pré-registro?

- Pré-registro é o ato de registrar publicamente o plano de pesquisa antes da coleta de dados ou análise.[?]
- O pré-registro é um elemento central das práticas abertas que aumentam a severidade de testes estatísticos e reduzem a flexibilidade analítica oportunista.²⁵

3.8 Reprodutibilidade

3.8.1 O que é reprodutibilidade?

- Reprodutibilidade é a habilidade de se obter resultados iguais ou similares quando uma análise ou teste estatístico é repetido.²⁶⁻²⁸

3.8.2 Por que reprodutibilidade é importante?

- Analisar a reprodutibilidade pode fornecer evidências a respeito da objetividade e confiabilidade dos achados, em detrimento de terem sido obtidos devido a vieses ou ao acaso.²⁶
- A reprodutibilidade não é apenas uma questão metodológica, mas também ética, uma vez que pode envolver mal práticas científicas como fabricação e/ou falsificação de dados.²⁶
- Reprodutibilidade pode ser considerada um padrão mínimo em pesquisa científica.²⁷

3.8.3 Como contribuir para a reproduzibilidade?

- Disponibilize publicamente os bancos de dados, respeitando as considerações éticas vigentes (ex.: autorização dos participantes e do Comitê de Ética em Pesquisa) e internacionalmente.²⁸
- Produza manuscritos reproduzíveis — manuscritos executáveis ou relatórios dinâmicos — que permitem a integração do banco de dados da(s) amostra(s), do(s) script(s) de análise estatística (incluindo comentários para sua interpretação), dos pacotes ou bibliotecas utilizados, das fontes e referências bibliográficas citadas, além dos demais elementos textuais (tabelas, gráficos) - todos gerados dinamicamente.²⁹
- Ao adotar delineamentos confirmatórios e justificar previamente hipóteses e tamanhos de efeito, o pesquisador contribui não apenas para a reproduzibilidade, mas também para a “severidade” inferencial.²⁵

3.9 Robustez

3.9.1 O que é robustez?

- ?

3.10 Replicabilidade

3.10.1 O que é replicabilidade?

- Replicabilidade é a habilidade de se obter conclusões iguais ou similares quando um experimento é repetido.^{27,28}

3.11 Generalização

3.11.1 O que é generalização?

- Generalização refere-se à extração das conclusões do estudo, observados na amostra, para a população.¹⁵

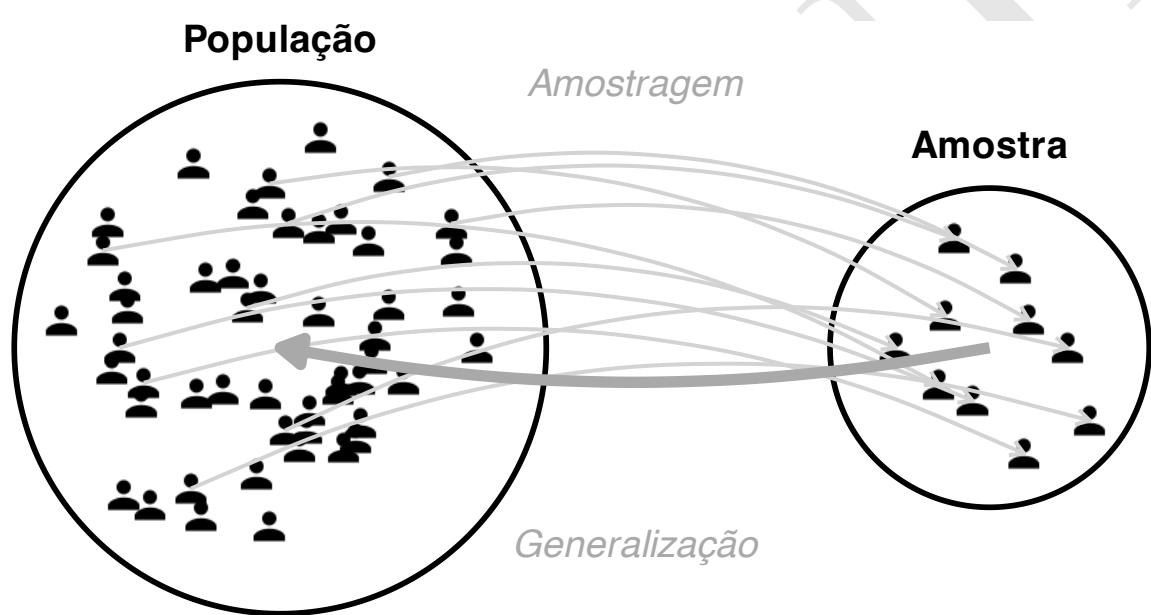


Figura 3.2: Representação esquemática da generalização de uma amostra para a população.

Capítulo 4

Pensamento computacional

4.1 R

4.1.1 O que é R?

- R é um programa de computador de código aberto com linguagem computacional direcionada para análise estatística.^{30,31}
- R version 4.5.2 (2025-10-31) está disponível gratuitamente em *Comprehensive R Archive Network* (CRAN).³²

4.1.2 Por que usar R?

- R é o software de maior abrangência de métodos estatísticos, possui sintaxe que permite análises estatísticas reproduzíveis e está disponível gratuitamente no *Comprehensive R Archive Network* (CRAN).^{28,32}

4.1.3 O que é R Markdown?

- R Markdown³³ é uma ferramenta que permite a integração de texto, código e saída em um único documento.[?]
- O R Markdown é uma extensão do Markdown, que é uma linguagem de marcação simples e fácil de aprender, que é usada para formatar texto.[?]
- O R Markdown permite a inclusão de blocos de código R, Python, SQL, C++, entre outros, e a saída desses blocos de código é incorporada ao documento final.[?]
- O R Markdown é uma ferramenta poderosa para a criação de relatórios dinâmicos, que podem ser facilmente atualizados com novos dados ou análises.[?]
- O R Markdown é amplamente utilizado na comunidade científica para a criação de relatórios de pesquisa, artigos científicos, apresentações, livros, entre outros.[?]
- O trabalho com RMarkdown³³ permite um fluxo de dados totalmente transparente, desde o conjunto de dados coletados até o manuscrito finalizado. Todos os aspectos do fluxo de dados podem ser incorporados em blocos de R script (*chunk*), exibindo tanto o R script quanto o respectivo texto, tabelas e figuras formatadas no estilo científico de interesse.³⁴
- O RMarkdown³³ foi projetado especificamente para relatórios dinâmicos onde a análise é realizada em R e oferece uma flexibilidade incrível por meio de uma linguagem de marcação.²⁸

4.1.4 Que programas de computador podem ser usados para análise estatística com R?

- JASP¹.³⁵
- jamovi².³⁶



Os pacotes *jmv*³⁷ e *jmvconnect*³⁸ fornecem funções para análise descritiva e inferencial com interface com jamovi.

4.2 RStudio

4.2.1 O que é RStudio?

- RStudio é um ambiente de desenvolvimento integrado (*integrated development environment*, IDE) desenvolvido visando a reproduzibilidade e a simplicidade para a criação e disseminação de conhecimento.^{31,39}
- O ambiente do RStudio é dividido em painéis:
 - *Source/Script editor*: para edição de R scripts.³¹
 - *Console*: para execução de códigos simples.³¹
 - *Environments*: para visualização de objetos criados durante a sessão de trabalho.³¹
 - *Output*: para visualização de gráficos criados durante a sessão de trabalho.³¹
- As principais características do RStudio incluem um ambiente de edição com abas para acesso rápido a arquivos, comandos e resultados; histórico de comandos previamente utilizados; ferramentas para visualização de bancos de dados e elaboração de scripts e gráficos e tabelas.^{31,39}
- RStudio está disponível gratuitamente em Posit³.



O pacote *learnr*⁴⁰ fornece tutoriais interativos para RStudio.

4.3 Scripts

4.3.1 O que são R scripts?

- “Scripts são dados”.²⁹
- Scripts permitem ao usuário se concentrar nas tarefas mais importantes da computação e utilizar pacotes ou bibliotecas para executar as funções mais básicas com maior eficiência.²⁹
- Um script é um arquivo de texto contendo (quase) os mesmos comandos que você digitaria na linha de comando do R. O “quase” refere-se ao fato de que se você estiver usando *sink()* para enviar a saída para um arquivo, você terá que incluir alguns comandos em *print()* para obter a mesma saída da linha de comando.[?]

```
# Exemplo de R script
```

```
# Este é um comentário
```

```
# Esta é uma variável
```

¹<https://jasp-stats.org>

²<https://www.jamovi.org>

³<https://posit.co/download/rstudio-desktop/>

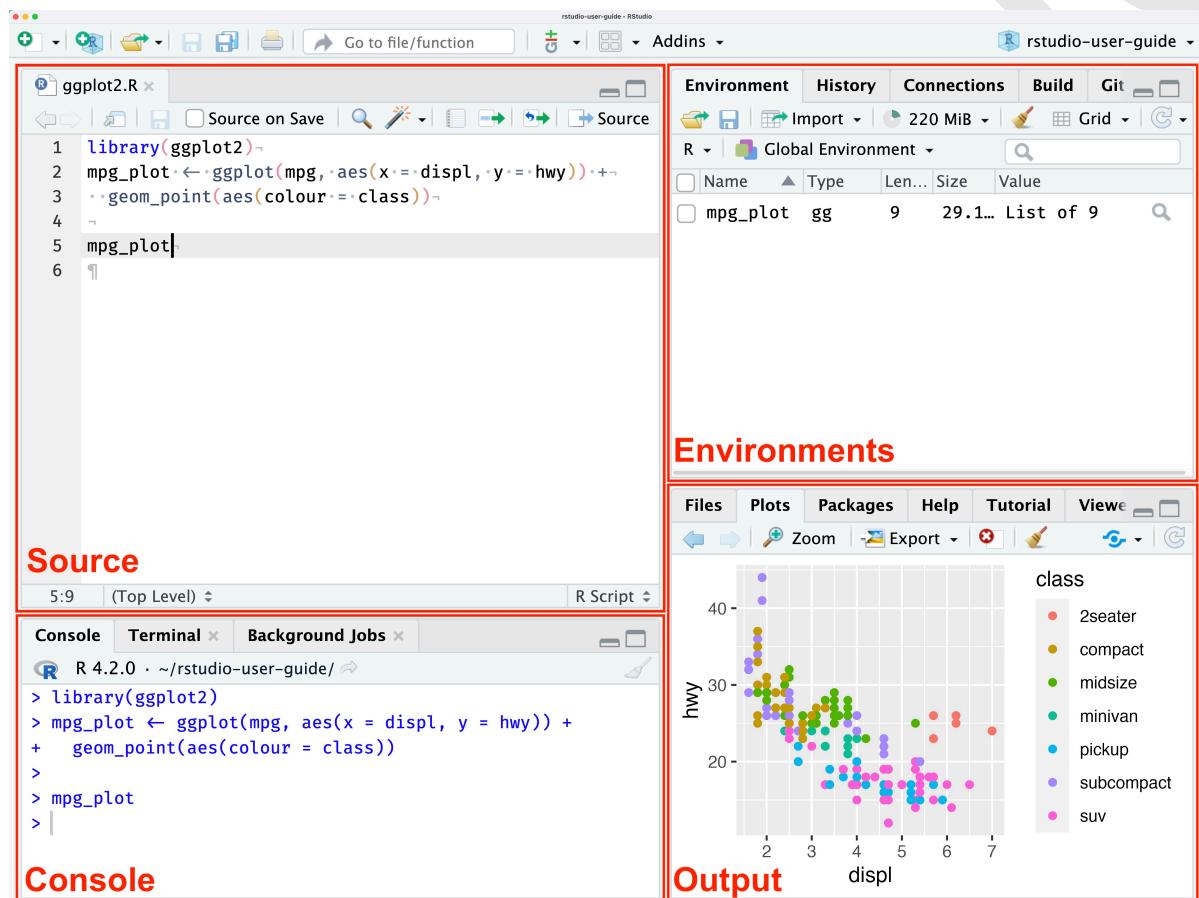


Figura 4.1: Interface do RStudio. Fonte: <https://docs.posit.co/ide/user/>

```

variavel <- 3.14 # Atribui o valor 3.14 à variável

# Esta é uma função
f <- function(x) {
  return(x^2) # Retorna o quadrado do valor de x
}

# Esta é uma chamada de função
resultado <- f(variavel) # Chama a função f com a variável como argumento

# Exibe o resultado da função
print(resultado) # Exibe o resultado na saída padrão

# Este é um vetor
vetor <- c(1, 2, 3, 4, 5) # Cria um vetor com os valores de 1 a 5
# Exibe o vetor
print(vetor) # Exibe o vetor na saída padrão

# Esta é uma matriz
matriz <- matrix(1:9, nrow=3, ncol=3) # Cria uma matriz 3x3 com os valores de 1 a 9
# Exibe a matriz
print(matriz) # Exibe a matriz na saída padrão

# Esta é uma lista
lista <- list(nome="João", idade=30, altura=1.75) # Cria uma lista com nome, idade e altura
# Exibe a lista
print(lista) # Exibe a lista na saída padrão

# Este é um dataframe
dataframe <- data.frame(nome=c("João", "Maria", "José"), idade=c(30, 25, 40), altura=c(1.75, 1.60, 1.8))
# Exibe o dataframe
print(dataframe) # Exibe o dataframe na saída padrão

# Este é um loop for
for (i in 1:5) {
  print(i) # Exibe os valores de 1 a 5 na saída padrão
}

# Este é um loop while
j <- 1
while (j <= 5) {
  print(j) # Exibe os valores de 1 a 5 na saída padrão
  j <- j + 1 # Incrementa o valor de j em 1
}

# Este é um condicional if-else
k <- 3
if (k > 0) {
  print("k é positivo") # Exibe "k é positivo" se k for maior que 0
} else if (k < 0) {
  print("k é negativo") # Exibe "k é negativo" se k for menor que 0
} else {

```

```

print("k é zero") # Exibe "k é zero" se k for igual a 0
}

# Fim do exemplo de R script

```

4.3.2 Quais são as boas práticas na redação de scripts?

- Use nomes consistentes para as variáveis.⁴¹
- Defina os tipos de variáveis adequadamente no banco de dados.⁴¹
- Defina constantes — isto é, variáveis de valor fixo — ao invés de digitar valores.⁴¹
- Use e cite os pacotes disponíveis para suas análises.⁴¹
- Controle as versões do script.^{41,42}
- Teste o script antes de sua utilização.⁴¹
- Conduza revisão por pares do código durante a redação (digitação em dupla).⁴¹



O pacote *formatR*⁴³ fornece a função *tidy_source*^a para formatar um R script.

^ahttps://www.rdocumentation.org/packages/formatR/versions/1.14/topics/tidy_source



O pacote *styler*⁴⁴ fornece a função *style_file*^a para formatar um R script.

^ahttps://www.rdocumentation.org/packages/styler/versions/1.10.1/topics/style_file



O pacote *lintr*⁴⁵ fornece a função *lint*^a para verificar a adesão de um script a um determinado estilo, identificando erros de sintaxe e possíveis problemas semânticos.

^a<https://www.rdocumentation.org/packages/lintr/versions/3.1.0/topics/lint>

4.4 Pacotes

4.4.1 O que são pacotes?

- Pacotes são conjuntos de scripts programados pela comunidade e compartilhados para uso público.³¹
- Os pacotes ficam armazenados no *Comprehensive R Archive Network* (CRAN) e podem ser instalados diretamente no RStudio.^{31,32}
- Na mais recente atualização deste livro, o [Comprehensive R Archive Network (CRAN) possui 390966 pacotes disponíveis.^{31,32}
- Os pacotes disponíveis podem ser encontrados em *R PACKAGES DOCUMENTATION*.⁴⁶



O pacote *utils*⁴⁷ fornece a função *install.packages*^a para instalar os pacotes no computador.

^a<https://www.rdocumentation.org/packages/utils/versions/3.6.2/topics/install.packages>

R

O pacote *utils*⁴⁷ fornece a função *library*^a para carregar os pacotes instalados no computador.

^a<https://www.rdocumentation.org/packages/utils/versions/3.6.2/topics/library>

R

O pacote *utils*⁴⁷ fornece a função *require*^a para indicar se o pacote requisitado está disponível.

^a<https://www.rdocumentation.org/packages/utils/versions/3.6.2/topics/require>

R

O pacote *utils*⁴⁷ fornece a função *installed.packages*^a para listar os pacotes instalados no computador.

^a<https://www.rdocumentation.org/packages/utils/versions/3.6.2/topics/installed.packages>

R

O pacote *utils*⁴⁷ fornece a função *update.packages*^a para atualizar os pacotes instalados no computador.

^a<https://www.rdocumentation.org/packages/utils/versions/3.6.2/topics/update.packages>

R

O pacote *roxygen2*⁴⁸ fornece a função *roxygenize*^a para criar arquivos .Rd para documentar pacotes.

^a<https://cran.r-project.org/web/packages/roxygen2/index.html>

4.5 Aplicativos

4.5.1 O que são Shiny Apps?

- Shiny Apps são aplicativos web interativos que permitem a criação de interfaces gráficas para visualização e análise de dados em tempo real, utilizando o R como backend.²⁷

4.6 Manuscritos reproduzíveis

4.6.1 O que são manuscritos reproduzíveis?

- Manuscritos reproduzíveis — manuscritos executáveis ou relatórios dinâmicos — permitem a produção de um manuscrito completo a partir da integração do banco de dados da(s) amostra(s), do(s) script(s) de análise estatística (incluindo comentários para sua interpretação), dos pacotes ou bibliotecas utilizados, das fontes e referências bibliográficas citadas, além dos demais elementos textuais (tabelas, gráficos) - todos gerados dinamicamente.²⁹

4.6.2 Por que usar manuscritos reproduzíveis?

- No processo tradicional de redação científica há muitas etapas de copiar e colar não reproduzíveis envolvidas. Documentos dinâmicos combinam uma ferramenta de processamento de texto com o R script que produz o texto/tabela/figura a ser incorporado no manuscrito.²⁸

- Ao trabalhar com relatórios dinâmicos, é possível extrair o mesmo script usado para análise estatística. Os documentos podem ser compilados em vários formatos de saída e salvos como DOCX, PPTX e PDF.²⁸
- Muitos erros de análise poderiam ser evitados com a adoção de boas práticas de programação em manuscritos reproduzíveis.⁴⁹

R

O pacote *rmarkdown*³³ fornece as funções *render*^a para criar manuscritos reproduzíveis a partir de arquivos .Rmd.

^a<https://www.rdocumentation.org/packages/rmarkdown/versions/2.24/topics/render>

R

O pacote *officedown*⁵⁰ fornece as funções *rdocx_document*^a e *rpptx_document*^b para criar arquivos DOCX e PPTX, respectivamente, com o conteúdo criado no manuscrito reproduzível.

^ahttps://www.rdocumentation.org/packages/officedown/versions/0.3.0/topics/rdocx_document

^bhttps://www.rdocumentation.org/packages/officedown/versions/0.3.0/topics/rpptx_document

R

O pacote *bookdown*⁵¹ fornece as funções *gitbook*^a, *pdf_book*^b, *epub_book*^c e *html_document*^d para criar documentos reproduzíveis em diversos formatos (Git, PDF, EPUB e HTML, respectivamente).

^a<https://www.rdocumentation.org/packages/bookdown/versions/0.35/topics/gitbook>

^bhttps://www.rdocumentation.org/packages/bookdown/versions/0.35/topics/pdf_book

^chttps://www.rdocumentation.org/packages/bookdown/versions/0.35/topics/epub_book

^dhttps://www.rdocumentation.org/packages/bookdown/versions/0.35/topics/html_document2

4.6.3 Como manuscritos reproduzíveis contribuem para a ciência?

- O compartilhamento de bancos de dados e seus scripts de análise estatística permitem a adoção de práticas reproduzíveis, tais como a reanálise dos dados.⁵²

R

O pacote *projects*⁵³ fornece a função *setup_projects*^a para criar um projeto com arquivos organizados em diretórios.

^ahttps://www.rdocumentation.org/packages/projects/versions/2.1.3/topics/setup_projects

R

O pacote *rmarkdown*³³ fornece a função *render*^a para criar manuscritos reproduzíveis a partir de arquivos .Rmd.

^a<https://www.rdocumentation.org/packages/rmarkdown/versions/2.24/topics/render>

R

O pacote *bookdown*⁵¹ fornece as funções *gitbook*^a, *pdf_book*^b, *epub_book*^c e *html_document2*^d para criar documentos reprodutíveis em diversos formatos (Git, PDF, EPUB e HTML, respectivamente).

^a<https://www.rdocumentation.org/packages/bookdown/versions/0.35/topics/gitbook>

^bhttps://www.rdocumentation.org/packages/bookdown/versions/0.35/topics/pdf_book

^chttps://www.rdocumentation.org/packages/bookdown/versions/0.35/topics/epub_book

^dhttps://www.rdocumentation.org/packages/bookdown/versions/0.35/topics/html_document2

4.7 Compartilhamento

4.7.1 Por que compartilhar scripts?

- Compartilhar o script — principalmente junto aos dados — pode facilitar a replicação direta do estudo, a detecção de eventuais erros de análise, a detecção de pesquisas fraudulentas.⁵⁴

4.7.2 O que pode ser compartilhado?

- Idealmente, todos os scripts, pacotes/bibliotecas e dados necessários para outros reproduzirem seus dados.⁴²
- Minimamente, partes importantes incluindo implementações de novos algoritmos e dados que permitam reproduzir um resultado importante.⁴²

4.7.3 Como preparar dados para compartilhamento?

- ?

4.7.4 Como preparar scripts para compartilhamento?

- Providencie a documentação sobre seu script (ex.: arquivo README).⁴²
- Inclua a versão dos pacotes usados no seu script por meio de um script inicial para instalação de pacotes (ex.: ‘instalar.R’).⁴⁹
- Documente em um arquivo README os arquivos disponíveis e os pré-requisitos necessários para executar o código (ex.: pacotes e respectivas versões). Uma lista de configurações (hardware e software) que foram usadas para rodar o código pode ajudar na reprodução dos resultados.²⁷
- Use endereços de arquivos relativos.⁴⁹
- Crie links persistentes para versões do seu script.⁴²
- Defina uma semente para o gerador de números aleatórios em scripts com métodos computacionais que dependem da geração de números pseudoaleatórios.²⁷

R

O pacote *base*⁵⁵ fornece a função *set.seed*^a para especificar uma semente para reprodutibilidade de computações que envolvem números aleatórios.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/Random>

- Escolha uma licença apropriada para garantir os direitos de criação e como outros poderão usar seus scripts.⁴²
- Teste o script em uma nova sessão antes de compartilhar.⁴⁹

- Cite todos os pacotes relacionados à sua análise.⁵⁶



O pacote *utils*⁴⁷ fornece a função *citation*^a para citar o programa R e os pacotes da sessão atual.

^a<https://www.rdocumentation.org/packages/utils/versions/3.6.2/topics/citation>



O pacote *grateful*⁵⁷ fornece a função *cite_packages*^a para citar os pacotes utilizados em um projeto R.

^ahttps://www.rdocumentation.org/packages/grateful/versions/0.2.0/topics/cite_packages

- Inclua a informação da sessão em que os scripts foram rodados.⁴⁹



O pacote *utils*⁴⁷ fornece a função *sessionInfo*^a para descrever as características do programa, pacotes e plataforma da sessão atual.

^a<https://www.rdocumentation.org/packages/utils/versions/3.6.2/topics/sessionInfo>

4.7.5 O que incluir no arquivo README?

- Título do trabalho.²⁷
- Autores do trabalho.²⁷
- Principais responsáveis pela escrita do script e quaisquer outras pessoas que fizeram contribuições substanciais para o desenvolvimento do script.²⁷
- Endereço de e-mail do autor ou contribuidor a quem devem ser direcionadas dúvidas, comentários, sugestões e bugs sobre o script.²⁷
- Lista de configurações nas quais o script foi testado, tais com nome e versão do programa, pacotes e plataforma.²⁷

RASCUNHO

Capítulo 5

Letramento estatístico

5.1 Introdução ao letramento estatístico

5.1.1 O que é letramento estatístico?

- Letramento em informação: Capacidade de reconhecer quando a informação é necessária e de localizá-la, avaliá-la criticamente (qualidade, validade, relevância, completude, imparcialidade) e usá-la de forma eficaz e ética. Abrange qualquer tipo de informação, em texto ou números.⁵⁸
- Letramento em dados: Competência técnica para acessar, manipular, resumir e apresentar dados, utilizando ferramentas e métodos (SQL, planilhas, softwares estatísticos), com foco na preparação e organização de conjuntos de dados para análise e comunicação.⁵⁸
- Letramento estatístico é a competência para compreender, interpretar e avaliar informações baseadas em dados, integrando conhecimentos técnicos (linguagem, estatística, matemática) e contextuais com postura crítica, crenças e atitudes que sustentem o uso ético e fundamentado da estatística.⁵⁹⁻⁶¹
- Letramento estatístico é parte essencial do letramento informacional (fornecer a capacidade de reconhecer, acessar e avaliar informações) e do letramento em dados (envolve acessar, manipular e apresentar dados de forma adequada).⁵⁸

5.1.2 Por que o letramento estatístico é importante?

- A presença dos dados no cotidiano deixou de ser restrita a decisões políticas ou relatórios técnicos: hoje, todos estamos expostos e interagimos com dados de forma constante, seja por dispositivos móveis, redes sociais ou sistemas automatizados de recomendação.⁶²
- Ferramentas para coletar e analisar dados estão mais acessíveis e baratas, o que amplia a possibilidade de qualquer pessoa atuar não só como consumidora, mas também como produtora de informações.⁶²

5.1.3 Quais são exemplos de armadilhas comuns na interpretação de estatísticas?

- Escolha do indicador: usar média ou mediana pode levar a conclusões muito diferentes sobre o mesmo fenômeno (por exemplo, renda média vs. mediana antes e depois de impostos).⁵⁸
- Confusão entre taxas e contagens: comparar números absolutos sem considerar proporções populacionais pode distorcer a realidade.⁵⁸
- Fatores de confusão: diferenças observadas podem ser explicadas por variáveis não consideradas, como idade média da população ao comparar taxas de mortalidade.⁵⁸

5.2 Elementos centrais do letramento estatístico

5.2.1 Quais são os elementos de conhecimento que sustentam o letramento estatístico?

- O modelo de letramento estatístico é composto por cinco elementos de conhecimento e dois elementos disposicionais.^{59–61}

5.2.2 Quais são os cinco elementos de conhecimento que sustentam o letramento estatístico?

- Competências de letramento, incluindo leitura de textos, gráficos e tabelas.⁵⁹
- Conhecimento estatístico básico, incluindo conceitos, métodos, interpretação de dados e probabilidade.⁵⁹
- Conhecimento matemático sobre percentagens, médias e raciocínio numérico.⁵⁹
- Conhecimento de contexto/mundo, com entendimento do cenário e origem dos dados.⁵⁹
- Questões críticas (lista de *worry questions* para avaliar a validade da informação).⁵⁹

5.2.3 Quais são os dois elementos de disposição que facilitam a ação estatisticamente letrada?

- Postura crítica: propensão para questionar e analisar mensagens quantitativas.⁵⁹
- Crenças e atitudes: visão positiva sobre a capacidade de pensar estatisticamente; valorização de dados bem produzidos.⁵⁹

5.2.4 Que tipo de perguntas críticas devemos fazer ao interpretar informação estatística?

- De onde vêm os dados? Que tipo de estudo foi feito?⁵⁹
- A amostra é representativa e suficientemente grande?⁵⁹
- Os instrumentos de medição são confiáveis?⁵⁹
- As estatísticas e gráficos são apropriados e não distorcem?⁵⁹
- Há relação causal ou apenas correlação? Há informação em falta?⁵⁹
- Existem interpretações alternativas plausíveis?⁵⁹

5.3 Hierarquia de letramento estatístico

5.3.1 Quais são os níveis da hierarquia de letramento estatístico?

- Nível 6 – Crítico Matemático: É o nível mais alto. A pessoa questiona e analisa as informações de forma profunda, usando cálculos e raciocínio proporcional (como comparar porcentagens e proporções). Reconhece que previsões sempre envolvem algum grau de incerteza e percebe detalhes sutis na forma como os dados são apresentados.⁶³
- Nível 5 – Crítico: Também envolve uma postura questionadora, mas sem exigir cálculos complexos de proporção. Usa corretamente a linguagem estatística, entende o significado de termos ligados à probabilidade e percebe que os resultados podem variar.⁶³
- Nível 4 – Consistente, mas Não Crítico: Consegue interpretar dados e usar termos estatísticos corretamente, mas não chega a questionar a forma como as informações são apresentadas. Reconhece a variação apenas

em situações que envolvem sorte ou acaso, e sabe lidar com conceitos como média, probabilidades simples e leitura de gráficos.⁶³

- Nível 3 – Inconsistente: Analisa partes do problema, mas de forma irregular. Pode identificar conclusões corretas, mas sem explicá-las. Usa ideias estatísticas de maneira mais descritiva do que numérica.⁶³
- Nível 2 – Informal: A interpretação é mais baseada no senso comum do que em conceitos estatísticos. Utiliza apenas alguns termos corretos e consegue fazer cálculos muito simples com tabelas, gráficos ou situações de probabilidade.⁶³
- Nível 1 – Idiossincrático: Responde de forma muito pessoal ou confusa, usando termos de maneira incorreta ou limitada. Realiza apenas contagens diretas e leituras simples de dados.⁶³

5.3.2 Quais são os componentes centrais do letramento estatístico com literacia de dados?

- Compreender quem coleta dados, por que e como essa coleta é feita.⁶²
- Saber interpretar dados de amostras aleatórias e não aleatórias, avaliando limitações e potencial.⁶²
- Conhecer e aplicar práticas de proteção de dados e direitos de propriedade sobre informações coletadas.⁶²
- Produzir representações descritivas (tabelas, gráficos, mapas, *dashboards*) para responder perguntas sobre fenômenos reais.⁶²
- Reconhecer a importância da proveniência e do armazenamento dos dados, bem como a necessidade de pré-processamento antes da análise.⁶²
- Entender fundamentos de modelagem preditiva e algoritmos, como árvores de classificação e regressão, especialmente no contexto de dados massivos (*big data*).⁶²

5.4 Habilidades de letramento estatístico baseadas no pensamento crítico

5.4.1 Quais são as habilidades de letramento estatístico?

- Identificar: Descobrir qual é a principal afirmação de um texto ou relatório e separar o que é opinião do que é realmente evidência ou dado.⁶⁴
- Questionar: Fazer perguntas sobre os dados: de onde vieram, como foram coletados, qual o tamanho da amostra, se houve erros, se os gráficos estão claros e se o questionário foi bem feito.⁶⁴
- Julgar: Avaliar se a afirmação é bem sustentada pelos dados ou se está exagerando, por exemplo, dizendo que algo causa quando só foi encontrada uma relação.⁶⁴
- Esclarecer: Entender e explicar palavras técnicas e expressões que podem confundir, além de saber como foi feita a pesquisa e a análise.⁶⁴
- Avaliar: Decidir se a afirmação é confiável comparando com outras informações disponíveis e verificando se faz sentido.⁶⁴
- Investigar mais: Procurar informações que não foram mostradas, como quem fez a pesquisa, por que foi feita, detalhes do processo ou fatores escondidos que podem influenciar os resultados.⁶⁴
- Considerar alternativas: Pensar em outras explicações possíveis ou diferentes interpretações para os mesmos dados.⁶⁴
- Concluir: Chegar à sua própria conclusão sobre o assunto, usando as informações e o raciocínio de forma clara e bem fundamentada.⁶⁴

RASCUNHO

PARTE 2: AMEAÇAS À QUALIDADE DA EVIDÊNCIA CIENTÍFICA

Vieses Metodológicos, Erros de Interpretação e Condutas Questionáveis

RASCUNHO

Capítulo 6

Vieses metodológicos

6.1 Vieses metodológicos

6.1.1 O que são vieses metodológicos?

- ?

6.2 Tipos de vieses metodológicos

6.2.1 Quais são os tipos de vieses metodológicos?

- ?

6.3 Efeitos relacionados aos vieses metodológicos

6.3.1 Quais são os efeitos relacionados aos vieses metodológicos?

- ?

6.3.2 O que é efeito placebo?

- ?

6.3.3 O que é efeito nocebo?

- ?

6.3.4 O que é efeito Hawthorne?

- ?

6.3.5 O que é efeito Rosenthal?

- ?

6.4 Diretrizes para redação

6.4.1 Quais são as diretrizes para redação de análises de vieses metodológicos?

- Visite a rede *Enhancing the QUAlity and Transparency Of health Research* (EQUATOR Network¹) para encontrar diretrizes específicas.
- *PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies*.⁶⁵
- *RoB 2: A Revised Tool for Assessing Risk of Bias in Randomized Trials*.⁶⁶
- *AMSTAR 2: A Critical Appraisal Tool for Systematic Reviews that Include Randomised or Non-Randomised Studies of Healthcare Interventions*⁶⁷
- *ROBINS-I: A Tool for Assessing Risk of Bias in Non-randomized Studies of Interventions*.⁶⁸
- *ROBIS: A New Tool to Assess Risk of Bias in Systematic Reviews*⁶⁹
- *QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies*⁷⁰

¹<https://www.equator-network.org/>

Capítulo 7

Falácia estatísticas

7.1 Falácia

7.1.1 O que são falácia estatísticas?

- Falácia estatísticas são erros de raciocínio que ocorrem em situações que envolvem dados e estatísticas. Elas podem ocorrer em qualquer etapa do processo de análise de dados, desde a coleta até a interpretação dos resultados.⁷⁰
- Falácia podem ser intencionais ou não intencionais, e podem ser usadas para manipular, enganar ou confundir as pessoas.⁷¹
- As falácia estatísticas podem ser difíceis de detectar, pois muitas vezes são sutis e podem parecer plausíveis à primeira vista. No entanto, é importante estar ciente delas e saber como identificá-las para evitar erros de interpretação e tomada de decisão.⁷²

7.1.2 O que é a falácia do jogador?

- A falácia do jogador é a crença de que eventos independentes têm uma influência sobre eventos futuros. Por exemplo, se uma moeda é lançada várias vezes e cai cara em todas as vezes, a falácia do jogador sugere que a próxima jogada será coroa, pois a moeda “deve” se equilibrar. No entanto, cada lançamento da moeda é independente e não afeta o resultado do próximo lançamento.⁷³

7.1.3 O que é a falácia da mão quente?

- A falácia da mão quente é a crença de que um jogador que teve sucesso em um jogo de azar terá mais chances de sucesso no futuro. Por exemplo, se uma moeda é lançada várias vezes e cai cara em todas as vezes, a falácia da mão quente sugere que a próxima jogada será cara, pois o jogador está “quente”. No entanto, cada lançamento da moeda é independente e não afeta o resultado do próximo lançamento.⁷⁴

RASCUNHO

Capítulo 8

Paradoxos estatísticos

8.1 Paradoxos

8.1.1 O que são paradoxos estatísticos?

- Paradoxos podem originar da incompreensão ou mal informação da nossa intuição a respeito do fenômeno.⁷²

8.1.2 O que é o paradoxo de Abelson?

- ⁷³

8.1.3 O que é o paradoxo de Berkson?

- ⁷⁴

8.1.4 O que é o paradoxo de *Big Data*?

- “Quanto maior a quantidade de dados, maior a certeza de que vamos nos enganar”.⁷²

8.1.5 O que é o paradoxo de Ellsberg?

- ⁷⁵

8.1.6 O que é o paradoxo de Freedman?

- ^{76,77}

8.1.7 O que é o paradoxo de Hand?

- ⁷⁸

8.1.8 O que é o paradoxo de Kelley?

- [?]

8.1.9 O que é o paradoxo de Lindley?

- ⁷⁹

8.1.10 O que é o paradoxo de Lord?

- 80,81

8.1.11 O que é o paradoxo de Probsting?

- ?

8.1.12 O que é o paradoxo de Simpson?

- O paradoxo de Simpson ocorre quando a associação entre duas variáveis X e Y desaparece ou mesmo reverte sua direção quando condicionadas em uma terceira variável Z .^{82,83}
- Para decisão do paradoxo de Simpson pode-se utilizar o conceito de ‘back-door’, o qual considera os ‘caminhos’ (isto é, associações) no gráfico acíclico direcionado e assegura que todos as associações espúrias do tratamento X para o desfecho Y nesse diagrama causal sejam interceptados pela variável Z .⁸⁴
- Dependendo do contexto em que os dados foram obtidos — delineamento do estudo, escolha dos instrumentos e dos tipos de variáveis — a melhor escolha para a análise pode variar entre a análise da população agregada ou da subpopulação desagregada.⁸⁴
- É possível que em alguns contextos nem a análise agregada ou a desagregada podem oferecer a resposta correta, sendo necessário o uso de outras (mais) covariáveis.⁸⁴

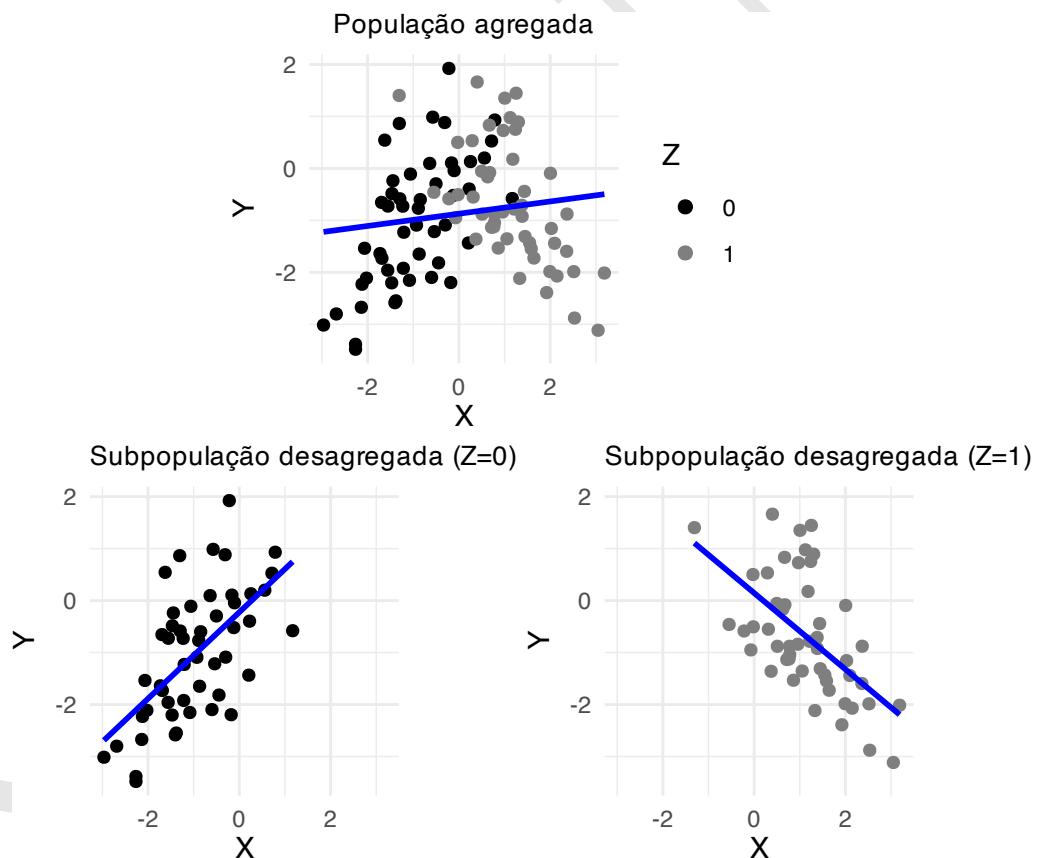


Figura 8.1: Paradoxo de Simpson representado com dados simulados. Os pontos no gráfico representam observações individuais e as linhas de tendência representam as regressões lineares ajustadas para os dados desagregados da população e agregados por subpopulação.

8.1.13 O que é o paradoxo de Stein?

• ⁸⁵

8.1.14 O que é o paradoxo de Okie?

• ?

8.1.15 O que é o paradoxo da acurácia?

• ?

8.1.16 O que é o paradoxo do falso positivo?

• ?

8.1.17 O que é o paradoxo da caixa de Bertrand?

• ?

8.1.18 O que é o paradoxo do elevador?

• ⁸⁶

8.1.19 O que é o paradoxo da amizade?

• ⁸⁷

8.1.20 O que é o paradoxo do menino ou menina?

• ⁸⁶

8.1.21 O que é o paradoxo do aniversário?

• ?

8.1.22 O que é o paradoxo do teste surpresa?

• ?

8.1.23 O que é o paradoxo do nó da gravata?

• ?

8.1.24 O que é o paradoxo de Monty Hall?

• ?

8.1.25 O que é o paradoxo da Bela Adormecida?

• ?

RASCUNHO

Capítulo 9

Práticas questionáveis em pesquisa

9.1 Práticas Questionáveis em Pesquisa

9.1.1 O que são práticas questionáveis em pesquisa?

- Práticas questionáveis em pesquisa são más condutas ou comportamentos impróprios, realizados desde o planejamento até a publicação dos resultados.^{88,89}

9.1.2 Por que práticas questionáveis em pesquisa devem ser combatidas?

- Práticas questionáveis em pesquisa são prevalentes.⁹⁰
- Práticas questionáveis em pesquisa comprometem a integridade científica, a confiabilidade dos resultados e a confiança do público na ciência.^{88,89}
- Práticas questionáveis em pesquisa inflam artificialmente o tamanho do efeito e poder estatístico.⁸⁹
- Práticas questionáveis em pesquisa parecem contribuir para a crise da replicação na ciência, onde muitos estudos não conseguem ser replicados ou reproduzidos.⁸⁹

9.2 Prática não intencional e má conduta

9.2.1 Quais são as categorias de práticas questionáveis em pesquisa?

- Práticas questionáveis podem ser classificadas em más condutas e não intencionais.⁹¹
- MÁS condutas são aquelas que são deliberadamente realizadas com o objetivo de enganar ou manipular os resultados, enquanto práticas não intencionais são aquelas que ocorrem devido a falta de conhecimento, treinamento inadequado ou outras razões.⁹²
- Práticas na zona cinzenta são aquelas que podem ser interpretadas de diferentes maneiras, dependendo do contexto e da intenção do pesquisador.⁹³

9.2.2 Quais práticas questionáveis podem ocorrer durante o planejamento do estudo?

- *Hypothesizing After Results are Known* (HARKing) consiste em formular hipóteses após a análise dos dados, o que pode levar a resultados enviesados e não replicáveis.⁹²
- *Storytelling* é a prática de criar narrativas convincentes para justificar os resultados, mesmo que não sejam suportados pelos dados.⁹³

Tabela 9.1: Classificação das práticas questionáveis em pesquisa segundo sua intencionalidade.

Prática	Intencionalidade	Definição
Data fabrication	Má conduta	Inventar dados inexistentes
Data falsification	Má conduta	Alterar ou manipular dados reais
Fake authorship	Má conduta	Inserir autores fictícios ou inexistentes
Fake peer review	Má conduta	Criar revisões falsas para facilitar publicação
Honorary authorship	Má conduta	Incluir autores sem contribuição real
Gold authorship	Má conduta	Atribuir autoria como forma de prestígio ou recompensa
Ghost authorship	Má conduta	Omitir autores que participaram do estudo
Duplicate publication	Má conduta	Publicar o mesmo estudo em mais de uma revista
Spin (doloso)	Má conduta	Apresentar os resultados de forma a exagerar efeitos positivos
Data distortion	Má conduta	Modificar dados ou gráficos para torná-los mais convincentes
SPARKing	Má conduta	Ajustar o tamanho da amostra após a coleta dos dados para obter significância estatística
HARKing	Zona cinzenta	Criar hipóteses após ver os dados (sem pré-registro)
Storytelling	Zona cinzenta	Construir uma narrativa forçada para justificar os achados
Selective reporting	Zona cinzenta	Relatar apenas os resultados favoráveis ou positivos
P-hacking	Zona cinzenta	Testar múltiplas análises até encontrar $p < 0.05$
Data peeking	Zona cinzenta	Analizar dados antes do término da coleta, parando quando um efeito aparece
Cherry picking	Zona cinzenta	Selecionar apenas os resultados que apoiam a hipótese
Salami slicing	Zona cinzenta	Dividir artificialmente um estudo em vários artigos para inflar publicações
Beautification	Zona cinzenta	Embelezar tabelas, gráficos ou resultados para torná-los mais atraentes
P-hacking reverso	Não intencional	Forçar análises para que não haja significância estatística
Fishing expedition	Não intencional	Procurar achados sem plano prévio
Data dredging	Não intencional	Explorar excessivamente os dados para encontrar associações irrelevantes
File drawer problem	Não intencional	Não publicar estudos com resultados negativos ou nulos
Publication bias	Não intencional	Tendência geral das revistas em favorecer publicações com resultados positivos

9.2.3 Quais práticas questionáveis podem ocorrer durante a coleta de dados?

- *Data falsification* é a prática de manipular ou inventar dados para obter resultados desejados.⁹⁷
- *Data fabrication* é a prática de inventar dados ou resultados que nunca foram coletados.⁹⁸

9.2.4 Quais práticas questionáveis podem ocorrer durante a análise dos dados?

- *P-hacking* é a prática de manipular os dados ou análises para obter resultados estatisticamente significativos, como realizar múltiplos testes sem correção adequada.⁹³⁻⁹⁵
- *P-hacking* reverso é a prática de manipular os dados ou análises para obter resultados não estatisticamente significativos, como realizar múltiplos testes sem correção adequada, o que pode levar a conclusões enviesadas e enganosas.⁹⁶
- *SPARKing (Sample size Planning After the Results are Known)* é uma mal prática que envolve o ajuste do tamanho da amostra após a coleta dos dados, com o objetivo de obter resultados estatisticamente significativos.⁹⁷
- *Data peeking* é a prática de analisar os dados repetidamente antes de completar a coleta, visando interromper a coleta quando um resultado desejado é alcançado.⁹⁸
- *Fishing expedition* refere-se à exploração dos dados sem uma hipótese pré-definida, o que pode levar a conclusões enganosas e enviesadas, uma vez que os resultados podem ser meramente acidentais.⁹⁴
- *Data dredging* refere-se à exploração excessiva dos dados para encontrar padrões ou relações que não são teoricamente fundamentados, o que pode resultar em conclusões enganosas e enviesadas.⁹⁴
- *Selective reporting* é a prática de relatar apenas os resultados que suportam uma hipótese específica, ignorando aqueles que não a apoiam, o que pode levar a conclusões enganosas e enviesadas.⁹⁹

9.2.5 Quais práticas questionáveis podem ocorrer durante a apresentação dos resultados?

- *Cherry picking* consiste em selecionar apenas os resultados que suportam uma hipótese específica, ignorando aqueles que não a apoiam, o que pode levar a conclusões enganosas e enviesadas.⁹⁴
- *Spin* é a prática de apresentar os resultados de forma a enfatizar aspectos positivos ou minimizar aspectos negativos, o que pode levar a interpretações enganosas e enviesadas dos dados.^{100,101}
- *Beautification* é a prática de embelezar visualmente gráficos ou tabelas para aumentar impacto visual.⁹⁷
- *Data distortion* é a prática de modificar ou omitir informações nos dados para induzir interpretações específicas.⁹⁷

9.2.6 Quais práticas questionáveis podem ocorrer durante a publicação e revisão por pares?

- *Honorary authorship* refere-se à inclusão de autores que não contribuíram significativamente para o estudo, o que pode distorcer a atribuição de crédito e responsabilidade.¹⁰²
- *Ghost authorship* é a prática de não reconhecer autores que contribuíram significativamente para o estudo, o que pode distorcer a atribuição de crédito e responsabilidade.¹⁰²
- *Gold authorship* é a prática de atribuir autoria em troca de prestígio, recursos ou favorecimento político, independentemente da contribuição acadêmica.¹⁰²
- *Fake authorship* refere-se à inclusão de autores fictícios ou inexistentes em uma publicação.¹⁰²

- *Fake peer review* refere-se à prática de criar revisões por pares falsas ou fraudulentas para apoiar a publicação de um estudo, o que compromete a integridade do processo de revisão por pares e pode levar a conclusões enganosas.[?]
- *File drawer problem* refere-se à tendência de não publicar estudos com resultados negativos ou não significativos, o que pode levar a uma visão distorcida da literatura científica e dificultar a replicação de estudos.[?]
- *Salami slicing* é a prática de dividir os resultados em múltiplas publicações para aumentar o número de publicações, o que pode levar a uma má interpretação dos dados e à fragmentação do conhecimento.[?]
- *Publication bias* é a tendência de publicar apenas resultados positivos ou significativos, o que pode levar a uma visão distorcida da literatura científica e dificultar a replicação de estudos.¹⁰³
- *Duplicate publication* é a prática de publicar o mesmo estudo ou resultados em mais de uma revista, o que pode levar a uma superestimação da importância dos resultados e à confusão na literatura científica.[?]

9.3 Prevenindo práticas questionáveis em pesquisa

9.3.1 Como prevenir práticas questionáveis?

- Educação formal em integridade científica e estatística.[?]
- Pré-registro do protocolo do estudo de ensaios clínicos (ex.: ReBEC¹, ClinicalTrials.gov², revisões sistemáticas (ex.: PROSPERO³), ou outras plataformas (ex.: OSF⁴).^{104,105}
- Planos de análise detalhados.[?]
- Compartilhamento de dados/scripts (reprodutibilidade).[?]
 - Dryad Digital Repository⁵
 - figshare⁶
 - Harvard Dataverse⁷
 - Mendeley Data⁸
 - Open Science Framework⁹
 - Zenodo¹⁰
- Manuscritos reproduzíveis (RMarkdown, bookdown, etc.).[?]
- Adoção de diretrizes para redação de manuscritos (CONSORT, STROBE, PRISMA).[?]

9.4 Reações éticas e institucionais práticas questionáveis em pesquisa

- *Post-publication peer review* é a prática de revisar e criticar publicações após sua publicação, o que pode levar a uma melhor compreensão dos resultados e à correção de erros, mas também pode ser usada para desacreditar estudos sem justificativa adequada.[?]

¹<https://ensaiosclinicos.gov.br>

²<https://clinicaltrials.gov>

³<https://www.crd.york.ac.uk/prospero/>

⁴<https://osf.io>

⁵<http://datadryad.org/>

⁶<http://figshare.com/>

⁷<http://dataverse.harvard.edu/>

⁸<https://data.mendeley.com/>

⁹<https://osf.io/>

¹⁰<http://zenodo.org/>

- *Corrigendum* é uma correção publicada para corrigir erros ou imprecisões em um artigo já publicado, o que pode levar a uma melhor compreensão dos resultados e à correção de erros, mas também pode ser usada para desacreditar estudos sem justificativa adequada.[?]
- *Expression of concern* é uma declaração emitida por uma revista científica para alertar os leitores sobre preocupações com a integridade de um estudo, sem necessariamente retirar o artigo.[?]
- *Retraction* é a prática de retirar uma publicação devido a erros, fraudes ou práticas questionáveis, o que pode levar a uma melhor compreensão dos resultados e à correção de erros, mas também pode ser usada para desacreditar estudos sem justificativa adequada.[?]
- Retraction Watch¹¹ é um blog que monitora e relata casos de retratações e preocupações éticas em publicações científicas, fornecendo informações sobre práticas questionáveis e promovendo a transparência na pesquisa.[?]



O pacote *retractcheck*¹⁰⁶ fornece a função *retractcheck*^a para verificar se um artigo foi retratado usando a Open Retractions^b.

^a<https://rdrr.io/github/chartgerink/retractcheck/man/retractcheck.html>

^b<http://openretractions.com>

¹¹<https://retractionwatch.com>

RASCUNHO

PARTE 3: DO MUNDO REAL À TABELA

Da coleta à organização: estruturando dados para análises

RASCUNHO

Capítulo 10

Variáveis e fatores

10.1 Variáveis

10.1.1 O que são variáveis?

- Variáveis são informações que podem variar entre medidas em diferentes indivíduos e/ou repetições.¹⁰⁷
- Variáveis definem características de uma amostra extraída da população, tipicamente observados por aplicação de métodos de amostragem (isto é, seleção) da população de interesse.¹⁰⁸

10.1.2 Como são classificadas as variáveis?

- Quanto à informação:^{108–111}
 - Quantitativa
 - Qualitativa
- Quanto ao conteúdo:^{108–112}
 - Contínua: representam ordem e magnitude entre valores.
 - * Discreta (valores inteiros ou contáveis) vs. Contínua (valores em escala contínua).
 - * Intervalo (valor ‘0’ é arbitrário) vs. Razão (valor ‘0’ verdadeiro).
 - Categórica ordinal (numérica ou nominal): representam ordem, mas não magnitude entre valores.
 - Categórica nominal (multinomial ou dicotômica): não representam ordem ou magnitude, apenas categorias.
- Quanto à interpretação:^{108–111}
 - Dependente (desfecho)
 - Independente (preditora, covariável, confundidora, controle)
 - Mediadora
 - Moderadora
 - Modificadora
 - Auxiliar
 - Indicadora

R O pacote *base*⁵⁵ fornece a função *class*^a para identificar qual é o tipo do objeto.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/class>

R O pacote *base*⁵⁵ fornece as funções *as.numeric*^a e *as.character*^b para criar objetos numéricos e categóricos, respectivamente.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/numeric>

^b<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/character>

R O pacote *base*⁵⁵ fornece as funções *as.Date*^a e *as.logical*^b para criar objetos em formato de data e lógicos (VERDADEIRO, FALSO), respectivamente.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/as.Date>

^b<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/logical>

10.2 Transformação de variáveis

10.2.1 Por que é importante classificar as variáveis?

- Identificar corretamente os tipos de variáveis da pesquisa é uma das etapas da escolha dos métodos estatísticos adequados para as análises e representações no texto, tabelas e gráficos.¹¹⁰

10.2.2 O que é transformação de variáveis?

- Transformação significa aplicar uma função matemática à variável medida em sua unidade original.¹¹³
- A transformação visa atender aos pressupostos dos modelos estatísticos quanto à distribuição da variável, em geral a distribuição gaussiana.^{108,113}
- A dicotomização pode ser interpretada como um caso particular de agrupamento.¹¹⁴

10.2.3 Por que transformar variáveis?

- Muitos procedimentos estatísticos supõem que as variáveis — ou seus termos de erro, mais especificamente — são normalmente distribuídas. A violação dessa suposição pode aumentar suas chances de cometer um erro do tipo I ou II.¹¹⁵
- Mesmo quando se está usando análises consideradas robustas para violações dessas suposições ou testes não paramétricos (que não assumem explicitamente termos de erro normalmente distribuídos), atender a essas questões pode melhorar os resultados das análises (por exemplo, Zimmerman, 1995).¹¹⁵

10.2.4 Quais transformações de variáveis podem ser aplicadas?

- Distribuições com assimetria à direita: raiz quadrada, logaritmo natural, logaritmo base 10, transformação inversa.¹¹⁵
- Distribuições com assimetria à esquerda: reflexão e raiz quadrada, reflexão e logaritmo natural, reflexão e logaritmo base 10, reflexão e transformação inversa.¹¹⁵
- Transformação *z* de Fisher (10.1).[REF]

Transformações (Raiz, Log, Log10, Inversa)

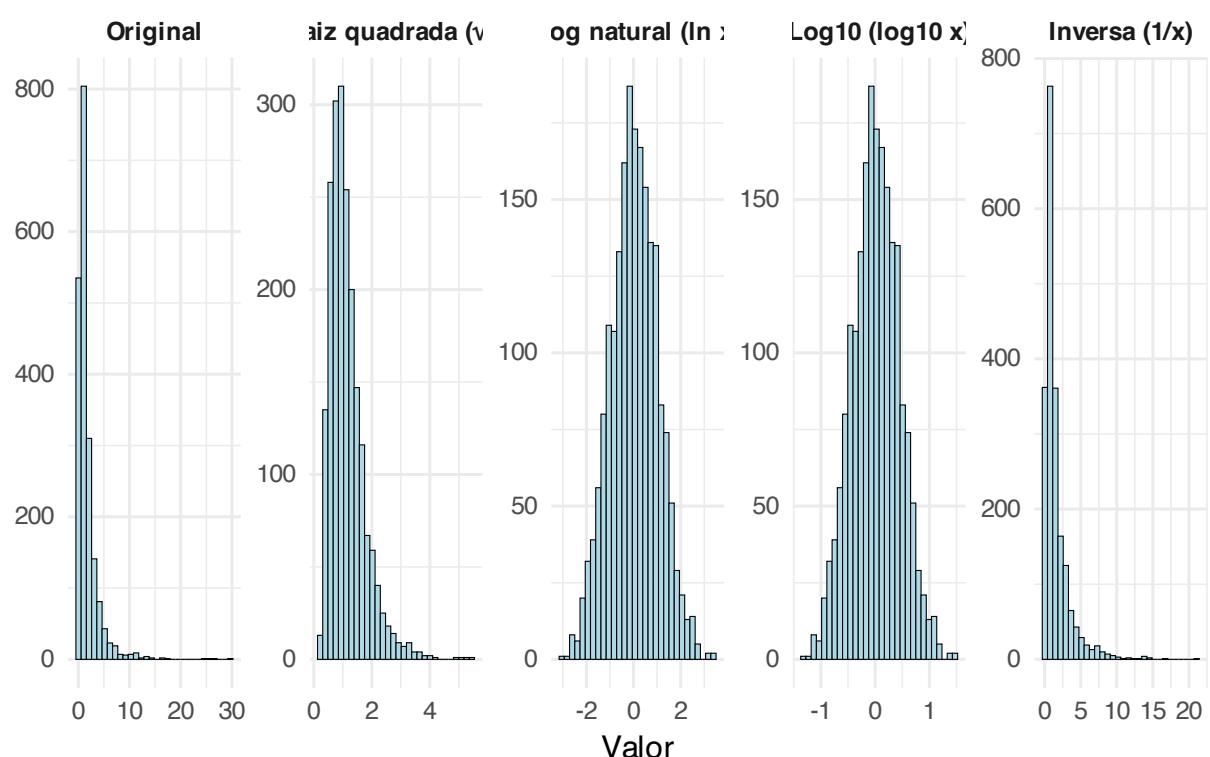


Figura 10.1: Transformações de variáveis com assimetria à direita (Original, Raiz quadrada, Log natural, Log10, Inversa).

Transformações para assimetria à esquerda (via reflexão)

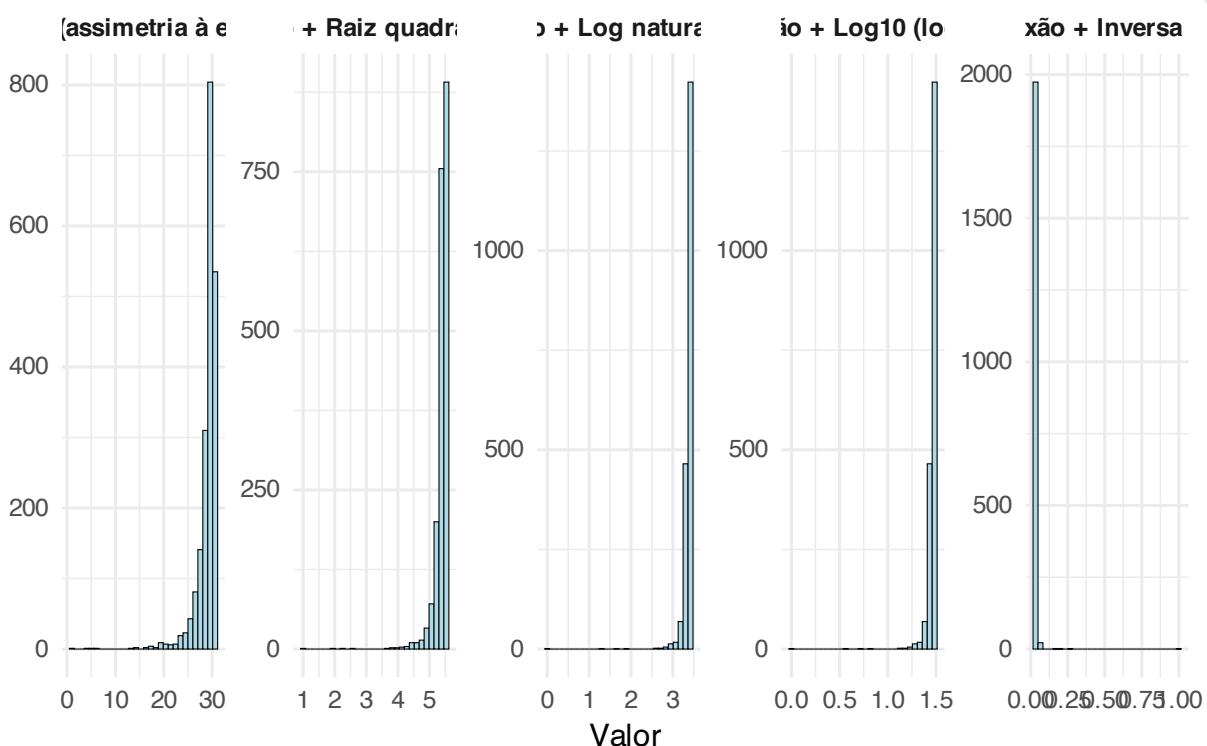


Figura 10.2: Transformações de variáveis com assimetria à esquerda (Original, Reflexão + Raiz quadrada, Reflexão + Log natural, Reflexão + Log10, Reflexão + Inversa).

$$Z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) \quad (10.1)$$

- Transformação de Box-Cox (10.2).¹¹⁶

$$Y(\lambda) = \begin{cases} \frac{Y^\lambda - 1}{\lambda}, & \text{se } \lambda \neq 0 \\ \ln(Y), & \text{se } \lambda = 0 \end{cases} \quad (10.2)$$

- Transformação arco-seno (10.3).¹¹⁵

$$Y' = \arcsin(\sqrt{Y}) \quad (10.3)$$

- Diferenciação.
- Categorização.
- Dicotomização.



O pacote *MASS*¹¹⁷ fornece a função *boxcox*^a para executar a transformação de Box-Cox.¹¹⁶

^a<https://www.rdocumentation.org/packages/MASS/versions/7.3-58.3/topics/boxcox>

10.3 Centralização de variáveis (*centering*)

10.3.1 O que é centralização?

- É uma transformação linear em que se subtrai a média da variável de cada observação. O objetivo é recentrar a variável em torno de zero, sem alterar a sua variabilidade.[?]

10.3.2 Por que centralizar?

- Facilita a interpretação dos coeficientes de regressão, especialmente em modelos com termos de interação.[?]
- Reduz a multicolinearidade entre variáveis e seus termos de interação ou polinomiais.[?]
- Mantém a escala original (apenas desloca a média).[?]

10.4 Padronização de variáveis

10.4.1 O que é padronização?

- Padronização é a transformação de uma variável contínua para uma escala comum, permitindo comparações entre variáveis medidas em diferentes unidades ou magnitudes.[?]

10.4.2 Por que padronizar?

- Facilita a interpretação em análises multivariadas.[?]
- Evita que variáveis em escalas maiores dominem os resultados de algoritmos que dependem de distância.[?]
- Melhora a comparabilidade entre estudos e bases de dados diferentes.[?]

10.4.3 Quais são os métodos de padronização mais comuns?

- Escore-Z (Z-score) (10.4): subtrair a média e dividir pelo desvio-padrão.⁵⁴

$$Z = \frac{X - \mu}{\sigma} \quad (10.4)$$

- Escala Min-Max (10.5): transformar para o intervalo [0,1].⁵⁵

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (10.5)$$

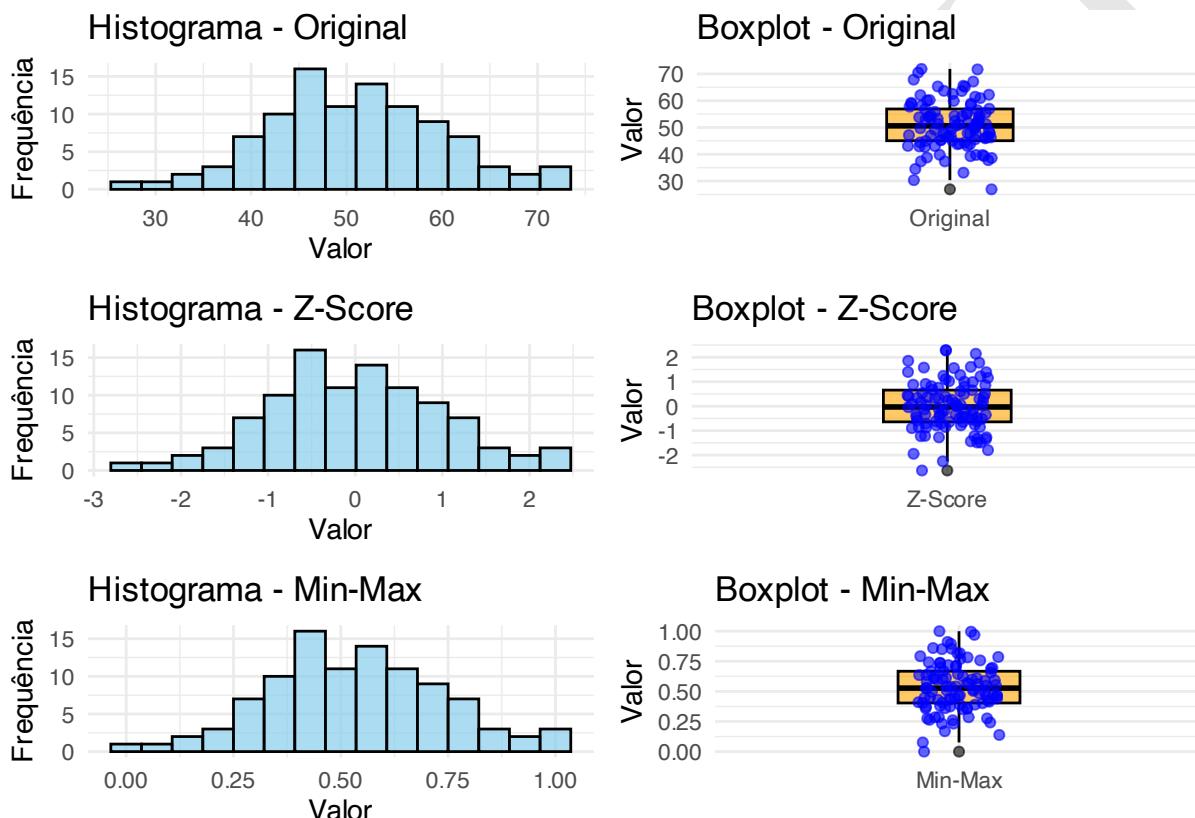


Figura 10.3: Comparação entre variáveis originais e padronizadas (Z-score e Min-Max).

10.4.4 Quais são as boas práticas de nomenclatura ao padronizar variáveis?

- Usar sufixos como `_z` ou `_std` para indicar padronização (`altura_z`, `peso_std`).⁵⁶
- Documentar no dicionário de dados como cada variável foi transformada.⁵⁷
- Evitar substituir a variável original: manter sempre a versão bruta e a padronizada.⁵⁸

R

O pacote `base`⁵⁵ fornece a função `scale`^a para calcular automaticamente a padronização (média = 0, desvio padrão = 1).

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/scale>

10.5 Categorização de variáveis contínuas

10.5.1 O que é categorização de uma variável?

- ?

10.5.2 Por que não é recomendado categorizar variáveis contínuas?

- Nenhum dos argumentos usados para defender a categorização de variáveis se sustenta sob uma análise técnica rigorosa.¹¹⁸
- Categorizar variáveis não é necessário para conduzir análises estatísticas. Ao invés de categorizar, priorize as variáveis contínuas.¹¹⁹⁻¹²¹
- Em geral, não existe uma justificativa racional (plausibilidade biológica) para assumir que as categorias artificiais subjacentes existam.¹¹⁹⁻¹²¹
- Caso exista um ponto de corte ou limiar verdadeiro que discrimine três ou mais grupos independentes, identificar tal ponto de corte ainda é um desafio.¹²²
- Categorização de variáveis contínuas aumenta a quantidade de testes de hipótese para comparações pareadas entre os quantis, inflando, portanto, o erro tipo I.¹²³
- Categorização de variáveis contínuas requer uma função teórica que pressupõe a homogeneidade da variável dentro dos grupos, levando tanto a uma perda de poder como a uma estimativa imprecisa.¹²³
- Categorização de variáveis contínuas pode dificultar a comparação de resultados entre estudos devido aos pontos de corte baseados em dados de um banco usados para definir as categorias.¹²³



O pacote *questionr*¹²⁴ fornece a função *irec*^a para executar uma interface interativa para codificação de variáveis categóricas.

^a<https://www.rdocumentation.org/packages/questionr/versions/0.7.8/topics/irec>

10.5.3 Quais são as alternativas à categorização de variáveis contínuas?

- Análise com os dados das variáveis na escala de medida original.¹¹⁸
- Análise com modelos de regressão com pesos locais (*lowess*) tais como *splines* e polinômios fracionais.¹¹⁸

10.6 Dicotomização de variáveis contínuas

10.6.1 O que são variáveis dicotônicas?

- Variáveis dicotônicas (ou binárias) podem representar categorias naturais tipo “presente/ausente”, “sim/não”?
- Variáveis dicotônicas podem representar categorias fictícias, criadas a partir de variáveis multinominais, em que cada nível é convertido em uma variável dicotônica indicadora (*dummy*).?
- Dicotomização é considerado um artefato da análise de dados, uma vez que é realizada após a coleta de dados.¹²⁵
- Geralmente são representadas por “1” (presente, sucesso) e “0” (ausente, falha).?

10.6.2 Quais argumentos são usados para defender a categorização ou dicotomização de variáveis contínuas?

- O argumento principal para dicotomização de variáveis é que tal procedimento facilita e simplifica a apresentação dos resultados, principalmente para o público em geral.¹¹⁴
- Os pesquisadores não conhecem as consequências estatísticas da dicotomização.¹¹⁸
- Os pesquisadores não conhecem os métodos adequados de análise não-paramétrica, não-linear e robusta.¹¹⁸
- As categorias representam características existentes dos participantes da pesquisa, de modo que as análises devam ser feitas por grupos e não por indivíduos.¹¹⁸
- A confiabilidade da(s) variável(eis) medida(s) é baixa e, portanto, categorizar os participantes resultaria em uma medida mais confiável.¹¹⁸

10.6.3 Por que não é recomendado dicotomizar variáveis contínuas?

- Nenhum dos argumentos usados para defender a dicotomização de variáveis se sustenta sob uma análise técnica rigorosa.¹¹⁸
- Dicotomizar variáveis não é necessário para conduzir análises estatísticas. Ao invés de dicotomizar, priorize as variáveis contínuas.^{119–121}
- Em geral, não existe uma justificativa racional (plausibilidade biológica) para assumir que as categorias artificiais subjacentes existam.^{119–121}
- Dicotomização causa perda de informação e consequentemente perda de poder estatístico para detectar efeitos.^{118,119}
- Dicotomização também classifica indivíduos com valores próximos na variável contínua como indivíduos em pontos opostos e extremos, artificialmente sugerindo que são muito diferentes.¹¹⁹
- Dicotomização pode diminuir a variabilidade das variáveis.¹¹⁹
- Dicotomização pode ocultar não-linearidades presentes na variável contínua.^{118,119}
- A média ou a mediana, embora amplamente utilizadas, não são bons parâmetros para dicotomizar variáveis.^{114,119}
- Caso exista um ponto de corte ou limiar verdadeiro que discrimine dois grupos independentes, identificar tal ponto de corte ainda é um desafio.¹²²

10.6.4 Quais cenários legitimam a dicotomização das variáveis contínuas?

- Quando existem dados e/ou análises que suportem a existência — não apenas a suposição ou teorização — de categorias com um ponto de corte claro e com significado entre elas.¹¹⁸
- Quando a distribuição da variável contínua é muito assimétrica, de modo que uma grande quantidade de observações está em um dos extremos da escala.¹¹⁸

10.6.5 Quais métodos são usados para dicotomizar variáveis contínuas?

- Em termos de tabelas de contingência 2x2, os seguintes métodos permitem¹²² a identificação do limiar verdadeiro:
 - Youden.¹²⁶
 - Gini Index.¹²⁷
 - Estatística qui-quadrado (χ^2).¹²⁸

Tabela 10.1: Tabela de variáveis indicadoras (dummy variables) criadas a partir de variáveis categóricas Sexo e Grupo.

ID	Sexo	Grupo	Sexo_Feminino	Grupo_TratA	Grupo_TratB
1	Masculino	Tratamento A	1	1	0
2	Masculino	Controle	1	0	0
3	Masculino	Tratamento A	1	1	0
4	Feminino	Tratamento B	0	0	1
5	Masculino	Controle	1	0	0
6	Feminino	Tratamento B	0	0	1
7	Feminino	Tratamento B	0	0	1
8	Feminino	Controle	0	0	0
9	Masculino	Controle	1	0	0
10	Masculino	Controle	1	0	0
11	Feminino	Controle	0	0	0
12	Feminino	Tratamento B	0	0	1

- Risco relativo (RR).¹²⁹
- Kappa (κ).¹³⁰.

10.7 Representação de variáveis categóricas

10.7.1 O que são variáveis indicadoras (dummy variables)?

- Variáveis indicadoras são variáveis dicotômicas criadas a partir dos níveis de um fator.[?]
- Cada variável indicadora assume o valor 1 quando a observação pertence àquela categoria e 0 caso contrário.[?]
- Variáveis indicadoras não representam magnitude ou ordem, apenas a presença ou ausência de uma categoria.[?]



O pacote *stats*¹³¹ fornece a função *model.matrix*^a para expandir variáveis categóricas em variáveis indicadoras.

^a<https://www.rdocumentation.org/packages/stats/versions/3.5.2/topics/model.matrix>

10.7.2 Por que variáveis indicadoras são importantes?

- Permitem a inclusão de fatores em modelos estatísticos.[?]
- Tornam explícitas as comparações entre categorias.[?]
- Garantem coerência matemática sem perder o significado conceitual das categorias.[?]

10.7.3 Quantas variáveis indicadoras são necessárias para um fator?

- Um fator com k níveis é representado por $k - 1$ variáveis indicadoras.[?]
- O nível que não gera uma variável indicadora explícita é chamado de nível de referência.[?]

10.7.4 O que é o nível de referência?

- O nível de referência é a categoria usada como base de comparação para as demais.⁵⁴
- Os coeficientes associados às variáveis indicadoras representam diferenças em relação a esse nível de referência.⁵⁴

10.7.5 Por que não se usam k variáveis indicadoras para k níveis?

- Utilizar k variáveis indicadoras gera redundância perfeita entre as variáveis.⁵⁴
- Essa redundância causa problemas de identificabilidade nos modelos, fenômeno conhecido como *dummy trap*.⁵⁴

10.7.6 Variáveis indicadoras são uma forma de dicotomização?

- Variáveis indicadoras são dicotômicas, mas não resultam da dicotomização de variáveis contínuas.⁵⁴
- Variáveis indicadoras são criadas a partir de variáveis categóricas multinominais, preservando toda a informação original do fator.⁵⁴
- Variáveis indicadoras não reduzem informação, enquanto a dicotomização de variáveis contínuas descarta informação por construção.⁵⁴

10.7.7 Variáveis indicadoras alteram os dados originais?

- Não. Variáveis indicadoras apenas representam os níveis do fator de forma numérica.⁵⁴
- A variável categórica original permanece intacta no conjunto de dados.⁵⁴

10.8 Fatores

10.8.1 O que são fatores?

- Fator é um sinônimo de variável categórica.⁵⁵
- Na modelagem, fator é sinônimo de variável preditora, em particular quando se refere à modelagem de efeitos fixos e aleatórios – os fatores (variáveis) são fatores fixos ou fatores aleatórios.⁵⁵
- Fatores são variáveis controladas pelos pesquisadores em um experimento para determinar seu efeito na(s) variável(ies) de resposta. Um fator pode assumir apenas um pequeno número de valores, conhecidos como níveis. Os fatores podem ser uma variável categórica ou baseados em uma variável contínua, mas usam apenas um número limitado de valores escolhidos pelos experimentadores.⁵⁵

O pacote *base*⁵⁵ fornece a função *as.factor*^a para converter uma variável em fator.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/factor>

10.8.2 O que são níveis de um fator?

- Níveis de um fator são as possíveis categorias que descrevem um fator.⁵⁵



O pacote *base*⁵⁵ fornece as funções *levels*^a e *nlevels*^b para listar os níveis e a quantidade deles em um fator.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/levels>

^b<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/nlevels>

RASCUNHO

Capítulo 11

Dados e metadados

11.1 Dados

11.1.1 O que são dados?

- “Tudo são dados”.¹³²
- Dados coletados em um estudo geralmente contêm erros de mensuração e/ou classificação, dados perdidos e são agrupados por alguma unidade de análise.¹³³

11.1.2 O que são *big data*?

- *Big data* refere-se a bancos de dados muito grandes com um mecanismo “R” — aleatório (*Random*), auto-reportado (*self-Reported*), reportado administrativamente (*administratively Reported*), seletivamente respondido (*selectively Responded*) — descontrolado ou desconhecido.⁷²

11.1.3 Quais são as fontes de dados?

- Experimentos.?
- Mundo real.?
- Simulação.?

11.1.4 O que são dados primários e secundários?

- Dados primários são dados originais coletados intencionalmente para uma determinada análise exploratória ou inferencial planejada a priori.¹⁰⁸
- Dados secundários compreendem dados coletados inicialmente para análises de um estudo, e são subsequentemente utilizados para outras análises.¹⁰⁸

11.1.5 O que são dados quantitativos e qualitativos?

- .?

11.2 Metadados

11.2.1 O que são metadados?

- Metadados são informações técnicas relacionadas às variáveis do estudo, tais como rótulos, limites de valores plausíveis, códigos para dados perdidos e unidades de medida.¹³⁴
- Metadados também são informações relacionadas ao delineamento e/ou protocolo do estudo, recrutamento dos participantes, e métodos para realização das medidas.¹³⁴

11.2.2 Quais são as recomendações para os metadados de um banco de dados?

- Utilize rótulos padronizados para variáveis e fatores para facilitar o reuso (reprodutibilidade) do conjuntos de dados e scripts de análise.¹³⁵
- Crie rótulos de variáveis concisos, claros e mutuamente exclusivos.¹³⁵
- Evite muitas letras maiúsculas ou outros caracteres especiais que usam a *shift*.¹³⁵
- Na existência de versões de instrumentos publicadas em diferentes anos, use o ano de publicação das escalas no rótulo.¹³⁵
- Divida o rótulo da variável ou fator em partes e ordene-as do mais geral para o mais particular geral (ex.: experimento -> repetição -> escala -> item).¹³⁵



O pacote *base*⁵⁵ fornece a função *names*^a para declarar o nome de uma variável.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/names>



O pacote *base*⁵⁵ fornece a função *labels*^a para declarar o rótulo de uma variável.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/labels>



O pacote *units*¹³⁶ fornece a função *units*^a para declarar as unidades de medida de uma variável.

^a<https://www.rdocumentation.org/packages/units/versions/0.8-3/topics/units>



O pacote *units*¹³⁶ fornece a função *valid_udunits*^a para listar as opções de unidades de medida de uma variável.

^ahttps://www.rdocumentation.org/packages/units/versions/0.8-3/topics/valid_udunits



O pacote *janitor*¹³⁷ fornece a função *clean_names*^a para formatar de modo padronizado o nome das variáveis utilizando apenas caracteres, números e o símbolo ‘_’.

^ahttps://www.rdocumentation.org/packages/janitor/versions/2.2.0/topics/clean_names



O pacote *Hmisc*¹³⁸ fornece a função *contents*^a para criar um objeto com os metadados (nomes, rótulos, unidades, quantidade e níveis das variáveis categóricas, e quantidade de dados perdidos) de um dataframe.

^a<https://www.rdocumentation.org/packages/Hmisc/versions/5.1-0/topics/contents>

RASCUNHO

Capítulo 12

Medidas e instrumentos

12.1 Escalas

12.1.1 O que são escalas?

- Uma escala de medição grosseira representa um construto de natureza contínua medido por itens tais que diferentes pontuações são agrupadas na mesma categoria no ato da coleta de dados.¹²⁵
- Em escalas grosseiras, erros são introduzidos porque as variações contínuas do constructo são colapsadas em uma mesma categorias ou separadas entre categorias próximas.¹²⁵
- Escalas tipo Likert com 5 categorias tipo “discordo totalmente”, “discordo parcialmente”, “nem concordo nem discordo”, “concordo parcialmente”, e “concordo totalmente” são escalas grosseira porque as diferenças entre as categorias não são iguais. Por exemplo, a diferença entre “discordo totalmente” e “discordo parcialmente” não é a mesma que a diferença entre “concordo parcialmente” e “concordo totalmente”.¹²⁵

R

O pacote *likert*¹³⁹ fornece a função *likert*^a para analisar respostas de instrumentos em escala tipo Likert.

^a<https://www.rdocumentation.org/packages/likert/versions/1.3.5/topics/likert>

R

O pacote *ggstats*¹⁴⁰ fornece a função *gglikert*^a para gerar um gráfico em escalas tipo Likert.

^a<https://www.rdocumentation.org/packages/ggstats/versions/0.10.0/topics/gglikert>

- O erro em escalas grosseiras é considerado sistemático mas não pode ser corrigido em nível da unidade de análise.¹²⁵

Tabela 12.1: Descrição dos itens tipo Likert do instrumento.

Itens	Discordância	Neutro	Concordância	Média	DP
Item1	40	22	38	2.94	1.38
Item2	36	20	44	3.12	1.42
Item3	38	34	28	2.82	1.32



Figura 12.1: Exemplo de instrumento com 3 itens tipo Likert com 5 categorias cada.

Tabela 12.2: Dados brutos com medidas únicas.

Unidade de análise	Pressão arterial, braço esquerdo (mmHg)
1	118
2	113
3	116
4	110
5	111
6	116
7	120
8	111
9	120
10	112

12.2 Medição e Medidas

12.2.1 O que é medição?

- Processo empírico, realizado por meio de um instrumento, que estabelece uma correspondência rigorosa e objetiva entre uma observação e uma categoria em um modelo da observação.¹⁴¹
- Esse processo tem como objetivo distinguir de maneira substantiva a manifestação observada de outras possíveis manifestações que também possam ser diferenciadas.¹⁴¹

12.2.2 O que são medidas diretas?

- ?

12.2.3 O que são medidas derivadas?

- ?

12.2.4 O que são medidas por teoria?

- ?

12.2.5 O que são medidas únicas?

- A medida única da pressão arterial sistólica no braço esquerdo resulta em um valor pontual.⁷
- Medidas únicas obtidas de diferentes unidades de análise podem ser consideradas independentes se observadas outras condições na coleta de dados.⁷
- O valor pontual será considerado representativo da variável para a unidade de análise (ex.: **120 mmHg** para o participante #9).

12.2.6 O que são medidas repetidas?

- As medidas repetidas podem ser tabuladas separadamente, por exemplo para análise da confiabilidade de obtenção dessa medida.⁷
- A medida repetida da pressão arterial no braço esquerdo resulta em um conjunto de valores pontuais (ex.: **110 mmHg**, **118 mmHg** e **116 mmHg** para o participante #5).

Tabela 12.3: Dados brutos com medidas repetidas.

Unidade de análise	Pressão arterial, braço esquerdo (mmHg) #1	Pressão arterial, braço esquerdo (mmHg) #2	Pressão arterial, braço esquerdo (mmHg) #3
1	114	112	112
2	115	120	113
3	115	110	120
4	117	116	114
5	110	118	116
6	110	120	113
7	118	114	117
8	111	112	119
9	120	112	117
10	110	115	115

Tabela 12.4: Dados brutos com medidas repetidas agregadas.

Unidade de análise	Pressão arterial, braço esquerdo (mmHg) média
1	113
2	116
3	115
4	116
5	115
6	114
7	116
8	114
9	116
10	113

- As medidas repetidas podem ser agregadas por algum parâmetro — ex.: média, mediana, máximo, mínimo, entre outros —, observando-se a relevância biológica, clínica e/ou metodológica desta escolha.[?]
- Medidas agregadas obtidas de diferentes unidades de análise podem ser consideradas independentes se observadas outras condições na coleta de dados.[?]
- O valor agregado será considerado representativo da variável para a unidade de análise (ex.: média = **115 mmHg** para o participante #5).

R

O pacote *stats*¹³¹ fornece a função *aggregate*^a para agregar medidas repetidas utilizando uma função personalizada.

^a<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/aggregate>

12.2.7 O que são medidas seriadas?

- Medidas seriadas são possivelmente relacionadas e, portanto, dependentes na mesma unidade de análise.[?]
- Por exemplo, a medida seriada da pressão arterial no braço esquerdo, em intervalos tipicamente regulares (ex.: **114 mmHg**, **120 mmHg** e **110 mmHg** em **1 min**, **2 min** e **3 min**, respectivamente, para o participante #1).

Tabela 12.5: Dados brutos com medidas seriadas não agregadas.

Unidade de análise	Tempo (min)	Pressão arterial, braço esquerdo (mmHg)
1	1	114
1	2	120
1	3	110
2	1	119
2	2	120
2	3	114
3	1	116
3	2	114
3	3	116
4	1	113

Tabela 12.6: Dados brutos com medidas seriadas não agregadas.

Unidade de análise	Pressão arterial, braço esquerdo (mmHg) amplitude
1	10
2	6
3	2
4	6
5	1
6	8
7	9
8	10
9	7
10	5

- Medidas seriadas também agregadas por parâmetros — ex.: máximo, mínimo, amplitude — são consideradas representativas da variação temporal ou de uma característica de interesse (ex.: amplitude = **10 mmHg** para o participante #1).



O pacote *stats*¹³¹ fornece a função *aggregate*^a para agregar medidas repetidas utilizando uma função personalizada.

^a<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/aggregate>

12.2.8 O que são medidas múltiplas?

- Medidas múltiplas também são possivelmente relacionadas e, portanto, são dependentes na mesma unidade de análise. Medidas múltiplas podem ser obtidas de modo repetido para análise agregada ou seriada.[?]
- A medida de pressão arterial bilateral resulta em um conjunto de valores pontuais (ex.: braço esquerdo = **114 mmHg**, braço direito = **118 mmHg** para o participante #8). Neste caso, ambos os valores pontuais são considerados representativos daquela unidade de análise.

Tabela 12.7: Dados brutos com medidas múltiplas.

Unidade de análise	Pressão arterial, braço esquerdo (mmHg)	Pressão arterial, braço direito (mmHg)
1	117	115
2	120	118
3	112	118
4	112	112
5	116	112
6	112	118
7	115	113
8	114	118
9	119	114
10	112	116

R

O pacote *stats*¹³¹ fornece a função *aggregate*^a para agrregar medidas repetidas utilizando uma função personalizada.

^a<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/aggregate>

12.3 Erro de medida

12.3.1 O que são erros de medida?

- A natureza dos erros de medida são em geral atribuídos aos (1) instrumentos utilizados e variações no protocolo, na medida em que o seu tamanho médio pode ser reduzido por modificações e melhorias nesses instrumentos; e (2) variações genuínas medida em de curto prazo.¹⁴²
- Estimativa pontual (um número exato) é um evento de probabilidade 0 sob um modelo contínuo.[?]
- Precisão como faixa $\pm \epsilon$ tem probabilidade > 0 , mensurável e dependente de σ .[?]
- Isso motiva trabalhar com intervalos (faixas) em vez de pontos.[?]

12.3.2 Quais fontes de variabilidade são comumente investigadas?

- Intra/Entre participantes (isto é, unidades de análise).¹⁴³
- Intra/Entre repetições.¹⁴³
- Intra/Entre observadores.¹⁴³

12.4 Instrumentos

12.4.1 O que são instrumentos?

- ?

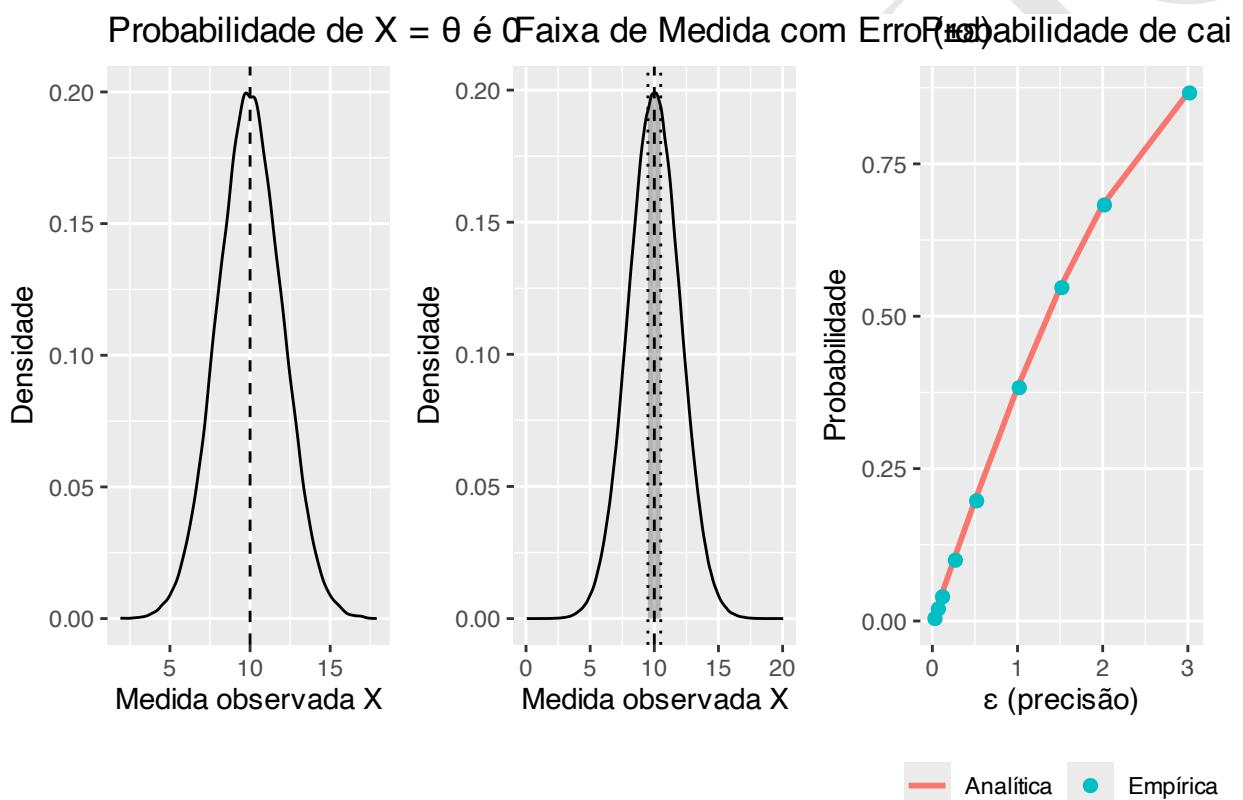


Figura 12.2: Ilustração do conceito de erro de medida em um modelo simples com erro normal. A linha tracejada indica o valor verdadeiro (desconhecido na prática). A área sombreada representa a probabilidade de cair na faixa $|X - \theta| \leq \epsilon$, que é > 0 . A probabilidade de 'acertar no ponto' $X = \theta$ é 0 (linha tracejada).

12.5 Acurácia e precisão

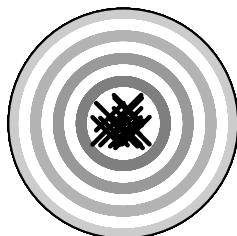
12.5.1 O que é acurácia?

- Acurácia expressa a proximidade de concordância entre uma mensuração e o valor real.¹⁴⁴
- Acurácia está para medidas como validade está para instrumentos de medida.[?]

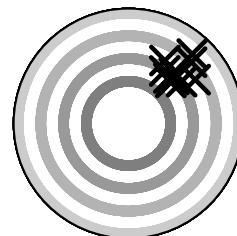
12.5.2 O que é precisão?

- Precisão se refere à proximidade de concordância entre resultados de testes independentes obtidos nas mesmas condições de teste.¹⁴⁴
- Precisão é um índice de quão próximo os resultados podem ser repetidos entre mensurações repetidas.¹⁴⁵
- Precisão está para medidas como confiabilidade está para instrumentos de medida.[?]

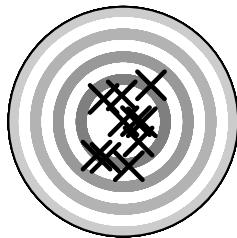
Acurácia alta, Precisão alta



Acurácia baixa, Precisão alta



Acurácia alta, Precisão baixa



Acurácia baixa, Precisão baixa

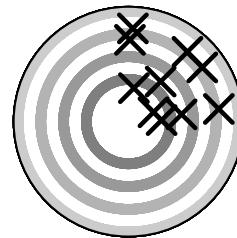


Figura 12.3: Acurácia e precisão como propriedades de uma medida.

12.6 Viés e variabilidade

12.6.1 Qual é a relação entre viés e variabilidade?

- ?

Viés alto, Variância baixa Viés baixo, Variância alta Viés baixo, Variância baixa

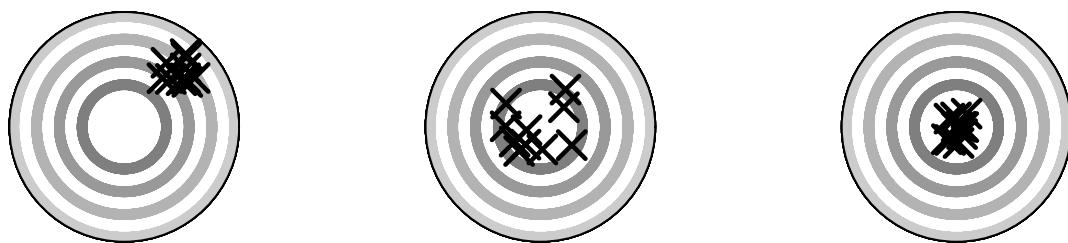


Figura 12.4: Viés e variabilidade de uma medida.

RASCUNHO

Capítulo 13

Tabulação de dados

13.1 Planilhas eletrônicas

13.1.1 Qual a organização de uma tabela de dados?

- As informações podem ser organizadas em formato de dados retangulares (ex.: matrizes, tabelas, quadro de dados) ou não retangulares (ex.: listas).¹⁴⁵
- Cada variável possui sua própria coluna (vertical).¹⁴⁶
- Cada observação possui sua própria linha (horizontal).¹⁴⁶
- Cada valor possui sua própria célula especificada em um par (linha, coluna).¹⁴⁶
- Cada célula possui seu próprio dado.¹⁴⁶



O pacote *DataEditR*¹⁴⁷ fornece a função *data_edit*^a para interativamente criar, editar e salvar a tabela de dados.

^a<https://www.rdocumentation.org/packages/DataEditR/versions/0.1.5/topics/dataInput>

13.1.2 Qual a estrutura básica de uma tabela para análise estatística?

- Use apenas 1 (uma) planilha eletrônica para conter todas as informações coletadas. Evite múltiplas abas no mesmo arquivo, assim como múltiplos arquivos quando possível.¹⁴⁸
- Use apenas 1 (uma) linha de cabeçalho para nomear os fatores e variáveis do seu estudo.¹⁴⁸

Tabela 13.1: Estrutura básica de uma tabela de dados.

V1	V2	V3	V4
$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,4}$
$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	$x_{2,4}$
$x_{3,1}$	$x_{3,2}$	$x_{3,3}$	$x_{3,4}$
$x_{4,1}$	$x_{4,2}$	$x_{4,3}$	$x_{4,4}$
$x_{5,1}$	$x_{5,2}$	$x_{5,3}$	$x_{5,4}$

Tabela 13.2: Formato longo para medidas repetidas (múltiplas linhas por sujeito; colunas = variáveis)

Linha	Variável	Valor
1	V1	$x_{1,1}$
1	V2	$x_{1,2}$
1	V3	$x_{1,3}$
1	V4	$x_{1,4}$
2	V1	$x_{2,1}$
2	V2	$x_{2,2}$
2	V3	$x_{2,3}$
2	V4	$x_{2,4}$
3	V1	$x_{3,1}$
3	V2	$x_{3,2}$
3	V3	$x_{3,3}$
3	V4	$x_{3,4}$
4	V1	$x_{4,1}$
4	V2	$x_{4,2}$
4	V3	$x_{4,3}$
4	V4	$x_{4,4}$
5	V1	$x_{5,1}$
5	V2	$x_{5,2}$
5	V3	$x_{5,3}$
5	V4	$x_{5,4}$

Tabela 13.3: Formato largo para medidas repetidas (1 variável; colunas = tempos)

id	T1	T2	T3	T4
1	$x_{1,T1}$	$x_{1,T2}$	$x_{1,T3}$	$x_{1,T4}$
2	$x_{2,T1}$	$x_{2,T2}$	$x_{2,T3}$	$x_{2,T4}$
3	$x_{3,T1}$	$x_{3,T2}$	$x_{3,T3}$	$x_{3,T4}$
4	$x_{4,T1}$	$x_{4,T2}$	$x_{4,T3}$	$x_{4,T4}$
5	$x_{5,T1}$	$x_{5,T2}$	$x_{5,T3}$	$x_{5,T4}$

- Tipicamente, cada linha representa um participante e cada coluna representa uma variável ou fator do estudo. Estudos com medidas repetidas dos participantes podem conter múltiplas linhas para o mesmo participante (repetindo os dados na mesma coluna, conhecido como *formato curto*) ou só uma linha para o participante (repetindo os dados em colunas separadas, conhecido como *formato longo*).¹⁴⁹

13.1.3 O que usar para organizar tabelas para análise computadorizada?

- Seja consistente em: códigos para as variáveis categóricas; códigos para dados perdidos; nomes das variáveis; identificadores de participantes; nome dos arquivos; formato de datas; uso de caracteres de espaço.^{148,149}
- Crie um dicionário de dados (metadados) em um arquivo separado contendo: nome da variável, descrição da variável, unidades de medida e valores extremos possíveis.¹⁴⁸
- Use recursos para validação de dados antes e durante a digitação de dados.^{148,149}

Tabela 13.4: Formatação recomendada para tabela de dados.

ID	Data.Coleta	Estado.Civil	Numero.Filhos
1	13-01-2026	casado	NA
2	14-01-2026	casado	1
3	15-01-2026	casado	NA
4	16-01-2026	solteiro	NA
5	17-01-2026	casado	NA
6	18-01-2026	solteiro	0
7	19-01-2026	solteiro	NA
8	20-01-2026	solteiro	NA
9	21-01-2026	casado	NA
10	22-01-2026	solteiro	NA

Tabela 13.5: Formatação não recomendada para tabela de dados.

ID	Data de Coleta	Estado Civil	Número de Filhos
1	13-01-2026	casado	NA
2	14-01-2026	Casado	1
3	15-01-2026	casado	NaN
4	16-01-2026	Solteiro	N/A
5	17-01-2026	Casado	N.A.
6	18-01-2026	solteiro	0
7	19-01-2026	solteiro	
8	20-01-2026	Solteiro	na
9	21-01-2026	casado	n.a.
10	22-01-2026	Solteiro	999



O pacote *data.table*¹⁵⁰ fornece a função *melt.data.table*^a para reorganizar a tabela em diferentes formatos.

^a<https://www.rdocumentation.org/packages/data.table/versions/1.14.8/topics/melt.data.table>

13.1.4 O que não usar para organizar tabelas para análise computadorizada?

- Não deixe células em branco: substitua dados perdidos por um código sistemático (ex.: NA [*not available*]).¹⁴⁸
- Não inclua análises estatísticas ou gráficos nas tabelas de dados brutos.¹⁴⁸
- Não utilize cores como informação. Se necessário, crie colunas adicionais — variáveis instrumentais ou auxiliares — para identificar a informação de modo que possa ser analisada.¹⁴⁸
- Não use células mescladas.
- Delete linhas e/ou colunas totalmente em branco (sem unidades de análise e/ou sem variáveis).

13.1.5 O que é recomendado e o que deve ser evitado na organização das tabelas para análise?

RASCUNHO

Capítulo 14

Dados perdidos e imputados

14.1 Dados perdidos

14.1.1 O que são dados perdidos?

- Dados perdidos são dados não coletados de um ou mais participantes, para uma ou mais variáveis.¹⁵¹



O pacote *base*⁵⁵ fornece a função *is.na*^a para identificar que elementos de um objeto são dados perdidos.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/na>

14.1.2 Qual o problema de um estudo ter dados perdidos?

- Uma grande quantidade de dados perdidos pode comprometer a integridade científica do estudo, considerando-se que o tamanho da amostra foi estimado para observar um determinado tamanho de efeito mínimo.¹⁵¹
- Perda de participantes no estudo por dados perdidos pode reduzir o poder estatístico (erro tipo II).¹⁵¹
- Não existe solução globalmente satisfatória para o problema de dados perdidos.¹⁵¹

Tabela 14.1: Simulação de uma amostra (n=10) de um ensaio clínico aleatorizado (dados com perdas aleatórias).

id	Grupo	Idade	Sexo	Desfecho (pré)	Desfecho (pós)
1	Controle	53	F	57.0	41.3
2	Controle	64	F	45.3	70.0
3	Controle	65	M	39.3	NA
4	Intervenção	66	F	47.8	NA
5	Controle	44	M	39.7	65.7
6	Intervenção	NA	F	42.7	NA
7	Intervenção	67	M	43.7	64.9
8	Intervenção	NA	F	33.1	63.3
9	Controle	68	F	58.4	61.6
10	Controle	74	M	51.5	54.3

14.2 Mecanismos geradores de dados perdidos

14.2.1 Quais os mecanismos geradores de dados perdidos?

- Dados perdidos completamente ao acaso (*missing completely at random*, MCAR), em que os dados perdidos estão distribuídos aleatoriamente nos dados da amostra.^{152,153}

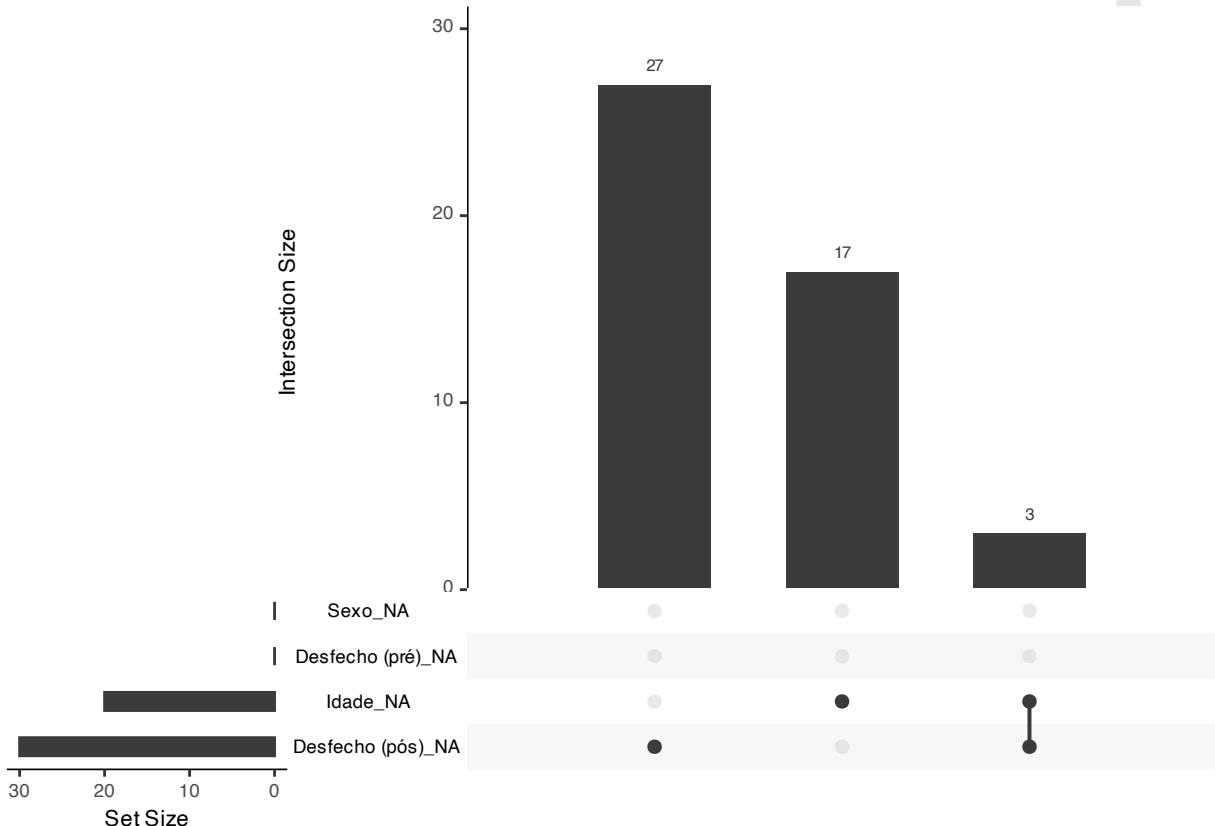


Figura 14.1: Representação gráfica de dados perdidos completamente ao acaso (MCAR) em um estudo randomizado controlado (RCT).

- Dados perdidos ao acaso (*missing at random*, MAR), em que a probabilidade de ocorrência de dados perdidos é relacionada a outras variáveis medidas.^{152,153}
- Dados perdidos não ao acaso (*missing not at random*, MNAR), em que a probabilidade da ocorrência de dados perdidos é relacionada com a própria variável.^{152,153}

14.2.2 Como identificar o mecanismo gerador de dados perdidos em um banco de dados?

- Por definição, não é possível avaliar se os dados foram perdidos ao acaso (MAR) ou não (MNAR).¹⁵²
- Testes t e regressões logísticas podem ser aplicados para identificar relações entre variáveis com e sem dados perdidos, criando um fator de análise ('dado perdido' = 1, 'dado observado' = 0).¹⁵²

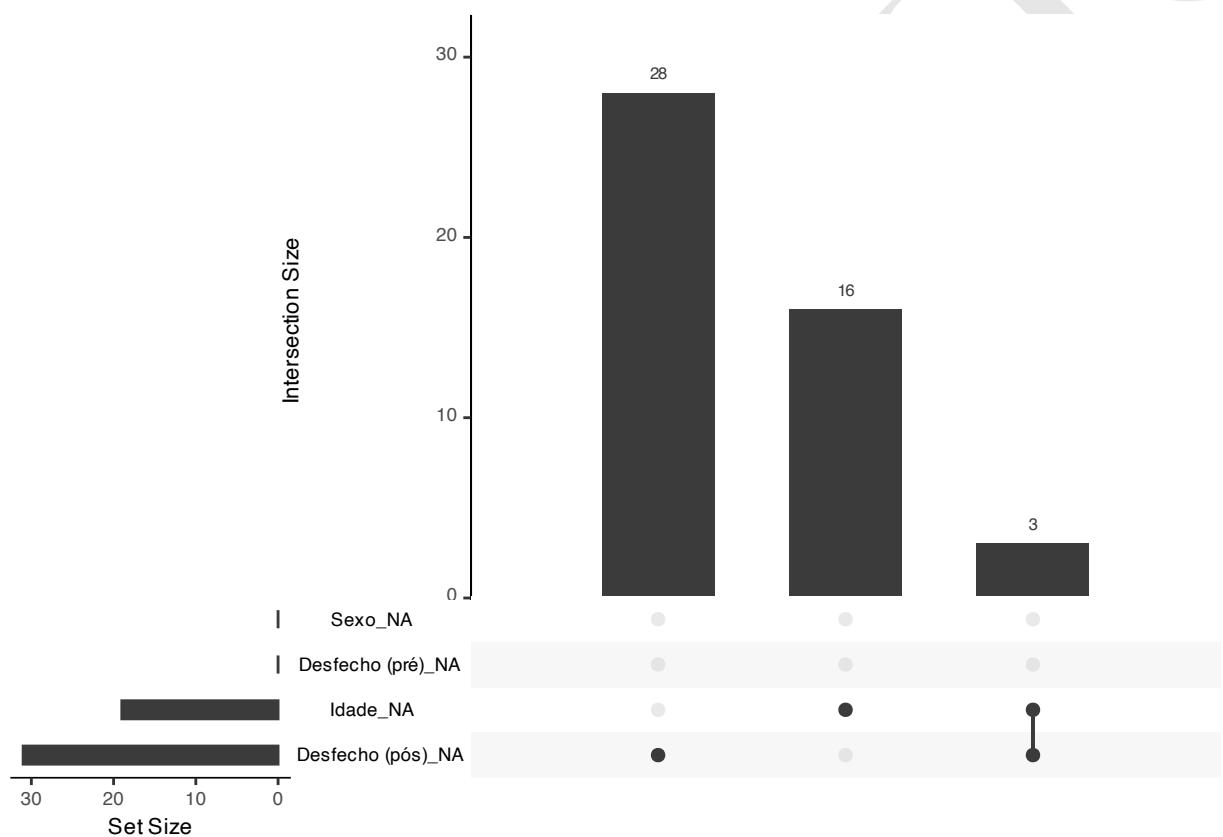


Figura 14.2: Representação gráfica de dados perdidos ao acaso (MAR) em um estudo randomizado controlado (RCT).

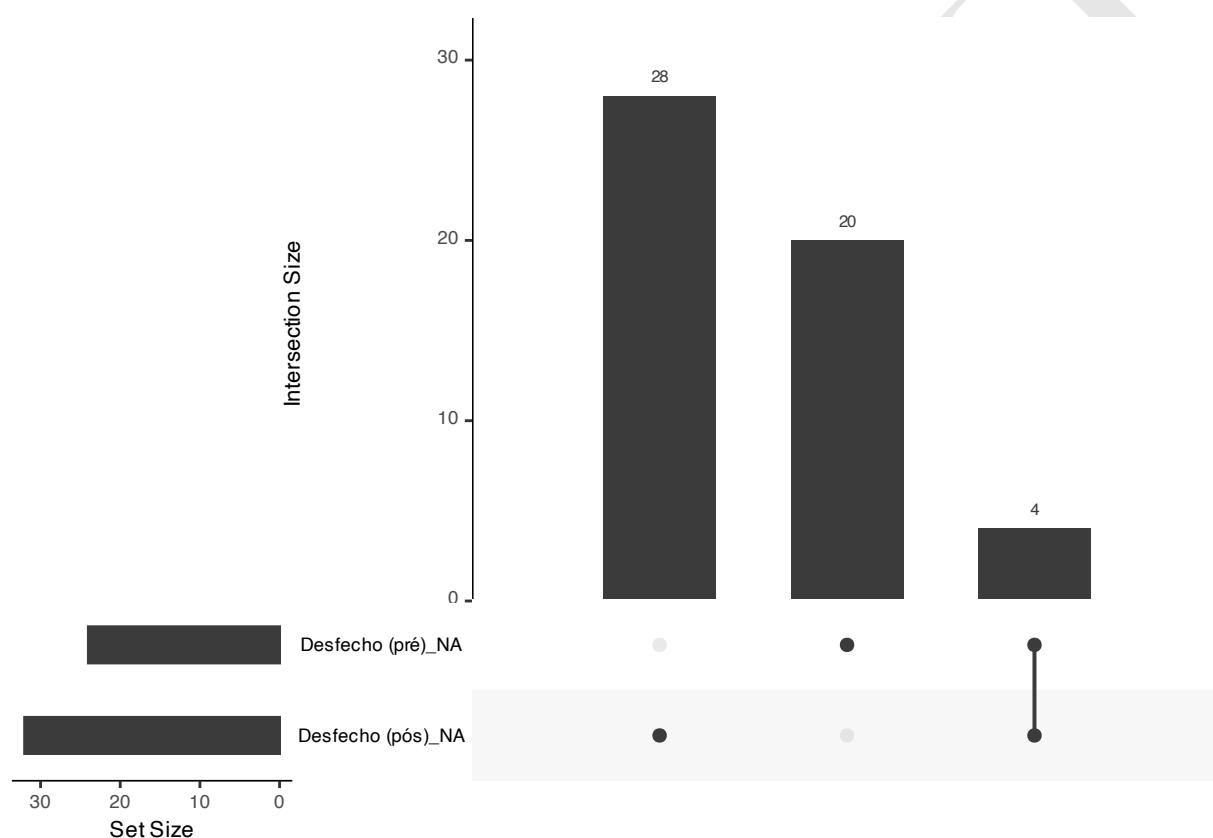


Figura 14.3: Representação gráfica de dados perdidos não ao acaso (MNAR) em um estudo randomizado controlado (RCT).



O pacote *misty*¹⁵⁴ fornece a função *na.test*^a para executar o Little's Missing Completely at Random (MCAR) test¹⁵⁵.

^a<https://www.rdocumentation.org/packages/misty/versions/0.5.0/topics/na.test>



O pacote *naniar*¹⁵⁶ fornece a função *mcar_test*^a para executar o Little's Missing Completely at Random (MCAR) test¹⁵⁵.

^ahttps://www.rdocumentation.org/packages/naniar/versions/1.0.0/topics/mcar_test



O pacote *naniar*¹⁵⁶ fornece a função *gg_miss_upset*^a para gerar o gráfico Upset para visualizar padrões de dados perdidos.

^ahttps://www.rdocumentation.org/packages/naniar/versions/1.0.0/topics/gg_miss_upset

14.3 Estratégias para lidar com dados perdidos

14.3.1 Que estratégias podem ser utilizadas na coleta de dados quando há expectativa de perda amostral?

- Na expectativa de ocorrência de perda amostral, com consequente ocorrência de dados perdidos, recomenda-se ampliar o tamanho da amostra com um percentual correspondente a tal estimativa (ex.: 10%), embora ainda não corrija potenciais vieses pela perda.¹⁵¹

14.3.2 Que estratégias podem ser utilizadas na análise quando há dados perdidos?

- Na ocorrência de dados perdidos, a análise mais comum compreende apenas os ‘casos completos’, com exclusão de participantes com algum dado perdido nas variáveis do estudo. Em casos de grande quantidade de dados perdidos, pode-se perder muito poder estatístico (erro tipo II elevado).¹⁵¹
- A análise de dados completos é válida quando pode-se argumentar que a probabilidade de o participante ter dados completos depende apenas das covariáveis e não dos desfechos.¹⁵³
- A análise de dados completos é eficiente quando todos os dados perdidos estão no desfecho, ou quando cada participante com dados perdidos nas covariáveis também possui dados perdidos nos desfechos.¹⁵³



O pacote *base*⁵⁵ fornece a função *na.omit*^a para remover dados perdidos de um objeto em um banco de dados.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/na.fail>



O pacote *stats*¹³¹ fornece a função *complete.cases*^a para identificar os casos completos - isto é, sem dados perdidos - em um banco de dados.

^a<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/complete.cases>

14.3.3 Que estratégias podem ser utilizadas na redação de estudos em que há dados perdidos?

- Informar: o número de participantes com dados perdidos; diferenças nas taxas de dados perdidos entre os braços do estudo; os motivos dos dados perdidos; quaisquer diferenças entre os participantes com e sem dados perdidos; o padrão de perda; os métodos para tratamento de dados perdidos das variáveis em análise; os resultados de análises de sensibilidade; as implicações dos dados perdidos na interpretação do resultados.¹⁵⁷

14.4 Dados imputados

14.4.1 O que são dados imputados?

- ?

14.4.2 Quando a imputação de dados é indicada?

- A análise com imputação de dados pode ser útil quando pode-se argumentar que os dados foram perdidos ao acaso (MAR); quando o desfecho foi observado e os dados perdidos estão nas covariáveis; e variáveis auxiliares — preditoras do desfecho e não dos dados perdidos — estão disponíveis.¹⁵³
- Na ocorrência de dados perdidos, a imputação de dados (substituição por dados simulados plausíveis preditos pelos dados presentes) pode ser uma alternativa para manter o erro tipo II estipulado no plano de análise.¹⁵¹

14.4.3 Quais são os métodos de imputação de dados?

- Modelos lineares e logísticos podem ser utilizados para imputar dados perdidos em variáveis contínuas e dicotômicas, respectivamente.¹⁵⁸
- Os métodos de imputação de dados mais robustos incluem a imputação multivariada por equações encadeadas (*multivariate imputation by chained equations*, MICE)¹⁵⁹ e a correspondência média preditiva (*predictive mean matching*, PMM).^{160,161}

R

Os pacotes *mice*¹⁵⁹ e *miceadds*¹⁶² fornecem funções *mice*^a e *mi.anova*^b para imputação multivariada por equações encadeadas, respectivamente, para imputação de dados.

^a<https://www.rdocumentation.org/packages/mice/versions/3.16.0/topics/mice>

^b<https://www.rdocumentation.org/packages/miceadds/versions/3.16-18/topics/mi.anova>

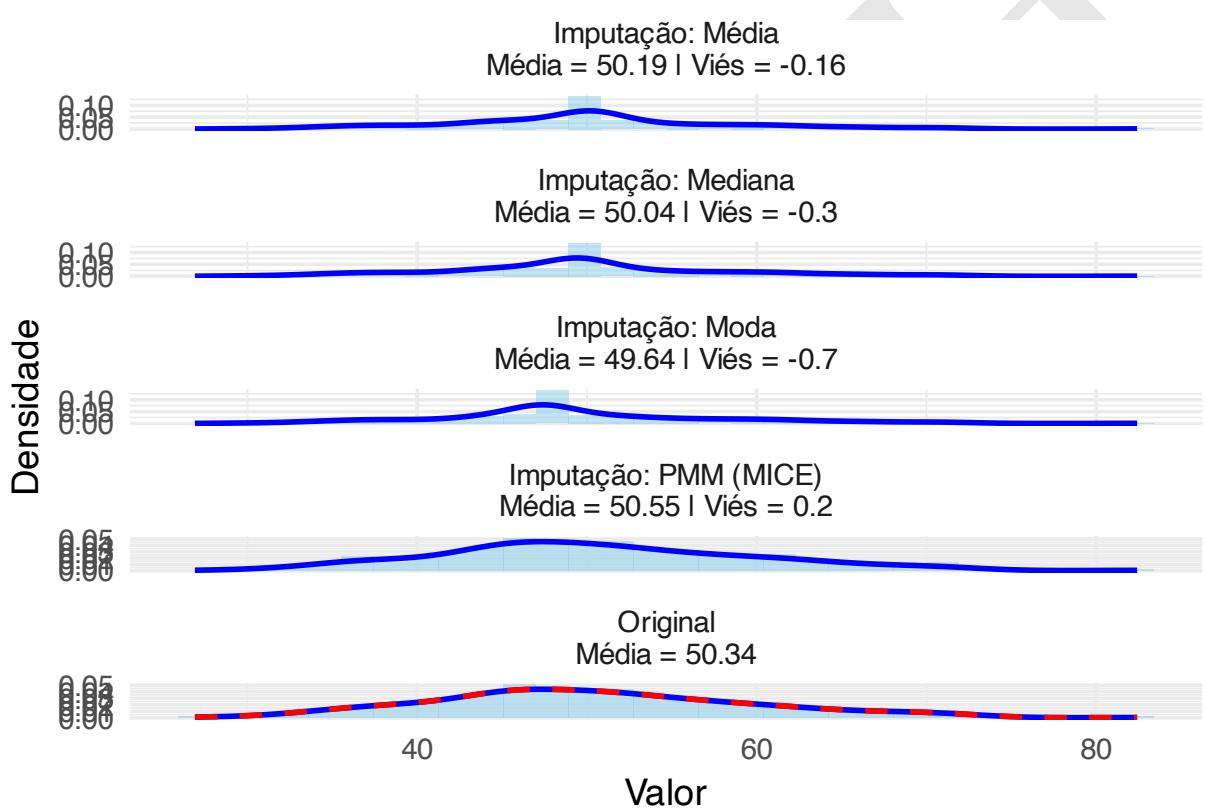


Figura 14.4: Impacto de métodos de imputação na distribuição de uma variável contínua com dados perdidos.

RASCUNHO

Capítulo 15

Dados anonimizados e sintéticos

15.1 Dados anonimizados

15.1.1 O que são dados anonimizados?

- ?

15.1.2 Com anonimizar os dados de um banco?

- ?



O pacote *ids*¹⁶³ fornece a função *random_id*^a para criar identificadores aleatórios por criptografia.

^ahttps://www.rdocumentation.org/packages/ids/versions/1.0.1/topics/random_id



O pacote *hash*¹⁶⁴ fornece a função *hash*^a para criar identificadores por objetos *hash*.

^a<https://www.rdocumentation.org/packages/hash/versions/3.0.1/topics/hash>



O pacote *anonymizer*¹⁶⁵ fornece a função *anonymize*^a para criar uma versão anônima de variáveis em um banco de dados.

^a<https://www.rdocumentation.org/packages/anonymizer/versions/0.2.0/topics/anonymize>



O pacote *digest*¹⁶⁶ fornece a função *digest*^a para criar identificadores por objetos *hash* criptografados ou não.

^a<https://www.rdocumentation.org/packages/digest/versions/0.6.33/topics/digest>

15.2 Dados sintéticos

15.2.1 O que são dados sintéticos?

• ?



O pacote *synthpop*¹⁶⁷ fornece a função *syn*^a para criar bancos de dados sintéticos a partir de um banco de dados real.

^a<https://www.rdocumentation.org/packages/synthpop/versions/1.8-0/topics/syn>

PARTE 4: ANÁLISES DESCRIPTIVAS E EXPLORATÓRIAS

Primeiros passos na análise: descrever, visualizar e explorar padrões nos dados

RASCUNHO

Capítulo 16

Distribuições e parâmetros

16.1 Distribuições de probabilidade

16.1.1 O que são distribuições de probabilidade?

- Uma distribuição de probabilidade é uma função que descreve os valores possíveis ou o intervalo de valores de uma variável (eixo horizontal) e a frequência com que cada valor é observado (eixo vertical).¹⁰⁸

16.1.2 Como representar distribuições de probabilidade?

- Tabelas de frequência, polígonos de frequência, gráficos de barras, histogramas e *boxplots* são formas de representar distribuições de probabilidade.¹⁶⁸
- Tabelas de frequência mostram as categorias de medição e o número de observações em cada uma. É necessário conhecer o intervalo de valores (mínimo e máximo), que é dividido em intervalos arbitrários chamados “intervalos de classe”.¹⁶⁸
- Se houver muitos intervalos, não haverá redução significativa na quantidade de dados, e pequenas variações serão perceptíveis. Se houver poucos intervalos, a forma da distribuição não poderá ser adequadamente determinada.¹⁶⁸
- A quantidade de intervalos pode ser determinada pelo método de Sturges, que é dado pela fórmula $k = 1 + 3.322 \times \log_{10}(n)$, onde k é o número de intervalos e n é o número de observações.¹⁶⁹
- A quantidade de intervalos pode ser determinada pelo método de Scott, que é dado pela fórmula $h = 3.5 \times \text{sd}(x) \times n^{-1/3}$, onde h é a largura do intervalo, $\text{sd}(x)$ é o desvio-padrão e n é o número de observações.¹⁷⁰
- A quantidade de intervalos pode ser determinada pelo método de Freedman-Diaconis, que é dado pela fórmula $h = 2 \times \text{IQR}(x) \times n^{-1/3}$, onde h é a largura do intervalo, $\text{IQR}(x)$ é o intervalo interquartil e n é o número de observações.¹⁷¹
- A largura das classes pode ser determinada dividindo o intervalo total de observações pelo número de classes. Recomenda-se larguras iguais, mas larguras desiguais podem ser usadas quando existirem grandes lacunas nos dados ou em contextos específicos. Os intervalos devem ser mutuamente exclusivos e não sobrepostos, evitando intervalos abertos (ex.: $<5, >10$).¹⁶⁸
- Polígonos de frequência são gráficos de linhas que conectam os pontos médios de cada barra do histograma. Eles são úteis para comparar duas ou mais distribuições de frequência.¹⁶⁸
- Gráficos de barra verticais ou horizontais representam a distribuição de frequências de uma variável categórica. A altura de cada barra é proporcional à frequência da classe. A largura da barra é igual à largura da

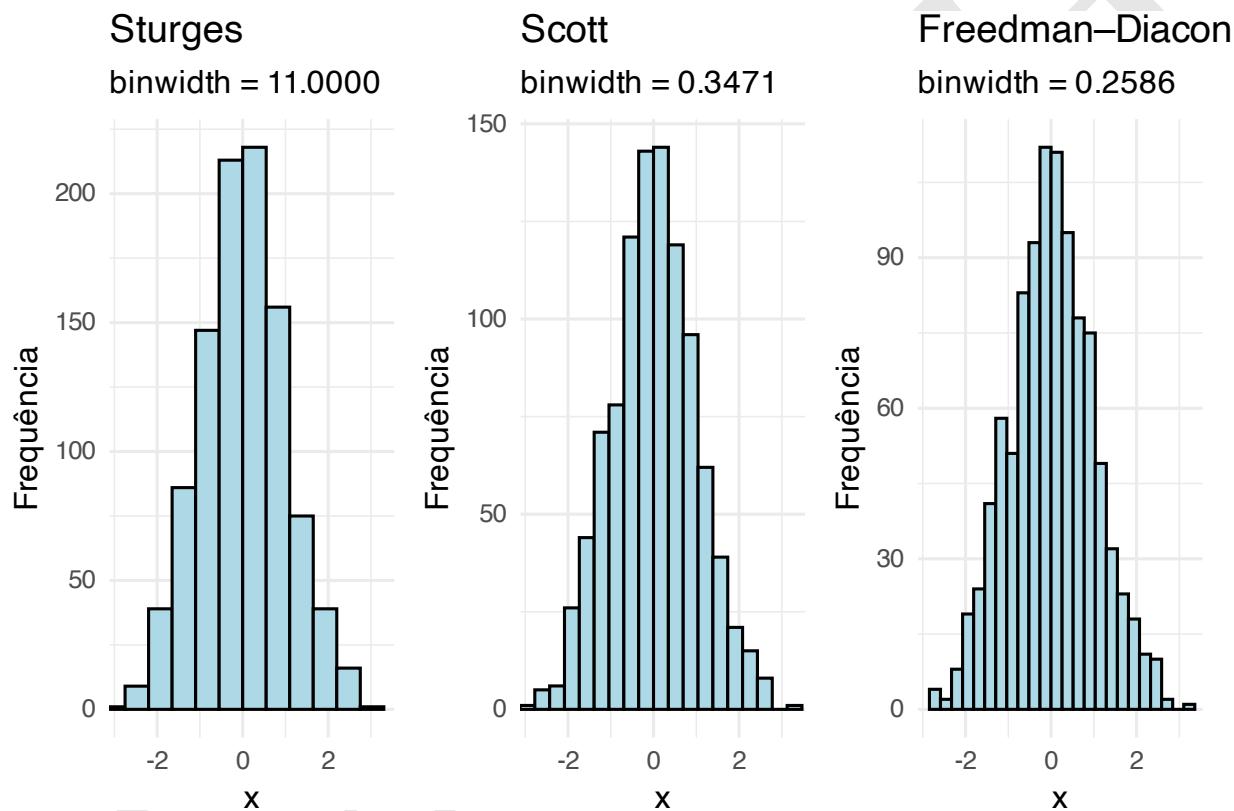


Figura 16.1: Histogramas com diferentes métodos de binning.: Sturges, Scott e Freedman-Diaconis.

classe. A área de cada barra é proporcional à frequência da classe. A área total do gráfico de barras é igual ao número total de observações.¹⁶⁸

- Histogramas representam a distribuição de frequências de uma variável contínua. A altura de cada barra é proporcional à frequência da classe. A largura da barra é igual à largura da classe. A área de cada barra é proporcional à frequência da classe. A área total do histograma é igual ao número total de observações.¹⁶⁸
- *Boxplots* representam a distribuição de frequências de uma variável contínua. A linha central divide os dados em duas partes iguais (mediana ou Q2). A caixa inferior representa o primeiro quartil (Q1) e a caixa superior representa o terceiro quartil (Q3). A linha inferior é o mínimo e a linha superior é o máximo. Os valores atípicos são representados por pontos individuais.¹⁶⁸



O pacote *grDevices*¹⁷² fornece a função `nclass`^a para determinar a quantidade de classes de um histograma com os métodos de Sturge¹⁶⁹, Scott¹⁷⁰ ou Freedman-Diaconis¹⁷¹.

^a<https://www.rdocumentation.org/packages/grDevices/versions/3.6.2/topics/nclass>



O pacote *ggplot2*¹⁷³ fornece a função `geom_freqpoly`^a para criar histogramas.

^ahttps://www.rdocumentation.org/packages/ggplot2/versions/3.5.2/topics/geom_freqpoly

16.1.3 Quais características definem uma distribuição?

- Uma distribuição pode ser definida por modelos matemáticos e caracterizada por parâmetros de tendência central, dispersão, simetria e curtose.[?]

16.1.4 Quais são as distribuições mais comuns?

- Distribuições discretas:
 - Bernoulli: resultado de um único teste com dois possíveis desfechos (sucesso ou fracasso).[?]
 - Binomial: número de sucessos em k tentativas.[?]
 - Geométrica: número de testes até o 1º sucesso.[?]
 - Binomial negativa: número de testes até o k -ésimo sucesso.[?]
 - Hipergeométrica: número de indivíduos na amostra tomados sem reposição.[?]
 - Poisson: número de eventos em um intervalo de tempo fixo.[?]
 - Uniforme: resultados (finitos) que são igualmente prováveis.[?]
 - Multinomial: resultados de múltiplos testes com mais de dois possíveis desfechos.[?]
- Distribuições contínuas:
 - Uniforme: .[?]
 - Exponencial: .[?]
 - Normal: .[?]
 - Aproximação binomial: .[?]
 - Aproximação Poisson: .[?]
 - t-Student: .[?]

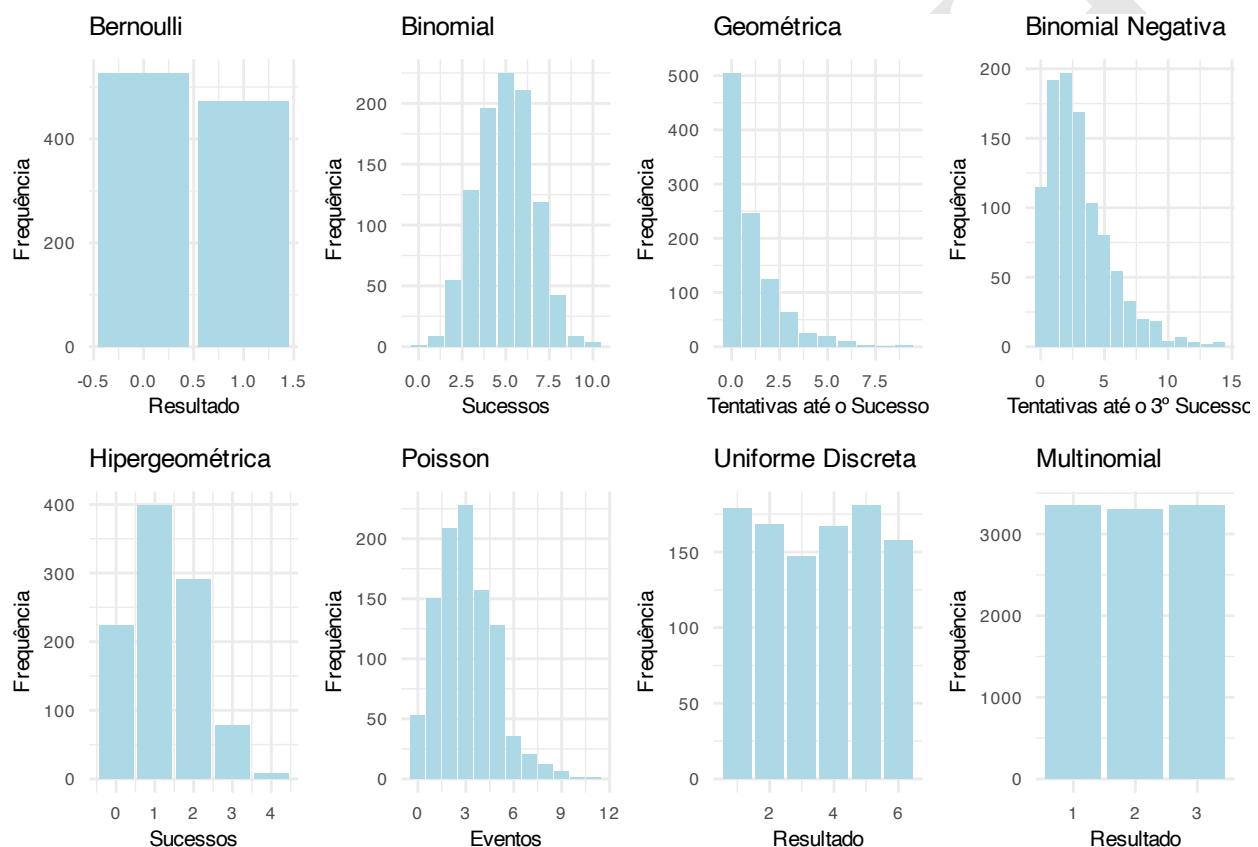


Figura 16.2: Distribuições discretas e suas funções de probabilidade.

- Qui-quadrado: ?
- Weibull: ?
- Gama: ?
- Log-normal: ?
- Beta: ?
- Logística: ?
- Pareto.?

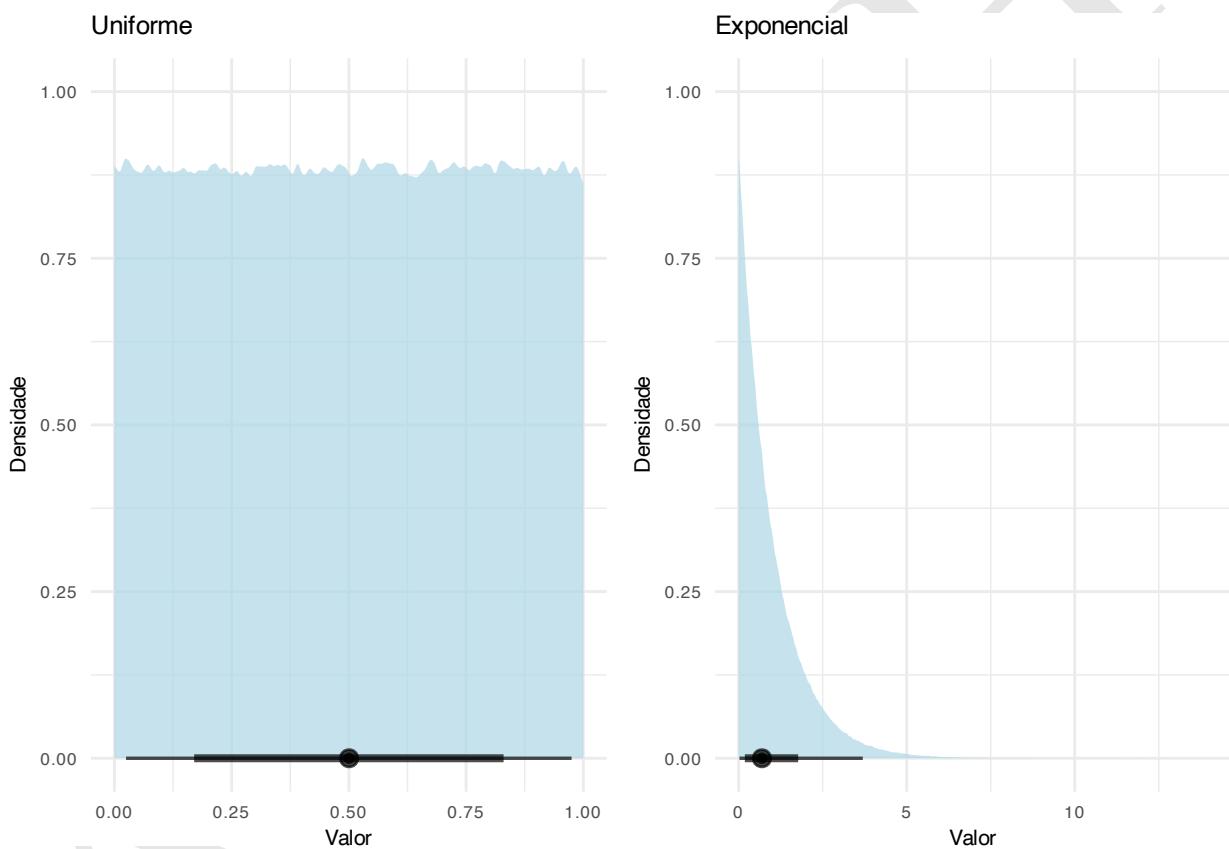


Figura 16.3: Distribuições contínuas básicas e suas funções de densidade.

16.1.5 Quais são as funções de uma distribuição?

- Função de massa de probabilidade (*probability mass function, pmf*).?
- Função de distribuição cumulativa (*cumulative distribution function, cdf*).?
- Função quantílicas (*quantile function, qf*).?
- Função geradora de números aleatórios (*random function, rf*).?

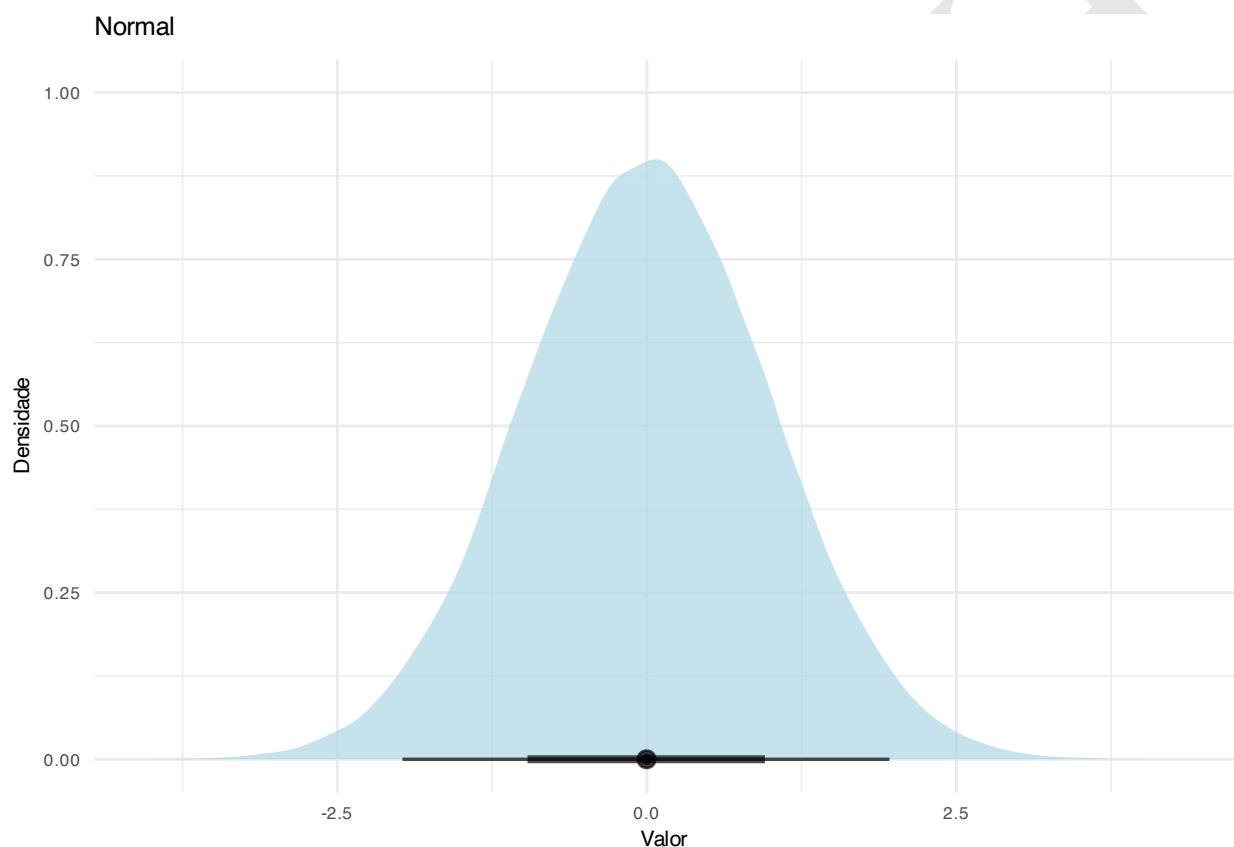


Figura 16.4: Distribuições contínuas aproximadas e suas funções de densidade.

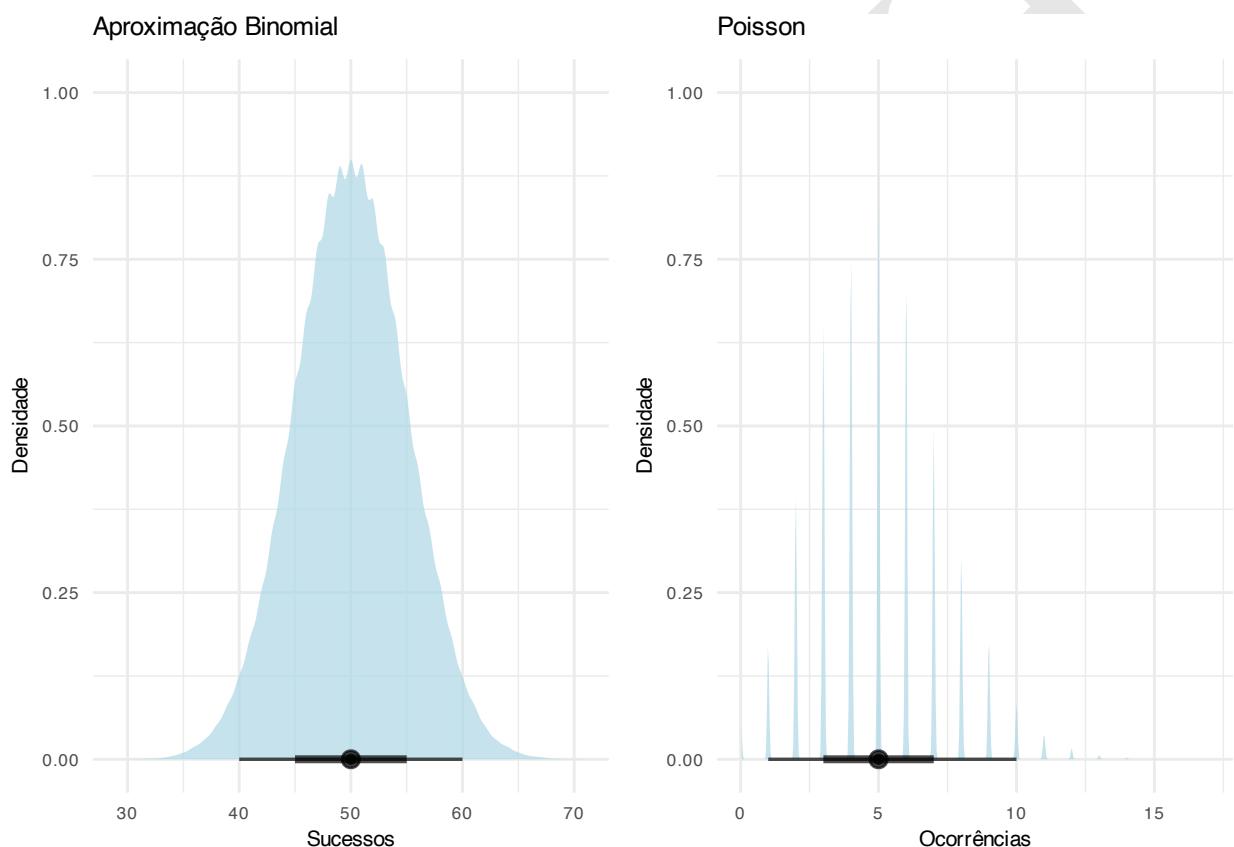


Figura 16.5: Distribuições contínuas aproximadas e suas funções de densidade.

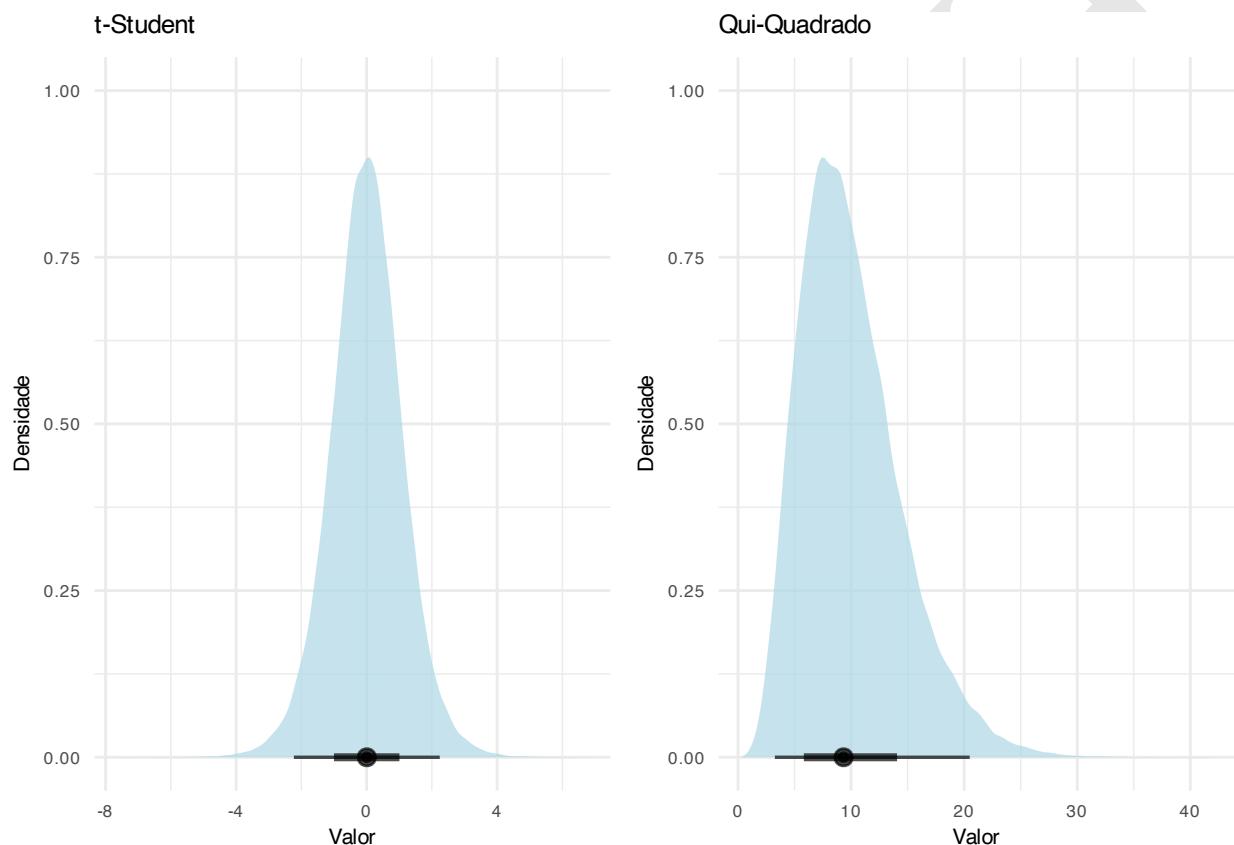


Figura 16.6: Distribuições contínuas para inferência e suas funções de densidade.

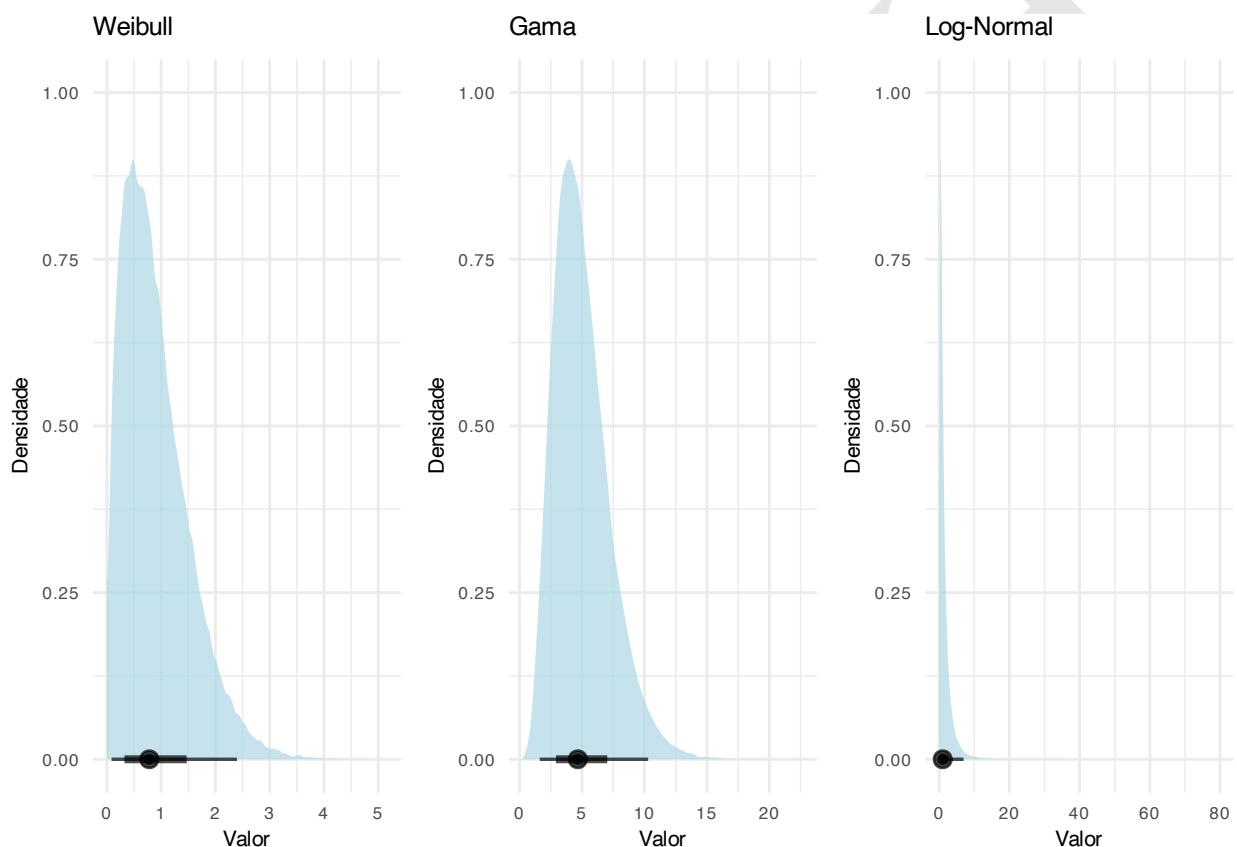


Figura 16.7: Distribuições contínuas para dados específicos e suas funções de densidade.

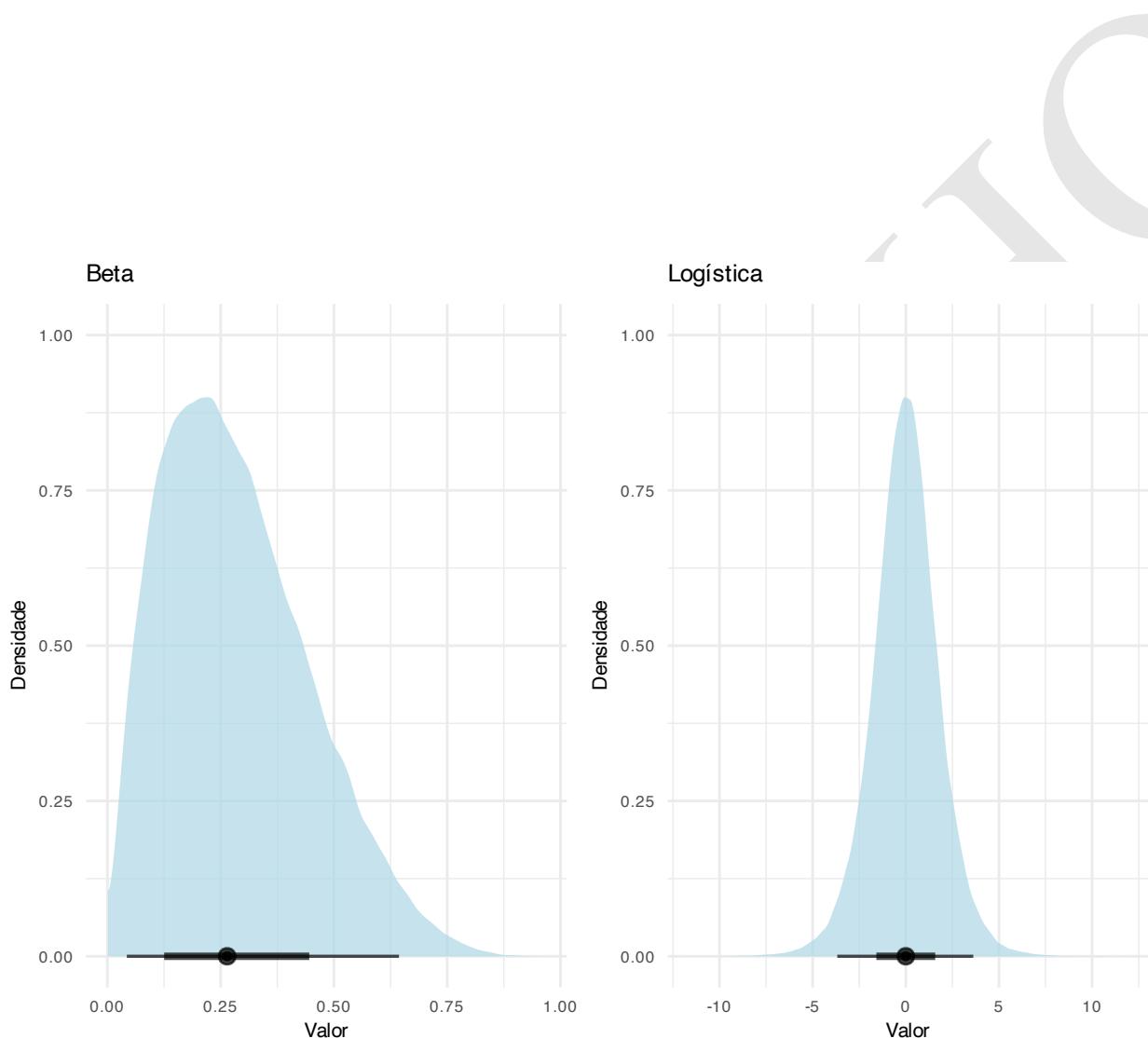


Figura 16.8: Distribuições contínuas para probabilidades e proporções e suas funções de densidade.

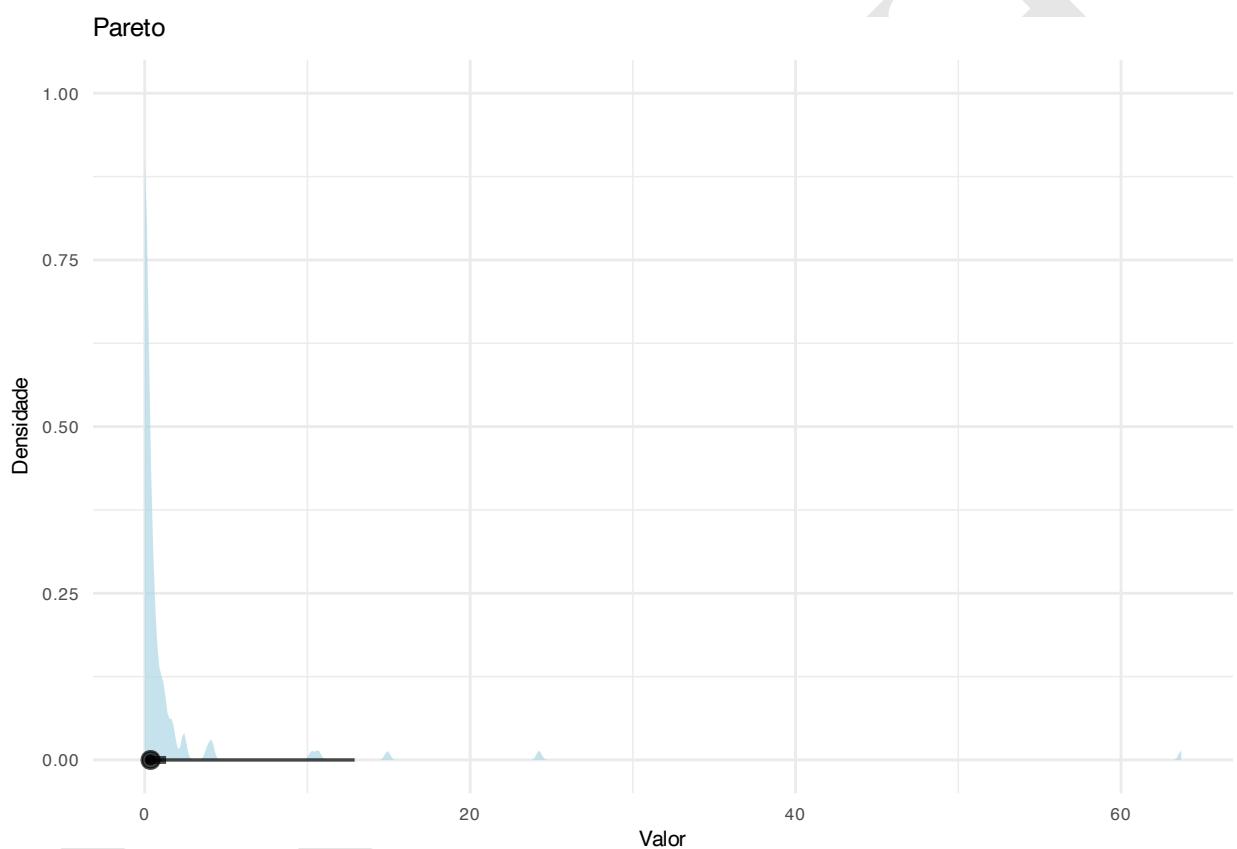


Figura 16.9: Distribuições contínuas com caudas pesadas e suas funções de densidade.

R

O pacote *stats*¹³¹ fornece funções de distribuição de probabilidade (p), funções de densidade (d), funções quantílicas (q) e funções geradoras de números aleatórios (r) para as distribuições normal^a, Student t^b, binomial^c, qui-quadrado^d, uniforme^e, dentre outras.

^a<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/Normal>

^b<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/TDist>

^c<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/Binomial>

d<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/Chisquare>e<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/Uniform>**R**

O pacote *ggdist*¹⁷⁴ fornece a função *geom_slabinterval*^a para criar gráficos de distribuição de probabilidade (p) e funções de densidade (d) as distribuições.

^ahttps://www.rdocumentation.org/packages/ggdist/versions/3.3.0/topics/geom_slabinterval

R

O pacote *ggfortify*¹⁷⁵ fornece a função *ggdistribution*^a para criar gráficos de distribuição de probabilidade (p), funções de densidade (d), funções quantílicas (q) e funções geradoras de números aleatórios (r) para as distribuições.

^a<https://www.rdocumentation.org/packages/ggfortify/versions/0.4.16/topics/ggdistribution>

16.1.6 O que é a distribuição normal?

- A distribuição normal (ou gaussiana) é uma distribuição com desvios simétricos positivos e negativos em torno de um valor central.¹⁰⁹
- Em uma distribuição normal, o intervalo de 1 desvio-padrão ($\pm 1DP$) inclui cerca de 68% dos dados; de 2 desvios-padrão ($\pm 2DP$) cerca de 95% dos dados; e no intervalo de 3 desvios-padrão ($\pm 3DP$) cerca de 99% dos dados.¹⁰⁹

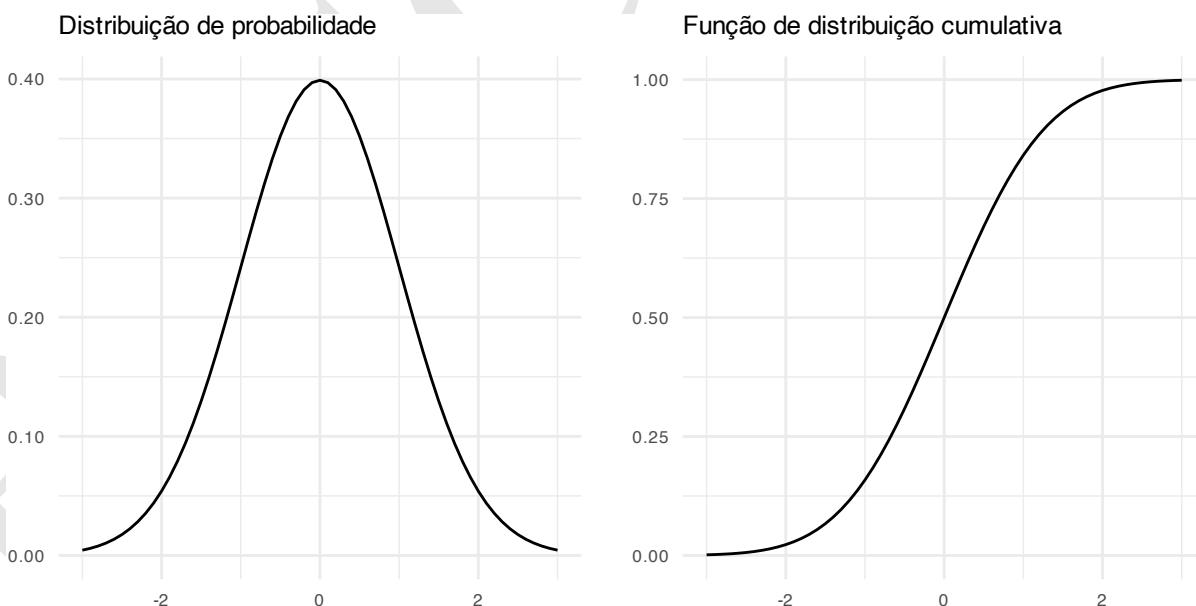


Figura 16.10: Distribuições e funções de probabilidade.

16.1.7 Que métodos podem ser utilizados para identificar a normalidade da distribuição?

- Histogramas.¹⁰⁸
- Gráficos Q-Q.¹⁰⁸
- Testes de hipótese nula:¹⁰⁸
 - Kolmogorov-Smirnov
 - Shapiro-Wilk
 - Anderson-Darling

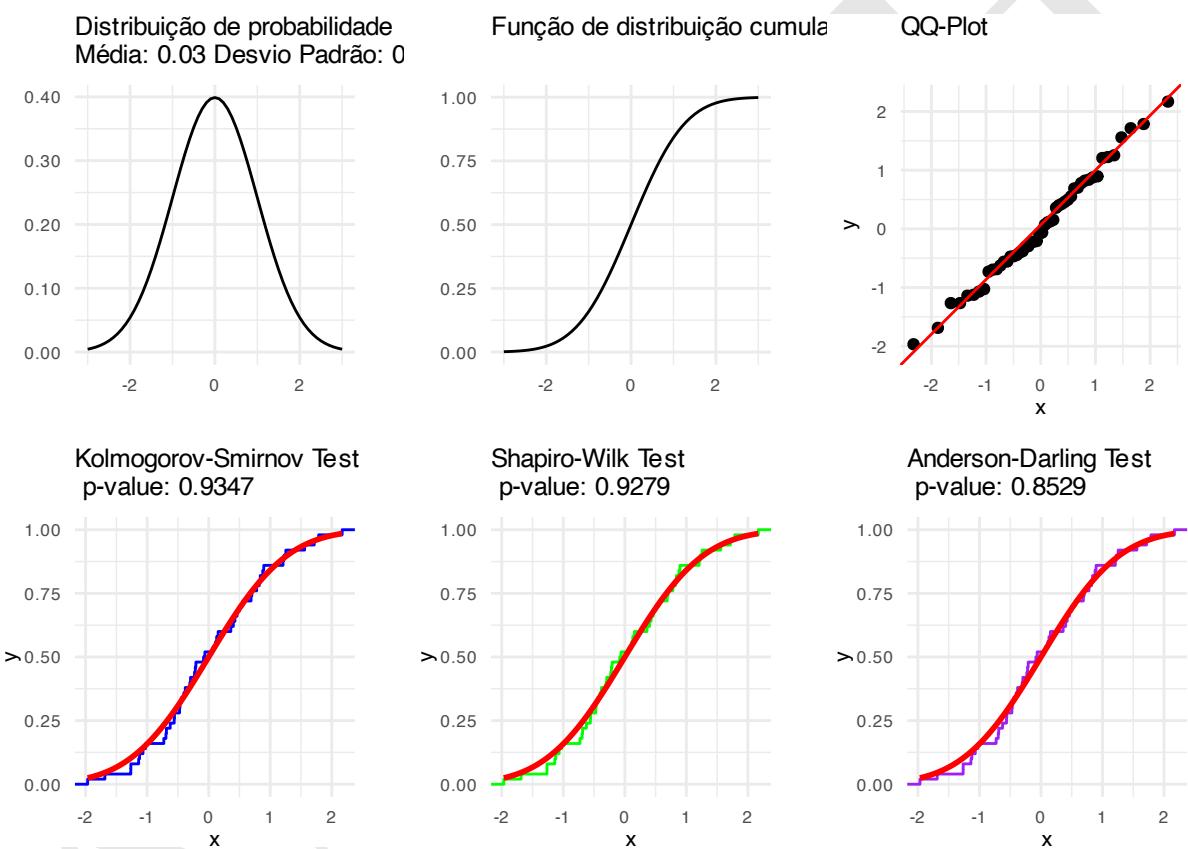


Figura 16.11: Distribuição normal e métodos de visualização e testes de normalidade.

16.1.8 O que são distribuições não-normais?

- ?

16.2 Distribuições multivariadas

16.2.1 O que são distribuições multivariadas?

- Distribuições multivariadas descrevem a probabilidade conjunta de duas ou mais variáveis aleatórias.[?]

- Exemplos de distribuições multivariadas incluem a distribuição normal multivariada, a distribuição t multivariada, a distribuição binomial multinomial e a distribuição de Dirichlet.⁷

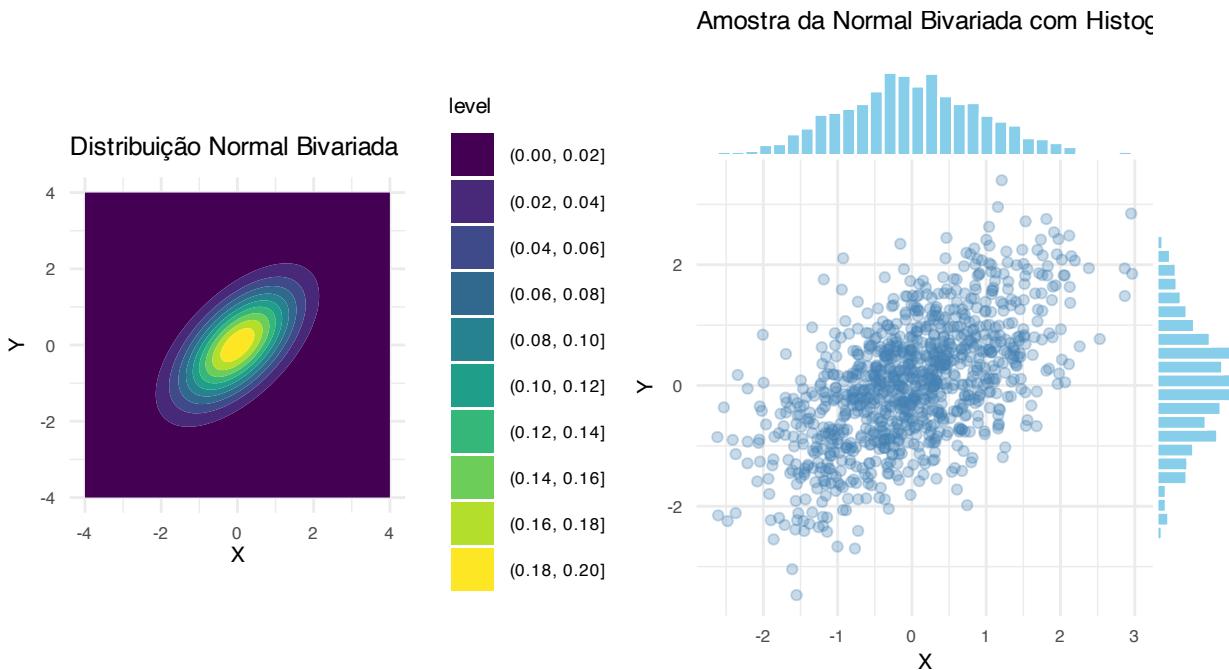


Figura 16.12: Distribuição normal bivariada e amostra simulada com histogramas marginais.

16.3 Parâmetros

16.3.1 O que são parâmetros?

- Parâmetros são informações que definem um modelo teórico, como propriedades de uma coleção de indivíduos.¹⁰⁷
- Parâmetros definem características de uma população inteira, tipicamente não observados por ser inviável ter acesso a todos os indivíduos que constituem tal população.¹⁰⁸



O pacote *base*⁵⁵ fornece a função *summary*^a para calcular diversos parâmetros descritivos.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/summary>

16.3.2 O que é uma análise paramétrica?

- Testes paramétricos possuem suposições sobre as características e/ou parâmetros da distribuição dos dados na população.¹⁰⁸
- Testes paramétricos assumem que: a variável é quantitativa numérica (contínua); os dados foram amostrados de uma população com distribuição normal; a variância da(S) amostra(s) é igual à da população; as amostras foram selecionadas de modo aleatório na população; os valores de cada amostra são independentes entre si.^{108,109}
- Testes paramétricos são baseados na suposição de que os dados amostrais provêm de uma população com parâmetros fixos determinando sua distribuição de probabilidade.⁸

16.3.3 O que é uma análise não paramétrica?

- Testes não-paramétricos fazem poucas suposições, ou menos rigorosas, sobre as características e/ou parâmetros da distribuição dos dados na população.^{108,109}
- Testes não-paramétricos são úteis quando as suposições de normalidade não podem ser sustentadas.¹⁰⁹

16.3.4 Devemos testar as suposições de normalidade?

- Testes preliminares de normalidade não são necessários para a maioria dos testes paramétricos de comparação, pois eles são robustos contra desvios moderados da normalidade. Normalidade da distribuição deve ser estabelecida para a população.¹⁷⁶

16.3.5 Por que as análises paramétricas são preferidas?

- Em geral, testes paramétricos são mais robustos (isto é, possuem menores erros tipo I e II) que seus testes não-paramétricos correspondentes.^{108,177,178}
- Testes não-paramétricos apresentam menor poder estatístico (maior erro tipo II) comparados aos testes paramétricos correspondentes.¹⁰⁹

16.3.6 Que parâmetros podem ser estimados?

- Parâmetros de tendência central.^{109,179}
- Parâmetros de dispersão.^{109,179,180}
- Parâmetros de proporção.^{109,179,181,181}
- Parâmetros de distribuição.¹⁷⁹
- Parâmetros de extremos.¹⁰⁹



O pacote *base*⁵⁵ fornece a função *summary*^a para calcular diversos parâmetros descritivos.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/summary>

16.4 Erro

16.4.1 Que parâmetros de erro podem ser estimados?

- Margem de erro (ME).^(16.1)

$$ME = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad (16.1)$$

- Erro-padrão da média (EPM) (16.2) (*sigma* conhecido) e (16.3) (*sigma* desconhecido).^{180,182}

$$EPM = \frac{\sigma}{\sqrt{n}} \quad (16.2)$$

$$\widehat{EPM} = \frac{s}{\sqrt{n}} \quad (16.3)$$

- Intervalo de confiança para variância conhecida (16.4) e desconhecida (16.5): Captura a média populacional correspondente ao nível de significância α pré-estabelecido^{109,179,182,183}.

$$IC = \bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad (16.4)$$

$$IC = \bar{x} \pm t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}} \quad (16.5)$$

16.5 Tendência central

16.5.1 Que parâmetros de tendência central podem ser estimados?

- Média aritmética (16.6), ponderada (16.7), geométrica (16.8) ou harmônica (16.9).^{109,179,184}

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (16.6)$$

$$\bar{x}_p = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad (16.7)$$

$$\bar{x}_g = \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}} \quad (16.8)$$

$$\bar{x}_h = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} \quad (16.9)$$

- Mediana (16.10).^{109,179,185}

$$\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{se } n \text{ é ímpar} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se } n \text{ é par} \end{cases} \quad (16.10)$$

- Moda (16.11), onde $f(x)$ é a função de frequência absoluta ou relativa e x_1, x_2, \dots, x_n são os valores observados.^{109,179,185}

$$Mo \in \arg \max_{x \in \{x_1, \dots, x_n\}} f(x) \quad (16.11)$$

- Moda (dados agrupados) (16.12), onde: L = limite inferior da classe modal; f_1 = frequência da classe modal; f_0 = frequência da classe anterior à classe modal; f_2 = frequência da classe posterior à classe modal; h = amplitude da classe modal.

$$Mo = L + \frac{(f_1 - f_0)}{(f_1 - f_0) + (f_1 - f_2)} \cdot h \quad (16.12)$$

- A posição relativa das medidas de tendência central (média, mediana e moda) depende da forma da distribuição.¹⁸⁵
- Em uma distribuição normal, as três medidas são idênticas.¹⁸⁵
- A média é sempre puxada para os valores extremos, por isso é deslocada para a cauda em distribuições assimétricas.¹⁸⁵
- A mediana fica entre a média e a moda em distribuições assimétricas.¹⁸⁵

- A moda é o valor mais frequente e, portanto, se localiza no pico da distribuição assimétrica.¹⁸⁵
- Uma distribuição pode uma moda (unimodal), duas modas (bimodal) ou três ou mais modas (multimodal), indicando a presença de mais de um valor com alta frequência.¹⁸⁵

Distribuições Unimodal, Bimodal e Multimodal

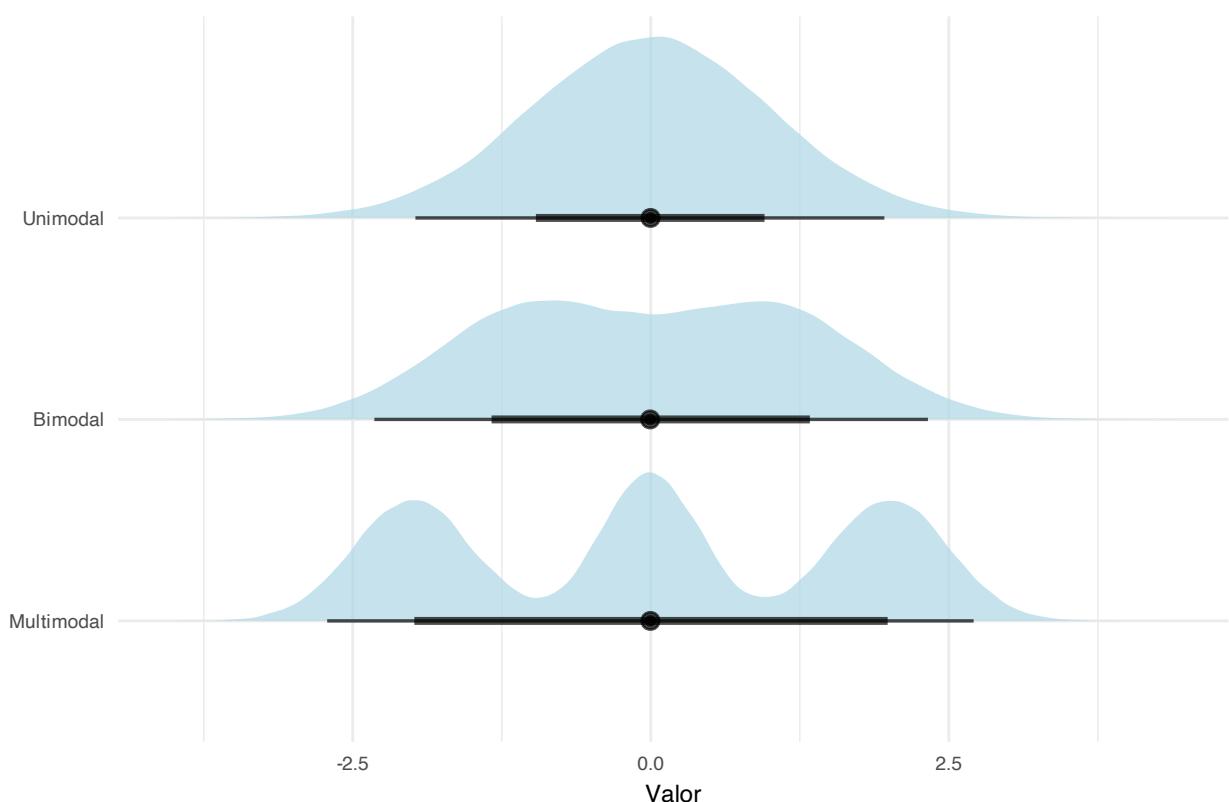


Figura 16.13: Distribuições unimodal, bimodal e multimodal.



O pacote *base*⁵⁵ fornece a função *summary*^a para calcular diversos parâmetros descritivos.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/summary>

16.5.2 Como escolher o parâmetro de tendência central?

- A mediana é preferida à média quando existem poucos valores extremos na distribuição, alguns valores são indeterminados, ou há uma distribuição aberta, ou os dados são medidos em uma escala ordinal.¹⁸⁵
- A moda é preferida quando os dados são medidos em uma escala nominal.¹⁸⁵
- A média geométrica é preferida quando os dados são medidos em uma escala logarítmica.¹⁸⁵

16.6 Dispersão

16.6.1 Que parâmetros de dispersão podem ser estimados?

- Variância (16.13).^{109,179}

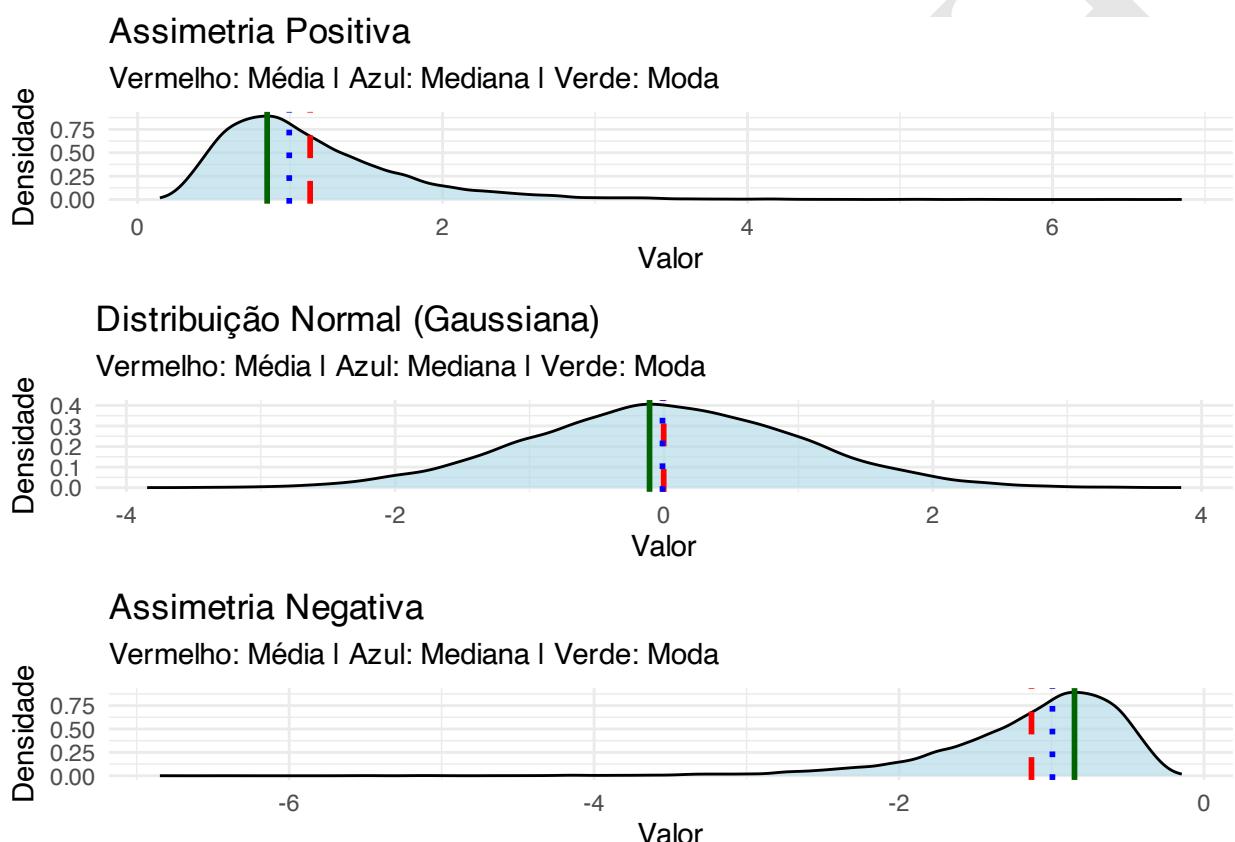


Figura 16.14: Parâmetros de tendência central em distribuições assimétricas e normais.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (16.13)$$

- Desvio-padrão (16.14).^{180,182,186}

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (16.14)$$

- Amplitude (16.15).^{109,179,186}

$$A = x_{\max} - x_{\min} \quad (16.15)$$

- Intervalo interquartil (16.16).^{109,179,186}

$$IQR = Q_3 - Q_1 \quad (16.16)$$

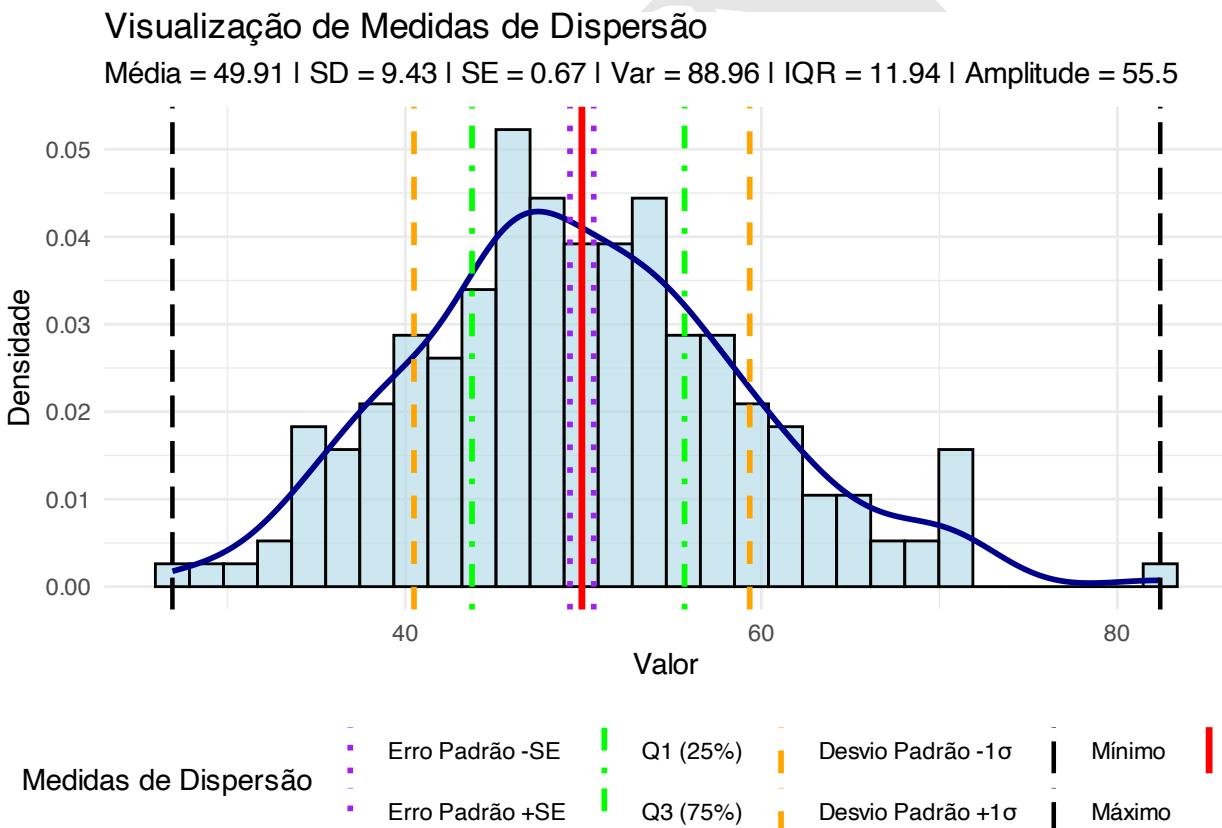


Figura 16.15: Parâmetros de dispersão em distribuições normais.

O pacote *base*⁵⁵ fornece a função *summary*^a para calcular diversos parâmetros descritivos.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/summary>



O pacote *stats*¹³¹ fornece a função *confint*^a para calcular o intervalo de confiança em um nível de significância α .

^a<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/confint>

16.6.2 Como escolher o parâmetro de dispersão?

- Desvio-padrão é apropriado quando a média é utilizada como parâmetro de tendência central em distribuições simétricas.¹⁸⁶
- Amplitude ou intervalo interquartil são apropriadas para variáveis ordinais ou distribuições assimétricas.¹⁸⁶

16.6.3 O que é a correção de Bessel para variância?

- Correção de Bessel é um ajuste feito no denominador da fórmula de variância da amostra — ou seja, o número de graus de liberdade — para evitar que a variância amostral seja menor do que a variância populacional.¹⁸⁷
- A correção de Bessel é feita subtraindo-se 1 do número de observações da amostra, ou seja, $n - 1$ (16.17).¹⁸⁷

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (16.17)$$

16.6.4 Por que a correção de Bessel para variância é importante?

- A correção de Bessel é importante porque a variância amostral tende a ser menor do que a variância populacional, especialmente em amostras pequenas.¹⁸⁷
- A correção de Bessel ajuda a garantir que a variância amostral seja uma estimativa mais precisa da variância populacional, o que é fundamental para a validade dos testes estatísticos e das inferências feitas a partir da amostra.¹⁸⁷

16.7 Proporção

16.7.1 Que parâmetros de proporção podem ser estimados?

- Frequência absoluta (16.18).^{109,179,181}

$$f_i = n_i \quad (16.18)$$

- Frequência relativa (16.19).^{109,179,181}

$$fr_i = \frac{n_i}{N} \quad (16.19)$$

- Percentil (16.20), onde k é o percentil desejado (0 a 100) e n é o número total de observações na amostra.^{109,179,181}

$$P_k = x_{(\frac{k}{100} \cdot (n+1))} \quad (16.20)$$

- Quantil: é o ponto de corte que define a divisão da amostra em grupos de tamanhos iguais. Portanto, não se referem aos grupos em si, mas aos valores que os dividem.¹⁸¹

- Tercil: 2 valores que dividem a amostra em 3 grupos de tamanhos iguais.¹⁸¹

- Quartil: 3 valores que dividem a amostra em 4 grupos de tamanhos iguais.¹⁸¹
- Quintil: 4 valores que dividem a amostra em 5 grupos de tamanhos iguais.¹⁸¹
- Decil: 9 valores que dividem a amostra em 10 grupos de tamanhos iguais.¹⁸¹

R

O pacote *base*⁵⁵ fornece a função *summary*^a para calcular diversos parâmetros descritivos.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/summary>

R

O pacote *base*⁵⁵ fornece a função *table*^a para calcular proporções.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/table>

R

O pacote *stats*⁵⁵ fornece a função *quantile*^a para executar análise de percentis.

^a<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/quantile>

16.8 Extremos

16.8.1 O que são valores extremos?

- Valores extremos podem constituir valores legítimos ou ilegítimos de uma distribuição.¹⁸⁸

16.8.2 Que parâmetros extremos podem ser estimados?

- Mínimo (16.21).¹⁰⁹

$$\text{Mínimo} = \min(x_1, x_2, \dots, x_n) \quad (16.21)$$

- Máximo (16.26).¹⁰⁹

$$\text{Máximo} = \max(x_1, x_2, \dots, x_n) \quad (16.22)$$

16.9 Distribuição

16.9.1 Que parâmetros de distribuição podem ser estimados?

- Assimetria (16.23).¹⁷⁹

$$\gamma_1 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}} \quad (16.23)$$

- Curtose (16.24).¹⁷⁹

$$\gamma_2 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2} \quad (16.24)$$

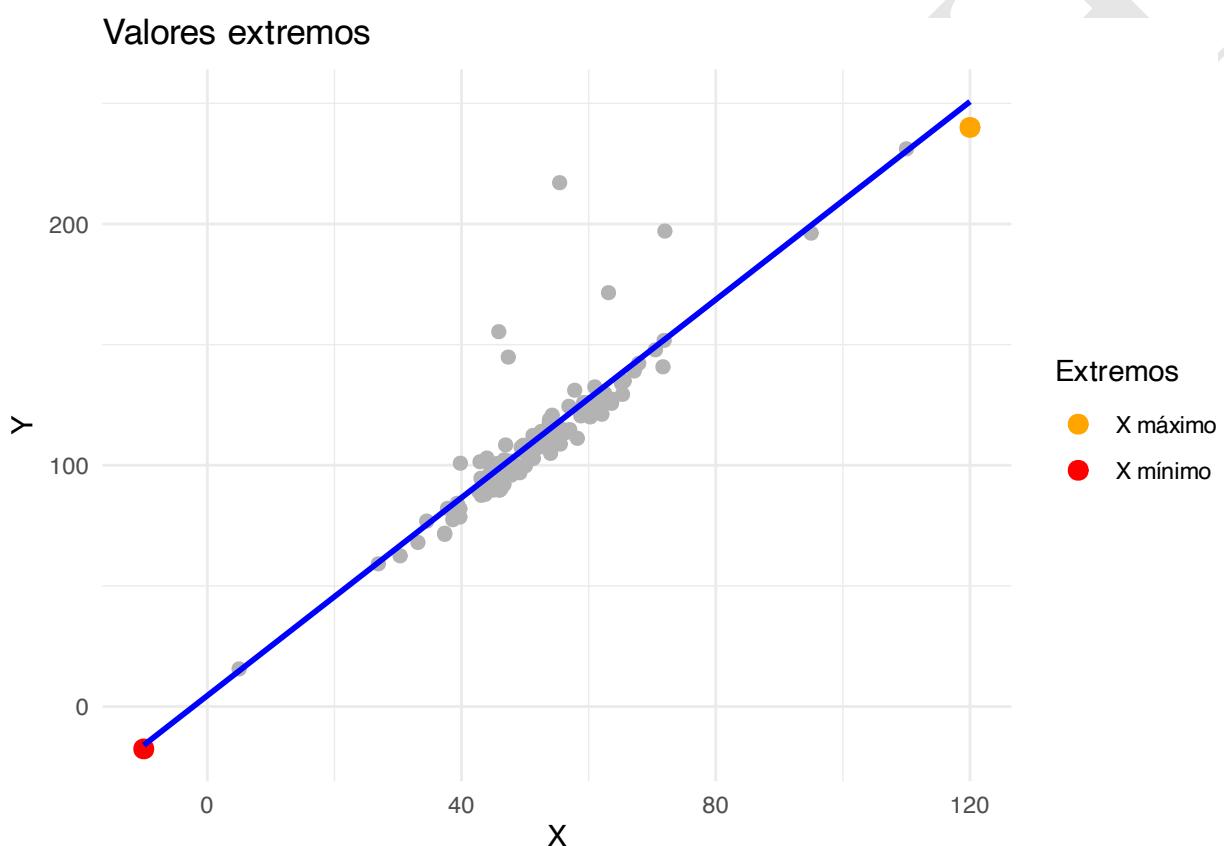


Figura 16.16: Regressão linear com valores extremos.

- Excesso de Curtose (16.25).¹⁷⁹

$$\kappa = \gamma_2 - 3$$

(16.25)

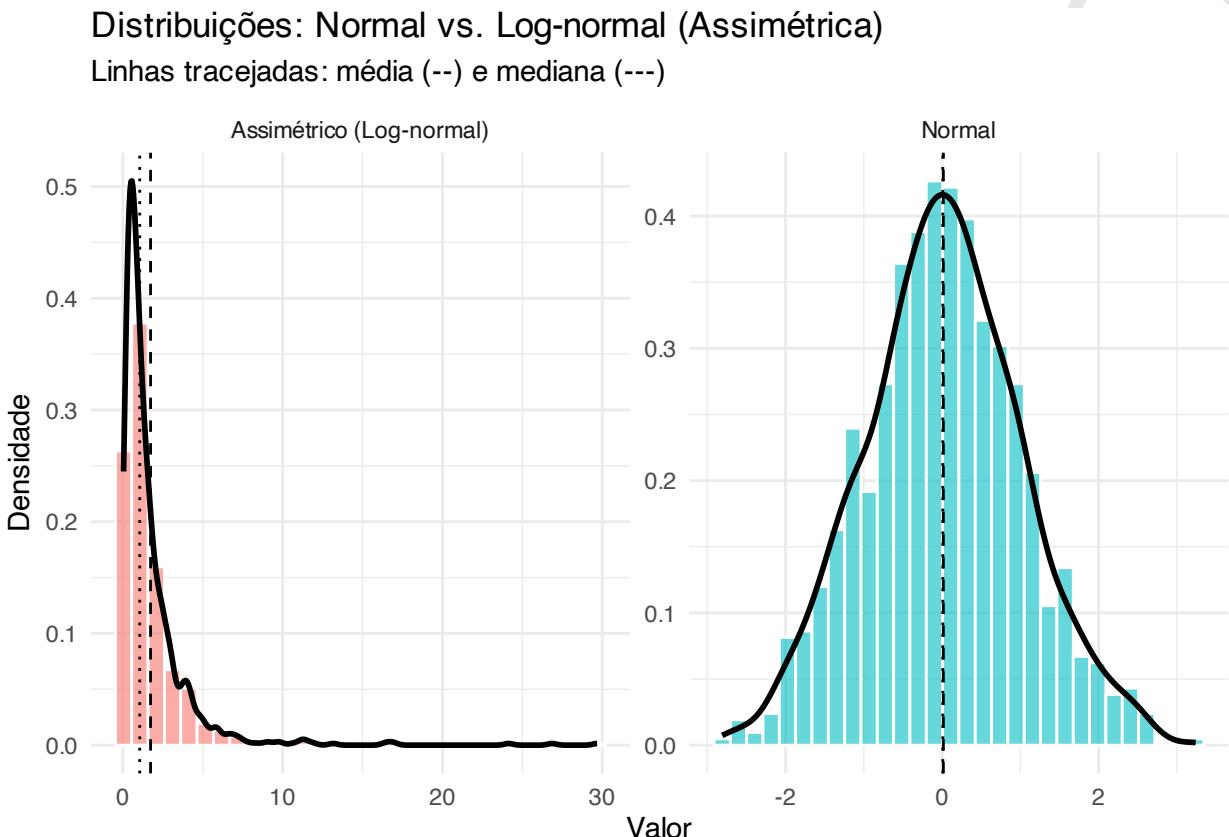


Figura 16.17: Parâmetros de distribuição: Assimetria e Curtose.

16.10 Robustez em medidas de localização

16.10.1 O que é ponto de quebra?

- É a menor proporção de contaminação que pode levar o estimador a resultados arbitrariamente errados; quanto maior, mais robusto.¹⁸⁹

16.10.2 Por que a média não é robusta?

- Porque tem ponto de quebra 0 e função influência não limitada; um único *outlier* pode distorcer a média arbitrariamente.¹⁸⁹

16.10.3 Qual a alternativa robusta para localização?

- Mediana, com 50 de ponto de quebra e função influência limitada.¹⁸⁹

Curtose na prática: Normal vs. t(df=5) vs. Uniforme

Todas padronizadas (média = 0, sd = 1). Curtose difere apenas nas caudas.

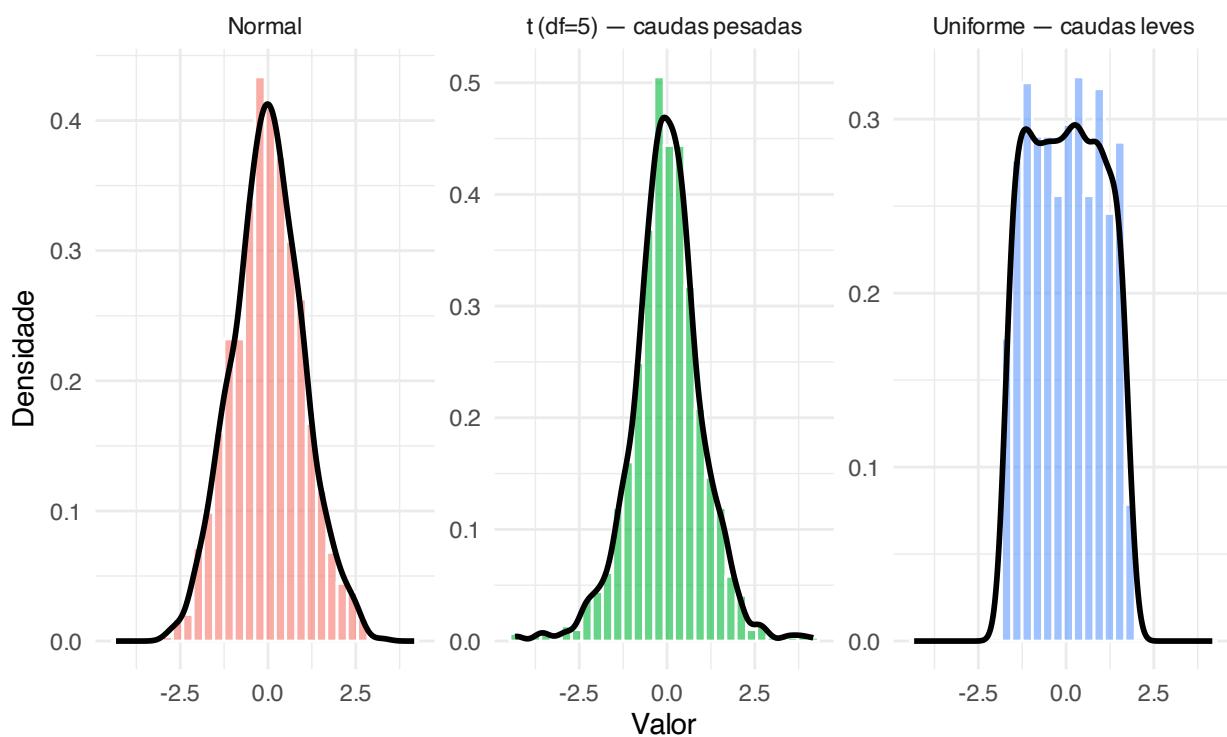


Figura 16.18: Parâmetros de distribuição: Curtose em distribuições simétricas (normal vs. uniforme).

16.10.4 Como estimar escala de forma robusta?

- *Median Absolute Deviation (MAD)* (??), com correção 1,483 para normalidade, com 50 de ponto de quebra.¹⁸⁹

$$MAD = 1.483 \cdot \text{median}(|x_i - \text{median}(x)|) \quad (16.26)$$

- Primeiro quartil das diferenças pareadas (Q_n) (16.27), com 50 de ponto de quebra.¹⁸⁹

$$Q_n = 2.2219 \cdot \text{first quartile}(|x_i - x_j|; i < j) \quad (16.27)$$

- O intervalo interquartil (IQR) (16.16) é robusto, com ponto de quebra 25, sendo simples de interpretar e útil em boxplots.¹⁸⁹

16.11 Parâmetros robustos

16.11.1 O que são parâmetros robustos?

- Parâmetros robustos são medidas de posição e dispersão que permanecem estáveis mesmo na presença de valores discrepantes.¹⁹⁰

16.11.2 Que parâmetros robustos podem ser estimados?

- Mediana em vez da média aritmética, pois é menos sensível a valores extremos.¹⁹⁰
- *Median Absolute Deviation (MAD)* em vez do desvio-padrão σ , que pode ser escalonado por 1,483 para comparabilidade.¹⁹⁰
- Q_n e S_n como estimadores alternativos de dispersão robusta.¹⁹⁰
- Média e variância Winsorizadas como opções intermediárias, reduzindo a influência dos *outliers*.¹⁹⁰

16.11.3 Por que utilizar parâmetros robustos?

- Parâmetros robustos garantem maior confiabilidade quando os dados não seguem a normalidade ou apresentam contaminação por *outliers*.¹⁹⁰
- Parâmetros robustos permitem análises mais estáveis em estudos exploratórios, evitando decisões equivocadas sobre variabilidade ou tendência central.¹⁹⁰

RASCUNHO

Capítulo 17

Análise inicial de dados

17.1 Análise inicial de dados

17.1.1 O que é análise inicial de dados?

- Análise inicial de dados¹⁹¹ é uma sequência de procedimentos que visam principalmente a transparência e integridade das pré-condições do estudo para conduzir a análise estatística apropriada de modo responsável para responder aos problemas da pesquisa.¹³⁴
- O objetivo da análise inicial de dados é propiciar dados prontos para análise estatística, incluindo informações confiáveis sobre as propriedades dos dados.¹³⁴
- A análise inicial de dados pode ser dividida nas seguintes etapas:¹³⁴
 - Configuração dos metadados
 - Limpeza dos dados
 - Verificação dos dados
 - Relatório inicial dos dados
 - Refinamento e atualização do plano de análise estatística
 - Documentação e relatório da análise inicial de dados
- A análise inicial de dados não deve ser confundida com análise exploratória,¹⁹² nem deve ser utilizada para hipotetizar após os dados serem coletados (conhecido como *Hypothesizing After Results are Known*, HARKing).⁹²

17.1.2 Como conduzir uma análise inicial de dados?

- Desenvolva um plano de análise inicial de dados consistente com os objetivos da pesquisa. Por exemplo, verifique a distribuição e escala das variáveis, procure por observações não-usuais ou improváveis, avalie possíveis padrões de dados perdidos.¹³⁴
- Não altere diretamente os dados de uma tabela obtida de uma fonte. Use scripts para implementar eventuais alterações, de modo a manter o registro de todas as modificações realizadas no banco de dados.¹³⁴
- Use os metadados do estudo para guiar a análise inicial dos dados e compartilhe com os dados para maior transparência e reproduzibilidade.¹³⁴

- Representação gráfica dos dados pode ajudar a identificar características e padrões no banco de dados, tais como suposições e tendências.¹³⁴
- Verifique a frequência e proporção de dados perdidos em cada variável, e depois examine por padrões de dados perdidos simultaneamente por duas ou mais variáveis.¹³⁴
- Verifique a frequência e proporção de dados perdidos em cada variável, e depois examine por padrões de dados perdidos simultaneamente por duas ou mais variáveis.¹³⁴
- Exclusão de dados *ad hoc* baseada no desfecho pode influenciar os resultados do estudo, portanto os critérios de exclusão de dados antes da análise estatística (descritiva e/ou inferencial) devem ser reportados.¹⁹³

17.1.3 Quais problemas podem ser detectados na análise inicial de dados?

- Ocorrência de dados perdidos, que podem ser excluídos ou imputados para não reduzir o poder do estudo.[?]



O pacote *stats*¹³¹ fornece a função *na.omit*^a para retornar os dados sem os dados perdidos.

^a<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/na.fail>



O pacote *stats*¹³¹ fornece a função *complete.cases*^a para identificar os casos completos — isto é, sem dados perdidos — em um banco de dados.

^a<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/complete.cases>

- Registros duplicados, que devem ser excluídos para não inflar a amostra.¹⁹⁴



O pacote *base*⁵⁵ fornece a função *duplicated*^a para identificar elementos duplicados de um banco de dados.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/duplicated>

- Codificação 0 ou 1 para variáveis dicotômicas para representar a direção esperada da associação entre elas.¹⁹⁴
- Ordenação cronológica de variáveis com registros temporais (retrospectivos ou prospectivos).¹⁹⁴
- A distribuição das variáveis para verificação das suposições das análises planejadas.¹⁹⁴
- Ocorrência de efeitos teto e piso nas variáveis.¹⁹⁴

Capítulo 18

Análise exploratória de dados

18.1 Análise exploratória de dados

18.1.1 O que é análise exploratória de dados?

- Análise exploratória de dados consiste em um processo iterativo de elaboração e interpretação da síntese de dados, tabelas e gráficos, considerando os aspectos teóricos do estudo.¹⁹²
- Análise exploratória deve ser separada da análise inferencial de testes de hipóteses; a decisão sobre os modelos a testar deve ser feita *a priori*.¹⁹³

18.1.2 Quais são os objetivos centrais da análise exploratória de dados?

- A análise exploratória de dados (EDA) tem dois objetivos principais: (a) descrição dos dados e (b) formulação de modelos. A descrição envolve resumir os dados e destacar características essenciais, enquanto a formulação de modelos auxilia na geração de hipóteses e na escolha de procedimentos estatísticos adequados.¹⁹⁴

18.1.3 Por que conduzir a análise exploratória de dados?

- A condução de análise exploratória de dados pode ajudar a identificar padrões e pode orientar trabalhos futuros, mas os resultados não devem ser interpretados como inferências sobre uma população.¹⁹⁵
- A análise exploratória não deve ser usada para definir as questões e hipóteses científicas do estudo.¹⁹⁶

R

O pacote *explore*¹⁹⁶ fornece a função *explore*^a para análise exploratória de um banco de dados.

^a<https://www.rdocumentation.org/packages/explore/versions/1.0.2/topics/explore>

R

O pacote *dataMaid*¹⁹⁷ fornece a função *makeDataReport*^a para criar um relatório de análise exploratória de um banco de dados.

^a<https://www.rdocumentation.org/packages/dataMaid/versions/1.4.1/topics/makeDataReport>

R

O pacote *DataExplorer*¹⁹⁸ fornece a função *create_report*^a para criar um relatório de análise exploratória de um banco de dados.

^ahttps://www.rdocumentation.org/packages/DataExplorer/versions/0.8.2/topics/create_report

R

O pacote *SmartEDA*¹⁹⁹ fornece a função *ExpReport*^a para criar um relatório de análise exploratória de um banco de dados.

^a<https://www.rdocumentation.org/packages/SmartEDA/versions/0.3.9/topics/ExpReport>

R

O pacote *gtExtras*²⁰⁰ fornece a função *gt_plt_summary*^a para criar uma tabela descritiva síntese com histogramas ou gráficos de barra a partir de um banco de dados.

^ahttps://www.rdocumentation.org/packages/gtExtras/versions/0.5.0/topics/gt_plt_summary

R

O pacote *radiant*²⁰¹ fornece a função *radiant*^a para executar uma interface interativa para análise exploratória de dados.

^a<https://www.rdocumentation.org/packages/radiant/versions/1.5.0/topics/radiant>

18.2 Ingredientes da análise exploratória de dados

18.2.1 Quais são os principais elementos que compõem a análise exploratória de dados?

- Verificação da qualidade dos dados (erros, ausências, *outliers*), o cálculo de estatísticas descritivas (média, desvio-padrão, intervalos, correlações) e o uso de representações gráficas como histogramas, diagramas de dispersão, *boxplots* e gráficos de séries temporais.¹⁹¹
- Técnicas multivariadas exploratórias, como análise de componentes principais e análise de clusters, podem revelar padrões em dados complexos.¹⁹¹

18.2.2 Quais etapas constituem a análise exploratória de dados?

- Cada combinação de problema de pesquisa e delineamento de estudo pode demandar um plano de análise exploratório distinto.¹⁹⁵
- Verifique a existência e/ou influência de valores discrepantes (“fora da curva” ou *outliers*):^{191,192,195}
 - Boxplots
 - Gráficos quantil-quantil (Q-Q)
- A análise exploratória valoriza o uso de gráficos interativos e técnicas de *brushing* e *linking*, que permitem explorar padrões ocultos, relacionar múltiplas variáveis e destacar subconjuntos de observações.²⁰²

R

O pacote *ggplot2*¹⁷³ fornece a função *geom_boxplot*^a para construção de gráficos *boxplot*.

^ahttps://www.rdocumentation.org/packages/ggplot2/versions/3.5.2/topics/geom_boxplot

- Verifique a homocedasticidade (homogeneidade da variância).¹⁹⁵

- Boxplots condicionais (por fator de análise)
- Análise dos resíduos do modelo de regressão
- Gráfico resíduos vs. valores ajustados
- Verifique a normalidade da distribuição dos dados:^{191,195}
 - Histograma das variáveis (por fator de análise)
 - Histograma dos resíduos da regressão
- Verifique a existência de grande quantidade de valores nulos (=0).¹⁹⁵
 - Histograma das variáveis (por fator de análise)
- Verifique a existência de colinearidade entre variáveis independentes de um modelo de regressão:¹⁹⁵
 - Fator de inflação de variância (*variance inflation factor*, VIF)
 - Coeficiente de correlação de Pearson (r)
 - Gráfico de dispersão entre variáveis
- Verifique possíveis relações entre as variáveis dependente(s) e independente(s) de um modelo de regressão:¹⁹⁵
 - Gráfico de dispersão entre variável independente e dependente
- Verifique possíveis interações entre as variáveis dependente(s) de um modelo de regressão:¹⁹⁵
 - Gráfico *coplot* de dispersão entre variáveis dependentes



O pacote *ggcleveland*²⁰³ fornece a função *gg_coplot*^a para construção de gráficos *boxplot* condicionais.

^ahttps://www.rdocumentation.org/packages/ggcleveland/versions/0.1.0/topics/gg_coplot

- Verifique por dependência entre variáveis de um modelo de regressão:¹⁹⁵
 - Gráfico de série temporal das variáveis
 - Gráfico de autocorrelação entre as variáveis
- Medidas como mediana, *trimean*, distância absoluta mediana e procedimentos de *winsorizing* ou *trimming* são preferidos, pois reduzem a influência de valores extremos e oferecem resumos mais fiéis.²⁰²
- A análise exploratória adota o esquema `dados = ajuste + resíduo`, no qual o analista ajusta modelos provisórios, examina resíduos e refina os modelos em ciclos sucessivos de aproximação.²⁰²
- Valores discrepantes (*outliers*) não devem ser ignorados; eles podem indicar erros de coleta ou fenômenos relevantes. *Fringeliers*, casos menos extremos mas recorrentes, também merecem atenção.²⁰²
- Transformar variáveis em novas formas (por exemplo, log ou inverso) pode revelar simetrias ocultas e tornar relações mais claras e lineares.²⁰²

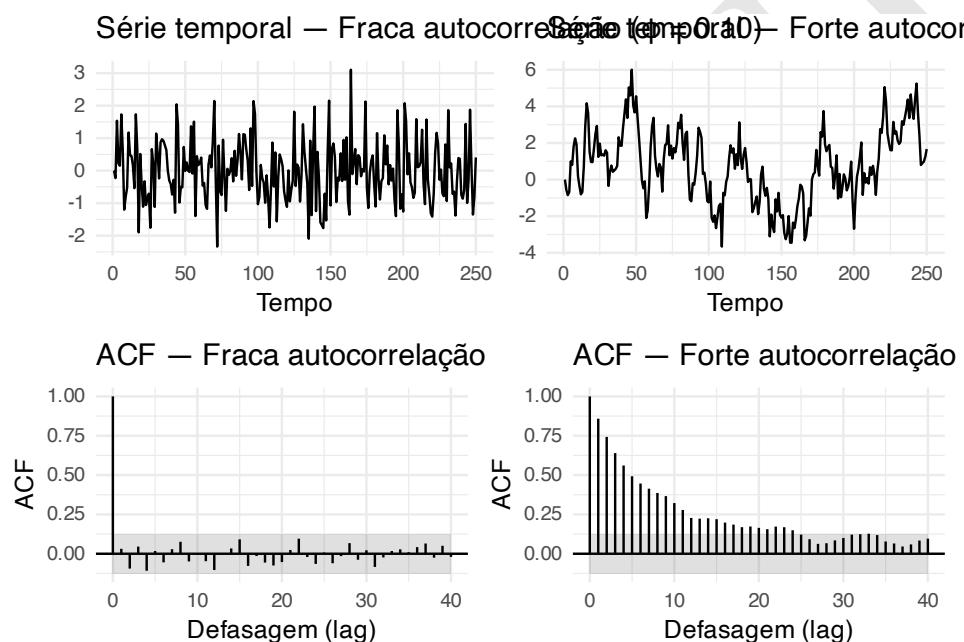


Figura 18.1: Séries temporais e autocorrelação de duas séries simuladas com fraca e forte autocorrelação.

Capítulo 19

Análise descritiva

19.1 Análise descritiva

19.1.1 O que é análise descritiva?

- Análise descritiva é usada para compreendermos algum aspecto de um conjunto de dados, respondendo a perguntas do tipo “quando?”, “onde?”, “quem?”, “o quê?”, “como?” e “e daí?”.^{108,204}



O pacote *base*⁵⁵ fornece a função *summary*^a para calcular diversos parâmetros descritivos.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/summary>

19.1.2 Como apresentar os resultados descritivos?

- Variáveis categóricas: Reporte valores de frequência absoluta e relativa (n, percentual).²⁰⁵
- Organização das tabelas: as variáveis são exibidas em linhas e os grupos são exibidos em colunas.²⁰⁵
- Calcule percentagens para as colunas (isto é, entre grupos) e não entre linhas.²⁰⁵
- Em caso de dados perdidos, não inclua uma linha com total de dados perdidos, pois distorce as proporções entre colunas e as análises de tabela de contingência. Indique no texto ou em uma coluna separada o total de dados perdidos por variável.²⁰⁵

19.2 Apresentação de resultados numéricos

19.2.1 O que são casas decimais?

- O número de casas decimais refere-se à quantidade de dígitos que aparecem após a vírgula decimal.^{206,207}
- Para tamanhos de efeito: use 2–3 dígitos significativos.²⁰⁸
- Para medidas de variabilidade (desvio-padrão/erro-padrão/intervalo de confiança): use 1–2 dígitos significativos.²⁰⁸

19.2.2 O que são dígitos significativos?

- O termo “dígitos significativos” é preferido a “algarismos significativos” ou “dígitos efetivos” e não se relaciona com significância estatística.^{206,207}

Tabela 19.1: Quantidade de casas decimais e dígitos significativos.

Valor	Casas Decimais	Dígitos Significativos
0,00789	5	0
0,0456	4	0
45,6	1	2
123,456	3	3
7890,0000	4	4

Tabela 19.2: Valores originais, arredondamentos e erros de arredondamento por casas decimais.

Valor	Casas Decimais	Dígitos Significativos	2 Casas decimais [Margem de erro]	1 Casa decimal [Margem de erro]	Sem casa decimal [Margem de erro]
0,00789	5	0	0,01 [0,005, 0,015]	0,0 [-0,05, 0,05]	0 [-0,5, 0,5]
0,0456	4	0	0,05 [0,045, 0,055]	0,0 [-0,05, 0,05]	0 [-0,5, 0,5]
45,6	1	2	45,60 [45,595, 45,605]	45,6 [45,55, 45,65]	46 [45,5, 46,5]
123,456	3	3	123,46 [123,455, 123,465]	123,5 [123,45, 123,55]	123 [122,5, 123,5]
7890,0000	4	4	7890,00 [7889,995, 7890,005]	7890,0 [7889,95, 7890,05]	7890 [7889,5, 7890,5]

- O número de dígitos significativos é a soma total de dígitos, desconsiderando a vírgula decimal e os zeros à esquerda; os zeros à direita são considerados informativos, salvo exceções.^{206,207}

19.2.3 Como arredondar dados numéricos?

- Apresentar dados com quantidade excessiva de casas decimais pode dificultar a interpretação e induzir erroneamente uma precisão espúria.^{206,207}
- A precisão é determinada pelo grau de arredondamento aplicado, medido em casas decimais ou dígitos significativos.^{206,207}
- O arredondamento também introduz erros, uma vez que aumenta a imprecisão (isto é, incerteza) em torno do valor original.^{206,207}
- A regra geral é utilizar 2 ou 3 dígitos significativos para tamanhos de efeito e 1 ou 2 dígitos significativos para medidas de variabilidade.²⁰⁷
- Regra dos 3 dígitos significativos para proporção de risco: em média, o erro de arredondamento é menor que os 0,5% exigidos, de modo que três dígitos significativos são mais precisos do que o necessário.²⁰⁶
- Regra dos 4 dígitos significativos para proporção de risco: divida a proporção de risco por quatro e arredonde para dois dígitos significativos e, em seguida, relate a proporção para esse número de casas decimais.²⁰⁶

19.3 Tabelas

19.3.1 Por que usar tabelas?

- Tabelas complementam o texto (e vice-versa), e podem apresentar os dados de modo mais acessível e informativo.²⁰⁹

19.3.2 Que informações incluir nas tabelas?

- Título ou legenda, uma síntese descritiva (geralmente por meio de parâmetros descritivos), intervalos de confiança e/ou P-valores conforme necessário para adequada interpretação.^{209,210}

19.3.3 Quais são os tipos de tabelas?

- Tabela de frequência: apresenta a quantidade de ocorrências (frequência absoluta e relativa) de cada categoria de uma variável; usada com variáveis qualitativas ou quantitativas discretas.²¹¹
- Tabela de frequência agrupada: organiza dados contínuos em intervalos de classe (ex: faixas etárias) e mostra as frequências correspondentes.²¹²
- Tabela de contingência (ou tabela cruzada): cruza duas variáveis categóricas, permitindo observar possíveis associações entre elas.²¹³
- Tabela de medidas descritivas: resume variáveis quantitativas com estatísticas como média, mediana, desvio-padrão, mínimo, máximo e quartis.²¹⁴
- Tabela de comparação entre grupos: apresenta médias, desvios-padrão e ocasionalmente resultados de testes de inferência estatística para comparar dois ou mais grupos.²¹⁵
- Tabela de resultados de testes estatísticos: exibe valores de estatísticas de teste, P valores e intervalos de confiança; usada para mostrar inferências.²¹⁶
- Tabela de regressão (ou de modelos estatísticos): mostra os coeficientes de regressão, erros padrão, intervalos de confiança e P valores para cada variável de um modelo.²¹⁷
- Tabela de séries temporais ou longitudinais: organiza dados medidos em diferentes momentos no tempo, permitindo visualizar tendências ou variações longitudinais.²¹⁸



O pacote *gtsummary*²¹¹ fornece a função *tbl_summary*^a para construção da ‘Tabela 1’ com dados descritivos.

^ahttps://search.r-project.org/CRAN/refmans/gtsummary/html/tbl_summary.html



O pacote *table1*²¹² fornece a função *table1*^a para construção de tabelas.

^a<https://search.r-project.org/CRAN/refmans/table1/html/table1.html>



O pacote *flextable*²¹³ fornece as funções *flextable*^a, *as_flextable*^b e *save_as_docx*^c para criar e salvar tabelas formatadas em DOCX.

^a<https://search.r-project.org/CRAN/refmans/flextable/html/flextable.html>

^bhttps://search.r-project.org/CRAN/refmans/flextable/html/as_flextable.html

^chttps://search.r-project.org/CRAN/refmans/flextable/html/save_as_docx.html



O pacote *rempscy*²¹⁴ fornece a função *nice_table*^a para criar tabelas formatadas.

^ahttps://search.r-project.org/CRAN/refmans/rempscy/html/nice_table.html

19.3.4 Quais são os erros mais comuns de preenchimento de tabelas?

- Erros tipográficos.²¹⁵
- Ausência de rótulos ou unidades nas variáveis.²¹⁵
- Relatar estatísticas incorretamente, tais como rotular variáveis contínuas como porcentagens.²¹⁵
- Estatísticas descritivas de tendência central (ex.: médias) relatadas sem a estatística de dispersão correspondente (ex.: desvio-padrão).²¹⁵
- Desvio-padrão nulo ($\sigma = 0$).²¹⁵
- Valores porcentuais que não correspondem ao numerador dividido pelo denominador.²¹⁵

19.4 Tabela 1

19.4.1 O que é a ‘Tabela 1’?

- A ‘Tabela 1’ descreve as características demográficas, sociais e clínicas da amostra, completa ou agrupada por algum fator, geralmente por meio de parâmetros de tendência central e dispersão.^{216,217}

19.4.2 Qual a utilidade da ‘Tabela 1’?

- Descrever (conhecer) as características da amostra e dos grupos sendo comparados, quando aplicável.²¹⁷
- Verificar aderência ao protocolo do estudo, incluindo critérios de inclusão/exclusão, tamanho da amostra e perdas amostrais.²¹⁷
- Permitir a replicação do estudo.²¹⁷
- Meta-analisar os dados junto a estudos similares.²¹⁷
- Avaliar a generalização (validade externa) das conclusões do estudo.²¹⁷

19.4.3 O que é a falácia da ‘Tabela 1’?

- Falácia da Tabela 1 ocorre pela interpretação errônea dos P-valores na comparação entre grupos, na linha de base, de um ensaio clínico aleatorizado.²¹⁸
- Não interprete P da linha de base em ensaios clínicos como “desequilíbrio” (falácia da Tabela 1). Mantenha P-valor apenas como descritivo (ou omita), enfatizando desenho e aleatorização.²⁰⁸

19.4.4 Como construir a ‘Tabela 1’?

- A Tabela 1 geralmente é utilizada para descrever as características da amostra estudada, possibilitando a análise de ameaças à validade interna e/ou externa ao estudo.^{177,219}



O pacote *table1*²¹² fornece a função *table1*^a para construção de tabelas.

^a<https://search.r-project.org/CRAN/refmans/table1/html/table1.html>

Tabela 19.3: Características da amostra por grupo.

Características	N	Controle N = 103 ¹	Intervenção N = 97 ¹	Valor-p ²
Sexo	200			0.060
F		49 (48%)	59 (61%)	
M		54 (52%)	38 (39%)	
Idade	200			0.8
Média (Desvio Padrão)		61 (12)	60 (12)	
Mediana [Q1, Q3]		61 [53, 69]	60 [53, 69]	
IMC	200			0.2
Média (Desvio Padrão)		26.8 (3.7)	27.5 (4.0)	
Mediana [Q1, Q3]		26.6 [24.5, 29.7]	27.6 [25.6, 29.9]	

¹ n (%)² Teste qui-quadrado de independência; Teste de soma de postos de Wilcoxon

O pacote *gtsummary*²¹¹ fornece a função *tbl_summary*^a para construção da ‘Tabela 1’ com dados descritivos.

^ahttps://search.r-project.org/CRAN/refmans/gtsummary/html/tbl_summary.html

19.5 Tabela 2

19.5.1 Qual a utilidade da ‘Tabela 2’?

- A Tabela 2 mostra associações ajustadas multivariadas com o resultado para variáveis resumidas na Tabela 1.²¹⁶

19.5.2 O que é a falácia da ‘Tabela 2’?

- A Tabela 2 pode induzir ao erro de interpretação pelas estimativas de efeitos para covariáveis do modelo também serem utilizados para controlar a confusão da exposição.^{216,220}
- Ao apresentar estimativas de efeito ajustadas para covariáveis juntamente com a estimativa de efeito ajustada para a exposição primária, a Tabela 2 sugere implicitamente que todas estas estimativas podem ser interpretadas de forma semelhante, se não de forma idêntica, como estimativa do efeito total.^{216,220}
- A falácia da Tabela 2 pode ser evitada limitando-se a tabela a estimativas das medidas primárias do efeito de exposição nos diferentes modelos, com as covariáveis secundárias de “ajuste” relatadas em uma nota de rodapé, juntamente com a forma como foram categorizadas ou modeladas.²¹⁶

19.5.3 Como construir a ‘Tabela 2’?

- A Tabela 2 pode ser utilizada para apresentar estimativas de múltiplos efeitos ajustados de um mesmo modelo estatístico.²¹⁶

Tabela 19.4: Exemplo clássico de apresentação suscetível à Falácia da ‘Tabela 2’.

Características	Sem ajuste			Ajustado		
	OR	95% IC	Valor-p	OR	95% IC	Valor-p
Grupo						
Controle	—	—	—	—	—	—
Intervenção	1.71	0.98, 3.02	0.061	1.70	0.97, 3.03	0.067
Idade						
IMC	—	—	—	1.02	1.00, 1.05	0.087
Abreviações: IC = Intervalo de Confiança, OR = Razão de chances						

Tabela 19.5: Exposição (OR; 95% IC) com e sem ajuste.

Características	Sem ajuste			Ajustado		
	OR	95% IC	Valor-p	OR	95% IC	Valor-p
Grupo						
Controle	—	—	—	—	—	—
Intervenção	1.71	0.98, 3.02	0.061	1.70	0.97, 3.03	0.067

Abreviações: IC = Intervalo de Confiança, OR = Razão de chances

Nota. Modelo ajustado por Idade (contínua) e IMC (contínuo). Covariáveis são usadas apenas para controle de confusão e não devem ser interpretadas como efeitos causais (*Falácia da Tabela 2*).



O pacote *table1*²¹² fornece a função *table1*^a para construção de tabelas.

^a<https://search.r-project.org/CRAN/refmans/table1/html/table1.html>



O pacote *gtsummary*²¹¹ fornece a função *tbl_summary*^a para construção da ‘Tabela 1’ com dados descritivos.

^ahttps://search.r-project.org/CRAN/refmans/gtsummary/html/tbl_summary.html

19.6 Visualização efetiva de dados

19.6.1 Por que começar pela mensagem antes do gráfico?

- A figura deve responder a uma pergunta clara (comparação? tendência? composição?) e isso orienta a escolha do tipo de gráfico, dados e anotações. Esboce a mensagem e a pergunta antes de abrir o software.²²¹

19.6.2 Como escolher a geometria e “mostrar os dados”?

- Prefira geometrias que revelem distribuição/variabilidade (pontos, *boxplots*, violinos) em vez de médias sozinhas. Sempre que possível, exiba os dados brutos (pontos com *jitter*) junto da estatística-resumo.²²¹

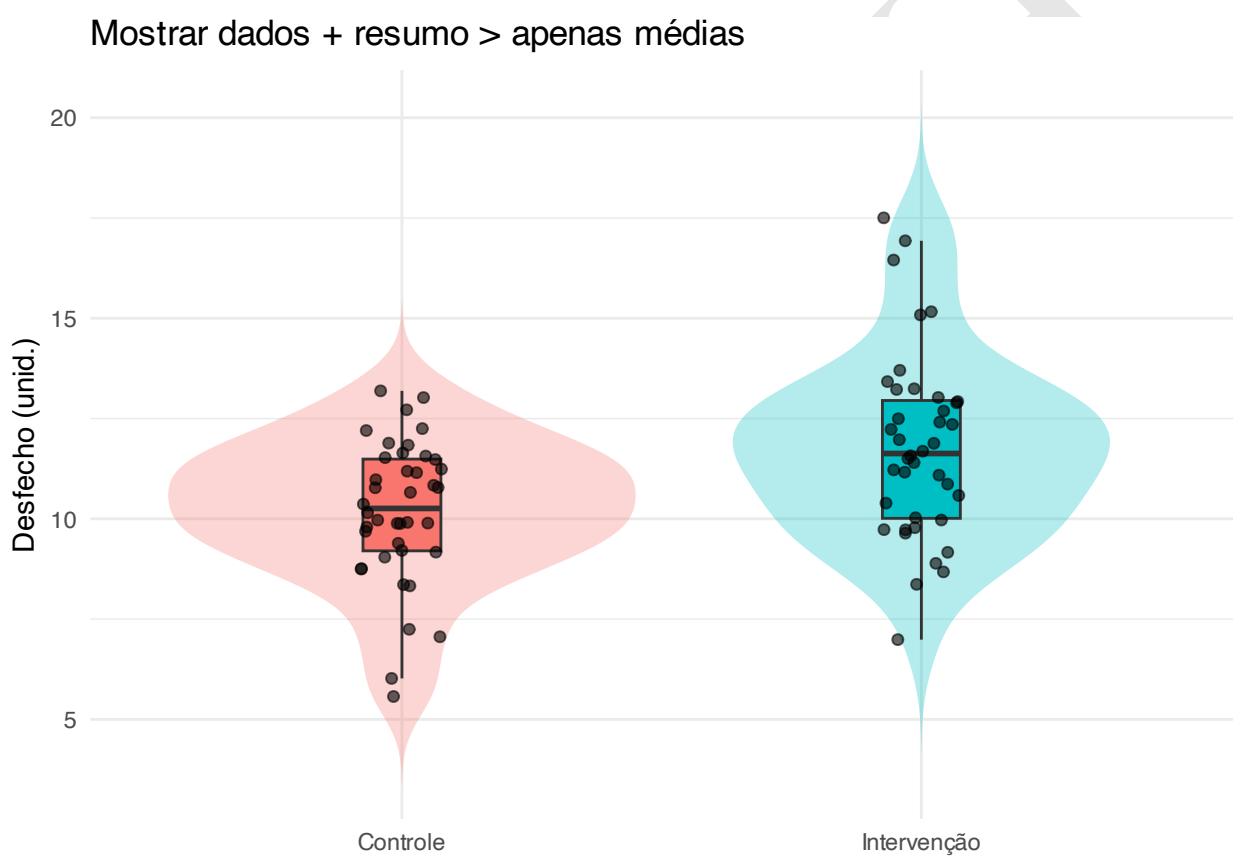


Figura 19.1: Exemplo de gráfico que mostra os dados brutos junto com um resumo estatístico (média e dispersão).

19.7 Gráficos

19.7.1 O que são gráficos?

- Gráficos são utilizados para apresentar dados (geralmente em grande quantidade) de modo mais intuitivo e fácil de compreender.²²²

19.7.2 O que torna um bom gráfico tão poderoso?

- “Não há ferramenta estatística tão poderosa quanto um gráfico bem escolhido”: gráficos ajudam a explorar dados, comunicar resultados e suportar decisões de forma clara e rápida.²²³

19.7.3 Que elementos incluir em gráficos?

- Título, eixos horizontal e vertical com respectivas unidades, escalas em intervalos representativos das variáveis, legenda com símbolos, síntese descritiva dos valores e respectiva margem de erro, conforme necessário para adequada interpretação.²²²

19.7.4 Para que servem as barras de erro em gráficos?

- Barras de erro ajudam ao autor a apresentar as informações que descrevem os dados (por exemplo, em uma análise descritiva) ou sobre as inferências ou conclusões tomadas a partir de dados.^{182,183}
- Barras de erro mais longas representam mais imprecisão (maiores erros), enquanto barras mais curtas representam mais precisão na estimativa.¹⁸³
- Barras de erro descritivas geralmente apresentam a amplitude (mínimo-máximo) ou desvio-padrão.¹⁸³
- Barras de erro inferenciais geralmente apresentam o erro-padrão ou intervalo de confiança no nível de significância α pré-estabelecido.^{182,183}
- Barras de erro com desvio-padrão são úteis para descrever a variabilidade dos dados, enquanto as barras de erro com erro padrão da média são úteis para descrever a precisão do parâmetro estimado (média) e sua relação com o tamanho da amostra.¹⁸²
- Barras de erro com intervalo de confiança são úteis para fornecer uma estimativa da incerteza da estimativa do parâmetro populacional.¹⁸²
- O comprimento das barras de erro sugere graficamente a imprecisão dos dados do estudo, uma vez que o valor verdadeiro da população pode estar em qualquer nível do intervalo da barra.¹⁸³
- De modo contraintuitivo, um espaço entre as barras não garante significância, nem a sobreposição a descarta—depende do tipo de barra.¹⁸²
- Para amostras pequenas é preferível apresentar os dados brutos, uma vez que as barras de erro não serão muito informativas.¹⁸²

R

Os pacotes *ggplot2*¹⁷³, *plotly*²²⁴ e *corrplot*²²⁵ fornecem diversas funções para construção de gráficos tais como *ggplot^a*, *plot_ly^b* e *corrplot^c* respectivamente.

^a<https://www.rdocumentation.org/packages/ggplot2/versions/3.4.3/topics/ggplot>

^bhttps://www.rdocumentation.org/packages/plotly/versions/4.10.2/topics/plot_ly

^c<https://www.rdocumentation.org/packages/corrplot/versions/0.92/topics/corrplot>

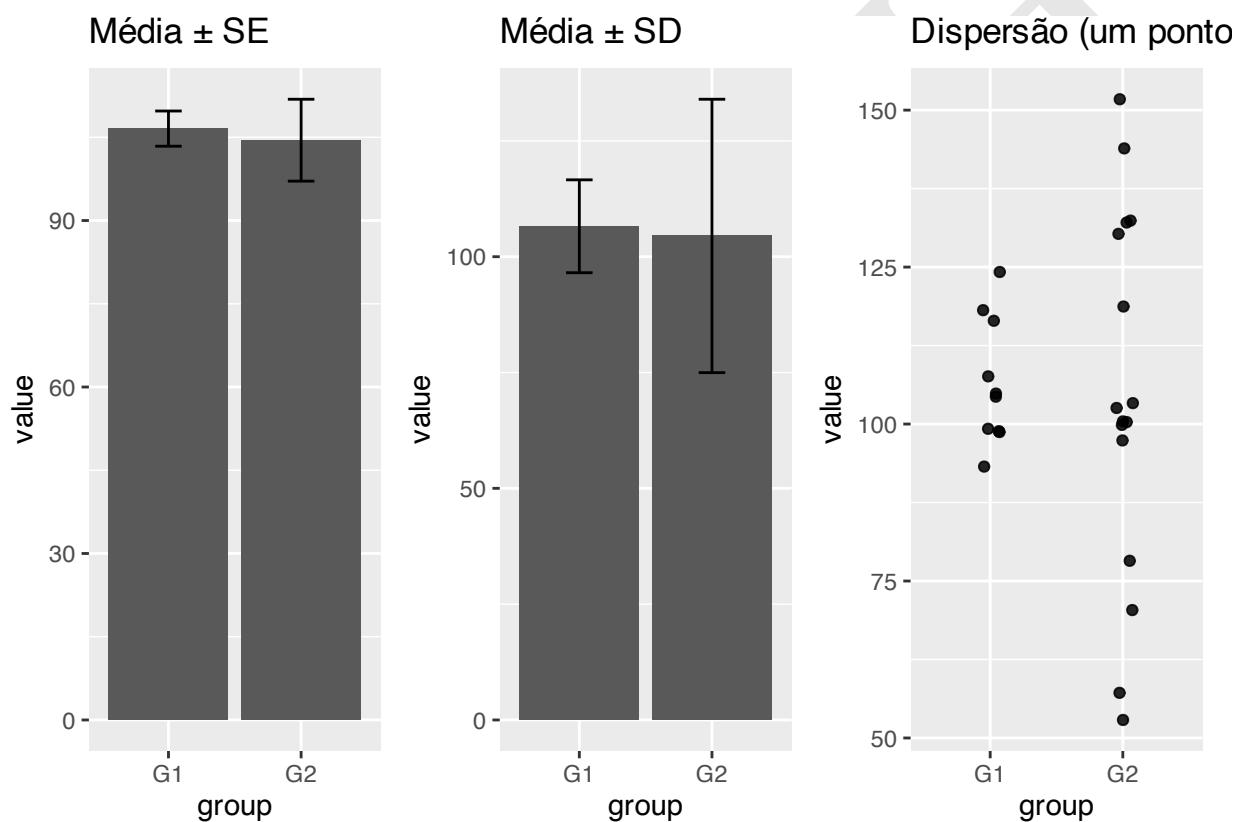


Figura 19.2: Exemplos de gráficos com barras de erro e dados brutos.

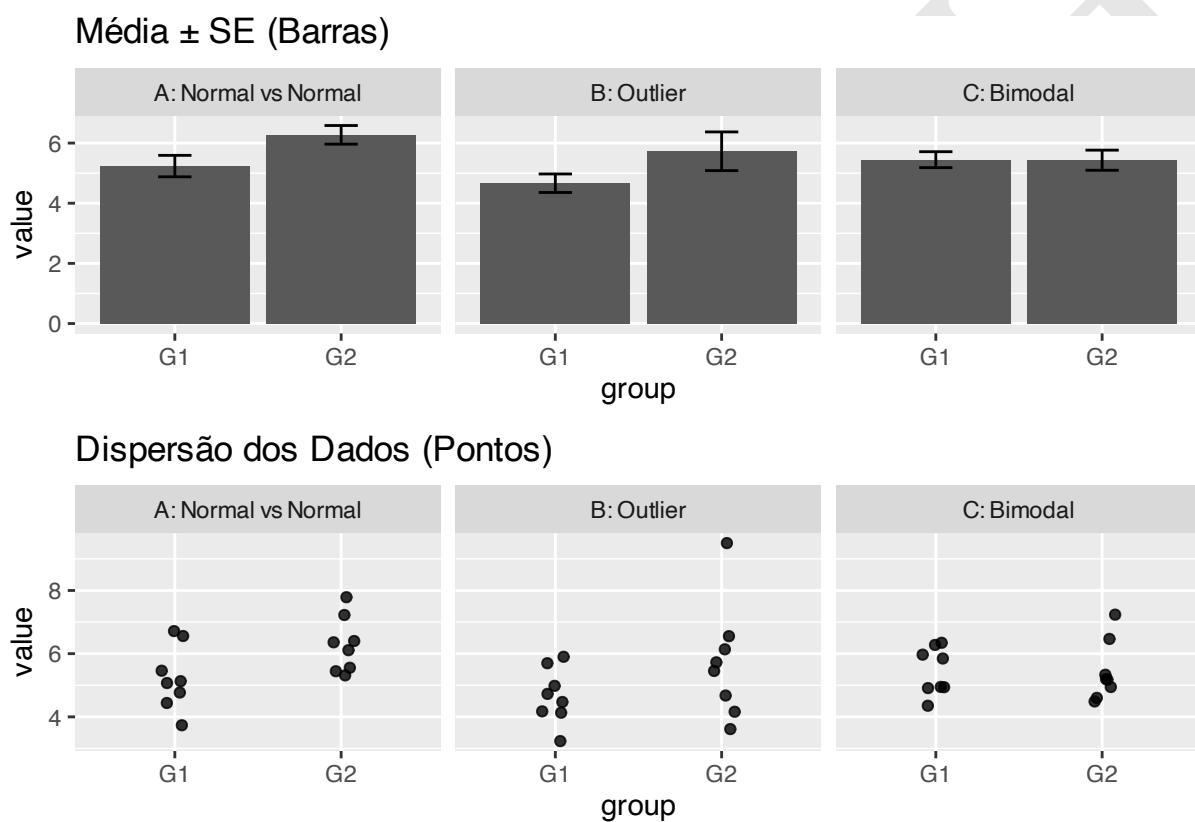


Figura 19.3: Exemplos de gráficos com barras de erro e dados brutos em diferentes cenários.

19.7.5 Quais são os principais obstáculos para bons gráficos?

- Dificuldade técnica, negligência no ensino tradicional e o foco em “beleza” sem clareza podem levar a gráficos ruins, mesmo quando bem intencionados.²²³

19.8 Tipos de gráficos

19.8.1 Quais são os tipos de gráficos para variáveis categóricas?

- Gráfico de barras: Mais usado para comparar frequências absolutas ou relativas entre categorias.²²⁴

Gráfico de Barras Simples

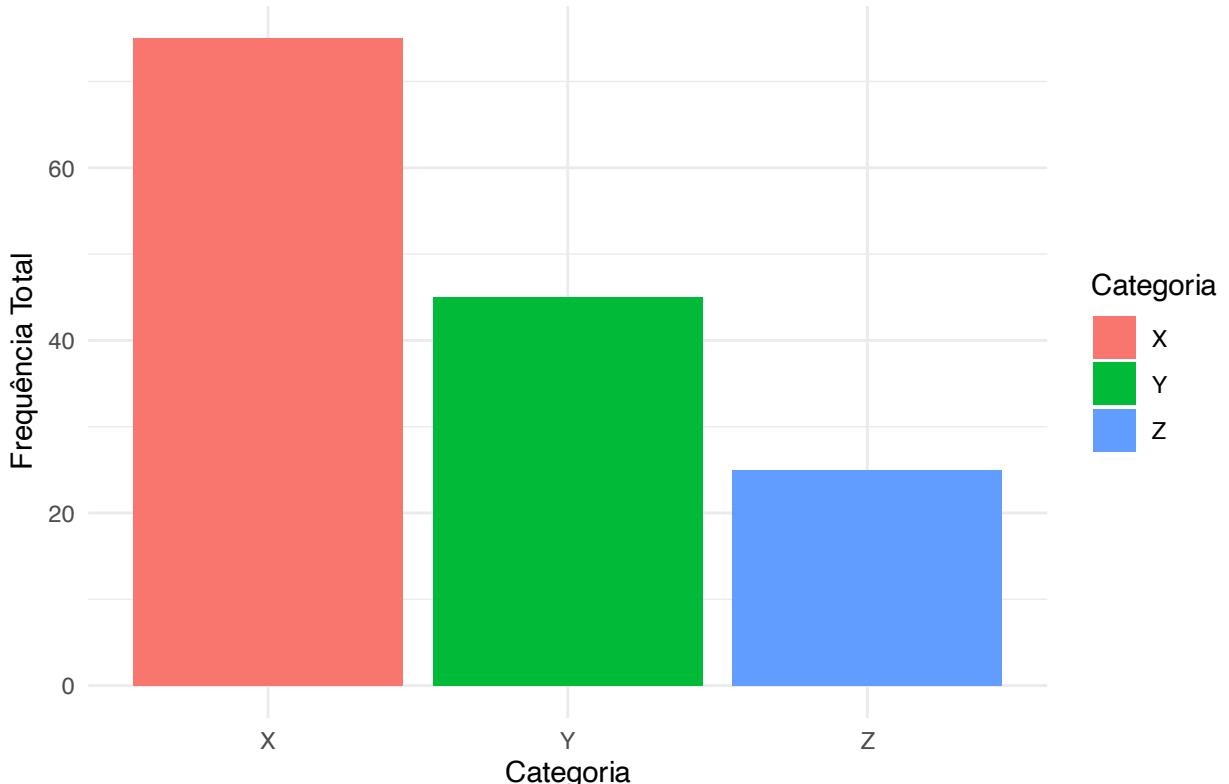


Figura 19.4: Gráfico de barras simples representando frequências por categoria.

- Gráfico de barras empilhadas: Útil para comparar proporções entre grupos em mais de uma variável categórica.²²⁵

19.8.2 Quais são os tipos de gráficos para variáveis numéricas?

- Histograma: Distribuição de frequência de uma variável contínua. Mostra a forma da distribuição (simétrica, assimétrica, bimodal).²²⁶
- Gráfico de densidade: Similar ao histograma, mas mais suave. Útil para avaliar a distribuição.²²⁷
- Diagrama de caixa (*boxplot*): Resume mediana, quartis e valores extremos. Excelente para comparar grupos.²²⁸

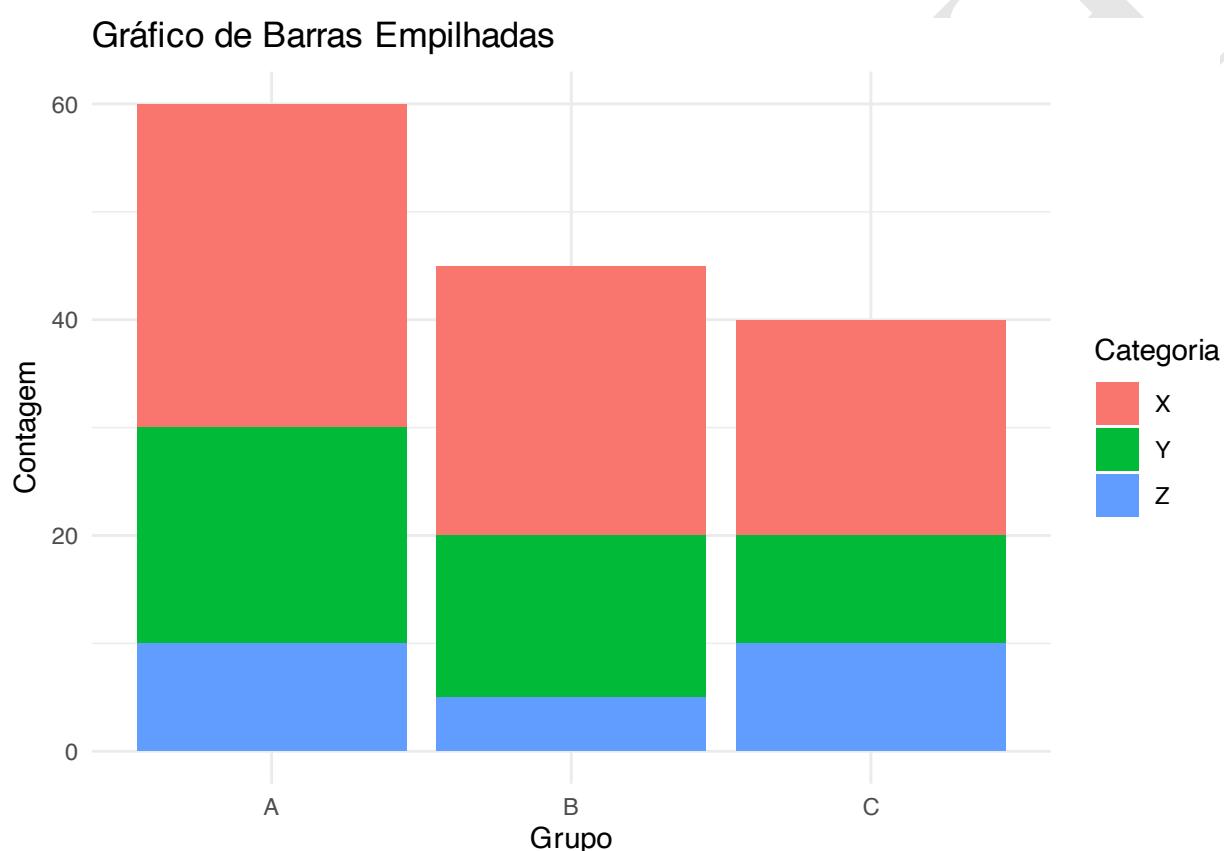


Figura 19.5: Gráfico de barras empilhadas representando frequências por categoria.

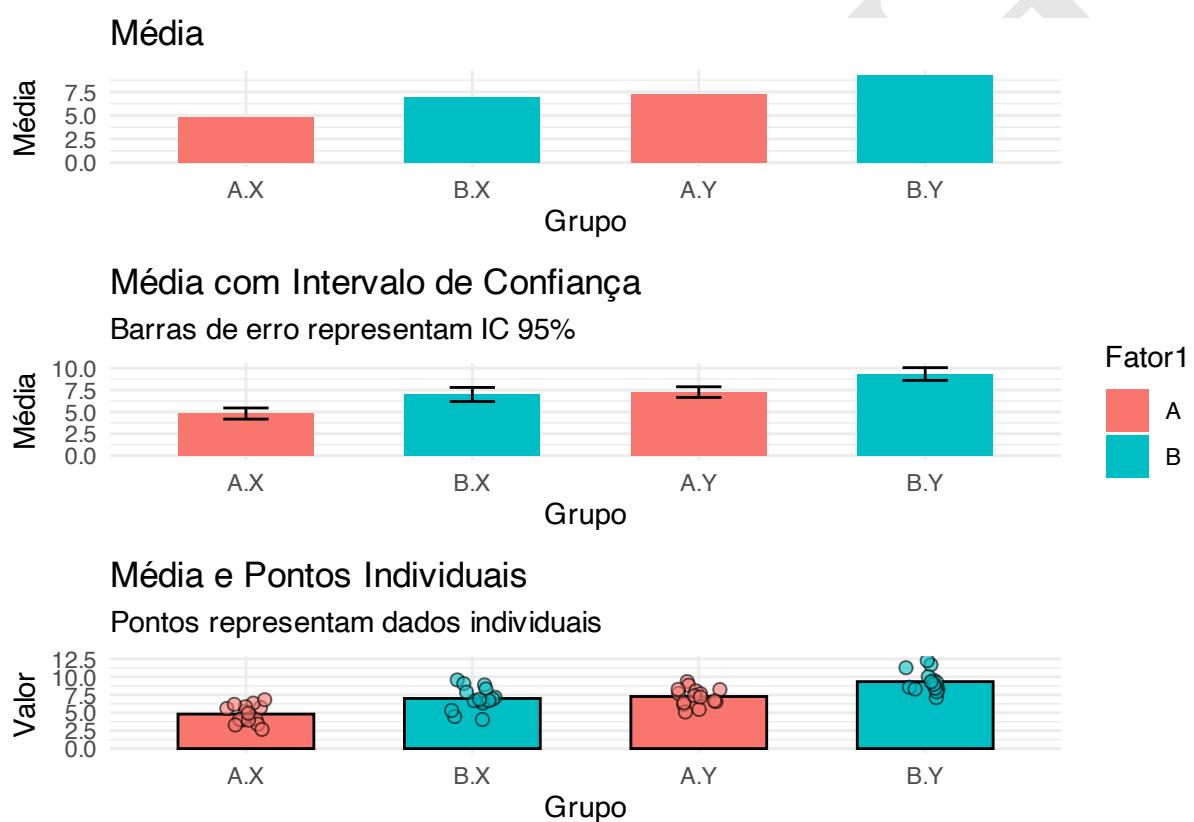


Figura 19.6: Gráficos de barras representando médias, barras de erro e dados individuais.

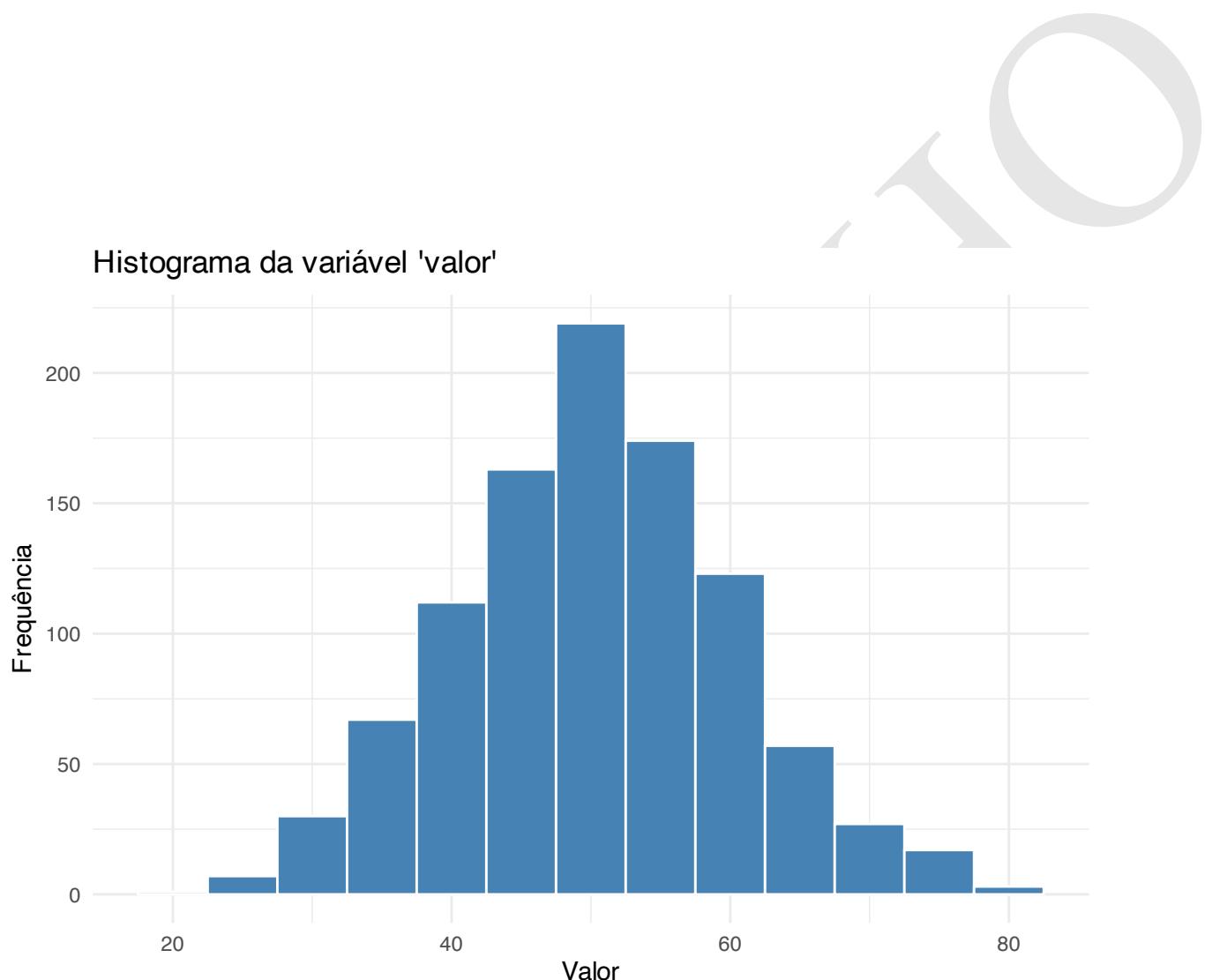


Figura 19.7: Histograma da variável 'valor'.

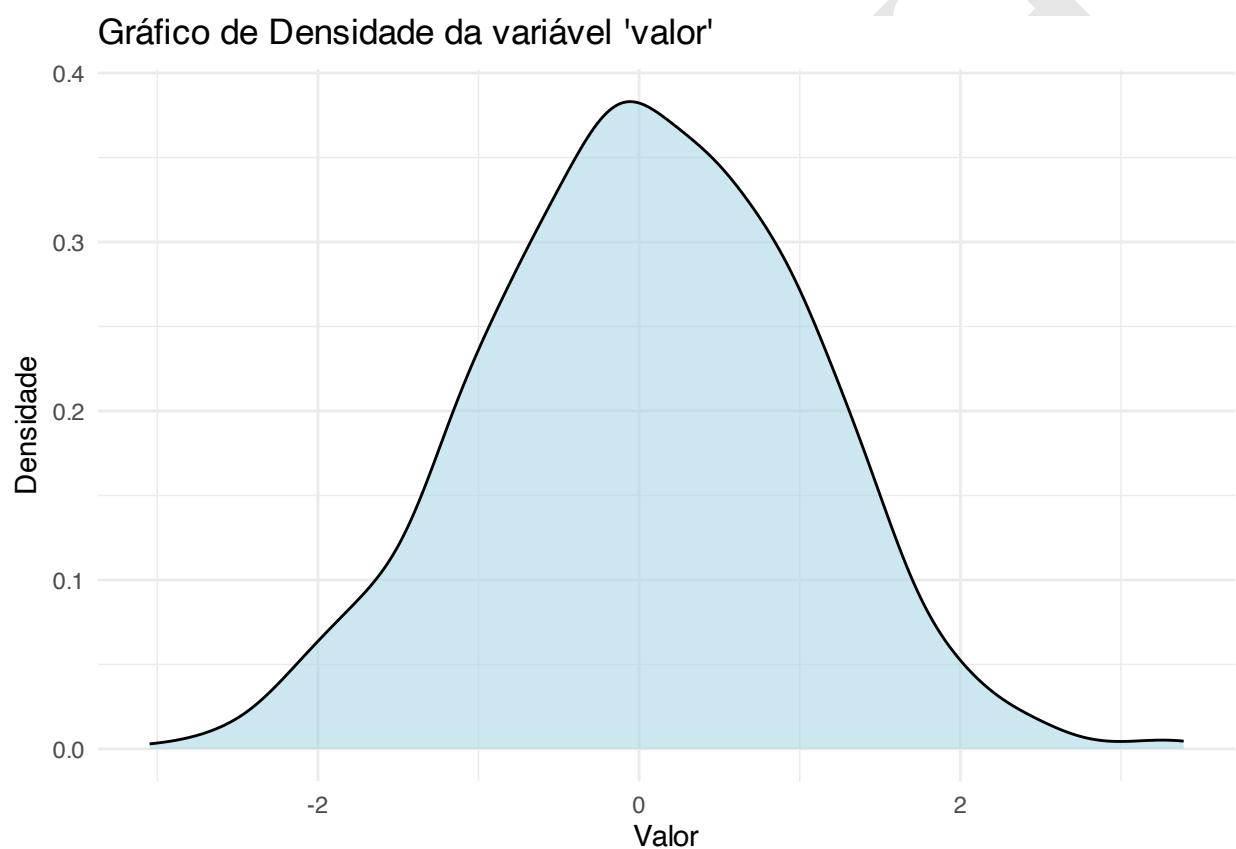


Figura 19.8: Gráfico de densidade da variável 'valor'.

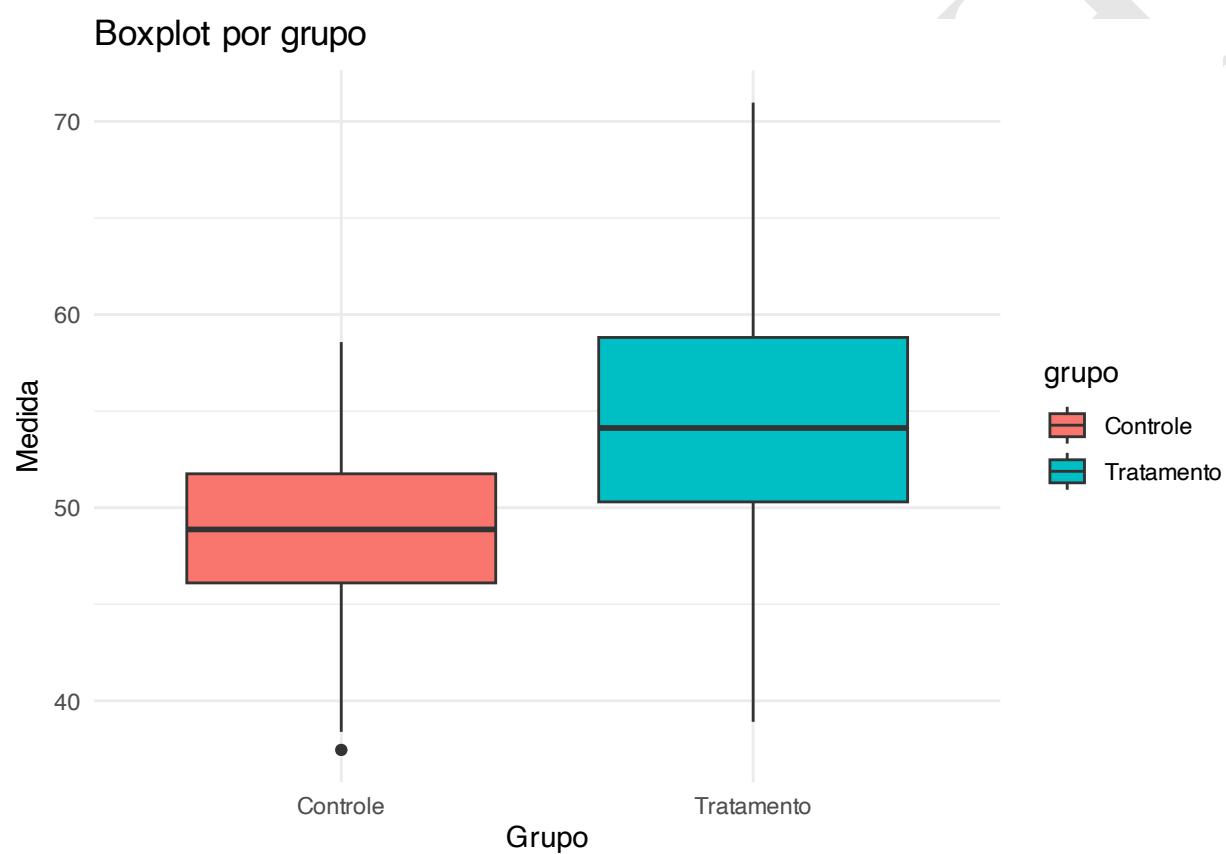


Figura 19.9: Boxplot por grupo.

- Gráfico de violino: Combina boxplot e densidade, mostrando a distribuição da variável. Útil para comparar grupos.[?]

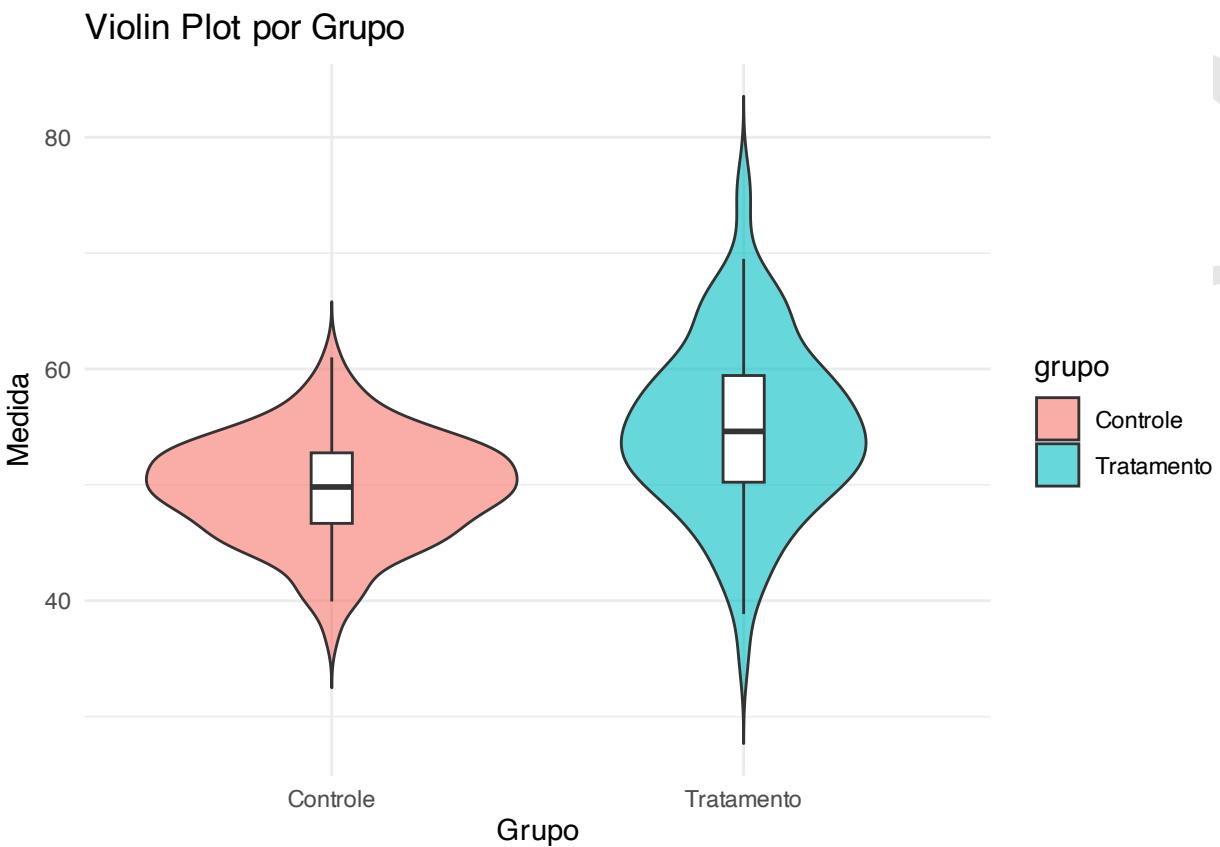


Figura 19.10: Violin plot por grupo.

- Gráfico de pontos (*dot plot*): Mostra cada valor individualmente, útil para pequenas amostras e para visualizar a distribuição.[?]

19.8.3 Quais são os tipos de gráficos para relações entre variáveis?

- Gráfico de dispersão (*scatter plot*): Mostra a relação entre duas variáveis quantitativas. Ideal para investigar correlações.[?]
- Gráfico de bolhas (*bubble chart*): Expande o gráfico de dispersão adicionando uma terceira variável (tamanho da bolha).[?]
- Gráfico Sankey: Visualiza fluxos entre categorias em diferentes etapas ou grupos. Útil para mostrar proporções e transições.[?]
- Gráfico de *parcats*: Mostra relações entre múltiplas variáveis categóricas em paralelo. Útil para visualizar fluxos e proporções.[?]
- Gráfico de *parts*: Mostra a composição percentual de uma variável categórica. Útil para visualizar proporções.[?]

19.8.4 Quais são os tipos de gráficos para dados longitudinais?

- Gráfico de *spaghetti*: Mostra trajetórias individuais ao longo do tempo, útil para dados longitudinais.[?]

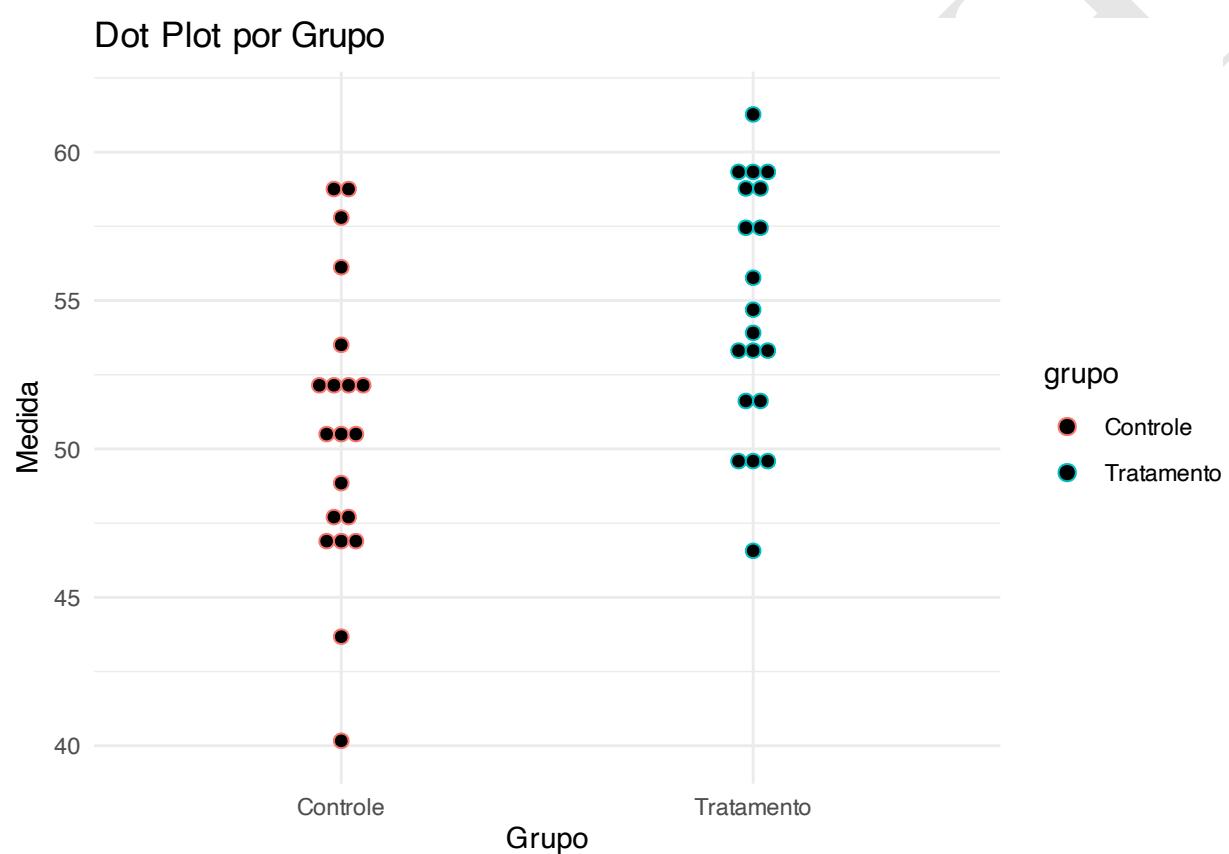


Figura 19.11: Gráfico de pontos da variável 'valor'.

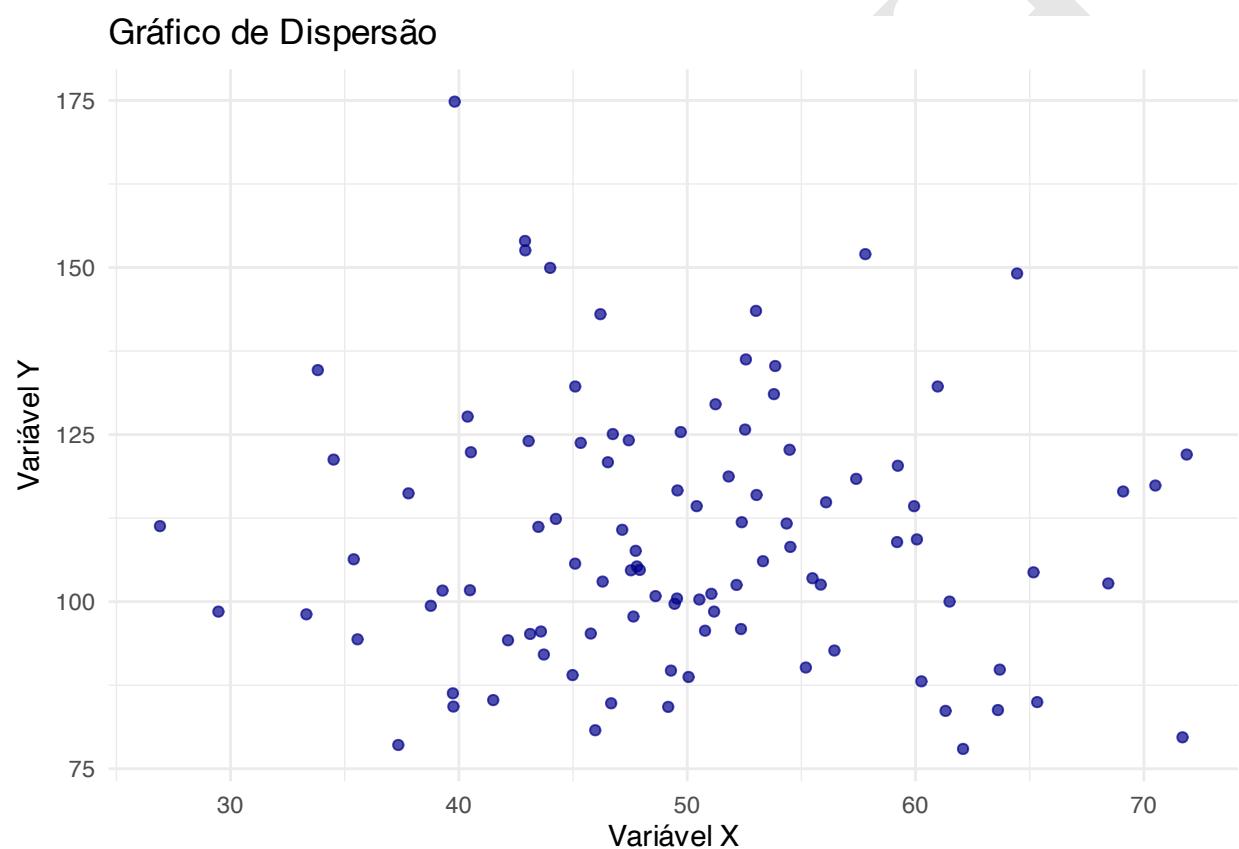


Figura 19.12: Gráfico de dispersão representando a relação entre duas variáveis.

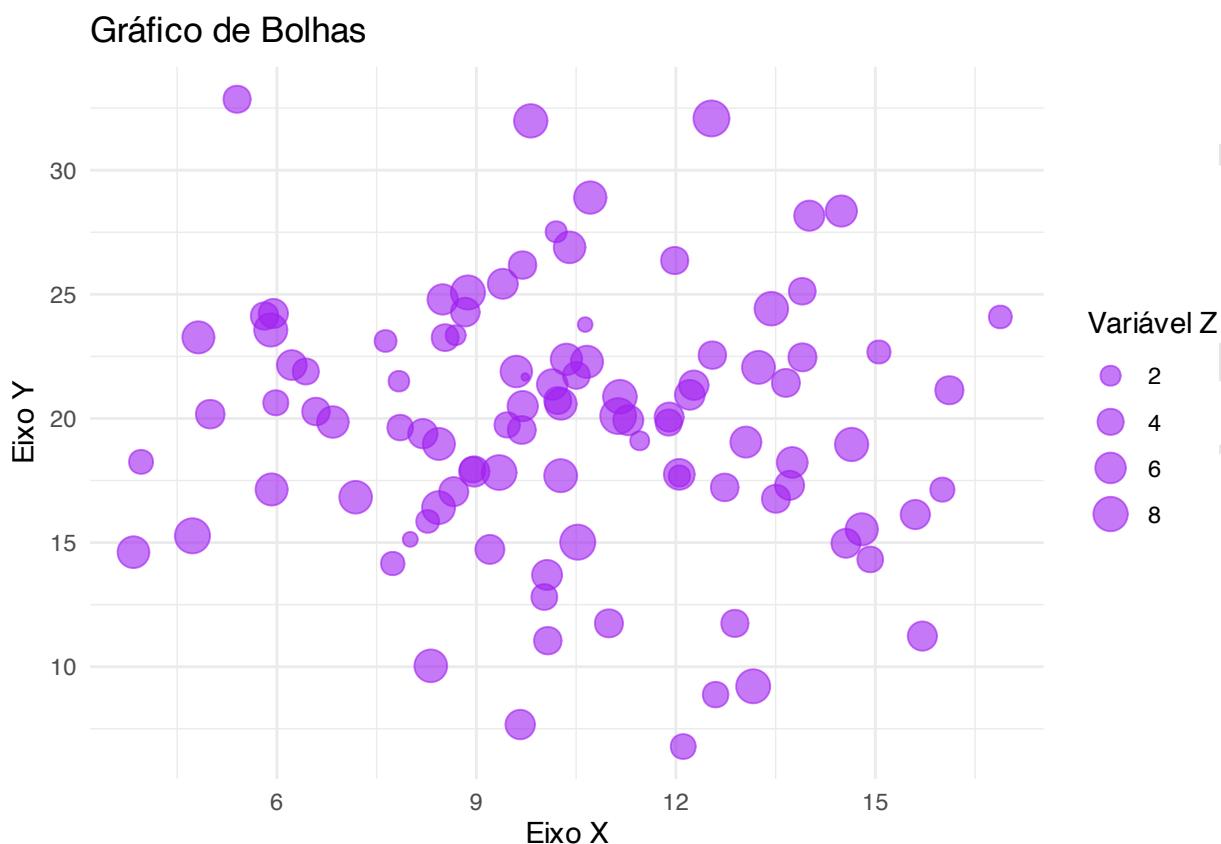


Figura 19.13: Gráfico de bolhas representando a relação entre três variáveis.

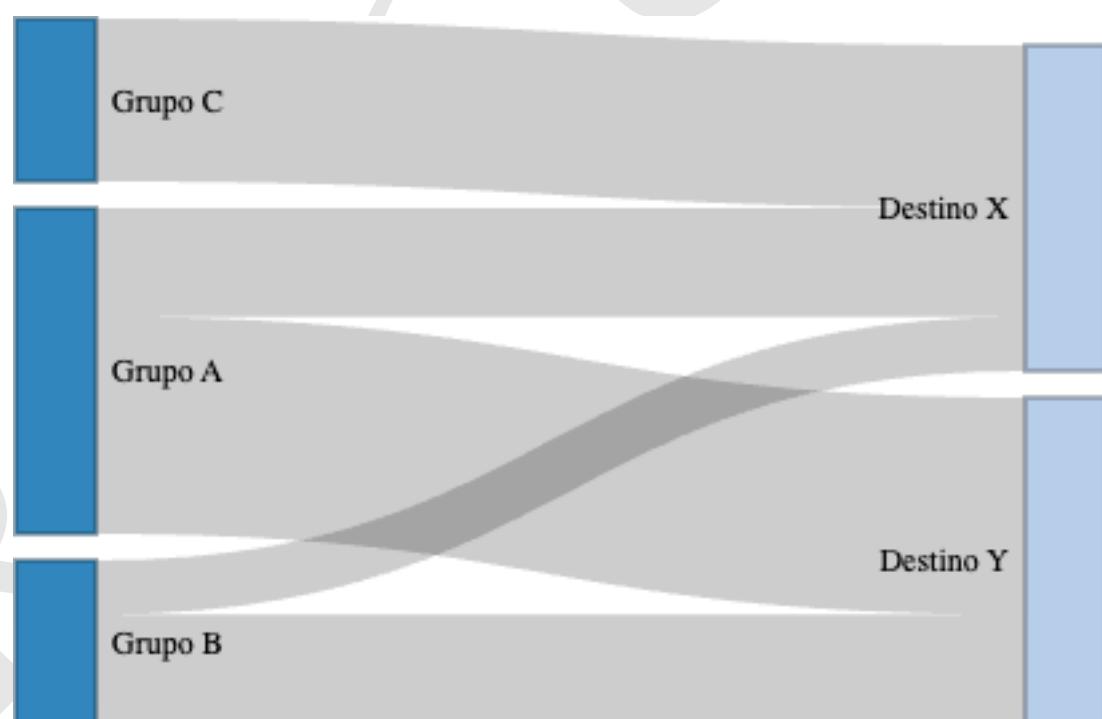


Figura 19.14: Sankey plot representando fluxos entre categorias.

Parallel Categories • parcats (plotly)

vs in categorical data — ideal for longitudinal data, surveys, multi-stage classification

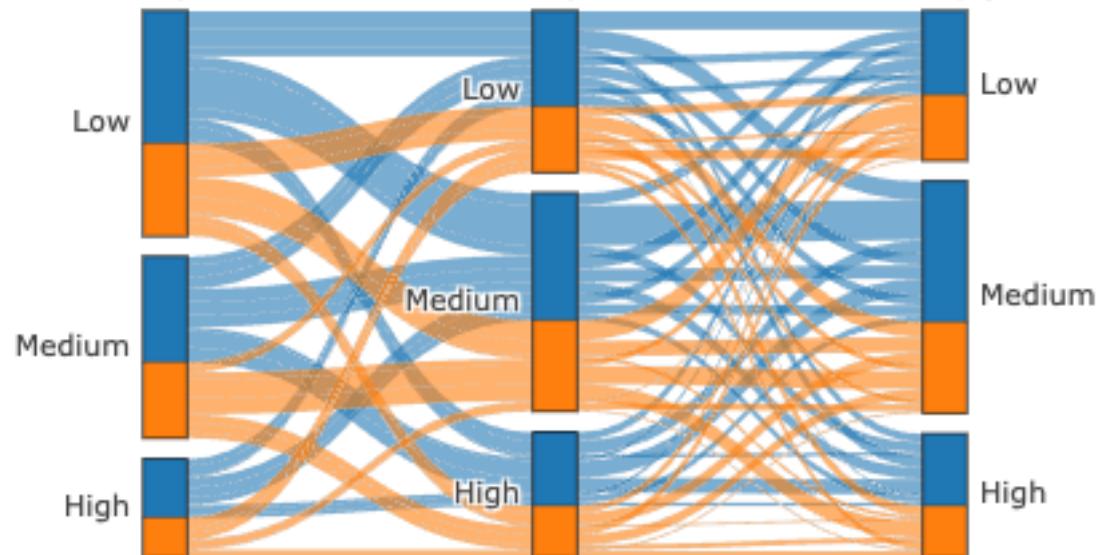


Figura 19.15: Gráfico de categorias paralelas (parcats) representando transições entre categorias ao longo do tempo.

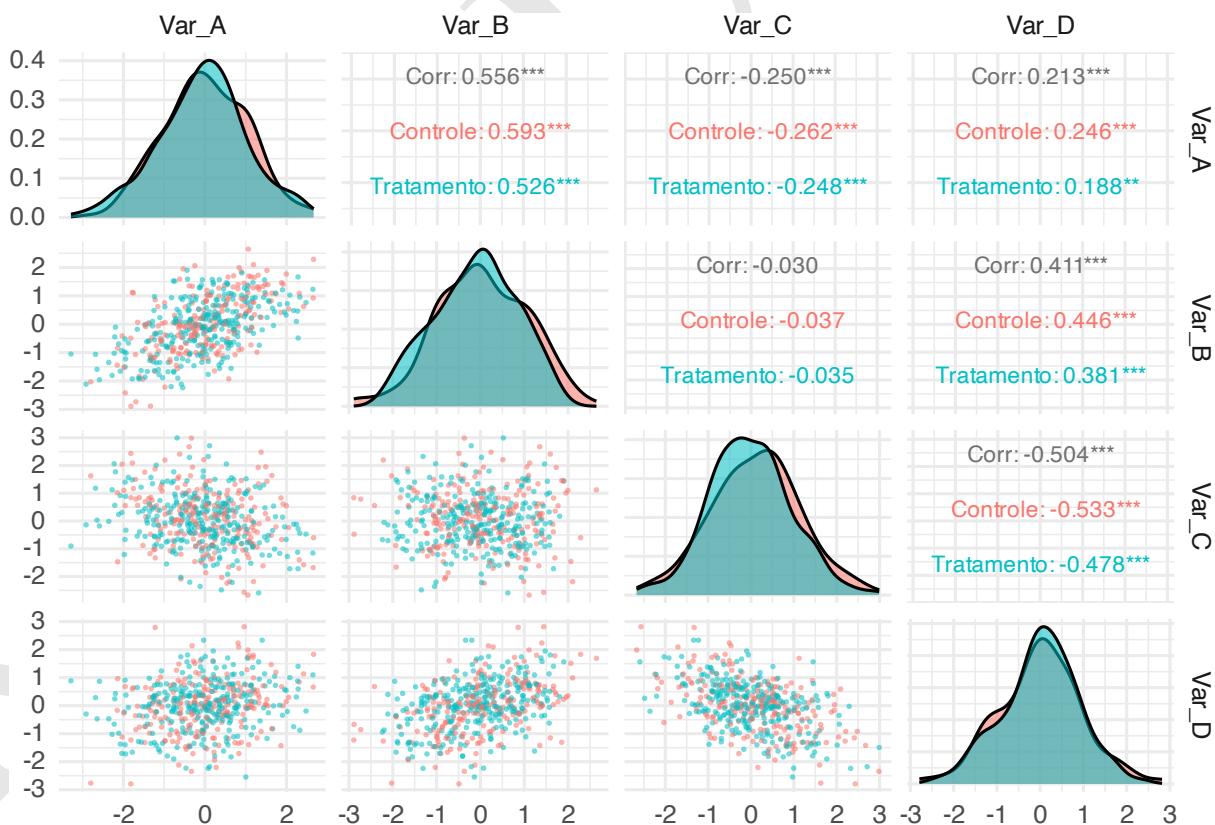


Figura 19.16: Gráfico de pares representando correlações entre múltiplas variáveis.

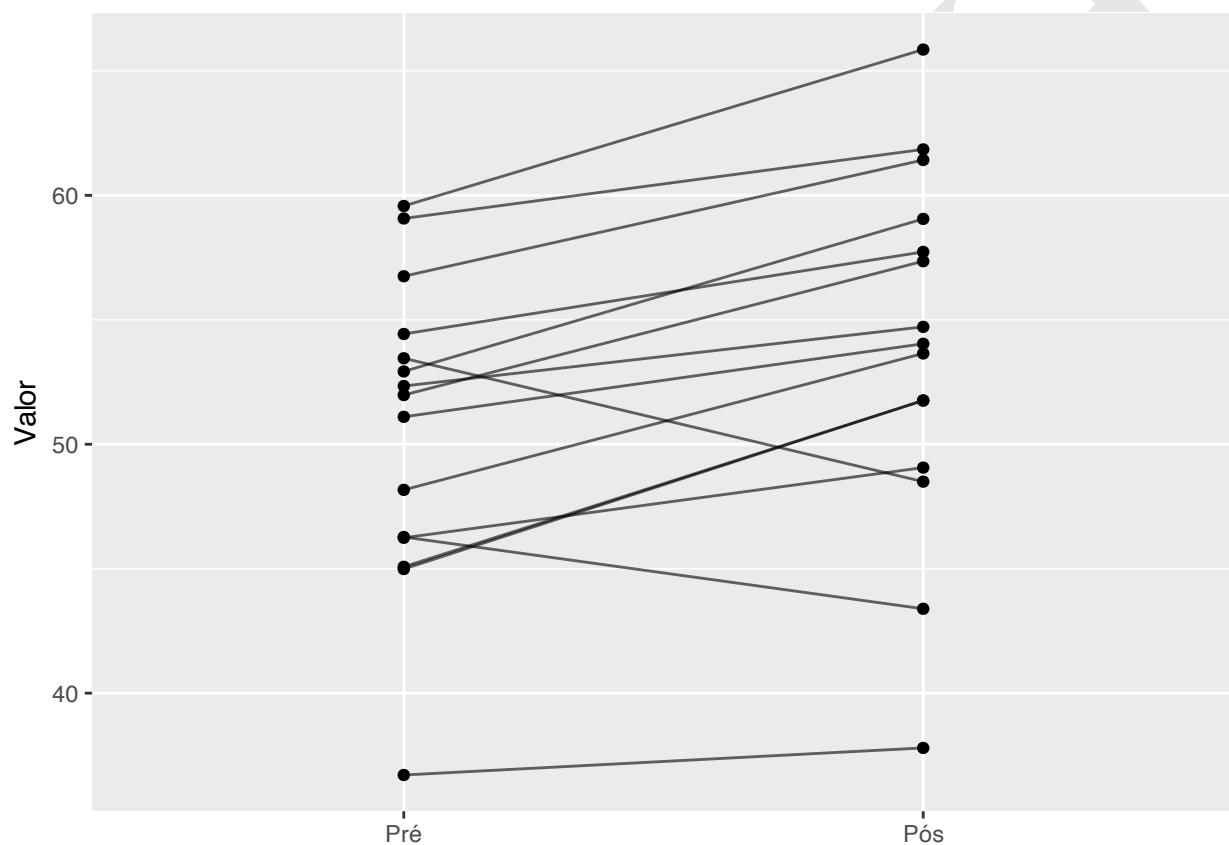


Figura 19.17: Gráfico spaghetti representando dados longitudinais.

19.8.5 Quais são os tipos de gráficos para séries temporais?

- Gráfico de linhas: Mostra a evolução de uma variável ao longo do tempo, com pontos conectados por linhas.²

Série Temporal Simulada

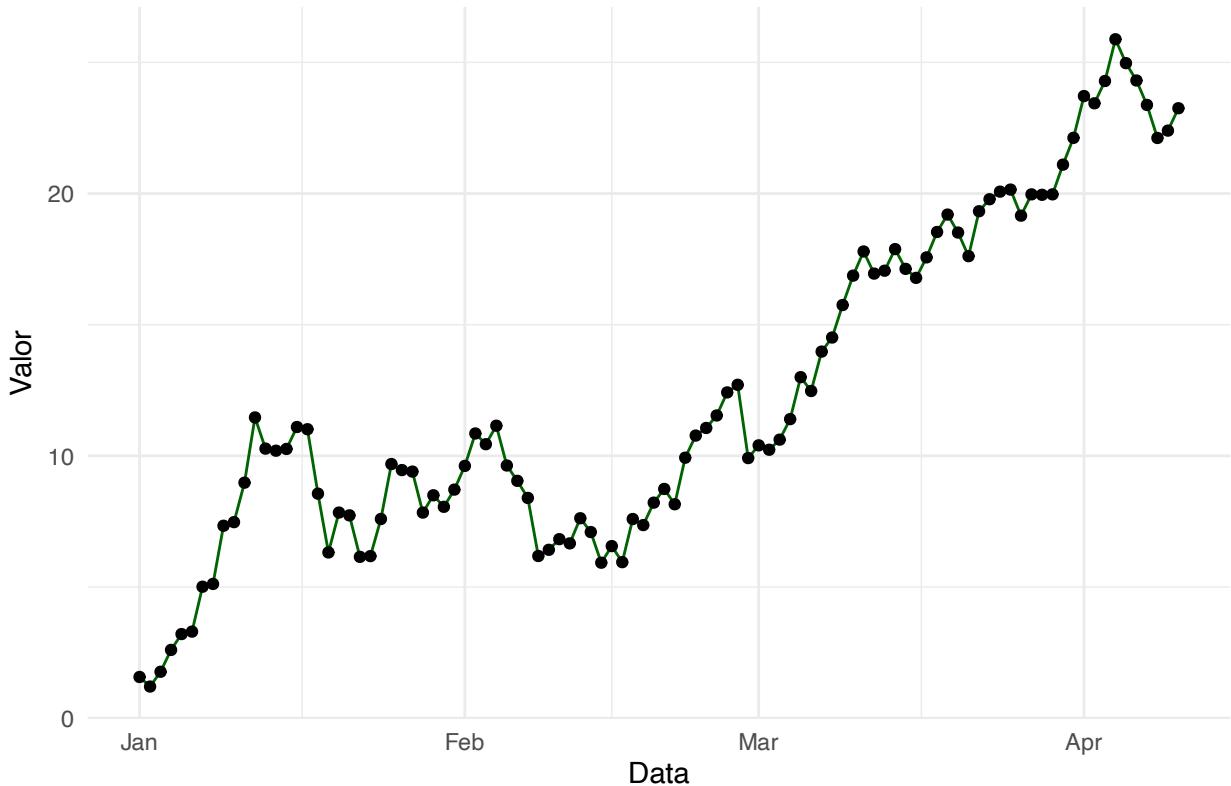


Figura 19.18: Gráfico de linha representando uma série temporal.

19.8.6 Quais são os tipos de gráficos para dados multivariados?

- Gráfico de dispersão: Representa a relação entre duas variáveis, com pontos e uma linha de tendência.²
- Gráfico de matriz de dispersão: Mostra relações entre múltiplas variáveis quantitativas, útil para identificar padrões.²
- Gráfico de calor (*heatmap*): Representa dados em uma matriz, com cores indicando intensidade ou frequência.²
- Gráfico de radar (ou gráfico de aranha): Representa várias variáveis em um único gráfico, útil para comparar perfis.²

19.8.7 Quais são as boas práticas na elaboração de gráficos?

- O tamanho da amostra total e subgrupos, se houver, deve estar descrito na figura ou na sua legenda.¹⁸³
- Para análise inferencial de figuras, as barras de erro representadas por erro-padrão ou intervalo de confiança no nível de significância α pré-estabelecido são preferíveis à amplitude ou desvio-padrão.^{182,183}
- Evite gráficos de barra e mostre a distribuição dos dados sempre que possível.²⁰⁸

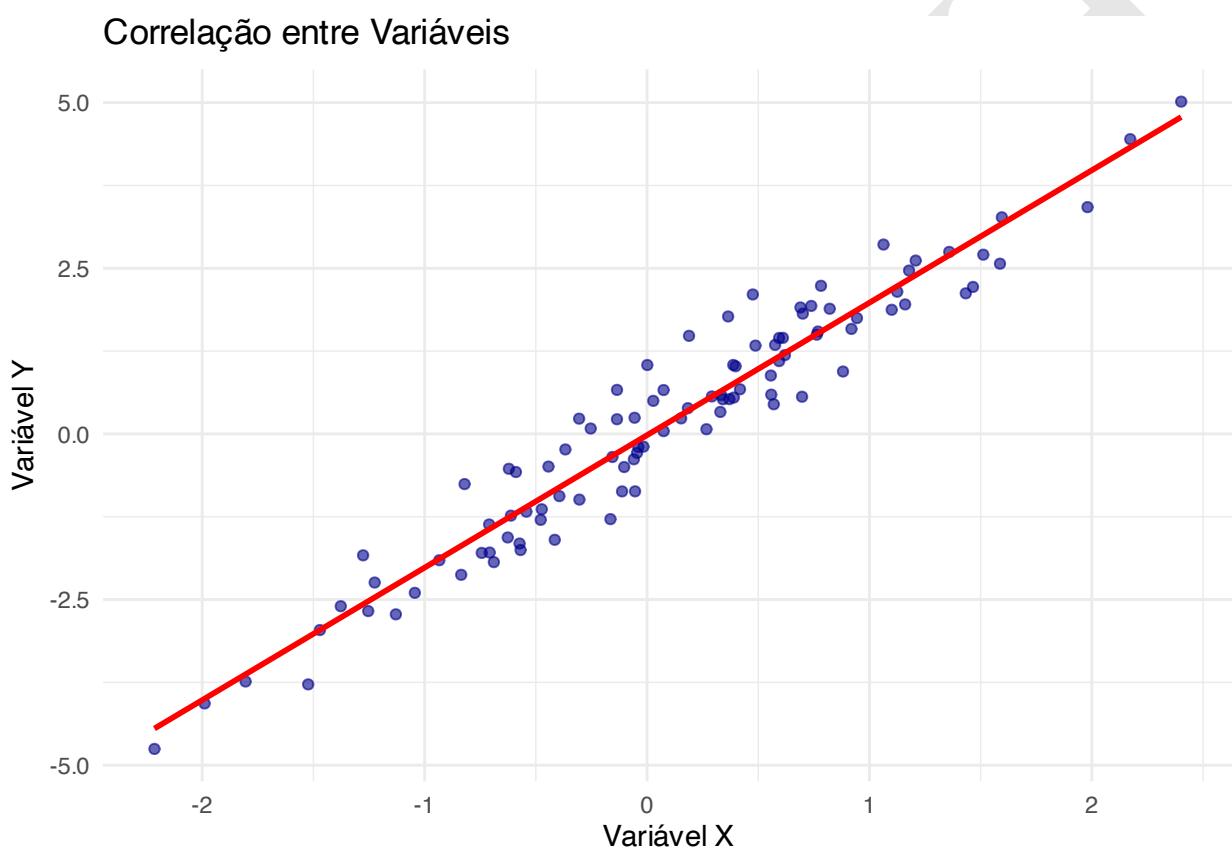


Figura 19.19: Gráfico de correlação entre duas variáveis com linha de tendência.

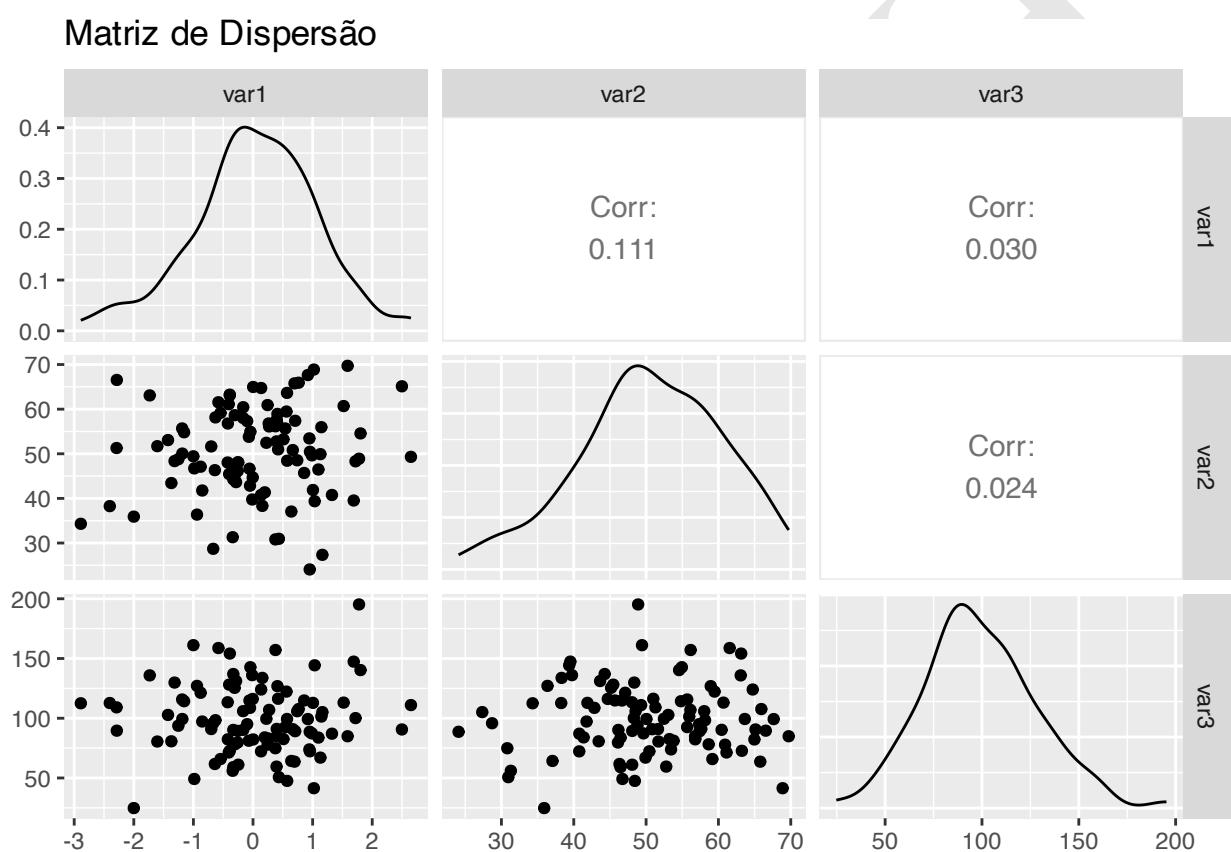


Figura 19.20: Matriz de dispersão representando relações entre múltiplas variáveis.

Mapa de Calor da Correlação entre Variáveis

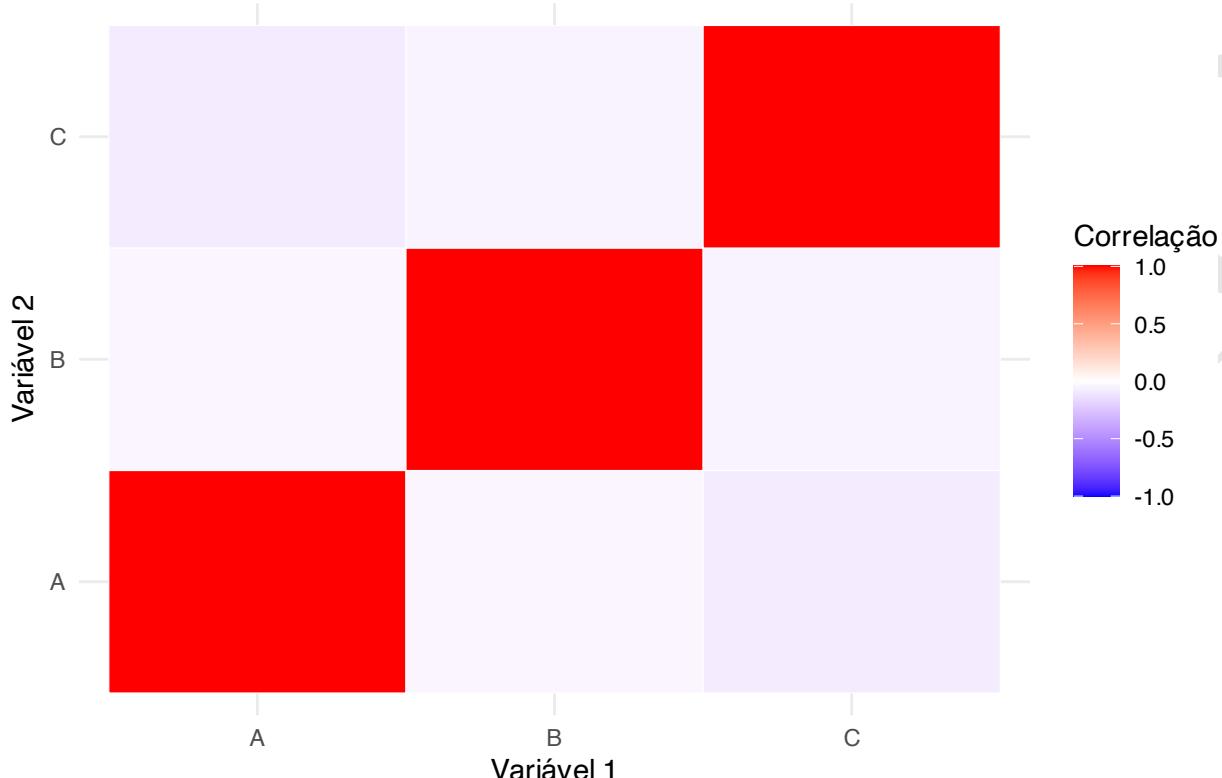


Figura 19.21: Mapa de calor da correlação entre variáveis.

Gráfico de Radar - Perfis Simulados

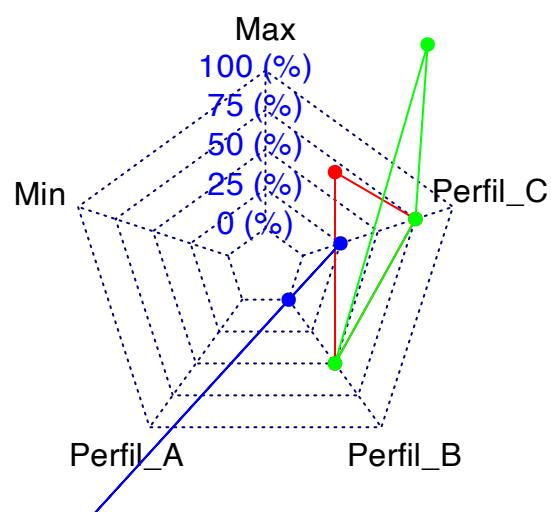


Figura 19.22: Gráfico radar representando múltiplas variáveis.

- Exiba os pontos de dados em boxplots.²⁰⁸
- Use *jitter* simétrico em gráficos de pontos para permitir a visualização de todos os dados.²⁰⁸
- Prefira palhetas de cor adaptadas para daltônicos.²⁰⁸
- Uma boa legenda torna a figura autossuficiente: descreva amostra (n), geometrias, métricas de incerteza, escalas/unidades e mensagem principal. Se houver modelo, indique fórmula/ajustes em nota.²²¹
- Evite gráficos de barras com médias para variáveis contínuas; prefira pontos/box/violino e, em amostras pequenas, exiba todos os dados.²⁰⁸
- Antes de finalizar um gráfico, faça as seguintes perguntas: (1) Mensagem está explícita? (2) Geometria adequada e dados visíveis? (3) Incerteza correta e rotulada? (4) Cores informativas e acessíveis? (5) Escalas comparáveis (se facetou)? (6) Legenda/caption autossuficiente? (7) Diferença clara entre dados e modelos? (8) Arquivo exportado na resolução/tamanho exigidos?²²¹



O pacote *ggsci*²²⁶ fornece palhetas de cores tais como *pal_lancet*^a, *pal_nejm*^b e *pal_npg*^c inspiradas em publicações científicas para uso em gráficos.

^ahttps://www.rdocumentation.org/packages/ggsci/versions/3.0.0/topics/pal_lancet

^bhttps://www.rdocumentation.org/packages/ggsci/versions/3.0.0/topics/pal_nejm

^chttps://www.rdocumentation.org/packages/ggsci/versions/3.0.0/topics/pal_npg



O pacote *grDevices*¹⁷² fornece a função *dev.new*^a para controlar diversos aspectos do gráfico, tais como tamanho e resolução.

^a<https://www.rdocumentation.org/packages/grDevices/versions/3.6.2/topics/dev>



O pacote *tiff*²²⁷ fornece a função *writeTIFF*^a para exportar gráficos em formato TIFF.

^a<https://www.rdocumentation.org/packages/tiff/versions/0.1-11/topics/writeTIFF>

19.9 Gráficos dinâmicos

19.9.1 O que são visualizações dinâmicas?

- Visualizações dinâmicas combinam interatividade (exploração ativa pelo leitor) e animação (mudanças ao longo do tempo/iterações) para empacotar informação rica em exibições simples, tornando comunicação e exploração mais transparentes.²²⁸

19.9.2 Quando preferir interatividade?

- Durante exploração de dados em equipe: destacar pontos/linhas por participante, filtrar subconjuntos e inspecionar impactos de escolhas analíticas (p.ex., outliers) sem gerar múltiplas figuras novas.²²⁸



O pacote *plotly*²²⁴ fornece a função *plot_ly*^a para gerar gráficos interativos.

^ahttps://www.rdocumentation.org/packages/plotly/versions/4.11.0/topics/plot_ly

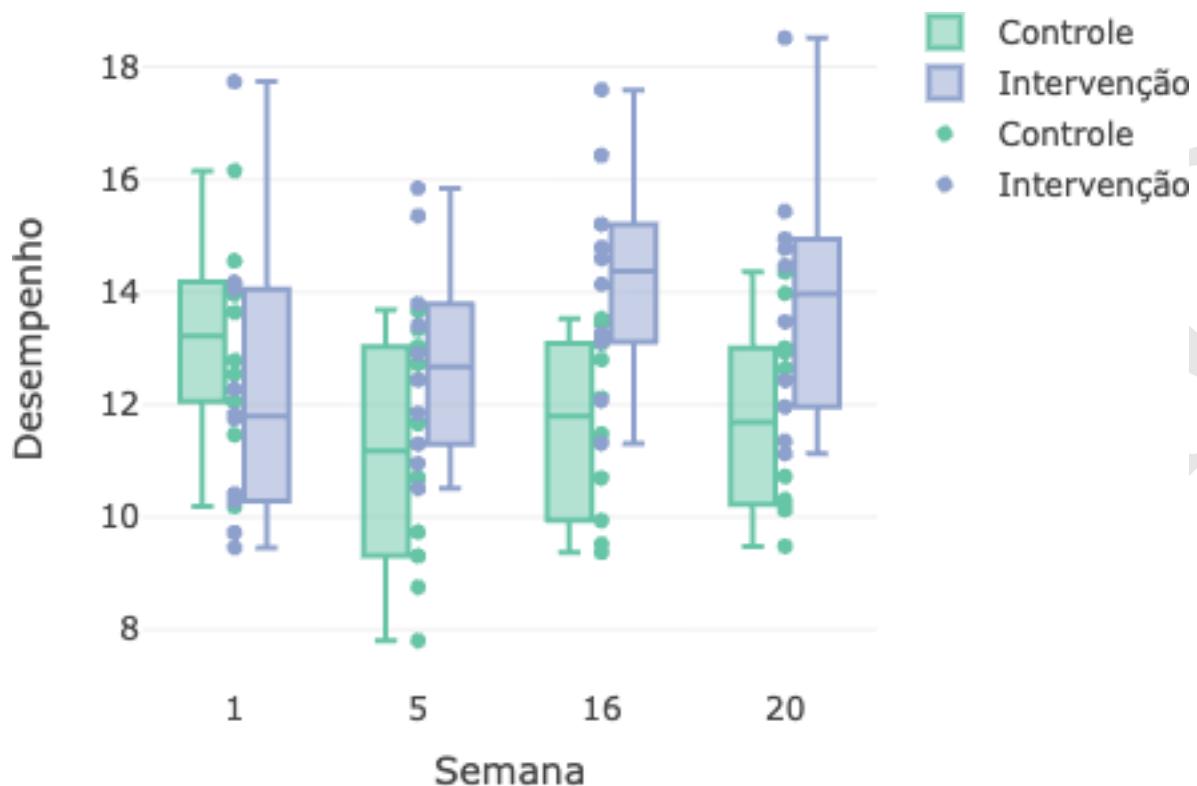


Figura 19.23: Exemplo de gráfico interativo com Plotly.

19.9.3 Quando preferir animação?

- Em apresentações e para ilustrar variação ao longo de tempo/condição/algoritmo, evitando painéis 3D ou facets excessivos. A animação guia a atenção e revela mudanças de forma passiva e fluida.²²⁸



O pacote *ganimate*²²⁹ fornece a função *transition_states*^a para criar gráficos animados a partir de gráficos estáticos do *ggplot2*¹⁷³.

^ahttps://www.rdocumentation.org/packages/gganimate/versions/1.0.7/topics/transition_states

Capítulo 20

Análise robusta

20.1 Raciocínio inferencial robusto

20.1.1 O que é análise robusta?

- Análise robusta é uma abordagem estatística que busca fornecer resultados confiáveis mesmo quando as suposições clássicas dos modelos estatísticos são violadas.²³⁰

20.1.2 Por que usar análise robusta?

- Métodos clássicos como ANOVA e regressão por mínimos quadrados assumem normalidade e homocedasticidade — suposições frequentemente violadas na prática. Violações dessas suposições podem comprometer os resultados, reduzindo o poder estatístico, distorcendo os intervalos de confiança e obscurecendo as reais diferenças entre grupos.²³⁰
- Testar previamente as suposições não é suficiente: testes de homocedasticidade têm baixo poder e não garantem segurança analítica.²³⁰
- Métodos estatísticos robustos oferecem uma solução mais segura e eficaz, lidando melhor com dados não ideais.²³⁰

20.1.3 Quando usar análise robusta?

- Em alguns casos, os métodos robustos confirmam os resultados clássicos; em outros, revelam interpretações completamente diferentes. A única forma de saber o impacto real dos métodos robustos é usá-los e comparar com os métodos tradicionais.²³⁰
- Mínimos e máximos são parâmetros descritivos, mas em certas condições podem se tornar discrepantes ou influentes, distorcendo análises. Nesses casos, a análise robusta oferece alternativas mais seguras.²³⁰

20.1.4 Por que métodos robustos são preferíveis?

- Métodos robustos têm a vantagem de resistir à influência de valores extremos, fornecendo medidas de posição e dispersão mais estáveis.¹⁹⁰
- Estimadores robustos oferecem maior segurança na presença de até 50% de contaminação nos dados, o que representa um ganho significativo em relação aos métodos clássicos.¹⁹⁰

20.2 Valores discrepantes

20.2.1 O que são valores discrepantes (*outliers*)?

- Em termos gerais, um valor discrepante — “fora da curva” ou *outlier* — é uma observação que possui um valor relativamente grande ou pequeno em comparação com a maioria das observações.¹⁹⁵
- Um valor discrepante é uma observação incomum que exerce influência indevida em uma análise.¹⁹⁵
- Valores discrepantes são dados com valores altos de resíduos.¹⁸⁸
- Nem todo valor extremo é um valor discrepante, e nem todo valor discrepante será influente.[?]
- Alguns valores discrepantes são apenas pontos incomuns, outros de fato mudam os resultados e por isso são chamados de influentes.

Régressão linear com valores discrepantes

Discrepante = Iresíduo > 2

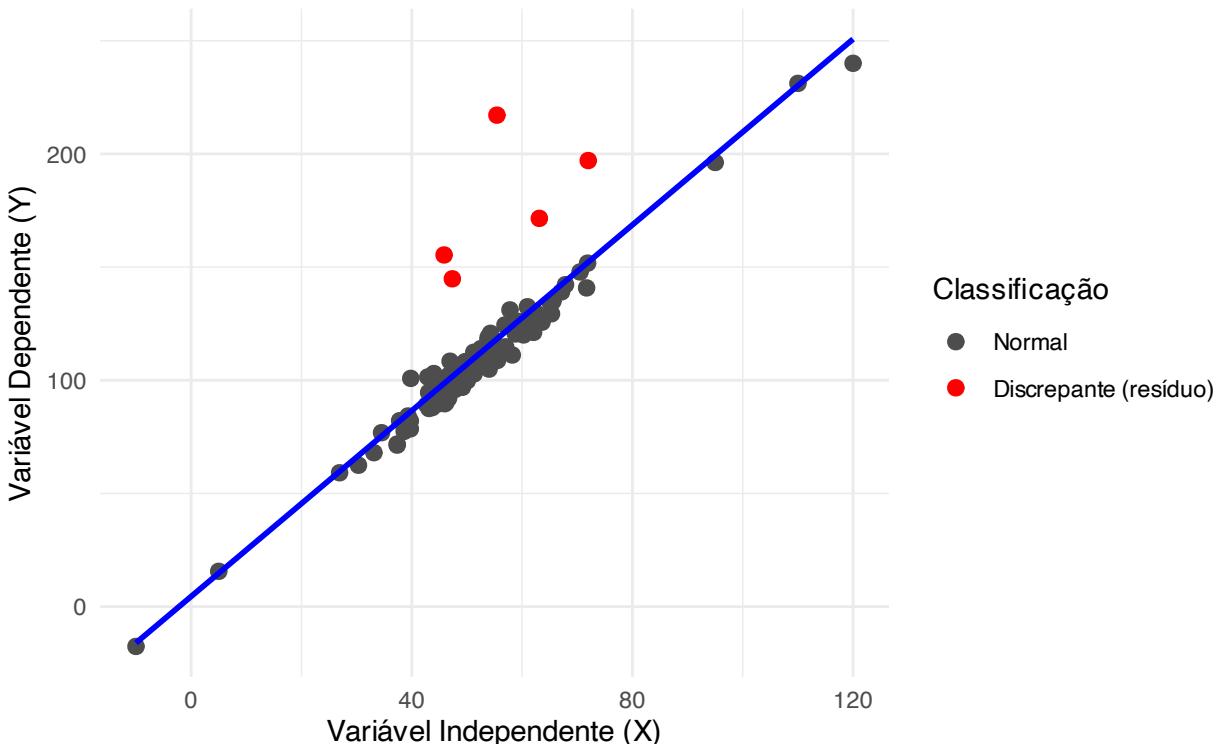


Figura 20.1: Régressão linear com valores discrepantes

20.2.2 Quais são os tipos de valores discrepantes?

- Valores discrepantes podem ser categorizados em três subtipos: *outliers* de erro, *outliers* interessantes e *outliers* aleatórios.¹⁸⁸
- Os valores discrepantes de erro são observações claramente não legítimas, distantes de outros dados devido a imprecisões por erro de mensuração e/ou codificação.¹⁸⁸
- Os valores discrepantes interessantes não são claramente erros, mas podem refletir um processo/mecanismo potencialmente interessante para futuras pesquisas.¹⁸⁸

- Os valores discrepantes aleatórios são observações que resultam por acaso, sem qualquer padrão ou tendência conhecida.¹⁸⁸
- Valores discrepantes podem ser univariados ou multivariados.¹⁸⁸

20.2.3 Por que é importante avaliar valores discrepantes?

- Excluir o valor discrepante implica em reduzir inadequadamente a variância, ao remover um valor que de fato pertence à distribuição considerada.¹⁸⁸
- Manter os dados inalterados (mantendo o valor discrepante) implica em aumentar inadequadamente a variância, pois a observação não pertence à distribuição que fundamenta o experimento.¹⁸⁸
- Em ambos os casos, uma decisão errada pode influenciar o erro do tipo I (α — rejeitar uma hipótese verdadeira) ou o erro do tipo II (β — não rejeitar uma hipótese falsa).¹⁸⁸

20.2.4 Como detectar valores discrepantes?

- Na maioria das vezes, não há como saber de qual distribuição uma observação provém. Por isso, não é possível ter certeza se um valor é legítimo ou não dentro do contexto do experimento.¹⁸⁸
- Recomenda-se seguir um procedimento em duas etapas: detectar possíveis candidatos a *outliers* usando ferramentas quantitativas; e gerenciar os outliers, decidindo manter, remover ou recodificar os valores, com base em informações qualitativas.¹⁸⁸
- A detecção de outliers deve ser aplicada apenas uma vez no conjunto de dados; um erro comum é identificar e tratar os outliers (como remover ou recodificar) e, em seguida, reaplicar o procedimento no conjunto de dados já modificado.¹⁸⁸
- A detecção ou o tratamento dos *outliers* não deve ser realizada após a análise dos resultados, pois isso introduz viés nos resultados.¹⁸⁸

20.2.5 Quais são os métodos para detectar valores discrepantes?

- Valores univariados são comumente considerados *outliers* quando são mais extremos do que a média \pm (desvio padrão \times constante), podendo essa constante ser 3 (99,7% das observações estão dentro de 3 desvios-padrão da média) ou 3,29 (99,9% estão dentro de 3,29 desvios-padrão).¹⁸⁸
- Para detectar *outliers* univariados, recomenda-se o uso da Mediana da Desviação Absoluta (*Median Absolute Deviation*, MAD), calculado a partir de um intervalo em torno da mediana, multiplicado por uma constante (valor padrão: 1,4826).^{188,231}
- Para detectar *outliers* multivariados, comumente utiliza-se a distância de Mahalanobis, que identifica valores muito distantes do centróide formado pela maioria dos dados (por exemplo, 99%).¹⁸⁸
- Para detectar *outliers* multivariados, recomenda-se o Determinante de Mínima Covariância (*Minimum Covariance Determinant*, MCD), pois possui o maior ponto de quebra possível e utiliza a mediana, que é o indicador mais robusto em presença de *outliers*.^{188,232}

20.2.6 Quais testes são apropriados para detectar valores discrepantes?

- A escolha do método de detecção depende da natureza do outlier, se univariado ou multivariado.¹⁹⁰
- Para valores univariados, podem ser usados box-plots (com pontos além de 1,5 vezes o intervalo interquartílico), z-scores clássicos ($|z| > 2,5$ ou $|z| > 3$) ou z-scores robustos, que substituem média por mediana e desvio-padrão por estimadores robustos.¹⁹⁰

- Para valores multivariados, recomenda-se a distância de Mahalanobis para medir o afastamento em relação ao centróide, com ajustes robustos de covariância como MCD (*Minimum Covariance Determinant*) ou MVE (*Minimum Volume Ellipsoid*).¹⁹⁰
- Técnicas baseadas em PCA robusta (ROBPCA, PP-PCA, SPCA, EPCA) também podem ser aplicadas para reduzir dimensionalidade e expor *outliers* mascarados.¹⁹⁰
- Métodos de *trimming* multivariado (MVT) podem iterativamente remover observações mais distantes, mas apresentam limitações em alta dimensionalidade.¹⁹⁰
- Estimadores com alto ponto de quebra, como o MCD, permitem detectar até 50% de *outliers* antes de comprometer a análise.¹⁹⁰

20.2.7 Como manejar os valores discrepantes?

- Manter *outliers* pode ser uma boa decisão se a maioria desses valores realmente pertence à distribuição de interesse. Manter *outliers* que pertencem a uma distribuição alternativa pode ser problemático, pois um teste pode se tornar significativo apenas por causa de um ou poucos *outliers*.¹⁸⁸
- Remover *outliers* pode ser eficaz quando eles distorcem a estimativa dos parâmetros da distribuição. Remover *outliers* que pertencem legitimamente à distribuição pode reduzir artificialmente a estimativa do erro.¹⁸⁸
- Remover *outliers* leva à perda de observações, especialmente em conjuntos de dados com muitas variáveis, quando *outliers* univariados são excluídos em cada variável.¹⁸⁸
- Recodificar *outliers* evita a perda de uma grande quantidade de dados, mas deve ser baseada em argumentos razoáveis e convincentes.¹⁸⁸
- Erros de observação e de medição são uma justificativa válida para descartar observações discrepantes.¹⁹⁵

20.2.8 Como conduzir análises com valores discrepantes?

- É importante reportar se existem valores discrepantes e como foram tratados.¹⁹⁵
- Valores discrepantes na variável de desfecho podem exigir uma abordagem mais refinada, especialmente quando representam uma variação real na variável que está sendo medida.¹⁹⁵
- Valores discrepantes em uma (co)variável podem surgir devido a um projeto experimental inadequado; nesse caso, abandonar a observação ou transformar a covariável são opções adequadas.¹⁹⁵
- Valores discrepantes podem ser recodificados usando a Winsorização,²³³ que transforma os *outliers* em valores de percentis específicos (como o 5º e o 95º).¹⁸⁸

R

O pacote *outliers*²³⁴ fornece a função *outlier*^a para identificar os valores mais distantes da média.

^a<https://www.rdocumentation.org/packages/outliers/versions/0.15/topics/outlier>

R

O pacote *outliers*²³⁴ fornece a função *rm.outlier*^a para remover os valores mais distantes da média detectados por testes de hipótese e/ou substitui-los pela média ou mediana.

^a<https://www.rdocumentation.org/packages/outliers/versions/0.15/topics/rm.outlier>

20.2.9 Como lidar com *outliers* na análise exploratória de dados?

- Após a detecção, três estratégias principais podem ser adotadas: (1) manter os outliers, (2) removê-los ou (3) recodificá-los (por exemplo, com Winsorização). A escolha deve ser justificada com base no contexto teórico e nas características do banco de dados. Idealmente, erros devem ser corrigidos ou removidos, enquanto outliers interessantes podem gerar novas hipóteses de pesquisa.¹⁸⁸
- A decisão sobre como lidar com outliers deve ser definida *a priori* e preferencialmente registrada em plataformas de pré-registro. Essa prática aumenta a transparência, reduz a flexibilidade analítica e evita inflar taxas de erro tipo I.¹⁸⁸

20.3 Valores influentes

20.3.1 O que são valores influentes?

- Valores influentes são observações que, se removidas, causariam uma mudança significativa nos resultados da análise estatística.¹⁸⁹

Régressão linear com valores influentes

Influente = Cook > 0.036

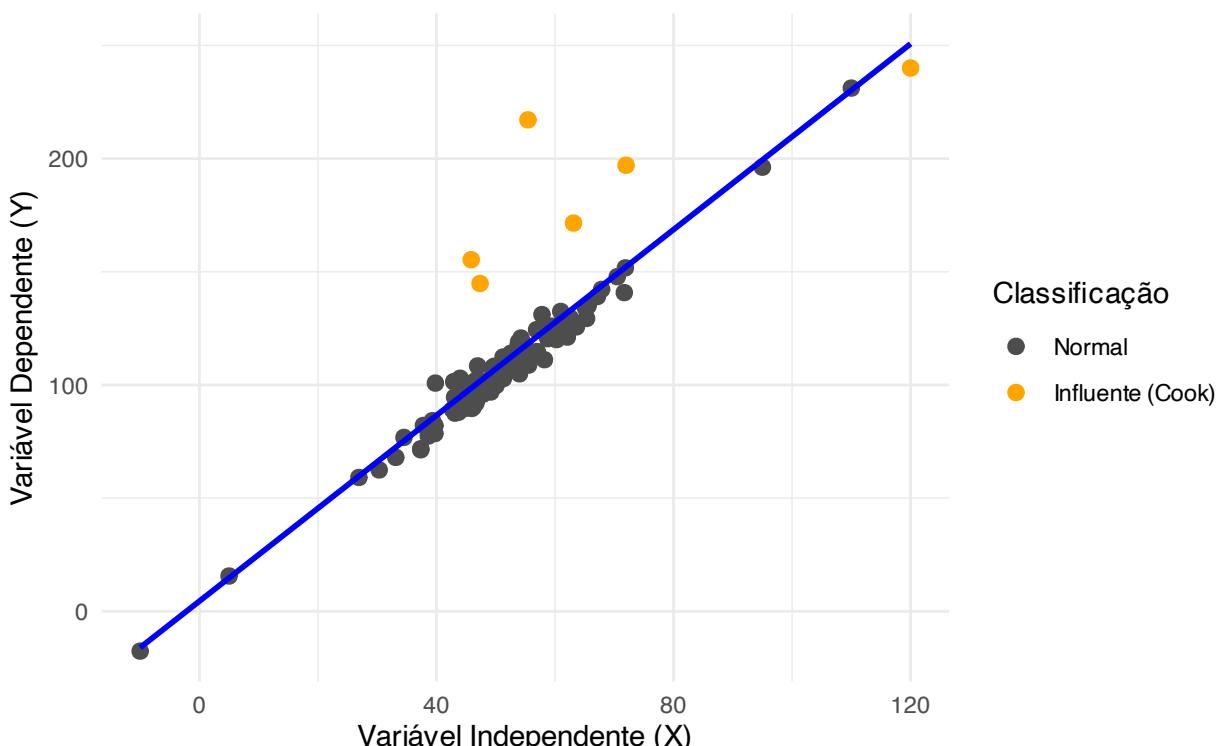


Figura 20.2: Regressão linear com valores influentes.

20.3.2 O que é função de influência?

- A função de influência mede a sensibilidade de um estimador a pequenas contaminações nos dados. Um estimador é considerado robusto se sua função de influência for limitada, indicando que valores extremos não exercem impacto desproporcional.²³⁵

20.3.3 O que é ponto de quebra?

- O ponto de quebra representa a fração mínima de observações contaminadas necessária para distorcer um estimador até o infinito. Por exemplo, a média tem ponto de quebra 0, enquanto a mediana atinge o ponto de quebra máximo (50%).²³⁵

20.3.4 Como detectar valores influentes?

- A alavancagem (*leverage*) mede o quanto distante uma observação está dos valores médios das variáveis independentes. Observações com alta alavancagem têm o potencial de influenciar significativamente a linha de regressão.²³⁶

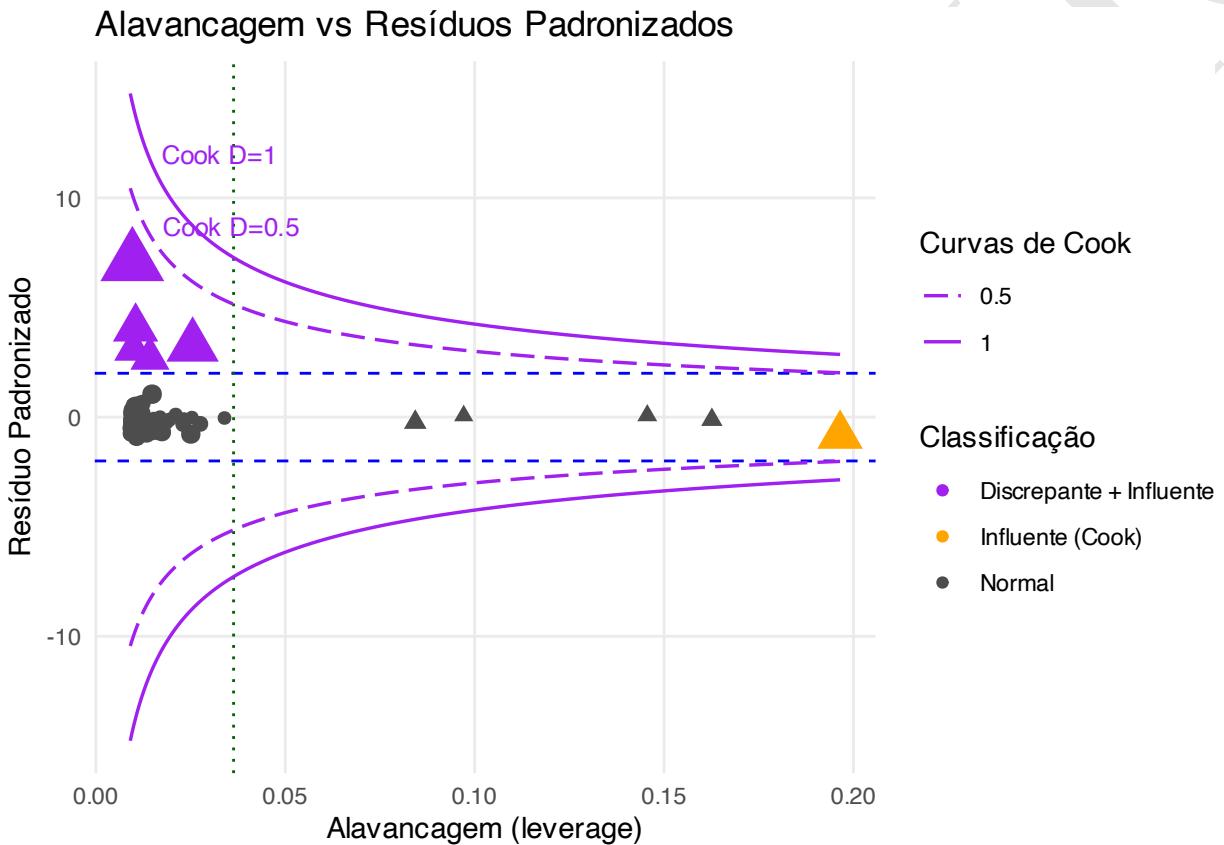


Figura 20.3: Alavancagem vs Resíduos Padronizados com distância de Cook para análise da influência de pontos.

20.4 Métodos robustos de tratamento de *outliers*

20.4.1 O que é Winsorização?

- Winsorização é uma técnica que substitui os valores extremos (*outliers*) por valores menos extremos, preservando a estrutura dos dados. Isso é feito definindo limites superior e inferior e substituindo os valores que ultrapassam esses limites pelos próprios limites.²³⁰

20.4.2 Quais são as alternativas à Winsorização?

- Podar (*trimming*): remove diretamente uma fração fixa das observações mais extremas.²³⁷

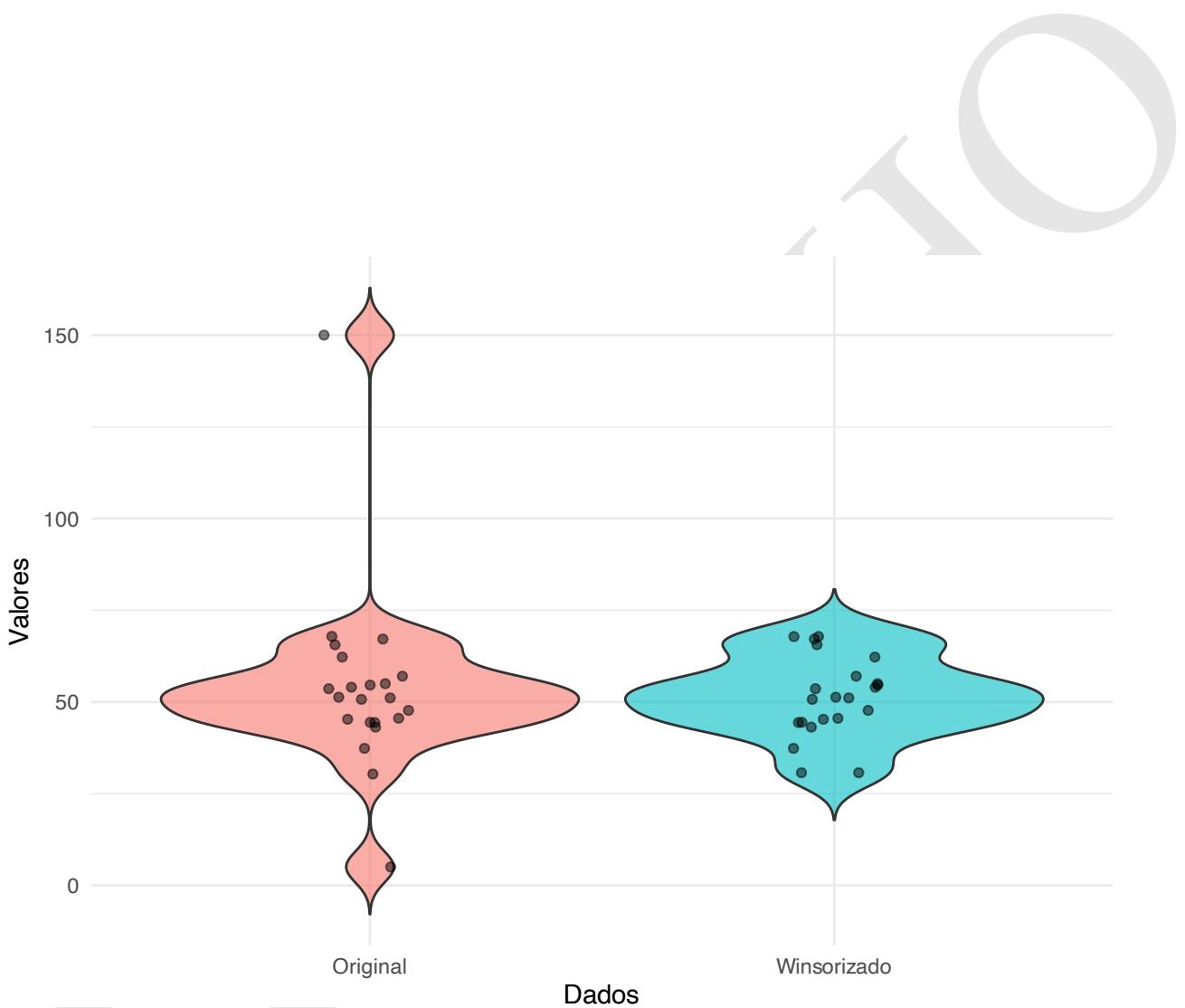


Figura 20.4: Boxplots comparando dados originais e dados Winsorizados.

- Estimadores robustos: resistem à influência de outliers sem transformar os dados.[?]
- Transformações de variáveis: reduzem a assimetria e impacto de valores extremos, mas mudam a escala interpretativa.[?]

R

O pacote *WRS2*²³⁶ fornece as funções *winmean*^a e *winvar*^b para calcular a média e variância Winsorizadas.

^a<https://www.rdocumentation.org/packages/WRS2/versions/1.1-6/topics/trimse>

^b<https://www.rdocumentation.org/packages/WRS2/versions/1.1-6/topics/trimse>

R

O pacote *WRS2*²³⁶ fornece a função *yuen*^a para realizar o teste de comparação de Yuen de médias Winsorizadas para amostras independentes ou dependentes.

^a<https://www.rdocumentation.org/packages/WRS2/versions/1.1-6/topics/yuen>

R

O pacote *WRS2*²³⁶ fornece a função *wincor*^a para calcular a correlação Winsorizada.

^a<https://www.rdocumentation.org/packages/WRS2/versions/1.1-6/topics/pbcor>

R

O pacote *WRS2*²³⁶ fornece as funções *t1way*^a, *t2way*^b e *t3way*^c para realizar testes de comparação de médias Winsorizadas para análise de variância para 1, 2 ou 3 fatores, respectivamente.

^a<https://www.rdocumentation.org/packages/WRS2/versions/1.1-6/topics/t1way>

^b<https://www.rdocumentation.org/packages/WRS2/versions/1.1-6/topics/t2way>

^c<https://www.rdocumentation.org/packages/WRS2/versions/1.1-6/topics/t3way>

Capítulo 21

Análise preditiva

21.1 Predição

21.1.1 O que são previsões?

- ?



O pacote *ggeffects*²³⁷ fornece a função *predict_response*^a para calcular valores preditos marginais ou ajustados das variáveis de desfecho.

^ahttps://www.rdocumentation.org/packages/ggeffects/versions/1.6.0/topics/predict_response



O pacote *ggeffects*²³⁷ fornece a função *test_response*^a para testar valores preditos marginais ou ajustados das variáveis de desfecho.

^ahttps://www.rdocumentation.org/packages/ggeffects/versions/1.6.0/topics/test_response

21.1.2 Como árvores de decisão são usadas para previsão?

- Utilizam padrões históricos para prever resultados futuros em novos registros.²³⁸ Identificam combinações de fatores que elevam ou reduzem o risco de um evento clínico.²³⁸
- Podem ser aplicadas em diagnósticos médicos para prever subtipos de condições ou diferentes respostas terapêuticas.²³⁸

21.2 Interpretação e aplicação

21.2.1 Quais são as implicações do uso de árvores de decisão em previsão?

- Tornam os resultados mais compreensíveis para clínicos e pesquisadores, com regras “se-então” claras.²³⁸
- Auxiliam na formulação de hipóteses clínicas e direcionamento de futuras pesquisas.²³⁸
- Podem ser estendidas para previsão de sobrevida, subgrupos de tratamento e análise de custo-benefício.²³⁸

21.3 Análise de curva de decisão

21.3.1 O que é a análise de curva de decisão?

- Análise de curva de decisão é um método para avaliar a utilidade clínica de modelos preditivos em comparação às estratégias padrão “tratar todos” ou “tratar ninguém”.²³⁹

21.3.2 O que significam os eixos da curva de decisão?

- O eixo vertical mostra o benefício líquido, ganho obtido ao usar o modelo em relação às estratégias de referência. O eixo horizontal mostra a preferência, que corresponde ao limiar de decisão: a probabilidade mínima de evento a partir da qual se recomenda intervir.²⁴⁰
- O limiar reflete o equilíbrio entre o risco de perder casos da doença e o risco de intervenções desnecessárias.²⁴⁰

21.3.3 Como interpretar o benefício líquido?

- O benefício líquido é o número de verdadeiros positivos obtidos ajustado pelo “custo” dos falsos positivos, expresso na mesma unidade dos verdadeiros positivos.²⁴⁰
- Quando a estratégia padrão é “tratar todos”, o benefício líquido também pode ser expresso como intervenções desnecessárias evitadas, facilitando a interpretação clínica.²⁴⁰

21.3.4 Por que é importante comparar sempre com “tratar todos” e “tratar nenhum”?

- Porque essas duas estratégias são frequentemente plausíveis em contextos reais: tratar todos os pacientes de risco ou tratar nenhum.²⁴⁰
- Um modelo ou teste só é considerado útil se apresentar maior benefício líquido do que ambas as estratégias, garantindo que realmente agrega valor clínico.²⁴⁰
- Essa comparação evita que um modelo com bom AUC mas má calibração seja usado, já que pode apresentar net benefit inferior às estratégias padrão.²⁴⁰

21.3.5 Quais são os limites e usos da análise de curva de decisão?

- A análise de curva de decisão não substitui uma análise de decisão completa ou uma avaliação de custo-efetividade, mas é mais simples e prática, exigindo apenas a definição de um intervalo de limiares de decisão razoáveis.²⁴⁰
- Em situações claras (modelo com nenhum benefício ou benefício amplo), pode dispensar análises complexas.²⁴⁰
- Em casos ambíguos, serve como primeiro passo antes de análises mais detalhadas.²⁴⁰

21.3.6 A análise de curva de decisão pode ser conduzida sem dados individuais de pacientes?

- Em modelos bem calibrados e com tamanho amostral adequado, a análise de curva de decisão pode ser realizada utilizando apenas estimativas sumárias (média e desvio-padrão).²³⁹

21.3.7 Como funciona o cálculo do benefício líquido?

- O benefício líquido é estimado contrastando verdadeiros positivos e falsos positivos em diferentes limiares de decisão.²³⁹

Capítulo 22

Análise causal

22.1 Causalidade

22.1.1 O que é análise causal?

- Análise causal é usada para explicar a relação entre causa e efeito em um conjunto de dados, respondendo a perguntas do tipo “por quê?”.²⁰⁴
- Análise causal implica em contrafactual, no sentido de que a análise causal é baseada na comparação entre o que realmente aconteceu e o que teria acontecido se uma ou mais variáveis tivessem sido diferentes.²⁰⁴
- Análise causal refere-se ao processo conceitual e analítico, enquanto inferência causal diz respeito às conclusões extraídas a partir de dados.²⁰⁴

22.1.2 Quais os dois grandes tipos de causalidade?

- Baseada em experiência: conhecimento empírico, muitas vezes sem compreensão dos mecanismos. Estatística tem papel central na causalidade baseada em experiência.²⁴¹
- Mecanicista: busca entender processos internos e mecanismos. Estatística tem seu papel ainda limitado, mas crescente, especialmente em sistemas complexos.²⁴¹

22.1.3 Como realizar uma análise causal?

- Formular explicitamente uma pergunta causal, definindo qual intervenção hipotética está sendo considerada e qual efeito se deseja estimar.²⁰⁴
- Especificar a ordem temporal entre exposição, desfecho e demais variáveis, assegurando que a causa preceda o efeito.²⁰⁴
- Tornar explícitas as suposições causais sobre o sistema em estudo, preferencialmente por meio de um diagrama causal (DAG).²⁰⁴
- Identificar confundidores, mediadores e colisores com base na estrutura causal assumida, e não apenas em critérios estatísticos.²⁰⁴
- Definir o conjunto mínimo de ajuste necessário para bloquear caminhos espúrios entre exposição e desfecho.²⁰⁴
- Escolher métodos estatísticos compatíveis com a pergunta causal (ex.: regressão ajustada, ponderação por escore de propensão, g-methods), reconhecendo que o método não cria causalidade.²⁰⁴

- Avaliar limitações como confundimento residual, viés de seleção e erro de medida, discutindo seu impacto potencial nas inferências.²⁴⁰
- Interpretar os resultados em termos causais explícitos, distinguindo claramente associação de efeito causal.²⁴¹

22.2 Abordagens filosóficas e estatísticas da causalidade

22.2.1 O que é realidade causal?

- A estatística assume tanto a presença do acaso quanto de causalidade. Entretanto, a natureza de cada um (se essencial ou apenas reflexo de ignorância) é raramente debatida.²⁴¹

22.2.2 Por que estatísticos historicamente evitaram falar em causalidade?

- Pearson e Fisher defenderam que estatística trata apenas de associação, não de causalidade, o que gerou cautela excessiva e paralisou avanços em áreas como economia e ciências sociais.²⁴¹
- Autores como Judea Pearl, Robins e Rubin trouxeram definições mais precisas, especialmente via modelos contrafactuals.²⁴¹
- O uso de ensaios clínicos randomizados consolidou o papel da estatística em inferência causal aplicada.²⁴¹

22.3 Ilusões de causalidade

22.3.1 O que são ilusões de causalidade?

- Ocorrem quando acreditamos que há uma relação causal entre dois eventos que, na realidade, são independentes. São comuns em superstições, pseudociências e crenças do cotidiano.²⁴²

22.3.2 Quais fatores favorecem a ilusão?

- Alta frequência do desfecho: quando o resultado ocorre frequentemente por acaso, as pessoas superestimam a eficácia da causa (ex.: melhora espontânea de sintomas atribuída a um tratamento).²⁴²
- Alta frequência da causa: quanto mais vezes um comportamento ou tratamento é aplicado, mais coincidências com o desfecho ocorrem, aumentando a crença no efeito.²⁴²
- Coincidências causa–desfecho: damos peso desproporcional a casos em que causa e efeito ocorrem juntos, mesmo que sejam apenas coincidências.²⁴²

22.3.3 Como reduzir ilusões de causalidade?

- Ensinar princípios de controle científico, observando casos em que a causa está ausente (comparação necessária para detectar ausência de relação).²⁴²
- Diminuir a frequência da causa (ex.: reduzir uso de um “remédio ineficaz” ajuda a perceber que o resultado ocorre independentemente).²⁴²
- Instruções explícitas para testar hipóteses: orientar a aplicar a causa em apenas 50% das vezes favorece a detecção correta da ausência de efeito.²⁴²
- Promover educação científica prática, mostrando às pessoas como seus próprios julgamentos podem ser enviesados e oferecendo ferramentas para avaliação crítica.²⁴²

22.4 Inferência causal em estudos observacionais

22.4.1 Como diferenciar associação de causalidade?

- Associação descreve que duas variáveis variam juntas, mas não garante que uma afete a outra.²⁴³
- Causalidade exige evidências (diretas ou indiretas) de que modificar a variável de exposição altera o desfecho.²⁴³

22.4.2 Quais critérios ajudam a sustentar inferência causal?

- Existência de um mecanismo plausível.²⁴³
- Controle adequado de confundidores (medidos e não medidos).²⁴³
- Consistência com literatura prévia e plausibilidade do tamanho do efeito.²⁴³
- Avaliação de alternativas explicativas (ex.: viés de seleção, mediadores não controlados).²⁴³

22.4.3 Qual o papel dos caminhos causais (DAGs)?

- Ajudam a identificar quais variáveis precisam ser medidas e ajustadas.²⁴³
- Evitam ajustes indevidos (ex.: em colisores), que podem introduzir viés.²⁴³

22.4.4 Como lidar com confundimento residual?

- Reconhecer que modelos multivariados e escores de propensão não eliminam completamente o confundimento.²⁴³
- Comparar características basais entre grupos para identificar diferenças persistentes.²⁴³
- Considerar análises de sensibilidade, mas com cautela na interpretação.²⁴³

22.5 Critérios de Hill para inferência causal

22.5.1 Quais são os nove critérios?

- Temporalidade: A exposição deve preceder o desfecho. Único critério considerado essencial por Hill.²⁴⁴
- Força da associação: Associações mais fortes são mais prováveis de refletir causalidade.²⁴⁴
- Consistência: A associação é observada em diferentes estudos, populações e contextos.²⁴⁴
- Especificidade: Uma exposição leva a um efeito específico (embora nem sempre aplicável).²⁴⁴
- Gradiente biológico (dose–resposta): Aumentos na exposição acompanham aumentos no risco.²⁴⁴
- Plausibilidade biológica: Compatibilidade com o conhecimento científico da época.²⁴⁴
- Coerência: A associação não deve contradizer a história natural ou biologia da doença.²⁴⁴
- Evidência experimental: Reduções na exposição devem reduzir o risco observado.²⁴⁴
- Analogia: Comparação com relações causais já conhecidas.²⁴⁴

22.5.2 Hill propôs um checklist rígido?

- Nenhum critério, isoladamente, prova ou refuta causalidade. Devem ser usados como guias para reflexão científica, não como requisitos obrigatórios.²⁴⁴

22.6 Críticas contemporâneas aos critérios de Hill

22.6.1 Qual critério é indispensável?

- A temporalidade: a exposição deve preceder o desfecho. Mesmo assim, observar uma ordem temporal inversa apenas invalida a hipótese em casos específicos, não em todos.²⁴⁵

22.6.2 A força da associação garante causalidade?

- Não. Associações fortes podem ainda ser não-causais e associações fracas podem ser causais.²⁴⁵

22.6.3 A consistência é indispensável?

- Não. A ausência de consistência não elimina causalidade, pois alguns efeitos só se manifestam em condições específicas (ex.: transfusão só causa HIV se o vírus estiver presente).²⁴⁵
- A consistência ajuda apenas a afastar a hipótese de viés ou erro em um estudo isolado:contentReference.²⁴⁵

22.6.4 O critério da especificidade é válido?

- Não. É considerado um critério inválido e enganoso. Uma causa pode ter múltiplos efeitos (tabagismo → vários desfechos) e um efeito pode ter múltiplas causas.²⁴⁵

22.6.5 O gradiente biológico (dose–resposta) é confiável?

- Nem sempre. Pode ser distorcido por confundimento. A ausência de gradiente não invalida a causalidade.²⁴⁵

22.6.6 A plausibilidade e a coerência são objetivas?

- Não. Ambas são fortemente dependentes do conhecimento científico da época. O que parecia implausível no passado depois se confirmou como verdadeiro.²⁴⁵

22.6.7 Evidência experimental é decisiva?

- Pode ser útil, mas raramente está disponível em epidemiologia. Mesmo quando disponível, pode ter explicações alternativas.²⁴⁵

22.6.8 Analogia é útil?

- Tem pouco valor. Analogias podem sempre ser inventadas e, na prática, funcionam mais como fonte de hipóteses do que como prova.²⁴⁵

22.7 Visão atual sobre os critérios de Hill

22.7.1 Como os critérios de Hill foram revisitados?

- Estudos recentes propõem integrá-los a três abordagens modernas: DAG (destacam estrutura causal e confundimento), modelos de causa suficiente (enfatizam multifatorialidade) e GRADE (orienta sobre a certeza da evidência em corpos de estudos).²⁴⁶

22.7.2 Quais mudanças na interpretação?

- Temporalidade e experimentos: seguem centrais, mas analisados com mais sofisticação.²⁴⁶
- Força da associação: relevante, mas não garante causalidade (pode haver confundimento).²⁴⁶
- Consistência: pensada como transportabilidade entre populações.²⁴⁶
- Especificidade: pouco útil hoje; substituída por falsificação (controles negativos).²⁴⁶
- Dose–resposta: pode ser espúria, cautela é necessária.²⁴⁶
- Coerência e analogia: utilidade limitada.²⁴⁶

22.8 Linguagem causal em estudos observacionais

22.8.1 Quais são as principais recomendações para relatar causalidade?

- Usar termos causais de forma explícita e criteriosa (“causa”, “efeito”, “reduzir”, “aumentar”), evitando expressões ambíguas como “fator de risco”.²⁴³
- Contextualizar a causalidade em termos práticos, explicando por que identificar a causa é relevante para intervenções.²⁴³
- Declarar claramente na introdução se existe hipótese causal, justificando quando não houver.²⁴³
- Descrever caminhos causais (mediadores, confundidores, colisores) em texto claro ou com diagramas.²⁴³
- Justificar a seleção de covariáveis com base nas relações causais previstas.²⁴³
- Avaliar o controle de confundimento, reconhecendo limitações e possível confundimento residual.²⁴³
- Discutir as inferências causais considerando estimativas, vieses e plausibilidade biológica.²⁴³
- Indicar recomendações específicas para pesquisas futuras ou prática clínica baseadas nas conclusões causais.²⁴³

22.9 Efeitos diretos e indiretos

22.9.1 Como distinguir efeitos diretos de indiretos?

- Um efeito direto ocorre quando uma variável influencia outra sem mediação.²⁴¹
- Um efeito indireto acontece quando a influência é mediada por variáveis intermediárias.²⁴¹

22.10 O papel do tempo e a causalidade dinâmica

22.10.1 O que é causalidade de Granger?

- É um conceito estatístico que analisa como processos passados influenciam o futuro, indo além da simples associação.²⁴¹
- Permite identificar relações direcionais entre processos ao longo do tempo (ex.: cérebro controlando contrações musculares).²⁴¹
- A estatística, nesse contexto, busca “olhar dentro da caixa”, aproximando-se de uma visão mecanicista.²⁴¹

22.10.2 Por que o tempo é essencial na análise causal?

- Processos causais não ocorrem de forma estática: efeitos diretos e indiretos se acumulam em cadeias temporais.²⁴¹
- Modelos tradicionais (ex.: regressões estáticas ou DAGs sem tempo) podem falhar em capturar a dinâmica.²⁴¹
- A integração de séries temporais e processos estocásticos é fundamental para compreender mecanismos.²⁴¹

22.11 Diagrama acíclico direcionado (DAG)

22.11.1 O que são DAGs?

- DAGs são representações gráficas de relações causais entre variáveis, usando nós (variáveis) e arestas direcionadas (relações causais).?
- DAGs ajudam a identificar confundidores, mediadores e colisores, orientando a seleção de variáveis para ajuste em análises estatísticas.?
- DAGs são acíclicos, ou seja, não permitem ciclos ou loops, refletindo a natureza unidirecional das relações causais.?

22.11.2 Quais são os padrões causais básicos?

- Independência: duas variáveis não têm relação causal direta ou indireta.?
- Cadeia: uma variável causa outra, que por sua vez causa uma terceira ($X \rightarrow M \rightarrow Y$).?
- Garfo: uma variável causa duas outras ($X \leftarrow Z \rightarrow Y$), onde Z é um confundidor.?
- Colisor: duas variáveis causam uma terceira ($X \rightarrow Z \leftarrow Y$), onde Z é um colisor.?

1) Independência



2) Cadeia ($X \rightarrow M \rightarrow Y$)



3) Garfo ($X \leftarrow Z \rightarrow Y$)



4) Colisor ($X \rightarrow Z \leftarrow Y$)



Figura 22.1: Padrões causais básicos: independência, cadeia, garfo e colisor.

22.11.3 Quais são as regras básicas para análise causal?

- Variáveis causalmente independentes tendem a não ser correlacionadas, exceto na presença de confundimento ou viés de seleção.²⁴⁶
- Influência causal pode criar correlação estatística, mas correlação não é condição necessária nem suficiente para causalidade.
- Confundimento cria correlação espúria entre exposição e desfecho, mesmo na ausência de efeito causal direto.
- Aleatorização protege a variável de exposição contra influência causal de fatores antecedentes, reduzindo confundimento pré-exposição.²⁴⁷

22.11.4 Quais são as regras básicas para ajuste?

- Controlar por um confundidor bloqueia caminhos causais espúrios e reduz viés na estimativa do efeito causal.²⁴⁸
- Controlar por um mediador bloqueia parte ou todo o efeito causal total, sendo inadequado quando o objetivo é estimar o efeito total da exposição.²⁴⁹
- Controlar por um colisor cria correlação espúria entre as variáveis que o causam, introduzindo viés mesmo quando não havia associação prévia.²⁵⁰
- Controlar por uma variável descendente de um colisor pode reabrir parcialmente caminhos espúrios, produzindo viés de seleção.²⁵¹



O pacote *dagitty*²⁴⁷ fornece a função *dagitty*^a para criar um objeto grafo a partir de uma descrição textual.

^a<https://cran.r-project.org/web/packages/dagitty/index.html>



O pacote *ggdag*²⁴⁸ fornece a função *ggdag*^a para criar figuras de grafos.

^a<https://www.rdocumentation.org/packages/ggdag/versions/0.2.10/topics/ggdag>



O pacote *performance*²⁴⁹ fornece a função *check_dag*^a para criar, verificar e visualizar os modelos em grafos.

^ahttps://easystats.github.io/performance/reference/check_dag.html

RASCUNHO

Capítulo 23

Análise qualitativa

23.1 Análise qualitativa

23.1.1 O que é análise qualitativa?

- ?

23.2 Representação de texto

23.2.1 O que é tokenização?

- Tokenização é o processo de dividir texto contínuo em unidades menores (tokens), como palavras, pontuação, subpalavras ou caracteres. O objetivo é criar uma representação discreta sobre a qual modelos podem calcular frequências, probabilidades e relações.²⁵⁰
- É comum combinar tokenização com normalização (*lowercase*), remoção de *stopwords*, lematização/*stemming* e regras para números e pontuação.²⁵¹

23.2.2 Modelagem com N-gramas

23.2.3 O que são n-gramas?

- Um n-grama é uma sequência contígua de n tokens, tais como: 1-gramas (unigramas), 2-gramas (bigramas), 3-gramas (trigramas).²⁵²
- Contagens de n-gramas aproximam dependências locais no texto e servem de base para DTM/TF-IDF, modelos de linguagem clássicos e detecção de coligações.²⁵³



O pacote *tidytext*²⁵⁰ fornece a função *unnest_token*^a para transformar um texto em um *data frame* com uma coluna para cada palavra.

^ahttps://www.rdocumentation.org/packages/tidytext/versions/0.4.3/topics/unnest_tokens

R

O pacote *tidytext*²⁵⁰ fornece a função *stop_words*^a para remover palavras comuns que não agregam significado.

^ahttps://www.rdocumentation.org/packages/tidytext/versions/0.4.3/topics/stop_words

R

O pacote *tidytext*²⁵⁰ fornece a função *get_sentiments*^a para obter listas de palavras com sentimentos associados.

^ahttps://www.rdocumentation.org/packages/tidytext/versions/0.4.3/topics/get_sentiments

PARTE 5: ANÁLISES INFERENCIAIS

Testando hipóteses e estimando parâmetros para responder perguntas de pesquisa

RASCUNHO

Capítulo 24

Suposições inferenciais

24.1 Suposições gerais em análises inferenciais

24.1.1 Quais são as suposições ao nível dos dados (condicionais ao modelo)?

- Independência (ou dependência corretamente modelada) das observações: .[REF]
- Forma da distribuição dos erros ou resíduos (normalidade, assimetria, caudas): .[REF]
- Homocedasticidade (igualdade de variâncias condicionais): .[REF]

24.1.2 Quais são as suposições ao nível do modelo?

- Linearidade da relação entre variáveis: .[REF]
- Multicolinearidade ausente ou controlada: .[REF]
- Especificação funcional correta do modelo: .[REF]

24.1.3 Quais são as suposições ao nível do estudo?

- Ausência de confundimento relevante não controlado: .[REF]
- Estabilidade do processo gerador de dados (invariância temporal, populacional ou contextual): .[REF]

24.2 Suposições implícitas e explícitas nos testes

24.2.1 Quais suposições implícitas são feitas nos testes estatísticos?

- Amostragem aleatória ou ignorabilidade condicional: .[REF]
- Medição sem erro relevante: .[REF]
- Correspondência entre modelo estatístico e processo gerador de dados: .[REF]
- Ausência de múltiplas comparações não ajustadas: .[REF]

24.2.2 Quais suposições explícitas são feitas nos testes estatísticos?

- Normalidade dos erros ou da estatística de teste: .[REF]
- Homocedasticidade: .[REF]

- Independência das observações: .[REF]

24.3 Suposições causais que conectam dados observados a efeitos causais

24.3.1 Quais são as suposições causais que conectam dados observados a efeitos causais?

- Ausência de correlação espúria: associações observadas refletem relações sistemáticas e não flutuações aleatórias; quanto maior a amostra, mais plausível essa condição.?
- Consistência: os valores observados do tratamento correspondem a intervenções bem definidas e coincidem com os valores dos contrafactual relevantes.?
- Intercambialidade: condicionalmente às covariáveis medidas, a atribuição do tratamento é independente dos desfechos potenciais.?
- Positividade: para todos os valores das covariáveis consideradas, a probabilidade de receber cada nível do tratamento é maior que zero.?
- Fidelidade: efeitos causais não se cancelam sistematicamente no agregado populacional, de modo que efeitos médios nulos correspondem à ausência de efeito causal relevante.?

24.3.2 Qual a relação dessas suposições com as demais suposições inferenciais?

- Essas suposições operam antes do modelo estatístico.?
- Não são verificáveis por diagnóstico residual ou testes de ajuste.?
- Mesmo com todas as suposições estatísticas satisfeitas, a inferência causal pode falhar se qualquer uma dessas suposições não for atendida.?

24.4 Diagnóstico e verificação

24.4.1 O que fazer quando suposições gerais falham?

- Transformações: .[REF]
- Métodos robustos (estimadores e testes): .[REF]
- Reamostragem: .[REF]
- Modelos alternativos: .[REF]

24.4.2 O que fazer quando as suposições causais falham?

- Clarificar o alvo causal: redefinir a população, o tratamento ou o efeito de interesse.[REF]
- Análise de sensibilidade: avaliar quanto confundimento não medido seria necessário para invalidar as conclusões.[REF]
- Restringir o suporte: limitar a análise a regiões com positividade plausível (suporte comum).[REF]
- Estratificação ou ajuste Enriquecido: incluir covariáveis adicionais relevantes, quando disponíveis.[REF]
- Modelagem causal explícita: usar DAGs para tornar suposições transparentes e discutíveis.[REF]
- Estimativas parciais ou locais: reportar efeitos condicionais ou locais quando o efeito médio não é identificável.[REF]
- Conclusões mais fracas: interpretar resultados como associações ajustadas, não como efeitos causais.[REF]

- Relato explícito das falhas: documentar quais suposições não são plausíveis e por quê.[REF]



O pacote *performance*²⁴⁹ fornece a função *check_model*^a para analisar a colinearidade entre variáveis, a normalidade da distribuição das variáveis e a heteroscedasticidade.

^ahttps://www.rdocumentation.org/packages/Performance/versions/0.10.4/topics/check_model

24.4.3 Como avaliar as suposições de uma regressão?

- Usando diagnóstico de regressão (ex.: análise de resíduos, gráficos de valores observados vs. preditos) e comparação com análises estratificadas.²⁵¹

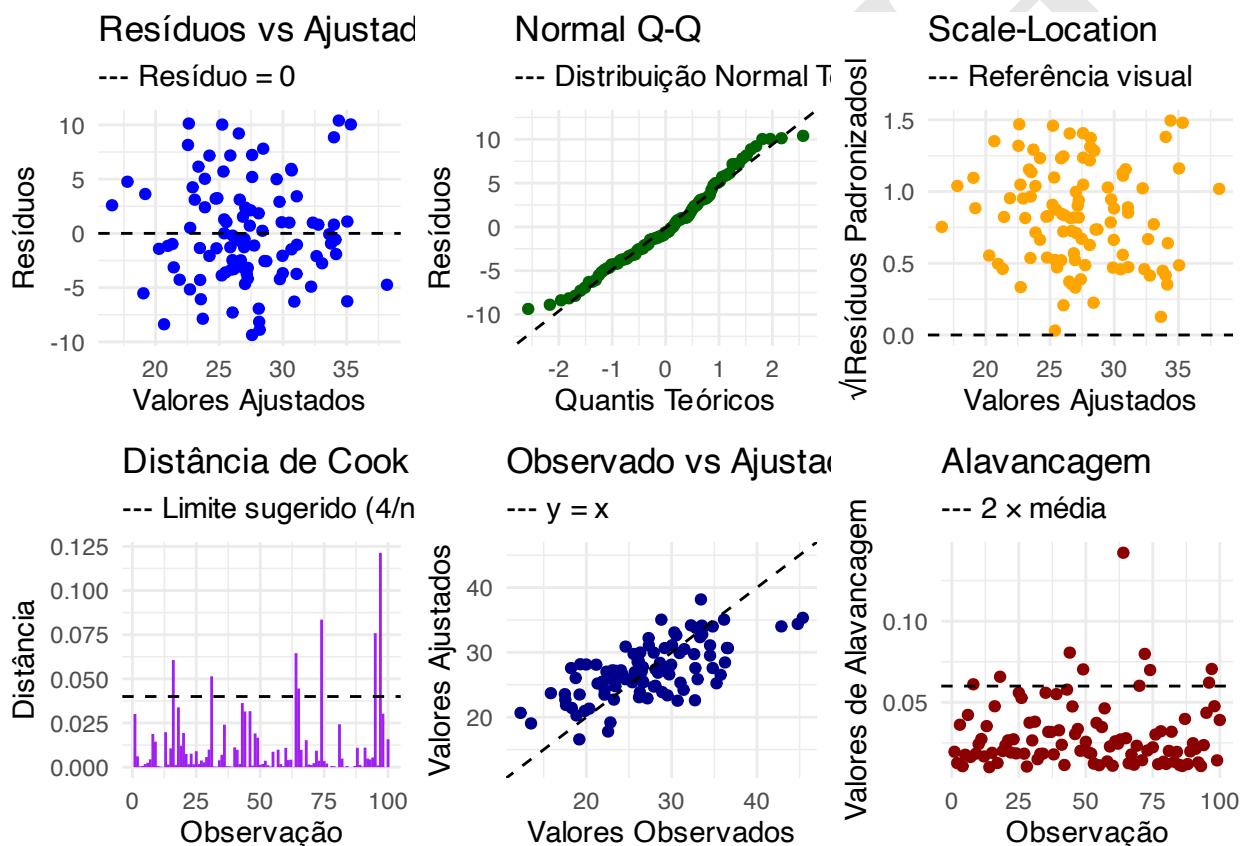


Figura 24.1: Diagnóstico de regressão para avaliar suposições do modelo: linearidade, normalidade dos resíduos, homocedasticidade e alavancagem.

RASCUNHO

Capítulo 25

Seleção de testes

25.1 Multiverso de análises estatísticas

25.1.1 Por que escolher o teste é um problema?

- Analisar a mesma hipótese com o mesmo banco de dados pode resultar em diferenças substanciais nas estimativas estatísticas e nas conclusões.²⁵²
- As decisões para especificação das análises estatísticas podem ser tão minuciosas que muitas vezes nem sequer são registradas como decisões e, assim, podem impactar negativamente na reproduzibilidade do estudo.²⁵²

25.2 Escolha de testes para análise inferencial

25.2.1 Como selecionar os testes para a análise estatística inferencial?

- ²⁵³
- ²⁵⁴
- ²⁵⁵
- ²⁵⁶
- ²⁵⁷
- ²⁵⁸
- ²⁵⁹
- ²⁶⁰

Tabela 25.1: Tabela de seleção de testes estatísticos: Descrição.

Delineamento do estudo	Qtd. de variáveis / fatores	Níveis do fator	Relação entre amostras	Tipo de variáveis	Teste estatístico
Transversal	1 variável	–	–	Contínua	Média (DP) IC95% / Mediana (IIQ) IC95%
Transversal	1 variável	–	–	Categórica	Frequências e proporções

Tabela 25.2: Tabela de seleção de testes estatísticos: Comparação.

Delineamento do estudo	Qtd. de variáveis / fatores	Níveis do fator	Relação entre amostras	Tipo de variáveis	Teste estatístico
Transversal	1 variável	–	–	Contínua	t de Student (1 amostra) / Wilcoxon one-sample
Transversal	1 variável	≥ 2	–	Categórica	Qui-quadrado de aderência
Transversal	1 variável	–	–	Categórica dicotômica	Teste binomial (1 amostra)
Experimental / observacional	1 fator + 1 resposta	2	Independentes	Contínua categórica	t de Student independente / t de Welch
Pareado / longitudinal	1 fator + 1 resposta	2	Dependentes	Contínua categórica	t pareado
Experimental / observacional	1 fator + 1 resposta	≥ 3	Independentes	Contínua categórica	ANOVA one-way
Experimental / observacional	1 fator + ≥ 1 covariável + 1 resposta	≥ 2	Independentes	Contínua categórica + covariável	ANCOVA
Longitudinal	1 fator + 1 resposta	≥ 3	Dependentes	Contínua categórica	ANOVA de medidas repetidas

Continuado na próxima página

Tabela 25.2: Tabela de seleção de testes estatísticos: Comparação. (Continuado)

Delineamento do estudo	Qtd. de variáveis / fatores	Níveis do fator	Relação entre amostras	Tipo de variáveis	Teste estatístico
Experimental / observacional	1 fator + 1 resposta	2	Independentes	Ordinal / não normal	Mann–Whitney U / Kruskal–Wallis
Longitudinal	1 fator + 1 resposta	≥ 3	Dependentes	Ordinal / não normal	Wilcoxon pareado / Friedman
Experimental / observacional	1 fator + ≥ 2 respostas	≥ 2	Independentes	Contínuas múltiplas	MANOVA
Experimental / observacional	1 fator + 1 resposta	2	Independentes	Contínua / ordinal	Brunner–Munzel
Transversal	2 correlações	–	Independentes	Contínua \times contínua	Fisher r-to-z (correlações independentes)
Transversal	≥ 3 variáveis contínuas	–	Dependentes	Contínua \times contínua	Steiger / Meng–Rosenthal–Rubin (correlações dependentes)

Tabela 25.3: Tabela de seleção de testes estatísticos: Associação.

Delineamento do estudo	Qtd. de variáveis / fatores	Níveis do fator	Relação entre amostras	Tipo de variáveis	Teste estatístico
Transversal	2 variáveis	–	–	Contínua \times contínua	Correlação de Pearson
Transversal	2 variáveis	–	–	Ordinal / não normal	Correlação de Spearman / Kendall
Transversal	2 variáveis	–	–	Ordinal \times ordinal	Gamma de Goodman–Kruskal
Transversal	2 variáveis	–	–	Ordinal \times ordinal	Tau-b de Kendall
Transversal	2 variáveis	–	–	Ordinal \times ordinal	Somers' D

Continuado na próxima página

Tabela 25.3: Tabela de seleção de testes estatísticos: Associação. (Continuado)

Delineamento do estudo	Qtd. de variáveis / fatores	Níveis do fator	Relação entre amostras	Tipo de variáveis	Teste estatístico
Transversal	2 variáveis	–	–	Categórica × categórica (2×2)	Qui-quadrado / Fisher + Phi (ϕ)
Transversal	2 variáveis	–	–	Categórica × categórica (freq. pequenas)	Exato de Fisher
Transversal	2 variáveis	–	–	Categórica × categórica ($R \times C$)	Qui-quadrado + V de Cramér

Tabela 25.4: Tabela de seleção de testes estatísticos: Predição.

Delineamento do estudo	Qtd. de variáveis / fatores	Níveis do fator	Relação entre amostras	Tipo de variáveis	Teste estatístico
Observacional / experimental	≥ 1 preditor + 1 resposta	–	–	Contínua	Régressão linear
Observacional / experimental	≥ 1 preditor + 1 resposta	–	–	Binária	Régressão logística

Tabela 25.5: Tabela de seleção de testes estatísticos: Contagem.

Delineamento do estudo	Qtd. de variáveis / fatores	Níveis do fator	Relação entre amostras	Tipo de variáveis	Teste estatístico
Observacional	≥ 1 preditor + 1 resposta	–	–	Contagem	Poisson / Binomial negativa

Tabela 25.6: Tabela de seleção de testes estatísticos: Sobrevida.

Delineamento do estudo	Qtd. de variáveis / fatores	Níveis do fator	Relação entre amostras	Tipo de variáveis	Teste estatístico
Longitudinal	≥ 1 preditor + tempo	≥ 2	Independentes	Tempo	Log-rank / Modelo de Cox
Longitudinal	≥ 1 preditor + tempo	≥ 2	Independentes	Tempo (riscos competitivos)	Modelo de riscos competitivos de Fine–Gray

Tabela 25.7: Tabela de seleção de testes estatísticos: Concordância.

Delineamento do estudo	Qtd. de variáveis / fatores	Níveis do fator	Relação entre amostras	Tipo de variáveis	Teste estatístico
Metodológico	2 avaliadores	–	Dependentes	Categórica nominal	Kappa de Cohen
Metodológico	≥ 2 avaliadores	–	Dependentes	Categórica nominal	Kappa de Fleiss
Metodológico	2 avaliadores	≥ 3	Dependentes	Categórica ordinal	Kappa de Light
Metodológico	2 medidas	–	Dependentes	Contínua	Coeficiente de Correlação Intraclass (ICC)

Tabela 25.8: Tabela de seleção de testes estatísticos: Ajuste multivariado.

Delineamento do estudo	Qtd. de variáveis / fatores	Níveis do fator	Relação entre amostras	Tipo de variáveis	Teste estatístico
Longitudinal / clusterizado	≥ 1 fator + resposta	–	Dependentes / clusters	Contínua / categórica	Modelos lineares mistos (LMM / GLMM)

Tabela 25.9: Tabela de seleção de testes estatísticos: Desempenho diagnóstico.

Delineamento do estudo	Qtd. de variáveis / fatores	Níveis do fator	Relação entre amostras	Tipo de variáveis	Teste estatístico
Transversal / longitudinal	1 teste + 1 padrão-ouro	2	Pareadas	Binária × binária	Sensibilidade, Especificidade, VPP, VPN, Acurácia
Transversal / longitudinal	1 teste + 1 padrão-ouro	2	Pareadas	Binária × binária	Razões de verossimilhança (LR^+ / LR^-)
Transversal / longitudinal	1 escore contínuo + 1 desfecho	–	Pareadas	Contínua × binária	Curva ROC + AUC (IC95%)
Transversal / longitudinal	1 modelo + 1 desfecho	–	Pareadas	Probabilidade × binária	ROC / AUC + calibração (Brier, Hosmer–Lemeshow)

Tabela 25.10: Tabela de seleção de testes estatísticos: Diagnóstico.

Delineamento do estudo	Qtd. de variáveis / fatores	Níveis do fator	Relação entre amostras	Tipo de variáveis	Teste estatístico
Transversal	1 variável	–	–	Contínua	Shapiro–Wilk
Transversal	1 variável	–	–	Contínua	Kolmogorov–Smirnov / Lilliefors
Transversal	1 variável	–	–	Contínua	Anderson–Darling
Transversal	1 variável	–	–	Contínua	Jarque–Bera
Transversal	≥ 2 variáveis	–	–	Contínuas multiplas	Teste de Mardia (assimetria e curtose)
Transversal	≥ 2 variáveis	–	–	Contínuas multiplas	Henze–Zirkler
Transversal	≥ 2 variáveis	–	–	Contínuas multiplas	Royston (Shapiro–Wilk multivariado)
Transversal	1 fator + 1 resposta	≥ 2	Independentes	Contínua categórica	Levene

Continuado na próxima página

Tabela 25.10: Tabela de seleção de testes estatísticos: Diagnóstico. (Continuado)

Delineamento do estudo	Qtd. de variáveis / fatores	Níveis do fator	Relação entre amostras	Tipo de variáveis	Teste estatístico
Transversal	1 fator + 1 resposta	≥ 2	Independentes	Contínua categórica	Brown–Forsythe
Transversal	1 fator + 1 resposta	≥ 2	Independentes	Contínua categórica	Bartlett (quando normalidade é plausível)

Tabela 25.11: Tabela de seleção de testes estatísticos: Transição de estados.

Delineamento do estudo	Qtd. de variáveis / fatores	Níveis do fator	Relação entre amostras	Tipo de variáveis	Teste estatístico
Longitudinal	≥ 2 estados ao longo do tempo	≥ 2	Dependentes (medidas repetidas)	Categórica (estados)	Modelo de Markov (tempo discreto)
Longitudinal	Estados + covariáveis	≥ 2	Dependentes (medidas repetidas)	Estados categóricos + preditores	Modelo de Markov com covariáveis / HMM
Longitudinal	Estados + tempo contínuo	≥ 2	Dependentes (medidas repetidas)	Estados categóricos ao longo do tempo	Modelo de Markov em tempo contínuo (CTMC)

RASCUNHO

Capítulo 26

Análise inferencial

26.1 Raciocínio inferencial

26.1.1 O que é análise inferencial?

- Na análise inferencial são utilizados dados da(s) amostra(s) para fazer uma inferência válida (isto é, estimativa) sobre os parâmetros populacionais desconhecidos.¹⁰⁸
- No paradigma de Jerzy Neyman e Egon Pearson, um teste de hipótese científica envolve a tomada de decisão sobre hipóteses nulas (H_0) e alternativa (H_1) concorrentes e mutuamente exclusivas.²⁶¹

26.1.2 Quais são os tipos de raciocínio inferencial?

- Inferência dedutiva: Uma dada hipótese inicial é utilizada para prever o que seria observado caso tal hipótese fosse verdadeira.²⁶²
- Inferência induutiva: Com base nos dados observados, avalia-se qual hipótese é mais defensável (isto é, mais provável).²⁶²

26.1.3 Quais são as questões fundamentais da análise inferencial?

- A direção do efeito²⁶³
- A magnitude do efeito²⁶³
- A importância do efeito²⁶³

26.2 Hipóteses científicas

26.2.1 O que é hipótese científica?

- Hipótese científica é uma ideia que pode ser testada.²⁶¹
- Definir claramente os problemas e os objetivos da pesquisa são o ponto de partida de todos os estudos científicos.¹³³
- Além do papel técnico, os testes de hipótese carregam uma dimensão interpretativa que molda como os pesquisadores comunicam descobertas. Estudos recentes destacam o caráter pragmático e dicotômico dessas decisões.²⁶⁴

26.2.2 Quais são as fontes de ideias para gerar hipóteses científicas?

- Revisão das práticas atuais.²⁶⁵
- Desafio a ideias aceitas.²⁶⁵
- Conflito entre ideias divergentes.²⁶⁵
- Variações regionais, temporais e populacionais.²⁶⁵
- Experiências dos próprios pesquisadores.²⁶⁵
- Imaginação sem fronteiras ou limites convencionais.²⁶⁵

26.3 Hipóteses estatísticas

26.3.1 O que é hipótese nula?

- A hipótese nula (H_0) é uma expressão que representa o estado atual do conhecimento (*status quo*), em geral a não existência de um determinado efeito.¹⁷⁹

26.3.2 O que é hipótese alternativa?

- A hipótese alternativa (H_1) é uma expressão que contém as situações que serão testadas, de modo que um resultado positivo indique alguma ação a ser conduzida.¹⁷⁹

26.3.3 Qual hipótese está sendo testada?

- A hipótese nula (H_0) é a hipótese sob teste em análises inferenciais.¹⁰⁹
- Pode-se concluir sobre rejeitar ou não rejeitar a hipótese nula (H_0).¹⁰⁹
- Não se conclui sobre a hipótese alternativa (H_1).¹⁷⁹
- Para testar a hipótese nula, deve-se selecionar o nível de significância crítica (P-valor de corte); a probabilidade de rejeitarmos uma hipótese nula verdadeira (α); e a probabilidade de não rejeitarmos uma hipótese nula falsa (β).²⁶¹

26.4 Testes de hipóteses

26.4.1 Quais são os tipos de teste de hipóteses?

- Teste (clássico) de significância da hipótese nula: verifica evidência contra H_0 usando P-valor.²⁶⁶
- Teste de mínimos efeitos (MOTE/MOST/SESOI): testa se o efeito é pelo menos tão grande quanto um limiar de relevância (SESOI). Rejeitar H_0 sugere efeito grande o suficiente.²⁶⁶
- Teste de equivalência (TOST): testa se o efeito está dentro de uma margem de equivalência clinicamente irrelevante (entre Δ e $-\Delta$). Rejeitar H_0 sugere equivalência prática.²⁶⁶
- Teste de superioridade: avalia se um tratamento/intervenção supera o controle por uma margem > 0 ou $> \Delta$.²⁶⁶
- Teste de não-inferioridade: avalia se o tratamento não é pior que o controle por mais do que $-\Delta$.?
- Teste de inferioridade: avalia se o tratamento é pior que o controle (por exemplo, para checar segurança).?

26.4.2 O que reportar após um teste de hipótese?

- P-valores, como estimativa da significância estatística.²⁶⁷
- Tamanho do efeito, como estimativa de significância substantiva.²⁶⁷

26.5 Intervalos de confiança e raciocínio de longo prazo

26.5.1 O que é um intervalo de confiança?

- Um intervalo de confiança é um procedimento inferencial utilizado para estimar um parâmetro populacional desconhecido a partir de dados amostrais, levando em conta a variabilidade inerente ao processo de amostragem.²⁶⁸
- Diferentemente de uma estatística descritiva, o intervalo de confiança não é uma propriedade fixa do parâmetro, mas uma propriedade do procedimento estatístico utilizado para estimá-lo.²⁶⁹
- O parâmetro populacional é considerado fixo (embora desconhecido), enquanto o intervalo de confiança é uma quantidade aleatória, pois depende da amostra observada.²⁷⁰

26.5.2 Intervalos de confiança e a lógica frequentista

- No paradigma frequentista, a probabilidade é interpretada como uma frequência relativa observável no longo prazo, associada a um processo repetível.²⁶⁸
- Um intervalo de confiança de nível $(1 - \alpha)$ é construído de modo que, se o mesmo procedimento de amostragem e análise fosse repetido indefinidamente sob as mesmas condições, aproximadamente $(1 - \alpha) \times 100\%$ dos intervalos assim gerados conteriam o verdadeiro valor do parâmetro populacional.²⁶⁹
- Portanto, a probabilidade associada ao intervalo de confiança refere-se ao desempenho do procedimento no longo prazo, e não à probabilidade de o parâmetro estar contido em um intervalo específico observado.

26.5.3 O que um intervalo de confiança não representa

- Um intervalo de confiança não deve ser interpretado como a probabilidade de que o parâmetro esteja contido naquele intervalo específico.²⁷⁰
- Após os dados terem sido observados e o intervalo calculado, o parâmetro populacional ou está dentro do intervalo ou não está; não há incerteza probabilística sobre isso no sentido frequentista.²⁷⁰
- A incerteza expressa pelo intervalo de confiança refere-se à variabilidade do processo inferencial, e não a uma distribuição de probabilidade do parâmetro.²⁷⁰

26.5.4 Relação entre intervalos de confiança e testes de hipóteses

- Intervalos de confiança e testes de hipótese frequentistas são derivados do mesmo modelo probabilístico subjacente e utilizam as mesmas suposições estatísticas.²⁷¹
- Em testes bicaudais, qualquer valor hipotético do parâmetro que esteja fora do intervalo de confiança de nível $(1 - \alpha)$ corresponde a uma hipótese nula que seria rejeitada ao nível de significância α .²⁷¹
- De forma análoga, valores do parâmetro que estejam dentro do intervalo de confiança correspondem a hipóteses nulas para as quais não haveria evidência suficiente para rejeição ao nível α .²⁷¹
- Apesar dessa equivalência formal, intervalos de confiança e testes de hipótese respondem a perguntas distintas: testes enfatizam decisões dicotômicas, enquanto intervalos de confiança enfatizam estimação e incerteza.²⁷¹

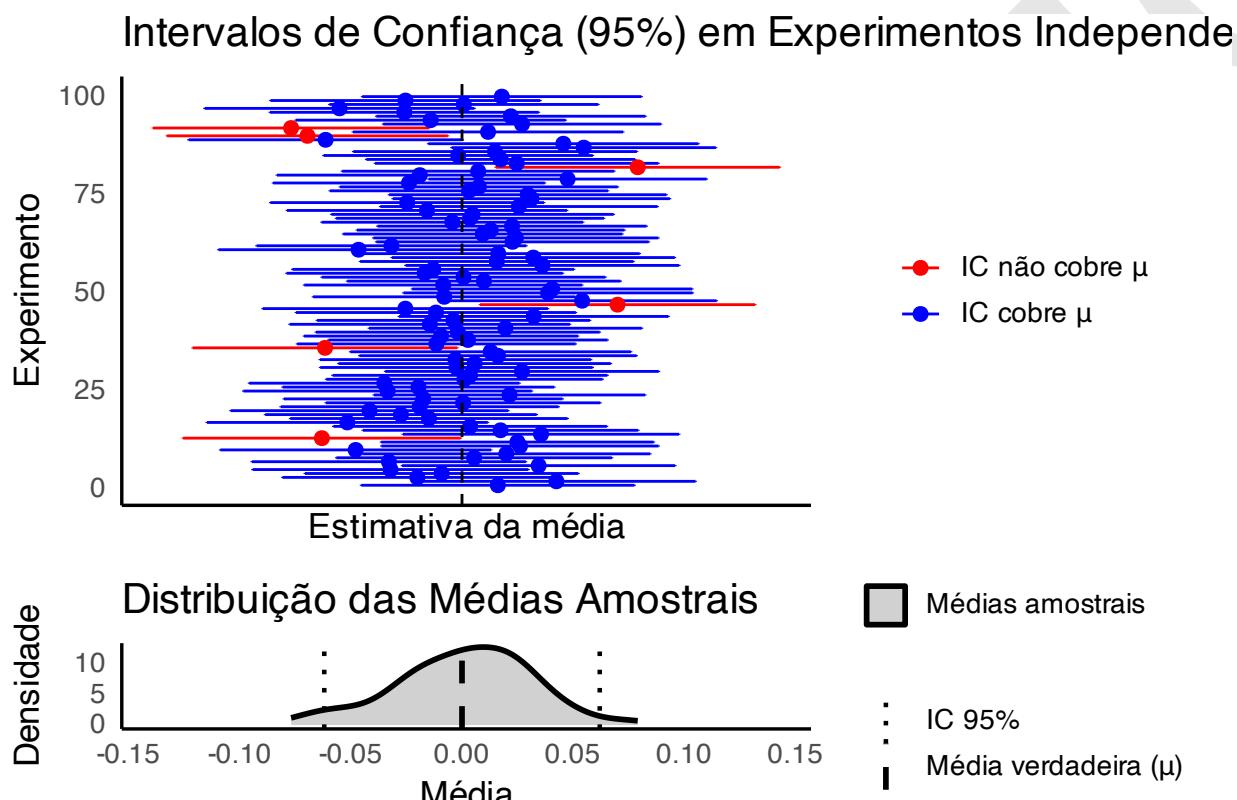


Figura 26.1: Simulação ilustrativa de intervalos de confiança (IC) em 100 experimentos independentes, cada um com 1.000 observações amostradas de uma população normal padrão (média = 0, desvio-padrão = 1). Os ICs são construídos no nível de 95%. O gráfico superior mostra os ICs individuais para cada experimento, indicando se o IC cobre ou não a média verdadeira ($\mu = 0$). O gráfico inferior apresenta a distribuição das médias amostrais obtidas nos experimentos, juntamente com o IC teórico para a média populacional. Observe que aproximadamente 95% dos ICs individuais cobrem a média verdadeira, ilustrando o conceito de cobertura no longo prazo associado aos intervalos de confiança.

26.5.5 Por que intervalos de confiança são centrais na inferência científica

- Intervalos de confiança permitem avaliar simultaneamente a magnitude do efeito e a incerteza associada à sua estimativa, na mesma unidade de medida da variável de interesse.²⁷¹
- Essa abordagem favorece interpretações substantivas e científicas dos resultados, em oposição a decisões puramente dicotômicas baseadas em pontos de corte arbitrários.²⁷¹
- Quando corretamente interpretados, intervalos de confiança promovem uma comunicação mais informativa da evidência científica e reduzem interpretações equivocadas associadas ao uso exclusivo de P-valores.²⁷⁰

26.6 Comparações múltiplas

26.6.1 O que é uma família de hipóteses?

- Família de hipóteses é um conjunto de comparações/inferências que, por desenho ou análise, devem ser consideradas juntas para controle do erro tipo I global (ex.: todas as comparações de um desfecho primário, todos os subgrupos pré-especificados, todos os desfechos coprimários).²
- O controle do *family-wise error rate* (FWER) ou do *false discovery rate* (FDR) deve considerar a família pertinente, não comparações isoladas.

26.6.2 O que são testes *ad hoc* e *post hoc*?

- *Ad hoc*: análises/décisões não planejadas a priori, motivadas por inspeção dos dados. Úteis para gerar hipóteses, não para confirmá-las.²
- *Post hoc*: procedimentos de comparações múltiplas aplicados após um teste global ter indicado efeito significativo. Visam controlar o erro tipo I em múltiplas comparações.²

26.6.3 Como ajustar a análise inferencial para hipóteses múltiplas?

- Defina a família (o que entra no ajuste) e priorize desfechos (primário, coprimários, secundários).²
- Aplique métodos de controle FWER (Bonferroni, Holm, Hochberg, Dunnett para múltiplos vs. controle) ou controle FDR (Benjamini–Hochberg) conforme o objetivo (confirmação vs. exploração).²
- Em planos confirmatórios, use hierarquização/gatekeeping: testa-se em sequência; a alocação de α passa adiante apenas se houver significância no nível anterior.²



O pacote *stats*¹³¹ fornece a função *p.adjust*^a para ajustar o P-valor utilizando diversos métodos.

^a<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/p.adjust>

26.6.4 O que são testes unicaudais e bicaudais?

- Teste unicaudal procura evidência em uma direção específica (ex.: “maior que 0”). Toda a região crítica está numa só cauda; tem maior poder para aquela direção, mas não detecta sinal oposto.²
- Teste bicaudal procura evidência em qualquer direção (ex.: “diferente de 0”). Divide α em duas caudas (direita e esquerda). É a escolha padrão quando ambas as direções são plausíveis.²

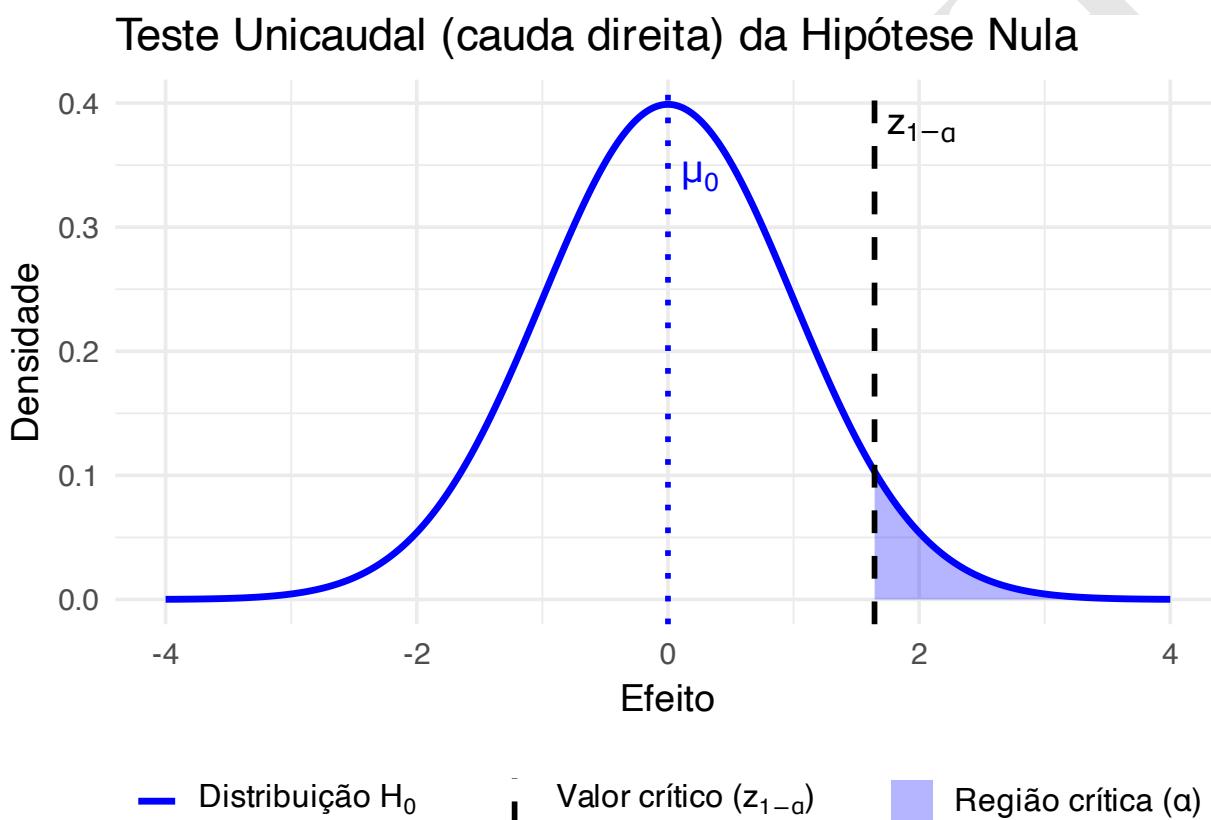


Figura 26.2: Representação gráfica de um teste de hipótese unicaudal à direita, aplicado quando se busca evidência de efeitos positivos (valores significativamente maiores que o esperado sob H_0).

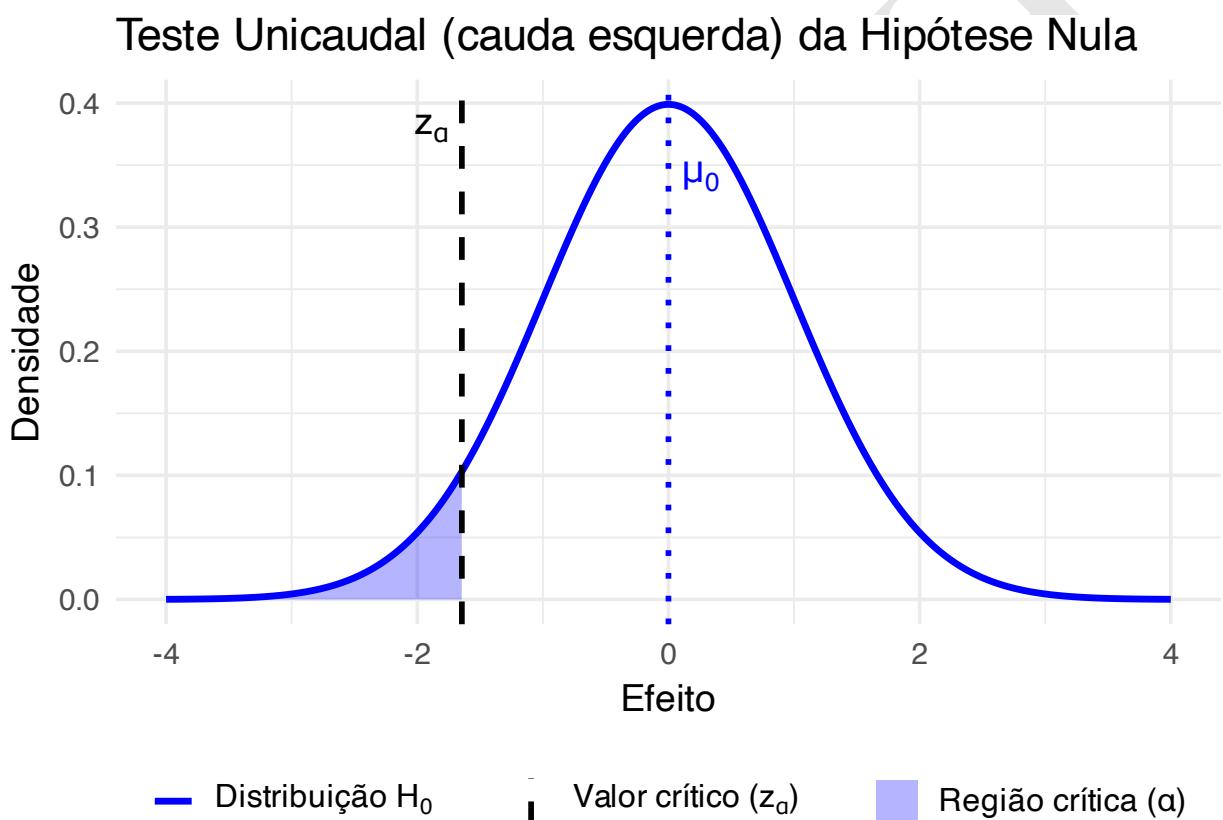


Figura 26.3: Representação gráfica de um teste de hipótese unicaudal à esquerda, aplicado quando se busca evidência de efeitos negativos (valores significativamente menores que o esperado sob H_0).

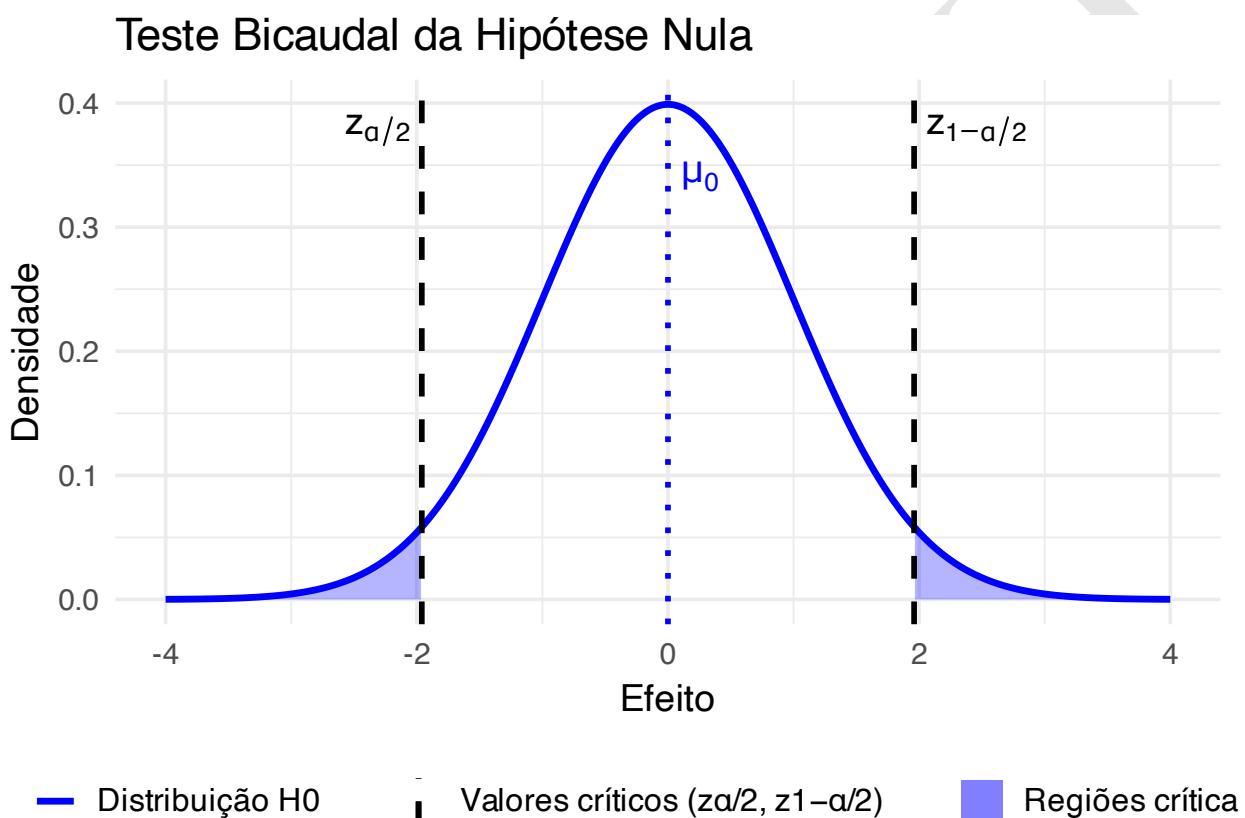


Figura 26.4: Representação gráfica de um teste de hipótese bicaudal, aplicado quando se busca evidência de efeitos positivos ou negativos (valores significativamente diferentes do esperado sob H_0).

26.7 Inferência visual

26.7.1 O que é inferência visual?

- Inferência visual consiste na interpretação de dados apresentados em gráficos.²⁷¹
- Para inferência visual, recomenda-se a apresentação dos dados em gráficos com estimativas de tendência central e seu intervalo (preferencialmente intervalo de confiança no nível de significância α pré-estabelecido).²⁷¹

26.7.2 Por que usar intervalos de confiança para inferência visual?

- Intervalos de confiança fornecem estimativas pontuais e intervalares na mesma unidade de medida da variável.²⁷¹
- Existe uma relação entre o intervalo de confiança e o P-valor obtido pelo teste de significância de hipótese nula H_0 , em que ambos consideram o mesmo nível de significância α pré-estabelecido.²⁷¹

26.7.3 Como interpretar intervalos de confiança em uma figura?

- Identifique o que as tendências centrais e as barras de erro representam. Qual é a variável dependente? É expressa em unidades originais ou é padronizada? A figura mostra intervalos de confiança, erro-padrão ou desvio-padrão? Qual é o desenho experimental?²⁷¹
- Faça uma interpretação substantiva dos valores de tendência central e dos intervalos de confiança.²⁷¹
- O intervalo de confiança é uma faixa de valores plausíveis para a tendência central. Valores fora do intervalo são relativamente implausíveis, no nível de significância α pré-estabelecido.²⁷¹
- Qualquer valor fora do intervalo de confiança, quando considerado como hipótese nula (H_0), equivale a $P < \alpha$ pré-estabelecido (bicaudal).²⁷¹
- Qualquer valor dentro do intervalo, quando considerado como hipótese nula (H_0), equivale a $P > \alpha$ pré-estabelecido (bicaudal).²⁷¹

26.8 Interpretação de análise inferencial

26.8.1 Como interpretar uma análise inferencial?

- Testes de hipótese nula (H_0) vs. alternativa (H_1) a partir de um nível de significância (α) pré-especificado.²⁶⁹
- P-valor como evidência estatística sobre (H_0).²⁶⁹
- Estimação de intervalos de confiança de um nível de significância (α) pré-especificado bicaudal ($IC_{1-\alpha/2}$) ou unicaudal ($IC_{1-\alpha}$).²⁶⁹
- Análise Bayesiana.²⁶⁹

26.8.2 O que são resultados positivos” e “negativos” (inconclusivos) em teste de hipótese?

- Resultados “positivos” compreendem um P-valor dentro da zona crítica estatisticamente significativa (ex.: $P < 0,05$ ou outro ponto de corte) e sugerem que os autores rejeitem a hipótese nula (H_0), confirmando assim sua hipótese científica.²⁷²

Tabela 26.1: Tabela de erros tipos I e II de inferência estatística.

	Hipótese nula H_0 é falsa	Hipótese nula H_0 é verdadeira
Hipótese nula H_0 foi rejeitada	Decisão correta	Decisão incorreta (erro tipo I)
Hipótese nula H_0 não foi rejeitada	Decisão incorreta (erro tipo II)	Decisão correta

- Resultados “negativos” ou inconclusivos compreendem um P-valor fora da zona crítica estatisticamente significativa (ex.: $P \geq 0,05$ ou outro ponto de corte) e sugerem que os autores não rejeitem a hipótese nula (H_0) porque o efeito observado é nulo (logo, “negativo”), ou porque o estudo não possui poder suficiente para detectá-lo, não permitindo portanto afirmar a hipótese científica (logo, inconclusivo).²⁷²

26.8.3 Qual a importância de resultados “negativos”?

- Conhecer resultados negativos contribui com uma visão mais ampla do campo de estudo junto aos resultados positivos.²⁷³
- Resultados negativos permitem um melhor planejamento das pesquisas futuras e pode aumentar suas chances de sucesso.²⁷³

26.8.4 Resultados inconclusivos: Ausência de evidência ou evidência de ausência?

- Em estudos (geralmente com amostras grandes), resultados estatisticamente significativos (com P-valores menores do limiar pré-estabelecido, $P < \alpha$) podem não ser clinicamente relevantes.²⁷⁴
- Em estudos (geralmente com amostras pequenas), resultados estatisticamente não significativos (com P-valores iguais ou maiores do limiar pré-estabelecido, $P \geq \alpha$) não devem ser interpretados como evidência de inexistência do efeito.²⁷⁴
- Geralmente é razoável aceitar uma nova conclusão apenas quando há dados a seu favor (“resultados positivos”). Também é razoável questionar se apenas a ausência de dados a seu favor (“resultados negativos”) justifica suficientemente a rejeição de tal conclusão.²⁷⁴
- A prática estatística convencional tende a reduzir a incerteza científica a decisões docotômicas. Essa simplificação possui implicações epistemológicas importantes: os testes de hipótese produzem não apenas juízos empíricos, mas também atos pragmáticos que comunicam graus de confiança e orientam ações.²⁶⁴

26.9 Erros de inferência I, II, S e M

26.9.1 O que são erros de inferência estatística?

- Um erro de inferência é a tomada de decisão incorreta, seja a favor ou contra a hipótese nula (H_0).²⁶¹

26.9.2 O que são erros Tipo I e Tipo II?

- Erro Tipo I significa a rejeição de uma hipótese nula (H_0) quando esta é verdadeira.²⁶¹
- Erro Tipo II significa a não rejeição de uma hipótese nula (H_0) quando esta é falsa.²⁶¹

26.9.3 O que são erros Tipo S e Tipo M?

- Erro Tipo S (do inglês *sign*) significa a identificação errônea da direção — positiva ou negativa — do efeito observado.^{275,276}

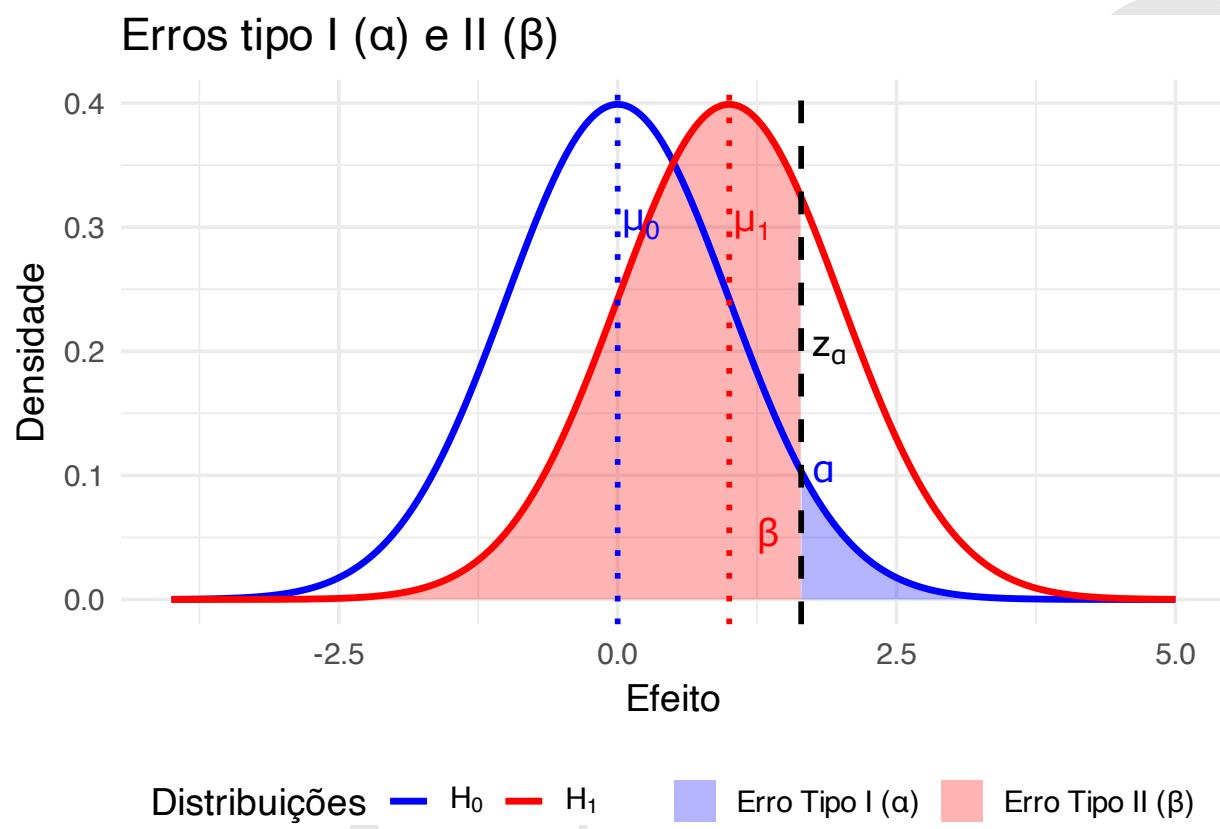


Figura 26.5: Representação gráfica dos erros tipo I e tipo II em um teste de hipótese (bicaudal).

Tabela 26.2: Tabela de erro tipo S de inferência estatística.

	Sinal positivo	Sinal negativo
Sinal positivo	Decisão correta	Decisão incorreta (erro tipo S)
Sinal negativo	Decisão incorreta (erro tipo S)	Decisão correta

Distribuição dos p-valores com H_0 verdadeira

Número de testes com $p < 0,05$ (erro tipo I): 5 de 100

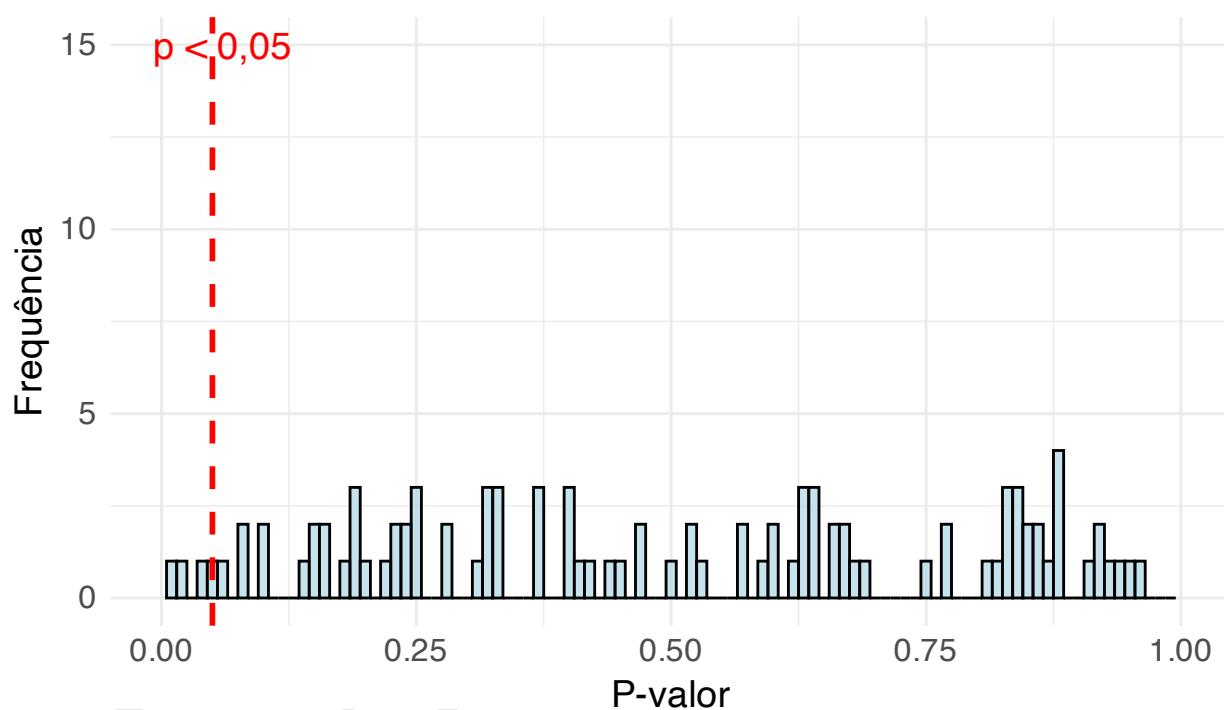


Figura 26.6: Erro tipo I: Distribuição dos p-valores em 100 testes de hipótese de amostras aleatórias de tamanho 30. A linha vermelha pontilhada indica o nível de significância estatística de 0,05.

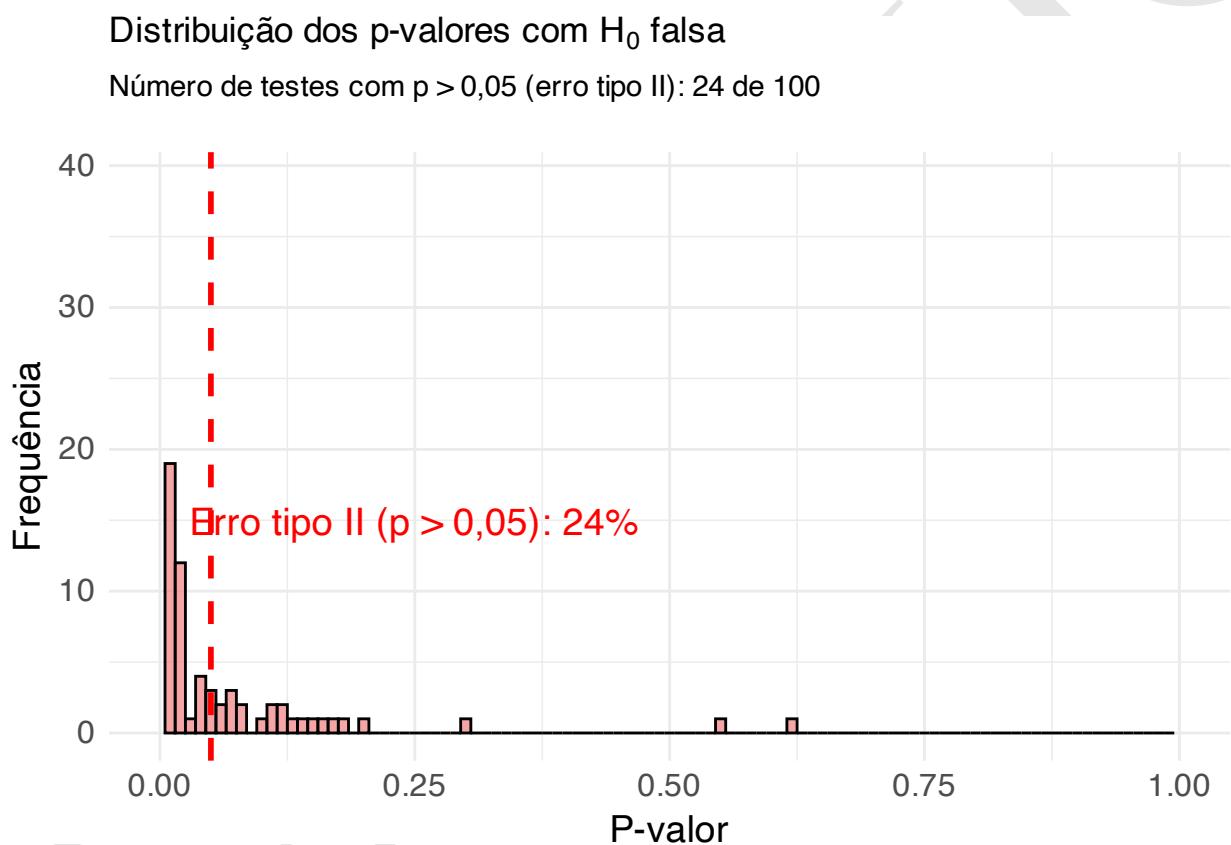


Figura 26.7: Erro tipo II: Distribuição dos p-valores em 100 testes de hipótese de amostras aleatórias de tamanho 10. A linha vermelha pontilhada indica o nível de significância estatística de 0,05.

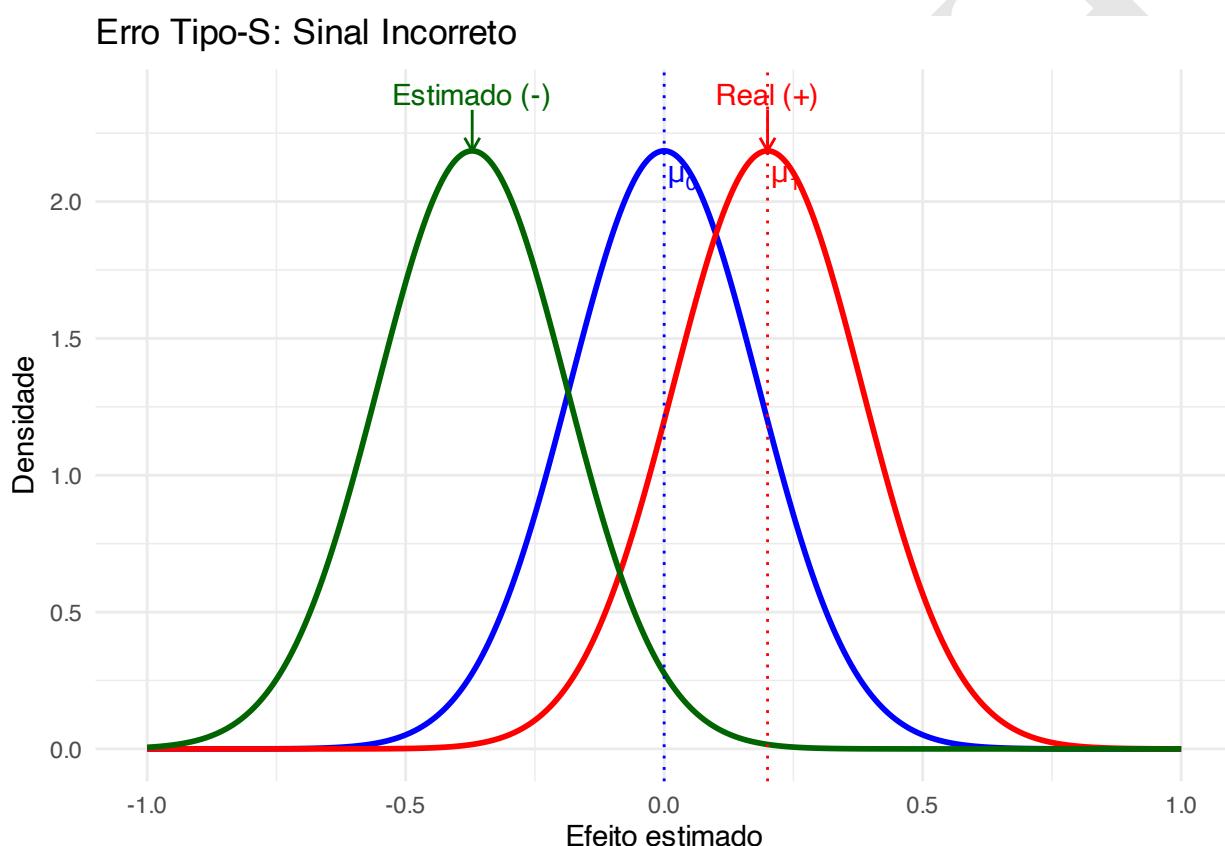


Figura 26.8: Representação gráfica do erro tipo S (sinal) em um teste de hipótese (bicaudal).

Tabela 26.3: Tabela de erro tipo M de inferência estatística.

		Magnitude alta	Magnitude baixa
Magnitude alta	Decisão correta	Decisão incorreta (erro tipo M)	
Magnitude baixa	Decisão incorreta (erro tipo M)	Decisão correta	

- Erro Tipo M (do inglês *magnitude*) significa a identificação errônea — em geral, exagerada — da magnitude do efeito observado.^{275,276}

Erro Tipo-M: Superestimação da Magnitude

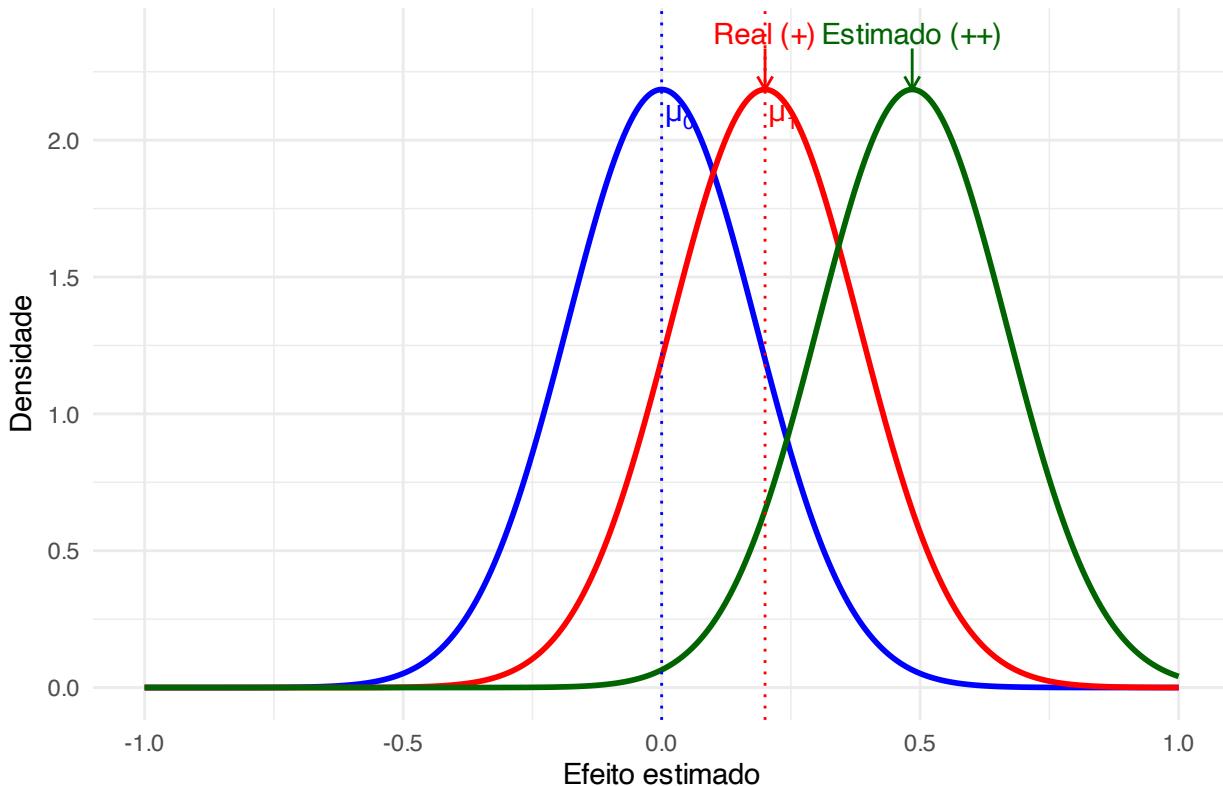


Figura 26.9: Representação gráfica do erro tipo M (magnitude) em um teste de hipótese (bicaudal).

RASCUNHO

Capítulo 27

Tamanho do efeito e P-valor

27.1 Tamanho do efeito

27.1.1 O que é o tamanho do efeito?

- Tamanho do efeito quantifica a magnitude de um efeito real da análise, expressando uma importância descriptiva dos resultados.²⁷⁷

27.2 Tipos de tamanho do efeito

- Diferenças padronizadas entre grupos:^{267,277}
 - Cohen's d
 - Glass's Δ
 - Razão de chances (RC ou OR)
 - Risco relativo ou razão de risco (RR)

R O pacote *epitools*²⁷⁸ fornece a função *oddsratio.wald*^a para calcular a razão de chances.

^a<https://www.rdocumentation.org/packages/epitools/versions/0.09/topics/odds.ratio>

R O pacote *epitools*²⁷⁸ fornece a função *riskratio.wald*^a para calcular a razão de risco.

^a<https://www.rdocumentation.org/packages/epitools/versions/0.09/topics/riskratio.wald>

- Medidas de associação:^{267,277}
 - Coeficiente de correlação de Pearson (r), ponto-bisserial (r_s), Spearman (ρ), Kendall (τ), Cramér (V) e ϕ .
 - Coeficiente de determinação (R^2)

27.2.1 Como interpretar um tamanho do efeito?

- Tamanhos de efeito podem ser comparadores entre diferentes estudos.²⁶⁷

R

O pacote *effectsize*²⁷⁹ fornece a função *rules*^a para criar regras de interpretação de tamanhos de efeito.

^a<https://www.rdocumentation.org/packages/effectsize/versions/0.8.3/topics/rules>

R

O pacote *effectsize*²⁷⁹ fornece a função *interpret*^a para interpretar os tamanhos de efeito com base em uma lista de regras pré-definidas.

^a<https://www.rdocumentation.org/packages/effectsize/versions/0.8.3/topics/interpret>

R

O pacote *pwr*²⁸⁰ fornece a função *cohen.ES*^a para obter os tamanhos de efeito “pequeno”, “médio” e “grande” para diversos testes de hipóteses.

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/cohen.ES>

27.2.2 O que é a diferença de média bruta?

- A diferença de média bruta representa a diferença absoluta entre as médias de dois grupos, expressa na unidade original da variável.[?]
- Trata-se de uma medida não padronizada, sendo particularmente útil quando a escala possui interpretação clínica ou substantiva direta (por exemplo, mmHg, pontos de escore).[?]
- Por depender da unidade de medida, não permite comparações diretas entre estudos com métricas diferentes.[?]

27.2.3 Correlações podem ser consideradas tamanhos de efeito?

- Sim. Coeficientes de correlação podem ser interpretados diretamente como tamanhos de efeito, pois expressam a força e a direção da associação entre variáveis.[?]
- O coeficiente de Spearman (ρ) mede associações monotônicas e é robusto a violações de normalidade.[?]
- Kendall (τ) é especialmente indicado para amostras pequenas ou dados com empates.[?]
- Esses coeficientes são frequentemente utilizados como tamanhos de efeito em testes não paramétricos.[?]

27.2.4 O que é o q de Cohen?

- O tamanho de efeito q quantifica a diferença entre dois coeficientes de correlação, após transformação de Fisher (z).[?]
- É utilizado principalmente para comparar associações observadas em grupos independentes.[?]
- Cohen propôs valores de referência para interpretação (pequeno, médio e grande), reforçando seu caráter descritivo.[?]

27.2.5 O que o g no teste do sinal?

- O coeficiente g é utilizado como tamanho de efeito no teste do sinal.[?]
- Representa a diferença padronizada entre a proporção de observações positivas e negativas.[?]
- Aplica-se a delineamentos pareados em que apenas a direção do efeito é considerada.[?]

- É uma medida robusta, porém menos informativa do que medidas baseadas em magnitude contínua.[?]

27.2.6 O que é o h de Cohen?

- O tamanho de efeito h mede a diferença entre duas proporções, após transformação angular para estabilizar a variância.[?]
- É indicado para comparações entre desfechos binários.[?]
- Por ser padronizado, permite comparações entre estudos com diferentes proporções absolutas.[?]

27.2.7 O que representa o tamanho de efeito w ?

- O coeficiente w é utilizado como tamanho de efeito em testes do qui-quadrado, tanto de aderência quanto de independência.[?]
- Quantifica o grau global de discrepância entre frequências observadas e esperadas.[?]
- Sua interpretação depende do número de categorias e do tamanho da amostra, devendo ser feita com cautela.[?]

27.2.8 O que é o tamanho de efeito f em ANOVA?

- O coeficiente f é utilizado como tamanho de efeito em análises de variância (ANOVA).[?]
- Está relacionado à proporção da variância explicada pelo fator em relação à variância residual.[?]
- É amplamente empregado em cálculos de poder estatístico e no planejamento amostral.[?]

27.2.9 O que é o tamanho de efeito f^2 em regressão?

- O coeficiente f^2 mede o impacto incremental de um conjunto de preditores em modelos de regressão.[?]
- É definido como a razão entre a variância explicada adicional e a variância não explicada.[?]
- É particularmente útil para avaliar contribuições locais em modelos hierárquicos ou multivariados.[?]

27.2.10 O que é a estatística Λ de Wilks na MANOVA?

- O Lambda de Wilks (Λ) é utilizado como estatística global em análises multivariadas de variância (MANOVA).[?]
- Representa a proporção da variância multivariada não explicada pelo modelo.[?]
- Embora menos intuitivo como medida de magnitude, é amplamente utilizado e pode ser convertido em outras estatísticas.[?]

27.2.11 Como escolher o tamanho de efeito adequado?

- Não existe um tamanho de efeito universalmente superior; a escolha depende da pergunta científica, do delineamento, do tipo de variável e do modelo estatístico.[?]
- A boa prática estatística recomenda reportar estimativa pontual, intervalo de confiança e tamanho de efeito, evitando decisões baseadas exclusivamente em significância estatística.[?]
- Sempre que possível, a interpretação deve considerar relevância prática, contexto científico e incerteza associada.[?]

27.3 Conversão entre tamanhos do efeito

27.3.1 Como converter um tamanho de efeito em outro?

- ²⁷⁷



O pacote *effectsize*²⁷⁹ fornece diversas funções para conversão de diferentes estimativas de tamanhos de efeito.

27.4 Efeitos bruto e padronizado

27.4.1 O que é efeito bruto?

- ²⁸¹
- ²⁸²

27.4.2 O que é efeito padronizado?

- ²⁸¹
- ²⁸²

27.5 P-valor

27.5.1 O que é significância estatística?

- A expressão “significância estatística”²⁸³ ou “evidência estatística de significância” sugere apenas que um experimento merece ser repetido, uma vez que um baixo P-valor (calculado a partir dos dados, modelos e demais suposições do estudo) sugere ser improvável que os dados coletados sejam coletados no contexto de que a hipótese nula (H_0) assumida é verdadeira.²⁸⁴

27.5.2 Como justificar o nível de significância estatística de um teste?

- ?



O pacote *Superpower*²⁸⁵ fornece a função *optimal_alpha*^a para calcular e justificar o nível de significância α por balanço dos erros tipo I e II.

^ahttps://www.rdocumentation.org/packages/Superpower/versions/0.2.0/topics/optimal_alpha



O pacote *Superpower*²⁸⁵ fornece a função *ANOVA_compromise*^a para calcular e justificar o nível de significância α por balanço dos erros tipo I e II em análise de variância (ANOVA).

^ahttps://www.rdocumentation.org/packages/Superpower/versions/0.2.0/topics/ANOVA_compromise

27.5.3 O que é o P-valor?

- P-valor é a probabilidade, assumindo-se um dado modelo estatístico, de que um efeito calculado a partir dos dados seria igual ou mais extremo do que o seu valor observado.²⁸⁶

- P-valor é uma variável aleatória que possui distribuição uniforme quando a hipótese nula (H_0) é verdadeira.²⁸⁷

27.5.4 Como interpretar o P-valor?

- P-valores abaixo de um nível de significância estatística pré-especificado representam que um experimento merece ser repetido, com a rejeição da hipótese nula (H_0) justificada apenas quando experimentos adicionais frequentemente reportem igualmente resultados positivos (rejeição da hipótese nula (H_0)).²⁶⁹
- P-valor resulta da coleta e análise de dados, e assim quantifica a plausibilidade dos dados observados sob a hipótese nula (H_0).²⁸⁸
- P-valores podem indicar quantitativamente a incompatibilidade entre os dados obtidos e o modelo estatístico especificado a priori (geralmente constituído pela hipótese nula (H_0)).²⁸⁶
- P-valores menores/maiores do que o nível de significância estatístico pré-estabelecido não devem ser utilizados como única fonte de informação para tomada de decisão em ciência.²⁸⁶

27.5.5 O que o P-valor não é?

- P-valor não representa a probabilidade de que a hipótese nula (H_0) seja verdadeira, nem a probabilidade de que os dados tenham sido produzidos pelo acaso.²⁸⁶
- P-valor não mede o tamanho do efeito ou a relevância da sua observação.²⁸⁶
- P-valor sozinho não provê informação suficiente sobre a evidência sobre um modelo teórico. A sua interpretação correta requer uma descrição ampla sobre o delineamento, métodos e análises estatísticas aplicados no estudo.²⁸⁶
- Evidência estatística de significância não provê informação sobre a magnitude do efeito observado e não necessariamente implica que o efeito é robusto.^{193,287}

27.5.6 Qual a origem do ‘P<0,05’?

- A origem do P<0,05 remonta aos trabalhos de R. A. Fisher nas décadas de 1920 e 1930. Fisher introduziu o conceito de P-valor dentro de uma abordagem frequentista de inferência estatística.²⁶⁹
- O P<0,05 foi sugerido por Ronald A. Fisher como um limiar prático para indicar que um resultado era “estatisticamente significativo”.²⁶⁹
- Para Ronald A. Fisher, a significância estatística não era prova definitiva, mas um sinal de que o resultado merecia investigação adicional. A rejeição da hipótese nula só deveria ocorrer após repetidas observações significativas, e não com base em um único teste.²⁶⁹

27.5.7 Quais são os complementos ou alternativas ao P-valor?

- Intervalos de confiança, credibilidade ou predição.²⁸⁶
- Razão de verossimilhança.²⁸⁶
- Métodos Bayesianos, fator Bayes.²⁸⁶

27.6 P-valor de 2^a geração

27.6.1 O que é o P-valor de 2^a geração?

- O P-valor de 2^a geração (SGPV) resume a fração das hipóteses apoiadas pelos dados que também pertencem à hipótese nula intervalar (intervalo de equivalência previamente especificado). Quantifica quanto do inter-

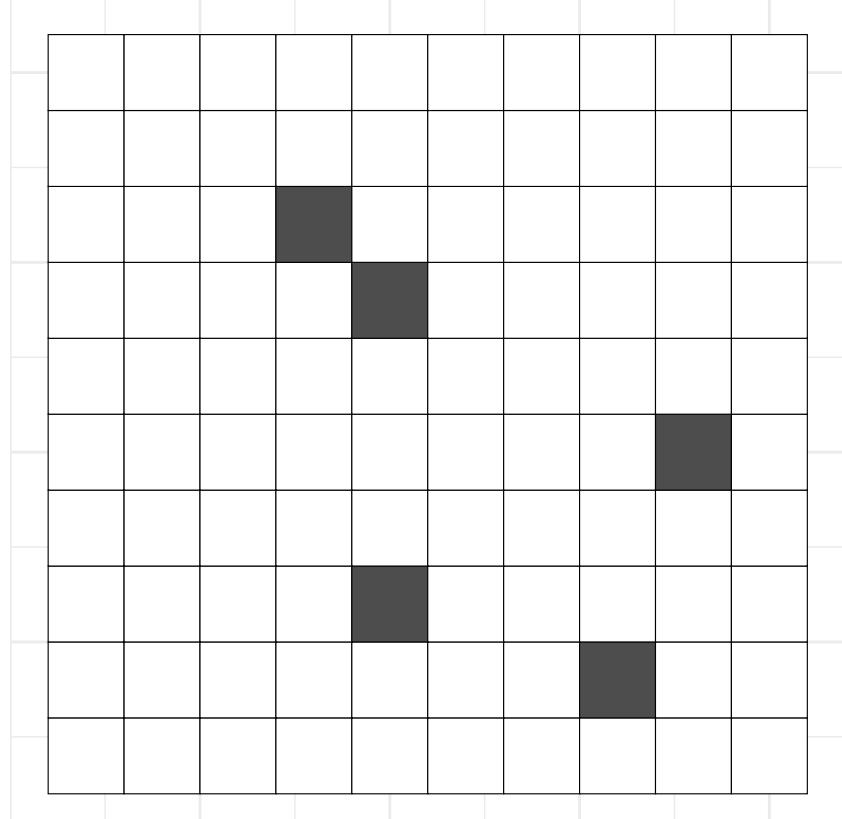


Figura 27.1: Visualização espacial de $p < 0,05$ (5 quadrados aleatórios em 100).

Tabela 27.1: Comparação entre p-valor (bicaudal, inferido do IC 95%) e SGPV (p_δ) nos cenários simulados.

Cenário	a	b	H_0^-	H_0^+	$\hat{\theta}$	SE	p-valor (bicaudal)	p_δ	Conclusão (SGPV)
1	0.350	0.550	-0.100	0.100	0.450	0.0510	<0,001	0.000	Apoia alternativas (SGPV=0)
2	-0.050	0.080	-0.100	0.100	0.015	0.0332	0.651	1.000	Equivalência (SGPV=1)
3	-0.500	0.700	-0.100	0.100	0.100	0.3061	0.744	0.500	Inconclusivo ($0 < p_\delta < 1$)
4	0.050	0.250	-0.100	0.100	0.150	0.0510	0.003	0.250	Inconclusivo ($0 < p_\delta < 1$)
5	-0.250	-0.050	-0.100	0.100	-0.150	0.0510	0.003	0.250	Inconclusivo ($0 < p_\delta < 1$)
6	0.150	0.550	-0.100	0.100	0.350	0.1020	0.001	0.000	Apoia alternativas (SGPV=0)
7	-0.550	-0.150	-0.100	0.100	-0.350	0.1020	0.001	0.000	Apoia alternativas (SGPV=0)

valo de estimativa (p.ex., IC95%) recai dentro da zona de indiferença científica/clinicamente irrelevante.²⁸⁹

- Essa abordagem exige declarar a hipótese nula como intervalo (e não um ponto), incorporando o que é considerado “efeito sem relevância prática” segundo o contexto científico (precisão de medida, relevância clínica etc.).²⁸⁹

27.6.2 Como definir a hipótese nula intervalar e δ ?

- Especifique H_0 como um intervalo de equivalência $[H_0^-, H_0^+]$ que contém efeitos considerados praticamente nulos. Defina δ como a meia-largura do intervalo de equivalência ($\delta = (H_0^+ - H_0^-)/2$).²⁸⁹
- A escolha deve ser a priori e justificada por critérios científicos (p.ex., MCID, precisão de medida).²⁸⁹

27.6.3 Como calcular o SGPV?

- Seja $I = [a, b]$ o intervalo apoiado pelos dados (p.ex., IC 95%) e H_0 o intervalo nulo. O SGPV é (27.1), onde $|I|$ é a largura do intervalo de estimativa, $|H_0|$ é a largura do intervalo nulo e $|I \cap H_0|$ é a largura da sobreposição entre os dois intervalos. O SGPV é restrito ao intervalo $[0, 1]$.²⁸⁹

$$p_\delta = \frac{|I \cap H_0|}{|I|} \times \max\left\{\frac{|I|}{2|H_0|}, 1\right\} \quad (27.1)$$

- Quando $|I| < 2|H_0|$, p_δ é apenas a fração de sobreposição $|I \cap H_0|/|I|$.²⁸⁹
- Quando $|I| > 2|H_0|$, o SGPV reduz-se a $\frac{1}{2} \times \frac{|I \cap H_0|}{|H_0|} \leq \frac{1}{2}$, sinalizando inconclusão por imprecisão.²⁸⁹

27.6.4 Como interpretar o SGPV?

- $p_\delta = 0$: dados apoiam apenas hipóteses alternativas relevantes (IC totalmente fora da equivalência).²⁸⁹
- $p_\delta = 1$: dados apoiam apenas hipóteses nulas (equivalentes) (IC totalmente dentro da equivalência).²⁸⁹ $0 < p_\delta < 1$: inconclusivo; o valor expressa o grau de inconclusão. Em particular, $p_\delta = \frac{1}{2}$ indica inconclusão estrita.²⁸⁹
- O SGPV é descritivo (não é probabilidade posterior de H_0).²⁸⁹

27.6.5 Relação com testes de equivalência (TOST)

- Tanto SGPV quanto TOST compararam o IC com os limites de equivalência. Se o IC $(1 - 2\alpha)$ (p.ex., 90% quando $\alpha = 0,05$) cai inteiro dentro dos limites, TOST conclui equivalência — situação análoga a $p_\delta = 1$.²⁹⁰
- Com ICs simétricos, há pontos de ancoragem em que as estatísticas coincidem: quando $p_{\text{TOST}} = 0,5$, então SGPV = 0,5; quando o IC toca o limite mas fica inteiramente dentro (fronteira), $p_{\text{TOST}} = 0,025$ e SGPV = 1; quando o IC fica inteiramente fora tocando o limite, $p_{\text{TOST}} = 0,975$ e SGPV = 0.²⁹⁰
- Em ICs assimétricos ou quando $|I| > 2|H_0|$, o SGPV fica difícil de interpretar quando $0 < p_\delta < 1$; nesses cenários, o TOST costuma diferenciar melhor os resultados.²⁹⁰

27.6.6 Propriedades freqüenciais e múltiplas comparações

- Usando ICs $100(1 - \alpha)$, sob qualquer hipótese em H_0 , $\Pr(p_\delta = 0) \leq \alpha$ e $\rightarrow 0$ com o aumento de n ; fora de H_0 , $\Pr(p_\delta = 0) \rightarrow 1$ quando n cresce.²⁸⁹
- O SGPV mitiga naturalmente inflação de erro Tipo I em muitas comparações e prioriza relevância científica (não requer ajustes ad hoc).²⁸⁹

27.7 Distribuição de confiança

27.7.1 O que é distribuição de confiança?

- Distribuição de confiança é uma representação contínua da evidência inferencial sobre um parâmetro de interesse. Ela mostra, para cada valor possível do tamanho do efeito, o nível de confiança associado, sendo uma generalização visual do intervalo de confiança e do P-valor.²⁹¹

27.8 Boas práticas

- Defina H_0 intervalar e δ a priori com justificativa científica.^{289,290}
- Reporte: estimativa pontual, IC, limites de equivalência e p_δ ; interprete $p_\delta \in [0, 1]$ de forma dicotômica e $0 < p_\delta < 1$ como inconclusivo; quando necessário, complemente com TOST.^{289,290}

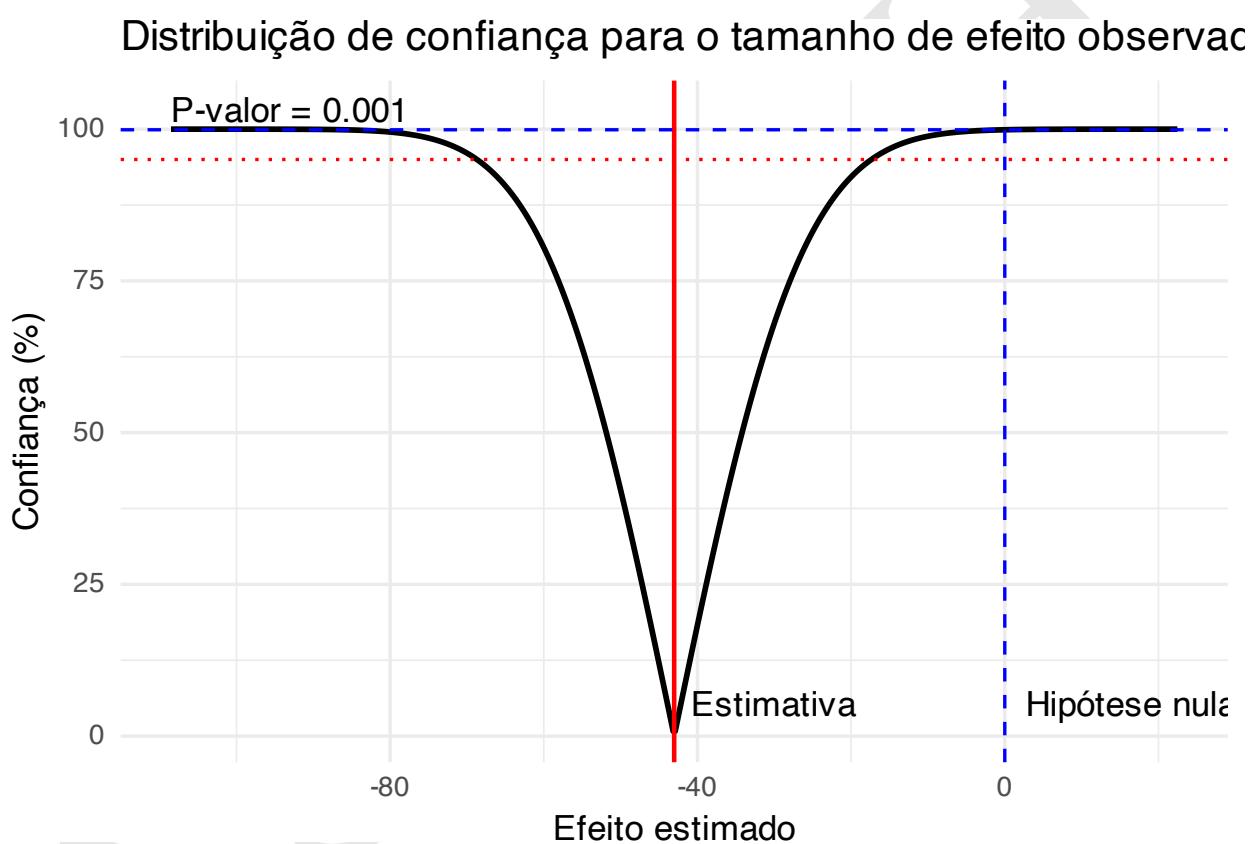


Figura 27.2: Distribuição de confiança para o tamanho do efeito estimado.

RASCUNHO

Capítulo 28

Testes estatísticos

28.1 Variáveis categóricas

- 28.1.1 Testes de Qui-quadrado (χ^2)
- 28.1.2 Teste exato de Fisher
- 28.1.3 Teste de McNemar
- 28.1.4 Teste Q de Cochran
- 28.1.5 Teste de Cochran–Armitage
- 28.1.6 Odds ratio (OR) e risco relativo (RR)

28.2 Variáveis contínuas

- 28.2.1 Teste t de Student
- 28.2.2 Teste t de Welch
- 28.2.3 Teste de Mann-Whitney
- 28.2.4 Teste de Wilcoxon
- 28.2.5 Análise de variância
- 28.2.6 Análise de variância (Welch)
- 28.2.7 Teste de Kruskal-Wallis

Tabela 28.1: Teste Qui-quadrado (com correção de Yates)

Tratamento	Resposta		Total	P-valor¹
	Respondeu	Não respondeu		
Total	95 (48%)	105 (53%)	200 (100%)	<0.001
A	60 (30%)	37 (19%)	97 (49%)	
B	35 (18%)	68 (34%)	103 (52%)	

¹ Teste qui-quadrado de independência

Tamanho do efeito (Cramér's V): 0.269

Tabela 28.2: Teste Qui-quadrado (sem correção de Yates)

Tratamento	Resposta		Total	P-valor¹
	Respondeu	Não respondeu		
Total	95 (48%)	105 (53%)	200 (100%)	<0.001
A	60 (30%)	37 (19%)	97 (49%)	
B	35 (18%)	68 (34%)	103 (52%)	

¹ Teste qui-quadrado de independência

Tamanho do efeito (Cramér's V): 0.269

Tabela 28.3: Teste exato de Fisher

Tratamento	Resposta		Total	P-valor¹
	Respondeu	Não respondeu		
Total	95 (48%)	105 (53%)	200 (100%)	<0.001
A	60 (30%)	37 (19%)	97 (49%)	
B	35 (18%)	68 (34%)	103 (52%)	

¹ Teste exato de Fisher

Tamanho do efeito (Cramér's V): 0.269

Tabela 28.4: Teste de McNemar

	Pos		Total
	Sim	Não	
Pre			
Sim	41 (41%)	25 (25%)	66 (66%)
Não	11 (11%)	23 (23%)	34 (34%)
Total	52 (52%)	48 (48%)	100 (100%)

McNemar $\chi^2 = 4.694$; gl = 1; p = 0.030

Tabela 28.5: Teste Q de Cochran

	Resposta		
	Positivo	Negativo	
Tempo			
T1	38 (16%)	42 (18%)	
T2	52 (22%)	28 (12%)	
T3	56 (23%)	24 (10%)	
Total	146 (61%)	94 (39%)	

Cochran Q $\chi^2(2) = 8.645$, p = 0.013

Tabela 28.6: Teste de Cochran–Armitage

	Status		
	Sucesso	Fracasso	
Grupo			
Alta	1 (50%)	1 (50%)	
Baixa	1 (50%)	1 (50%)	
Média	1 (50%)	1 (50%)	
Total	3 (50%)	3 (50%)	

$\chi^2(1) = 12.047$; p = <0.001

Tabela 28.7: Medidas de associação

Tratamento	Resposta		Total	Valor-p¹
	Não	Sim		
A	1 (50%)	1 (50%)	2 (100%)	>0.9
B	1 (50%)	1 (50%)	2 (100%)	
Total	2 (50%)	2 (50%)	4 (100%)	

¹ Teste qui-quadrado de independência

OR = 2.1 [1.17, 3.79]; RR = 1.61 [1.1, 2.35].

Tabela 28.8: Teste t de Student

Características	A N = 45¹	B N = 55¹	P-valor²
Desfecho	51.30 (8.88)	55.38 (10.05)	0.034

¹ Média (Desvio Padrão)² Teste t com correção de Welch

Tamanho do efeito (d de Cohen): -0.432

Tabela 28.9: Teste t de Welch

Características	A N = 45¹	B N = 55¹	P-valor²
Desfecho	51.30 (8.88)	55.38 (10.05)	0.034

¹ Média (Desvio Padrão)² Teste t com correção de Welch

Tamanho do efeito (d de Cohen): -0.432

Tabela 28.10: Teste de Mann–Whitney (Wilcoxon rank-sum)

Características	A N = 51¹	B N = 69¹	P-valor²
Desfecho	50.07 [46.34, 56.48]	54.47 [48.51, 61.73]	0.023

¹ Mediana [Q1, Q3]² Teste de soma de postos de Wilcoxon

Tamanho do efeito (r): 0.208

Tabela 28.11: Teste de Wilcoxon (signed-rank)

Características	Pré N = 60 ¹	Pós N = 60 ¹	P-valor ²
Desfecho	44.44 [39.93, 50.25]	53.00 [48.35, 59.65]	<0.001

¹ Mediana [Q1, Q3]² Teste de soma de postos de Wilcoxon

Tamanho do efeito (r): 0.616

Tabela 28.12: Análise de variância de um fator

Características	A N = 25 ¹	B N = 34 ¹	C N = 31 ¹	P-valor ²
Desfecho	50.58 (7.80)	55.64 (9.52)	61.65 (9.55)	<0.001

¹ Média (Desvio Padrão)² One-way analysis of meansTamanho do efeito (eta²): 0.193

Post hoc de Tukey

Comparação	Diferença de médias	IC95% inferior	IC95% superior	p (ajustado)
B-A	5.06	-0.65	10.77	0.093
C-A	11.07	5.25	16.89	<0.001
C-B	6.01	0.63	11.39	0.025

Tabela 28.13: Análise de variância de Welch

Características	A N = 43 ¹	B N = 57 ¹	C N = 50 ¹	P-valor (Welch)
Desfecho	53.42 (8.43)	55.90 (10.17)	58.10 (14.91)	0.137

¹ Média (Desvio Padrão)Tamanho do efeito (eta², via SS): 0.025

Post hoc de Games-Howell

Grupo 1	Grupo 2	Diferença de médias	IC95% inferior	IC95% superior	p (ajustado)
A	B	2.48	-1.95	6.91	0.381
A	C	4.68	-1.21	10.58	0.146
B	C	2.20	-3.76	8.17	0.654

Tabela 28.14: Teste de Kruskal-Wallis

Características	A N = 54 ¹	B N = 67 ¹	C N = 59 ¹	P-valor ²
Desfecho	40.98 [31.12, 51.85]	40.48 [31.75, 62.18]	64.96 [45.37, 77.91]	<0.001

¹ Mediana [Q1, Q3]² Teste de Kruskal-WallisTamanho do efeito: epsilon² = 0.14

Post hoc de Dunn (Bonferroni)

Grupo 1	Grupo 2	Z	p	p (ajustado)
A	B	0.742	0.458	1.000
A	C	4.714	<0.001	<0.001
B	C	4.213	<0.001	<0.001

Capítulo 29

Descrição

29.1 Análise de descrição

29.1.1 O que é análise de descrição de dados?

- A análise descritiva utiliza métodos para calcular, descrever e resumir os dados coletados da(s) amostra(s) de modo que sejam interpretadas adequadamente.¹⁰⁸
- As análises descritivas geralmente compreendem a apresentação quantitativa (numérica) em tabelas e/ou gráficos.¹⁰⁸



O pacote *explore*¹⁹⁶ fornece a função *explore*^a para análise exploratória de um banco de dados.

^a<https://www.rdocumentation.org/packages/explore/versions/1.0.2/topics/explore>



O pacote *dataMaid*¹⁹⁷ fornece a função *makeDataReport*^a para criar um relatório de análise exploratória de um banco de dados.

^a<https://www.rdocumentation.org/packages/dataMaid/versions/1.4.1/topics/makeDataReport>



O pacote *DataExplorer*¹⁹⁸ fornece a função *create_report*^a para criar um relatório de análise exploratória de um banco de dados.

^ahttps://www.rdocumentation.org/packages/DataExplorer/versions/0.8.2/topics/create_report



O pacote *SmartEDA*¹⁹⁹ fornece a função *ExpReport*^a para criar um relatório de análise exploratória de um banco de dados.

^a<https://www.rdocumentation.org/packages/SmartEDA/versions/0.3.9/topics/ExpReport>



O pacote *esquisse*²⁹¹ fornece a função *esquisser*^a para executar uma interface interativa para visualização de dados.

^a<https://www.rdocumentation.org/packages/esquisse/versions/1.1.2/topics/esquisser>

29.2 Estimação

29.2.1 O que é estimativa?

- Estimativa é o valor de uma variável de interesse calculado a partir de uma amostra.[?]

29.2.2 O que é estimativa pontual?

- Estimativa pontual é o valor único de uma variável de interesse calculado a partir de uma amostra.[?]

29.2.3 O que é estimativa intervalar?

- Estimativa intervalar é um intervalo de valores de uma variável de interesse calculado a partir de uma amostra.[?]

29.2.4 O que é estimativa de parâmetro?

- Estimativa de parâmetro é o valor de uma variável de interesse calculado a partir de uma amostra que representa o valor da população.[?]

Capítulo 30

Comparação

30.1 Análise inferencial de comparação

30.1.1 O que é análise de comparação de dados?

- ?

R

O pacote *cocor*²⁹² fornece as funções *cocor.indep.groups*^a, *cocor.dep.groups.overlap*^b e *cocor.dep.groups.nonoverlap*^c para comparar 2 coeficientes de correlação entre grupos independentes, grupos sobrepostos ou independentes, respectivamente.²⁹²

^a<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/cor.test>

^b<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/cor.test>

^c<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/cor.test>

30.2 F-teste

30.2.1 O que é o F-teste?

- O F-teste é uma estatística que compara a variabilidade entre grupos com a variabilidade dentro dos grupos.[?]
- A estatística é calculada como (30.1), onde QM são “quadrados médios”, com gl_1 e gl_2 definidos pelo desenho (ex.: fatores e resíduos).[?]

$$F = \frac{QM_{entre}}{QM_{dentro}} \quad (30.1)$$

30.2.2 Quando usar o F-teste?

- Análise de variância de um fator (≥ 3 grupos) e ANOVA multifatorial (efeitos principais e interações).[?]
- Modelo linear generalizado / regressão linear: teste global $H_0 : \beta_1 = \dots = \beta_p = 0$.[?]
- Análise de covariância (comparação de grupos ajustando covariáveis).[?]
- Contrastos planejados ou pós-hoc (usando a razão F correspondente).[?]

30.2.3 Quais são os pressupostos?

- Observações independentes?
- Normalidade (aproximada) dos resíduos?
- Homogeneidade de variâncias entre grupos (homoscedasticidade)?
- Se houver violações importantes: considerar ANOVA de Welch, transformações apropriadas ou alternativas não paramétricas (ex.: Kruskal–Wallis para um fator).?

30.2.4 Como interpretar o resultado?

- Valor de F elevado com $P < \alpha$ indica evidência contra H_0 (diferenças entre grupos/modelo com ajuste significativo)?
- Relate sempre gl_1 , gl_2 , F e P , além de um tamanho de efeito (ex.: η^2 , η_p^2 ou ω^2) e intervalo de confiança quando possível?
- Após rejeitar H_0 , use contrastes ou pós-hoc com ajuste para múltiplas comparações para localizar as diferenças?

30.2.5 O que reportar em publicações?

- Estrutura do desenho (fatores, níveis, balanceamento)?
- Verificação/diagnóstico dos pressupostos?
- Estatística F com gl e P ?
- Tamanho de efeito e intervalo de confiança?
- Método de ajuste para múltiplas comparações quando aplicável?

Capítulo 31

Associação

31.1 Análise inferencial de associação

31.1.1 O que é análise de associação?

- ?

31.2 Associação bivariada

31.2.1 O que são análises de associação bivariada?

- ?

31.2.2 Quais testes podem ser usados para análises de associação bivariada?

- Teste Qui-quadrado (χ^2).^{293,294}
 - O teste qui-quadrado (χ^2) avalia uma hipótese global se a relação entre duas variáveis e/ou fatores é independente ou associada.²⁹⁴
 - O teste qui-quadrado é utilizado para comparar a distribuição de uma variável categórica em uma amostra ou grupo com a distribuição em outro. Se a distribuição da variável categórica não for muito diferente nos diferentes grupos, pode-se concluir que a distribuição da variável categórica não está relacionada com a variável dos grupos. Pode-se também concluir que a variável categórica e os grupos são independentes.²⁹⁴
 - Tipo: não paramétrico.^{293,294}
 - Suposições:^{293,294}
 - * As variáveis são ordinais ou categóricas nominais, de modo que as células representem frequência.
 - * Os níveis dos fatores (variáveis categóricas) são mutuamente exclusivos.
 - * Tamanho de amostra grande e adequado porque é baseado em uma abordagem de aproximação.
 - * Menos de 20% das células com frequências esperadas < 5
 - * Nenhuma célula com frequência esperada < 1 .
 - Hipóteses:²⁹⁴

- * Nula (H_0): independente (sem associação)
- * Alternativa (H_1): não independente (associação)
- Tamanho do efeito:²⁹⁴
 - * Phi (ϕ), para tabelas de contingência 2x2
 - * Razão de chances (RC ou OR), para tabelas de contingência 2x2
 - * Cramer V (V), para tabelas de contingência NxM

R O pacote *gtsummary*²¹¹ fornece a função *tbl_cross*^a para criar uma tabela NxM.

^ahttps://www.rdocumentation.org/packages/gtsummary/versions/1.6.3/topics/tbl_cross

- Teste Exato de Fisher.^{293,294}
 - O teste exato de Fisher avalia a hipótese nula de independência aplicando a distribuição hipergeométrica dos números nas células da tabela.²⁹⁴
 - Hipóteses:^{293,294}
 - * Nula (H_0): independente (sem associação)
 - * Alternativa (H_1): não independente (associação)
 - Tamanho do efeito:^{293,294}
 - * Phi (ϕ), para tabelas de contingência 2x2
 - * Razão de chances (RC ou OR), para tabelas de contingência 2x2
 - * Cramer V (V), para tabelas de contingência NxM

R O pacote *gtsummary*²¹¹ fornece a função *tbl_cross*^a para criar uma tabela NxM.

^ahttps://www.rdocumentation.org/packages/gtsummary/versions/1.6.3/topics/tbl_cross

31.3 Associação multivariada

31.3.1 O que são análises de associação multivariada?

- ?

31.3.2 Quais testes podem ser usados para análises de associação multivariada?

- ?

Capítulo 32

Correlação

32.1 Análise inferencial de correlação

32.1.1 O que é covariância?

- ?

32.1.2 O que é correlação?

- ?

32.1.3 Qual é a interpretação das medidas de correlação?

- Os valores de correlação estão no intervalo $[-1; 1]$.^{112,295,296}
- Valores de correlação positivos representam uma relação direta entre as variáveis, tal que valores maiores de uma variável estão associados a valores maiores de outra variável.^{295,296}
- Valores de correlação negativos representam uma relação indireta (ou inversa) entre as variáveis, tal que valores maiores (menores) de uma variável estão associados a valores maiores (menores) de outra variável.^{295,296}
- Valores de correlação próximos de 0 representam a inexistência de relação entre as variáveis.^{295,296}

32.1.4 Quais precauções devem ser tomadas na interpretação de medidas de correlação?

- Tamanhos de efeito grande (ou qualquer outro) não representam necessariamente uma relação causa-efeito entre as variáveis.²⁹⁵
- Tamanhos de efeito grande (ou qualquer outro) não representam necessariamente uma relação de concordância ou confiabilidade entre as variáveis.²⁹⁵
- Uma escala de medição com representação agregada do constructo na coleta de dados pode subestimar o tamanho do efeito da correlação r em de cerca de 13% e do coeficiente de determinação R^2 de cerca de 30%.¹²⁵ Neste caso, a correlação desatenuada $r_{x'y'}$ pode ser calculada por (32.1), utilizando a correlação observada r_{xy} e os fatores de correção $r_{xx'}$ e $r_{yy'}$ para o número de intervalos nas variáveis X e Y, respectivamente:¹²⁵

$$r_{x'y'} = \frac{r_{xy}}{r_{xx'} r_{yy'}} \quad (32.1)$$

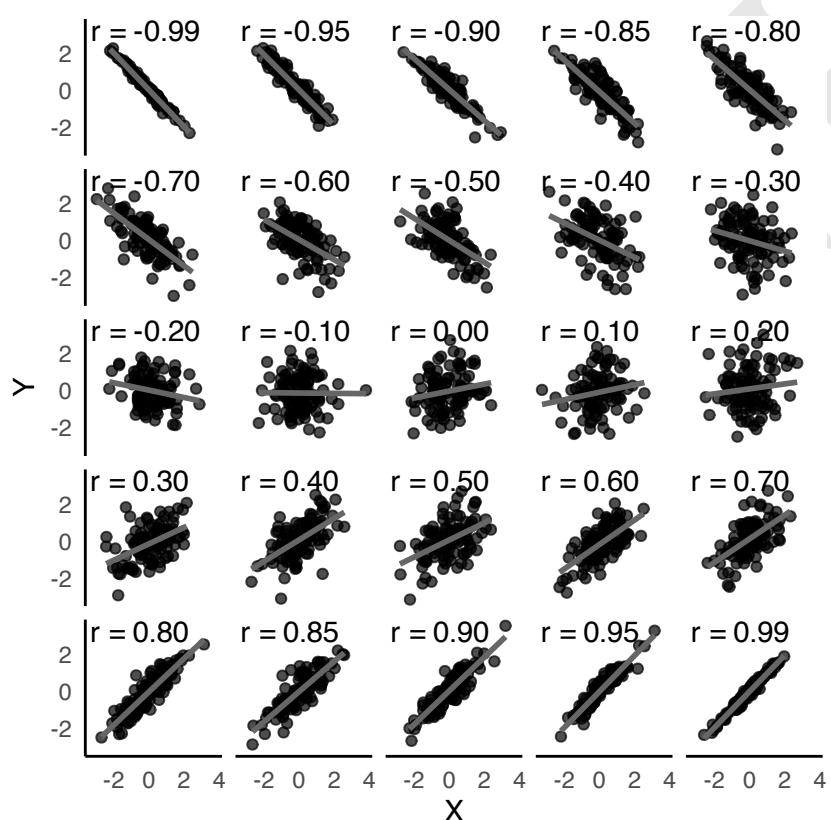


Figura 32.1: Exemplo de diferentes forças e direção de correlação entre duas variáveis X e Y.

Tabela 32.1: Quarteto de Anscombe.

ID	x1	x2	x3	x4	y1	y2	y3	y4
1	10	10	10	8	8.04	9.14	7.46	6.58
2	8	8	8	8	6.95	8.14	6.77	5.76
3	13	13	13	8	7.58	8.74	12.74	7.71
4	9	9	9	8	8.81	8.77	7.11	8.84
5	11	11	11	8	8.33	9.26	7.81	8.47
6	14	14	14	8	9.96	8.10	8.84	7.04
7	6	6	6	8	7.24	6.13	6.08	5.25
8	4	4	4	19	4.26	3.10	5.39	12.50
9	12	12	12	8	10.84	9.13	8.15	5.56
10	7	7	7	8	4.82	7.26	6.42	7.91
11	5	5	5	8	5.68	4.74	5.73	6.89

Tabela 32.2: Análise descritiva do Quarteto de Anscombe demonstrando os conjuntos de dados bivariados com parâmetros quase idênticos.

	X1Y1	X2Y2	X3Y3	X4Y4
Observações	11.00	11.00	11.00	11.00
Média x	9.00	9.00	9.00	9.00
Média y	7.50	7.50	7.50	7.50
Variância x	11.00	11.00	11.00	11.00
Variância y	4.13	4.13	4.12	4.12
Correlação	0.82	0.82	0.82	0.82
Coeficiente angular	0.50	0.50	0.50	0.50
Coeficiente linear	3.00	3.00	3.00	3.00
Coeficiente de determinação	0.67	0.67	0.67	0.67



O pacote *psychmeta*²⁹⁷ fornece a função *correct_r_coarseness*^a para calcular o coeficiente de correlação desatenuado ($r_{x'y'}$).

^ahttps://www.rdocumentation.org/packages/psychmeta/versions/2.7.0/topics/correct_r_coarseness



O pacote *psychmeta*²⁹⁷ fornece a função *correct_r*^a para calcular o coeficiente de correlação em escala restrita e/ou com erro de mensuração ($r_{x'y'}$).

^ahttps://www.rdocumentation.org/packages/psychmeta/versions/2.7.0/topics/correct_r

- Os coeficientes de correlação possuem suposições que, se violadas, podem levar a interpretações equivocadas. Nestes cenários, visualizar os dados e as relações entre as variáveis pode contribuir com a interpretação e utilidade dos coeficientes de correlação.²⁹⁸
- O quarteto de Anscombe é um conjunto de quatro bancos de dados bivariados que possuem a mesma média, variância, correlação e regressão linear (até a 2a casa decimal), mas que são visualmente diferentes e, assim, demonstram a importância da análise gráfica da correlação.²⁹⁸

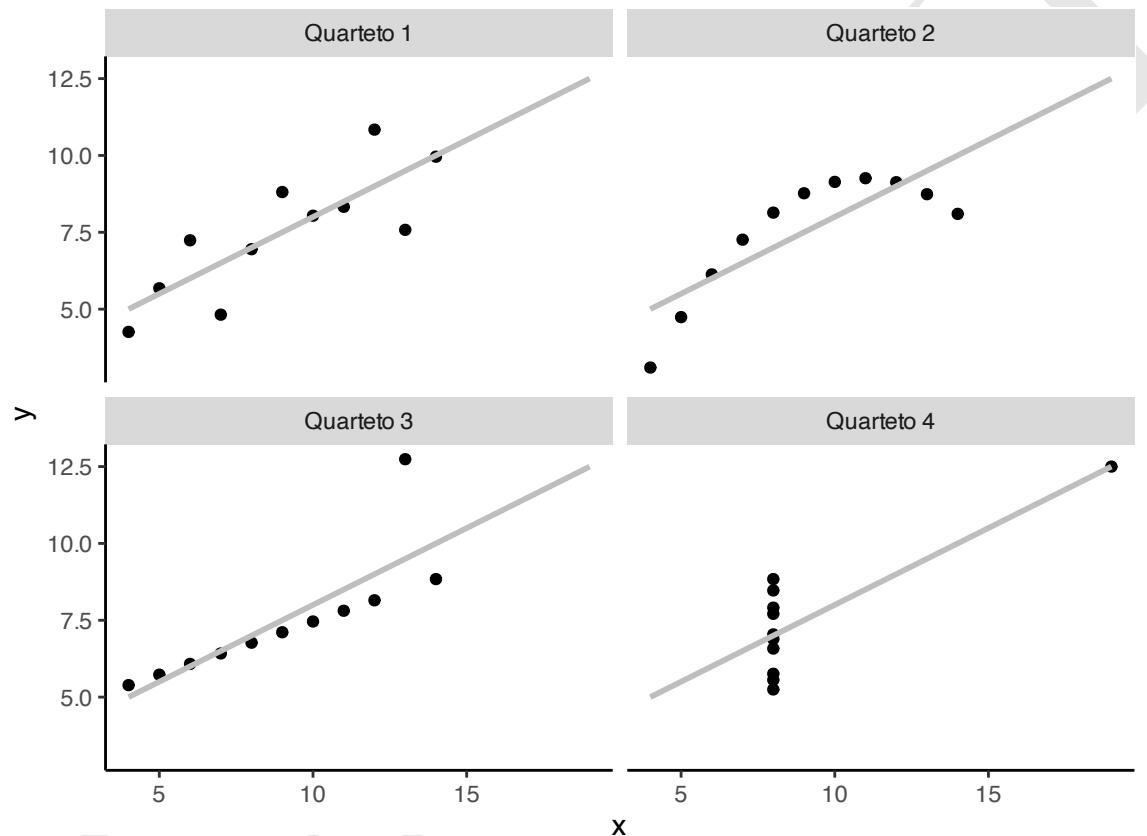


Figura 32.2: Gráfico de dispersão do Quarteto de Anscombe para representação gráfica de conjuntos de dados bivariados com parâmetros quase idênticos e relações muito distintas.



O pacote *anscombiner*²⁹⁹ fornece a função *anscombise*^a para gerar bancos de dados que compartilham os mesmos valores de parâmetros do Quarteto de Anscombe.

^a<https://www.rdocumentation.org/packages/anscombiner/versions/1.1.0/topics/anscombise>

32.2 Coeficientes de correlação

32.2.1 Quais coeficientes podem ser usados em análises de correlação?

- Coeficiente de correlação de Pearson (r)^{295,296}

$$r = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2] [n \sum y_i^2 - (\sum y_i)^2]}} \quad (32.2)$$

- O coeficiente de correlação de Pearson (r) avalia a força e direção da relação linear entre duas variáveis quantitativas.^{295,296}
- Tipo: paramétrico.^{295,296}
- Hipóteses:²⁹⁶
- Nula (H_0): $r = 0$
- Alternativa (H_1): $r \neq 0$
- Tamanho do efeito:^{295,296}
- Coeficiente de correlação de Pearson (r)



O pacote *stats*¹³¹ fornece a função *cor.test*^a para calcular o coeficiente de correlação de Pearson (r).

^a<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/cor.test>



O pacote *correlation*³⁰⁰ do projeto *easystats*³⁰¹ fornece a função *correlation*^a para calcular o coeficiente de correlação de Pearson (r).

^a<https://cloud.r-project.org/web/packages/correlation/index.html>

- Coeficiente de correlação ponto-bisserial (r_s)²⁹⁵

$$r_s = \frac{M_1 - M_0}{s_y} \sqrt{\frac{n_1 n_0}{n^2}} \quad (32.3)$$

- O coeficiente de correlação ponto-bisserial (r_s) avalia a força e direção da relação linear entre uma variável quantitativa e outra dicotômica.²⁹⁵
- Tipo: paramétrico.²⁹⁵
- Hipóteses:²⁹⁵
- Nula (H_0): $r_s = 0$
- Alternativa (H_1): $r_s \neq 0$

- Tamanho do efeito.²⁹⁵
- Coeficiente de correlação ponto-bisserial (r_s)

R

O pacote *stats*¹³¹ fornece a função *cor.test*^a para calcular o coeficiente de correlação ponto-bisserial (r_s).

^a<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/cor.test>

R

O pacote *correlation*³⁰⁰ do projeto *easystats*³⁰¹ fornece a função *correlation*^a para calcular o coeficiente de correlação ponto-bisserial (r_s).

^a<https://cloud.r-project.org/web/packages/correlation/index.html>

- Coeficiente de correlação de Spearman (ρ) (32.4).^{295,296}

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \quad (32.4)$$

- O coeficiente de correlação de Spearman (ρ) avalia a força e direção da relação monotônica entre duas variáveis quantitativas.^{295,296}
- O coeficiente de correlação de Spearman (ρ) pode ser também definida como a correlação de Pearson (r) entre as classificações (*ranks*) das duas variáveis quantitativas.^{295,296}
- Tipo: não-paramétrico.^{295,296}
- Hipóteses:^{295,296}
- Nula (H_0): $\rho = 0$
- Alternativa (H_1): $\rho \neq 0$
- Tamanho do efeito.^{295,296}
- Coeficiente de correlação de Spearman (ρ)

R

O pacote *stats*¹³¹ fornece a função *cor.test*^a para calcular o coeficiente de correlação de Spearman (ρ).

^a<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/cor.test>

R

O pacote *correlation*³⁰⁰ do projeto *easystats*³⁰¹ fornece a função *correlation*^a para calcular o coeficiente de correlação de Spearman (ρ).

^a<https://cloud.r-project.org/web/packages/correlation/index.html>

- Coeficiente de Kendall (τ) (32.5).^{295,296}

$$\tau = \frac{(n_c - n_d)}{\frac{1}{2}n(n - 1)} \quad (32.5)$$

- O coeficiente Kendall τ avalia a força e direção da relação monotônica entre duas variáveis quantitativas ou qualitativas.^{295,296}
- O coeficiente Kendall τ é definido como a proporção de todos os pares concordantes menos a proporção de todos os pares discordantes.^{295,296}
- Tipo: não-paramétrico.^{295,296}
- Hipóteses:^{295,296}
- Nula (H_0): $\tau = 0$
- Alternativa (H_1): $\tau \neq 0$
- Tamanho do efeito:^{295,296}
- Kendall τ



O pacote *stats*¹³¹ fornece a função *cor.test*^a para calcular o coeficiente Kendall τ .

^a<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/cor.test>



O pacote *correlation*³⁰⁰ do projeto *easystats*³⁰¹ fornece a função *correlation*^a para calcular o coeficiente Kendall τ .

^a<https://cloud.r-project.org/web/packages/correlation/index.html>

- Coeficiente de Cramér (V) (32.6).[?]

$$V = \sqrt{\frac{\chi^2/n}{\min(k-1, r-1)}} \quad (32.6)$$

- O coeficiente Cramér (V) avalia a força e direção da relação entre duas variáveis qualitativas.[?]
- Tipo: não-paramétrico.[?]
- Hipóteses:[?]
- Nula (H_0): $V = 0$
- Alternativa (H_1): $V \neq 0$
- Tamanho do efeito:[?]
- Coeficiente Cramer (V)
- Coeficiente de Sheperd (ϕ) (32.7).[?]

$$\phi = \sqrt{\frac{\chi^2}{n}} \quad (32.7)$$

- O coeficiente Phi (ϕ) avalia a força e direção da relação entre duas variáveis dicotômicas.[?]
- Tipo: não-paramétrico.[?]
- Hipóteses:[?]
- Nula (H_0): $\phi = 0$
- Alternativa (H_1): $\phi \neq 0$

- Tamanho do efeito.²
- Coeficiente Phi (ϕ)

R

O pacote *correlation*³⁰⁰ do projeto *easystats*³⁰¹ fornece a função *correlation^a* para calcular o coeficiente coeficiente Sheperd ϕ .

^a<https://cloud.r-project.org/web/packages/correlation/index.html>

R

O pacote *corrplot*²²⁵ fornece a função *cor.mtest^a* para calcular os P-valores e intervalos de confiança da matriz de correlação.

^a<https://www.rdocumentation.org/packages/corrplot/versions/0.92/topics/cor.mtest>

R

O pacote *corrplot*²²⁵ fornece a função *corrplot^a* para visualização da matriz de correlação.

^a<https://www.rdocumentation.org/packages/corrplot/versions/0.92/topics/corrplot>

32.3 Colinearidade

32.3.1 O que é colinearidade?

- Colinearidade representa a correlação entre duas variáveis.³⁰²
- Colinearidade exata indica uma relação linear perfeita entre duas variáveis.³⁰²

32.3.2 Como identificar colinearidade na matriz de correlação?

- A colinearidade pode ser identificada na matriz de correlação por meio da análise dos coeficientes de correlação entre as variáveis.³⁰²
- Valores de correlação próximos de 1 ou -1 indicam colinearidade entre as variáveis.³⁰²

R

O pacote *GGally*³⁰³ fornece a função *ggally_cor^a* para estimar a correlação bivariada e exibir o coeficiente de correlação e o P-valor na matriz de correlação.³⁰³

^ahttps://www.rdocumentation.org/packages/GGally/versions/2.2.1/topics/ggally_cor

32.4 Correlação entre conjuntos de variáveis

32.4.1 O que é correlação entre conjuntos de variáveis?

- A Correlação Canônica (CCA) analisa a relação entre dois conjuntos de variáveis simultaneamente.²
- Busca combinações lineares que maximizam a correlação entre os dois blocos.²

32.4.2 Quando usar CCA?

- Quando existem dois blocos distintos de variáveis.²
- Quando a correlação bivariada é insuficiente para captar padrões multivariados.²

32.4.3 Quais são os principais resultados?

- Correlação canônica (ρ_1, ρ_2, \dots): força da associação entre os escores dos blocos.[?]
- Escores canônicos (U e V): novas variáveis representando os blocos.[?]
- *Loadings* e *cross-loadings*: indicam quais variáveis mais contribuem para cada eixo.[?]

32.4.4 Como interpretar a CCA?

- ρ_1 indica a força do primeiro eixo canônico ($U_1 \leftrightarrow V_1$).[?]
- Gráficos de U_1 vs V_1 podem revelar padrões por grupo ou gradiente.[?]
- *Loadings/cross-loadings* mostram quais variáveis explicam a correlação.[?]

32.4.5 Quais suposições e cuidados?

- As variáveis devem estar padronizadas (escalas comparáveis).[?]
- Preferível $n >$ número de variáveis em cada bloco.[?]
- Atenção a multicolinearidade alta (pode exigir CCA regularizada).[?]

32.4.6 O que reportar nos resultados?

- Valores de ρ_1, ρ_2, \dots com teste de Wilks e p-valores.[?]
- Figura U_1 vs. V_1 com interpretação.[?]
- Tabela de *loadings* ou *cross-loadings* destacando contribuições relevantes.[?]
- Uma interpretação substantiva da relação entre os blocos.[?]

```
# Reproducibilidade
set.seed(123)

# Suponha X: variáveis ambientais; Y: traços de plantas
n <- 120
X <- scale(cbind(pH = rnorm(n, 6.5, .4),
temp = rnorm(n, 20, 3),
rain = rnorm(n, 1000, 120)))
Y <- scale(cbind(height = 0.4*X[, "temp"] - 0.3*X[, "pH"] + rnorm(n, 0, .6),
leaf = 0.3*X[, "rain"] + 0.25*X[, "temp"] + rnorm(n, 0, .6),
chl = -0.35*X[, "pH"] + 0.3*X[, "rain"] + rnorm(n, 0, .6)))

# (Opcional) fator de cor para o scatter
classe <- factor(sample(c("Forest", "Herb", "Planted", "Shrub"), n, TRUE))

# ----- CCA (base R) -----
fit <- cancor(X, Y)    # stats::cancor
rho <- fit$cor           # correlações canônicas

# Escores canônicos (U e V)

U <- scale(X) %*% fit$xcoef
V <- scale(Y) %*% fit$ycoef
U1 <- U[, 1]; V1 <- V[, 1]
```

```

# ---- Teste sequencial (Wilks) ----

# wilks <- CCP::p.asym(rho, N=n, p=ncol(X), q=ncol(Y), tstat="Wilks")
# wilks_tab <- transform(wilks, rho=round(rho,3),
# p.value=signif(p.value,3))

# ---- Loadings e cross-loadings (correlações com escores) ----

# Loadings: var-X com U1.., var-Y com V1..

loadX <- cor(X, U[,1:min(ncol(X), ncol(Y))])
loadY <- cor(Y, V[,1:min(ncol(X), ncol(Y))])

# Cross-loadings: var-X com V1.., var-Y com U1..

crossX <- cor(X, V[,1:min(ncol(X), ncol(Y))])
crossY <- cor(Y, U[,1:min(ncol(X), ncol(Y))])

# ---- Visualizações ----

p1 <- ggplot2::ggplot(data.frame(U1=U1, V1=V1, classe=classe),
ggplot2::aes(U1, V1, shape=classe)) +
ggplot2::geom_point(alpha=.75) +
ggplot2::geom_smooth(method="lm", se=FALSE) +
ggplot2::labs(x="U1 (X → a1)", y="V1 (Y → b1)",
subtitle=paste0(" 1 = ", round(rho[1], 3))) +
ggplot2::theme_minimal()

# Heatmap de cross-loadings (quais variáveis de X e Y conectam-se ao outro bloco)

cx <- reshape2::melt(round(crossX, 2), varnames=c("VarX","CompV"),
value.name="cross")
cy <- reshape2::melt(round(crossY, 2), varnames=c("VarY","CompU"),
value.name="cross")

print(p1)

```

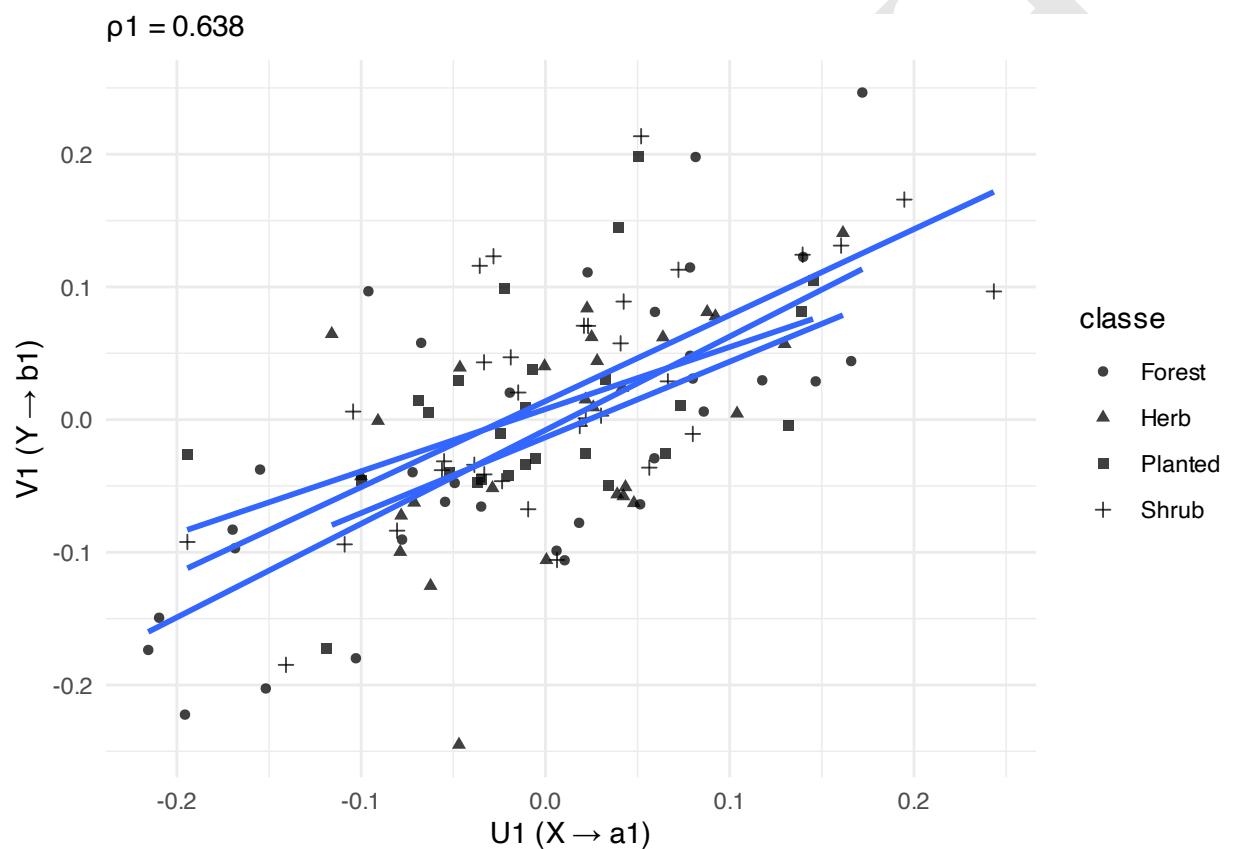


Figura 32.3: Exemplo de análise de correlação canônica (CCA) entre dois conjuntos de variáveis.

RASCUNHO

Capítulo 33

Regressão

33.1 Análise de regressão

33.1.1 O que é regressão?

- Regressão refere-se a uma equação matemática que permite que uma ou mais variável(is) de desfecho (dependentes) seja(m) prevista(s) a partir de uma ou mais variável(is) independente(s). A regressão implica em uma direção de efeito, mas não garante causalidade.²⁷²
- Para estimar os efeitos imparciais de um fator de exposição primária sobre uma variável de desfecho, frequentemente constroem-se modelos estatísticos de regressão.²²⁰



O pacote *modelsummary*³⁰⁴ fornece as funções *modelsummary*^a e *modelplot*^b para gerar tabelas e gráficos de coeficientes de regressão.

^a<https://www.rdocumentation.org/packages/modelsummary/versions/1.4.1/topics/modelsummary>

^b<https://www.rdocumentation.org/packages/modelsummary/versions/1.4.1/topics/modelplot>



O pacote *gtsummary*²¹¹ fornece a função *tbl_regression*^a para construção da ‘Tabela 2’ com dados do modelo de regressão.

^ahttps://www.rdocumentation.org/packages/gtsummary/versions/1.6.3/topics/tbl_regression

33.1.2 Quais são os algoritmos de regressão?

- Linear: Simples, Múltipla, Polinomial.[?]
- Linear generalizado: Binomial (logística), Multinomial, Ordinal, Poisson, Binomial negativa, Gama.[?]
- Não-linear (nos parâmetros).[?]
- Aditivo generalizado.[?]
- Efeitos mistos: Linear, Generalizado.[?]
- Sobrevida: Cox, Weibull, Exponencial, Log-normal, Log-logístico.[?]
- Regularização: Ridge, LASSO[?]

33.2 Estruturas de análise de regressão

33.2.1 O que são análises de regressão simples?

- A análise de regressão simples consiste em modelos estatísticos com uma variável dependente (desfecho) e uma variável independente (preditor).³⁰⁵
- A equação de regressão simples é expressa como (33.1), onde Y é a variável dependente, X é a variável independente, β_0 é o intercepto (constante), β_1 é o coeficiente de regressão da variável independente e ϵ representa o erro aleatório do modelo.³⁰⁵

$$Y = \beta_0 + \beta_1 X + \epsilon \quad (33.1)$$

33.2.2 O que são análises de regressão multivariável?

- A análise multivariável (ou múltiplo) consiste em modelos estatísticos com uma variável dependente (desfecho) e duas ou mais variáveis independentes.³⁰⁵
- A equação de regressão multivariável é expressa como (33.2), onde Y é a variável dependente, X_1, X_2, \dots, X_n são as variáveis independentes, β_0 é o intercepto (constante), $\beta_1, \beta_2, \dots, \beta_n$ são os coeficientes de regressão das variáveis independentes e ϵ representa o erro aleatório do modelo.³⁰⁵

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \quad (33.2)$$

33.2.3 O que são análises de regressão multivariada?

- A análise multivariada consiste em modelos estatísticos com duas ou mais variáveis dependentes (desfechos) e duas ou mais variáveis independentes.³⁰⁵
- A equação de regressão multivariada é expressa como (33.3), onde Y_1, Y_2, \dots, Y_m são as variáveis dependentes, X_1, X_2, \dots, X_n são as variáveis independentes, β_{0j} é o intercepto (constante) da variável dependente Y_j , β_{ij} são os coeficientes de regressão das variáveis independentes para a variável dependente Y_j e ϵ_j representa o erro aleatório do modelo para a variável dependente Y_j .³⁰⁵

$$Y_1 = \beta_{01} + \beta_{11} X_1 + \beta_{12} X_2 + \dots + \beta_{1n} X_n + \epsilon_1 \quad (33.3)$$

$$Y_2 = \beta_{02} + \beta_{21} X_1 + \beta_{22} X_2 + \dots + \beta_{2n} X_n + \epsilon_2 \quad (33.4)$$

$$\vdots \quad (33.5)$$

$$Y_m = \beta_{0m} + \beta_{m1} X_1 + \beta_{m2} X_2 + \dots + \beta_{mn} X_n + \epsilon_m \quad (33.6)$$

33.3 Tipos e famílias de regressão

33.3.1 O que são modelos de regressão linear?

- Modelos lineares (33.7) descrevem uma relação linear nos parâmetros entre um desfecho contínuo Y e um ou mais preditores X .?

$$Y = \beta_0 + \sum_{i=1}^n \beta_i X_i + \epsilon \quad (33.7)$$

- Assumem erros independentes, de média zero e variância constante (homocedasticidade).?

- A normalidade dos resíduos é uma hipótese comum para inferência estatística, mas não obrigatória para estimação dos coeficientes?

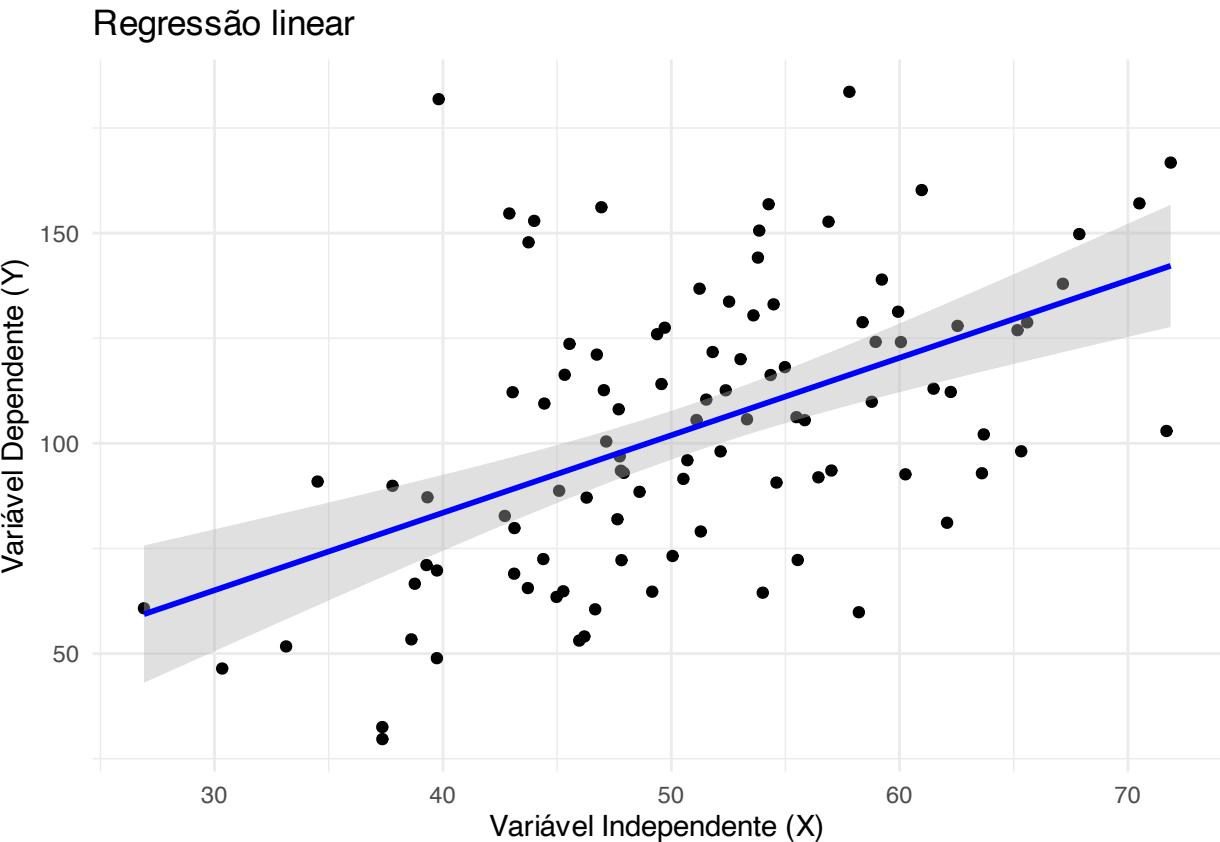


Figura 33.1: Regressão linear.

33.3.2 O que são modelos de regressão polinomial?

- São extensões da regressão linear em que se incluem termos elevados a potências das variáveis independentes (ex.: X^2 , X^3), permitindo capturar relações curvas?
- Modelos de regressão polinomial continuam sendo lineares nos parâmetros, por isso ainda se enquadram como um caso particular da regressão linear?

33.3.3 O que são modelos de regressão não-linear?

- São modelos em que a relação entre os parâmetros e a variável resposta não é linear.
- Podem assumir formas funcionais mais complexas (ex.: exponencial, logarítmica, logística)?
- Importante diferenciar “não-linear na variável” (ex.: polinomial) de “não-linear no parâmetro” (ex.: modelos logísticos de crescimento).

33.3.4 O que são modelos de regressão logística?

- Modelos logísticos são casos de regressão linear generalizada em que a resposta Y é binária?
- A equação (33.8) modela a razão de chances (*odds*) em função dos preditores?

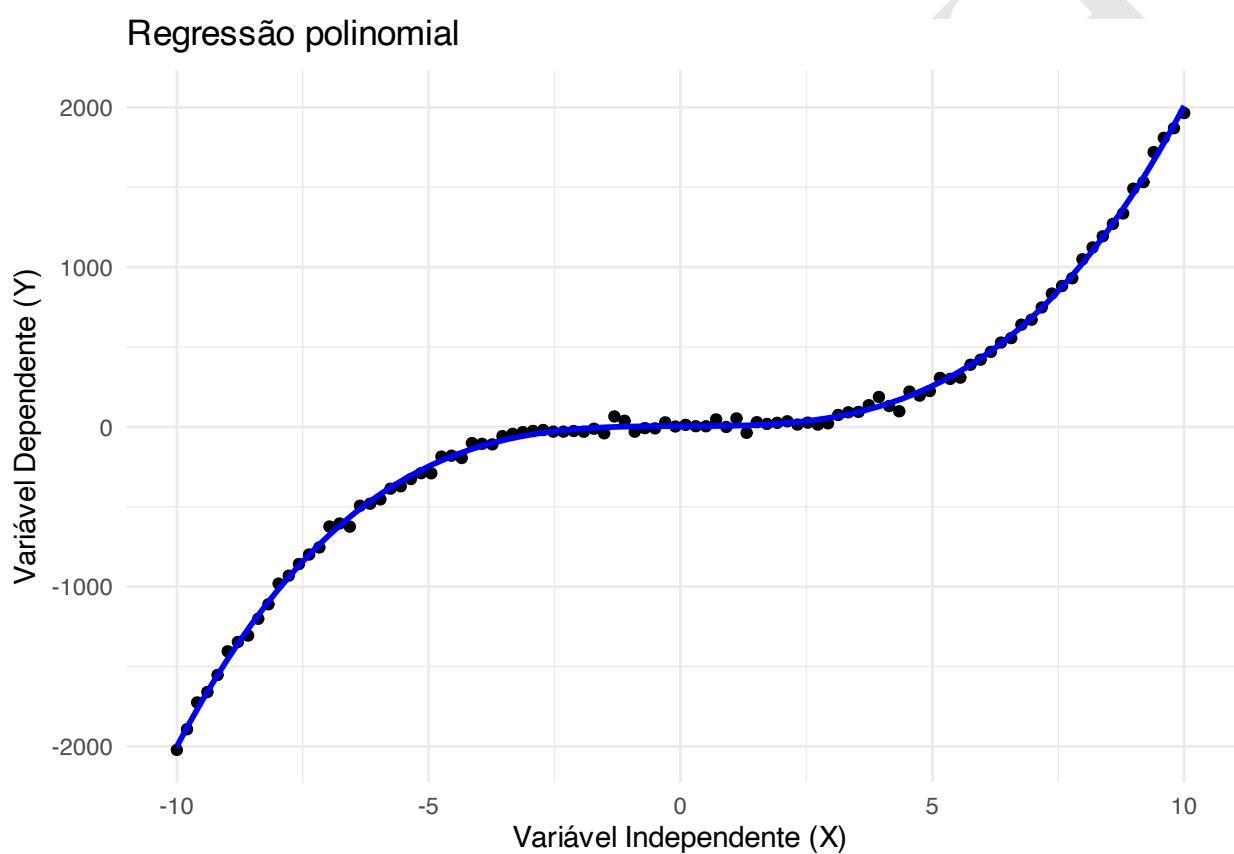


Figura 33.2: Regressão polinomial.

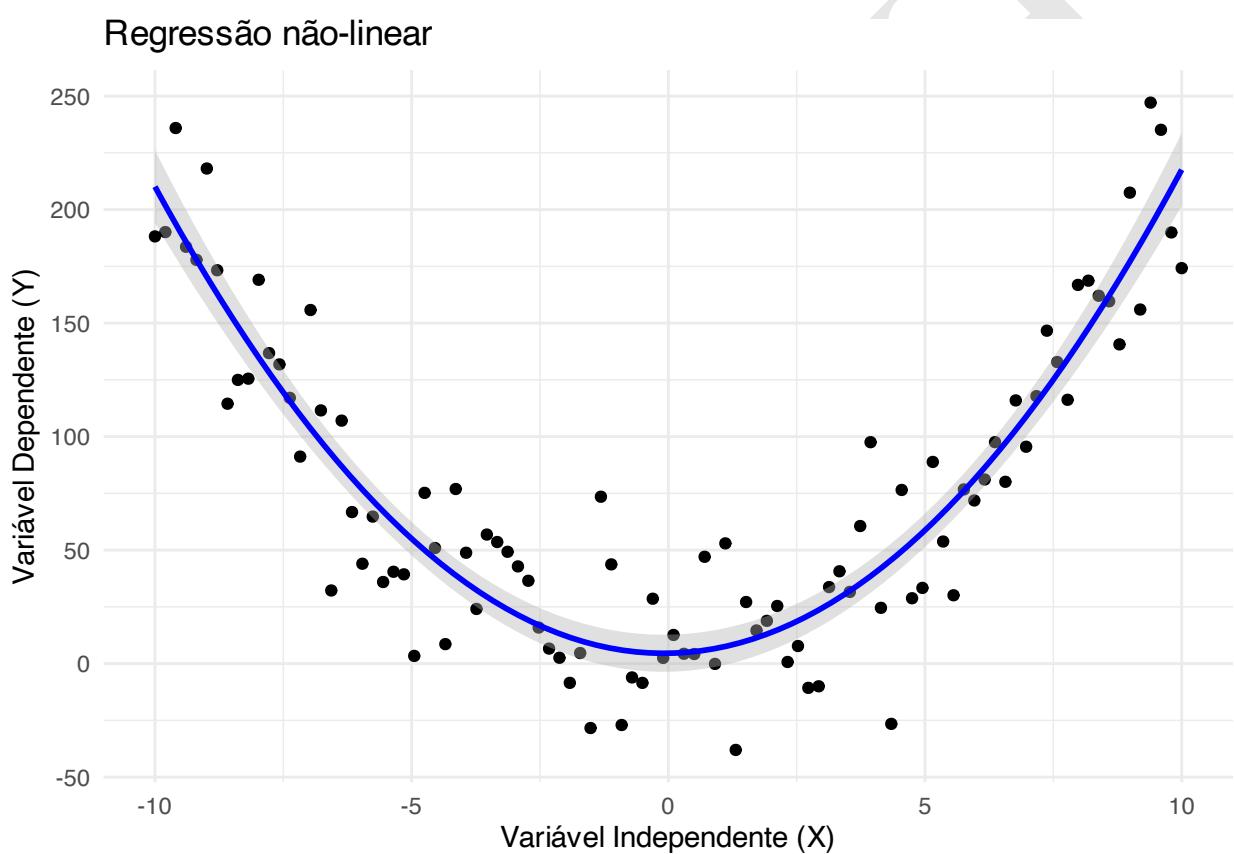


Figura 33.3: Regressão não-linear.

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X + \dots + \beta_n X_n \quad (33.8)$$

- A ligação (*link*) usada é o logit (33.9).?

$$g(p) = \log\left(\frac{p}{1-p}\right) \quad (33.9)$$

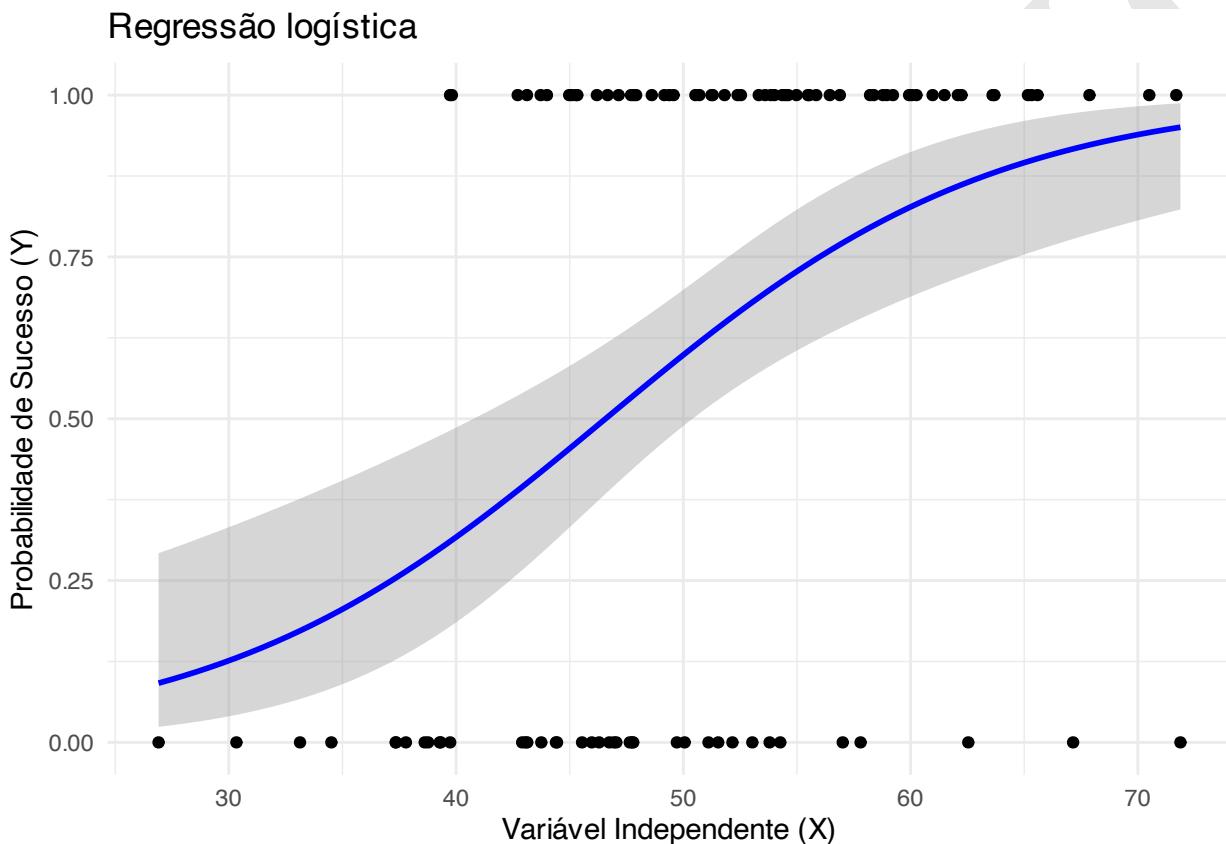


Figura 33.4: Regressão logística.

33.3.5 O que são modelos de regressão multinomial?

- Modelos de regressão multinomial são usados quando a variável resposta é categórica com mais de dois níveis não ordenados.?
- Estendem a regressão logística binária, modelando as razões de chances (*odds ratios*) de cada categoria em relação a uma categoria de referência.?

33.3.6 O que são modelos de regressão ordinal?

- Modelos de regressão ordinal são usados quando a variável resposta é categórica com mais de dois níveis ordenados.?
- Modelam a probabilidade acumulada de estar em ou abaixo de cada categoria, usando uma função de ligação logit, probit ou log-log.?

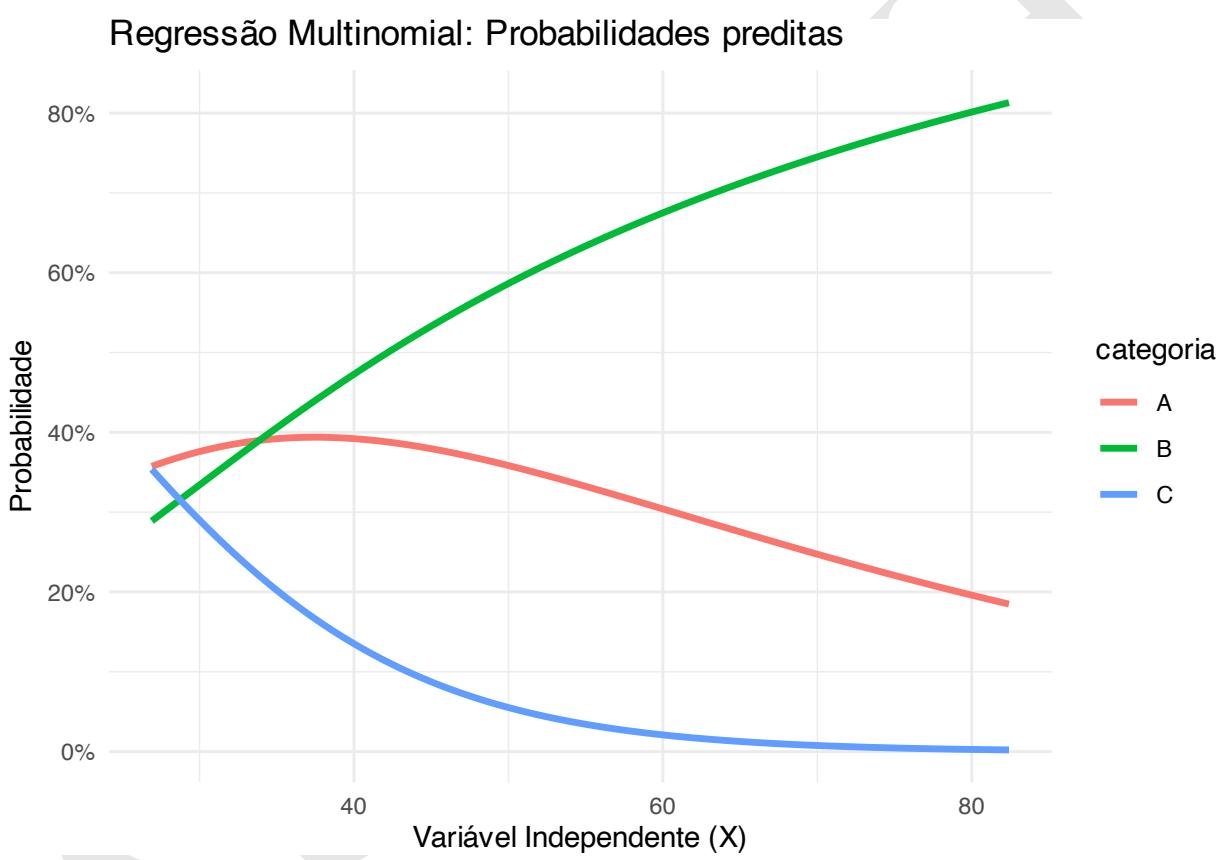


Figura 33.5: Regressão multinomial

- Assumem a proporcionalidade dos coeficientes entre as categorias (*proportional odds*).?

33.3.7 O que são modelos de regressão de Poisson?

- Modelos de regressão de Poisson são usados quando a variável resposta é uma contagem de eventos não negativos?
- Assumem que $Y \sim Poisson(\mu)$, com $\mu = E[Y|X]$ relacionado aos preditores via função de ligação log.?
- A sobre-dispersão (variância maior que a média) pode exigir modelos alternativos como a regressão binomial negativa.?

Régressão de Poisson (link log)

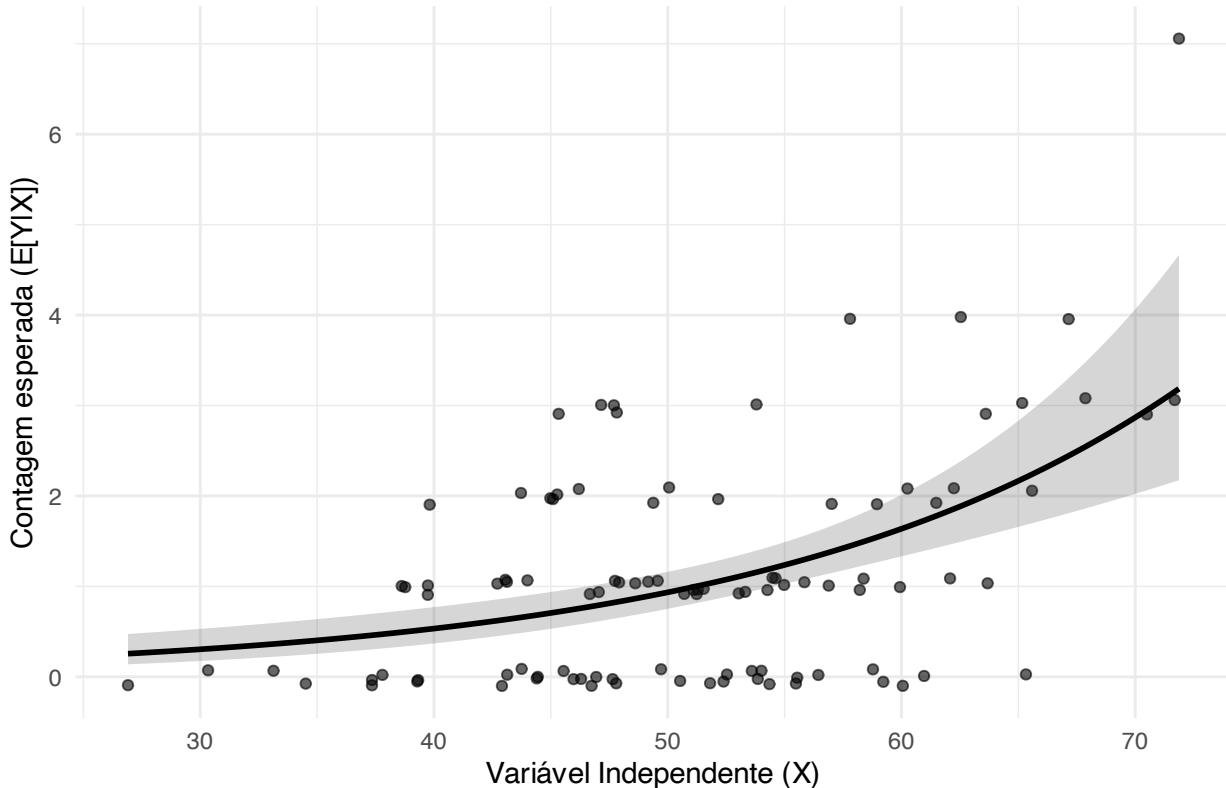


Figura 33.6: Regressão de Poisson.

33.3.8 O que são modelos de regressão binomial negativa?

- Modelos de regressão binomial negativa são usados para contagens superdispersas, onde a variância excede a média.?
- Introduzem um parâmetro de dispersão adicional para modelar a variabilidade extra.?
- A função de ligação log é comumente usada, semelhante à regressão de Poisson.?

33.3.9 O que são modelos de regressão Gama?

- Modelos de regressão Gama são usados para variáveis resposta contínuas e positivas, frequentemente com distribuição assimétrica.?

- A função de ligação log é comumente usada para garantir previsões positivas.³⁰⁵

33.3.10 O que são modelos de regressão com efeitos mistos?

- Modelos de efeitos mistos incorporam efeitos fixos (coeficientes comuns a todos os indivíduos) e efeitos aleatórios (variações específicas de grupos ou indivíduos).³⁰⁶
- Usados para dados hierárquicos ou longitudinais, onde observações estão agrupadas.³⁰⁷
- Permitem modelar correlações intra-grupo e variabilidade entre grupos.³⁰⁸

33.3.11 O que são modelos de regressão com efeitos mistos generalizados?

- Modelos de efeitos mistos generalizados (GLMM) estendem os modelos de efeitos mistos para variáveis resposta que seguem distribuições da família exponencial (ex.: binomial, Poisson).³⁰⁹
- Combinam a flexibilidade dos modelos lineares generalizados com a capacidade de modelar correlações e variabilidade entre grupos.³¹⁰
- Usados para dados hierárquicos ou longitudinais com desfechos não normais.³¹¹

33.3.12 O que são modelos de regressão ridge?

- Regressão ridge é um modelo linear regularizado que adiciona uma penalização L2 à soma dos quadrados dos coeficientes.³¹²
- Ajuda a reduzir multicolinearidade e overfitting, encolhendo os coeficientes em direção a zero, mas nunca os tornando exatamente nulos.³¹³
- O hiperparâmetro de regularização é λ , controlando a intensidade da penalização. Valores maiores de λ resultam em maior encolhimento dos coeficientes.³¹⁴

33.3.13 O que são modelos de regressão LASSO?

- Regressão LASSO (*Least Absolute Shrinkage and Selection Operator*) utiliza penalização L1, que pode zerar coeficientes.³¹⁵
- Além de reduzir overfitting, também realiza seleção automática de variáveis.³¹⁶
- Enquanto a regressão ridge mantém todos os preditores, a LASSO pode excluir variáveis irrelevantes.³¹⁷

33.4 Preparação de variáveis

33.4.1 Como preparar as variáveis categóricas para análise de regressão?

- Variáveis fictícias (*dummy*) compreendem variáveis criadas para introduzir, nos modelos de regressão, informações contidas em outras variáveis que não podem ser medidas em escala numérica.³⁰⁶
- Variáveis categóricas nominais, com 2 ou mais níveis, devem ser subdivididas em variáveis fictícias dicotômicas para ser usada em modelos de regressão.³⁰⁷
- Cada nível da variável categórica nominal será convertido em uma nova variável fictícia dicotômica, tal que a nova variável dicotômica assume valor 1 para a presença do nível correspondente e 0 em qualquer outro caso.³⁰⁷

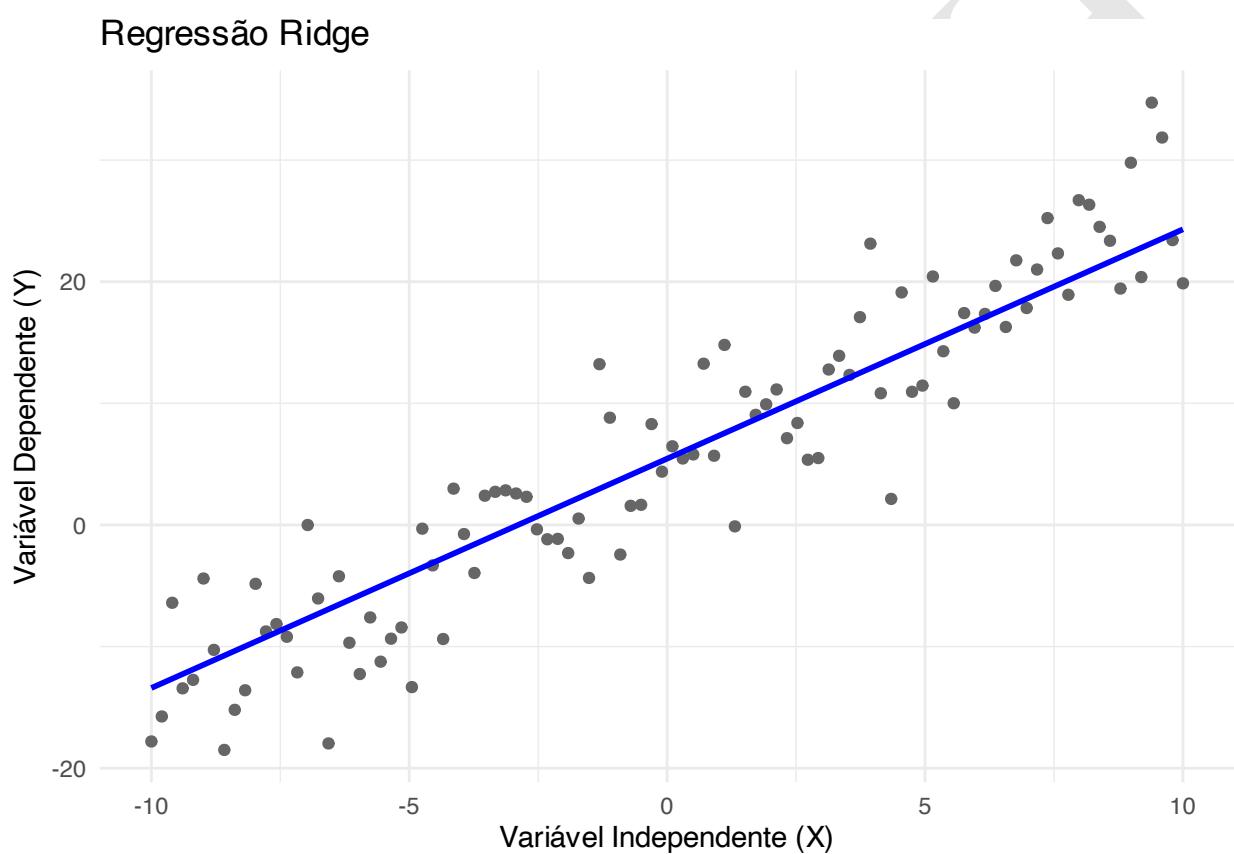


Figura 33.7: Regressão ridge.



O pacote *fastDummies*³⁰⁸ fornece a função *dummy_cols*^a para preparar as variáveis categóricas fictícias para análise de regressão.

^ahttps://www.rdocumentation.org/packages/fastDummies/versions/1.7.3/topics/dummy_columns

33.4.2 Por que é comum escolher a categoria mais frequente como referência em modelos epidemiológicos?

- Maior estabilidade estatística: a categoria mais frequente costuma gerar estimativas mais estáveis, com menor erro padrão nos coeficientes das demais categorias.³
- A escolha da referência não altera o ajuste nem o valor predito pelo modelo — apenas muda o ponto de comparação.³

33.5 Multicolinearidade

33.5.1 O que é multicolinearidade?

- Multicolinearidade representa a intercorrelação entre as variáveis independentes (explanatórias) de um modelo.³⁰²

33.5.2 Como diagnosticar multicolinearidade de forma quantitativa?

- Verifique a existência de multicolinearidade entre as variáveis candidatas.³⁰⁹
- O Coeficiente de determinação (R^2) é uma medida de quão bem as variáveis independentes explicam a variabilidade da variável dependente. Valores próximos a 1 indicam que as variáveis independentes estão fortemente correlacionadas entre si, o que pode indicar multicolinearidade.³⁰²
- O Fator de Inflação da Variância (*variance inflation factor*, VIF) é uma medida que quantifica o quanto a variância de um coeficiente de regressão é inflacionada devido à multicolinearidade. Valores de VIF maiores que 10 são frequentemente considerados indicativos de multicolinearidade significativa.³⁰²
- O recíproco da VIF é chamado de Tolerância, que mede a proporção da variância de uma variável independente que não é explicada pelas outras variáveis independentes. Valores baixos de Tolerância (geralmente abaixo de 0.1) indicam multicolinearidade.³⁰²
- O número de condições (*Condition Number*) é uma medida que avalia a estabilidade numérica de um modelo de regressão. Valores altos (entre 10 de 30) indicam multicolinearidade, e valores maiores que 30 indicam forte multicolinearidade.³⁰²



O pacote *GGally*³⁰³ fornece a função *ggpairs*^a para criar uma matriz gráfica de correlações bivariadas.

^a<https://www.rdocumentation.org/packages/GGally/versions/2.2.1/topics/ggpairs>



O pacote *car*³¹⁰ fornece a função *vif*^a para calcular o fator de inflação da variância (VIF).

^a<https://www.rdocumentation.org/packages/car/versions/3.1-3/topics/vif>

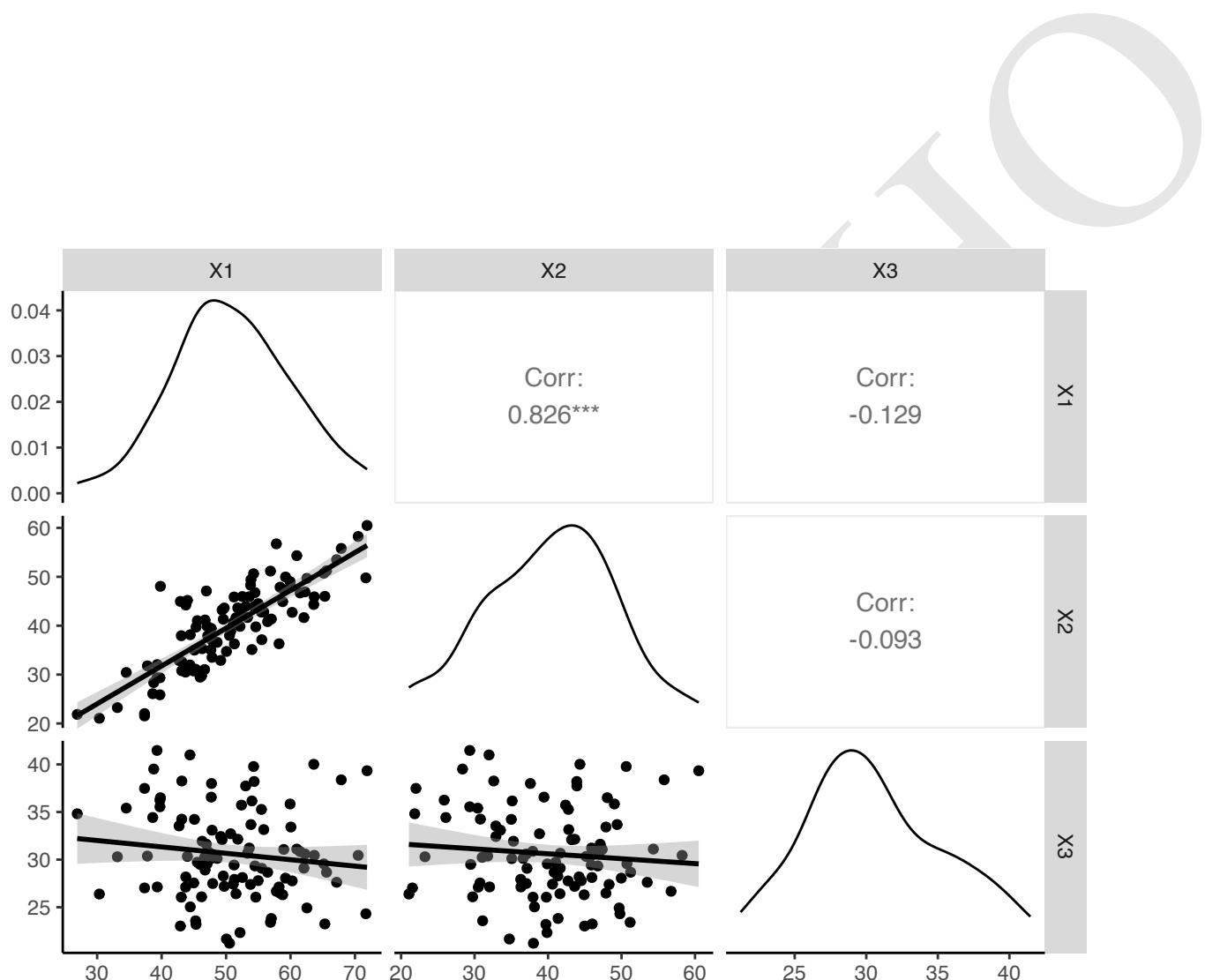


Figura 33.8: Multicolinearidade entre variáveis candidatas em modelos de regressão multivariável.

33.5.3 O que fazer em caso de multicolinearidade elevada?

- Verifique a transformação (codificação) de variáveis numéricas em categóricas.³⁰²
- Aumente o tamanho da amostra, se possível, para reduzir a multicolinearidade.³⁰²
- Combine níveis de variáveis categóricas com baixa frequência de ocorrência.³⁰²
- Combine variáveis numéricas altamente correlacionadas em uma única variável composta, como a média ou soma das variáveis.³⁰²
- Considere a exclusão de variáveis altamente correlacionadas do modelo, especialmente se elas não forem essenciais para a análise.³⁰²
- Use técnicas de seleção de variáveis, como seleção passo a passo, para identificar e remover variáveis redundantes.³⁰²
- Use técnicas de regularização, como regressão ridge ou LASSO, que podem lidar com multicolinearidade ao penalizar coeficientes de regressão.³⁰²

33.6 Redução de dimensionalidade

33.6.1 Correlação bivariada pode ser usada para seleção de variáveis em modelos de regressão multivariável?

- Seleção bivariada de variáveis — isto é, aplicação de testes de correlação em pares de variáveis candidatas e variável de desfecho afim de selecionar quais serão incluídas no modelo multivariável — é um dos erros mais comuns na literatura.^{288,309,311}
- A seleção bivariada de variáveis torna o modelo mais suscetível a otimismo no ajuste se as variáveis de confundimento não são adequadamente controladas.^{309,311}

33.6.2 Variáveis sem significância estatística devem ser excluídas do modelo final?

- Eliminar uma variável de um modelo significa anular o seu coeficiente de regressão ($\beta = 0$), mesmo que o valor estimado pelos dados seja outro. Desta forma, os resultados se afastam de uma solução de máxima verossimilhança (que tem fundamento teórico) e o modelo resultante é intencionalmente subótimo.²⁸⁸
- Os coeficientes de regressão geralmente dependem do conjunto de variáveis do modelo e, portanto, podem mudar de valor (“mudança na estimativa” positiva ou negativa) se uma (ou mais) variável(is) for(em) eliminada(s) do modelo.²⁸⁸

33.6.3 Por que métodos de regressão gradual não são recomendados para seleção de variáveis em modelos de regressão multivariável?

- Métodos diferentes de regressão gradual podem produzir diferentes seleções de variáveis de um mesmo banco de dados.³⁰⁷
- Nenhum método de regressão gradual garante a seleção ótima de variáveis de um banco de dados.³⁰⁷
- As regras de término da regressão baseadas em P-valor tendem a ser arbitrárias.³⁰⁷

33.6.4 O que pode ser feito para reduzir o número de variáveis candidatas em modelos de regressão multivariável?

- Em caso de uma proporção baixa entre o número de participantes e de variáveis, use o conhecimento prévio da literatura para selecionar um pequeno conjunto de variáveis candidatas.³⁰⁹

- Colapse categorias com contagem nula (células com valor igual a 0) de variáveis candidatas.³⁰⁹
- Use simulações de dados para identificar qual(is) variável(is) está(ão) causando problemas de convergência do ajuste do modelo.³⁰⁹
- A eliminação retroativa tem sido recomendada como a abordagem de regressão gradual mais confiável entre aquelas que podem ser facilmente alcançadas com programas de computador.²⁸⁸

33.6.5 Quando devemos forçar uma variável no modelo?

- Sempre que houver base teórica ou evidência prévia forte, ou se for a variável de exposição principal.²⁵¹

33.7 Seleção de variáveis em regressão

33.7.1 O que é seleção de variáveis em regressão?

- Seleção de variáveis em regressão consiste em identificar, dentre um conjunto de preditores disponíveis, quais devem ser incluídos no modelo para otimizar o equilíbrio entre ajuste e parcimônia.³¹²

33.7.2 Quais são os principais critérios de informação usados na seleção de variáveis?

- Critérios de informação avaliam o ajuste do modelo penalizando a complexidade (número de preditores), ajudando a evitar *overfitting*.³¹²
- R^2_{adj} (33.10) penaliza o R^2 pelo número de preditores, reduzindo o viés em modelos com muitas variáveis, onde n é o tamanho amostral, k o número de preditores, RSS a soma dos quadrados dos resíduos e SST a soma total dos quadrados.

$$R^2_{adj} = 1 - \frac{(n-1)}{(n-k-1)} \cdot \frac{RSS}{SST} \quad (33.10)$$

- *AIC* (Akaike Information Criterion) (33.11) mede o equilíbrio entre ajuste e complexidade:

$$AIC = n \cdot \log\left(\frac{RSS}{n}\right) + 2k + n + n \cdot \log(2\pi) \quad (33.11)$$

- *AICc* (33.12) é uma versão corrigida do AIC, preferida para amostras pequenas:

$$AIC_c = AIC + \frac{2(k+2)(k+3)}{n-(k+2)-1} \quad (33.12)$$

- C_p de Mallows compara o erro do modelo reduzido com o modelo completo, idealmente satisfazendo $C_p \approx p$, onde m é o número total de preditores disponíveis, p o número de parâmetros (incluindo o intercepto), e RSS_{FULL} o erro quadrático residual do modelo completo:

$$C_p = (n-m-1) \frac{RSS}{RSS_{FULL}} - (n-2p) \quad (33.13)$$

- *BIC* (Bayesian Information Criterion) (33.14) penaliza fortemente modelos complexos:

$$BIC = n \cdot \log\left(\frac{RSS}{n}\right) + k \cdot \log(n) + n + n \cdot \log(2\pi) \quad (33.14)$$

33.7.3 Quais algoritmos podem ser usados para seleção automática?

- Seleção progressiva (*forward selection*): começa com o modelo nulo e adiciona, a cada iteração, a variável que mais melhora o critério escolhido. O processo para quando nenhuma nova variável melhora o modelo.³¹²
- Eliminação retrógrada (*backward elimination*): parte do modelo completo e remove, a cada iteração, a variável cuja exclusão mais melhora o critério. O processo para quando nenhuma remoção melhora o ajuste.³¹²
- *Leaps-and-bounds*: método exato que examina apenas uma fração dos 2^m modelos possíveis, determinando os melhores subconjuntos para cada tamanho de preditor (usando os critérios AIC, BIC, AICC, R² ajustado e Cp).³¹²
- Esses métodos podem divergir em presença de alta multicolinearidade ou amostras pequenas, e devem ser acompanhados de diagnóstico de resíduos e validação cruzada.³¹²



O pacote *leaps*³¹³ fornece a função *regsubsets*^a para realizar os métodos de seleção de variáveis.

^a<https://www.rdocumentation.org/packages/leaps/versions/3.1/topics/regsubsets>



O pacote *olsrr*³¹⁴ fornece a função *ols_step_all_possible*^a para testar todos os subconjuntos de potenciais preditores de uma regressão.

^ahttps://www.rdocumentation.org/packages/olsrr/versions/0.6.0/topics/ols_step_all_possible



O pacote *olsrr*³¹⁴ fornece a função *ols_step_best_subset*^a para selecionar o melhor de todos os subconjuntos de potenciais preditores de uma regressão, de acordo com critérios objetivos.

^ahttps://www.rdocumentation.org/packages/olsrr/versions/0.6.0/topics/ols_step_best_subset

RASCUNHO

Capítulo 34

Redes

34.1 Análise de redes

34.1.1 O que é análise de rede?

- ?

RASCUNHO

PARTE 6: MODELAGEM

Estratégias para entender relações complexas, prever resultados e explorar padrões ocultos

RASCUNHO

Capítulo 35

Modelos

35.1 Modelos

35.1.1 O que são modelos?

- Modelos são representações simplificadas de um sistema real, usados para entender, prever ou controlar fenômenos complexos.³

35.1.2 O que é modelagem?

- Modelagem é o processo de usar dados para selecionar um modelo matemático explícito que represente o processo gerador dos dados.²⁵¹

35.1.3 Por que a escolha do modelo é complexa?

- Há inúmeras combinações possíveis de variáveis, formas funcionais (lineares, quadráticas, transformações), interações e formas do desfecho, o que torna o espaço de possibilidades muito amplo.²⁵¹
- Todos os modelos são errados, mas alguns são úteis.³¹⁵



O pacote *equatiomatic*³¹⁶ fornece a função *extract_eq*^a para extrair a equação dos modelos em formato LaTeX para visualização.

^ahttps://www.rdocumentation.org/packages/equatiomatic/versions/0.3.1/topics/extract_eq

35.1.4 O que diferencia modelos clássicos e modernos em predição?

- Modelos clássicos, como a regressão logística e as árvores de decisão, contrastam com os modelos modernos, como máquinas de vetor de suporte, redes neurais e *random forests*, principalmente pela maior flexibilidade e capacidade destes últimos de capturar não linearidades e interações.³¹⁷

35.2 Modelos estocásticos

35.2.1 O que são modelos estocásticos?

- ?

35.2.2 O que são cadeias de Markov?

- As cadeias de Markov descrevem processos em que o estado futuro depende apenas do estado presente, e não da trajetória passada.³¹⁸

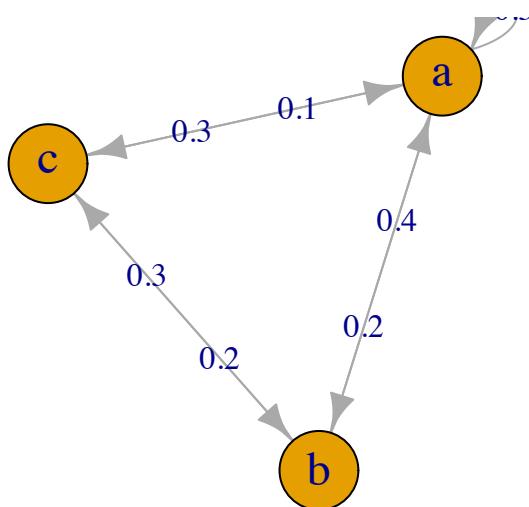


Figura 35.1: Cadeia de Markov com 3 estados (a, b, c) e suas probabilidades de transição.

R O pacote *markovchain*³¹⁹ fornece a função *markovchainFit*^a ajusta uma cadeia com base em dados observados.

^a<https://www.rdocumentation.org/packages/markovchain/versions/0.9.5/topics/createSequenceMatrix>

35.3 Efeito fixo

35.3.1 O que é efeito fixo?

- Efeito fixo é a relação média entre variáveis assumida como igual para todos os grupos ou indivíduos, representando o comportamento populacional esperado.[?]
- Ele descreve tendências sistemáticas e reproduzíveis que não dependem de pertencer a um grupo específico.[?]
- Em modelos estatísticos, corresponde aos parâmetros estimados globalmente a partir de todos os dados.[?]

35.4 Efeito aleatório

35.4.1 O que é efeito aleatório?

- Efeito aleatório representa desvios específicos de grupos ou unidades em relação ao efeito fixo.[?]
- Ele modela a variabilidade entre grupos, assumindo que esses desvios são amostras de uma distribuição comum.[?]
- Não busca estimar cada grupo isoladamente, mas sim quantificar a variabilidade entre eles.[?]

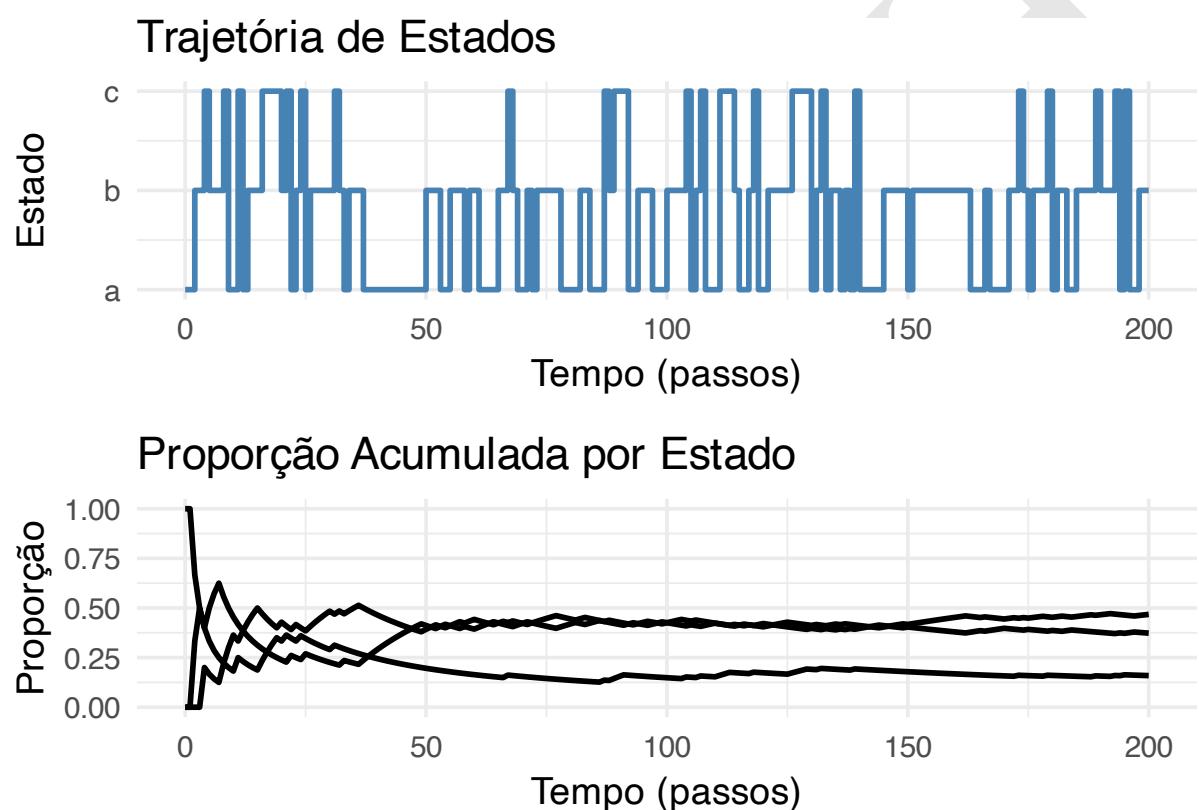


Figura 35.2: Trajetória de estados e proporção acumulada por estado em uma cadeia de Markov com 3 estados (a, b, c).

35.5 Efeito misto

35.5.1 O que é efeito misto?

- Um modelo de efeitos mistos combina efeitos fixos e aleatórios em uma única estrutura estatística.⁷
- Ele permite estimar tendências globais ao mesmo tempo em que ajusta variações específicas por grupo.⁸
- Essa combinação possibilita inferência correta mesmo na presença de heterogeneidade, evitando armadilhas como o Paradoxo de Simpson.⁹

Simpson's Paradox and Mixed Effects Models

Within-group trends, pooled inference, and partial pooling

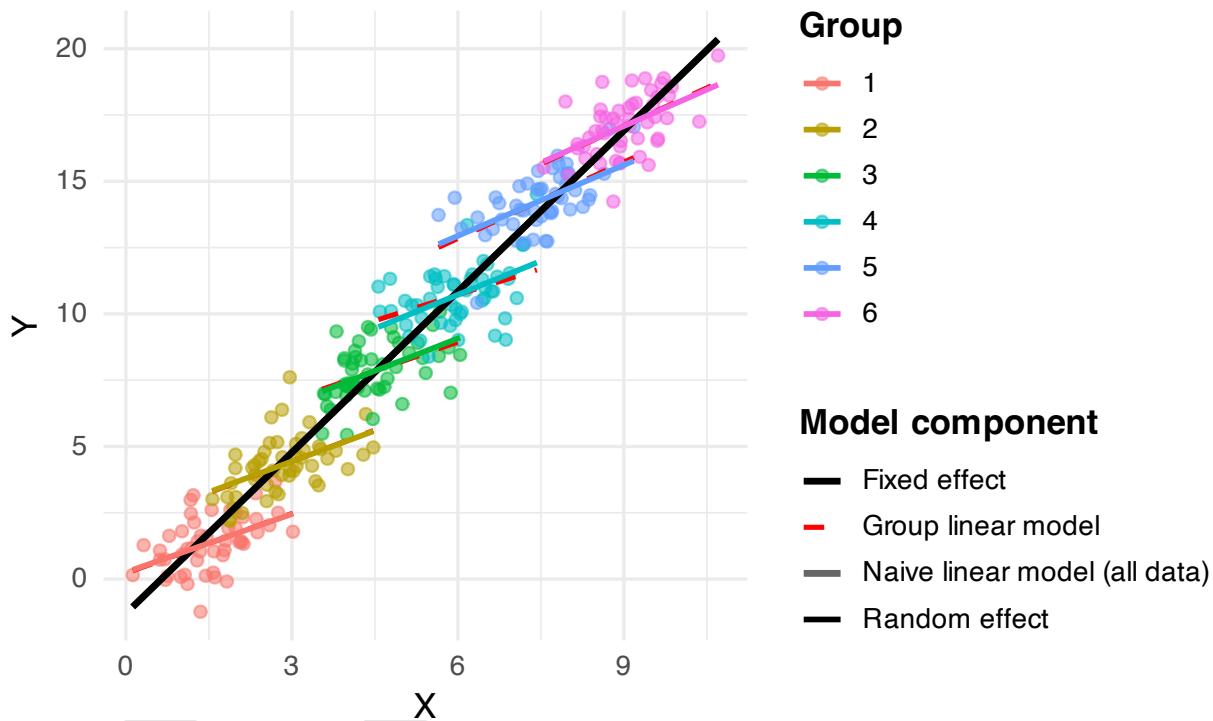


Figura 35.3: Efeitos fixos, aleatórios e mistos em dados simulados com paradoxo de Simpson. As linhas vermelhas representam os efeitos dentro dos grupos, enquanto as linhas cinza e preta representam os efeitos globais (naive e fixo, respectivamente). O modelo misto (linhas coloridas) captura os efeitos dentro dos grupos sem extrapolar além dos dados observados.

35.6 Efeito principal

35.6.1 O que é efeito principal?

- 320

35.7 Efeito de interação

35.7.1 O que é efeito de interação?

- A interação - representada pelo símbolo * - é o termo estatístico empregado para representar a heterogeneidade de um determinado efeito.³²¹
- ³²⁰

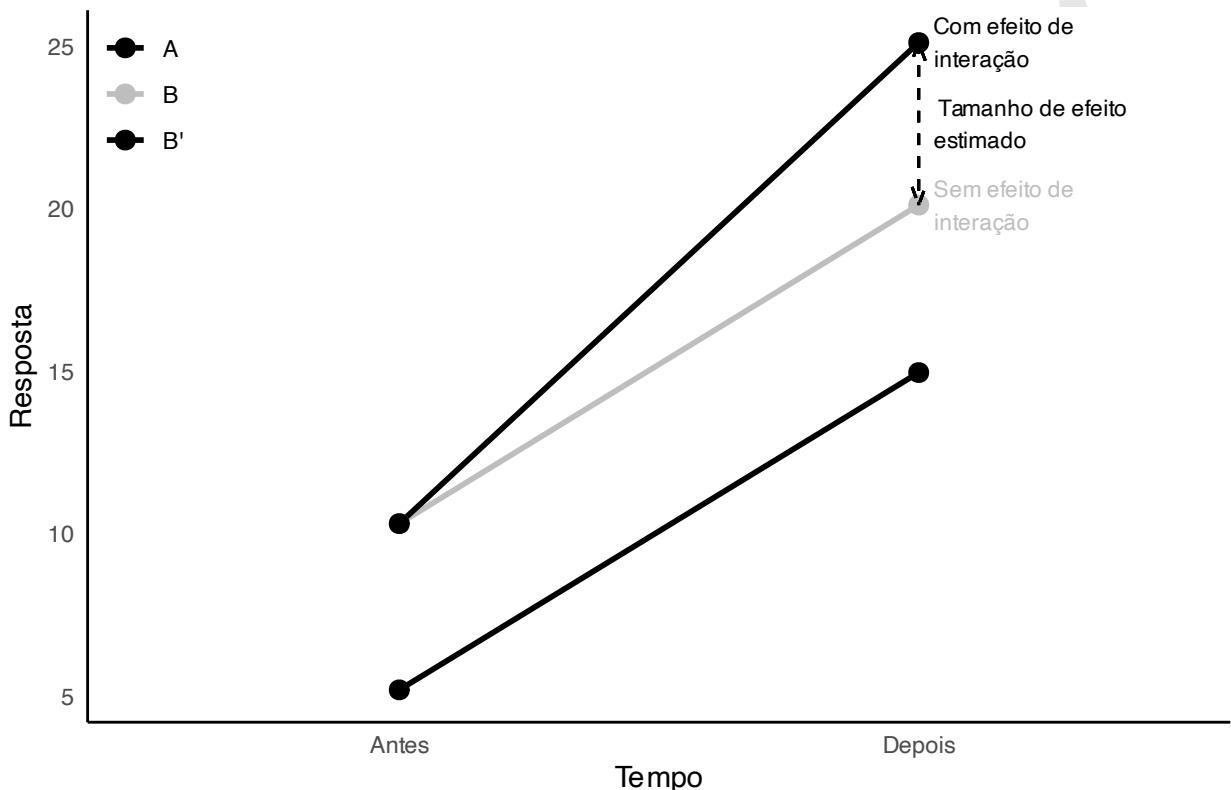


Figura 35.4: Análise de efeito de interação (direta) entre grupos e tempo. Retas paralelas sugerem ausência de efeito de interação.

R O pacote *nlme*³²² fornece a função *nlme*^a para ajustar um modelo de regressão misto não linear.

^a<https://www.rdocumentation.org/packages/nlme/versions/3.1-163/topics/nlme>

R O pacote *mmrm*³²³ fornece a função *mmrm*^a para ajuste de um modelo de regressão misto linear.

^a<https://rdrr.io/cran/mmrm/man/mmrm.html>

R O pacote *emmeans*³²⁴ fornece a função *emmeans*^a para calcular as médias marginais dos fatores e suas combinações de um modelo de regressão misto linear.

^a<https://www.rdocumentation.org/packages/emmeans/versions/1.8.7/topics/emmeans>

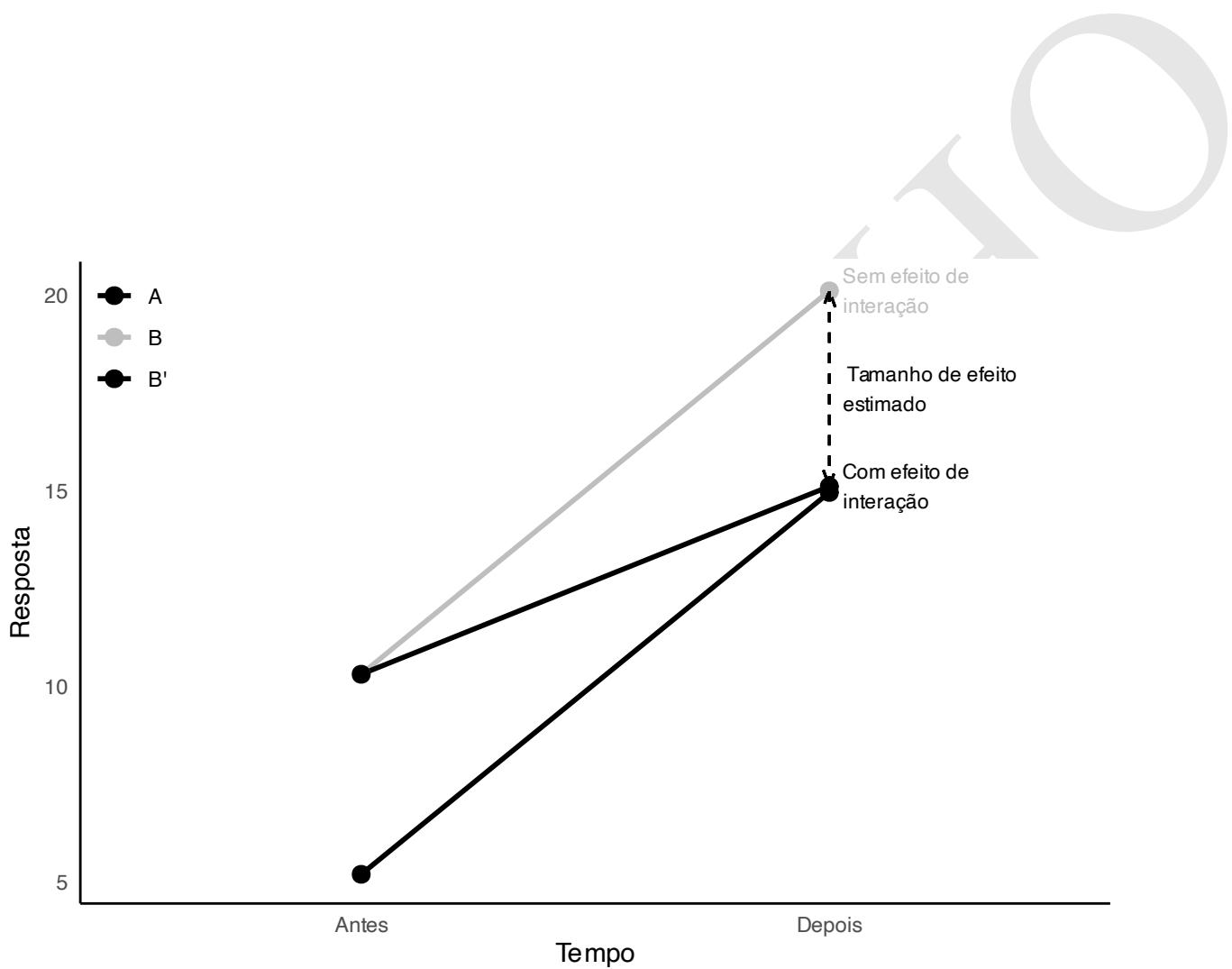


Figura 35.5: Análise de efeito de interação (inversa) entre grupos e tempo. Retas paralelas sugerem ausência de efeito de interação.

35.8 Efeito de mediação

35.8.1 O que é um mediador de efeito?

- ³²⁵
- ³²⁰

35.8.2 O que é efeito de mediação?

- ³²⁵
- ³²⁰

35.8.3 O que é efeito direto?

- ³²⁵
- ³²⁰

35.8.4 O que é efeito indireto?

- ³²⁵
- ³²⁰

35.8.5 O que é efeito total?

- ³²⁵
- ³²⁰

35.9 Efeito de modificação

35.9.1 O que é um modificador de efeito?

- ³²⁰

35.9.2 O que é efeito de modificação?

- ³²⁰

35.10 Preditores e desfechos

35.10.1 O que são desfechos de um modelo?

- ?

35.10.2 O que são preditores de um modelo?

- ?

35.10.3 Como selecionar preditores para um modelo?

- ?

35.11 Desempenho e estabilidade de modelos

35.11.1 Como avaliar o desempenho dos modelos?

- Pela área sob a curva ROC em conjunto com o otimismo (diferença entre AUC aparente e validada).³¹⁷
- O desempenho melhora com maior tamanho amostral, mas de forma desigual entre técnicas.³¹⁷

35.11.2 Qual modelo alcança estabilidade mais rapidamente?

- Regressão logística é o mais estável e menos *data hungry*.³¹⁷
- Árvore de decisão para classificação e regressão estabiliza rápido, mas em nível de desempenho baixo.³¹⁷
- Máquina de vetores de suporte, redes neurais e *random forests* apresentam instabilidade mesmo em amostras muito grandes.³¹⁷

35.12 Comparação de modelos

35.12.1 Como comparar modelos estatísticos?

- ?

35.12.2 Como comparar modelos de aprendizagem de máquina?

- ?



O pacote *correctR*³²⁶ fornece funções para comparar o desempenho e a qualidade do ajuste de diversos modelos de aprendizagem de máquina em amostras correlacionadas.

35.13 Avaliação de modelos

35.13.1 Como avaliar a qualidade de ajuste de um modelo?

- Coeficiente de determinação (R^2) (35.1) e R^2 ajustado (35.2): Medem a proporção da variabilidade dos dados explicada pelo modelo. O R^2 ajustado penaliza a inclusão de variáveis irrelevantes.³

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (35.1)$$

$$R^2_{ajustado} = 1 - (1 - R^2) \frac{n - 1}{n - p - 1} \quad (35.2)$$

- Erro quadrático médio (*RMSE*) (35.3): Mede a média dos erros ao quadrado entre os valores observados e os valores previstos pelo modelo, onde y_i são os valores observados, \hat{y}_i são os valores previstos pelo modelo, e n é o número de observações. Valores menores indicam melhor ajuste.³

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (35.3)$$

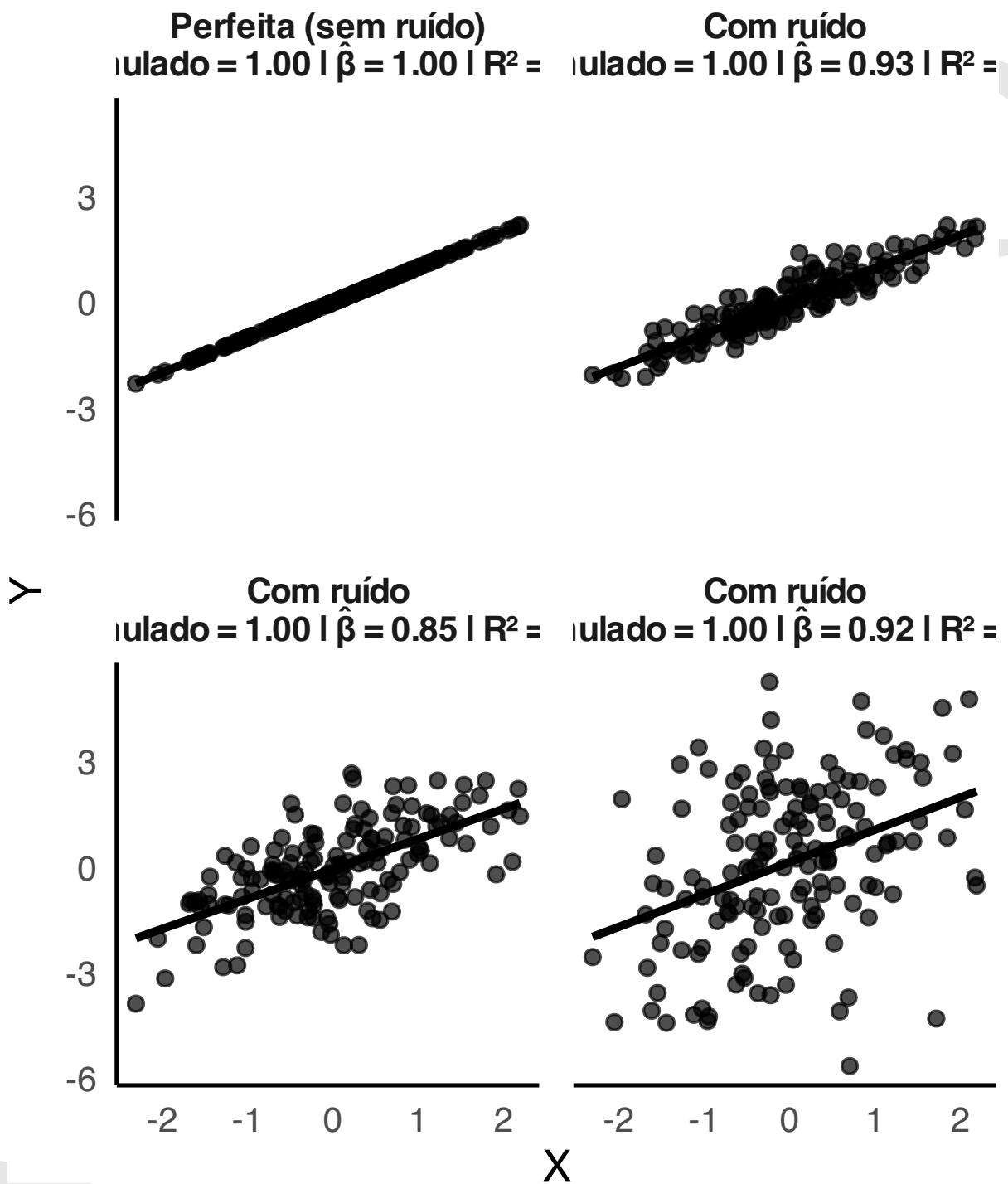


Figura 35.6: Exemplos de ajuste de modelos de regressão linear simples ($y \times x$) com diferentes níveis de ruído (R^2). Cada painel mostra a reta ajustada (cinza) e os valores observados (pontos). Os valores anotados indicam o coeficiente angular simulado (β), o coeficiente angular estimado ($\hat{\beta}$) e o R^2 observado.

Tabela 35.1: Métricas de desempenho do modelo de regressão linear.

Métrica	Valor
AIC	513.017
AIC corrigido	513.267
BIC	520.833
R^2	0.007
R^2 ajustado	-0.003
Erro quadrático médio (RMSE)	3.053
Desvio residual (sigma)	3.084

- Critério de Informação Akaike (*AIC*) (35.4) e Critério de Informação Bayesiano (*BIC*) (35.5): Avaliam o ajuste do modelo penalizando a complexidade (número de parâmetros), onde k é o número de parâmetros do modelo, L é a verossimilhança máxima do modelo, e n é o tamanho da amostra. Modelos com menor AIC ou BIC são preferíveis.²⁴⁸

$$AIC = 2k - 2 \ln(L) \quad (35.4)$$

$$BIC = \ln(n)k - 2 \ln(L) \quad (35.5)$$

- Desvio residual (σ): Mede a variabilidade dos resíduos do modelo. Valores menores indicam melhor ajuste.²⁴⁹



O pacote *performance*²⁴⁹ fornece a função *model_performance*^a para calcular as métricas de ajuste da regressão adequadas ao modelo pré-especificado.

^ahttps://www.rdocumentation.org/packages/performance/versions/0.10.4/topics/model_performance



O pacote *performance*²⁴⁹ fornece a função *compare_performance*^a para comparar o desempenho e a qualidade do ajuste de diversos modelos de regressão pré-especificados.

^ahttps://www.rdocumentation.org/packages/performance/versions/0.10.4/topics/compare_performance

35.14 Validação de modelos

35.14.1 Como validar modelos estatísticos?

- ?

35.15 Calibração de modelos

35.15.1 Como calibrar modelos estatísticos?

- ?

Capítulo 36

Modelagem temporal

36.1 Modelos temporais

36.1.1 O que são modelos temporais?

• ?



O pacote *UComp*³²⁷ fornece a função *UCsetup*^a para preparar modelos temporais gerais univariados.

^a<https://www.rdocumentation.org/packages/UComp/versions/5.1.5/topics/UCsetup>

RASCUNHO

Capítulo 37

Modelagem espacial

37.1 Modelos espaciais

37.1.1 O que são modelos espaciais?

• ?



Os pacotes *sf*³²⁸ e *ggplot*¹⁷³ fornecem a função *geom_sf*^a para visualização de dados espaciais em R.

^a<https://www.rdocumentation.org/packages/ggplot2/versions/3.4.3/topics/CoordSf>



O pacote *leaflet*^[329] fornece a função *leaflet*^a para criar um mapa interativo.

^a<https://www.rdocumentation.org/packages/leaflet/versions/2.2.2/topics/leaflet>

RASCUNHO

Capítulo 38

Modelagem de sobrevida

38.1 Sobrevida

38.1.1 O que é sobrevida?

- A sobrevida é um termo utilizado em estatística e análise de dados para descrever o tempo que decorre até a ocorrência de um evento específico, como a morte, a falha de um equipamento ou a recidiva de uma doença. Em estudos clínicos, por exemplo, a sobrevida pode referir-se ao tempo que um paciente vive após o diagnóstico de uma doença ou após o início de um tratamento.²

38.2 Análise de sobrevida

38.2.1 O que é análise de sobrevida?

- A análise de sobrevida é uma área da estatística que se concentra no estudo desses tempos até o evento, levando em consideração que nem todos os indivíduos podem ter experimentado o evento durante o período de estudo (censura).²
- Métodos comuns de análise de sobrevida incluem a estimativa da função de sobrevivência, a análise de Kaplan-Meier e modelos de regressão como o modelo de riscos proporcionais de Cox.²



O pacote *survival*³³⁰ fornece a função [survfit]<https://www.rdocumentation.org/packages/survival/versions/3.8-3/topics/survfit>) para criar curvas de sobrevida.

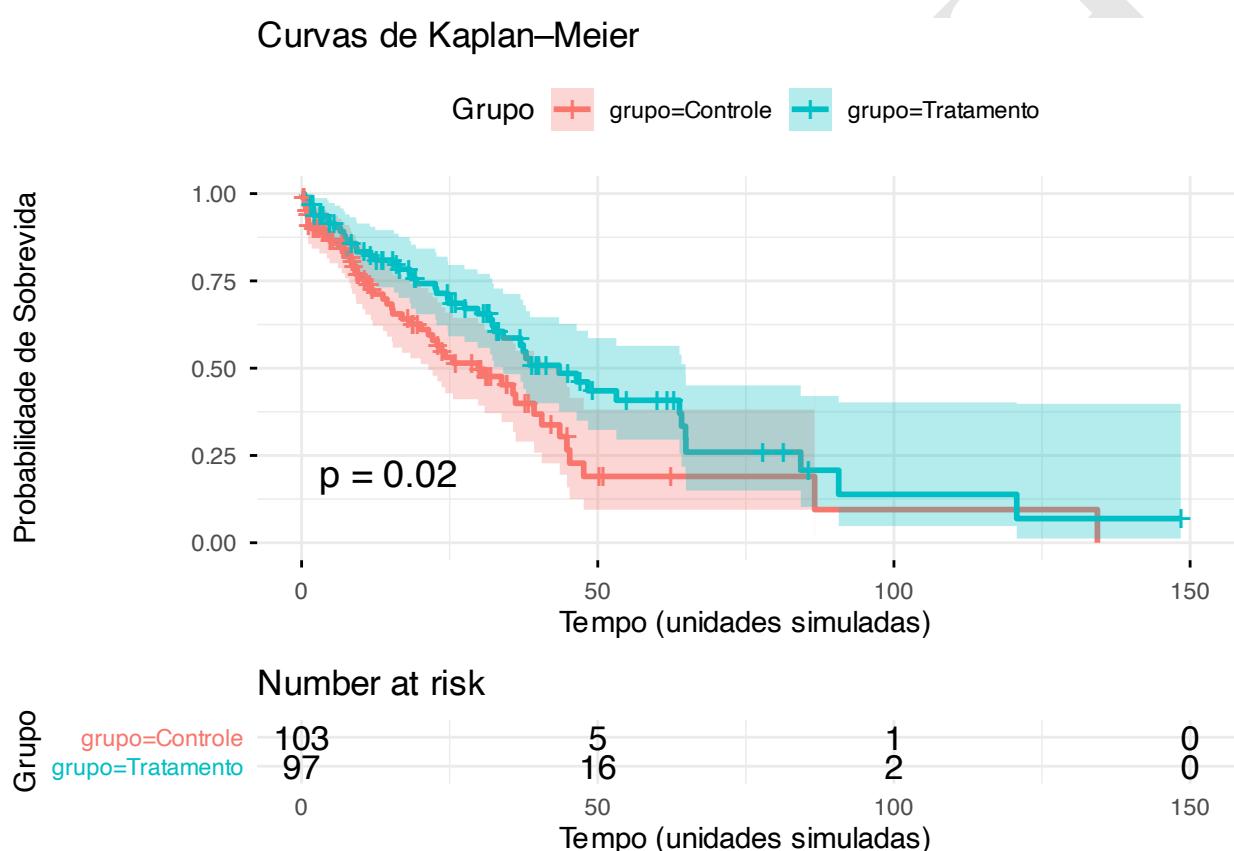


Figura 38.1: Curvas de Kaplan–Meier simuladas para dois grupos (controle e tratamento).

PARTE 7: REPRESENTAÇÃO, APRENDIZADO DE MÁQUINA E INTELIGÊNCIA ARTIFICIAL

Do avanço estatístico ao poder computacional: Métodos modernos para problemas complexos

RASCUNHO

Capítulo 39

Representações

39.1 Representações de dados e extração de atributos

39.1.1 Por que modelos não podem acessar o fenômeno diretamente?

• ?

39.1.2 Como um fenômeno do mundo real é traduzido em uma estrutura matemática?

• ?

39.1.3 De que forma a representação limita ou expande o que um modelo pode aprender?

• ?

39.1.4 O que é um atributo e por que ela não é um algoritmo?

• ?

39.1.5 Quais são as principais formas de representar diferentes tipos de dados?

• ?

39.1.6 O que se perde — e o que se ganha — ao escolher uma representação?

• ?

39.1.7 Como essas escolhas antecipam os desafios da inteligência artificial e do aprendizado de máquina?

• ?

RASCUNHO

Capítulo 40

Inteligência artificial

40.1 Inteligência artificial

40.1.1 O que é inteligência artificial (IA)?

- ?

40.1.2 Como ela se relaciona com estatística, ciência de dados e aprendizado de máquina?

- ?

40.2 IA generativa e grandes modelos de linguagem

40.2.1 O que são grandes modelos de linguagem (*large language models, LLM*)?

- ?

40.2.2 Como funcionam modelos como GPT, BERT e similares?

- ?

R

O pacote *keras@keras*^a possuu funções para criar, treinar e avaliar modelos de redes neurais.

^a<https://cran.r-project.org/web/packages/keras/index.html>

R

O pacote *tensorflow*^{331a} fornece uma interface para o TensorFlow, uma biblioteca de código aberto amplamente utilizada para aprendizado de máquina e redes neurais.

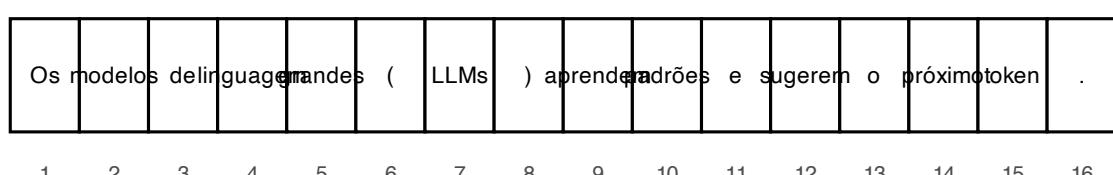
^a<https://cran.r-project.org/web/packages/tensorflow/index.html>

R

O pacote *torch*^{332a} permite criar e treinar redes neurais com alto desempenho.

^a<https://cran.r-project.org/web/packages/torch/index.html>

A) Sequência de tokens (após tokenização)



B) Atenção do último token: ":"



C) Próximo token (top-10) – distribuição toy

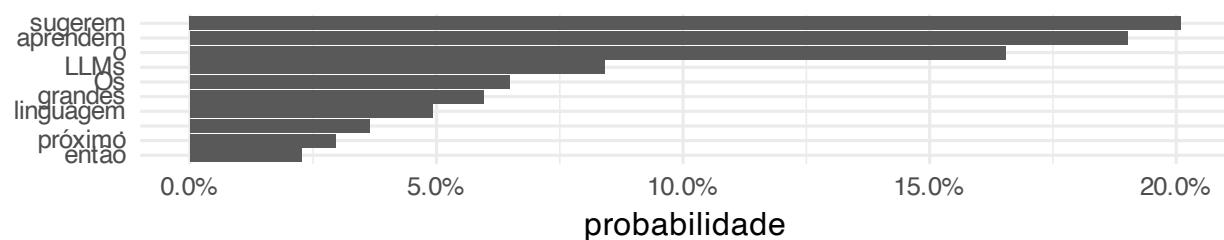


Figura 40.1: Representação esquemática de um modelo de linguagem grande (LLM)



O pacote *reticulate*^a integra R e Python em um mesmo ambiente de trabalho, permitindo chamar funções Python a partir de R e facilitar o uso de bibliotecas de IA disponíveis nesse ecossistema.

^a<https://cran.r-project.org/web/packages/reticulate/index.html>

RASCUNHO

Capítulo 41

Aprendizado de máquina

41.1 Aprendizado de máquina

41.1.1 O que é aprendizado de máquina?

- ?

41.2 Tipos de aprendizado

41.2.1 O que é aprendizado supervisionado?

- ?

41.2.2 O que é aprendizado não supervisionado?

- ?

41.2.3 O que é aprendizado semi-supervisionado?

- ?

41.2.4 O que é aprendizado por reforço?

- ?

41.2.5 O que é aprendizado profundo?

- ?

41.2.6 Quais são os limites do progresso em classificadores supervisionados?

- Os maiores ganhos de acurácia vêm de modelos simples, como análise discriminante linear; métodos mais sofisticados oferecem apenas ganhos marginais.³³³
- O aumento da complexidade do modelo traz retornos decrescentes em termos de redução da taxa de erro.³³³

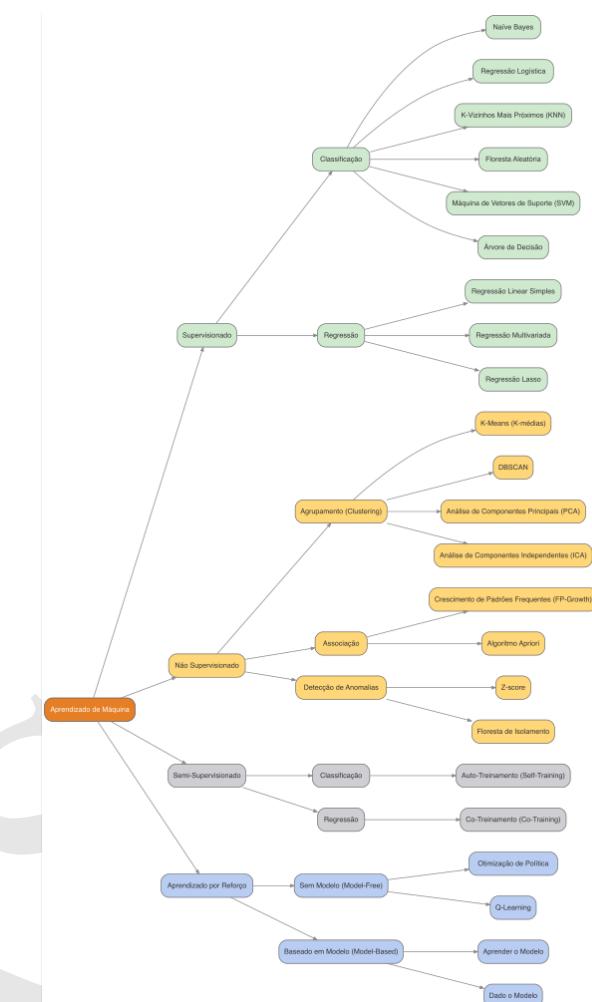


Figura 41.1: Mapa mental de algoritmos de aprendizado de máquina.

41.2.7 Quais problemas práticos limitam a generalização de classificadores?

- *Population drift*: mudanças na distribuição dos dados ao longo do tempo degradam a performance de modelos.³³³
- *Sample selectivity bias*: amostras de treino podem não representar a população futura, levando a superestimação de desempenho.³³³
- Erros de rótulo e definições arbitrárias de classes comprometem a validade dos modelos.³³³

41.2.8 Por que estudos comparativos entre classificadores podem ser enganosos?

- Resultados dependem da experiência do pesquisador com cada método, da escolha dos conjuntos de dados e do critério de avaliação usado.³³³
- Diferenças pequenas em acurácia frequentemente desaparecem quando se consideram incertezas reais de aplicação.³³³

41.3 Principais algoritmos

41.3.1 Quais são os principais algoritmos de aprendizado de máquina?

- Modelos de regressão não penalizados, modelos de regressão penalizados, modelos baseados em árvores, modelos baseados em vizinhos, redes neurais, máquinas de vetores de suporte, Naive Bayes e enssembles do tipo Superlearner.³³⁴
- Do ponto de vista matemático, redes neurais não contradizem a estatística clássica; elas a estendem, substituindo modelos explícitos por representações aprendidas.[?]

Modelos de regressão	Redes neurais artificiais	Papel conceitual
Variável preditora (x)	Neurônio de entrada	Informação observada fornecida ao modelo
Coeficiente (β)	Peso (w)	Intensidade e direção da influência da variável
Intercepto (β_0)	Viés (b)	Deslocamento da fronteira de decisão
Combinação linear ($\beta_0 + \sum \beta_i x_i$)	Soma ponderada ($\sum w_i x_i + b$)	Agregação das entradas antes da não linearidade
Função de ligação (link)	Função de ativação	Introdução de não linearidade
Regressão linear	Neurônio linear	Modelo puramente linear
Regressão logística	Perceptron com ativação sigmoide	Classificação binária probabilística
Log-odds	Entrada da função sigmoide	Escala interna antes da probabilidade
Predição (\hat{y})	Saída do neurônio	Resposta estimada do modelo
Função de perda	Função de perda (loss)	Quantificação do erro de predição
Máxima verossimilhança	Otimização da função de perda	Ajuste dos parâmetros do modelo
Gradiente da verossimilhança	Retropropagação (backpropagation)	Direção de atualização dos parâmetros
Regularização (L1, L2)	Penalização de pesos (weight decay)	Controle de complexidade e overfitting
Interações explícitas	Interações aprendidas implicitamente	Modelagem de efeitos combinados
Modelo interpretável	Modelo geralmente opaco	Trade-off entre interpretação e flexibilidade

41.4 Regressão logística

41.4.1 O que são é regressão logística?

• ?

41.5 Máquina de vetores de suporte

41.5.1 O que são máquinas de vetores de suporte?

• ?

41.6 *K-nearest neighbours*

41.6.1 O que é *K-nearest neighbours*?

• ?

41.7 *K-means Clustering*

41.7.1 O que é *K-means clustering*?

• ?

41.8 Árvores de decisão

41.8.1 O que são árvores de decisão?

- São modelos de aprendizado supervisionado que dividem os dados em ramos e folhas, representando regras de decisão de forma hierárquica.²³⁹
- Podem lidar eficientemente com grandes conjuntos de dados sem pressupor estrutura paramétrica complexa.²³⁸
- São aplicáveis a variáveis contínuas e discretas, tanto como preditoras quanto como desfechos.²³⁸

41.8.2 Quais são os principais usos de árvores de decisão?

- Seleção de variáveis relevantes em cenários com muitos preditores, como registros clínicos eletrônicos.²³⁸
- Avaliação da importância relativa das variáveis, com base na redução da pureza dos nós ou da acurácia ao remover variáveis.²³⁸
- Tratamento de valores ausentes, seja classificando-os como categoria própria ou imputando-os por previsão dentro da árvore.²³⁸
- Predição de novos casos a partir de dados históricos.²³⁸
- Manipulação de dados, colapsando categorias muito numerosas ou subdividindo variáveis contínuas assimétricas.²³⁸

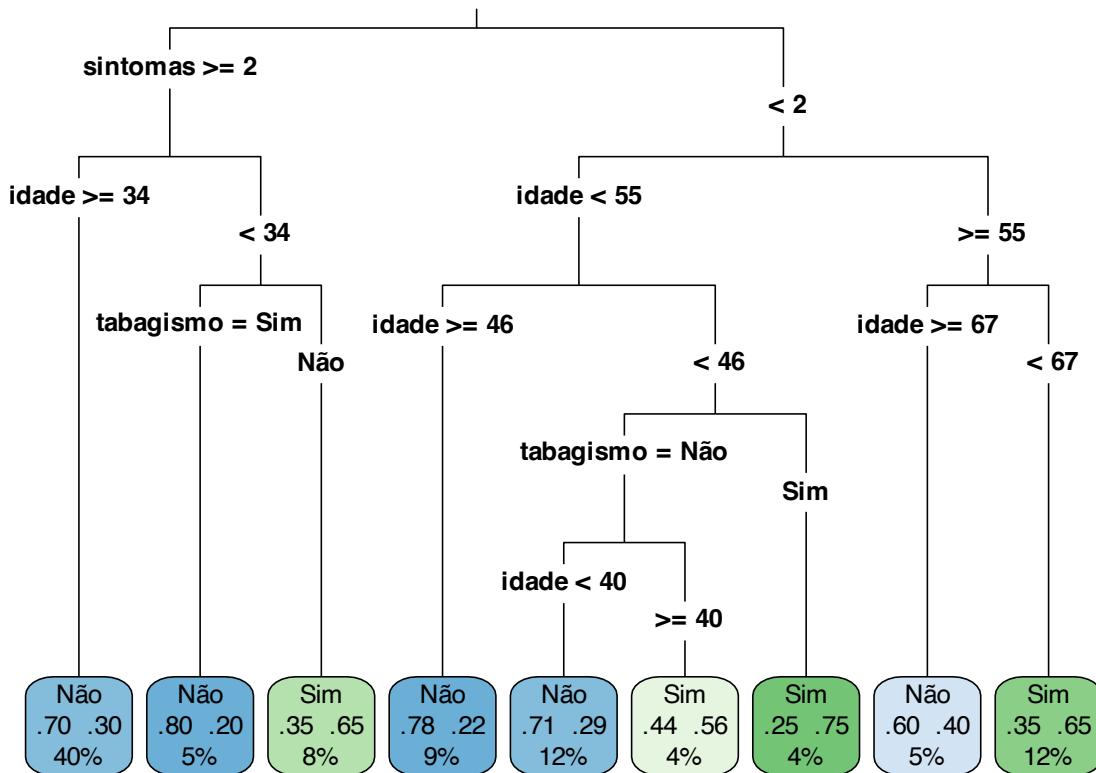


Figura 41.2: Exemplo de árvore de decisão para predizer depressão a partir de idade, tabagismo e sintomas.

41.8.3 Quais são os componentes básicos de uma árvore de decisão?

- Nós raiz (ou de decisão): subdividem todos os registros iniciais.²³⁸
 - Nós internos (ou de chance): representam subdivisões intermediárias.²³⁸
 - Nós folha (ou finais): resultados finais após sucessivas divisões.²³⁸
 - Ramos: representam condições “se-então”, ligando nós em sequência até a classificação final.²³⁸

41.8.4 Como funcionam splitting, stopping e pruning?

- *Splitting*: divide registros em subconjuntos mais homogêneos com base em métricas como entropia, índice de Gini e ganho de informação.²³⁸
 - *Stopping*: evita árvores excessivamente complexas ao definir parâmetros como número mínimo de registros por nó ou profundidade máxima.²³⁸
 - *Pruning*: reduz árvores grandes eliminando ramos pouco informativos, usando validação ou métodos como qui-quadrado.²³⁸

41.8.5 Quais são as vantagens e limitações de árvores de decisão?

- Vantagens: simplificam relações complexas; são intuitivas e fáceis de interpretar; não exigem pressupostos de distribuição; lidam bem com valores ausentes e dados enviesados; são robustas a *outliers*.²³⁸
 - Limitações: podem sofrer *overfitting* ou *underfitting* em amostras pequenas; podem selecionar variáveis correlacionadas sem relação causal real.²³⁸

41.8.6 Espaço de decisão em árvores de decisão vs. regressão logística

- A regressão logística assume relações lineares entre variáveis e log-odds.²³⁹
- Árvores de decisão permitem capturar relações não lineares e interações de forma automática.²³⁹

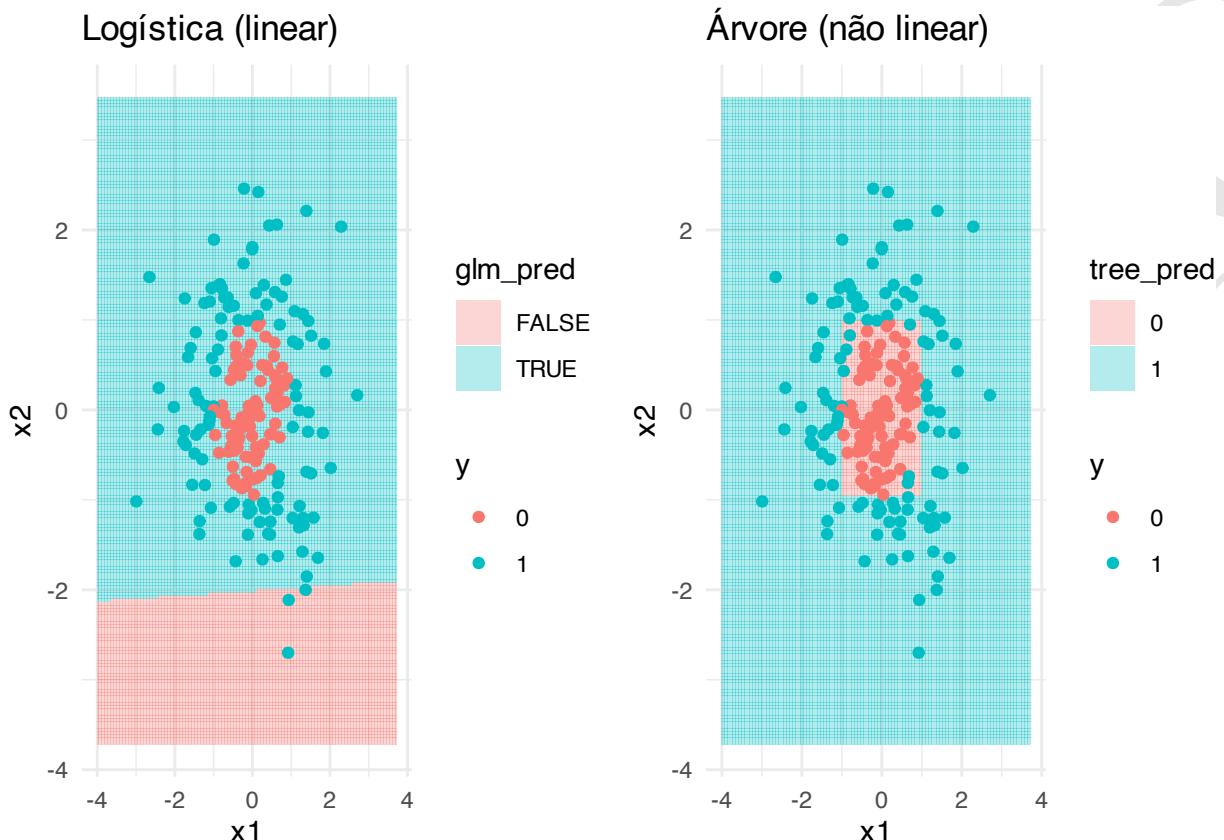


Figura 41.3: Comparação entre modelos de regressão logística e árvore de decisão.

R O pacote *h2o@correctR*^a fornece funções construir modelos de aprendizado de máquina.

^a<https://cran.r-project.org/web/packages/h2o/index.html>

R O pacote *correctR*³²⁶ fornece as funções *kfold_ttest*^a, *repkfold_ttest*^b e *resampled_ttest*^c para calcular estatística para comparação de modelos de aprendizado de máquina em amostras dependentes.

^a<https://cloud.r-project.org/web/packages/correctR/correctR.pdf>

^b<https://cloud.r-project.org/web/packages/correctR/correctR.pdf>

^c<https://cloud.r-project.org/web/packages/correctR/correctR.pdf>

R O pacote *caret@caret*^a fornece um conjunto de funções para pré-processamento, ajuste, avaliação e comparação de modelos de aprendizado de máquina.

^a<https://cran.r-project.org/web/packages/caret/index.html>



O pacote *mlr3@mlr3^a* fornece funções para fluxos de trabalho complexos, incluindo pré-processamento, ajuste de hiperparâmetros e integração com diversos algoritmos.

^a<https://cran.r-project.org/web/packages/mlr3/index.html>

41.9 Análise de componentes principais

41.9.1 O que é análise de componentes principais?

- ?

41.10 Random forests

41.10.1 O que são random forests?

- ?

41.11 Ensemble

41.11.1 O que são ensemble?

- ?

41.12 Desbalanceamento de classes

41.12.1 O que é desbalanceamento de classes (*class imbalance*)?

- Ocorre quando as classes do desfecho (por exemplo, presença vs. ausência de um evento) não estão igualmente representadas nos dados de treinamento.[?]

41.12.2 Por que o desbalanceamento é um problema?

- Modelos podem aprender a priorizar a classe mais frequente, obtendo alta acurácia global, mas baixo desempenho para a classe minoritária.[?]
- Isso pode comprometer métricas como sensibilidade, especificidade e, em alguns casos, a calibração.[?]

41.12.3 Quais são as abordagens mais comuns para lidar com desbalanceamento de classes?

- Reamostragem aleatória: superamostragem da classe minoritária; subamostragem da classe majoritária).[?]
- Ajuste de pesos: penaliza mais os erros na classe menos frequente.[?]
- Alteração do limiar de decisão: muda o ponto de corte de probabilidade para otimizar métricas específicas.[?]

41.12.4 Qual é o impacto do desbalanceamento de classes na calibração de modelos?

- Corrigir o desbalanceamento de classes nem sempre melhora a calibração e, em alguns casos, pode piorá-la.³³⁵

- Em simulações computacionais, modelos sem correção tiveram calibração igual ou superior aos corrigidos.³³⁵
- A piora observada foi caracterizada por superestimação do risco, nem sempre reversível com recalibração.³³⁵

Capítulo 42

Redes neurais

42.1 Neurônios artificiais

42.1.1 O que são neurônios artificiais?

- Neuronios artificiais (ou perceptrons) são modelos matemáticos que imitam o funcionamento dos neurônios biológicos, recebendo entradas, aplicando pesos e uma função de ativação para produzir uma saída.³³⁶⁻³³⁸
- A equação geral de um neurônio artificial é dada por (42.1), onde x_i são as entradas, w_i os pesos, b o viés e ϕ a função de ativação:

$$y = \phi \left(\sum_{i=1}^d w_i x_i + b \right) \quad (42.1)$$

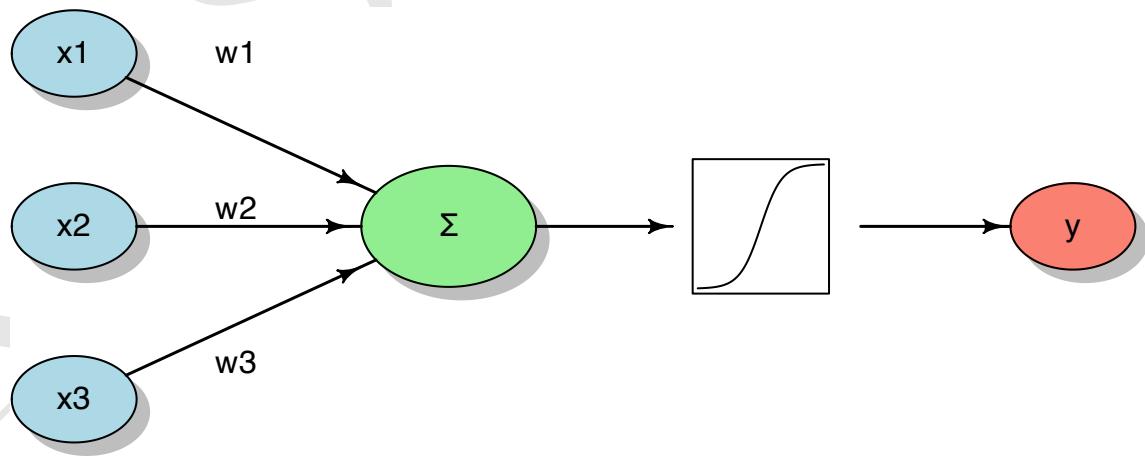


Figura 42.1: Representação esquemática de um neurônio computacional.

42.2 Rede neural artificial

42.2.1 O que é uma rede neural artificial?

- Redes neurais artificiais são modelos computacionais compostos por camadas de neurônios artificiais interconectados, nos quais cada camada aplica transformações lineares seguidas de funções não lineares, permitindo a aproximação de relações complexas entre variáveis de entrada e saída.⁷

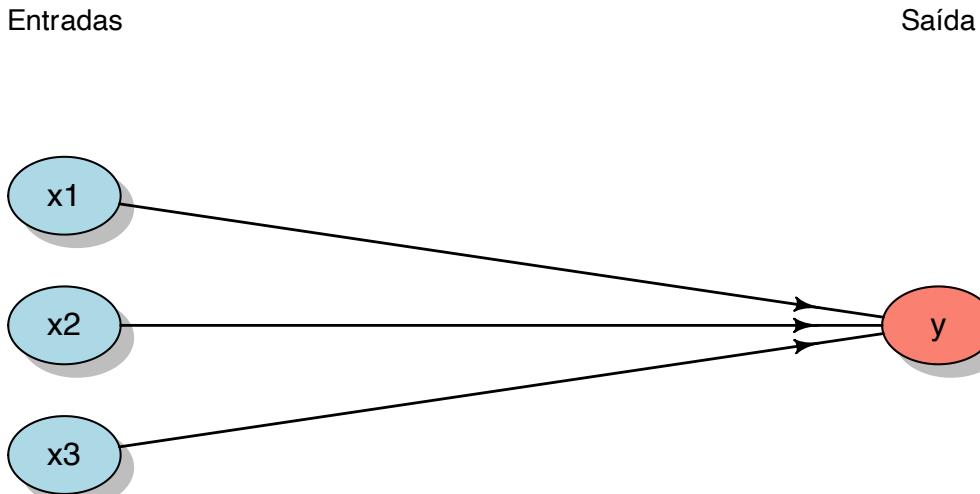


Figura 42.2: Representação esquemática de uma rede neural simples com camada de entrada e saída.

R O pacote *neuralnet*³³⁹ fornece a função *neuralnet*^a para treinar redes neurais artificiais.

^a<https://www.rdocumentation.org/packages/neuralnet/versions/1.44.2/topics/neuralnet>

42.3 Funções de ativação

42.3.1 Quais são as funções de ativação mais comuns?

- As funções de ativação introduzem não-linearidades nas redes neurais, permitindo que aprendam padrões complexos, como sigmoide (42.2), tangente hiperbólica (42.3) e unidade linear retificada (ReLU) (42.4).⁷

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (42.2)$$

$$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (42.3)$$

$$\text{ReLU}(z) = \max(0, z) \quad (42.4)$$

- Ao manter gradientes constantes na região positiva, a ReLU favorece estabilidade numérica e eficiência computacional em redes multicamadas.⁷
- Diferentemente das funções sigmoide e tangente hiperbólica, a ReLU preserva gradientes úteis em regiões amplas do espaço de entrada.⁷
- Sem funções de ativação não lineares, uma rede neural profunda se reduz a um modelo linear equivalente.⁷

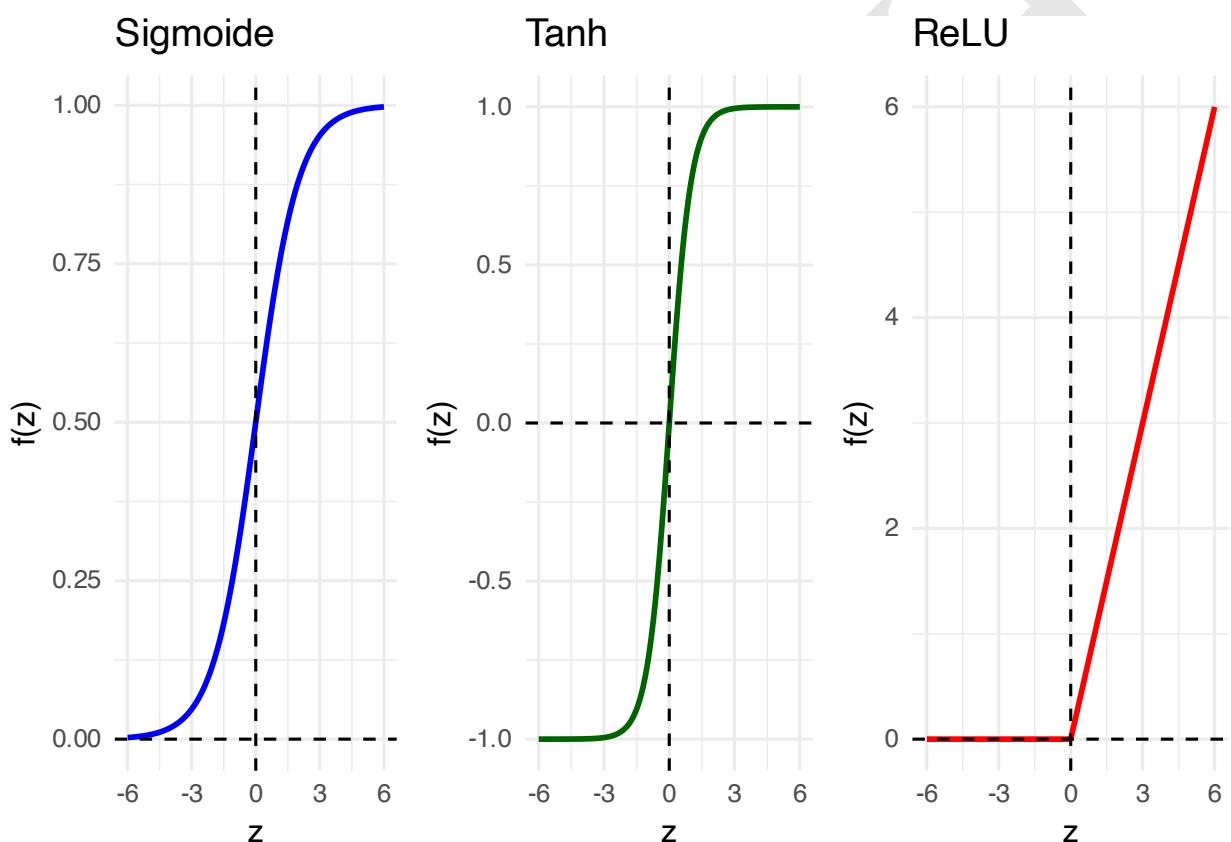


Figura 42.3: Gráficos das funções de ativação mais comuns.

42.4 Espaço de decisão

42.4.1 O que é espaço de decisão?

- O espaço de decisão é a região do espaço de entrada onde o modelo classifica as entradas em diferentes categorias. Ele é definido pelas fronteiras de decisão aprendidas pelo modelo durante o treinamento.⁷

42.4.2 Como ele é visualizado?

- O espaço de decisão pode ser visualizado graficamente, especialmente em problemas de classificação binária ou multiclasse, onde as regiões correspondem às classes previstas pelo modelo.⁸

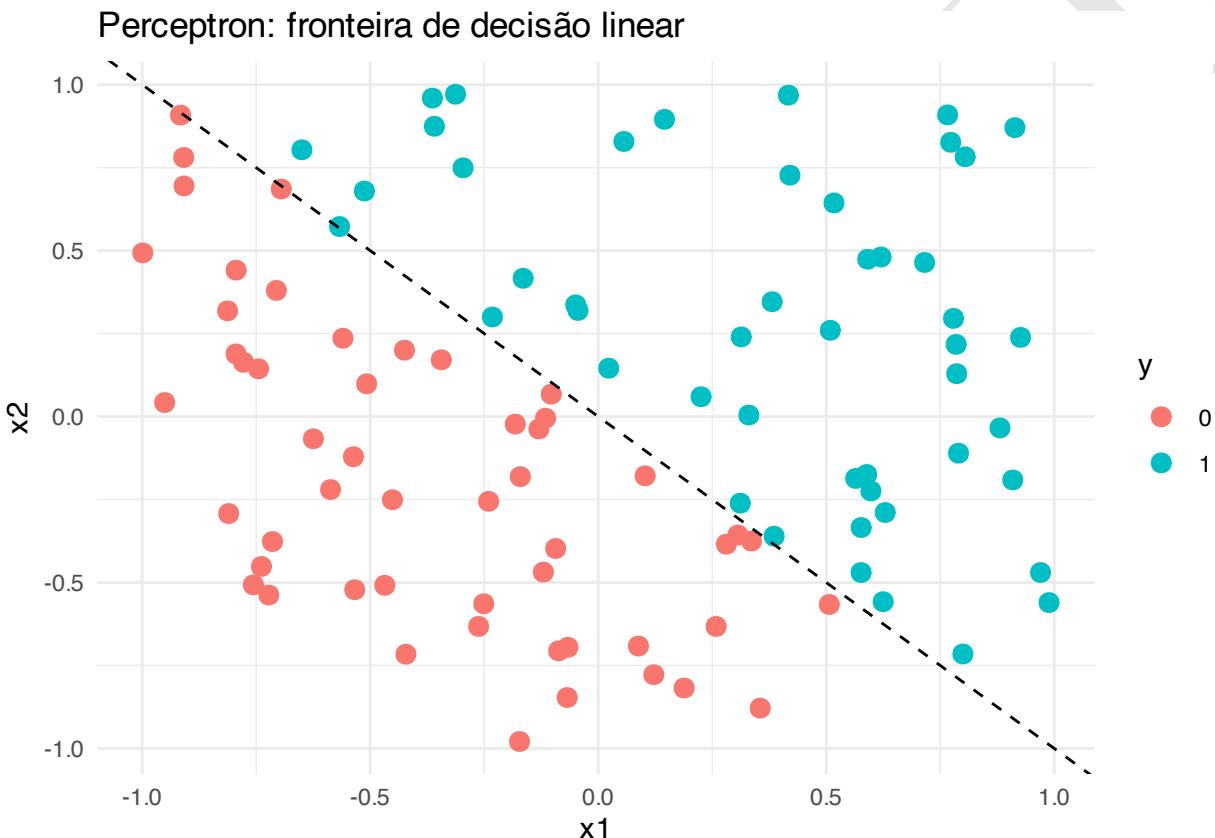


Figura 42.4: Espaço de decisão de um perceptron (regressão logística).

42.5 Redes neurais multicamadas

42.5.1 O que são redes neurais multicamadas?

- Redes neurais multicamadas são redes neurais artificiais que contêm uma ou mais camadas escondidas entre a camada de entrada e a camada de saída, permitindo a composição sucessiva de transformações não lineares e, consequentemente, maior capacidade de representação de funções complexas.⁹
- Essa estrutura amplia o espaço de decisão do modelo sem exigir especificação explícita de interações ou transformações entre variáveis.¹⁰
- À medida que a profundidade da rede aumenta, passa-se do ajuste de parâmetros para o aprendizado de re-

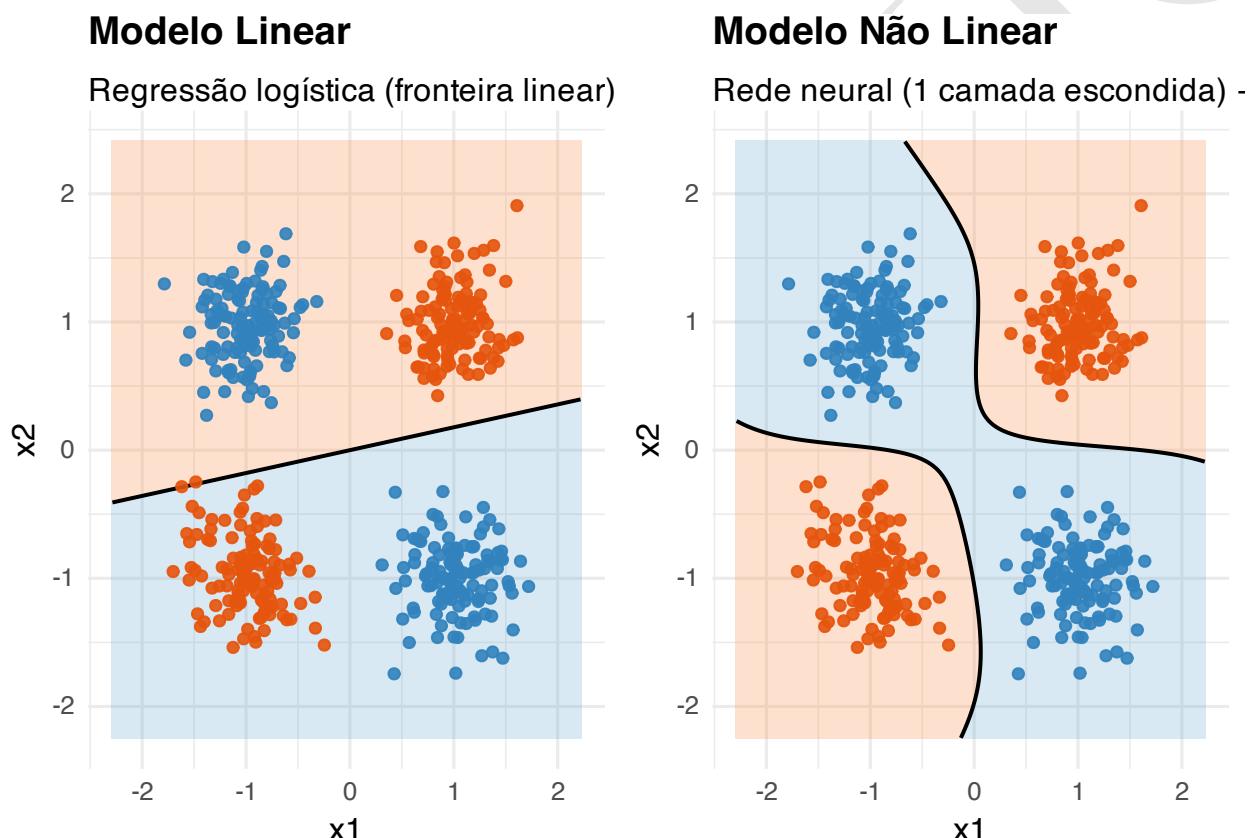


Figura 42.5: Comparação do espaço de decisão entre um modelo linear (regressão logística) e um modelo não linear (MLP).

presentações internas dos dados, caracterizando um regime distinto de modelagem conhecido como aprendizado profundo.?

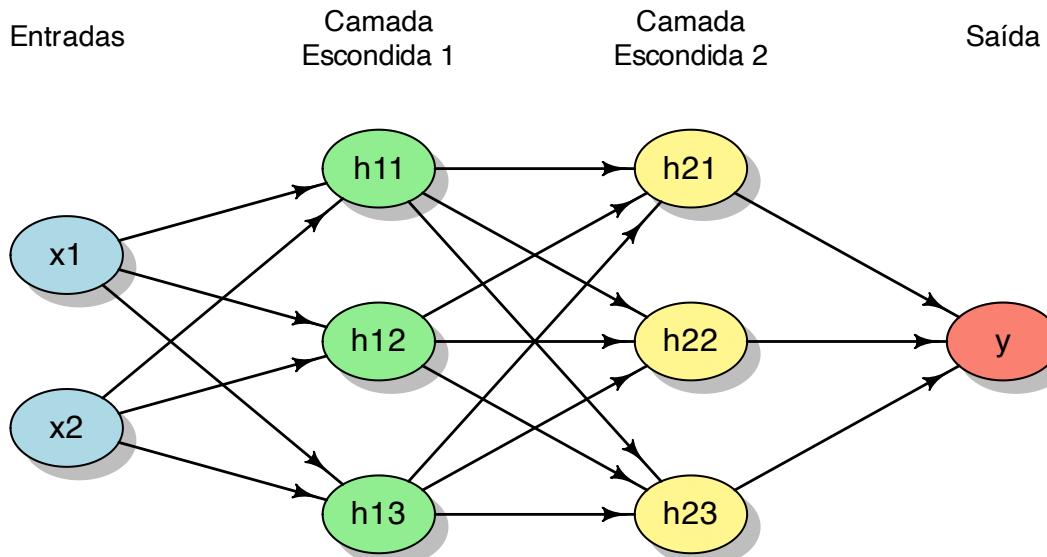


Figura 42.6: Representação esquemática de uma rede neural multicamadas com 2 camadas escondidas além das camadas de entrada e saída.

PARTE 8: PLANEJAMENTO DE ESTUDOS

Definindo poder, tamanho amostral e plano de análise

RASCUNHO

Capítulo 43

Poder estatístico

43.1 Poder do teste

43.1.1 O que é poder do teste?

- Poder do teste é a probabilidade de rejeitar corretamente a hipótese nula (H_0) quando esta é falsa.²⁶¹
- Poder do teste pode ser calculado como $1 - \beta$.²⁶¹

43.1.2 O que é análise de poder do teste?

- Poder é a probabilidade de que um dado tamanho de efeito será observado em um experimento futuro sob um conjunto de hipóteses — tamanho de efeito real e erro tipo I — para um dado tamanho de amostra.³⁴⁰
- O objetivo geral da análise de poder ao projetar um estudo é escolher um tamanho de amostra que controle os 2 tipos de erros de inferência estatística: tipo I (α , resultado falso-positivo) e tipo II (β , resultado falso-negativo).³⁴⁰
- Numericamente, o poder de um estudo é calculado como $1 - \beta$ e reportado em valor percentual.³⁴⁰

43.1.3 Quando realizar a análise de poder do teste?

- Na fase de projeto de pesquisa: a análise de poder para determinar o tamanho da amostra objetiva que o tamanho da amostra permita uma probabilidade razoável de detectar um efeito significativo pré-especificado.³⁴⁰
- Após a coleta de dados: a análise de poder objetiva informar estudos futuros a respeito do tamanho da amostra necessário para a detecção de um efeito significativo pré-especificado.³⁴⁰



O pacote *pwr*²⁸⁰ fornece a função *pwr.2p.test*^a para cálculo do poder do teste de proporção balanceado (2 amostras com mesmo número de participantes).

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.2p.test>



O pacote *pwr*²⁸⁰ fornece a função *pwr.2p2n.test*^a para cálculo do poder do teste de proporção não balanceado (2 amostras com diferente número de participantes).

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.2p.test>

R

O pacote *pwr*²⁸⁰ fornece a função *pwr.anova.test*^a para cálculo do poder do teste de análise de variância balanceado (3 ou mais amostras com mesmo número de participantes).

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.anova.test>

R

O pacote *pwr*²⁸⁰ fornece a função *pwr.chisq.test*^a para cálculo do poder do teste de qui-quadrado χ^2 .

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.chisq.test>

R

O pacote *pwr*²⁸⁰ fornece a função *pwr.f2.test*^a para cálculo do poder do teste com modelo linear geral.

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.f2.test>

R

O pacote *pwr*²⁸⁰ fornece a função *pwr.norm.test*^a para cálculo do poder do teste de média de uma distribuição normal com variância conhecida.

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.norm.test>

R

O pacote *pwr*²⁸⁰ fornece a função *pwr.p.test*^a para cálculo do poder do teste de proporção (1 amostra).

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.p.test>

R

O pacote *pwr*²⁸⁰ fornece a função *pwr.r.test*^a para cálculo do poder do teste de correlação (1 amostra).

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.r.test>

R

O pacote *pwr*²⁸⁰ fornece a função *pwr.t.test*^a para cálculo do poder do teste *t* de diferença de 1 amostra, 2 amostras dependentes ou 2 amostras independentes (grupos balanceados).

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.t.test>

R

O pacote *pwr*²⁸⁰ fornece a função *pwr.t2n.test*^a para cálculo do poder do teste *t* de diferença de 2 amostras independentes (grupos não balanceados).

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.t2n.test>



O pacote *longpower*³⁴¹ fornece a função *power:mmrm*^a para calcular o poder de testes com análises por modelo de regressão linear misto.

^a<https://www.rdocumentation.org/packages/longpower/versions/1.0.24/topics/power.mmrm>



O pacote *Superpower*²⁸⁵ fornece a função *power_ftest*^a para calcular o poder do teste por análise de testes F.

^ahttps://www.rdocumentation.org/packages/Superpower/versions/0.2.0/topics/power_ftest



O pacote *Superpower*²⁸⁵ fornece a função *power_oneway_between*^a para calcular o poder do teste por análise de variância (ANOVA) de 1 fator entre-sujeitos.

^ahttps://www.rdocumentation.org/packages/Superpower/versions/0.2.0/topics/power_oneway_between



O pacote *Superpower*²⁸⁵ fornece a função *power_oneway_within*^a para calcular o poder do teste por análise de variância (ANOVA) de 1 fator intra-sujeitos.

^ahttps://www.rdocumentation.org/packages/Superpower/versions/0.2.0/topics/power_oneway_within



O pacote *Superpower*²⁸⁵ fornece a função *power_oneway_ancova*^a para calcular o poder do teste por análise de covariância (ANCOVA).

^ahttps://www.rdocumentation.org/packages/Superpower/versions/0.2.0/topics/power_oneway_ancova



O pacote *Superpower*²⁸⁵ fornece a função *power_twoway_between*^a para calcular o poder do teste por análise de covariância (ANOVA) de 2 fatores entre-sujeitos.

^ahttps://www.rdocumentation.org/packages/Superpower/versions/0.2.0/topics/power_twoway_between



O pacote *Superpower*²⁸⁵ fornece a função *power_threeway_between*^a para calcular o poder do teste por análise de covariância (ANOVA) de 3 fatores entre-sujeitos.

^ahttps://www.rdocumentation.org/packages/Superpower/versions/0.2.0/topics/power_threeway_between



O pacote *InteractionPowerR*³⁴² fornece a função *power_interaction*^a para calcular o poder do teste por análise de efeito de interações.

^ahttps://www.rdocumentation.org/packages/InteractionPowerR/versions/0.2.1/topics/power_interaction

43.1.4 Por que a análise de poder do teste *post hoc* é inadequada?

- A análise do poder é teoricamente incorreta, uma vez que a probabilidade calculada $1 - \beta$ expressa a probabilidade de um evento futuro, o que não é mais relevante quando o evento de interesse já ocorreu.^{205,340}

43.1.5 O que pode ser realizado ao invés da análise de poder?

- Após a coleta e análise de dados, recomenda-se realizar a análise e interpretação dos resultados a partir do tamanho do efeito e do seu intervalo de confiança no nível de significância α pré-estabelecido.³⁴⁰

Capítulo 44

Tamanho da amostra

44.1 Tamanho da amostra

44.1.1 O que é tamanho da amostra?

- Tamanho da amostra n é a quantidade de participantes (ou unidades de análise) necessárias para conduzir um estudo a fim de testar uma hipótese.³⁴³
- O cálculo do tamanho da amostra depende de quatro pilares interligados — tamanho de efeito esperado, variabilidade, nível de significância (α) e poder ($1 - \beta$) — cuja combinação determina o n necessário para detectar efeitos de interesse com precisão adequada.¹⁵

44.1.2 Por que determinar o tamanho da amostra é importante?

- É virtualmente impossível, devido a limitações de recursos — tempo, acesso, custo, dentre outros — coletar dados da população completa.⁸
- Uma amostra muito pequena para o estudo pode resultar em ajuste exagerado, imprecisão e baixo poder do teste.¹³³

44.1.3 Quais fatores devem ser considerados para determinar o tamanho da amostra?

- Tamanho da população (N): O tamanho da amostra depende parcialmente do tamanho da população de origem. Geralmente assume-se que a população tem tamanho desconhecido ou infinito. Em alguns estudos serão amostradas populações de tamanho finito (inferior a 100.000 indivíduos), geralmente em pesquisas descritivas, em que esse tamanho deve ser incorporado nos cálculos.³⁴³
- Delineamento do estudo.³⁴³
- Quantidade e características (dependente vs. independente) dos grupos de participantes do estudo.³⁴³
- Erros tipo I (α) e tipo II (β).³⁴³
- Tipo de variável a ser observada (contínua, intervalo, ordinal, nominal, dicotômica).³⁴³
- Tamanho de efeito mínimo a ser observado.³⁴³
- Variabilidade da(s) variável(eis) coletada(s).³⁴³
- Lateralidade do teste de hipótese (uni- ou bicaudais).³⁴³
- Perdas de dados durante a coleta e/ou acompanhamento dos participantes do estudo.³⁴³

R

O pacote *pwr*²⁸⁰ fornece a função *plot.power.htest*^a para apresentar graficamente a relação entre o tamanho da amostra e o poder de testes de hipóteses.

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/plot.power.htest>

44.1.4 Quais aspectos éticos estão envolvidos no tamanho da amostra?

- Determinar a priori o tamanho da amostra pode diminuir o risco de realizar testes ou intervenções desnecessários, de desperdício de recursos (tempo e dinheiro) associados e, por outro lado, de coletar dados insuficientes para testar as hipóteses do estudo.³⁴³
- O tratamento ético dos participantes do estudo, portanto, não exige que se considere se o poder do estudo é inferior à meta convencional de 80% ou 90%.³⁴⁴
- Estudos com poder <80% não são necessariamente antiéticos.³⁴⁴
- Metas convencionais de poder (80–90%) são guias pragmáticos e não regras morais rígidas; estudos com poder <80% não são automaticamente antiéticos quando bem justificados.³⁴⁴
- Grandes estudos podem ser desejáveis por outras razões que não as éticas.³⁴⁴

44.2 Saturação em pesquisas qualitativas

44.2.1 O que é saturação de dados em pesquisas qualitativas?

- Saturação é o ponto em que a coleta de dados não produz novas informações, categorias ou temas, indicando que o fenômeno investigado já foi suficientemente explorado.³⁴⁵
- Essa noção surgiu na teoria fundamentada com o termo “saturação teórica”, mas hoje é amplamente usada em diferentes tradições qualitativas, incluindo fenomenologia, etnografia e análise temática.³⁴⁶

44.2.2 Quais tipos de saturação existem?

- Saturação de códigos: ocorre quando não emergem novos códigos relevantes nos dados³⁴⁶
- Saturação de significados: atinge-se quando a profundidade e a variação dos significados de um tema foram plenamente exploradas.³⁴⁶
- Saturação teórica: quando categorias estão suficientemente desenvolvidas e suas relações esclarecidas.³⁴⁵
- Saturação de metatemas: em pesquisas multicêntricas, quando os grandes temas transversais já foram identificados.³⁴⁷

44.2.3 Quantas entrevistas ou grupos focais são necessários para alcançar saturação?

- Estudos empíricos mostram que a saturação de códigos pode ser atingida com 9 a 17 entrevistas em populações homogêneas e objetivos específicos.³⁴⁶
- Para saturação de significados, podem ser necessárias entre 16 e 24 entrevistas.³⁴⁶
- Em grupos focais, a saturação temática pode ocorrer com 4 a 8 grupos homogêneos.³⁴⁶
- Revisões recentes sugerem que a saturação teórica exige 20 a 30 entrevistas ou mais, dependendo da complexidade do estudo.³⁴⁷

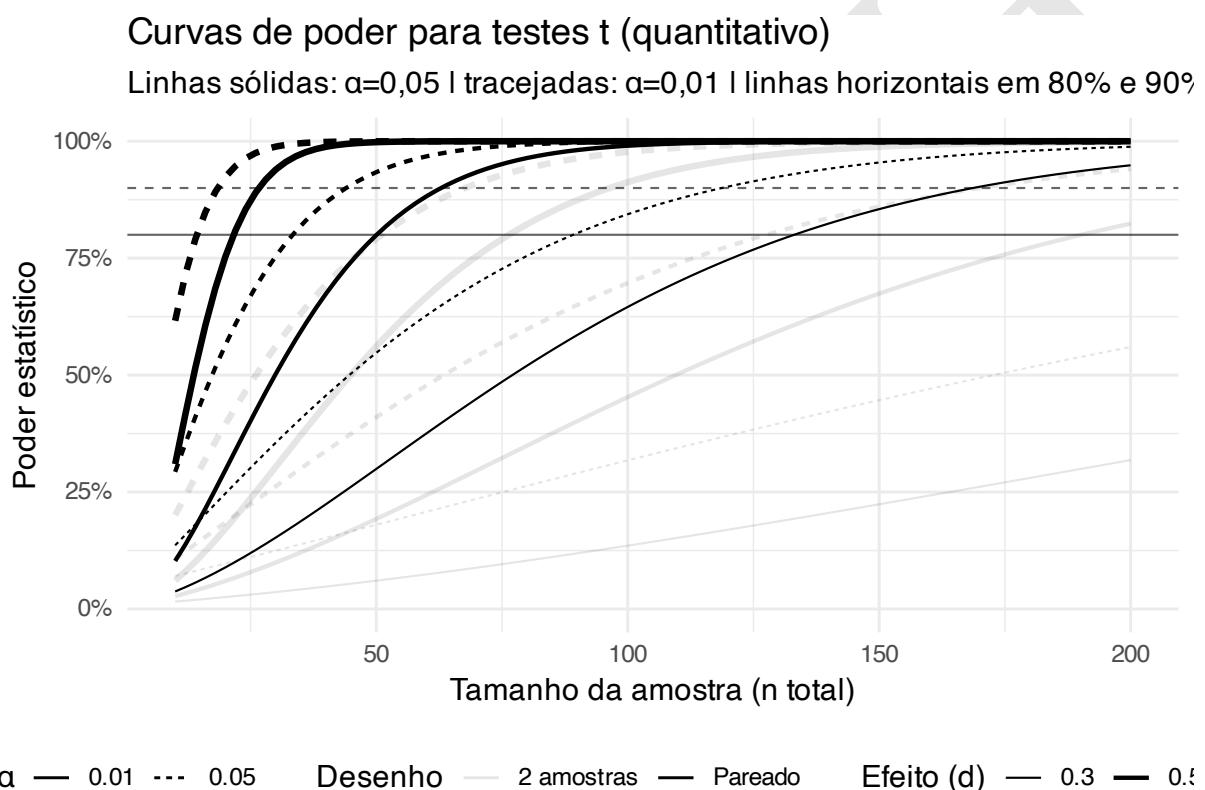


Figura 44.1: Curvas de poder para testes t (quantitativo). Linhas sólidas: $\alpha=0,05$ | tracejadas: $\alpha=0,01$ | linhas horizontais em 80% e 90% de poder.

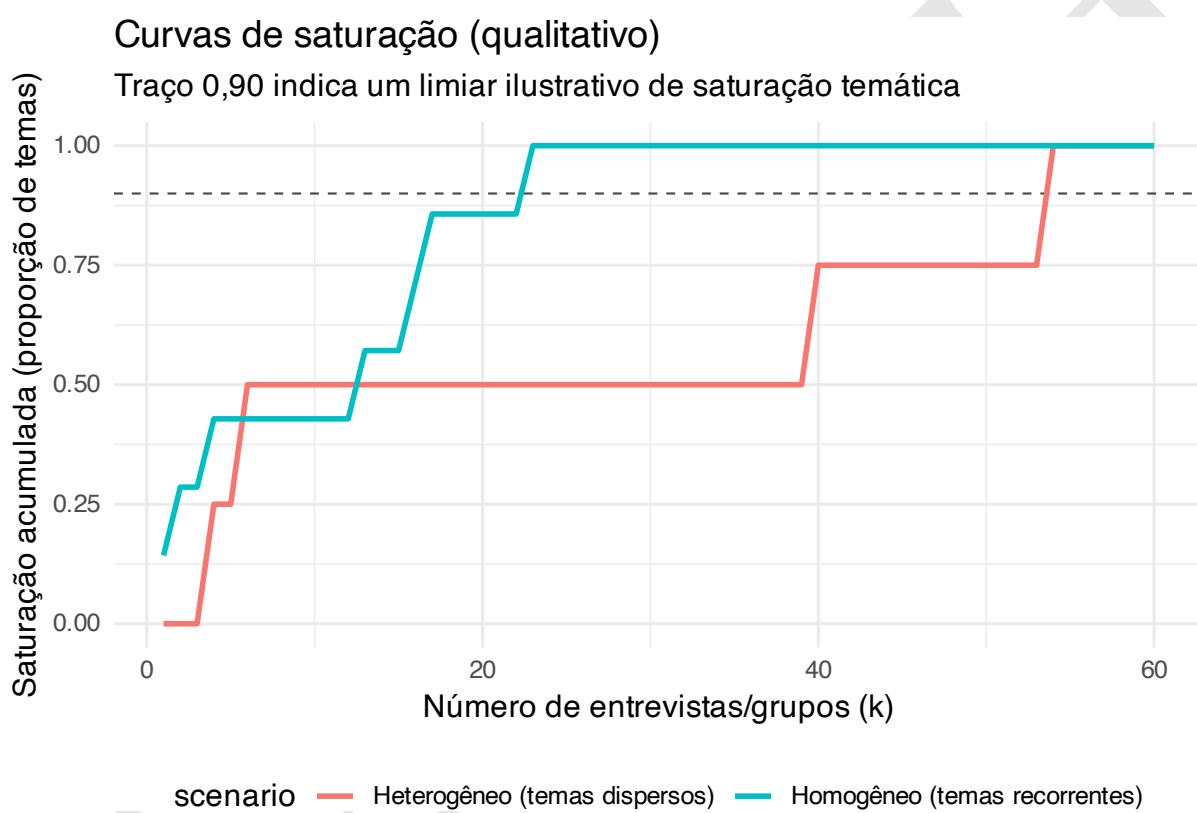


Figura 44.2: Curvas de saturação para estudos qualitativos de descoberta de temas.

44.2.4 Quais debates existem sobre o conceito de saturação?

- Defensores argumentam que a saturação é central para garantir rigor e confiança nos resultados qualitativos.³⁴⁵
- Críticos sugerem que o conceito pode ser usado de forma rígida, levando a coletas excessivas ou pouco sensíveis a perspectivas únicas.³⁴⁵
- Pesquisadores contemporâneos recomendam usar a saturação de forma flexível, adaptada ao contexto, método e população estudada.³⁴⁷

44.2.5 Quais recomendações práticas para tamanho de amostras de estudos qualitativos?

- Para entrevistas individuais: 9–12 entrevistas podem ser suficientes para saturação temática em contextos homogêneos, mas estudos heterogêneos ou multicêntricos exigem mais casos.^{346,347}
- Para grupos focais: 4–8 grupos são geralmente adequados.³⁴⁶
- Para estudos multicêntricos: recomenda-se 20–40 entrevistas por local para alcançar saturação de metatemas.³⁴⁷
- É importante relatar não apenas o número de entrevistas, mas também como e quando a saturação foi avaliada.³⁴⁸

44.3 “Fome de dados”

44.3.1 O que significa “fome de dados”?

- *Data hungry* descreve a necessidade de um modelo contar com muitos eventos por variável (EPV) para alcançar estabilidade estatística.
- Enquanto a regressão logística (LR) atinge desempenho estável com cerca de 20–50 EPV, modelos como random forest (RF), redes neurais (NN) e máquinas de vetor de suporte (SVM) podem demandar >200 EPV para reduzir o otimismo e estabilizar a AUC.

44.3.2 Por que a “fome de dados” é relevante?

- Em bases de dados pequenas, modelos clássicos tendem a ser mais robustos e menos suscetíveis a superajuste.³¹⁷
- O uso de modelos modernos só se justifica quando há grandes bases de dados, caso contrário o ganho em acurácia é marginal.³¹⁷
- Esse conceito conecta diretamente a escolha do modelo ao planejamento amostral.³¹⁷

44.4 Eventos por variável (EPV) em modelos preditivos

44.4.1 Quantos eventos por variável (EPV) são necessários?

- Regressão logística: entre 20 e 50 EPV.³¹⁷
- Árvore de decisão para classificação e regressão: cerca de 60 EPV.³¹⁷
- Máquina de vetores de suporte, redes neurais e *random forests*: muitas vezes >200 EPV e ainda instáveis.³¹⁷

44.4.2 O que acontece se não houver eventos suficientes?

- Modelos modernos podem apresentar alto otimismo (desempenho inflado no treino, mas ruim na validação).³¹⁷
- Pequenos bancos de dados favorecem o uso de modelos clássicos.³¹⁷

44.5 Cálculo do tamanho da amostra

44.5.1 Como calcular o tamanho da amostra?

- O tamanho amostral pode ser calculado por meio de fórmulas matemáticas que tendem a assegurar margens de erros tipos I (α) e II (β) para a estimativa dos parâmetros populacionais (tamanho de efeito) a partir dos dados amostrais.³⁴³
- O tamanho da amostra deve ser calculado para cada um dos objetivos primários e/ou secundários, sendo escolhido o maior tamanho de amostra calculado para o estudo.³⁴³
- Geralmente é recomendado ser cético em relação às regras práticas para o tamanho da amostra, tais como a proporção entre o número de variáveis (ou eventos) e de participantes.¹³³

44.5.2 Como especificar o tamanho do efeito esperado?

- Estudo-piloto — realizados nas mesmas condições do estudo, mas envolvendo um tamanho de amostra limitado — pode ser útil na estimativa do tamanho da amostra a partir do tamanho do efeito estimado.³⁴³
- Utilizar os limites dos intervalos de confiança de estudos-piloto de ensaios clínicos como estimativa do tamanho do efeito pode aumentar o poder estatístico da análise se comparado ao uso das estimativas pontuais obtidas no mesmo piloto.³⁴⁹
- Embora os testes de hipótese considerem efeito nulo para a hipótese nula — ex.: diferença de média ($H_0 : \mu_1 - \mu_2 = 0$), correlação ($H_0 : r = 0$), associação ($H_0 : \beta = 0$ ou $H_0 : OR = 1$) —, em geral é improvável que os efeitos populacionais sejam de fato nulos (isto é, exatamente 0).³⁵⁰

R

O pacote *pwr*²⁸⁰ fornece a função *pwr.2p.test*^a para cálculo do tamanho da amostra para testes de proporção balanceados (2 amostras com mesmo número de participantes).

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.2p.test>

R

O pacote *pwr*²⁸⁰ fornece a função *pwr.2p2n.test*^a para cálculo do tamanho da amostra para testes de proporção não balanceados (2 amostras com diferente número de participantes).

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.2p2n.test>

R

O pacote *pwr*²⁸⁰ fornece a função *pwr.anova.test*^a para cálculo do tamanho da amostra para testes de análise de variância balanceados (3 ou mais amostras com mesmo número de participantes).

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.anova.test>



O pacote *pwr*²⁸⁰ fornece a função *pwr.chisq.test*^a para cálculo do tamanho da amostra para testes de qui-quadrado χ^2 .

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.chisq.test>



O pacote *pwr*²⁸⁰ fornece a função *pwr.f2.test*^a para cálculo do tamanho da amostra para testes com modelo linear geral.

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.f2.test>



O pacote *pwr*²⁸⁰ fornece a função *pwr.norm.test*^a para cálculo do tamanho da amostra para a média de uma distribuição normal com variância conhecida.

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.norm.test>



O pacote *pwr*²⁸⁰ fornece a função *pwr.p.test*^a para cálculo do tamanho da amostra para testes de proporção (1 amostra).

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.p.test>



O pacote *pwr*²⁸⁰ fornece a função *pwr.r.test*^a para cálculo do tamanho da amostra para testes de correlação (1 amostra).

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.r.test>



O pacote *pwr*²⁸⁰ fornece a função *pwr.t.test*^a para cálculo do tamanho da amostra para testes *t* de diferença de 1 amostra, 2 amostras dependentes ou 2 amostras independentes (grupos balanceados).

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.t.test>



O pacote *pwr*²⁸⁰ fornece a função *pwr.t2n.test*^a para cálculo do tamanho da amostra para testes *t* de diferença de 2 amostras independentes (grupos não balanceados).

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.t2n.test>



O pacote *longpower*³⁴¹ fornece a função *power.mmrm*^a para calcular o tamanho da amostra para estudos com análises por modelo de regressão linear misto.

^a<https://www.rdocumentation.org/packages/longpower/versions/1.0.24/topics/power.mmrm>

44.6 Perdas de amostra

44.6.1 O que é perda de amostra?

- Perda de amostra(s) — isto é, participante(s) ou unidade(s) de análise — pode ocorrer durante a coleta e/ou acompanhamento dos participantes do estudo.³⁴³
- Perda amostral pode ocorrer por: abandono ou desistência do participante, perda de contato com o participante, perda de informação, ocorrência de eventos adversos, morte do participante, entre outros.³⁴³

44.6.2 Por que a perda de amostra é um problema?

- A perda de amostra reduz o tamanho efetivo de n e, portanto, o poder estatístico do estudo, elevando a probabilidade de erro tipo II (β).^{133,343}
- A atração diferencial também pode introduzir viés de seleção (ou de atrito), quando as características dos participantes que permanecem diferem sistematicamente das daqueles que se perdem ao seguimento.³⁴³

44.6.3 Como evitar perda de amostra?

- A perda de amostra pode ser evitada por meio de um planejamento cuidadoso do estudo, incluindo a definição de critérios de inclusão e exclusão claros e apropriados, bem como a definição de estratégias para minimizar a perda de amostra.³⁴³
- A perda de amostra pode ser compensada pelo aumento do tamanho da amostra, desde que o aumento seja suficiente para manter o poder do estudo.³⁴³

44.7 Ajustes no tamanho da amostra

44.7.1 Por que ajustar o tamanho da amostra?

- O tamanho da amostra pode ser ajustado durante o estudo para compensar a perda de amostra, desde que o aumento seja suficiente para manter o poder do estudo.³⁴³

44.7.2 Como ajustar para perda amostral?

- Aumentar o tamanho da amostra estimada n pela porcentagem d de perdas esperada ou prevista, para obter o tamanho da amostra efetiva n' (44.1):³⁴³

$$n' = \frac{n}{1 - d} \quad (44.1)$$

44.8 Justificativa do tamanho da amostra

44.8.1 Como justificar o tamanho da amostra de um estudo?

- Em estudos que envolvem condições raras, pode ser difícil recrutar o número necessário de participantes devido à limitada disponibilidade de casos da população. Mesmo assim, é aconselhável determinar o tamanho da amostra.³⁴³
- Quando um estudo deste tipo não é possível, as considerações referentes ao tamanho da amostra são justificadas de acordo com o número máximo de pacientes que podem ser recrutados no decorrer do estudo.³⁴³

44.8.2 Como justificar o tamanho da amostra em estudos qualitativos?

- Pesquisas qualitativas devem apresentar uma justificativa explícita da amostra, relacionando-a à estratégia de coleta, aos objetivos e ao critério de saturação adotado.³⁴⁸
- A noção de “poder da informação” (*information power*) indica que quanto mais relevante e focada é a amostra em relação à pergunta de pesquisa, menor pode ser o número de participantes.³⁴⁸
- Relatar claramente o processo de decisão aumenta a transparência e a credibilidade da pesquisa.³⁴⁸

RASCUNHO

Capítulo 45

Plano de análise

45.1 Plano de análise estatística

45.1.1 O que é plano de análise estatística?

• ?

45.2 Diretrizes para redação

45.2.1 Quais são as diretrizes para redação de planos de análise estatística?

- Visite a rede *Enhancing the QUAlity and Transparency Of health Research* (EQUATOR Network¹) para encontrar diretrizes específicas.
- *Guidelines for the Content of Statistical Analysis Plans in Clinical Trials:*³⁵¹ <https://www.equator-network.org/reporting-guidelines/guidelines-for-the-content-of-statistical-analysis-plans-in-clinical-trials/>

¹<https://www.equator-network.org/>

RASCUNHO

PARTE 9: DELINEAMENTOS E SÍNTESE DE EVIDÊNCIAS

Tipos de estudo e integração de resultados: observacionais, experimentais e revisões

RASCUNHO

Capítulo 46

Delineamento de estudos

46.1 Critérios de delineamento

46.1.1 Quais critérios são utilizados para classificar os delineamentos de estudos?

- ?

46.2 Alocação

46.2.1 O que é alocação?

- ?

46.3 Cegamento

- ?

46.3.1 O que é cegamento?

46.4 Pareamento

46.4.1 O que é pareamento?

- Pareamento significa que para cada participante de um grupo (por exemplo, com alguma condição clínica), existe um (ou mais) participantes (por exemplo, grupo controle) que possui características iguais ou similares relativas a algumas variáveis de interesse.³⁵²
- As variáveis escolhidas para pareamento devem ter relação com as variáveis de desfecho, mas não são de interesse elas mesmas.³⁵²
- O ajuste por pareamento deve ser incluído nas análises estatísticas mesmo que as variáveis de pareamento não sejam consideradas prognósticas ou confundidores na amostra estudada.³⁵²
- A ausência de evidência estatística de diferença entre grupos não é considerada pareamento.³⁵²

Alocação Aleatória (1:1)

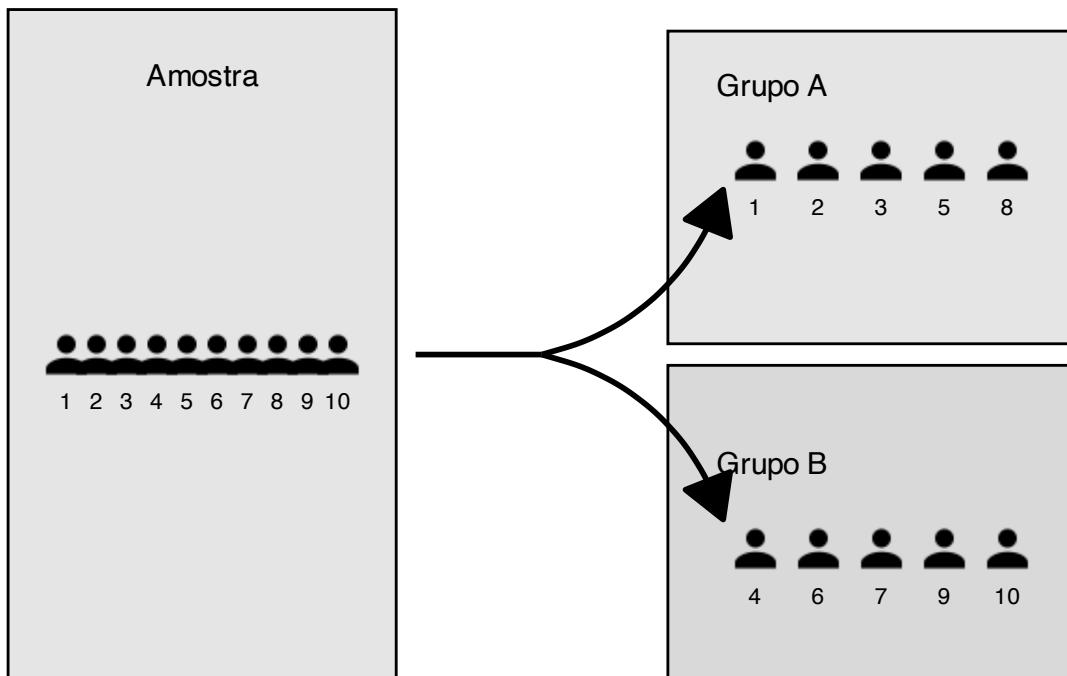


Figura 46.1: Alocação 1:1 entre dois grupos de participantes

46.5 Aleatorização

46.5.1 O que é aleatorização?

- ?

46.6 Taxonomia de estudos

46.6.1 Como podem ser classificados os estudos científicos?

- Estudos científicos podem ser classificados em *básicos, observacionais, experimentais, acurácia diagnóstica, propriedades psicométricas, avaliação econômica e revisões de literatura*.³⁵³⁻³⁶²
- *Estudos básicos*^{354,359}
 - Genética
 - Celular
 - Experimentos com animais
 - Desenvolvimento de métodos
- *Estudos de simulação computacional*^{360,362}
- *Estudos de propriedades psicométricas*^{355,357}
 - Validade
 - Concordância

- Confiabilidade
- *Estudos de desempenho diagnóstico*^{358,361}
 - Transversal
 - Caso-Controle
 - Comparativo
 - Totalmente pareado
 - Parcialmente pareado com subgrupo aleatório
 - Parcialmente pareado com subgrupo não aleatório
 - Não pareado aleatório
 - Não pareado não aleatório
- *Estudos observacionais*^{354,359}
 - Descritivo
 - * Estudo de caso
 - * Série de casos
 - * Transversal
 - Analítico
 - * Transversal
 - * Caso-Controle
 - Caso-Controle aninhado
 - Caso-Coorte
 - Coorte prospectiva ou retrospectiva
- *Estudos quase-experimentais*³⁵⁶
 - Quase-aleatorizado controlado
 - Estimação de variável instrumental
 - Descontinuidade de regressão
 - Série temporal interrompida controlada
 - Série temporal interrompida
 - Diferença
- *Estudos experimentais*^{354,359}
 - Fases I a IV
 - * Aleatorizado controlado
 - * Não-aleatorizado controlado
 - * Autocontrolado
 - * Cruzado
 - * Fatorial

- Campo
- Comunitário
- *Estudos de avaliação econômica*³⁵⁴
 - Análise de custo
 - Análise de minimização de custo
 - Análise de custo-utilidade
 - Análise de custo-efetividade
 - Análise de custo-benefício
- *Estudos de revisão*³⁵³
 - Estado-da-arte
 - Narrativa
 - Crítica
 - Mapeamento
 - Escopo
 - Busca e revisão sistemática
 - Sistematizada
 - Sistemática
 - * Meta-análise
 - * Bibliométrica.^{363,364}
 - Sistemática qualitativa
 - Mista
 - Visão geral
 - Rápida
 - Guarda-chuva

Capítulo 47

Simulação computacional

47.1 Simulações computacionais

47.1.1 O que são simulações computacionais?

- Simulações computacionais consistem na geração de dados artificiais baseados em regras matemáticas e estatísticas, permitindo testar hipóteses, validar métodos e explorar cenários complexos sem necessidade de dados reais.²⁹
- A simulação é frequentemente usada em estatística para avaliar o desempenho de testes, estimadores e modelos sob diferentes condições.³⁰

47.1.2 Por que usar simulações?

- Testar o comportamento de métodos estatísticos sob diferentes premissas (ex: normalidade, homocedasticidade, tamanho amostral).³¹
- Avaliar a robustez de algoritmos computacionais.³²
- Reproduzir processos naturais ou sociais para compreensão teórica.³³

47.1.3 Quais são as boas práticas em simulações computacionais?

- Defina claramente o objetivo da simulação e as hipóteses a serem testadas, incluindo quais aspectos do fenômeno ou do método você pretende avaliar.⁴²
- Use uma semente para o gerador de números aleatórios com `set.seed()` para garantir a reproduzibilidade dos resultados.³⁴
- Documente detalhadamente o processo de simulação, incluindo os parâmetros utilizados, a lógica do algoritmo e as suposições feitas.³⁶⁵
- Realize múltiplas simulações (ex.: 1000 ou mais) para obter estimativas estáveis e resultados mais robustos e confiáveis.³⁵
- Analise os resultados de forma crítica, considerando a variabilidade, as limitações do modelo e possíveis vieses do processo de simulação.³⁶
- Use funções vetorizadas para otimizar o desempenho e reduzir o tempo de execução da simulação.³⁷



O pacote *base*⁵⁵ fornece a função *set.seed*^a para especificar uma semente e garantir a reproduzibilidade de computações que envolvem números aleatórios.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/Random>

47.2 Características

47.2.1 Quais são as características de estudos de simulação computacional?

- ?

47.3 Métodos de simulação

47.3.1 Simulações computacionais dependem da distribuição Normal?

- Não. Simulações computacionais não dependem da distribuição Normal. Qualquer distribuição de probabilidade pode ser usada para gerar dados artificiais, desde que represente adequadamente o mecanismo gerador do fenômeno em estudo.³
- A escolha da distribuição depende da natureza da variável (contínua ou discreta), do domínio dos valores possíveis e da estrutura probabilística do processo observado.³
- Distribuições como Binomial, Poisson e Exponencial são frequentemente utilizadas em simulações para representar proporções, contagens de eventos e tempos até a ocorrência de eventos, respectivamente.³
- Simulações baseadas em diferentes distribuições permitem avaliar o desempenho de métodos estatísticos e algoritmos inferenciais sob condições realistas e variadas, incluindo assimetria, discretação e presença de valores extremos.³

47.3.2 Como escolher a distribuição adequada em um estudo de simulação?

- A escolha da distribuição deve ser guiada pelo mecanismo gerador dos dados do fenômeno de interesse, e não por conveniência matemática.³
- Aspectos como o tipo da variável, assimetria, limites naturais e frequência de eventos devem ser considerados na especificação da distribuição.³
- Em estudos de simulação, é comum avaliar múltiplas distribuições para investigar a sensibilidade dos resultados às suposições do modelo.⁴²

47.3.3 Como simular dados de diferentes distribuições?

- Use funções específicas para cada distribuição, disponíveis nas bibliotecas estatísticas da linguagem utilizada.³



O pacote *stats*¹³¹ fornece a função *rnorm*^a para simular dados de uma distribuição normal.

^a<https://www.rdocumentation.org/packages/stats/versions/3.5.2/topics/Normal>



O pacote *stats*¹³¹ fornece a função *rbinom*^a para simular dados de uma distribuição Binomial.

^a<https://www.rdocumentation.org/packages/stats/versions/3.5.2/topics/rbinom>



O pacote *stats*¹³¹ fornece a função *rpois*^a para simular dados de uma distribuição Poisson.

^a<https://www.rdocumentation.org/packages/stats/versions/3.5.2/topics/rpois>



O pacote *stats*¹³¹ fornece a função *rexp*^a para simular dados de uma distribuição Exponencial.

^a<https://www.rdocumentation.org/packages/stats/versions/3.5.2/topics/rexp>

Dados simulados a partir de diferentes distribuições

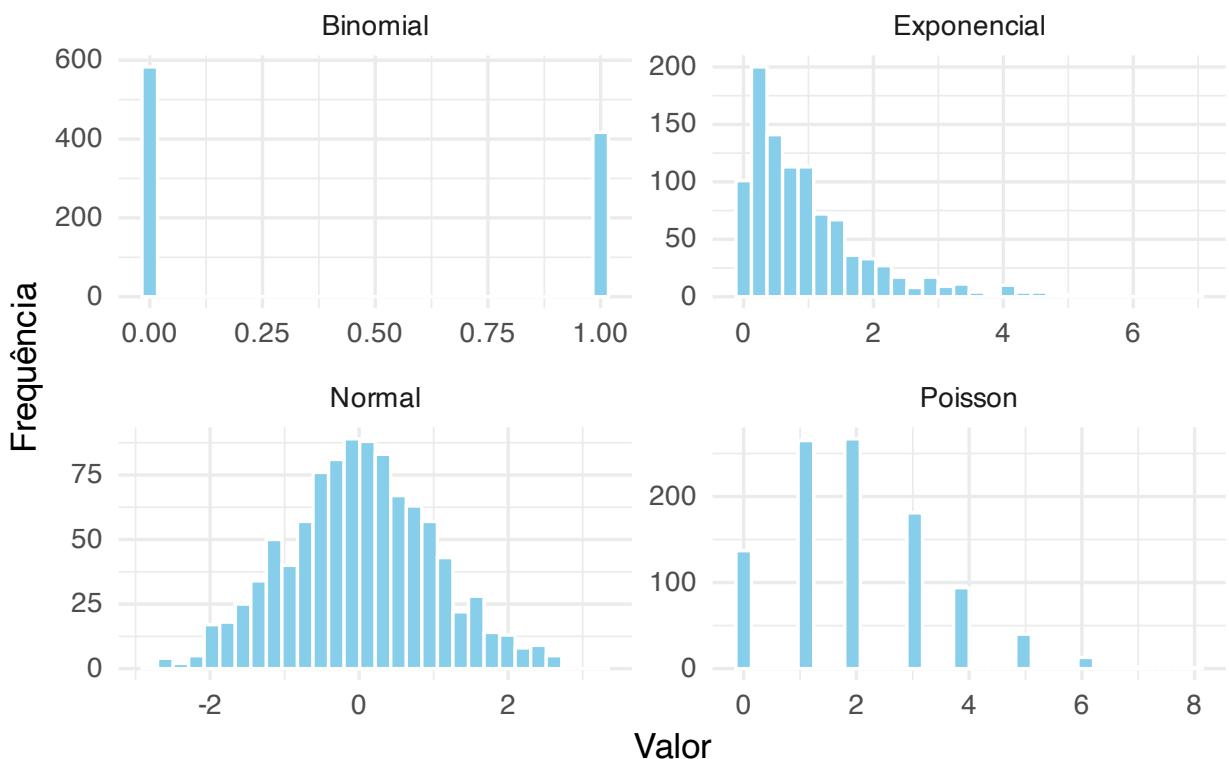


Figura 47.1: Dados simulados a partir de diferentes distribuições: $\text{Normal}(0,1)$, $\text{Binomial}(1,0.4)$, $\text{Poisson}(2)$ e $\text{Exponencial}(1)$.

- Independentemente da distribuição utilizada, o procedimento de simulação segue os mesmos princípios computacionais, diferindo apenas na função geradora dos números aleatórios.[?]

47.3.4 O que é o método de Monte Carlo?

- ³⁶⁶

- No método Markov Chain Monte Carlo (MCMC), o modelo de Markov é usado para gerar amostras de distribuições complexas a partir da simulação de cadeias com distribuição estacionária prescrita.³¹⁸

Convergência do histograma \rightarrow Normal(0,1)

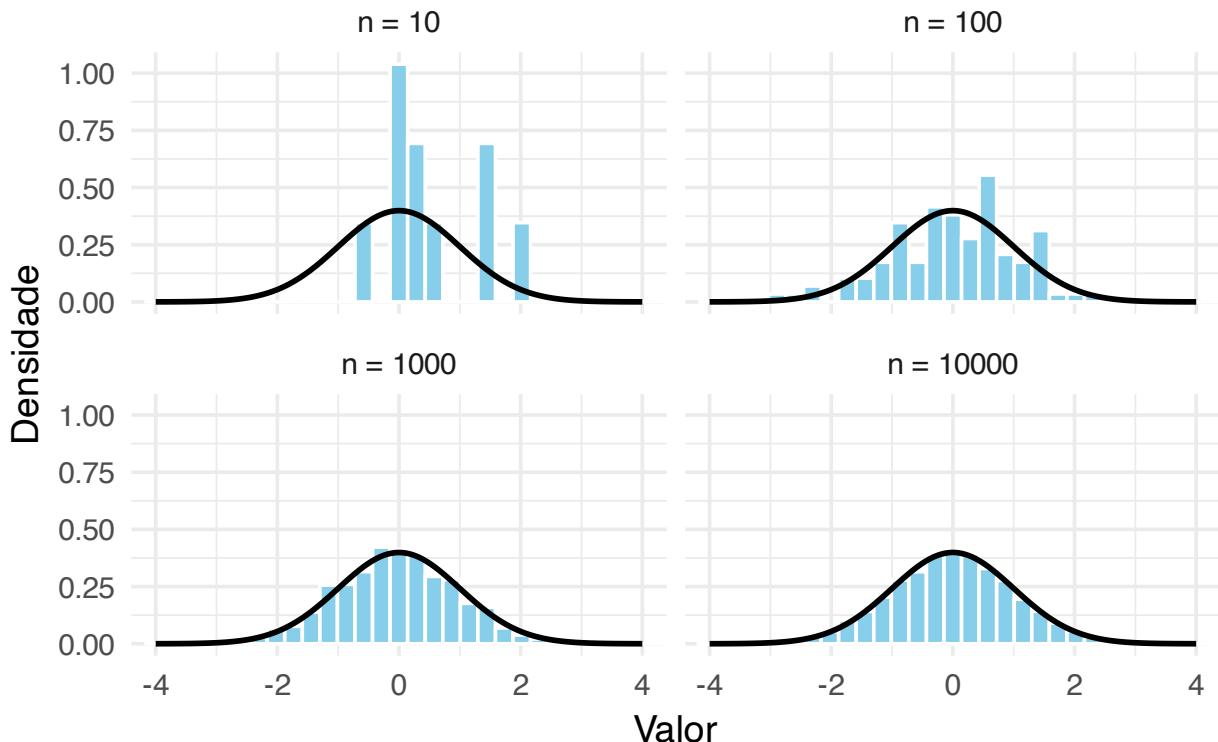


Figura 47.2: Convergência do histograma para a PDF teórica da $Normal(0,1)$ com o aumento do tamanho amostral ($n = 10, 100, 1000, 10000$).

R O pacote *base*⁵⁵ fornece a função *set.seed*^a para especificar uma semente para reproduzibilidade de computações que envolvem números aleatórios.

^a<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/Random>

R O pacote *simstudy*³⁶⁷ fornece as funções *defData*^a e *genData*^b para criar variáveis e simular um banco de dados de acordo com o delineamento pré-especificado, respectivamente.

^a<https://www.rdocumentation.org/packages/simstudy/versions/0.7.0/topics/defData>

^b<https://www.rdocumentation.org/packages/simstudy/versions/0.7.0/topics/genData>

R O pacote *faux*³⁶⁸ fornece a função *sim_design*^a para simular um banco de dados de acordo com o delineamento pré-especificado.

^ahttps://www.rdocumentation.org/packages/faux/versions/1.2.1/topics/sim_design

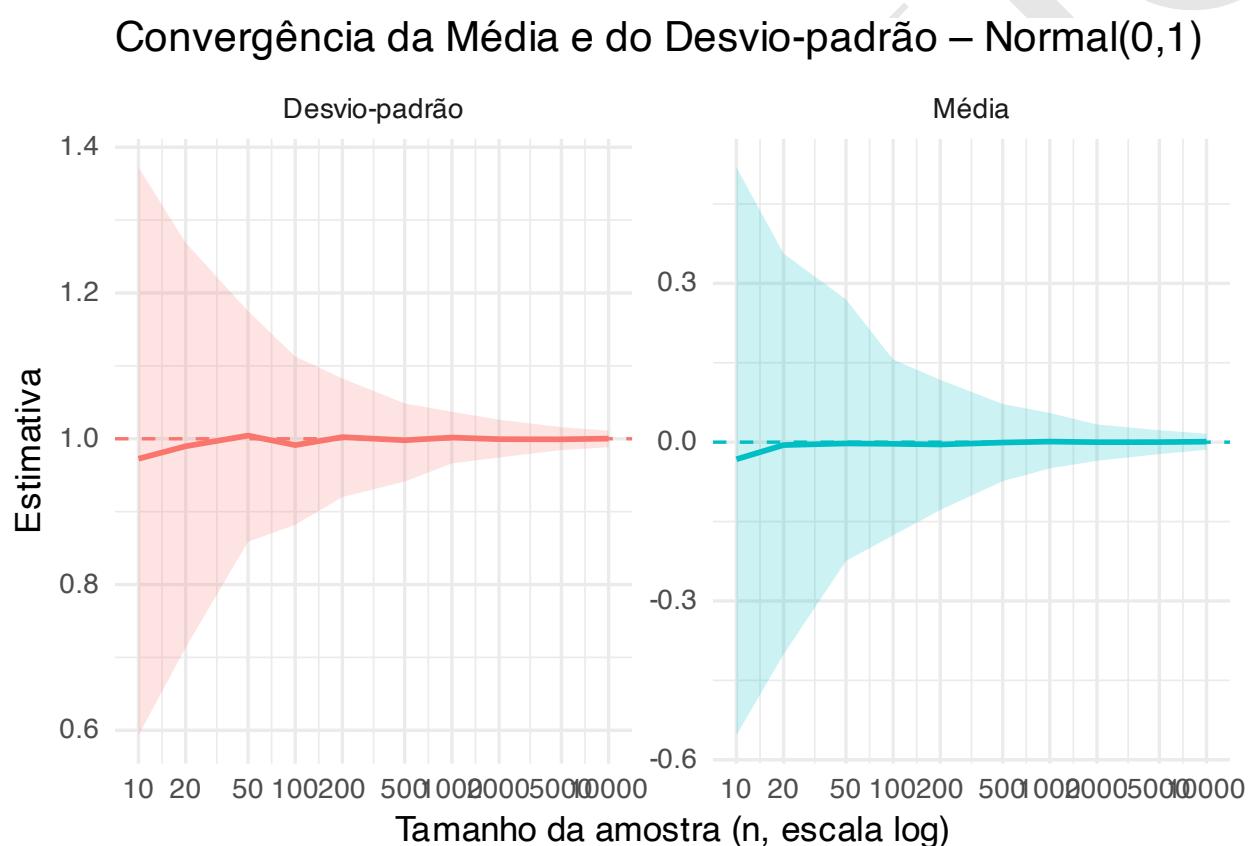


Figura 47.3: Convergência da média e do desvio-padrão amostral para os valores teóricos (0 e 1, respectivamente) com o aumento do tamanho amostral ($n = 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000$).



O pacote *InteractionPowerR*³⁴² fornece a função *generate_interaction*^a para simular bancos de dados com efeitos de interação.

^ahttps://www.rdocumentation.org/packages/InteractionPowerR/versions/0.2.1/topics/generate_interaction

47.4 Diretrizes para redação

47.4.1 Quais são as diretrizes para redação de estudos de simulação computacional?

- Visite a rede *Enhancing the QUAlity and Transparency Of health Research* (EQUATOR Network¹) para encontrar diretrizes específicas.
- *Strengthening the reporting of empirical simulation studies: Introducing the STRESS guidelines:*³⁶⁹ <https://www.equator-network.org/reporting-guidelines/strengthening-the-reporting-of-empirical-simulation-studies-introducing-the-stress-guidelines/>

¹<https://www.equator-network.org/>

Capítulo 48

Estudos observacionais

48.1 Características

48.1.1 Quais são as características de estudos observacionais?

- ?

48.2 Diretrizes para redação

48.2.1 Quais são as diretrizes para redação de estudos observacionais?

- Visite a rede *Enhancing the QUAlity and Transparency Of health Research* (EQUATOR Network¹) para encontrar diretrizes específicas.
- *The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies:*³⁷⁰ <https://www.equator-network.org/reporting-guidelines/strobe/>

¹<https://www.equator-network.org/>

RASCUNHO

Capítulo 49

Propriedades psicométricas

49.1 Características

49.1.1 O que são propriedades psicométricas?

- ?



O pacote *lavaan*³⁷¹ fornece a função *modificationIndices*^a para calcular os índices de modificação.

^a<https://www.rdocumentation.org/packages/lavaan/versions/0.6-16/topics/modificationIndices>

49.2 Análise factorial exploratória

49.2.1 O que é análise factorial exploratória?

- ?

49.3 Análise factorial confirmatória

49.3.1 O que é análise factorial confirmatória?

- ?

Tabela 49.1: Tabela de confusão sobre propriedades psicométricas de instrumentos.

	Concordância alta	Concordância baixa
Validade alta	Adequado	Inadequado
Validade baixa	Inadequado	Inadequado



O pacote *lavaan*³⁷¹ fornece a função *cfa*^a para implementar modelos de análise fatorial confirmatória.

^a<https://www.rdocumentation.org/packages/lavaan/versions/0.6-16/topics/cfa>

49.4 Validade de conteúdo

49.4.1 O que é validade interna?

- ³⁷²

49.4.2 O que é validade externa?

- ³⁷²

49.4.3 Que fatores afetam a validade?

- A amostragem não probabilística pode dificultar a generalização dos achados da amostra para a população, diminuindo assim a validade externa do estudo.¹⁵
- Quando as características da amostra obtida por seleção não probabilística forem similares às da população, a validade externa pode ser maior.¹⁵

49.4.4 Como avaliar a validade de um estudo?

- As características da amostra apresentadas na Tabela 1 são úteis para interpretação da validade interna e externa dos achados do estudo.²¹⁶

49.5 Validade de face

49.5.1 O que é validade de face?

- .[RF]

49.6 Validade do construto

49.6.1 O que é construto?

- .[RF]

49.7 Validade fatorial

49.7.1 O que é validade fatorial?

- .[RF]

49.8 Validade convergente

49.8.1 O que é validade convergente?

- .[RF]

Tabela 49.2: Tabela de confusão 2x2 para análise de concordância de testes e variáveis dicotômicas.

	Teste positivo	Teste negativo	Total
Teste positivo	a	b	$g = a + b$
Teste negativo	c	d	$h = c + d$
Total	$e = a + c$	$f = b + d$	$N = a + b + c + d$

49.9 Validação discriminante

49.9.1 O que é validade discriminante?

- .[RF]

49.10 Validação de critério

49.10.1 O que é validade de critério?

- .[RF]

49.11 Validação concorrente

49.11.1 O que é concorrente?

- .[RF]

49.11.2 O que é validade concorrente?

- .[RF]

49.11.3 O que é validade preditiva?

- .[RF]

49.12 Responsividade

49.12.1 O que é responsividade?

- ?

49.13 Concordância

49.13.1 O que é concordância?

- ?

49.13.2 Quais métodos são adequados para análise de concordância de variáveis dicotômicas?

- Coeficiente de Cohen κ : mede a concordância corrigida pelo acaso.^{373,374}

R

O pacote *irr*³⁷⁵ fornece a função *kappa2*^a para estimar a confiabilidade utilizando coeficiente κ de Cohen para 2 examinadores.

^a<https://www.rdocumentation.org/packages/irr/versions/0.84.1/topics/kappa2>

R

O pacote *irr*³⁷⁵ fornece a função *kappam.fleiss*^a para estimar a confiabilidade utilizando coeficiente κ de Fleiss para mais de 2 examinadores.

^a<https://www.rdocumentation.org/packages/irr/versions/0.84.1/topics/kappam.fleiss>

R

O pacote *irr*³⁷⁵ fornece a função *kappam.light*^a para estimar a confiabilidade utilizando coeficiente κ de Light para mais de 2 examinadores em dados categóricos.

^a<https://www.rdocumentation.org/packages/irr/versions/0.84.1/topics/kappam.light>

- Coeficiente de correlação tetracórica r_{tet} .^{376,377}

R

O pacote *psych*³⁷⁸ fornece a função *tetrachoric*^a para calcular o coeficiente de correlação tetracórica (r_{tet}).

^a<https://www.rdocumentation.org/packages/psych/versions/2.3.6/topics/tetrachoric>

49.13.3 Quais métodos não são adequados para análise de concordância de variáveis dicotômicas?

- Concordância absoluta C_A — quantidade de casos em que examinadores concordam — não é recomendada porque não corrige a estimativa para possíveis concordâncias ao acaso.³⁷⁷
- Concordância percentual $C\%$ — proporção de casos em que examinadores concordam pela quantidade total de casos — não é recomendada porque não corrige a estimativa para possíveis concordâncias ao acaso.³⁷⁷
- Qui-quadrado χ^2 a partir da tabela de contigência não é recomendado porque tal teste analisa associação.³⁷⁷
- A família de coeficientes de Cohen κ não é adequada para analisar concordância quando as variáveis são aparentemente (e não originalmente) dicotômicas.³⁷⁷

R

O pacote *irr*³⁷⁵ fornece a função *agree*^a para estimar a concordância percentual entre examinadores.

^a<https://www.rdocumentation.org/packages/irr/versions/0.84.1/topics/agree>

49.13.4 Quais métodos são adequados para análise de concordância de variáveis categóricas?

- Coeficiente de Cohen κ : mede a concordância corrigida pelo acaso.^{373,374}
- Coeficiente de Cohen ponderado κ_w : mede a concordância corrigida pelo acaso.^{373,374}
- Coeficiente de correlação policórica r_{pol} .³⁷⁷

Tabela 49.3: Tabela de confusão 3x3 para análise de concordância de testes e variáveis dicotômicas.

	Grave	Moderado	Leve	Total
Grave	a	b	c	$j = a + b + c$
Moderado	d	e	f	$k = d + e + f$
Leve	g	h	i	$l = g + h + i$
Total	$j = a + d + g$	$k = b + e + h$	$l = c + f + i$	$N = a + b + c + d + e + f + g + h + i$



O pacote *psych*³⁷⁸ fornece a função *tetrachoric*^a para calcular o coeficiente de correlação polí-córica (r_{pol}).

^a<https://www.rdocumentation.org/packages/psych/versions/2.3.6/topics/tetrachoric>

49.13.5 Quais métodos são adequados para análise de concordância de variáveis categóricas e contínuas?

- Coeficiente de correlação bisserial r_s .³⁷⁷



O pacote *psych*³⁷⁸ fornece a função *tetrachoric*^a para calcular o coeficiente de correlação bisserial (r_s).

^a<https://www.rdocumentation.org/packages/psych/versions/2.3.6/topics/tetrachoric>

49.13.6 Quais métodos são adequados para análise de concordância de variáveis ordinais?

- Coeficiente de Cohen ponderado κ_w : mede a concordância corrigida pelo acaso.^{373,374}

49.13.7 Quais métodos são adequados para análise de concordância de variáveis contínuas?

- Gráfico de dispersão com a reta de regressão.¹⁴³
- Gráfico de limites de concordância (média dos testes vs. diferença entre testes) com a reta de regressão do viés e respectivo no nível de significância α pré-estabelecido.¹⁴³



O pacote *BlandAltmanLeh*³⁷⁹ fornece as funções *bland.altman.stats*^a e *bland.altman.plot*^b para calcular e apresentar, respectivamente, o gráfico com os limites de concordância entre dois métodos.

^a<https://www.rdocumentation.org/packages/BlandAltmanLeh/versions/0.3.1/topics/bland.altman.stats>

^b<https://www.rdocumentation.org/packages/BlandAltmanLeh/versions/0.3.1/topics/bland.altman.plot>

49.13.8 Quais métodos não são adequados para análise de concordância de variáveis contínuas?

- Comparação de médias: dois métodos apresentarem médias similares — isto é, ‘sem diferença estatística’ após um teste inferencial de hipótese nula $H_0 : \mu_1 = \mu_2$ — não informa sobre a concordância deles. Métodos com maior erro de medida tendem a ter menos chance de rejeição da hipótese nula.¹⁴³
- Correlação bivariada: o coeficiente de correlação dependente tanto da variação entre indivíduos (isto é, entre os valores verdadeiros) quanto da variação intraindividual (isto é, erro de medida). Se a variância dos erros de medida de ambos os métodos não for pequena comparadas à variância dos valores verdadeiros, o tamanho do efeito da correlação será pequeno mesmo que os métodos possuam boa concordância.¹⁴³
- Regressão linear: o teste da hipótese nula da inclinação da reta de regressão ($H_0 : \beta = 0$) é equivalente a testar a correlação bivariada ($H_0 : \rho = 0$).¹⁴³

49.13.9 Quais métodos são adequados para modelagem de concordância?

- Modelo log-linear.³⁷⁷

49.14 Confiabilidade

49.14.1 O que é confiabilidade?

- ?

49.14.2 Quais métodos são adequados para análise de confiabilidade?

- ?

R O pacote *semTools*³⁸⁰ fornece a função *reliability*^a para analisar a confiabilidade de um instrumento.

^a<https://www.rdocumentation.org/packages/semTools/versions/0.5-6/topics/reliability-deprecated>

R O pacote *psych*³⁷⁸ fornece a função *icc*^a para estimar a confiabilidade utilizando coeficientes de correlação intraclasse.

^a<https://www.rdocumentation.org/packages/psych/versions/2.3.6/topics/ICC>

R O pacote *multilevel*³⁸¹ fornece a função *mult.icc*^a para estimar a confiabilidade utilizando diversos coeficientes de correlação intraclasse.

^a<https://www.rdocumentation.org/packages/multilevel/versions/2.5/topics/mult.icc>

R O pacote *multilevel*³⁸¹ fornece a função *cronbach*^a para estimar a confiabilidade utilizando o α de Cronbach.

^a<https://www.rdocumentation.org/packages/multilevel/versions/2.5/topics/cronbach>



O pacote *irr*³⁷⁵ fornece a função *kripp.alpha*^a para estimar a confiabilidade utilizando coeficiente α de Krippendorff.

^a<https://www.rdocumentation.org/packages/irr/versions/0.84.1/topics/kripp.alpha>



O pacote *irr*³⁷⁵ fornece a função *iota*^a para estimar a confiabilidade utilizando coeficiente ι .

^a<https://www.rdocumentation.org/packages/irr/versions/0.84.1/topics/iota>

49.15 Diretrizes para redação

49.15.1 Quais são as diretrizes para redação de estudos de propriedades psicométricas?

- Visite a rede *Enhancing the QUAlity and Transparency Of health Research* (EQUATOR Network¹) para encontrar diretrizes específicas.
- COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures*:³⁸² <https://www.equator-network.org/reporting-guidelines/cosmin-reporting-guideline-for-studies-on-measurement-properties-of-patient-reported-outcome-measures/>
- Recommendations for reporting the results of studies of instrument and scale development and testing*:³⁸³ <https://www.equator-network.org/reporting-guidelines/recommendations-for-reporting-the-results-of-studies-of-instrument-and-scale-development-and-testing/>
- Guidelines for reporting reliability and agreement studies (GRRAS) were proposed*:³⁸⁴ <https://www.equator-network.org/reporting-guidelines/guidelines-for-reporting-reliability-and-agreement-studies-grras-were-proposed/>

¹<https://www.equator-network.org/>

RASCUNHO

Capítulo 50

Desempenho diagnóstico

50.1 Características

50.1.1 Quais são as características de estudos de desempenho diagnóstico?

- ?

50.2 Tabelas 2x2

50.2.1 O que é uma tabela de confusão 2x2?

- Tabela de confusão é uma matriz de 2 linhas por 2 colunas que permite analisar o desempenho de classificação de uma variável dicotômica (padrão-ouro ou referência) versus outra variável dicotômica (novo teste).³⁸⁵

50.2.2 Como analisar o desempenho diagnóstico em tabelas 2x2?

- Verdadeiro-positivo (*VP*): caso com a condição presente e corretamente identificado como tal.³⁸⁶
- Falso-negativo (*FN*): caso com a condição presente e erroneamente identificado como ausente.³⁸⁶
- Verdadeiro-negativo (*VN*): controle sem a condição presente e corretamente identificados como tal.³⁸⁶
- Falso-positivo (*FP*): controle sem a condição presente e erroneamente identificado como presente.³⁸⁶
- Tabelas de confusão também podem ser visualizadas em formato de árvores de frequência.³⁸⁵

Tabela 50.1: Tabela de confusão 2x2 para análise de desempenho diagnóstico de testes e variáveis dicotômicas.

	Condição presente	Condição ausente	Total
Teste positivo	VP	FP	$VP + FP$
Teste negativo	FN	VN	$FN + VN$
Total	$VP + FN$	$FP + VN$	$N = VP + VN + FP + FN$

Scenario

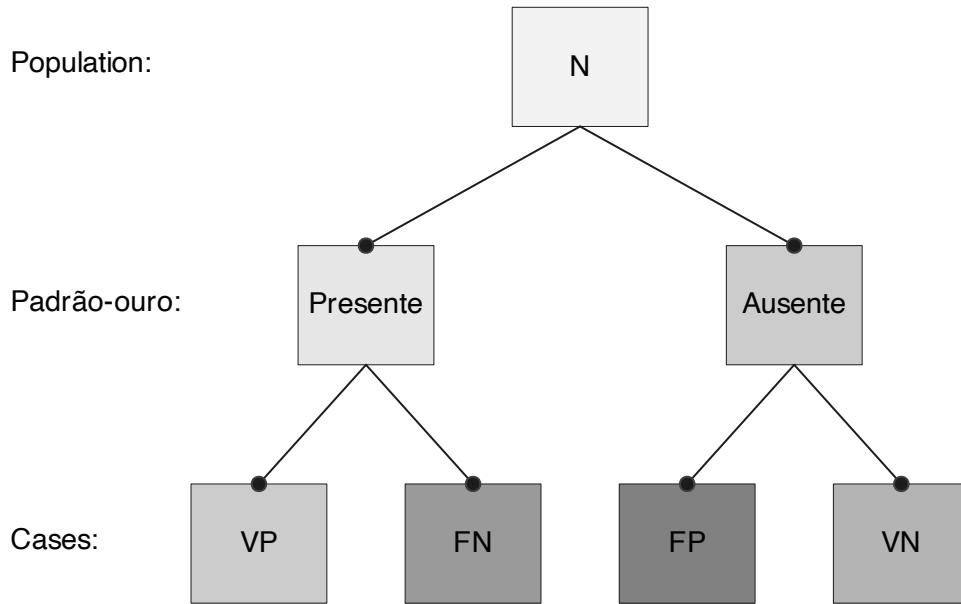


Figura 50.1: Árvore de frequência do desempenho diagnóstico de uma tabela de confusão 2x2 representando um método novo (dicotômico) comparado ao método padrão-ouro ou referência (dicotômico).

R

O pacote *riskyR*³⁸⁷ fornece a função *plot_prism*^a para construir árvores de frequência a partir de diferentes cenários.

^ahttps://www.rdocumentation.org/packages/riskyR/versions/0.4.0/topics/plot_prism

50.2.3 Quais probabilidades caracterizam o desempenho diagnóstico de um teste em tabelas 2x2?

- Sensibilidade (*SEN*) (50.1): Proporção de verdadeiro-positivos dentre aqueles com a condição.³⁸⁶

$$SEN = \frac{VP}{VP + FN} \quad (50.1)$$

- Especificidade (*ESP*) (50.2): Proporção de verdadeiro-negativos dentre aqueles sem a condição.³⁸⁶

$$ESP = \frac{VN}{VN + FP} \quad (50.2)$$

- Valor predictivo positivo (*VPP*) (50.3): Proporção de casos corretamente identificados como verdadeiro-positivos.³⁸⁶

$$VPP = \frac{VP}{VP + FP} \quad (50.3)$$

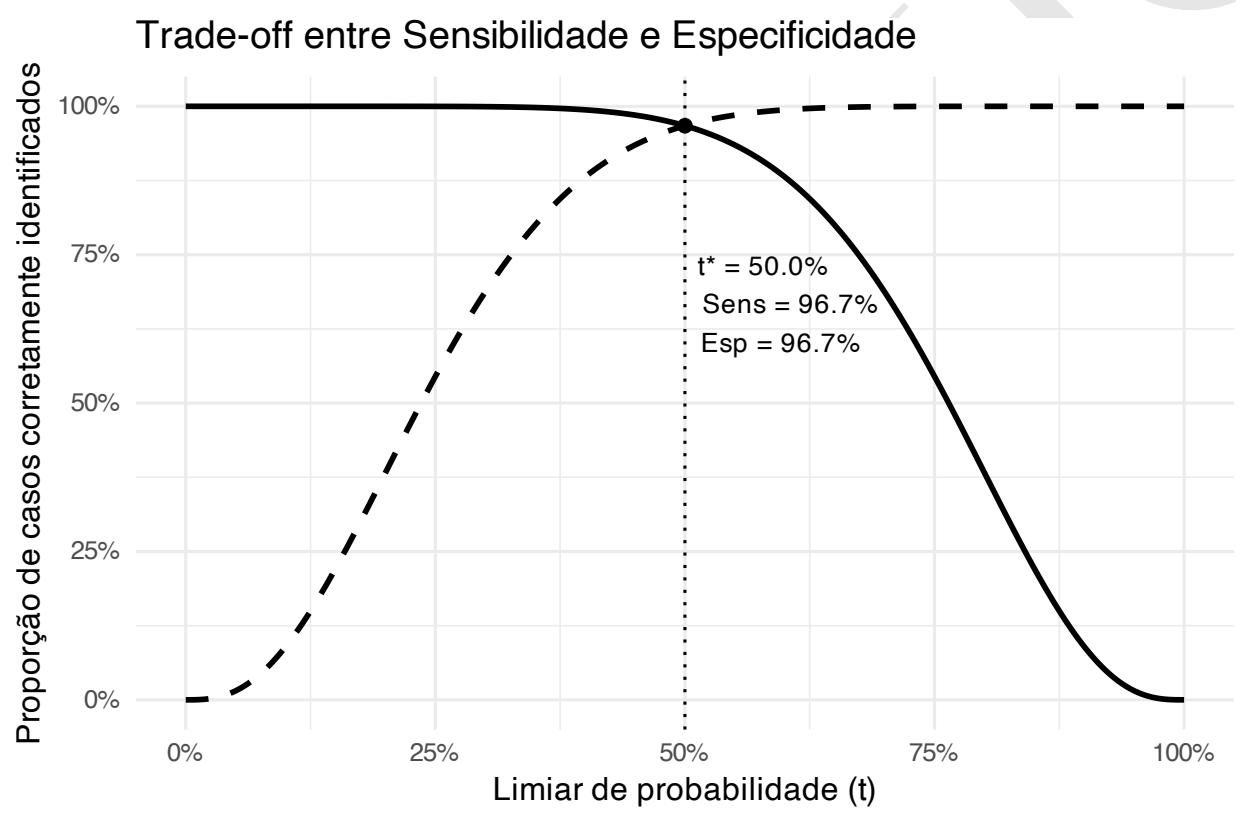


Figura 50.2: Trade-off entre sensibilidade e especificidade em função do limiar de probabilidade (t) para um modelo de classificação.

Tabela 50.2: Probabilidades calculados a partir da tabela de confusão 2x2 para análise de desempenho diagnóstico de testes e variáveis dicotômicas.

	Condição presente	Condição ausente	Total	Probabilidades
Teste positivo	VP	FP	$VP + FP$	$VPP = \frac{VP}{VP+FP}$
Teste negativo	FN	VN	$FN + VN$	$VPN = \frac{VN}{VN+FN}$
Total	$VP + FN$	$FP + VN$	$N = VP + VN + FP + FN$	
Probabilidades	$SEN = \frac{VP}{VP+FN}$	$ESP = \frac{VN}{VN+FP}$		$ACU = \frac{VP+VN}{VP+VN+FP+FN}$ $GOR = \frac{VP \cdot VN}{FP \cdot FN}$

- Valor preditivo negativo (VPN) (50.4): Proporção de controles corretamente identificados como verdadeiro-negativos.³⁸⁶

$$VPN = \frac{VN}{VN + FN} \quad (50.4)$$

- Razão de verossimilhança positiva ($LR+$) (50.5): Quantifica o quanto a probabilidade de a condição estar presente aumenta quando o teste é positivo.³

$$LR+ = \frac{SEN}{1 - ESP} = \frac{VP/(VP + FN)}{FP/(FP + VN)} \quad (50.5)$$

- Razão de verossimilhança negativa ($LR-$) (50.6): Quantifica o quanto a probabilidade de a condição estar presente diminui quando o teste é negativo.³

$$LR- = \frac{1 - SEN}{ESP} = \frac{FN/(VP + FN)}{VN/(FP + VN)} \quad (50.6)$$

- Acurácia (ACU), (50.7): Proporção de casos e controles corretamente identificados.³⁸⁶

$$ACU = \frac{VP + VN}{VP + VN + FP + FN} \quad (50.7)$$

- Razão de chances diagnóstica (GOR) (50.8), (50.9) e (50.10): Razão entre a chance de um teste ser positivo quando a condição está presente e a chance de um teste ser positivo quando a condição está ausente.³⁸⁸

$$GOR = \frac{VP}{FN} \div \frac{FP}{VN} = \frac{VP \cdot VN}{FP \cdot FN} \quad (50.8)$$

$$GOR = \frac{SEN/(1 - SEN)}{(1 - ESP)/ESP} = \frac{SEN \cdot ESP}{(1 - SEN) \cdot (1 - ESP)} \quad (50.9)$$

$$GOR = \frac{LR+}{LR-} \quad (50.10)$$



O pacote *riskyR*³⁸⁷ fornece a função *comp_prob*^a para estimar 13 probabilidades relacionadas ao desempenho diagnóstico em tabelas 2x2.

^ahttps://www.rdocumentation.org/packages/riskyR/versions/0.4.0/topics/comp_prob

Tabela 50.3: Tabela de confusão 3-vias (2×3) com totais: referência vs decisão (3WD).

	POS (aceitar)	BND (deferir)	NEG (rejeitar)	Total
Condição presente (C)	$ POS \cap C $	$ BND \cap C $	$ NEG \cap C $	$ POS \cap C + BND \cap C + NEG \cap C $
Condição ausente (C^c)	$ POS \cap C^c $	$ BND \cap C^c $	$ NEG \cap C^c $	$ POS \cap C^c + BND \cap C^c + NEG \cap C^c $
Total	$ POS \cap C + POS \cap C^c $	$ BND \cap C + BND \cap C^c $	$ NEG \cap C + NEG \cap C^c $	N



O pacote *caret*³⁸⁹ fornece a função *confusionMatrix*^a para estimar 11 probabilidades relacionadas ao desempenho diagnóstico em tabelas 2x2.

^a<https://www.rdocumentation.org/packages/caret/versions/3.45/topics/confusionMatrix>

50.3 Tabelas 2x3

50.3.1 O que é uma tabela de confusão 2x3?

- É a extensão da tabela 2×2 que inclui uma terceira decisão (deferimento/*boundary*) além de aceitar (positivo) e rejeitar (negativo).³⁹⁰
- As colunas** representam as decisões** (*POS*, *BND*, *NEG*) e as linhas representam a verdade de referência (condição presente vs ausente).³⁹⁰
- Essa formulação vem do arcabouço de *Three-Way Decisions (3WD)*, que partitiona o universo em três regiões por dois limiares α e β .³⁹⁰

50.3.2 Como as regiões POS, BND e NEG são definidas?

- Dado um escore ou probabilidade condicional $Pr(C | [x])$ para a classe C , classifica-se como *POS* (aceitar) quando $Pr(C | [x]) \geq \alpha$, como *BND* (deferir) quando $\beta < Pr(C | [x]) < \alpha$ e como *NEG* (rejeitar) quando $Pr(C | [x]) \leq \beta$, sendo que os limiares (α, β) determinam simultaneamente as três regiões e os *trade-offs* entre acurácia e comprometimento.³⁹⁰

50.3.3 Qual é o formato de uma tabela 2x3?

- Estrutura geral (linhas = condição real; colunas = decisão):

50.3.4 Quais são as medidas básicas na 2x3?

- Acurácia em *POS* (M_{PT}), equação (50.11): Proporção de positivos corretamente identificados na região *POS*.³⁹⁰

$$M_{PT} = \frac{|POS \cap C|}{|POS|} \quad (50.11)$$

- Erro em *POS* (M_{PF}), equação (50.12): Proporção de negativos incorretamente classificados na região *POS*.³⁹⁰

$$M_{PF} = \frac{|POS \cap C^c|}{|POS|} \quad (50.12)$$

- Acurácia em NEG (M_{NF}), equação (50.13): Proporção de negativos corretamente identificados na região NEG.³⁹⁰

$$M_{NF} = \frac{|NEG \cap C^c|}{|NEG|} \quad (50.13)$$

- Erro em NEG (M_{NT}), equação (50.14): Proporção de positivos incorretamente classificados na região NEG.³⁹⁰

$$M_{NT} = \frac{|NEG \cap C|}{|NEG|} \quad (50.14)$$

- Frações em BND (M_{BT} e M_{BF}), equações (50.15) e (50.16): Proporção de deferimentos verdadeiros e falsos.³⁹⁰

$$M_{BT} = \frac{|BND \cap C|}{|BND|} \quad (50.15)$$

$$M_{BF} = \frac{|BND \cap C^c|}{|BND|} \quad (50.16)$$

50.3.5 Como escolher os limiares α e β ?

- Os limiares (α, β) controlam o tamanho das regiões POS , NEG e BND e, portanto, os *trade-offs* entre “acertar mais” (acurácia nas regiões) e “decidir mais” (comprometimento; menos deferimentos).³⁹⁰

50.3.6 Quando preferir 3-vias em vez de 2×2?

- Quando o custo de erro é assimétrico e/ou há incerteza relevante.³⁹⁰
- O deferimento (BND) evita decisões precipitadas e permite avaliação adicional, equilibrando acurácia e cobertura.³⁹⁰
- É particularmente útil em triagens diagnósticas com etapas confirmatórias.³⁹⁰

50.4 Curvas ROC

50.4.1 O que representa a curva ROC?

- A relação entre sensibilidade (SEN) no eixo vertical e $1 - ESP$ no eixo horizontal.³⁹¹
- Cada ponto na curva corresponde a um ponto de corte possível do teste.³⁹¹

50.4.2 Quais são os tipos de curva ROC?

- Curva empírica: conecta diretamente os pontos obtidos a partir dos diferentes pontos de corte observados.³⁹²
- Curva suavizada (paramétrica): assume uma distribuição binomial e gera uma curva ajustada por máxima verossimilhança.³⁹²

50.4.3 Como definir o melhor ponto de corte?

- O ponto de corte em uma curva ROC representa um balanço entre sensibilidade e especificidade, ou seja, a taxa de verdadeiros positivos e a taxa de falsos positivos.^{391,392}
- O método de Youden (equação (50.17)) maximiza a diferença entre a taxa de verdadeiros positivos e a taxa de falsos positivos. O ponto de corte ideal será aquele com maior valor de Y .¹²⁶

$$Y = SEN + ESP - 1 \quad (50.17)$$

- O método da distância Euclidiana ((50.18)) minimiza a distância entre um ponto da curva ROC e o ponto (0,1), que representa sensibilidade perfeita ($SEN = 100$) e especificidade perfeita ($ESP = 100$). O ponto de corte ideal será aquele com menor valor de D .³⁹³

$$D = \sqrt{(1 - SEN)^2 + (1 - ESP)^2} \quad (50.18)$$

50.4.4 O que é a área sob a curva (AUROC)?

- A área sob a curva ROC (AUC ou AUROC) quantifica o poder de discriminação ou desempenho diagnóstico na classificação de uma variável dicotômica.³⁹⁴
- A área sob a curva (AUC) resume o desempenho global e representa a probabilidade de o teste classificar corretamente um caso positivo selecionado aleatoriamente em relação a um caso negativo selecionado aleatoriamente.³⁹¹

50.4.5 Como calcular a AUC?

- Método não paramétrico: soma das áreas trapezoidais sob a curva empírica (50.19). Pode subestimar AUC quando os dados são discretos.³⁹²

$$AUC = \sum_{i=1}^{n-1} (x_{i+1} - x_i) \cdot \frac{y_i + y_{i+1}}{2} \quad (50.19)$$

- Método paramétrico (binormal): mais robusto para dados em escala ordinal, com viés reduzido (50.20), onde Φ é a função de distribuição acumulada da Normal padrão, μ_1 e μ_0 são as médias dos escores para os grupos positivo e negativo, respectivamente, e σ_1^2 e σ_0^2 são as variâncias dos escores para os grupos positivo e negativo, respectivamente.³⁹²

$$AUC = \Phi \left(\frac{\mu_1 - \mu_0}{\sqrt{\sigma_1^2 + \sigma_0^2}} \right) \quad (50.20)$$

- AUC deve sempre vir acompanhada de intervalo de confiança (IC95%).³⁹²



O pacote *proc*³⁹⁵ fornece a função *plot.roc*^a para criar uma curva ROC.

^a<https://www.rdocumentation.org/packages/pROC/versions/1.18.4/topics/plot.roc>

50.4.6 Como interpretar a área sob a curva (ROC)?

- A área sob a curva AUC varia no intervalo [0.5; 1], com valores mais elevados indicando melhor discriminação ou desempenho do modelo de classificação.³⁹⁴

- As interpretações qualitativas (isto é: pobre/fraca/baixa, moderada/razoável/aceitável, boa ou muito boa/alta/excelente) dos valores de área sob a curva são arbitrários e não devem ser considerados isoladamente.³⁹⁴
- Modelos de classificação com valores altos de área sob a curva podem ser enganosos se os valores preditos por esses modelos não estiverem adequadamente calibrados.³⁹⁴
- Diferenças pequenas entre AUCs podem não ser estatisticamente significativas.³⁹¹
- A interpretação clínica pode ser equivocada se não houver teste estatístico adequado.³⁹¹
- Se as curvas vêm do mesmo conjunto de pacientes, aplique o teste de DeLong.³⁹¹
- Se as curvas vêm de amostras independentes, use métodos como Dorfman-Alf.³⁹¹

50.4.7 Por que uma AUC menor que 0.5 está errada?

- Porque indica desempenho pior que o acaso.³⁹¹
- Geralmente decorre de seleção incorreta da direção do teste ou da variável de estado.³⁹¹
- Verifique se o software está configurado para maiores valores indicam presença do evento ou o inverso.³⁹¹
- Ajuste a direção do teste para que $AUC \geq 0.5$.³⁹¹

Curva Característica de Operação do Receptor (Curva ROC)

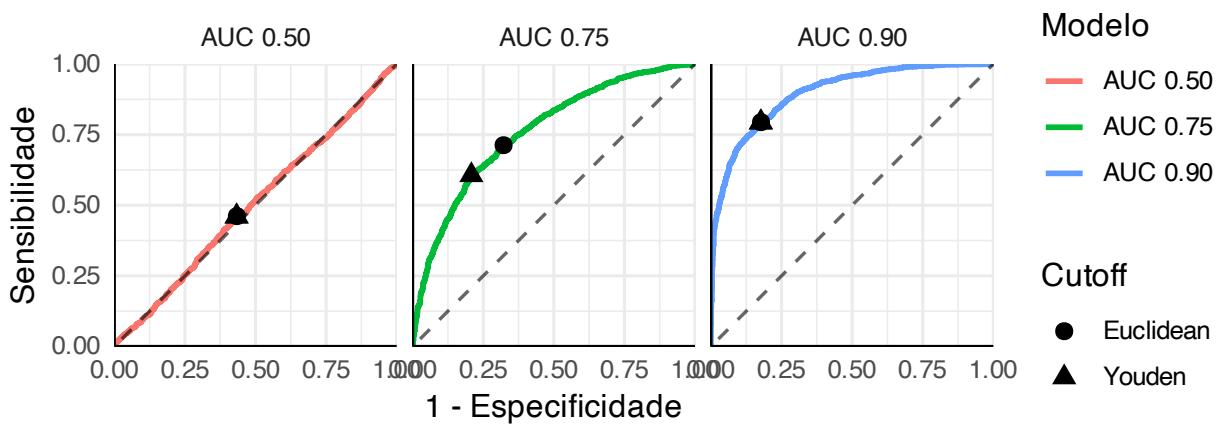


Figura 50.3: Curva ROC (Receiver Operating Characteristic) para um modelos de classificação com diferentes desempenhos diagnósticos.

50.4.8 Como analisar o desempenho diagnóstico em desfechos com distribuição trimodal na população?

- Limiares duplos podem ser utilizados para análise de desempenho diagnóstico de testes com distribuição trimodal.³⁹⁶

50.5 Gráficos crosshair

50.5.1 O que um gráfico crosshair?

- ³⁹⁷

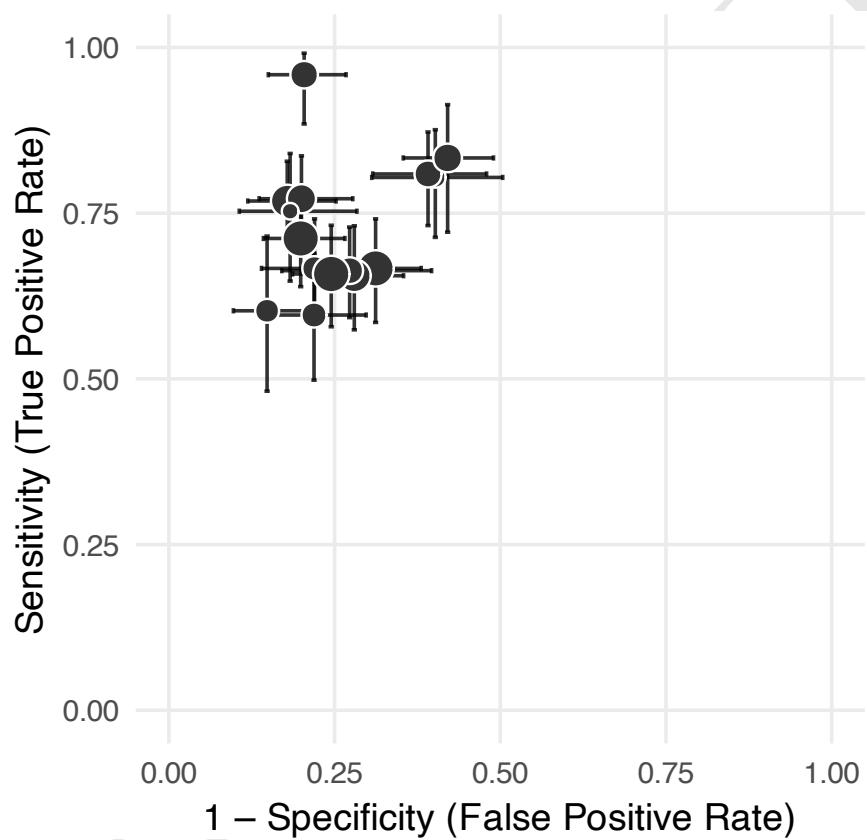


Figura 50.4: Gráfico *crosshair* em espaço ROC (Receiver Operating Characteristic) para 15 estudos simulados de desempenho diagnóstico.

O pacote *mada*³⁹⁸ fornece a função *crosshair*^a para criar um gráfico *crosshair*³⁹⁷ a partir de dados de verdadeiro-positivo, falso-positivo, verdadeiro-negativo e verdadeiro-positivo de tabelas de confusão 2x2.

^a<https://www.rdocumentation.org/packages/mada/versions/0.5.11/topics/crosshair>

50.6 Interpretação da validade de um teste

50.6.1 Que itens devem ser verificados na interpretação de um estudo de validade?

- O novo teste foi comparado junto ao método padrão-ouro.³⁸⁶
 - As probabilidades pontuais estimadas que caracterizam o desempenho diagnóstico do novo teste são altas e adequadas para sua aplicação clínica.³⁸⁶
 - Os intervalos de confiança estimados para as probabilidades do novo teste são estreitos e adequadas para sua aplicação clínica.³⁸⁶
 - O novo teste possui adequada confiabilidade intra/inter examinadores.³⁸⁶
 - O estudo de validação incluiu um espectro adequado da amostra.³⁸⁶
 - Todos os participantes realizaram ambos o novo teste e o padrão-ouro no estudo de validação.³⁸⁶
 - Os examinadores do novo teste estavam cegados para o resultado do teste padrão-ouro.³⁸⁶

50.7 Diretrizes para redação

50.7.1 Quais são as diretrizes para redação de estudos diagnósticos?

- Visite a rede *Enhancing the Quality and Transparency Of health Research* (EQUATOR Network¹) para encontrar diretrizes específicas.
 - *STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies:*³⁹⁹ <https://www.equator-network.org/reporting-guidelines/stard/>

¹<https://www.equator-network.org/>

Capítulo 51

Ensaios quase-experimentais

51.1 Características

51.1.1 Quais são as características de ensaios quase-experimentais?

- ?

51.2 Diretrizes para redação

51.2.1 Quais são as diretrizes para redação de ensaios quase-experimentais?

- Visite a rede *Enhancing the QUAlity and Transparency Of health Research* (EQUATOR Network¹) para encontrar diretrizes específicas.
- *Guidelines for reporting non-randomised studies:*⁴⁰⁰ <https://www.equator-network.org/reporting-guidelines/guidelines-for-reporting-non-randomised-studies/>

¹<https://www.equator-network.org/>

RASCUNHO

Capítulo 52

Ensaios experimentais

52.1 Ensaio clínico aleatorizado

52.1.1 Quais são as características de ensaios clínicos aleatorizados?

- A característica essencial de um ensaio clínico aleatorizado é a comparação entre grupos.⁴⁰¹
- Quanto à unidade de alocação:⁴⁰²
 - Individual
 - Agrupado
- Quanto ao número de braços.⁴⁰²
 - Múltiplos
- Quanto ao número de centros:⁴⁰²
 - Único
 - Múltiplos
- Quanto ao cegamento:⁴⁰²
 - Aberto
 - Simples-cego
 - Duplo-cego
 - Triplo-cego
 - Quádruplo-cego
- Quanto à alocação:⁴⁰²
 - Sem sorteio
 - Estratificada (centro apenas)
 - Estratificada
 - Minimizada
 - Estratificada e minimizada

52.1.2 Quais são as estratégias metodológicas para reduzir vieses?

- Grupo controle: comparar a intervenção a um cuidado usual ou controle ativo ajuda a isolar o efeito específico do tratamento, reduzindo vieses de confusão e maturação.³
- Grupo placebo: prepara uma intervenção indistinguível da ativa para mitigar expectativas de participantes e profissionais, reduzindo viés de desempenho e detecção.³
- Controle sham: em intervenções de procedimento (p.ex., cirúrgicas/fisioterapêuticas), um comparador que reproduz etapas não-específicas do procedimento controla efeitos placebo e da atenção.³
- Cegamento: mascarar participantes, profissionais, avaliadores e/ou analistas diminui vieses de desempenho e detecção; deve-se explicitar quem foi cegado e como a manutenção do cegamento foi assegurada.³

52.2 Modelos de análise de comparação

52.2.1 Que modelos podem ser utilizados para comparações?

- As abordagens compreendem a comparação da variável de desfecho medida entre os momentos antes e depois ou da sua mudança (pré - pós) entre os momentos.⁴⁰³
- Se a média da variável é igual entre grupos no início do acompanhamento, ambas abordagens estimam o mesmo efeito. Caso contrário, o efeito será influenciado pela correlação entre as medidas antes e depois. A análise da mudança não controla para desbalanços no início do estudo.⁴⁰³
- A abordagem mais recomendada é a análise de covariância (ANCOVA) (52.1), pois ajusta os valores pós-intervenção (Y_{ij}) aos valores pré-intervenção (X_{ij}) para cada participante (i) de cada grupo $\{Z_{ij}\}$, e portanto não é afetada pelas diferenças entre grupos no início do estudo.^{10,403}

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 Z_j + \epsilon_{ij} \quad (52.1)$$

- A ANCOVA modelando seja a mudança (pré - pós: $\Delta = X_{ij} - Y_{ij}$) quando o desfecho no pós-tratamento parece ser o método mais efetivo considerando-se o viés de estimativa dos parâmetros, a precisão das estimativas, a cobertura nominal (isto é, intervalo de confiança) e o poder do teste.⁴⁰⁴
- Quando a ANCOVA (52.2) é utilizada com a mudança (pré - pós) com variável de desfecho (Y_{ij}), o coeficiente de regressão β_1 é diminuído em 1 unidade.^{10,405}

$$(X_{ij} - Y_{ij}) = \beta_0 + \beta_1 Z_j + \epsilon_{ij} \quad (52.2)$$

- Análise de variância (ANOVA) e modelos lineares mistos (MLM) são outras opções de métodos, embora apresentem maior variância, menor poder, e cobertura nominal comparados à ANCOVA.⁴⁰⁴
- Em desenhos com múltiplas medições por participante, modelos lineares mistos (efeitos aleatórios para indivíduo e, se pertinente, para centro) permitem lidar com correlação intra-sujeito e dados ausentes sob MAR, oferecendo estimativas válidas do efeito de tratamento no tempo.⁴⁰⁶
- Para dados longitudinais com desfechos contínuos, estratégias de modelo de efeitos mistos com medidas repetidas evitam a imputação explícita e, sob suposições de MAR, tendem a melhor cobertura e controle de erro tipo I do que abordagens tipo “última observação transportada”.⁴⁰⁷

52.3 Comparação na linha de base

52.3.1 O que são dados na linha de base?

- Dados sociodemográficos, clínicos e funcionais são coletados na linha de base sobre cada participante no momento da aleatorização.⁴⁰⁸
- Os dados de linha de base são usados para caracterizar os pacientes incluídos no estudo e para mostrar que os grupos de tratamento estão bem equilibrados.⁴⁰⁸
- Dados da linha de base podem ser utilizados para a aleatorização de modo a equilibrar ou estratificar os grupos considerando alguns fatores-chave.⁴⁰⁸
- Dados da linha de base podem ser utilizados como ajuste de covariável para análise do resultado por grupo de tratamento.⁴⁰⁸

52.3.2 O que é comparação entre grupos na linha de base em ensaios clínicos aleatorizados?

- A comparação se refere ao teste de hipótese nula de não haver diferença ('balanço' ou 'equilíbrio') entre grupos de tratamento nas (co)variáveis na linha de base, geralmente apresentadas apenas de modo descritivo na 'Tabela 1'.⁴⁰⁹
- A interpretação isolada do P-valor da comparação entre grupos na linha de base não permite identificar as razões para eventuais diferenças.⁴⁰⁹

52.3.3 Para quê comparar grupos na linha de base em ensaios clínicos aleatorizados?

- Os P-valores estão relacionados à aleatorização dos participantes em grupos.⁴¹⁰
- Em ensaios clínicos aleatorizados, a comparação de (co)variáveis na linha de base é usada para avaliar se aleatorização foi 'bem sucedida'.⁴¹⁰

52.3.4 Quais são as razões para diferenças entre grupos de tratamento nas (co)variáveis na linha de base?

- Acaso.^{217,409}
- Viés.^{217,409}
- Tamanho da amostra.^{217,409}
- Má conduta científica.²¹⁷

52.3.5 Quais cenários permitem a comparação entre grupos na linha de base em ensaios clínicos aleatorizados?

- Em ensaios clínicos aleatorizados agregados, os P-valores possuem interpretação diferente de estudos aleatorizados individualmente.⁴¹⁰
- Em ensaios clínicos com agrupamento, nos quais o recrutamento ocorreu após a aleatorização, os P-valores já não estão inteiramente relacionados ao processo de aleatorização, mas sim ao método de recrutamento, o que pode resultar na comparação de amostras não aleatórias.⁴¹⁰

52.3.6 Por que não se deve comparar grupos na linha de base em ensaios clínicos aleatorizados?

- A interpretação errônea dos P-valores na comparação entre grupos, na linha de base, de um ensaio clínico aleatorizado constitui a ‘falácia da Tabela 1’.²¹⁸
- Quando a aleatorização é bem-sucedida, a hipótese nula de diferença entre grupos na linha de base é verdadeira.⁴¹¹
- Testes de significância estatística na linha de base avaliam a probabilidade de que as diferenças observadas possam ter ocorrido por acaso; no entanto, já sabemos — pelo delineamento do experimento — que quaisquer diferenças são causadas pelo acaso.⁴¹²

52.3.7 Quais estratégias podem ser adotadas para substituir a comparação entre grupos na linha de base em ensaios clínicos aleatorizados?

- Na fase de projeto: identifique as variáveis prognósticas do desfecho de acordo com a literatura.⁴¹¹
- Na fase de análise: inclua as variáveis prognósticas nos modelos para ajuste.⁴¹¹

52.4 Comparação intragrupos

52.4.1 Por que não se deve comparar intragrupos (pré - pós) em ensaios clínicos aleatorizados?

- Testar por mudanças a partir da linha de base separadamente em cada grupo aleatorizado não permite concluir sobre diferenças entre grupos; não se pode fazer inferências a partir da comparação de P-valores.⁴⁰¹

52.5 Comparação entre grupos

52.5.1 O que é comparação entre grupos em ensaios clínicos aleatorizados?

- A comparação se refere ao teste de hipótese nula de não haver diferença (‘alteração’ ou ‘mudança’) pós-tratamento entre grupos de tratamento.⁴⁰¹

52.5.2 O que pode ser comparado entre grupos?

- Valores pós-tratamento; mudança entre linha de base e pós-tratamento; mudança percentual da linha de base.⁴¹³

52.5.3 Qual é a comparação entre grupos mais adequada em ensaios clínicos aleatorizados?

- Análise de covariância (ANCOVA) que analisa o pós-tratamento com a linha de base como covariável é a opção que possui maior poder estatístico.⁴¹³
- Mudança entre linha de base e pós-tratamento tem poder adequado apenas quando a correlação entre linha de base e pós-tratamento é alta.⁴¹³
- Mudança percentual da linha de base é a opção menos eficiente em termos de poder estatístico.⁴¹³

52.6 Comparação de subgrupos

52.6.1 O que é comparação de subgrupos em ensaios clínicos aleatorizados?

- Análises de subgrupos podem ser realizadas para avaliar se as diferenças no resultado do tratamento (ou a falta delas) dependem de algumas características na linha de base dos pacientes.⁴⁰⁸

52.6.2 Como realizar a comparação de subgrupos em ensaios clínicos aleatorizados?

- Testes estatísticos de interação (que avaliam se um efeito de tratamento difere entre subgrupos) devem ser usados, e não apenas a inspeção dos P-valores do subgrupo. Somente se o teste de interação estatística apoiar um efeito de subgrupo as conclusões poderão ser influenciadas.^{408,414}

52.6.3 Como interpretar a comparação de subgrupos em ensaios clínicos aleatorizados?

- Análises de subgrupos devem ser consideradas de natureza exploratória e raramente elas afetam as conclusões obtidas a partir do estudo.^{408,414}
- A credibilidade das análises de subgrupos é melhor se restrita ao desfecho primário e a alguns subgrupos predefinidos e baseadas em hipóteses biologicamente plausíveis.⁴⁰⁸
- Deve-se verificar se o estudo possui poder estatístico suficiente para detectar tamanhos de efeitos realistas em subgrupos e interpretar com cautela uma diferença de tratamento em um subgrupo quando a comparação global do tratamento não é significativa.⁴⁰⁸

52.7 Efeito de interação

52.7.1 Por que analisar o efeito de interação?

- Em ensaios clínicos aleatorizados, o principal problema de pesquisa é se há uma diferença pré - pós maior em um grupo do que em outro(s).⁴⁰¹
- A comparação de subgrupos por meio de testes de significância de hipótese nula separados é enganosa por não testar (comparar) diretamente os tamanhos dos efeitos dos tratamentos.⁴¹⁵
- Revisões recentes destacam que a interpretação de interações requer parcimônia (predefinição, plausibilidade biológica e controle do error-rate), e recomendam relatar estimativas e intervalos de confiança por subgrupo junto com o teste formal de interação.³²⁰

52.7.2 Quando usar o termo de interação?

- Análise de efeito de interação pode ser usada para testar se o efeito de um tratamento varia entre dois ou mais subgrupos de indivíduos, ou seja, se um efeito é modificado pelo(s) outros(s) efeito(s).³²¹
- A interação entre duas (ou mais) variáveis pode ser utilizada para comparar efeitos do tratamento em subgrupos de ensaios clínicos.⁴¹⁶
- O poder estatístico para detectar efeitos de interação é limitado.⁴¹⁶

52.8 Ajuste de covariáveis

52.8.1 Quais variáveis devem ser utilizadas no ajuste de covariáveis?

- A escolha das características de linha de base pelas quais uma análise é ajustada deve ser determinada pelo conhecimento prévio de uma possível influência no resultado, em vez da evidência de desequilíbrio entre os grupos de tratamento no estudo.⁴¹¹

52.8.2 Quais os benefícios do ajuste de covariáveis?

- Ajustar por covariáveis ajuda a estimar os efeitos do tratamento para o indivíduo, assim como aumenta a eficiência dos testes para hipótese nula e a validade externa do estudo.⁴¹⁷
- Incluir a variável de desfecho medida na linha de base como covariável — independentemente de a análise ser realizada com a medida pós-tratamento da mesma variável ou a diferença para a linha de base — pode aumentar o poder estatístico do estudo.⁴¹⁸
- Incluir outras variáveis medidas na linha de base, com potencial para serem desbalanceadas entre grupos após a aleatorização, diminui a chance de afetar as estimativas de efeito dos tratamentos.⁴¹⁸

52.8.3 Quais os riscos do ajuste de covariáveis?

- Incluir covariáveis que não são prognósticas do desfecho pode reduzir o poder estatístico do estudo.⁴¹⁸
- Incluir covariáveis com dados perdidos pode reduzir o tamanho amostral e consequentemente o poder estatístico do estudo (análise de casos completos) ou levar a desvios do plano de análise por exclusão de covariáveis prognósticas.⁴¹⁸

52.9 Imputação de dados perdidos

52.9.1 Como lidar com os dados perdidos em desfechos?

- Em dados longitudinais com um pequeno número de ‘ondas’ (medidas repetidas) e poucas variáveis, para análise com modelos de regressão univariados, a imputação paramétrica via especificação condicional completa - também conhecido como imputação multivariada por equações encadeadas (*multivariate imputation by chained equations*, MICE) — é eficiente do ponto de vista computacional e possui acurácia e precisão para estimação de parâmetros.^{152,419}
- Para dados perdidos em desfechos dicotômicos, o desempenho dos métodos de imputação multivariada por equações encadeadas (*multivariate imputation by chained equations*, MICE)¹⁵⁹ e por correspondência média preditiva (*predictive mean matching*, PMM)^{160,161} é similar.¹⁵⁸

52.9.2 Como lidar com os dados perdidos em covariáveis?

- Imputação de dados perdidos de uma covariável pela média dos dados do respectivo grupo permite estimativas não enviesadas dos efeitos do tratamento, preserva o erro tipo I e aumenta o poder estatístico comparado à análise de dados completos.⁴¹⁸
- Para desfechos ausentes, recomenda-se evitar transportar a última observação e, quando aplicável, preferir modelos lineares mistos ou imputação múltipla consistentes com o estimando de interesse.⁴⁰⁷

R Os pacotes *mice*¹⁵⁹ e *miceadds*¹⁶² fornecem funções *mice*^a e *mi.anova*^b para imputação multivariada por equações encadeadas, respectivamente, para imputação de dados.

^a<https://www.rdocumentation.org/packages/mice/versions/3.16.0/topics/mice>

^b<https://www.rdocumentation.org/packages/miceadds/versions/3.16-18/topics/mi.anova>

52.10 Diretrizes para redação

52.10.1 Quais são as diretrizes para redação de ensaios experimentais?

- Visite a rede *Enhancing the QUAlity and Transparency Of health Research* (EQUATOR Network¹) para encontrar diretrizes específicas.
- *CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials:*⁴²⁰ <https://www.equator-network.org/reporting-guidelines/consort/>



O pacote *consort*⁴²¹ fornece a função *consort_plot*^a para elaboração do fluxograma de ensaios experimentais no formato padrão.

^a%60r%20paste0(%22https://search.r-project.org/CRAN/refmans/%22,%20%22consort%22,%20%22/html/%22,%20%22consort_plot%22,%20%22.html%22)%60

¹<https://www.equator-network.org/>

RASCUNHO

Capítulo 53

Ensaios cruzados

53.1 Características

53.1.1 Quais são as características de ensaios cruzados?

- ?

53.2 Diretrizes para redação

53.2.1 Quais são as diretrizes para redação de ensaios cruzados?

- Visite a rede *Enhancing the QUAlity and Transparency Of health Research* (EQUATOR Network¹) para encontrar diretrizes específicas.
- *CONSORT 2010 statement: extension to randomised crossover trials:*⁴²² <https://www.equator-network.org/reporting-guidelines/consort-2010-statement-extension-to-randomised-crossover-trials/>

¹<https://www.equator-network.org/>

RASCUNHO

Capítulo 54

N de 1

54.1 Ensaio N-de-1

54.1.1 O que são ensaios N-de-1?

- Ensaios N-de-1 são delineamentos experimentais em que um único paciente recebe, em períodos alternados, duas ou mais intervenções (ex.: tratamento A e tratamento B).⁴²³
- Cada ciclo é formado por dois períodos (AB ou BA), cuja ordem é randomizada, garantindo controle temporal e redução de vieses.⁴²³
- O foco está na comparação intraindivíduo, permitindo avaliar diretamente se o paciente em questão responde melhor a uma intervenção.⁴²³

54.1.2 Quando usar ensaios N-de-1?

- Doenças crônicas estáveis, em que o desfecho pode ser observado repetidamente.⁴²³
- Condições raras ou com grande heterogeneidade de resposta entre pacientes.⁴²³
- Situações clínicas de incerteza, quando se deseja personalizar o tratamento.⁴²³

54.1.3 Qual a relevância dos ensaios N-de-1?

- Os ensaios N-de-1 permitem decisões clínicas personalizadas e baseadas em evidência direta.⁴²³
- Quando combinados, podem gerar estimativas comparáveis às de ensaios clínicos convencionais, mantendo o foco centrado no paciente.⁴²³
- Representam uma alternativa metodológica robusta para cenários de incerteza terapêutica.⁴²³

54.2 Aspectos metodológicos

54.2.1 Como é feita a randomização?

- A ordem dos tratamentos em cada ciclo é definida aleatoriamente (ex.: AB, BA, AB...).⁴²³
- Essa randomização reduz viés de período e efeito de expectativa.⁴²³

54.2.2 Como são feitas as análises?

- Comparações intraindivíduo (testes pareados ou estimativas de efeito médio por paciente).⁴²³
- Combinação de múltiplos N-de-1 por meio de meta-análises ou modelos mistos para inferências em nível populacional.⁴²³

54.2.3 Quais perguntas de inferência podem ser respondidas?

- Q1: Há efeito do tratamento dentro dos ciclos de um paciente?⁴²³
- Q2: Qual é o efeito médio observado nos pacientes estudados?⁴²³
- Q3: O efeito é homogêneo ou heterogêneo entre pacientes?⁴²³
- Q4: Qual é o efeito específico em cada paciente individual?⁴²³
- Q5: Qual é o efeito esperado em populações semelhantes?⁴²³

54.3 Limitações e cuidados

54.3.1 Quais são os principais desafios dos ensaios N-de-1?

- Baixo poder estatístico quando poucos ciclos são realizados.⁴²³
- Necessidade de períodos de *washout* para evitar efeito de *carry-over*.⁴²³
- Interpretação dependente de pressupostos sobre homogeneidade ou heterogeneidade dos efeitos.⁴²³
- Em amostras muito pequenas, pode ser necessário usar variâncias externas ou modelos mistos.⁴²³

Capítulo 55

Revisão sistemática

55.1 Tipologia de revisões

55.1.1 O que é a tipologia de revisões?

- Foram mapeados 14 tipos de revisão e suas metodologias, organizando-as pelo modelo SALSA (*Search, Appraisal, Synthesis, Analysis*). Essa tipologia demonstra que diferentes revisões têm diferentes níveis de rigor, sistematicidade e propósito, ajudando pesquisadores a escolher o delineamento mais adequado.³⁵³

55.2 Revisão sistemática de literatura

55.2.1 O que é revisão sistemática?

- Uma revisão sistemática é um tipo de estudo secundário que utiliza métodos explícitos, transparentes e reproduzíveis para identificar, selecionar, avaliar criticamente e sintetizar todas as evidências relevantes sobre uma pergunta claramente definida. Trata-se de uma estratégia formal para reduzir vieses e aumentar a confiabilidade das conclusões, distinguindo-se de revisões narrativas tradicionais por seu rigor metodológico.^{353,424}

R

O pacote *easyPubMed*⁴²⁵ fornece a função *get_pubmed_ids*^a e *fetch_pubmed_data*^b para buscar artigos no PubMed^c.

^a<https://cran.r-project.org/web/packages/easyPubMed/index.html>

^b<https://cran.r-project.org/web/packages/easyPubMed/index.html>

^c<https://pubmed.ncbi.nlm.nih.gov>

R

O pacote *rcrossref*⁴²⁶ fornece a função *cr_works*^a para buscar artigos.

^a<https://cloud.r-project.org/web/packages/rcrossref/index.html>

R

O pacote *roadoi*⁴²⁷ fornece a função *oadoi_fetch*^a para recuperar dados de acesso aberto do Unpaywall^b.

^a<https://cloud.r-project.org/web/packages/roadoi/index.html>

^b<https://unpaywall.org>

55.3 Tipos de revisão sistemática

55.3.1 Quais são os principais tipos de revisão sistemática?

- Revisão sistemática tradicional (*systematic review*): Síntese estruturada de estudos primários (ex.: ensaios clínicos, estudos observacionais) para responder a uma pergunta específica. Caracteriza-se por protocolo explícito, critérios de inclusão definidos e busca exaustiva.^{353,424}
- *Overview of Reviews* (também chamada de *review of reviews*): Tipo de síntese que reúne resultados de revisões sistemáticas previamente publicadas. O foco não está em estudos primários, mas nos achados das próprias revisões. É útil quando já existem muitas revisões sistemáticas sobre um mesmo tema e há necessidade de sumarizar o conjunto da evidência disponível de forma organizada.^{353,424,428,429}
- *Umbrella Review*: Um subtipo de *overview* em que a unidade de análise é a própria revisão sistemática. Tem escopo mais amplo e busca integrar achados de múltiplas revisões sobre um tópico, permitindo identificar consistências e discrepâncias entre revisões. São cada vez mais comuns devido ao aumento do número de revisões sistemáticas publicadas anualmente.^{353,430}
- Revisões sistemáticas qualitativas (*Qualitative evidence synthesis*): Focam em estudos qualitativos para responder questões sobre experiências, percepções ou processos. Podem empregar métodos como meta-aggregation, meta-ethnography ou thematic synthesis. Têm crescido em importância porque muitas questões da prática clínica não podem ser respondidas apenas por estudos experimentais.^{353,424}
- Revisões mistas (*Mixed-methods review*): Integram evidências quantitativas e qualitativas. São úteis quando a compreensão de um fenômeno exige tanto medidas objetivas quanto interpretações subjetivas, mas apresentam desafios metodológicos adicionais pela necessidade de combinar técnicas analíticas distintas.^{353,424}
- Revisão rápida (*Rapid review*): Uma forma condensada de revisão sistemática, com métodos acelerados (redução de bases de dados, um único revisor em algumas etapas). É útil em contextos que exigem respostas em curto prazo, embora apresente maior risco de viés.^{353,424}

55.3.2 Quais delineamentos de revisão parecem mas não são revisões sistemáticas?

- Revisão narrativa estruturada (*Structured narrative review*): Narrativas estruturadas não seguem métodos reproduzíveis como buscas explícitas e critérios formais de inclusão, portanto não devem ser confundidas com revisões sistemáticas, como alerta Baker et al. (2014), que destaca a importância de diferenciar revisões críticas, integrativas e narrativas de revisões sistemáticas verdadeiras.^{353,424}
- Revisão integrativa (*Integrative review*): Embora possa incluir estudos primários, não segue os rigorosos métodos de uma revisão sistemática, como busca exaustiva e critérios de inclusão pré-definidos. Portanto, não deve ser confundida com uma revisão sistemática.^{353,424}
- Revisão crítica (*Critical review*): Foca na avaliação crítica da literatura, mas não necessariamente segue os métodos sistemáticos de uma revisão sistemática, como busca estruturada e critérios de inclusão claros.^{353,424}

55.4 Diretrizes para redação

55.4.1 Quais são as diretrizes para revisão sistemática?

- Visite a rede *Enhancing the QUAlity and Transparency Of health Research* (EQUATOR Network¹) para encontrar diretrizes específicas.
- *Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA)*.⁴³¹ <https://www.equator-network.org/reporting-guidelines/tripod-srma/>
- *Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist*.⁴³² <https://doi.org/10.1371/journal.pmed.1001744>

¹<https://www.equator-network.org/>

RASCUNHO

Capítulo 56

Meta-análise

56.1 Características

56.1.1 O que é meta-análise?

- Meta-análise é um método estatístico que combina quantitativamente os resultados de múltiplos estudos independentes sobre uma mesma questão de pesquisa, aumentando o poder estatístico e a precisão das estimativas de efeito.⁴³³
- Meta-análise sintetiza evidências considerando o peso de cada estudo (geralmente inversamente proporcional à variância) e permite avaliar a consistência dos resultados, identificar fontes de heterogeneidade e estimar um efeito global.⁴³³

56.2 Modelos de meta-análise

56.2.1 Quais são os principais modelos de meta-análise?

- Modelo de efeitos fixos: assume que todos os estudos avaliam o mesmo efeito verdadeiro, e a variação observada é apenas devido ao erro de amostragem. É adequado quando os estudos são homogêneos e as diferenças entre eles são pequenas.⁴³³
- Modelo de efeitos aleatórios: assume que os estudos avaliam efeitos verdadeiros diferentes, com uma distribuição normal. É mais apropriado quando há heterogeneidade entre os estudos, pois considera a variação entre eles além do erro de amostragem.⁴³³
- Modelo de efeitos de rede: estende a meta-análise para comparar múltiplas intervenções simultaneamente, mesmo que não tenham sido comparadas diretamente em estudos. É útil para avaliar a eficácia relativa de várias intervenções.⁴³³

O pacote *metafor*⁴³⁴ fornece a função *forest*^a para criar figuras tipo *forest plot*.

^a<https://cran.r-project.org/web/packages/metafor/index.html>

R

O pacote *netmeta*⁴³⁵ fornece a função *netmeta*^a para realizar meta-análise de rede usando método de grafo.

^a<https://CRAN.R-project.org/package=netmeta>

R

O pacote *gemtc*⁴³⁶ fornece a função *mtc.model*^a para criar modelos de meta-análise de rede.

^a<https://CRAN.R-project.org/package=gemtc>

Curvas Normais por Estudo: FE vs RE

Curvas por estudo ($FE = \sigma$; $RE = \sqrt{(\sigma^2 + \tau^2)}$). Linhas verticais: pooled FE

Modelo — FE — RE

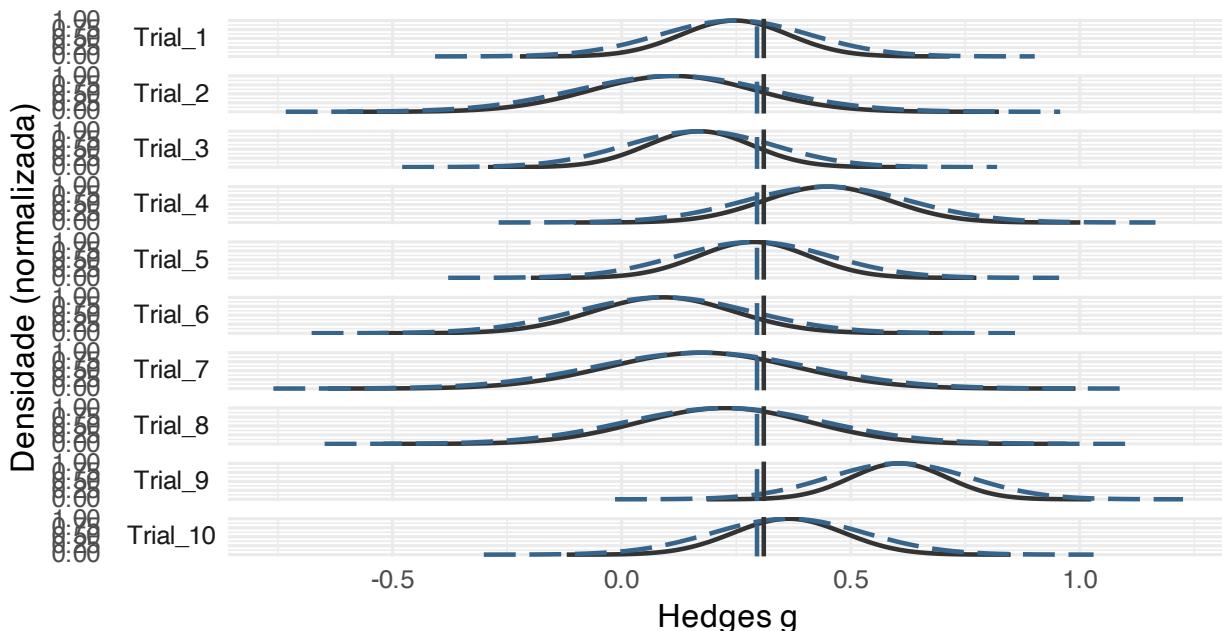


Figura 56.1: Comparação entre modelos de efeito fixo e aleatório com 10 ensaios clínicos simulados.

56.3 Conversão de Medidas em Meta-análises

56.3.1 O que fazer quando os estudos apresentam resultados com diferentes parâmetros?

- Quando os estudos reportam médias e desvios-padrão, os dados podem ser usados diretamente na metanálise.⁴³⁸
- Quando apresentam mediana e intervalo interquartil (ou mínimo–máximo), existem métodos estatísticos para converter em média e DP.⁴³⁷
- Hozo et al. (2005) propuseram fórmulas para estimar a média e o desvio-padrão a partir da mediana, amplitude e tamanho da amostra.⁴³⁷

- Wan et al. (2014) aperfeiçoaram essas estimativas, oferecendo métodos mais precisos para converter mediana e IQR em média e DP.⁴³⁸



O pacote *metafor*⁴³⁴ fornece a função *conv.fivenum*^a para converter mínimo/mediana/máximo ou Q1/mediana/Q2 em média e desvio-padrão.

^a<https://cran.r-project.org/web/packages/metafor/refman/metafor.html#conv.fivenum>

56.4 Interpretação de efeitos em meta-análise

56.4.1 Como avaliar a variação do tamanho do efeito?

- O intervalo de predição contém informação sobre a variação do tamanho do efeito.⁴³⁹
- Se o intervalo de predição não contém a hipótese nula (H_0), podemos concluir que (a) o tratamento funciona igualmente bem em todas as populações, ou que ele funciona melhor em algumas populações do que em outras.⁴³⁹
- Se o intervalo de predição contém a hipótese nula (H_0), podemos concluir que o tratamento pode ser benéfico em algumas populações, mas prejudicial em outras, de modo que a estimativa pontual (geralmente a média) torna-se amplamente irrelevante. Nesse caso, é recomendado investigar em que populações o tratamento seria benéfico e em quais causaria danos.⁴³⁹

56.4.2 Como avaliar a heterogeneidade entre os estudos?

- A heterogeneidade — variação não-aleatória — no efeito do tratamento entre os estudos incluídos em uma meta-análise pode ser avaliada pelo I^2 (56.1).^{439,440}

$$I^2 = \max \left(0, \frac{Q - df}{Q} \right) \times 100\% \quad (56.1)$$

- I^2 representa qual proporção da variância observada reflete a variância nos efeitos verdadeiros em vez do erro de amostragem.⁴³⁹
- I^2 mede a proporção da variância total que pode ser atribuída à heterogeneidade entre os estudos incluídos.⁴⁴⁰
- I^2 não depende da quantidade de estudos incluídos na meta-análise. Entretanto, I^2 aumenta com a quantidade de participantes incluídos nos estudos meta-analisisados.⁴⁴⁰
- A heterogeneidade entre estudos é explicada de modo mais confiável utilizando dados de pacientes individuais, uma vez que a direção verdadeira da modificação de efeito não pode ser observada a partir de dados agregados no estudo.⁴⁴¹



O pacote *psychmeta*²⁹⁷ fornece a função *ma_d*^a para meta-analisar valores d .

^ahttps://www.rdocumentation.org/packages/psychmeta/versions/2.7.0/topics/ma_d



O pacote *psychmeta*²⁹⁷ fornece a função *ma_r*^a para meta-analisar correlações.

^ahttps://www.rdocumentation.org/packages/psychmeta/versions/2.7.0/topics/ma_r

56.5 Interpretação do *forest plot*

56.5.1 O que é um *forest plot*?

- Um *forest plot* é uma representação gráfica dos achados de uma meta-análise. Ele resume os resultados de estudos individuais e apresenta uma estimativa combinada do efeito, permitindo interpretação visual da magnitude, direção e significância estatística dos resultados.⁴⁴²

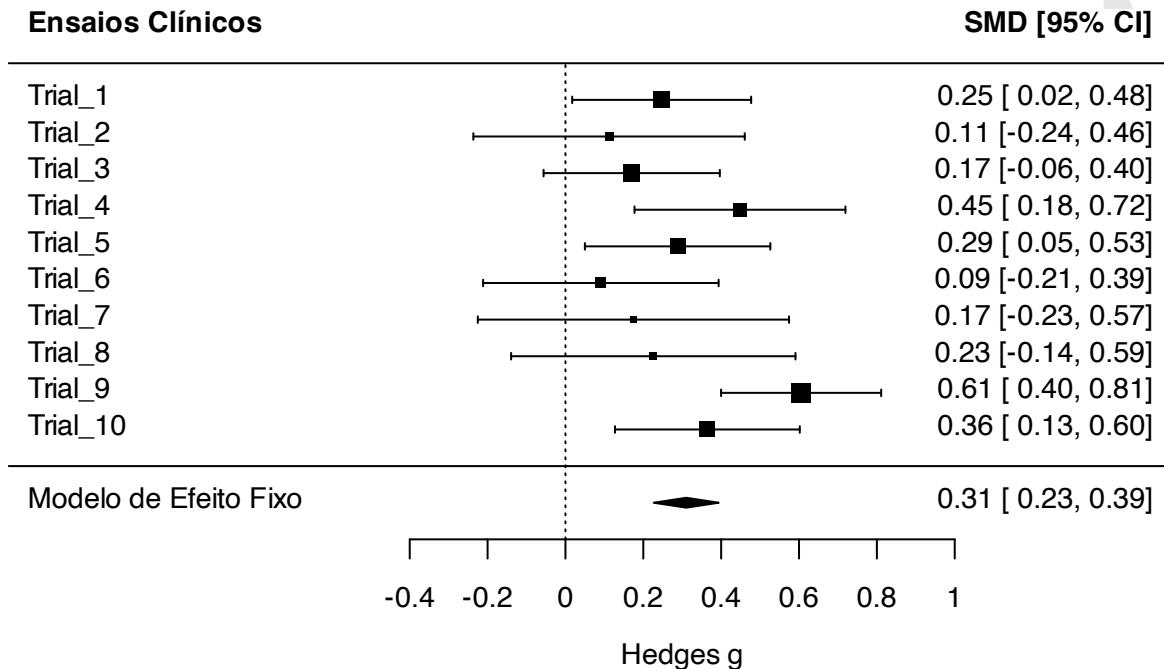


Figura 56.2: Forest plot de uma meta-análise de efeito fixo com 10 ensaios clínicos simulados.

56.5.2 Quais são as seis colunas básicas que um *forest plot* geralmente apresenta?

- As seis colunas básicas incluem: estudos incluídos (e subgrupos, se analisados); dados do grupo de intervenção; dados do grupo controle; peso de cada estudo; medida numérica do efeito; representação gráfica do efeito.⁴⁴²

56.5.3 Como diferenciar um desfecho binário de um contínuo em um *forest plot*?

- Em desfechos binários, são mostrados número de eventos e total da amostra, sendo o efeito medido por *risk ratio* (*RR*) ou *odds ratio* (*OR*).⁴⁴²
- Em desfechos contínuos, apresentam-se médias, desvios-padrão e tamanhos amostrais, com o efeito medido pela diferença de médias.⁴⁴²

56.5.4 O que representa o ponto central da caixa e o tamanho desta no gráfico?

- O ponto central indica a estimativa pontual do efeito (melhor estimativa para o efeito real).⁴⁴²
- O tamanho da caixa é proporcional ao peso do estudo na meta-análise, geralmente maior para estudos com amostras maiores.⁴⁴²

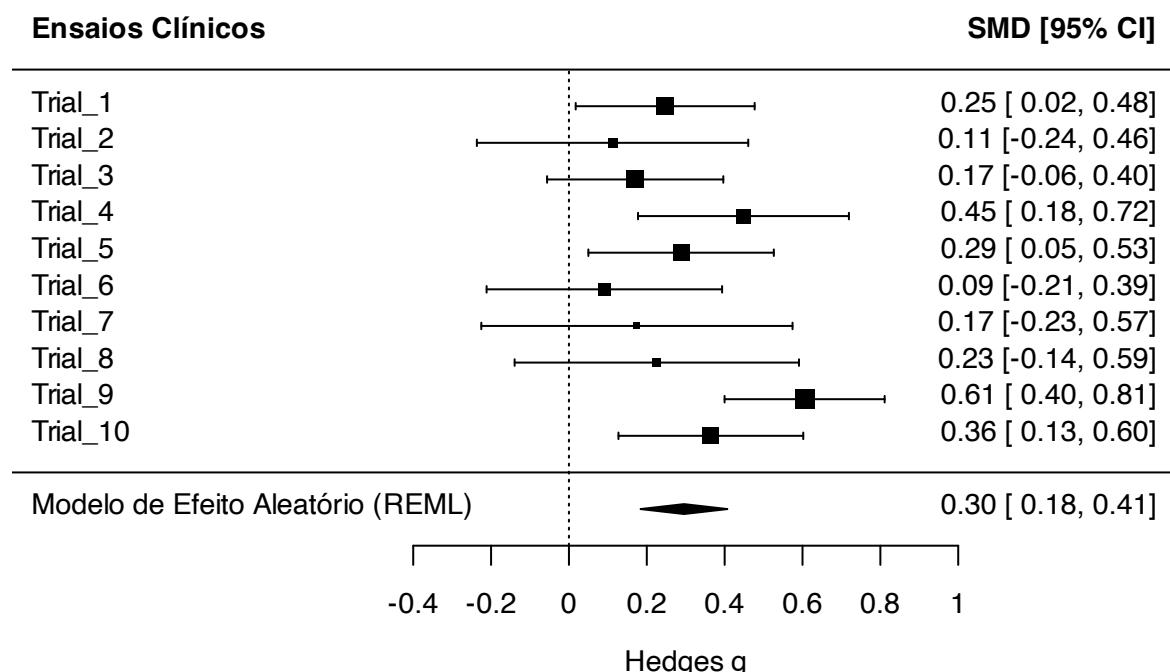


Figura 56.3: Forest plot de uma meta-análise de efeito aleatório com 10 ensaios clínicos simulados.

56.5.5 Qual é o significado da linha vertical do “nenhum efeito”?

- É a linha de referência que indica efeito nulo.⁴⁴²
- Para desfechos binários, corresponde ao valor 1 (RR ou $OR = 1$).⁴⁴²
- Para desfechos contínuos, corresponde ao valor 0 (diferença de médias = 0).⁴⁴²
- Se o intervalo de confiança de um estudo ou do resultado combinado cruza essa linha, o resultado não é estatisticamente significativo.⁴⁴²

56.5.6 Como interpretar o diamante na parte inferior do *forest plot*?

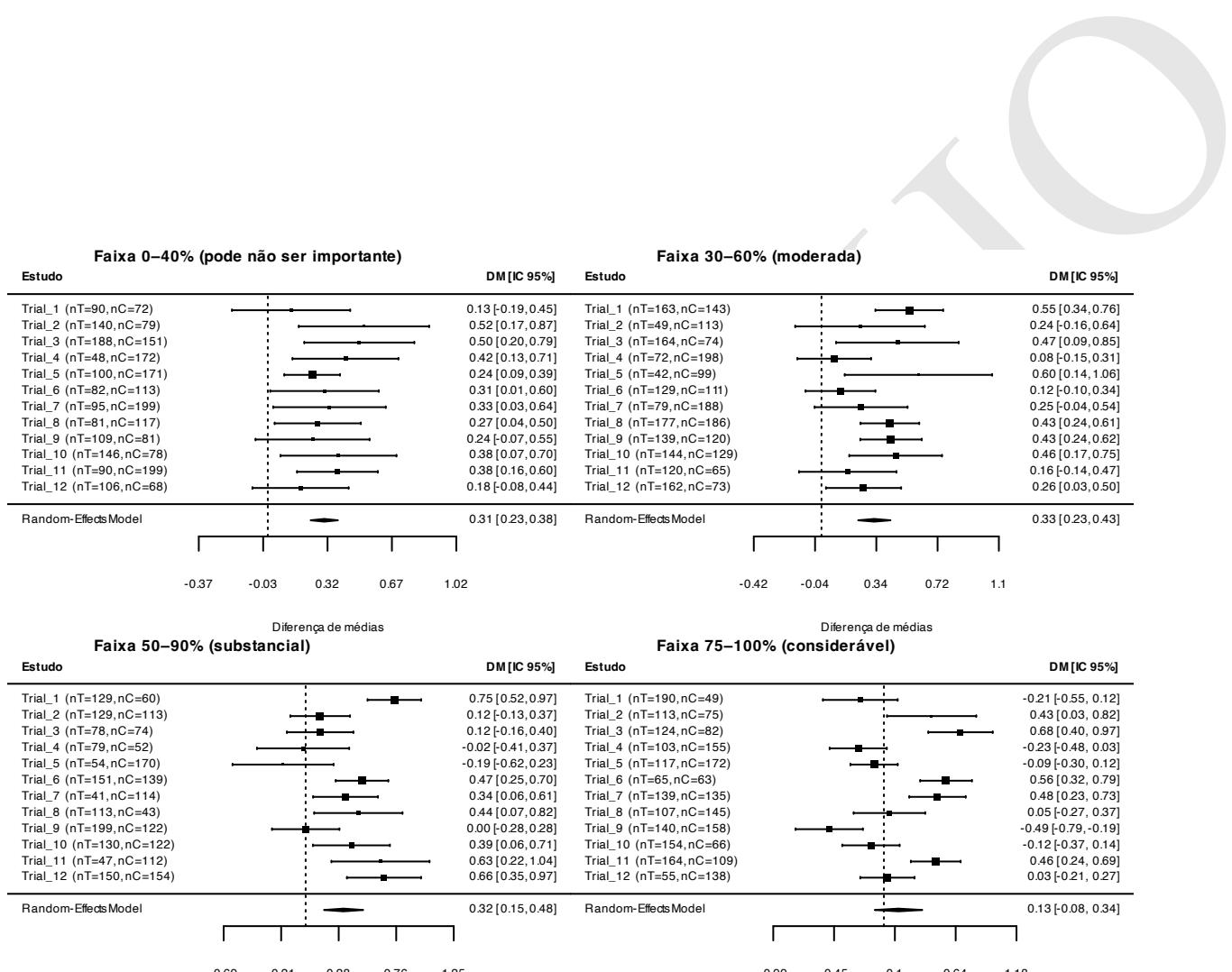
- O diamante representa o efeito combinado dos estudos incluídos.⁴⁴²
- O ponto central do diamante é a estimativa global.⁴⁴²
- A largura do diamante representa o intervalo de confiança de 95% para o efeito combinado.⁴⁴²

56.5.7 Como a heterogeneidade pode ser avaliada no *forest plot*?

- A variabilidade nos resultados dos estudos incluídos é avaliada pela sobreposição dos intervalos de confiança dos estudos; pelo teste do qui-quadrado (χ^2) e pelo valor de I^2 .⁴⁴²

56.5.8 Quais são as interpretações usuais para os valores de heterogeneidade?

- I^2 de 0% a 40%: pode não ser importante; 30% a 60%: heterogeneidade moderada; 50% a 90%: heterogeneidade substancial; 75% a 100%: heterogeneidade considerável.⁴⁴²

Figura 56.4: Forest plots ilustrativos para faixas usuais de I^2 .

56.6 Viés de publicação em meta-análises

56.6.1 O que é viés de publicação?

- O viés de publicação ocorre quando estudos com resultados não significativos ou contrários à hipótese tendem a não ser publicados, afetando a estimativa final da meta-análise e podendo levar a conclusões incorretas.⁴⁴³

56.6.2 Quais métodos podem ser usados para identificar viés de publicação?

- O método mais simples é o *funnel plot*, que representa a estimativa de efeito de cada estudo em função de sua precisão ($1/SE$).⁴⁴⁴
- Na ausência de viés, espera-se uma distribuição simétrica (“forma de funil”). Assimetria pode indicar viés de publicação, heterogeneidade entre estudos ou efeitos de tamanho de estudo (*small-study effects*).⁴⁴⁴
- Para *odds ratios (OR)*, a correlação entre $\ln(OR)$ e seu erro padrão pode gerar assimetria mesmo sem viés, por isso recomenda-se, nesses casos, plotar em função do tamanho amostral.⁴⁴⁵

56.6.3 O que é um gráfico de funil (*funnel plot*)?

- É um gráfico de dispersão que relaciona a estimativa de efeito de cada estudo com uma medida de seu tamanho ou precisão (por exemplo, erro-padrão no eixo vertical, em escala invertida). Em condições ideais (ausência de viés e heterogeneidade), os estudos se distribuem de forma simétrica, formando um “funil invertido”.⁴⁴⁶

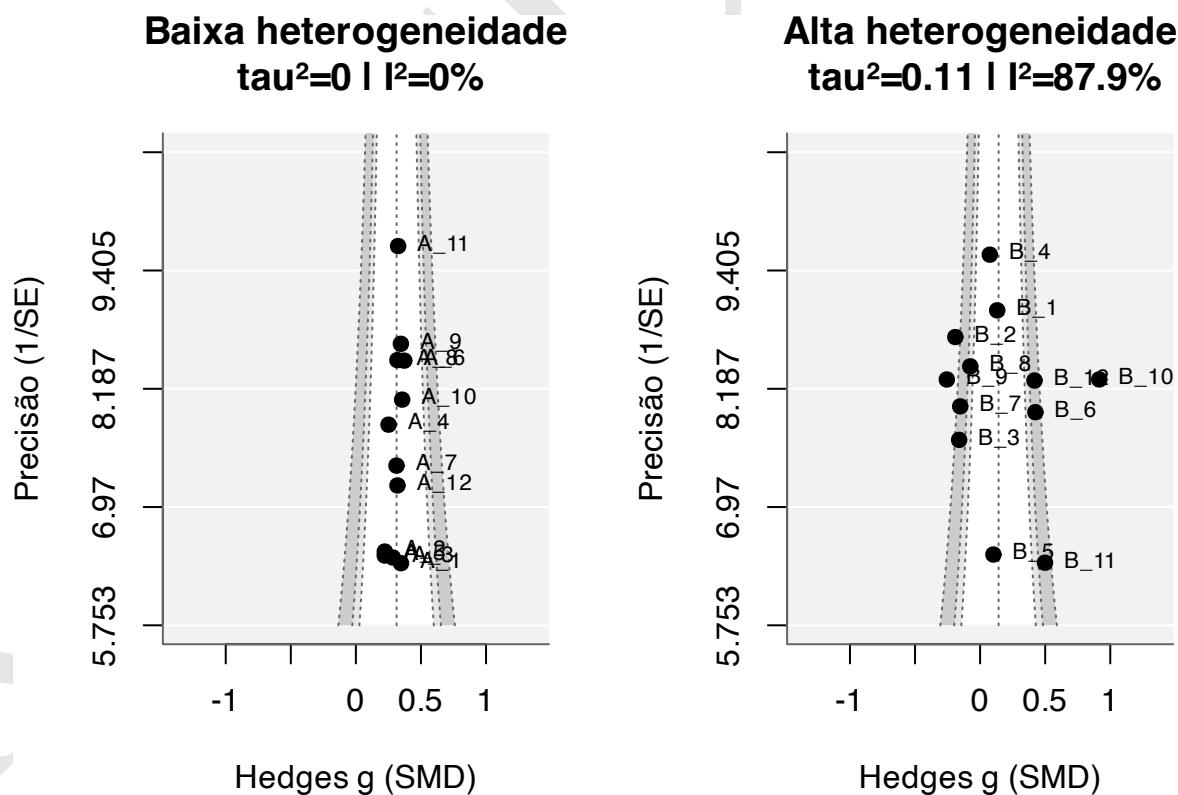


Figura 56.5: Gráficos de funil simulados com baixa e alta heterogeneidade.

56.6.4 A assimetria do *funnel plot* indica sempre viés de publicação?

- Viéses de relato (*reporting biases*), como viés de publicação, viés de linguagem ou de citação.⁴⁴⁶
- Diferenças metodológicas entre estudos pequenos e grandes.⁴⁴⁶
- Heterogeneidade verdadeira (diferença real no efeito conforme o tamanho ou o contexto do estudo).⁴⁴⁶
- Artefatos estatísticos ou mero acaso.⁴⁴⁶

56.6.5 O que é *trim and fill*?

- O método *trim and fill* “apara” (trim) os estudos que causam assimetria no funnel plot, estima o número de estudos possivelmente ausentes (não publicados) e “preenche” (fill) o gráfico com esses estudos simulados, recalculando o efeito combinado.⁴⁴⁷
- O método assume que a assimetria é causada unicamente por viés de publicação, podendo levar a conclusões equivocadas quando há outras causas, como heterogeneidade.⁴⁴⁷

56.6.6 O que é o teste de Egger?

- É um teste estatístico amplamente utilizado que avalia a relação entre o efeito padronizado (*efeito/SE*) e a precisão ($1/SE$).⁴⁴⁴
- No entanto, para meta-análises com *OR*, apresenta taxas de erro tipo I excessivas, especialmente quando o efeito é grande ou há alta heterogeneidade.⁴⁴⁵

56.6.7 O que é o teste de Peters?

- Uma regressão linear ponderada com $\ln(OR)$ como variável dependente e o inverso do tamanho total da amostra como variável independente (modificação do teste de Macaskill).⁴⁴⁵
- Essa abordagem reduz a correlação entre $\ln(OR)$ e seu *SE*, resultando em taxas de erro tipo I mais adequadas (~10%) independentemente do tamanho do efeito, número de estudos ou heterogeneidade.⁴⁴⁵
- O teste de Peters é preferível ao teste de Egger quando o desfecho é expresso como *OR*, pois mantém taxas de erro tipo I adequadas e ainda apresenta poder comparável para detectar viés em condições de baixa heterogeneidade.⁴⁴⁵
- Em casos de alta heterogeneidade, o teste de Egger pode apresentar maior poder, mas sua alta taxa de falsos positivos compromete a interpretação.⁴⁴⁵

56.6.8 Quais são as recomendações para testar a assimetria?

- Evitar testes quando há menos de 10 estudos, devido ao baixo poder estatístico.⁴⁴⁶
- Inspeccionar visualmente o *funnel plot* junto com os resultados dos testes.⁴⁴⁶
- Para desfechos contínuos (diferença de médias), o teste de Egger pode ser usado.⁴⁴⁶
- Para desfechos dicotômicos expressos como *odds ratio* (*OR*) com baixa heterogeneidade ($\tau^2 < 0,1$), considerar os testes de Harbord, Peters ou Rücker.⁴⁴⁶
- Para desfechos dicotômicos expressos como *odds ratio* (*OR*) com alta heterogeneidade ($\tau^2 > 0,1$), o teste de Rücker com transformação *arcsine* é mais indicado.⁴⁴⁶

56.6.9 Como interpretar os resultados de testes de viés de publicação?

- Um resultado não significativo não garante ausência de viés.⁴⁴⁵
- Recomenda-se complementar com inspeção visual do *funnel plot* e considerar métodos adicionais como *trim and fill*.^{445,447}
- Quando há suspeita de viés, discutir as implicações e interpretar o efeito combinado com cautela.⁴⁴⁵



O pacote *psychmeta*²⁹⁷ fornece a função *plot_funnel*^a para criar figuras tipo *funnel plot*.

^ahttps://www.rdocumentation.org/packages/psychmeta/versions/2.7.0/topics/plot_funnel

56.7 Diretrizes para redação

56.7.1 Quais são as diretrizes para redação de meta-análises?

- Visite a rede *Enhancing the QUAlity and Transparency Of health Research* (EQUATOR Network¹) para encontrar diretrizes específicas.
- *The PRISMA 2020 statement: An updated guideline for reporting systematic reviews*:⁴⁴⁸ <https://www.equator-network.org/reporting-guidelines/prisma/>
- *Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA)*.⁴³¹ <https://www.equator-network.org/reporting-guidelines/tripod-srma/>



O pacote *metagear*⁴⁴⁹ fornece a função *plot_PRISMA*^a para gerar o fluxograma de uma revisão sistemática de acordo com o *Preferred Reporting Items for Systematic Reviews and Meta-Analyses*⁴⁵⁰.

^ahttps://www.rdocumentation.org/packages/metagear/versions/0.7/topics/plot_PRISMA



O pacote *PRISMA2020*⁴⁵¹ fornece a função *PRISMA_flowdiagram*^a para elaboração do fluxograma de revisões sistemáticas no formato padrão.

^ahttps://www.rdocumentation.org/packages/PRISMA2020/versions/1.1.1/topics/PRISMA_flowdiagram

¹<https://www.equator-network.org/>

RASCUNHO

Capítulo 57

Revisão guarda-chuva

57.1 Revisão guarda-chuva

57.1.1 O que é revisão guarda-chuva?

• ?

57.2 Diretrizes para redação

57.2.1 Quais são as diretrizes para revisão sistemática?

- Visite a rede *Enhancing the QUAlity and Transparency Of health Research* (EQUATOR Network¹) para encontrar diretrizes específicas.
- *Reporting guideline for overviews of reviews of healthcare interventions: development of the PRIOR statement.*⁴⁵² <https://www.equator-network.org/reporting-guidelines/reporting-guideline-for-overviews-of-reviews-of-healthcare-interventions-development-of-the-prior-statement/>

¹<https://www.equator-network.org/>

RASCUNHO

Capítulo 58

Pesquisa qualitativa

58.1 Pesquisa qualitativa

58.1.1 O que é pesquisa qualitativa?

- No contexto de ensaios clínicos randomizados, pesquisa qualitativa é usada para compreender a complexidade das intervenções e dos contextos sociais em que são testadas, contribuindo para gerar evidências de efetividade.⁴⁵³
- O valor potencial da pesquisa qualitativa inclui: otimizar a intervenção e procedimentos dos ensaios clínicos randomizados; facilitar a interpretação dos achados; fortalecer a condução ética; ampliar a validade externa ; e economizar recursos ao direcionar futuros ensaios clínicos randomizados para intervenções mais promissoras.⁴⁵³

58.2 Diretrizes para redação

58.2.1 Quais são as diretrizes para redação de estudos de propriedades psicométricas?

- Visite a rede *Enhancing the QUAlity and Transparency Of health Research* (EQUATOR Network¹) para encontrar diretrizes específicas.
- *Standards for reporting qualitative research: a synthesis of recommendations:*⁴⁵⁴ <https://www.equator-network.org/reporting-guidelines/srqr/>
- *Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ:*⁴⁵⁵ <https://www.equator-network.org/reporting-guidelines/entreq/>
- *Consolidated criteria for reporting qualitative research (COREQ) : a 32-item checklist for interviews and focus groups:*⁴⁵⁶ <https://www.equator-network.org/reporting-guidelines/coreq/>

¹<https://www.equator-network.org/>

RASCUNHO

PARTE 10: COMUNICAÇÃO E RELATO CIENTÍFICO

Transformando resultados em narrativas claras, completas e alinhadas às boas práticas

RASCUNHO

Capítulo 59

Redação de resultados

59.1 Resultados da análise estatística

59.1.1 Como redigir os resultados da análise estatística?

- ?



O pacote *report*⁴⁵⁷ fornece a função *report*^a para redigir a descrição de diversas análises estatísticas.

^a<https://www.rdocumentation.org/packages/report/versions/0.5.8/topics/report>



O pacote *statcheck*⁴⁵⁸ fornece a função *statcheck*^a para extrair resultados de testes de significância de hipótese nula.

^a<https://www.rdocumentation.org/packages/statcheck/versions/1.4.0/topics/statcheck>

59.2 Diretrizes e Listas

59.2.1 Quais diretrizes estão disponíveis para redação estatística?

- *Review of guidance papers on regression modeling in statistical series of medical journals.*⁴⁵⁹
- *Principles and recommendations for incorporating estimands into clinical study protocol templates.*⁴⁶⁰
- *How to write statistical analysis section in medical research.*²⁵⁴
- *Recommendations for Statistical Reporting in Cardiovascular Medicine: A Special Report From the American Heart Association.*⁴⁶¹
- *Framework for the treatment and reporting of missing data in observational studies: The Treatment And Reporting of Missing data in Observational Studies framework.*⁴⁶²
- *Guidelines for reporting of figures and tables for clinical research in urology.*⁴⁶³
- *Who is in this study, anyway? Guidelines for a useful Table 1.*²¹⁹
- *Guidelines for Reporting of Statistics for Clinical Research in Urology.*⁴⁶⁴

- *Reveal, Don't Conceal: Transforming Data Visualization to Improve Transparency.*²⁰⁸
- *Guidelines for the Content of Statistical Analysis Plans in Clinical Trials.*³⁵¹
- *Basic statistical reporting for articles published in Biomedical Journals: The ‘‘Statistical Analyses and Methods in the Published Literature’’ or the SAMPL Guidelines.*⁴⁶⁵
- *Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm.*⁴⁶⁶
- *STRengthening analytical thinking for observational studies: the STRATOS initiative.*⁴⁶⁷
- *Research methods and reporting.*⁴⁶⁸
- *How to ensure your paper is rejected by the statistical reviewer.*⁴⁶⁹

59.2.2 Quais listas de verificação estão disponíveis para redação estatística?

- *A CHecklist for statistical Assessment of Medical Papers (the CHAMP statement): explanation and elaboration.*⁴⁷⁰
- *Checklist for clinical applicability of subgroup analysis.*⁴⁷¹
- *Evidence-based statistical analysis and methods in biomedical research (SAMBR) checklists according to design features.*²⁵³

Capítulo 60

Diretrizes e Listas

60.1 Diretrizes

60.1.1 Quais são as diretrizes para relatórios estatísticos em pesquisas?

- Visite a rede *Enhancing the QUAlity and Transparency Of health Research* (EQUATOR Network¹) para encontrar diretrizes específicas.
- *Guidelines for Reporting Observational Research in Urology: The Importance of Clear Reference to Causality.*²⁴³
- *Review of guidance papers on regression modeling in statistical series of medical journals.*⁴⁵⁹
- *Principles and recommendations for incorporating estimands into clinical study protocol templates.*⁴⁶⁰
- *How to write statistical analysis section in medical research.*²⁵⁴
- *A Guideline for Reporting Mediation Analyses of Randomized Trials and Observational Studies: The AGReMA Statement.*⁴⁷²
- *Recommendations for Statistical Reporting in Cardiovascular Medicine: A Special Report From the American Heart Association.*⁴⁶¹
- *Framework for the treatment and reporting of missing data in observational studies: The Treatment And Reporting of Missing data in Observational Studies framework.*⁴⁶²
- *Guidelines for reporting of figures and tables for clinical research in urology.*⁴⁶³
- *Who is in this study, anyway? Guidelines for a useful Table 1.*²¹⁹
- *Guidelines for Reporting of Statistics for Clinical Research in Urology.*⁴⁶⁴
- *Reveal, Don't Conceal: Transforming Data Visualization to Improve Transparency.*²⁰⁸
- *Guidelines for the Content of Statistical Analysis Plans in Clinical Trials.*³⁵¹
- *Basic statistical reporting for articles published in Biomedical Journals: The ‘‘Statistical Analyses and Methods in the Published Literature’’ or the SAMPL Guidelines.*⁴⁶⁵
- *Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm.*⁴⁶⁶
- *STRengthening analytical thinking for observational studies: the STRATOS initiative.*⁴⁶⁷
- *Research methods and reporting.*⁴⁶⁸

¹<https://www.equator-network.org/>

- *How to ensure your paper is rejected by the statistical reviewer.*⁴⁶⁹

60.2 Listas de verificação

60.2.1 Quais são as listas de verificação para relatórios estatísticos em pesquisas?

- *A Checklist for statistical Assessment of Medical Papers (the CHAMP statement): explanation and elaboration.*⁴⁷⁰
- *Checklist for clinical applicability of subgroup analysis.*⁴⁷¹
- *Evidence-based statistical analysis and methods in biomedical research (SAMBR) checklists according to design features.*²⁵³

Capítulo 61

Fontes externas

61.1 Fontes de informação externas

61.1.1 American Heart Association

- *Statistical Reporting Recommendations - AHA/ASA journals*¹

61.1.2 American Physiological Society

- *Statistics*²
- *Exploration in Statistics*³
- *General Statistics*⁴
- *Reporting Statistics*⁵

61.1.3 American Statistical Association

- *Statistical Inference in the 21st Century: A World Beyond $p < 0.05$* - The American Statistical Association⁶

61.1.4 British Medicine Journal

- *Statistics - Latest from The BMJ*⁷
- *Statistics notes - Latest from The BMJ*⁸
- *Statistics and research methods - Latest from The BMJ*⁹
- *Statistics at Square One*¹⁰

¹<https://www.ahajournals.org/statistical-recommendations>

²<https://journals.physiology.org/topic/advances-collections/statistics?seriesKey=&tagCode=&>

³<https://journals.physiology.org/topic/advances-collections/explorations-in-statistics?seriesKey=&tagCode=&>

⁴<https://journals.physiology.org/topic/advances-collections/general-statistics?seriesKey=&tagCode=&>

⁵<https://journals.physiology.org/topic/advances-collections/reporting-statistics?seriesKey=&tagCode=&>

⁶<https://www.tandfonline.com/toc/utas20/73/sup1?nav=tocList>

⁷<https://www.bmjjournals.org/journal/bmjspecialties>

⁸[https://www.bmjjournals.org/journal/bmjspecialties/statistics-notes](https://www.bmjjournals.org/journal/bmjspecialties)

⁹[https://www.bmjjournals.org/journal/bmjspecialties/statistics-and-research-methods](https://www.bmjjournals.org/journal/bmjspecialties)

¹⁰[https://www.bmjjournals.org/journal/bmjspecialties/about-bmj/resources-readers/publications/statistics-square-one](https://www.bmjjournals.org/journal/bmjspecialties)

- *Research methods & reporting*¹¹

61.1.5 Enhancing the QUality And Transparency Of health Research Network

- *Enhancing the Quality and Transparency of health research EQUATOR Network*¹²

61.1.6 Journal of the American Medical Association

- *JAMA Guide to Statistics and Methods - JAMA*¹³

61.1.7 Nature Publishing Group

- *Statistics for Biologists - Nature Publishing Group*¹⁴

61.1.8 Oxford Reference

- *A Dictionary of Statistics*¹⁵

61.1.9 Royal Statistical Society

- *Best Practices for Data Visualisation - Royal Statistical Society*¹⁶

61.1.10 Statistics in Medicine

- *Tutorials in Biostatistics Papers*¹⁷

61.1.11 BMC Trials

- *Design and analysis of n-of-1 trials*¹⁸

61.1.12 The Journal of Applied Statistics in the Pharmaceutical Industry

- *Tutorial Papers*¹⁹

¹¹<https://www.bmjjournals.org/research/research-methods-and-reporting>

¹²<https://www.equator-network.org>

¹³<https://jamanetwork.com/collections/44042/jama-guide-to-statistics-and-methods>

¹⁴<https://www.nature.com/collections/qghqm>

¹⁵<https://www.oxfordreference.com/display/10.1093/acref/9780199679188.001.0001/acref-9780199679188>

¹⁶<https://royal-statistical-society.github.io/datavisguide>

¹⁷<https://onlinelibrary.wiley.com/page/journal/10970258/homepage/tutorials.htm>

¹⁸<https://www.biomedcentral.com/collections/DANT>

¹⁹https://onlinelibrary.wiley.com/journal/15391612/homepage/tutorial_papers.htm

REFERÊNCIAS

RASCUNHO

Referências

1. Grami A. Discrete Probability. Em: *Discrete Mathematics: Essentials and Applications*. Elsevier; 2023:285–305. doi:10.1016/b978-0-12-820656-0.00016-2
2. Viti A, Terzi A, Bertolaccini L. A practical overview on probability distributions. *Journal of Thoracic Disease*. 2015;7(3). <https://jtd.amegroups.org/article/view/4086>.
3. Benford F. The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society*. 1938;78(4):551–572. <http://www.jstor.org/stable/984802>. Acessado novembro 24, 2024.
4. Tversky A, Kahneman D. Belief in the law of small numbers. *Psychological Bulletin*. 1971;76(2):105–110. doi:10.1037/h0031322
5. Bishop DVM, Thompson J, Parker AJ. Can we shift belief in the ‘Law of Small Numbers’? *Royal Society Open Science*. 2022;9(3). doi:10.1098/rsos.211028
6. Guy RK. The Strong Law of Small Numbers. *The American Mathematical Monthly*. 1988;95(8):697. doi:10.2307/2322249
7. Guy RK. The Second Strong Law of Small Numbers. *Mathematics Magazine*. 1990;63(1):3–20. doi:10.1080/0025570x.1990.11977475
8. Kwak SG, Kim JH. Central limit theorem: the cornerstone of modern statistics. *Korean Journal of Anesthesiology*. 2017;70(2):144. doi:10.4097/kjae.2017.70.2.144
9. Galton F. Regression Towards Mediocrity in Hereditary Stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*. 1886;15:246. doi:10.2307/2841583
10. Barnett AG. Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology*. 2004;34(1):215–220. doi:10.1093/ije/dyh299
11. Senn S. Francis Galton and Regression to the Mean. *Significance*. 2011;8(3):124–126. doi:10.1111/j.1740-9713.2011.00509.x
12. Recchia D. *regtomean: Regression Toward the Mean.*; 2022. <https://CRAN.R-project.org/package=regtomean>.
13. Altman DG, Bland JM. Statistics Notes: Units of analysis. *BMJ*. 1997;314(7098):1874–1874. doi:10.1136/bmj.314.7098.1874

14. Matthews JN, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. *BMJ*. 1990;300(6719):230–235. doi:10.1136/bmj.300.6719.230
15. Banerjee A, Chaudhury S. Statistics without tears: Populations and samples. *Industrial Psychiatry Journal*. 2010;19(1):60. doi:10.4103/0972-6748.77642
16. Martínez-Mesa J, González-Chica DA, Duquia RP, Bonamigo RR, Bastos JL. Sampling: how to select participants in my research study? *Anais Brasileiros de Dermatologia*. 2016;91(3):326–330. doi:10.1590/abd1806-4841.20165254
17. Bland JM, Altman DG. Statistics Notes: Bootstrap resampling methods. *BMJ*. 2015;350(jun02 13):h2622–h2622. doi:10.1136/bmj.h2622
18. Amatuzzi MLL, Barreto M do CC, Litvoc J, Leme LEG. Linguagem metodológica: parte 1. *Acta Ortopédica Brasileira*. 2006;14(1):53–56. doi:10.1590/s1413-78522006000100012
19. Amatuzzi MLL, Barreto M do CC, Litvoc J, Leme LEG. Linguagem metodológica: parte 2. *Acta Ortopédica Brasileira*. 2006;14(2):108–112. doi:10.1590/s1413-78522006000200012
20. Munafò MR, Nosek BA, Bishop DVM, et al. A manifesto for reproducible science. *Nature Human Behaviour*. 2017;1(1). doi:10.1038/s41562-016-0021
21. Wood M, Welch C. Are ‘Qualitative’ and ‘Quantitative’ Useful Terms for Describing Research? *Methodological Innovations Online*. 2010;5(1):56–71. doi:10.4256/mio.2010.0010
22. Lall D. Mixed-Methods Research. *Indian Journal of Continuing Nursing Education*. 2021;22(2):143–147. doi:10.4103/ijcn.ijcn_107_21
23. Schoonenboom J, Johnson RB. How to Construct a Mixed Methods Research Design. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*. 2017;69(S2):107–131. doi:10.1007/s11577-017-0454-1
24. Rubin M, Donkin C. Exploratory hypothesis tests can be more compelling than confirmatory hypothesis tests. *Philosophical Psychology*. 2022;37(8):2019–2047. doi:10.1080/09515089.2022.2113771
25. Späth C. From best practices to severe testing: A methodological response to Büsch and Loffing (2024). *German Journal of Exercise and Sport Research*. outubro 2025. doi:10.1007/s12662-025-01072-7
26. Resnik DB, Shamoo AE. Reproducibility and Research Integrity. *Accountability in Research*. 2016;24(2):116–123. doi:10.1080/08989621.2016.1257387
27. Hofner B, Schmid M, Edler L. Reproducible research in statistics: A review and guidelines for the *Biometrical Journal*. *Biometrical Journal*. 2015;58(2):416–427. doi:10.1002/bimj.201500156
28. Mair P. Thou Shalt Be Reproducible! A Technology Perspective. *Frontiers in Psychology*. 2016;7. doi:10.3389/fpsyg.2016.01079
29. Hinsen K. A data and code model for reproducible research and executable papers. *Procedia Computer Science*. 2011;4:579–588. doi:10.1016/j.procs.2011.04.061
30. Ihaka R, Gentleman R. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*. 1996;5(3):299. doi:10.2307/1390807

31. Nwanganga F, Chapple M. Introduction to R and RStudio. Em: Nwanganga F, Chapple M, orgs. *Practical Machine Learning in R*. John Wiley & Sons, Ltd; 2020:25–52. doi:10.1002/9781119591542.ch2
32. R Core Team. The Comprehensive R Archive Network. 2021. <https://cran.r-project.org>.
33. Allaire J, Xie Y, Dervieux C, et al. *rmarkdown: Dynamic Documents for R*; 2023. <https://CRAN.R-project.org/package=rmarkdown>.
34. Holmes DT, Mobini M, McCudden CR. Reproducible manuscript preparation with RMarkdown application to JMSACL and other Elsevier Journals. *Journal of Mass Spectrometry and Advances in the Clinical Lab*. 2021;22:8–16. doi:10.1016/j.jmsacl.2021.09.002
35. Love J, Selker R, Marsman M, et al. **JASP**: Graphical Statistical Software for Common Statistical Designs. *Journal of Statistical Software*. 2019;88(2). doi:10.18637/jss.v088.i02
36. ŞAHİN M, AYBEK E. Jamovi: An Easy to Use Statistical Software for the Social Scientists. *International Journal of Assessment Tools in Education*. 2020;6(4):670–692. doi:10.21449/ijate.661803
37. Selker R, Love J, Dropmann D. *jmv: The {jamovi} Analyses*; 2023. <https://CRAN.R-project.org/package=jmv>.
38. Love J. *jmvconnect: Connect to the {jamovi} Statistical Spreadsheet*; 2022. <https://CRAN.R-project.org/package=jmvconnect>.
39. Racine JS. RStudio: A Platform-Independent IDE for R and Sweave. *Journal of Applied Econometrics*. 2011;27(1):167–172. doi:10.1002/jae.1278
40. Aden-Buie G, Schloerke B, Allaire J, Rossell Hayes A. *learnr: Interactive Tutorials for R*; 2023. <https://CRAN.R-project.org/package=learnr>.
41. Schwab, Simon, Held, Leonhard. Statistical programming: Small mistakes, big impacts. *Wiley-Blackwell Publishing, Inc.* 2021. doi:10.5167/UZH-205154
42. Eglen SJ, Marwick B, Halchenko YO, et al. Toward standard practices for sharing computer code and programs in neuroscience. *Nature Neuroscience*. 2017;20(6):770–773. doi:10.1038/nn.4550
43. Xie Y. *formatR: Format R Code Automatically*; 2022. <https://CRAN.R-project.org/package=formatR>.
44. Müller K, Walthert L. *styler: Non-Invasive Pretty Printing of R Code*; 2023. <https://CRAN.R-project.org/package=styler>.
45. Hester J, Angly F, Hyde R, et al. *lintr: A {Linter} for R Code*; 2023. <https://CRAN.R-project.org/package=lintr>.
46. All R CRAN packages [Full List]. 2025. <https://r-packages.io/packages>. Acessado fevereiro 11, 2025.
47. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2023. <https://www.R-project.org/>.
48. Wickham H, Danenberg P, Csárdi G, Eugster M. *roxygen2: In-Line Documentation for R*; 2024. doi:10.32614/CRAN.package.roxygen2

49. Trisovic A, Lau MK, Pasquier T, Crosas M. A large-scale study on research code quality and execution. *Scientific Data*. 2022;9(1). doi:10.1038/s41597-022-01143-6
50. Gohel D, Ross N. *officedown: Enhanced {R Markdown} Format for {Word} and {PowerPoint}*; 2023. <https://CRAN.R-project.org/package=officedown>.
51. Xie Y. *bookdown: Authoring Books and Technical Documents with R Markdown*. Chapman; Hall/CRC; 2023. <https://bookdown.org/yihui/bookdown/>.
52. Ioannidis JPA. How to Make More Published Research True. *PLoS Medicine*. 2014;11(10):e1001747. doi:10.1371/journal.pmed.1001747
53. Krieger N, Perzynski A, Dalton J. *projects: A Project Infrastructure for Researchers*; 2021. <https://CRAN.R-project.org/package=projects>.
54. Schultze A, Tazare J. The role of programming code sharing in improving the transparency of medical research. *BMJ*. outubro 2023:p2402. doi:10.1136/bmj.p2402
55. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2023. <https://www.R-project.org/>.
56. Zhao Y, Xiao N, Anderson K, Zhang Y. Electronic common technical document submission with analysis using R. *Clinical Trials*. 2022;20(1):89–92. doi:10.1177/17407745221123244
57. Francisco Rodríguez-Sánchez, Connor P. Jackson, Shaurita D. Hutchins. *grateful: Facilitate citation of R packages*; 2023. <https://github.com/Pakillo/grateful>.
58. Shields M. Information Literacy, Statistical Literacy, Data Literacy. *IASSIST Quarterly*. 2005;28(2):6. doi:10.29173/iq790
59. Gal I. Adults' Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review*. 2002;70(1):1–25. doi:10.1111/j.1751-5823.2002.tb00336.x
60. Sharma S. Definitions and models of statistical literacy: a literature review. *Open Review of Educational Research*. 2017;4(1):118–133. doi:10.1080/23265507.2017.1354313
61. Hidayati NA, Waluya SB, Rochmad, Wardono. Statistics literacy: what, why and how? *Journal of Physics: Conference Series*. 2020;1613(1):012080. doi:10.1088/1742-6596/1613/1/012080
62. GOULD R. DATA LITERACY IS STATISTICAL LITERACY. *STATISTICS EDUCATION RESEARCH JOURNAL*. 2017;16(1):22–25. doi:10.52041/serj.v16i1.209
63. CALLINGHAM R, WATSON JM. THE DEVELOPMENT OF STATISTICAL LITERACY AT SCHOOL. *STATISTICS EDUCATION RESEARCH JOURNAL*. 2017;16(1):181–201. doi:10.52041/serj.v16i1.223
64. Koga S. Characteristics of statistical literacy skills from the perspective of critical thinking. *Teaching Statistics*. 2022;44(2):59–67. doi:10.1111/test.12302
65. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine*. 2019;170(1):51–58. doi:10.7326/m18-1376

66. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. agosto 2019;i4898. doi:10.1136/bmj.i4898
67. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. setembro 2017;j4008. doi:10.1136/bmj.j4008
68. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. outubro 2016;i4919. doi:10.1136/bmj.i4919
69. Whiting P, Savović J, Higgins JPT, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology*. 2016;69:225–234. doi:10.1016/j.jclinepi.2015.06.005
70. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Annals of Internal Medicine*. 2011;155(8):529–536. doi:10.7326/0003-4819-155-8-201110180-00009
71. Polin BA, Benisaac E. A longitudinal analysis of the hot hand and gambler's fallacy biases. *Judgment and Decision Making*. 2023;18. doi:10.1017/jdm.2023.23
72. Meng XL. Statistical paradeses and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election. *The Annals of Applied Statistics*. 2018;12(2). doi:10.1214/18-aos1161sf
73. Abelson RP. A variance explanation paradox: When a little is a lot. *Psychological Bulletin*. 1985;97(1):129–133. doi:10.1037/0033-2909.97.1.129
74. Berkson J. Limitations of the Application of Fourfold Table Analysis to Hospital Data. *Biometrics Bulletin*. 1946;2(3):47. doi:10.2307/3002000
75. Ellsberg D. Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics*. 1961;75(4):643. doi:10.2307/1884324
76. Freedman DA, Freedman DA. A Note on Screening Regression Equations. *The American Statistician*. 1983;37(2):152–155. doi:10.1080/00031305.1983.10482729
77. Freedman LS, Pee D. Return to a Note on Screening Regression Equations. *The American Statistician*. 1989;43(4):279. doi:10.2307/2685389
78. Hand DJ. On Comparing Two Treatments. *The American Statistician*. 1992;46(3):190–192. doi:10.1080/00031305.1992.10475881
79. LINDLEY DV. A STATISTICAL PARADOX. *Biometrika*. 1957;44(1-2):187–192. doi:10.1093/biomet/44.1-2.187
80. Lord FM. A paradox in the interpretation of group comparisons. *Psychological Bulletin*. 1967;68(5):304–305. doi:10.1037/h0025105
81. Lord FM. Statistical adjustments when comparing preexisting groups. *Psychological Bulletin*. 1969;72(5):336–337. doi:10.1037/h0028108

82. Simpson EH. The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1951;13(2):238–241. doi:10.1111/j.2517-6161.1951.tb00088.x
83. Blyth CR. On Simpson's Paradox and the Sure-Thing Principle. *Journal of the American Statistical Association*. 1972;67(338):364–366. doi:10.1080/01621459.1972.10482387
84. Pearl J. Comment: Understanding Simpson's Paradox. *The American Statistician*. 2014;68(1):8–13. doi:10.1080/00031305.2014.876829
85. Stein C. INADMISSIBILITY OF THE USUAL ESTIMATOR FOR THE MEAN OF A MULTIVARIATE NORMAL DISTRIBUTION. Em: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume I*. University of California Press; 1956:197–206. doi:10.1525/9780520313880-018
86. De S, Sen A. The generalised Gamow-Stern problem. *The Mathematical Gazette*. 1996;80(488):345–348. doi:10.2307/3619568
87. Feld SL. Why Your Friends Have More Friends Than You Do. *American Journal of Sociology*. 1991;96(6):1464–1477. doi:10.1086/229693
88. John LK, Loewenstein G, Prelec D. Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*. 2012;23(5):524–532. doi:10.1177/0956797611430953
89. Bausell RB. Too Much Medicine: Not Enough Health. Em: *The Problem with Science: The Reproducibility Crisis and What to do About It*. New York: Oxford University Press; 2021:56–C3.P203. doi:10.1093/oso/9780197536537.003.0004
90. Neoh MJY, Carollo A, Lee A, Esposito G. Fifty years of research on questionable research practises in science: quantitative analysis of co-citation patterns. *Royal Society Open Science*. 2023;10(10). doi:10.1098/rsos.230677
91. Kleinert S. COPE's retraction guidelines. *The Lancet*. 2009;374(9705):1876–1877. doi:10.1016/s0140-6736(09)62074-2
92. Kerr NL. HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*. 1998;2(3):196–217. doi:10.1207/s15327957pspr0203_4
93. Groot AD de. The meaning of “significance” for different types of research [translated and annotated by Eric-Jan Wagenmakers, Denny Borsboom, Josine Verhagen, Rogier Kievit, Marjan Bakker, Angelique Cramer, Dora Matzke, Don Mellenbergh, and Han L. J. van der Maas]. *Acta Psychologica*. 2014;148:188–194. doi:10.1016/j.actpsy.2014.02.001
94. Andrade C. HARKing, Cherry-Picking, P-Hacking, Fishing Expeditions, and Data Dredging and Mining as Questionable Research Practices. *The Journal of Clinical Psychiatry*. 2021;82(1). doi:10.4088/jcp.20f13804
95. Stefan AM, Schönbrodt FD. Big little lies: a compendium and simulation of p-hacking strategies. *Royal Society Open Science*. 2023;10(2). doi:10.1098/rsos.220346

96. Chuard PJC, Vrtílek M, Head ML, Jennions MD. Evidence that nonsignificant results are sometimes preferred: Reverse P-hacking or selective reporting? *PLOS Biology*. 2019;17(1):e3000127. doi:10.1371/journal.pbio.3000127
97. Sasaki K, Yamada Y. SPARKing: Sample-size planning after the results are known. *Frontiers in Human Neuroscience*. 2023;17. doi:10.3389/fnhum.2023.912338
98. Armitage P, McPherson CK, Rowe BC. Repeated Significance Tests on Accumulating Data. *Journal of the Royal Statistical Society Series A (General)*. 1969;132(2):235. doi:10.2307/2343787
99. Hutton JL, Williamson PR. Bias in Meta-Analysis Due to Outcome Variable Selection Within Studies. *Journal of the Royal Statistical Society Series C: Applied Statistics*. 2000;49(3):359–370. doi:10.1111/1467-9876.00197
100. Horton R. The rhetoric of research. *BMJ*. 1995;310(6985):985–987. doi:10.1136/bmj.310.6985.985
101. Chiu K, Grundy Q, Bero L. ‘Spin’ in published biomedical literature: A methodological systematic review. Boutron I, org. *PLOS Biology*. 2017;15(9):e2002173. doi:10.1371/journal.pbio.2002173
102. Picano E. Who is the author: genuine, honorary, ghost, gold, and fake authors? *Exploration of Cardiology*. 2024;2(3):88–96. doi:10.37349/ec.2024.00024
103. Montori VM, Smieja M, Guyatt GH. Publication Bias: A Brief Review for Clinicians. *Mayo Clinic Proceedings*. 2000;75(12):1284–1288. doi:10.4065/75.12.1284
104. Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. The preregistration revolution. *Proceedings of the National Academy of Sciences*. 2018;115(11):2600–2606. doi:10.1073/pnas.1708274114
105. P. Simmons J, D. Nelson L, Simonsohn U. Pre-registration: Why and How. *Journal of Consumer Psychology*. 2021;31(1):151–162. doi:10.1002/jcpy.1208
106. Hartgerink C, Aust F. *retractcheck: Retraction Scanner*; 2025. <https://github.com/chartgerink/retractcheck>.
107. Altman DG, Bland JM. Statistics notes Variables and parameters. *BMJ*. 1999;318(7199):1667–1667. doi:10.1136/bmj.318.7199.1667
108. Vetter TR. Fundamentals of Research Data and Variables. *Anesthesia & Analgesia*. 2017;125(4):1375–1380. doi:10.1213/ane.0000000000002370
109. Ali Z, Bhaskar Sb. Basic statistical tools in research and data analysis. *Indian Journal of Anaesthesia*. 2016;60(9):662. doi:10.4103/0019-5049.190623
110. Dettori JR, Norvell DC. The Anatomy of Data. *Global Spine Journal*. 2018;8(3):311–313. doi:10.1177/2192568217746998
111. Kaliyadan F, Kulkarni V. Types of variables, descriptive statistics, and sample size. *Indian Dermatology Online Journal*. 2019;10(1):82. doi:10.4103/idoj.idoj_468_18
112. Barkan H. Statistics in clinical research: Important considerations. *Annals of Cardiac Anaesthesia*. 2015;18(1):74. doi:10.4103/0971-9784.148325

113. Bland JM, Altman DG. Statistics Notes: Transforming data. *BMJ*. 1996;312(7033):770–770. doi:10.1136/bmj.312.7033.770
114. Fedorov V, Mannino F, Zhang R. Consequences of dichotomization. *Pharmaceutical Statistics*. 2009;8(1):50–61. doi:10.1002/pst.331
115. Osborne J. Improving your data transformations: Applying the Box-Cox transformation. *University of Massachusetts Amherst*. 2010. doi:10.7275/QBPC-GK17
116. Box GEP, Cox DR. An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1964;26(2):211–243. doi:10.1111/j.2517-6161.1964.tb00553.x
117. Venables WN, Ripley BD. *Modern Applied Statistics with S*. Springer; 2002. <https://www.stats.ox.ac.uk/pub/MASS4/>.
118. MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. *Psychological Methods*. 2002;7(1):19–40. doi:10.1037/1082-989x.7.1.19
119. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*. 2006;332(7549):1080.1. doi:10.1136/bmj.332.7549.1080
120. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine*. 2005;25(1):127–141. doi:10.1002/sim.2331
121. Collins GS, Ogundimu EO, Cook JA, Manach YL, Altman DG. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. *Statistics in Medicine*. 2016;35(23):4124–4135. doi:10.1002/sim.6986
122. Nelson SLP, Ramakrishnan V, Nietert PJ, Kamen DL, Ramos PS, Wolf BJ. An evaluation of common methods for dichotomization of continuous variables to discriminate disease status. *Communications in Statistics – Theory and Methods*. 2017;46(21):10823–10834. doi:10.1080/03610926.2016.1248783
123. Bennette C, Vickers A. Against quantiles: categorization of continuous variables in epidemiologic research, and its discontents. *BMC Medical Research Methodology*. 2012;12(1). doi:10.1186/1471-2288-12-21
124. Barnier J, Briatte F, Larmarange J. *questionr: Functions to Make Surveys Processing Easier*; 2023. <https://CRAN.R-project.org/package=questionr>.
125. Aguinis H, Pierce CA, Culpepper SA. Scale Coarseness as a Methodological Artifact. *Organizational Research Methods*. 2008;12(4):623–652. doi:10.1177/1094428108318065
126. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32–35. doi:10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3
127. Strobl C, Boulesteix AL, Augustin T. Unbiased split selection for classification trees based on the Gini Index. *Computational Statistics & Data Analysis*. 2007;52(1):483–501. doi:10.1016/j.csda.2006.12.030

128. Pearson K. X. *On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling.* *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science.* 1900;50(302):157–175. doi:10.1080/14786440009463897
129. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive Veterinary Medicine.* 2000;45(1-2):23–41. doi:10.1016/s0167-5877(00)00115-x
130. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological Bulletin.* 1971;76(5):378–382. doi:10.1037/h0031619
131. R Core Team. *R: A Language and Environment for Statistical Computing*; 2025. <https://www.R-project.org/>.
132. Olson K. What Are Data? *Qualitative Health Research.* 2021;31(9):1567–1569. doi:10.1177/10497323211015960
133. Smeden M van. A Very Short List of Common Pitfalls in Research Design, Data Analysis, and Reporting. *PRiMER.* 2022;6. doi:10.22454/PRiMER.2022.511416
134. Baillie M, Cessie S le, Schmidt CO, Lusa L, Huebner M. Ten simple rules for initial data analysis. *PLOS Computational Biology.* 2022;18(2):e1009819. doi:10.1371/journal.pcbi.1009819
135. Buttliere B. Adopting standard variable labels solves many of the problems with sharing and reusing data. *Methodological Innovations.* 2021;14(2):205979912110266. doi:10.1177/20597991211026616
136. Pebesma E, Mailund T, Hiebert J. Measurement Units in {R}. *The R Journal.* 2016;8. doi:10.32614/RJ-2016-061
137. Firke S. *janitor: Simple Tools for Examining and Cleaning Dirty Data.*; 2023. <https://CRAN.R-project.org/package=janitor>.
138. Harrell Jr FE. *Hmisc: Harrell Miscellaneous.*; 2023. <https://CRAN.R-project.org/package=Hmisc>.
139. Bryer J, Speerschneider K. *likert: Analysis and Visualization Likert Items.*; 2016. <https://CRAN.R-project.org/package=likert>.
140. Larmarange J. *ggstats: Extension to ggplot2 for Plotting Stats.*; 2025. doi:10.32614/CRAN.package.ggstats
141. Ferris TLJ. A new definition of measurement. *Measurement.* 2004;36(1):101–109. doi:10.1016/j.measurement.2004.03.001
142. Healy MJR, Goldstein H. Regression to the mean. *Annals of Human Biology.* 1978;5(3):277–280. doi:10.1080/03014467800002891
143. Altman DG, Bland JM. Measurement in Medicine: The Analysis of Method Comparison Studies. *The Statistician.* 1983;32(3):307. doi:10.2307/2987937

144. Menditto A, Patriarca M, Magnusson B. Understanding the meaning of accuracy, trueness and precision. *Accreditation and Quality Assurance*. 2006;12(1):45–47. doi:10.1007/s00769-006-0191-z
145. Streiner DL, Norman GR. “Precision” and “Accuracy”: Two Terms That Are Neither. *Journal of Clinical Epidemiology*. 2006;59(4):327–330. doi:10.1016/j.jclinepi.2005.09.005
146. Tierney N, Cook D. Expanding Tidy Data Principles to Facilitate Missing Data Exploration, Visualization and Assessment of Imputations. *Journal of Statistical Software*. 2023;105(7). doi:10.18637/jss.v105.i07
147. Hammill D. *DataEditR: An Interactive Editor for Viewing, Entering, Filtering & Editing Data.*; 2022. <https://CRAN.R-project.org/package=DataEditR>.
148. Broman KW, Woo KH. Data Organization in Spreadsheets. *The American Statistician*. 2018;72(1):2–10. doi:10.1080/00031305.2017.1375989
149. Juluru K, Eng J. Use of Spreadsheets for Research Data Collection and Preparation: *Academic Radiology*. 2015;22(12):1592–1599. doi:10.1016/j.acra.2015.08.024
150. Dowle M, Srinivasan A. *data.table: Extension of ‘data.frame’*; 2023. <https://CRAN.R-project.org/package=data.table>.
151. Altman DG, Bland JM. Missing data. *BMJ*. 2007;334(7590):424–424. doi:10.1136/bmj.38977.682025.2c
152. Heymans MW, Twisk JWR. Handling missing data in clinical research. *Journal of Clinical Epidemiology*. setembro 2022. doi:10.1016/j.jclinepi.2022.08.016
153. Carpenter JR, Smuk M. Missing data: A statistical framework for practice. *Biometrical Journal*. 2021;63(5):915–947. doi:10.1002/bimj.202000196
154. Yanagida T. *misty: Miscellaneous Functions.*; 2023. <https://CRAN.R-project.org/package=misty>.
155. Little RJA. A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*. 1988;83(404):1198–1202. doi:10.1080/01621459.1988.10478722
156. Tierney N, Cook D. Expanding Tidy Data Principles to Facilitate Missing Data Exploration, Visualization and Assessment of Imputations. *Journal of Statistical Software*. 2023;105(7):1–31. doi:10.18637/jss.v105.i07
157. Akl EA, Shawwa K, Kahale LA, et al. Reporting missing participant data in randomised trials: systematic survey of the methodological literature and a proposed guide. *BMJ Open*. 2015;5(12):e008431. doi:10.1136/bmjopen-2015-008431
158. Austin PC, Buuren S van. Logistic regression vs. predictive mean matching for imputing binary covariates. *Statistical Methods in Medical Research*. setembro 2023. doi:10.1177/09622802231198795
159. Buuren S van, Groothuis-Oudshoorn K. {mice}: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. 2011;45:1–67. doi:10.18637/jss.v045.i03
160. Rubin DB. Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations. *Journal of Business & Economic Statistics*. 1986;4(1):87. doi:10.2307/1391390

161. Little RJA. Missing-Data Adjustments in Large Surveys. *Journal of Business & Economic Statistics*. 1988;6(3):287–296. doi:10.1080/07350015.1988.10509663
162. Robitzsch A, Grund S. *miceadds: Some Additional Multiple Imputation Functions, Especially for {mice}*; 2023. <https://CRAN.R-project.org/package=miceadds>.
163. FitzJohn R. *ids: Generate Random Identifiers*; 2017. <https://CRAN.R-project.org/package=ids>.
164. Brown C. *hash: Full Featured Implementation of Hash Tables/Associative Arrays/Dictionaries*; 2023. <https://CRAN.R-project.org/package=hash>.
165. Hendricks P. *anonymizer: Anonymize Data Containing Personally Identifiable Information*; 2023. <https://github.com/paulhendricks/anonymizer>.
166. Lucas DE with contributions by A, Tuszyński J, Bengtsson H, et al. *digest: Create Compact Hash Digests of R Objects*; 2023. <https://CRAN.R-project.org/package=digest>.
167. Nowok B, Raab GM, Dibben C. *{synthpop}: Bespoke Creation of Synthetic Data in {R}*. *Journal of Statistical Software*. 2016;74. doi:10.18637/jss.v074.i11
168. S M. Frequency distribution. *Journal of Pharmacology and Pharmacotherapeutics*. 2011;2(1):54–56. doi:10.4103/0976-500x.77120
169. Sturges HA. The Choice of a Class Interval. *Journal of the American Statistical Association*. 1926;21(153):65–66. doi:10.1080/01621459.1926.10502161
170. SCOTT DW. On optimal and data-based histograms. *Biometrika*. 1979;66(3):605–610. doi:10.1093/biomet/66.3.605
171. Freedman D, Diaconis P. On the histogram as a density estimator:L 2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*. 1981;57(4):453–476. doi:10.1007/bf01025868
172. R Core Team. *R: A Language and Environment for Statistical Computing*; 2023. <https://www.R-project.org/>.
173. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer; 2016. <https://ggplot2.tidyverse.org>.
174. Kay M. *{ggdist}: Visualizations of Distributions and Uncertainty in the Grammar of Graphics*. *IEEE Transactions on Visualization and Computer Graphics*. 2024;30(1):414–424. doi:10.1109/TVCG.2023.3327195
175. Tang Y, Horikoshi M, Li W. *ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages*. Vol 8.; 2016. doi:10.32614/RJ-2016-060
176. Rochon J, Gondan M, Kieser M. To test or not to test: Preliminary assessment of normality when comparing two independent samples. *BMC Medical Research Methodology*. 2012;12(1). doi:10.1186/1471-2288-12-81
177. Greenhalgh T. How to read a paper: Statistics for the non-statistician. I: Different types of data need different statistical tests. *BMJ*. 1997;315(7104):364–366. doi:10.1136/bmj.315.7104.364

178. Schmider E, Ziegler M, Danay E, Beyer L, Bühner M. Is It Really Robust? *Methodology*. 2010;6(4):147–151. doi:10.1027/1614-2241/a000016
179. Kanji G. *100 Statistical Tests*. SAGE Publications Ltd; 2006. doi:10.4135/9781849208499
180. Curran-Everett D. Explorations in statistics: standard deviations and standard errors. *Advances in Physiology Education*. 2008;32(3):203–208. doi:10.1152/advan.90123.2008
181. Altman DG, Bland JM. Statistics Notes: Quartiles, quintiles, centiles, and other quantiles. *BMJ*. 1994;309(6960):996–996. doi:10.1136/bmj.309.6960.996
182. Krzywinski M, Altman N. Error bars. *Nature Methods*. 2013;10(10):921–922. doi:10.1038/nmeth.2659
183. Cumming G, Fidler F, Vaux DL. Error bars in experimental biology. *The Journal of Cell Biology*. 2007;177(1):7–11. doi:10.1083/jcb.200611141
184. S. M. Measures of central tendency: The mean. *Journal of Pharmacology and Pharmacotherapeutics*. 2011;2(2):140–142. doi:10.4103/0976-500x.81920
185. S. M. Measures of central tendency: Median and mode. *Journal of Pharmacology and Pharmacotherapeutics*. 2011;2(3):214–215. doi:10.4103/0976-500x.83300
186. Manikandan S. Measures of dispersion. *Journal of Pharmacology and Pharmacotherapeutics*. 2011;2(4):315–316. doi:10.4103/0976-500x.85931
187. Sahai H, Misra S. Definitions of Sample Variance: Some Teaching Problems to be Overcome. *The Statistician*. 1992;41(1):55. doi:10.2307/2348636
188. Leys C, Delacre M, Mora YL, Lakens D, Ley C. How to Classify, Detect, and Manage Univariate and Multivariate Outliers, With Emphasis on Pre-Registration. *International Review of Social Psychology*. 2019;32(1). doi:10.5334/irsp.289
189. Rousseeuw PJ, Hubert M. Robust statistics for outlier detection. *WIREs Data Mining and Knowledge Discovery*. 2011;1(1):73–79. doi:10.1002/widm.2
190. Daszykowski M, Kaczmarek K, Vander Heyden Y, Walczak B. Robust statistics in data analysis — A review. *Chemometrics and Intelligent Laboratory Systems*. 2007;85(2):203–219. doi:10.1016/j.chemolab.2006.06.016
191. Chatfield C. Exploratory data analysis. *European Journal of Operational Research*. 1986;23(1):5–13. doi:10.1016/0377-2217(86)90209-2
192. Ferketich S, Verran J. Technical Notes. *Western Journal of Nursing Research*. 1986;8(4):464–466. doi:10.1177/019394598600800409
193. Landis SC, Amara SG, Asadullah K, et al. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature*. 2012;490(7419):187–191. doi:10.1038/nature11556
194. Huebner M, Vach W, Cessie S le. A systematic approach to initial data analysis is good research practice. *The Journal of Thoracic and Cardiovascular Surgery*. 2016;151(1):25–27. doi:10.1016/j.jtcvs.2015.09.085

195. Zuur AF, Ieno EN, Elphick CS. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*. 2009;1(1):3–14. doi:10.1111/j.2041-210x.2009.00001.x
196. Krasser R. *explore: Simplifies Exploratory Data Analysis.*; 2023. <https://CRAN.R-project.org/package=explore>.
197. Petersen AH, Ekstrøm CT. {dataMaid}: Your Assistant for Documenting Supervised Data Quality Screening in {R}. *Journal of Statistical Software*. 2019;90. doi:10.18637/jss.v090.i06
198. Cui B. *DataExplorer: Automate Data Exploration and Treatment.*; 2020. <https://CRAN.R-project.org/package=DataExplorer>.
199. Dayanand Ubrangala, R K, Prasad Kondapalli R, Putatunda S. *SmartEDA: Summarize and Explore the Data.*; 2022. <https://CRAN.R-project.org/package=SmartEDA>.
200. Mock T. *gtExtras: Extending {gt} for Beautiful HTML Tables.*; 2023. <https://CRAN.R-project.org/package=gtExtras>.
201. Nijs V. *radiant: Business Analytics using R and Shiny.*; 2023. <https://CRAN.R-project.org/package=radiant>.
202. Behrens JT. Principles and procedures of exploratory data analysis. *Psychological Methods*. 1997;2(2):131–160. doi:10.1037/1082-989x.2.2.131
203. Prunello M, Mari G. *ggcleveland: Implementation of Plots from Cleveland's Visualizing Data Book.*; 2021. doi:10.32614/CRAN.package.ggcleveland
204. Gerring J. Mere Description. *British Journal of Political Science*. 2012;42(4):721–746. doi:10.1017/s0007123412000130
205. Cummings P, Rivara FP. Reporting Statistical Information in Medical Journal Articles. *Archives of Pediatrics & Adolescent Medicine*. 2003;157(4):321. doi:10.1001/archpedi.157.4.321
206. Cole TJ. Setting number of decimal places for reporting risk ratios: rule of four. *BMJ*. 2015;350(apr27 3):h1845–h1845. doi:10.1136/bmj.h1845
207. Cole TJ. Too many digits: the presentation of numerical data. *Archives of Disease in Childhood*. 2015;100(7):608–609. doi:10.1136/archdischild-2014-307149
208. Weissgerber TL, Winham SJ, Heinzen EP, et al. Reveal, Don't Conceal. *Circulation*. 2019;140(18):1506–1518. doi:10.1161/circulationaha.118.037777
209. Inskip H, Ntani G, Westbury L, et al. Getting started with tables. *Archives of Public Health*. 2017;75(1). doi:10.1186/s13690-017-0180-1
210. Kwak SG, Kang H, Kim JH, et al. The principles of presenting statistical results: Table. *Korean Journal of Anesthesiology*. 2021;74(2):115–119. doi:10.4097/kja.20582
211. Sjoberg DD, Whiting K, Curry M, Lavery JA, Larmarange J. Reproducible Summary Tables with the gtsummary Package. *The R Journal*. 2021;13:570–580. doi:10.32614/RJ-2021-053

212. Rich B. *table1: Tables of Descriptive Statistics in HTML.*; 2023. <https://CRAN.R-project.org/package=table1>.
213. Gohel D, Skintzos P. *flextable: Functions for Tabular Reporting.*; 2023. <https://CRAN.R-project.org/package=flextable>.
214. Thériault R. {rempscy}: Convenience functions for psychology. *Journal of Open Source Software*. 2023;8:5466. doi:10.21105/joss.05466
215. Barnett A. Automated detection of over- and under-dispersion in baseline tables in randomised controlled trials. *F1000Research*. 2023;11:783. doi:10.12688/f1000research.123002.2
216. Westreich D, Greenland S. The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients. *American Journal of Epidemiology*. 2013;177(4):292–298. doi:10.1093/aje/kws412
217. Chen H, Lu Y, Slye N. Testing for baseline differences in clinical trials. *International Journal of Clinical Trials*. 2020;7(2):150. doi:10.18203/2349-3259.ijct20201720
218. Pijls BG. The Table I Fallacy: P Values in Baseline Tables of Randomized Controlled Trials. *Journal of Bone and Joint Surgery*. 2022;104(16):e71. doi:10.2106/jbjs.21.01166
219. Hayes-Larson E, Kezios KL, Mooney SJ, Lovasi G. Who is in this study, anyway? Guidelines for a useful Table 1. *Journal of Clinical Epidemiology*. 2019;114:125–132. doi:10.1016/j.jclinepi.2019.06.011
220. Bandoli G, Palmsten K, Chambers CD, Jelliffe-Pawlowski LL, Baer RJ, Thompson CA. Revisiting the Table 2 fallacy: A motivating example examining preeclampsia and preterm birth. *Paediatric and Perinatal Epidemiology*. 2018;32(4):390–397. doi:10.1111/ppe.12474
221. Midway SR. Principles of Effective Data Visualization. *Patterns*. 2020;1(9):100141. doi:10.1016/j.patter.2020.100141
222. Park JH, Lee DK, Kang H, et al. The principles of presenting statistical results using figures. *Korean Journal of Anesthesiology*. 2022;75(2):139–150. doi:10.4097/kja.21508
223. Vandemeulebroecke M, Baillie M, Carr D, et al. How can we make better graphs? An initiative to increase the graphical expertise and productivity of quantitative scientists. *Pharmaceutical Statistics*. 2018;18(1):106–114. doi:10.1002/pst.1912
224. Sievert C. *Interactive Web-Based Data Visualization with R, plotly, and shiny*. Chapman; Hall/CRC; 2020. <https://plotly-r.com>.
225. Wei T, Simko V. *R package corrplot: Visualization of a Correlation Matrix.*; 2024. <https://github.com/taiyun/corrplot>.
226. Xiao N. *ggsci: Scientific Journal and Sci-Fi Themed Color Palettes for {ggplot2}.*; 2023. <https://CRAN.R-project.org/package=ggsci>.
227. Urbanek S, Johnson K. *tiff: Read and Write TIFF Images.*; 2022. <https://CRAN.R-project.org/package=tiff>.

228. Wiebels K, Moreau D. Dynamic Data Visualizations to Enhance Insight and Communication Across the Life Cycle of a Scientific Project. *Advances in Methods and Practices in Psychological Science*. 2023;6(3). doi:10.1177/25152459231160103
229. Pedersen TL, Robinson D. *gganimate: A Grammar of Animated Graphics.*; 2025. doi:10.32614/CRAN.package.gganimate
230. Mair P, Wilcox R. Robust Statistical Methods in R Using the WRS2 Package. *Behavior Research Methods*. 2020;52:464–488. doi:10.3758/s13428-019-01246-w
231. Leys C, Ley C, Klein O, Bernard P, Licata L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*. 2013;49(4):764–766. doi:10.1016/j.jesp.2013.03.013
232. Leys C, Klein O, Dominicy Y, Ley C. Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance. *Journal of Experimental Social Psychology*. 2018;74:150–156. doi:10.1016/j.jesp.2017.09.011
233. Tukey JW, McLaughlin DH. Less Vulnerable Confidence and Significance Procedures for Location Based on a Single Sample: Trimming/Winsorization 1. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*. 1963;25(3):331–352. <http://www.jstor.org/stable/25049278>. Acessado abril 11, 2025.
234. Komsta L. *outliers: Tests for Outliers.*; 2022. <https://CRAN.R-project.org/package=outliers>.
235. Loh PL. A Theoretical Review of Modern Robust Statistics. *Annual Review of Statistics and Its Application*. 2025;12(1):477–496. doi:10.1146/annurev-statistics-112723-034446
236. Mair P, Wilcox R, Indrajeet P. *A Collection of Robust Statistical Methods.*; 2025. <https://CRAN.R-project.org/package=WRS2>.
237. Lüdecke D. *ggeffects: Tidy Data Frames of Marginal Effects from Regression Models*. *Journal of Open Source Software*. 2018;3:772. doi:10.21105/joss.00772
238. Song YY, Lu Y. Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*. 2015;27(2):130–135. doi:10.11919/j.issn.1002-0829.215044
239. Hozo I, Guyatt G, Djulbegovic B. Decision curve analysis based on summary data. *Journal of Evaluation in Clinical Practice*. 2023;30(2):281–289. doi:10.1111/jep.13945
240. Vickers AJ, Calster B van, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagnostic and Prognostic Research*. 2019;3(1). doi:10.1186/s41512-019-0064-7
241. AALEN OO, FRIGESSI A. What can Statistics Contribute to a Causal Understanding? *Scandinavian Journal of Statistics*. 2007;34(1):155–168. doi:10.1111/j.1467-9469.2006.00549.x
242. Matute H, Blanco F, Yarritu I, Díaz-Lago M, Vadillo MA, Barbería I. Illusions of causality: how they bias our everyday thinking and how they could be reduced. *Frontiers in Psychology*. 2015;6. doi:10.3389/fpsyg.2015.00888

243. Vickers AJ, Assel M, Dunn RL, et al. Guidelines for Reporting Observational Research in Urology: The Importance of Clear Reference to Causality. *European Urology*. 2023;84(2):147–151. doi:10.1016/j.eururo.2023.04.027
244. Hill AB. The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*. 1965;58(5):295–300. doi:10.1177/003591576505800503
245. Rothman KJ, Greenland S. Hill's Criteria for Causality. *Encyclopedia of Biostatistics*. fevereiro 2005. doi:10.1002/0470011815.b2a03072
246. Shimonovich M, Pearce A, Thomson H, Keyes K, Katikireddi SV. Assessing causality in epidemiology: revisiting Bradford Hill to incorporate developments in causal thinking. *European Journal of Epidemiology*. 2020;36(9):873–887. doi:10.1007/s10654-020-00703-7
247. Textor J, Zander B van der, Gilthorpe MS, Liskiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package {dagitty}. *International Journal of Epidemiology*. 2016;45:1887–1894. doi:10.1093/ije/dyw341
248. Barrett M. *ggdag: Analyze and Create Elegant Directed Acyclic Graphs.*; 2024. <https://CRAN.R-project.org/package=ggdag>.
249. Lüdecke D, Ben-Shachar MS, Patil I, Waggoner P, Makowski D. {performance}: An {R} Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*. 2021;6:3139. doi:10.21105/joss.03139
250. Silge J, Robinson D. tidytext: Text Mining and Analysis Using Tidy Data Principles in R. *The Journal of Open Source Software*. 2016;1. doi:10.21105/joss.00037
251. Greenland S. Modeling and variable selection in epidemiologic analysis. *American Journal of Public Health*. 1989;79(3):340–349. doi:10.2105/ajph.79.3.340
252. Breznau N, Rinke EM, Wuttke A, et al. Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proceedings of the National Academy of Sciences*. 2022;(44):e2203150119. doi:10.1073/pnas.2203150119
253. Dwivedi AK, Shukla R. Evidence-based statistical analysis and methods in biomedical research (SAMBR) checklists according to design features. *CANCER REPORTS*. 2019;3(4). doi:10.1002/cnr2.1211
254. Dwivedi AK. How to Write Statistical Analysis Section in Medical Research. *Journal of Investigative Medicine*. 2022;70(8):1759–1770. doi:10.1136/jim-2022-002479
255. Kim N, Fischer AH, Dyring-Andersen B, Rosner B, Okoye GA. Research Techniques Made Simple: Choosing Appropriate Statistical Methods for Clinical Research. *Journal of Investigative Dermatology*. 2017;137(10):e173–e178. doi:10.1016/j.jid.2017.08.007
256. Marusteri M, Bacarea V. Comparing groups for statistical differences: how to choose the right statistical test? *Biochimia Medica*. 2010;15–32. doi:10.11613/bm.2010.004
257. Mishra P, Pandey C, Singh U, Keshri A, Sabaretnam M. Selection of appropriate statistical methods for data analysis. *Annals of Cardiac Anaesthesia*. 2019;22(3):297. doi:10.4103/aca.aca_248_18

258. Ray A, Najmi A, Sadasivam B. How to choose and interpret a statistical test? An update for budding researchers. *Journal of Family Medicine and Primary Care*. 2021;10(8):2763. doi:10.4103/jfmpc.jfmpc_433_21
259. Nayak B, Hazra A. How to choose the right statistical test? *Indian Journal of Ophthalmology*. 2011;59(2):85. doi:10.4103/0301-4738.77005
260. Shankar S, Singh R. Demystifying statistics: How to choose a statistical test? *Indian Journal of Rheumatology*. 2014;9(2):77–81. doi:10.1016/j.injr.2014.04.002
261. Curran-Everett D. Explorations in statistics: hypothesis tests and *P* values. *Advances in Physiology Education*. 2009;33(2):81–86. doi:10.1152/advan.90218.2008
262. Goodman SN. Toward Evidence-Based Medical Statistics. 1: The *P* Value Fallacy. *Annals of Internal Medicine*. 1999;130(12):995. doi:10.7326/0003-4819-130-12-199906150-00008
263. McCaskey K, Rainey C. Substantive Importance and the Veil of Statistical Significance. *Statistics, Politics and Policy*. 2015;6(1-2). doi:10.1515/spp-2015-0001
264. Uygun Tunç D, Tunç MN, Lakens D. The epistemic and pragmatic function of dichotomous claims based on statistical hypothesis tests. *Theory & Psychology*. 2023;33(3):403–423. doi:10.1177/09593543231160112
265. Vandenbroucke JP, Pearce N. From ideas to studies: how to get ideas and sharpen them into research questions. *Clinical Epidemiology*. 2018;Volume 10:253–264. doi:10.2147/clep.s142940
266. Lakens D, Scheel AM, Isager PM. Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*. 2018;1(2):259–269. doi:10.1177/2515245918770963
267. Sullivan GM, Feinn R. Using Effect Size—or Why the *P* Value Is Not Enough. *Journal of Graduate Medical Education*. 2012;4(3):279–282. doi:10.4300/jgme-d-12-00156.1
268. Neyman J. Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. *Philosophical Transactions of the Royal Society of London Series A, Mathematical and Physical Sciences*. 1937;236(767):333–380. doi:10.1098/rsta.1937.0005
269. Goodman SN. Aligning statistical and scientific reasoning. *Science*. 2016;352(6290):1180–1181. doi:10.1126/science.aaf5406
270. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*. 2016;31(4):337–350. doi:10.1007/s10654-016-0149-3
271. Cumming G, Finch S. Inference by Eye: Confidence Intervals and How to Read Pictures of Data. *American Psychologist*. 2005;60(2):170–180. doi:10.1037/0003-066x.60.2.170
272. Greenhalgh T. How to read a paper: Statistics for the non-statistician. II: Significant relations and their pitfalls. *BMJ*. 1997;315(7105):422–425. doi:10.1136/bmj.315.7105.422
273. Weintraub PG. The Importance of Publishing Negative Results. *Journal of Insect Science*. 2016;16(1):109. doi:10.1093/jisesa/iew092

274. Altman DG, Bland JM. Statistics notes: Absence of evidence is not evidence of absence. *BMJ*. 1995;311(7003):485–485. doi:10.1136/bmj.311.7003.485
275. Gelman A, Carlin J. Beyond Power Calculations. *Perspectives on Psychological Science*. 2014;9(6):641–651. doi:10.1177/1745691614551642
276. Lu J, Qiu Y, Deng A. A note on Type S/M errors in hypothesis testing. *British Journal of Mathematical and Statistical Psychology*. 2018;72(1):1–17. doi:10.1111/bmsp.12132
277. Kim HY. Statistical notes for clinical researchers: effect size. *Restorative Dentistry & Endodontics*. 2015;40(4):328. doi:10.5395/rde.2015.40.4.328
278. Aragon TJ. *epitools: Epidemiology Tools*; 2020. doi:10.32614/CRAN.package.epitools
279. Ben-Shachar MS, Lüdecke D, Makowski D. {e}ffectsizer: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software*. 2020;5:2815. doi:10.21105/joss.02815
280. Champely S. *pwr: Basic Functions for Power Analysis*; 2020. <https://CRAN.R-project.org/package=pwr>.
281. GREENLAND S, SCHLESSELMAN JJ, CRIQUI MH. THE FALLACY OF EMPLOYING STANDARDIZED REGRESSION COEFFICIENTS AND CORRELATIONS AS MEASURES OF EFFECT. *American Journal of Epidemiology*. 1986;123(2):203–208. doi:10.1093/oxfordjournals.aje.a114229
282. Greenland S, Maclure M, Schlesselman JJ, Poole C, Morgenstern H. Standardized Regression Coefficients. *Epidemiology*. 1991;2(5):387–392. doi:10.1097/00001648-199109000-00015
283. LATTER OH. THE EGG OF CUCULUS CANORUS: AN ENQUIRY INTO THE DIMENSIONS OF THE CUCKOO'S EGO AND THE RELATION OF THE VARIATIONS TO THE SIZE OF THE EGGS OF THE FOSTER-PARENT, WITH NOTES ON COLORATION, &c. *Biometrika*. 1902;1(2):164–176. doi:10.1093/biomet/1.2.164
284. Aylmer Fisher R. The arrangement of field experiments. *Ministry of Agriculture and Fisheries*. 1926. doi:10.23637/ROTHAMSTED.8V61Q
285. Lakens D, Caldwell A. Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. *Advances in Methods and Practices in Psychological Science*. 2021;4:251524592095150. doi:10.1177/2515245920951503
286. Wasserstein RL, Lazar NA. The ASA Statement on *p*-Values: Context, Process, and Purpose. *The American Statistician*. 2016;70(2):129–133. doi:10.1080/00031305.2016.1154108
287. Altman N, Krzywinski M. P values and the search for significance. *Nature Methods*. 2017;14(1):3–4. doi:10.1038/nmeth.4120
288. Heinze G, Dunkler D. Five myths about variable selection. *Transplant International*. 2016;30(1):6–10. doi:10.1111/tri.12895
289. Blume JD, D'Agostino McGowan L, Dupont WD, Greevy RA. Second-generation *p*-values: Improved rigor, reproducibility, & transparency in statistical analyses. Smalheiser NR, org. *PLOS ONE*. 2018;13(3):e0188299. doi:10.1371/journal.pone.0188299

290. Lakens D, Delacre M. Equivalence Testing and the Second Generation P-Value. *Meta-Psychology*. 2020;4. doi:10.15626/mp.2018.933
291. Meyer F, Perrier V. *esquisse: Explore and Visualize Your Data Interactively.*; 2022. <https://CRAN.R-project.org/package=esquisse>.
292. Diedenhofen B, Musch J. cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLOS ONE*. 2015;10:e0121945. doi:10.1371/journal.pone.0121945
293. McHugh ML. The Chi-square test of independence. *Biochimia Medica*. 2013;143–149. doi:10.11613/bm.2013.018
294. Kim HY. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restorative Dentistry & Endodontics*. 2017;42(2):152. doi:10.5395/rde.2017.42.2.152
295. Khamis H. Measures of Association: How to Choose? *Journal of Diagnostic Medical Sonography*. 2008;24(3):155–162. doi:10.1177/8756479308317006
296. Allison JS, Santana L, (Jaco) Visagie IJH. A primer on simple measures of association taught at undergraduate level. *Teaching Statistics*. 2022;44(3):96–103. doi:10.1111/test.12307
297. Dahlke JA, Wiernik BM. {psychmeta}: An R Package for Psychometric Meta-Analysis. *Applied Psychological Measurement*. 2018;43(3):415–416. doi:10.1177/0146621618795933
298. Anscombe FJ. Graphs in Statistical Analysis. *The American Statistician*. 1973;27(1):17–21. doi:10.1080/00031305.1973.10478966
299. Northrop PJ. *anscombiser: Create Datasets with Identical Summary Statistics.*; 2022. <https://CRAN.R-project.org/package=anscombiser>.
300. Makowski D, Wiernik BM, Patil I, Lüdecke D, Ben-Shachar MS. {{correlation}}: Methods for Correlation Analysis.; 2022. <https://CRAN.R-project.org/package=correlation>.
301. Lüdecke D, Ben-Shachar MS, Patil I, et al. *easystats: Framework for Easy Statistical Modeling, Visualization, and Reporting.*; 2022. <https://easystats.github.io/easystats/>.
302. Kim JH. Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*. 2019;72(6):558–569. doi:10.4097/kja.19087
303. Schloerke B, Cook D, Larmerange J, et al. *GGally: Extension to ggplot2*; 2024. doi:10.32614/CRAN.package.GGally
304. Arel-Bundock V. {modelsummary}: Data and Model Summaries in {R}. *Journal of Statistical Software*. 2022;103. doi:10.18637/jss.v103.i01
305. Hidalgo B, Goodman M. Multivariate or Multivariable Regression? *American Journal of Public Health*. 2013;103(1):39–40. doi:10.2105/ajph.2012.300897
306. Suits DB. Use of Dummy Variables in Regression Equations. *Journal of the American Statistical Association*. 1957;52(280):548–551. doi:10.1080/01621459.1957.10501412

307. Healy MJ. Statistics from the inside. 16. Multiple regression (2). *Archives of Disease in Childhood*. 1995;73(3):270–274. doi:10.1136/adc.73.3.270
308. Kaplan J. *fastDummies: Fast Creation of Dummy (Binary) Columns and Rows from Categorical Variables.*; 2023. <https://CRAN.R-project.org/package=fastDummies>.
309. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *Journal of Clinical Epidemiology*. 1996;49(8):907–916. doi:10.1016/0895-4356(96)00025-x
310. Fox J, Weisberg S. *An {R} Companion to Applied Regression*. Sage Publications, Inc.; 2019. <https://www.john-fox.ca/Companion/>.
311. DALES LG, URY HK. An Improper Use of Statistical Significance Testing in Studying Covariates. *International Journal of Epidemiology*. 1978;7(4):373–376. doi:10.1093/ije/7.4.373
312. Lindsey C, Sheather S. Variable Selection in Linear Regression. *The Stata Journal: Promoting communications on statistics and Stata*. 2011;10(4):650–669. doi:10.1177/1536867x1001000407
313. Miller TL based on F code by A. *leaps: Regression Subset Selection.*; 2024. doi:10.32614/CRAN.package.leaps
314. Hebbali A. *olsrr: Tools for Building OLS Regression Models.*; 2024. doi:10.32614/CRAN.package.olrr
315. Box GEP. Science and Statistics. *Journal of the American Statistical Association*. 1976;71(356):791–799. doi:10.1080/01621459.1976.10480949
316. Anderson D, Heiss A, Sumners J. *equatiomatic: Transform Models into {LaTeX} Equations.*; 2024. <https://CRAN.R-project.org/package=equatiomatic>.
317. Ploeg T van der, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Medical Research Methodology*. 2014;14(1). doi:10.1186/1471-2288-14-137
318. HÄGGSTRÖM O. Problem Solving is Often a Matter of Cooking Up an Appropriate Markov Chain*. *Scandinavian Journal of Statistics*. 2007;34(4):768–780. doi:10.1111/j.1467-9469.2007.00561.x
319. Spedicato GA. Discrete Time Markov Chains with R. *The R Journal*. 2017;9(2):84–104. doi:10.32614/RJ-2017-036
320. Bours MJL. Using mediators to understand effect modification and interaction. *Journal of Clinical Epidemiology*. setembro 2023. doi:10.1016/j.jclinepi.2023.09.005
321. Altman DG, Matthews JNS. Statistics Notes: Interaction 1: heterogeneity of effects. *BMJ*. 1996;313(7055):486–486. doi:10.1136/bmj.313.7055.486
322. Pinheiro J, Bates D, R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models.*; 2023. <https://CRAN.R-project.org/package=nlme>.
323. Sabanes Bove D, Dedic J, Kelkhoff D, et al. *mmrm: Mixed Models for Repeated Measures.*; 2022. <https://CRAN.R-project.org/package=mmrm>.

324. Lenth RV. *emmeans: Estimated Marginal Means, aka Least-Squares Means.*; 2023. <https://CRAN.R-project.org/package=emmeans>
325. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*. 1986;51(6):1173–1182. doi:10.1037/0022-3514.51.6.1173
326. Henderson T. *correctR: Corrected Test Statistics for Comparing Machine Learning Models on Correlated Samples.*; 2025. <https://CRAN.R-project.org/package=correctR>
327. Diego J. Pedregal. *UComp: Automatic Univariate Time Series Modelling of many Kinds.*; 2025. doi:10.32614/CRAN.package.UComp
328. Pebesma E, Bivand R. *{Spatial Data Science: With applications in R}*. CRC Press, Taylor & Francis Group; 2023. doi:10.1201/9780429459016
329. Cheng J, Schloerke B, Karambelkar B, Xie Y, Aden-Buie G. *leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library.*; 2025. doi:10.32614/CRAN.package.leaflet
330. Therneau TM. *A Package for Survival Analysis in R.*; 2024. <https://CRAN.R-project.org/package=survival>
331. Allaire J, Tang Y. *tensorflow: R Interface to 'TensorFlow'*; 2025. doi:10.32614/CRAN.package.tensorflow
332. Falbel D, Luraschi J. *torch: Tensors and Neural Networks with GPU Acceleration.*; 2025. doi:10.32614/CRAN.package.torch
333. Hand DJ. Classifier Technology and the Illusion of Progress. *Statistical Science*. 2006;21(1). doi:10.1214/088342306000000060
334. Andaur Navarro CL, Damen JAA, Smeden M van, et al. Systematic review identifies the design and methodological conduct of studies on machine learning-based prediction models. *Journal of Clinical Epidemiology*. 2023;154:8–22. doi:10.1016/j.jclinepi.2022.11.015
335. Carriero A, Luijken K, Hond A de, Moons KGM, Calster B van, Smeden M van. The Harms of Class Imbalance Corrections for Machine Learning Based Prediction Models: A Simulation Study. *Statistics in Medicine*. 2025;44(3-4). doi:10.1002/sim.10320
336. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*. 1943;5(4):115–133. doi:10.1007/bf02478259
337. Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*. 1958;65(6):386–408. doi:10.1037/h0042519
338. Rosenblatt F. Perceptron Simulation Experiments. *Proceedings of the IRE*. 1960;48(3):301–309. doi:10.1109/jrproc.1960.287598
339. Fritsch S, Guenther F, Wright MN. *neuralnet: Training of Neural Networks.*; 2019. doi:10.32614/CRAN.package.neuralnet

340. Heckman MG, Davis JM, Crowson CS. Post Hoc Power Calculations: An Inappropriate Method for Interpreting the Findings of a Research Study. *The Journal of Rheumatology*. 2022;49(8):867–870. doi:10.3899/jrheum.211115
341. Iddi S, Donohue MC. Power and Sample Size for Longitudinal Models in R – The longpower Package and Shiny App. *The R Journal*. 2022;14:264–282.
342. Baranger DAA, Finsaas MC, Goldstein BL, Vize CE, Lynam DR, Olino TM. Tutorial: Power Analyses for Interaction Effects in Cross-Sectional Regressions. *Advances in Methods and Practices in Psychological Science*. 2023;6(3):25152459231187531. doi:10.1177/25152459231187531
343. Rodríguez del Águila M, González-Ramírez A. Sample size calculation. *Allergologia et Immunopathologia*. 2014;42(5):485–492. doi:10.1016/j.aller.2013.03.008
344. Bacchetti P. Ethics and Sample Size. *American Journal of Epidemiology*. 2005;161(2):105–110. doi:10.1093/aje/kwi014
345. Ahmed SK. Sample size for saturation in qualitative research: Debates, definitions, and strategies. *Journal of Medicine, Surgery, and Public Health*. 2025;5:100171. doi:10.1016/j.gmedi.2024.100171
346. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine*. 2022;292:114523. doi:10.1016/j.socscimed.2021.114523
347. Wutich A, Beresford M, Bernard HR. Sample Sizes for 10 Types of Qualitative Data Analysis: An Integrative Review, Empirical Guidance, and Next Steps. *International Journal of Qualitative Methods*. 2024;23. doi:10.1177/16094069241296206
348. Vasileiou K, Barnett J, Thorpe S, Young T. Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology*. 2018;18(1). doi:10.1186/s12874-018-0594-7
349. Ying X, Robinson KA, Ehrhardt S. Re-evaluating the role of pilot trials in informing effect and sample size estimates for full-scale trials: a meta-epidemiological study. *BMJ Evidence-Based Medicine*. 2023;28(6):383–391. doi:10.1136/bmjebm-2023-112358
350. Andrade C. Sample Size and its Importance in Research. *Indian Journal of Psychological Medicine*. 2020;42(1):102–103. doi:10.4103/ijpsym.ijpsym_504_19
351. Gamble C, Krishan A, Stocken D, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. *JAMA*. 2017;318(23):2337. doi:10.1001/jama.2017.18556
352. Bland JM, Altman DG. Statistics notes: Matching. *BMJ*. 1994;309(6962):1128–1128. doi:10.1136/bmj.309.6962.1128
353. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*. 2009;26(2):91–108. doi:10.1111/j.1471-1842.2009.00848.x
354. Sut N. Study Designs in Medicine. *Balkan Medical Journal*. 2015;31(4):273–277. doi:10.5152/balkanmedj.2014.1408

355. Souza AC de, Alexandre NMC, Guirardello E de B, Souza AC de, Alexandre NMC, Guirardello E de B. Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. *Epidemiologia e Serviços de Saúde*. 2017;26(3):649–659. doi:10.5123/s1679-49742017000300022
356. Reeves BC, Wells GA, Waddington H. Quasi-experimental study designs series—paper 5: a checklist for classifying studies evaluating the effects on health interventions—a taxonomy without labels. *Journal of Clinical Epidemiology*. 2017;89:30–42. doi:10.1016/j.jclinepi.2017.02.016
357. Echevarría-Guanilo ME, Gonçalves N, Romanoski PJ. PSYCHOMETRIC PROPERTIES OF MEASUREMENT INSTRUMENTS: CONCEPTUAL BASIS AND EVALUATION METHODS – PART II. *Texto & Contexto – Enfermagem*. 2019;28. doi:10.1590/1980-265x-tce-2017-0311
358. Chassé M, Fergusson DA. Diagnostic Accuracy Studies. *Seminars in Nuclear Medicine*. 2019;49(2):87–93. doi:10.1053/j.semnuclmed.2018.11.005
359. Chidambaram AG, Josephson M. Clinical research study designs: The essentials. *PEDIATRIC INVESTIGATION*. 2019;3(4):245–252. doi:10.1002/ped4.12166
360. Erdemir A, Mulugeta L, Ku JP, et al. Credible practice of modeling and simulation in healthcare: ten rules from a multidisciplinary perspective. *Journal of Translational Medicine*. 2020;18(1). doi:10.1186/s12967-020-02540-4
361. Yang B, Olsen M, Vali Y, et al. Study designs for comparative diagnostic test accuracy: A methodological review and classification scheme. *Journal of Clinical Epidemiology*. 2021;138:128–138. doi:10.1016/j.jclinepi.2021.04.013
362. Chipman H, Bingham D. Let's practice what we preach: Planning and interpreting simulation studies with design and analysis of experiments. *Canadian Journal of Statistics*. 2022;50(4):1228–1249. doi:10.1002/cjs.11719
363. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*. 2021;133:285–296. doi:10.1016/j.jbusres.2021.04.070
364. Lim WM, Kumar S. Guidelines for interpreting the results of bibliometric analysis: A sensemaking approach. *Global Business and Organizational Excellence*. agosto 2023. doi:10.1002/joe.22229
365. Trisovic A, Lau MK, Pasquier T, Crosas M. A large-scale study on research code quality and execution. *Scientific Data*. 2022;9(1). doi:10.1038/s41597-022-01143-6
366. Metropolis N, Ulam S. The Monte Carlo Method. *Journal of the American Statistical Association*. 1949;44(247):335–341. doi:10.1080/01621459.1949.10483310
367. Goldfeld K, Wujciak-Jens J. simstudy: Illuminating research methods through data generation. *Journal of Open Source Software*. 2020;5:2763. doi:10.21105/joss.02763
368. DeBruine L. *faux: Simulation for Factorial Designs*; 2023. doi:10.5281/zenodo.2669586
369. Monks T, Currie CSM, Onggo BS, Robinson S, Kunc M, Taylor SJE. Strengthening the reporting of empirical simulation studies: Introducing the STRESS guidelines. *Journal of Simulation*. 2018;13(1):55–67. doi:10.1080/17477778.2018.1442155

370. Elm E von, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandebroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *Annals of Internal Medicine*. 2007;147(8):573. doi:10.7326/0003-4819-147-8-200710160-00010
371. Rosseel Y. {lavaan}: An {R} Package for Structural Equation Modeling. *Journal of Statistical Software*. 2012;48. doi:10.18637/jss.v048.i02
372. Findley MG, Kikuta K, Denly M. External Validity. *Annual Review of Political Science*. 2021;24(1):365–393. doi:10.1146/annurev-polisci-041719-102556
373. Scott WA. Reliability of Content Analysis: The Case of Nominal Scale Coding. *Public Opinion Quarterly*. 1955;19(3):321. doi:10.1086/266577
374. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*. 1960;20(1):37–46. doi:10.1177/001316446002000104
375. Gamer M, Lemon J, Ian Fellows Puspendra Singh. *irr: Various Coefficients of Interrater Reliability and Agreement.*; 2019. doi:10.32614/CRAN.package.irr
376. Mathews I, Pearson K. I. Mathematical contributions to the theory of evolution. —VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London Series A, Containing Papers of a Mathematical or Physical Character*. 1901;195(262-273):1–47. doi:10.1098/rsta.1900.0022
377. Banerjee M, Capozzoli M, McSweeney L, Sinha D. Beyond kappa: A review of interrater agreement measures. *Canadian Journal of Statistics*. 1999;27(1):3–23. doi:10.2307/3315487
378. William Revelle. *psych: Procedures for Psychological, Psychometric, and Personality Research.*; 2023. <https://CRAN.R-project.org/package=psych>.
379. Lehnert B. *BlandAltmanLeh: Plots (Slightly Extended) Bland-Altman Plots.*; 2015. <https://CRAN.R-project.org/package=BlandAltmanLeh>.
380. Contributors semTools. *semTools: Useful tools for structural equation modeling.*; 2016. <https://CRAN.R-project.org/package=semTools>.
381. Quartagno M, Carpenter J. *{jomo}: A package for Multilevel Joint Modelling Multiple Imputation.*; 2023. <https://CRAN.R-project.org/package=jomo>.
382. Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Quality of Life Research*. 2021;30(8):2197–2218. doi:10.1007/s11136-021-02822-4
383. Streiner DL, Kottner J. Recommendations for reporting the results of studies of instrument and scale development and testing. *Journal of Advanced Nursing*. 2014;70(9):1970–1979. doi:10.1111/jan.12402
384. Kottner J, Audigé L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Journal of Clinical Epidemiology*. 2011;64(1):96–106. doi:10.1016/j.jclinepi.2010.03.002

385. Steckelberg A, Balgenorth A, Berger J, Mühlhauser I. Explaining computation of predictive values: 2×2 table versus frequency tree. A randomized controlled trial [ISRCTN74278823]. *BMC Medical Education.* 2004;4(1). doi:10.1186/1472-6920-4-13
386. Greenhalgh T. How to read a paper: Papers that report diagnostic or screening tests. *BMJ.* 1997;315(7107):540–543. doi:10.1136/bmj.315.7107.540
387. Neth H, Gaisbauer F, Gradwohl N, Gaissmaier W. *riskyR: Rendering Risk Literacy more Transparent.*; 2022. <https://CRAN.R-project.org/package=riskyR>.
388. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical Epidemiology.* 2003;56(11):1129–1135. doi:10.1016/s0895-4356(03)00177-x
389. Kuhn, Max. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software.* 2008;28(5):1–26. doi:10.18637/jss.v028.i05
390. Xu J, Zhang Y, Miao D. Three-way confusion matrix for classification: A measure driven view. *Information Sciences.* 2020;507:772–794. doi:10.1016/j.ins.2019.06.064
391. He Z, Zhang Q, Song M, Tan X, Wang W. Four overlooked errors in ROC analysis: how to prevent and avoid. *BMJ Evidence-Based Medicine.* 2024;30(3):208–211. doi:10.1136/bmjebm-2024-113078
392. Park SH, Goo JM, Jo CH. Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists. *Korean Journal of Radiology.* 2004;5(1):11. doi:10.3348/kjr.2004.5.1.11
393. Park SH, Goo JM, Jo CH. UniODA vs ROC Analysis: Computing the “optimal” cut-point. *Optimal Data Analysis.* 2014;3(14):117–120. <https://odajournal.com/wp-content/uploads/2019/01/v3a29.pdf>.
394. Hond AAH de, Steyerberg EW, Calster B van. Interpreting area under the receiver operating characteristic curve. *The Lancet Digital Health.* 2022;4(12):e853–e855. doi:10.1016/s2589-7500(22)00188-1
395. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011;12:77. doi:10.1186/1471-2105-12-77
396. Ferreira ADS, Meziat-Filho N, Ferreira APA. Double threshold receiver operating characteristic plot for three-modal continuous predictors. *Computational Statistics.* 2021;36(3):2231–2245. doi:10.1007/s00180-021-01080-9
397. Phillips B, Stewart LA, Sutton AJ. ‘Cross hairs’ plots for diagnostic meta-analysis. *Research Synthesis Methods.* 2010;1(3-4):308–315. doi:10.1002/jrsm.26
398. Sousa-Pinto PD with contributions from B. *mada: Meta-Analysis of Diagnostic Accuracy.*; 2022. <https://CRAN.R-project.org/package=mada>.
399. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ.* outubro 2015:h5527. doi:10.1136/bmj.h5527
400. Reeves BC, Gaus W. Guidelines for Reporting Non-Randomised Studies. *Complementary Medicine Research.* 2004;11(1):46–52. doi:10.1159/000080576

401. Bland JM, Altman DG. Comparisons within randomised groups can be very misleading. *BMJ*. 2011;342(may06 2):d561–d561. doi:10.1136/bmj.d561
402. Bruce CL, Juszczak E, Ogollah R, Partlett C, Montgomery A. A systematic review of randomisation method use in RCTs and association of trial design characteristics with method selection. *BMC Medical Research Methodology*. 2022;22(1). doi:10.1186/s12874-022-01786-4
403. Vickers AJ, Altman DG. Statistics Notes: Analysing controlled trials with baseline and follow up measurements. *BMJ*. 2001;323(7321):1123–1124. doi:10.1136/bmj.323.7321.1123
404. O Connell NS, Dai L, Jiang Y, et al. Methods for Analysis of Pre-Post Data in Clinical Research: A Comparison of Five Common Methods. *Journal of Biometrics & Biostatistics*. 2017;08(01). doi:10.4172/2155-6180.1000334
405. Laird N. Further Comparative Analyses of Pretest-Posttest Research Designs. *The American Statistician*. 1983;37(4a):329–330. doi:10.1080/00031305.1983.10483133
406. Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine*. 1997;16(20):2349–2380. doi:10.1002/(sici)1097-0258(19971030)16:20<2349::aid-sim667>3.0.co;2-e
407. Mallinckrodt CH, Lane PW, Schnell D, Peng Y, Mancuso JP. Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials. *Drug Information Journal*. 2008;42(4):303–319. doi:10.1177/009286150804200402
408. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *The Lancet*. 2000;355(9209):1064–1069. doi:10.1016/s0140-6736(00)02039-0
409. Stang A, Baethge C. Imbalance p values for baseline covariates in randomized controlled trials: a last resort for the use of p values? A pro and contra debate. *Clinical Epidemiology*. 2018;Volume 10:531–535. doi:10.2147/cepl.s161508
410. Bolzern JE, Mitchell A, Torgerson DJ. Baseline testing in cluster randomised controlled trials: should this be done? *BMC Medical Research Methodology*. 2019;19(1). doi:10.1186/s12874-019-0750-8
411. Roberts C, Torgerson DJ. Understanding controlled trials: Baseline imbalance in randomised controlled trials. *BMJ*. 1999;319(7203):185–185. doi:10.1136/bmj.319.7203.185
412. Gruijters SLK. Baseline comparisons and covariate fishing: Bad statistical habits we should have broken yesterday. julho 2020. <http://dx.doi.org/10.31234/osf.io/qftwg>.
413. Vickers AJ. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. *BMC Medical Research Methodology*. 2001;1(1). doi:10.1186/1471-2288-1-6
414. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; *Journal of Clinical Epidemiology*. 2004;57(3):229–236. doi:10.1016/j.jclinepi.2003.08.009
415. Matthews JNS, Altman DG. Statistics Notes: Interaction 2: compare effect sizes not P values. *BMJ*. 1996;313(7060):808–808. doi:10.1136/bmj.313.7060.808

416. Altman DG. Statistics Notes: Interaction revisited: the difference between two estimates. *BMJ*. 2003;326(7382):219–219. doi:10.1136/bmj.326.7382.219
417. Hauck WW, Anderson S, Marcus SM. Should We Adjust for Covariates in Nonlinear Regression Analyses of Randomized Trials? *Controlled Clinical Trials*. 1998;19(3):249–256. doi:10.1016/s0197-2456(97)00147-5
418. Kahan BC, Jairath V, Doré CJ, Morris TP. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials*. 2014;15(1). doi:10.1186/1745-6215-15-139
419. Cao Y, Allore H, Vander Wyk B, Gutman R. Review and evaluation of imputation methods for multivariate longitudinal data with mixed-type incomplete variables. *Statistics in Medicine*. outubro 2022. doi:10.1002/sim.9592
420. Schulz KF. CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomized Trials. *Annals of Internal Medicine*. 2010;152(11):726. doi:10.7326/0003-4819-152-11-201006010-00232
421. Dayim A. *consort: Create Consort Diagram.*; 2023. <https://CRAN.R-project.org/package=consort>.
422. Dwan K, Li T, Altman DG, Elbourne D. CONSORT 2010 statement: extension to randomised crossover trials. *BMJ*. julho 2019;l4378. doi:10.1136/bmj.l4378
423. Senn S. The analysis of continuous data from n-of-1 trials using paired cycles: a simple tutorial. *Trials*. 2024;25(1). doi:10.1186/s13063-024-07964-7
424. Baker KA, Weeks SM. An Overview of Systematic Review. *Journal of PeriAnesthesia Nursing*. 2014;29(6):454–458. doi:10.1016/j.jopan.2014.07.002
425. Fantini D. *easyPubMed: Search and Retrieve Scientific Publication Records from PubMed.*; 2019. doi:10.32614/CRAN.package.easyPubMed
426. Chamberlain S, Zhu H, Jahn N, Boettiger C, Ram K. *rcrossref: Client for Various CrossRef APIs.*; 2022. doi:10.32614/CRAN.package.rcrossref
427. Jahn N. *roadoi: Find Free Versions of Scholarly Publications via Unpaywall.*; 2024. doi:10.32614/CRAN.package.roadoi
428. Silva V, Grande AJ, Martimbiano ALC, Riera R, Carvalho APV. Overview of systematic reviews - a new type of study: part I: why and for whom? *Sao Paulo Medical Journal*. 2012;130(6):398–404. doi:10.1590/s1516-31802012000600007
429. Silva V, Grande AJ, Carvalho APV de, Martimbiano ALC, Riera R. Overview of systematic reviews - a new type of study. Part II. *Sao Paulo Medical Journal*. 2014;133(3):206–217. doi:10.1590/1516-3180.2013.8150015
430. Stern C, Li J, Stone J, et al. Data analysis and presentation methods in umbrella reviews/overviews of reviews in health care: A cross-sectional study. *Research Synthesis Methods*. outubro 2025;1–15. doi:10.1017/rsm.2025.10040

431. Snell KIE, Levis B, Damen JAA, et al. Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA). *BMJ*. maio 2023:e073538. doi:10.1136/bmj-2022-073538
432. Moons KGM, Groot JAH de, Bouwmeester W, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLoS Medicine*. 2014;11(10):e1001744. doi:10.1371/journal.pmed.1001744
433. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*. 2010;1(2):97–111. doi:10.1002/jrsm.12
434. Viechtbauer W. *Conducting meta-analyses in {R} with the {metafor} package*. Vol 36.; 2010. doi:10.18637/jss.v036.i03
435. Balduzzi S, Rücker G, Nikolakopoulou A, et al. netmeta: An R Package for Network Meta-Analysis Using Frequentist Methods. *Journal of Statistical Software*. 2023;106(2):1–40. doi:10.18637/jss.v106.i02
436. Valkenhoef G van, Kuiper J. *gemtc: Network Meta-Analysis Using Bayesian Methods.*; 2025. doi:10.32614/CRAN.package.gemtc
437. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Medical Research Methodology*. 2005;5(1). doi:10.1186/1471-2288-5-13
438. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Medical Research Methodology*. 2014;14(1). doi:10.1186/1471-2288-14-135
439. Borenstein M. In a meta-analysis, the I-squared statistic does not tell us how much the effect size varies. *Journal of Clinical Epidemiology*. outubro 2022. doi:10.1016/j.jclinepi.2022.10.003
440. Rücker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on I² in assessing heterogeneity may mislead. *BMC Medical Research Methodology*. 2008;8(1). doi:10.1186/1471-2288-8-79
441. Grooth HJ de, Parienti JJ. Heterogeneity between studies can be explained more reliably with individual patient data. *Intensive Care Medicine*. julho 2023. doi:10.1007/s00134-023-07163-z
442. Dettori JR, Norvell DC, Chapman JR. Seeing the Forest by Looking at the Trees: How to Interpret a Meta-Analysis Forest Plot. *Global Spine Journal*. 2021;11(4):614–616. doi:10.1177/21925682211003889
443. Song, Eastwood, Gilbody, Duley, Sutton. Publication and related biases. *Health Technology Assessment*. 2000;4(10). doi:10.3310/hta4100
444. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629–634. doi:10.1136/bmj.315.7109.629
445. Peters JL. Comparison of Two Methods to Detect Publication Bias in Meta-analysis. *JAMA*. 2006;295(6):676. doi:10.1001/jama.295.6.676
446. Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011;343(jul22 1):d4002–d4002. doi:10.1136/bmj.d4002

447. Duval S, Tweedie R. Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*. 2000;56(2):455–463. doi:10.1111/j.0006-341x.2000.00455.x
448. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *PLOS Medicine*. 2021;18(3):e1003583. doi:10.1371/journal.pmed.1003583
449. Lajeunesse MJ. Facilitating systematic reviews, data extraction, and meta-analysis with the metagear package for R. *Methods in Ecology and Evolution*. 2016;7(3):323–330. doi:10.1111/2041-210X.12472
450. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*. 2015;4(1). doi:10.1186/2046-4053-4-1
451. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews*. 2022;18:e1230. doi:10.1002/cl2.1230
452. Gates M, Gates A, Pieper D, et al. Reporting guideline for overviews of reviews of healthcare interventions: development of the PRIOR statement. *BMJ*. agosto 2022:e070849. doi:10.1136/bmj-2022-070849
453. O'Cathain A, Thomas KJ, Drabble SJ, Rudolph A, Hewison J. What can qualitative research do for randomised controlled trials? A systematic mapping review. *BMJ Open*. 2013;3(6):e002889. doi:10.1136/bmjopen-2013-002889
454. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for Reporting Qualitative Research. *Academic Medicine*. 2014;89(9):1245–1251. doi:10.1097/acm.0000000000000388
455. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Medical Research Methodology*. 2012;12(1). doi:10.1186/1471-2288-12-181
456. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*. 2007;19(6):349–357. doi:10.1093/intqhc/mzm042
457. Makowski D, Lüdecke D, Patil I, Thériault R, Ben-Shachar MS, Wiernik BM. *Automated Results Reporting as a Practical Tool to Improve Reproducibility and Methodological Best Practices Adoption.*; 2023. <https://easystats.github.io/report/>.
458. Nuijten MB, Epskamp S. *statcheck: Extract Statistics from Articles and Recompute P-Values.*; 2024. doi:10.32614/CRAN.package.statcheck
459. Wallisch C, Bach P, Hafermann L, et al. Review of guidance papers on regression modeling in statistical series of medical journals. Mathes T, org. *PLOS ONE*. 2022;17(1):e0262918. doi:10.1371/journal.pone.0262918
460. Lynggaard H, Bell J, Lösch C, et al. Principles and recommendations for incorporating estimands into clinical study protocol templates. *Trials*. 2022;23(1). doi:10.1186/s13063-022-06515-2

461. Althouse AD, Below JE, Claggett BL, et al. Recommendations for Statistical Reporting in Cardiovascular Medicine: A Special Report From the American Heart Association. *Circulation*. 2021;144(4). doi:10.1161/circulationaha.121.055393
462. Lee KJ, Tilling KM, Cornish RP, et al. Framework for the treatment and reporting of missing data in observational studies: The Treatment And Reporting of Missing data in Observational Studies framework. *Journal of Clinical Epidemiology*. 2021;134:79–88. doi:10.1016/j.jclinepi.2021.01.008
463. Vickers AJ, Assel MJ, Sjoberg DD, et al. Guidelines for Reporting of Figures and Tables for Clinical Research in Urology. *Urology*. 2020;142:1–13. doi:10.1016/j.urology.2020.05.002
464. Assel M, Sjoberg D, Elders A, et al. Guidelines for Reporting of Statistics for Clinical Research in Urology. *Journal of Urology*. 2019;201(3):595–604. doi:10.1097/ju.0000000000000001
465. Lang TA, Altman DG. Basic statistical reporting for articles published in Biomedical Journals: The “Statistical Analyses and Methods in the Published Literature” or the SAMPL Guidelines. *International Journal of Nursing Studies*. 2015;52(1):5–9. doi:10.1016/j.ijnurstu.2014.09.006
466. Weissgerber TL, Milic NM, Winham SJ, Garovic VD. Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. *PLOS Biology*. 2015;13(4):e1002128. doi:10.1371/journal.pbio.1002128
467. Sauerbrei W, Abrahamowicz M, Altman DG, Cessie S, Carpenter J. STREngthening Analytical Thinking for Observational Studies: the STRATOS initiative. *Statistics in Medicine*. 2014;33(30):5413–5432. doi:10.1002/sim.6265
468. Groves T. Research methods and reporting. *BMJ*. 2008;337(oct22 1):a2201–a2201. doi:10.1136/bmj.a2201
469. Stratton IM, Neil A. How to ensure your paper is rejected by the statistical reviewer. *Diabetic Medicine*. 2005;22(4):371–373. doi:10.1111/j.1464-5491.2004.01443.x
470. Mansournia MA, Collins GS, Nielsen RO, et al. A CHecklist for statistical Assessment of MediCal Papers (the CHAMP statement): explanation and elaboration. *British Journal of Sports Medicine*. 2021;55(18):1009–1017. doi:10.1136/bjsports-2020-103652
471. Gil-Sierra MD, Fénix-Caballero S, Abdel kader-Martin L, et al. Checklist for clinical applicability of subgroup analysis. *Journal of Clinical Pharmacy and Therapeutics*. 2019;45(3):530–538. doi:10.1111/jcpt.13102
472. Lee H, Cashin AG, Lamb SE, et al. A Guideline for Reporting Mediation Analyses of Randomized Trials and Observational Studies. *JAMA*. 2021;326(11):1045. doi:10.1001/jama.2021.14075

Ciência com R

Você está pronto para desbloquear o poder da análise estatística de dados e elevar sua pesquisa a novos patamares? Não procure mais. Em “Ciência com R”, o Dr. Arthur de Sá Ferreira, um pesquisador experiente, oferece um guia indispensável que capacitará pesquisadores, analistas de dados e estudantes a tomarem decisões informadas e baseadas em evidências em seus empreendimentos científicos.

ORIENTAÇÃO ESPECIALIZADA: Beneficie-se da ampla experiência do Dr. Arthur de Sá Ferreira enquanto ele responde às perguntas mais fundamentais: *O que é isso? Por que usá-lo? Quando usar? Quando não usar? Como fazer?* Cada capítulo se aprofunda em questões específicas, oferecendo explicações claras e concisas e exemplos práticos.

FORMATO DE PERGUNTAS E RESPOSTAS: Mantenha uma conversa direta e objetiva com o autor. Descubra respostas para as perguntas comumente feitas por estudantes, pesquisadores e profissionais em todas as fases de sua jornada acadêmica e científica.

APRENDIZADO PROGRESSIVO: Navegue por uma progressão de conceitos e aplicações. Capítulos são estruturados didaticamente para maior clareza educacional, com referências cruzadas para garantir uma compreensão coesa dos tópicos inter-relacionados, reduzindo a fragmentação do conteúdo.

INSIGHTS ATUALIZADOS: Fique à frente da curva com as referências e insights mais recentes. Dr. [Seu nome] lança luz sobre preconceitos, paradoxos, mitos e práticas ilícitas na área, oferecendo uma clareza inestimável até mesmo para os pesquisadores mais experientes.

SUPORTE DA LITERATURA: Com um total de 511 referências, cada capítulo é apoiado por uma extensa bibliografia, permitindo que você aprofunde seu conhecimento e explore tópicos adicionais.

Quer você seja um estudante de pós-graduação em busca de métodos para analisar seus projetos de pesquisa, um pesquisador que precisa de informações e referências para o desenvolvimento de projetos ou um analista de dados experiente que deseja se manter atualizado, este livro é seu melhor companheiro. Além disso, pessoas de diversas áreas encontrarão neste livro uma porta de entrada para compreender a importância de fazer e responder perguntas no mundo da ciência.

Tome decisões informadas, evite armadilhas e destaque-se em sua pesquisa científica com “Ciência com R”. Os insights profundos do Dr. Arthur de Sá Ferreira permitirão que você transforme seus dados em descobertas significativas, colocando você no caminho da excelência em pesquisa.