

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

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.
465 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	5
1.5 Independência e probabilidade	5
1.6 Leis dos números anômalos	7
1.7 Leis dos pequenos números	8
1.8 Leis dos grandes números	8
1.9 Teorema central do limite	8
1.10 Regressão para a média	11
2 Pensamento estatístico	13
2.1 Unidade de análise	13
2.2 População	13
2.3 Amostra	14
2.4 Amostragem	14
2.5 Reamostragem	15
2.6 Subamostragem	18
2.7 Superamostragem	18
3 Pensamento metodológico	21
3.1 Metodologia da pesquisa	21
3.2 Relação Estatística-Metodologia	21

3.3	Pesquisa quantitativa vs. qualitativa	21
3.4	Pesquisa de métodos mistos	23
3.5	Pesquisa exploratória vs. confirmatória	24
3.6	Pesquisa translacional	24
3.7	Pré-registro	24
3.8	Reprodutibilidade e Ciência Aberta	25
3.9	Robustez	25
3.10	Replicabilidade	25
3.11	Generalização	25
4	Pensamento computacional	29
4.1	R	29
4.2	RStudio	30
4.3	Scripts	30
4.4	Pacotes	33
4.5	Aplicativos	34
4.6	Manuscritos reproduzíveis	34
4.7	Compartilhamento	36
5	Letramento estatístico	39
5.1	Introdução ao letramento estatístico	39
5.2	Elementos centrais do letramento estatístico	40
5.3	Hierarquia de letramento estatístico	40
5.4	Habilidades de letramento estatístico baseadas no pensamento crítico	41
<hr/>		
<i>PARTE 2: AMEAÇAS À QUALIDADE DA EVIDÊNCIA CIENTÍFICA</i>		43
6	Vieses metodológicos	45
6.1	Vieses metodológicos	45
6.2	Tipos de vieses metodológicos	45
6.3	Efeitos relacionados aos vieses metodológicos	45
6.4	Diretrizes para redação	46
7	Falácia estatísticas	47
7.1	Falácia	47
8	Paradoxos estatísticos	49
8.1	Paradoxos	49
9	Práticas questionáveis em pesquisa	53
9.1	Práticas Questionáveis em Pesquisa	53
9.2	Prática não intencional e má conduta	53
9.3	Prevenindo práticas questionáveis em pesquisa	56
9.4	Reações éticas e institucionais práticas questionáveis em pesquisa	56
<hr/>		
<i>PARTE 3: DO MUNDO REAL À TABELA</i>		59
10	Variáveis e fatores	61
10.1	Variáveis	61
10.2	Transformação de variáveis	62

10.3 Centralização de variáveis (<i>centering</i>)	64
10.4 Padronização de variáveis	64
10.5 Categorização de variáveis contínuas	66
10.6 Dicotomização de variáveis contínuas	66
10.7 Representação de variáveis categóricas	68
10.8 Fatores	69
11 Dados e metadados	71
11.1 Dados	71
11.2 <i>Big data</i>	71
11.3 Metadados	72
11.4 Armazenamento de dados	73
12 Medidas e instrumentos	75
12.1 Escalas	75
12.2 Medição e Medidas	77
12.3 Erro de medida	80
12.4 Instrumentos	80
12.5 Acurácia e precisão	80
12.6 Viés e variabilidade	81
13 Tabulação de dados	83
13.1 Planilhas eletrônicas	83
14 Dados perdidos e imputados	87
14.1 Dados perdidos	87
14.2 Mecanismos geradores de dados perdidos	88
14.3 Estratégias para lidar com dados perdidos	91
14.4 Dados imputados	92
15 Dados anonimizados e sintéticos	95
15.1 Dados anonimizados	95
15.2 Dados sintéticos	96
16 Ciência cidadã	97
16.1 Introdução	97
<hr/>	
PARTE 4: ANÁLISES DESCRIPTIVAS E EXPLORATÓRIAS	99
17 Distribuições e parâmetros	101
17.1 Distribuições de probabilidade	101
17.2 Distribuições multivariadas	107
17.3 Parâmetros	108
17.4 Erro	109
17.5 Tendência central	110
17.6 Dispersão	111
17.7 Proporção	114
17.8 Extremos	115
17.9 Distribuição	115
17.10 Robustez em medidas de localização	117
17.11 Parâmetros robustos	118

18 Análise inicial de dados	119
18.1 Análise inicial de dados	119
19 Análise exploratória de dados	121
19.1 Análise exploratória de dados	121
19.2 Ingredientes da análise exploratória de dados	122
20 Análise descritiva	125
20.1 Análise descritiva	125
20.2 Apresentação de resultados numéricos	125
20.3 Tabelas	127
20.4 Tabela 1	128
20.5 Tabela 2	129
20.6 Visualização efetiva de dados	130
20.7 Gráficos	132
20.8 Tipos de gráficos	134
20.9 Gráficos dinâmicos	150
21 Análise robusta	153
21.1 Raciocínio inferencial robusto	153
21.2 Valores discrepantes	154
21.3 Valores influentes	157
21.4 Métodos robustos de tratamento de <i>outliers</i>	158
22 Análise preditiva	161
22.1 Predição	161
22.2 Interpretação e aplicação	161
22.3 Análise de curva de decisão	162
23 Análise causal	163
23.1 Causalidade	163
23.2 Abordagens filosóficas e estatísticas da causalidade	164
23.3 Ilusões de causalidade	164
23.4 Inferência causal em estudos observacionais	165
23.5 Critérios de Hill para inferência causal	165
23.6 Críticas contemporâneas aos critérios de Hill	166
23.7 Visão atual sobre os critérios de Hill	166
23.8 Linguagem causal em estudos observacionais	167
23.9 Efeitos diretos e indiretos	167
23.10 O papel do tempo e a causalidade dinâmica	167
23.11 Diagrama acíclico direcionado (DAG)	168
24 Análise qualitativa	171
24.1 Análise qualitativa	171
24.2 Representação de texto	171
<hr/>	
PARTE 5: ANÁLISES INFERENCIAIS	173
25 Suposições inferenciais	175
25.1 Suposições gerais em análises inferenciais	175
25.2 Suposições implícitas e explícitas nos testes	175
25.3 Suposições causais que conectam dados observados a efeitos causais	176

25.4 Diagnóstico e verificação	176
26 Seleção de testes	179
26.1 Multiverso de análises estatísticas	179
26.2 Escolha de testes para análise inferencial	179
27 Análise inferencial	187
27.1 Raciocínio inferencial	187
27.2 Hipóteses científicas	187
27.3 Hipóteses estatísticas	188
27.4 Testes de hipóteses	188
27.5 Intervalos de confiança e raciocínio de longo prazo	189
27.6 Comparações múltiplas	191
27.7 Inferência visual	195
27.8 Interpretação de análise inferencial	195
27.9 Erros de inferência I, II, S e M	196
28 Tamanho do efeito	201
28.1 Tamanho do efeito	201
28.2 Tipos de tamanho do efeito	201
28.3 Conversão entre tamanhos do efeito	204
28.4 Efeitos bruto e padronizado	204
29 P-valor	205
29.1 P-valor	205
29.2 Significância estatística	205
29.3 Interpretação do P-valor	206
29.4 P-valor de 2 ^a geração	208
29.5 Distribuição de confiança	209
29.6 Boas práticas	209
30 Testes estatísticos	211
30.1 Variáveis categóricas	211
30.2 Variáveis contínuas	211
31 Descrição	217
31.1 Análise de descrição	217
31.2 Estimação	218
32 Comparação	219
32.1 Análise inferencial de comparação	219
32.2 F-teste	219
33 Associação	221
33.1 Análise inferencial de associação	221
33.2 Associação bivariada	221
33.3 Associação multivariada	223
34 Correlação	225
34.1 Análise inferencial de correlação	225
34.2 Coeficientes de correlação	229
34.3 Colinearidade	232
34.4 Correlação entre conjuntos de variáveis	232

35 Redes	235
35.1 Análise de redes	235
35.2 Matriz de incidência	236
35.3 Elementos da rede	236
35.4 Tipos de redes	237
35.5 Métricas de rede	238
36 Regressão	241
36.1 Análise de regressão	241
36.2 Estruturas de análise de regressão	242
36.3 Tipos e famílias de regressão	242
36.4 Preparação de variáveis	249
36.5 Multicolinearidade	251
36.6 Redução de dimensionalidade	253
36.7 Seleção de variáveis em regressão	254
37 Mediação	257
37.1 Análise inferencial de mediação	257
<hr/>	
PARTE 6: MODELAGEM	259
38 Modelos	261
38.1 Modelos	261
38.2 Modelos estocásticos	261
38.3 Efeito fixo	262
38.4 Efeito aleatório	262
38.5 Efeito misto	264
38.6 Efeito principal	264
38.7 Efeito de interação	265
38.8 Efeito de mediação	267
38.9 Efeito de modificação	267
38.10 Preditores e desfechos	267
38.11 Desempenho e estabilidade de modelos	268
38.12 Comparação de modelos	268
38.13 Avaliação de modelos	268
38.14 Validação de modelos	270
38.15 Calibração de modelos	270
39 Modelagem temporal	271
39.1 Modelos temporais	271
40 Modelagem espacial	273
40.1 Modelos espaciais	273
41 Modelagem de sobrevida	275
41.1 Análise de sobrevida	275
41.2 Eventos	275
41.3 Dados censurados	275
41.4 Medidas de associação em análise de sobrevida	276
41.5 Modelo de Kaplan–Meier	277
41.6 Modelos de Cox	278

PARTE 7: REPRESENTAÇÃO, APRENDIZADO DE MÁQUINA E INTELIGÊNCIA ARTIFICIAL	281
42 Representações	283
42.1 Representações de dados e extração de atributos	283
43 Inteligência artificial	285
43.1 Inteligência artificial	285
43.2 Inteligência artificial gerativa	285
43.3 Inteligência artificial explicável (<i>eXplainable Artificial Intelligence, XAI</i>)	285
43.4 Limitações fundamentais de modelos gerativos	287
44 Aprendizado de máquina	291
44.1 Aprendizado de máquina	291
44.2 Representação de dados e engenharia de atributos	291
44.3 Métricas de distância e similaridade	294
44.4 Desbalanceamento de classes	296
44.5 Tipos de aprendizado	297
44.6 Avaliação de modelos de classificação	297
44.7 Principais algoritmos	298
44.8 Regressão logística	299
44.9 Máquina de vetores de suporte	299
44.10 <i>K-nearest neighbours</i>	299
44.11 <i>K-means Clustering</i>	301
44.12 Árvores de decisão	301
44.13 Análise de componentes principais	303
44.14 <i>Random forests</i>	305
44.15 <i>Ensemble</i>	305
45 Redes neurais	307
45.1 Neurônios artificiais	307
45.2 Rede neural artificial	308
45.3 Funções de ativação	308
45.4 Funções de perda	309
45.5 Treinamento de redes neurais	310
45.6 Espaço de decisão	311
45.7 Redes neurais multicamadas	312
45.8 Redes neurais profundas	313
45.9 Redes neurais convolucionais (Convolutional Neural Networks)	313
PARTE 8: PLANEJAMENTO DE ESTUDOS	315
46 Poder estatístico	317
46.1 Poder do teste	317
47 Tamanho da amostra	321
47.1 Tamanho da amostra	321
47.2 Saturação em pesquisas qualitativas	322
47.3 “Fome de dados”	325
47.4 Eventos por variável (EPV) em modelos preditivos	325
47.5 Cálculo do tamanho da amostra	326

47.6 Perdas de amostra	328
47.7 Ajustes no tamanho da amostra	328
47.8 Justificativa do tamanho da amostra	328
48 Plano de análise	331
48.1 Plano de análise estatística	331
48.2 Estimandos	332
48.3 Diretrizes para redação	333
<hr/>	
PARTE 9: DELINEAMENTOS E SÍNTSEDE EVIDÊNCIAS	335
49 Delineamento de estudos	337
49.1 Critérios de delineamento	337
49.2 Alocação	337
49.3 Cegamento	337
49.4 Pareamento	337
49.5 Aleatorização	338
49.6 Taxonomia de estudos	338
50 Simulação computacional	341
50.1 Simulações computacionais	341
50.2 Características	342
50.3 Métodos de simulação	342
50.4 Diretrizes para redação	346
51 Estudos observacionais	347
51.1 Características	347
51.2 Diretrizes para redação	347
52 Concordância e confiabilidade	349
52.1 Concordância	349
52.2 Confiabilidade	351
52.3 Diretrizes para redação	352
53 Propriedades psicométricas	353
53.1 Características	353
53.2 Análise fatorial exploratória	353
53.3 Análise fatorial confirmatória	353
53.4 Validade de conteúdo	354
53.5 Validade de face	354
53.6 Validade do construto	354
53.7 Validade fatorial	354
53.8 Validade convergente	354
53.9 Validade discriminante	355
53.10 Validade de critério	355
53.11 Validade concorrente	355
53.12 Responsividade	355
53.13 Diretrizes para redação	355
54 Desempenho diagnóstico	357
54.1 Características	357
54.2 Tabelas 2x2	357

54.3 Tabelas 2x3	361
54.4 Curvas ROC	362
54.5 Interpretação da validade de um teste	364
54.6 Diretrizes para redação	365
55 Ensaios quase-experimentais	367
55.1 Características	367
55.2 Diretrizes para redação	367
56 Ensaios experimentais	369
56.1 Ensaio experimental aleatorizado	369
56.2 Modelos de análise de comparação	370
56.3 Comparação na linha de base	371
56.4 Comparação intragrupos	372
56.5 Comparação entre grupos	372
56.6 Comparação de subgrupos	373
56.7 Efeito de interação	373
56.8 Ajuste de covariáveis	374
56.9 Imputação de dados perdidos	374
56.10 Diretrizes para redação	375
57 Ensaios cruzados	377
57.1 Características	377
57.2 Diretrizes para redação	377
58 n-de-1	379
58.1 Ensaio n-de-1	379
58.2 Aspectos metodológicos	380
58.3 Análise de dados	381
58.4 Abordagem meta-analítica	383
58.5 Limitações e cuidados	383
59 Revisão sistemática	385
59.1 Tipologia de revisões	385
59.2 Revisão sistemática de literatura	385
59.3 Tipos de revisão sistemática	386
59.4 Diretrizes para redação	387
60 Meta-análise	389
60.1 Características	389
60.2 Modelos de meta-análise	389
60.3 Conversão de Medidas em Meta-análises	390
60.4 Interpretação de efeitos em meta-análise	391
60.5 <i>Forest plot</i>	392
60.6 <i>Crosshair</i>	395
60.7 <i>Funnel plot</i>	395
60.8 Testes de assimetria do <i>funnel plot</i>	396
60.9 Diretrizes para redação	397
61 Revisão guarda-chuva	399
61.1 Revisão guarda-chuva	399
61.2 Diretrizes para redação	399
62 Pesquisa qualitativa	401

62.1 Pesquisa qualitativa	401
62.2 Diretrizes para redação	401
 PARTE 10: COMUNICAÇÃO E RELATO CIENTÍFICO	
63 Redação de resultados 403	
63.1 Comunicação de resultados da análise estatística	405
63.2 Diretrizes e Listas	406
64 Diretrizes e Listas 409	
64.1 Diretrizes	409
64.2 Listas de verificação	410
 REFERÊNCIAS 411	
Produção do autor 413	
Artigos em periódicos científicos	413
Preprints	427
Resumos publicados em eventos científicos	427
Livros editorados	428
Capítulos de livro	428
Aulas	428
Programas de computador	428
Bancos de dados	429
Outros	429
Fontes externas 431	
Fontes de informação externas	431
Referências 433	

Lista de Figuras

1.1	Exemplos de espaço amostral discreto. Superior: Todas as faces de uma moeda. Inferior: Todas as faces de um dado.	3
1.2	Exemplos de evento de experimento. Superior: 1 lançamento de 1 moeda. Inferior: 1 lançamento de 1 dado.	4
1.3	Espaço de eventos: União dos eventos face = 3 e face = 4 de um dado.	4
1.4	Superior: Eventos independentes. Inferior: Eventos dependentes.	6
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).	7
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.	9
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.	10
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.	10
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).	12
2.1	Representação esquemática da amostragem: seleção de uma população para a amostra.	15
2.2	Representação esquemática do erro de amostragem: seleção de várias amostras independentes de uma população.	16
2.3	Representação esquemática da amostragem de uma população para a amostra.	17
2.4	Representação esquemática da reamostragem de uma amostra.	18
2.5	Representação esquemática da subamostragem de uma amostra.	19
2.6	Representação esquemática da superamostragem de uma população.	19
3.1	Mapa mental da relação entre o pensamento estatístico e o pensamento metodológico.	22
3.2	Representação esquemática da generalização de uma amostra para a população.	26
4.1	Interface do RStudio. Fonte: https://docs.posit.co/ide/user/	31
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.	51
8.2	Simulação do paradoxo de Stein. Comparação do erro médio quadrático entre o estimador clássico (média amostral) e o estimador de James-Stein para diferentes números de médias sendo estimadas simultaneamente. Estimadores aparentemente piores localmente podem ser melhores globalmente quando o objetivo é reduzir o erro total.	51

10.1 Transformações de variáveis com assimetria à direita (Original, Raiz quadrada, Log natural, Log10, Inversa)	63
10.2 Transformações de variáveis com assimetria à esquerda (Original, Reflexão + Raiz quadrada, Reflexão + Log natural, Reflexão + Log10, Reflexão + Inversa).	63
10.3 Comparação entre variáveis originais e padronizadas (Z-score e Min-Max).	65
12.1 Exemplo de instrumento com 3 itens tipo Likert com 5 categorias cada.	76
12.2 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.	80
12.3 Acurácia e precisão como propriedades de uma medida.	81
12.4 Viés e variabilidade de uma medida.	82
14.1 Representação gráfica de dados perdidos completamente ao acaso (MCAR) em um estudo randomizado controlado (RCT).	88
14.2 Representação gráfica de dados perdidos ao acaso (MAR) em um estudo randomizado controlado (RCT).	89
14.3 Representação gráfica de dados perdidos não ao acaso (MNAR) em um estudo randomizado controlado (RCT).	90
14.4 Impacto de métodos de imputação na distribuição de uma variável contínua com dados perdidos.	93
17.1 Histogramas com diferentes métodos de binning.: Sturges, Scott e Freedman-Diaconis.	102
17.2 Distribuições discretas e suas funções de probabilidade.	103
17.3 Distribuições contínuas básicas e suas funções de densidade.	104
17.4 Distribuições contínuas aproximadas e suas funções de densidade.	104
17.5 Distribuições contínuas aproximadas e suas funções de densidade.	104
17.6 Distribuições contínuas para inferência e suas funções de densidade.	105
17.7 Distribuições contínuas para dados específicos e suas funções de densidade.	105
17.8 Distribuições contínuas para probabilidades e proporções e suas funções de densidade.	105
17.9 Distribuições contínuas com caudas pesadas e suas funções de densidade.	105
17.10 Distribuições e funções de probabilidade.	106
17.11 Distribuição normal e métodos de visualização e testes de normalidade.	107
17.12 Distribuição normal bivariada e amostra simulada com histogramas marginais.	108
17.13 Distribuições unimodal, bimodal e multimodal.	111
17.14 Parâmetros de tendência central em distribuições assimétricas e normais.	112
17.15 Parâmetros de dispersão em distribuições normais.	113
17.16 Regressão linear com valores extremos.	116
17.17 Parâmetros de distribuição: Assimetria e Curtose.	117
17.18 Parâmetros de distribuição: Curtose em distribuições simétricas (normal vs. uniforme).	117
19.1 Séries temporais e autocorrelação de duas séries simuladas com fraca e forte autocorrelação.	124
20.1 Exemplo de gráfico que mostra os dados brutos junto com um resumo estatístico (média e dispersão).	131
20.2 Exemplos de gráficos com barras de erro e dados brutos.	133
20.3 Exemplos de gráficos com barras de erro e dados brutos em diferentes cenários.	133
20.4 Gráfico de barras simples representando frequências por categoria.	134
20.5 Gráfico de barras empilhadas representando frequências por categoria.	135
20.6 Gráficos de barras representando médias, barras de erro e dados individuais.	136
20.7 Histograma da variável 'valor'.	137
20.8 Gráfico de densidade da variável 'valor'.	137
20.9 Boxplot por grupo.	138
20.10 Violin plot por grupo.	139
20.11 Gráfico de pontos da variável 'valor'.	140

20.12 Gráfico de dispersão representando a relação entre duas variáveis.	141
20.13 Gráfico de bolhas representando a relação entre três variáveis.	142
20.14 Sankey plot representando fluxos entre categorias.	142
20.15 Gráfico de categorias paralelas (parcats) representando transições entre categorias ao longo do tempo.	143
20.16 Gráfico de pares representando correlações entre múltiplas variáveis.	143
20.17 Gráfico spaghetti representando dados longitudinais.	144
20.18 Gráfico spaghetti representando dados longitudinais.	145
20.19 Gráfico de linha representando uma série temporal.	146
20.20 Gráfico de correlação entre duas variáveis com linha de tendência.	147
20.21 Matriz de dispersão representando relações entre múltiplas variáveis.	148
20.22 Mapa de calor da correlação entre variáveis.	149
20.23 Gráfico radar representando múltiplas variáveis.	149
20.24 Exemplo de gráfico interativo com Plotly.	151
21.1 Regressão linear com valores discrepantes	154
21.2 Regressão linear com valores influentes.	157
21.3 Alavancagem vs Resíduos Padronizados com distância de Cook para análise da influência de pontos.	158
21.4 Boxplots comparando dados originais e dados Winsorizados.	159
23.1 Padrões causais básicos: independência, cadeia, garfo e colisor.	168
25.1 Diagnóstico de regressão para avaliar suposições do modelo: linearidade, normalidade dos resíduos, homocedasticidade e alavancagem.	177
27.1 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.	190
27.2 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).	192
27.3 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).	193
27.4 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).	194
27.5 Representação gráfica dos erros tipo I e tipo II em um teste de hipótese (bicaudal).	197
27.6 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.	197
27.7 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.	198
27.8 Representação gráfica do erro tipo S (sinal) em um teste de hipótese (bicaudal).	198
27.9 Representação gráfica do erro tipo M (magnitude) em um teste de hipótese (bicaudal).	199
29.1 Visualização espacial de $p < 0,05$ (5 quadrados aleatórios em 100).	207
29.2 Distribuição de confiança para o tamanho do efeito estimado.	210
34.1 Exemplo de diferentes forças e direção de correlação entre duas variáveis X e Y.	226
34.2 Gráfico de dispersão do Quarteto de Anscombe para representação gráfica de conjuntos de dados bivariados com parâmetros quase idênticos e relações muito distintas.	228

34.3 Exemplo de análise de correlação canônica (CCA) entre dois conjuntos de variáveis.	234
35.1 Exemplo de grafo de rede com nodos e arestas ponderadas.	237
35.2 Exemplo de grafo de rede com tamanho dos nodos proporcional à força.	238
36.1 Regressão linear.	243
36.2 Regressão polinomial.	244
36.3 Regressão não-linear.	245
36.4 Regressão logística.	246
36.5 Regressão multinomial.	247
36.6 Regressão de Poisson.	248
36.7 Regressão ridge.	250
36.8 Multicolinearidade entre variáveis candidatas em modelos de regressão multivariável.	252
38.1 Cadeia de Markov com 3 estados (a, b, c) e suas probabilidades de transição.	262
38.2 Trajetória de estados e proporção acumulada por estado em uma cadeia de Markov com 3 estados (a, b, c).	263
38.3 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.	264
38.4 Análise de efeito de interação (direta) entre grupos e tempo. Retas paralelas sugerem ausência de efeito de interação.	265
38.5 Análise de efeito de interação (inversa) entre grupos e tempo. Retas paralelas sugerem ausência de efeito de interação.	266
38.6 Exemplos de ajuste de modelos de regressão linear simples ($y \sim x$) 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.	269
41.1 Curvas de Kaplan–Meier simuladas para dois grupos (controle e tratamento).	277
41.2 Curvas ajustadas pelo modelo de Cox para dois grupos (controle e tratamento).	279
43.1 Representação esquemática de um modelo de linguagem grande (LLM)	286
43.2 Funcionamento conceitual de modelos de linguagem e origem das alucinações. O painel A ilustra o fluxo abstrato de um LLM até a predição do próximo token; o painel B mostra a escolha baseada em plausibilidade linguística; o painel C evidencia que essa otimização probabilística pode gerar respostas fluentes, porém factualmente falsas.	288
44.1 Mapa mental de algoritmos de aprendizado de máquina.	292
44.2 Visualização de diferentes métricas de distância e similaridade.	295
44.3 Regressão logística.	300
44.4 Exemplo de árvore de decisão para predizer depressão a partir de idade, tabagismo e sintomas.	301
44.5 Comparação entre modelos de regressão logística e árvore de decisão.	303
44.6 Análise de Componentes Principais (PCA). O PC1 maximiza variância total, mas pode não alinhar com o fator latente real (z_1).	304
45.1 Representação esquemática de um neurônio computacional.	307
45.2 Representação esquemática de uma rede neural simples com camada de entrada e saída.	308
45.3 Gráficos das funções de ativação mais comuns.	309
45.4 Espaço de decisão de um perceptron (regressão logística).	312
45.5 Comparação do espaço de decisão entre um modelo linear (regressão logística) e um modelo não linear (MLP).	313

45.6 Representação esquemática de uma rede neural multicamadas com 2 camadas escondidas além das camadas de entrada e saída.	314
47.1 Curvas de poder para testes t (quantitativo). Linhas sólidas: $\alpha = 0,05$ tracejadas: $\alpha = 0,01$ linhas horizontais em 80% e 90% de poder.	323
47.2 Curvas de saturação para estudos qualitativos de descoberta de temas.	324
49.1 Alocação 1:1 entre dois grupos de participantes	338
50.1 Dados simulados a partir de diferentes distribuições: Normal(0,1), Binomial(1,0.4), Poisson(2) e Exponencial(1).	343
50.2 Convergência do histograma para a PDF teórica da Normal(0,1) com o aumento do tamanho amostral ($n = 10, 100, 1000, 10000$).	344
50.3 Convergência da média e do desvio-padrão amostral para os valores teóricos (0 e 1, respectivamente) com o aumento do tamanho amostral ($n = 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000$).	345
54.1 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).	358
54.2 Trade-off entre sensibilidade e especificidade em função do limiar de probabilidade (t) para um modelo de classificação.	359
54.3 Curva ROC (Receiver Operating Characteristic) para um modelos de classificação com diferentes desempenhos diagnósticos.	364
58.1 Trellis plot para dados simulados de ensaio n-de-1. Cada painel representa um paciente, cada ponto azul um ciclo, e o asterisco vermelho indica a média dos ciclos para aquele paciente. A linha tracejada representa a linha de identidade (igualdade entre tratamentos).	382
60.1 Comparação entre modelos de efeito fixo e aleatório com 10 ensaios clínicos simulados.	390
60.2 Forest plot de uma meta-análise de efeito fixo com 10 ensaios clínicos simulados.	392
60.3 Forest plot de uma meta-análise de efeito aleatório com 10 ensaios clínicos simulados.	393
60.4 Forest plots ilustrativos para faixas usuais de I^2	394
60.5 Gráfico de cruzes (*crosshair*) em espaço ROC (*Receiver Operating Characteristic*) para 15 estudos simulados de desempenho diagnóstico.	395
60.6 Gráficos de funil simulados com baixa e alta heterogeneidade.	396

RASCUNHO

Lista de Tabelas

8.1	Modelo micro (por tentativa): Desempenho ~ Habilidade	50
9.1	Classificação das práticas questionáveis em pesquisa segundo sua intencionalidade.	54
10.1	Tabela de variáveis indicadoras (dummy variables) criadas a partir de variáveis categóricas Sexo e Grupo.	68
12.1	Descrição dos itens tipo Likert do instrumento.	75
12.2	Dados brutos com medidas únicas.	77
12.3	Dados brutos com medidas repetidas.	78
12.4	Dados brutos com medidas repetidas agregadas.	78
12.5	Dados brutos com medidas seriadas não agregadas.	79
12.6	Dados brutos com medidas seriadas não agregadas.	79
13.1	Estrutura básica de uma tabela de dados.	83
13.2	Formato longo para medidas repetidas (múltiplas linhas por sujeito; colunas = variáveis)	84
13.3	Formato largo para medidas repetidas (1 variável; colunas = tempos)	84
13.4	Formatação recomendada para tabela de dados.	85
13.5	Formatação não recomendada para tabela de dados.	85
14.1	Simulação de uma amostra (n=10) de um ensaio clínico aleatorizado (dados com perdas aleatórias).	87
20.1	Quantidade de casas decimais e dígitos significativos.	126
20.2	Valores originais, arredondamentos e erros de arredondamento por casas decimais.	126
20.3	Características da amostra por grupo.	129
20.4	Exemplo clássico de apresentação suscetível à Falácia da ‘Tabela 2’.	130
20.5	Exposição (OR; 95% IC) com e sem ajuste.	130
26.1	Tabela de seleção de testes estatísticos: Descrição.	180
26.2	Tabela de seleção de testes estatísticos: Comparação.	180
26.3	Tabela de seleção de testes estatísticos: Associação.	181
26.4	Tabela de seleção de testes estatísticos: Predição.	182
26.5	Tabela de seleção de testes estatísticos: Contagem.	182
26.6	Tabela de seleção de testes estatísticos: Sobrevida.	183
26.7	Tabela de seleção de testes estatísticos: Concordância.	183
26.8	Tabela de seleção de testes estatísticos: Ajuste multivariado.	183
26.9	Tabela de seleção de testes estatísticos: Desempenho diagnóstico.	184
26.10	Tabela de seleção de testes estatísticos: Diagnóstico.	184
26.11	Tabela de seleção de testes estatísticos: Transição de estados.	185
27.1	Tabela de erros tipos I e II de inferência estatística.	196

27.2 Tabela de erro tipo S de inferência estatística.	197
27.3 Tabela de erro tipo M de inferência estatística.	199
29.1 Comparação entre p-valor (bicaudal, inferido do IC95%) e SGPV (p_δ) nos cenários simulados.	209
30.1 Teste Qui-quadrado (com correção de Yates)	212
30.2 Teste Qui-quadrado (sem correção de Yates)	212
30.3 Teste exato de Fisher	212
30.4 Teste de McNemar	213
30.5 Teste Q de Cochran	213
30.6 Teste de Cochran–Armitage	213
30.7 Medidas de associação	214
30.8 Teste t de Student	214
30.9 Teste t de Welch	214
30.10 Teste de Mann–Whitney (Wilcoxon rank-sum)	214
30.11 Teste de Wilcoxon (signed-rank)	215
30.12 Análise de variância de um fator	215
30.13 Análise de variância de Welch	215
30.14 Teste de Kruskal–Wallis	216
34.1 Quarteto de Anscombe.	227
34.2 Análise descritiva do Quarteto de Anscombe demonstrando os conjuntos de dados bivariados com parâmetros quase idênticos.	227
35.1 Exemplo de dados categóricos para criação de matriz de incidência.	236
35.2 Matriz de incidência criada a partir de dados categóricos.	236
35.3 Tabela com principais métricas de centralidade dos nodos.	238
38.1 Métricas de desempenho do modelo de regressão linear.	270
44.1 Exemplo de codificação one-hot, multi-hot e count encoding para representação de texto.	293
44.2 Tabela de confusão 2x2 para análise de desempenho diagnóstico de testes e variáveis dicotômicas.	298
52.1 Tabela de confusão 2x2 para análise de concordância de testes e variáveis dicotômicas.	349
52.2 Tabela de confusão 3x3 para análise de concordância de testes e variáveis dicotômicas.	350
53.1 Tabela de confusão sobre propriedades psicométricas de instrumentos.	353
54.1 Tabela de confusão 2x2 para análise de desempenho diagnóstico de testes e variáveis dicotômicas.	357
54.2 Probabilidades calculados a partir da tabela de confusão 2x2 para análise de desempenho diagnóstico de testes e variáveis dicotômicas.	360
54.3 Tabela de confusão 3-vias (2×3) com totais: referência vs decisão (3WD).	361

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 545 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

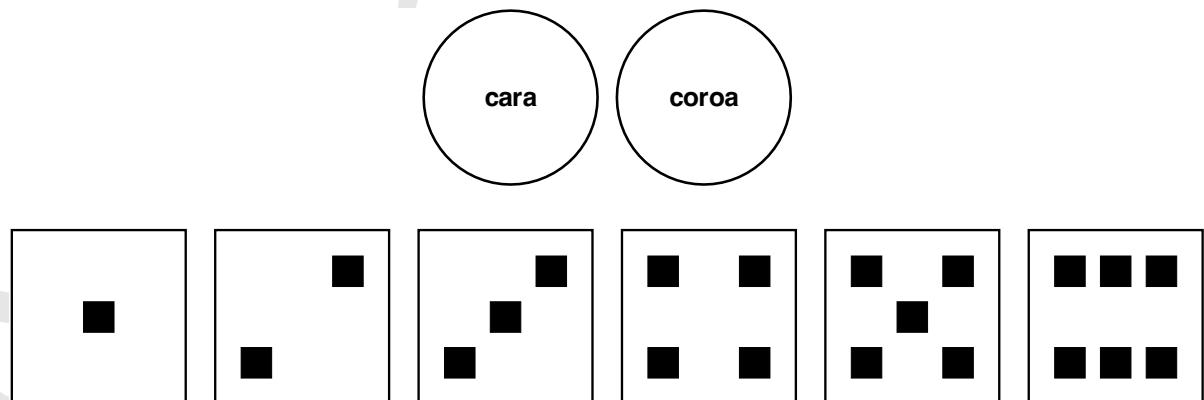


Figura 1.1: Exemplos de espaço amostral discreto. Superior: Todas as faces de uma moeda. Inferior: Todas as faces de um dado.

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

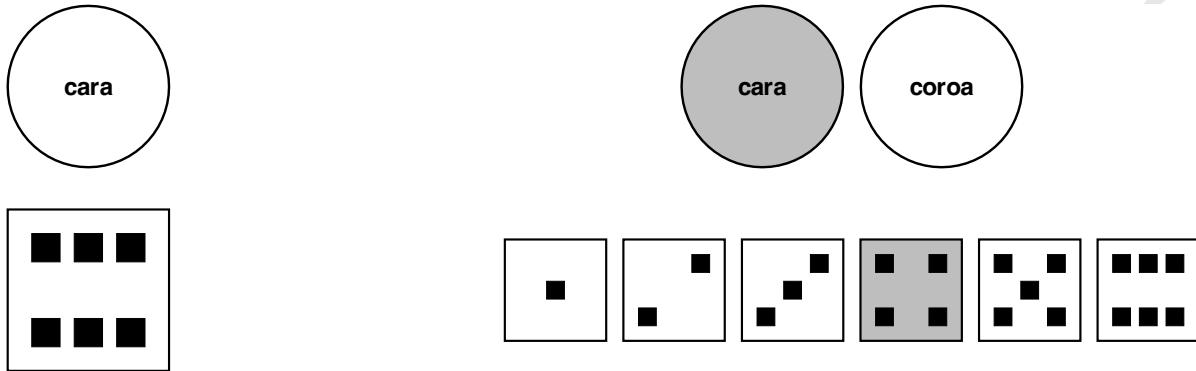


Figura 1.2: Exemplos de evento de experimento. Superior: 1 lançamento de 1 moeda. Inferior: 1 lançamento de 1 dado.

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

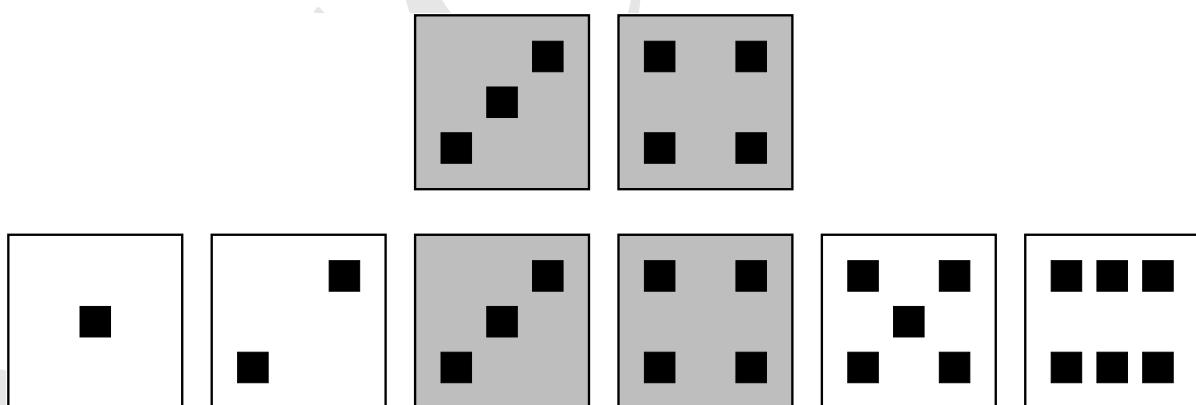


Figura 1.3: Espaço de eventos: União dos eventos face = 3 e face = 4 de um dado.

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}
- 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: $P(E) \geq 0$.^{1,2}
- Axioma II. Probabilidade de todos os eventos do espaço amostral A ocorrerem é 100%: $P(S) = 1$.^{1,2}
- Axioma III. A probabilidade de um conjunto k de eventos mutuamente exclusivos é igual a soma da probabilidade de cada evento: $P(E_1 \cup E_2 \cup \dots \cup E_k) = P(E_1) + P(E_2) + \dots + P(E_k)$.^{1,2}

1.4.3 Quais as consequências dos axiomas da probabilidade?

- A soma da probabilidade de dois eventos que dividem o espaço amostral é 100%: $P(E) + P(E^C) = 1$.¹
- O valor máximo de probabilidade de um evento é 100%: $P(S) \leq 1$.¹
- A probabilidade é uma função não decrescente do número de desfechos de um evento.¹

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

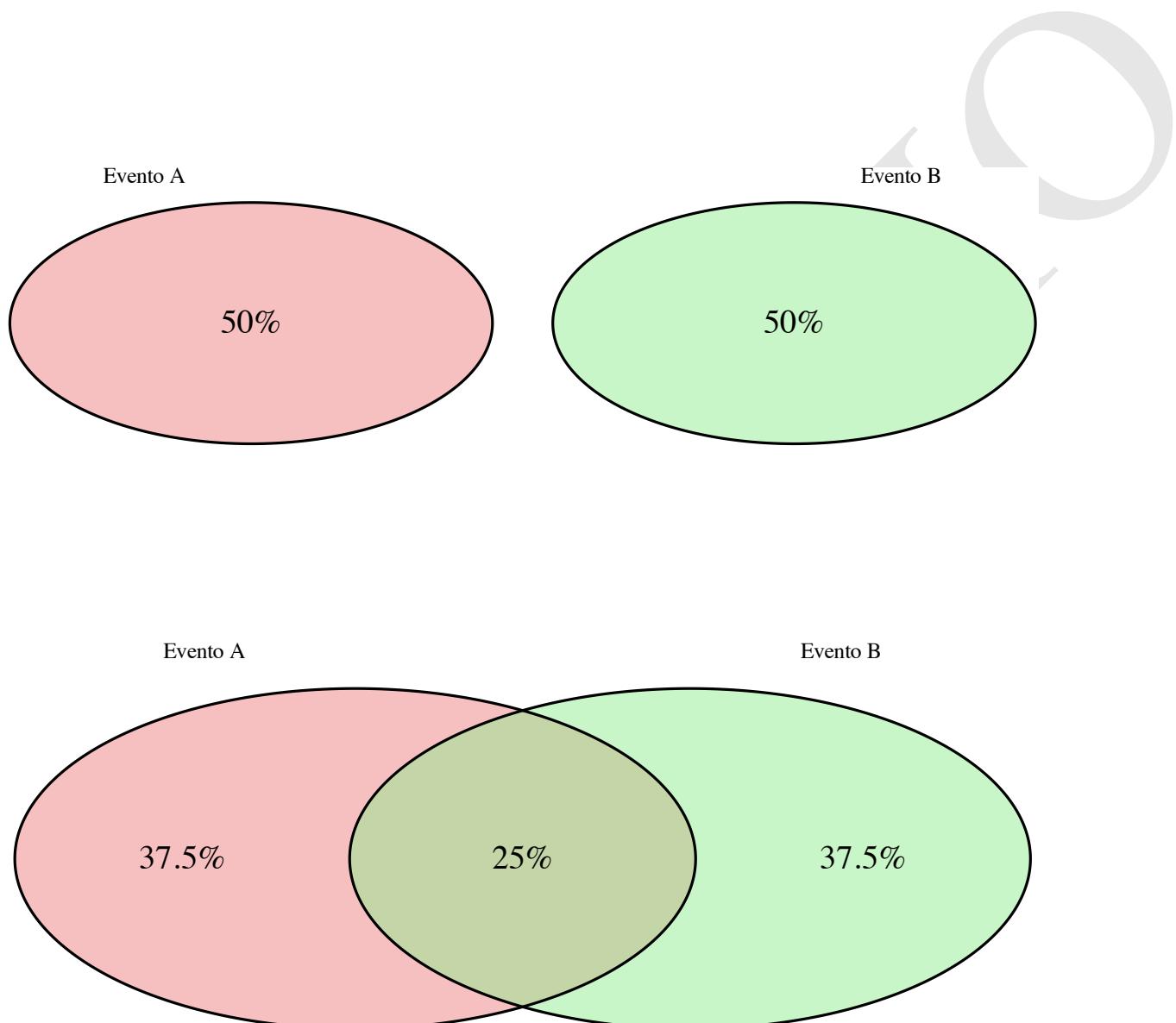


Figura 1.4: Superior: Eventos independentes. Inferior: Eventos dependentes.

- 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

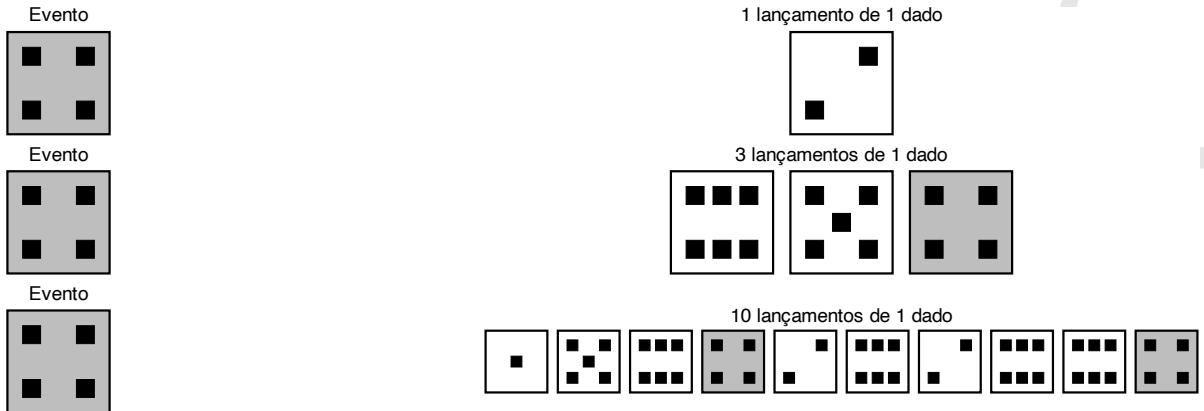


Figura 1.5: Esquerda: Evento (face = 4). Direita: Experimentos de 1 lançamento de 1 dado (superior), 3 lançamentos de 1 dado (central), 10 lançamentos de 1 dado (inferior).

1.5.2 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 ($E_1 \cup E_2 = \emptyset$), esses dois eventos E_1 e E_2 são mutuamente exclusivos ou disjuntos.¹

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

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:

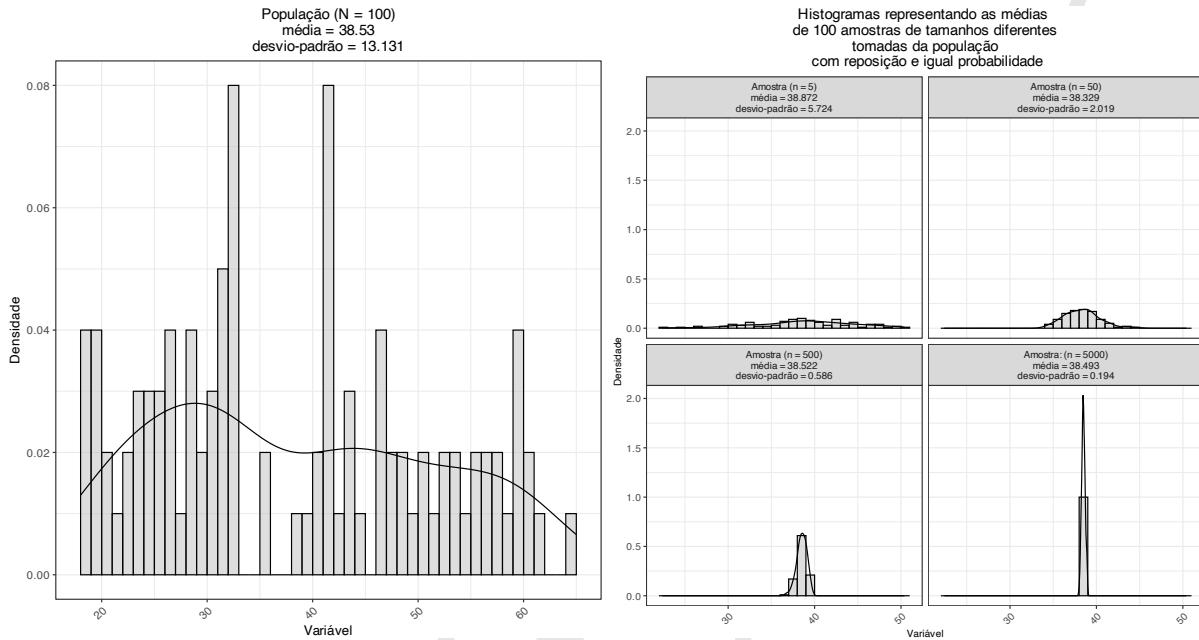


Figura 1.6: Esquerda: Histogramas de uma variável aleatória com distribuição uniforme ($N = 100$). Direita: Histogramas da média de 100 amostras de tamanhos 5, 50, 500 e 5000 tomadas da população com reposição e igual probabilidade.

- 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;
- Mais um exemplo, o lançamento de uma moeda com distribuição uniforme no espaço amostral $S = \{0, 1\}$ — codificado para *sucesso* = 1 e *insucesso* = 0 — tem média $\mu = 0.59$ e variância $\sigma^2 = 0.244$. 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 $\mu = 0.591$ e variância $\sigma^2 = 0$, independentemente da distribuição da população;

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, $n \geq 30$).

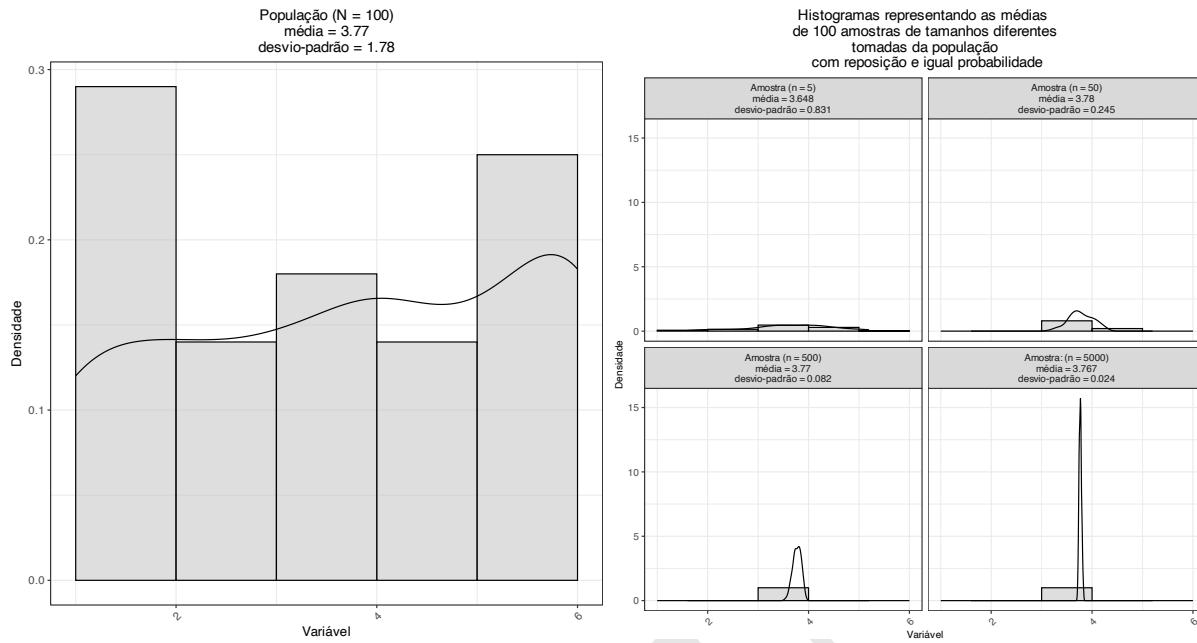


Figura 1.7: Esquerda: Histogramas de lançamento de 1 dado com distribuição uniforme ($N = 100$). Direita: Histogramas da média de 100 amostras de tamanhos 5, 50, 500 e 5000 tomadas da população com reposição e igual probabilidade.

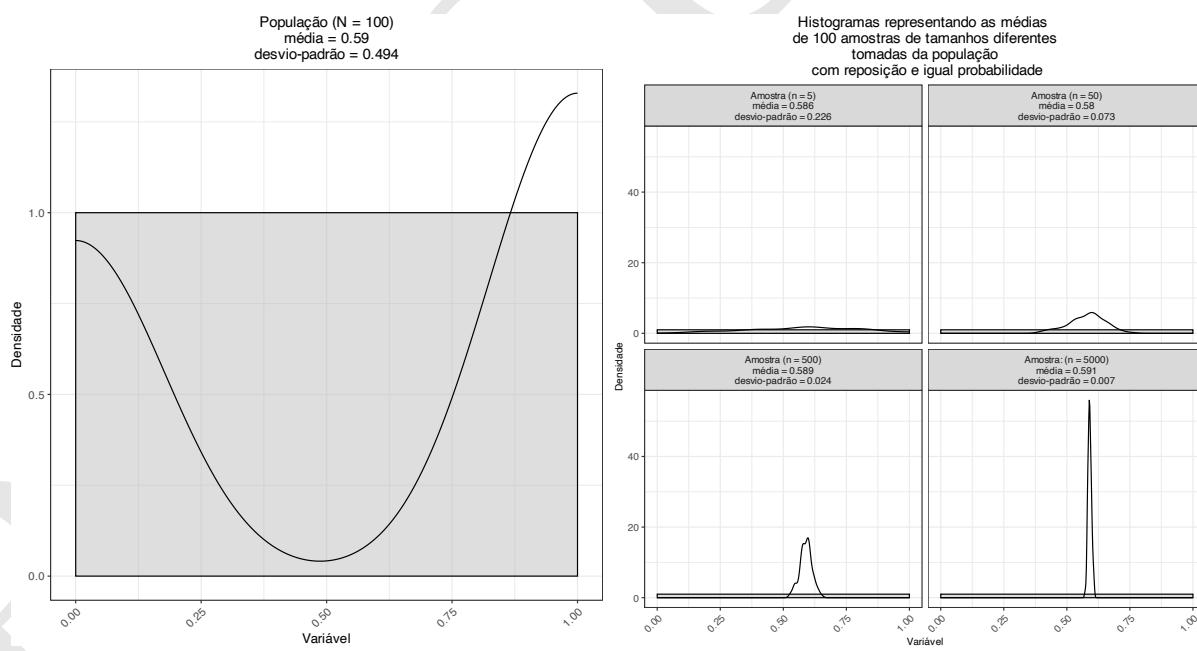


Figura 1.8: Esquerda: Histogramas de lançamento de 1 moeda com distribuição uniforme ($N = 100$). Direita: Histogramas da média de 100 amostras de tamanhos 5, 50, 500 e 5000 tomadas da população com reposição e igual probabilidade.

1.9.3 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.⁷

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

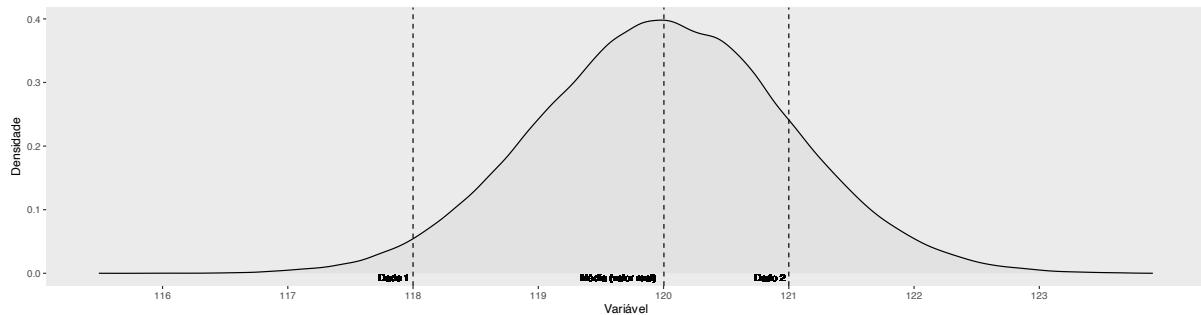


Figura 1.9: Representação gráfica da regressão para a média em medidas repetidas. A segunda medida (dado 2) é mais próxima da média (valor real) do que a primeira medida (dado 1).

- 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

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

R

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

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.

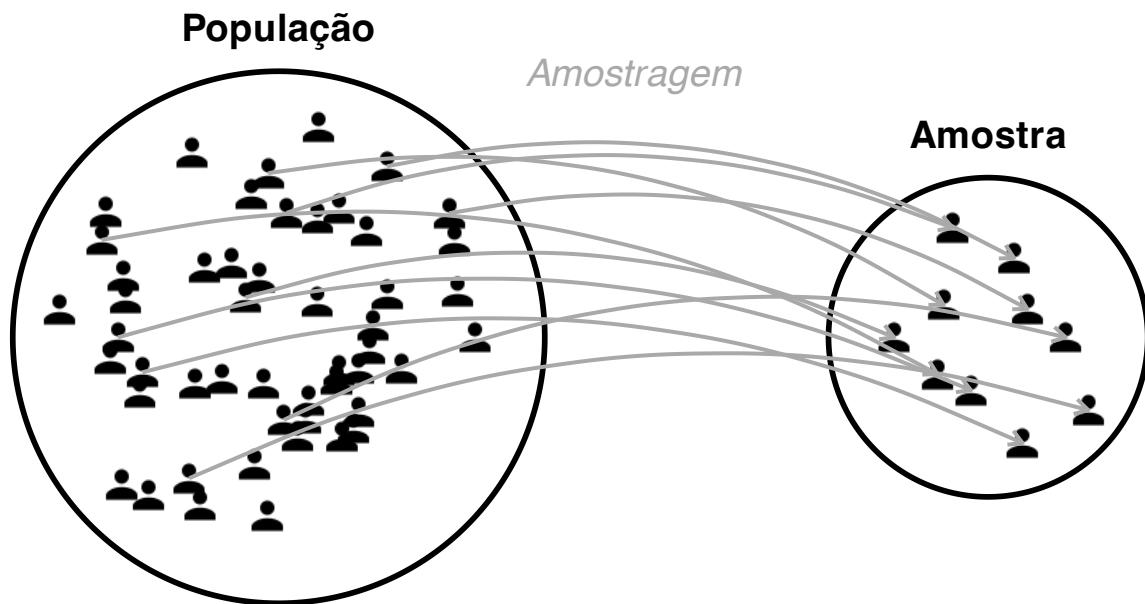


Figura 2.1: Representação esquemática da amostragem: seleção de uma população para a amostra.

- 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

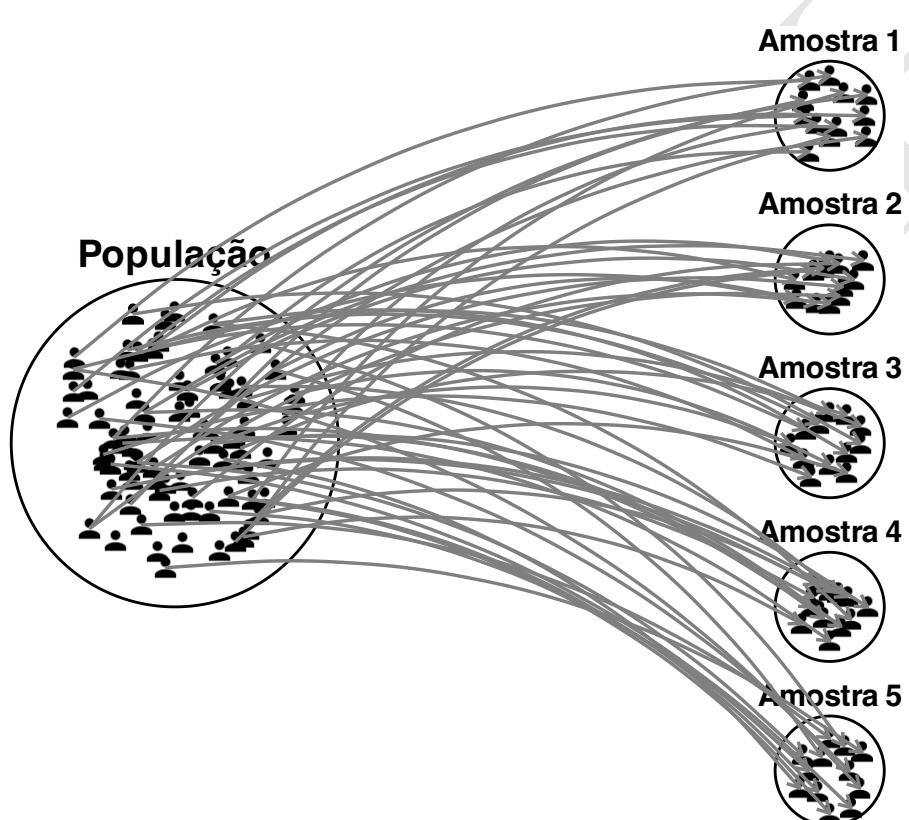


Figura 2.2: Representação esquemática do erro de amostragem: seleção de várias amostras independentes de uma população.

Mesma população → amostras diferentes
Cada amostra tem sua própria média

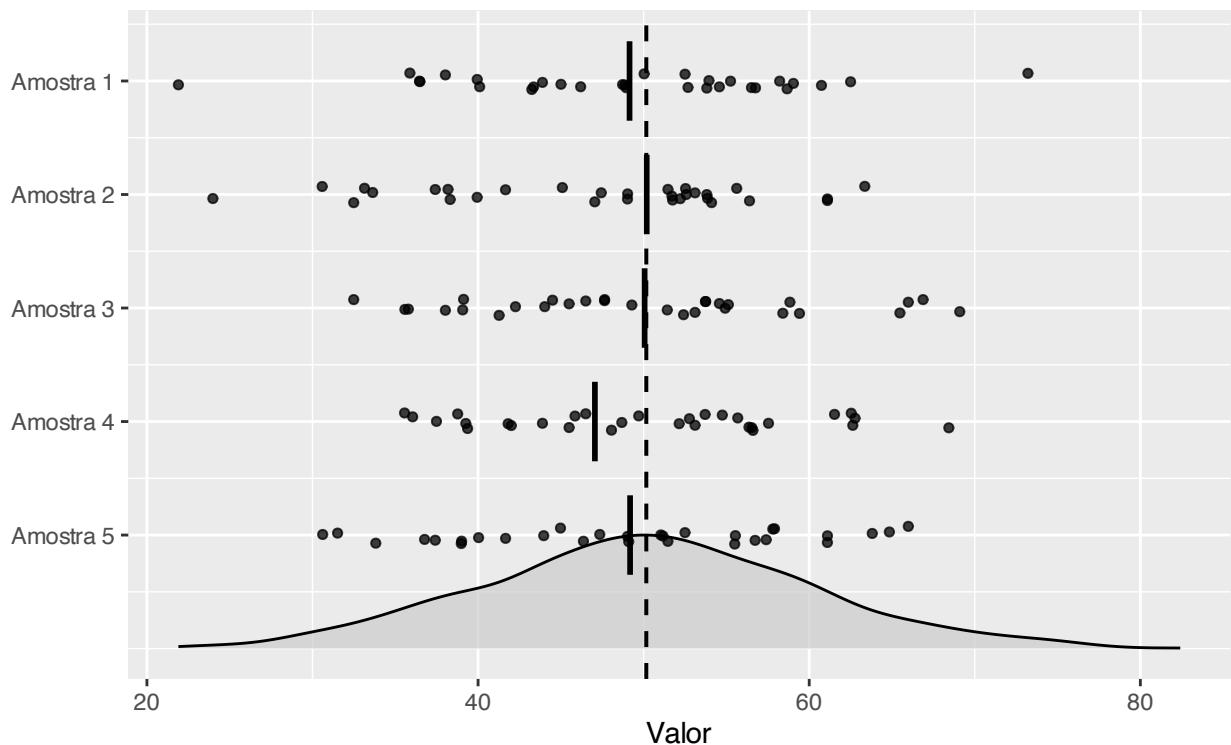


Figura 2.3: Representação esquemática da amostragem de uma população para a amostra.

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

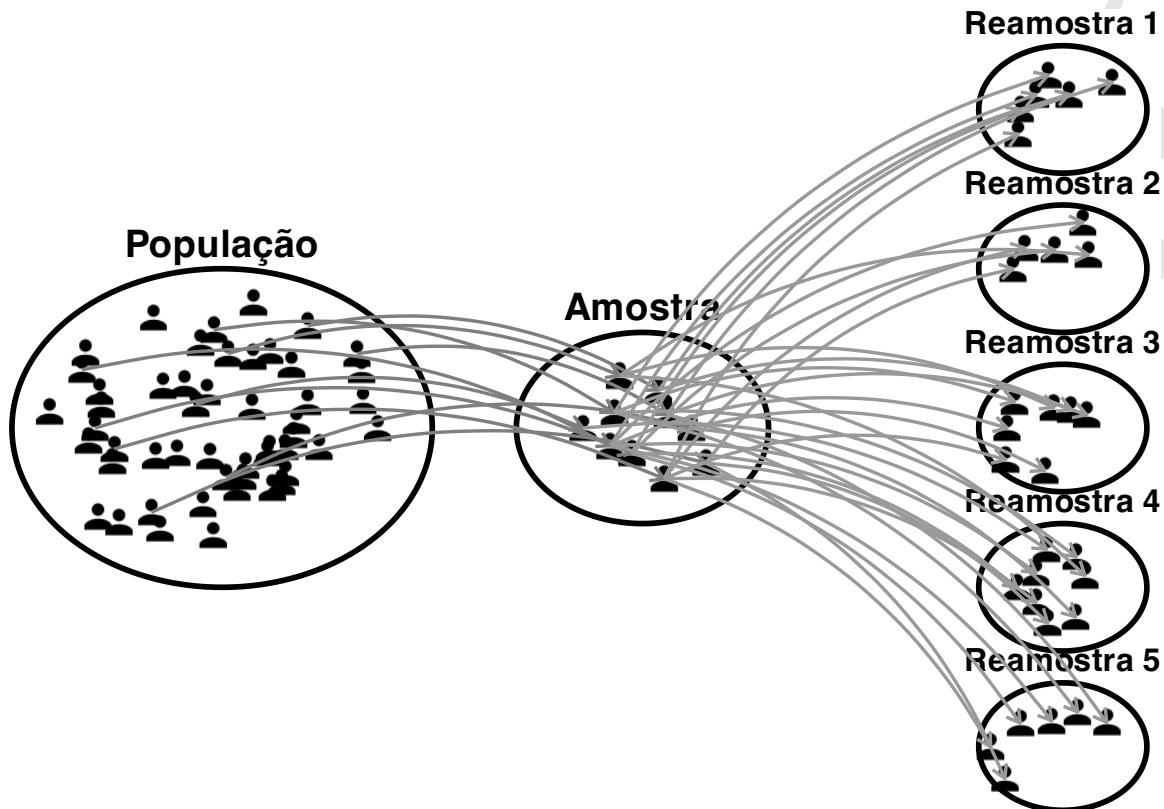


Figura 2.4: Representação esquemática da reamostragem de uma amostra.

2.6 Subamostragem

2.6.1 O que é subamostragem?

- ?

2.7 Superamostragem

2.7.1 O que é superamostragem?

- ?

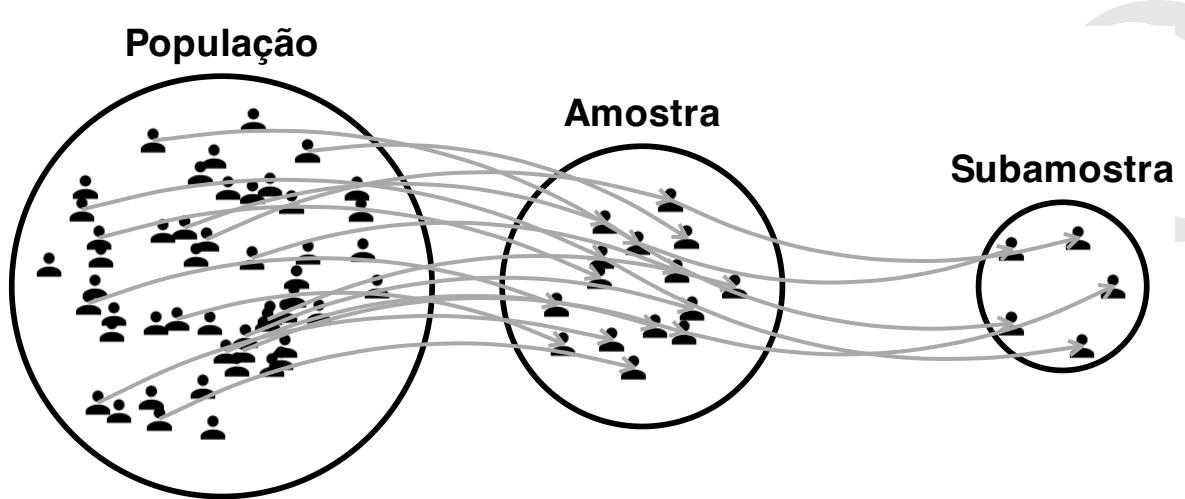


Figura 2.5: Representação esquemática da subamostragem de uma amostra.

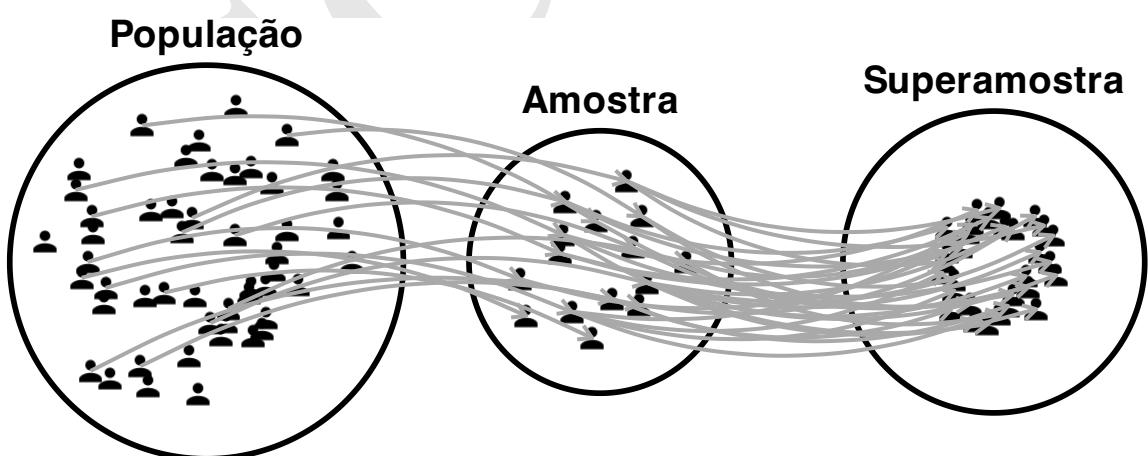


Figura 2.6: Representação esquemática da superamostragem de uma população.

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 diferencia pesquisa qualitativa da quantitativa?

- A divisão entre quantitativo e qualitativo é amplamente usada, mas é considerada por muitos autores como superficial ou imprecisa.²¹
- 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

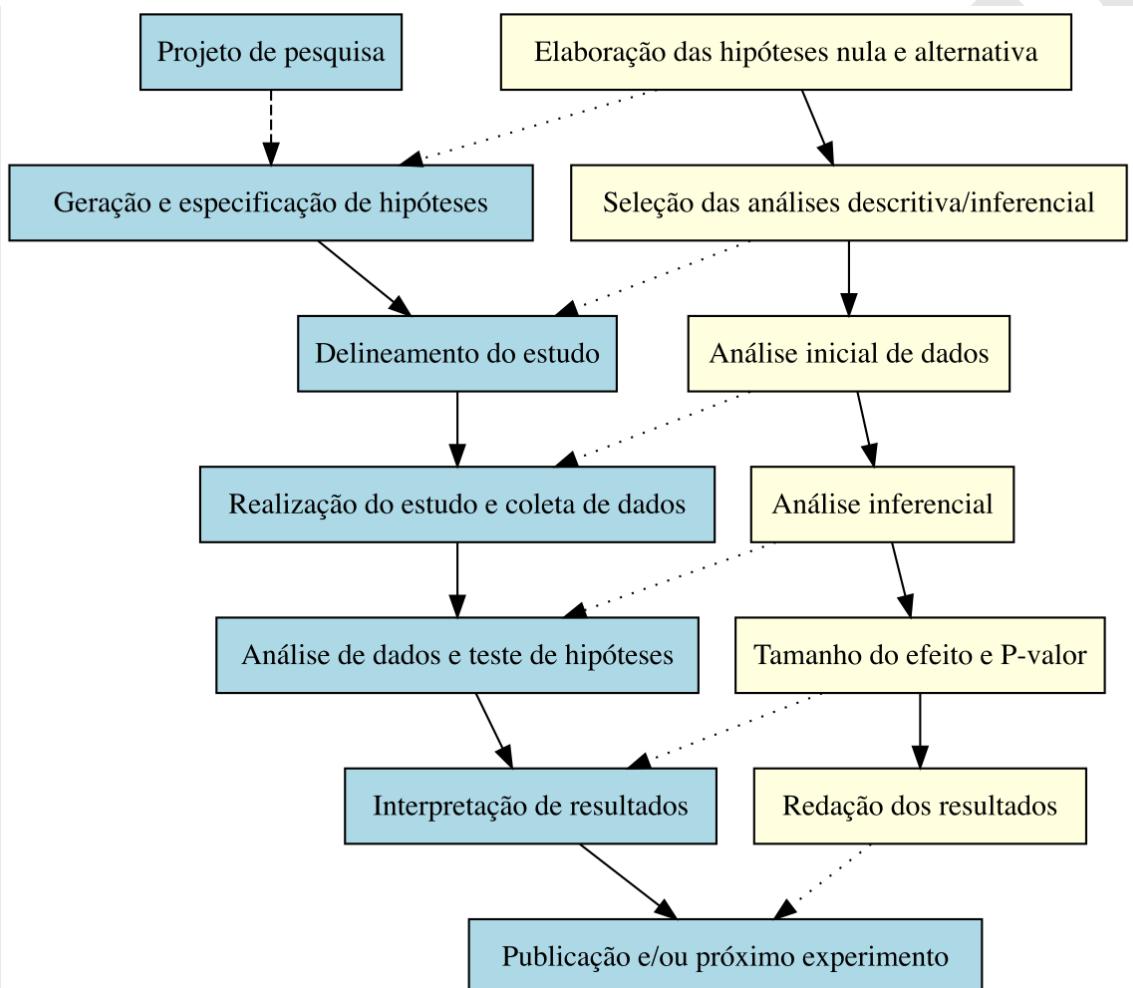


Figura 3.1: Mapa mental da relação entre o pensamento estatístico e o pensamento metodológico.

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
- A adoção de métodos mistos não deve decorrer de uma busca por completude ou sofisticação aparente, mas de coerência entre a pergunta de pesquisa, a posição epistemológica do pesquisador e as técnicas empregadas.²³

3.4.2 Métodos mistos significam misturar paradigmas epistemológicos?

- Não necessariamente. Combinações no nível paradigmático não constituem fusões reais de visões de mundo, pois paradigmas envolvem pressupostos ontológicos e epistemológicos distintos e, muitas vezes, incompatíveis.²³
- A combinação ocorre concretamente no nível técnico da pesquisa — isto é, na articulação de estratégias de amostragem, coleta e análise de dados — e não como fusão de ontologias ou epistemologias.²³
- Técnicas não são inherentemente vinculadas a paradigmas específicos; o que varia é a orientação do pesquisador e a forma como essas técnicas são empregadas e interpretadas.²³

3.4.3 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.²⁴
- 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).²⁴

3.4.4 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.²⁴

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.²⁵
- 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.²⁵

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).²⁵
- Análises exploratórias não são inherentemente inferiores: quando bem justificadas e comparando explicações alternativas, podem aumentar a rigorosidade do teste e produzir inferências informativas.²⁵
- 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 falsificação metodológico.²⁶

3.5.3 Quais são as boas práticas de transparência?

- Rotular claramente quais análises são confirmatórias e quais são exploratórias.²⁵
- Pré-Registrar hipóteses e plano confirmatório; documentar desvios e justificá-los.²⁵
- Relatar análises de sensibilidade (modelos alternativos, decisões analíticas razoáveis) para avaliar robustez.²⁵
- Disponibilizar dados e código sempre que eticamente possível, distinguindo scripts confirmatórios de scripts exploratórios.²⁵

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.²⁷
- O pré-registro é um elemento central das práticas abertas que aumentam a severidade de testes estatísticos e reduzem a flexibilidade analítica oportunista.²⁶

3.8 Reprodutibilidade e Ciência Aberta

3.8.1 O que é reproduzibilidade?

- Reproduzibilidade é a habilidade de se obter resultados iguais ou similares quando uma análise ou teste estatístico é repetido.²⁷⁻²⁹

3.8.2 Por que reproduzibilidade é importante?

- Analisar a reproduzibilidade pode fornecer evidências a respeito da objetividade e confiabilidade dos achados, em detrimento de terem sido obtidos devido a vieses ou ao acaso.²⁷
- A reproduzibilidade não é apenas uma questão metodológica, mas também ética, uma vez que pode envolver mal práticas científicas como fabricação e/ou falsificação de dados.²⁷
- Reproduzibilidade pode ser considerada um padrão mínimo em pesquisa científica.²⁸

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.²⁹
- 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.³⁰
- Adote delineamentos confirmatórios e justifique previamente hipóteses e tamanhos de efeito.²⁶
- Um plano de análise estatística bem definido contribui para a transparência, a reproduzibilidade e a credibilidade da inferência científica, reduzindo a flexibilidade analítica pós-hoc.^{31,32}

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.^{28,29}

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

3.11.2 O que é generalização estatística?

- É o modelo clássico em que inferências são feitas da amostra para a população-alvo, idealmente por meio de amostragem probabilística e inferência estatística.³³
- Esse modelo raramente é plenamente atendido na prática, pois a maioria dos estudos utiliza amostras de conveniência e populações acessíveis.³³

3.11.3 O que é generalização analítica?

- É o processo de extrapolar achados empíricos para conceitos, categorias ou teorias mais amplas, por meio de abstração e interpretação rigorosa.³³
- Esse modelo é especialmente relevante na pesquisa qualitativa, mas também ocorre em estudos quantitativos orientados por teoria.³³

3.11.4 Quais são os mitos mais comuns sobre generalização?

- Mito 1: Estudos quantitativos sempre utilizam amostragem aleatória.³³
- Mito 2: Pesquisa qualitativa não pode gerar generalizações.³³
- Mito 3: Transferibilidade depende apenas do leitor.³³

3.11.5 Generalizações são conclusões definitivas?

- Não, toda generalização deve ser entendida como uma “hipótese de trabalho”, sujeita a novas verificações em diferentes contextos.³³
- A generalização existe em um continuum — a questão não é “se” podemos generalizar, mas “quanto” podemos generalizar.³³

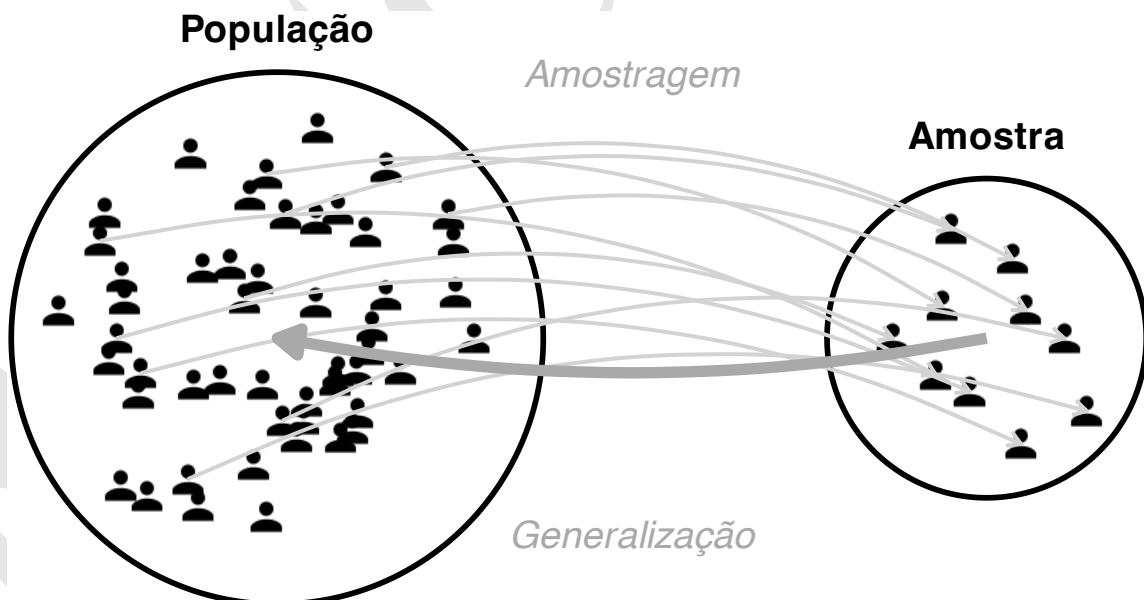


Figura 3.2: Representação esquemática da generalização de uma amostra para a população.

3.11.6 Como a replicação contribui para a generalização?

- Replicações deliberadas — em diferentes contextos, populações e momentos — fortalecem a robustez das inferências.³³
- A replicação pode ocorrer tanto na amostragem (variação deliberada de casos) quanto na repetição de estudos independentes.³³

3.11.7 Como revisões sistemáticas e meta-análises ampliam a generalização?

- A integração de evidências de múltiplos estudos permite examinar padrões consistentes e limites das inferências, fortalecendo conclusões generalizáveis.³³
- Metassínteses qualitativas podem produzir abstrações conceituais mais robustas do que estudos isolados.³³

RASCUNHO

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.^{34,35}
- R version 4.5.2 (2025-10-31) está disponível gratuitamente em *Comprehensive R Archive Network* (CRAN).³⁶

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).^{29,36}

4.1.3 O que é R Markdown?

- R Markdown³⁷ 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 RMarkdown³⁷ 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.³⁸
- O RMarkdown³⁷ 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.²⁹

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

- JASP¹.³⁹
- jamovi².⁴⁰



Os pacotes *jmv*⁴¹ e *jmvconnect*⁴² 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.^{35,43}
- O ambiente do RStudio é dividido em painéis:
 - *Source/Script editor*: para edição de R scripts.³⁵
 - *Console*: para execução de códigos simples.³⁵
 - *Environments*: para visualização de objetos criados durante a sessão de trabalho.³⁵
 - *Output*: para visualização de gráficos criados durante a sessão de trabalho.³⁵
- 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.^{35,43}
- RStudio está disponível gratuitamente em Posit³.



O pacote *learnr*⁴⁴ fornece tutoriais interativos para RStudio.

4.3 Scripts

4.3.1 O que são R scripts?

- “Scripts são dados”.³⁰
- 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.³⁰
- 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.⁷

```
# Exemplo de R script
```

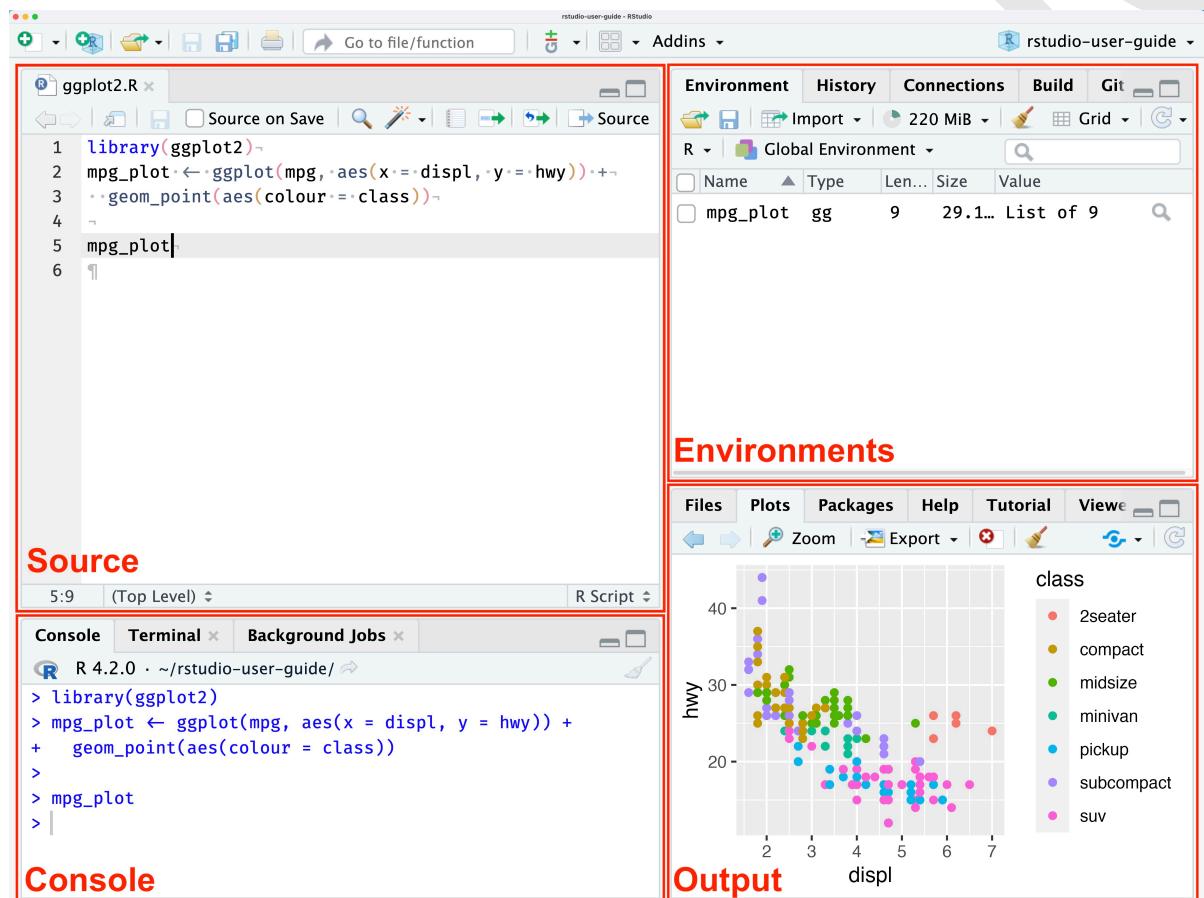
```
# Este é um comentário
```

```
# 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.⁴⁵
- Defina os tipos de variáveis adequadamente no banco de dados.⁴⁵
- Defina constantes — isto é, variáveis de valor fixo — ao invés de digitar valores.⁴⁵
- Use e cite os pacotes disponíveis para suas análises.⁴⁵
- Controle as versões do script.^{45,46}
- Teste o script antes de sua utilização.⁴⁵
- Conduza revisão por pares do código durante a redação (digitação em dupla).⁴⁵



O pacote *formatR*⁴⁷ 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*⁴⁸ 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*⁴⁹ 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.³⁵
- Os pacotes ficam armazenados no *Comprehensive R Archive Network* (CRAN) e podem ser instalados diretamente no RStudio.^{35,36}
- Na mais recente atualização deste livro, o [Comprehensive R Archive Network (CRAN) possui 394859 pacotes disponíveis.^{35,36}
- Os pacotes disponíveis podem ser encontrados em *R PACKAGES DOCUMENTATION*.⁵⁰



O pacote *utils*⁵¹ 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*⁵¹ 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*⁵¹ 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*⁵¹ 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*⁵¹ 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*⁵² 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.²⁸

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.³⁰

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.²⁹

- 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.²⁹
- Muitos erros de análise poderiam ser evitados com a adoção de boas práticas de programação em manuscritos reproduzíveis.⁵³

R

O pacote *rmarkdown*³⁷ 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*⁵⁴ 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*⁵⁵ 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.⁵⁶

R

O pacote *projects*⁵⁷ 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*³⁷ 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*⁵⁵ 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.⁵⁸

4.7.2 O que pode ser compartilhado?

- Idealmente, todos os scripts, pacotes/bibliotecas e dados necessários para outros reproduzirem seus dados.⁴⁶
- Minimamente, partes importantes incluindo implementações de novos algoritmos e dados que permitam reproduzir um resultado importante.⁴⁶

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).⁴⁶
- Inclua a versão dos pacotes usados no seu script por meio de um script inicial para instalação de pacotes (ex.: ‘instalar.R’).⁵³
- 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.²⁸
- Use endereços de arquivos relativos.⁵³
- Crie links persistentes para versões do seu script.⁴⁶
- Defina uma semente para o gerador de números aleatórios em scripts com métodos computacionais que dependem da geração de números pseudoaleatórios.²⁸

R

O pacote *base*⁵⁹ 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.⁴⁶
- Teste o script em uma nova sessão antes de compartilhar.⁵³

- Cite todos os pacotes relacionados à sua análise.⁶⁰



O pacote *utils*⁵¹ 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*⁶¹ 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.⁵³



O pacote *utils*⁵¹ 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.²⁸
- Autores do trabalho.²⁸
- Principais responsáveis pela escrita do script e quaisquer outras pessoas que fizeram contribuições substanciais para o desenvolvimento do script.²⁸
- Endereço de e-mail do autor ou contribuidor a quem devem ser direcionadas dúvidas, comentários, sugestões e bugs sobre o script.²⁸
- Lista de configurações nas quais o script foi testado, tais com nome e versão do programa, pacotes e plataforma.²⁸

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.⁶²
- 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.⁶²
- 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.^{63–65}
- 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).⁶²

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.⁶⁶
- 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.⁶⁶

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).⁶²
- Confusão entre taxas e contagens: comparar números absolutos sem considerar proporções populacionais pode distorcer a realidade.⁶²
- Fatores de confusão: diferenças observadas podem ser explicadas por variáveis não consideradas, como idade média da população ao comparar taxas de mortalidade.⁶²

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.^{63–65}

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.⁶³
- Conhecimento estatístico básico, incluindo conceitos, métodos, interpretação de dados e probabilidade.⁶³
- Conhecimento matemático sobre percentagens, médias e raciocínio numérico.⁶³
- Conhecimento de contexto/mundo, com entendimento do cenário e origem dos dados.⁶³
- Questões críticas (lista de *worry questions* para avaliar a validade da informação).⁶³

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.⁶³
- Crenças e atitudes: visão positiva sobre a capacidade de pensar estatisticamente; valorização de dados bem produzidos.⁶³

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?⁶³
- A amostra é representativa e suficientemente grande?⁶³
- Os instrumentos de medição são confiáveis?⁶³
- As estatísticas e gráficos são apropriados e não distorcem?⁶³
- Há relação causal ou apenas correlação? Há informação em falta?⁶³
- Existem interpretações alternativas plausíveis?⁶³

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.⁶⁷
- 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.⁶⁷
- 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.⁶⁷

- 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.⁶⁷
- 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.⁶⁷
- 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.⁶⁷

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.⁶⁶
- Saber interpretar dados de amostras aleatórias e não aleatórias, avaliando limitações e potencial.⁶⁶
- Conhecer e aplicar práticas de proteção de dados e direitos de propriedade sobre informações coletadas.⁶⁶
- Produzir representações descritivas (tabelas, gráficos, mapas, dashboards) para responder perguntas sobre fenômenos reais.⁶⁶
- Reconhecer a importância da proveniência e do armazenamento dos dados, bem como a necessidade de pré-processamento antes da análise.⁶⁶
- Entender fundamentos de modelagem preditiva e algoritmos, como árvores de classificação e regressão, especialmente no contexto de dados massivos (*big data*).⁶⁶

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.⁶⁸
- 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.⁶⁸
- 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.⁶⁸
- Esclarecer: Entender e explicar palavras técnicas e expressões que podem confundir, além de saber como foi feita a pesquisa e a análise.⁶⁸
- Avaliar: Decidir se a afirmação é confiável comparando com outras informações disponíveis e verificando se faz sentido.
- 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.⁶⁸
- Considerar alternativas: Pensar em outras explicações possíveis ou diferentes interpretações para os mesmos dados.⁶⁸
- Concluir: Chegar à sua própria conclusão sobre o assunto, usando as informações e o raciocínio de forma clara e bem fundamentada.⁶⁸

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*.⁶⁹
- *RoB 2: A Revised Tool for Assessing Risk of Bias in Randomized Trials*.⁷⁰
- *AMSTAR 2: A Critical Appraisal Tool for Systematic Reviews that Include Randomised or Non-Randomised Studies of Healthcare Interventions*⁷¹
- *ROBINS-I: A Tool for Assessing Risk of Bias in Non-randomized Studies of Interventions*.⁷²
- *ROBIS: A New Tool to Assess Risk of Bias in Systematic Reviews*⁷³
- *QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies*⁷⁴

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.⁷⁴
- Falácia podem ser intencionais ou não intencionais, e podem ser usadas para manipular, enganar ou confundir as pessoas.⁷⁵
- 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.⁷⁶

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.⁷⁵

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.⁷⁵

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.⁷⁶

8.1.2 O que é o paradoxo de Abelson?

- Um baixo percentual de variância explicada não implica que o fator causal seja irrelevante. Em processos cumulativos, efeitos pequenos podem produzir consequências grandes.⁷⁷
- Esse paradoxo alerta contra o uso ingênuo de medidas como R^2 ou “% de explicação” para julgar a importância prática de um fator.⁷⁸

8.1.3 O que é o paradoxo de Berkson?

- ⁷⁹

8.1.4 O que é o paradoxo de Ellsberg?

- ⁸⁰

8.1.5 O que é o paradoxo de Freedman?

- ⁸¹

8.1.6 O que é o paradoxo de Hand?

- ⁸²

8.1.7 O que é o paradoxo de Kelley?

- ⁸³

8.1.8 O que é o paradoxo de Lindley?

- ⁸⁴

Tabela 8.1: **Modelo micro (por tentativa): Desempenho ~ Habilidade**

Variável explicativa	Estimativa	IC 95% ¹	P-valor
Nível micro — eventos individuais			
Habilidade	1.527	-0.229, 3.282	0.088
R ² (variância explicada)	0.001		
Nível macro — resultado acumulado			
Habilidade	152.675	0.918, 304.432	0.049
R ² (variância explicada)	0.199		

¹IC = Intervalo de confiança

8.1.9 O que é o paradoxo de Lord?

- ^{84,85}

8.1.10 O que é o paradoxo de Meng?

- *Big Data*: “Quanto maior a quantidade de dados, maior a certeza de que vamos nos enganar”⁷⁶

8.1.11 O que é o paradoxo de Proebsting?

- ?

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 .^{86,87}
- 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 .⁸⁸
- 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.⁸⁸
- 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.⁸⁸

8.1.13 O que é o paradoxo de James-Stein?

- O paradoxo de James-Stein mostra que, ao estimar simultaneamente 3 ou mais médias de variáveis normais independentes (com perda quadrática), o estimador ‘óbvio’ X_i — que é ótimo para cada média isoladamente — deixa de ser ótimo no conjunto, existindo estimadores que têm erro médio total menor.^{89,90}
- O resultado é paradoxal porque essa melhoria exige ‘misturar’ as estimativas entre si (como no estimador de James-Stein), introduzindo um viés controlado que reduz o erro global, algo impossível quando $n \leq 2$.^{89,90}

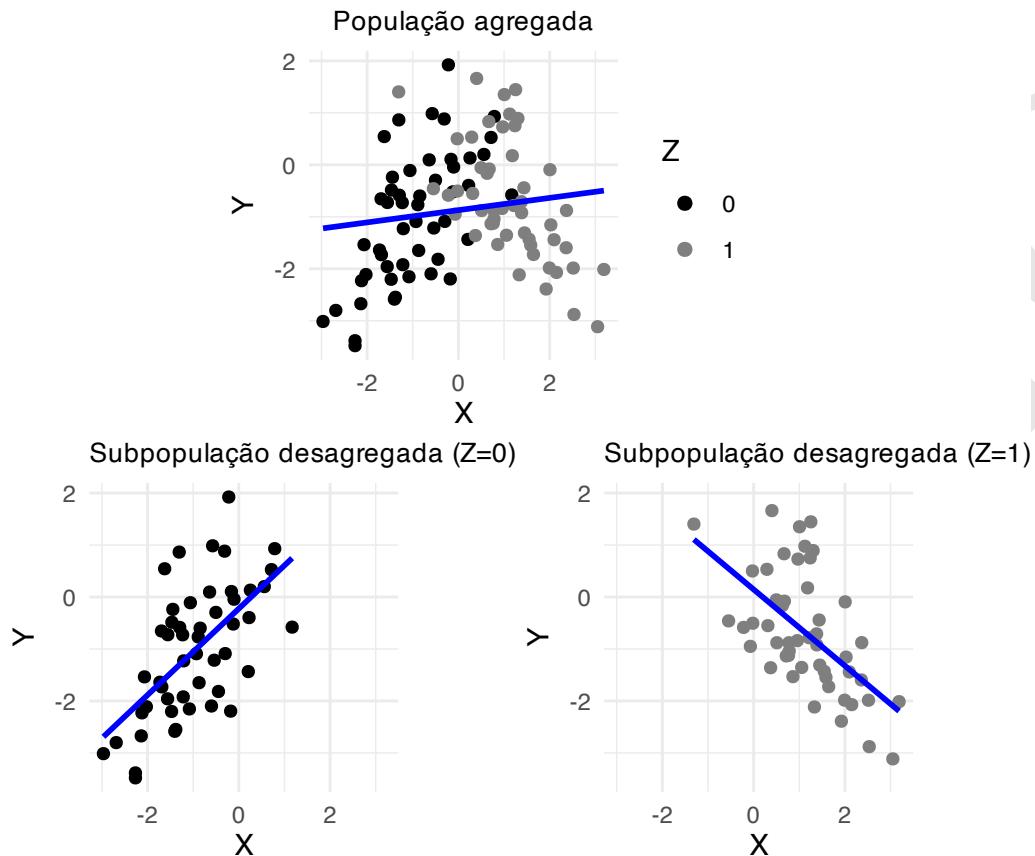


Figura 8.1: Paradoxo de Simpson representado com dados simulados. Os pontos no gráfico representam observações individuais e as linhas de tendência representam as regressões lineares ajustadas para os dados desagregados da população e agregados por subpopulação.

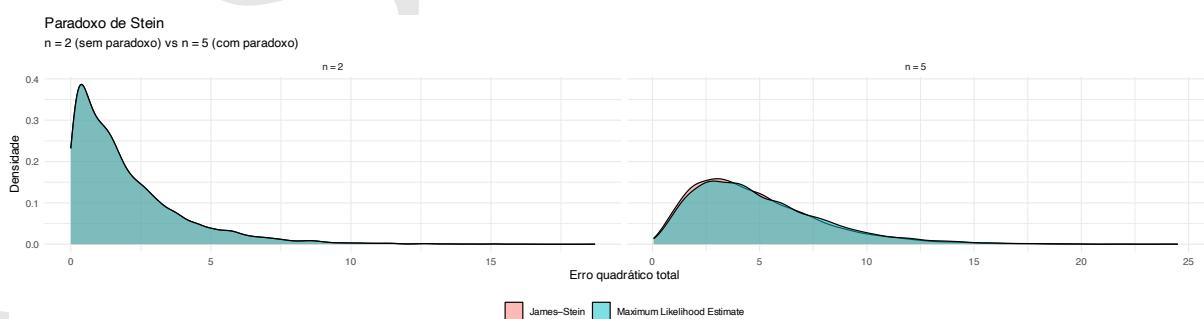


Figura 8.2: Simulação do paradoxo de Stein. Comparação do erro médio quadrático entre o estimador clássico (média amostral) e o estimador de James-Stein para diferentes números de médias sendo estimadas simultaneamente. Estimadores aparentemente piores localmente podem ser melhores globalmente quando o objetivo é reduzir o erro total.

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?

• ⁹¹

8.1.19 O que é o paradoxo da amizade?

• ⁹²

8.1.20 O que é o paradoxo do menino ou menina?

• ⁹¹

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?

• ?

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.^{93,94}

9.1.2 Por que práticas questionáveis em pesquisa devem ser combatidas?

- Práticas questionáveis em pesquisa são prevalentes.⁹⁵
- Práticas questionáveis em pesquisa comprometem a integridade científica, a confiabilidade dos resultados e a confiança do público na ciência.^{93,94}
- Práticas questionáveis em pesquisa inflam artificialmente o tamanho do efeito e poder estatístico.⁹⁴
- Práticas questionáveis em pesquisa parecem contribuir para a crise da replicação na ciência, onde muitos estudos não conseguem ser replicados ou reproduzidos.⁹⁴

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.⁹⁶
- 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.⁹⁷
- Práticas na zona cinzenta são aquelas que podem ser interpretadas de diferentes maneiras, dependendo do contexto e da intenção do pesquisador.⁹⁸

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.⁹⁷
- *Storytelling* é a prática de criar narrativas convincentes para justificar os resultados, mesmo que não sejam suportados pelos dados.⁹⁸

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.^{98–100}
- *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.¹⁰¹
- *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.¹⁰²
- *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.¹⁰³
- *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.⁹⁹
- *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.⁹⁹
- *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.¹⁰⁴

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.⁹⁹
- *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.^{105,106}
- *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.¹⁰⁷
- *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.¹⁰⁷
- *Gold authorship* é a prática de atribuir autoria em troca de prestígio, recursos ou favorecimento político, independentemente da contribuição acadêmica.¹⁰⁷
- *Fake authorship* refere-se à inclusão de autores fictícios ou inexistentes em uma publicação.¹⁰⁷

- *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.¹⁰⁸
- *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).^{109,110}
- 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.^a
- *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.^b
- *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.^b
- 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.^b



O pacote *retractcheck*¹¹¹ 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.¹¹²
- Variáveis definem características de uma amostra extraída da população, tipicamente observados por aplicação de métodos de amostragem (isto é, seleção) da população de interesse.¹¹³

10.1.2 Como são classificadas as variáveis?

- Quanto à informação:^{113–116}
 - Quantitativa
 - Qualitativa
- Quanto ao conteúdo:^{113–117}
 - 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:^{113–116}
 - Dependente (desfecho)
 - Independente (preditora, covariável, confundidora, controle)
 - Mediadora
 - Moderadora
 - Modificadora
 - Auxiliar
 - Indicadora

R

O pacote *base*⁵⁹ 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*⁵⁹ 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*⁵⁹ 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.¹¹⁵

10.2.2 O que é transformação de variáveis?

- Transformação significa aplicar uma função matemática à variável medida em sua unidade original.¹¹⁶
- A transformação visa atender aos pressupostos dos modelos estatísticos quanto à distribuição da variável, em geral a distribuição gaussiana.^{113,118}
- A dicotomização pode ser interpretada como um caso particular de agrupamento.¹¹⁹

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.¹²⁰
- 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).¹²⁰

10.2.4 Quais transformações de variáveis podem ser aplicadas?

- Distribuições com assimetria à direita: raiz quadrada, logaritmo natural, logaritmo base 10, transformação inversa.¹²⁰
- Distribuições com assimetria à esquerda: reflexão e raiz quadrada, reflexão e logaritmo natural, reflexão e logaritmo base 10, reflexão e transformação inversa.¹²⁰
- Transformação *z* de Fisher (10.1).[REF]

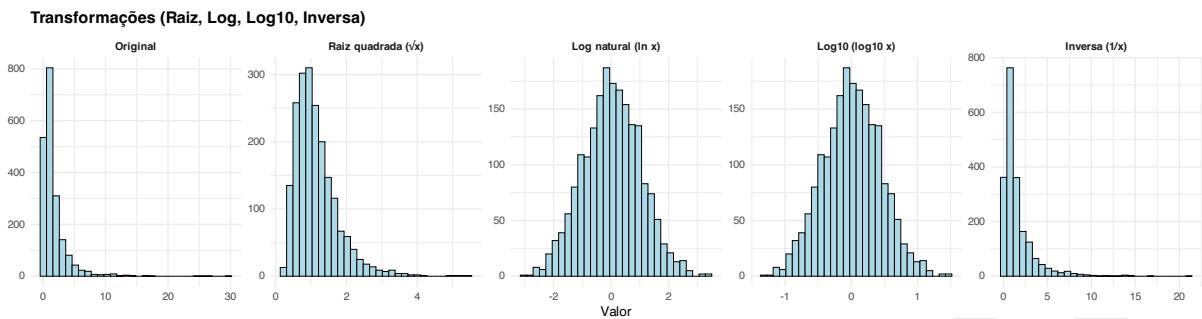


Figura 10.1: Transformações de variáveis com assimetria à direita (Original, Raiz quadrada, Log natural, Log10, Inversa).

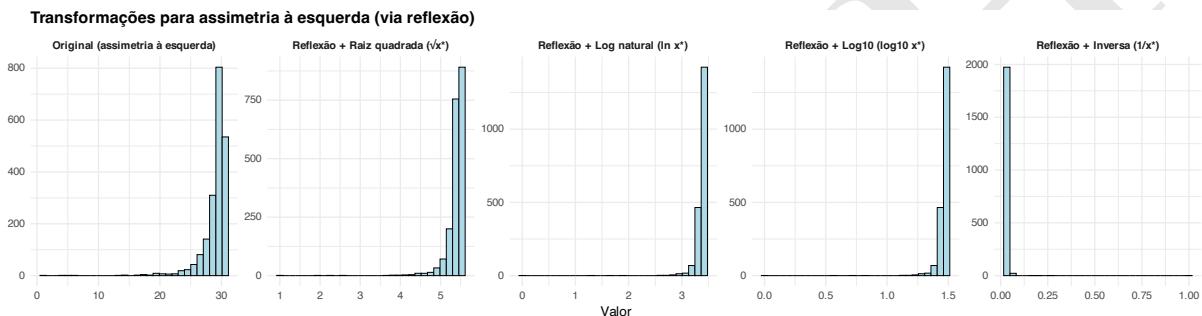


Figura 10.2: Transformações de variáveis com assimetria à esquerda (Original, Reflexão + Raiz quadrada, Reflexão + Log natural, Reflexão + Log10, Reflexão + Inversa).

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

- Transformação de Box-Cox (10.2).¹²¹

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

- Transformação arco-seno (10.3).¹²⁰

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

- Diferenciação.
- Categorização.
- Dicotomização.

O pacote *MASS*¹²² fornece a função *boxcox*^a para executar a transformação de Box-Cox.¹²¹

^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 recenterar a variável em torno de zero, sem alterar a sua variabilidade.⁵⁸

10.3.2 Por que centralizar?

- Facilita a interpretação dos coeficientes de regressão, especialmente em modelos com termos de interação.⁵⁹
- Reduz a multicolinearidade entre variáveis e seus termos de interação ou polinomiais.⁶⁰
- Mantém a escala original (apenas desloca a média).⁶¹

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.⁶²

10.4.2 Por que padronizar?

- Facilita a interpretação em análises multivariadas.⁶³
- Evita que variáveis em escalas maiores dominem os resultados de algoritmos que dependem de distância.⁶⁴
- Melhora a comparabilidade entre estudos e bases de dados diferentes.⁶⁵

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.⁶⁶

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

- Escala Min-Max (10.5): transformar para o intervalo [0,1].⁶⁷

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

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`).⁶⁸
- Documentar no dicionário de dados como cada variável foi transformada.⁶⁹
- Evitar substituir a variável original: manter sempre a versão bruta e a padronizada.⁷⁰

R O pacote `base`⁵⁹ 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>

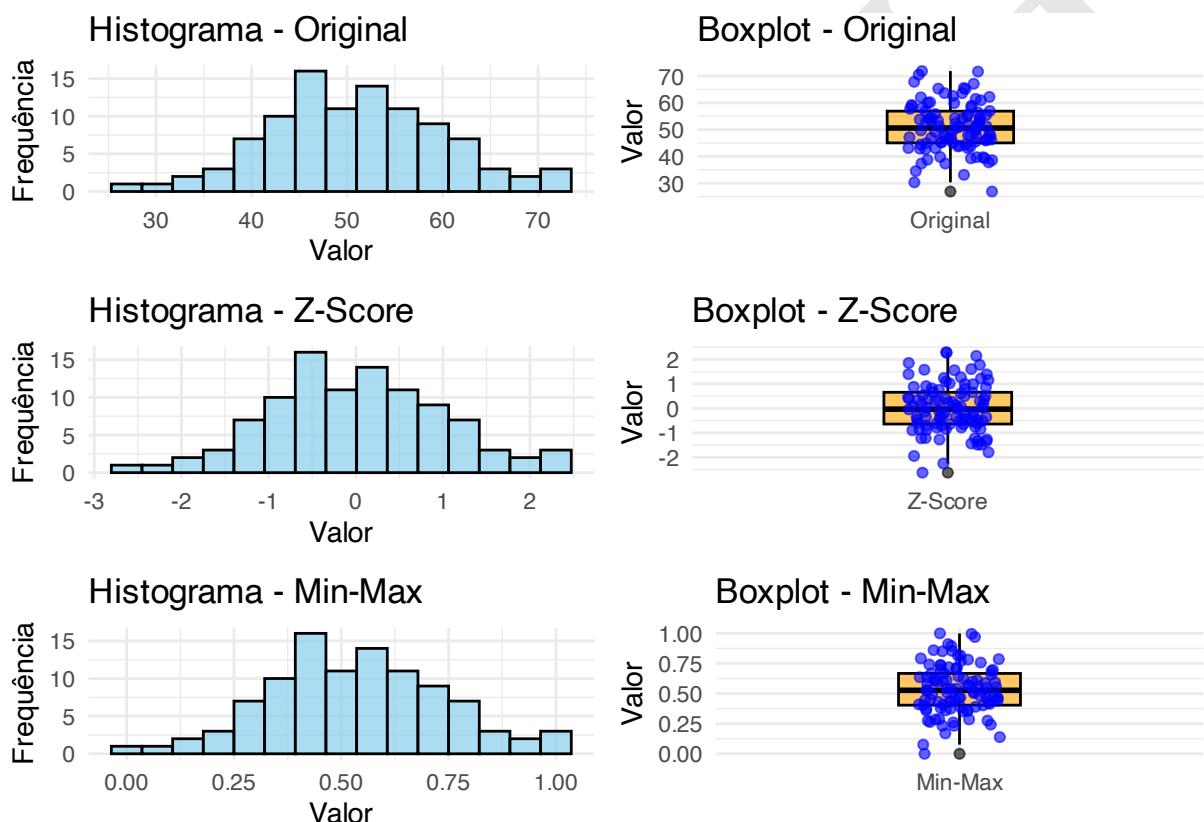


Figura 10.3: Comparaçāo entre variáveis originais e padronizadas (Z-score e Min-Max).

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.¹²³
- Categorizar variáveis não é necessário para conduzir análises estatísticas. Ao invés de categorizar, priorize as variáveis contínuas.^{124–126}
- Em geral, não existe uma justificativa racional (plausibilidade biológica) para assumir que as categorias artificiais subjacentes existam.^{124–126}
- 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.¹²⁷
- Categorização de variáveis contínuas aumenta a quantidade de testes de hipótese para comparações pareadas entre os quantis, inflando, portanto, o erro tipo I.¹²⁸
- 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.¹²⁸
- 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.¹²⁸

R

O pacote *questionr*¹²⁹ 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.¹²³
- Análise com modelos de regressão com pesos locais (*lowess*) tais como *splines* e polinômios fracionais.¹²³

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.¹³⁰
- 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.¹¹⁹
- Os pesquisadores não conhecem as consequências estatísticas da dicotomização.¹²³
- Os pesquisadores não conhecem os métodos adequados de análise não-paramétrica, não-linear e robusta.¹²³
- As categorias representam características existentes dos participantes da pesquisa, de modo que as análises devam ser feitas por grupos e não por indivíduos.¹²³
- A confiabilidade da(s) variável(eis) medida(s) é baixa e, portanto, categorizar os participantes resultaria em uma medida mais confiável.¹²³

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.¹²³
- Dicotomizar variáveis não é necessário para conduzir análises estatísticas. Ao invés de dicotomizar, priorize as variáveis contínuas.¹²⁴⁻¹²⁶
- Em geral, não existe uma justificativa racional (plausibilidade biológica) para assumir que as categorias artificiais subjacentes existam.¹²⁴⁻¹²⁶
- Dicotomização causa perda de informação e consequentemente perda de poder estatístico para detectar efeitos.^{123,124}
- Dicotomização também classifica indivíduos com valores próximos na variável contínua como indivíduos em pontos opostos e extremos, artificialmente sugerindo que são muito diferentes.¹²⁴
- Dicotomização pode diminuir a variabilidade das variáveis.¹²⁴
- Dicotomização pode ocultar não-linearidades presentes na variável contínua.^{123,124}
- A média ou a mediana, embora amplamente utilizadas, não são bons parâmetros para dicotomizar variáveis.^{119,124}
- Caso exista um ponto de corte ou limiar verdadeiro que discrimine dois grupos independentes, identificar tal ponto de corte ainda é um desafio.¹²⁷

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.¹²³
- Quando a distribuição da variável contínua é muito assimétrica, de modo que uma grande quantidade de observações está em um dos extremos da escala.¹²³

10.6.5 Quais métodos são usados para dicotomizar variáveis contínuas?

- Em termos de tabelas de contingência 2x2, os seguintes métodos permitem¹²⁷ a identificação do limiar verdadeiro:
 - Youden.¹³¹
 - Gini Index.¹³²
 - Estatística qui-quadrado (χ^2).¹³³

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).¹³⁴
- Kappa (κ).¹³⁵.

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.[?]

R O pacote *stats*¹³⁶ 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.⁵⁸
- Os coeficientes associados às variáveis indicadoras representam diferenças em relação a esse nível de referência.⁵⁹

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

- Utilizar k variáveis indicadoras gera redundância perfeita entre as variáveis.⁶⁰
- Essa redundância causa problemas de identificabilidade nos modelos, fenômeno conhecido como *dummy trap*.⁶¹

10.7.6 Variáveis indicadoras são uma forma de dicotomização?

- Variáveis indicadoras são dicotômicas, mas não resultam da dicotomização de variáveis contínuas.⁶²
- Variáveis indicadoras são criadas a partir de variáveis categóricas multinominais, preservando toda a informação original do fator.⁶³
- Variáveis indicadoras não reduzem informação, enquanto a dicotomização de variáveis contínuas descarta informação por construção.⁶⁴

10.7.7 Variáveis indicadoras alteram os dados originais?

- Não. Variáveis indicadoras apenas representam os níveis do fator de forma numérica.⁶⁵
- A variável categórica original permanece intacta no conjunto de dados.⁶⁶

10.8 Fatores

10.8.1 O que são fatores?

- Fator é um sinônimo de variável categórica.⁶⁷
- 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.⁶⁸
- 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.⁶⁹



O pacote *base*⁵⁹ 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.⁷⁰

R

O pacote *base*⁵⁹ 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>

Capítulo 11

Dados e metadados

11.1 Dados

11.1.1 O que são dados?

- “Tudo são dados”.¹³⁷
- Dados coletados em um estudo geralmente contêm erros de mensuração e/ou classificação, dados perdidos e são agrupados por alguma unidade de análise.¹³⁸

11.1.2 O que são dados estruturados?

- Dados estruturados são dados organizados em um formato tabular, como planilhas eletrônicas ou bancos de dados relacionais, onde cada linha representa uma observação e cada coluna representa uma variável ou atributo.[?]

11.1.3 O que são dados não estruturados?

- Dados não estruturados são dados que não possuem um formato ou organização predefinidos, como textos, imagens, vídeos, áudios e sinais biomédicos, tornando sua análise mais complexa em comparação com dados estruturados.[?]

11.2 *Big data*

11.2.1 O que são grandes dados (*big data*)?

- Grandes dados (*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.⁷⁶

11.2.2 Quais são as fontes de dados?

- Experimentos.[?]
- Mundo real.[?]
- Simulação.[?]

11.2.3 O que são dados primários e secundários?

- Dados primários são dados originais coletados intencionalmente para uma determinada análise exploratória ou inferencial planejada a priori.¹¹³
- Dados secundários compreendem dados coletados inicialmente para análises de um estudo, e são subsequentemente utilizados para outras análises.¹¹³

11.2.4 O que são dados quantitativos e qualitativos?

- ?

11.3 Metadados

11.3.1 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.¹³⁹
- Metadados também são informações relacionadas ao delineamento e/ou protocolo do estudo, recrutamento dos participantes, e métodos para realização das medidas.¹³⁹

11.3.2 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.¹⁴⁰
- Crie rótulos de variáveis concisos, claros e mutuamente exclusivos.¹⁴⁰
- Evite muitas letras maiúsculas ou outros caracteres especiais que usam a *shift*.¹⁴⁰
- Na existência de versões de instrumentos publicadas em diferentes anos, use o ano de publicação das escalas no rótulo.¹⁴⁰
- 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).¹⁴⁰

 O pacote *base*⁵⁹ 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*⁵⁹ 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*¹⁴¹ 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*¹⁴¹ 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*¹⁴² 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*¹⁴³ 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>

11.4 Armazenamento de dados

11.4.1 Como armazenar dados?

- Dados devem ser armazenados em três partes: dados brutos, *codebooks* e metadados ISO.¹⁴⁴
- Os dados brutos são organizados em colunas com um identificador único — o cabeçalho da coluna.¹⁴⁴
- *Codebooks* são documentos que descrevem os dados brutos, incluindo seus cabeçalhos, e respectivos rótulos, tipos de variável, formatos, unidades de medida, códigos para dados categóricos e dados perdidos e limites de valores plausíveis.¹⁴⁴
- Metadados são informações sobre o estudo, como título, palavras-chave e outras informações relevantes, tais como delineamento, protocolo, recrutamento dos participantes e métodos de medição¹⁴⁴

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.¹³⁰
- Em escalas grosseiras, erros são introduzidos porque as variações contínuas do constructo são colapsadas em uma mesma categorias ou separadas entre categorias próximas.¹³⁰
- 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”.¹³⁰



O pacote *likert*¹⁴⁵ 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>



O pacote *ggstats*¹⁴⁶ 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.¹³⁰

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.¹⁴⁷
- Esse processo tem como objetivo distinguir de maneira substantiva a manifestação observada de outras possíveis manifestações que também possam ser diferenciadas.¹⁴⁷

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*¹³⁶ 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*¹³⁶ 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>

12.2.8 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.⁷
- 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.



O pacote *stats*¹³⁶ 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>

12.3 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.¹⁴⁸
- 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.[?]

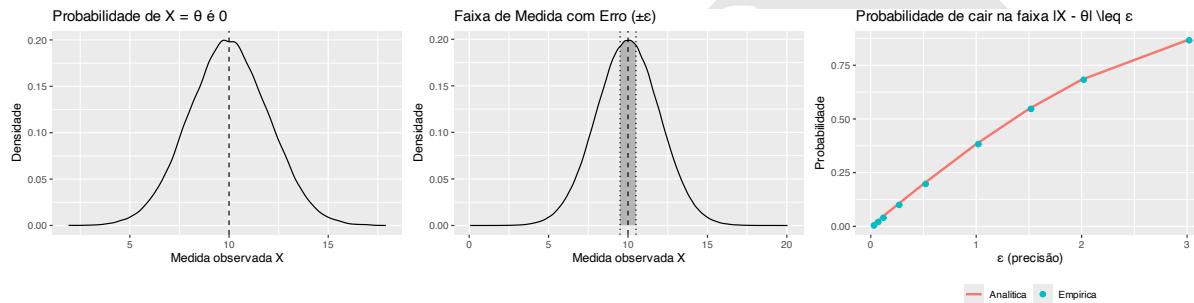


Figura 12.2: Erro de medida em um modelo simples com erro normal. A linha tracejada indica o valor verdadeiro (desconhecido na prática) A área sombreada representa a probabilidade de cair na faixa $|X - \theta| \leq \epsilon$, que é > 0 . A probabilidade de 'acertar no ponto' $X = \theta$ é 0.

12.3.2 Quais fontes de variabilidade são comumente investigadas?

- Intra/Entre participantes (isto é, unidades de análise).¹⁴⁹
- Intra/Entre repetições.¹⁴⁹
- Intra/Entre observadores.¹⁴⁹

12.4 Instrumentos

12.4.1 O que são instrumentos?

- ?

12.5 Acurácia e precisão

12.5.1 O que é acurácia?

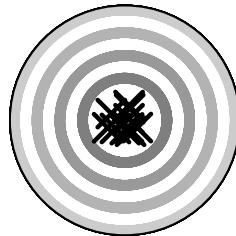
- Acurácia expressa a proximidade de concordância entre uma mensuração e o valor real.¹⁵⁰

- Acurácia está para medidas como validade está para instrumentos de medida.[?]

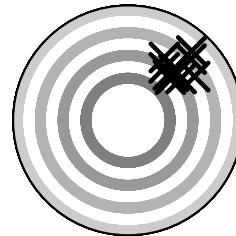
12.5.2 O que é precisão?

- Precisão se refere à proximidade de concordância entre resultados de testes independentes obtidos nas mesmas condições de teste.¹⁵⁰
- Precisão é um índice de quão próximo os resultados podem ser repetidos entre mensurações repetidas.¹⁵¹
- Precisão está para medidas como confiabilidade está para instrumentos de medida.[?]

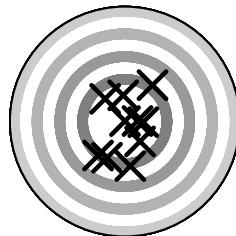
Acurácia alta, Precisão alta



Acurácia baixa, Precisão alta



Acurácia alta, Precisão baixa



Acurácia baixa, Precisão baixa

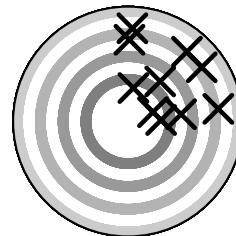


Figura 12.3: Acurácia e precisão como propriedades de uma medida.

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

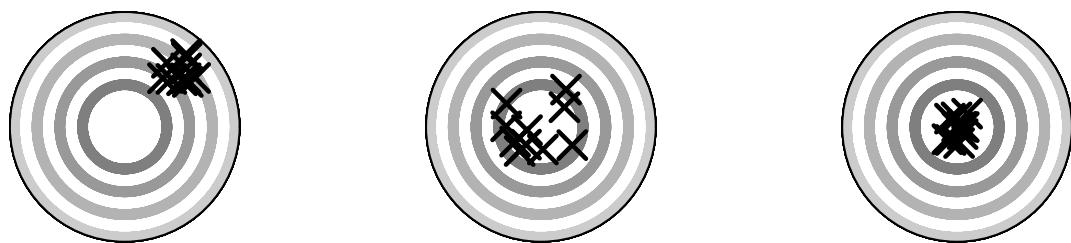


Figura 12.4: Viés e variabilidade de uma medida.

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).¹⁵¹
- Cada variável possui sua própria coluna (vertical).¹⁵²
- Cada observação possui sua própria linha (horizontal).¹⁵²
- Cada valor possui sua própria célula especificada em um par (linha, coluna).¹⁵²
- Cada célula possui seu próprio dado.¹⁵²



O pacote *DataEditR*¹⁵³ 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.¹⁵⁴
- Use apenas 1 (uma) linha de cabeçalho para nomear os fatores e variáveis do seu estudo.¹⁵⁴

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*).¹⁵⁵

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.^{154,155}
- Crie um dicionário de dados (metadados) em um arquivo separado contendo: nome da variável, descrição da variável, unidades de medida e valores extremos possíveis.¹⁵⁴
- Use recursos para validação de dados antes e durante a digitação de dados.^{154,155}

Tabela 13.4: Formatação recomendada para tabela de dados.

ID	Data.Coleta	Estado.Civil	Numero.Filhos
1	22-02-2026	casado	NA
2	23-02-2026	casado	1
3	24-02-2026	casado	NA
4	25-02-2026	solteiro	NA
5	26-02-2026	casado	NA
6	27-02-2026	solteiro	0
7	28-02-2026	solteiro	NA
8	01-03-2026	solteiro	NA
9	02-03-2026	casado	NA
10	03-03-2026	solteiro	NA

Tabela 13.5: Formatação não recomendada para tabela de dados.

ID	Data de Coleta	Estado Civil	Número de Filhos
1	22-02-2026	casado	NA
2	23-02-2026	Casado	1
3	24-02-2026	casado	NaN
4	25-02-2026	Solteiro	N/A
5	26-02-2026	Casado	N.A.
6	27-02-2026	solteiro	0
7	28-02-2026	solteiro	
8	01-03-2026	Solteiro	na
9	02-03-2026	casado	n.a.
10	03-03-2026	Solteiro	999



O pacote *data.table*¹⁵⁶ 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*]).¹⁵⁴
- Não inclua análises estatísticas ou gráficos nas tabelas de dados brutos.¹⁵⁴
- Não utilize cores como informação. Se necessário, crie colunas adicionais — variáveis instrumentais ou auxiliares — para identificar a informação de modo que possa ser analisada.¹⁵⁴
- 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.¹⁵⁷



O pacote *base*⁵⁹ 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.¹⁵⁷
- Perda de participantes no estudo por dados perdidos pode reduzir o poder estatístico (erro tipo II).¹⁵⁷
- Não existe solução globalmente satisfatória para o problema de dados perdidos.¹⁵⁷

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 são 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.^{158,159}

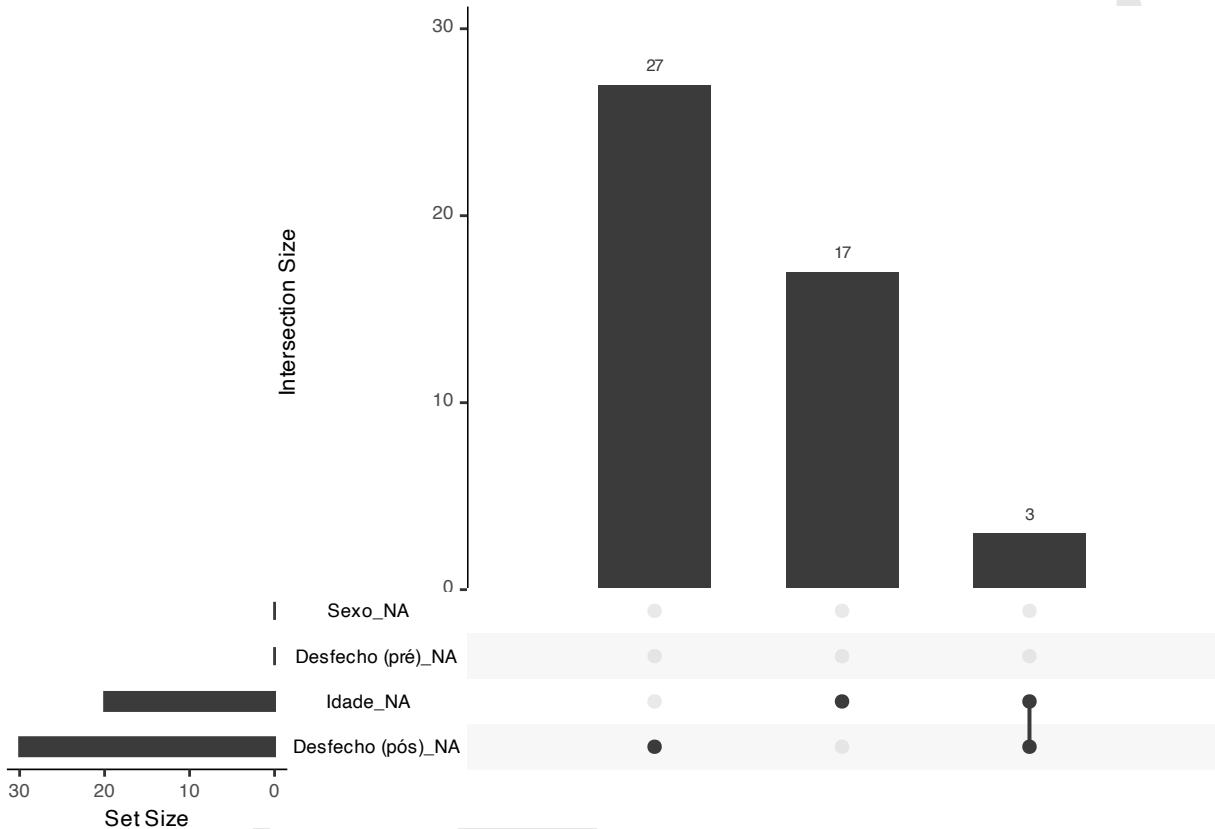


Figura 14.1: Representação gráfica de dados perdidos completamente ao acaso (MCAR) em um estudo randomizado controlado (RCT).

- Dados perdidos ao acaso (*missing at random*, MAR), em que a probabilidade de ocorrência de dados perdidos é relacionada a outras variáveis medidas.^{158,159}
- 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.^{158,159}

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).¹⁵⁸
- 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).¹⁵⁸

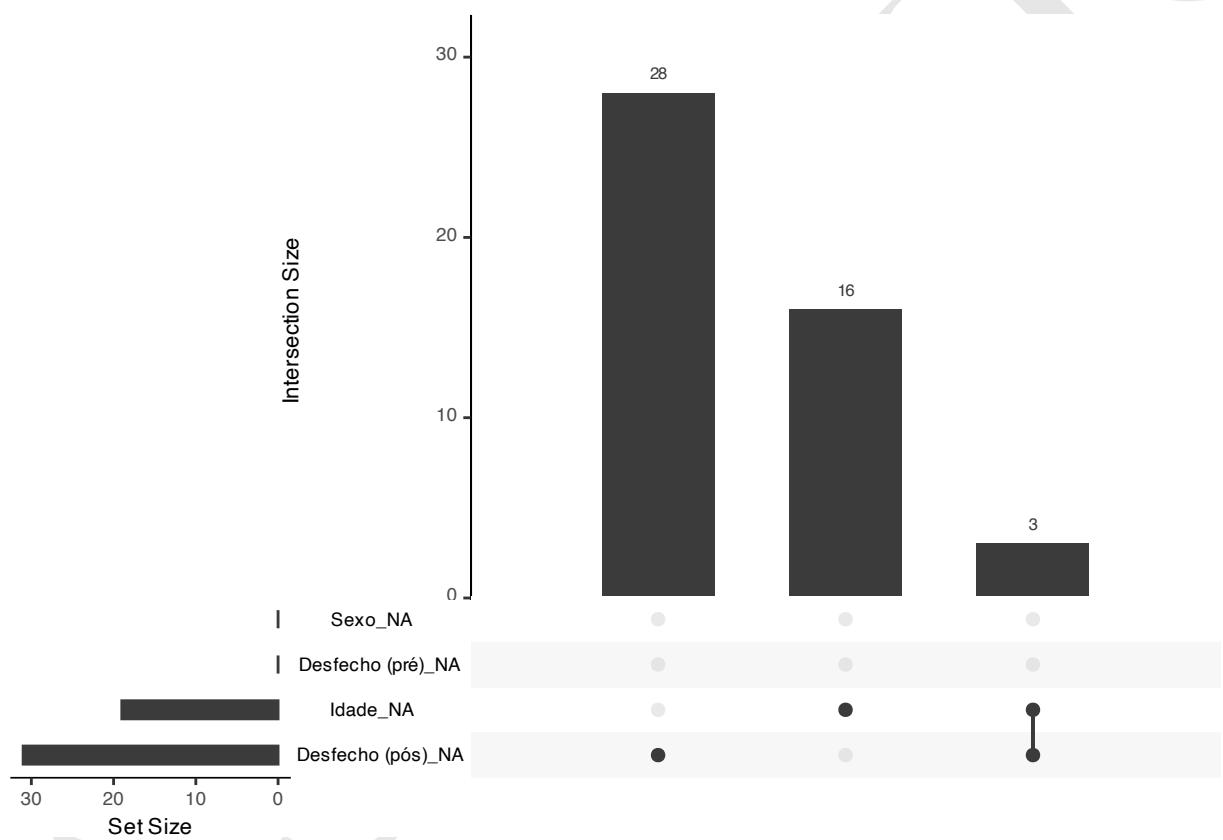


Figura 14.2: Representação gráfica de dados perdidos ao acaso (MAR) em um estudo randomizado controlado (RCT).

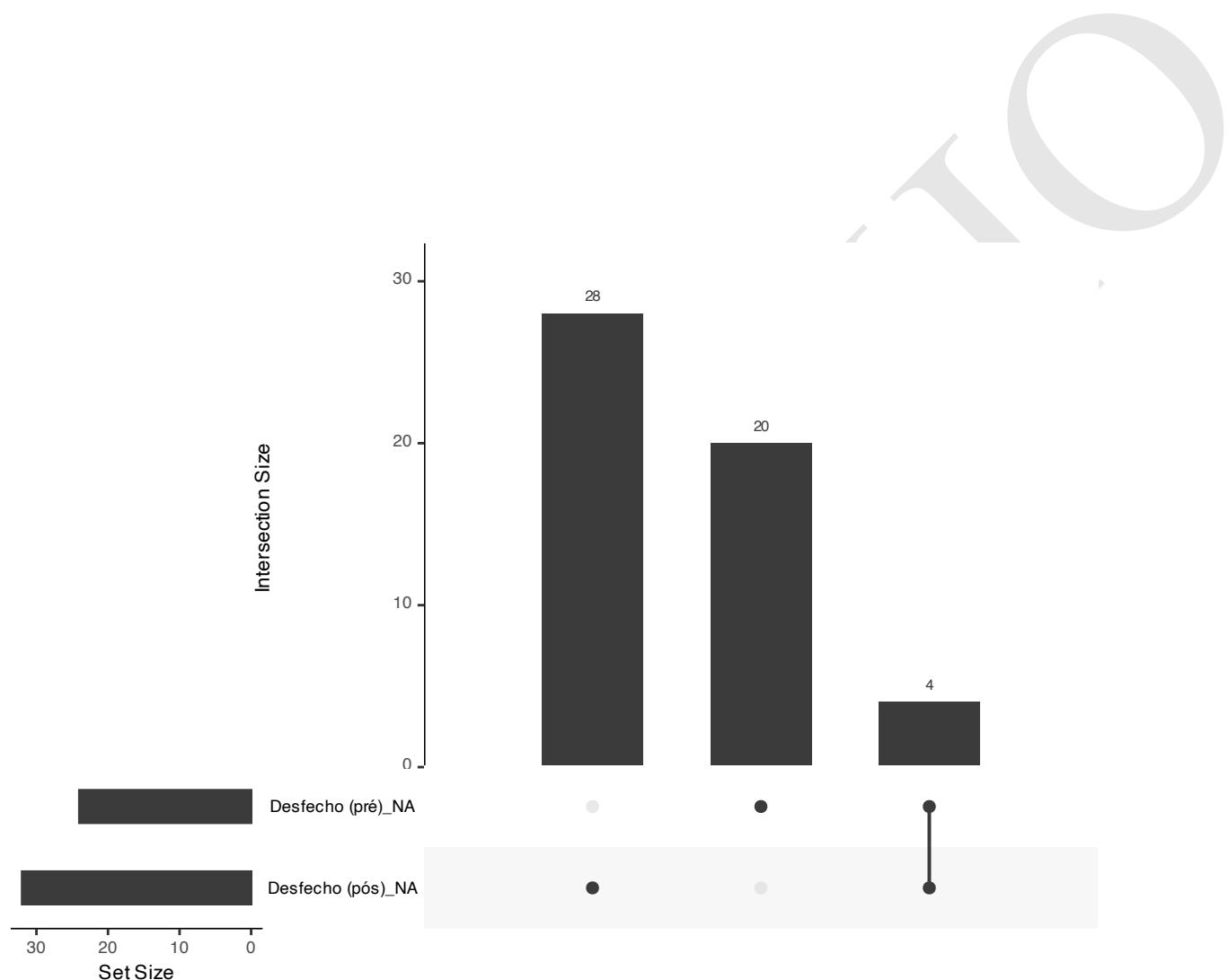


Figura 14.3: Representação gráfica de dados perdidos não ao acaso (MNAR) em um estudo randomizado controlado (RCT).



O pacote *misty*¹⁶⁰ fornece a função *na.test*^a para executar o Little's Missing Completely at Random (MCAR) test¹⁶¹.

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



O pacote *naniar*¹⁶² fornece a função *mcar_test*^a para executar o Little's Missing Completely at Random (MCAR) test¹⁶¹.

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



O pacote *naniar*¹⁶² 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.¹⁵⁷

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).¹⁵⁷
- 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.¹⁵⁹
- 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.¹⁵⁹



O pacote *base*⁵⁹ 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*¹³⁶ 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.¹⁶³

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.¹⁵⁹
- 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.¹⁵⁷

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.¹⁶⁴
- Os métodos de imputação de dados mais robustos incluem a imputação multivariada por equações encadeadas (*multivariate imputation by chained equations*, MICE)¹⁶⁵ e a correspondência média preditiva (*predictive mean matching*, PMM).^{166,167}

R

Os pacotes *mice*¹⁶⁵ e *miceadds*¹⁶⁸ 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>

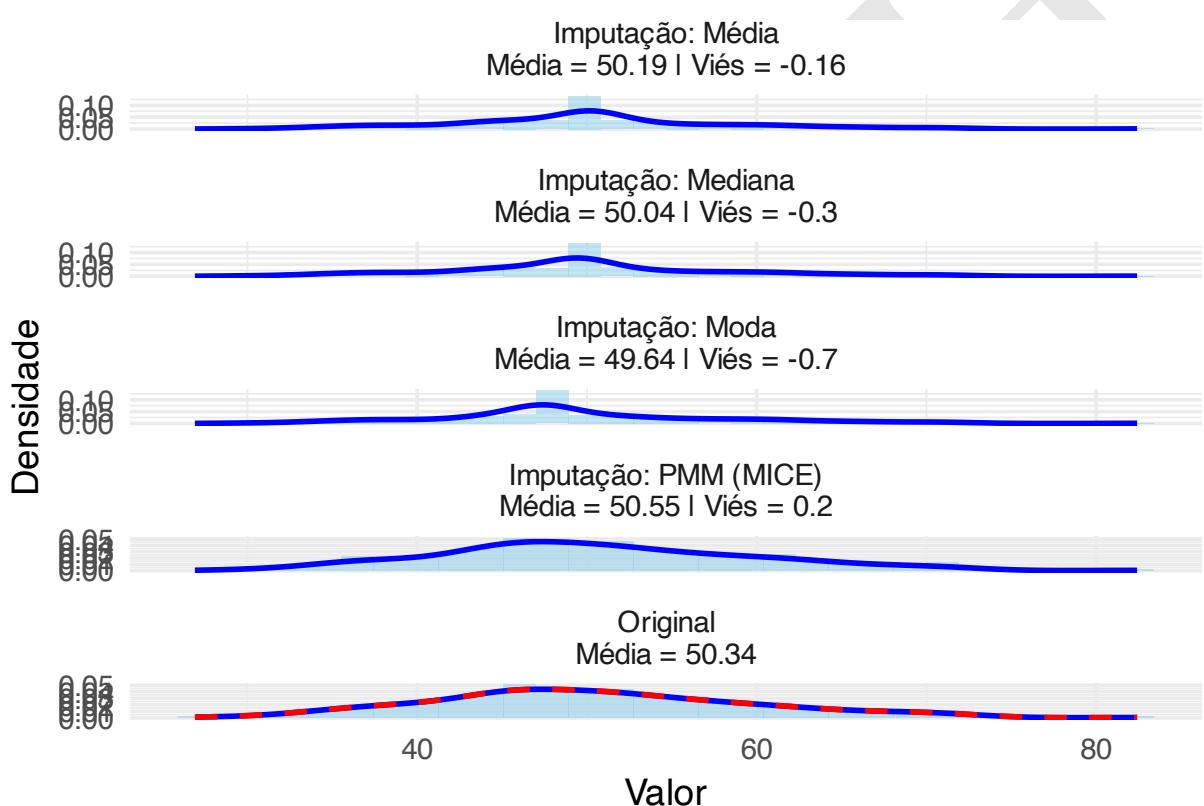


Figura 14.4: Impacto de métodos de imputação na distribuição de uma variável contínua com dados perdidos.

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*¹⁶⁹ 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*¹⁷⁰ 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*¹⁷¹ 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*¹⁷² 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*¹⁷³ 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>

Capítulo 16

Ciência cidadã

16.1 Introdução

16.1.1 O que é Ciência Cidadã?

• . ?

16.1.2 Que tipo de dados ela produz?

• . ?

16.1.3 Como ela difere de amostragem tradicional?

• . ?

16.1.4 Quais erros são mais comuns?

• . ?

16.1.5 Quando esses dados são válidos?

• . ?

16.1.6 Quando não são?

• . ?

16.1.7 Como documentar metadados?

• . ?

16.1.8 Que análises fazem sentido?

• . ?

16.1.9 Como relatar resultados com honestidade?

• . ?

RASCUNHO

PARTE 4: ANÁLISES DESCRIPTIVAS E EXPLORATÓRIAS

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

RASCUNHO

Capítulo 17

Distribuições e parâmetros

17.1 Distribuições de probabilidade

17.1.1 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).¹¹³

17.1.2 Como representar distribuições de probabilidade?

- Tabelas de frequência, polígonos de frequência, gráficos de barras, histogramas e *boxplots* são formas de representar distribuições de probabilidade.¹⁷⁴
- 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”.¹⁷⁴
- 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.¹⁷⁴
- 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.¹⁷⁵
- 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.¹⁷⁶
- 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.¹⁷⁷
- 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).¹⁷⁴
- 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.¹⁷⁴
- 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

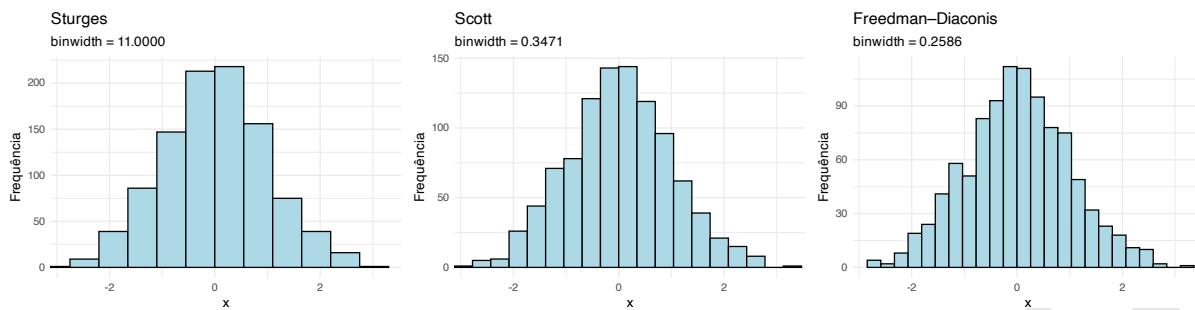


Figura 17.1: Histogramas com diferentes métodos de binning.: Sturges, Scott e Freedman-Diaconis.

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.¹⁷⁴

- 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.¹⁷⁴
- *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.¹⁷⁴

R O pacote *grDevices*¹⁷⁸ fornece a função *nclass*^a para determinar a quantidade de classes de um histograma com os métodos de Sturge¹⁷⁵, Scott¹⁷⁶ ou Freedman-Diaconis¹⁷⁷.

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

R O pacote *ggplot2*¹⁷⁹ fornece a função *geom_freqpoly*^a para criar histogramas.

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

17.1.3 Quais características definem uma distribuição?

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

17.1.4 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.?

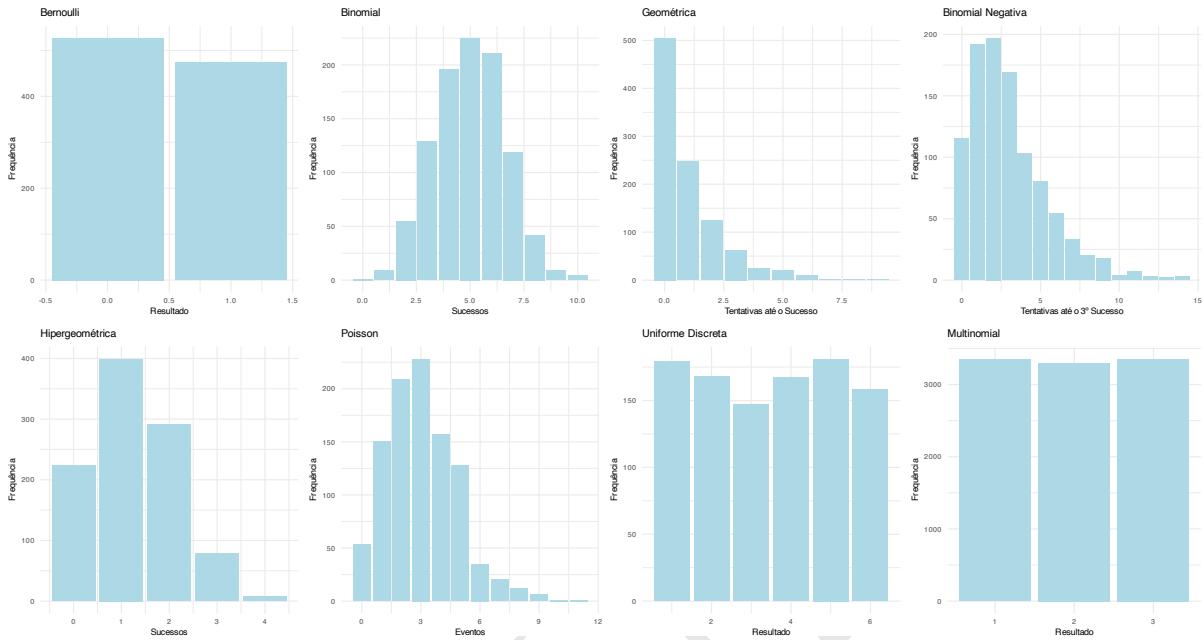


Figura 17.2: Distribuições discretas e suas funções de probabilidade.

- Distribuições contínuas:
 - Uniforme: .?
 - Exponencial: .?
 - Normal: .?
 - Aproximação binomial: .?
 - Aproximação Poisson: .?
 - t-Student: .?
 - Qui-quadrado: .?
 - Weibull: .?
 - Gama: .?
 - Log-normal: .?
 - Beta: .?
 - Logística: .?
 - Pareto.?

17.1.5 Quais são as funções de uma distribuição?

- Função de massa de probabilidade (*probability mass function, pmf*).?
- Função de distribuição cumulativa (*cumulative distribution function, cdf*).?

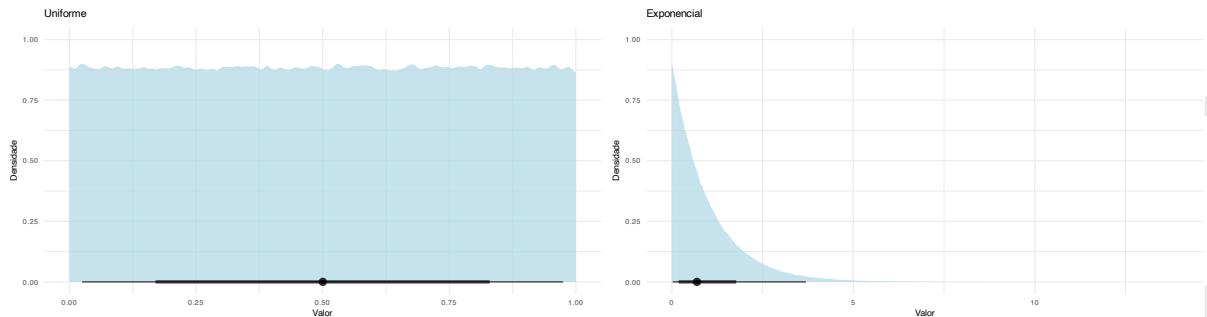


Figura 17.3: Distribuições contínuas básicas e suas funções de densidade.

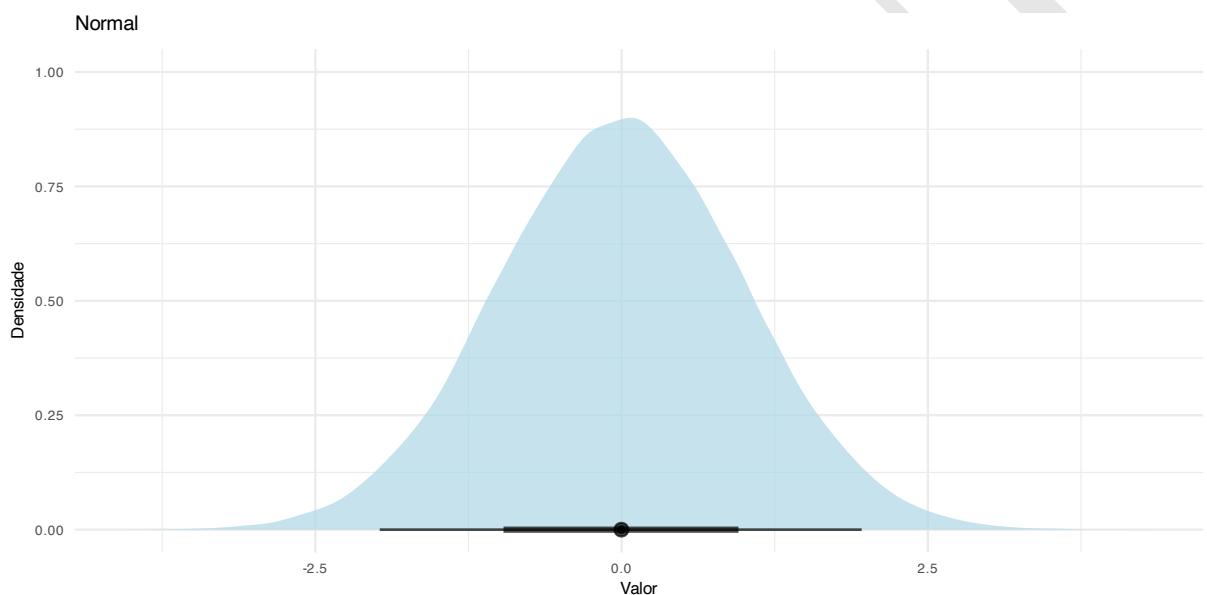


Figura 17.4: Distribuições contínuas aproximadas e suas funções de densidade.

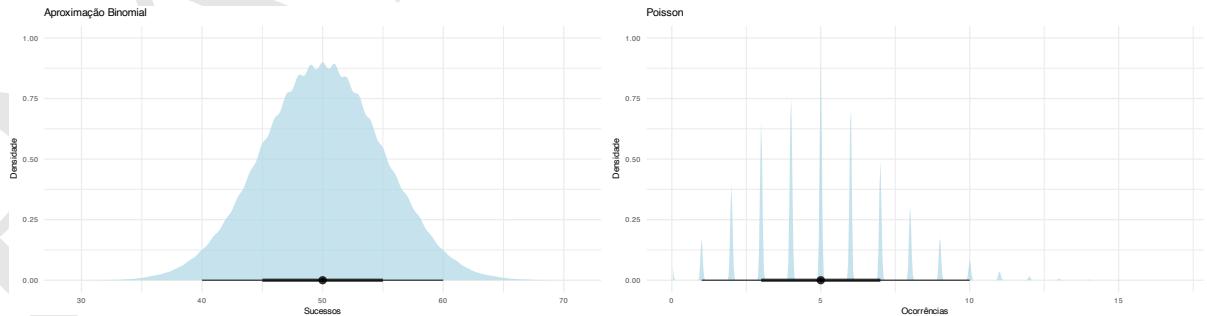


Figura 17.5: Distribuições contínuas aproximadas e suas funções de densidade.

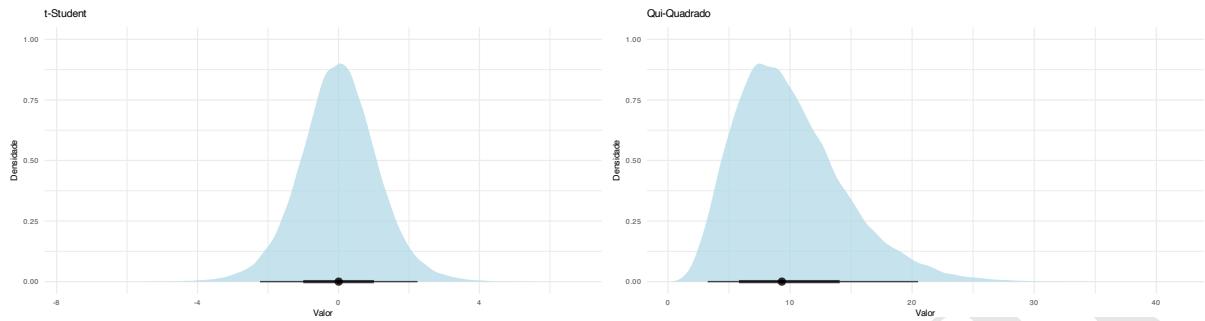


Figura 17.6: Distribuições contínuas para inferência e suas funções de densidade.

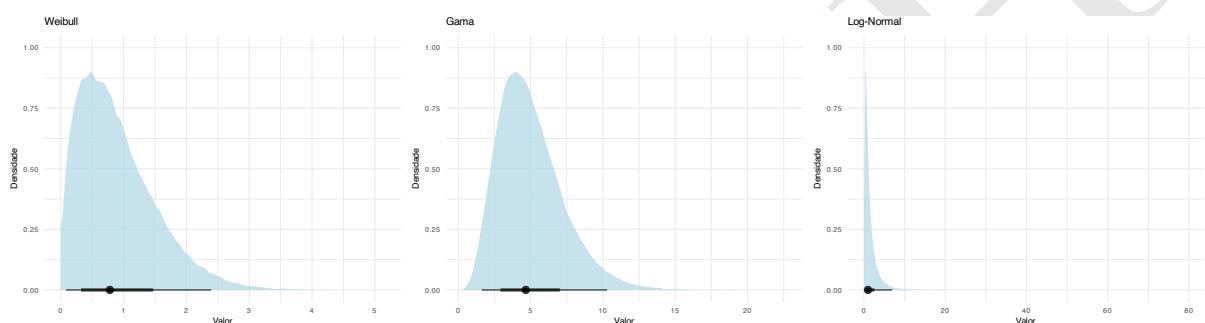


Figura 17.7: Distribuições contínuas para dados específicos e suas funções de densidade.

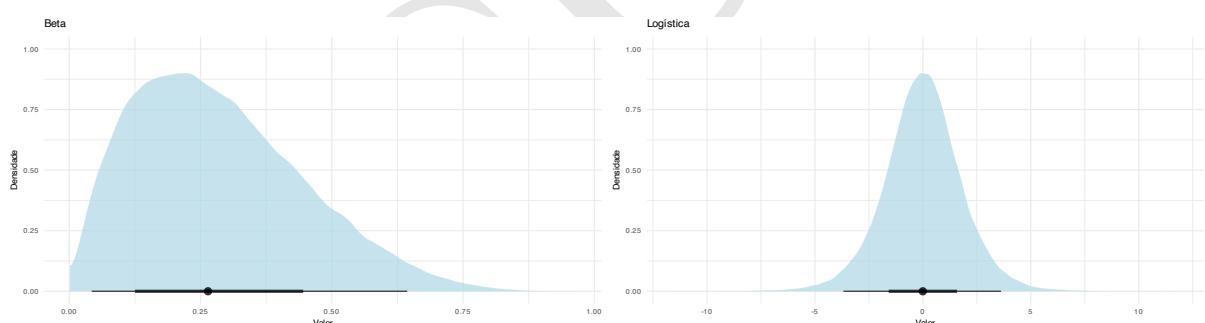


Figura 17.8: Distribuições contínuas para probabilidades e proporções e suas funções de densidade.

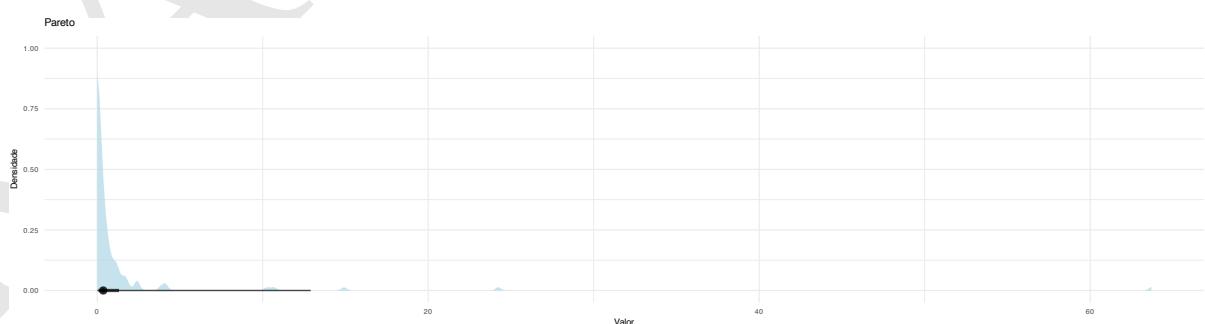


Figura 17.9: Distribuições contínuas com caudas pesadas e suas funções de densidade.

- Função quantílicas (*quantile function*, qf).?
- Função geradora de números aleatórios (*random function*, rf).?

R

O pacote *stats*¹³⁶ 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*¹⁸⁰ 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*¹⁸¹ 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>

17.1.6 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.¹¹⁴
- 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.¹¹⁴

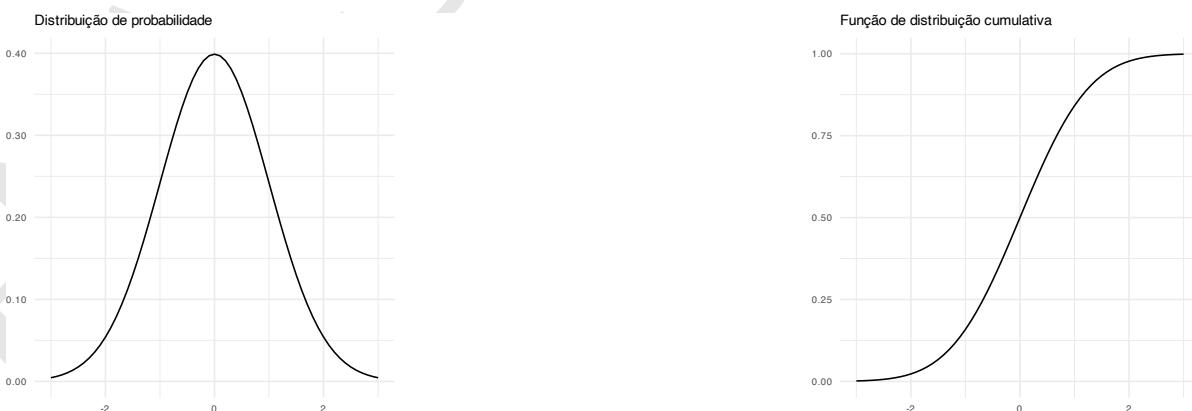


Figura 17.10: Distribuições e funções de probabilidade.

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

- Histogramas.¹¹³
- Gráficos Q-Q.¹¹³
- Testes de hipótese nula:¹¹³
 - Kolmogorov-Smirnov
 - Shapiro-Wilk
 - Anderson-Darling

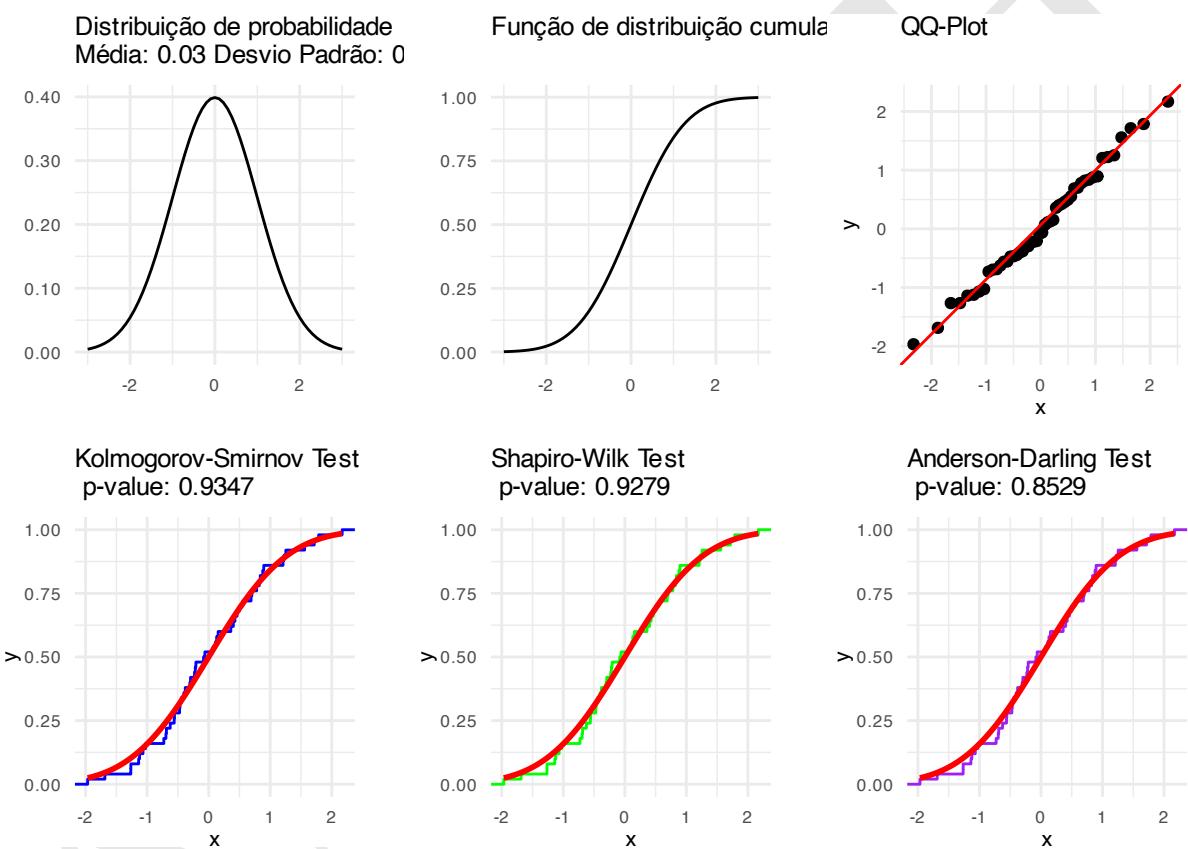


Figura 17.11: Distribuição normal e métodos de visualização e testes de normalidade.

17.1.8 O que são distribuições não-normais?

• ?

17.2 Distribuições multivariadas

17.2.1 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

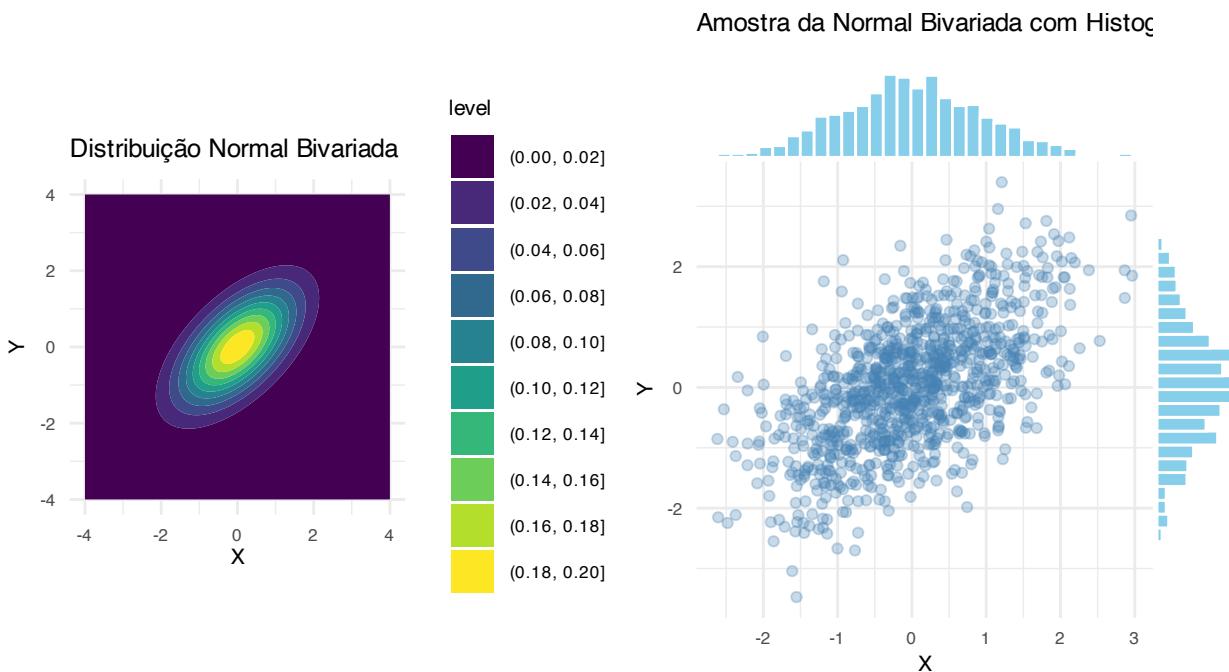


Figura 17.12: Distribuição normal bivariada e amostra simulada com histogramas marginais.

17.3 Parâmetros

17.3.1 O que são parâmetros?

- Parâmetros são informações que definem um modelo teórico, como propriedades de uma coleção de indivíduos.¹¹²
- Parâmetros definem características de uma população inteira, tipicamente não observados por ser inviável ter acesso a todos os indivíduos que constituem tal população.¹¹³



O pacote *base*⁵⁹ fornece a função *summary*^a para calcular diversos parâmetros descritivos.

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

17.3.2 O que é uma análise paramétrica?

- Testes paramétricos possuem suposições sobre as características e/ou parâmetros da distribuição dos dados na população.¹¹³
- 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.^{113,114}
- Testes paramétricos são baseados na suposição de que os dados amostrais provêm de uma população com parâmetros fixos determinando sua distribuição de probabilidade.⁸

17.3.3 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.^{113,114}
- Testes não-paramétricos são úteis quando as suposições de normalidade não podem ser sustentadas.¹¹⁴

17.3.4 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.¹⁸²

17.3.5 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.^{113,183,184}
- Testes não-paramétricos apresentam menor poder estatístico (maior erro tipo II) comparados aos testes paramétricos correspondentes.¹¹⁴

17.3.6 Que parâmetros podem ser estimados?

- Parâmetros de tendência central.^{114,185}
- Parâmetros de dispersão.^{114,185,186}
- Parâmetros de proporção.^{114,185,187,187}
- Parâmetros de distribuição.¹⁸⁵
- Parâmetros de extremos.¹¹⁴



O pacote *base*⁵⁹ fornece a função *summary*^a para calcular diversos parâmetros descritivos.

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

17.4 Erro

17.4.1 Que parâmetros de erro podem ser estimados?

- Margem de erro (ME).^(17.1).

$$ME = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad (17.1)$$

- Erro-padrão da média (EPM) (17.2) (*sigma* conhecido) e (17.3) (*sigma* desconhecido).^{186,188}

$$EPM = \frac{\sigma}{\sqrt{n}} \quad (17.2)$$

$$\widehat{EPM} = \frac{s}{\sqrt{n}} \quad (17.3)$$

- Intervalo de confiança para variância conhecida (17.4) e desconhecida (17.5): Captura a média populacional correspondente ao nível de significância α pré-estabelecido^{114,185,188,189}.

$$IC = \bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad (17.4)$$

$$IC = \bar{x} \pm t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}} \quad (17.5)$$

17.5 Tendência central

17.5.1 Que parâmetros de tendência central podem ser estimados?

- Média aritmética (17.6), ponderada (17.7), geométrica (17.8) ou harmônica (17.9).^{114,185,190}

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (17.6)$$

$$\bar{x}_p = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad (17.7)$$

$$\bar{x}_g = \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}} \quad (17.8)$$

$$\bar{x}_h = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} \quad (17.9)$$

- Mediana (17.10).^{114,185,191}

$$\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{se } n \text{ é ímpar} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se } n \text{ é par} \end{cases} \quad (17.10)$$

- Moda (17.11), onde $f(x)$ é a função de frequência absoluta ou relativa e x_1, x_2, \dots, x_n são os valores observados.^{114,185,191}

$$Mo \in \arg \max_{x \in \{x_1, \dots, x_n\}} f(x) \quad (17.11)$$

- Moda (dados agrupados) (17.12), 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.

$$Mo = L + \frac{(f_1 - f_0)}{(f_1 - f_0) + (f_1 - f_2)} \cdot h \quad (17.12)$$

- A posição relativa das medidas de tendência central (média, mediana e moda) depende da forma da distribuição.¹⁹¹
- Em uma distribuição normal, as três medidas são idênticas.¹⁹¹
- A média é sempre puxada para os valores extremos, por isso é deslocada para a cauda em distribuições assimétricas.¹⁹¹
- A mediana fica entre a média e a moda em distribuições assimétricas.¹⁹¹

- A moda é o valor mais frequente e, portanto, se localiza no pico da distribuição assimétrica.¹⁹¹
- 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.¹⁹¹

Distribuições Unimodal, Bimodal e Multimodal

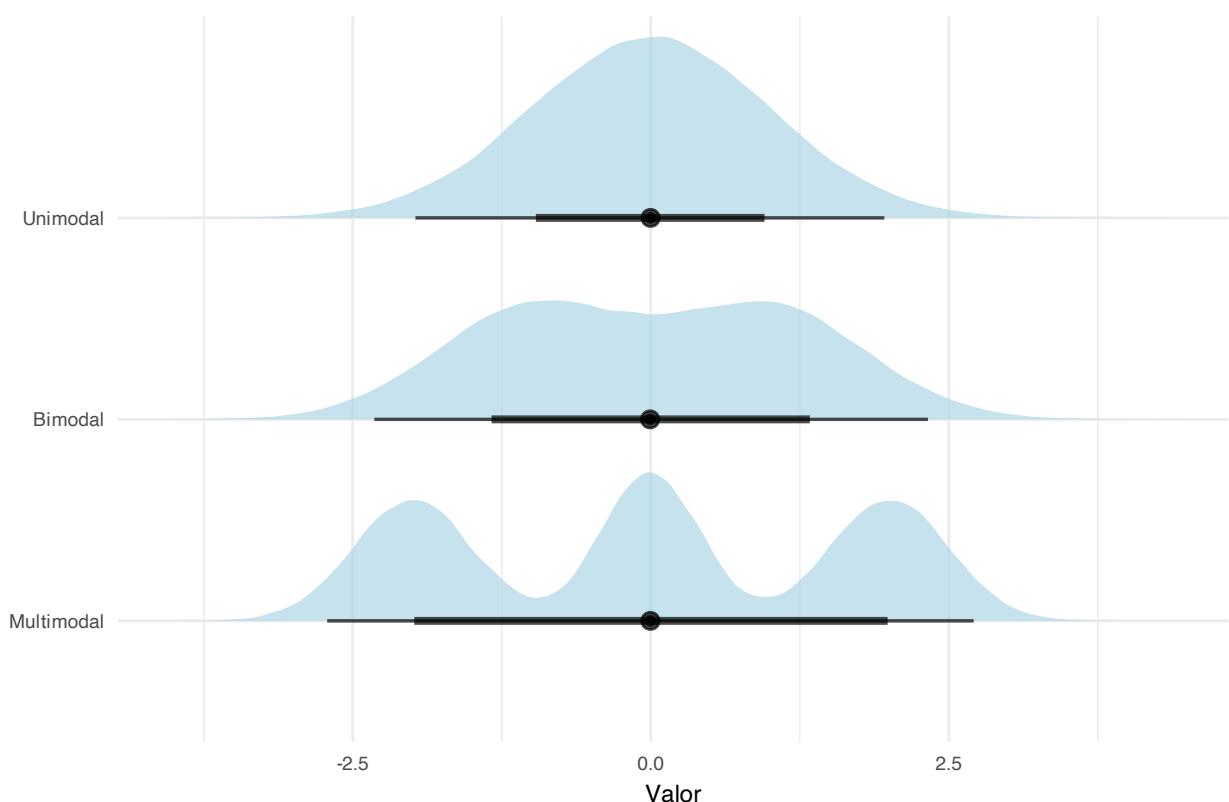


Figura 17.13: Distribuições unimodal, bimodal e multimodal.



O pacote *base*⁵⁹ fornece a função *summary*^a para calcular diversos parâmetros descritivos.

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

17.5.2 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.¹⁹¹
- A moda é preferida quando os dados são medidos em uma escala nominal.¹⁹¹
- A média geométrica é preferida quando os dados são medidos em uma escala logarítmica.¹⁹¹

17.6 Dispersão

17.6.1 Que parâmetros de dispersão podem ser estimados?

- Variância (17.13).^{114,185}

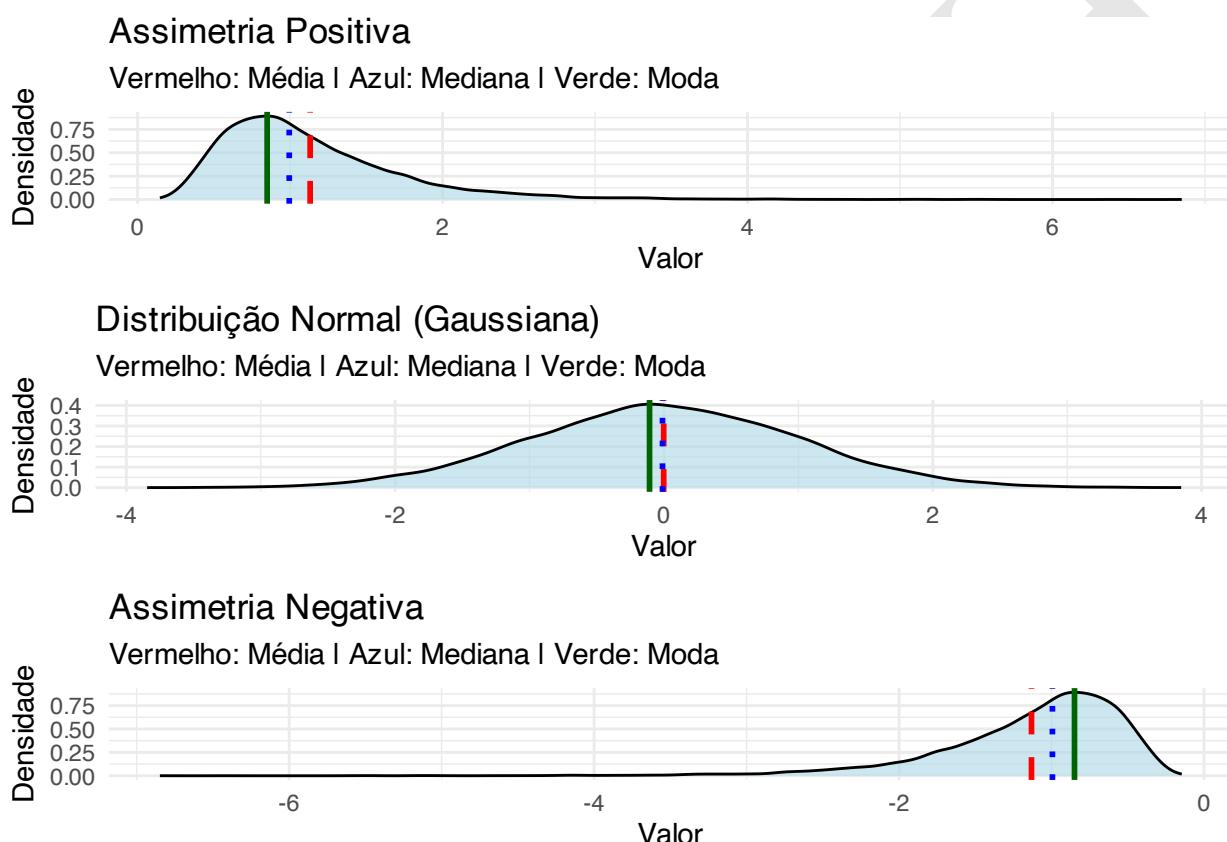


Figura 17.14: Parâmetros de tendência central em distribuições assimétricas e normais.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (17.13)$$

- Desvio-padrão (17.14).^{186,188,192}

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (17.14)$$

- Amplitude (17.15).^{114,185,192}

$$A = x_{\max} - x_{\min} \quad (17.15)$$

- Intervalo interquartil (17.16).^{114,185,192}

$$IQR = Q_3 - Q_1 \quad (17.16)$$

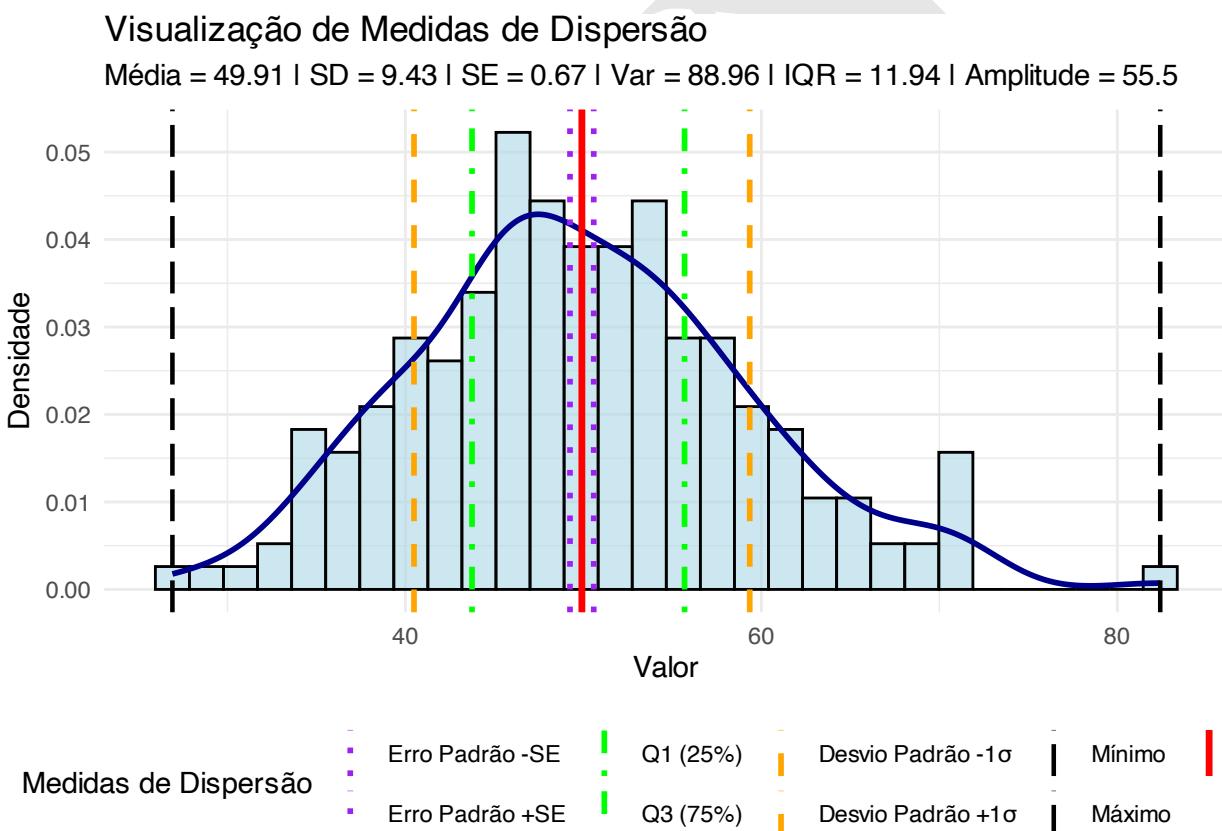


Figura 17.15: Parâmetros de dispersão em distribuições normais.

O pacote *base*⁵⁹ 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*¹³⁶ 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>

17.6.2 Como escolher o parâmetro de dispersão?

- Desvio-padrão σ é apropriado quando a média é utilizada como parâmetro de tendência central em distribuições simétricas.¹⁹²
- Amplitude ou intervalo interquartil são apropriadas para variáveis ordinais ou distribuições assimétricas.¹⁹²

17.6.3 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.¹⁹³
- A correção de Bessel é feita subtraindo-se 1 do número de observações da amostra, ou seja, $n - 1$ (17.17).¹⁹³

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (17.17)$$

17.6.4 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.¹⁹³
- 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.¹⁹³

17.7 Proporção

17.7.1 Que parâmetros de proporção podem ser estimados?

- Frequência absoluta (17.18).^{114,185,187}

$$f_i = n_i \quad (17.18)$$

- Frequência relativa (17.19).^{114,185,187}

$$fr_i = \frac{n_i}{N} \quad (17.19)$$

- Percentil (17.20), onde k é o percentil desejado (0 a 100) e n é o número total de observações na amostra.^{114,185,187}

$$P_k = x_{(\frac{k}{100} \cdot (n+1))} \quad (17.20)$$

- 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.¹⁸⁷

- Tercil: 2 valores que dividem a amostra em 3 grupos de tamanhos iguais.¹⁸⁷

- Quartil: 3 valores que dividem a amostra em 4 grupos de tamanhos iguais.¹⁸⁷
- Quintil: 4 valores que dividem a amostra em 5 grupos de tamanhos iguais.¹⁸⁷
- Decil: 9 valores que dividem a amostra em 10 grupos de tamanhos iguais.¹⁸⁷

R

O pacote *base*⁵⁹ 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*⁵⁹ 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*⁵⁹ fornece a função *quantile*^a para executar análise de percentis.

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

17.8 Extremos

17.8.1 O que são valores extremos?

- Valores extremos podem constituir valores legítimos ou ilegítimos de uma distribuição.¹⁹⁴

17.8.2 Que parâmetros extremos podem ser estimados?

- Mínimo (17.21).¹¹⁴

$$\text{Mínimo} = \min(x_1, x_2, \dots, x_n) \quad (17.21)$$

- Máximo (17.26).¹¹⁴

$$\text{Máximo} = \max(x_1, x_2, \dots, x_n) \quad (17.22)$$

17.9 Distribuição

17.9.1 Que parâmetros de distribuição podem ser estimados?

- Assimetria (17.23).¹⁸⁵

$$\gamma_1 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}} \quad (17.23)$$

- Curtose (17.24).¹⁸⁵

$$\gamma_2 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2} \quad (17.24)$$

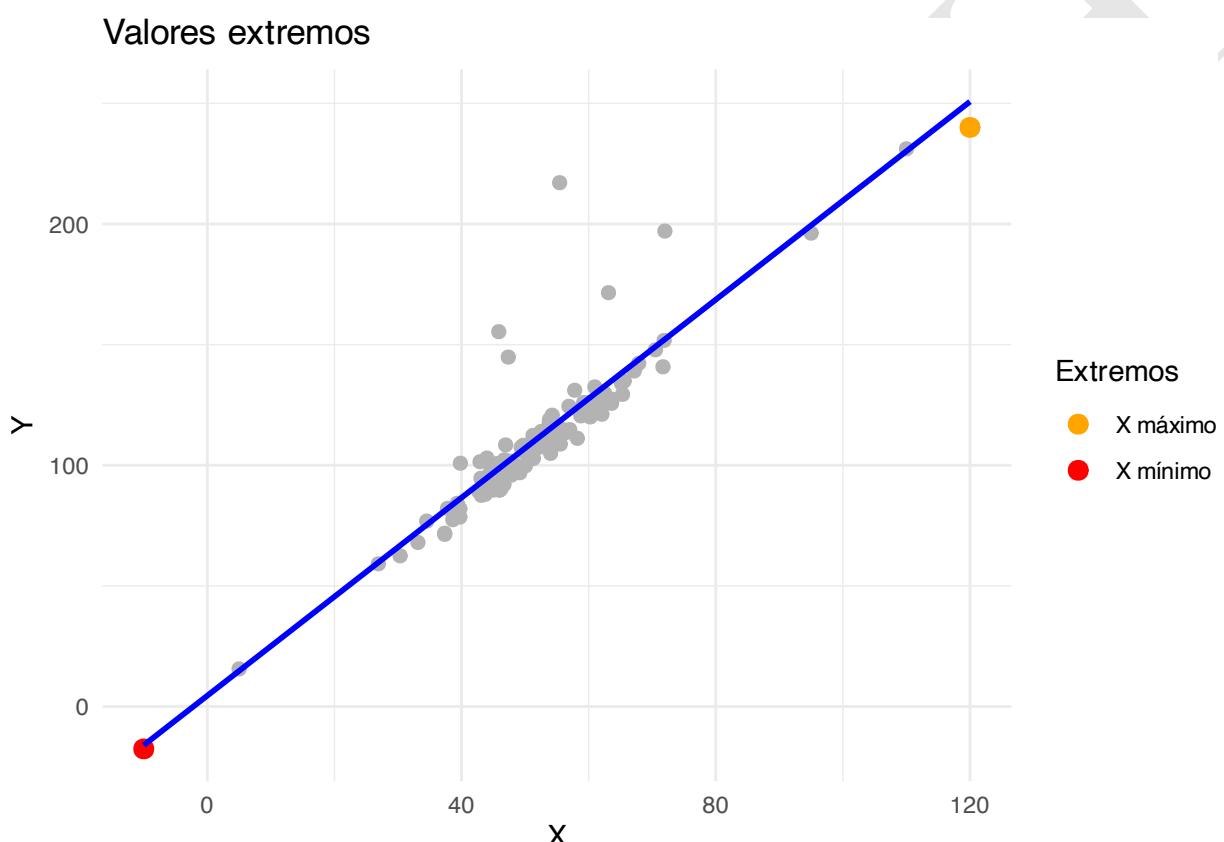


Figura 17.16: Regressão linear com valores extremos.

- Excesso de Curtose (17.25).¹⁸⁵

$$\kappa = \gamma_2 - 3 \quad (17.25)$$

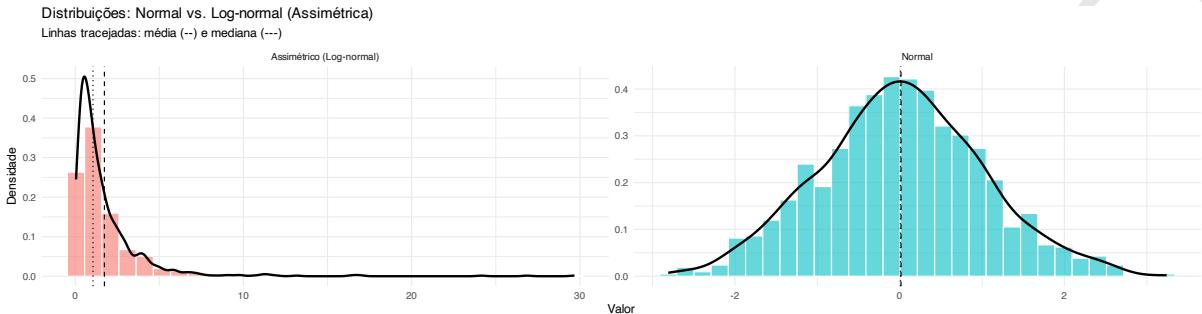


Figura 17.17: Parâmetros de distribuição: Assimetria e Curtose.

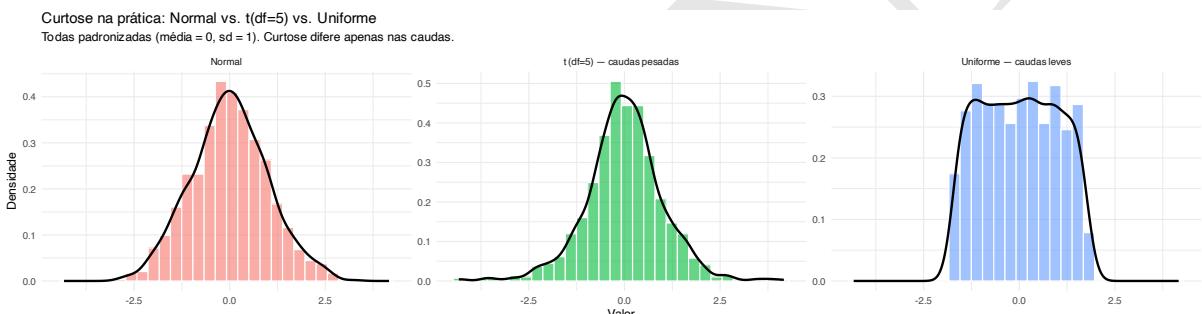


Figura 17.18: Parâmetros de distribuição: Curtose em distribuições simétricas (normal vs. uniforme).

17.10 Robustez em medidas de localização

17.10.1 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.¹⁹⁵

17.10.2 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.¹⁹⁵

17.10.3 Qual a alternativa robusta para localização?

- Mediana, com 50 de ponto de quebra e função influência limitada.¹⁹⁵

17.10.4 Como estimar escala de forma robusta?

- *Median Absolute Deviation* (MAD) (??), com correção 1,483 para normalidade, com 50 de ponto de quebra.¹⁹⁵

$$MAD = 1.483 \cdot \text{median}(|x_i - \text{median}(x)|) \quad (17.26)$$

- Primeiro quartil das diferenças pareadas (Qn) (17.27), com 50 de ponto de quebra.¹⁹⁵

$$Qn = 2.2219 \cdot \text{first quartile}(|x_i - x_j|; i < j) \quad (17.27)$$

- O intervalo interquartil (IQR) (17.16) é robusto, com ponto de quebra 25, sendo simples de interpretar e útil em boxplots.¹⁹⁵

17.11 Parâmetros robustos

17.11.1 O que são parâmetros robustos?

- Parâmetros robustos são medidas de posição e dispersão que permanecem estáveis mesmo na presença de valores discrepantes.¹⁹⁶

17.11.2 Que parâmetros robustos podem ser estimados?

- Mediana em vez da média aritmética, pois é menos sensível a valores extremos.¹⁹⁶
- *Median Absolute Deviation* (MAD) em vez do desvio-padrão σ , que pode ser escalonado por 1,483 para comparabilidade.¹⁹⁶
- Qn e Sn como estimadores alternativos de dispersão robusta.¹⁹⁶
- Média e variância Winsorizadas como opções intermediárias, reduzindo a influência dos *outliers*.¹⁹⁶

17.11.3 Por que utilizar parâmetros robustos?

- Parâmetros robustos garantem maior confiabilidade quando os dados não seguem a normalidade ou apresentam contaminação por *outliers*.¹⁹⁶
- Parâmetros robustos permitem análises mais estáveis em estudos exploratórios, evitando decisões equivocadas sobre variabilidade ou tendência central.¹⁹⁶

Capítulo 18

Análise inicial de dados

18.1 Análise inicial de dados

18.1.1 O que é análise inicial de dados?

- Análise inicial de dados¹⁹⁷ 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.¹³⁹
- O objetivo da análise inicial de dados é propiciar dados prontos para análise estatística, incluindo informações confiáveis sobre as propriedades dos dados.¹³⁹
- A análise inicial de dados pode ser dividida nas seguintes etapas:¹³⁹
 - 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,¹⁹⁸ nem deve ser utilizada para hipotetizar após os dados serem coletados (conhecido como *Hypothesizing After Results are Known*, HARKing).⁹⁷

18.1.2 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.¹³⁹
- 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.¹³⁹
- Use os metadados do estudo para guiar a análise inicial dos dados e compartilhe com os dados para maior transparência e reproduzibilidade.¹³⁹

- Representação gráfica dos dados pode ajudar a identificar características e padrões no banco de dados, tais como suposições e tendências.¹³⁹
- 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.¹³⁹
- 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.¹³⁹
- 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.¹⁹⁹

18.1.3 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*¹³⁶ 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*¹³⁶ 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.²⁰⁰



O pacote *base*⁵⁹ 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.²⁰⁰
- Ordenação cronológica de variáveis com registros temporais (retrospectivos ou prospectivos).²⁰⁰
- A distribuição das variáveis para verificação das suposições das análises planejadas.²⁰⁰
- Ocorrência de efeitos teto e piso nas variáveis.²⁰⁰

Capítulo 19

Análise exploratória de dados

19.1 Análise exploratória de dados

19.1.1 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.¹⁹⁸
- Análise exploratória deve ser separada da análise inferencial de testes de hipóteses; a decisão sobre os modelos a testar deve ser feita *a priori*.²⁰¹

19.1.2 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.¹⁹⁷

19.1.3 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.²⁰¹
- A análise exploratória não deve ser usada para definir as questões e hipóteses científicas do estudo.²⁰¹



O pacote *explore*²⁰² 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*²⁰³ 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*²⁰⁴ 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*²⁰⁵ 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*²⁰⁶ 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*²⁰⁷ 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>

19.2 Ingredientes da análise exploratória de dados

19.2.1 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.¹⁹⁷
- Técnicas multivariadas exploratórias, como análise de componentes principais e análise de clusters, podem revelar padrões em dados complexos.¹⁹⁷

19.2.2 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.²⁰¹
- Verifique a existência e/ou influência de valores discrepantes (“fora da curva” ou *outliers*):^{197,198,201}
 - 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.²⁰⁸

R

O pacote *ggplot2*¹⁷⁹ 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).²⁰¹

- 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:^{197,201}
 - Histograma das variáveis (por fator de análise)
 - Histograma dos resíduos da regressão
- Verifique a existência de grande quantidade de valores nulos (=0).²⁰¹
 - Histograma das variáveis (por fator de análise)
- Verifique a existência de colinearidade entre variáveis independentes de um modelo de regressão:²⁰¹
 - 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:²⁰¹
 - Gráfico de dispersão entre variável independente e dependente
- Verifique possíveis interações entre as variáveis dependente(s) de um modelo de regressão.²⁰¹
 - Gráfico *coplot* de dispersão entre variáveis dependentes



O pacote *ggcleveland*²⁰⁹ 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:²⁰¹
 - 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.²⁰⁸
- 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.²⁰⁸
- 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.²⁰⁸
- Transformar variáveis em novas formas (por exemplo, log ou inverso) pode revelar simetrias ocultas e tornar relações mais claras e lineares.²⁰⁸

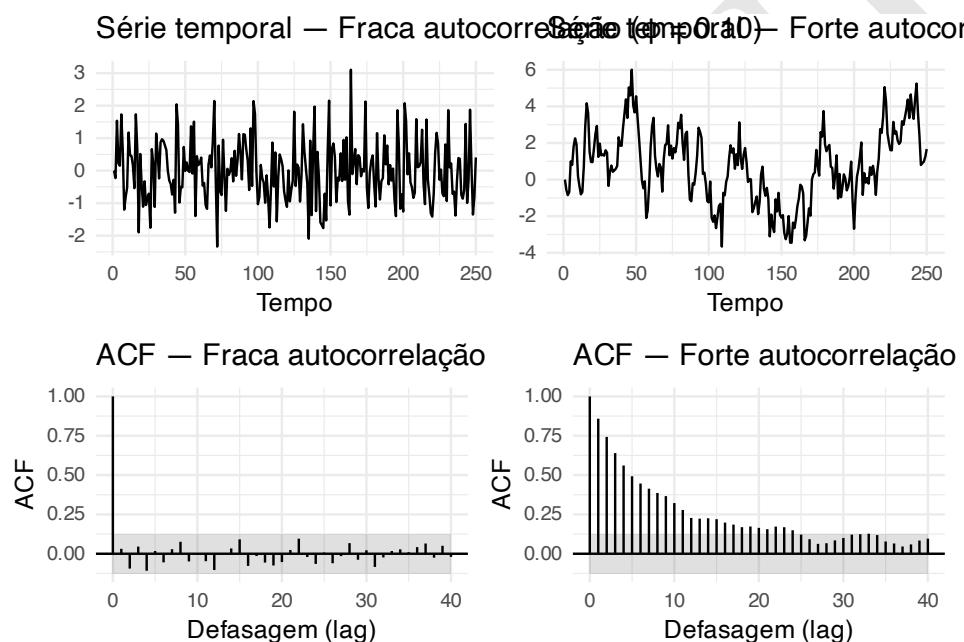


Figura 19.1: Séries temporais e autocorrelação de duas séries simuladas com fraca e forte autocorrelação.

Capítulo 20

Análise descritiva

20.1 Análise descritiva

20.1.1 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?”.^{113,210}

O pacote *base*⁵⁹ fornece a função *summary*^a para calcular diversos parâmetros descritivos.

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

20.1.2 Como apresentar os resultados descritivos?

- Variáveis categóricas: Reporte valores de frequência absoluta e relativa (n, percentual).²¹¹
- Organização das tabelas: as variáveis são exibidas em linhas e os grupos são exibidos em colunas.²¹¹
- Calcule percentagens para as colunas (isto é, entre grupos) e não entre linhas.²¹¹
- 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.²¹¹

20.2 Apresentação de resultados numéricos

20.2.1 O que são casas decimais?

- O número de casas decimais refere-se à quantidade de dígitos que aparecem após a vírgula decimal.^{212,213}
- Para tamanhos de efeito: use 2–3 dígitos significativos.²¹⁴
- Para medidas de variabilidade (desvio-padrão/erro-padrão/intervalo de confiança): use 1–2 dígitos significativos.²¹⁴

20.2.2 O que são dígitos significativos?

- O termo “dígitos significativos” é preferido a “algarismos significativos” ou “dígitos efetivos” e não se relaciona com significância estatística.^{212,213}

Tabela 20.1: 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 20.2: 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.^{212,213}

20.2.3 Como arredondar dados numéricos?

- Apresentar dados com quantidade excessiva de casas decimais pode dificultar a interpretação e induzir erroneamente uma precisão espúria.^{212,213}
- A precisão é determinada pelo grau de arredondamento aplicado, medido em casas decimais ou dígitos significativos.^{212,213}
- O arredondamento também introduz erros, uma vez que aumenta a imprecisão (isto é, incerteza) em torno do valor original.^{212,213}
- A regra geral é utilizar 2 ou 3 dígitos significativos para tamanhos de efeito e 1 ou 2 dígitos significativos para medidas de variabilidade.²¹³
- 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.²¹²
- 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.²¹²

20.3 Tabelas

20.3.1 Por que usar tabelas?

- Tabelas complementam o texto (e vice-versa), e podem apresentar os dados de modo mais acessível e informativo.²¹⁵

20.3.2 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.^{215,216}

20.3.3 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.[?]
- Tabela de frequência agrupada: organiza dados contínuos em intervalos de classe (ex: faixas etárias) e mostra as frequências correspondentes.[?]
- Tabela de contingência (ou tabela cruzada): cruza duas variáveis categóricas, permitindo observar possíveis associações entre elas.[?]
- Tabela de medidas descritivas: resume variáveis quantitativas com estatísticas como média, mediana, desvio-padrão, mínimo, máximo e quartis.[?]
- Tabela de comparação entre grupos: apresenta médias, desvios-padrão e ocasionalmente resultados de testes de inferência estatística para comparar dois ou mais grupos.[?]
- Tabela de resultados de testes estatísticos: exibe valores de estatísticas de teste, P valores e intervalos de confiança; usada para mostrar inferências.[?]
- 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.[?]
- Tabela de séries temporais ou longitudinais: organiza dados medidos em diferentes momentos no tempo, permitindo visualizar tendências ou variações longitudinais.[?]



O pacote *gtsummary*²¹⁷ 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*²¹⁸ fornece a função *table1*^a para construção de tabelas.

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



O pacote *flextable*²¹⁹ 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*²²⁰ fornece a função *nice_table*^a para criar tabelas formatadas.

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

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

- Erros tipográficos.²²¹
- Ausência de rótulos ou unidades nas variáveis.²²¹
- Relatar estatísticas incorretamente, tais como rotular variáveis contínuas como porcentagens.²²¹
- Estatísticas descritivas de tendência central (ex.: médias) relatadas sem a estatística de dispersão correspondente (ex.: desvio-padrão).²²¹
- Desvio-padrão nulo ($\sigma = 0$).²²¹
- Valores porcentuais que não correspondem ao numerador dividido pelo denominador.²²¹

20.4 Tabela 1

20.4.1 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.^{222,223}

20.4.2 Qual a utilidade da ‘Tabela 1’?

- Descrever (conhecer) as características da amostra e dos grupos sendo comparados, quando aplicável.²²³
- Verificar aderência ao protocolo do estudo, incluindo critérios de inclusão/exclusão, tamanho da amostra e perdas amostrais.²²³
- Permitir a replicação do estudo.²²³
- Meta-analisar os dados junto a estudos similares.²²³
- Avaliar a generalização (validade externa) das conclusões do estudo.²²³

20.4.3 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.²²⁴
- 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.²¹⁴

20.4.4 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.^{183,225}

 O pacote *table1*²¹⁸ fornece a função *table1*^a para construção de tabelas.

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

Tabela 20.3: 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]	

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

O pacote *gtsummary*²¹⁷ 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

20.5 Tabela 2

20.5.1 Qual a utilidade da ‘Tabela 2’?

- A Tabela 2 mostra associações ajustadas multivariadas com o resultado para variáveis resumidas na Tabela 1.²²²

20.5.2 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.^{222,226}
- 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.^{222,226}
- 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.²²²

20.5.3 Como construir a ‘Tabela 2’?

- A Tabela 2 pode ser utilizada para apresentar estimativas de múltiplos efeitos ajustados de um mesmo modelo estatístico.²²²

Tabela 20.4: 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 20.5: 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*²¹⁸ fornece a função *table1*^a para construção de tabelas.

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



O pacote *gtsummary*²¹⁷ 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

20.6 Visualização efetiva de dados

20.6.1 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.²²⁷

20.6.2 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.²²⁷

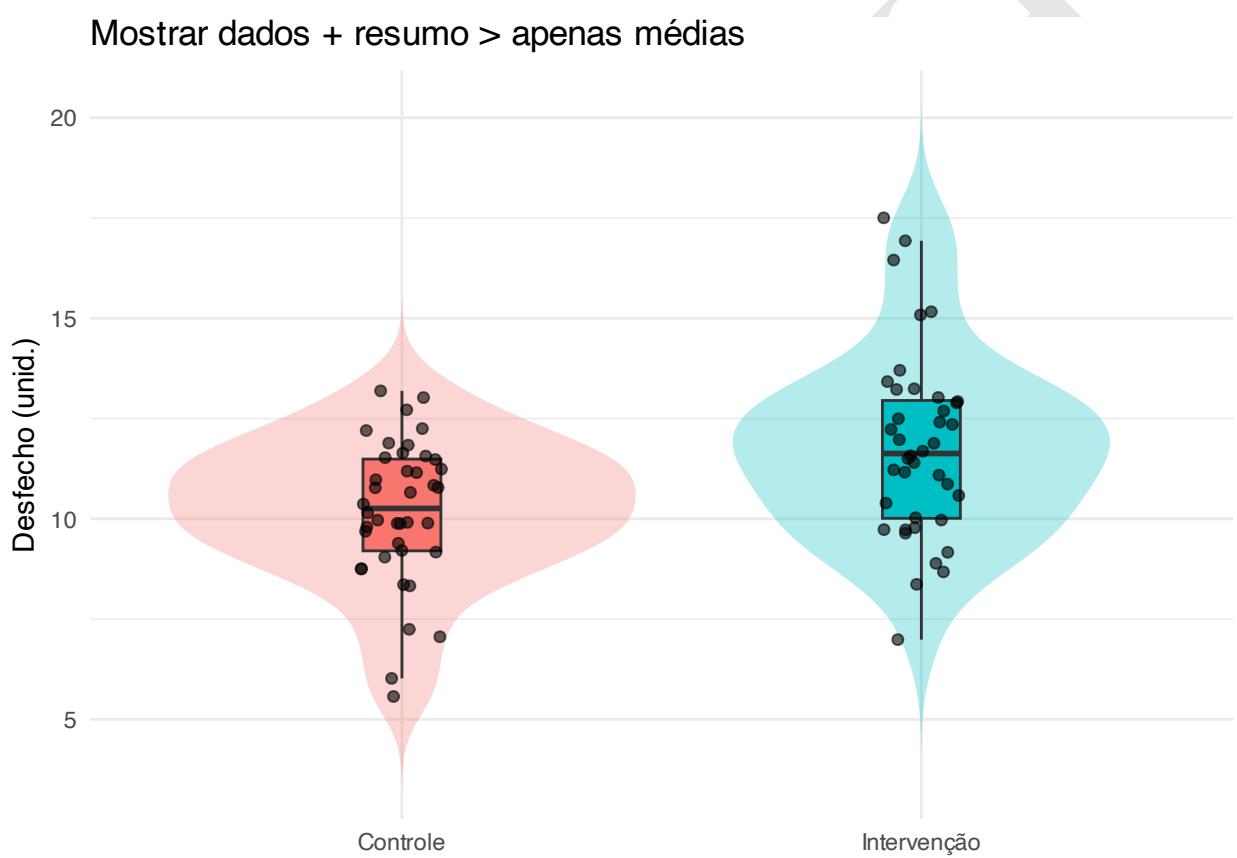


Figura 20.1: Exemplo de gráfico que mostra os dados brutos junto com um resumo estatístico (média e dispersão).

20.7 Gráficos

20.7.1 O que são gráficos?

- Gráficos são utilizados para apresentar dados (geralmente em grande quantidade) de modo mais intuitivo e fácil de compreender.²²⁸

20.7.2 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.²²⁹

20.7.3 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.²²⁸

20.7.4 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.^{188,189}
- Barras de erro mais longas representam mais imprecisão (maiores erros), enquanto barras mais curtas representam mais precisão na estimativa.¹⁸⁹
- Barras de erro descritivas geralmente apresentam a amplitude (mínimo-máximo) ou desvio-padrão.¹⁸⁹
- Barras de erro inferenciais geralmente apresentam o erro-padrão ou intervalo de confiança no nível de significância α pré-estabelecido.^{188,189}
- 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.¹⁸⁸
- Barras de erro com intervalo de confiança são úteis para fornecer uma estimativa da incerteza da estimativa do parâmetro populacional.¹⁸⁸
- 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.¹⁸⁹
- De modo contraintuitivo, um espaço entre as barras não garante significância, nem a sobreposição a descarta—depende do tipo de barra.¹⁸⁸
- Para amostras pequenas é preferível apresentar os dados brutos, uma vez que as barras de erro não serão muito informativas.¹⁸⁸

R

Os pacotes *ggplot2*¹⁷⁹, *plotly*²³⁰ e *corrplot*²³¹ 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>

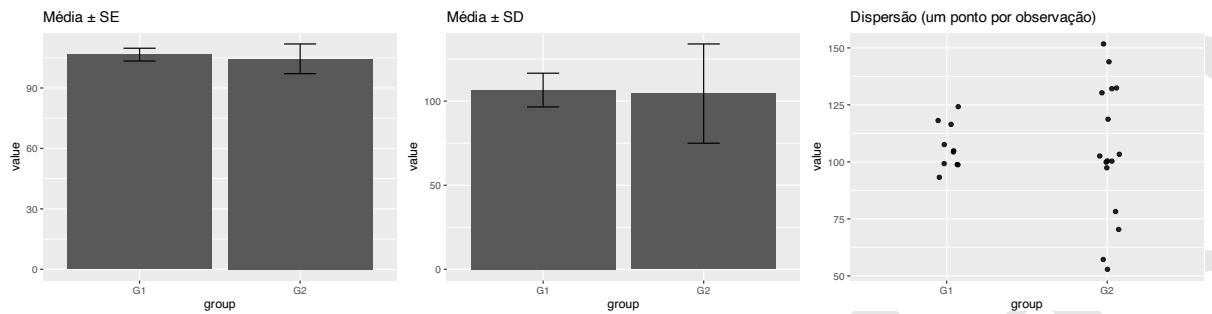


Figura 20.2: Exemplos de gráficos com barras de erro e dados brutos.

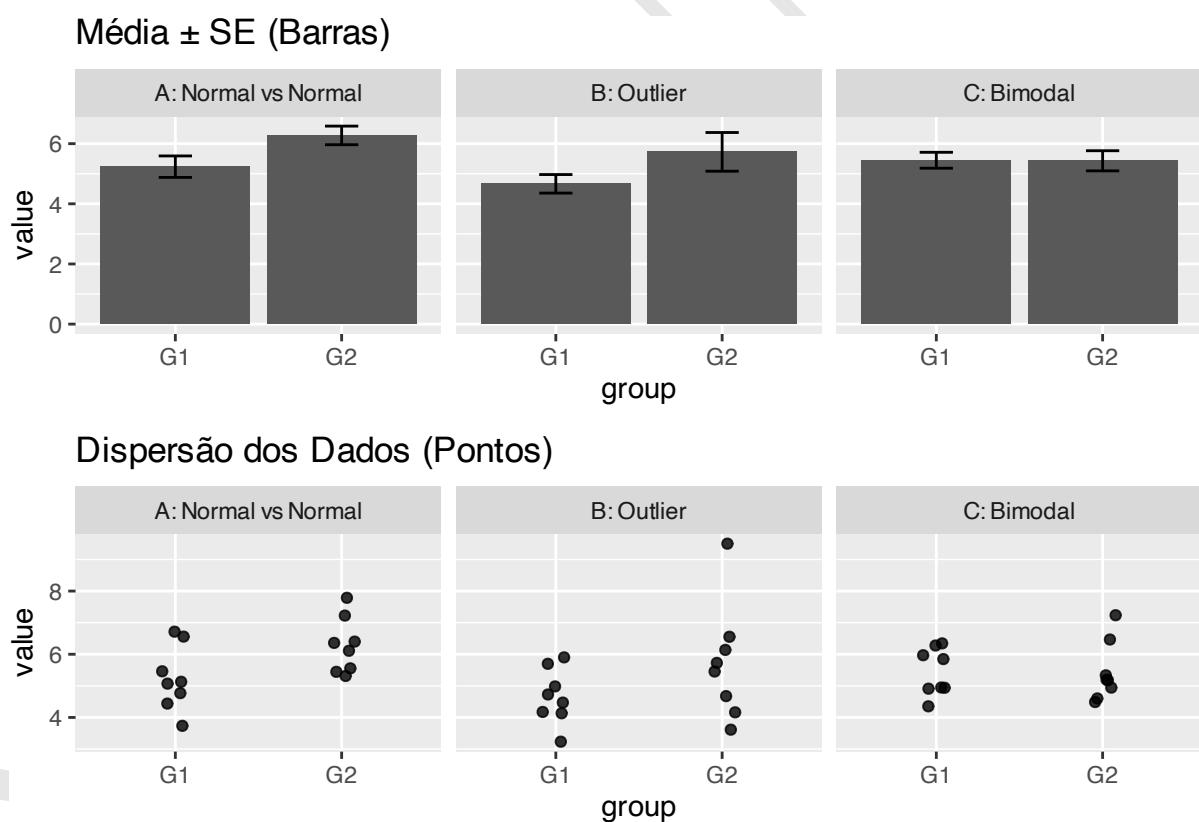


Figura 20.3: Exemplos de gráficos com barras de erro e dados brutos em diferentes cenários.

20.7.5 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.²²⁹

20.8 Tipos de gráficos

20.8.1 Quais são os tipos de gráficos para variáveis categóricas?

- Gráfico de barras: Mais usado para comparar frequências absolutas ou relativas entre categorias.²³⁰

Gráfico de Barras Simples

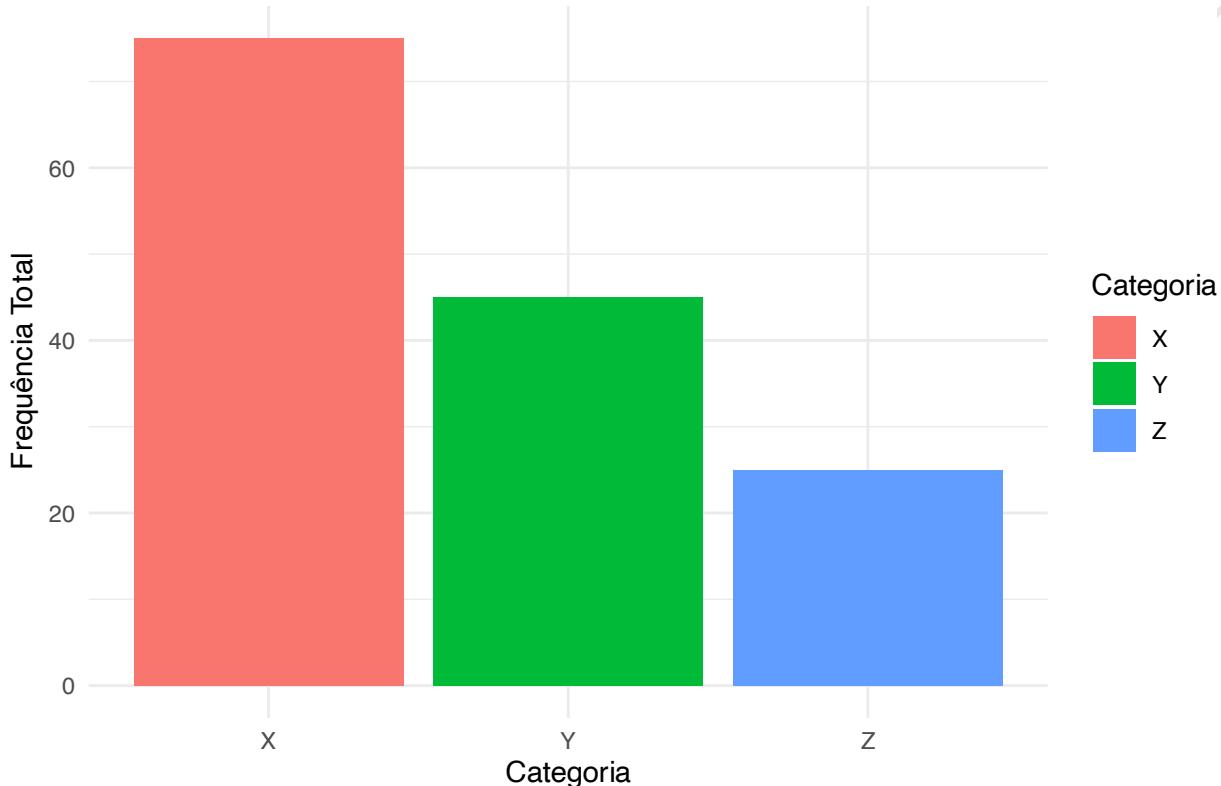


Figura 20.4: Gráfico de barras simples representando frequências por categoria.

- Gráfico de barras empilhadas: Útil para comparar proporções entre grupos em mais de uma variável categórica.²³¹

20.8.2 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).²³²
- Gráfico de densidade: Similar ao histograma, mas mais suave. Útil para avaliar a distribuição.²³³
- Diagrama de caixa (*boxplot*): Resume mediana, quartis e valores extremos. Excelente para comparar grupos.²³⁴

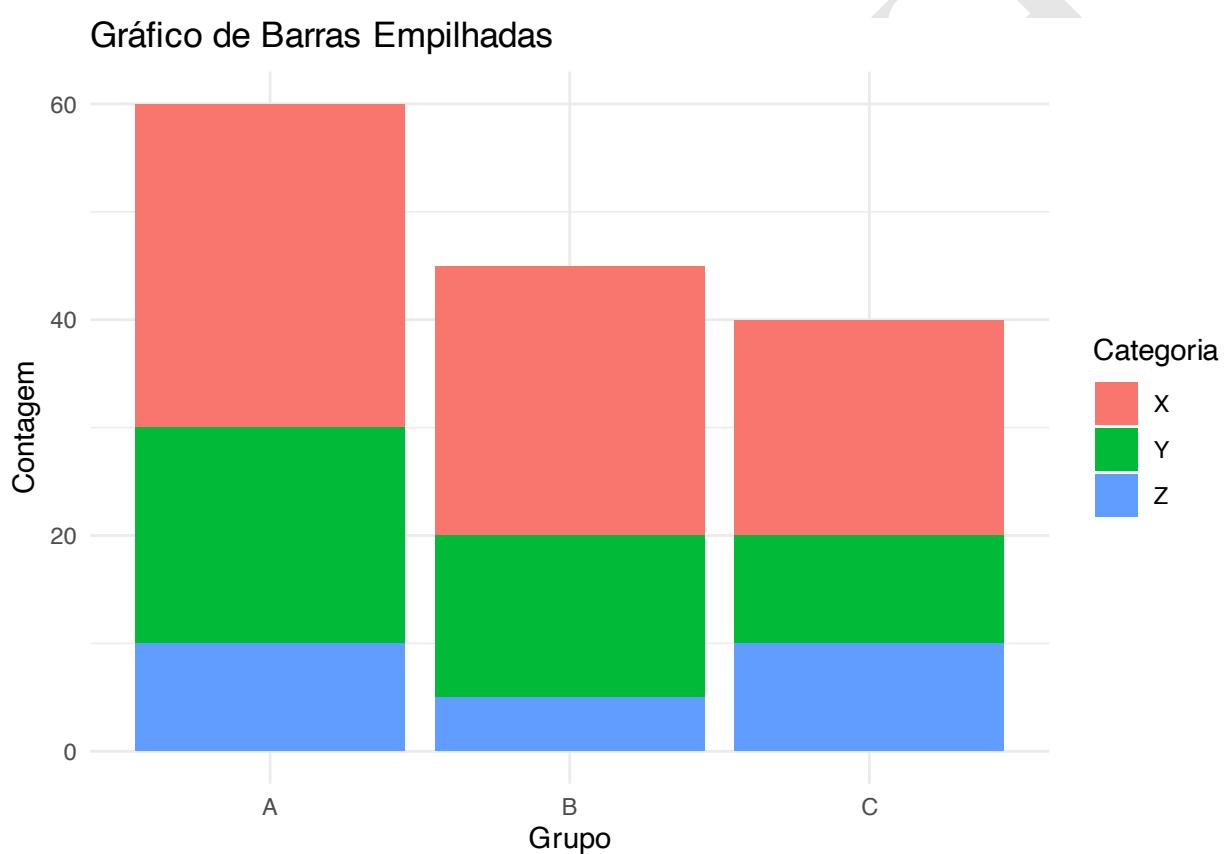


Figura 20.5: Gráfico de barras empilhadas representando frequências por categoria.

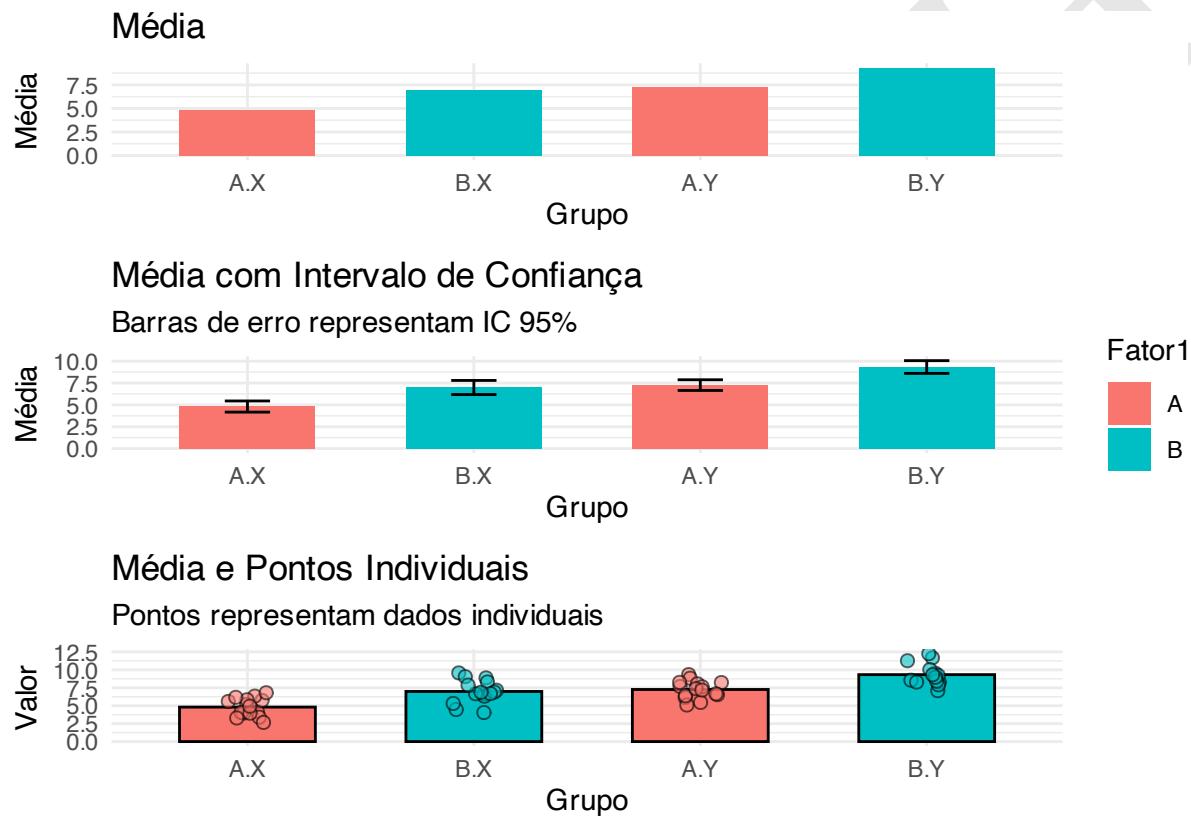


Figura 20.6: Gráficos de barras representando médias, barras de erro e dados individuais.

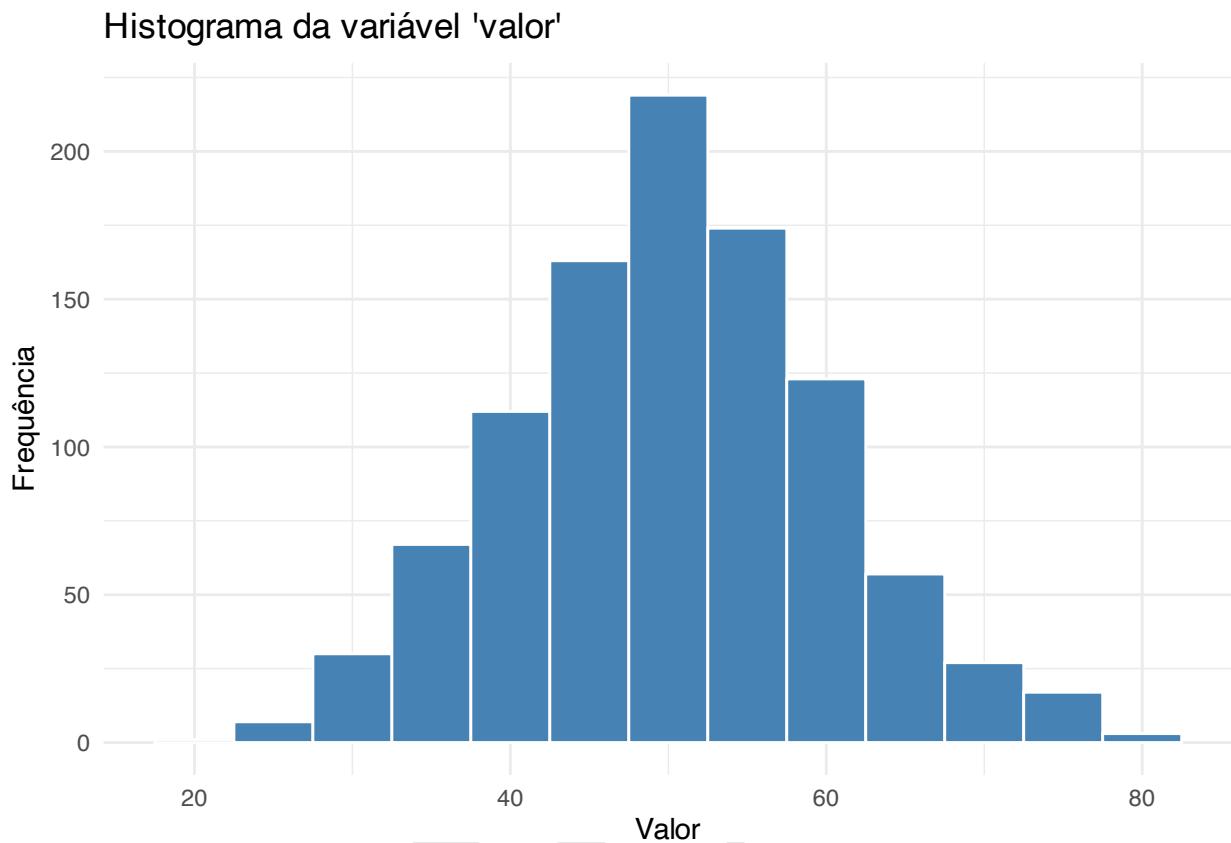


Figura 20.7: Histograma da variável 'valor'.

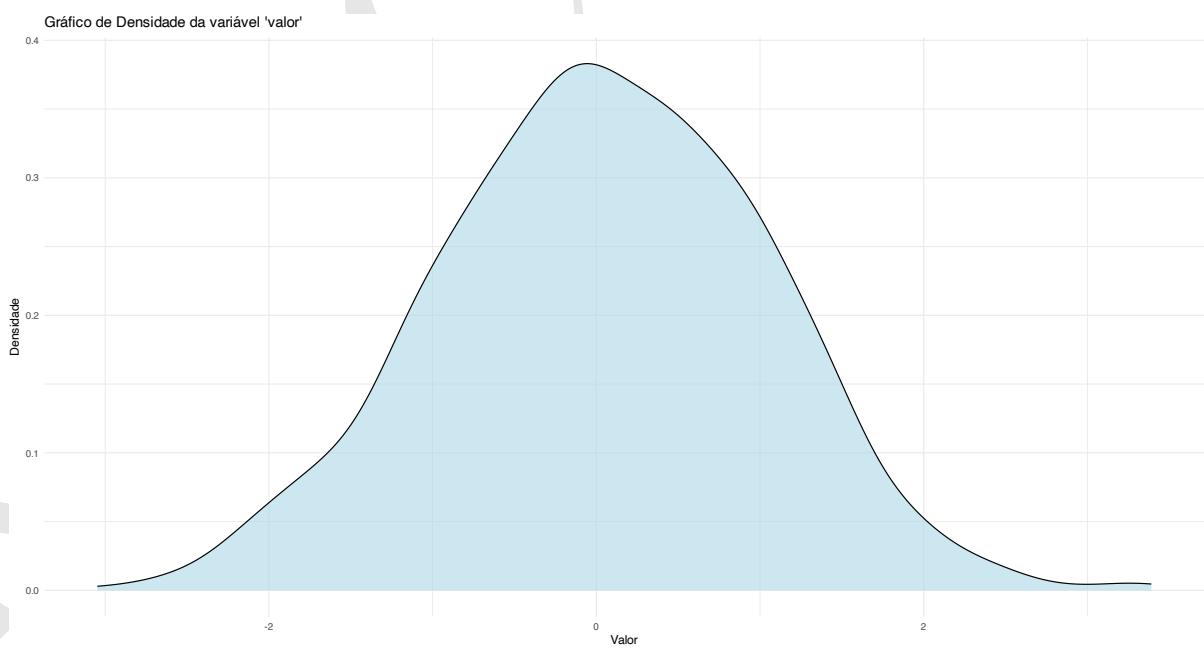


Figura 20.8: Gráfico de densidade da variável 'valor'.

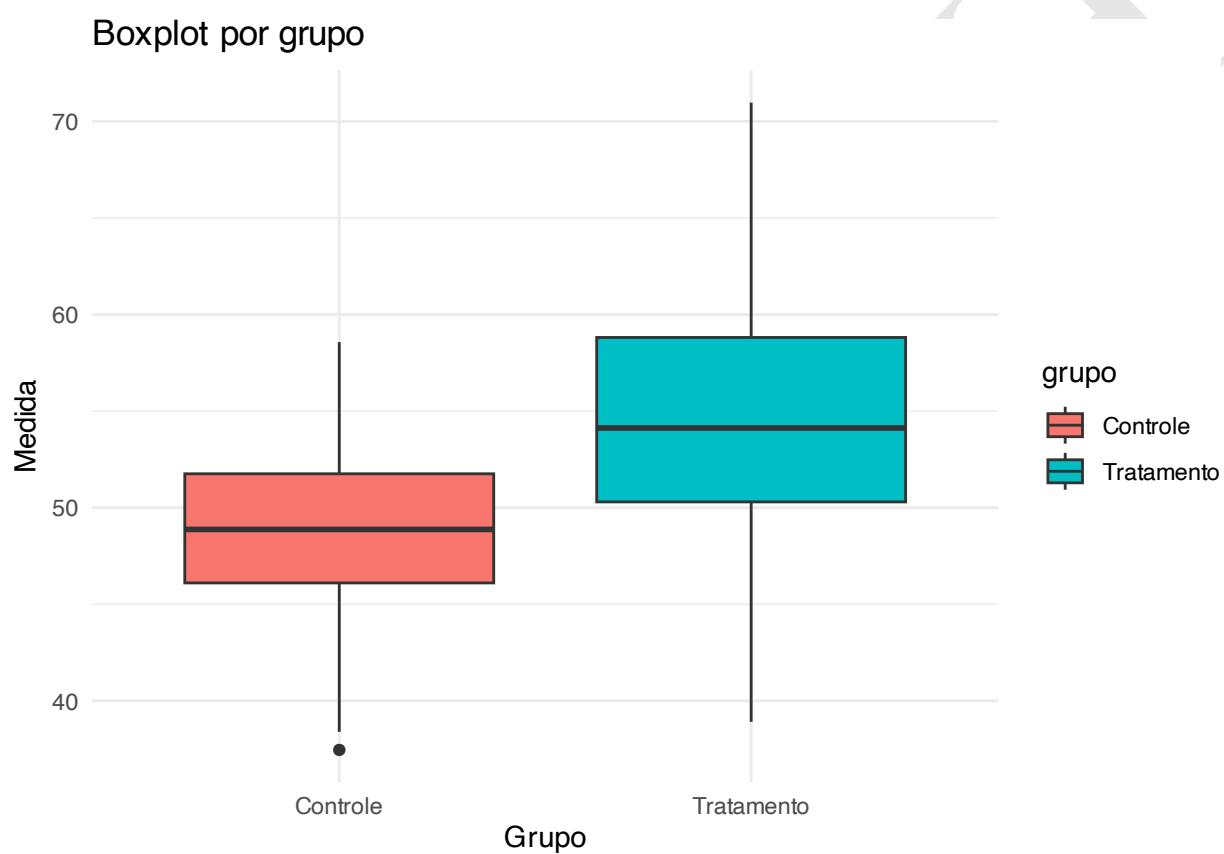


Figura 20.9: Boxplot por grupo.

- Gráfico de violino: Combina boxplot e densidade, mostrando a distribuição da variável. Útil para comparar grupos.²³¹

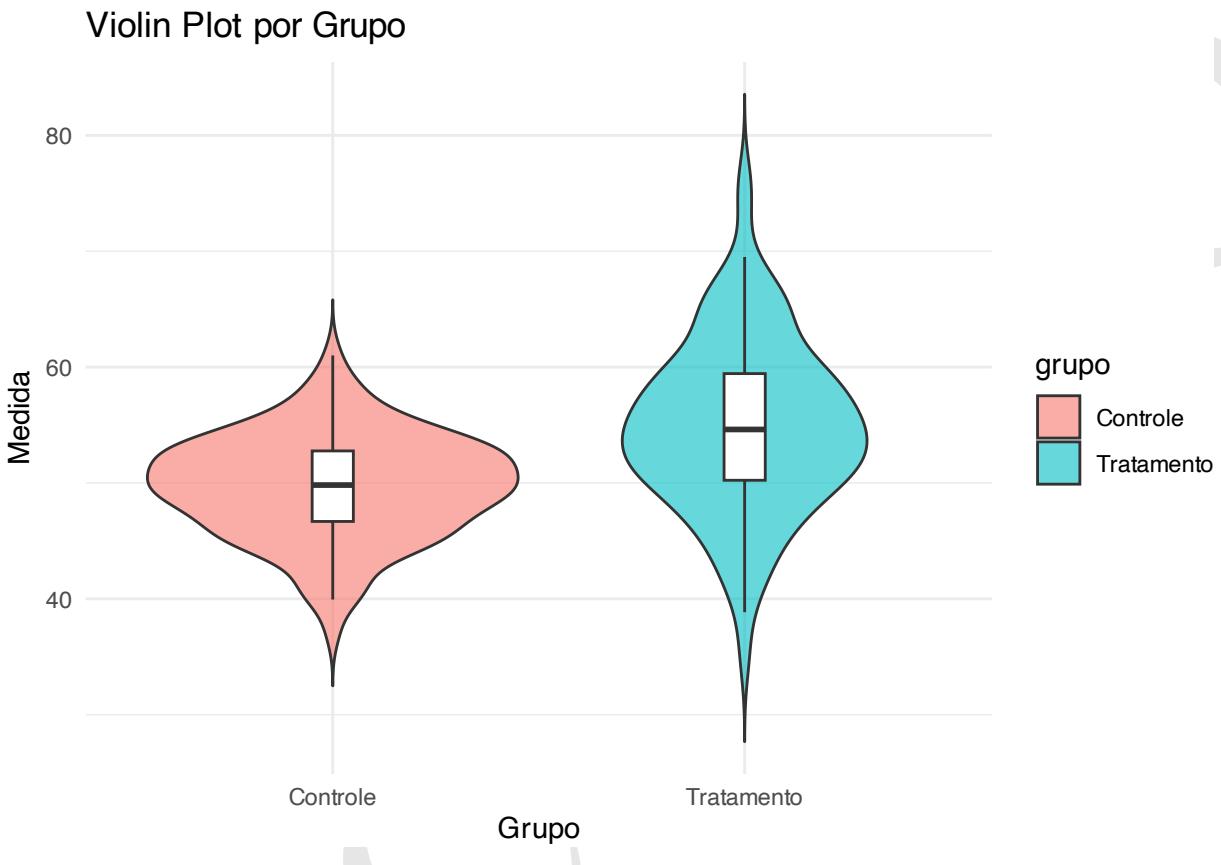


Figura 20.10: Violin plot por grupo.

- Gráfico de pontos (*dot plot*): Mostra cada valor individualmente, útil para pequenas amostras e para visualizar a distribuição.²³²

20.8.3 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.²³³
- Gráfico de bolhas (*bubble chart*): Expande o gráfico de dispersão adicionando uma terceira variável (tamanho da bolha).²³⁴
- Gráfico Sankey: Visualiza fluxos entre categorias em diferentes etapas ou grupos. Útil para mostrar proporções e transições.²³⁵
- Gráfico de *parcats*: Mostra relações entre múltiplas variáveis categóricas em paralelo. Útil para visualizar fluxos e proporções.²³⁶
- Gráfico de *parts*: Mostra a composição percentual de uma variável categórica. Útil para visualizar proporções.²³⁷

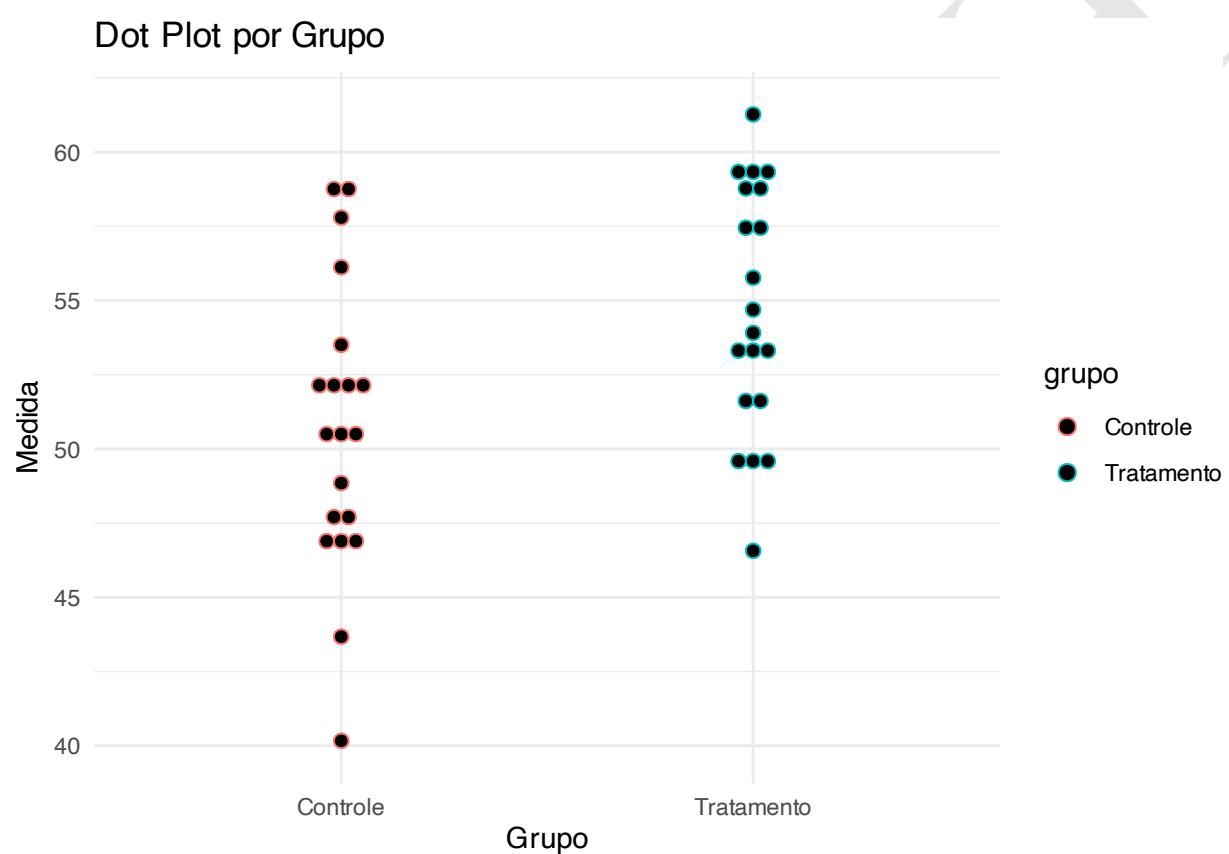


Figura 20.11: Gráfico de pontos da variável 'valor'.

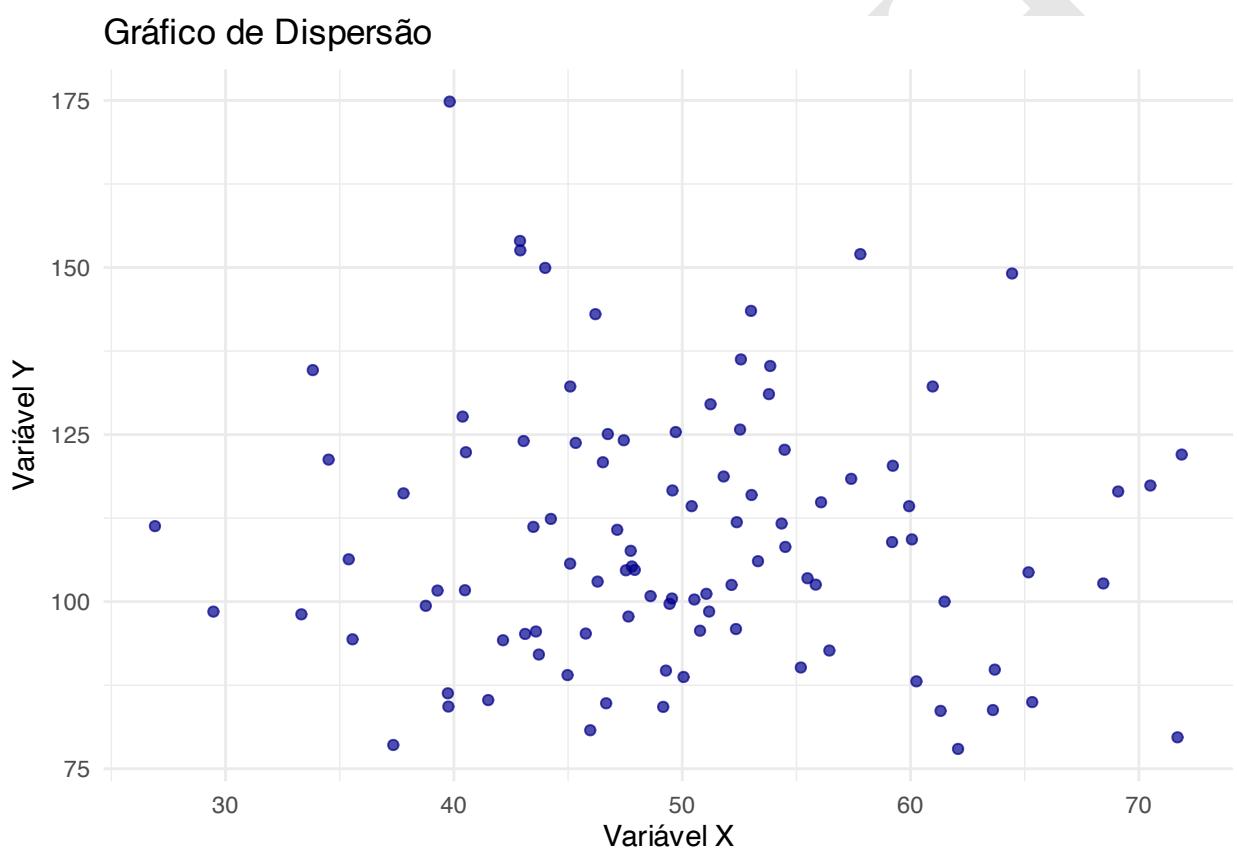


Figura 20.12: Gráfico de dispersão representando a relação entre duas variáveis.

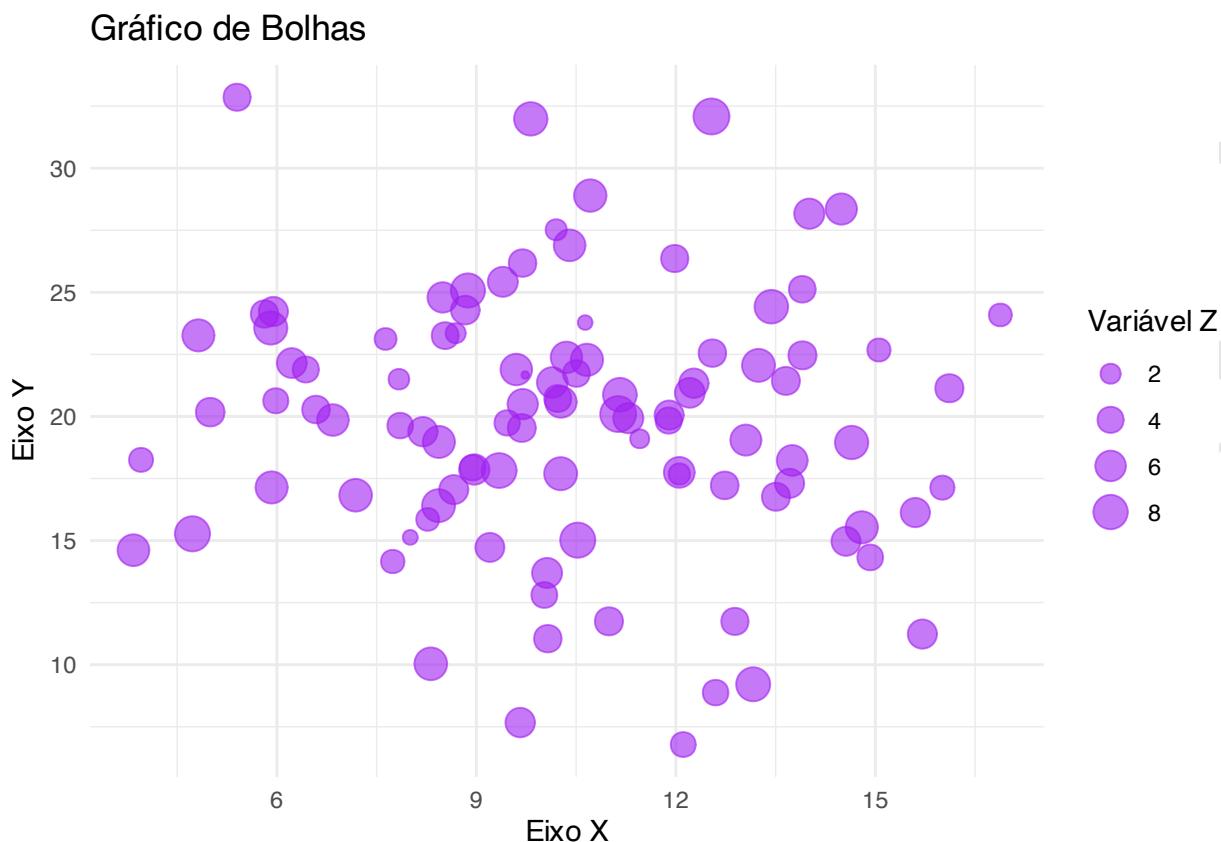


Figura 20.13: Gráfico de bolhas representando a relação entre três variáveis.

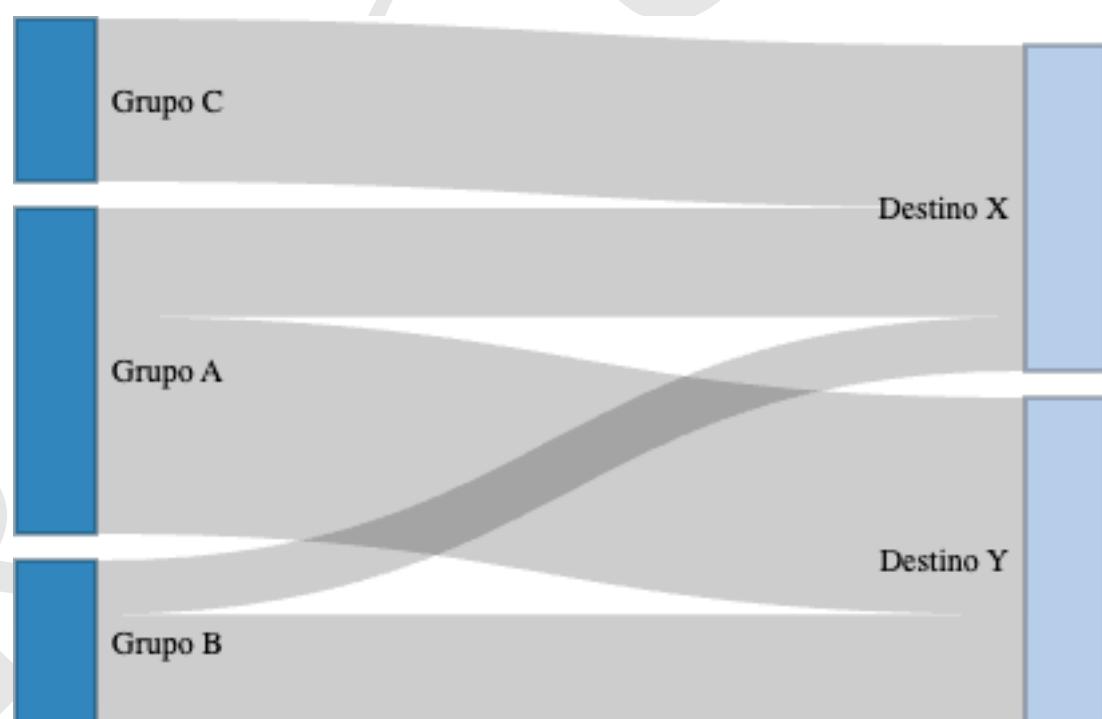


Figura 20.14: Sankey plot representando fluxos entre categorias.

Parallel Categories • parcats (plotly)

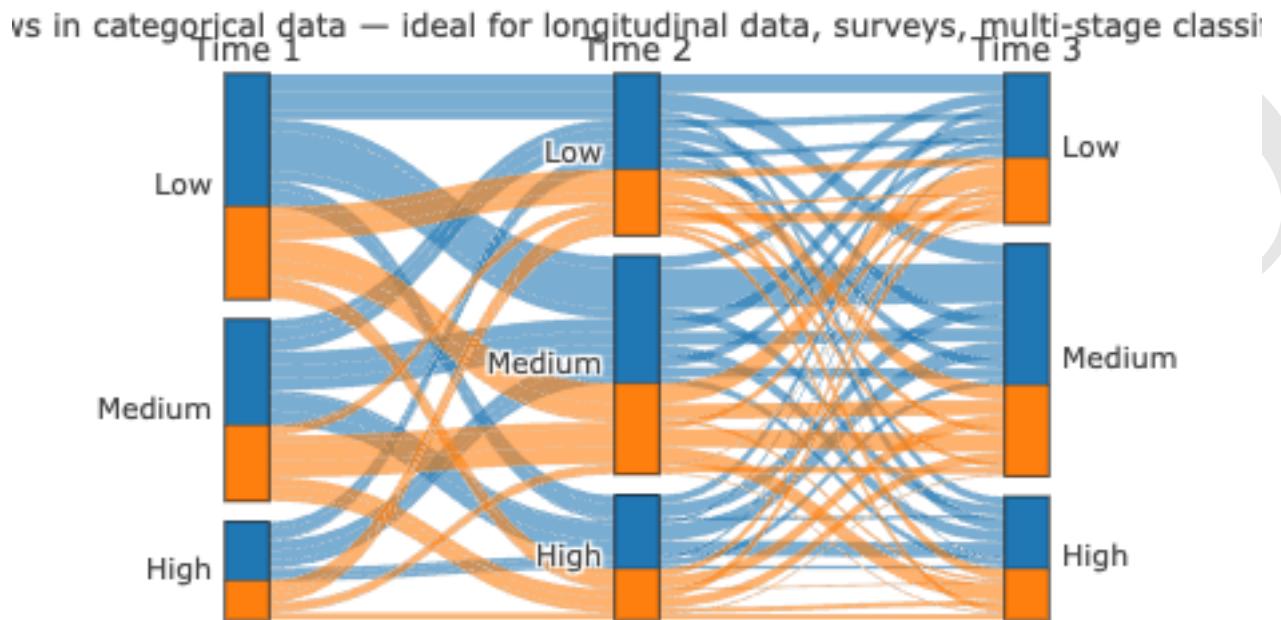


Figura 20.15: Gráfico de categorias paralelas (parcats) representando transições entre categorias ao longo do tempo.

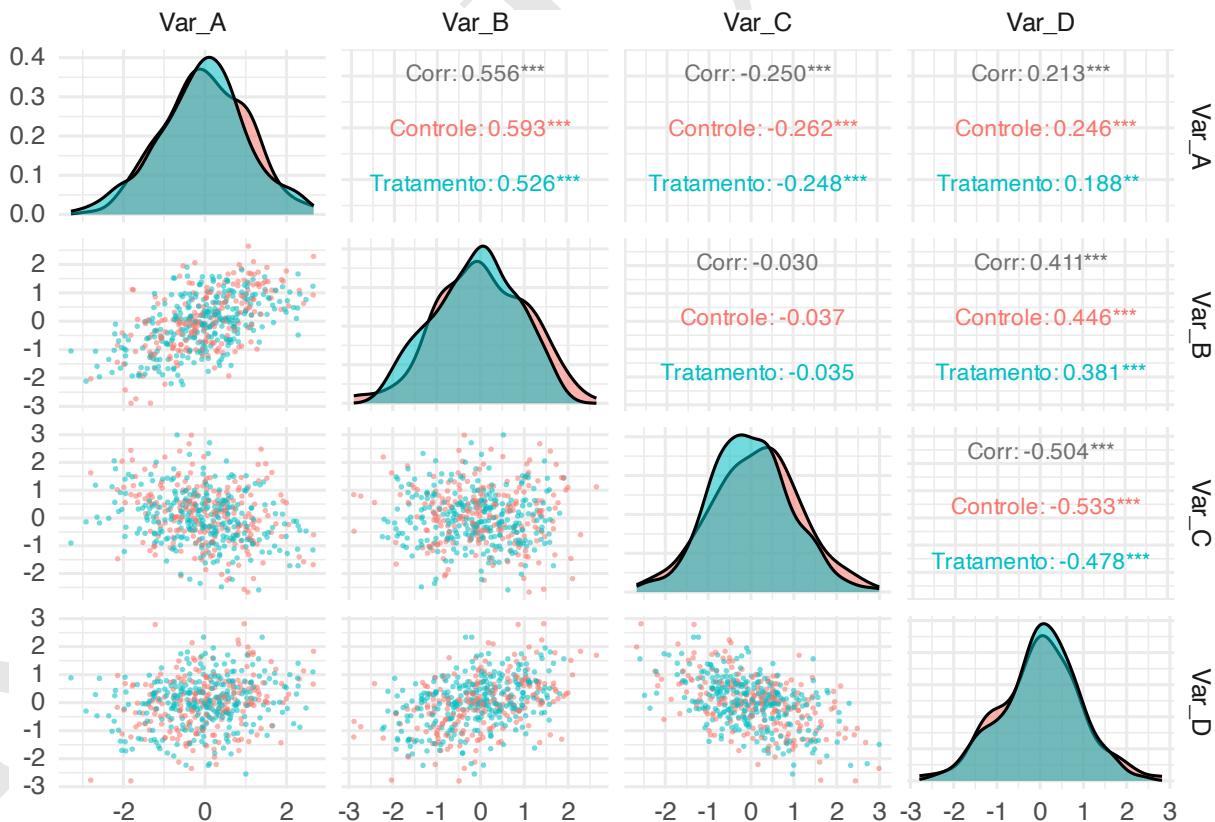


Figura 20.16: Gráfico de pares representando correlações entre múltiplas variáveis.

20.8.4 Quais são os tipos de gráficos para dados longitudinais?

- Gráfico de *spaghetti*: Mostra trajetórias individuais ao longo do tempo, útil para dados longitudinais de variáveis contínuas.^{232,233}

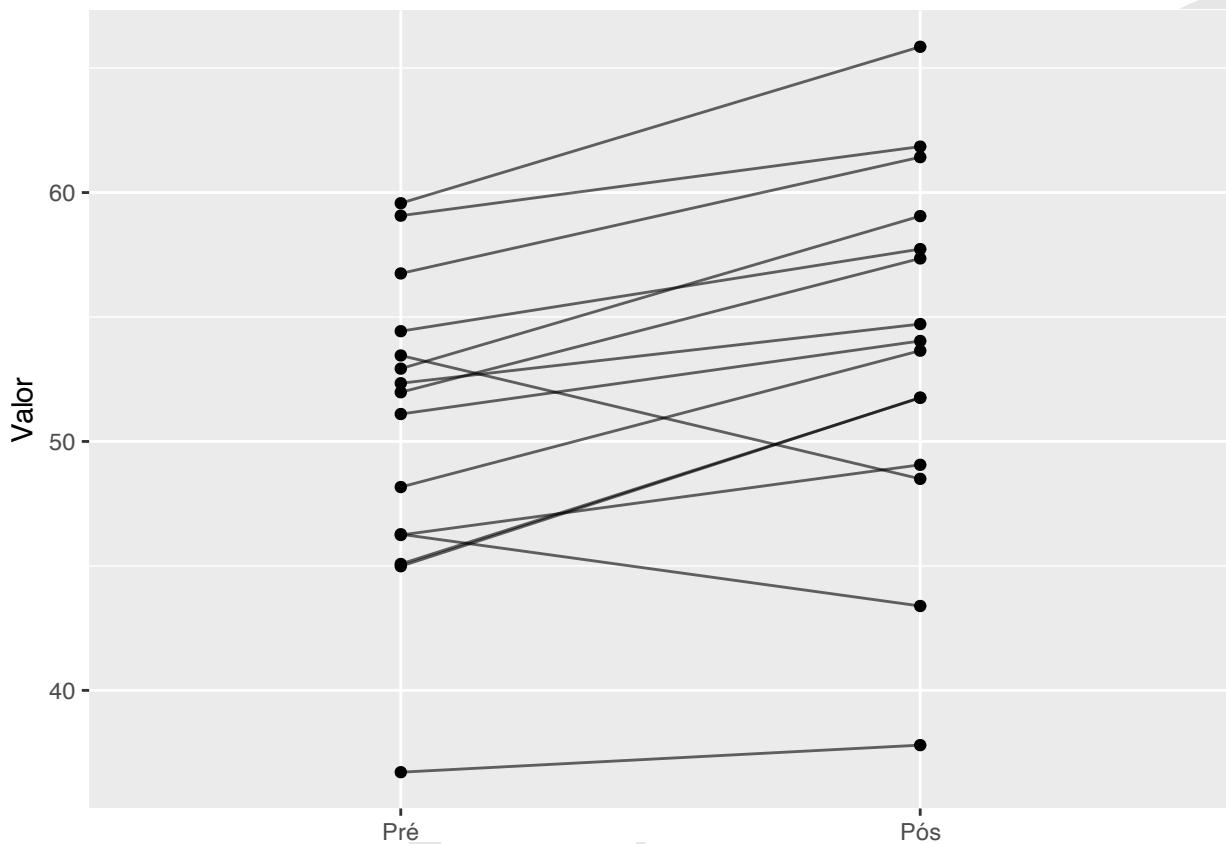


Figura 20.17: Gráfico spaghetti representando dados longitudinais.

- Gráfico de *lasagna*: Mostra trajetórias individuais ao longo do tempo, útil para dados longitudinais de variáveis categóricas.^{232,233}

20.8.5 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

20.8.6 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

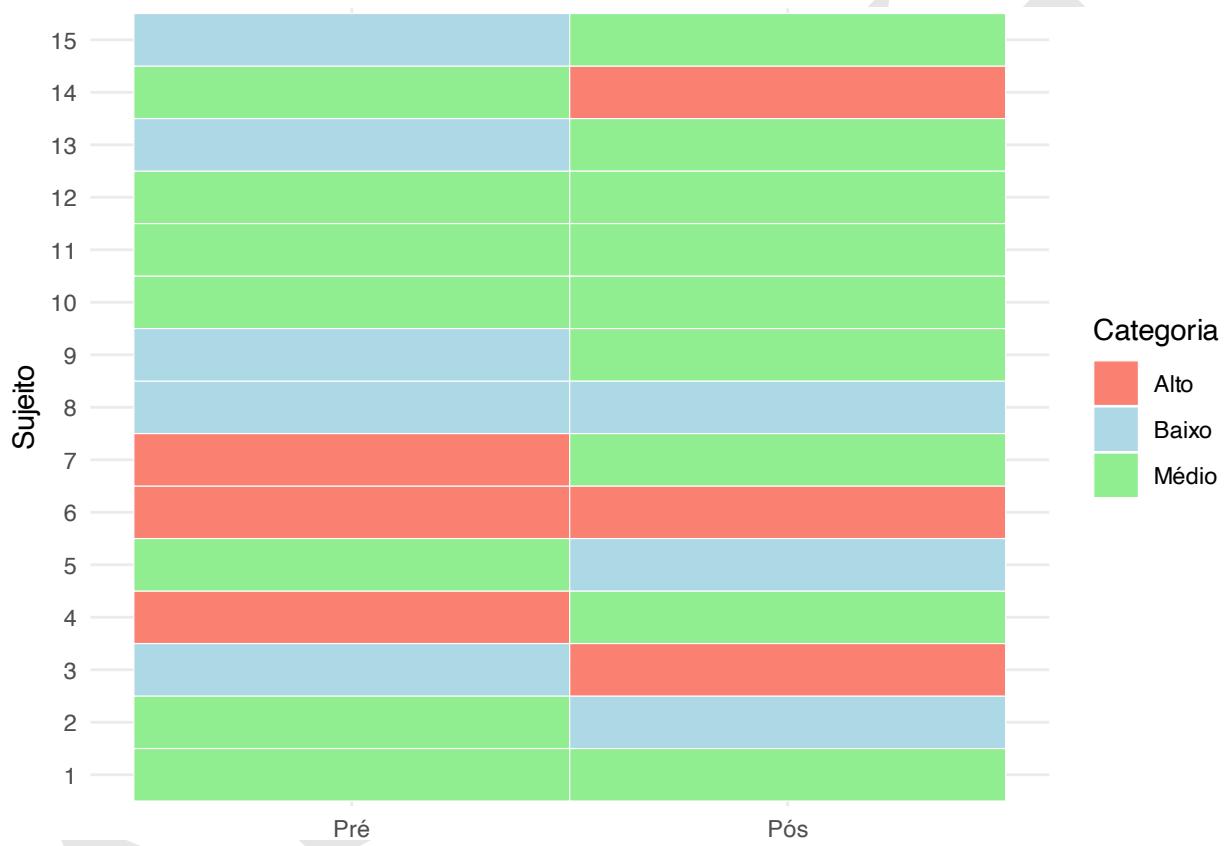


Figura 20.18: Gráfico spaghetti representando dados longitudinais.

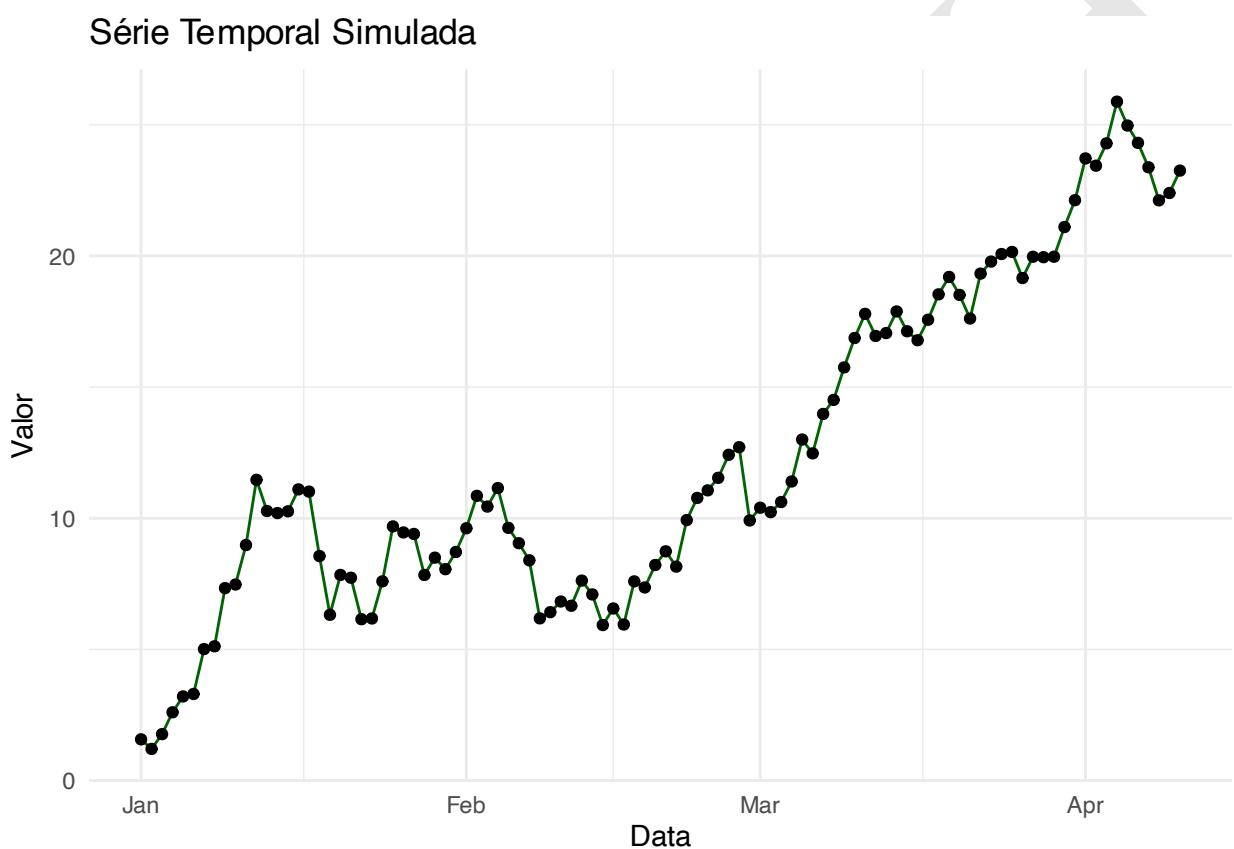


Figura 20.19: Gráfico de linha representando uma série temporal.

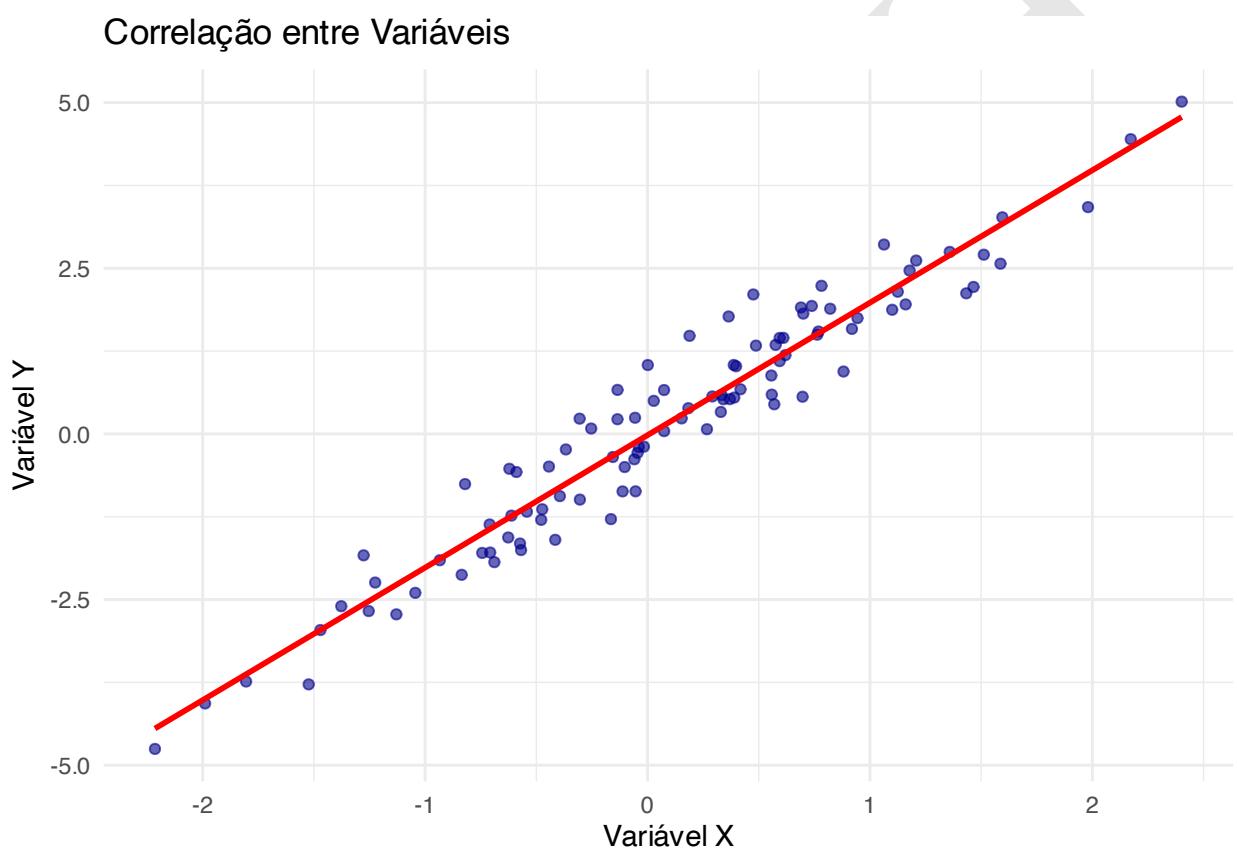


Figura 20.20: Gráfico de correlação entre duas variáveis com linha de tendência.

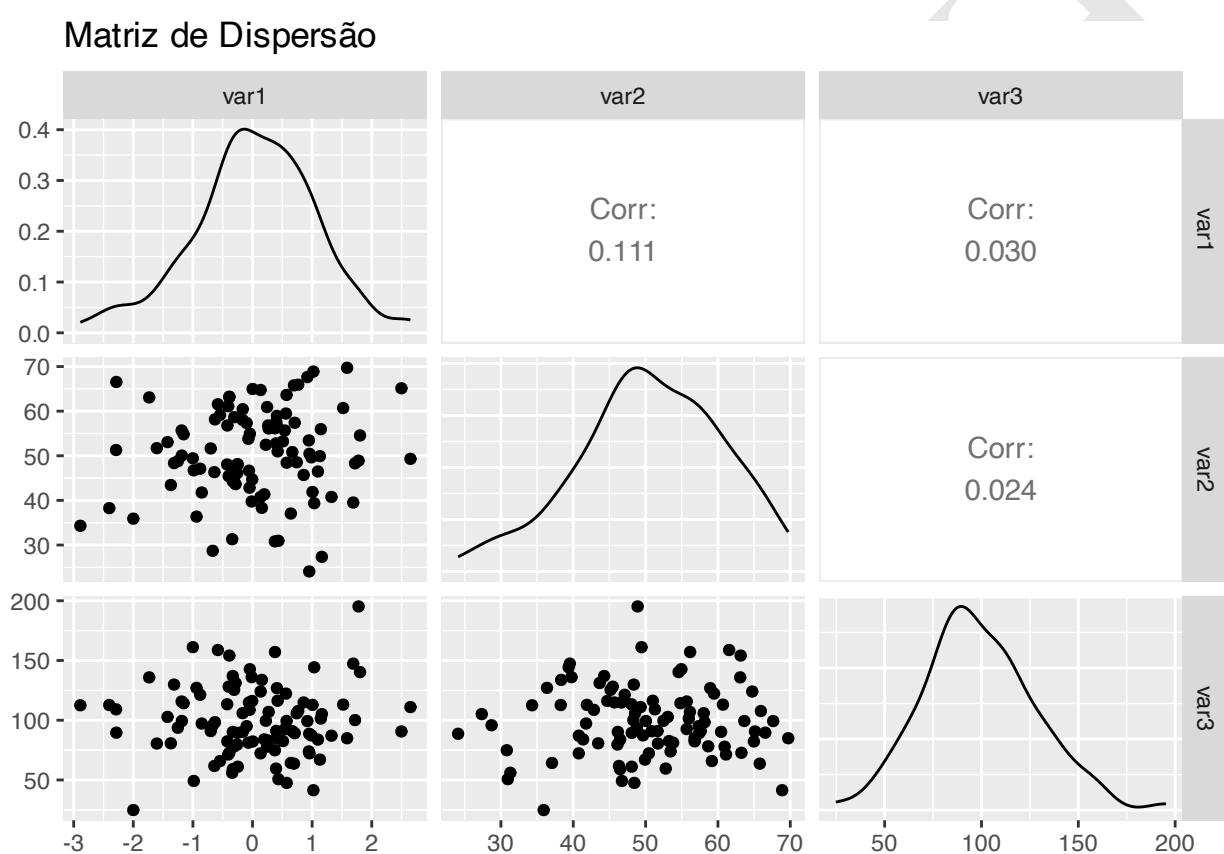


Figura 20.21: Matriz de dispersão representando relações entre múltiplas variáveis.

Mapa de Calor da Correlação entre Variáveis

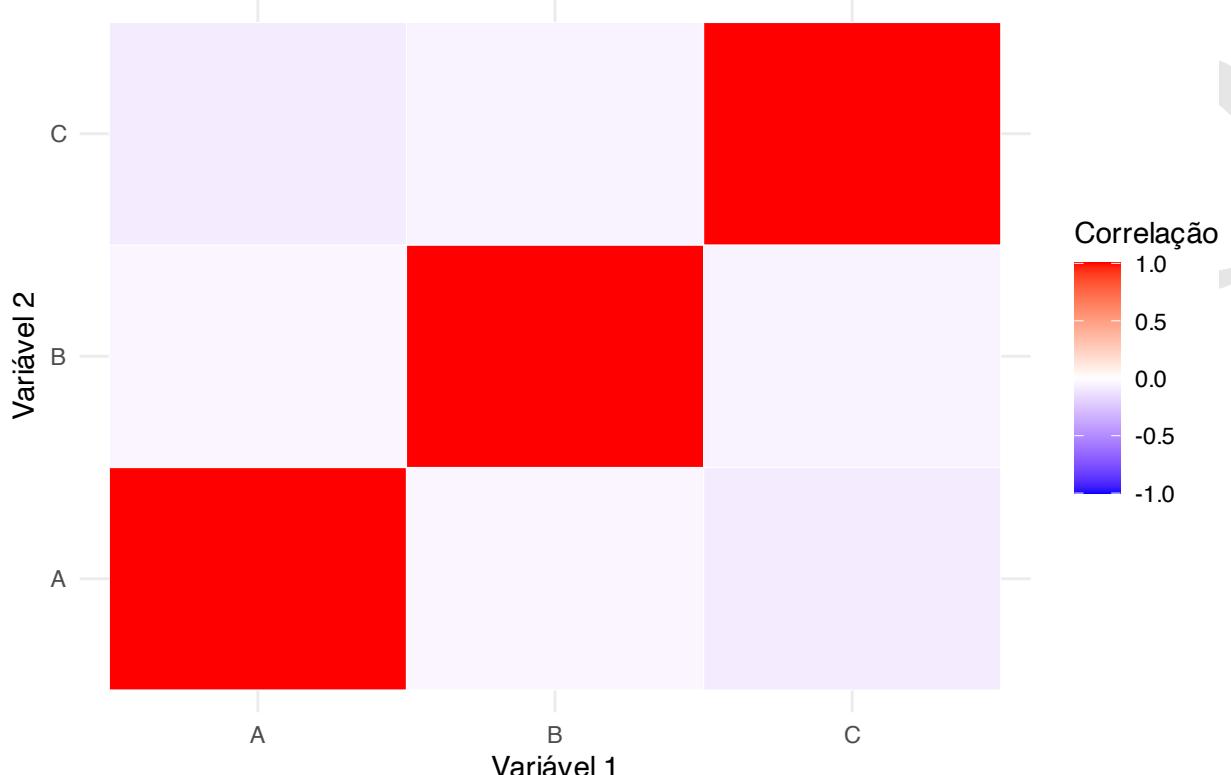


Figura 20.22: Mapa de calor da correlação entre variáveis.

Gráfico de Radar - Perfis Simulados

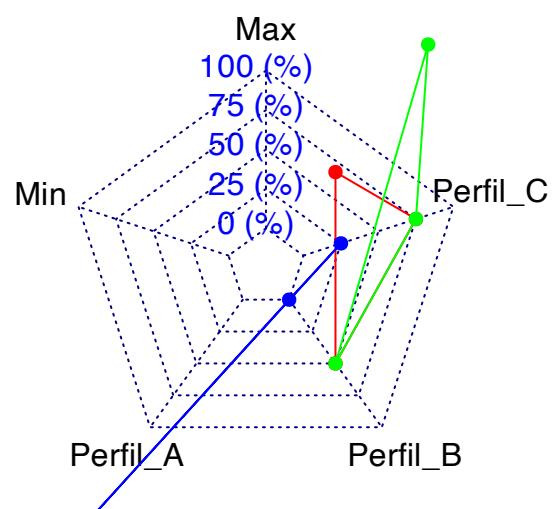


Figura 20.23: Gráfico radar representando múltiplas variáveis.

20.8.7 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.¹⁸⁹
- 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.^{188,189}
- Evite gráficos de barra e mostre a distribuição dos dados sempre que possível.²¹⁴
- Exiba os pontos de dados em boxplots.²¹⁴
- Use *jitter* simétrico em gráficos de pontos para permitir a visualização de todos os dados.²¹⁴
- Prefira palhetas de cor adaptadas para daltônicos.²¹⁴
- 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.²²⁷
- Evite gráficos de barras com médias para variáveis contínuas; prefira pontos/box/violino e, em amostras pequenas, exiba todos os dados.²¹⁴
- 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?²²⁷

R

O pacote *ggsci*²³⁴ 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

R

O pacote *grDevices*¹⁷⁸ 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>

R

O pacote *tiff*²³⁵ fornece a função *writeTIFF*^a para exportar gráficos em formato TIFF.

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

20.9 Gráficos dinâmicos

20.9.1 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.²³⁶

20.9.2 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.²³⁶

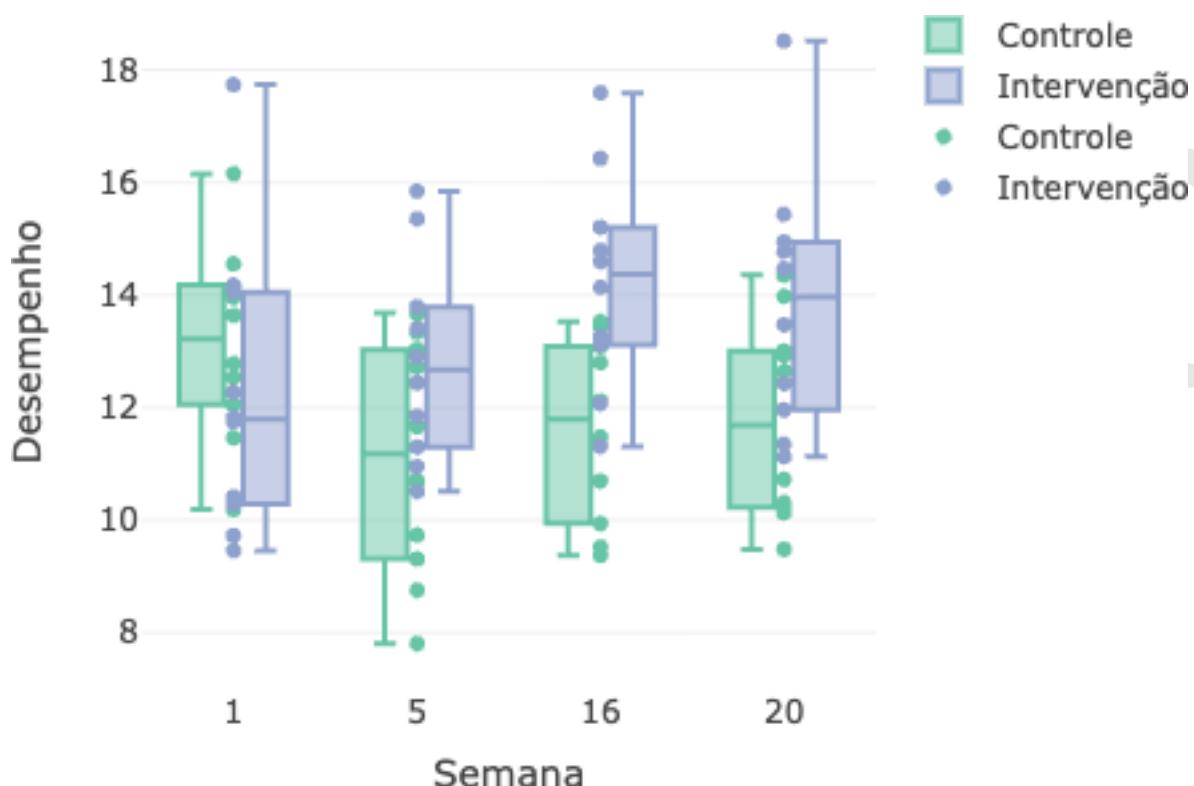


Figura 20.24: Exemplo de gráfico interativo com Plotly.

 O pacote *plotly*²³⁰ fornece a função *plot_ly*^a para gerar gráficos interativos.

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

20.9.3 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.²³⁶

 O pacote *ggnimate*²³⁷ fornece a função *transition_states*^a para criar gráficos animados a partir de gráficos estáticos do *ggplot2*¹⁷⁹.

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

RASCUNHO

Capítulo 21

Análise robusta

21.1 Raciocínio inferencial robusto

21.1.1 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.²³⁸

21.1.2 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.²³⁸
- Testar previamente as suposições não é suficiente: testes de homocedasticidade têm baixo poder e não garantem segurança analítica.²³⁸
- Métodos estatísticos robustos oferecem uma solução mais segura e eficaz, lidando melhor com dados não ideais.²³⁸

21.1.3 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.²³⁸
- 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.²³⁸

21.1.4 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.¹⁹⁶
- 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.¹⁹⁶

21.2 Valores discrepantes

21.2.1 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.²⁰¹
- Um valor discrepante é uma observação incomum que exerce influência indevida em uma análise.²⁰¹
- Valores discrepantes são dados com valores altos de resíduos.¹⁹⁴
- 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

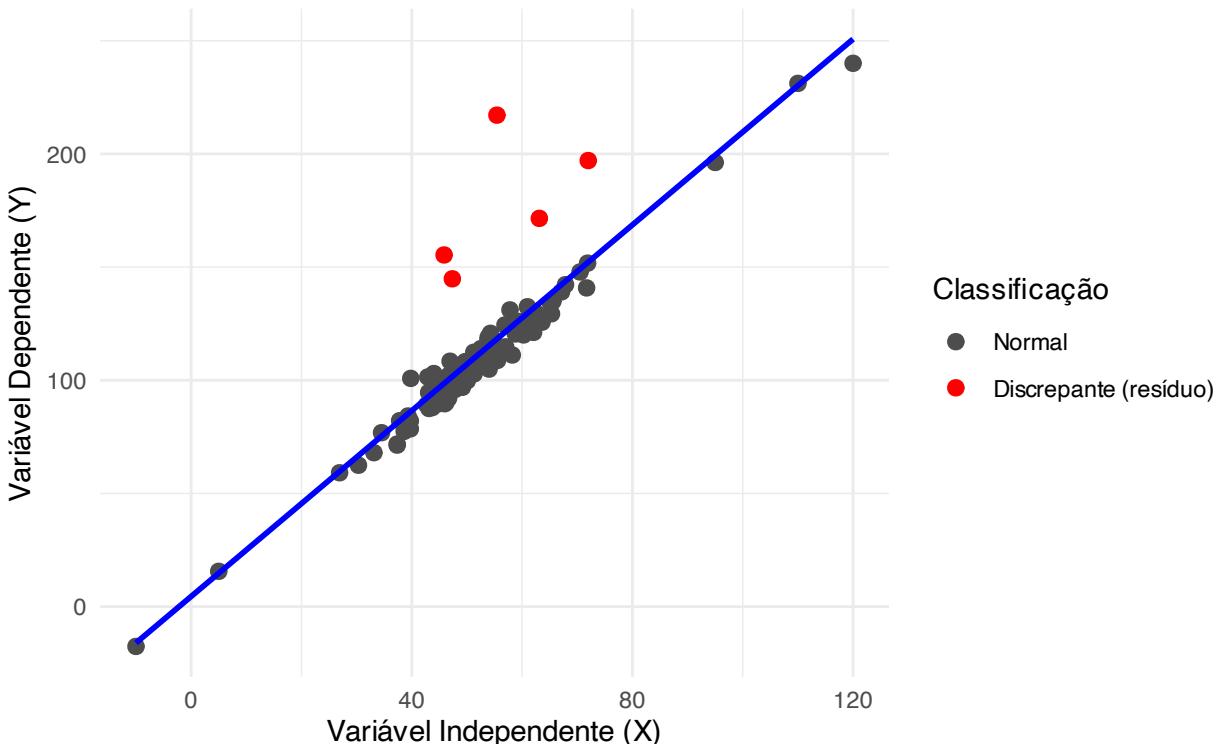


Figura 21.1: Régressão linear com valores discrepantes

21.2.2 Quais são os tipos de valores discrepantes?

- Valores discrepantes podem ser categorizados em três subtipos: *outliers* de erro, *outliers* interessantes e *outliers* aleatórios.¹⁹⁴
- Os valores discrepantes de erro são observações claramente não legítimas, distantes de outros dados devido a imprecisões por erro de mensuração e/ou codificação.¹⁹⁴
- Os valores discrepantes interessantes não são claramente erros, mas podem refletir um processo/mecanismo potencialmente interessante para futuras pesquisas.¹⁹⁴

- Os valores discrepantes aleatórios são observações que resultam por acaso, sem qualquer padrão ou tendência conhecida.¹⁹⁴
- Valores discrepantes podem ser univariados ou multivariados.¹⁹⁴

21.2.3 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.¹⁹⁴
- 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.¹⁹⁴
- 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).¹⁹⁴

21.2.4 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.¹⁹⁴
- 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.¹⁹⁴
- 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.¹⁹⁴
- A detecção ou o tratamento dos *outliers* não deve ser realizada após a análise dos resultados, pois isso introduz viés nos resultados.¹⁹⁴

21.2.5 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).¹⁹⁴
- 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).^{194,239}
- 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%).¹⁹⁴
- 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*.^{194,240}

21.2.6 Quais testes são apropriados para detectar valores discrepantes?

- A escolha do método de detecção depende da natureza do outlier, se univariado ou multivariado.¹⁹⁶
- 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.¹⁹⁶

- 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*).¹⁹⁶
- Técnicas baseadas em PCA robusta (ROBPCA, PP-PCA, SPCA, EPCA) também podem ser aplicadas para reduzir dimensionalidade e expor *outliers* mascarados.¹⁹⁶
- Métodos de *trimming* multivariado (MVT) podem iterativamente remover observações mais distantes, mas apresentam limitações em alta dimensionalidade.¹⁹⁶
- Estimadores com alto ponto de quebra, como o MCD, permitem detectar até 50% de *outliers* antes de comprometer a análise.¹⁹⁶

21.2.7 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*.¹⁹⁴
- 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.¹⁹⁴
- Remover *outliers* leva à perda de observações, especialmente em conjuntos de dados com muitas variáveis, quando *outliers* univariados são excluídos em cada variável.¹⁹⁴
- Recodificar *outliers* evita a perda de uma grande quantidade de dados, mas deve ser baseada em argumentos razoáveis e convincentes.¹⁹⁴
- Erros de observação e de medição são uma justificativa válida para descartar observações discrepantes.²⁰¹

21.2.8 Como conduzir análises com valores discrepantes?

- É importante reportar se existem valores discrepantes e como foram tratados.²⁰¹
- Valores discrepantes na variável de desfecho podem exigir uma abordagem mais refinada, especialmente quando representam uma variação real na variável que está sendo medida.²⁰¹
- 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.²⁰¹
- Valores discrepantes podem ser recodificados usando a Winsorização,²⁴¹ que transforma os *outliers* em valores de percentis específicos (como o 5º e o 95º).¹⁹⁴

R

O pacote *outliers*²⁴² 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*²⁴² 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>

21.2.9 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.¹⁹⁴
- 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.¹⁹⁴

21.3 Valores influentes

21.3.1 O que são valores influentes?

- Valores influentes são observações que, se removidas, causariam uma mudança significativa nos resultados da análise estatística.¹⁹⁵

Régressão linear com valores influentes

Influente = Cook > 0.036

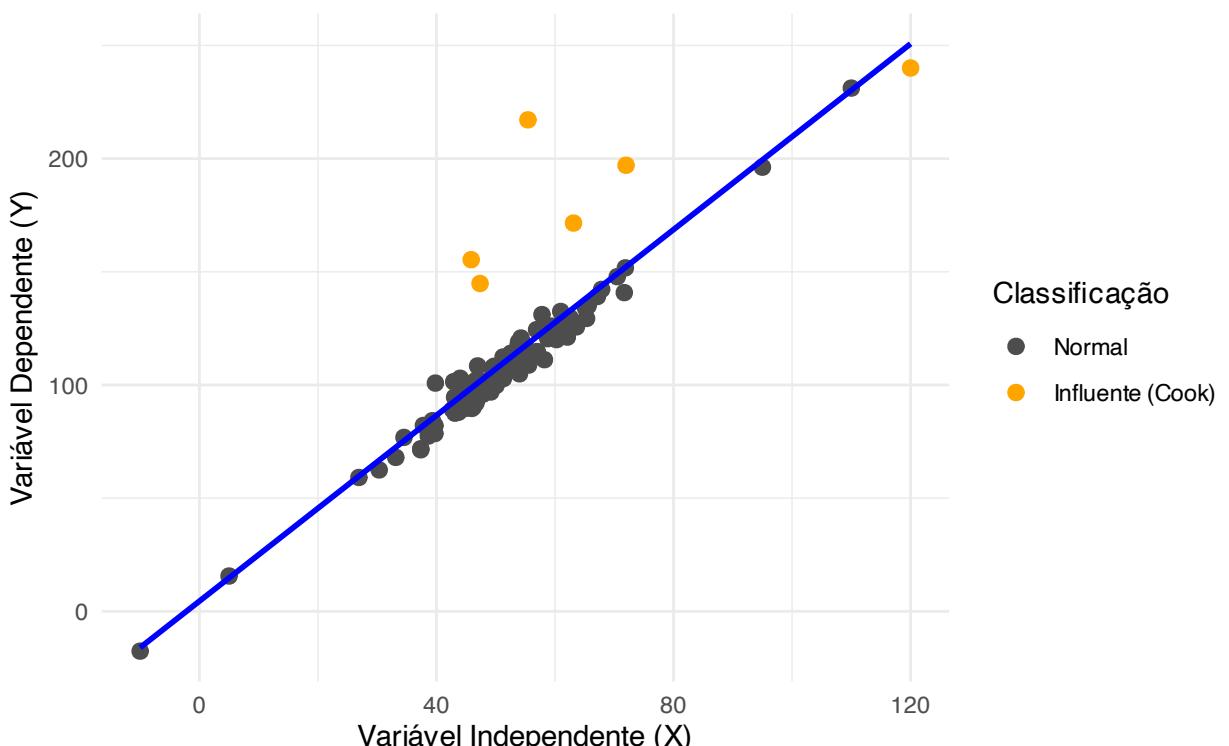


Figura 21.2: Regressão linear com valores influentes.

21.3.2 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.²⁴³

21.3.3 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%).²⁴³

21.3.4 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.²⁴⁴

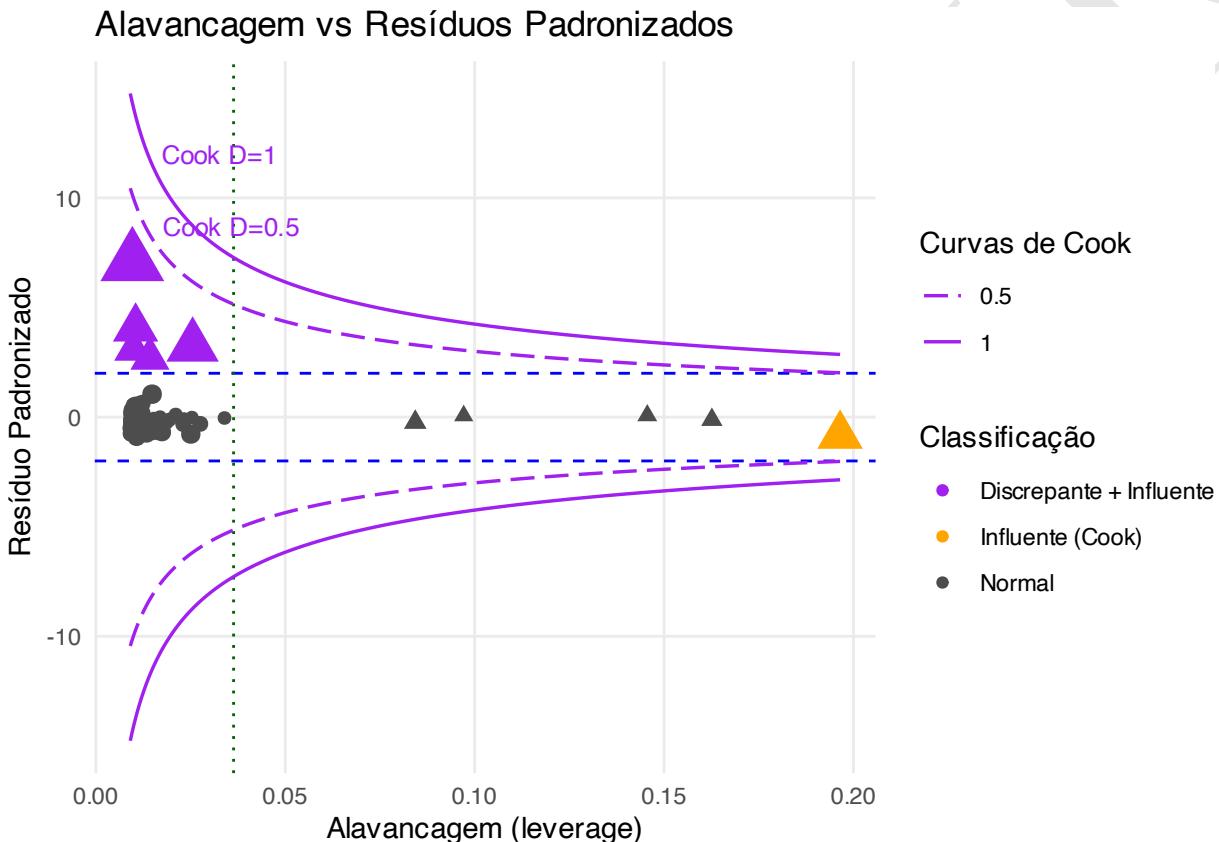


Figura 21.3: Alavancagem vs Resíduos Padronizados com distância de Cook para análise da influência de pontos.

21.4 Métodos robustos de tratamento de *outliers*

21.4.1 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.²³⁸

21.4.2 Quais são as alternativas à Winsorização?

- Podar (*trimming*): remove diretamente uma fração fixa das observações mais extremas.²⁴⁵

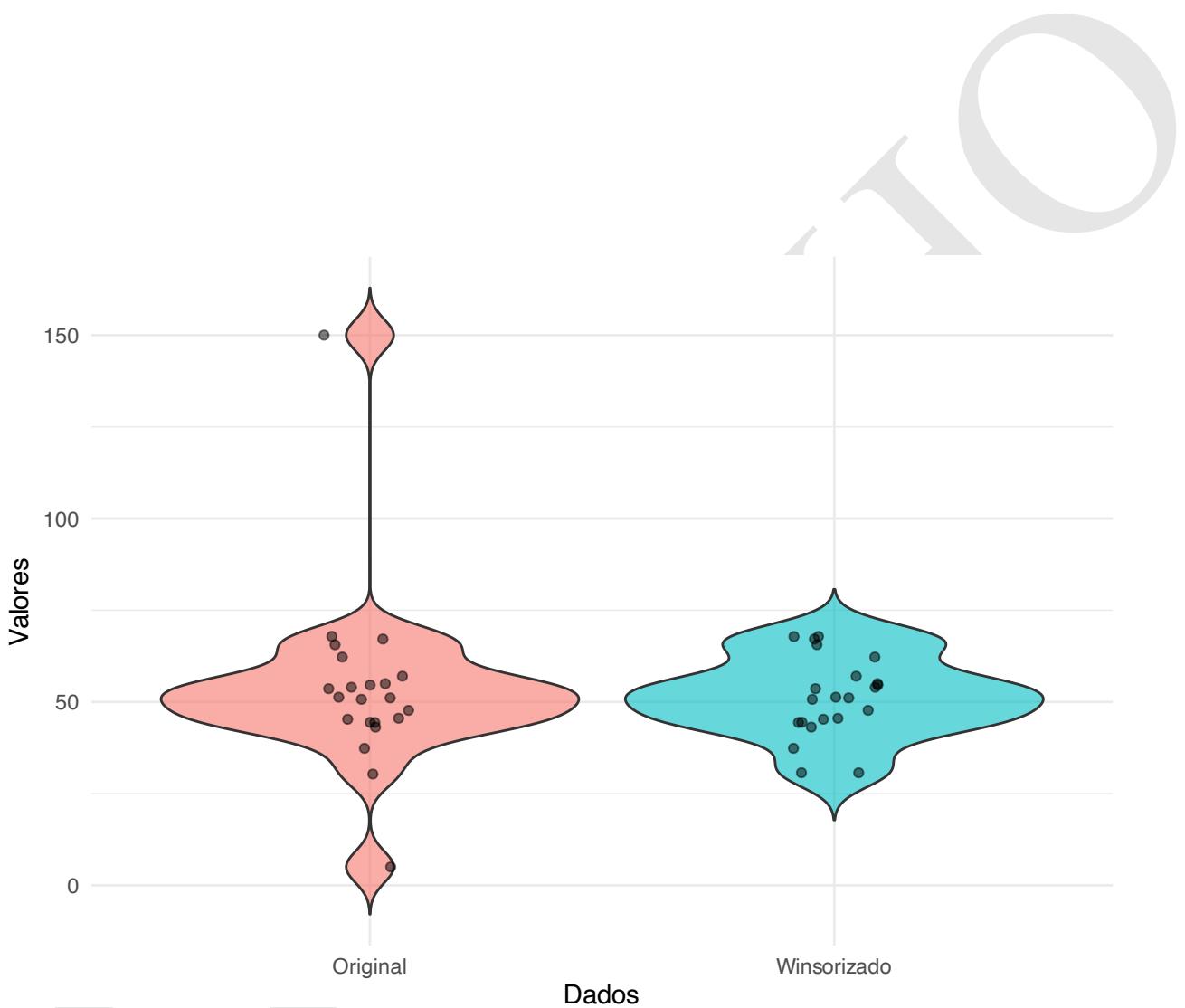


Figura 21.4: Boxplots comparando dados originais e dados Winsorizados.

- 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*²⁴⁴ 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*²⁴⁴ 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*²⁴⁴ 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*²⁴⁴ 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 22

Análise preditiva

22.1 Predição

22.1.1 O que são previsões?

- ?



O pacote *ggeffects*²⁴⁵ 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*²⁴⁵ 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

22.1.2 Como árvores de decisão são usadas para previsão?

- Utilizam padrões históricos para prever resultados futuros em novos registros.²⁴⁶ Identificam combinações de fatores que elevam ou reduzem o risco de um evento clínico.²⁴⁶
- Podem ser aplicadas em diagnósticos médicos para prever subtipos de condições ou diferentes respostas terapêuticas.²⁴⁶

22.2 Interpretação e aplicação

22.2.1 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.²⁴⁶
- Auxiliam na formulação de hipóteses clínicas e direcionamento de futuras pesquisas.²⁴⁶
- Podem ser estendidas para previsão de sobrevida, subgrupos de tratamento e análise de custo-benefício.²⁴⁶

22.3 Análise de curva de decisão

22.3.1 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”.²⁴⁷

22.3.2 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.²⁴⁸
- O limiar reflete o equilíbrio entre o risco de perder casos da doença e o risco de intervenções desnecessárias.²⁴⁸

22.3.3 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.²⁴⁸
- Quando a estratégia padrão é “tratar todos”, o benefício líquido também pode ser expresso como intervenções desnecessárias evitadas, facilitando a interpretação clínica.²⁴⁸

22.3.4 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.²⁴⁸
- Um modelo ou teste só é considerado útil se apresentar maior benefício líquido do que ambas as estratégias, garantindo que realmente agrega valor clínico.²⁴⁸
- Essa comparação evita que um modelo com bom AUC mas má calibração seja usado, já que pode apresentar net benefit inferior às estratégias padrão.²⁴⁸

22.3.5 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.²⁴⁸
- Em situações claras (modelo com nenhum benefício ou benefício amplo), pode dispensar análises complexas.²⁴⁸
- Em casos ambíguos, serve como primeiro passo antes de análises mais detalhadas.²⁴⁸

22.3.6 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).²⁴⁷

22.3.7 Como funciona o cálculo do benefício líquido?

- O benefício líquido é estimado contrastando verdadeiros positivos e falsos positivos em diferentes limiares de decisão.²⁴⁷

Capítulo 23

Análise causal

23.1 Causalidade

23.1.1 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?”.²¹⁰
- 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.²¹⁰
- Análise causal refere-se ao processo conceitual e analítico, enquanto inferência causal diz respeito às conclusões extraídas a partir de dados.²¹⁰

23.1.2 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.²⁴⁹
- Mecanicista: busca entender processos internos e mecanismos. Estatística tem seu papel ainda limitado, mas crescente, especialmente em sistemas complexos.²⁴⁹

23.1.3 Como realizar uma análise causal?

- Formular explicitamente uma pergunta causal, definindo qual intervenção hipotética está sendo considerada e qual efeito se deseja estimar.²¹⁰
- Especificar a ordem temporal entre exposição, desfecho e demais variáveis, assegurando que a causa preceda o efeito.²¹⁰
- Tornar explícitas as suposições causais sobre o sistema em estudo, preferencialmente por meio de um diagrama causal (DAG).²¹⁰
- Identificar confundidores, mediadores e colisores com base na estrutura causal assumida, e não apenas em critérios estatísticos.²¹⁰
- Definir o conjunto mínimo de ajuste necessário para bloquear caminhos espúrios entre exposição e desfecho.²¹⁰
- 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.²¹⁰

- Avaliar limitações como confundimento residual, viés de seleção e erro de medida, discutindo seu impacto potencial nas inferências.²⁴⁸
- Interpretar os resultados em termos causais explícitos, distinguindo claramente associação de efeito causal.²⁴⁹

23.2 Abordagens filosóficas e estatísticas da causalidade

23.2.1 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.²⁴⁹

23.2.2 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.²⁴⁹
- Autores como Judea Pearl, Robins e Rubin trouxeram definições mais precisas, especialmente via modelos contrafactuals.²⁴⁹
- O uso de ensaios clínicos randomizados consolidou o papel da estatística em inferência causal aplicada.²⁴⁹

23.3 Ilusões de causalidade

23.3.1 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.²⁵⁰

23.3.2 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).²⁵⁰
- Alta frequência da causa: quanto mais vezes um comportamento ou tratamento é aplicado, mais coincidências com o desfecho ocorrem, aumentando a crença no efeito.²⁵⁰
- Coincidências causa–desfecho: damos peso desproporcional a casos em que causa e efeito ocorrem juntos, mesmo que sejam apenas coincidências.²⁵⁰

23.3.3 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).²⁵⁰
- Diminuir a frequência da causa (ex.: reduzir uso de um “remédio ineficaz” ajuda a perceber que o resultado ocorre independentemente).²⁵⁰
- Instruções explícitas para testar hipóteses: orientar a aplicar a causa em apenas 50% das vezes favorece a detecção correta da ausência de efeito.²⁵⁰
- Promover educação científica prática, mostrando às pessoas como seus próprios julgamentos podem ser enviesados e oferecendo ferramentas para avaliação crítica.²⁵⁰

23.4 Inferência causal em estudos observacionais

23.4.1 Como diferenciar associação de causalidade?

- Associação descreve que duas variáveis variam juntas, mas não garante que uma afete a outra.²⁵¹
- Causalidade exige evidências (diretas ou indiretas) de que modificar a variável de exposição altera o desfecho.²⁵¹

23.4.2 Quais critérios ajudam a sustentar inferência causal?

- Existência de um mecanismo plausível.²⁵¹
- Controle adequado de confundidores (medidos e não medidos).²⁵¹
- Consistência com literatura prévia e plausibilidade do tamanho do efeito.²⁵¹
- Avaliação de alternativas explicativas (ex.: viés de seleção, mediadores não controlados).²⁵¹

23.4.3 Qual o papel dos caminhos causais (DAGs)?

- Ajudam a identificar quais variáveis precisam ser medidas e ajustadas.²⁵¹
- Evitam ajustes indevidos (ex.: em colisores), que podem introduzir viés.²⁵¹

23.4.4 Como lidar com confundimento residual?

- Reconhecer que modelos multivariados e escores de propensão não eliminam completamente o confundimento.²⁵¹
- Comparar características basais entre grupos para identificar diferenças persistentes.²⁵¹
- Considerar análises de sensibilidade, mas com cautela na interpretação.²⁵¹

23.5 Critérios de Hill para inferência causal

23.5.1 Quais são os nove critérios?

- Temporalidade: A exposição deve preceder o desfecho. Único critério considerado essencial por Hill.²⁵²
- Força da associação: Associações mais fortes são mais prováveis de refletir causalidade.²⁵²
- Consistência: A associação é observada em diferentes estudos, populações e contextos.²⁵²
- Especificidade: Uma exposição leva a um efeito específico (embora nem sempre aplicável).²⁵²
- Gradiente biológico (dose–resposta): Aumentos na exposição acompanham aumentos no risco.²⁵²
- Plausibilidade biológica: Compatibilidade com o conhecimento científico da época.²⁵²
- Coerência: A associação não deve contradizer a história natural ou biologia da doença.²⁵²
- Evidência experimental: Reduções na exposição devem reduzir o risco observado.²⁵²
- Analogia: Comparação com relações causais já conhecidas.²⁵²

23.5.2 Hill propôs um checklist rígido?

- Nenhum critério, isoladamente, prova ou refuta causalidade. Devem ser usados como guias para reflexão científica, não como requisitos obrigatórios.²⁵²

23.6 Críticas contemporâneas aos critérios de Hill

23.6.1 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.²⁵³

23.6.2 A força da associação garante causalidade?

- Não. Associações fortes podem ainda ser não-causais e associações fracas podem ser causais.²⁵³

23.6.3 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).²⁵³
- A consistência ajuda apenas a afastar a hipótese de viés ou erro em um estudo isolado:[contentReference](#).²⁵³

23.6.4 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.²⁵³

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

- Nem sempre. Pode ser distorcido por confundimento. A ausência de gradiente não invalida a causalidade.²⁵³

23.6.6 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.²⁵³

23.6.7 Evidência experimental é decisiva?

- Pode ser útil, mas raramente está disponível em epidemiologia. Mesmo quando disponível, pode ter explicações alternativas.²⁵³

23.6.8 Analogia é útil?

- Tem pouco valor. Analogias podem sempre ser inventadas e, na prática, funcionam mais como fonte de hipóteses do que como prova.²⁵³

23.7 Visão atual sobre os critérios de Hill

23.7.1 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).²⁵⁴

23.7.2 Quais mudanças na interpretação?

- Temporalidade e experimentos: seguem centrais, mas analisados com mais sofisticação.²⁵⁴
- Força da associação: relevante, mas não garante causalidade (pode haver confundimento).²⁵⁴
- Consistência: pensada como transportabilidade entre populações.²⁵⁴
- Especificidade: pouco útil hoje; substituída por falsificação (controles negativos).²⁵⁴
- Dose–resposta: pode ser espúria, cautela é necessária.²⁵⁴
- Coerência e analogia: utilidade limitada.²⁵⁴

23.8 Linguagem causal em estudos observacionais

23.8.1 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”.²⁵¹
- Contextualizar a causalidade em termos práticos, explicando por que identificar a causa é relevante para intervenções.²⁵¹
- Declarar claramente na introdução se existe hipótese causal, justificando quando não houver.²⁵¹
- Descrever caminhos causais (mediadores, confundidores, colisores) em texto claro ou com diagramas.²⁵¹
- Justificar a seleção de covariáveis com base nas relações causais previstas.²⁵¹
- Avaliar o controle de confundimento, reconhecendo limitações e possível confundimento residual.²⁵¹
- Discutir as inferências causais considerando estimativas, vieses e plausibilidade biológica.²⁵¹
- Indicar recomendações específicas para pesquisas futuras ou prática clínica baseadas nas conclusões causais.²⁵¹

23.9 Efeitos diretos e indiretos

23.9.1 Como distinguir efeitos diretos de indiretos?

- Um efeito direto ocorre quando uma variável influencia outra sem mediação.²⁴⁹
- Um efeito indireto acontece quando a influência é mediada por variáveis intermediárias.²⁴⁹

23.10 O papel do tempo e a causalidade dinâmica

23.10.1 O que é causalidade de Granger?

- É um conceito estatístico que analisa como processos passados influenciam o futuro, indo além da simples associação.²⁴⁹
- Permite identificar relações direcionais entre processos ao longo do tempo (ex.: cérebro controlando contrações musculares).²⁴⁹
- A estatística, nesse contexto, busca “olhar dentro da caixa”, aproximando-se de uma visão mecanicista.²⁴⁹

23.10.2 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.²⁴⁹
- Modelos tradicionais (ex.: regressões estáticas ou DAGs sem tempo) podem falhar em capturar a dinâmica.²⁴⁹
- A integração de séries temporais e processos estocásticos é fundamental para compreender mecanismos.²⁴⁹

23.11 Diagrama acíclico direcionado (DAG)

23.11.1 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.?

23.11.2 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 23.1: Padrões causais básicos: independência, cadeia, garfo e colisor.

23.11.3 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.²⁵⁴
- 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.²⁵⁵

23.11.4 Quais são as regras básicas para ajuste?

- Controlar por um confundidor bloqueia caminhos causais espúrios e reduz viés na estimativa do efeito causal.²⁵⁶
- Controlar por um mediador bloqueia parte ou todo o efeito causal total, sendo inadequado quando o objetivo é estimar o efeito total da exposição.²⁵⁷
- Controlar por um colisor cria correlação espúria entre as variáveis que o causam, introduzindo viés mesmo quando não havia associação prévia.²⁵⁸
- Controlar por uma variável descendente de um colisor pode reabrir parcialmente caminhos espúrios, produzindo viés de seleção.²⁵⁹



O pacote *dagitty*²⁵⁵ 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*²⁵⁶ 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*²⁵⁷ 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 24

Análise qualitativa

24.1 Análise qualitativa

24.1.1 O que é análise qualitativa?

- ?

24.2 Representação de texto

24.2.1 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.²⁵⁷
- É comum combinar tokenização com normalização (*lowercase*), remoção de *stopwords*, lematização/*stemming* e regras para números e pontuação.²⁵⁸

24.2.2 Modelagem com N-gramas

24.2.3 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).²⁵⁹
- 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.²⁶⁰



O pacote *tidytext*²⁵⁸ 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*²⁵⁸ 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*²⁵⁸ 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 25

Suposições inferenciais

25.1 Suposições gerais em análises inferenciais

25.1.1 Quais são as suposições ao nível dos dados (condicionais ao modelo)?

- Independência (ou dependência corretamente modelada) das observações.[?]
- Forma da distribuição dos erros ou resíduos (normalidade, assimetria, caudas).[?]
- Homocedasticidade (igualdade de variâncias condicionais).[?]

25.1.2 Quais são as suposições ao nível do modelo?

- Linearidade da relação entre variáveis.[?]
- Multicolinearidade ausente ou controlada.[?]
- Especificação funcional correta do modelo.[?]

25.1.3 Quais são as suposições ao nível do estudo?

- Ausência de confundimento relevante não controlado.[?]
- Estabilidade do processo gerador de dados (invariância temporal, populacional ou contextual).[?]

25.2 Suposições implícitas e explícitas nos testes

25.2.1 Quais suposições implícitas são feitas nos testes estatísticos?

- Amostragem aleatória ou ignorabilidade condicional.[?]
- Medição sem erro relevante.[?]
- Correspondência entre modelo estatístico e processo gerador de dados.[?]
- Ausência de múltiplas comparações não ajustadas.[?]

25.2.2 Quais suposições explícitas são feitas nos testes estatísticos?

- Normalidade dos erros ou da estatística de teste.[?]
- Homocedasticidade.[?]

- Independência das observações.[?]

25.3 Suposições causais que conectam dados observados a efeitos causais

25.3.1 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 contrafactualis 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.[?]

25.3.2 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.[?]

25.4 Diagnóstico e verificação

25.4.1 O que fazer quando suposições gerais falham?

- Transformações.[?]
- Métodos robustos (estimadores e testes).[?]
- Reamostragem.[?]
- Modelos alternativos.[?]

25.4.2 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*²⁵⁷ 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

25.4.3 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.²⁵⁹

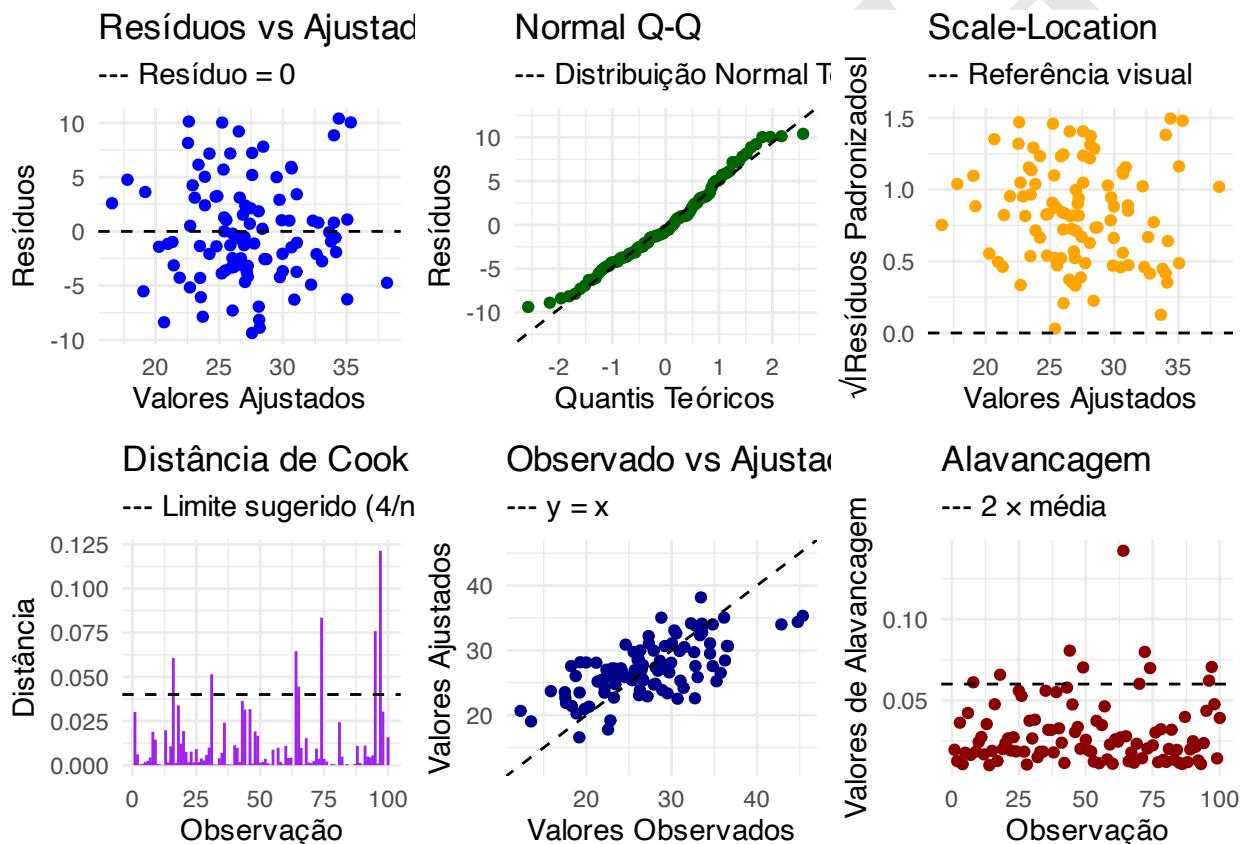


Figura 25.1: Diagnóstico de regressão para avaliar suposições do modelo: linearidade, normalidade dos resíduos, homocedasticidade e alavancagem.

RASCUNHO

Capítulo 26

Seleção de testes

26.1 Multiverso de análises estatísticas

26.1.1 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.²⁶⁰
- 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.²⁶⁰

26.2 Escolha de testes para análise inferencial

26.2.1 Como selecionar os testes para a análise estatística inferencial?

- ²⁶¹
- ²⁶²
- ²⁶³
- ²⁶⁴
- ²⁶⁵
- ²⁶⁶
- ²⁶⁷
- ²⁶⁸

Tabela 26.1: 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 26.2: 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
Experimental / observacional	1 fator + 1 resposta	2	Independentes	Ordinal / não normal	Mann-Whitney U / Kruskal-Wallis

Continuado na próxima página

Tabela 26.2: 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
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 26.3: 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
Transversal	2 variáveis	—	—	Categórica \times categórica (2×2)	Qui-quadrado / Fisher + Phi (ϕ)

Continuado na próxima página

Tabela 26.3: 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 (freq. pequenas)	Exato de Fisher
Transversal	2 variáveis	–	–	Categórica × categórica ($R \times C$)	Qui-quadrado + V de Cramér

Tabela 26.4: 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	Regressão linear
Observacional / experimental	≥ 1 preditor + 1 resposta	–	–	Binária	Regressão logística

Tabela 26.5: 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 26.6: 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 26.7: 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 Intraclasse (ICC)

Tabela 26.8: 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 26.9: 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 26.10: 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 múltiplas	Teste de Mardia (assimetria e curtose)
Transversal	≥ 2 variáveis	–	–	Contínuas múltiplas	Henze–Zirkler
Transversal	≥ 2 variáveis	–	–	Contínuas múltiplas	Royston (Shapiro–Wilk multivariado)
Transversal	1 fator + 1 resposta	≥ 2	Independentes	Contínua categórica	Levene

Continuado na próxima página

Tabela 26.10: 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 26.11: 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 27

Análise inferencial

27.1 Raciocínio inferencial

27.1.1 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.¹¹³
- 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.²⁶⁹

27.1.2 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.²⁷⁰
- Inferência induutiva: Com base nos dados observados, avalia-se qual hipótese é mais defensável (isto é, mais provável).²⁷⁰

27.1.3 Quais são as questões fundamentais da análise inferencial?

- A direção do efeito²⁷¹
- A magnitude do efeito²⁷¹
- A importância do efeito²⁷¹

27.2 Hipóteses científicas

27.2.1 O que é hipótese científica?

- Hipótese científica é uma ideia que pode ser testada.²⁶⁹
- Definir claramente os problemas e os objetivos da pesquisa são o ponto de partida de todos os estudos científicos.¹³⁸
- 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.²⁷²

27.2.2 Quais são as fontes de ideias para gerar hipóteses científicas?

- Revisão das práticas atuais.²⁷³
- Desafio a ideias aceitas.²⁷³
- Conflito entre ideias divergentes.²⁷³
- Variações regionais, temporais e populacionais.²⁷³
- Experiências dos próprios pesquisadores.²⁷³
- Imaginação sem fronteiras ou limites convencionais.²⁷³

27.3 Hipóteses estatísticas

27.3.1 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.¹⁸⁵

27.3.2 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.¹⁸⁵

27.3.3 Qual hipótese está sendo testada?

- A hipótese nula (H_0) é a hipótese sob teste em análises inferenciais.¹¹⁴
- Pode-se concluir sobre rejeitar ou não rejeitar a hipótese nula (H_0).¹¹⁴
- Não se conclui sobre a hipótese alternativa (H_1).¹⁸⁵
- 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).²⁶⁹

27.4 Testes de hipóteses

27.4.1 Quais são os tipos de teste de hipóteses?

- Teste (clássico) de significância da hipótese nula: verifica evidência contra H_0 usando P-valor.²⁷⁴
- 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.²⁷⁴
- 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.²⁷⁴
- Teste de superioridade: avalia se um tratamento/intervenção supera o controle por uma margem > 0 ou $> \Delta$.²⁷⁴
- 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).?

27.4.2 O que reportar após um teste de hipótese?

- P-valores, como estimativa da significância estatística.²⁷⁵
- Tamanho do efeito, como estimativa de significância substantiva.²⁷⁵

27.5 Intervalos de confiança e raciocínio de longo prazo

27.5.1 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.²⁷⁶
- 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.²⁷⁷
- O parâmetro populacional é considerado fixo (embora desconhecido), enquanto o intervalo de confiança é uma quantidade aleatória, pois depende da amostra observada.²⁷⁸

27.5.2 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.²⁷⁶
- 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.²⁷⁷
- 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.

27.5.3 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.²⁷⁸
- 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.²⁷⁸
- A incerteza expressa pelo intervalo de confiança refere-se à variabilidade do processo inferencial, e não a uma distribuição de probabilidade do parâmetro.²⁷⁸

27.5.4 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.²⁷⁹
- 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 .²⁷⁹
- 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 .²⁷⁹
- Apesar dessa equivalência formal, intervalos de confiança e testes de hipótese respondem a perguntas distintas: testes enfatizam decisões dicotômicas, enquanto intervalos de confiança enfatizam estimativa e incerteza.²⁷⁹

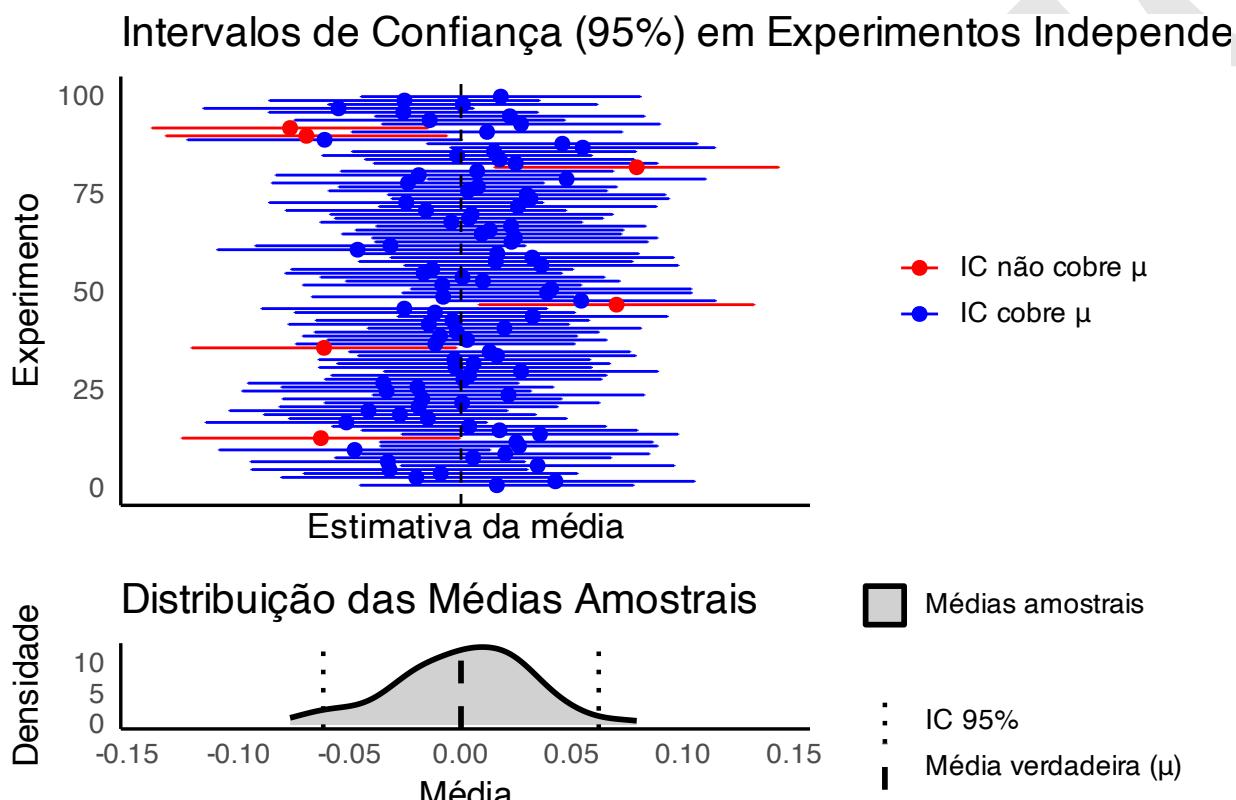


Figura 27.1: Simulação ilustrativa de intervalos de confiança (IC) em 100 experimentos independentes, cada um com 1.000 observações amostradas de uma população normal padrão (média = 0, desvio-padrão = 1). Os ICs são construídos no nível de 95%. O gráfico superior mostra os ICs individuais para cada 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.

27.5.5 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.²⁷⁹
- Essa abordagem favorece interpretações substantivas e científicas dos resultados, em oposição a decisões puramente dicotômicas baseadas em pontos de corte arbitrários.²⁷⁹
- 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.²⁷⁸

27.6 Comparações múltiplas

27.6.1 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.

27.6.2 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

27.6.3 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*¹³⁶ 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>

27.6.4 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

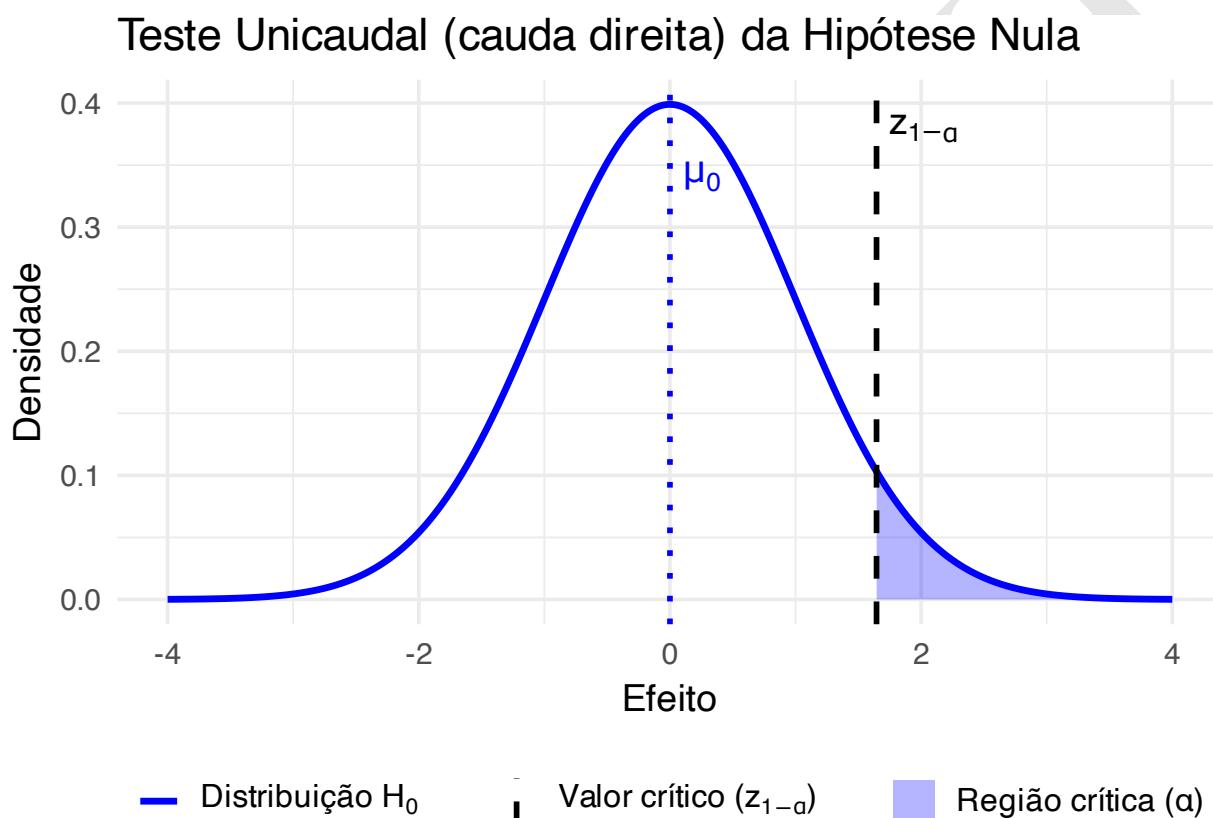


Figura 27.2: Representação gráfica de um teste de hipótese unicaudal à direita, aplicado quando se busca evidência de efeitos positivos (valores significativamente maiores que o esperado sob H_0).

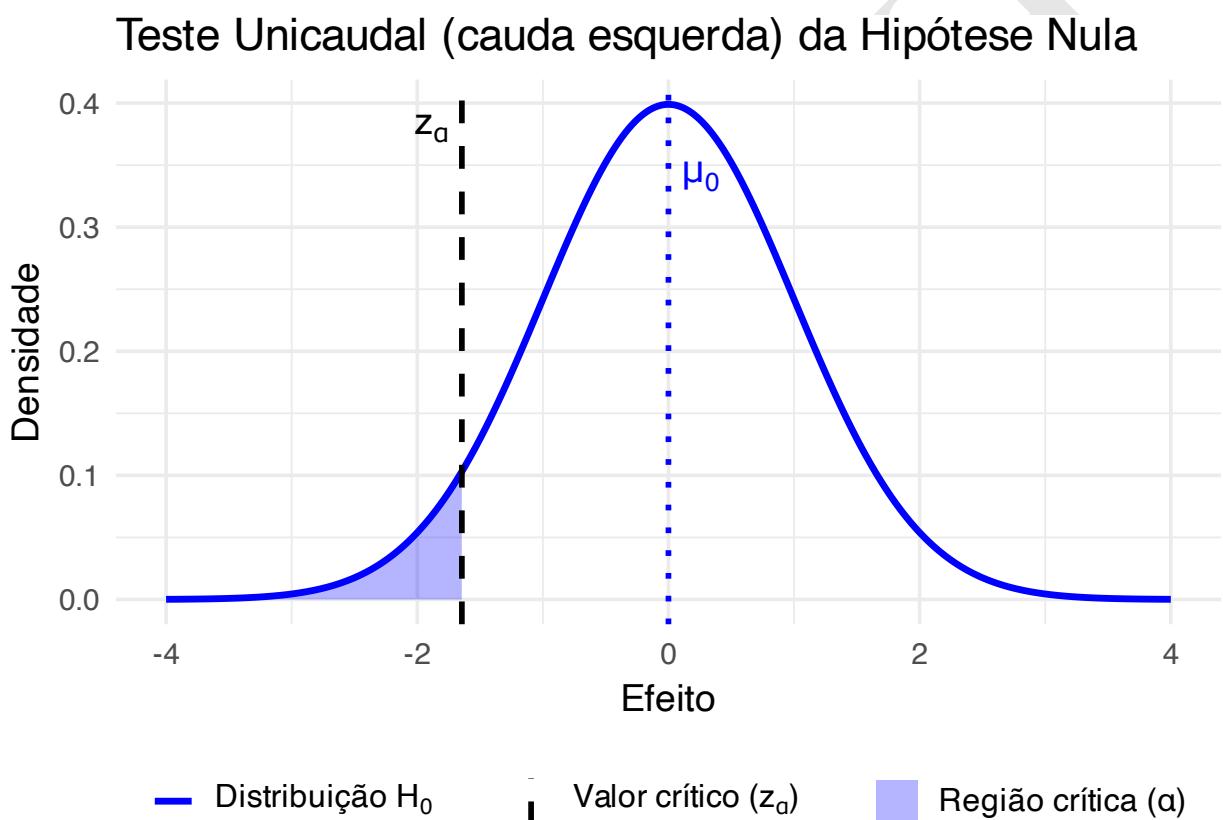


Figura 27.3: Representação gráfica de um teste de hipótese unicaudal à esquerda, aplicado quando se busca evidência de efeitos negativos (valores significativamente menores que o esperado sob H_0).

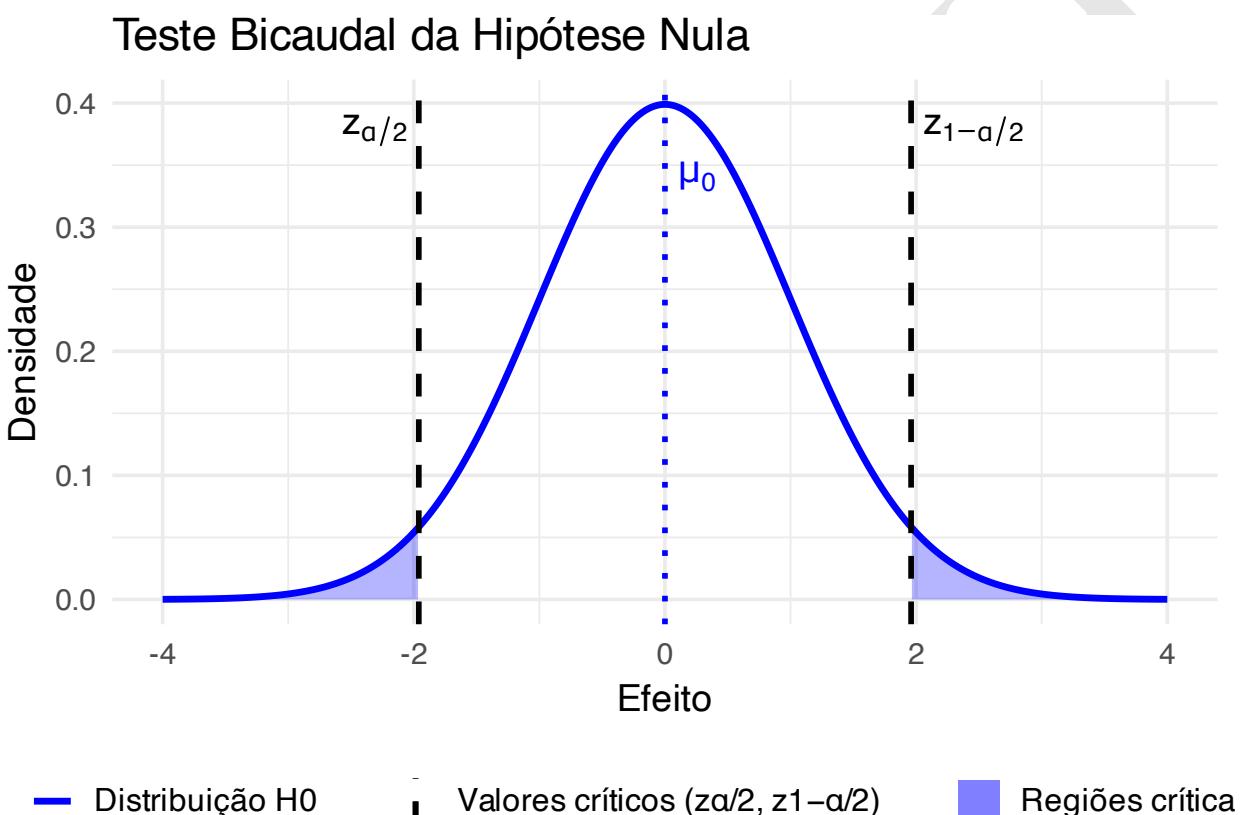


Figura 27.4: Representação gráfica de um teste de hipótese bicaudal, aplicado quando se busca evidência de efeitos positivos ou negativos (valores significativamente diferentes do esperado sob H_0).

27.7 Inferência visual

27.7.1 O que é inferência visual?

- Inferência visual consiste na interpretação de dados apresentados em gráficos.²⁷⁹
- 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).²⁷⁹

27.7.2 Por que usar intervalos de confiança para inferência visual?

- Intervalos de confiança fornecem estimativas pontuais e intervalares na mesma unidade de medida da variável.²⁷⁹
- 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.²⁷⁹

27.7.3 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?²⁷⁹
- Faça uma interpretação substantiva dos valores de tendência central e dos intervalos de confiança.²⁷⁹
- 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.²⁷⁹
- Qualquer valor fora do intervalo de confiança, quando considerado como hipótese nula (H_0), equivale a $P < \alpha$ pré-estabelecido (bicaudal).²⁷⁹
- Qualquer valor dentro do intervalo, quando considerado como hipótese nula (H_0), equivale a $P > \alpha$ pré-estabelecido (bicaudal).²⁷⁹

27.8 Interpretação de análise inferencial

27.8.1 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.²⁷⁷
- P-valor como evidência estatística sobre (H_0).²⁷⁷
- Estimação de intervalos de confiança de um nível de significância (α) pré-especificado bicaudal ($IC_{1-\alpha/2}$) ou unicaudal ($IC_{1-\alpha}$).²⁷⁷
- Análise Bayesiana.²⁷⁷

27.8.2 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.²⁸⁰

Tabela 27.1: 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).²⁸⁰

27.8.3 Qual a importância de resultados “negativos”?

- Conhecer resultados negativos contribui com uma visão mais ampla do campo de estudo junto aos resultados positivos.²⁸¹
- Resultados negativos permitem um melhor planejamento das pesquisas futuras e pode aumentar suas chances de sucesso.²⁸¹

27.8.4 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.²⁸²
- 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.²⁸²
- 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.²⁸²
- 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.²⁷²

27.9 Erros de inferência I, II, S e M

27.9.1 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).²⁶⁹

27.9.2 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.²⁶⁹
- Erro Tipo II significa a não rejeição de uma hipótese nula (H_0) quando esta é falsa.²⁶⁹

27.9.3 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.^{283,284}

Erros tipo I (α) e II (β)

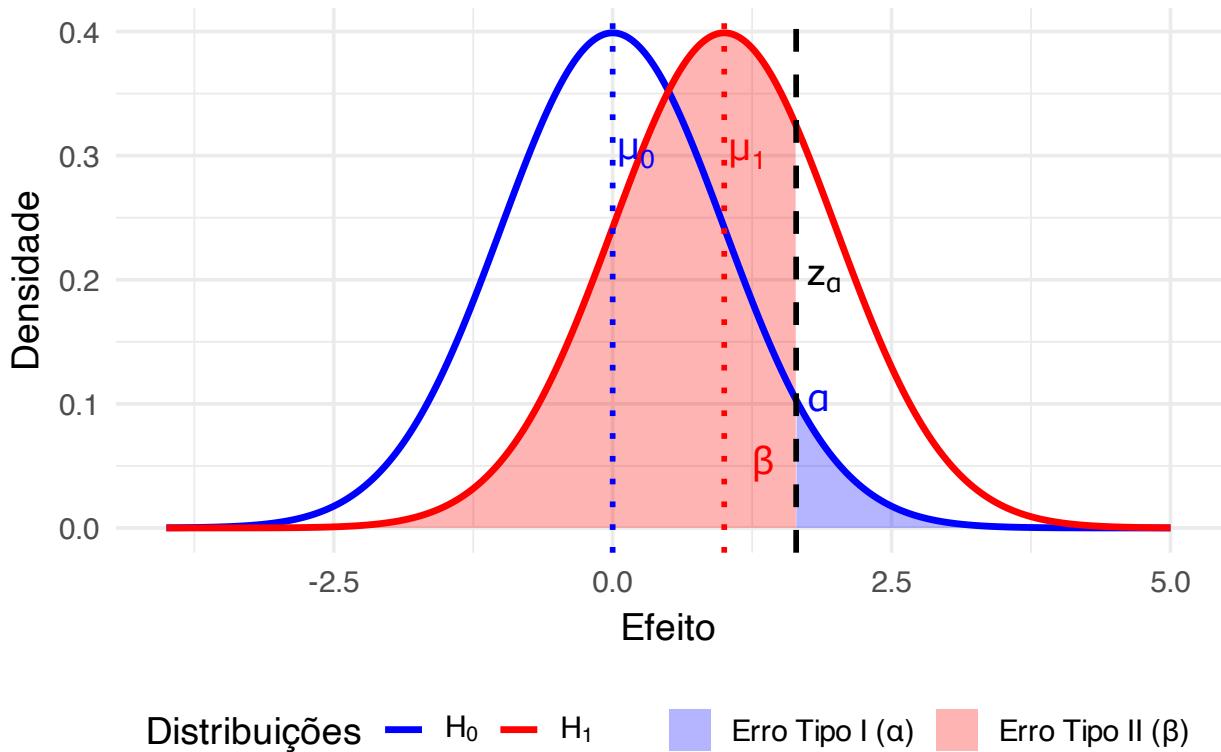


Figura 27.5: Representação gráfica dos erros tipo I e tipo II em um teste de hipótese (bicaudal).

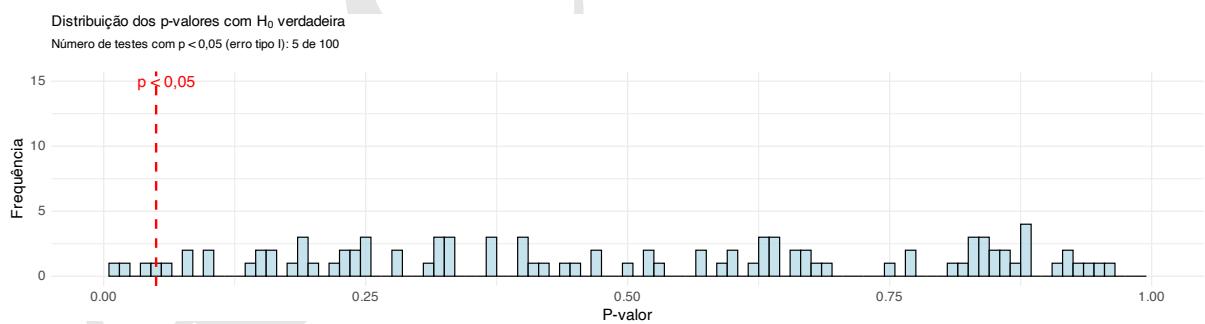


Figura 27.6: Erro tipo I: Distribuição dos p-valores em 100 testes de hipótese de amostras aleatórias de tamanho 30. A linha vermelha pontilhada indica o nível de significância estatística de 0,05.

Tabela 27.2: 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	

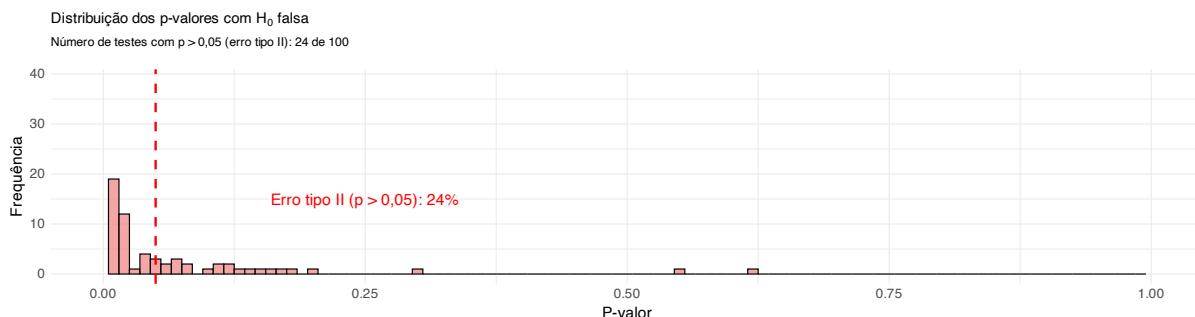


Figura 27.7: Erro tipo II: Distribuição dos p-valores em 100 testes de hipótese de amostras aleatórias de tamanho 10. A linha vermelha pontilhada indica o nível de significância estatística de 0,05.

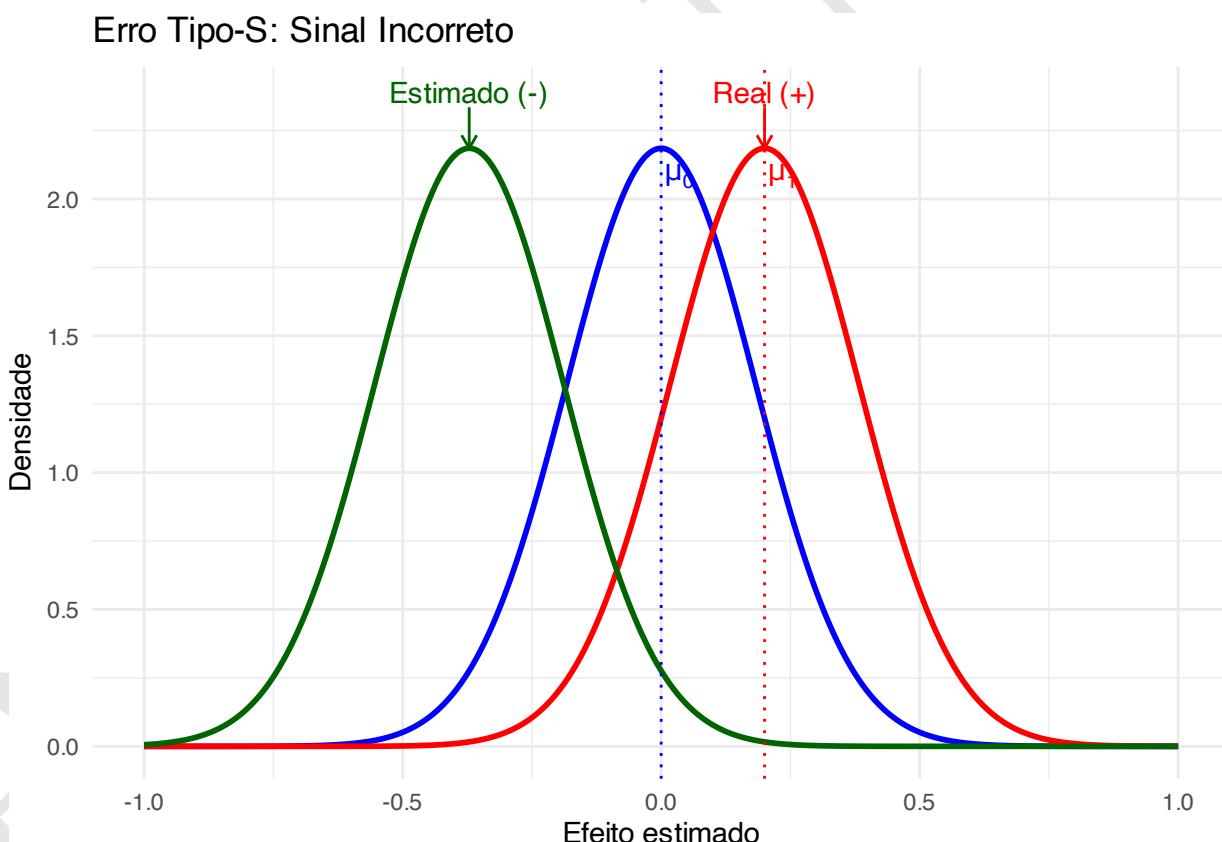


Figura 27.8: Representação gráfica do erro tipo S (sinal) em um teste de hipótese (bicaudal).

Tabela 27.3: 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.^{283,284}

Erro Tipo-M: Superestimação da Magnitude

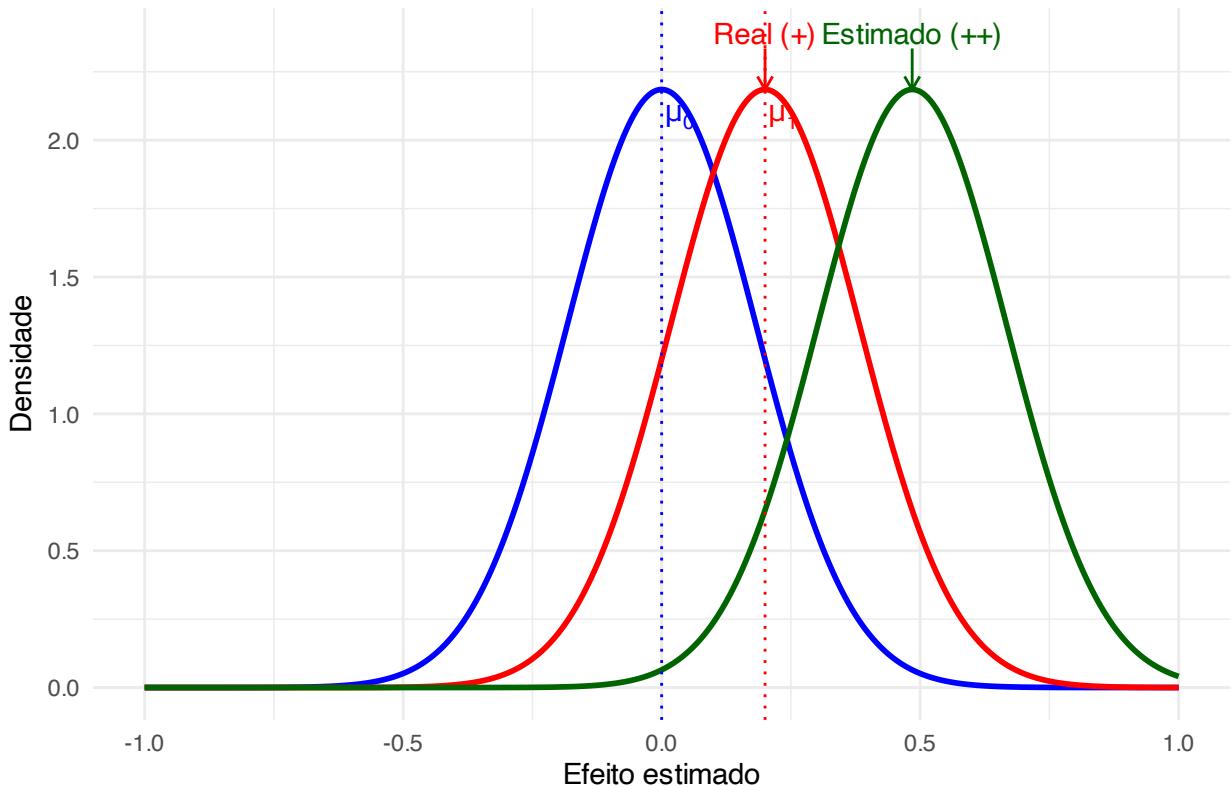


Figura 27.9: Representação gráfica do erro tipo M (magnitude) em um teste de hipótese (bicaudal).

RASCUNHO

Capítulo 28

Tamanho do efeito

28.1 Tamanho do efeito

28.1.1 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.²⁸⁵

28.2 Tipos de tamanho do efeito

- Diferenças padronizadas entre grupos:^{275,285}
 - Cohen's d
 - Glass's Δ
 - Razão de chances (RC ou OR)
 - Risco relativo ou razão de risco (RR)

R O pacote *epitools*²⁸⁶ 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*²⁸⁶ 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:^{275,285}
 - Coeficiente de correlação de Pearson (r), ponto-bisserial (r_s), Spearman (ρ), Kendall (τ), Cramér (V) e ϕ .
 - Coeficiente de determinação (R^2)

28.2.1 Como interpretar um tamanho do efeito?

- Tamanhos de efeito podem ser comparadores entre diferentes estudos.²⁷⁵

R

O pacote *effectsize*²⁸⁷ 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*²⁸⁷ 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*²⁸⁸ 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>

28.2.2 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.[?]

28.2.3 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.[?]

28.2.4 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.[?]

28.2.5 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.[?]

28.2.6 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.[?]

28.2.7 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.[?]

28.2.8 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.[?]

28.2.9 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.[?]

28.2.10 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.[?]

28.2.11 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.[?]

28.3 Conversão entre tamanhos do efeito

28.3.1 Como converter um tamanho de efeito em outro?

- ²⁸⁵

R

O pacote *effectsize*²⁸⁷ fornece diversas funções para conversão de diferentes estimativas de tamanhos de efeito.

28.4 Efeitos bruto e padronizado

28.4.1 O que é efeito bruto?

- ²⁸⁹
- ²⁹⁰

28.4.2 O que é efeito padronizado?

- ²⁸⁹
- ²⁹⁰

Capítulo 29

P-valor

29.1 P-valor

29.1.1 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.²⁹¹
- P-valor é uma variável aleatória que possui distribuição uniforme quando a hipótese nula (H_0) é verdadeira.²⁹²

29.1.2 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.²⁹¹
- P-valor não mede o tamanho do efeito ou a relevância da sua observação.²⁹¹
- 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.²⁹¹
- Evidência estatística de significância não provê informação sobre a magnitude do efeito observado e não necessariamente implica que o efeito é robusto.^{199,292}

29.2 Significância estatística

29.2.1 O que é significância estatística?

- A expressão “significância estatística”²⁹³ 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.²⁹⁴

29.3 Interpretação do P-valor

29.3.1 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)).²⁷⁷
- P-valor resulta da coleta e análise de dados, e assim quantifica a plausibilidade dos dados observados sob a hipótese nula (H_0).²⁹⁵
- 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)).²⁹¹
- 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.²⁹¹

29.3.2 Existe uma crítica lógica à significância estatística?

- Sim. Parte da crítica contemporânea argumenta que a significância estatística possui uma base lógica frágil, especialmente quando interpretada como “evidência contra” a hipótese nula.²⁹⁶
- O raciocínio subjacente ao P-valor pode ser entendido como uma forma probabilística de “prova por contradição”, cuja validade não se sustenta sob incerteza.²⁹⁶

29.3.3 O que é “prova probabilística por contradição”?

- Na lógica clássica, se um evento B é impossível sob A , então observar B implica que A é falso.²⁹⁶
- Entretanto, quando substituímos “impossível” por “improvável”, a conclusão não é logicamente válida.²⁹⁶
- O fato de B ser improvável sob H_0 não implica que H_0 seja improvável após observar B .²⁹⁶

29.3.4 Qual é o equívoco central?

- Confundir $P(\text{dados} | H_0)$ com $P(H_0 | \text{dados})$.²⁹⁶
- O P-valor mede a improbabilidade dos dados assumindo H_0 verdadeira.²⁹⁶
- Ele não mede a probabilidade de H_0 ser verdadeira.²⁹⁶

29.3.5 O que isso implica para a interpretação do P-valor?

- A significância estatística não equivale a “improbabilidade da hipótese nula”.²⁹⁶
- A expressão “evidência contra H_0 ” conceitualmente mais cautelosa do que “ H_0 é improvável” ou “dados são improváveis sob H_0 ”, mas ainda repousa em uma estrutura lógica debatida.²⁹⁶



O pacote *Superpower*²⁹⁷ 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*²⁹⁷ 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

29.3.6 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.²⁷⁷
- O P<0,05 foi sugerido por Ronald A. Fisher como um limiar prático para indicar que um resultado era “estatisticamente significativo”.²⁷⁷
- 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.²⁷⁷

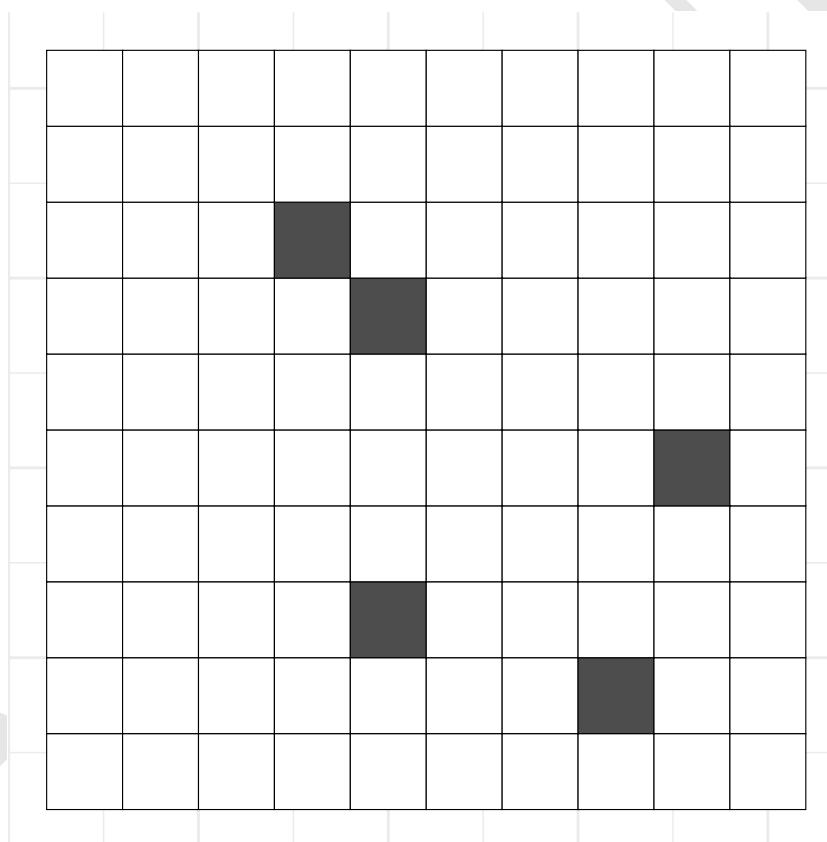


Figura 29.1: Visualização espacial de $p < 0,05$ (5 quadrados aleatórios em 100).

29.3.7 Quais são os complementos ou alternativas ao P-valor?

- Intervalos de confiança, credibilidade ou predição.²⁹¹
- Razão de verossimilhança.²⁹¹
- Métodos Bayesianos, fator Bayes.²⁹¹

29.4 P-valor de 2^a geração

29.4.1 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 intervalo de estimativa (p.ex., IC95%) recai dentro da zona de indiferença científica/clinicamente irrelevante.²⁹⁸
- 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.).²⁹⁸

29.4.2 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 ($\delta = (H_0^+ - H_0^-)/2$).²⁹⁸
- A escolha deve ser a priori e justificada por critérios científicos (p.ex., MCID, precisão de medida).²⁹⁸

29.4.3 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 (29.1), 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 $[0, 1]$.²⁹⁸

$$p_\delta = \frac{|I \cap H_0|}{|I|} \times \max\left\{\frac{|I|}{2|H_0|}, 1\right\} \quad (29.1)$$

- Quando $|I| < 2|H_0|$, p_δ é apenas a fração de sobreposição $|I \cap H_0|/|I|$.²⁹⁸
- 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.²⁹⁸

29.4.4 Como interpretar o SGPV?

- $p_\delta = 0$: dados apoiam apenas hipóteses alternativas relevantes (IC totalmente fora da equivalência).²⁹⁸
- $p_\delta = 1$: dados apoiam apenas hipóteses nulas (equivalentes) (IC totalmente dentro da equivalência).²⁹⁸
- $0 < p_\delta < 1$: inconclusivo; o valor expressa o grau de inconclusão. Em particular, $p_\delta = \frac{1}{2}$ indica inconclusão estrita.²⁹⁸
- O SGPV é descriptivo (não é probabilidade posterior de H_0).²⁹⁸

29.4.5 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 $p_\delta = 1$.²⁹⁹
- Com ICs simétricos, há pontos de ancoragem em que as estatísticas coincidem: quando $p_{TOST} = 0,5$, então SGPV = 0,5; quando o IC toca o limite mas fica inteiramente dentro (fronteira), $p_{TOST} = 0,025$ e SGPV = 1; quando o IC fica inteiramente fora tocando o limite, $p_{TOST} = 0,975$ e SGPV = 0.²⁹⁹
- 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.²⁹⁹

Tabela 29.1: Comparação entre p-valor (bicaudal, inferido do IC95%) 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)

29.4.6 Propriedades frequenciais 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.²⁹⁸
- O SGPV mitiga naturalmente inflação de erro Tipo I em muitas comparações e prioriza relevância científica (não requer ajustes ad hoc).²⁹⁸

29.5 Distribuição de confiança

29.5.1 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.²⁹⁹

29.6 Boas práticas

- Defina H_0 intervalar e δ a priori com justificativa científica.^{298,299}
- 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.^{298,299}

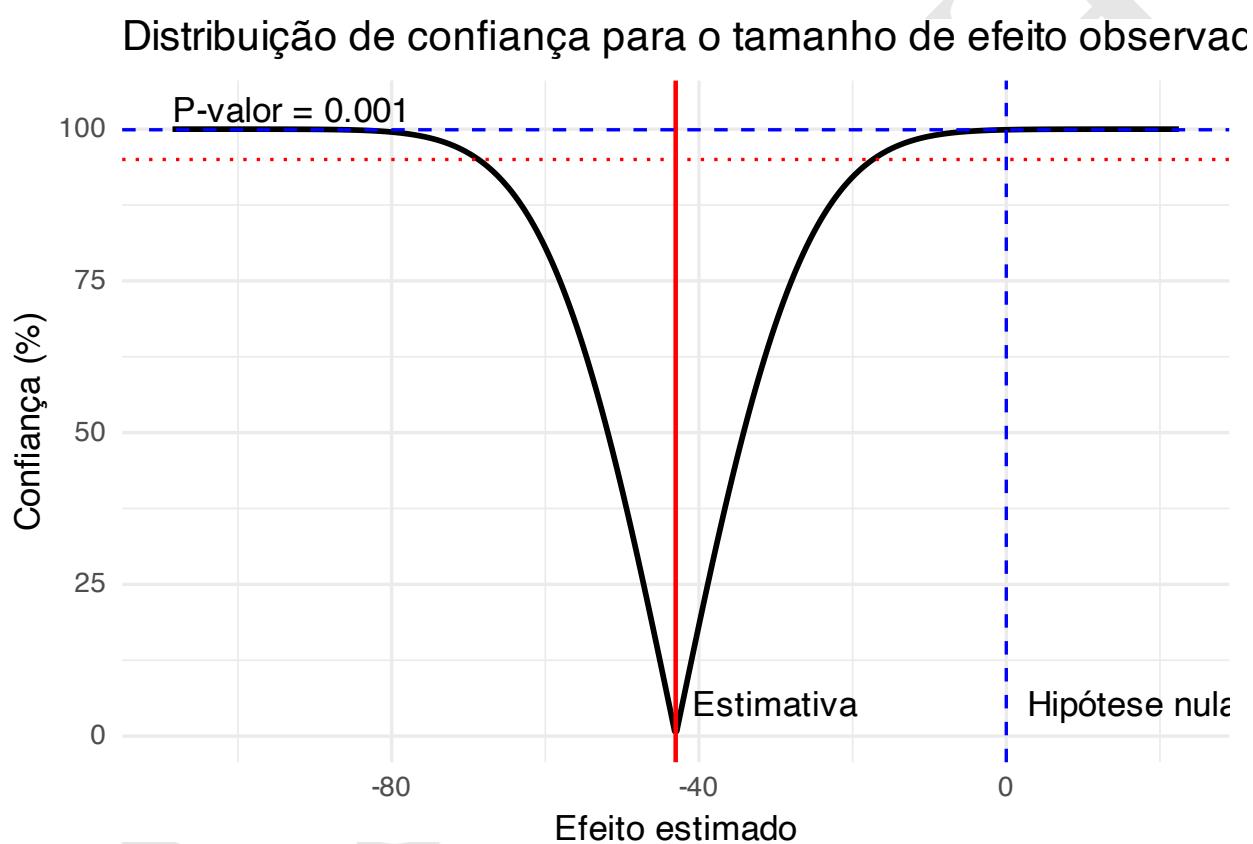


Figura 29.2: Distribuição de confiança para o tamanho do efeito estimado.

Capítulo 30

Testes estatísticos

30.1 Variáveis categóricas

- 30.1.1 Testes de Qui-quadrado (χ^2)
- 30.1.2 Teste exato de Fisher
- 30.1.3 Teste de McNemar
- 30.1.4 Teste Q de Cochran
- 30.1.5 Teste de Cochran–Armitage
- 30.1.6 Odds ratio (OR) e risco relativo (RR)

30.2 Variáveis contínuas

- 30.2.1 Teste t de Student
- 30.2.2 Teste t de Welch
- 30.2.3 Teste de Mann-Whitney
- 30.2.4 Teste de Wilcoxon
- 30.2.5 Análise de variância
- 30.2.6 Análise de variância (Welch)
- 30.2.7 Teste de Kruskal-Wallis

Tabela 30.1: 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%)	

¹Teste qui-quadrado de independência

Tamanho do efeito (Cramér's V): 0.269

Tabela 30.2: 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%)	

¹Teste qui-quadrado de independência

Tamanho do efeito (Cramér's V): 0.269

Tabela 30.3: 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%)	

¹Teste exato de Fisher

Tamanho do efeito (Cramér's V): 0.269

Tabela 30.4: 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 30.5: 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 30.6: 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 30.7: 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%)	

¹Teste qui-quadrado de independência

OR = 2.1 [1.17, 3.79]; RR = 1.61 [1.1, 2.35].

Tabela 30.8: Teste t de Student

Características	A N = 45¹	B N = 55¹	P-valor²
Desfecho	51.30 (8.88)	55.38 (10.05)	0.034

¹Média (Desvio Padrão)²Teste t com correção de Welch

Tamanho do efeito (d de Cohen): -0.432

Tabela 30.9: Teste t de Welch

Características	A N = 45¹	B N = 55¹	P-valor²
Desfecho	51.30 (8.88)	55.38 (10.05)	0.034

¹Média (Desvio Padrão)²Teste t com correção de Welch

Tamanho do efeito (d de Cohen): -0.432

Tabela 30.10: 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

¹Mediana [Q1, Q3]²Teste de soma de postos de Wilcoxon

Tamanho do efeito (r): 0.208

Tabela 30.11: 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

¹Mediana [Q1, Q3]²Teste de soma de postos de Wilcoxon

Tamanho do efeito (r): 0.616

Tabela 30.12: 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

¹Média (Desvio Padrão)²One-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 30.13: 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

¹Média (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 30.14: Teste de Kruskal-Wallis

Características	A N = 54¹	B N = 67¹	C N = 59¹	P-valor²
Desfecho	40.98 [31.12, 51.85]	40.48 [31.75, 62.18]	64.96 [45.37, 77.91]	<0.001

¹Mediana [Q1, Q3]²Teste 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 31

Descrição

31.1 Análise de descrição

31.1.1 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.¹¹³
- As análises descritivas geralmente compreendem a apresentação quantitativa (numérica) em tabelas e/ou gráficos.¹¹³



O pacote *explore*²⁰² 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*²⁰³ 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*²⁰⁴ 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*²⁰⁵ 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*³⁰⁰ 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>

31.2 Estimação

31.2.1 O que é estimativa?

- Estimativa é o valor de uma variável de interesse calculado a partir de uma amostra.[?]

31.2.2 O que é estimativa pontual?

- Estimativa pontual é o valor único de uma variável de interesse calculado a partir de uma amostra.[?]

31.2.3 O que é estimativa de intervalar?

- Estimativa intervalar é um intervalo de valores de uma variável de interesse calculado a partir de uma amostra.[?]

31.2.4 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 32

Comparação

32.1 Análise inferencial de comparação

32.1.1 O que é análise de comparação de dados?

- ?

R

O pacote *cocor*³⁰¹ 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.³⁰¹

^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>

32.2 F-teste

32.2.1 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 (32.1), onde QM são “quadrados médios”, com gl_1 e gl_2 definidos pelo desenho (ex.: fatores e resíduos).[?]

$$F = \frac{QM_{entre}}{QM_{dentro}} \quad (32.1)$$

32.2.2 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).[?]

32.2.3 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).?

32.2.4 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?

32.2.5 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 33

Associação

33.1 Análise inferencial de associação

33.1.1 O que é análise de associação?

- ?

33.2 Associação bivariada

33.2.1 O que são análises de associação bivariada?

- ?

33.2.2 Quais testes podem ser usados para análises de associação bivariada?

- Teste Qui-quadrado (χ^2).^{302,303}
 - O teste qui-quadrado (χ^2) avalia uma hipótese global se a relação entre duas variáveis e/ou fatores é independente ou associada.³⁰³
 - 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.³⁰³
 - Tipo: não paramétrico.^{302,303}
 - Suposições:^{302,303}
 - * 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:³⁰³

- * Nula (H_0): independente (sem associação)
- * Alternativa (H_1): não independente (associação)
- Tamanho do efeito:³⁰³
 - * 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*²¹⁷ 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.^{302,303}
 - O teste exato de Fisher avalia a hipótese nula de independência aplicando a distribuição hipergeométrica dos números nas células da tabela.³⁰³
 - Hipóteses:^{302,303}
 - * Nula (H_0): independente (sem associação)
 - * Alternativa (H_1): não independente (associação)
 - Tamanho do efeito:^{302,303}
 - * 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*²¹⁷ fornece a função *tbl_cross*^a para criar uma tabela NxM.

^ahttps://www.rdocumentation.org/packages/gtsummary/versions/1.6.3/topics/tbl_cross

33.2.3 Quais testes podem ser usados para análises de associação bivariada entre variáveis contínuas?

- Coeficiente de correlação de Pearson r .?
- Coeficiente de correlação de Spearman ρ .?
- Coeficiente de correlação de Kendall τ .?

33.2.4 Quais testes podem ser usados para análises de associação bivariada entre variáveis dicotômicas?

- Coeficiente de correlação tetracórica r_{tet} .^{304,305}

R

O pacote *psych*³⁰⁶ 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>

33.2.5 Quais testes podem ser usados para análises de associação bivariada entre variáveis ordinais sob modelo latente contínuo?

- Coeficiente de correlação policórica r_{pol} .³⁰⁵

R

O pacote *psych*³⁰⁶ fornece a função *tetrachoric*^a para calcular o coeficiente de correlação policórica (r_{pol}).

^a<https://www.rdocumentation.org/packages/psych/versions/2.3.6/topics/tetrachoric>

33.2.6 Quais testes podem ser usados para análises de associação bivariada entre variáveis dicotômica e contínua?

- Variável dicotômica real vs. contínua: coeficiente de correlação ponto-bisserial r_{pb} .
- Variável dicotômica artificial (corte em variável latente contínua): Coeficiente de correlação bisserial r_s .³⁰⁵

R

O pacote *psych*³⁰⁶ 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>

33.3 Associação multivariada

33.3.1 O que são análises de associação multivariada?

- ?

33.3.2 Quais testes podem ser usados para análises de associação multivariada?

- ?

RASCUNHO

Capítulo 34

Correlação

34.1 Análise inferencial de correlação

34.1.1 O que é covariância?

- ?

34.1.2 O que é correlação?

- ?

34.1.3 Qual é a interpretação das medidas de correlação?

- Os valores de correlação estão no intervalo $[-1; 1]$.^{117,307,308}
- 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.^{307,308}
- 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.^{307,308}
- Valores de correlação próximos de 0 representam a inexistência de relação entre as variáveis.^{307,308}

34.1.4 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.³⁰⁷
- Tamanhos de efeito grande (ou qualquer outro) não representam necessariamente uma relação de concordância ou confiabilidade entre as variáveis.³⁰⁷
- 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%.¹³⁰ Neste caso, a correlação desatenuada $r_{x'y'}$ pode ser calculada por (34.1), 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:¹³⁰

$$r_{x'y'} = \frac{r_{xy}}{r_{xx'} r_{yy'}} \quad (34.1)$$

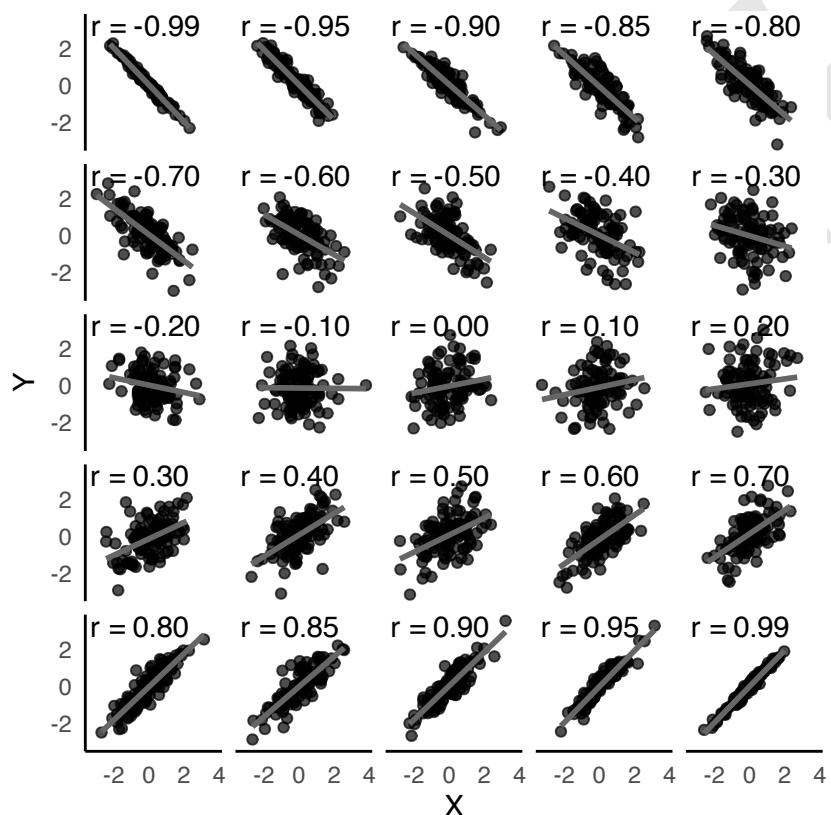


Figura 34.1: Exemplo de diferentes forças e direção de correlação entre duas variáveis X e Y.

Tabela 34.1: 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 34.2: 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*³⁰⁹ 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*³⁰⁹ 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.³¹⁰
- 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.³¹⁰

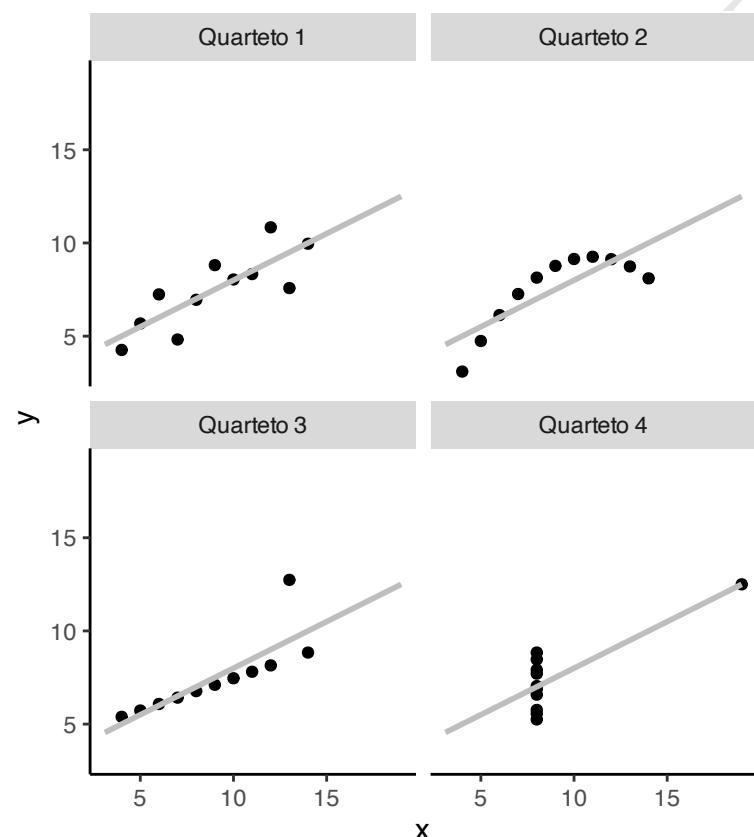


Figura 34.2: Gráfico de dispersão do Quarteto de Anscombe para representação gráfica de conjuntos de dados bivariados com parâmetros quase idênticos e relações muito distintas.



O pacote *anscombiner*³¹¹ 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>

34.2 Coeficientes de correlação

34.2.1 Quais coeficientes podem ser usados em análises de correlação?

- Coeficiente de correlação de Pearson (r) (34.2).^{307,308}

$$r = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2] [n \sum y_i^2 - (\sum y_i)^2]}} \quad (34.2)$$

- O coeficiente de correlação de Pearson (r) avalia a força e direção da relação linear entre duas variáveis quantitativas.^{307,308}
- Tipo: paramétrico.^{307,308}
- Hipóteses:³⁰⁸
- Nula (H_0): $r = 0$
- Alternativa (H_1): $r \neq 0$
- Tamanho do efeito:^{307,308}
- Coeficiente de correlação de Pearson (r)



O pacote *stats*¹³⁶ 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*³¹² do projeto *easystats*³¹³ 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) (34.3).³⁰⁷

$$r_s = \frac{M_1 - M_0}{s_y} \sqrt{\frac{n_1 n_0}{n^2}} \quad (34.3)$$

- O coeficiente de correlação ponto-bisserial (r_s) avalia a força e direção da relação linear entre uma variável quantitativa e outra dicotômica.³⁰⁷
- Tipo: paramétrico.³⁰⁷
- Hipóteses:³⁰⁷
- Nula (H_0): $r_s = 0$
- Alternativa (H_1): $r_s \neq 0$

- Tamanho do efeito.³⁰⁷
- Coeficiente de correlação ponto-bisserial (r_s)

R

O pacote *stats*¹³⁶ 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*³¹² do projeto *easystats*³¹³ 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) (34.4).^{307,308}

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \quad (34.4)$$

- O coeficiente de correlação de Spearman (ρ) avalia a força e direção da relação monotônica entre duas variáveis quantitativas.^{307,308}
- 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.^{307,308}
- Tipo: não-paramétrico.^{307,308}
- Hipóteses:^{307,308}
- Nula (H_0): $\rho = 0$
- Alternativa (H_1): $\rho \neq 0$
- Tamanho do efeito.^{307,308}
- Coeficiente de correlação de Spearman (ρ)

R

O pacote *stats*¹³⁶ 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*³¹² do projeto *easystats*³¹³ 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) (34.5).^{307,308}

$$\tau = \frac{(n_c - n_d)}{\frac{1}{2}n(n - 1)} \quad (34.5)$$

- O coeficiente Kendall τ avalia a força e direção da relação monotônica entre duas variáveis quantitativas ou qualitativas.^{307,308}
- O coeficiente Kendall τ é definido como a proporção de todos os pares concordantes menos a proporção de todos os pares discordantes.^{307,308}
- Tipo: não-paramétrico.^{307,308}
- Hipóteses:^{307,308}
- Nula (H_0): $\tau = 0$
- Alternativa (H_1): $\tau \neq 0$
- Tamanho do efeito:^{307,308}
- Kendall τ



O pacote *stats*¹³⁶ 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*³¹² do projeto *easystats*³¹³ 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) (34.6).[?]

$$V = \sqrt{\frac{\chi^2/n}{\min(k-1, r-1)}} \quad (34.6)$$

- 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) (34.7).[?]

$$\phi = \sqrt{\frac{\chi^2}{n}} \quad (34.7)$$

- 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*³¹² do projeto *easystats*³¹³ 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*²³¹ 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*²³¹ fornece a função *corrplot*^a para visualização da matriz de correlação.

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

34.3 Colinearidade

34.3.1 O que é colinearidade?

- Colinearidade representa a correlação entre duas variáveis.³¹⁴
- Colinearidade exata indica uma relação linear perfeita entre duas variáveis.³¹⁴

34.3.2 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.³¹⁴
- Valores de correlação próximos de 1 ou -1 indicam colinearidade entre as variáveis.³¹⁴

R

O pacote *GGally*³¹⁵ 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.³¹⁵

^ahttps://www.rdocumentation.org/packages/GGally/versions/2.2.1/topics/ggally_cor

34.4 Correlação entre conjuntos de variáveis

34.4.1 O que é correlação entre conjuntos de variáveis?

- A Correlação Canônica (CCA) analisa a relação entre dois conjuntos de variáveis simultaneamente.[?]
- Busca combinações lineares que maximizam a correlação entre os dois blocos.[?]

34.4.2 Quando usar CCA?

- Quando existem dois blocos distintos de variáveis.[?]
- Quando a correlação bivariada é insuficiente para captar padrões multivariados.[?]

34.4.3 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.[?]

34.4.4 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.[?]

34.4.5 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).[?]

34.4.6 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.[?]

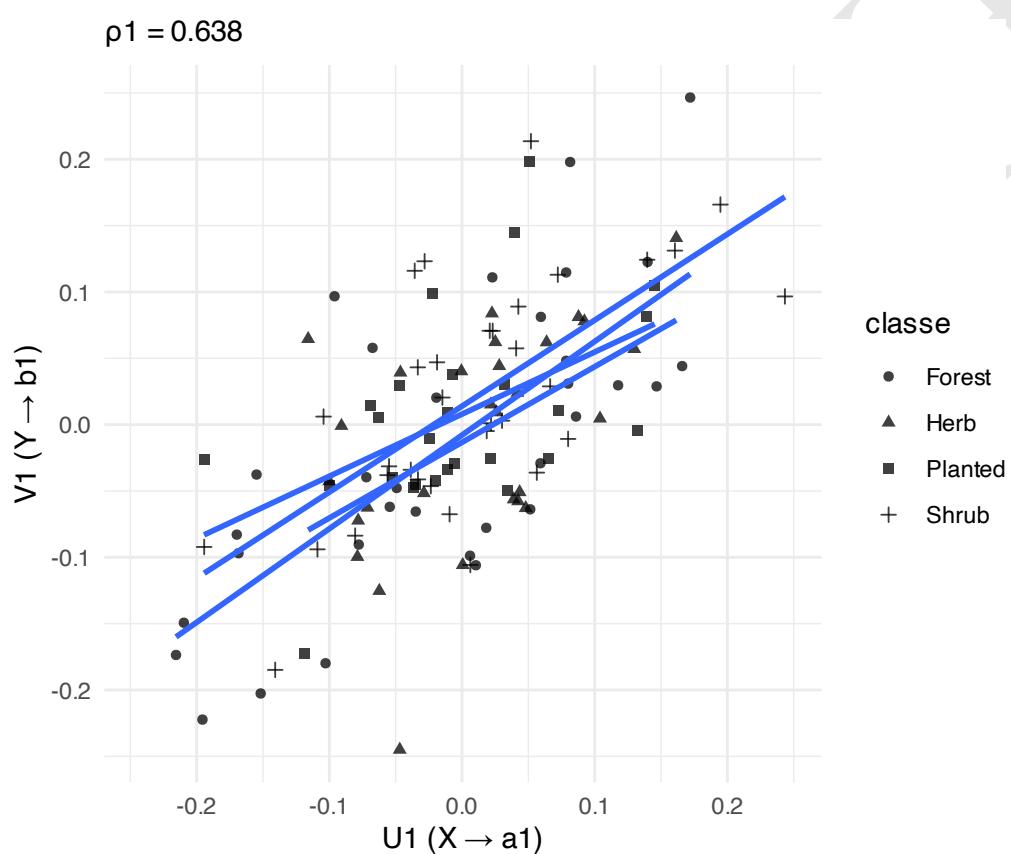


Figura 34.3: Exemplo de análise de correlação canônica (CCA) entre dois conjuntos de variáveis.

Capítulo 35

Redes

35.1 Análise de redes

35.1.1 O que é análise de rede?

- A análise de redes é uma abordagem estatística gráfica baseada na teoria dos grafos que permite representar, explorar e interpretar relações complexas entre múltiplas variáveis analisadas simultaneamente.³¹⁶
- Nessa abordagem, as variáveis são representadas por nodos (ou nós) e as relações entre elas por arestas, formando uma estrutura relacional que evidencia padrões de associação, interdependência e organização do sistema estudado.³¹⁶
- Diferentemente de métodos tradicionais, como análises univariadas ou modelos de regressão clássicos, a análise de redes não foca relações isoladas entre variáveis, mas sim o comportamento conjunto do sistema, permitindo observar fenômenos emergentes que não seriam detectáveis individualmente.³¹⁶
- A análise de redes representa uma mudança conceitual importante em relação às abordagens estatísticas tradicionais, ao enfatizar sistemas, interações e complexidade.³¹⁶

35.1.2 Por que a análise de redes é útil em pesquisa científica?

- Muitos fenômenos científicos, especialmente nas ciências da saúde, são multifatoriais, interdependentes e não lineares, envolvendo variáveis biológicas, comportamentais, psicológicas e sociais.³¹⁶
- Análises univariadas tendem a simplificar esses fenômenos ao avaliar efeitos médios ou relações diretas entre pares de variáveis, o que pode ocultar padrões relevantes.³¹⁶
- A análise de redes permite: visualizar associações simultâneas entre diversas variáveis; identificar variáveis centrais em um sistema; detectar subestruturas densas ou agrupamentos de variáveis fortemente associadas; explorar potenciais mecanismos intermediários ou mediadores.³¹⁶

35.1.3 Quais são as limitações da análise de redes?

Apesar de seu potencial, a interpretação dos resultados deve ser cuidadosa, especialmente em estudos observacionais, evitando inferências causais indevidas.³¹⁶

Tabela 35.1: Exemplo de dados categóricos para criação de matriz de incidência.

Sexo	Tabagismo	Atividade
F	Sim	Baixa
M	Não	Alta
F	Não	Moderada
F	Sim	Baixa
M	Não	Alta

Tabela 35.2: Matriz de incidência criada a partir de dados categóricos.

SexoF	SexoM	TabagismoSim	AtividadeBaixa	AtividadeModerada
1	0	1	1	0
0	1	0	0	0
1	0	0	0	1
1	0	1	1	0
0	1	0	0	0

35.2 Matriz de incidência

35.2.1 O que é uma matriz de incidência?

- Uma matriz de incidência é uma representação tabular que descreve a relação entre dois conjuntos distintos de entidades, como nodos e arestas em uma rede.³¹⁶
- Em uma matriz de incidência, as linhas representam um conjunto de entidades (nodos) e as colunas representam outro conjunto (arestas). A presença ou ausência de uma relação entre as entidades é indicada por valores binários (0 ou 1) ou por pesos numéricos que refletem a intensidade da relação.³¹⁶



O pacote *stats*¹³⁶ fornece a função *model.matrix*^a para criar uma matrix de incidência por expansão de variáveis indicadoras.

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

35.3 Elementos da rede

35.3.1 Quais são os principais elementos de uma rede?

- Nodos (nós): representam as variáveis do estudo, como sintomas, doenças, características clínicas, sociais ou psicológicas.³¹⁶
- Arestas: representam as relações entre os nodos. Podem indicar correlação, associação parcial, dependência condicional ou outro tipo de relação estatística.³¹⁶
- Peso das arestas: em redes ponderadas, a espessura da aresta indica a magnitude da relação; relações mais fortes são representadas por conexões mais espessas.³¹⁶
- Sinal das arestas: geralmente codificado por cores, indicando associações positivas ou negativas.³¹⁶

35.3.2 Como as redes podem ser classificadas?

- Redes não ponderadas: indicam apenas a presença ou ausência de relação entre os nodos.³¹⁶

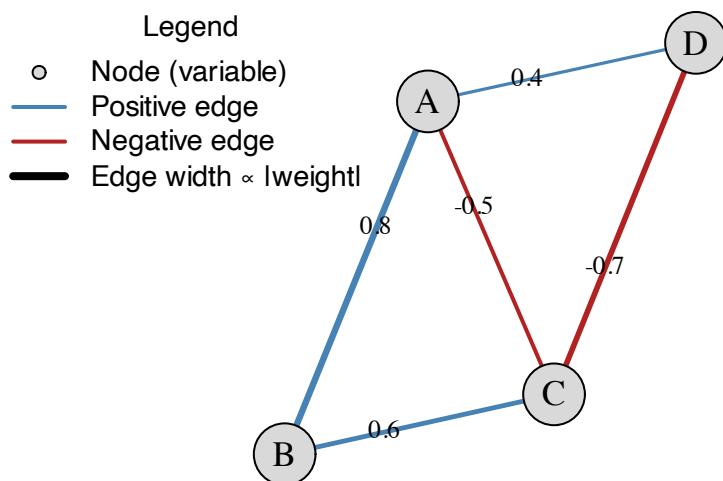


Figura 35.1: Exemplo de grafo de rede com nodos e arestas ponderadas.

- Redes ponderadas: representam também a intensidade da relação.³¹⁶
- Redes direcionais: possuem setas nas arestas, indicando direção da relação (por exemplo, causalidade hipotética).³¹⁶
- Redes não direcionais: não assumem direção causal e são mais apropriadas para estudos observacionais e transversais.³¹⁶
- Em pesquisas transversais, especialmente em saúde, redes não direcionais e ponderadas são geralmente preferidas, pois evitam inferências causais indevidas quando não há informação temporal adequada.³¹⁶

35.3.3 O que define a posição dos nodos em um grafo de rede?

- A posição espacial dos nodos é determinada por algoritmos de disposição (layout), que organizam a rede de modo a facilitar a interpretação visual.³¹⁶
- Um dos algoritmos mais utilizados é o Fruchterman–Reingold³¹⁷, que simula forças de atração e repulsão entre os nodos, posicionando nodos mais fortemente associados mais próximos entre si. A proximidade entre nodos em um grafo geralmente reflete maior associação estatística.³¹⁶

35.4 Tipos de redes

35.4.1 Quais são os principais tipos de redes estatísticas?

- Redes de correlação: baseadas em matrizes de correlação simples; são fáceis de interpretar, mas podem conter associações espúrias.³¹⁶
- Redes de correlação parcial: representam relações entre dois nodos controlando todas as demais variáveis do sistema, reduzindo associações indiretas.³¹⁶
- Modelos gráficos gaussianos: uma forma específica de rede de correlação parcial, muito utilizada em psicométrica e epidemiologia.³¹⁶
- Modelos gráficos mistos: permitem a análise conjunta de variáveis contínuas, ordinais e dicotômicas, comuns em dados reais de saúde.³¹⁶

Tabela 35.3: Tabela com principais métricas de centralidade dos nodos.

Nó		Grau	Força	Intermediação	Proximidade	Agrupamento
A	A	3	3	0.167	1.00	0.667
B	B	2	2	0.000	0.75	1.000
C	C	3	3	0.167	1.00	0.667
D	D	2	2	0.000	0.75	1.000

35.4.2 Como reduzir associações espúrias em redes?

- Em conjuntos de dados com muitas variáveis, redes podem se tornar densas e difíceis de interpretar. Para contornar esse problema, utiliza-se penalização estatística, como o método LASSO (Least Absolute Shrinkage and Selection Operator).³¹⁶
- Na análise de redes, o Graphical LASSO (gLASSO) elimina associações fracas, resultando em redes mais parcimoniosas e interpretáveis.³¹⁶

35.5 Métricas de rede

35.5.1 O que são medidas de centralidade?

- Medidas de centralidade quantificam a importância relativa de cada nodo no sistema.³¹⁶
- Grau: número de conexões diretas de um nodo.³¹⁶
- Força: soma dos pesos das conexões de um nodo, sendo uma das medidas mais utilizadas em redes ponderadas.³¹⁶
- Intermediação: frequência com que um nodo atua como ponte entre outros nodos.³¹⁶
- Proximidade: quanto próximo um nodo está de todos os outros.³¹⁶
- Agrupamento: tendência de um nodo formar grupos com seus vizinhos.³¹⁶

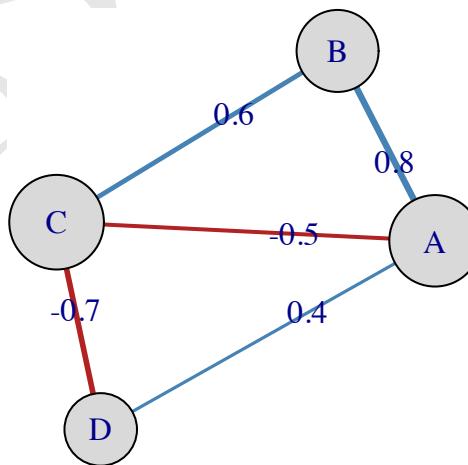


Figura 35.2: Exemplo de grafo de rede com tamanho dos nodos proporcional à força.



O pacote *igraph*³¹⁸ fornece a função *graph_from_incidence_matrix*^a para criar uma rede a partir de uma matriz de incidência.

^ahttps://www.rdocumentation.org/packages/igraph/versions/1.0.0/topics/graph_from_incidence_matrix



O pacote *bootnet*³¹⁹ fornece a função *bootnet*^a para avaliação da estabilidade e precisão das redes por reamostragem.

^a<https://www.rdocumentation.org/packages/bootnet/versions/0.2/topics/bootnet>



O pacote *mgm*³²⁰ fornece a função *mgm*^a para estimação de modelos gráficos mistos.

^a<https://www.rdocumentation.org/packages/mgm/versions/1.2-15/topics/mgm>

RASCUNHO

Capítulo 36

Regressão

36.1 Análise de regressão

36.1.1 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.²⁸⁰
- Para estimar os efeitos imparciais de um fator de exposição primária sobre uma variável de desfecho, frequentemente constroem-se modelos estatísticos de regressão.²²⁶



O pacote *modelsummary*³²¹ 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*²¹⁷ 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

36.1.2 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[?]

36.2 Estruturas de análise de regressão

36.2.1 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).³²²
- A equação de regressão simples é expressa como (36.1), 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.³²²

$$Y = \beta_0 + \beta_1 X + \epsilon \quad (36.1)$$

36.2.2 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.³²²
- A equação de regressão multivariável é expressa como (36.2), 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.³²²

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \quad (36.2)$$

36.2.3 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.³²²
- A equação de regressão multivariada é expressa como (36.3), 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 .³²²

$$Y_1 = \beta_{01} + \beta_{11} X_1 + \beta_{12} X_2 + \dots + \beta_{1n} X_n + \epsilon_1 \quad (36.3)$$

$$Y_2 = \beta_{02} + \beta_{21} X_1 + \beta_{22} X_2 + \dots + \beta_{2n} X_n + \epsilon_2 \quad (36.4)$$

$$\vdots \quad (36.5)$$

$$Y_m = \beta_{0m} + \beta_{m1} X_1 + \beta_{m2} X_2 + \dots + \beta_{mn} X_n + \epsilon_m \quad (36.6)$$

36.3 Tipos e famílias de regressão

36.3.1 O que são modelos de regressão linear?

- Modelos lineares (36.7) descrevem uma relação linear nos parâmetros entre um desfecho contínuo Y e um ou mais preditores X .?

$$Y = \beta_0 + \sum_{i=1}^n \beta_i X_i + \epsilon \quad (36.7)$$

- 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?

Régressão linear

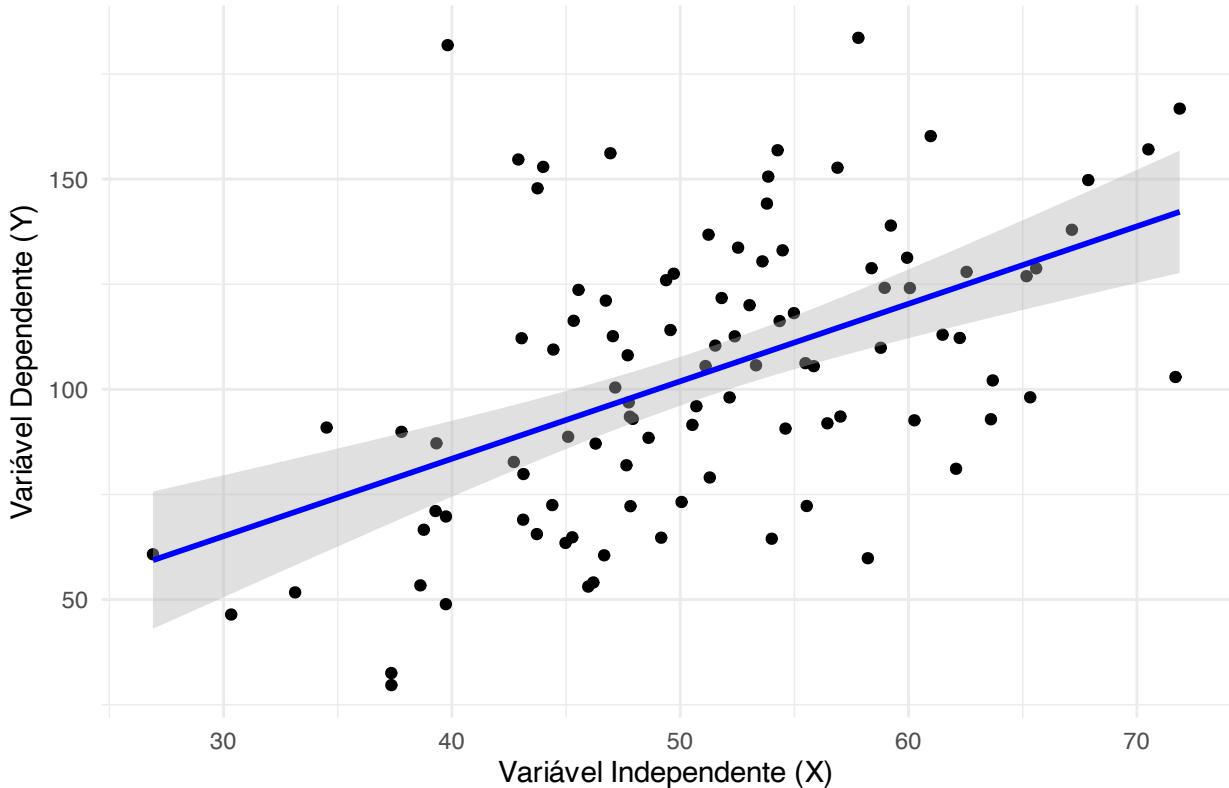


Figura 36.1: Régressão linear.

36.3.2 O que são modelos de régressão polinomial?

- São extensões da régressão linear em que se incluem termos elevados a potências das variáveis independentes (ex.: X^2 , X^3), permitindo capturar relações curvas?
- Modelos de régressão polinomial continuam sendo lineares nos parâmetros, por isso ainda se enquadram como um caso particular da régressão linear?

36.3.3 O que são modelos de régressão 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).

36.3.4 O que são modelos de régressão logística?

- Modelos logísticos são casos de régressão linear generalizada em que a resposta Y é binária.³²³
- A equação (44.12) modela a razão de chances (*odds*) em função dos preditores.³²³

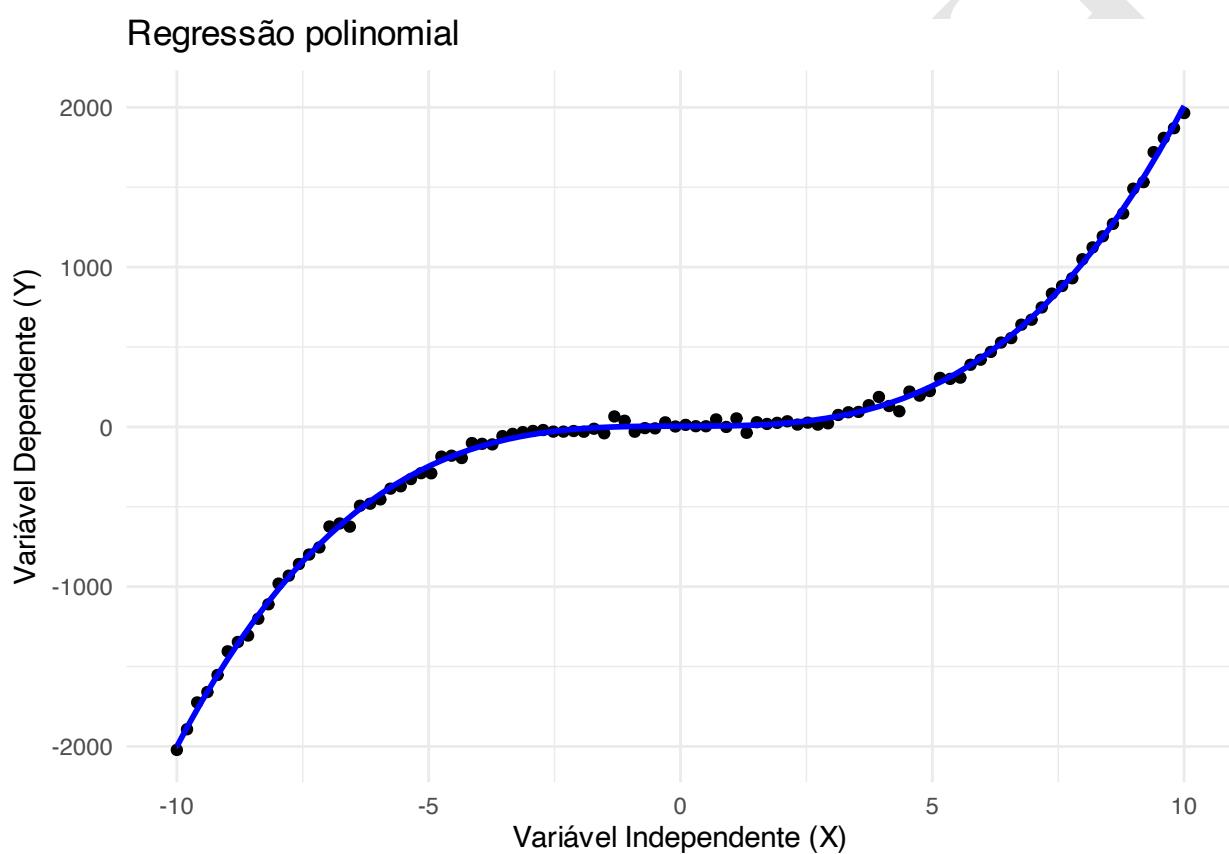


Figura 36.2: Regressão polinomial.

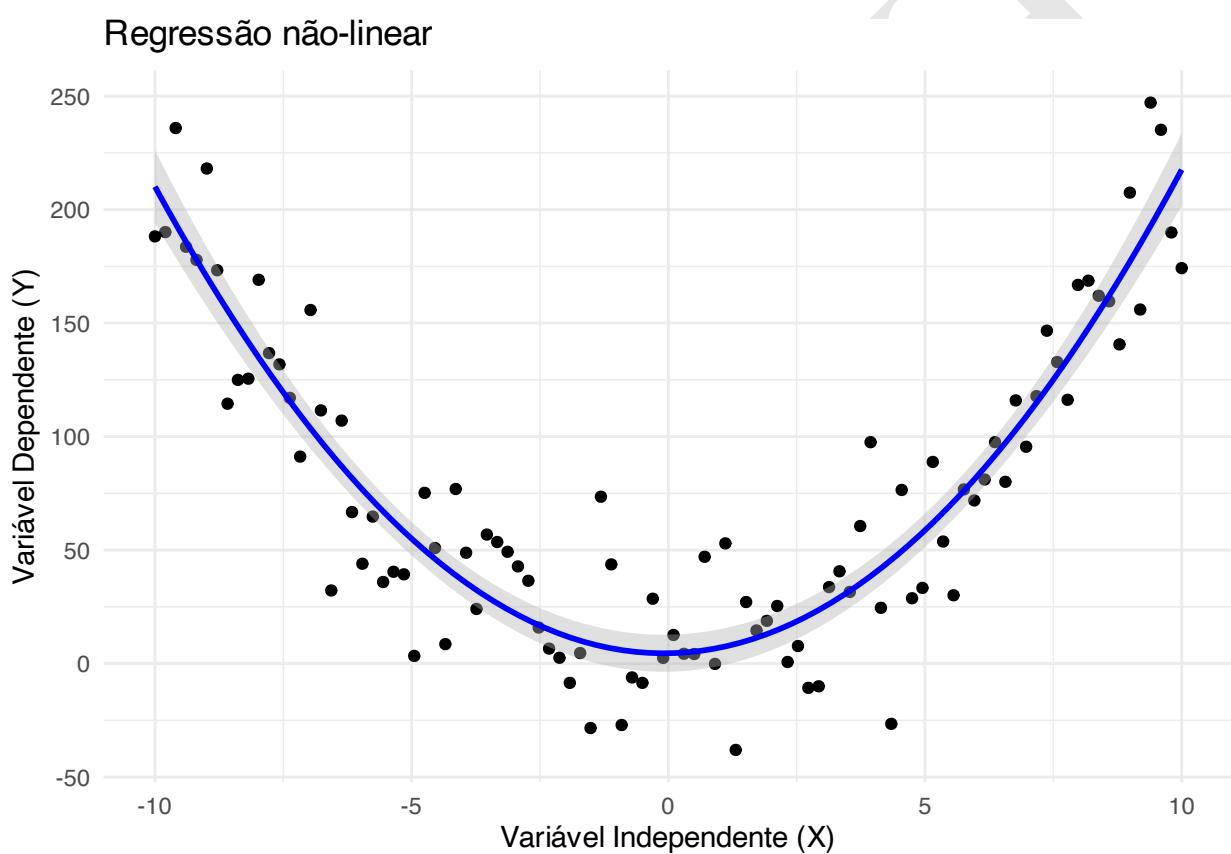


Figura 36.3: Regressão não-linear.

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X + \dots + \beta_n X_n \quad (36.8)$$

- A ligação (*link*) usada é o logit (44.13).³²³

$$g(p) = \log\left(\frac{p}{1-p}\right) \quad (36.9)$$

- A interpretação dos coeficientes β_i pode ser feita em termos de razões de chances (*odds ratios*, *OR*), por exponenciação dos coeficientes: $OR_i = e^{\beta_i}$, o que representa o fator multiplicativo na *OR* do desfecho para cada aumento de uma unidade em X_i (mantendo os demais preditores constantes).³²³
- A interpretação pode ser feita por estimativa da variação percentual na chance (*OR*) de ocorrência de Y , calculando $(e^b - 1) \times 100$, de modo que um aumento de 1 unidade em X está associado a um aumento de $(e^b - 1) \times 100\%$ na chance de Y ocorrer (mantidos os demais preditores constantes).

Regressão logística

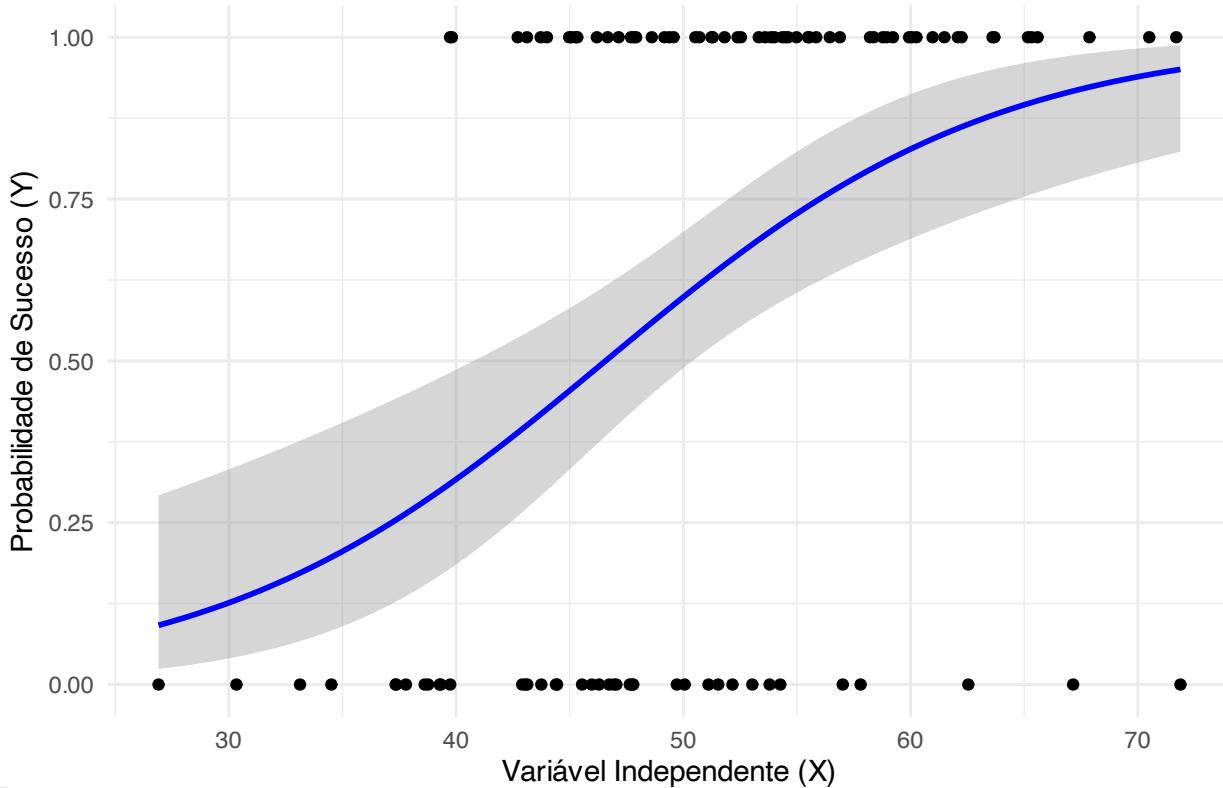


Figura 36.4: Regressão logística.

36.3.5 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.³
- Estendem a regressão logística binária, modelando as razões de chances (*odds ratios*) de cada categoria em relação a uma categoria de referência.³

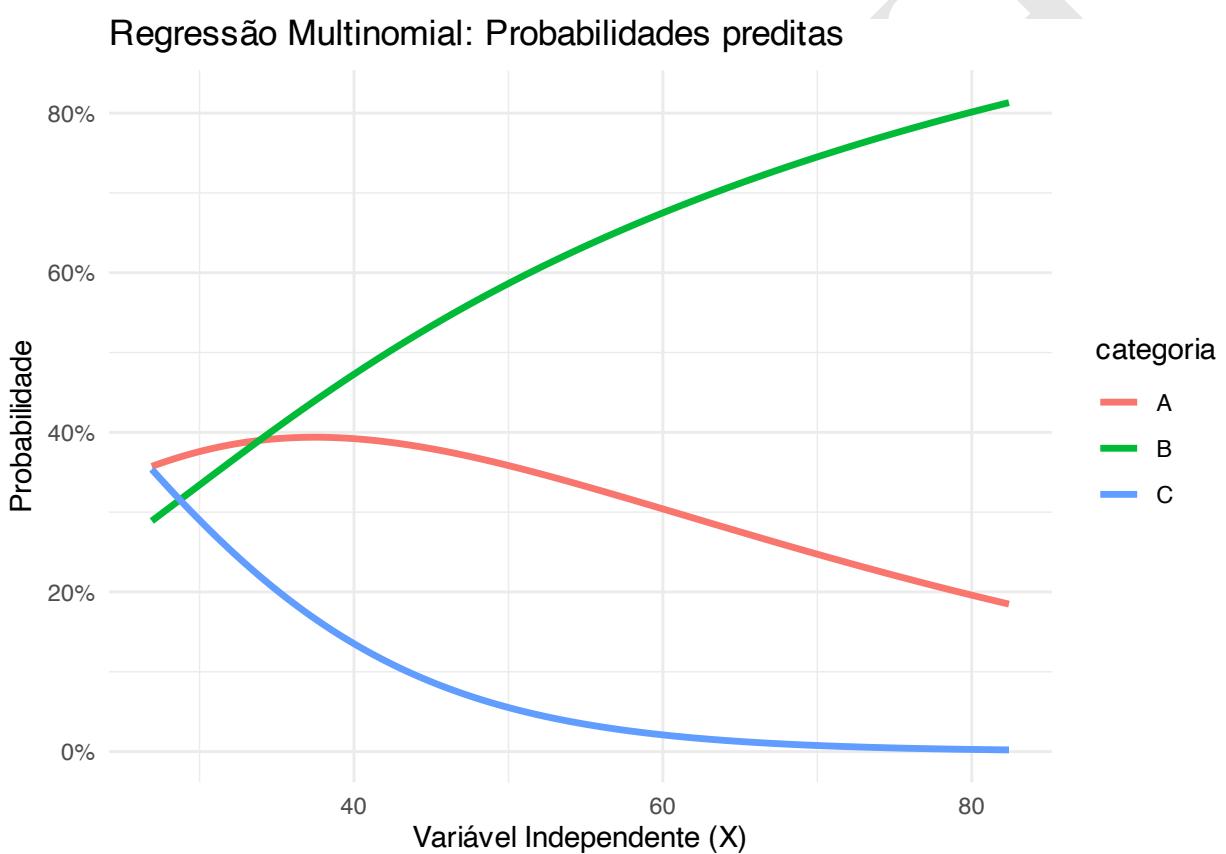


Figura 36.5: Regressão multinomial

36.3.6 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.⁹
- Modelam a probabilidade acumulada de estar em ou abaixo de cada categoria, usando uma função de ligação logit, probit ou log-log.⁹
- Assumem a proporcionalidade dos coeficientes entre as categorias (*proportional odds*).⁹

36.3.7 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.⁹
- Assumem que $Y \sim Poisson(\mu)$, com $\mu = E[Y|X]$ relacionado aos preditores via função de ligação log.⁹
- A sobre-dispersão (variância maior que a média) pode exigir modelos alternativos como a regressão binomial negativa.⁹

Régressão de Poisson (link log)

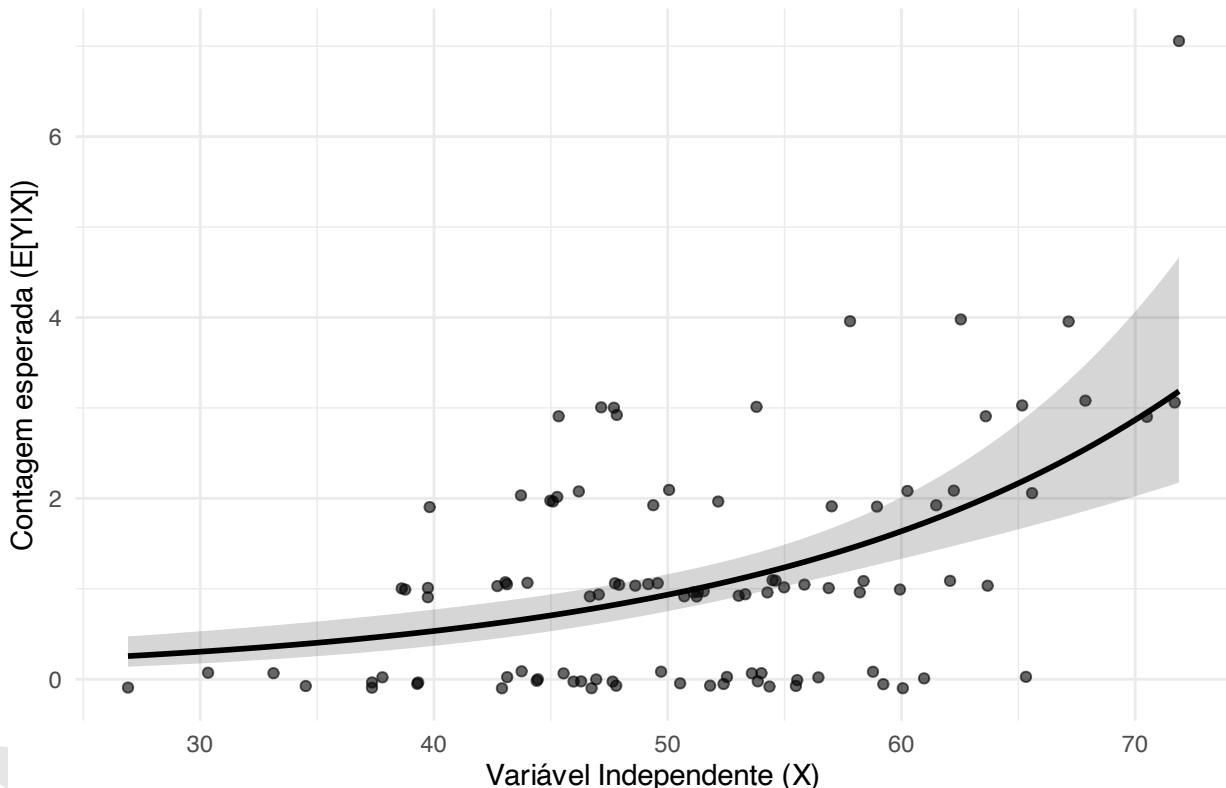


Figura 36.6: Regressão de Poisson.

36.3.8 O que são modelos de regressão binomial negativa?

- Modelos de regressão binomial negativa são usados para contagens superdispersas, onde a variância excede a média.⁹
- Introduzem um parâmetro de dispersão adicional para modelar a variabilidade extra.⁹

- A função de ligação log é comumente usada, semelhante à regressão de Poisson.[?]

36.3.9 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.[?]

36.3.10 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).[?]
- Usados para dados hierárquicos ou longitudinais, onde observações estão agrupadas.[?]
- Permitem modelar correlações intra-grupo e variabilidade entre grupos.[?]

36.3.11 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).[?]
- Combinam a flexibilidade dos modelos lineares generalizados com a capacidade de modelar correlações e variabilidade entre grupos.[?]
- Usados para dados hierárquicos ou longitudinais com desfechos não normais.[?]

36.3.12 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.[?]
- Ajuda a reduzir multicolinearidade e overfitting, encolhendo os coeficientes em direção a zero, mas nunca os tornando exatamente nulos.[?]
- O hiperparâmetro de regularização é λ , controlando a intensidade da penalização. Valores maiores de λ resultam em maior encolhimento dos coeficientes.[?]

36.3.13 O que são modelos de regressão LASSO?

- Regressão LASSO (*Least Absolute Shrinkage and Selection Operator*) utiliza penalização L1, que pode zerar coeficientes.[?]
- Além de reduzir overfitting, também realiza seleção automática de variáveis.[?]
- Enquanto a regressão ridge mantém todos os preditores, a LASSO pode excluir variáveis irrelevantes.[?]

36.4 Preparação de variáveis

36.4.1 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.³²⁴
- Variáveis categóricas nominais, com 2 ou mais níveis, devem ser subdivididas em variáveis fictícias dicotômicas para ser usada em modelos de regressão.³²⁵

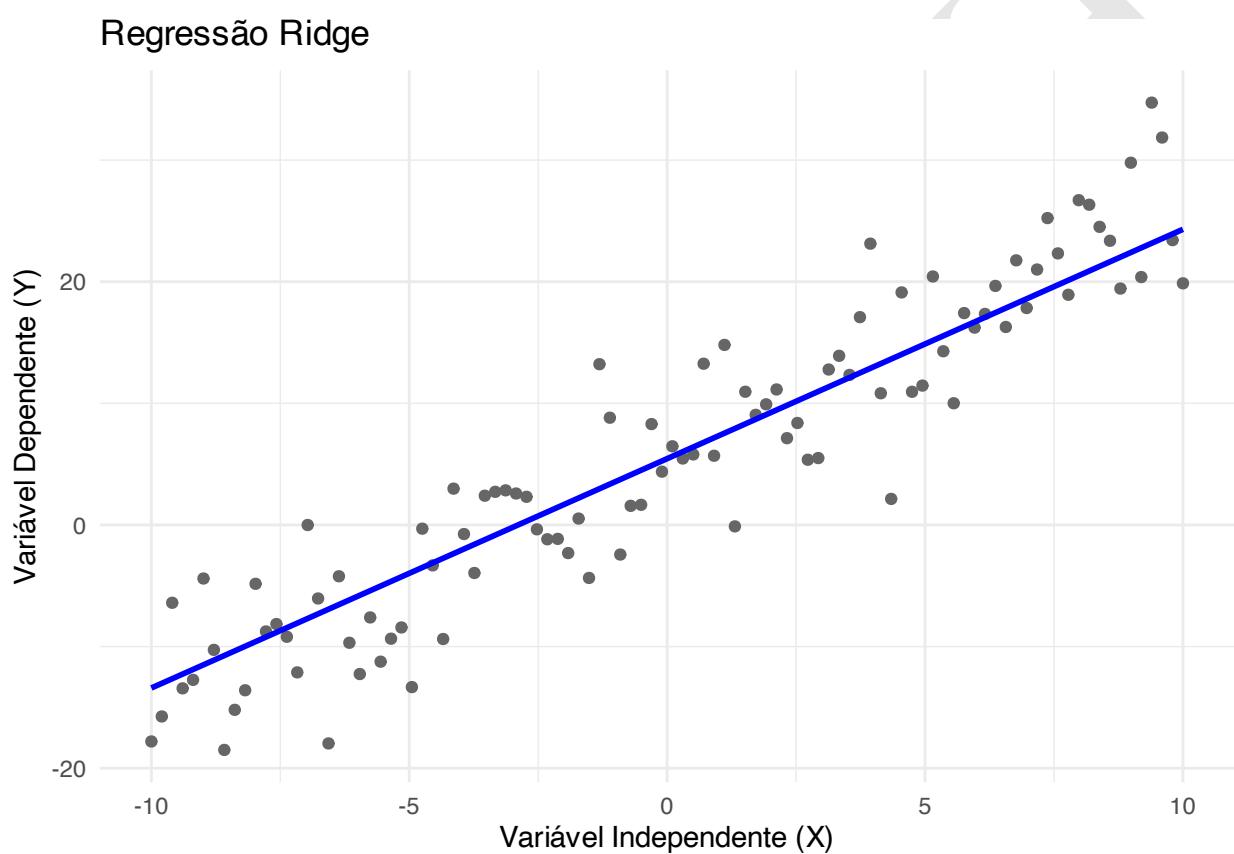


Figura 36.7: Regressão ridge.

- Cada nível da variável categórica nominal será convertido em uma nova variável fictícias dicotômica, tal que a nova variável dicotômica assume valor 1 para a presença do nível correspondente e 0 em qualquer outro caso.³²⁵



O pacote *fastDummies*³²⁶ 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

36.4.2 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

36.5 Multicolinearidade

36.5.1 O que é multicolinearidade?

- Multicolinearidade representa a intercorrelação entre as variáveis independentes (explanatórias) de um modelo.³¹⁴

36.5.2 Como diagnosticar multicolinearidade de forma quantitativa?

- Verifique a existência de multicolinearidade entre as variáveis candidatas.³²⁷
- 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.³¹⁴
- 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.³¹⁴
- 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.³¹⁴
- 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.³¹⁴



O pacote *GGally*³¹⁵ 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>

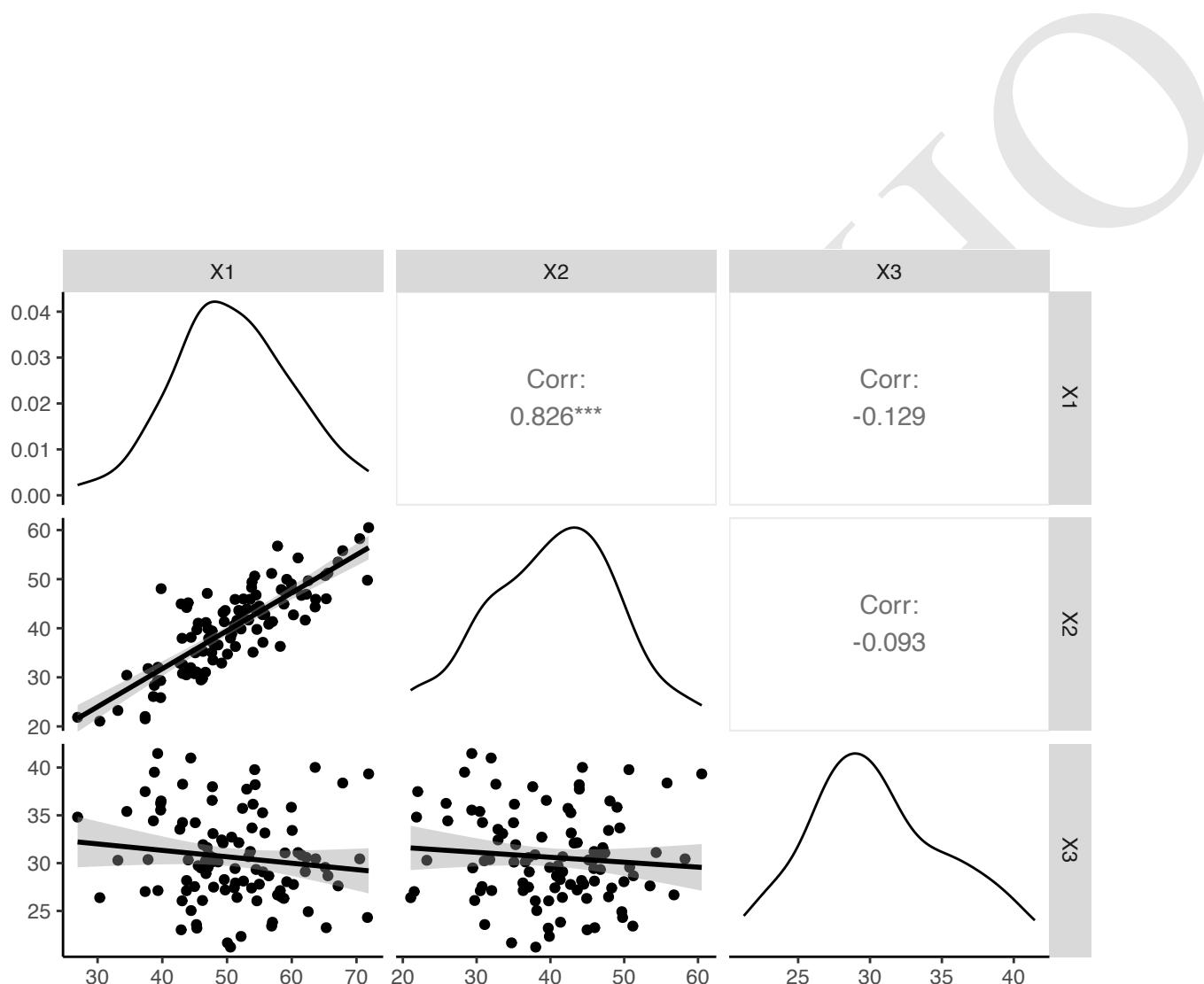


Figura 36.8: Multicolinearidade entre variáveis candidatas em modelos de regressão multivariável.



O pacote *car*³²⁸ 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>

36.5.3 O que fazer em caso de multicolinearidade elevada?

- Verifique a transformação (codificação) de variáveis numéricas em categóricas.³¹⁴
- Aumente o tamanho da amostra, se possível, para reduzir a multicolinearidade.³¹⁴
- Combine níveis de variáveis categóricas com baixa frequência de ocorrência.³¹⁴
- Combine variáveis numéricas altamente correlacionadas em uma única variável composta, como a média ou soma das variáveis.³¹⁴
- Considere a exclusão de variáveis altamente correlacionadas do modelo, especialmente se elas não forem essenciais para a análise.³¹⁴
- Use técnicas de seleção de variáveis, como seleção passo a passo, para identificar e remover variáveis redundantes.³¹⁴
- Use técnicas de regularização, como regressão ridge ou LASSO, que podem lidar com multicolinearidade ao penalizar coeficientes de regressão.³¹⁴

36.6 Redução de dimensionalidade

36.6.1 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.^{295,327,329}
- A seleção bivariada de variáveis torna o modelo mais suscetível a otimismo no ajuste se as variáveis de confundimento não são adequadamente controladas.^{327,329}

36.6.2 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.²⁹⁵
- 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.²⁹⁵

36.6.3 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.³²⁵
- Nenhum método de regressão gradual garante a seleção ótima de variáveis de um banco de dados.³²⁵
- As regras de término da regressão baseadas em P-valor tendem a ser arbitrárias.³²⁵

36.6.4 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.³²⁷
- Colapse categorias com contagem nula (células com valor igual a 0) de variáveis candidatas.³²⁷
- Use simulações de dados para identificar qual(is) variável(is) está(ão) causando problemas de convergência do ajuste do modelo.³²⁷
- 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.²⁹⁵

36.6.5 Quando devemos forçar uma variável no modelo?

- Sempre que houver base teórica ou evidência prévia forte, ou se for a variável de exposição principal.²⁵⁹

36.7 Seleção de variáveis em regressão

36.7.1 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.³³⁰

36.7.2 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*.³³⁰
- R^2_{adj} (36.10) 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.

$$R^2_{adj} = 1 - \frac{(n-1)}{(n-k-1)} \cdot \frac{RSS}{SST} \quad (36.10)$$

- *AIC* (Akaike Information Criterion) (36.11) mede o equilíbrio entre ajuste e complexidade:

$$AIC = n \cdot \log\left(\frac{RSS}{n}\right) + 2k + n + n \cdot \log(2\pi) \quad (36.11)$$

- *AICc* (36.12) é uma versão corrigida do AIC, preferida para amostras pequenas:

$$AIC_c = AIC + \frac{2(k+2)(k+3)}{n-(k+2)-1} \quad (36.12)$$

- 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:

$$C_p = (n-m-1) \frac{RSS}{RSS_{FULL}} - (n-2p) \quad (36.13)$$

- *BIC* (Bayesian Information Criterion) (36.14) penaliza fortemente modelos complexos:

$$BIC = n \cdot \log\left(\frac{RSS}{n}\right) + k \cdot \log(n) + n + n \cdot \log(2\pi) \quad (36.14)$$

36.7.3 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.³³⁰
- 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.³³⁰
- *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).³³⁰
- Esses métodos podem divergir em presença de alta multicolinearidade ou amostras pequenas, e devem ser acompanhados de diagnóstico de resíduos e validação cruzada.³³⁰



O pacote *leaps*³³¹ 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*³³² 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*³³² 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 37

Mediação

37.1 Análise inferencial de mediação

37.1.1 O que é análise de mediação?

- ?

RASCUNHO

PARTE 6: MODELAGEM

Estratégias para entender relações complexas, prever resultados e explorar padrões ocultos

RASCUNHO

Capítulo 38

Modelos

38.1 Modelos

38.1.1 O que são modelos?

- Modelos são representações simplificadas de um sistema real, usados para entender, prever ou controlar fenômenos complexos.³

38.1.2 O que é modelagem?

- Modelagem é o processo de usar dados para selecionar um modelo matemático explícito que represente o processo gerador dos dados.²⁵⁹

38.1.3 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.²⁵⁹
- Todos os modelos são errados, mas alguns são úteis.³³³



O pacote *equatiomatic*³³⁴ 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

38.1.4 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.³³⁵

38.2 Modelos estocásticos

38.2.1 O que são modelos estocásticos?

- ?

38.2.2 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.³³⁶

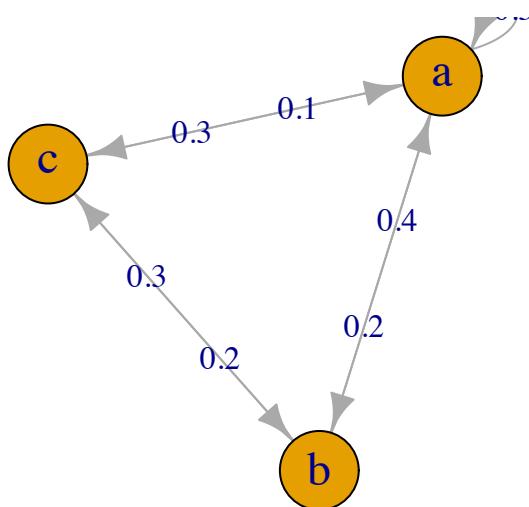


Figura 38.1: Cadeia de Markov com 3 estados (a, b, c) e suas probabilidades de transição.

R

O pacote *markovchain*³³⁷ 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>

38.3 Efeito fixo

38.3.1 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.[?]

38.4 Efeito aleatório

38.4.1 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.[?]

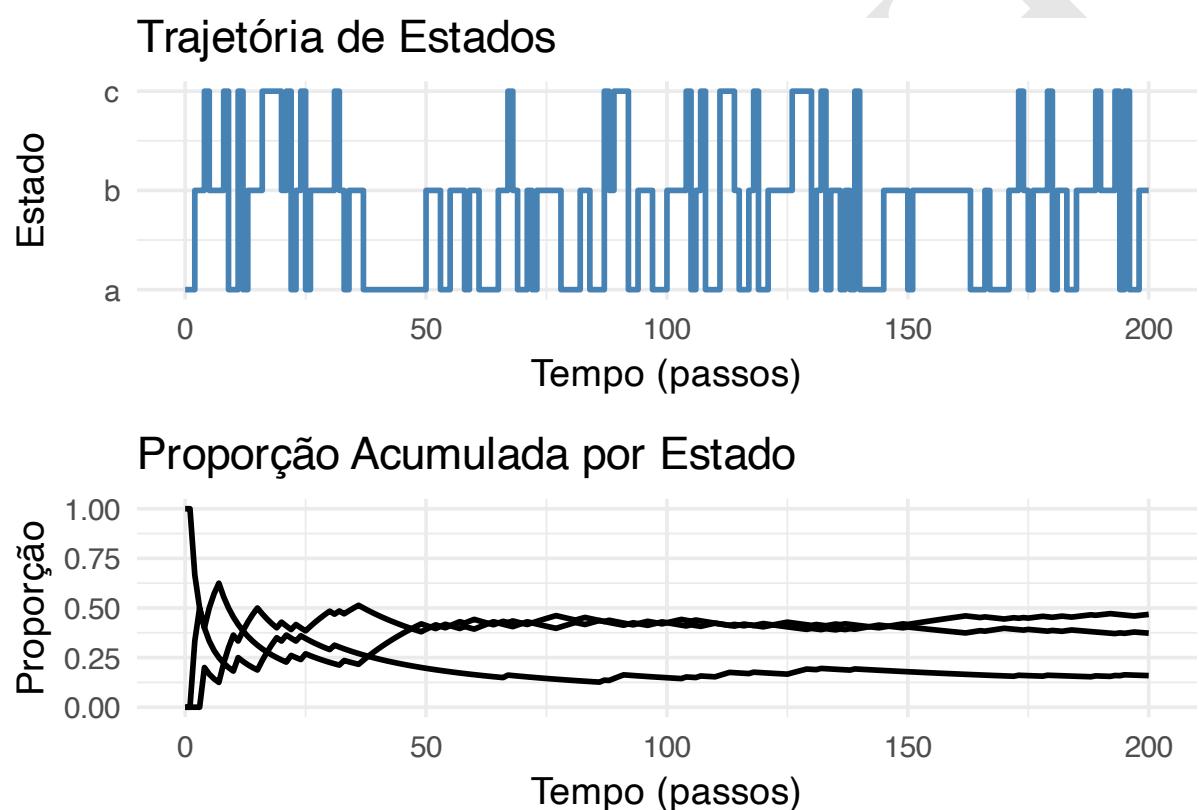


Figura 38.2: Trajetória de estados e proporção acumulada por estado em uma cadeia de Markov com 3 estados (a, b, c).

38.5 Efeito misto

38.5.1 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

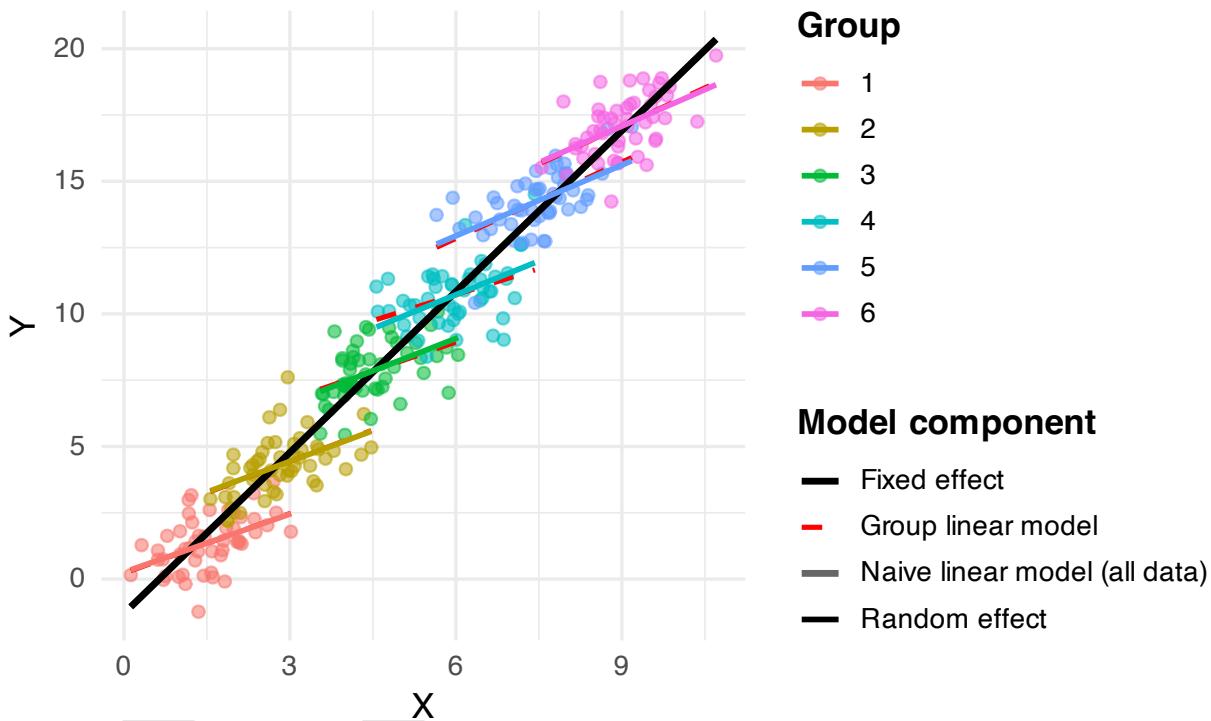


Figura 38.3: Efeitos fixos, aleatórios e mistos em dados simulados com paradoxo de Simpson. As linhas vermelhas representam os efeitos dentro dos grupos, enquanto as linhas cinza e preta representam os efeitos globais (naive e fixo, respectivamente). O modelo misto (linhas coloridas) captura os efeitos dentro dos grupos sem extrapolar além dos dados observados.

38.6 Efeito principal

38.6.1 O que é efeito principal?

- 338

38.7 Efeito de interação

38.7.1 O que é efeito de interação?

- A interação - representada pelo símbolo * - é o termo estatístico empregado para representar a heterogeneidade de um determinado efeito.³³⁹
- ³³⁸

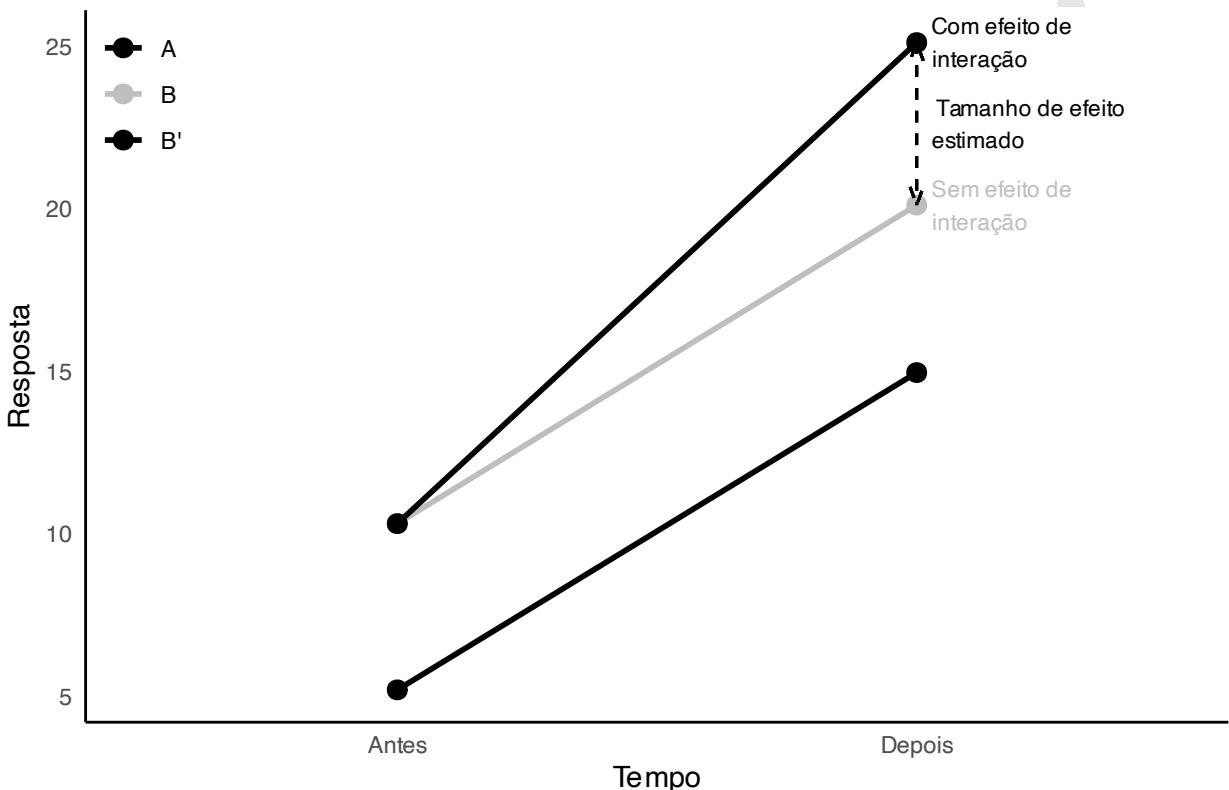


Figura 38.4: Análise de efeito de interação (direta) entre grupos e tempo. Retas paralelas sugerem ausência de efeito de interação.

R O pacote *nlme*³⁴⁰ 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*³⁴¹ 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*³⁴² 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>

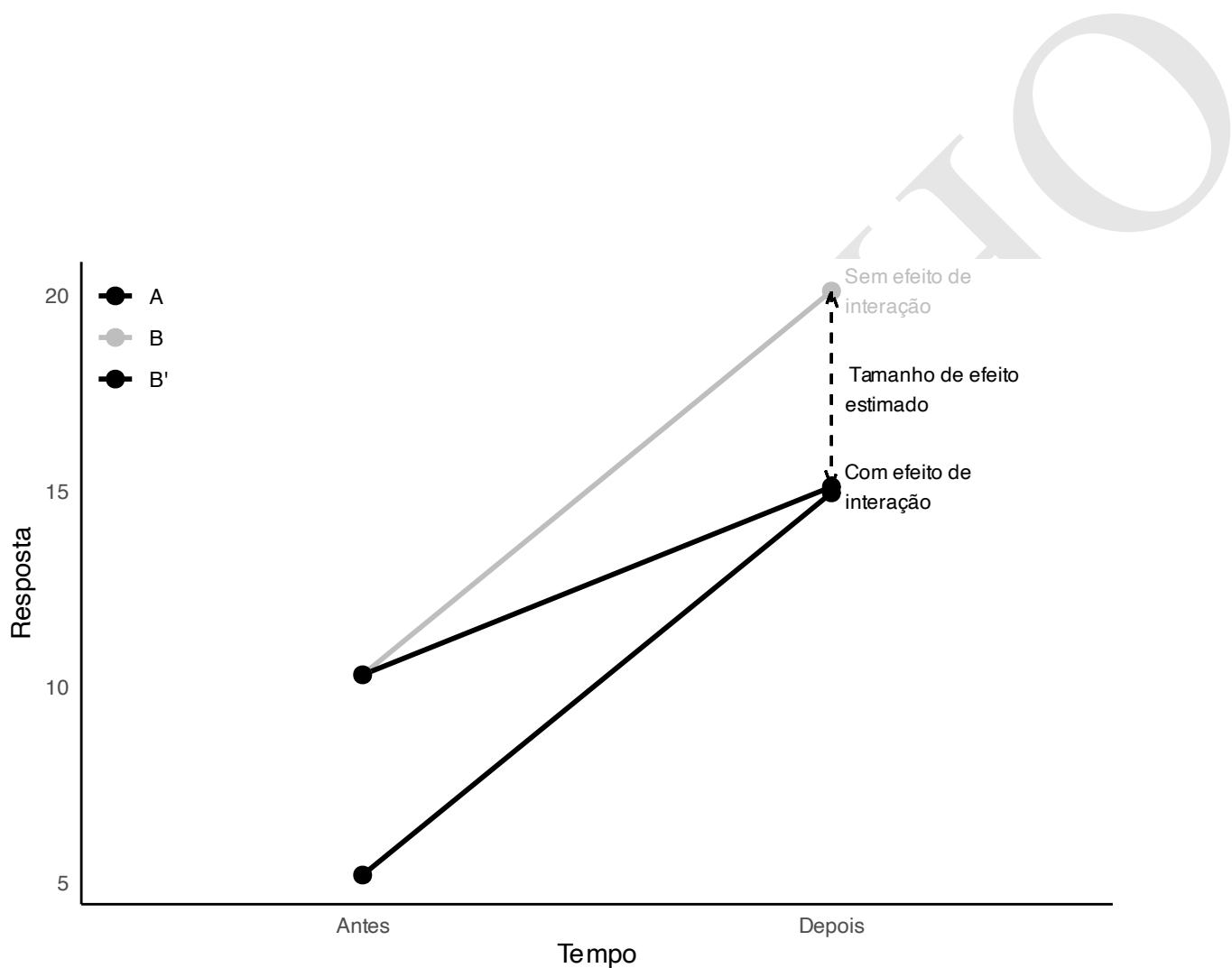


Figura 38.5: Análise de efeito de interação (inversa) entre grupos e tempo. Retas paralelas sugerem ausência de efeito de interação.

38.8 Efeito de mediação

38.8.1 O que é um mediador de efeito?

- 343
- 338

38.8.2 O que é efeito de mediação?

- 343
- 338

38.8.3 O que é efeito direto?

- 343
- 338

38.8.4 O que é efeito indireto?

- 343
- 338

38.8.5 O que é efeito total?

- 343
- 338

38.9 Efeito de modificação

38.9.1 O que é um modificador de efeito?

- 338

38.9.2 O que é efeito de modificação?

- 338

38.10 Preditores e desfechos

38.10.1 O que são desfechos de um modelo?

- ?

38.10.2 O que são preditores de um modelo?

- ?

38.10.3 Como selecionar preditores para um modelo?

- ?

38.11 Desempenho e estabilidade de modelos

38.11.1 Como avaliar o desempenho dos modelos?

- Pela área sob a curva ROC em conjunto com o otimismo (diferença entre AUC aparente e validada).³³⁵
- O desempenho melhora com maior tamanho amostral, mas de forma desigual entre técnicas.³³⁵

38.11.2 Qual modelo alcança estabilidade mais rapidamente?

- Regressão logística é o mais estável e menos *data hungry*.³³⁵
- Árvore de decisão para classificação e regressão estabiliza rápido, mas em nível de desempenho baixo.³³⁵
- Máquina de vetores de suporte, redes neurais e *random forests* apresentam instabilidade mesmo em amostras muito grandes.³³⁵

38.12 Comparação de modelos

38.12.1 Como comparar modelos estatísticos?

- ?

38.12.2 Como comparar modelos de aprendizagem de máquina?

- ?



O pacote *correctR*³⁴⁴ fornece funções para comparar o desempenho e a qualidade do ajuste de diversos modelos de aprendizagem de máquina em amostras correlacionadas.

38.13 Avaliação de modelos

38.13.1 Como avaliar a qualidade de ajuste de um modelo?

- Coeficiente de determinação (R^2) (38.1) e R^2 ajustado (38.2): Medem a proporção da variabilidade dos dados explicada pelo modelo. O R^2 ajustado penaliza a inclusão de variáveis irrelevantes.³

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (38.1)$$

$$R^2_{ajustado} = 1 - (1 - R^2) \frac{n - 1}{n - p - 1} \quad (38.2)$$

- Erro quadrático médio (*RMSE*) (38.3): 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

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (38.3)$$

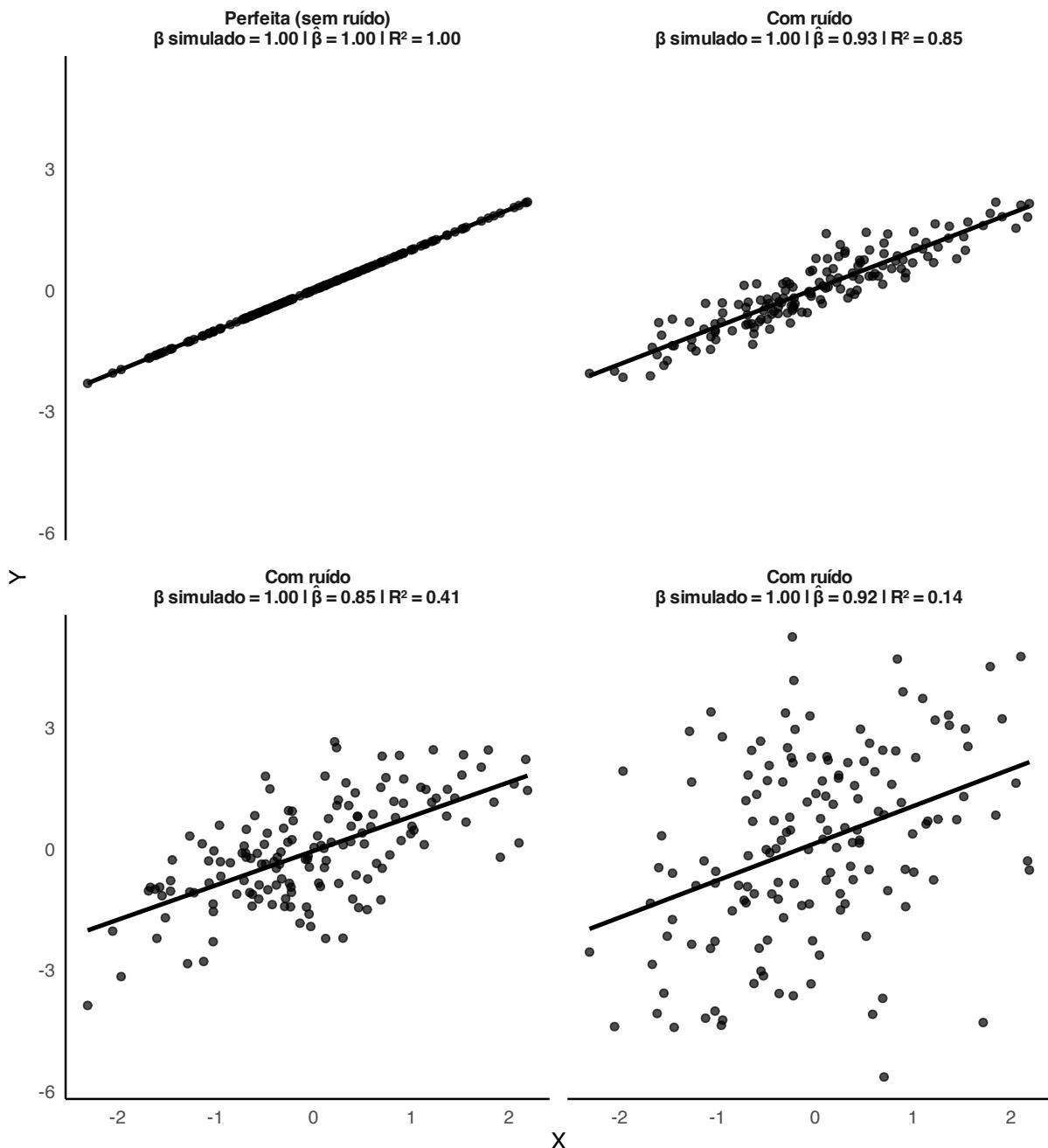


Figura 38.6: Exemplos de ajuste de modelos de regressão linear simples ($y \sim x$) com diferentes níveis de ruído (R^2). Cada painel mostra a reta ajustada (cinza) e os valores observados (pontos). Os valores anotados indicam o coeficiente angular simulado (β), o coeficiente angular estimado ($\hat{\beta}$) e o R^2 observado.

Tabela 38.1: 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*) (38.4) e Critério de Informação Bayesiano (*BIC*) (38.5): 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.²⁵⁶

$$AIC = 2k - 2 \ln(L) \quad (38.4)$$

$$BIC = \ln(n)k - 2 \ln(L) \quad (38.5)$$

- Desvio residual (σ): Mede a variabilidade dos resíduos do modelo. Valores menores indicam melhor ajuste.²⁵⁷



O pacote *performance*²⁵⁷ 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*²⁵⁷ 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

38.14 Validação de modelos

38.14.1 Como validar modelos estatísticos?

- ?

38.15 Calibração de modelos

38.15.1 Como calibrar modelos estatísticos?

- ?

Capítulo 39

Modelagem temporal

39.1 Modelos temporais

39.1.1 O que são modelos temporais?

• ?



O pacote *UComp*³⁴⁵ 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 40

Modelagem espacial

40.1 Modelos espaciais

40.1.1 O que são modelos espaciais?

• ?



Os pacotes *sf*³⁴⁶ e *ggplot*¹⁷⁹ 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*^[347] 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 41

Modelagem de sobrevida

41.1 Análise de sobrevida

41.1.1 O que é análise de sobrevida?

- É um conjunto de métodos estatísticos utilizados para analisar o tempo até a ocorrência de um evento de interesse.³⁴⁸

41.1.2 Quando usar análise de sobrevida?

- Use quando o desfecho envolve tempo até o evento, nem todos os indivíduos apresentam o evento, há diferentes tempos de seguimento e existem perdas de acompanhamento (censura).³⁴⁸

41.2 Eventos

- O evento não precisa ser morte — pode ser qualquer desfecho binário que ocorra (ou não) durante o período de observação, tais como: tempo até recidiva de doença; até ocorrência de um sintoma; até falha terapêutica.³⁴⁸

41.2.1 O que caracteriza um “evento” em análise de sobrevida?

- É um desfecho binário que indica se o evento ocorreu (= 1) ou foi censurado/não ocorreu até o tempo observado (= 0).³⁴⁸

41.3 Dados censurados

41.3.1 O que são dados censurados?

- São observações em que o evento não ocorreu até o fim do estudo, geralmente porque o participante foi perdido no seguimento ou o estudo terminou antes do evento ocorrer.³⁴⁸
- Na análise tradicional, esses casos seriam excluídos. Na análise de sobrevida, eles são incorporados ao modelo.³⁴⁸

41.3.2 Quais são os tipos de censura de dados?

- Censura à direita: O evento não ocorreu até o fim do estudo ou perda de seguimento.³⁴⁸

- Censura à esquerda: O evento ocorreu antes do início do estudo.³⁴⁸
- Censura intervalar: O evento ocorreu entre dois pontos de tempo, mas o momento exato é desconhecido.³⁴⁸

41.4 Medidas de associação em análise de sobrevida

41.4.1 O que é a função de sobrevida?

- A função de sobrevida, $S(t)$, representa a probabilidade de o evento não ter ocorrido até o tempo t , isto é, $S(t) = P(T > t)$.³⁴⁸

41.4.2 O que é a função de risco?

- A função de risco, $h(t)$, representa a taxa instantânea de ocorrência do evento no tempo t , condicional ao fato de o indivíduo ainda não ter apresentado o evento até esse momento.³⁴⁸

41.4.3 O que é a razão de risco (*hazard ratio*)?

- A razão de risco (hazard ratio, HR) é a razão entre as taxas instantâneas de ocorrência do evento em dois grupos ao longo do tempo.³⁴⁸
- O hazard ratio compara riscos instantâneos e não probabilidades acumuladas.³⁴⁸
- $HR > 1$: O grupo de tratamento tem um risco maior de evento em comparação ao grupo controle.³⁴⁸
- $HR = 1$: Não há diferença no risco de evento entre os grupos.³⁴⁸
- $HR < 1$: O grupo de tratamento tem um risco menor de evento em comparação ao grupo controle.³⁴⁸

41.4.4 Qual é a diferença entre modelos de Kaplan–Meier e Cox?

- Kaplan–Meier estima a função de sobrevida sem ajustar para covariáveis.³⁴⁸
- O teste log-rank compara curvas entre grupos.³⁴⁸
- O modelo de Cox permite ajustar para múltiplas covariáveis e estima razões de risco ajustadas.³⁴⁸

41.4.5 O que é o tempo médio de sobrevida restrito?

- O tempo médio de sobrevida restrito (*Restricted Mean Survival Time*, $RMST$) é definido como a área sob a curva de sobrevida até um tempo pré-especificado τ (41.1).³⁴⁹

$$\mu(\tau) = \int_0^\tau S(t)dt \quad (41.1)$$

- Diferentemente do HR , o $RMST$ não depende da suposição de riscos proporcionais.³⁴⁹
- O $RMST$ corresponde à média truncada do tempo até o evento e por isso possui interpretação clínica direta (diferença média de tempo).³⁴⁹
- Essa diferença corresponde à área entre as curvas de sobrevida até o tempo τ .³⁴⁹
- $RMST$ é robusto quando há violação da suposição de riscos proporcionais.³⁴⁹
- $RMST$ é recomendado quando há separação tardia das curvas.³⁴⁹
- Enquanto o HR mede uma razão instantânea de riscos, o $RMST$ mede uma diferença média acumulada de tempo.

41.5 Modelo de Kaplan–Meier

41.5.1 O que é a curva de Kaplan–Meier?

- A curva de Kaplan–Meier é uma estimativa não paramétrica da função de sobrevida ao longo do tempo, que leva em consideração os dados censurados.³⁴⁸
- Ela é construída a partir dos tempos de evento e censura, e mostra a probabilidade de sobrevivência em diferentes pontos no tempo.³⁴⁸
- A curva de Kaplan–Meier é útil para comparar a sobrevida entre grupos e para visualizar a distribuição do tempo até o evento.³⁴⁸

41.5.2 Como interpretar as curvas de Kaplan–Meier?

- A curva mostra a probabilidade de sobrevivência ao longo do tempo para um grupo específico.³⁴⁸
- A distância entre as curvas de diferentes grupos pode indicar diferenças na sobrevida.³⁴⁸
- O teste log-rank pode ser usado para avaliar se as diferenças entre as curvas são estatisticamente significativas.³⁴⁸

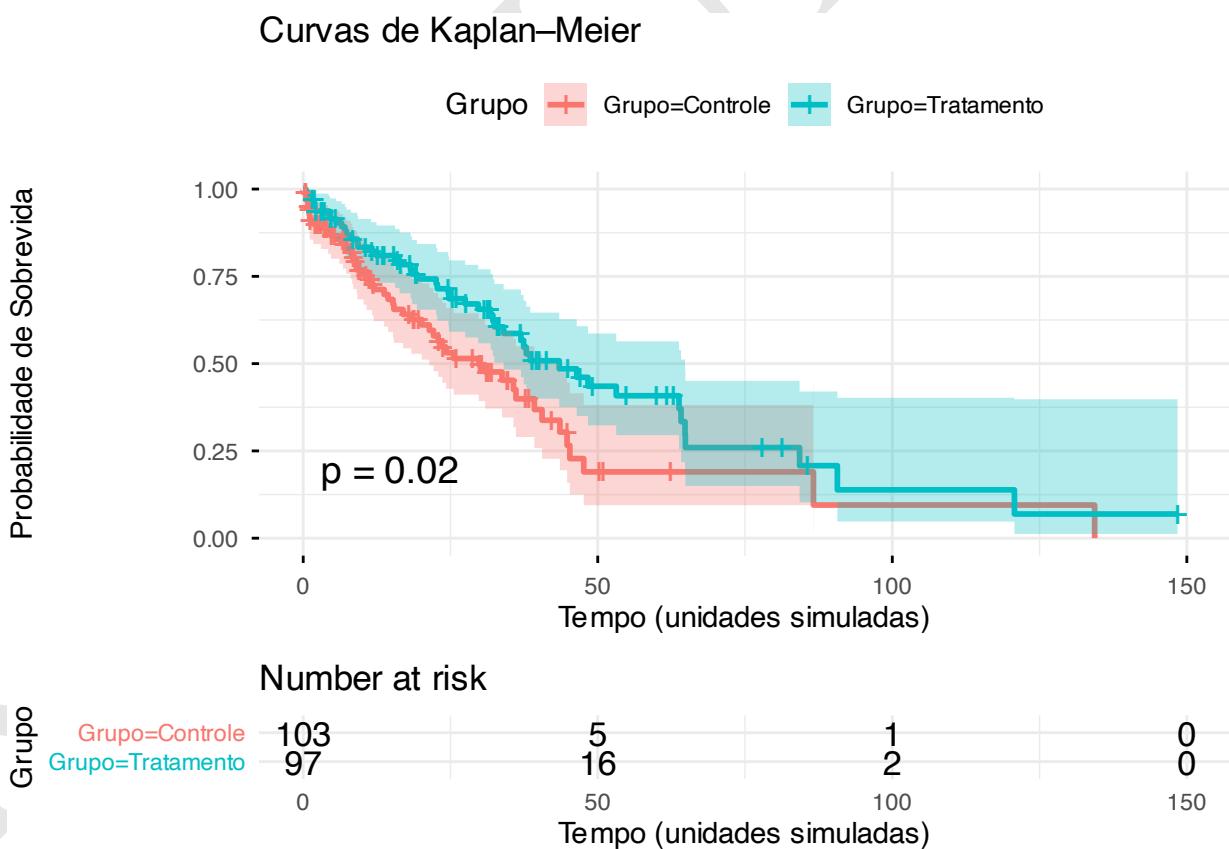


Figura 41.1: Curvas de Kaplan–Meier simuladas para dois grupos (controle e tratamento).

Grupo	N	Eventos	Censuras	Mediana (IC 95%)
Controle	103	51	52	29.81 (21.41 – 43.58)
Tratamento	97	47	50	43.47 (34.03 – 64.82)

Teste log-rank: $p = 0.02$

Características	HR	IC 95% (inf)	IC 95% (sup)	p-valor
Grupo				0.021
Controle	—	—		
Tratamento	0.62	0.42, 0.93	0.93	

Abreviações: HR = Taxa de risco, IC = Intervalo de Confiança

41.6 Modelos de Cox

41.6.1 O que é o modelo de Cox?

- O modelo de Cox, ou modelo de riscos proporcionais de Cox, é um modelo de regressão semiparamétrico utilizado para analisar a relação entre o tempo até o evento e um conjunto de covariáveis.³⁴⁸
- O modelo de Cox assume que a razão de riscos entre os grupos é constante ao longo do tempo (proporcionalidade dos riscos).³⁴⁸

41.6.2 Como interpretar o coeficiente do modelo de Cox?

- O modelo estima o logaritmo da razão de riscos ($\log(HR)$).³⁴⁸
- A exponenciação do coeficiente fornece o HR .³⁴⁸
- Valores de $HR < 1$ indicam efeito protetor.³⁴⁸
- Valores de $HR > 1$ indicam aumento do risco.³⁴⁸

R O pacote *survival*³⁵⁰ fornece a função *survfit*^a para criar curvas de sobrevida.

^a<https://www.rdocumentation.org/packages/survival/versions/3.8-3/topics/survfit>

R O pacote *survminer*³⁵¹ fornece a função *ggsurvplot*^a para criar curvas de sobrevida.

^a<https://www.rdocumentation.org/packages/survminer/versions/0.4.2/topics/ggsurvplot>

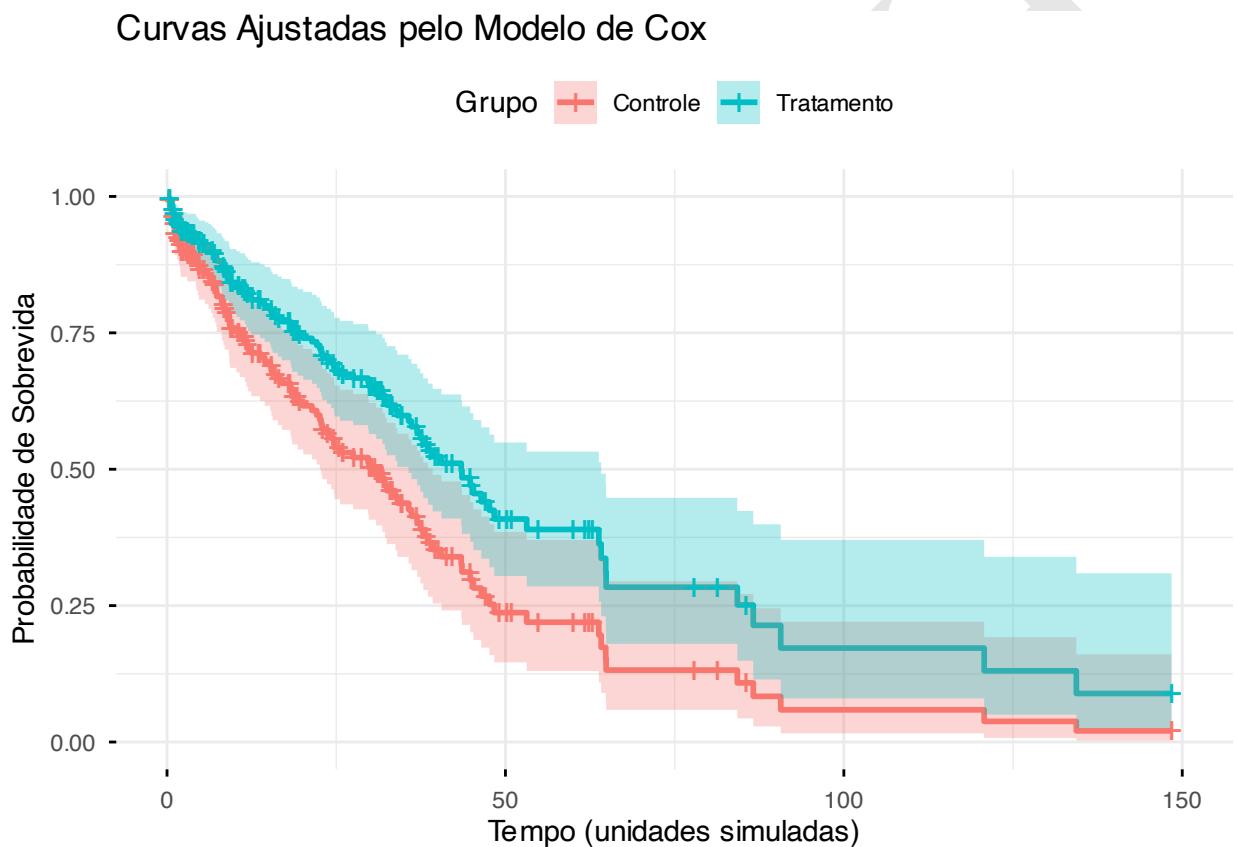


Figura 41.2: Curvas ajustadas pelo modelo de Cox para dois grupos (controle e tratamento).

RASCUNHO

PARTE 7: REPRESENTAÇÃO, APRENDIZADO DE MÁQUINA E INTELIGÊNCIA ARTIFICIAL

Do avanço estatístico ao poder computacional: Métodos modernos para problemas complexos

RASCUNHO

Capítulo 42

Representações

42.1 Representações de dados e extração de atributos

42.1.1 Por que modelos não podem acessar o fenômeno diretamente?

• ?

42.1.2 Como um fenômeno do mundo real é traduzido em uma estrutura matemática?

• ?

42.1.3 De que forma a representação limita ou expande o que um modelo pode aprender?

• ?

42.1.4 O que é um atributo e por que ela não é um algoritmo?

• ?

42.1.5 Quais são as principais formas de representar diferentes tipos de dados?

• ?

42.1.6 O que se perde — e o que se ganha — ao escolher uma representação?

• ?

42.1.7 Como essas escolhas antecipam os desafios da inteligência artificial e do aprendizado de máquina?

• ?

RASCUNHO

Capítulo 43

Inteligência artificial

43.1 Inteligência artificial

43.1.1 O que é inteligência artificial (IA)?

- ?

43.1.2 Como ela se relaciona com estatística, ciência de dados e aprendizado de máquina?

- ?

43.2 Inteligência artificial generativa

43.2.1 O que é IA generativa?

- ?

43.2.2 O que são grandes modelos de linguagem (*large language models, LLM*)?

- ?

43.2.3 Como funcionam os grandes modelos de linguagem?

- ?

43.3 Inteligência artificial explicável (*eXplainable Artificial Intelligence, XAI*)

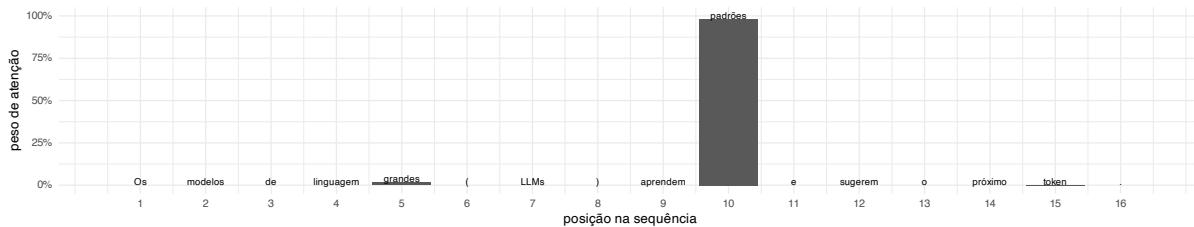
43.3.1 Quais princípios são utilizados para descrever explicabilidade de IA?

- Explicabilidade: capacidade de revelar e resumir, em termos comprehensíveis, os motivos por trás das decisões de um modelo, incluindo métodos post-hoc que explicam resultados após a predição.³⁵²
- Interpretabilidade: capacidade de compreender o funcionamento interno do modelo a partir de suas próprias estruturas, por meio de técnicas intrínsecas que tornam o processo decisório inteligível.³⁵²

A) Sequência de tokens (após tokenização)

Os	modelos	de	linguagem	grandes	(LLMs)	aprendem	padrões	e	sugarem	o	próximo	token	.
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

B) Atenção do último token: "."



C) Próximo token (top-10) – distribuição toy

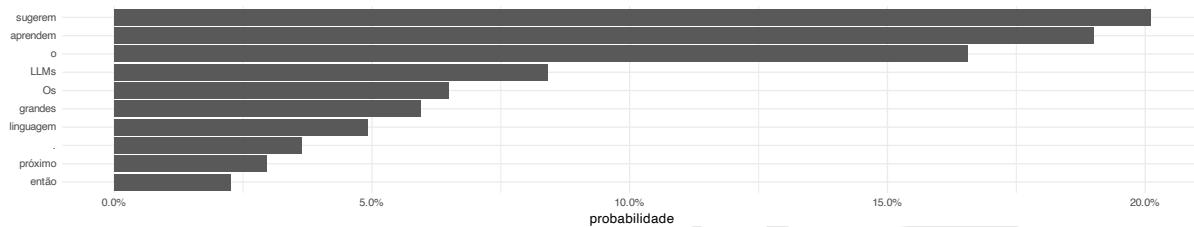


Figura 43.1: Representação esquemática de um modelo de linguagem grande (LLM)

- Transparência: geração de explicações claras e legíveis por humanos sobre as decisões do modelo, essencial para avaliar sua qualidade e resistir a usos adversariais.³⁵²
- Equidade: capacidade do modelo de tomar decisões imparciais, sem favorecer ou discriminar grupos, mitigando vieses associados a características sociais, econômicas ou demográficas.³⁵²
- Robustez: capacidade do modelo de manter desempenho estável diante de incertezas ou pequenas perturbações nos dados de entrada, incluindo ataques adversariais.³⁵²
- Satisfação: grau em que técnicas de explicabilidade aumentam a utilidade, a usabilidade e a aceitação do sistema baseado em aprendizado de máquina.³⁵²
- Estabilidade: capacidade de produzir explicações consistentes para entradas semelhantes, evitando variações arbitrárias nas interpretações.³⁵²
- Responsabilidade: alinhamento do modelo a valores sociais, éticos e morais, sustentado por transparência, responsabilização, equidade e ética, como base para uma IA responsável.³⁵²

43.3.2 Por que explicar modelos de IA?

- Para tornar as decisões defensáveis: compreender como e por que um modelo chega a determinado resultado é essencial para avaliar a legitimidade de suas decisões e sustentar seu uso em contextos críticos.^{353,354}
- Para tornar o sistema governável: explicações permitem enxergar o funcionamento interno do modelo, facilitando sua supervisão, auditoria, correção de erros e detecção de vieses ou fragilidades.^{353,354}
- Para aperfeiçoar o desempenho: ao revelar como o modelo utiliza informações, as explicações orientam ajustes que aumentam sua precisão, eficiência e robustez.^{353,354}
- Para ampliar o conhecimento: modelos explicáveis não apenas produzem previsões, mas também ajudam a identificar padrões, relações e hipóteses novas, contribuindo para a geração de conhecimento científico.^{353,354}



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>



O pacote *tensorflow@tensorflow*^a 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>



O pacote *torch@torch*^a permite criar e treinar redes neurais com alto desempenho.

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



O pacote *reticulate@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>

43.4 Limitações fundamentais de modelos generativos

43.4.1 O que são alucinações em IA generativa?

- Alucinações em IA generativa ocorrem quando um modelo produz saídas linguisticamente plausíveis, porém factualmente incorretas, inexistentes ou não verificáveis, sem que haja erro numérico ou falha no processo de treinamento.[?]

43.4.2 Por que modelos generativos alucinam?

- Modelos generativos são treinados para maximizar a probabilidade do próximo token dado um contexto, e não para verificar fatos ou manter correspondência com a realidade externa.[?]
- O critério de otimização privilegia a verossimilhança estatística da linguagem, não a veracidade epistemológica do conteúdo.[?]

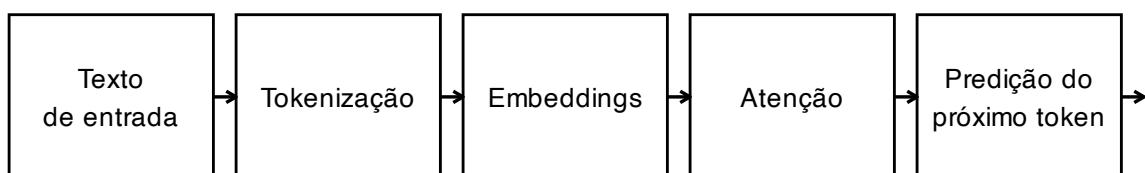
43.4.3 Alucinações indicam erro do modelo?

- Alucinações não indicam que o modelo “aprendeu errado”, mas que ele aprendeu exatamente o que foi solicitado: produzir linguagem plausível, não conhecimento verdadeiro.[?]
- Do ponto de vista matemático, uma resposta alucinatória pode ser ótima segundo a função de perda, mesmo sendo incorreta no mundo real.[?]

43.4.4 Qual a diferença entre erro estatístico e alucinação?

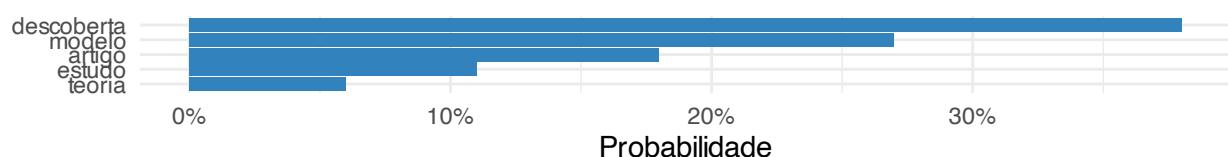
- Erros estatísticos decorrem de limitações de generalização, ruído nos dados ou inadequação do modelo ao problema.[?]
- Alucinações decorrem da ausênciade mecanismos internos de checagem factual e constituem um fenômeno semântico e comunicacional, não um erro numérico.[?]

A) Funcionamento conceitual de um LLM



B) Próximo token mais provável

Escolha baseada em plausibilidade linguística



C) Alucinação em modelos generativos

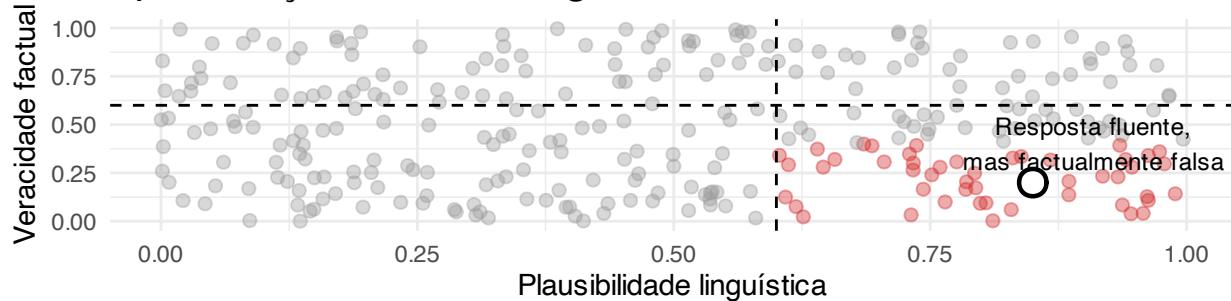


Figura 43.2: Funcionamento conceitual de modelos de linguagem e origem das alucinações. O painel A ilustra o fluxo abstrato de um LLM até a predição do próximo token; o painel B mostra a escolha baseada em plausibilidade linguística; o painel C evidencia que essa otimização probabilística pode gerar respostas fluentes, porém factualmente falsas.

43.4.5 Quais tipos de alucinação são mais comuns?

- Alucinação factual: afirmação de fatos inexistentes ou incorretos.⁷
- Alucinação de fonte: citação de autores, artigos ou documentos inexistentes.⁷
- Alucinação causal: inferência de relações de causa e efeito sem fundamento empírico.⁷

43.4.6 Por que alucinações são um problema prático?

- Porque a fluência linguística do modelo pode induzir confiança indevida em informações erradas, especialmente em contextos científicos, clínicos, jurídicos e educacionais.⁷

43.4.7 É possível eliminar completamente as alucinações?

- Até o momento, não. Alucinações são uma consequência estrutural do paradigma generativo probabilístico e podem apenas ser mitigadas, não eliminadas.⁷
- Estratégias como *prompting* controlado, ajuste de temperatura, recuperação de informações externas e validação pós-geração reduzem sua frequência, mas não suprimem o fenômeno.⁷

RASCUNHO

Capítulo 44

Aprendizado de máquina

44.1 Aprendizado de máquina

44.1.1 O que é aprendizado de máquina?

- Treinar um modelo significa resolver um problema matemático no qual um conjunto de observações (dados) é usado para ajustar um modelo. Esse modelo busca capturar tendências gerais dos dados, ignorando particularidades excessivas para evitar sobreajuste (*overfitting*).³⁵⁵
- O processo deriva do conceito estatístico de regressão e corresponde, em essência, à solução de um problema em que há mais restrições do que graus de liberdade.³⁵⁵



O pacote *fastml*³⁵⁶ fornece a função *train_models*^a para treinar algoritmos de aprendizado de máquina em dados de treinamento pré-processados.

^a<https://cran.r-project.org/web/packages/fastml/fastml.pdf>

44.2 Representação de dados e engenharia de atributos

44.2.1 Como representar texto como vetores?

- ?

44.2.2 O que é *one-hot*, *multi-hot* e *count encoding*?

- *One-hot encoding*: cria uma coluna binária para cada categoria de uma variável categórica, indicando a presença (1) ou ausência (0) da categoria em cada observação.[?]
- *Multi-hot encoding*: semelhante ao one-hot, mas permite que múltiplas categorias sejam representadas simultaneamente em uma única observação, útil para variáveis com múltiplas seleções.[?]
- *Count encoding*: substitui categorias por contagens de sua ocorrência no conjunto de dados, capturando a frequência relativa de cada categoria.[?]

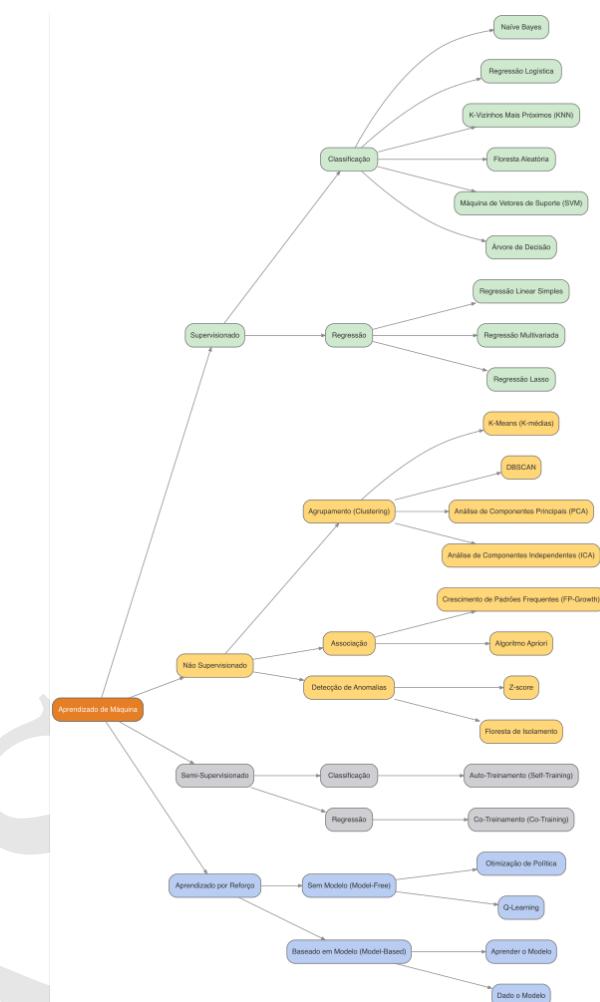


Figura 44.1: Mapa mental de algoritmos de aprendizado de máquina.

Tabela 44.1: Exemplo de codificação one-hot, multi-hot e count encoding para representação de texto.

Termo	avalia	cuidado	equipe	executa	final	planeja	projeto	revisa	tarefas
equipe	0	0	1	0	0	0	0	0	0
planeja	0	0	0	0	0	1	0	0	0
projeto	0	0	0	0	0	0	1	0	0
cuidado	0	1	0	0	0	0	0	0	0
equipe	0	0	1	0	0	0	0	0	0
executa	0	0	0	1	0	0	0	0	0
tarefas	0	0	0	0	0	0	0	0	1
projeto	0	0	0	0	0	0	1	0	0
revisa	0	0	0	0	0	0	0	1	0
tarefas	0	0	0	0	0	0	0	0	1
avalia	1	0	0	0	0	0	0	0	0
projeto	0	0	0	0	0	0	1	0	0
final	0	0	0	0	1	0	0	0	0
One-hot encoding	1								
Multi-hot encoding	1								
Count encoding	1	1	2	1	1	1	3	1	2

Frase original:

A equipe planeja o projeto com cuidado, a equipe executa as tarefas do projeto, revisa as tarefas e avalia o projeto final.

44.3 Métricas de distância e similaridade

44.3.1 O que é uma métrica?

- Uma métrica é uma função que quantifica a distância entre dois objetos em um espaço vetorial, obedecendo propriedades como não-negatividade, identidade, simetria e desigualdade triangular.⁷

44.3.2 Quais são as principais métricas?

- Distância Manhattan (44.1): soma das diferenças absolutas entre as coordenadas dos pontos.⁷

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (44.1)$$

- Distância Euclidiana (44.2): raiz quadrada da soma dos quadrados das diferenças entre as coordenadas dos pontos.⁷

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (44.2)$$

- Distância Minkowski: (44.3): generalização das distâncias de Manhattan e Euclidiana, onde o parâmetro p determina a ordem da métrica. Para $p = 1$, é equivalente à distância de Manhattan; para $p = 2$, é equivalente à distância Euclidiana.⁷

$$d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}} \quad (44.3)$$

- Distância Chebyshev (44.4): máximo das diferenças absolutas entre as coordenadas dos pontos, representando a distância em um espaço onde o movimento é permitido em todas as direções.⁷

$$d(x, y) = \max_i |x_i - y_i| \quad (44.4)$$

- Similaridade cosseno (44.5): medida de similaridade entre dois vetores que calcula o cosseno do ângulo entre eles, variando de -1 (totalmente opostos) a 1 (totalmente semelhantes).⁷

$$\text{similaridade}(x, y) = \frac{x \cdot y}{\|x\| \|y\|} \quad (44.5)$$

- Distância de Hamming (44.6): número de posições em que os símbolos correspondentes de duas sequências de igual comprimento são diferentes, frequentemente usada para medir a diferença entre strings ou códigos binários.⁷

$$d(x, y) = \sum_{i=1}^n \mathbf{1}(x_i \neq y_i) \quad (44.6)$$

- Índice de Jaccard (44.7): medida de similaridade entre conjuntos, calculada como a razão entre a interseção e a união dos conjuntos, variando de 0 (sem elementos em comum) a 1 (conjuntos idênticos).⁷

$$\text{similaridade}(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (44.7)$$

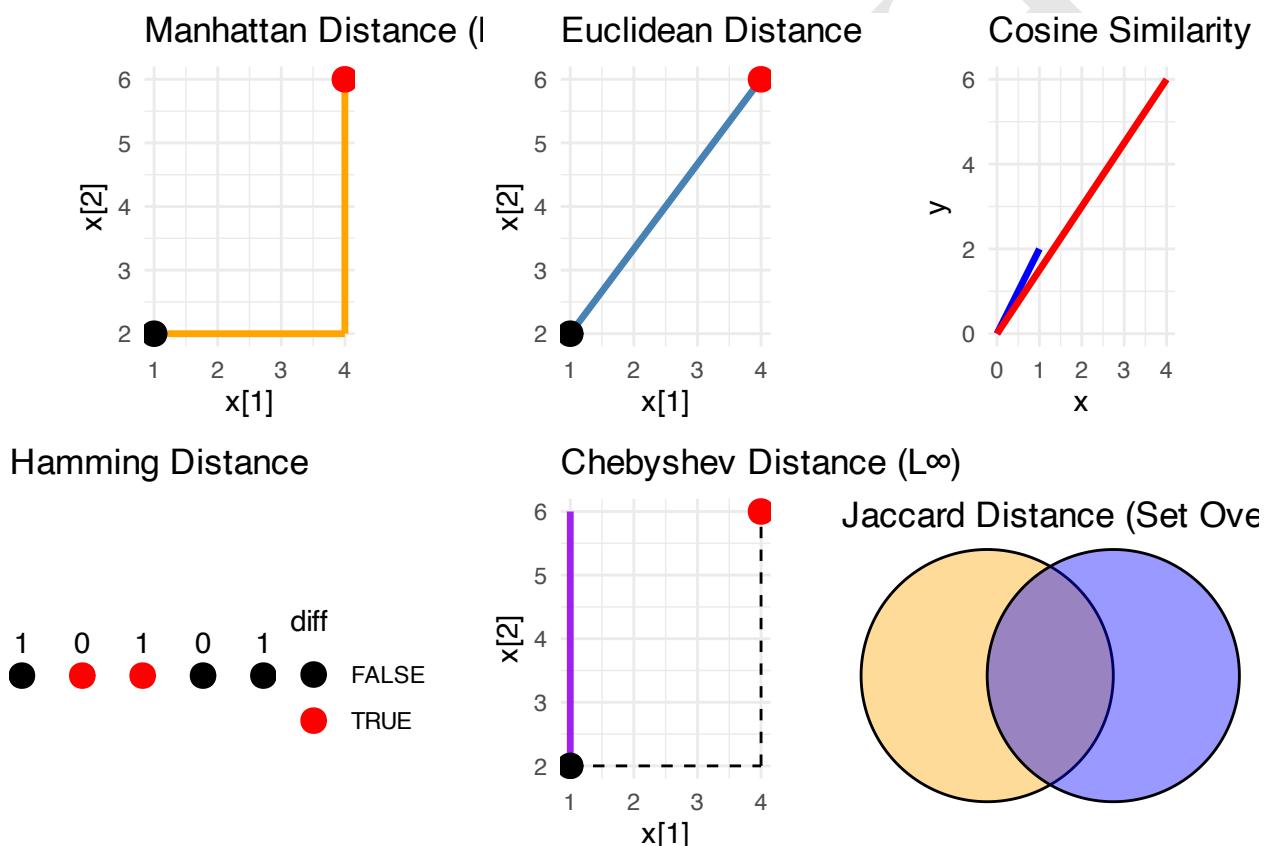


Figura 44.2: Visualização de diferentes métricas de distância e similaridade.

44.3.3 Como escolher a métrica adequada?

- A escolha da métrica depende do tipo de dados (contínuos, categóricos, binários), da escala das variáveis, da presença de outliers e do objetivo da análise (agrupamento, classificação, etc.).?

44.3.4 Quais métricas são indicadas para avaliar modelos preditivos?

- A área sob a curva ROC (AUROC) como medida de discriminação, um gráfico de calibração para avaliar a concordância entre probabilidades estimadas e proporções observadas, uma medida de utilidade clínica como o *net benefit* por meio de *decision curve analysis*, e um gráfico que mostre a distribuição das probabilidades preditas por categoria de desfecho.³⁵⁷

44.3.5 Quais métricas são consideradas inadequadas ou problemáticas?

- O F1-score ($F1$) é explicitamente desencorajado como métrica principal porque não tem um foco claro em desempenho estatístico e é uma medida inadequada dentro de sua própria categoria conceitual.³⁵⁷
- Todas as métricas puramente baseadas em classificação dependente de limiar são consideradas impróprias para limiares clinicamente relevantes, exceto quando o limiar é 0,5 ou igual à prevalência verdadeira.³⁵⁷
- Métricas derivadas da matriz de confusão, como acurácia, sensibilidade, especificidade e medidas relacionadas, não devem ser usadas isoladamente como critério principal de avaliação, pois dependem fortemente da escolha do limiar e não refletem diretamente a utilidade clínica.³⁵⁷

44.4 Desbalanceamento de classes

44.4.1 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.?

44.4.2 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.?

44.4.3 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.?

44.4.4 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.³⁵⁸
- Em simulações computacionais, modelos sem correção tiveram calibração igual ou superior aos corrigidos.³⁵⁸

- A piora observada foi caracterizada por superestimação do risco, nem sempre reversível com recalibração.³⁵⁸

44.5 Tipos de aprendizado

44.5.1 O que é aprendizado supervisionado?

- ?

44.5.2 O que é aprendizado não supervisionado?

- ?

44.5.3 O que é aprendizado por reforço?

- ?

44.5.4 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.³⁵⁹
- O aumento da complexidade do modelo traz retornos decrescentes em termos de redução da taxa de erro.³⁵⁹

44.5.5 Quais problemas práticos limitam a generalização de classificadores?

- *Population drift*: mudanças na distribuição dos dados ao longo do tempo degradam a performance de modelos.³⁵⁹
- *Sample selectivity bias*: amostras de treino podem não representar a população futura, levando a superestimação de desempenho.³⁵⁹
- Erros de rótulo e definições arbitrárias de classes comprometem a validade dos modelos.³⁵⁹

44.5.6 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.³⁵⁹
- Diferenças pequenas em acurácia frequentemente desaparecem quando se consideram incertezas reais de aplicação.³⁵⁹

44.6 Avaliação de modelos de classificação

44.6.1 O que é uma matriz de confusão 2x2?

- A matriz de confusão é uma tabela de contingência 2×2 utilizada para avaliar o desempenho de um classificador binário, comparando as classes verdadeiras observadas com as classes previstas pelo modelo.³⁶⁰
- Em aprendizado supervisionado, ela permite decompor os erros de classificação e analisar como o modelo se comporta em relação a cada classe.³⁶⁰

Tabela 44.2: Tabela de confusão 2x2 para análise de desempenho diagnóstico de testes e variáveis dicotômicas.

	Classe predita correta	Classe predita incorreta	Total
Classe verdadeira correta	VP	FP	$VP + FP$
Classe verdadeira incorreta	FN	VN	$FN + VN$
Total	$VP + FN$	$FP + VN$	$N =$ $VP + VN + FP + FN$

44.6.2 Como interpretar uma matriz de confusão 2x2?

- Verdadeiro-positivo (VP): instância cuja classe verdadeira é positiva e foi corretamente classificada como positiva.³⁶⁰
- Falso-negativo (FN): instância positiva que foi incorretamente classificada como negativa.³⁶⁰
- Verdadeiro-negativo (VN): instância negativa corretamente classificada como negativa.³⁶⁰
- Falso-positivo (FP): instância negativa incorretamente classificada como positiva.³⁶⁰

44.6.3 Quais métricas caracterizam o desempenho de um classificador?

- Acurácia (ACU) (54.7): Proporção de classificações corretas (verdadeiros-positivos e verdadeiros-negativos) em relação ao total de casos.³⁶⁰

$$ACU = \frac{VP + VN}{VP + FN + VN + FP} \quad (44.8)$$

- Precisão (PRE) (44.9): Proporção de verdadeiro-positivos dentre aqueles classificados como positivos.³⁶⁰

$$PRE = \frac{VP}{VP + FP} \quad (44.9)$$

- Revocação (REV) (44.10): Proporção de verdadeiro-positivos dentre aqueles com a condição, equivalente à sensibilidade.³⁶⁰

$$REV = \frac{VP}{VP + FN} \quad (44.10)$$

- F1-score ($F1$) (44.11): Média harmônica entre precisão e revocação, balanceando ambos os aspectos do desempenho.³⁶⁰

$$F1 = 2 \times \frac{PRE \times REC}{PRE + REC} \quad (44.11)$$

44.7 Principais algoritmos

44.7.1 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 ensembles do tipo Superlearner.³⁶¹
- 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

44.8 Regressão logística

44.8.1 O que é regressão logística?

- Modelos logísticos são casos de regressão linear generalizada em que a resposta Y é binária.³²³
- A equação (44.12) modela a razão de chances (*odds*) em função dos preditores.³²³

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n \quad (44.12)$$

- A ligação (*link*) usada é o logit (44.13).³²³

$$g(p) = \log\left(\frac{p}{1-p}\right) \quad (44.13)$$

44.9 Máquina de vetores de suporte

44.9.1 O que são máquinas de vetores de suporte?

- ?

44.10 K-nearest neighbours

44.10.1 O que é K-nearest neighbours?

- ?

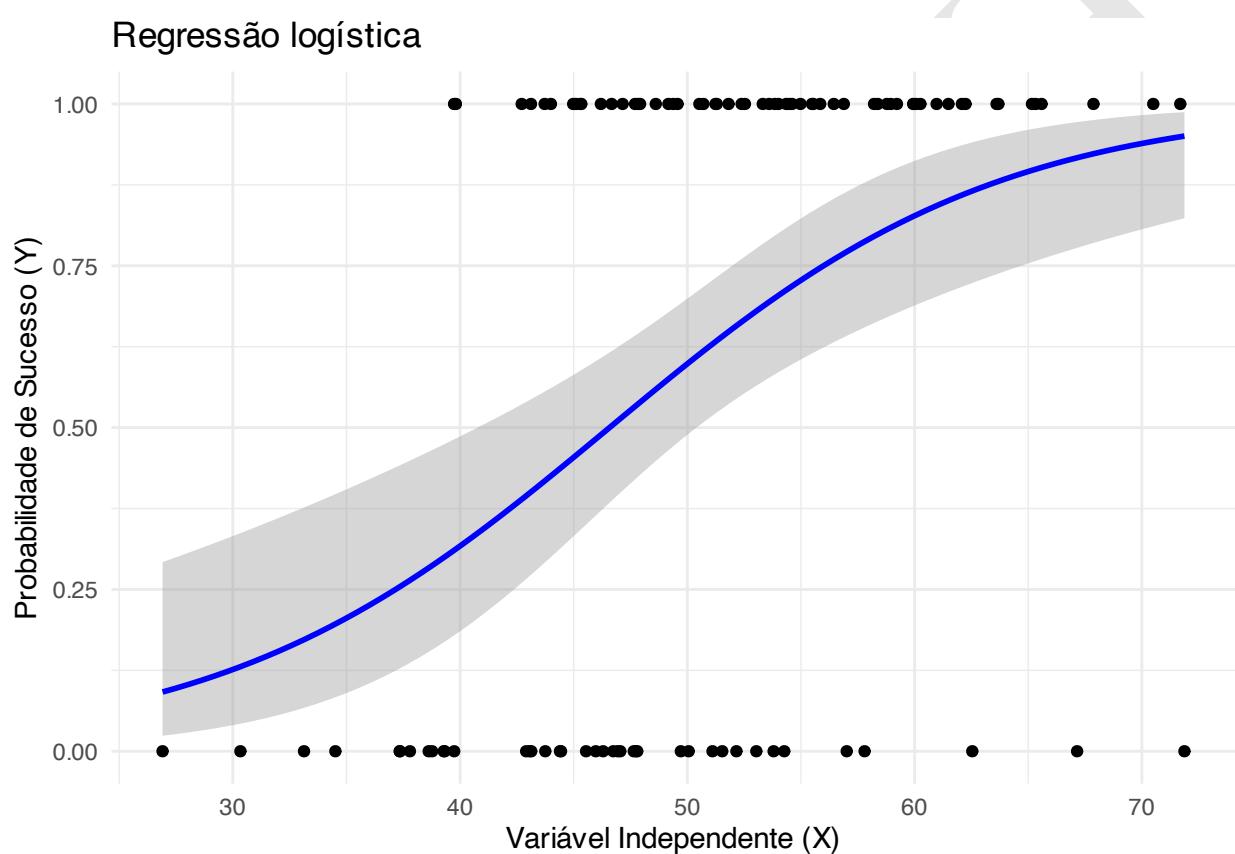


Figura 44.3: Regressão logística.

44.11 *K-means Clustering*

44.11.1 O que é *K-means clustering*?

- ?

44.12 Árvores de decisão

44.12.1 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.²⁴⁷
- Podem lidar eficientemente com grandes conjuntos de dados sem pressupor estrutura paramétrica complexa.²⁴⁸
- São aplicáveis a variáveis contínuas e discretas, tanto como preditoras quanto como desfechos.²⁴⁹

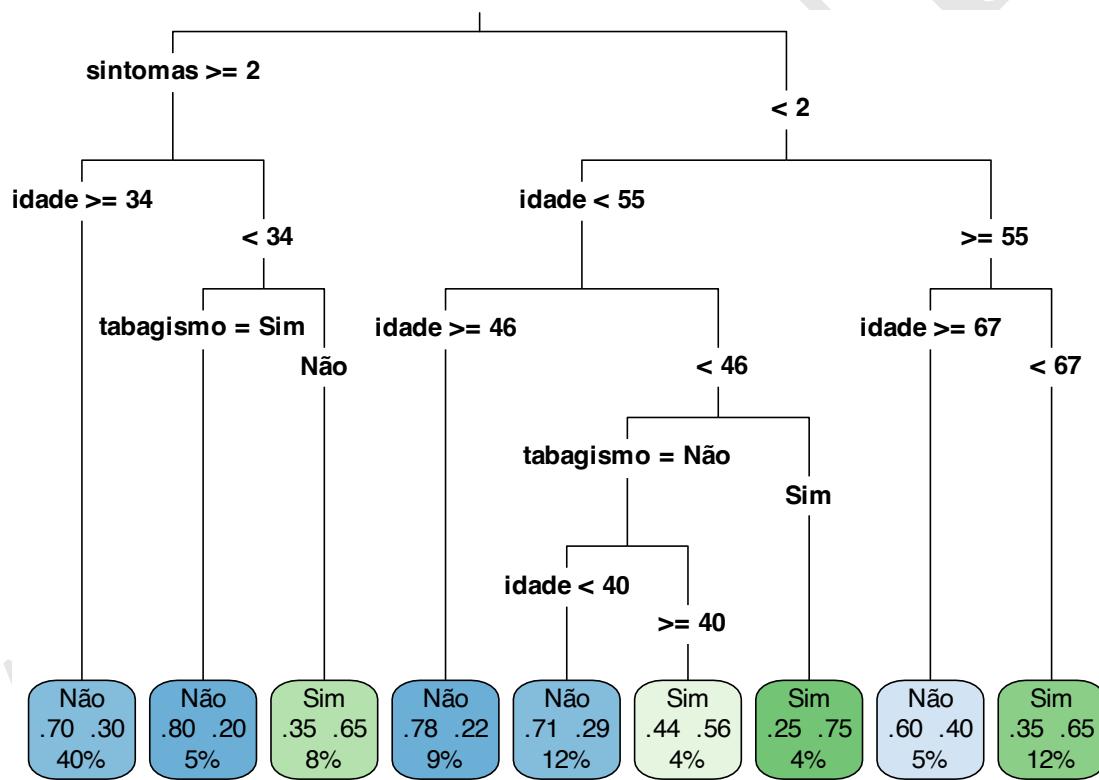


Figura 44.4: Exemplo de árvore de decisão para predizer depressão a partir de idade, tabagismo e sintomas.

44.12.2 Quais são os principais usos de árvores de decisão?

- Seleção de variáveis relevantes em cenários com muitos preditores, como registros clínicos eletrônicos.²⁴⁶
- Avaliação da importância relativa das variáveis, com base na redução da pureza dos nós ou da acurácia ao remover variáveis.²⁴⁶
- Tratamento de valores ausentes, seja classificando-os como categoria própria ou imputando-os por previsão dentro da árvore.²⁴⁶

- Predição de novos casos a partir de dados históricos.²⁴⁶
- Manipulação de dados, colapsando categorias muito numerosas ou subdividindo variáveis contínuas assimétricas.²⁴⁶

44.12.3 Quais são os componentes básicos de uma árvore de decisão?

- Nós raiz (ou de decisão): subdividem todos os registros iniciais.²⁴⁶
- Nós internos (ou de chance): representam subdivisões intermediárias.²⁴⁶
- Nós folha (ou finais): resultados finais após sucessivas divisões.²⁴⁶
- Ramos: representam condições “se-então”, ligando nós em sequência até a classificação final.²⁴⁶

44.12.4 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.²⁴⁶
- *Stopping*: evita árvores excessivamente complexas ao definir parâmetros como número mínimo de registros por nó ou profundidade máxima.²⁴⁶
- *Pruning*: reduz árvores grandes eliminando ramos pouco informativos, usando validação ou métodos como qui-quadrado.²⁴⁶

44.12.5 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*.²⁴⁶
- Limitações: podem sofrer *overfitting* ou *underfitting* em amostras pequenas; podem selecionar variáveis correlacionadas sem relação causal real.²⁴⁶

44.12.6 Espaço de decisão em árvores de decisão vs. regressão logística

- A regressão logística assume relações lineares entre variáveis e log-odds.²⁴⁷
- Árvores de decisão permitem capturar relações não lineares e interações de forma automática.²⁴⁷

R

O pacote *h2o*³⁶² fornece funções construir modelos de aprendizado de máquina.

R

O pacote *correctR*³⁴⁴ 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>

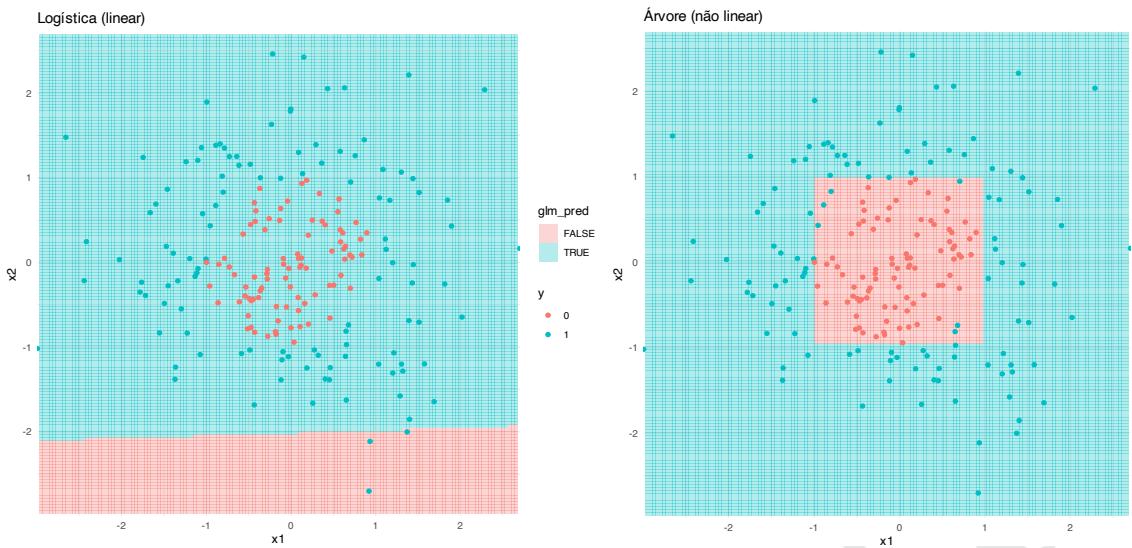


Figura 44.5: Comparação entre modelos de regressão logística e árvore de decisão.



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>

44.13 Análise de componentes principais

44.13.1 O que é análise de componentes principais?

- A análise de componentes principais (*Principal Component Analysis*, PCA) é uma técnica estatística amplamente utilizada para redução de dimensionalidade, para representar dados de alta dimensão por um conjunto menor de variáveis, preservando o máximo possível da variabilidade original.³⁶³
- O primeiro componente principal é definido como a direção, de comprimento unitário, que maximiza a variância dos dados projetados. Ele corresponde ao eixo ao longo do qual os dados apresentam a maior dispersão, concentrando a maior quantidade de informação estatística disponível.³⁶³
- O segundo componente principal é a direção que maximiza a variância restante, sob a condição de ser ortogonal ao primeiro componente. Essa restrição garante que cada novo componente adicione informação nova, não redundante, à representação dos dados.³⁶³
- Esse procedimento é repetido para os componentes subsequentes, de forma que cada componente principal seja ortogonal aos anteriores e capture a maior variância possível ainda não explicada, resultando em uma ordenação natural dos componentes por importância.³⁶³
- Ao final do processo, a PCA produz uma base ortogonal que impõe uma geometria específica à representação dos dados, restringindo a forma como os dados podem ser reconstruídos a partir dos componentes principais.³⁶³
- Embora os componentes principais descrevam de maneira eficiente a variabilidade dos dados, eles nem sempre correspondem aos fatores geradores subjacentes do fenômeno estudado. A PCA pode ainda introduzir padrões artificiais, criando uma aparência de estrutura que não reflete necessariamente os processos reais de geração dos dados.³⁶³

PCA: máxima variância \neq fator gerador

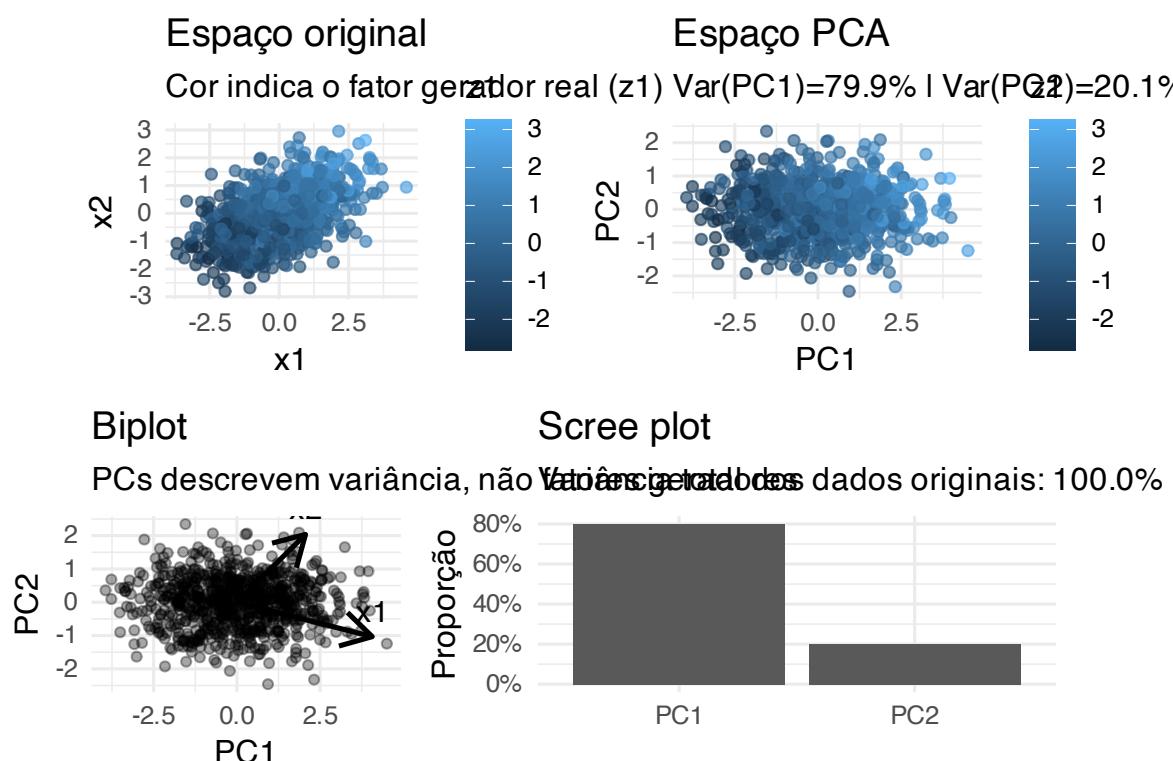


Figura 44.6: Análise de Componentes Principais (PCA). O PC1 maximiza variância total, mas pode não alinhar com o fator latente real (z_1).



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>

44.14 *Random forests*

44.14.1 O que são *random forests*?

- ?

44.15 *Ensemble*

44.15.1 O que são *ensemble*?

- ?

RASCUNHO

Capítulo 45

Redes neurais

45.1 Neurônios artificiais

45.1.1 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.³⁶⁴⁻³⁶⁶
- A equação geral de um neurônio artificial é dada por (45.1), onde x_i são as entradas, w_i os pesos, b o viés e ϕ a função de ativação:

$$y = \phi \left(\sum_{i=1}^d w_i x_i + b \right) \quad (45.1)$$

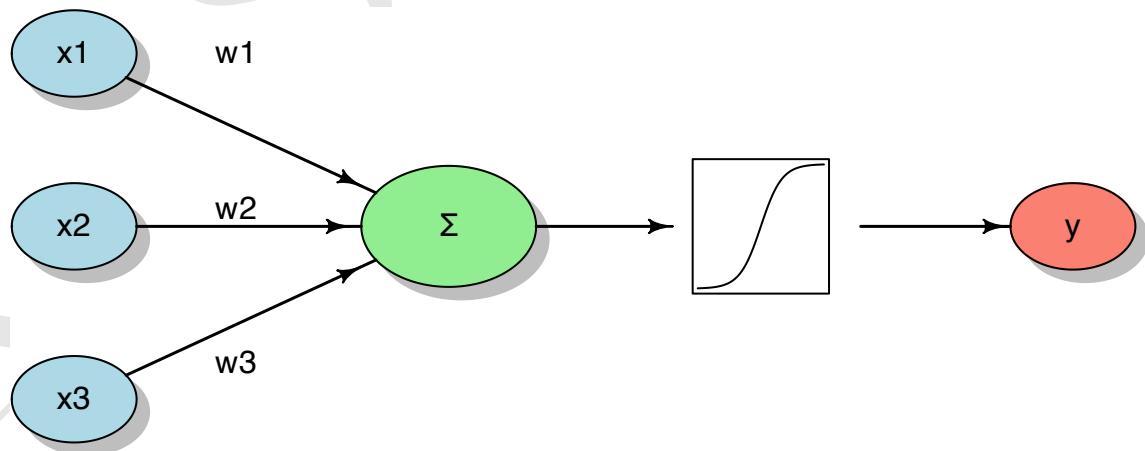


Figura 45.1: Representação esquemática de um neurônio computacional.

45.2 Rede neural artificial

45.2.1 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

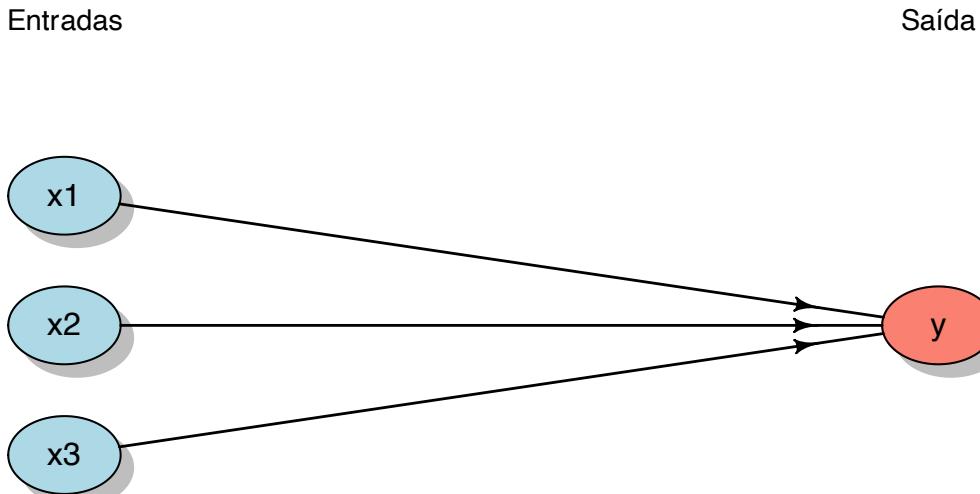


Figura 45.2: Representação esquemática de uma rede neural simples com camada de entrada e saída.

R O pacote *neuralnet*³⁶⁷ fornece a função *neuralnet*^a para treinar redes neurais artificiais.

^a<https://www.rdocumentation.org/packages/neuralnet/versions/1.44.2/topics/neuralnet>

45.3 Funções de ativação

45.3.1 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 (45.2), tangente hiperbólica (45.3) e unidade linear retificada (ReLU) (45.4).⁷

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (45.2)$$

$$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (45.3)$$

$$\text{ReLU}(z) = \max(0, z) \quad (45.4)$$

- 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

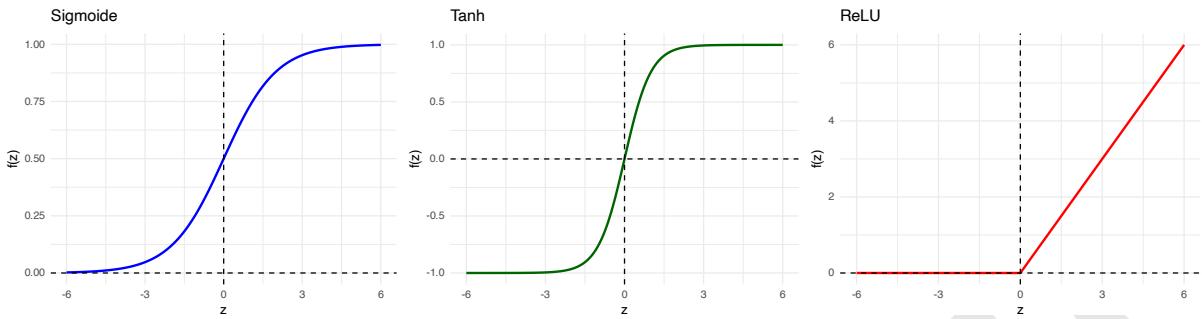


Figura 45.3: Gráficos das funções de ativação mais comuns.

45.4 Funções de perda

45.4.1 O que são funções de perda?

- Funções de perda (*loss functions*) quantificam o erro cometido por um modelo ao comparar suas previsões com os valores reais observados.⁷
- Funções de perda definem formalmente o objetivo do aprendizado, indicando o que significa “errar pouco” ou “errar muito” em um problema específico.⁸
- Durante o treinamento de modelos supervisionados, a função de perda orienta o ajuste dos parâmetros ao medir a discrepância entre saída prevista e desfecho verdadeiro.⁹
- Em redes neurais, a minimização da função de perda é realizada por métodos iterativos baseados em gradientes, como a retropropagação do erro.¹⁰
- A escolha da função de perda está intimamente ligada à natureza do problema (regressão, classificação, probabilidade, ranking) e influencia diretamente o espaço de decisão aprendido pelo modelo.¹¹

45.4.2 Quais são as funções de perda mais comuns?

- Erro quadrático médio (Mean Squared Error, MSE (45.5)): Essa função penaliza erros grandes de forma mais severa, sendo adequada quando desvios elevados são indesejáveis e a média do erro quadrático é uma medida relevante de desempenho.¹²

$$\mathcal{L}_{\text{MSE}}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (45.5)$$

- Erro absoluto médio (Mean Absolute Error, MAE(45.6)): Essa função atribui peso linear aos erros, tornando-se mais robusta a valores extremos quando comparada ao erro quadrático médio.¹³

$$\mathcal{L}_{\text{MAE}}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (45.6)$$

- Erro quadrático médio logarítmico (Mean Squared Logarithmic Error, MSLE (45.7)): Essa função enfatiza erros relativos, sendo particularmente útil quando diferenças proporcionais são mais relevantes do que diferenças absolutas.¹⁴

$$\mathcal{L}_{\text{MSLE}}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (\log(1 + y_i) - \log(1 + \hat{y}_i))^2 \quad (45.7)$$

- Entropia cruzada binária (Binary Cross-Entropy, BCE (45.8)): Essa função mede a discrepância entre probabilidades previstas e observadas, sendo o critério padrão em problemas de classificação binária probabilística.

$$\mathcal{L}_{\text{BCE}}(y, \hat{y}) = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (45.8)$$

- Entropia cruzada categórica (Categorical Cross-Entropy (45.9)): Essa função generaliza a entropia cruzada para múltiplas classes, penalizando previsões probabilísticas inconsistentes com a classe verdadeira.⁷

$$\mathcal{L}_{\text{CCE}}(y, \hat{y}) = -\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K y_{ik} \log(\hat{y}_{ik}) \quad (45.9)$$

- Função logística (log-verossimilhança negativa (45.10)): Essa função expressa o critério de máxima verossimilhança da regressão logística, conectando inferência estatística e aprendizado supervisionado.⁸

$$\mathcal{L}_{\text{log}}(y, \hat{p}) = -\sum_{i=1}^n [y_i \log(\hat{p}_i) + (1 - y_i) \log(1 - \hat{p}_i)] \quad (45.10)$$

- Função hinge (*Support Vector Machines* (45.11)): Essa função busca maximizar a margem entre classes, penalizando classificações incorretas ou pouco confiantes em relação à fronteira de decisão.

$$\mathcal{L}_{\text{hinge}}(y, f(x)) = \frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i f(x_i)) \quad (45.11)$$

45.5 Treinamento de redes neurais

45.5.1 O que significa treinar uma rede neural?

- ?

45.5.2 O que são os pesos em uma rede neural e por que eles são tratados como parâmetros do modelo?

- ?

45.5.3 Como ocorre o ajuste iterativo dos pesos ao longo do treinamento?

- ?

45.5.4 Qual é o objetivo formal do treinamento em termos de minimização da função de perda?

- ?

45.5.5 O que é o gradiente descendente (*gradient descent*)?

- ?

45.5.6 Qual é a interpretação geométrica do gradiente descendente no espaço dos parâmetros?

• ?

45.5.7 Por que o gradiente indica a direção de maior redução da função de perda?

• ?

45.5.8 O que é o gradiente descendente estocástico (*stochastic gradient descent*, SGD)?

• ?

45.5.9 Por que o uso de todo o conjunto de dados em cada atualização pode ser computacionalmente inviável?

• ?

45.5.10 O que são *mini-batches* e como eles são utilizados no treinamento?

• ?

45.5.11 De que forma o ruído introduzido pelo SGD pode atuar como um regularizador implícito?

• ?

45.5.12 O que é uma época (*epoch*) no treinamento de redes neurais?

• ?

45.5.13 O que define formalmente uma época durante o processo de treinamento?

• ?

45.5.14 Qual é a relação entre época, tamanho do batch e número de atualizações dos pesos?

• ?

45.5.15 Como o número de épocas influencia a convergência do modelo e o risco de *overfitting*?

• ?

45.6 Espaço de decisão

45.6.1 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

45.6.2 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.⁷

Perceptron: fronteira de decisão linear

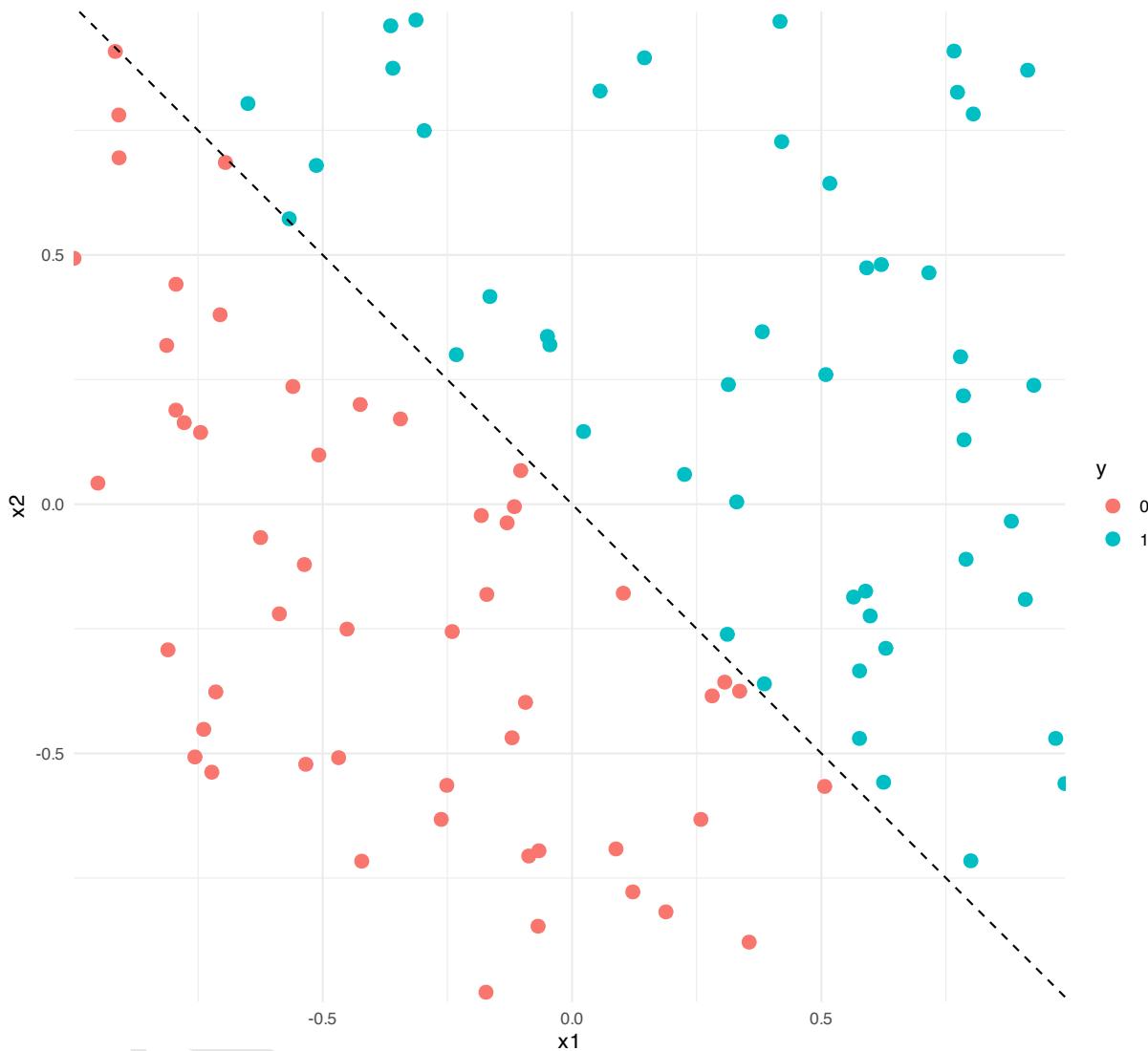


Figura 45.4: Espaço de decisão de um perceptron (regressão logística).

45.7 Redes neurais multicamadas

45.7.1 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.⁷
- Essa estrutura amplia o espaço de decisão do modelo sem exigir especificação explícita de interações ou transformações entre variáveis.⁷

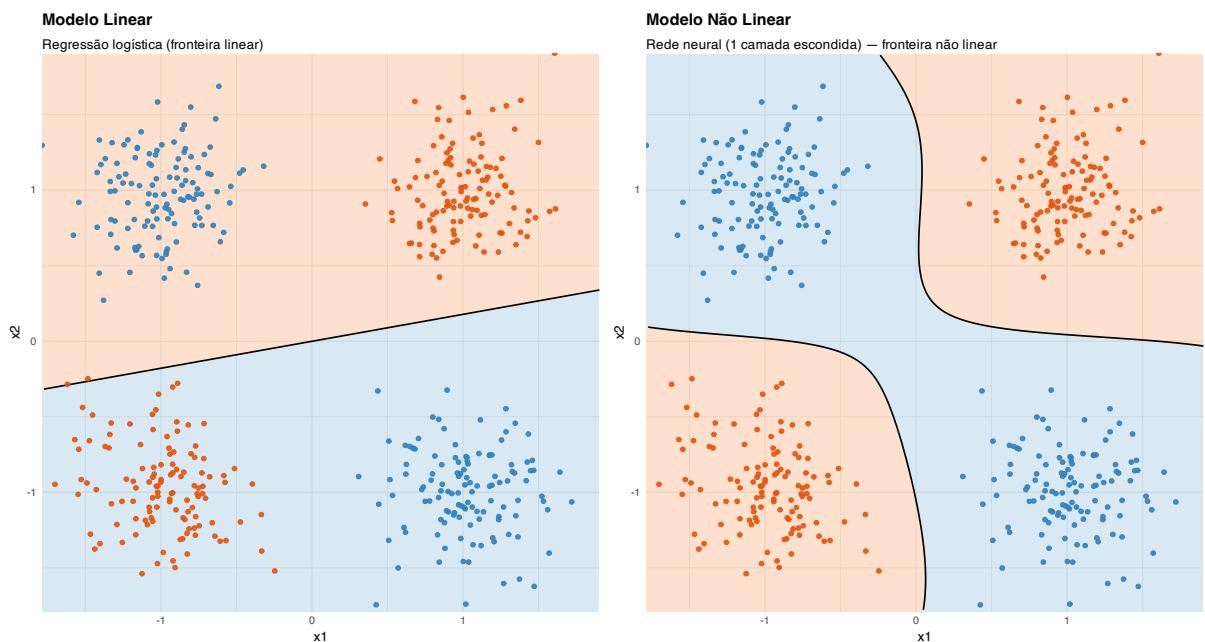


Figura 45.5: Comparação do espaço de decisão entre um modelo linear (regressão logística) e um modelo não linear (MLP).

- À medida que a profundidade da rede aumenta, passa-se do ajuste de parâmetros para o aprendizado de representações internas dos dados, caracterizando um regime distinto de modelagem conhecido como aprendizado profundo.⁷

45.8 Redes neurais profundas

45.8.1 O que são redes neurais profundas?

- Redes neurais profundas (*Deep Learning*, DL) são redes neurais multicamadas com várias camadas escondidas, permitindo a modelagem de relações altamente complexas e abstratas nos dados.⁷

45.9 Redes neurais convolucionais (Convolutional Neural Networks)

45.9.1 O que são redes neurais convolucionais?

- Redes neurais convolucionais (*Convolutional Neural Networks*, CNNs) são arquiteturas de redes neurais especialmente projetadas para processar dados com estrutura de grade, como imagens, utilizando operações de convolução para extraírem características locais e hierárquicas.⁷

45.9.2 O que é uma convolução?

- A convolução é uma operação matemática que combina duas funções para produzir uma terceira função, representando a sobreposição de uma função (o filtro ou kernel) sobre outra (a entrada), permitindo a extração de características locais em dados estruturados espacialmente.⁷

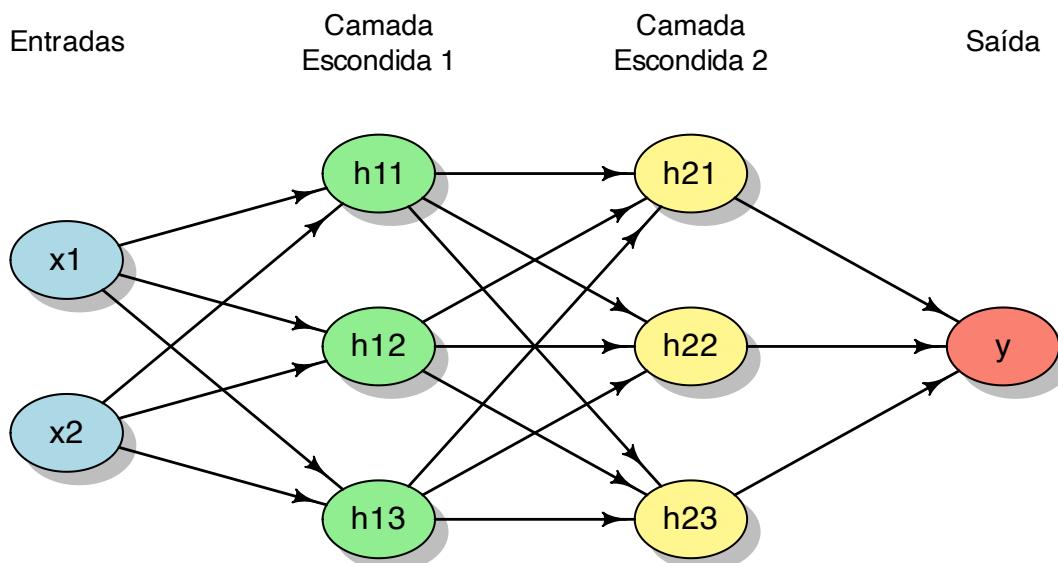


Figura 45.6: Representação esquemática de uma rede neural multicamadas com 2 camadas escondidas além das camadas de entrada e saída.

45.9.3 O que é um filtro convolucional?

- Um filtro convolucional (ou kernel) é uma pequena matriz de pesos utilizada na operação de convolução para detectar padrões específicos em dados de entrada, como bordas ou texturas em imagens, ao ser aplicada localmente sobre a entrada.⁷

45.9.4 Por que convoluções reduzem o número de parâmetros?

- Convoluções reduzem o número de parâmetros ao compartilhar pesos através de diferentes regiões da entrada, permitindo que o mesmo filtro seja aplicado em múltiplas posições, o que diminui a complexidade do modelo e melhora a eficiência computacional.⁷

PARTE 8: PLANEJAMENTO DE ESTUDOS

Definindo poder, tamanho amostral e plano de análise

RASCUNHO

Capítulo 46

Poder estatístico

46.1 Poder do teste

46.1.1 O que é poder do teste?

- Poder do teste é a probabilidade de rejeitar corretamente a hipótese nula (H_0) quando esta é falsa.²⁶⁹
- Poder do teste pode ser calculado como $1 - \beta$.²⁶⁹

46.1.2 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.³⁶⁸
- 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).³⁶⁸
- Numericamente, o poder de um estudo é calculado como $1 - \beta$ e reportado em valor percentual.³⁶⁸

46.1.3 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.³⁶⁸
- 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.³⁶⁸



O pacote *pwr*²⁸⁸ 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*²⁸⁸ 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*²⁸⁸ 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*²⁸⁸ 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*²⁸⁸ 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*²⁸⁸ 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*²⁸⁸ 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*²⁸⁸ 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*²⁸⁸ 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平衡ados).

^a<https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.t.test>

R

O pacote *pwr*²⁸⁸ 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*³⁶⁹ 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*²⁹⁷ fornece a função *power:ftest*^a para calcular o poder do teste por análise de testes F.

^a<https://www.rdocumentation.org/packages/Superpower/versions/0.2.0/topics/power.ftest>



O pacote *Superpower*²⁹⁷ 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*²⁹⁷ 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*²⁹⁷ 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*²⁹⁷ 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*²⁹⁷ 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*³⁷⁰ 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

46.1.4 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.^{211,368}

46.1.5 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.³⁶⁸

Capítulo 47

Tamanho da amostra

47.1 Tamanho da amostra

47.1.1 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.³⁷¹
- 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
- Em estudos de métodos mistos, o tamanho amostral não é único nem determinado por um único critério. Componentes quantitativos seguem lógica inferencial (poder estatístico), enquanto componentes qualitativos seguem lógica informacional (saturação, variação relevante).²³

47.1.2 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.¹³⁸

47.1.3 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.³⁷¹
- Delineamento do estudo.³⁷¹
- Quantidade e características (dependente vs. independente) dos grupos de participantes do estudo.³⁷¹
- Erros tipo I (α) e tipo II (β).³⁷¹
- Tipo de variável a ser observada (contínua, intervalo, ordinal, nominal, dicotômica).³⁷¹
- Tamanho de efeito mínimo a ser observado.³⁷¹
- Variabilidade da(s) variável(eis) coletada(s).³⁷¹

- Lateralidade do teste de hipótese (uni- ou bicaudais).³⁷¹
- Perdas de dados durante a coleta e/ou acompanhamento dos participantes do estudo.³⁷¹

R

O pacote *pwr*²⁸⁸ 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>

47.1.4 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.³⁷¹
- 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%.³⁷²
- Estudos com poder <80% não são necessariamente antiéticos.³⁷²
- 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.³⁷²
- Grandes estudos podem ser desejáveis por outras razões que não as éticas.³⁷²

47.2 Saturação em pesquisas qualitativas

47.2.1 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.³⁷³
- 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.³⁷⁴

47.2.2 Quais tipos de saturação existem?

- Saturação de códigos: ocorre quando não emergem novos códigos relevantes nos dados³⁷⁴
- Saturação de significados: atinge-se quando a profundidade e a variação dos significados de um tema foram plenamente exploradas.³⁷⁴
- Saturação teórica: quando categorias estão suficientemente desenvolvidas e suas relações esclarecidas.³⁷³
- Saturação de metatemas: em pesquisas multicêntricas, quando os grandes temas transversais já foram identificados.³⁷⁵

47.2.3 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.³⁷⁴
- Para saturação de significados, podem ser necessárias entre 16 e 24 entrevistas.³⁷⁴
- Em grupos focais, a saturação temática pode ocorrer com 4 a 8 grupos homogêneos.³⁷⁴

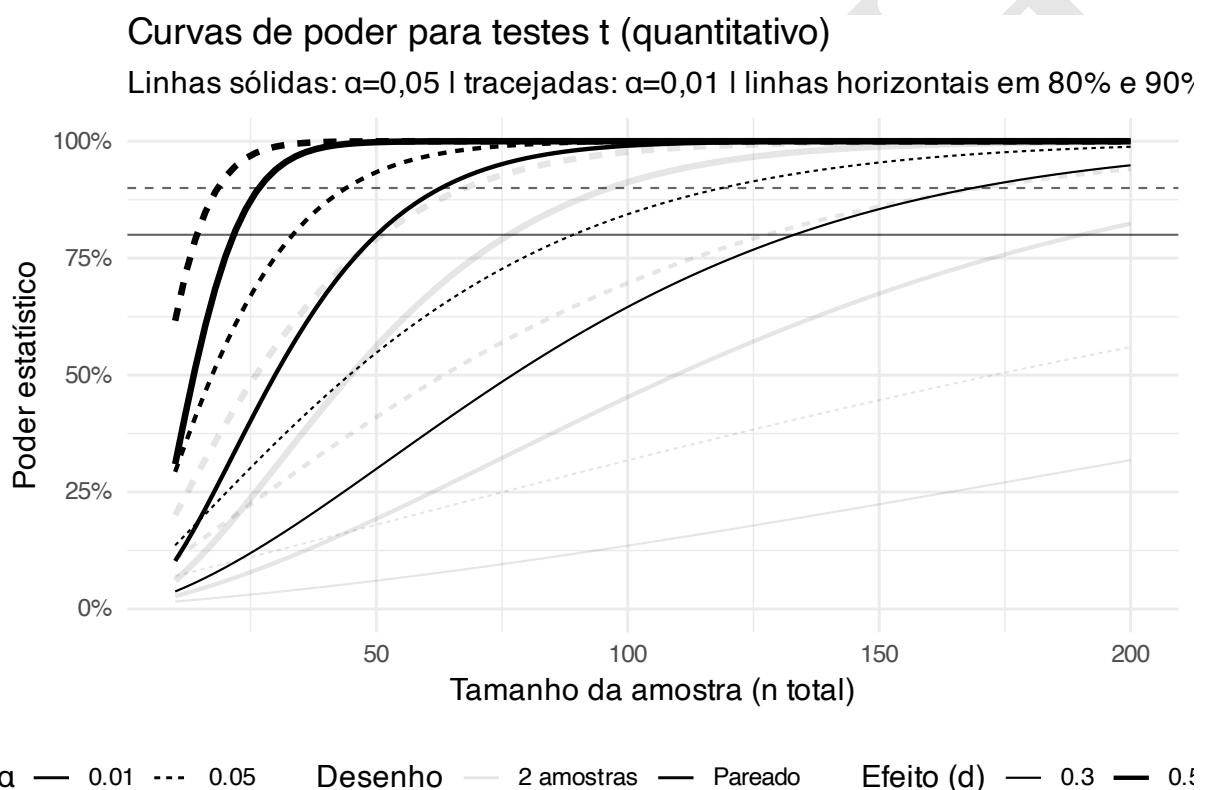


Figura 47.1: Curvas de poder para testes t (quantitativo). Linhas sólidas: $\alpha = 0,05$ | tracejadas: $\alpha = 0,01$ | linhas horizontais em 80% e 90% de poder.

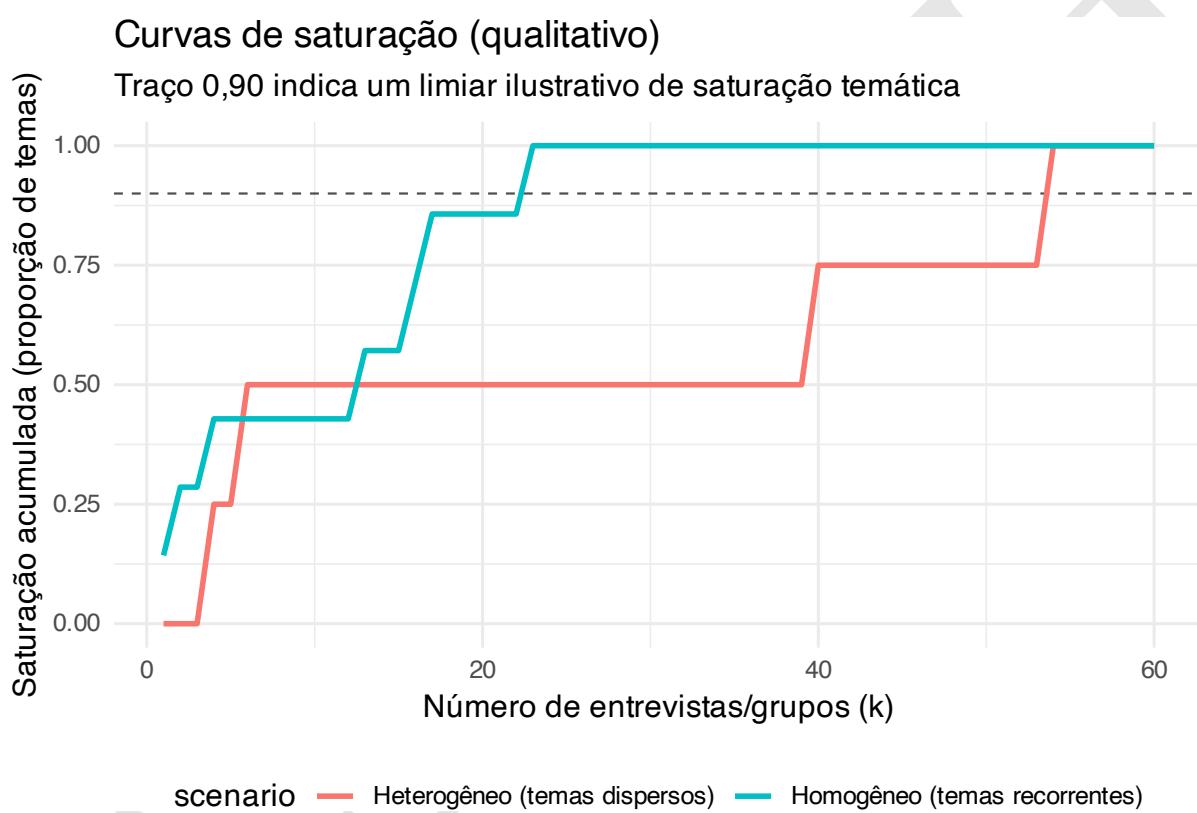


Figura 47.2: Curvas de saturação para estudos qualitativos de descoberta de temas.

- Revisões recentes sugerem que a saturação teórica exige 20 a 30 entrevistas ou mais, dependendo da complexidade do estudo.³⁷⁵

47.2.4 Quais debates existem sobre o conceito de saturação?

- Defensores argumentam que a saturação é central para garantir rigor e confiança nos resultados qualitativos.³⁷³
- Críticos sugerem que o conceito pode ser usado de forma rígida, levando a coletas excessivas ou pouco sensíveis a perspectivas únicas.³⁷³
- Pesquisadores contemporâneos recomendam usar a saturação de forma flexível, adaptada ao contexto, método e população estudada.³⁷⁵

47.2.5 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.^{374,375}
- Para grupos focais: 4–8 grupos são geralmente adequados.³⁷⁴
- Para estudos multicêntricos: recomenda-se 20–40 entrevistas por local para alcançar saturação de metatemas.³⁷⁵
- É importante relatar não apenas o número de entrevistas, mas também como e quando a saturação foi avaliada.³⁷⁶

47.3 “Fome de dados”

47.3.1 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.

47.3.2 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.³³⁵
- O uso de modelos modernos só se justifica quando há grandes bases de dados, caso contrário o ganho em acurácia é marginal.³³⁵
- Esse conceito conecta diretamente a escolha do modelo ao planejamento amostral.³³⁵

47.4 Eventos por variável (EPV) em modelos preditivos

47.4.1 Quantos eventos por variável (EPV) são necessários?

- Regressão logística: entre 20 e 50 EPV.³³⁵
- Árvore de decisão para classificação e regressão: cerca de 60 EPV.³³⁵

- Máquina de vetores de suporte, redes neurais e *random forests*: muitas vezes >200 EPV e ainda instáveis.³³⁵

47.4.2 O que acontece se não houver eventos suficientes?

- Modelos modernos podem apresentar alto otimismo (desempenho inflado no treino, mas ruim na validação).³³⁵
- Pequenos bancos de dados favorecem o uso de modelos clássicos.³³⁵

47.5 Cálculo do tamanho da amostra

47.5.1 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.³⁷¹
- 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.³⁷¹
- 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.¹³⁸

47.5.2 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.³⁷¹
- 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.³⁷⁷
- 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).³⁷⁸

R

O pacote *pwr*²⁸⁸ 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*²⁸⁸ 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.2p.test>

R

O pacote *pwr*²⁸⁸ 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*²⁸⁸ 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*²⁸⁸ 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*²⁸⁸ 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*²⁸⁸ 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*²⁸⁸ 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*²⁸⁸ 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*²⁸⁸ 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*³⁶⁹ 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>

47.6 Perdas de amostra

47.6.1 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.³⁷¹
- 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.³⁷¹

47.6.2 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).^{138,371}
- 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.³⁷¹

47.6.3 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.³⁷¹
- 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.³⁷¹

47.7 Ajustes no tamanho da amostra

47.7.1 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.³⁷¹

47.7.2 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' (47.1):³⁷¹

$$n' = \frac{n}{1 - d} \quad (47.1)$$

47.8 Justificativa do tamanho da amostra

47.8.1 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.³⁷¹
- 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.³⁷¹

47.8.2 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.³⁷⁶
- 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.³⁷⁶
- Relatar claramente o processo de decisão aumenta a transparência e a credibilidade da pesquisa.³⁷⁶

RASCUNHO

Capítulo 48

Plano de análise

48.1 Plano de análise estatística

48.1.1 O que é plano de análise estatística?

- O plano de análise estatística é um documento que descreve, de forma pré-especificada, os métodos estatísticos que serão utilizados para analisar os dados de um estudo científico.³¹
- O plano de análise estatística define como os dados serão analisados, quais estimativas serão produzidas, quais populações serão consideradas, como eventos intercurrentes serão tratados e quais procedimentos serão usados para lidar com dados incompletos.³¹

48.1.2 Qual é a relação entre pergunta científica e plano de análise?

- O plano de análise estatística deve ser consequência direta da pergunta científica do estudo e não apenas da disponibilidade dos dados ou da familiaridade com determinados métodos analíticos.³²
- Perguntas cientificamente distintas exigem estimativas distintas, mesmo quando os dados observados são os mesmos.³²
- A ausência de uma definição explícita da pergunta científica pode resultar em análises tecnicamente corretas, porém conceitualmente desalinhadas com o objetivo do estudo.³²

48.1.3 Por que a pré-especificação do plano de análise é fundamental?

- A pré-especificação do plano de análise estatística reduz o risco de flexibilidade analítica, data dredging e p-hacking.³¹
- Ela permite distinguir análises confirmatórias de análises exploratórias, preservando a interpretação inferencial dos resultados.³¹
- Em estudos clínicos e observacionais, a existência de um SAP claro fortalece a avaliação crítica da qualidade metodológica do estudo.^{31,32}

48.1.4 Como o plano de análise lida com dados perdidos?

- O tratamento de dados perdidos deve ser especificado no plano de análise estatística de forma explícita e justificada.³²

- A escolha do método de tratamento de dados perdidos está diretamente relacionada ao estimando de interesse e às suposições feitas sobre o mecanismo de ausência dos dados.³²
- Métodos como imputação múltipla, análise de casos completos ou modelos baseados em máxima verossimilhança não são intercambiáveis do ponto de vista conceitual e podem corresponder a estimandos distintos.³²

48.2 Estimandos

48.2.1 O que são estimandos e por que eles são importantes?

- Um estimando define precisamente o efeito ou quantidade de interesse que se pretende estimar em um estudo.³²
- A definição explícita do estimando reduz ambiguidades frequentes em análises estatísticas, como interpretações divergentes entre análises por intenção de tratar, por protocolo ou análises com imputação de dados.³²

48.2.2 Qual é a diferença entre estimando teórico e estimando empírico?

- O estimando teórico é a quantidade de interesse definida em termos conceituais, podendo envolver conatractuais ou quantidades não observáveis (por exemplo, o efeito médio de um tratamento na população-alvo).³⁷⁹
- O estimando empírico é a quantidade que pode ser expressa apenas com dados observáveis e que se conecta ao estimando teórico por meio de suposições de identificação.³⁷⁹
- A distinção entre estimando teórico e empírico torna explícitas as suposições necessárias para interpretar uma estimativa como evidência sobre uma questão científica.³⁷⁹
- Essa distinção evita que o leitor confunda uma associação observada com o efeito causal pretendido.³⁷⁹

48.2.3 Por que definir o estimando antes do método estatístico?

- Todo estudo quantitativo deve ser capaz de responder explicitamente à pergunta: qual é o seu estimando?³⁷⁹
- O estimando é o objeto de interesse — a quantidade precisa sobre a qual desejamos fazer inferência — e não deve ser confundido com o coeficiente de um modelo específico.³⁷⁹
- Definir o objetivo dentro de um modelo (por exemplo, “o coeficiente da regressão logística”) restringe artificialmente a pergunta científica ao formato daquele modelo estatístico.³⁷⁹
- Ao declarar o estimando fora do modelo estatístico (por exemplo, uma diferença média populacional ou um efeito causal médio), ampliamos as possibilidades metodológicas e fortalecemos a conexão entre teoria e evidência empírica.³⁷⁹

48.2.4 Quais componentes devem ser definidos em um estimando?

- Quantidade específica por unidade (por exemplo, desfecho observado ou potencial sob intervenção).³⁷⁹
- População-alvo à qual a inferência se refere.^{32,379}
- Variável de interesse (desfecho).^{32,379}
- Forma de tratamento de eventos intercurrentes (por exemplo, abandono do tratamento, uso de terapias concomitantes ou morte).^{32,379}
- Medida resumo que expressa o efeito de interesse (por exemplo, diferença de médias, razão de riscos ou odds ratio).^{32,379}

48.2.5 Qual é a relação entre estimando, estimador e método estatístico?

- O estimando descreve o que se deseja estimar.³²
- O estimador descreve como o estimando será estimado a partir dos dados observados.³²
- O método estatístico descreve o procedimento matemático e computacional usado para calcular o estimador.³²
- Diferentes métodos estatísticos podem produzir estimativas válidas para o mesmo estimando, assim como o mesmo método pode estimar estimandos diferentes, dependendo das decisões analíticas adotadas.³²

48.2.6 Como a definição explícita do estimando melhora o Plano de Análise Estatística?

- Obriga o pesquisador a declarar a pergunta científica em termos formais antes da escolha do método.³⁷⁹
- Separa claramente três decisões distintas: definir o estimando teórico, estabelecer o estimando empírico sob hipóteses de identificação e escolher a estratégia de estimação.³⁷⁹
- Reduz ambiguidades interpretativas, especialmente em estudos observacionais.³⁷⁹
- Facilita a avaliação crítica por revisores e leitores, pois torna explícitas as suposições necessárias para que a evidência empírica sustente a conclusão teórica.³⁷⁹

48.2.7 Por que o estimando deve existir fora do modelo estatístico?

- O estimando deve ser definido independentemente do modelo estatístico que será utilizado para estimá-lo.³⁷⁹
- Quando o objetivo científico é formulado como um parâmetro específico de um modelo (por exemplo, um coeficiente de regressão), a pergunta científica passa a depender das suposições daquele modelo.³⁷⁹
- A definição do estimando fora do modelo amplia o espaço de possíveis estratégias analíticas e permite avaliar métodos alternativos sob os mesmos objetivos científicos.³⁷⁹
- No contexto de um plano de análise estatística, isso significa que o método é escolhido para estimar o estimando — e não o estimando para justificar o método.

48.3 Diretrizes para redação

48.3.1 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:*³¹ <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 49

Delineamento de estudos

49.1 Critérios de delineamento

49.1.1 Quais critérios são utilizados para classificar os delineamentos de estudos?

- ?

49.2 Alocação

49.2.1 O que é alocação?

- ?

49.3 Cegamento

- ?

49.3.1 O que é cegamento?

49.4 Pareamento

49.4.1 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.³⁸⁰
- As variáveis escolhidas para pareamento devem ter relação com as variáveis de desfecho, mas não são de interesse elas mesmas.³⁸⁰
- 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.³⁸⁰
- A ausência de evidência estatística de diferença entre grupos não é considerada pareamento.³⁸⁰

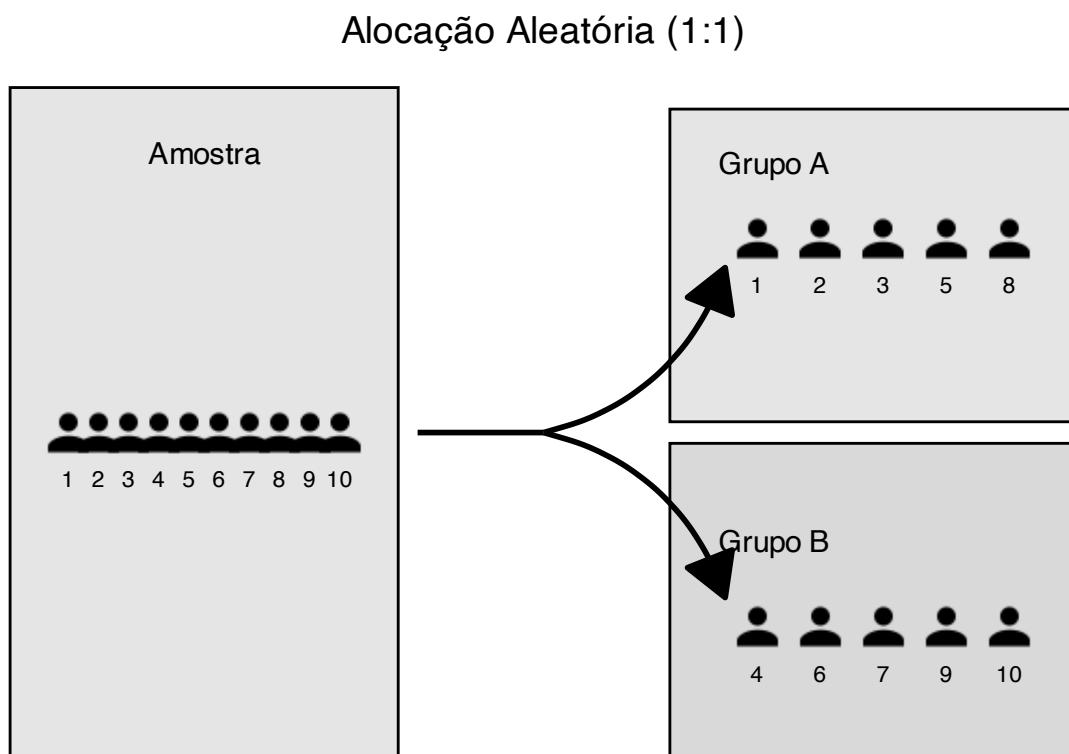


Figura 49.1: Alocação 1:1 entre dois grupos de participantes

49.5 Aleatorização

49.5.1 O que é aleatorização?

- ?

49.6 Taxonomia de estudos

49.6.1 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*.³⁸¹⁻³⁹⁰
- *Estudos básicos*^{382,387}
 - Genética
 - Celular
 - Experimentos com animais
 - Desenvolvimento de métodos
- *Estudos de simulação computacional*^{388,390}
- *Estudos de propriedades psicométricas*^{383,385}
 - Validade
 - Concordância

- Confiabilidade
- *Estudos de desempenho diagnóstico*^{386,389}
 - 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*^{382,387}
 - Descritivo
 - * Estudo de caso
 - * Série de casos
 - * Transversal
 - Analítico
 - * Transversal
 - * Caso-Controle
 - Caso-Controle aninhado
 - Caso-Coorte
 - Coorte prospectiva ou retrospectiva
- *Estudos quase-experimentais*³⁸⁴
 - Quase-aleatorizado controlado
 - Estimação de variável instrumental
 - Descontinuidade de regressão
 - Série temporal interrompida controlada
 - Série temporal interrompida
 - Diferença
- *Estudos experimentais*^{382,387}
 - Fases I a IV
 - * Aleatorizado controlado
 - * Não-aleatorizado controlado
 - * Autocontrolado
 - * Cruzado
 - * Fatorial

- Campo
- Comunitário
- *Estudos de avaliação econômica*³⁸²
 - Análise de custo
 - Análise de minimização de custo
 - Análise de custo-utilidade
 - Análise de custo-efetividade
 - Análise de custo-benefício
- *Estudos de revisão*³⁸¹
 - Estado-da-arte
 - Narrativa
 - Crítica
 - Mapeamento
 - Escopo
 - Busca e revisão sistemática
 - Sistematizada
 - Sistemática
 - * Meta-análise
 - * Bibliométrica.^{391,392}
 - Sistemática qualitativa
 - Mista
 - Visão geral
 - Rápida
 - Guarda-chuva

Capítulo 50

Simulação computacional

50.1 Simulações computacionais

50.1.1 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.³⁰
- A simulação é frequentemente usada em estatística para avaliar o desempenho de testes, estimadores e modelos sob diferentes condições.³¹

50.1.2 Por que usar simulações?

- Testar o comportamento de métodos estatísticos sob diferentes premissas (ex: normalidade, homocedasticidade, tamanho amostral).³²
- Avaliar a robustez de algoritmos computacionais.³³
- Reproduzir processos naturais ou sociais para compreensão teórica.³⁴

50.1.3 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.⁴⁶
- Use uma semente para o gerador de números aleatórios com `set.seed()` para garantir a reproduzibilidade dos resultados.³⁵
- Documente detalhadamente o processo de simulação, incluindo os parâmetros utilizados, a lógica do algoritmo e as suposições feitas.³⁹³
- Realize múltiplas simulações (ex.: 1000 ou mais) para obter estimativas estáveis e resultados mais robustos e confiáveis.³⁶
- Analise os resultados de forma crítica, considerando a variabilidade, as limitações do modelo e possíveis vieses do processo de simulação.³⁷
- Use funções vetorizadas para otimizar o desempenho e reduzir o tempo de execução da simulação.³⁸



O pacote *base*⁵⁹ 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>

50.2 Características

50.2.1 Quais são as características de estudos de simulação computacional?

- ?

50.3 Métodos de simulação

50.3.1 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.⁵
- A escolha da distribuição depende da natureza da variável (contínua ou discreta), do domínio dos valores possíveis e da estrutura probabilística do processo observado.⁵
- 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.⁵
- 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.⁵

50.3.2 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.⁵
- Aspectos como o tipo da variável, assimetria, limites naturais e frequência de eventos devem ser considerados na especificação da distribuição.⁵
- Em estudos de simulação, é comum avaliar múltiplas distribuições para investigar a sensibilidade dos resultados às suposições do modelo.⁴⁶

50.3.3 Como simular dados de diferentes distribuições?

- Use funções específicas para cada distribuição, disponíveis nas bibliotecas estatísticas da linguagem utilizada.⁵



O pacote *stats*¹³⁶ 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*¹³⁶ 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*¹³⁶ 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*¹³⁶ 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

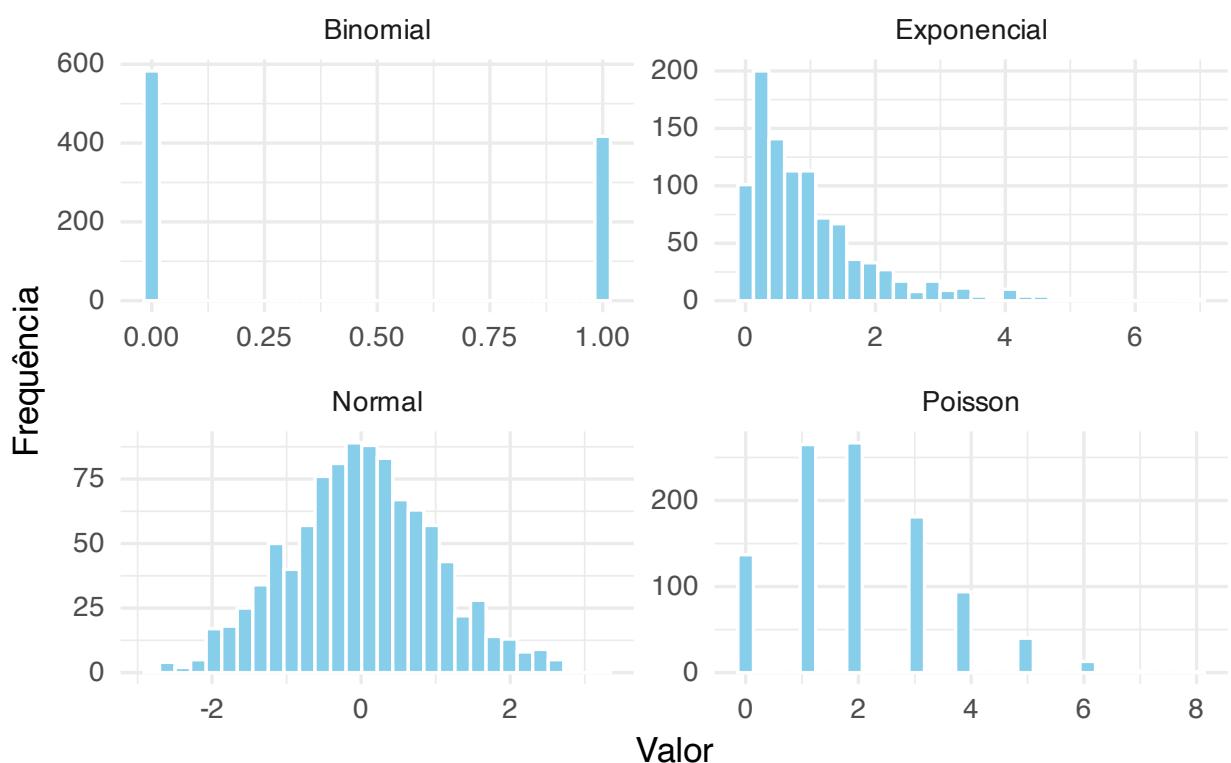


Figura 50.1: Dados simulados a partir de diferentes distribuições: $\text{Normal}(0,1)$, $\text{Binomial}(1,0.4)$, $\text{Poisson}(2)$ e $\text{Exponencial}(1)$.

- Independentemente da distribuição utilizada, o procedimento de simulação segue os mesmos princípios computacionais, diferindo apenas na função geradora dos números aleatórios.³⁹⁴

50.3.4 O que é o método de Monte Carlo?

- ³⁹⁴

- 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.³³⁶

Convergência do histograma \rightarrow Normal(0,1)

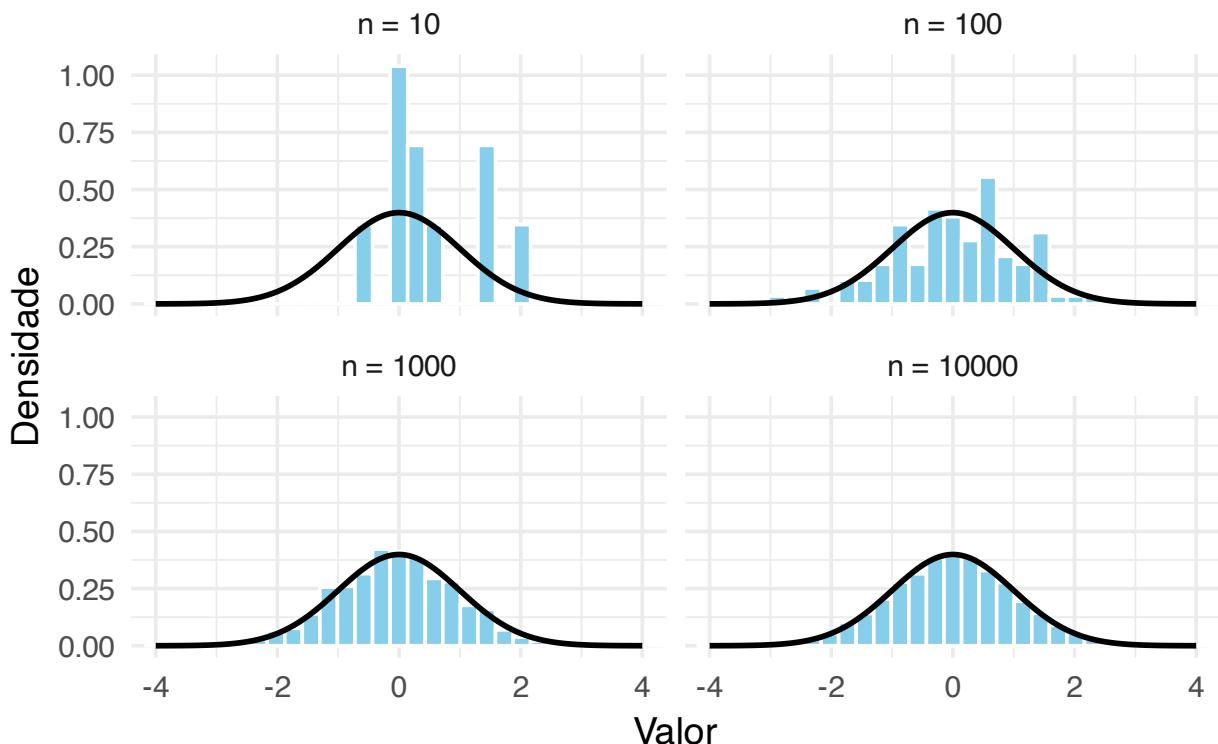


Figura 50.2: Convergência do histograma para a PDF teórica da $Normal(0,1)$ com o aumento do tamanho amostral ($n = 10, 100, 1000, 10000$).

R O pacote *base*⁵⁹ 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*³⁹⁵ 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*³⁹⁶ 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

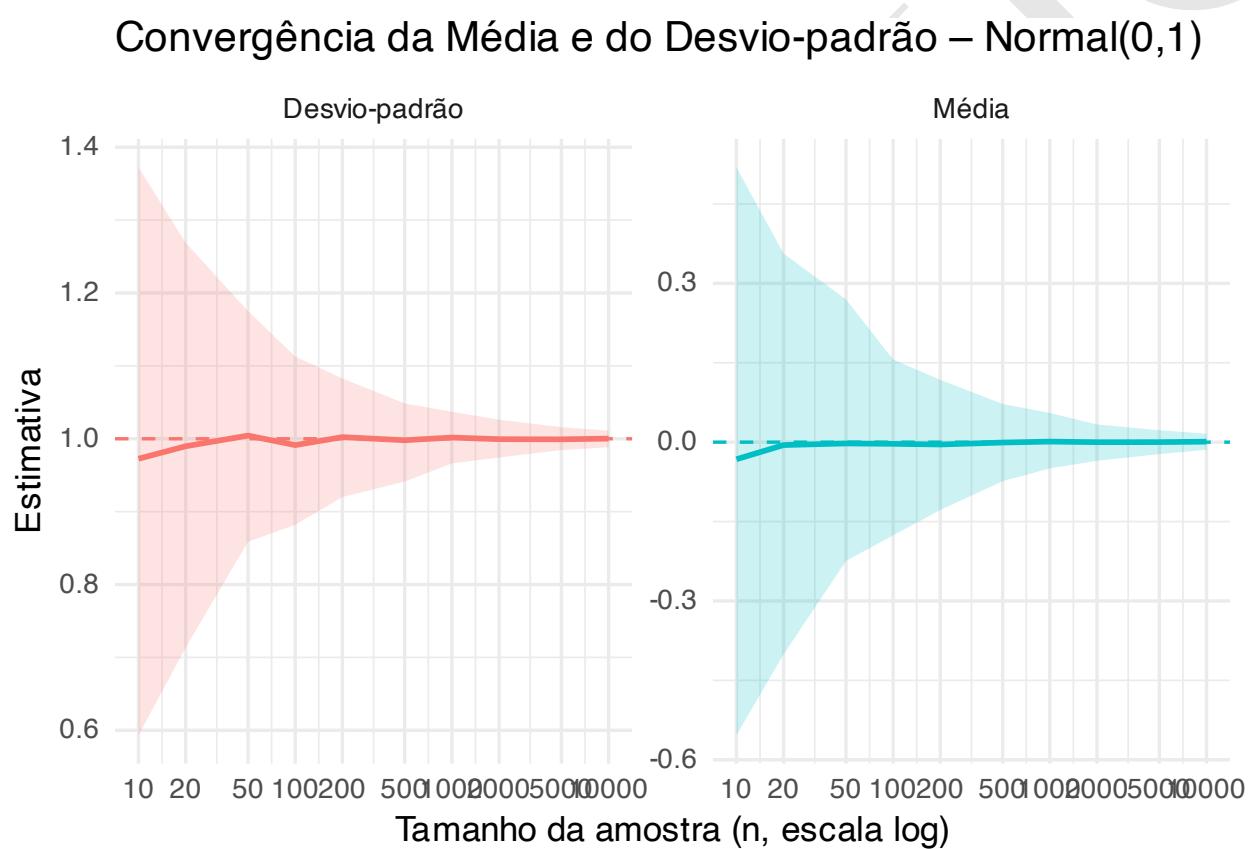


Figura 50.3: Convergência da média e do desvio-padrão amostral para os valores teóricos (0 e 1, respectivamente) com o aumento do tamanho amostral ($n = 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000$).



O pacote *InteractionPowerR*³⁷⁰ 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

50.4 Diretrizes para redação

50.4.1 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:*³⁹⁷ <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 51

Estudos observacionais

51.1 Características

51.1.1 Quais são as características de estudos observacionais?

- ?

51.2 Diretrizes para redação

51.2.1 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:*³⁹⁸ <https://www.equator-network.org/reporting-guidelines/strobe/>

¹<https://www.equator-network.org/>

RASCUNHO

Capítulo 52

Concordância e confiabilidade

52.1 Concordância

52.1.1 O que é concordância?

- ?

52.1.2 Quais métodos são adequados para análise de concordância de variáveis dicotômicas?

- Coeficiente de Cohen κ : mede a concordância corrigida pelo acaso.^{399,400}

 O pacote *irr*⁴⁰¹ 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>

 O pacote *irr*⁴⁰¹ 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>

Tabela 52.1: 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	$e = a + c$	$f = b + d$	$N = a + b + c + d$

Tabela 52.2: 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	$j = a + d + g$	$k = b + e + h$	$l = c + f + i$	$N = a + b + c + d + e + f + g + h + i$

R

O pacote *irr*⁴⁰¹ 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>

52.1.3 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.³⁰⁵
- 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.³⁰⁵
- Qui-quadrado χ^2 a partir da tabela de contigência não é recomendado porque tal teste analisa associação.³⁰⁵
- A família de coeficientes de Cohen κ não é adequada para analisar concordância quando as variáveis são aparentemente (e não originalmente) dicotômicas.³⁰⁵

R

O pacote *irr*⁴⁰¹ 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>

52.1.4 Quais métodos são adequados para análise de concordância de variáveis categóricas?

- Coeficiente de Cohen κ : mede a concordância corrigida pelo acaso.^{399,400}
- Coeficiente de Cohen ponderado κ_w : mede a concordância corrigida pelo acaso.^{399,400}

52.1.5 Quais métodos são adequados para análise de concordância de variáveis ordináis?

- Coeficiente de Cohen ponderado κ_w : mede a concordância corrigida pelo acaso.^{399,400}

52.1.6 Quais métodos são adequados para análise de concordância de variáveis contínuas?

- Gráfico de dispersão com a reta de regressão.¹⁴⁹

- 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.¹⁴⁹



O pacote *BlandAltmanLeh*⁴⁰² 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>

52.1.7 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.¹⁴⁹
- 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.¹⁴⁹
- Regressão linear: o teste da hipótese nula da inclinação da reta de regressão ($H_0 : \beta = 0$) é equivalente a testar a correlação bivariada ($H_0 : \rho = 0$).¹⁴⁹

52.1.8 Quais métodos são adequados para modelagem de concordância?

- Modelo log-linear.³⁰⁵

52.2 Confiabilidade

52.2.1 O que é confiabilidade?

- ?

52.2.2 Quais métodos são adequados para análise de confiabilidade?

- ?



O pacote *semTools*⁴⁰³ 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>



O pacote *psych*³⁰⁶ fornece a função *icc*^a para estimar a confiabilidade utilizando coeficientes de correlação intraclass.

^a<https://www.rdocumentation.org/packages/psych/versions/2.3.6/topics/ICC>

R

O pacote *multilevel*⁴⁰⁴ 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*⁴⁰⁴ 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>

R

O pacote *irr*⁴⁰¹ 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>

R

O pacote *irr*⁴⁰¹ 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>

52.3 Diretrizes para redação

52.3.1 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.
- *Guidelines for reporting reliability and agreement studies (GRRAS) were proposed:*⁴⁰⁵ <https://www.equator-network.org/reporting-guidelines/guidelines-for-reporting-reliability-and-agreement-studies-grras-were-proposed/>

¹<https://www.equator-network.org/>

Capítulo 53

Propriedades psicométricas

53.1 Características

53.1.1 O que são propriedades psicométricas?

- ?



O pacote *lavaan*⁴⁰⁶ fornece a função *modificationIndices*^a para calcular os índices de modificação.

^a<https://www.rdocumentation.org/packages/lavaan/versions/0.6-16/topics/modificationIndices>

53.2 Análise factorial exploratória

53.2.1 O que é análise factorial exploratória?

- ?

53.3 Análise factorial confirmatória

53.3.1 O que é análise factorial confirmatória?

- ?

Tabela 53.1: Tabela de confusão sobre propriedades psicométricas de instrumentos.

	Concordância alta	Concordância baixa
Validade alta	Adequado	Inadequado
Validade baixa	Inadequado	Inadequado

R

O pacote *lavaan*⁴⁰⁶ 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>

53.4 Validade de conteúdo

53.4.1 O que é validade interna?

- ⁴⁰⁷

53.4.2 O que é validade externa?

- ⁴⁰⁷

53.4.3 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

53.4.4 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.²²²

53.5 Validade de face

53.5.1 O que é validade de face?

- ?

53.6 Validade do construto

53.6.1 O que é construto?

- ?

53.7 Validade fatorial

53.7.1 O que é validade fatorial?

- ?

53.8 Validade convergente

53.8.1 O que é validade convergente?

- ?

53.9 Validação discriminante

53.9.1 O que é validade discriminante?

- ?

53.10 Validação de critério

53.10.1 O que é validade de critério?

- ?

53.11 Validação concorrente

53.11.1 O que é concorrente?

- ?

53.11.2 O que é validade concorrente?

- ?

53.11.3 O que é validade preditiva?

- ?

53.12 Responsividade

53.12.1 O que é responsividade?

- ?

53.13 Diretrizes para redação

53.13.1 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*:⁴⁰⁸ <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*:⁴⁰⁹ <https://www.equator-network.org/reporting-guidelines/recommendations-for-reporting-the-results-of-studies-of-instrument-and-scale-development-and-testing/>

¹<https://www.equator-network.org/>

RASCUNHO

Capítulo 54

Desempenho diagnóstico

54.1 Características

54.1.1 Quais são as características de estudos de desempenho diagnóstico?

- ?

54.2 Tabelas 2x2

54.2.1 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).⁴¹⁰

54.2.2 Como analisar o desempenho diagnóstico em tabelas 2x2?

- Verdadeiro-positivo (*VP*): caso com a condição presente e corretamente identificado como tal.⁴¹¹
- Falso-negativo (*FN*): caso com a condição presente e erroneamente identificado como ausente.⁴¹¹
- Verdadeiro-negativo (*VN*): controle sem a condição presente e corretamente identificados como tal.⁴¹¹
- Falso-positivo (*FP*): controle sem a condição presente e erroneamente identificado como presente.⁴¹¹
- Tabelas de confusão também podem ser visualizadas em formato de árvores de frequência.⁴¹⁰

Tabela 54.1: 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

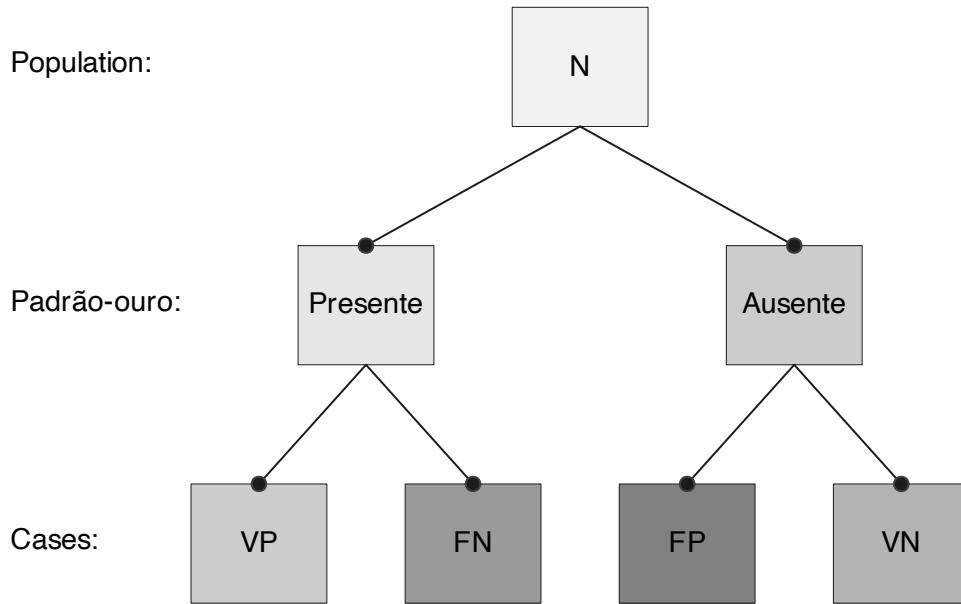


Figura 54.1: Árvore de frequência do desempenho diagnóstico de uma tabela de confusão 2x2 representando um método novo (dicotômico) comparado ao método padrão-ouro ou referência (dicotômico).

R O pacote *riskyR*⁴¹² 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

54.2.3 Quais probabilidades caracterizam o desempenho diagnóstico de um teste em tabelas 2x2?

- Sensibilidade (*SEN*) (54.1): Proporção de verdadeiro-positivos dentre aqueles com a condição.⁴¹¹

$$SEN = \frac{VP}{VP + FN} \quad (54.1)$$

- Especificidade (*ESP*) (54.2): Proporção de verdadeiro-negativos dentre aqueles sem a condição.⁴¹¹

$$ESP = \frac{VN}{VN + FP} \quad (54.2)$$

- Valor predictivo positivo (*VPP*) (54.3): Proporção de casos corretamente identificados como verdadeiro-positivos.⁴¹¹

$$VPP = \frac{VP}{VP + FP} \quad (54.3)$$

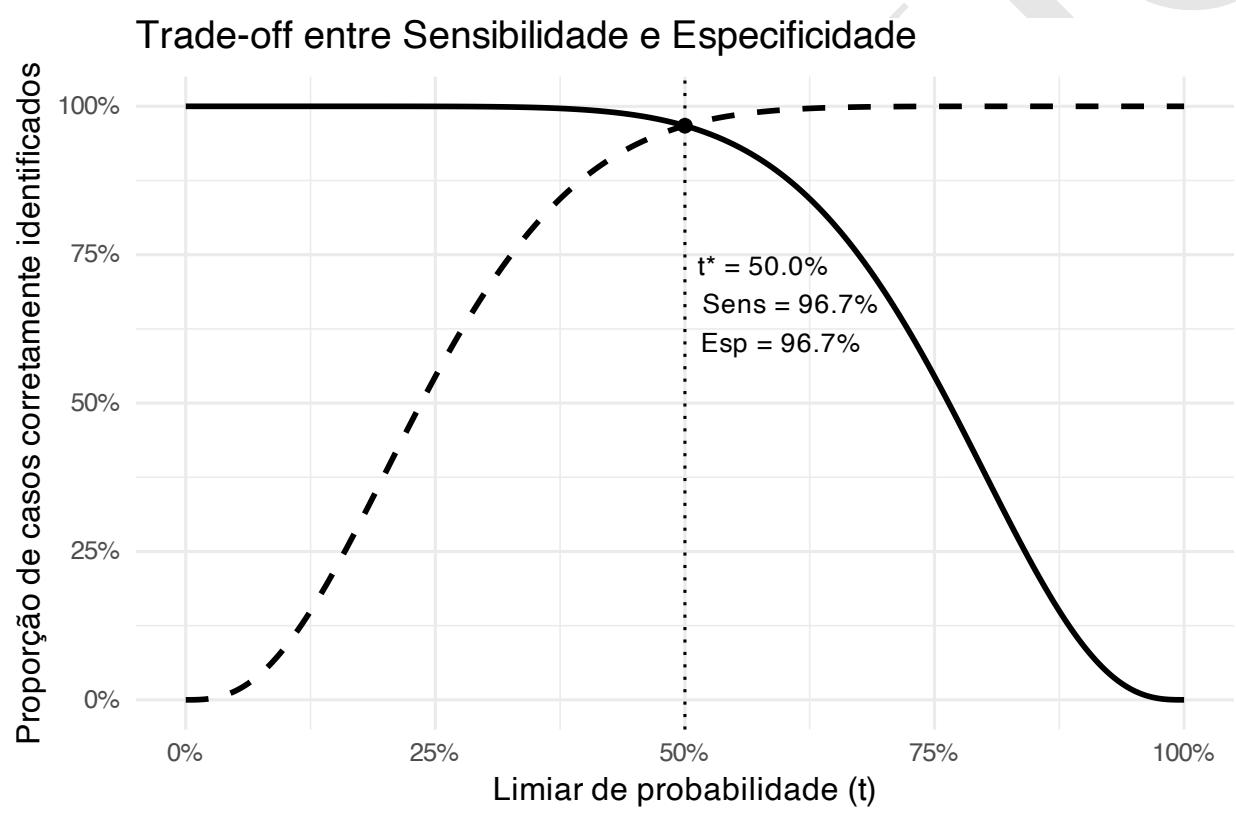


Figura 54.2: Trade-off entre sensibilidade e especificidade em função do limiar de probabilidade (t) para um modelo de classificação.

Tabela 54.2: 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	$SEN = \frac{VP}{VP+FN}$	$ESP = \frac{VN}{VN+FP}$		$ACU = \frac{VP+VN}{VP+VN+FP+FN}$ $DOR = \frac{VP \cdot VN}{FP \cdot FN}$

- Valor preditivo negativo (VPN) (54.4): Proporção de controles corretamente identificados como verdadeiro-negativos.⁴¹¹

$$VPN = \frac{VN}{VN + FN} \quad (54.4)$$

- Razão de verossimilhança positiva ($LR+$) (54.5): Quantifica o quanto a probabilidade de a condição estar presente aumenta quando o teste é positivo.⁴¹²

$$LR+ = \frac{SEN}{1 - ESP} = \frac{VP/(VP + FN)}{FP/(FP + VN)} \quad (54.5)$$

- Razão de verossimilhança negativa ($LR-$) (54.6): Quantifica o quanto a probabilidade de a condição estar presente diminui quando o teste é negativo.⁴¹³

$$LR- = \frac{1 - SEN}{ESP} = \frac{FN/(VP + FN)}{VN/(FP + VN)} \quad (54.6)$$

- Acurácia (ACU), (54.7): Proporção de casos e controles corretamente identificados.⁴¹¹

$$ACU = \frac{VP + VN}{VP + VN + FP + FN} \quad (54.7)$$

- Razão de chances diagnóstica (DOR) (54.8), (54.9) e (54.10): 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.⁴¹³

$$DOR = \frac{VP}{FN} \div \frac{FP}{VN} = \frac{VP \cdot VN}{FP \cdot FN} \quad (54.8)$$

$$DOR = \frac{SEN/(1 - SEN)}{(1 - ESP)/ESP} = \frac{SEN \cdot ESP}{(1 - SEN) \cdot (1 - ESP)} \quad (54.9)$$

$$DOR = \frac{LR+}{LR-} \quad (54.10)$$



O pacote *riskyR*⁴¹² 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 54.3: 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*⁴¹⁴ 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>

54.3 Tabelas 2x3

54.3.1 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).⁴¹⁵
- As colunas** representam as decisões** (*POS*, *BND*, *NEG*) e as linhas representam a verdade de referência (condição presente vs ausente).⁴¹⁵
- Essa formulação vem do arcabouço de *Three-Way Decisions (3WD)*, que partitiona o universo em três regiões por dois limiares α e β .⁴¹⁵

54.3.2 Como as regiões POS, BND e NEG são definidas?

- Dado um escore ou probabilidade condicional $Pr(C \mid [x])$ para a classe C , classifica-se como *POS* (aceitar) quando $Pr(C \mid [x]) \geq \alpha$, como *BND* (deferir) quando $\beta < Pr(C \mid [x]) < \alpha$ e como *NEG* (rejeitar) quando $Pr(C \mid [x]) \leq \beta$, sendo que os limiares (α, β) determinam simultaneamente as três regiões e os *trade-offs* entre acurácia e comprometimento.⁴¹⁵

54.3.3 Qual é o formato de uma tabela 2×3 ?

- Estrutura geral (linhas = condição real; colunas = decisão):

54.3.4 Quais são as medidas básicas na 2×3 ?

- Acurácia em *POS* (M_{PT}), equação (54.11): Proporção de positivos corretamente identificados na região *POS*.⁴¹⁵

$$M_{PT} = \frac{|POS \cap C|}{|POS|} \quad (54.11)$$

- Erro em *POS* (M_{PF}), equação (54.12): Proporção de negativos incorretamente classificados na região *POS*.⁴¹⁵

$$M_{PF} = \frac{|POS \cap C^c|}{|POS|} \quad (54.12)$$

- Acurácia em NEG (M_{NF}), equação (54.13): Proporção de negativos corretamente identificados na região NEG.⁴¹⁵

$$M_{NF} = \frac{|NEG \cap C^c|}{|NEG|} \quad (54.13)$$

- Erro em NEG (M_{NT}), equação (54.14): Proporção de positivos incorretamente classificados na região NEG.⁴¹⁵

$$M_{NT} = \frac{|NEG \cap C|}{|NEG|} \quad (54.14)$$

- Frações em BND (M_{BT} e M_{BF}), equações (54.15) e (54.16): Proporção de deferimentos verdadeiros e falsos.⁴¹⁵

$$M_{BT} = \frac{|BND \cap C|}{|BND|} \quad (54.15)$$

$$M_{BF} = \frac{|BND \cap C^c|}{|BND|} \quad (54.16)$$

54.3.5 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).⁴¹⁵

54.3.6 Quando preferir 3-vias em vez de 2×2?

- Quando o custo de erro é assimétrico e/ou há incerteza relevante.⁴¹⁵
- O deferimento (BND) evita decisões precipitadas e permite avaliação adicional, equilibrando acurácia e cobertura.⁴¹⁵
- É particularmente útil em triagens diagnósticas com etapas confirmatórias.⁴¹⁵

54.4 Curvas ROC

54.4.1 O que representa a curva ROC?

- A relação entre sensibilidade (SEN) no eixo vertical e $1 - ESP$ no eixo horizontal.⁴¹⁶
- Cada ponto na curva corresponde a um ponto de corte possível do teste.⁴¹⁶

54.4.2 Quais são os tipos de curva ROC?

- Curva empírica: conecta diretamente os pontos obtidos a partir dos diferentes pontos de corte observados.⁴¹⁷
- Curva suavizada (paramétrica): assume uma distribuição binomial e gera uma curva ajustada por máxima verossimilhança.⁴¹⁷

54.4.3 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.^{416,417}
- O método de Youden (equação (54.17)) 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 .¹³¹

$$Y = SEN + ESP - 1 \quad (54.17)$$

- O método da distância Euclidiana ((54.18)) 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 .⁴¹⁸

$$D = \sqrt{(1 - SEN)^2 + (1 - ESP)^2} \quad (54.18)$$

54.4.4 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.⁴¹⁹
- 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.⁴¹⁶

54.4.5 Como calcular a AUC?

- Método não paramétrico: soma das áreas trapezoidais sob a curva empírica (54.19). Pode subestimar AUC quando os dados são discretos.⁴¹⁷

$$AUC = \sum_{i=1}^{n-1} (x_{i+1} - x_i) \cdot \frac{y_i + y_{i+1}}{2} \quad (54.19)$$

- Método paramétrico (binormal): mais robusto para dados em escala ordinal, com viés reduzido (54.20), 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.⁴¹⁷

$$AUC = \Phi \left(\frac{\mu_1 - \mu_0}{\sqrt{\sigma_1^2 + \sigma_0^2}} \right) \quad (54.20)$$

- AUC deve sempre vir acompanhada de intervalo de confiança (IC95%).⁴¹⁷



O pacote *proc*⁴²⁰ 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>

54.4.6 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.⁴¹⁹

- 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.⁴¹⁹
- 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.⁴¹⁹
- Diferenças pequenas entre AUCs podem não ser estatisticamente significativas.⁴¹⁶
- A interpretação clínica pode ser equivocada se não houver teste estatístico adequado.⁴¹⁶
- Se as curvas vêm do mesmo conjunto de pacientes, aplique o teste de DeLong.⁴¹⁶
- Se as curvas vêm de amostras independentes, use métodos como Dorfman-Alf.⁴¹⁶

54.4.7 Por que uma AUC menor que 0.5 está errada?

- Porque indica desempenho pior que o acaso.⁴¹⁶
- Geralmente decorre de seleção incorreta da direção do teste ou da variável de estado.⁴¹⁶
- Verifique se o software está configurado para maiores valores indicam presença do evento ou o inverso.⁴¹⁶
- Ajuste a direção do teste para que $AUC \geq 0.5$.⁴¹⁶

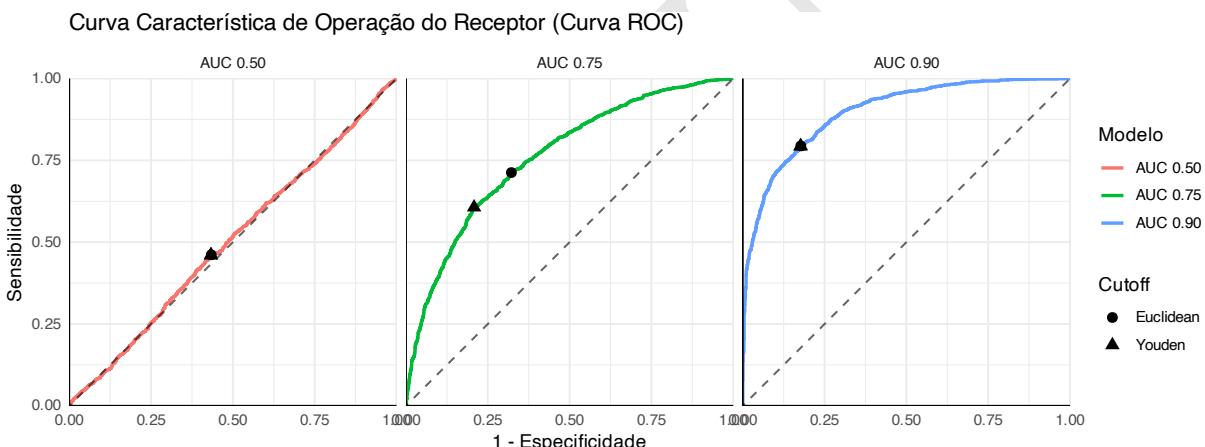


Figura 54.3: Curva ROC (Receiver Operating Characteristic) para um modelo de classificação com diferentes desempenhos diagnósticos.

54.4.8 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.⁴²¹

54.5 Interpretação da validade de um teste

54.5.1 Que itens devem ser verificados na interpretação de um estudo de validade?

- O novo teste foi comparado junto ao método padrão-ouro.⁴¹¹
- As probabilidades pontuais estimadas que caracterizam o desempenho diagnóstico do novo teste são altas e adequadas para sua aplicação clínica.⁴¹¹

- Os intervalos de confiança estimados para as probabilidades do novo teste são estreitos e adequadas para sua aplicação clínica.⁴¹¹
- O novo teste possui adequada confiabilidade intra/inter examinadores.⁴¹¹
- O estudo de validação incluiu um espectro adequado da amostra.⁴¹¹
- Todos os participantes realizaram ambos o novo teste e o padrão-ouro no estudo de validação.⁴¹¹
- Os examinadores do novo teste estavam cegados para o resultado do teste padrão-ouro.⁴¹¹

54.6 Diretrizes para redação

54.6.1 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:*⁴²² <https://www.equator-network.org/reporting-guidelines/stard/>

¹<https://www.equator-network.org/>

RASCUNHO

Capítulo 55

Ensaios quase-experimentais

55.1 Características

55.1.1 Quais são as características de ensaios quase-experimentais?

- ?

55.2 Diretrizes para redação

55.2.1 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:*⁴²³ <https://www.equator-network.org/reporting-guidelines/guidelines-for-reporting-non-randomised-studies/>

¹<https://www.equator-network.org/>

RASCUNHO

Capítulo 56

Ensaios experimentais

56.1 Ensaio experimental aleatorizado

56.1.1 Quais são as características de ensaios experimentais aleatorizados?

- A característica essencial de um ensaio experimental aleatorizado é a comparação entre grupos.⁴²⁴
- Quanto à unidade de alocação:⁴²⁵
 - Individual
 - Agrupado
- Quanto ao número de braços.⁴²⁵
 - Múltiplos
- Quanto ao número de centros:⁴²⁵
 - Único
 - Múltiplos
- Quanto ao cegamento:⁴²⁵
 - Aberto
 - Simples-cego
 - Duplo-cego
 - Triplo-cego
 - Quádruplo-cego
- Quanto à alocação:⁴²⁵
 - Sem sorteio
 - Estratificada (centro apenas)
 - Estratificada
 - Minimizada
 - Estratificada e minimizada

56.1.2 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.⁴²⁵
- Grupo placebo: prepara uma intervenção indistinguível da ativa para mitigar expectativas de participantes e profissionais, reduzindo viés de desempenho e detecção.⁴²⁶
- 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.⁴²⁷
- 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.⁴²⁸

56.2 Modelos de análise de comparação

56.2.1 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.⁴²⁶
- 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.⁴²⁶
- A abordagem mais recomendada é a análise de covariância (ANCOVA) (56.1), 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.^{10,426}

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 Z_j + \epsilon_{ij} \quad (56.1)$$

- 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.⁴²⁷
- Quando a ANCOVA (56.2) é utilizada com a mudança (pré - pós) com variável de desfecho (Y_{ij}), o coeficiente de regressão β_1 é diminuído em 1 unidade.^{10,428}

$$(X_{ij} - Y_{ij}) = \beta_0 + \beta_1 Z_j + \epsilon_{ij} \quad (56.2)$$

- 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.⁴²⁷
- 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.⁴²⁹
- 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”.⁴³⁰

56.3 Comparação na linha de base

56.3.1 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.⁴³¹
- 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.⁴³¹
- Dados da linha de base podem ser utilizados para a aleatorização de modo a equilibrar ou estratificar os grupos considerando alguns fatores-chave.⁴³¹
- Dados da linha de base podem ser utilizados como ajuste de covariável para análise do resultado por grupo de tratamento.⁴³¹

56.3.2 O que é comparação entre grupos na linha de base em ensaios experimentais 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'.⁴³²
- A interpretação isolada do P-valor da comparação entre grupos na linha de base não permite identificar as razões para eventuais diferenças.⁴³²

56.3.3 Quais são as razões para diferenças entre grupos de tratamento nas (co)variáveis na linha de base?

- Acaso.^{223,432}
- Viés.^{223,432}
- Tamanho da amostra.^{223,432}
- Má conduta científica.²²³

56.3.4 Para quê comparar grupos na linha de base em ensaios experimentais aleatorizados?

- Os P-valores estão relacionados à aleatorização dos participantes em grupos.⁴³³
- Em ensaios experimentais aleatorizados, a comparação de (co)variáveis na linha de base é usada para avaliar se aleatorização foi 'bem sucedida'.⁴³³

56.3.5 Quais cenários permitem a comparação entre grupos na linha de base em ensaios experimentais aleatorizados?

- Em ensaios experimentais aleatorizados agregados, os P-valores possuem interpretação diferente de estudos aleatorizados individualmente.⁴³³
- Em ensaios experimentais 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.⁴³³

56.3.6 Por que não se deve comparar grupos na linha de base em ensaios experimentais aleatorizados?

- Se a randomização tiver sido conduzida de forma adequada, a hipótese nula de que ambos os grupos provêm da mesma população é, por definição, verdadeira; assim, seria esperado que cerca de 5% dessas comparações apresentassem significância estatística ao nível de 5%.⁴³⁴⁻⁴³⁶
- Esses testes avaliam indiretamente a adequação do processo de randomização, e não se os dois grupos possuem características semelhantes.⁴³⁴⁻⁴³⁶
- Testes de hipóteses não são uma forma válida de avaliar a similaridade. Essa avaliação deve basear-se na consideração do força prognóstica das variáveis e a magnitude do desequilíbrio.⁴³⁴⁻⁴³⁶
- 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’.²²⁴
- Quando a aleatorização é bem-sucedida, a hipótese nula de diferença entre grupos na linha de base é verdadeira.⁴³⁷
- 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.⁴³⁸

56.3.7 Quais estratégias podem ser adotadas para substituir a comparação entre grupos na linha de base em ensaios experimentais aleatorizados?

- Na fase de projeto: identifique as variáveis prognósticas do desfecho de acordo com a literatura.⁴³⁷
- Na fase de análise: inclua as variáveis prognósticas nos modelos para ajuste.⁴³⁷

56.4 Comparação intragrupos

56.4.1 Por que não se deve comparar intragrupos (pré - pós) em ensaios experimentais 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.⁴²⁴

56.5 Comparação entre grupos

56.5.1 O que é comparação entre grupos em ensaios experimentais aleatorizados?

- A comparação se refere ao teste de hipótese nula de não haver diferença (‘alteração’ ou ‘mudança’) pós-tratamento entre grupos de tratamento.⁴²⁴

56.5.2 O que pode ser comparado entre grupos?

- Valores pós-tratamento; mudança entre linha de base e pós-tratamento; mudança percentual da linha de base.⁴³⁹

56.5.3 Qual é a comparação entre grupos mais adequada em ensaios experimentais 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.⁴³⁹

- Mudança entre linha de base e pós-tratamento tem poder adequado apenas quando a correlação entre linha de base e pós-tratamento é alta.⁴³⁹
- Mudança percentual da linha de base é a opção menos eficiente em termos de poder estatístico.⁴³⁹

56.6 Comparação de subgrupos

56.6.1 O que é comparação de subgrupos em ensaios experimentais 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.⁴³¹

56.6.2 Como realizar a comparação de subgrupos em ensaios experimentais 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 apoia um efeito de subgrupo as conclusões poderão ser influenciadas.^{431,440}

56.6.3 Como interpretar a comparação de subgrupos em ensaios experimentais aleatorizados?

- Análises de subgrupos devem ser consideradas de natureza exploratória e raramente elas afetam as conclusões obtidas a partir do estudo.^{431,440}
- A credibilidade das análises de subgrupos é melhor se restrita ao desfecho primário e a alguns subgrupos predefinidos e baseadas em hipóteses biologicamente plausíveis.⁴³¹
- 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.⁴³¹

56.7 Efeito de interação

56.7.1 Por que analisar o efeito de interação?

- Em ensaios experimentais aleatorizados, o principal problema de pesquisa é se há uma diferença pré - pós maior em um grupo do que em outro(s).⁴²⁴
- 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.⁴⁴¹
- 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.³³⁸

56.7.2 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).³³⁹
- A interação entre duas (ou mais) variáveis pode ser utilizada para comparar efeitos do tratamento em subgrupos de ensaios experimentais.⁴⁴²
- O poder estatístico para detectar efeitos de interação é limitado.⁴⁴²

56.8 Ajuste de covariáveis

56.8.1 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.⁴³⁷

56.8.2 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.⁴⁴³
- 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.⁴⁴⁴
- 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.⁴⁴⁴

56.8.3 Quais os riscos do ajuste de covariáveis?

- Incluir covariáveis que não são prognósticas do desfecho pode reduzir o poder estatístico do estudo.⁴⁴⁴
- 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.⁴⁴⁴

56.9 Imputação de dados perdidos

56.9.1 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.^{158,445}
- Para dados perdidos em desfechos dicotômicos, o desempenho dos métodos de imputação multivariada por equações encadeadas (*multivariate imputation by chained equations*, MICE)¹⁶⁵ e por correspondência média preditiva (*predictive mean matching*, PMM)^{166,167} é similar.¹⁶⁴

56.9.2 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.⁴⁴⁴
- 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.⁴³⁰



Os pacotes *mice*¹⁶⁵ e *miceadds*¹⁶⁸ fornecem funções *mice*^a e *mi.anova*^b para imputação multi-variada 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>

56.10 Diretrizes para redação

56.10.1 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*:⁴⁴⁶ <https://www.equator-network.org/reporting-guidelines/consort/>



O pacote *consort*⁴⁴⁷ 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 57

Ensaios cruzados

57.1 Características

57.1.1 Quais são as características de ensaios cruzados?

- ?

57.2 Diretrizes para redação

57.2.1 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:*⁴⁴⁸ <https://www.equator-network.org/reporting-guidelines/consort-2010-statement-extension-to-randomised-crossover-trials/>

¹<https://www.equator-network.org/>

RASCUNHO

Capítulo 58

n-de-1

58.1 Ensaio n-de-1

58.1.1 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).⁴⁴⁹
- Cada ciclo é formado por dois períodos (AB ou BA), cuja ordem é randomizada, garantindo controle temporal e redução de vieses.⁴⁴⁹
- O foco está na comparação intraindivíduo, permitindo avaliar diretamente se o paciente em questão responde melhor a uma intervenção.⁴⁴⁹

58.1.2 Quando é apropriado conduzir ensaios n-de-1?

- Doenças crônicas estáveis, em que o desfecho pode ser observado repetidamente.⁴⁴⁹
- Condições raras ou com grande heterogeneidade de resposta entre pacientes.⁴⁴⁹
- Situações clínicas de incerteza, quando se deseja personalizar o tratamento.⁴⁴⁹
- Ensaios n-de-1 devem ser utilizados principalmente para apoiar decisões clínicas sobre o cuidado de um paciente individual, e não para estimar efeitos médios em populações.⁴⁵⁰
- Eles são uma opção viável quando se lidam com intervenções de muito baixo volume, como em doenças raras ou ultrarraras, nas quais ensaios tradicionais são inviáveis.⁴⁵⁰
- Uma ampla variedade de tecnologias em saúde pode ser avaliada por ensaios n-de-1, incluindo medicamentos, dispositivos, intervenções dietéticas e comportamentais, desde que atendam aos critérios de início de efeito e ausência de efeitos residuais prolongados.⁴⁵⁰
- Para que uma tecnologia seja adequada a um ensaio n-de-1, seu efeito deve surgir em um intervalo de tempo que possa ser medido dentro de um período de estudo factível.⁴⁵⁰
- Tecnologias avaliadas em ensaios n-de-1 não devem apresentar efeitos de “carryover” prolongados; quando isso não é possível, é necessário empregar estratégias adequadas de washout.⁴⁵⁰
- Ensaios n-de-1 são particularmente úteis para avaliar tecnologias caras ou com efeitos adversos relevantes, permitindo verificar se, para aquele paciente específico, os benefícios superam os riscos e os custos.⁴⁵⁰

- Esses ensaios podem responder perguntas sobre se o tratamento funciona, se funciona melhor do que alternativas existentes, qual tratamento é mais adequado para um paciente específico ou se a resposta ao tratamento varia entre indivíduos.⁴⁵⁰

58.1.3 Qual a relevância dos ensaios n-de-1?

- Os ensaios n-de-1 permitem decisões clínicas personalizadas e baseadas em evidência direta.⁴⁴⁹
- Quando combinados, podem gerar estimativas comparáveis às de ensaios clínicos convencionais, mantendo o foco centrado no paciente.⁴⁴⁹
- Representam uma alternativa metodológica robusta para cenários de incerteza terapêutica.⁴⁴⁹

58.2 Aspectos metodológicos

58.2.1 Quais são os principais aspectos do delineamento de ensaios n-de-1?

- A escolha do desfecho primário deve ser guiada diretamente pela pergunta clínica que o ensaio pretende responder.⁴⁵⁰
- Sempre que possível, recomenda-se combinar desfechos relatados pelo paciente com medidas mais objetivas, especialmente em ensaios envolvendo tratamentos caros ou de maior risco.⁴⁵⁰
- Ensaios n-de-1 podem avaliar não apenas eficácia, mas também outros desfechos relevantes para o paciente, como efeitos adversos, qualidade de vida e preferências pessoais.⁴⁵⁰
- O comparador do tratamento deve ser escolhido de acordo com a pergunta de pesquisa, podendo ser placebo, tratamento ativo ou cuidado padrão.⁴⁵⁰
- O foco do tratamento em um ensaio n-de-1 pode ser a doença em si, sintomas específicos, efeitos colaterais ou a satisfação do paciente com a intervenção.⁴⁵⁰
- Na prática, a maioria dos ensaios n-de-1 compara duas tecnologias; a comparação de três ou mais aumenta significativamente a complexidade operacional.⁴⁵⁰
- O número de períodos do estudo deve equilibrar precisão e viabilidade, considerando duração total do ensaio e carga para o paciente.⁴⁵⁰
- O cegamento deve ser utilizado sempre que viável; quando não for possível, outras formas de minimizar viés, como avaliação cega de desfechos, são recomendadas.⁴⁵⁰
- A randomização da sequência de tratamentos é recomendada, preferencialmente em blocos, para reduzir vieses e efeitos relacionados ao tempo.⁴⁵⁰
- Análises interinas podem ser consideradas, permitindo interromper o ensaio precocemente caso haja evidência suficiente de benefício ou dano.⁴⁵⁰
- Quando há risco de efeitos residuais entre períodos, devem ser empregados washout tradicionais ou washout ativo, sempre ponderando rigor metodológico e segurança do paciente.⁴⁵⁰
- A interpretação dos resultados não deve se basear apenas em significância estatística; a relevância clínica do efeito é essencial.⁴⁵⁰
- A análise central de um ensaio n-de-1 é a análise intra-paciente, que avalia se houve um efeito clinicamente importante para aquele indivíduo.⁴⁵⁰
- Quando vários ensaios n-de-1 são conduzidos, seus resultados podem ser combinados para estimar efeitos médios e avaliar a consistência entre pacientes.⁴⁵⁰

- O envolvimento de pacientes e do público é fundamental em todas as etapas do ensaio n-de-1, incluindo planejamento, definição de desfechos, interpretação, disseminação e implementação dos resultados.⁴⁵⁰

58.2.2 Quantos períodos são usualmente utilizados?

- A mediana observada em ensaios recentes foi de 6 períodos (intervalo interquartil 4–8).⁴⁵¹
- Esse número parece representar um equilíbrio entre maior precisão estatística (mais períodos) e viabilidade prática e adesão do paciente (menos períodos).⁴⁵¹

58.2.3 Como é feita a randomização?

- A ordem dos tratamentos em cada ciclo é definida aleatoriamente (ex.: AB, BA, AB...).⁴⁴⁹
- A randomização reduz viés de período e efeito de expectativa.⁴⁴⁹

58.2.4 Quais perguntas de inferência podem ser respondidas?

- Q1: Há efeito do tratamento dentro dos ciclos de um paciente?⁴⁴⁹
- Q2: Qual é o efeito médio observado nos pacientes estudados?⁴⁴⁹
- Q3: O efeito é homogêneo ou heterogêneo entre pacientes?⁴⁴⁹
- Q4: Qual é o efeito específico em cada paciente individual?⁴⁴⁹
- Q5: Qual é o efeito esperado em populações semelhantes?⁴⁴⁹

58.3 Análise de dados

58.3.1 Como são feitas as análises?

- Comparações intraindivíduo (testes pareados ou estimativas de efeito médio por paciente).⁴⁴⁹
- Combinação de múltiplos n-de-1 por meio de meta-analises ou modelos mistos para inferências em nível populacional.⁴⁴⁹

58.3.2 Quais métodos estatísticos têm sido utilizados recentemente?

- Diversos métodos têm sido aplicados, incluindo testes *t*, modelos de regressão, modelos Bayesianos, testes não paramétricos, modelos mistos, e inspeção gráfica.⁴⁵¹

58.3.3 É adequado utilizar P-valores em ensaios n-de-1?

- Em ensaios n-de-1 a interpretação deve ser cautelosa, pois um P-valor não significativo não exclui a presença de efeito clinicamente relevante.⁴⁵¹
- Estimativas pontuais e intervalos de confiança podem ser mais informativos para decisões clínicas individuais.⁴⁵¹

58.3.4 Como os dados de ensaios n-de-1 podem ser visualizados?

- *Trellis plot*, no qual cada painel representa um paciente e cada ponto corresponde a um ciclo, plotando-se o desfecho sob B no eixo Y e sob A no eixo X. A linha de identidade indica igualdade entre os tratamentos; a concentração de pontos acima e à esquerda dessa linha sugere maior efeito do tratamento B.⁴⁴⁹

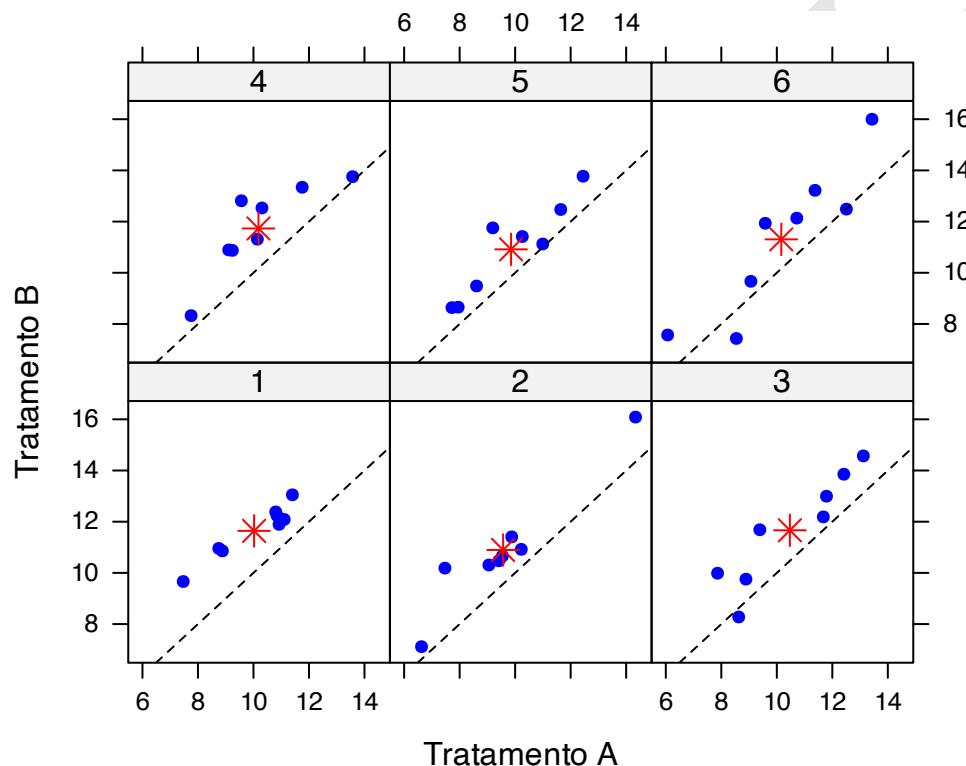


Figura 58.1: Trellis plot para dados simulados de ensaio n-de-1. Cada painel representa um paciente, cada ponto azul um ciclo, e o asterisco vermelho indica a média dos ciclos para aquele paciente. A linha tracejada representa a linha de identidade (igualdade entre tratamentos).



O pacote *lattice*⁴⁵² fornece a função *xyplot*^a para criar gráficos *trellis plot*.

^ahttps://www.rdocumentation.org/packages/lattice/versions/0.22-7/topics/B_01_xyplot.ts

58.3.5 Quais métodos estatísticos são usados na análise?

- A hipótese nula assume ausência de diferença entre os tratamentos para todos os pacientes, implicando independência das diferenças entre A e B e permitindo a agregação dos dados entre ciclos e indivíduos.⁴⁴⁹
- Os dados são reduzidos a diferenças emparelhadas por ciclo e paciente, resultando em uma única amostra de diferenças.⁴⁴⁹
- As diferenças observadas podem ser analisadas por meio de um teste t de uma amostra ou Mann-Whitney-Wilcoxon.⁴⁴⁹

58.4 Abordagem meta-analítica

58.4.1 Como um conjunto de ensaios n-de-1 pode ser analisado conjuntamente?

- Ele pode ser tratado de forma análoga a uma meta-análise de ensaios clínicos independentes, utilizando métodos e softwares padrão com pequenas adaptações.⁴⁴⁹
- Devido ao pequeno número de graus de liberdade por paciente, a estimativa ingênua de variâncias é inadequada, sendo necessário assumir variância intraindivíduo constante entre pacientes e utilizar uma variância agrupada.⁴⁴⁹
- A variância agrupada é estimada a partir da soma dos quadrados corrigida das estimativas individuais, dividida pelo total de graus de liberdade, produzindo uma estimativa comum da variância intraindivíduo.⁴⁴⁹

58.4.2 Por que a meta-análise com efeitos aleatórios produz intervalos de confiança mais amplos?

- Porque ela estima um efeito de tratamento aplicável a uma população mais ampla e “semelhante”, tratando os pacientes estudados como uma amostra aleatória e incorporando a variabilidade entre indivíduos à incerteza do efeito estimado.⁴⁴⁹

58.4.3 Qual é a principal vantagem da abordagem meta-analítica?

- A vantagem é responder a uma pergunta inferencial mais relevante, sobre efeitos esperados em populações futuras.⁴⁴⁹

58.5 Limitações e cuidados

58.5.1 Quais são os principais desafios dos ensaios n-de-1?

- Baixo poder estatístico quando poucos ciclos são realizados.⁴⁴⁹
- Necessidade de períodos de *washout* para evitar efeito de *carry-over*.⁴⁴⁹
- Interpretação dependente de pressupostos sobre homogeneidade ou heterogeneidade dos efeitos.⁴⁴⁹
- Em amostras muito pequenas, pode ser necessário usar variâncias externas ou modelos mistos.⁴⁴⁹
- A abordagem meta-analítica assume forte similaridade entre os pacientes estudados e a população-alvo.⁴⁴⁹

RASCUNHO

Capítulo 59

Revisão sistemática

59.1 Tipologia de revisões

59.1.1 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.³⁸¹

59.2 Revisão sistemática de literatura

59.2.1 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.^{381,453}

R

O pacote *easyPubMed*⁴⁵⁴ 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*⁴⁵⁵ fornece a função *cr_works*^a para buscar artigos.

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

R

O pacote *roadoi*⁴⁵⁶ 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>

59.3 Tipos de revisão sistemática

59.3.1 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.^{381,453}
- *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.^{381,453,457,458}
- *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.^{381,459}
- 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.^{381,453}
- 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.^{381,453}
- 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.^{381,453}

59.3.2 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.^{381,453}
- 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.^{381,453}
- 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.^{381,453}

59.4 Diretrizes para redação

59.4.1 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)*.⁴⁶⁰ <https://www.equator-network.org/reporting-guidelines/tripod-srma/>
- *Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist*.⁴⁶¹ <https://doi.org/10.1371/journal.pmed.1001744>

¹<https://www.equator-network.org/>

RASCUNHO

Capítulo 60

Meta-análise

60.1 Características

60.1.1 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.⁴⁶²
- 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.⁴⁶²

60.2 Modelos de meta-análise

60.2.1 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.⁴⁶²
- 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.⁴⁶²
- 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.⁴⁶²

O pacote *metafor*⁴⁶³ 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*⁴⁶⁴ 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*⁴⁶⁵ 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

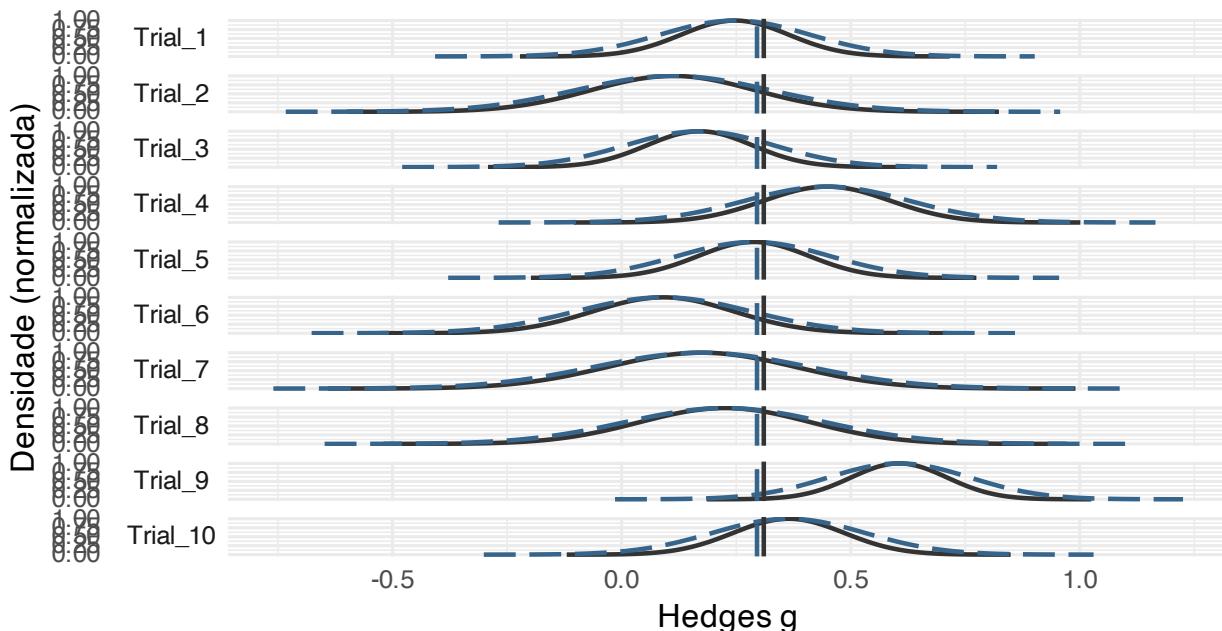


Figura 60.1: Comparação entre modelos de efeito fixo e aleatório com 10 ensaios clínicos simulados.

60.3 Conversão de Medidas em Meta-análises

60.3.1 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.⁴⁶⁷
- Quando apresentam mediana e intervalo interquartil (ou mínimo–máximo), existem métodos estatísticos para converter em média e DP.⁴⁶⁶
- Hozo et al. (2005) propuseram fórmulas para estimar a média e o desvio-padrão a partir da mediana, amplitude e tamanho da amostra.⁴⁶⁶

- Wan et al. (2014) aperfeiçoaram essas estimativas, oferecendo métodos mais precisos para converter mediana e IQR em média e DP.⁴⁶⁷



O pacote *metafor*⁴⁶³ 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>

60.4 Interpretação de efeitos em meta-análise

60.4.1 Como avaliar a variação do tamanho do efeito?

- O intervalo de predição contém informação sobre a variação do tamanho do efeito.⁴⁶⁸
- 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.⁴⁶⁸
- 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.⁴⁶⁸

60.4.2 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 (60.1).^{468,469}

$$I^2 = \max \left(0, \frac{Q - df}{Q} \right) \times 100\% \quad (60.1)$$

- I^2 representa qual proporção da variância observada reflete a variância nos efeitos verdadeiros em vez do erro de amostragem.⁴⁶⁸
- I^2 mede a proporção da variância total que pode ser atribuída à heterogeneidade entre os estudos incluídos.⁴⁶⁹
- 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.⁴⁶⁹
- 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.⁴⁷⁰



O pacote *psychmeta*³⁰⁹ 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*³⁰⁹ fornece a função *ma_r*^a para meta-analisar correlações.

^ahttps://www.rdocumentation.org/packages/psychmeta/versions/2.7.0/topics/ma_r

60.5 Forest plot

60.5.1 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.⁴⁷¹

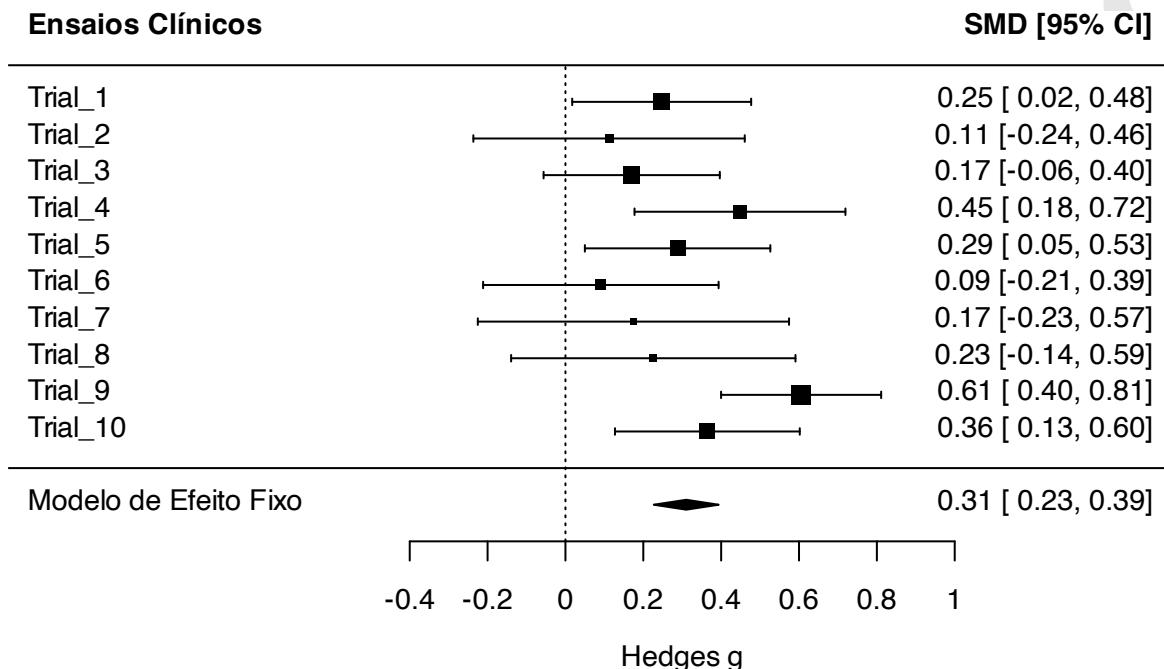


Figura 60.2: Forest plot de uma meta-análise de efeito fixo com 10 ensaios clínicos simulados.

60.5.2 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.⁴⁷¹

60.5.3 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*).⁴⁷¹
- Em desfechos contínuos, apresentam-se médias, desvios-padrão e tamanhos amostrais, com o efeito medido pela diferença de médias.⁴⁷¹

60.5.4 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).⁴⁷¹
- O tamanho da caixa é proporcional ao peso do estudo na meta-análise, geralmente maior para estudos com amostras maiores.⁴⁷¹

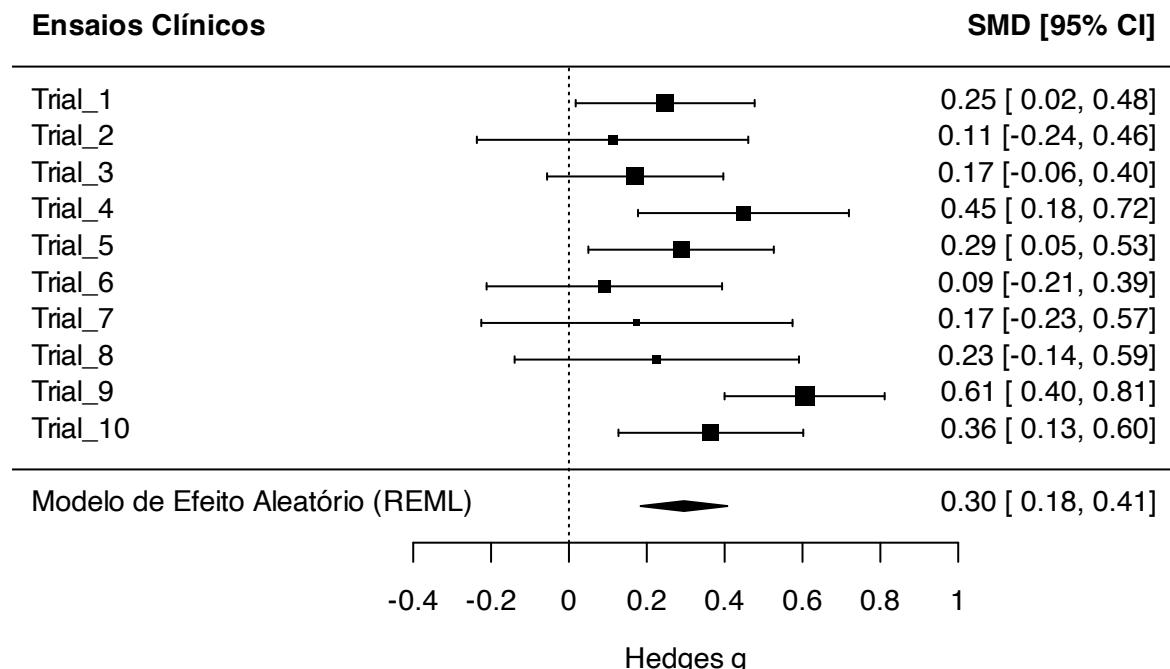


Figura 60.3: Forest plot de uma meta-análise de efeito aleatório com 10 ensaios clínicos simulados.

60.5.5 Qual é o significado da linha vertical do “nenhum efeito”?

- É a linha de referência que indica efeito nulo.⁴⁷¹
- Para desfechos binários, corresponde ao valor 1 (RR ou $OR = 1$).⁴⁷¹
- Para desfechos contínuos, corresponde ao valor 0 (diferença de médias = 0).⁴⁷¹
- Se o intervalo de confiança de um estudo ou do resultado combinado cruza essa linha, o resultado não é estatisticamente significativo.⁴⁷¹

60.5.6 Como interpretar o diamante na parte inferior do *forest plot*?

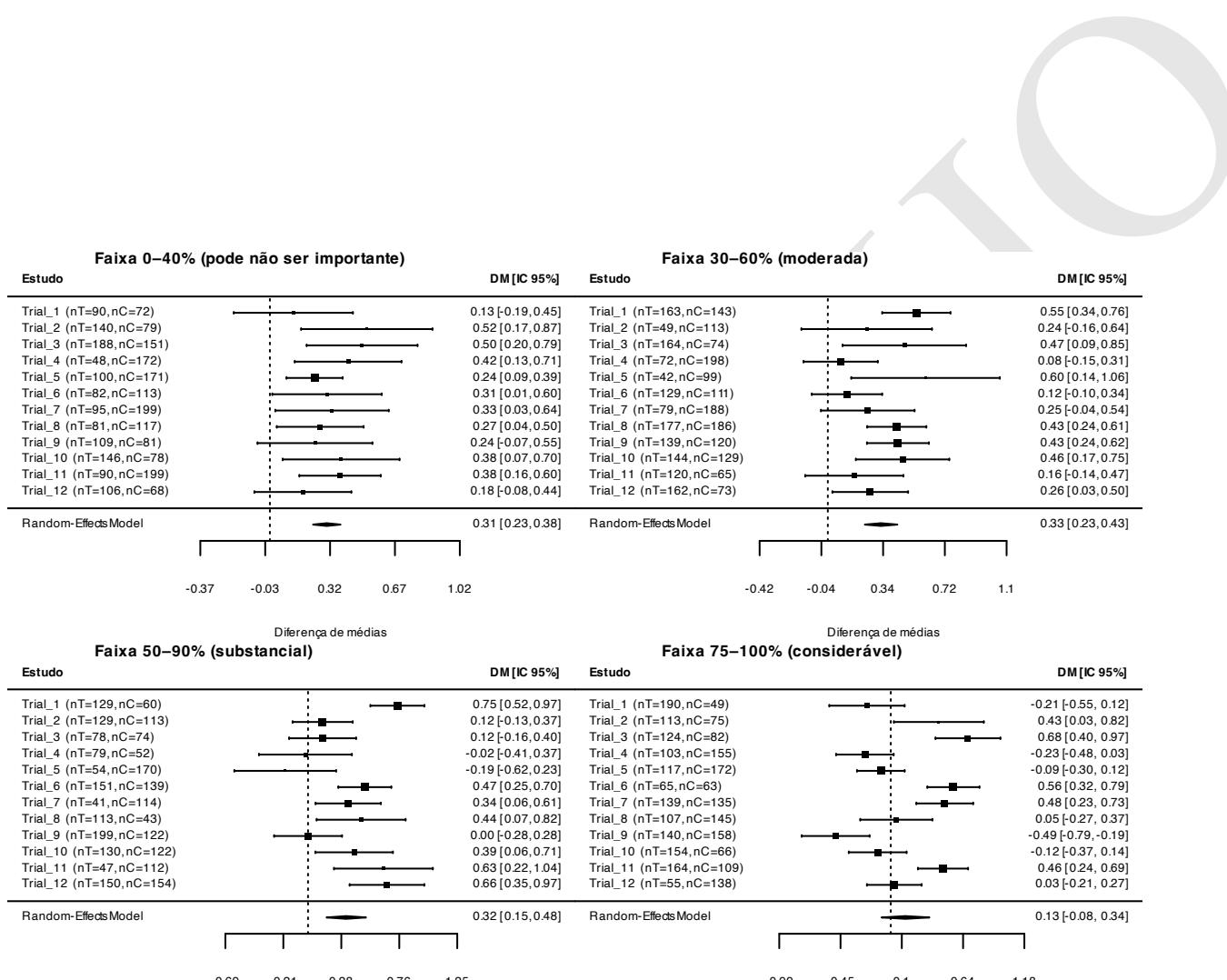
- O diamante representa o efeito combinado dos estudos incluídos.⁴⁷¹
- O ponto central do diamante é a estimativa global.⁴⁷¹
- A largura do diamante representa o intervalo de confiança de 95% para o efeito combinado.⁴⁷¹

60.5.7 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 .⁴⁷¹

60.5.8 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.⁴⁷¹

Figura 60.4: Forest plots ilustrativos para faixas usuais de I^2 .

60.6 *Crosshair*

60.6.1 O que um gráfico de cruzes (*crosshair*)?

• ⁴⁷²

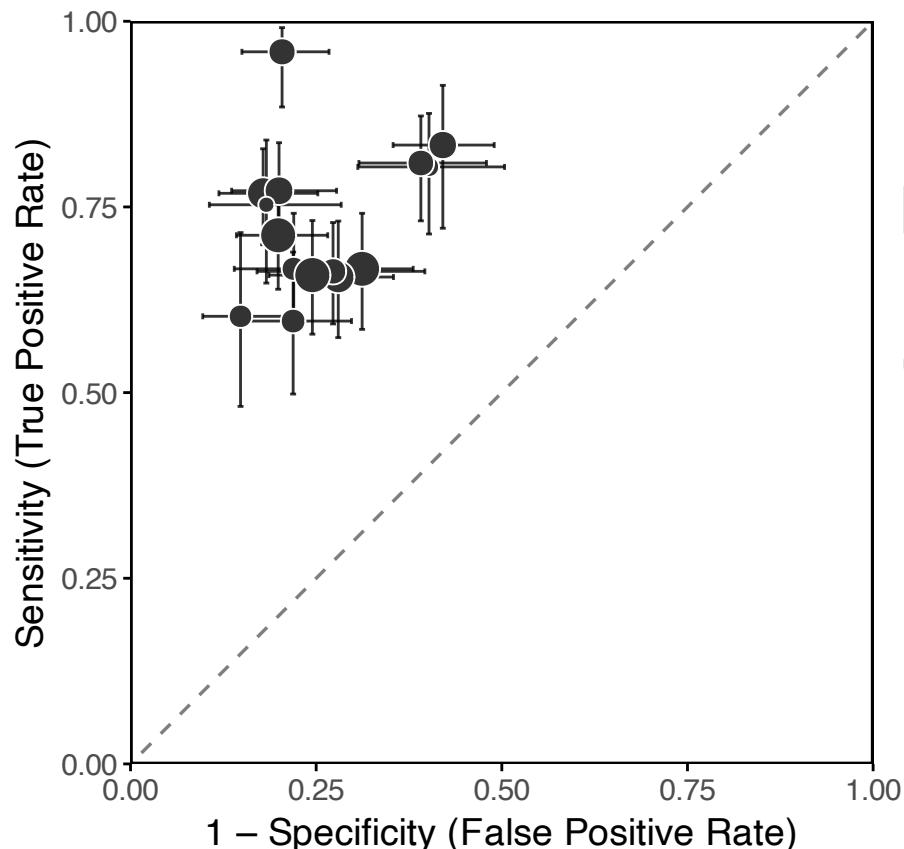


Figura 60.5: Gráfico de cruzes (*crosshair*) em espaço ROC (*Receiver Operating Characteristic*) para 15 estudos simulados de desempenho diagnóstico.



O pacote *mada*⁴⁷³ fornece a função *crosshair*^a para criar um gráfico *crosshair*⁴⁷² 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>

60.7 *Funnel plot*

60.7.1 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”.⁴⁷⁴

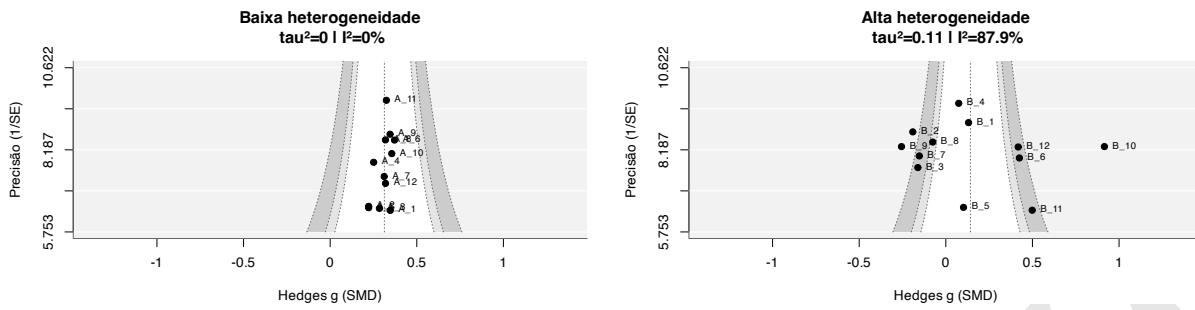


Figura 60.6: Gráficos de funil simulados com baixa e alta heterogeneidade.

60.7.2 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.⁴⁷⁵

60.7.3 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 ($1/SE$).⁴⁷⁶
- 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*).⁴⁷⁶
- 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.⁴⁷⁷

60.7.4 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.⁴⁷⁴
- Diferenças metodológicas entre estudos pequenos e grandes.⁴⁷⁴
- Heterogeneidade verdadeira (diferença real no efeito conforme o tamanho ou o contexto do estudo).⁴⁷⁴
- Artefatos estatísticos ou mero acaso.⁴⁷⁴

60.7.5 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.⁴⁷⁸
- O método assume que a assimetria é causada unicamente por viés de publicação, podendo levar a conclusões equivocadas quando há outras causas, como heterogeneidade.⁴⁷⁸

60.8 Testes de assimetria do *funnel plot*

60.8.1 O que é o teste de Egger?

- É um teste estatístico amplamente utilizado que avalia a relação entre o efeito padronizado ($e\!f\!e\!i\!t\!o\!/SE$) e a precisão ($1/SE$).⁴⁷⁶

- No entanto, para meta-análises com *OR*, apresenta taxas de erro tipo I excessivas, especialmente quando o efeito é grande ou há alta heterogeneidade.⁴⁷⁷

60.8.2 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).⁴⁷⁷
- 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.⁴⁷⁷
- 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.⁴⁷⁷
- Em casos de alta heterogeneidade, o teste de Egger pode apresentar maior poder, mas sua alta taxa de falsos positivos compromete a interpretação.⁴⁷⁷

60.8.3 Quais são as recomendações para testar a assimetria?

- Evitar testes quando há menos de 10 estudos, devido ao baixo poder estatístico.⁴⁷⁴
- Inspecionar visualmente o *funnel plot* junto com os resultados dos testes.⁴⁷⁴
- Para desfechos contínuos (diferença de médias), o teste de Egger pode ser usado.⁴⁷⁴
- Para desfechos dicotômicos expressos como *odds ratio* (*OR*) com baixa heterogeneidade ($\tau^2 < 0,1$), considerar os testes de Harbord, Peters ou Rücker.⁴⁷⁴
- 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.⁴⁷⁴

60.8.4 Como interpretar os resultados de testes de viés de publicação?

- Um resultado não significativo não garante ausência de viés.⁴⁷⁷
- Recomenda-se complementar com inspeção visual do *funnel plot* e considerar métodos adicionais como *trim and fill*.^{477,478}
- Quando há suspeita de viés, discutir as implicações e interpretar o efeito combinado com cautela.⁴⁷⁷



O pacote *psychmeta*³⁰⁹ 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

60.9 Diretrizes para redação

60.9.1 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:*⁴⁷⁹ <https://www.equator-network.org/reporting-guidelines/prisma/>

¹<https://www.equator-network.org/>

- *Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA)*.⁴⁶⁰ <https://www.equator-network.org/reporting-guidelines/tripod-srma/>

R

O pacote *metagear*⁴⁸⁰ 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*⁴⁸¹.

^ahttps://www.rdocumentation.org/packages/metagear/versions/0.7/topics/plot_PRISMA

R

O pacote *PRISMA2020*⁴⁸² 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

Capítulo 61

Revisão guarda-chuva

61.1 Revisão guarda-chuva

61.1.1 O que é revisão guarda-chuva?

• ?

61.2 Diretrizes para redação

61.2.1 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.*⁴⁸³ <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 62

Pesquisa qualitativa

62.1 Pesquisa qualitativa

62.1.1 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.⁴⁸⁴
- 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.⁴⁸⁴

62.2 Diretrizes para redação

62.2.1 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:*⁴⁸⁵ <https://www.equator-network.org/reporting-guidelines/srqr/>
- *Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ:*⁴⁸⁶ <https://www.equator-network.org/reporting-guidelines/entreq/>
- *Consolidated criteria for reporting qualitative research (COREQ) : a 32-item checklist for interviews and focus groups:*⁴⁸⁷ <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 63

Redação de resultados

63.1 Comunicação de resultados da análise estatística

63.1.1 Como devo comunicar probabilidades e riscos?

- Sempre que possível, utilize números, mesmo que aproximados.⁴⁸⁸
- A comunicação exclusivamente verbal (por exemplo, “raro” ou “comum”) é ambígua e interpretada de maneiras muito diferentes pelos leitores.⁴⁸⁸
- Os números permitem compreender a ordem de grandeza do risco, facilitam comparações e reduzem interpretações equivocadas.⁴⁸⁸
- Quando apropriado, a combinação de números com descrições verbais pode ser usada para melhorar a compreensão.⁴⁸⁸

63.1.2 Qual é o formato numérico mais adequado para apresentar riscos?

- Prefira percentuais ou taxas com denominadores fixos, como por 100 ou por 1.000 indivíduos. Esse padrão facilita a leitura e a comparação entre grupos.⁴⁸⁸
- Evite o formato “1 em X”, pois ele tende a inflar a percepção de risco e dificulta a avaliação comparativa.⁴⁸⁸

63.1.3 Como devo apresentar diferenças entre grupos ou intervenções?

- Dê prioridade às diferenças absolutas.⁴⁸⁸
- Apresente as taxas observadas em cada grupo e, em seguida, a diferença entre elas. Esse formato responde melhor às perguntas práticas do leitor, como “quanto isso realmente muda o risco?”.⁴⁸⁸
- Medidas relativas podem ser incluídas como complemento, mas não devem ser apresentadas isoladamente.⁴⁸⁸

63.1.4 Como evitar distorções na interpretação de benefícios e riscos?

- Utilize o mesmo tipo de medida para benefícios e para danos.⁴⁸⁸
- Apresentar benefícios como reduções relativas e riscos como aumentos absolutos pode induzir uma falsa impressão de vantagem.⁴⁸⁸
- A simetria na apresentação favorece interpretações mais equilibradas e transparentes.⁴⁸⁸

63.1.5 Como devo usar gráficos para apresentar probabilidades e efeitos?

- Escolha gráficos que evidenciem a relação parte–todo, mostrando claramente numerador e denominador.⁴⁸⁸
- Esses formatos ajudam o leitor a compreender a magnitude real do risco ou do efeito.⁴⁸⁸
- Evite gráficos que apresentem apenas o numerador ou utilizem escalas truncadas, pois podem exagerar visualmente diferenças pequenas.⁴⁸⁸

63.1.6 O que fazer ao apresentar valores pouco familiares ao leitor?

- Inclua intervalos de referência, valores-alvo, pontos de corte ou limiares de decisão. Isso permite que o leitor compreenda não apenas o valor numérico, mas também sua relevância prática ou clínica.⁴⁸⁸

63.1.7 Qual é o objetivo final da seção de resultados estatísticos?

- Permitir que o leitor entenda, de forma clara e honesta: o tamanho do efeito, a direção da associação e a importância prática dos achados.⁴⁸⁸
- A significância estatística, por si só, não é suficiente sem uma comunicação adequada da magnitude e do contexto dos resultados.⁴⁸⁸

R

O pacote *report*⁴⁸⁹ 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>

R

O pacote *statcheck*⁴⁹⁰ 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>

63.2 Diretrizes e Listas

63.2.1 Quais diretrizes estão disponíveis para redação estatística?

- *Review of guidance papers on regression modeling in statistical series of medical journals.*⁴⁹¹
- *Principles and recommendations for incorporating estimands into clinical study protocol templates.*⁴⁹²
- *How to write statistical analysis section in medical research.*²⁶²
- *Recommendations for Statistical Reporting in Cardiovascular Medicine: A Special Report From the American Heart Association.*⁴⁹³
- *Framework for the treatment and reporting of missing data in observational studies: The Treatment And Reporting of Missing data in Observational Studies framework.*⁴⁹⁴
- *Guidelines for reporting of figures and tables for clinical research in urology.*⁴⁹⁵
- *Who is in this study, anyway? Guidelines for a useful Table 1.*²²⁵
- *Guidelines for Reporting of Statistics for Clinical Research in Urology.*⁴⁹⁶
- *Reveal, Don't Conceal: Transforming Data Visualization to Improve Transparency.*²¹⁴
- *Guidelines for the Content of Statistical Analysis Plans in Clinical Trials.*³¹

- Basic statistical reporting for articles published in Biomedical Journals: The ‘‘Statistical Analyses and Methods in the Published Literature’’ or the SAMPL Guidelines.⁴⁹⁷
- Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm.⁴⁹⁸
- STRengthening analytical thinking for observational studies: the STRATOS initiative.⁴⁹⁹
- Research methods and reporting.⁵⁰⁰
- How to ensure your paper is rejected by the statistical reviewer.⁵⁰¹

63.2.2 Quais listas de verificação estão disponíveis para redação estatística?

- A CHecklist for statistical Assessment of Medical Papers (the CHAMP statement): explanation and elaboration.⁵⁰²
- Checklist for clinical applicability of subgroup analysis.⁵⁰³
- Evidence-based statistical analysis and methods in biomedical research (SAMBR) checklists according to design features.²⁶¹

RASCUNHO

Capítulo 64

Diretrizes e Listas

64.1 Diretrizes

64.1.1 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.*²⁵¹
- *Review of guidance papers on regression modeling in statistical series of medical journals.*⁴⁹¹
- *Principles and recommendations for incorporating estimands into clinical study protocol templates.*⁴⁹²
- *How to write statistical analysis section in medical research.*²⁶²
- *A Guideline for Reporting Mediation Analyses of Randomized Trials and Observational Studies: The AGReMA Statement.*⁵⁰⁴
- *Recommendations for Statistical Reporting in Cardiovascular Medicine: A Special Report From the American Heart Association.*⁴⁹³
- *Framework for the treatment and reporting of missing data in observational studies: The Treatment And Reporting of Missing data in Observational Studies framework.*⁴⁹⁴
- *Guidelines for reporting of figures and tables for clinical research in urology.*⁴⁹⁵
- *Who is in this study, anyway? Guidelines for a useful Table 1.*²²⁵
- *Guidelines for Reporting of Statistics for Clinical Research in Urology.*⁴⁹⁶
- *Reveal, Don't Conceal: Transforming Data Visualization to Improve Transparency.*²¹⁴
- *Guidelines for the Content of Statistical Analysis Plans in Clinical Trials.*³¹
- *Basic statistical reporting for articles published in Biomedical Journals: The ‘‘Statistical Analyses and Methods in the Published Literature’’ or the SAMPL Guidelines.*⁴⁹⁷
- *Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm.*⁴⁹⁸
- *STRengthening analytical thinking for observational studies: the STRATOS initiative.*⁴⁹⁹
- *Research methods and reporting.*⁵⁰⁰

¹<https://www.equator-network.org/>

- *How to ensure your paper is rejected by the statistical reviewer.*⁵⁰¹

64.2 Listas de verificação

64.2.1 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.*⁵⁰²
- *Checklist for clinical applicability of subgroup analysis.*⁵⁰³
- *Evidence-based statistical analysis and methods in biomedical research (SAMBR) checklists according to design features.*²⁶¹

REFERÊNCIAS

RASCUNHO

Produção do autor

Artigos em periódicos científicos

1. Barros I de A, Ferreira A de S, Horsth T de S, Holmes T de J, dos Santos AA, Lunkes LC. Type of health locus of control predicting pain, function, and global perceived effect in patients with chronic low back pain receiving active versus passive interventions: an observational study. *Brazilian Journal of Physical Therapy.* 2026;30(2):101560. doi:10.1016/j.bjpt.2025.101560
2. Ferreira A de S, Parisotto G. Insights from Ungureanu et al.: interpreting AI-clinician disagreement in ECG diagnostics. *Acta Cardiologica.* February 2026;1-3. doi:10.1080/00015385.2026.2630134
3. Ferreira A de S, dos Anjos FV. Applying Symmetry to Motor Control in Sports and Rehabilitation. *Symmetry.* 2026;18(1):166. doi:10.3390/sym18010166
4. Petrov I, Stoichev K, Aliman O, et al. Effects of differing nutritional supplementation combined with high-intensity aerobic interval training on functional exercise capacity, cardiac function, and quality of life in patients with heart failure and reduced ejection fraction: a randomized trial. *American Journal of Physical Medicine & Rehabilitation.* December 2025. doi:10.1097/phm.0000000000002932
5. Amaravadi SK, Ferreira A de S, Vigário P dos S. Comparative effects of combined aerobic and resistance training versus high-intensity interval training on insulin resistance, glycaemic control, body composition and quality of life in type 2 diabetes: A 12-week randomised controlled trial. Cobucci RNO, ed. *PLOS One.* 2025;20(12):e0336898. doi:10.1371/journal.pone.0336898
6. Parisotto G, Sant'Anna Junior M, Papathanasiou J, Reis LF da F, Ferreira A de S. Functional determinants and perceived barriers to cardiac rehabilitation as predictors of short-term hospital readmission in acute coronary syndrome: an observational longitudinal cohort study. *Acta Cardiologica.* 2025;80(10):1102-1111. doi:10.1080/00015385.2025.2576437
7. Correia IMT, Ferreira A de S, Gomes JFM, Reis FJJ, Nogueira LAC, Meziat-Filho N. Cervical flexion posture during smartphone use was not a risk factor for neck pain, but low sleep quality and insufficient levels of physical activity were. A longitudinal investigation. *Brazilian Journal of Physical Therapy.* 2025;29(6):101258. doi:10.1016/j.bjpt.2025.101258
8. de Souza Horsth T, de Araujo Barros I, de Souza RC, de Sá Ferreira A, de Almeida RS. Heart rate variability responses to instrument-assisted atlas (C1) chiropractic manipulation: A randomized placebo-controlled trial. *Journal of Bodywork and Movement Therapies.* 2025;44:784-788. doi:10.1016/j.jbmt.2025.06.027
9. Ferreira ALNV, Ferreira A de S. POP do “quarto terapêutico” proposto pela equipe profissional aos pacientes com neoplasia de tireoide com Iodoradioativo I-131. *Revista JRG de Estudos Acadêmicos.* 2025;8(18):e082222. doi:10.55892/jrg.v8i18.2222
10. Monteiro ER, Aguilera LM, Ruá-Alonso M, et al. Effect of Manual Massage, Foam Rolling, and Strength Training on Hemodynamic and Autonomic Responses in Adults: A Scoping Review. *Healthcare.* 2025;13(12):1371. doi:10.3390/healthcare13121371

11. Monteiro ER, de Oliveira Muniz Cunha JC, de Souza Horsth T, et al. Effect of cervical manipulation on blood pressure and heart rate variability responses in adults: A scoping review. *Journal of Bodywork and Movement Therapies.* 2025;42:1120-1127. doi:10.1016/j.jbmt.2025.03.023
12. Resende PA, Tavares Correia IM, de Sá Ferreira A, Meziat-Filho N, Lunkes LC. Neck pain and text neck using Hill's criteria of causation: A scoping review. *Journal of Bodywork and Movement Therapies.* 2025;42:132-138. doi:10.1016/j.jbmt.2024.12.016
13. Ferreira AS, Parisotto G. Rethinking methodologies in nocturnal blood pressure dipping research: insights from Lopez et al. *Acta Cardiologica.* 2025;80(5):529-530. doi:10.1080/00015385.2025.2453799
14. Da Costa CCM, Olímpio Júnior H, Da Silva Pinto PVL, et al. Contribution of small airway disease to dynamic hyperinflation in patients with chronic obstructive pulmonary disease. *Monaldi Archives for Chest Disease.* May 2025. doi:10.4081/monaldi.2025.3402
15. Reis FJJ, Neves G de A, Carvalho MBL de, et al. Mapping global research on artificial intelligence in physical therapy: a bibliometric analysis from 1990 to 2023. *European Journal of Physiotherapy.* May 2025;1-11. doi:10.1080/21679169.2025.2497780
16. Ferreira ALNV, De Sá Ferreira A. Contribuições do POP à assistência de enfermagem ao tratamento da neoplasia de tireoide no quarto terapêutico. *Revista JRG de Estudos Acadêmicos.* 2025;8(18):e082001. doi:10.55892/jrg.v8i18.2001
17. Shanmugam S, Anjos FV dos, Ferreira A de S, Muthukrishnan R, Kandakurti PK, Durairaj S. Effectiveness of intramuscular electrical stimulation using conventional and inverse electrode placement methods on pressure pain threshold and electromyographic activity of the upper trapezius muscle with myofascial trigger points: a randomized clinical trial. *The Korean Journal of Pain.* 2025;38(2):187-197. doi:10.3344/kjp.24332
18. Alaparthi GK, Moustafa IM, Lopes AJ, Ferreira A de S. Pulmonary function, body posture and balance in young adults with asthma: A cross-sectional study. Kweh B, ed. *PLOS ONE.* 2025;20(3):e0316663. doi:10.1371/journal.pone.0316663
19. Silva M de C, Ferreira A de S, Baldon R de M, et al. Immediate Effects of Manual Therapy Techniques on Ankle Dorsiflexion: A Randomized Clinical Trial. *Journal of Manipulative and Physiological Therapeutics.* 2025;48(1-5):166-176. doi:10.1016/j.jmpt.2025.09.002
20. Amaral J, Silva GRC da, Thimoteo LM, Reis LF da F, Silva CA, Ferreira A de S. Explainable machine learning reveals key predictors of ICU mortality in COVID-19: functional outcomes and physiotherapy interventions in cardiovascular patients. *Brazilian Journal of Respiratory, Cardiovascular and Critical Care Physiotherapy.* 2025;16:e00072025. doi:10.47066/2966-4837.e00072025en
21. Bastos JRM, Ferreira AS, Lopes AJ, Pinto TP, Rodrigues E, dos Anjos FV. The Tinetti Balance Test Is an Effective Predictor of Functional Decline in Non-Hospitalized Post-COVID-19 Individuals: A Cross-Sectional Study. *Journal of Clinical Medicine.* 2024;13(21):6626. doi:10.3390/jcm13216626
22. Vieira L, De Sá Ferreira A, Meziat Filho NA, Frare JC, Zaidan de Barros P, Santos de Almeida R. Body composition does not interfere with urinary incontinence in adult women with grade III obesity. *Revista de Ciências Médicas e Biológicas.* 2024;23(2):389-393. doi:10.9771/cmbio.v23i2.57359
23. Kumar P, Ferreira A de S, Nogueira LAC, Arulsingh W, Patil MrS. Influence of Kinesiophobia on muscle endurance in patients with chronic low back pain- A case-control study. *F1000Research.* 2024;13:1016. doi:10.12688/f1000research.152751.1
24. Banks HC, Lemos T, Oliveira LAS, Ferreira AS. Short-term effects of Pilates-based exercise on upper limb strength and function in people with Parkinson's disease. *Journal of Bodywork and Movement Therapies.* 2024;39:237-242. doi:10.1016/j.jbmt.2024.02.032
25. Ali NM, Alaparthi GK, de Sá Ferreira A, Arumugam A, Bairapareddy KC. A national survey of physiotherapists' assessment and management practices for patients with COVID-19 in acute and rehabilitation care

- in the United Arab Emirates. *Fizjoterapia Polska*. 2024;24(2):309-317. doi:10.56984/8zg5608sr5
26. Mocarzel R, Kornin A, Tesser C, Ferreira A de S. Quem pode atuar com acupuntura no Brasil? *Saúde e Sociedade*. 2024;33(2). doi:10.1590/s0104-12902024230197pt
 27. Farias JP, Ferreira A de S. Evidence map on burnout syndrome in higher education teachers and its relationship with ergonomic and biopsychosocial factors: a scoping review. *International Journal of Occupational Safety and Ergonomics*. 2024;30(2):579-586. doi:10.1080/10803548.2024.2325819
 28. Al Yammahi RJ, Alaparthi GK, de Sá Ferreira A, Bairapareddy KC, Hegazy FA. Cardiopulmonary Response in Post-COVID-19 Individuals: A Cross-Sectional Study Comparing the Londrina Activities of Daily Living Protocol, 6-Minute Walk Test, and Glittre Activities of Daily Living Test. *Healthcare*. 2024;12(7):712. doi:10.3390/healthcare12070712
 29. Leivas EG, Corrêa LA, Ferreira A de S, De Almeida RS, Nogueira LAC. Falta de conhecimento sobre os fatores de risco não ocupacionais da dor lombar entre profissionais de saúde ocupacional brasileiros: um estudo observacional transversal. *Revista Pesquisa em Fisioterapia*. 2024;14:e5427. doi:10.17267/2238-2704rpf.2024.e5427
 30. Santos LE, de Sá Ferreira A, Vilella RC, Lunkes LC. The Importance of Physical Therapy in the Evaluation of Fall Prevention Programs in Older Adults. *Topics in Geriatric Rehabilitation*. 2024;40(1):83-92. doi:10.1097/tgr.0000000000000426
 31. Avila L, da Silva MD, Neves ML, et al. Effectiveness of Cognitive Functional Therapy Versus Core Exercises and Manual Therapy in Patients With Chronic Low Back Pain After Spinal Surgery: Randomized Controlled Trial. *Physical Therapy*. 2023;104(1). doi:10.1093/ptj/pzad105
 32. Ferreira APA, Maddaluno MLM, Curi ACC, Ferreira A de S. Interrater agreement and reliability of a palpation method for locating C1 transverse process in the cervical spine. *International Journal of Osteopathic Medicine*. 2024;51:100699. doi:10.1016/j.ijosm.2023.100699
 33. Vieira JE de A, Ferreira A de S, Monnerat LB, et al. Prediction models for physical function in COVID-19 survivors. *Journal of Bodywork and Movement Therapies*. 2024;37:70-75. doi:10.1016/j.jbmt.2023.11.002
 34. Reis LF da F, Oliveira JPA de, Ferreira A de S, Lopes AJ. Reply to: Factors associated with mortality in mechanically ventilated patients with severe acute respiratory syndrome due to COVID-19 evolution. *Critical Care Science*. 2024;36. doi:10.62675/2965-2774.20240213-en
 35. Reis LF da F, Oliveira JPA de, Ferreira A de S, Lopes AJ. Resposta para: Fatores associados à mortalidade em pacientes ventilados mecanicamente com síndrome respiratória aguda grave por evolução da COVID-19. *Critical Care Science*. 2024;36. doi:10.62675/2965-2774.20240213-pt
 36. Ribeiro Moço VJ, Gular AA, Lopes AJ, de Sá Ferreira A, da Fonseca Reis LF. Minimal-Resource Home Exercise Program Improves Activities of Daily Living, Perceived Health Status, and Shortness of Breath in Individuals with COPD Stages GOLD II to IV. *COPD: Journal of Chronic Obstructive Pulmonary Disease*. 2023;20(1):298-306. doi:10.1080/15412555.2023.2253907
 37. Silva AL dos S, Collange LA, Ferreira A de S. Hybrid maneuver for benign paroxysmal positional vertigo in individuals with limited neck mobility: Case series. *Journal of Bodywork and Movement Therapies*. 2024;37:386-391. doi:10.1016/j.jbmt.2023.11.056
 38. Rafagnin CZ, Ferreira A de S, Telles GF, Lemos de Carvalho T, Alexandre DJ de A, Nogueira LAC. Anterior component of Y-Balance test is correlated to ankle dorsiflexion range of motion in futsal players: A cross-sectional study. *Physiotherapy Research International*. 2023;28(4). doi:10.1002/pri.2028
 39. Casagrande CMZ, Farias JP, Meziat-Filho N, Nogueira LAC, Ferreira AS. Better Work Ability Is Associated With Lower Levels of Both Occupational Stress and Occupational Physical Activity in Professional Drivers.

- Journal of Occupational & Environmental Medicine. 2023;65(10):846-852. doi:10.1097/jom.00000000000
002918
40. Ferreira A de S, Miranda MG de, Vianna MA, Nascimento WR do. Tecnologias da Informação e Comunicação: visão de moradores e gestores para implementação da política pública de inclusão digital. Ead & Tecnologias Digitais na Educação. 2023;11(13):128-140. doi:10.30612/eadtde.v11i13.17329
 41. Bittencourt JV, Leivas EG, de Sá Ferreira A, Nogueira LAC. Does the painDETECT questionnaire identify impaired conditioned pain modulation in people with musculoskeletal pain? – a diagnostic accuracy study. Archives of Physiotherapy. 2023;13(1). doi:10.1186/s40945-023-00171-8
 42. Custódio LA, Marques YA, de Toledo AM, et al. The care pathway of individuals with spinal disorders in a Health Care Network in the Federal District, Brazil: a retrospective study. Brazilian Journal of Physical Therapy. 2023;27(5):100553. doi:10.1016/j.bjpt.2023.100553
 43. Deucher RA de O, Reis LF da F, Papathanasiou JV, et al. Estimating cardiopulmonary fitness with a new sampling technology in patients with rheumatoid arthritis-associated interstitial lung disease. Physiotherapy Research International. 2023;28(3). doi:10.1002/pri.2005
 44. Balata MR, Ferreira AS, da Silva Sousa A, et al. Assessment of Functional Capacity in Patients with Nondialysis-Dependent Chronic Kidney Disease with the Glittre Activities of Daily Living Test. Healthcare. 2023;11(12):1809. doi:10.3390/healthcare11121809
 45. Omar A, Ferreira A de S, Hegazy FA, Alaparthi GK. Cardiorespiratory Response to Six-Minute Step Test in Post COVID-19 Patients—A Cross Sectional Study. Healthcare. 2023;11(10):1386. doi:10.3390/healthcare11101386
 46. dos Santos Bento AP, Filho NM, Ferreira A de S, Cassetta AP, de Almeida RS. Sleep quality and polysomnographic changes in patients with chronic pain with and without central sensitization signs. Brazilian Journal of Physical Therapy. 2023;27(3):100504. doi:10.1016/j.bjpt.2023.100504
 47. Costa M, Saldanha PEC, Ferreira AS, Felicio LR, Lemos T. Posturography measures in specific ballet stance position discriminate ballet dancers with different occurrences of musculoskeletal injuries. Journal of Bodywork and Movement Therapies. 2023;34:41-45. doi:10.1016/j.jbmt.2023.04.020
 48. Parisotto G, Reis LFF, Junior MS, Papathanasiou J, Lopes AJ, Ferreira AS. Association of Multiple Cardiovascular Risk Factors with Musculoskeletal Function in Acute Coronary Syndrome Ward Inpatients. Healthcare. 2023;11(7):954. doi:10.3390/healthcare11070954
 49. Pinto TP, Inácio JC, de Aguiar E, et al. Prefrontal tDCS modulates autonomic responses in COVID-19 inpatients. Brain Stimulation. 2023;16(2):657-666. doi:10.1016/j.brs.2023.03.001
 50. Cunha J de A, Silva MM, Casagrande CMZ, Ferreira A de S. AMBIENTE DE TRABALHO SEGURO E SUSTENTÁVEL: COMO A ERGONOMIA DE CONSCIENTIZAÇÃO E PARTICIPATIVA SE APLICA AOS SERVIDORES PÚBLICOS? Arquivos de Ciências da Saúde da UNIPAR. 2023;27(1). doi:10.25110/arqsaudade.v27i1.2023.9145
 51. Fonseca GF, Michalski AC, Ferreira AS, et al. Is postexercise hypotension a method-dependent phenomenon in chronic stroke? A crossover randomized controlled trial. Clinical Physiology and Functional Imaging. 2023;43(4):242-252. doi:10.1111/cpf.12812
 52. Lunkes LC, Dias Neto MA, Barra LF, de Castro LR, Ferreira AS, Meziat-Filho N. Education to keep the abdomen relaxed versus contracted during pilates in patients with chronic low back pain: study protocol for a randomised controlled trial. BMC Musculoskeletal Disorders. 2023;24(1). doi:10.1186/s12891-023-06160-z
 53. Silva S de O, Barbosa JB, Lemos T, Oliveira LAS, Ferreira A de S. Agreement and predictive performance of fall risk assessment methods and factors associated with falls in hospitalized older adults: A longitudinal study. Geriatric Nursing. 2023;49:109-114. doi:10.1016/j.gerinurse.2022.11.016

54. Lunkes LC, Neto MAD, Barra LF, de Castro LR, Ferreira AS, Meziat-Filho N. Correction: Education to keep the abdomen relaxed versus contracted during pilates in patients with chronic low back pain: study protocol for a randomised controlled trial. *BMC Musculoskeletal Disorders*. 2023;24(1). doi:10.1186/s12891-023-06224-0
55. Reis FJJ, Bittencourt JV, Calestini L, de Sá Ferreira A, Meziat-Filho N, Nogueira LC. Exploratory analysis of 5 supervised machine learning models for predicting the efficacy of the endogenous pain inhibitory pathway in patients with musculoskeletal pain. *Musculoskeletal Science and Practice*. 2023;66:102788. doi:10.1016/j.msksp.2023.102788
56. de Oliveira JPA, Costa ACT, Lopes AJ, Ferreira A de S, Reis LF da F. Factors associated with mortality in mechanically ventilated patients with severe acute respiratory syndrome due to COVID-19 evolution. *Critical Care Science*. 2023;35(1). doi:10.5935/2965-2774.20230203-en
57. Neto RBD, Reis LFF, Ferreira A de S, Alexandre DJ de A, Almeida RS de. Hospital admission is associated with disability and late musculoskeletal pain in individuals with long COVID. *Frontiers in Rehabilitation Sciences*. 2023;4. doi:10.3389/fresc.2023.1186499
58. Bittencourt JV, Bezerra MC, Pina MR, Reis FJJ, de Sá Ferreira A, Nogueira LAC. Use of the painDETECT to discriminate musculoskeletal pain phenotypes. *Archives of Physiotherapy*. 2022;12(1). doi:10.1186/s40945-022-00129-2
59. Michalski AC, Ferreira AS, Midgley AW, et al. Mixed circuit training acutely reduces arterial stiffness in patients with chronic stroke: a crossover randomized controlled trial. *European Journal of Applied Physiology*. 2022;123(1):121-134. doi:10.1007/s00421-022-05061-8
60. Oliveira FAF, Martins CP, de Oliveira LAS, Rodrigues EC, Ferreira AS, Lemos T. Poststroke consequences upon optimization properties of postural sway during upright stance: a cross-sectional study. *Topics in Stroke Rehabilitation*. 2022;30(7):663-671. doi:10.1080/10749357.2022.2130620
61. Ferreira APA, Zanier JFC, Santos EBG, Ferreira AS. Accuracy of Palpation Procedures for Locating the C1 Transverse Process and Masseter Muscle as Confirmed by Computed Tomography Images. *Journal of Manipulative and Physiological Therapeutics*. 2022;45(5):337-345. doi:10.1016/j.jmpt.2022.07.005
62. Xavier DD, Graf RM, Ferreira AS. Short-Term Changes in Posture and Pain of the Neck and Lower Back of Women Undergoing Lipoabdominoplasty: A Case Series Report. *Journal of Chiropractic Medicine*. 2023;22(2):138-147. doi:10.1016/j.jcm.2022.07.003
63. Telles GF, Ferreira A de S, Junior PMP, Lemos T, Bittencourt JV, Nogueira LAC. Concurrent validity of the inertial sensors for assessment of balance control during quiet standing in patients with chronic low back pain and asymptomatic individuals. *Journal of Medical Engineering & Technology*. 2022;46(5):354-362. doi:10.1080/03091902.2022.2043947
64. Paz T da SR, Rodrigues PTV, Silva BM, de Sá Ferreira A, Nogueira LAC. Mediation Analysis in Manual Therapy Research. *Journal of Chiropractic Medicine*. 2023;22(1):35-44. doi:10.1016/j.jcm.2022.04.007
65. Silva CA, Lopes AJ, Papathanasiou J, Reis LFF, Ferreira AS. Association of Functional Characteristics and Physiotherapy with COVID-19 Mortality in Intensive Care Unit in Inpatients with Cardiovascular Diseases. *Medicina*. 2022;58(6):823. doi:10.3390/medicina58060823
66. Leivas EG, Bittencourt JV, Ferreira AS, Nogueira LAC. Is it possible to discriminate workers with a higher prevalence of low back pain considering daily exposure time in a work-related lumbar posture? A diagnostic accuracy study. *Ergonomics*. 2021;65(6):877-885. doi:10.1080/00140139.2021.2001577
67. Galvão AF, Lemos T, Martins CP, Horszczaruk CHR, Oliveira LAS, Ferreira A de S. Body sway and movement strategies for control of postural stability in people with spinocerebellar ataxia type 3: A cross-sectional study. *Clinical Biomechanics*. 2022;97:105711. doi:10.1016/j.clinbiomech.2022.105711

68. Certain Curi AC, Antunes Ferreira AP, Calazans Nogueira LA, Meziat Filho NAM, Sá Ferreira A. Osteopathy and physiotherapy compared to physiotherapy alone on fatigue in long COVID: Study protocol for a pragmatic randomized controlled superiority trial. *International Journal of Osteopathic Medicine.* 2022;44:22-28. doi:10.1016/j.ijosm.2022.04.004
69. Tedla JS, Rodrigues E, Ferreira AS, et al. Transcranial direct current stimulation combined with trunk-targeted, proprioceptive neuromuscular facilitation in subacute stroke: a randomized controlled trial. *PeerJ.* 2022;10:e13329. doi:10.7717/peerj.13329
70. Castro J, Correia L, Donato B de S, et al. Cognitive functional therapy compared with core exercise and manual therapy in patients with chronic low back pain: randomised controlled trial. *Pain.* 2022;163(12):2430-2437. doi:10.1097/j.pain.0000000000002644
71. Tedla JS, Gular K, Reddy RS, et al. Effectiveness of Constraint-Induced Movement Therapy (CIMT) on Balance and Functional Mobility in the Stroke Population: A Systematic Review and Meta-Analysis. *Healthcare.* 2022;10(3):495. doi:10.3390/healthcare10030495
72. CASAGRANDE CMZ, FERREIRA A de S. Challenges and Perspectives for Research on Work Ability in Professional Drivers: A Scoping Review. *Journal of UOEH.* 2022;44(1):25-34. doi:10.7888/juoeh.44.25
73. Sá R de A, Ferreira A de S, Lemos T, de Oliveira LAS. Correlation Analysis of Lower-Limb Muscle Function With Clinical Status, Balance Tests, and Quality of Life in People With Parkinson Disease. *Topics in Geriatric Rehabilitation.* 2022;38(1):56-64. doi:10.1097/tgr.0000000000000343
74. Volpato MP, Menezes M, Prado TS, Piccini A, Ferreira AS, Botelho S. Electromyographic analysis of maximal voluntary contraction of female pelvic floor muscles: Intrarater and interrater reliability study. *Neurourology and Urodynamics.* 2021;41(1):383-390. doi:10.1002/nau.24834
75. Nascimento MM, Silva PRO, Felício LR, et al. Postural control in football players with vision impairment: Effect of sports adaptation or visual input restriction? *Motriz: Revista de Educação Física.* 2022;28. doi:10.1590/s1980-657420220010821
76. Corrêa LA, Mathieson S, Meziat-Filho NA de M, Reis FJ, Ferreira A de S, Nogueira LAC. Which psychosocial factors are related to severe pain and functional limitation in patients with low back pain? *Brazilian Journal of Physical Therapy.* 2022;26(3):100413. doi:10.1016/j.bjpt.2022.100413
77. Vilela AC, Nogueira LAC, Ferreira A de S, Kochem FB, de Almeida RS. Musculoskeletal Pain and Musical Performance in First and Second Violinists of Professional Youth Chamber Orchestras: A Comparative Study. *Medical Problems of Performing Artists.* 2021;36(4):263-268. doi:10.21091/mppa.2021.4029
78. Willuweit MGA, Lopes AJ, Ferreira AS. Development of a multivariable prediction model of functional exercise capacity in liver transplant recipients. *Journal of Liver Transplantation.* 2022;6:100067. doi:10.1016/j.liver.2021.100067
79. Reddy RS, Gautam AP, Tedla JS, et al. The Aftermath of the COVID-19 Crisis in Saudi Arabia: Respiratory Rehabilitation Recommendations by Physical Therapists. *Healthcare.* 2021;9(11):1560. doi:10.3390/healthcare9111560
80. Pena Junior PM, de Sá Ferreira A, Telles G, Lemos T, Calazans Nogueira LA. Concurrent validation of the centre of pressure displacement analyzed by baropodometry in patients with chronic non-specific low back pain during functional tasks. *Journal of Bodywork and Movement Therapies.* 2021;28:489-495. doi:10.1016/j.jbmt.2021.06.020
81. Varella NC, Almeida RS, Nogueira LAC, Ferreira AS. Cross-cultural adaptation of the Richards-Campbell Sleep Questionnaire for intensive care unit inpatients in Brazil: internal consistency, test-retest reliability, and measurement error. *Sleep Medicine.* 2021;85:38-44. doi:10.1016/j.sleep.2021.06.039
82. 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

83. Deucher RA de O, Ferreira A de S, Nascimento LPA da S, Cal MS da, Papathanasiou JV, Lopes AJ. Heart Rate Variability in Adults with Sickle Cell Anemia During a Multitasking Field Test. *Asian Journal of Sports Medicine*. 2021;12(3). doi:10.5812/asjsm.108537
84. PAPATHANASIOU JV, PETROV I, TSEKOURA D, et al. Does group-based high-intensity aerobic interval training improve the inflammatory status in patients with chronic heart failure? *European Journal of Physical and Rehabilitation Medicine*. 2022;58(2). doi:10.23736/s1973-9087.21.06894-5
85. Saraiva NAO, Ferreira AS, Papathanasiou JV, Guimarães FS, Lopes AJ. Kinematic evaluation of patients with chronic obstructive pulmonary disease during the 6-min walk test. *Journal of Bodywork and Movement Therapies*. 2021;27:134-140. doi:10.1016/j.jbmt.2021.01.005
86. Souza MF, Ferreira AS. Education of traditional medicine for people with visual impairments in Brazil: Challenges and strategies. *Integrative Medicine Research*. 2021;10(2):100687. doi:10.1016/j.imr.2020.100687
87. Reddy RS, Meziat-Filho N, Ferreira AS, Tedla JS, Kandakurti PK, Kakaraparthi VN. Comparison of neck extensor muscle endurance and cervical proprioception between asymptomatic individuals and patients with chronic neck pain. *Journal of Bodywork and Movement Therapies*. 2021;26:180-186. doi:10.1016/j.jbmt.2020.12.040
88. Maddaluno MLM, Ferreira APA, Tavares ACLC, Meziat-Filho N, Ferreira AS. Craniocervical Posture Assessed With Photogrammetry and the Accuracy of Palpation Methods for Locating the Seventh Cervical Spinous Process: A Cross-sectional Study. *Journal of Manipulative and Physiological Therapeutics*. 2021;44(3):196-204. doi:10.1016/j.jmpt.2020.07.012
89. Queiroz dos Santos AN, Lemos T, Duarte Carvalho PH, Ferreira AS, Silva JG. Immediate effects of myofascial release maneuver applied in different lower limb muscle chains on postural sway. *Journal of Bodywork and Movement Therapies*. 2021;25:151-156. doi:10.1016/j.jbmt.2020.10.024
90. Correia IMT, Ferreira A de S, Fernandez J, Reis FJJ, Nogueira LAC, Meziat-Filho N. Association Between Text Neck and Neck Pain in Adults. *Spine*. 2020;46(9):571-578. doi:10.1097/brs.0000000000003854
91. Ferreira N de A, Ferreira A de S, Guimarães FS. Cough peak flow to predict extubation outcome: a systematic review and meta-analysis. *Revista Brasileira de Terapia Intensiva*. 2021;33(3). doi:10.5935/0103-507x.20210060
92. Palugan MJA, Assis ACB, Bessa EJC, Ferreira AS, Lopes AJ. Predictors of functional capacity as measured by the Glittre activities of daily living test in women with rheumatoid arthritis. *Brazilian Journal of Medical and Biological Research*. 2021;54(5). doi:10.1590/1414-431x202010040
93. Castro P, Ferreira A de S, Lopes AJ, et al. Validity of the Polar V800 heart rate monitor for assessing cardiac autonomic control in individuals with spinal cord injury. *Motriz: Revista de Educação Física*. 2021;27. doi:10.1590/s1980-65742021003221
94. Menezes M, Meziat-Filho NAM, Lemos T, Ferreira AS. 'Believe the positive' aggregation of fall risk assessment methods reduces the detection of risk of falling in older adults. *Archives of Gerontology and Geriatrics*. 2020;91:104228. doi:10.1016/j.archger.2020.104228
95. Gomes AS, de Sá Ferreira A, Reis FJJ, de Jesus-Moraleida FR, Nogueira LAC, Meziat-Filho N. Association Between Low Back Pain and Biomedical Beliefs in Academics of Physiotherapy. *Spine*. 2020;45(19):1354-1359. doi:10.1097/brs.0000000000003487
96. PAPATHANASIOU JV, PETROV I, TOKMAKOVA MP, et al. Group-based cardiac rehabilitation interventions. A challenge for physical and rehabilitation medicine physicians: a randomized controlled trial. *European Journal of Physical and Rehabilitation Medicine*. 2020;56(4). doi:10.23736/s1973-9087.20.06013-x

97. Galvão TS, Magalhães Júnior ES, Orsini Neves MA, de Sá Ferreira A. Lower-limb muscle strength, static and dynamic postural stabilities, risk of falling and fear of falling in polio survivors and healthy subjects. *Physiotherapy Theory and Practice.* 2018;36(8):899-906. doi:10.1080/09593985.2018.1512178
98. Jeronymo BF, Silva PR de O, Mainenti M, et al. The Relationship Between Postural Stability, Anthropometry Measurements, Body Composition, and Sport Experience in Judokas with Visual Impairment. *Asian Journal of Sports Medicine.* 2020;11(3). doi:10.5812/asjsm.103030
99. Corrêa LA, Bittencourt JV, Ferreira A de S, Reis FJJ dos, de Almeida RS, Nogueira LAC. The Reliability and Concurrent Validity of PainMAP Software for Automated Quantification of Pain Drawings on Body Charts of Patients With Low Back Pain. *Pain Practice.* 2020;20(5):462-470. doi:10.1111/papr.12872
100. Ferreira AS, Maior AS. Two decades of research in soccer and acupuncture: to what point should we stick? *Longhua Chinese Medicine.* 2020;3:2-2. doi:10.21037/lcm.2020.02.01
101. de Andrade Junior AB, Ferreira A de S, Assis ACB, et al. Cardiac Autonomic Control in Women with Rheumatoid Arthritis During the Glitter Activities of Daily Living Test. *Asian Journal of Sports Medicine.* 2020;11(2). doi:10.5812/asjsm.101400
102. Sant'Anna do Carmo Aprigio P, Ramathur Telles de Jesus I, Porto C, Lemos T, de Sá Ferreira A. Lower limb muscle fatigability is not associated with changes in movement strategies for balance control in the upright stance. *Human Movement Science.* 2020;70:102588. doi:10.1016/j.humov.2020.102588
103. Porto C, Lemos T, Sá Ferreira A. Reliability and robustness of optimization properties for stabilization of the upright stance as determined using posturography. *Journal of Biomechanics.* 2020;103:109686. doi:10.1016/j.jbiomech.2020.109686
104. Menezes M, de Mello Meziat-Filho NA, Araújo CS, Lemos T, Ferreira AS. Agreement and predictive power of six fall risk assessment methods in community-dwelling older adults. *Archives of Gerontology and Geriatrics.* 2020;87:103975. doi:10.1016/j.archger.2019.103975
105. Papathanasiou J, Dimitrova D, Dzhafer N, et al. Are group-based high-intensity aerobic interval training modalities the future of the cardiac rehabilitation? *Hellenic Journal of Cardiology.* 2020;61(2):141-144. doi:10.1016/j.hjc.2019.10.015
106. Sam-Kit Tin T, Daniel Weng CH, Vigário P dos S, Ferreira A de S. Effects of A Short-term Cardio Tai Chi Program on Cardiorespiratory Fitness and Hemodynamic Parameters in Sedentary Adults: A Pilot Study. *Journal of Acupuncture and Meridian Studies.* 2020;13(1):12-18. doi:10.1016/j.jams.2019.12.002
107. Maior AS, Tannure M, Eiras F, de Sá Ferreira A. Effects of intermittent negative pressure and active recovery therapies in the post-match period in elite soccer players: A randomized, parallel arm, comparative study. *Biomedical Human Kinetics.* 2020;12(1):59-68. doi:10.2478/bhk-2020-0008
108. Almeida VP, Ferreira AS, Guimarães FS, Papathanasiou J, Lopes AJ. Predictive models for the six-minute walk test considering the walking course and physical activity level. *European Journal of Physical and Rehabilitation Medicine.* 2020;55(6). doi:10.23736/s1973-9087.19.05687-9
109. Fernandez J, Ferreira A de S, Castro J, Correia LCL, Meziat Filho N. Comment on the paper “Cognitive functional therapy in patients with non specific chronic low back pain”, by Vibe Fersum et al. *European Journal of Pain.* 2019;23(8):1574-1575. doi:10.1002/ejp.1441
110. Gomes BSQ, Coelho VK, Terra BS, et al. Patients with Subacromial Pain Syndrome Present no Reduction of Shoulder Proprioception: A Matched Case-Control Study. *PM&R.* 2019;11(9):972-978. doi:10.1002/pmrj.12055
111. Porto C, Lemos T, Ferreira AS. Analysis of the postural stabilization in the upright stance using optimization properties. *Biomedical Signal Processing and Control.* 2019;52:171-178. doi:10.1016/j.bspc.2019.04.009

112. Almeida VP, Ferreira AS, Guimarães FS, Papathanasiou J, Lopes AJ. The impact of physical activity level, degree of dyspnoea and pulmonary function on the performance of healthy young adults during exercise. *Journal of Bodywork and Movement Therapies.* 2019;23(3):494-501. doi:10.1016/j.jbmt.2018.05.005
113. Vieira ÉCN, Meziat-Filho NAM, Ferreira AS. Photogrammetric Variables Used by Physical Therapists to Detect Neck Pain and to Refer for Physiotherapeutic Intervention: A Cross-Sectional Study. *Journal of Manipulative and Physiological Therapeutics.* 2019;42(4):254-266. doi:10.1016/j.jmpt.2018.11.014
114. Michalski A da C, Ferreira A de S, Kasuki L, Gadelha MR, Lopes AJ, Guimarães FS. Clinical and functional variables can predict general fatigue in patients with acromegaly: an explanatory model approach. *Archives of Endocrinology and Metabolism.* April 2019. doi:10.20945/2359-3997000000127
115. Ramos R de A, Guimarães FS, Dionyssiotis Y, Tsekoura D, Papathanasiou J, Ferreira A de S. Development of a multivariate model of the six-minute walked distance to predict functional exercise capacity in hypertension. *Journal of Bodywork and Movement Therapies.* 2019;23(1):32-38. doi:10.1016/j.jbmt.2018.01.010
116. Ferreira AS, Cunha FA. The circadian blood pressure variability: There is a signal in the noise. *The Journal of Clinical Hypertension.* 2018;21(1):46-47. doi:10.1111/jch.13430
117. Lima TRL, Almeida VP, Ferreira AS, Guimarães FS, Lopes AJ. Handgrip Strength and Pulmonary Disease in the Elderly: What is the Link? *Aging and disease.* 2019;10(5):1109. doi:10.14336/ad.2018.1226
118. LOPES AJ, VIGÁRIO PS, HORA AL, et al. Ventilation Distribution, Pulmonary Diffusion and Peripheral Muscle Endurance as Determinants of Exercise Intolerance in Elderly Patients With Chronic Obstructive Pulmonary Disease. *Physiological Research.* December 2018:863-874. doi:10.33549/physiolres.933867
119. Alvim DT, Ferreira AS. Pragmatic Combinations of Acupuncture Points for Lateral Epicondylalgia are Unreliable in the Physiotherapy Setting. *Journal of Acupuncture and Meridian Studies.* 2018;11(6):367-374. doi:10.1016/j.jams.2018.07.006
120. da Silva DCL, Lemos T, de Sá Ferreira A, et al. Effects of Acute Transcranial Direct Current Stimulation on Gait Kinematics of Individuals With Parkinson Disease. *Topics in Geriatric Rehabilitation.* 2018;34(4):262-268. doi:10.1097/tgr.0000000000000203
121. Damasceno GM, Ferreira AS, Nogueira LAC, Reis FJJ, Lara RW, Meziat-Filho N. Reliability of two pragmatic tools for assessing text neck. *Journal of Bodywork and Movement Therapies.* 2018;22(4):963-967. doi:10.1016/j.jbmt.2018.01.007
122. Damasceno GM, Ferreira AS, Nogueira LAC, Reis FJJ, Andrade ICS, Meziat-Filho N. Text neck and neck pain in 18–21-year-old young adults. *European Spine Journal.* 2018;27(6):1249-1254. doi:10.1007/s00586-017-5444-5
123. Silva PO, Ferreira AS, Lima CM de A, Guimarães FS, Lopes AJ. Balance control is impaired in adults with sickle cell anaemia. *Somatosensory & Motor Research.* 2018;35(2):109-118. doi:10.1080/08990220.2018.1481829
124. Saraiva NAO, Guimarães FS, Lopes AJ, Papathanasiou J, Ferreira AS. Feasibility of whole-body gait kinematics to assess the validity of the six-minute walk test over a 10-m walkway in the elderly. *Biomedical Signal Processing and Control.* 2018;42:202-209. doi:10.1016/j.bspc.2018.02.002
125. Meziat-Filho N, Ferreira AS, Nogueira LAC, Reis FJJ. “Text-neck”: an epidemic of the modern era of cell phones? *The Spine Journal.* 2018;18(4):714-715. doi:10.1016/j.spinee.2017.11.022
126. Lima M, Ferreira AS, Reis FJJ, Paes V, Meziat-Filho N. Chronic low back pain and back muscle activity during functional tasks. *Gait & Posture.* 2018;61:250-256. doi:10.1016/j.gaitpost.2018.01.021
127. Fonseca GF, Farinatti PTV, Midgley AW, et al. Continuous and Accumulated Bouts of Cycling Matched by Intensity and Energy Expenditure Elicit Similar Acute Blood Pressure Reductions in Prehypertensive Men. *Journal of Strength and Conditioning Research.* 2018;32(3):857-866. doi:10.1519/jsc.0000000000002317

128. Alvim DT, Ferreira AS. Inter-expert agreement and similarity analysis of traditional diagnoses and acupuncture prescriptions in textbook- and pragmatic-based practices. *Complementary Therapies in Clinical Practice*. 2018;30:38-43. doi:10.1016/j.ctcp.2017.12.002
129. de Mello MC, de Sá Ferreira A, Felicio LR. Postural Control during Different Unipodal Positions in Professional Ballet Dancers. *Journal of Dance Medicine & Science*. 2017;21(4):151-155. doi:10.12678/1089-313x.21.4.151
130. Santos PBR, Vigário PS, Mainenti MRM, Ferreira AS, Lemos T. Seated limits of stability of athletes with disabilities with regard to competitive levels and sport classification. *Scandinavian Journal of Medicine & Science in Sports*. 2017;27(12):2019-2026. doi:10.1111/sms.12847
131. Lopes AJ, Justo AC, Ferreira AS, Guimaraes FS. Systemic sclerosis: Association between physical function, handgrip strength and pulmonary function. *Journal of Bodywork and Movement Therapies*. 2017;21(4):972-977. doi:10.1016/j.jbmt.2017.03.018
132. Ferreira APA, Póvoa LC, Zanier JFC, Machado DC, Ferreira AS. Sensitivity for palpating lumbopelvic soft-tissues and bony landmarks and its associated factors: A single-blinded diagnostic accuracy study. *Journal of Back and Musculoskeletal Rehabilitation*. 2017;30(4):735-744. doi:10.3233/bmr-150356
133. Justo AC, Guimarães FS, Ferreira AS, Soares MS, Bunn PS, Lopes AJ. Muscle function in women with systemic sclerosis: Association with fatigue and general physical function. *Clinical Biomechanics*. 2017;47:33-39. doi:10.1016/j.clinbiomech.2017.05.011
134. Facchinetti LD, Araújo AQ, Silva MT, et al. Home-based exercise program in TSP/HAM individuals: a feasibility and effectiveness study. *Arquivos de Neuro-Psiquiatria*. 2017;75(4):221-227. doi:10.1590/0004-282x20170022
135. Ferreira APA, Póvoa LC, Zanier JFC, Ferreira AS. Locating the Seventh Cervical Spinous Process: Accuracy of the Thorax-Rib Static Method and the Effects of Clinical Data on Its Performance. *Journal of Manipulative and Physiological Therapeutics*. 2017;40(2):98-105. doi:10.1016/j.jmpt.2016.10.011
136. Ferreira APA, Póvoa LC, Zanier JFC, Ferreira AS. Locating the Seventh Cervical Spinous Process: Development and Validation of a Multivariate Model Using Palpation and Personal Information. *Journal of Manipulative and Physiological Therapeutics*. 2017;40(2):89-97. doi:10.1016/j.jmpt.2016.10.012
137. Marques NLXR, de Sá Ferreira A, da Silva DPG, de Menezes SLS, Guimarães FS, Dias CM. Performance of National and Foreign Models for Predicting the 6-Minute Walk Distance for Assessment of Functional Exercise Capacity of Brazilian Elderly Women. *Topics in Geriatric Rehabilitation*. 2017;33(1):68-75. doi:10.1097/tgr.0000000000000134
138. Santos Neves R, Lopes AJ. Hand grip strength in healthy young and older Brazilian adults: Kinesiology. 2017;49(2):208-216. doi:10.26582/k.49.2.5
139. Lopes AJ, Ferreira AS, Walchan EM, Soares MS, Bunn PS, Guimarães FS. Explanatory models of muscle performance in acromegaly patients evaluated by knee isokinetic dynamometry: Implications for rehabilitation. *Human Movement Science*. 2016;49:160-169. doi:10.1016/j.humov.2016.07.005
140. Marinho C de L, Maioli MCP, Soares AR, et al. Predictive models of six-minute walking distance in adults with sickle cell anemia: Implications for rehabilitation. *Journal of Bodywork and Movement Therapies*. 2016;20(4):824-831. doi:10.1016/j.jbmt.2016.02.005
141. Costa MSS, Ferreira AS, Orsini M, Silva EB, Felicio LR. Characteristics and prevalence of musculoskeletal injury in professional and non-professional ballet dancers. *Brazilian Journal of Physical Therapy*. 2016;20(2):166-175. doi:10.1590/bjpt-rbf.2014.0142
142. Gomes Ribeiro Moura N, Sá Ferreira A. Pulse Waveform Analysis of Chinese Pulse Images and Its Association with Disability in Hypertension. *Journal of Acupuncture and Meridian Studies*. 2016;9(2):93-98. doi:10.1016/j.jams.2015.06.012

143. Ribeiro de Moura NG, Cordovil I, de Sá Ferreira A. Traditional Chinese medicine wrist pulse-taking is associated with pulse waveform analysis and hemodynamics in hypertension. *Journal of Integrative Medicine*. 2016;14(2):100-113. doi:10.1016/s2095-4964(16)60233-9
144. Papathanasiou J, Troev T, Ferreira AS, et al. Advanced Role and Field of Competence of the Physical and Rehabilitation Medicine Specialist in Contemporary Cardiac Rehabilitation. *Hellenic Journal of Cardiology*. 2016;57(1):16-22. doi:10.1016/s1109-9666(16)30013-6
145. Lopes AJ, Ferreira A de S, Lima TRL, Menezes SLS, Guimarães FS. An explanatory model of functional exercise capacity in patients with systemic sclerosis: considerations for rehabilitation programs. *Journal of Physical Therapy Science*. 2016;28(2):569-575. doi:10.1589/jpts.28.569
146. Ferreira NA, Lopes AJ, Ferreira AS, Ntoumenopoulos G, Dias J, Guimaraes FS. Determination of functional prognosis in hospitalized patients following an intensive care admission. *World Journal of Critical Care Medicine*. 2016;5(4):219. doi:10.5492/wjccm.v5.i4.219
147. Lopes AJ, Mafort TT, De SÁ; Ferreira A, Santos de Castro MC, De CÁssia Firmida M, De Andrade Marques E. Is the type of chronic pulmonary infection a determinant of lung function outcomes in adult patients with cystic fibrosis? *Monaldi Archives for Chest Disease*. 2015;77(3-4). doi:10.4081/monaldi.2012.145
148. Monteiro-Junior R, Ferreira A, Puell V, et al. Wii Balance Board: Reliability and Clinical Use in Assessment of Balance in Healthy Elderly Women. *CNS & Neurological Disorders - Drug Targets*. 2015;14(9):1165-1170. doi:10.2174/187152731566615111120403
149. Sá MRC, Ribeiro CT, Fracho FG, et al. Age of independent sitting posture acquisition for children with myelomeningocele. *Journal of the Neurological Sciences*. 2015;357:e198. doi:10.1016/j.jns.2015.08.685
150. Ferreira AS. Immunity, Inflammation, and Prehypertension: In What Order? *The Journal of Clinical Hypertension*. 2015;17(10):775-776. doi:10.1111/jch.12611
151. Orsini M, Reis CHM, Ferreira AS, et al. Postural balance in Machado-Joseph disease. *Journal of the Neurological Sciences*. 2015;357:e266. doi:10.1016/j.jns.2015.08.935
152. de Sá Ferreira A. Plasma Homocysteine and Arterial Stiffness: Risk Factors or Risk Markers for Cardiovascular Diseases? *The Journal of Clinical Hypertension*. 2015;17(8):601-602. doi:10.1111/jch.12549
153. Orsini M, De Souza JA, Leite MAA, et al. Previous acute polio and post-polio syndrome: recognizing the pathophysiology for the establishment of rehabilitation programs. *Neurology International*. 2015;7(1). doi:10.4081/ni.2015.5452
154. Gonçalves BL, Guimarães FS, Souza MLL de, Ferreira A de S, Mainenti MRM. Association among body composition, muscle performance and functional autonomy in older adults. *Fisioterapia em Movimento*. 2015;28(1):49-59. doi:10.1590/0103-5150.028.001.ao05
155. de Sá Ferreira A, Pacheco AG. SimTCM: A human patient simulator with application to diagnostic accuracy studies of Chinese medicine. *Journal of Integrative Medicine*. 2015;13(1):9-19. doi:10.1016/s2095-4964(15)60151-0
156. Oliveira IJ de AS, de Sá Ferreira A. Effects of Diagnostic Errors in Pattern Differentiation and Acupuncture Prescription: A Single-Blinded, Interrater Agreement Study. *Evidence-Based Complementary and Alternative Medicine*. 2015;2015:1-11. doi:10.1155/2015/469675
157. Lopes AJ, Guedes da Silva DP, Ferreira A de S, Kasuki L, Gadelha MR, Guimarães FS. What is the effect of peripheral muscle fatigue, pulmonary function, and body composition on functional exercise capacity in acromegalic patients? *Journal of Physical Therapy Science*. 2015;27(3):719-724. doi:10.1589/jpts.27.719
158. Concordar ou discordar: (eis) a questão da diversidade. *Fisioterapia em Movimento*. 2014;27(4):491-492. doi:10.1590/0103-5150.027.004.ed01

159. Portela FM, Ferreira AS. Kinematic Mapping Reveals Different Spatial Distributions of Center of Pressure High-Speed Regions Under Somatosensory Loss. *Journal of Motor Behavior*. 2014;46(5):369-379. doi: 10.1080/00222895.2014.916651
160. de Sá Ferreira A, Junqueira Ferraz Baracat P. Test-retest reliability for assessment of postural stability using center of pressure spatial patterns of three-dimensional statokinesigrams in young health participants. *Journal of Biomechanics*. 2014;47(12):2919-2924. doi:10.1016/j.jbiomech.2014.07.010
161. Ramos RA, Guimarães FS, Cordovil I, de Sa Ferreira A. The six-minute walk distance is a marker of hemodynamic-related functional capacity in hypertension: a case-control study. *Hypertension Research*. 2014;37(8):746-752. doi:10.1038/hr.2014.59
162. Ng SS, Fong SS, Lam SS, Lai CW, Chow LP, de Sá Ferreira A. Acupressure and task-related training after stroke: A case study. *International Journal of Therapy and Rehabilitation*. 2014;21(4):183-189. doi: 10.12968/ijtr.2014.21.4.183
163. Mainenti MRM, Rodrigues EDC, Ferreira ADS, Sousa RCM de, Silva DTR da. Alinhamento articular de membros inferiores e controle postural em idosas. *Revista Brasileira de Cineantropometria e Desempenho Humano*. 2014;16(3):287. doi:10.5007/1980-0037.2014v16n3p287
164. Pereira RB, Felício LR, Ferreira A de S, Menezes SL de, Freitas MRG de, Orsini M. Immediate effects of using ankle-foot orthoses in the kinematics of gait and in the balance reactions in Charcot-Marie-Tooth disease. *Fisioterapia e Pesquisa*. 2014;21(1):87-93. doi:10.1590/1809-2950/515210114
165. Lima TRL, Guimarães FS, Sá Ferreira A, Penafortes JTS, Almeida VP, Lopes AJ. Correlation between posture, balance control, and peripheral muscle function in adults with cystic fibrosis. *Physiotherapy Theory and Practice*. 2013;30(2):79-84. doi:10.3109/09593985.2013.820246
166. Portela FM, Rodrigues EC, Ferreira A de S. A critical review of position- and velocity-based concepts of postural control during upright stance. *Human Movement*. 2018;15(4):227-233. doi:10.1515/humo-2015-0016
167. Ferreira A de S, Moura NGR de. Asserted and neglected issues linking evidence-based and Chinese medicines for cardiac rehabilitation. *World Journal of Cardiology*. 2014;6(5):295. doi:10.4330/wjc.v6.i5.295
168. Xiong X, Borrelli F, de Sá Ferreira A, Ashfaq T, Feng B. Herbal Medicines for Cardiovascular Diseases. *Evidence-Based Complementary and Alternative Medicine*. 2014;2014(1). doi:10.1155/2014/809741
169. Baracat PJF, de Sá Ferreira A. Postural tasks are associated with center of pressure spatial patterns of three-dimensional statokinesigrams in young and elderly healthy subjects. *Human Movement Science*. 2013;32(6):1325-1338. doi:10.1016/j.humov.2013.06.005
170. de Almeida VP, Guimarães FS, Moço VJR, de Sá Ferreira A, de Menezes SLS, Lopes AJ. Is there an association between postural balance and pulmonary function in adults with asthma? *Clinics*. 2013;68(11):1421-1427. doi:10.6061/clinics/2013(11)07
171. de Sá Ferreira A. Evidence-based practice of Chinese medicine in physical rehabilitation science. *Chinese Journal of Integrative Medicine*. 2013;19(10):723-729. doi:10.1007/s11655-013-1451-5
172. Ferreira A de S, Guimarães FS, Magalhães MAR, Silva RCS e. Accuracy and learning curves of inexperienced observers for manual segmentation of electromyograms. *Fisioterapia em Movimento*. 2013;26(3):559-567. doi:10.1590/s0103-51502013000300009
173. Costa MS da S, Ferreira A de S, Felicio LR. Equilíbrio estático e dinâmico em bailarinos: revisão da literatura. *Fisioterapia e Pesquisa*. 2013;20(3):299-305. doi:10.1590/s1809-29502013000300016
174. Mainenti MRM, Sousa RCM de, Dias CM, et al. Body Composition and Chest Expansion of Type II and III Spinal Muscular Atrophy Patients. *Journal of Human Growth and Development*. 2013;23(2):164. doi: 10.7322/jhgd.61291

175. Ferreira A de S. Integrative medicine for hypertension: the earlier the better for treating who and what are not yet ill. *Hypertension Research*. 2013;36(7):583-585. doi:10.1038/hr.2013.15
176. Ferreira A de S. Resonance phenomenon during wrist pulse-taking: A stochastic simulation, model-based study of the ‘pressing with one finger’ technique. *Biomedical Signal Processing and Control*. 2013;8(3):229-236. doi:10.1016/j.bspc.2012.10.004
177. de Sá Ferreira A, Lopes AJ. Pulse waveform analysis as a bridge between pulse examination in Chinese medicine and cardiology. *Chinese Journal of Integrative Medicine*. 2013;19(4):307-314. doi:10.1007/s1655-013-1412-z
178. de Sá Ferreira A. Promoting integrative medicine by computerization of traditional Chinese medicine for scientific research and clinical practice: The SuiteTCM Project. *Journal of Integrative Medicine*. 2013;11(2):135-139. doi:10.3736/jintegrmed2013013
179. Ferreira AS, Luiz AB. Role of dermatomes in the determination of therapeutic characteristics of channel acupoints: a similarity-based analysis of data compiled from literature. *Chinese Medicine*. 2013;8(1):24. doi:10.1186/1749-8546-8-24
180. Oliveira PC, Silva MCS, Silva MCS, Ferreira ADS. Tratamento da fibromialgia por acupuntura baseado na diferenciação de padrões: Revisão sistemática. *Cadernos de Naturopatia e Terapias Complementares*. 2013;2(3):39. doi:10.19177/cntc.v2e3201339-47
181. Fragoso AP de S, Ferreira A de S. Statistical distribution of acupoint prescriptions for sensory-motor impairments in post-stroke subjects. *Chinese Journal of Integrative Medicine*. December 2012. doi:10.1007/s11655-012-1245-1
182. Pereira RB, Orsini M, Ferreira A de S, et al. Efeitos do uso de órteses na Doença de Charcot-Marie-Tooth: atualização da literatura. *Fisioterapia e Pesquisa*. 2012;19(4):388-393. doi:10.1590/s1809-29502012000400016
183. Ferreira A de S, Filho JB, Cordovil I, Souza MN de. Noninvasive pressure pulse waveform analysis of flow-mediated vasodilation evoked by post-occlusive reactive hyperemia maneuver. *Biomedical Signal Processing and Control*. 2012;7(6):616-621. doi:10.1016/j.bspc.2012.03.001
184. Lopes AJ, Costa W, Thomaz Mafort T, de Sá Ferreira A, Silveira de Menezes SL, Silva Guimarães F. Silicose em jateadores de areia de estaleiro versus silicose em escultores de pedra no Brasil: uma comparação dos achados de imagem, função pulmonar e teste de exercício cardiopulmonar. *Revista Portuguesa de Pneumologia*. 2012;18(6):260-266. doi:10.1016/j.rppneu.2012.04.006
185. Lopes AJ, Costa W, Thomaz Mafort T, de Sá Ferreira A, Silveira de Menezes SL, Silva Guimarães F. Silicosis in sandblasters of shipyard versus silicosis in stone carvers in Brazil: A comparison of imaging findings, lung function variables and cardiopulmonary exercise testing parameters. *Revista Portuguesa de Pneumologia (English Edition)*. 2012;18(6):260-266. doi:10.1016/j.rppnen.2012.06.002
186. Lima JGM, Oliveira Filho GR, Lima MTBRM, Ferreira AS, Silva JG. Influence of low intensity laser therapy (AsGa) on the cicatrization process of mechanic tendon injury in wistar rats. *Laser Physics*. 2012;22(9):1445-1448. doi:10.1134/s1054660x12090083
187. Fragoso A. Evaluation of the immediate effects of manual acupuncture on brachial bicep muscle function in healthy individuals and poststroke patients: a study protocol of a parallel-group randomized clinical trial. *Journal of Chinese Integrative Medicine*. 2012;10(3):303-309. doi:10.3736/jcim20120309
188. Fragoso APS, Ferreira AS. Immediate effects of acupuncture on biceps brachii muscle function in healthy and post-stroke subjects. *Chinese Medicine*. 2012;7(1):7. doi:10.1186/1749-8546-7-7
189. Ferreira A de S, Oliveira JF de, Cordovil I, Barbosa Filho J. Quadriceps short-term resistance exercise in subjects with resistant hypertension. *Fisioterapia em Movimento*. 2011;24(4):629-636. doi:10.1590/s0103-51502011000400006

190. Luiz AB, Cordovil I, Filho JB, Ferreira AS. Zangfu zheng (patterns) are associated with clinical manifestations of zang shang (target-organ damage) in arterial hypertension. *Chinese Medicine*. 2011;6(1). doi:10.1186/1749-8546-6-23
191. Ferreira AS, Lopes AJ. Chinese medicine pattern differentiation and its implications for clinical practice. *Chinese Journal of Integrative Medicine*. 2011;17(11):818-823. doi:10.1007/s11655-011-0892-y
192. Ferreira A de S, Filho JB, Souza MN de. Model for post-occlusive reactive hyperemia as measured noninvasively with pressure pulse waveform. *Biomedical Signal Processing and Control*. 2011;6(4):410-413. doi:10.1016/j.bspc.2010.11.003
193. Povoa LC, Vanuzzi FK, Ferreira APA, Ferreira A de S. Intervenção osteopática em idosos e o impacto na qualidade de vida. *Fisioterapia em Movimento*. 2011;24(3):429-436. doi:10.1590/s0103-515020110003 00007
194. Silva AL dos S, Marinho MRC, Gouveia FM de V, Silva JG, Ferreira A de S, Cal R. Benign Paroxysmal Positional Vertigo: comparison of two recent international guidelines. *Brazilian Journal of Otorhinolaryngology*. 2011;77(2):191-200. doi:10.1590/s1808-86942011000200009
195. Mainenti MRM, de Carvalho Rodrigues É, de Oliveira JF, de Sá Ferreira A, Dias CM, dos Santos Silva AL. Adiposity and postural balance control: Correlations between bioelectrical impedance and stabilometric signals in elderly Brazilian women. *Clinics*. 2011;66(9):1513-1518. doi:10.1590/s1807-59322011000900 001
196. Sá Ferreira A. Misdiagnosis and undiagnosis due to pattern similarity in Chinese medicine: a stochastic simulation study using pattern differentiation algorithm. *Chinese Medicine*. 2011;6(1):1. doi:10.1186/1749-8546-6-1
197. Ferreira A de S, Gave N de S, Abrahão F, Silva JG. Influência da morfologia de pés e joelhos no equilíbrio durante apoio bipodal. *Fisioterapia em Movimento*. 2010;23(2):193-200. doi:10.1590/s0103-5150201000 0200003
198. Silva JG, Antonioli R de S, Orsini M, Santos Júnior MAJ dos, Ferreira A de S. Mobilização do osso pisiforme no tratamento da neuropraxia do nervo ulnar no canal de Guyon: relato de caso. *Fisioterapia e Pesquisa*. 2009;16(4):363-367. doi:10.1590/s1809-29502009000400014
199. Ferre D. Statistical validation of strategies for Zang-Fu single patterndifferentiation. *Journal of Chinese Integrative Medicine*. 2008;6(11):1109-1116. doi:10.3736/jcim20081103
200. Peng Q. Effects of Huoxue Tongmai Lishui method on fundus fluorescein angiography of non-ischemic retinal vein occlusion: a randomized controlled trial. *Journal of Chinese Integrative Medicine*. 2009;7(11):1035-1041. doi:10.3736/jcim20091103
201. Ferreira A. Prophylactic effects of short-term acupuncture on Zusanli (ST36) in Wistar rats with lipopolysaccharide-induced acute lung injury. *Journal of Chinese Integrative Medicine*. 2009;7(10):969-975. doi:10.3736/jcim20091011
202. de Sá Ferreira A, Filho JB, Cordovil I, de Souza MN. Three-section transmission-line arterial model for noninvasive assessment of vascular remodeling in primary hypertension. *Biomedical Signal Processing and Control*. 2009;4(1):2-6. doi:10.1016/j.bspc.2008.07.001
203. Ferreira A. Diagnostic accuracy of pattern differentiation algorithm based on Chinese medicine theory: a stochastic simulation study. *Chinese Medicine*. 2009;4(1):24. doi:10.1186/1749-8546-4-24
204. Ferreira A de S, Souza MN de, Filho JB. Avaliação de um modelo de parâmetros distribuídos aplicado à estimativa da geometria arterial na hipertensão. *Revista Brasileira de Engenharia Biomédica*. 2008;24(3):193-200. doi:10.4322/rbeb.2012.057

205. Ferreira AS, Santos MAR, Filho JB, Cordovil I, Souza MN. Determination of radial artery compliance can increase the diagnostic power of pulse wave velocity measurement. *Physiological Measurement*. 2003;25(1):37-50. doi:10.1088/0967-3334/25/1/004
206. Oliveira CL, Ferreira AS. Reabilitação de pessoas com amputação de membros inferiores: uma revisão sistematizada. *Ciência em Movimento*. 2022;23(48):99-109. doi:10.15602/1983-9480/cm.v23n48p99-109

Preprints

1. Sancho AG, Ferreira AS. Validation of SAUDE: A Citizen Science App for Functional Balance Assessment in Older Adults. February 2026. doi:10.21203/rs.3.rs-8627904/v1
2. Kandakurti PK, arulsingh watson, Ferreira A de S, Patil SS. C5-C6 And Thoracic Spine Mobilization With Postural Correction Exercise Compared With Conventional Therapy In Patients With Adhesive Capsulitis , A Two-Group, Parallel-Arm, Single-Blinded, Randomized Clinical Trial- Study Protocol. November 2025. doi:10.21203/rs.3.rs-7551921/v1
3. Monteiro ER, Aguilera LM, Ruá-Alonso M, et al. Effect of Manual Massage, Foam Rolling, and Strength Training on Cardiovascular Responses in Adults: A Scoping Review. February 2025. doi:10.20944/preprints202502.1932.v1
4. Menezes K, B. Lima MA, Xerez DR, et al. RETURN OF VOLUNTARY MOTOR CONTRACTION AFTER COMPLETE SPINAL CORD INJURY: A PILOT HUMAN STUDY ON POLYLMININ. February 2024. doi:10.1101/2024.02.19.24301010
5. Alaparthi GK, Gatty AP, Isa H, et al. Reference Values for the Grocery Shelving Test among United Arab Emirates Population: A Cross Sectional Study. June 2023. doi:10.20944/preprints202305.1684.v3
6. Alaparthi GK, Gatty AP, Isa H, et al. Reference Values for the Grocery Shelving Test among United Arab Emirates Population: A Cross Sectional Study. May 2023. doi:10.20944/preprints202305.1684.v2
7. Alaparthi GK, Gatty AP, Isa H, et al. Reference Values for the Grocery Shelving Test among United Arab Emirates Population: A Cross Sectional Study. May 2023. doi:10.20944/preprints202305.1684.v1

Resumos publicados em eventos científicos

1. Ferreira AET, Almeida APGS de, Vidal F de B. Autonomous Vehicle Steering Wheel Estimation from a Video using Multichannel Convolutional Neural Networks. Proceedings of the 15th International Conference on Informatics in Control, Automation and Robotics. 2018:517-524. doi:10.5220/0006920605170524
2. de Araujo Ferreira N, Lopes AJ, de Sá Ferreira A, Oliveira Dias J, Silva Guimarães F. Fatores determinantes do prognóstico funcional do paciente crítico sob intervenção fisioterapeutica. Anais do Congresso Internacional de Qualidade em Serviços e Sistemas de Saúde. May 2017. doi:10.17648/qualihosp-2017-69645
3. Ferreira AS, Filho JB, Souza MN. Simplified Distributed-Parametermodel of Brachial-Radial Arteries for Noninvasive Determination of Mechanical Characteristics of Vessel. 2006 International Conference of the IEEE Engineering in Medicine and Biology Society. August 2006:1814-1817. doi:10.1109/emb.2006.260102
4. Ferreira AS, Filho JB, Souza MN. Comparison of segmental arterial compliance determined with three and four element Windkessel models. Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat No03CH37439).:3161-3164. doi:10.1109/emb.2003.1280813
5. Ferreira AS, Filho JB, Souza MN. Identification of vascular parameters based on the same pressure pulses waves used to measure pulse wave velocity. 2001 Conference Proceedings of the 23rd Annual International

Conference of the IEEE Engineering in Medicine and Biology Society. 4:3418-3421. doi:10.1109/emb.2001.1019564

Livros editorados

1. Vigário P dos S, Ferreira A de S. DESENVOLVIMENTO LOCAL E OS DESAFIOS PARA O ALCANCE DO DESENVOLVIMENTO SUSTENTÁVEL SOB O PRISMA DA AGENDA 2030. 2024. doi:10.47879/ed.ep.2024745

Capítulos de livro

1. Luna L de O, Tavares EM, Cezário BS, Guedes ALA, Ferreira A de S. A IMPORTÂNCIA DE MEDI-DAS PROTETIVAS NA FLORESTA AMAZÔNICA PARA A PRESERVAÇÃO DA BIODIVERSIDADE - ODS 15. DESENVOLVIMENTO LOCAL E OS DESAFIOS PARA O ALCANCE DO DESENVOLVIMENTO SUSTENTÁVEL SOB O PRISMA DA AGENDA 2030. 2024:135-148. doi:10.47879/ed.ep.2024745p135
2. Tavares EM, Luna L de O, Cezário BS, Guedes ALA, Ferreira A de S. CIDADES INTELIGENTES: UMA REVISÃO DA RELAÇÃO ENTRE A TECNOLOGIA E A MOBILIDADE URBANA SUSTENTÁVEL. DESENVOLVIMENTO LOCAL E OS DESAFIOS PARA O ALCANCE DO DESENVOLVIMENTO SUSTENTÁVEL SOB O PRISMA DA AGENDA 2030. 2024:106-114. doi:10.47879/ed.ep.2024745p106
3. Ferreira AS, Oliveira IJAS. Methods for Assessment of Interrater Reliability for Diagnosis and Intervention in Traditional Chinese Medicine Studies. Evidence-based Research Methods for Chinese Medicine. 2016:89-111. doi:10.1007/978-981-10-2290-6_7
4. Sa Ferreira A de. Advances in Chinese Medicine Diagnosis: From Traditional Methods to Computational Models. Recent Advances in Theories and Practice of Chinese Medicine. January 2012. doi:10.5772/27703

Aulas

1. Laett CT, de Sá Ferreira A. Instrumentação Virtual e Tecnologia em Reabilitação. OSF. 2025. doi:10.17605/OSF.IO/XR784
2. De Sá Ferreira A, Laett CT. Bioestatística I. 2024. doi:10.17605/OSF.IO/YZ8WG
3. Laett CT, de Sá Ferreira A. Bioestatística II. 2024. doi:10.17605/OSF.IO/C7VKP
4. Thiago Lemos, Da Hora AL, Felix BP, et al. Didática e Prática do Ensino Superior. 2024. doi:10.17605/OSF.IO/GHXUD

Programas de computador

1. Ferreira Ade S, Meziat Filho N. RCTapp. Zenodo; 2025. doi:10.5281/ZENODO.13848815
2. Ferreira A. UsIA | Ultrasound Image Analysis. Zenodo; 2023. doi:10.5281/ZENODO.10439719
3. Arthur de Sá Ferreira. Observatório. Zenodo; 2023. doi:10.5281/ZENODO.8322622
4. De Sá Ferreira A. Ciência com R. setembro 2023. doi:10.5281/ZENODO.8320233
5. Ferreira AS. SuiteEBG. Zenodo; 2023. doi:10.5281/ZENODO.8210025
6. Ferreira Ade S. SuiteMYO. Zenodo; 2023. doi:10.5281/ZENODO.8211266
7. Ferreira Ade S. SuiteTCM. Zenodo; 2023. doi:10.5281/ZENODO.8211409

8. Ferreira Ade S. wiiVIEW. Zenodo; 2019. doi:10.5281/ZENODO.8233112

Bancos de dados

1. Alaparthi G, Moustafa I, Lopes A, Ferreira A. Data and code for “Pulmonary function, body posture and balance in young adults with asthma: A cross-sectional study”. dezembro 2024. doi:10.17632/PT5SCMP33Z.1
2. Abrahão P, Meziat-Filho N, Rodrigues E, Pinto T, Ferreira A. Dataset and code for “Physiotherapists’ attitudes and beliefs about pain treatment influence their visual assessment of posture in individuals with neck pain”. outubro 2024. doi:10.17632/7CBWM57PCD.1
3. Arthur Ferreira. Data for “Locating the Seventh Cervical Spinous Process: Accuracy of the Thorax-Rib Static Method and the Effects of Clinical Data on Its Performance”. setembro 2022. doi:10.17632/DXJ78TYYHJ.2
4. Ferreira A. Dataset for “Accuracy and learning curves of inexperienced observers for manual segmentation of electromyograms”. janeiro 2021. doi:10.17632/5HBFNNJFFF.1
5. Ferreira A. Dataset for “Feasibility of whole-body gait kinematics to assess the validity of the six-minute walk test over a 10-m walkway in the elderly”. janeiro 2021. doi:10.17632/47S5JDW7VN.1
6. Ferreira A. Dataset for “Kinematic evaluation of patients with chronic obstructive pulmonary disease during the 6-min walk test”. janeiro 2021. doi:10.17632/5C97XJMCW2.1
7. Ferreira A. Data for “Agreement and predictive power of six fall risk assessment methods in community-dwelling older adults”. janeiro 2021. doi:10.17632/3D4VR4DWJS.3
8. Ferreira A. Dataset for “Lower limb muscle fatigability is not associated with changes in movement strategies for balance control in the upright stance”. janeiro 2021. doi:10.17632/XBGX8GBYMS.4
9. Ferreira A. Dataset for “Optimization of postural stability after repeated motor tasks with visual feedback in the upright stance”. janeiro 2021. doi:10.17632/TB5FMH8ZNJ.1
10. Ferreira A. Data for: Pragmatic combinations of acupuncture points for lateral epicondylalgia are unreliable in the physiotherapy setting. novembro 2018. doi:10.17632/6WYN9KTX7F.1
11. Ferreira A. Inter-expert agreement and similarity analysis of traditional diagnoses and acupuncture prescriptions in textbook- and pragmatic-based practices. dezembro 2017. doi:10.17632/KBGSNJTBVY.1

Outros

1. de Sá Ferreira A, dos Santos Vigário P. Livro de anais do IX Simpósio Paradesportivo Carioca. OSF. 2026. doi:10.17605/OSF.IO/92ZNR
2. Williane Nascimento, Rosa F, de Sá Ferreira A. Tecnologias digitais na periferia: Uma perspectiva para o desenvolvimento multidimensional de populações em situação de vulnerabilidade socioeconômica. 2025. doi:10.17605/OSF.IO/8CWJ5
3. de Sá Ferreira A. Advanced Computational Methods for Studying Postural Balance in Health and Neuromuscular Diseases. OSF. 2024. doi:10.17605/OSF.IO/UNPZW
4. de Sá Ferreira A. Cardiac Rehabilitation: Current Evidence and Future Perspectives. OSF. 2024. doi:10.17605/OSF.IO/SQ3WU
5. de Sá Ferreira A. Do balanço postural ao processamento e análise posturográfica. OSF. 2024. doi:10.17605/OSF.IO/V6U48
6. de Sá Ferreira A. Inteligência Artificial na Academia. OSF. 2024. doi:10.17605/OSF.IO/EYSPK

7. de Sá Ferreira A. Inteligência Artificial na Academia. 2024. doi:10.17605/OSF.IO/GFEQ5
8. de Sá Ferreira A. Inteligência Artificial na Academia: Da Teoria às Aplicações. OSF. 2024. doi:10.17605/OSF.IO/H8JBD
9. de Sá Ferreira A. Inteligência Artificial na Saúde. OSF. 2024. doi:10.17605/OSF.IO/HVY9W
10. de Sá Ferreira A, Bittencourt JV. Universidade Corporativa. OSF. 2024. doi:10.17605/OSF.IO/2FCVX
11. de Sá Ferreira A. Uso de Inteligência Artificial em Pesquisas. OSF. 2024. doi:10.17605/OSF.IO/XTYGW

Fontes externas

Fontes de informação externas

American Heart Association

- *Statistical Reporting Recommendations - AHA/ASA journals*²

American Physiological Society

- *Statistics*³
- *Exploration in Statistics*⁴
- *General Statistics*⁵
- *Reporting Statistics*⁶

American Statistical Association

- *Statistical Inference in the 21st Century: A World Beyond $p < 0.05$* - The American Statistical Association⁷

British Medicine Journal

- *Statistics - Latest from The BMJ*⁸
- *Statistics notes - Latest from The BMJ*⁹
- *Statistics and research methods - Latest from The BMJ*¹⁰
- *Statistics at Square One*¹¹
- *Research methods & reporting*¹²

²<https://www.ahajournals.org/statistical-recommendations>

³<https://journals.physiology.org/topic/advances-collections/statistics?seriesKey=&tagCode=&>

⁴<https://journals.physiology.org/topic/advances-collections/explorations-in-statistics?seriesKey=&tagCode=&>

⁵<https://journals.physiology.org/topic/advances-collections/general-statistics?seriesKey=&tagCode=&>

⁶<https://journals.physiology.org/topic/advances-collections/reporting-statistics?seriesKey=&tagCode=&>

⁷<https://www.tandfonline.com/toc/utas20/73/sup1?nav=tocList>

⁸<https://www.bmjjournals.org/journal/bmjspecialties>

⁹<https://www.bmjjournals.org/journal/bmjspecialties/section/statistics-notes>

¹⁰<https://www.bmjjournals.org/journal/bmjspecialties/section/statistics-and-research-methods>

¹¹<https://www.bmjjournals.org/journal/bmjspecialties/section/statistics-and-research-methods>

¹²<https://www.bmjjournals.org/journal/bmjspecialties/section/statistics-and-research-methods>

Enhancing the QUality And Transparency Of health Research Network

- *Enhancing the Quality and Transparency of health research EQUATOR Network*¹³

Journal of the American Medical Association

- *JAMA Guide to Statistics and Methods - JAMA*¹⁴

Nature Publishing Group

- *Statistics for Biologists - Nature Publishing Group*¹⁵

Oxford Reference

- *A Dictionary of Statistics*¹⁶

Royal Statistical Society

- *Best Practices for Data Visualisation - Royal Statistical Society*¹⁷

Statistics in Medicine

- *Tutorials in Biostatistics Papers*¹⁸

BMC Trials

- *Design and analysis of n-of-1 trials*¹⁹

The Journal of Applied Statistics in the Pharmaceutical Industry

- *Tutorial Papers*²⁰

¹³<https://www.equator-network.org>

¹⁴<https://jamanetwork.com/collections/44042/jama-guide-to-statistics-and-methods>

¹⁵<https://www.nature.com/collections/qghqm>

¹⁶<https://www.oxfordreference.com/display/10.1093/acref/9780199679188.001.0001/acref-9780199679188>

¹⁷<https://royal-statistical-society.github.io/datavisguide>

¹⁸<https://onlinelibrary.wiley.com/page/journal/10970258/homepage/tutorials.htm>

¹⁹<https://www.biomedcentral.com/collections/DANT>

²⁰https://onlinelibrary.wiley.com/journal/15391612/homepage/tutorial_papers.htm

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. Sandelowski M. Combining Qualitative and Quantitative Sampling, Data Collection, and Analysis Techniques in Mixed-Method Studies. *Research in Nursing & Health*. 2000;23(3):246–255. doi:10.1002/1098-240x(200006)23:3<246::aid-nur9>3.0.co;2-h
24. 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
25. 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
26. 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
27. Resnik DB, Shamoo AE. Reproducibility and Research Integrity. *Accountability in Research*. 2016;24(2):116–123. doi:10.1080/08989621.2016.1257387
28. 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
29. Mair P. Thou Shalt Be Reproducible! A Technology Perspective. *Frontiers in Psychology*. 2016;7. doi:10.3389/fpsyg.2016.01079

30. 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
31. 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
32. Kahan BC, Hindley J, Edwards M, Cro S, Morris TP. The estimands framework: a primer on the ICH E9(R1) addendum. *BMJ*. janeiro 2024:e076316. doi:10.1136/bmj-2023-076316
33. Polit DF, Beck CT. Generalization in quantitative and qualitative research: Myths and strategies. *International Journal of Nursing Studies*. 2010;47(11):1451–1458. doi:10.1016/j.ijnurstu.2010.06.004
34. 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
35. 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
36. R Core Team. The Comprehensive R Archive Network. 2021. <https://cran.r-project.org>.
37. Allaire J, Xie Y, Dervieux C, et al. *rmarkdown: Dynamic Documents for R*; 2023. <https://CRAN.R-project.org/package=rmarkdown>.
38. 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
39. 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
40. 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
41. Selker R, Love J, Dropmann D. *jmv: The {jamovi} Analyses*; 2023. <https://CRAN.R-project.org/package=jmv>.
42. Love J. *jmvconnect: Connect to the {jamovi} Statistical Spreadsheet*; 2022. <https://CRAN.R-project.org/package=jmvconnect>.
43. 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
44. Aden-Buie G, Schloerke B, Allaire J, Rossell Hayes A. *learnr: Interactive Tutorials for R*; 2023. <https://CRAN.R-project.org/package=learnr>.
45. Schwab, Simon, Held, Leonhard. Statistical programming: Small mistakes, big impacts. *Wiley-Blackwell Publishing, Inc.* 2021. doi:10.5167/UZH-205154
46. 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

47. Xie Y. *formatR: Format R Code Automatically.*; 2022. <https://CRAN.R-project.org/package=formatR>.
48. Müller K, Walthert L. *styler: Non-Invasive Pretty Printing of R Code.*; 2023. <https://CRAN.R-project.org/package=styler>.
49. Hester J, Angly F, Hyde R, et al. *lintr: A {Linter} for R Code.*; 2023. <https://CRAN.R-project.org/package=lintr>.
50. All R CRAN packages [Full List]. 2025. <https://r-packages.io/packages>. Acessado fevereiro 11, 2025.
51. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2023. <https://www.R-project.org/>.
52. Wickham H, Danenberg P, Csárdi G, Eugster M. *roxygen2: In-Line Documentation for R.*; 2024. doi:10.32614/CRAN.package.roxygen2
53. 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
54. Gohel D, Ross N. *officedown: Enhanced {R Markdown} Format for {Word} and {PowerPoint}.*; 2023. <https://CRAN.R-project.org/package=officedown>.
55. Xie Y. *bookdown: Authoring Books and Technical Documents with R Markdown*. Chapman; Hall/CRC; 2023. <https://bookdown.org/yihui/bookdown/>.
56. Ioannidis JPA. How to Make More Published Research True. *PLoS Medicine*. 2014;11(10):e1001747. doi:10.1371/journal.pmed.1001747
57. Krieger N, Perzynski A, Dalton J. *projects: A Project Infrastructure for Researchers.*; 2021. <https://CRAN.R-project.org/package=projects>.
58. 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
59. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2023. <https://www.R-project.org/>.
60. 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
61. Francisco Rodríguez-Sánchez, Connor P. Jackson, Shaurita D. Hutchins. *grateful: Facilitate citation of R packages.*; 2023. <https://github.com/Pakillo/grateful>.
62. Shields M. Information Literacy, Statistical Literacy, Data Literacy. *IASSIST Quarterly*. 2005;28(2):6. doi:10.29173/iq790
63. 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
64. 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

65. 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
66. GOULD R. DATA LITERACY IS STATISTICAL LITERACY. *STATISTICS EDUCATION RESEARCH JOURNAL*. 2017;16(1):22–25. doi:10.52041/serj.v16i1.209
67. 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
68. 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
69. 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
70. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. agosto 2019;l4898. doi:10.1136/bmj.l4898
71. 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
72. 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
73. 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
74. 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
75. 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
76. Meng XL. Statistical paradises 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
77. 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
78. Berkson J. Limitations of the Application of Fourfold Table Analysis to Hospital Data. *Biometrics Bulletin*. 1946;2(3):47. doi:10.2307/3002000
79. Ellsberg D. Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics*. 1961;75(4):643. doi:10.2307/1884324
80. Freedman DA, Freedman DA. A Note on Screening Regression Equations. *The American Statistician*. 1983;37(2):152–155. doi:10.1080/00031305.1983.10482729

81. Freedman LS, Pee D. Return to a Note on Screening Regression Equations. *The American Statistician*. 1989;43(4):279. doi:10.2307/2685389
82. Hand DJ. On Comparing Two Treatments. *The American Statistician*. 1992;46(3):190–192. doi:10.1080/00031305.1992.10475881
83. LINDLEY DV. A STATISTICAL PARADOX. *Biometrika*. 1957;44(1-2):187–192. doi:10.1093/biomet/44.1-2.187
84. Lord FM. A paradox in the interpretation of group comparisons. *Psychological Bulletin*. 1967;68(5):304–305. doi:10.1037/h0025105
85. Lord FM. Statistical adjustments when comparing preexisting groups. *Psychological Bulletin*. 1969;72(5):336–337. doi:10.1037/h0028108
86. 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
87. 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
88. Pearl J. Comment: Understanding Simpson's Paradox. *The American Statistician*. 2014;68(1):8–13. doi:10.1080/00031305.2014.876829
89. 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
90. James W, Stein C. Estimation with Quadratic Loss. Em: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. Berkeley, Calif.: University of California Press; 1961:361–379. <http://projecteuclid.org/euclid.bsmsp/1200512173>.
91. De S, Sen A. The generalised Gamow-Stern problem. *The Mathematical Gazette*. 1996;80(488):345–348. doi:10.2307/3619568
92. Feld SL. Why Your Friends Have More Friends Than You Do. *American Journal of Sociology*. 1991;96(6):1464–1477. doi:10.1086/229693
93. 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
94. 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
95. 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

96. Kleinert S. COPE's retraction guidelines. *The Lancet*. 2009;374(9705):1876–1877. doi:10.1016/s0140-6736(09)62074-2
97. Kerr NL. HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*. 1998;2(3):196–217. doi:10.1207/s15327957pspr0203_4
98. 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
99. 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
100. 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
101. 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
102. 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
103. 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
104. 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
105. Horton R. The rhetoric of research. *BMJ*. 1995;310(6985):985–987. doi:10.1136/bmj.310.6985.985
106. 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
107. 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
108. 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
109. 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
110. Simmons J, Nelson L, Simonsohn U. Pre-registration: Why and How. *Journal of Consumer Psychology*. 2021;31(1):151–162. doi:10.1002/jcpy.1208
111. Hartgerink C, Aust F. *retractcheck: Retraction Scanner*; 2025. <https://github.com/chartgerink/retractcheck>.

112. Altman DG, Bland JM. Statistics notes Variables and parameters. *BMJ*. 1999;318(7199):1667–1667. doi:10.1136/bmj.318.7199.1667
113. Vetter TR. Fundamentals of Research Data and Variables. *Anesthesia & Analgesia*. 2017;125(4):1375–1380. doi:10.1213/ane.0000000000002370
114. 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
115. Dettori JR, Norvell DC. The Anatomy of Data. *Global Spine Journal*. 2018;8(3):311–313. doi:10.1177/2192568217746998
116. 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
117. Barkan H. Statistics in clinical research: Important considerations. *Annals of Cardiac Anaesthesia*. 2015;18(1):74. doi:10.4103/0971-9784.148325
118. Bland JM, Altman DG. Statistics Notes: Transforming data. *BMJ*. 1996;312(7033):770–770. doi:10.1136/bmj.312.7033.770
119. Fedorov V, Mannino F, Zhang R. Consequences of dichotomization. *Pharmaceutical Statistics*. 2009;8(1):50–61. doi:10.1002/pst.331
120. Osborne J. Improving your data transformations: Applying the Box-Cox transformation. *University of Massachusetts Amherst*. 2010. doi:10.7275/QBPC-GK17
121. 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
122. Venables WN, Ripley BD. *Modern Applied Statistics with S*. Springer; 2002. <https://www.stats.ox.ac.uk/pub/MASS4/>.
123. 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
124. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*. 2006;332(7549):1080.1. doi:10.1136/bmj.332.7549.1080
125. 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
126. 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
127. 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

128. 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
129. Barnier J, Briatte F, Larmarange J. *questionr: Functions to Make Surveys Processing Easier.*; 2023. <https://CRAN.R-project.org/package=questionr>.
130. Aguinis H, Pierce CA, Culpepper SA. Scale Coarseness as a Methodological Artifact. *Organizational Research Methods*. 2008;12(4):623–652. doi:10.1177/1094428108318065
131. 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
132. 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
133. 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
134. 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
135. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological Bulletin*. 1971;76(5):378–382. doi:10.1037/h0031619
136. R Core Team. *R: A Language and Environment for Statistical Computing.*; 2025. <https://www.R-project.org/>.
137. Olson K. What Are Data? *Qualitative Health Research*. 2021;31(9):1567–1569. doi:10.1177/10497323211015960
138. 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
139. 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
140. 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
141. Pebesma E, Mailund T, Hiebert J. Measurement Units in {R}. *The R Journal*. 2016;8. doi:10.32614/RJ-2016-061
142. Firke S. *janitor: Simple Tools for Examining and Cleaning Dirty Data.*; 2023. <https://CRAN.R-project.org/package=janitor>.
143. Harrell Jr FE. *Hmisc: Harrell Miscellaneous.*; 2023. <https://CRAN.R-project.org/package=Hmisc>.

144. Reese A. Databases and Documenting Data. *Significance*. 2007;4(4):184–186. doi:10.1111/j.1740-9713.2007.00265.x
145. Bryer J, Speerschneider K. *likert: Analysis and Visualization Likert Items.*; 2016. <https://CRAN.R-project.org/package=likert>.
146. Larmarange J. *ggstats: Extension to ggplot2 for Plotting Stats.*; 2025. doi:10.32614/CRAN.package.ggstats
147. Ferris TLJ. A new definition of measurement. *Measurement*. 2004;36(1):101–109. doi:10.1016/j.measurement.2004.03.001
148. Healy MJR, Goldstein H. Regression to the mean. *Annals of Human Biology*. 1978;5(3):277–280. doi:10.1080/03014467800002891
149. Altman DG, Bland JM. Measurement in Medicine: The Analysis of Method Comparison Studies. *The Statistician*. 1983;32(3):307. doi:10.2307/2987937
150. 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
151. 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
152. 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
153. Hammill D. *DataEditR: An Interactive Editor for Viewing, Entering, Filtering & Editing Data.*; 2022. <https://CRAN.R-project.org/package=DataEditR>.
154. Broman KW, Woo KH. Data Organization in Spreadsheets. *The American Statistician*. 2018;72(1):2–10. doi:10.1080/00031305.2017.1375989
155. 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
156. Dowle M, Srinivasan A. *data.table: Extension of ‘data.frame’.*; 2023. <https://CRAN.R-project.org/package=data.table>.
157. Altman DG, Bland JM. Missing data. *BMJ*. 2007;334(7590):424–424. doi:10.1136/bmj.38977.682025.2c
158. Heymans MW, Twisk JWR. Handling missing data in clinical research. *Journal of Clinical Epidemiology*. setember 2022. doi:10.1016/j.jclinepi.2022.08.016
159. Carpenter JR, Smuk M. Missing data: A statistical framework for practice. *Biometrical Journal*. 2021;63(5):915–947. doi:10.1002/bimj.202000196
160. Yanagida T. *misty: Miscellaneous Functions.*; 2023. <https://CRAN.R-project.org/package=misty>.

161. 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
162. 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
163. 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
164. 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
165. 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
166. 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
167. Little RJA. Missing-Data Adjustments in Large Surveys. *Journal of Business & Economic Statistics*. 1988;6(3):287–296. doi:10.1080/07350015.1988.10509663
168. Robitzsch A, Grund S. *miceadds: Some Additional Multiple Imputation Functions, Especially for {mice}*; 2023. <https://CRAN.R-project.org/package=miceadds>.
169. FitzJohn R. *ids: Generate Random Identifiers*; 2017. <https://CRAN.R-project.org/package=ids>.
170. Brown C. *hash: Full Featured Implementation of Hash Tables/Associative Arrays/Dictionaries*; 2023. <https://CRAN.R-project.org/package=hash>.
171. Hendricks P. *anonymizer: Anonymize Data Containing Personally Identifiable Information*; 2023. <https://github.com/paulhendricks/anonymizer>.
172. 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>.
173. 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
174. S M. Frequency distribution. *Journal of Pharmacology and Pharmacotherapeutics*. 2011;2(1):54–56. doi:10.4103/0976-500x.77120
175. 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
176. SCOTT DW. On optimal and data-based histograms. *Biometrika*. 1979;66(3):605–610. doi:10.1093/biomet/66.3.605

177. 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
178. R Core Team. *R: A Language and Environment for Statistical Computing.*; 2023. <https://www.R-project.org/>.
179. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer; 2016. <https://ggplot2.tidyverse.org>.
180. 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
181. 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
182. 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
183. 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
184. 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
185. Kanji G. *100 Statistical Tests*. SAGE Publications Ltd; 2006. doi:10.4135/9781849208499
186. 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
187. 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
188. Krzywinski M, Altman N. Error bars. *Nature Methods*. 2013;10(10):921–922. doi:10.1038/nmeth.2659
189. 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
190. S. M. Measures of central tendency: The mean. *Journal of Pharmacology and Pharmacotherapeutics*. 2011;2(2):140–142. doi:10.4103/0976-500x.81920
191. 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
192. Manikandan S. Measures of dispersion. *Journal of Pharmacology and Pharmacotherapeutics*. 2011;2(4):315–316. doi:10.4103/0976-500x.85931
193. Sahai H, Misra S. Definitions of Sample Variance: Some Teaching Problems to be Overcome. *The Statistician*. 1992;41(1):55. doi:10.2307/2348636

194. 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
195. 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
196. 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
197. Chatfield C. Exploratory data analysis. *European Journal of Operational Research*. 1986;23(1):5–13. doi:10.1016/0377-2217(86)90209-2
198. Ferketich S, Verran J. Technical Notes. *Western Journal of Nursing Research*. 1986;8(4):464–466. doi:10.1177/019394598600800409
199. 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
200. 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
201. 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
202. Krasser R. *explore: Simplifies Exploratory Data Analysis.*; 2023. <https://CRAN.R-project.org/package=explore>.
203. 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
204. Cui B. *DataExplorer: Automate Data Exploration and Treatment.*; 2020. <https://CRAN.R-project.org/package=DataExplorer>.
205. Dayanand Ubrangala, R K, Prasad Kondapalli R, Putatunda S. *SmartEDA: Summarize and Explore the Data.*; 2022. <https://CRAN.R-project.org/package=SmartEDA>.
206. Mock T. *gtExtras: Extending {gt} for Beautiful HTML Tables.*; 2023. <https://CRAN.R-project.org/package=gtExtras>.
207. Nijs V. *radiant: Business Analytics using R and Shiny.*; 2023. <https://CRAN.R-project.org/package=radiant>.
208. Behrens JT. Principles and procedures of exploratory data analysis. *Psychological Methods*. 1997;2(2):131–160. doi:10.1037/1082-989x.2.2.131
209. Prunello M, Mari G. *ggcleveland: Implementation of Plots from Cleveland's Visualizing Data Book.*; 2021. doi:10.32614/CRAN.package.ggcleveland

210. Gerring J. Mere Description. *British Journal of Political Science*. 2012;42(4):721–746. doi:10.1017/s0007123412000130
211. 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
212. 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
213. 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
214. Weissgerber TL, Winham SJ, Heinzen EP, et al. Reveal, Don't Conceal. *Circulation*. 2019;140(18):1506–1518. doi:10.1161/circulationaha.118.037777
215. 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
216. 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
217. 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
218. Rich B. *table1: Tables of Descriptive Statistics in HTML.*; 2023. <https://CRAN.R-project.org/package=table1>.
219. Gohel D, Skintzos P. *flextable: Functions for Tabular Reporting.*; 2023. <https://CRAN.R-project.org/package=flextable>.
220. Thériault R. {rempscy}: Convenience functions for psychology. *Journal of Open Source Software*. 2023;8:5466. doi:10.21105/joss.05466
221. 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
222. 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
223. 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
224. 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
225. 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
226. 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

227. Midway SR. Principles of Effective Data Visualization. *Patterns.* 2020;1(9):100141. doi:10.1016/j.patter.2020.100141
228. 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
229. 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
230. Sievert C. *Interactive Web-Based Data Visualization with R, plotly, and shiny.* Chapman; Hall/CRC; 2020. <https://plotly-r.com>.
231. Wei T, Simko V. *R package corrplot: Visualization of a Correlation Matrix.*; 2024. <https://github.com/taiyun/corrplot>.
232. Wise EA, Adams RJ, Lyketsos CG, Leoutsakos JM. Graphical methods for understanding changes in states: Understanding medication use pathways. *International Journal of Methods in Psychiatric Research.* 2022;31(4). doi:10.1002/mpr.1932
233. Swihart BJ, Caffo B, James BD, Strand M, Schwartz BS, Punjabi NM. Lasagna Plots. *Epidemiology.* 2010;21(5):621–625. doi:10.1097/ede.0b013e3181e5b06a
234. Xiao N. *ggsci: Scientific Journal and Sci-Fi Themed Color Palettes for {ggplot2}.*; 2023. <https://CRAN.R-project.org/package=ggsci>.
235. Urbanek S, Johnson K. *tiff: Read and Write TIFF Images.*; 2022. <https://CRAN.R-project.org/package=tiff>.
236. 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
237. Pedersen TL, Robinson D. *ganimate: A Grammar of Animated Graphics.*; 2025. doi:10.32614/CRAN.package.ganimate
238. 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
239. 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
240. 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
241. 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.

242. Komsta L. *outliers: Tests for Outliers.*; 2022. <https://CRAN.R-project.org/package=outliers>.
243. 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
244. Mair P, Wilcox R, Indrajeet P. *A Collection of Robust Statistical Methods.*; 2025. <https://CRAN.R-project.org/package=WRS2>.
245. 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
246. 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
247. 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
248. 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
249. 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
250. 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
251. 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
252. 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
253. Rothman KJ, Greenland S. Hill's Criteria for Causality. *Encyclopedia of Biostatistics*. fevereiro 2005. doi:10.1002/0470011815.b2a03072
254. 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
255. 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
256. Barrett M. *ggdag: Analyze and Create Elegant Directed Acyclic Graphs.*; 2024. <https://CRAN.R-project.org/package=ggdag>.
257. 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

258. 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
259. 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
260. 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
261. 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
262. 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
263. 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
264. 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
265. 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
266. 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
267. 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
268. 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
269. 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
270. 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
271. 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
272. 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
273. 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

274. 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
275. 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
276. 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
277. Goodman SN. Aligning statistical and scientific reasoning. *Science*. 2016;352(6290):1180–1181. doi:10.1126/science.aaf5406
278. 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
279. 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
280. 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
281. Weintraub PG. The Importance of Publishing Negative Results. *Journal of Insect Science*. 2016;16(1):109. doi:10.1093/jisesa/iew092
282. 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
283. Gelman A, Carlin J. Beyond Power Calculations. *Perspectives on Psychological Science*. 2014;9(6):641–651. doi:10.1177/1745691614551642
284. 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
285. Kim HY. Statistical notes for clinical researchers: effect size. *Restorative Dentistry & Endodontics*. 2015;40(4):328. doi:10.5395/rde.2015.40.4.328
286. Aragon TJ. *epitools: Epidemiology Tools*; 2020. doi:10.32614/CRAN.package.epitools
287. 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
288. Champely S. *pwr: Basic Functions for Power Analysis*; 2020. <https://CRAN.R-project.org/package=pwr>.
289. 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
290. 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

291. 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
292. Altman N, Krzywinski M. P values and the search for significance. *Nature Methods*. 2017;14(1):3–4. doi:10.1038/nmeth.4120
293. 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
294. Aylmer Fisher R. The arrangement of field experiments. *Ministry of Agriculture and Fisheries*. 1926. doi:10.23637/ROTHAMSTED.8V61Q
295. Heinze G, Dunkler D. Five myths about variable selection. *Transplant International*. 2016;30(1):6–10. doi:10.1111/tri.12895
296. Zhu M. The rule of three, proof by contradiction, and uncertainty. *Math Horizons*. 2026;33(3):16–20. doi:10.1080/10724117.2025.2580363
297. 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
298. 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
299. Lakens D, Delacre M. Equivalence Testing and the Second Generation P-Value. *Meta-Psychology*. 2020;4. doi:10.15626/mp.2018.933
300. Meyer F, Perrier V. *esquisse: Explore and Visualize Your Data Interactively*; 2022. <https://CRAN.R-project.org/package=esquisse>.
301. 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
302. McHugh ML. The Chi-square test of independence. *Biochimia Medica*. 2013;143–149. doi:10.11613/bm.2013.018
303. 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
304. 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
305. 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

306. William Revelle. *psych: Procedures for Psychological, Psychometric, and Personality Research.*; 2023. <https://CRAN.R-project.org/package=psych>.
307. Khamis H. Measures of Association: How to Choose? *Journal of Diagnostic Medical Sonography*. 2008;24(3):155–162. doi:10.1177/8756479308317006
308. 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
309. Dahlke JA, Wiernik BM. {psychmeta}: An R Package for Psychometric Meta-Analysis. *Applied Psychological Measurement*. 2018;43(3):415–416. doi:10.1177/0146621618795933
310. Anscombe FJ. Graphs in Statistical Analysis. *The American Statistician*. 1973;27(1):17–21. doi:10.1080/00031305.1973.10478966
311. Northrop PJ. *anscombiner: Create Datasets with Identical Summary Statistics.*; 2022. <https://CRAN.R-project.org/package=anscombiner>.
312. Makowski D, Wiernik BM, Patil I, Lüdecke D, Ben-Shachar MS. {{correlation}}: Methods for Correlation Analysis.; 2022. <https://CRAN.R-project.org/package=correlation>.
313. Lüdecke D, Ben-Shachar MS, Patil I, et al. *easystats: Framework for Easy Statistical Modeling, Visualization, and Reporting.*; 2022. <https://easystats.github.io/easystats/>.
314. Kim JH. Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*. 2019;72(6):558–569. doi:10.4097/kja.19087
315. Schloerke B, Cook D, Larmarange J, et al. *GGally: Extension to ggplot2.*; 2024. doi:10.32614/CRAN.package.GGally
316. Leme DE da C, Alves EV da C, Lemos V do CO, Fattori A. NETWORK ANALYSIS: A MULTIVARIATE STATISTICAL APPROACH FOR HEALTH SCIENCE RESEARCH. *Geriatrics, Gerontology and Aging*. 2020;14(1):43–51. doi:10.5327/z2447-212320201900073
317. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. *Software: Practice and Experience*. 1991;21(11):1129–1164. doi:10.1002/spe.4380211102
318. Csárdi G, Nepusz T. *The igraph software package for complex network research*. Vol Complex Systems.; 2006:1695. <https://igraph.org>.
319. Epskamp S, Borsboom D, Fried EI. *Estimating Psychological Networks and their Accuracy: A Tutorial Paper*. Vol 50.; 2018.
320. Haslbeck JMB, Waldorp LJ. {mgm}: Estimating Time-Varying Mixed Graphical Models in High-Dimensional Data. Vol 93.; 2020. doi:10.18637/jss.v093.i08
321. Arel-Bundock V. {modelsummary}: Data and Model Summaries in {R}. *Journal of Statistical Software*. 2022;103. doi:10.18637/jss.v103.i01
322. Hidalgo B, Goodman M. Multivariate or Multivariable Regression? *American Journal of Public Health*. 2013;103(1):39–40. doi:10.2105/ajph.2012.300897

323. Fernandes AAT, Figueiredo Filho DB, Rocha EC da, Nascimento W da S. Read this paper if you want to learn logistic regression. *Revista de Sociologia e Política*. 2020;28(74). doi:10.1590/1678-987320287406en
324. 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
325. 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
326. Kaplan J. *fastDummies: Fast Creation of Dummy (Binary) Columns and Rows from Categorical Variables*; 2023. <https://CRAN.R-project.org/package=fastDummies>.
327. 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
328. Fox J, Weisberg S. *An {R} Companion to Applied Regression*. Sage Publications, Inc.; 2019. <https://www.john-fox.ca/Companion/>.
329. 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
330. 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
331. Miller TL based on F code by A. *leaps: Regression Subset Selection*; 2024. doi:10.32614/CRAN.package.leaps
332. Hebbali A. *olsrr: Tools for Building OLS Regression Models*; 2024. doi:10.32614/CRAN.package.olsrr
333. Box GEP. Science and Statistics. *Journal of the American Statistical Association*. 1976;71(356):791–799. doi:10.1080/01621459.1976.10480949
334. Anderson D, Heiss A, Sumners J. *equatiomatic: Transform Models into {LaTeX} Equations*; 2024. <https://CRAN.R-project.org/package=equatiomatic>.
335. 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
336. 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
337. Spedicato GA. Discrete Time Markov Chains with R. *The R Journal*. 2017;9(2):84–104. doi:10.32614/RJ-2017-036
338. Bours MJL. Using mediators to understand effect modification and interaction. *Journal of Clinical Epidemiology*. setembro 2023. doi:10.1016/j.jclinepi.2023.09.005

339. Altman DG, Matthews JNS. Statistics Notes: Interaction 1: heterogeneity of effects. *BMJ*. 1996;313(7055):486–486. doi:10.1136/bmj.313.7055.486
340. Pinheiro J, Bates D, R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models.*; 2023. <https://CRAN.R-project.org/package=nlme>.
341. Sabanes Bove D, Dedic J, Kelkhoff D, et al. *mmrm: Mixed Models for Repeated Measures.*; 2022. <https://CRAN.R-project.org/package=mmrm>.
342. Lenth RV. *emmeans: Estimated Marginal Means, aka Least-Squares Means.*; 2023. <https://CRAN.R-project.org/package=emmeans>.
343. 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
344. Henderson T. *correctR: Corrected Test Statistics for Comparing Machine Learning Models on Correlated Samples.*; 2025. <https://CRAN.R-project.org/package=correctR>.
345. Diego J. Pedregal. *UComp: Automatic Univariate Time Series Modelling of many Kinds.*; 2025. doi:10.32614/CRAN.package.UComp
346. Pebesma E, Bivand R. *{Spatial Data Science: With applications in R}*. CRC Press, Taylor & Francis Group; 2023. doi:10.1201/9780429459016
347. 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
348. In J, Lee DK. Survival analysis: Part I — analysis of time-to-event. *Korean Journal of Anesthesiology*. 2018;71(3):182–191. doi:10.4097/kja.d.18.00067
349. Hanada K, Moriya J, Kojima M. Comparison of baseline covariate adjustment methods for restricted mean survival time. *Contemporary Clinical Trials*. 2024;138:107440. doi:10.1016/j.cct.2024.107440
350. Therneau TM. *A Package for Survival Analysis in R.*; 2024. <https://CRAN.R-project.org/package=survival>.
351. Kassambara A, Kosinski M, Biecek P. *survminer: Drawing Survival Curves using 'ggplot2'.*; 2025. doi:10.32614/CRAN.package.survminer
352. Ali S, Abuhmed T, El-Sappagh S, et al. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion*. 2023;99:101805. doi:10.1016/j.inffus.2023.101805
353. Adadi A, Berrada M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*. 2018;6:52138–52160. doi:10.1109/access.2018.2870052
354. Vilone G, Longo L. Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion*. 2021;76:89–106. doi:10.1016/j.inffus.2021.05.009

355. Burger T. Keeping generative artificial intelligence reliable in omics biology. *Patterns*. 2026;7(1):101417. doi:10.1016/j.patter.2025.101417
356. Korkmaz S, Goksuluk D, Karaismailoglu E. *fastml: Guarded Resampling Workflows for Safe and Automated Machine Learning in R*; 2026. doi:10.32614/CRAN.package.fastml
357. Van Calster B, Collins GS, Vickers AJ, et al. Evaluation of performance measures in predictive artificial intelligence models to support medical decisions: overview and guidance. *The Lancet Digital Health*. 2025;7(12):100916. doi:10.1016/j.landig.2025.100916
358. 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
359. Hand DJ. Classifier Technology and the Illusion of Progress. *Statistical Science*. 2006;21(1). doi:10.1214/088342306000000060
360. Hicks SA, Strümke I, Thambawita V, et al. On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports*. 2022;12(1). doi:10.1038/s41598-022-09954-8
361. 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
362. Fryda T, LeDell E, Gill N, et al. *h2o: R Interface for the 'H2O' Scalable Machine Learning Platform*; 2024. doi:10.32614/CRAN.package.h2o
363. Dyer EL, Kording K. Why the simplest explanation isn't always the best. *Proceedings of the National Academy of Sciences*. 2023;120(52). doi:10.1073/pnas.2319169120
364. 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
365. 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
366. Rosenblatt F. Perceptron Simulation Experiments. *Proceedings of the IRE*. 1960;48(3):301–309. doi:10.1109/jrproc.1960.287598
367. Fritsch S, Guenther F, Wright MN. *neuralnet: Training of Neural Networks*; 2019. doi:10.32614/CRAN.package.neuralnet
368. 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
369. 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.

370. 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
371. 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
372. Bacchetti P. Ethics and Sample Size. *American Journal of Epidemiology*. 2005;161(2):105–110. doi:10.1093/aje/kwi014
373. 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
374. 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
375. 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
376. 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
377. 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
378. 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
379. Lundberg I, Johnson R, Stewart BM. What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory. *American Sociological Review*. 2021;86(3):532–565. doi:10.1177/00031224211004187
380. Bland JM, Altman DG. Statistics notes: Matching. *BMJ*. 1994;309(6962):1128–1128. doi:10.1136/bmj.309.6962.1128
381. 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
382. Sut N. Study Designs in Medicine. *Balkan Medical Journal*. 2015;31(4):273–277. doi:10.5152/balkanmedj.2014.1408
383. 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
384. 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

385. 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
386. Chassé M, Fergusson DA. Diagnostic Accuracy Studies. *Seminars in Nuclear Medicine*. 2019;49(2):87–93. doi:10.1053/j.semnuclmed.2018.11.005
387. Chidambaram AG, Josephson M. Clinical research study designs: The essentials. *PEDIATRIC INVESTIGATION*. 2019;3(4):245–252. doi:10.1002/ped4.12166
388. 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
389. 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
390. 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
391. 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
392. 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
393. 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
394. 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
395. 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
396. DeBruine L. *faux: Simulation for Factorial Designs.*; 2023. doi:10.5281/zenodo.2669586
397. 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
398. Elm E von, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke 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
399. Scott WA. Reliability of Content Analysis: The Case of Nominal Scale Coding. *Public Opinion Quarterly*. 1955;19(3):321. doi:10.1086/266577

400. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*. 1960;20(1):37–46. doi:10.1177/001316446002000104
401. Gamer M, Lemon J, Ian Fellows Puspendra Singh. *irr: Various Coefficients of Interrater Reliability and Agreement.*; 2019. doi:10.32614/CRAN.package.irr
402. Lehnert B. *BlandAltmanLeh: Plots (Slightly Extended) Bland-Altman Plots.*; 2015. <https://CRAN.R-project.org/package=BlandAltmanLeh>.
403. Contributors semTools. *semTools: Useful tools for structural equation modeling*; 2016. <https://CRAN.R-project.org/package=semTools>.
404. Quartagno M, Carpenter J. *{jomo}: A package for Multilevel Joint Modelling Multiple Imputation.*; 2023. <https://CRAN.R-project.org/package=jomo>.
405. 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
406. Rosseel Y. {lavaan}: An {R} Package for Structural Equation Modeling. *Journal of Statistical Software*. 2012;48. doi:10.18637/jss.v048.i02
407. 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
408. 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
409. 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
410. 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
411. 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
412. Neth H, Gaisbauer F, Gradwohl N, Gaissmaier W. *riskyR: Rendering Risk Literacy more Transparent.*; 2022. <https://CRAN.R-project.org/package=riskyR>.
413. 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
414. 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
415. 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

416. 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
417. 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
418. 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>.
419. 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
420. 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
421. 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
422. 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
423. Reeves BC, Gaus W. Guidelines for Reporting Non-Randomised Studies. *Complementary Medicine Research*. 2004;11(1):46–52. doi:10.1159/000080576
424. Bland JM, Altman DG. Comparisons within randomised groups can be very misleading. *BMJ*. 2011;342(may06 2):d561–d561. doi:10.1136/bmj.d561
425. 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
426. 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
427. 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
428. Laird N. Further Comparative Analyses of Pretest-Posttest Research Designs. *The American Statistician*. 1983;37(4a):329–330. doi:10.1080/00031305.1983.10483133
429. 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
430. 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

431. 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
432. 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/clep.s161508
433. 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
434. Lavori PW, Louis TA, Bailar JC, Polansky M. Designs for Experiments — Parallel Comparisons of Treatment. *New England Journal of Medicine*. 1983;309(21):1291–1299. doi:10.1056/nejm198311243092105
435. Altman DG. Comparability of Randomised Groups. *The Statistician*. 1985;34(1):125. doi:10.2307/2987510
436. Altman DG, Doré CJ. Randomisation and baseline comparisons in clinical trials. *The Lancet*. 1990;335(8682):149–153. doi:10.1016/0140-6736(90)90014-v
437. 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
438. 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>.
439. 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
440. 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
441. 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
442. Altman DG. Statistics Notes: Interaction revisited: the difference between two estimates. *BMJ*. 2003;326(7382):219–219. doi:10.1136/bmj.326.7382.219
443. 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
444. 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
445. 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

446. 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
447. Dayim A. *consort: Create Consort Diagram.*; 2023. <https://CRAN.R-project.org/package=consort>.
448. Dwan K, Li T, Altman DG, Elbourne D. CONSORT 2010 statement: extension to randomised crossover trials. *BMJ.* julho 2019;14378. doi:10.1136/bmj.14378
449. 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
450. Chatters R, Hawksworth O, Julious S, Cook A. The development of a set of key points to aid clinicians and researchers in designing and conducting n-of-1 trials. *Trials.* 2024;25(1). doi:10.1186/s13063-024-08261-z
451. Hawksworth O, Chatters R, Julious S, et al. A methodological review of randomised n-of-1 trials. *Trials.* 2024;25(1). doi:10.1186/s13063-024-08100-1
452. Sarkar D. *Lattice: Multivariate Data Visualization with R.*; 2008. <http://lmdvr.r-forge.r-project.org>.
453. 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
454. Fantini D. *easyPubMed: Search and Retrieve Scientific Publication Records from PubMed.*; 2019. doi:10.32614/CRAN.package.easyPubMed
455. Chamberlain S, Zhu H, Jahn N, Boettiger C, Ram K. *rcrossref: Client for Various CrossRef APIs.*; 2022. doi:10.32614/CRAN.package.rcrossref
456. Jahn N. *roadoi: Find Free Versions of Scholarly Publications via Unpaywall.*; 2024. doi:10.32614/CRAN.package.roadoi
457. 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
458. 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
459. 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
460. 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

461. 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
462. 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
463. Viechtbauer W. *Conducting meta-analyses in {R} with the {metafor} package*. Vol 36.; 2010. doi:10.18637/jss.v036.i03
464. 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
465. Valkenhoef G van, Kuiper J. *gemtc: Network Meta-Analysis Using Bayesian Methods*; 2025. doi:10.32614/CRAN.package.gemtc
466. 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
467. 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
468. 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
469. 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
470. Grooth HJ de, Parietti 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
471. 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
472. 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
473. Sousa-Pinto PD with contributions from B. *mada: Meta-Analysis of Diagnostic Accuracy*; 2022. <https://CRAN.R-project.org/package=mada>.
474. 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
475. Song, Eastwood, Gilbody, Duley, Sutton. Publication and related biases. *Health Technology Assessment*. 2000;4(10). doi:10.3310/hta4100
476. 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

477. 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
478. 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
479. 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
480. 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
481. 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
482. 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
483. 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
484. 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
485. 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
486. 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
487. 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
488. Zikmund-Fisher BJ, Thorpe A, Fagerlin A. How to Communicate Medical Numbers. *JAMA*. 2025;334(16):1474. doi:10.1001/jama.2025.13655
489. 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/>.
490. Nuijten MB, Epskamp S. *statcheck: Extract Statistics from Articles and Recompute P-Values.*; 2024. doi:10.32614/CRAN.package.statcheck
491. 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

492. 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
493. 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
494. 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
495. 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
496. 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
497. 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
498. 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
499. 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
500. Groves T. Research methods and reporting. *BMJ.* 2008;337(oct22 1):a2201–a2201. doi:10.1136/bmj.a2201
501. 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
502. 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
503. 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
504. 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.

Com um total de 725 perguntas, 1721 respostas, 132 figuras e 512 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.