



Estatística Aplicada a Negócios

Para
leigos

As estatísticas permitem analisar problemas empresariais reais com dados concretos, para que você possa determinar se uma estratégia de marketing está realmente funcionando, quanto uma empresa deve cobrar por seus produtos ou qualquer outra das inúmeras questões práticas. A ciência da estatística utiliza análise de regressão, testes de hipóteses, distribuições amostrais e muito mais para garantir uma análise precisa dos dados.

TRÊS FORMAS DE DESCREVER POPULAÇÕES E AMOSTRAS

População é um conjunto de medições que você tem interesse em estudar. Por exemplo, uma população poderia consistir nas rendas dos residentes dos Estados Unidos. Como muitas vezes não é prático analisar todos os elementos de uma população, uma amostra de dados pode ser selecionada a partir de uma população e usada como substituto da população.



LEMBRE-SE

Uma amostra pode ser selecionada de várias maneiras diferentes. É extremamente importante garantir que os membros de uma amostra reflitam verdadeiramente a população subjacente.

MEDIDAS DE TENDÊNCIA CENTRAL

Em estatística, a média, a mediana e a moda são conhecidas como medidas de tendência central. Elas são usadas para identificar o centro de um conjunto de dados:

- **Média:** o valor médio de um conjunto de dados.
- **Mediana:** o valor que divide um conjunto de dados ao meio, de modo que metade das observações seja menor que a mediana e metade seja maior que a mediana.
- **Moda:** o valor que ocorre com mais frequência em um conjunto de dados.

MEDIDAS DE DISPERSÃO

As medidas de dispersão são usadas para descobrir o quanto os elementos de um conjunto de dados estão espalhados. Três das medidas de dispersão mais comumente usadas são:

- Intervalo
- Variância
- Desvio-padrão



Estatística Aplicada a Negócios

Para
leigos

MEDIDAS DE ASSOCIAÇÃO

As *medidas de associação* são utilizadas para determinar o grau de relação entre duas variáveis. Duas das medidas de associação mais utilizadas são:

- Covariância
- Correlação

VARIÁVEIS ALEATÓRIAS E DISTRIBUIÇÕES DE PROBABILIDADE

Uma *variável aleatória* é usada para atribuir valores numéricos a todos os resultados possíveis de um experimento aleatório. Uma *distribuição de probabilidade* atribui probabilidades a esses valores numéricos. Os dois tipos básicos de distribuições de probabilidade são:

- Discreta
- Contínua

Com uma distribuição de probabilidade discreta, há apenas um número finito de valores possíveis; com uma distribuição de probabilidade contínua, o número de valores possíveis é infinito.

DISTRIBUIÇÕES DE PROBABILIDADE DISCRETAS

Duas das distribuições de probabilidade discretas mais utilizadas são:

- Binomial
- Poisson

DISTRIBUIÇÕES DE PROBABILIDADE CONTÍNUAS

Algumas das distribuições de probabilidade contínuas mais utilizadas são:

- Uniforme
- Normal
- T de Student
- Qui-quadrado
- F



Estatística Aplicada a Negócios

Para
leigos

EXPLORE O TESTE DE HIPÓTESES EM ESTATÍSTICA APLICADA A NEGÓCIOS

O teste de hipóteses é realizado para tirar conclusões sobre um parâmetro populacional, como a média, a proporção, a variância etc. O teste de hipóteses é um procedimento de seis etapas:

1. Hipótese nula
2. Hipótese alternativa
3. Nível de significância
4. Estatística de teste
5. Valor(es) crítico(s)
6. Decisão

AMOSTRA

AMOSTRA

Estatística Aplicada a Negócios

Para
leigos

AMOSTRA

AMOSTRA



Estatística Aplicada a Negócios

Para **leigos**

2ª Edição

Alan Anderson, PhD



ALTA BOOKS
E D I T O R A
Rio de Janeiro, 2026

Estatística Aplicada a Negócios – Tradução da 2ª Edição

Copyright © 2026 STARLIN ALTA EDITORA E CONSULTORIA LTDA.

Copyright ©2024 Alan Anderson.

ISBN: 978-85-508-2903-6

Translated from original Business Statistics For Dummies, 2nd Edition © 2024 by Wiley Publishing, Inc. ISBN 978-1-394-21992-6.

This translation is published and sold by John Wiley, the owner of all rights to publish and sell the same. PORTUGUESE language edition published by Starlin Alta Editora e Consultoria Eireli, Copyright © 2026 by STARLIN ALTA EDITORA E CONSULTORIA LTDA.

Impresso no Brasil – 1ª Edição, 2026 – Edição revisada conforme o Acordo Ortográfico da Língua Portuguesa de 2009.

Dados Internacionais de Catalogação na Publicação (CIP)

A53e
1.ed. - Anderson, Alan.
Estatística Aplicada a Negócios Para Leigos /
Alan Anderson. - 1ª Edição - Traduzido por Eveline
Machado. Rio de Janeiro, Alta Books, 2026.
400 p.; 16 x 23 cm.

Inclui bibliografia e índice.
ISBN 978-85-508-2903-6

1. Estatística Aplicada a Negócios. 2. Análise de
dados empresariais. 3. Probabilidade I. Título. II.
Machado, Eveline.

CDD: 519.5
CDU: 519.2

Índice para catálogo sistemático:
1. Estatística Matemática - 519.5

Todos os direitos estão reservados e protegidos por Lei. Nenhuma parte deste livro, sem autorização prévia por escrito da editora, poderá ser reproduzida ou transmitida. A violação dos Direitos Autorais é crime estabelecido na Lei nº 9.610/98 e com punição de acordo com o artigo 184 do Código Penal.

O conteúdo desta obra fora formulado exclusivamente pelo(s) autor(es).

Marcas Registradas: Todos os termos mencionados e reconhecidos como Marca Registrada e/ou Comercial são de responsabilidade de seus proprietários. A editora informa não estar associada a nenhum produto e/ou fornecedor apresentado no livro.

Material de apoio e erratas: Se parte integrante da obra e/ou por real necessidade, no site da editora o leitor encontrará os materiais de apoio (download), errata e/ou quaisquer outros conteúdos aplicáveis à obra. Acesse o site www.altabooks.com.br e procure pelo título do livro desejado para ter acesso ao conteúdo.

Suporte Técnico: A obra é comercializada na forma em que está, sem direito a suporte técnico ou orientação pessoal/exclusiva ao leitor.

A editora não se responsabiliza pela manutenção, atualização e idioma dos sites, programas, materiais complementares ou similares referidos pelos autores nesta obra.

Produção Editorial: Grupo Editorial Alta Books

Diretor Editorial: Anderson Vieira

Vendas Governamentais: Cristiane Mutús

Produtor Editorial: Eveline Machado

Diagramação: Lucia Quaresma

Sobre o Autor

Alan Anderson atualmente leciona finanças, economia, estatística e matemática em várias instituições de ensino, incluindo Fordham University, New York University, Manhattanville College, Purchase College e Fairfield University. Ele também passou muitos anos no “mundo real” como economista, gestor de risco e analista de renda fixa. (Ele prefere o meio acadêmico!)

Alan fez doutorado em economia na Fordham University e também tem mestrado em engenharia financeira pela New York University.

Agradecimentos do Autor

Gostaria de agradecer a várias pessoas que ajudaram a compilar este livro. Em primeiro lugar, agradeço a Lindsay Berg, minha editora de aquisições. Obrigado por me dar a oportunidade de escrever este livro. Agradeço também a Katharine Dvorak, minha editora de desenvolvimento, que me manteve focado. E agradeço a M. Higuera por fazer uma revisão técnica criativa e sólida. Vocês foram uma ótima equipe... obrigado.

AMOSTRA

Sumário Resumido

Introdução	1
Parte 1: Introdução da Estatística Aplicada a Negócios	5
CAPÍTULO 1: Arte e Ciência na Estatística Aplicada a Negócios	7
CAPÍTULO 2: Imagens Contam História: Representações Gráficas dos Dados	23
CAPÍTULO 3: Identificando o Centro do Conjunto de Dados	37
CAPÍTULO 4: Medindo a Variação em um Conjunto de Dados	55
CAPÍTULO 5: Medindo a Relação dos Conjuntos de Dados	75
Parte 2: Teoria e Distribuições de Probabilidade	99
CAPÍTULO 6: Teoria da Probabilidade: Medindo a Probabilidade dos Eventos.	101
CAPÍTULO 7: Distribuições de Probabilidade e Variáveis Aleatórias	121
CAPÍTULO 8: Distribuições Binomial e de Poisson	133
CAPÍTULO 9: Distribuição Normal: Tantas Possibilidades!	153
CAPÍTULO 10: Técnicas de Amostragem e Distribuições	175
Parte 3: Conclusões a partir de Amostras	195
CAPÍTULO 11: Intervalos de Confiança e Distribuição t de Student	197
CAPÍTULO 12: Testando Hipóteses sobre a Média Populacional	215
CAPÍTULO 13: Aplicações da Distribuição Qui-quadrado	255
CAPÍTULO 14: Aplicações da Distribuição F	285
Parte 4: Técnicas Mais Avançadas: Análise de Regressão e Modelagem em Planilhas	299
CAPÍTULO 15: Análise de Regressão Simples	301
CAPÍTULO 16: Principais Técnicas Estatísticas no Excel	331
Parte 5: A Parte dos Dez	357
CAPÍTULO 17: Dez Erros Comuns na Análise Estatística	359
CAPÍTULO 18: (Quase) Dez Fórmulas Principais para Estatística Aplicada a Negócios	367
Índice	379

AMOSTRA

Sumário

INTRODUÇÃO 1

 Sobre Este Livro. 1

 Ícones usados neste livro 3

 Além Deste Livro 3

 De Lá para Cá, Daqui para Lá 4

PARTE 1: INTRODUÇÃO DA ESTATÍSTICA
APLICADA A NEGÓCIOS 5

CAPÍTULO 1: **Arte e Ciência na Estatística
Aplicada a Negócios** 7

 Representando as Principais Propriedades dos Dados 8

 Analisando dados com gráficos 8

 Definindo propriedades e relações com
 medidas numéricas 12

 Encontrando o centro dos dados 12

 Medindo a dispersão dos dados 13

 Determinando a relação entre duas variáveis 14

 Probabilidade: A Base de Toda Análise Estatística 14

 Variáveis aleatórias 15

 Distribuições de probabilidade 16

 Distribuições de probabilidade discretas 17

 Distribuições de probabilidade contínuas 17

 Usando Técnicas de Amostragem e Distribuições
 Amostrais 18

 Inferência Estatística: Tirando Conclusões a
 Partir dos Dados 19

 Intervalos de confiança 19

 Teste de hipóteses 20

Análise de regressão simples 20

CAPÍTULO 2: **Imagens Contam História:
Representações Gráficas dos Dados** 23

 Analisando a Distribuição dos Dados por Classe
 ou Categoria 24

 Distribuição de frequência para valores qualitativos. ... 30

 Distribuições de frequência cumulativa 31

 Histogramas: Uma Imagem das Distribuições de
 Frequência 31

Verificando Outros Gráficos Úteis	33
Gráficos de linha: Mostrando os valores de uma série de dados	33
Gráficos de pizza: mostrando a composição de um conjunto de dados	34
Gráficos de dispersão: Mostrando a relação entre duas variáveis	35
CAPÍTULO 3: Identificando o Centro do Conjunto de Dados	37
Analisando Métodos para Encontrar a Média	38
Média aritmética	38
Média geométrica	40
Média ponderada	43
No Meio: A Mediana de um Conjunto de Dados	45
Determinando a Relação entre Média e Mediana	47
Simétrico	47
Assimetria negativa	48
Assimetria positiva	49
Descobrimos a Moda: O Elemento Mais Repetido	50
Média, Mediana e Moda com a Calculadora TI-84 Plus	51
CAPÍTULO 4: Medindo a Variação em um Conjunto de Dados	55
Determinando a Variância e o Desvio-Padrão	56
Encontrando a variância amostral	57
Encontrando o desvio-padrão amostral	58
Calculando a variância e o desvio-padrão populacional	61
Encontrando a Posição Relativa dos Dados	65
Percentis: Dividindo tudo em centésimos	65
Quartis: Dividindo tudo em quartas partes	67
Intervalo Interquartil: Identificando os 50% centrais	68
Medindo a Variação Relativa	69
Coeficiente de Variação: A dispersão de um conjunto de dados em relação à média	70
Comparando os riscos relativos dos dois portfólios	71
Calculando Medidas de Dispersão com a Calculadora TI-84 Plus	72

CAPÍTULO 5:	Medindo a Relação dos Conjuntos de Dados	75
	Entendendo Covariância e Correlação	76
	Covariância amostral e coeficiente de correlação	77
	Covariância populacional e coeficiente de correlação	82
	Comparando correlação e covariância	86
	Interpretando o Coeficiente da Correlação	90
	Mostrando a relação entre duas variáveis	91
	Aplicação: Correlação e benefícios da diversificação	93
	Covariância e Correlação com a Calculadora TI-84 Plus	95
PARTE 2:	TEORIA E DISTRIBUIÇÕES DE PROBABILIDADE	99
CAPÍTULO 6:	Teoria da Probabilidade: Medindo a Probabilidade dos Eventos	101
	Trabalhando com Conjuntos	102
	Pertencimento	102
	Subconjunto	102
	União	103
	Interseção	105
	Complemento	106
	Apostando em Resultados Incertos	107
	Espaço amostral: Tudo o que pode acontecer	107
	Evento: Um resultado possível	108
	Calculando as probabilidades dos eventos	109
	Analizando os Tipos de Probabilidades	110
	Probabilidades incondicionais (marginais):	
	Quando os eventos são independentes	111
	Probabilidades conjuntas: Quando duas coisas acontecem ao mesmo tempo	112
	Probabilidades condicionais: Quando um evento depende de outro	113
	Determinando a independência dos eventos	114
	Seguindo as Regras: Calculando Probabilidades	115
	Regra da adição	115
	Regra do complemento	117
	Regra da multiplicação	118

CAPÍTULO 7:	Distribuições de Probabilidade e Variáveis Aleatórias	121
	Definindo o Papel da Variável Aleatória	122
	Atribuindo Probabilidades a uma Variável Aleatória	125
	Calculando a distribuição de probabilidade	126
	Visualizando a distribuição de probabilidade com um histograma	127
	Caracterizando uma Distribuição de Probabilidade com Momentos	128
	Entendendo o operador de soma (Σ)	128
	Valor esperado	129
	Variância e desvio-padrão	131
CAPÍTULO 8:	Distribuições Binomial e de Poisson	133
	Duas Possibilidades com a Distribuição Binomial	134
	Verificando a distribuição binomial	135
	Calculando as probabilidades binomiais	135
	Momentos da distribuição binomial	141
	Representando a distribuição binomial graficamente	142
	Mantendo o Tempo: A Distribuição de Poisson	145
	Calculando as probabilidades de Poisson	145
	Representando graficamente a distribuição de Poisson	147
	Calculando as Probabilidades Binomiais e de Poisson com a Calculadora TI-84 Plus	149
	Calculando as probabilidades binomiais	149
	Calculando as probabilidades de Poisson	150
CAPÍTULO 9:	Distribuição Normal: Tantas Possibilidades!	153
	Comparando Distribuições Discretas e Contínuas	154
	Entendendo a Distribuição Normal	156
	Representando graficamente a distribuição normal	157
	Conhecendo a distribuição normal padrão	159
	Calculando as probabilidades normais padrão	160
	Calculando as probabilidades normais além da normal padrão	168
	Calculando as Probabilidades para a Distribuição Normal com a Calculadora TI-84 Plus	171

CAPÍTULO 10: Técnicas de Amostragem e Distribuições . . 175

Técnicas de Amostragem: Escolhendo Dados de uma População	176
Amostragem probabilística	177
Amostragem não probabilística	182
Distribuições Amostrais	184
Representando as distribuições amostrais graficamente	185
Momentos de uma distribuição amostral	188
Teorema do Limite Central	189
Convertendo $\bar{x}$ em uma variável aleatória normal padrão.	189

PARTE 3: CONCLUSÕES A PARTIR DE AMOSTRAS . . . 195

CAPÍTULO 11: Intervalos de Confiança e Distribuição t de Student 197

Quase Normal: Distribuição t de Student	198
Propriedades da distribuição t.	198
Graus de liberdade	199
Momentos da distribuição t	200
Criando o Gráfico da Distribuição t	201
Probabilidades e Tabela t	203
Estimativas Pontuais vs. Estimativas de Intervalo	204
Estimando Intervalos de Confiança para a Média Populacional	206
Desvio-padrão populacional conhecido.	206
Desvio-padrão populacional desconhecido	210
Calculando Intervalos de Confiança para a Média Populacional com a Calculadora TI-84 Plus	212
O desvio-padrão populacional é conhecido	212
O desvio-padrão populacional é desconhecido	213

CAPÍTULO 12: Testando Hipóteses sobre a Média Populacional 215

Etapas Principais no Teste de Hipóteses para a Média de uma Única População	216
Escrevendo a hipótese nula	216
Definindo uma hipótese alternativa	217
Escolhendo um nível de significância	220
Calculando a estatística de teste	222
Comparando valores críticos	223
Desvio-padrão populacional conhecido.	227

Usando a regra de decisão	230
Testando Hipóteses Sobre Duas Médias Populacionais. ...	233
Escrevendo a hipótese nula para duas médias populacionais	234
Definindo hipóteses alternativas para duas médias populacionais	234
Determinando as estatísticas de teste para duas médias populacionais	235
Testando Hipóteses sobre Médias Populacionais com a Calculadora TI-84 Plus.	246
Média de uma única população	246
Médias de duas populações	250
CAPÍTULO 13: Aplicações da Distribuição Qui-quadrado	255
Mantendo-se Positivo com a Distribuição Qui-quadrado	256
Representando a distribuição qui-quadrado graficamente	257
Definindo uma variável qui-quadrado aleatória	259
Conferindo os momentos da distribuição qui-quadrado.	259
Testando Hipóteses sobre a Variância Populacional.	260
Definindo o que você presume ser verdade:	
A hipótese nula	261
Estabelecendo a hipótese alternativa.	261
Escolhendo o nível de significância	264
Calculando a estatística de teste	264
Determinando o(s) valor(es) crítico(s)	265
Tomando a decisão.	268
Praticando com Testes de Aderência.	270
Comparando uma população com a distribuição de Poisson	270
Comparando uma população com a distribuição normal	276
Realizando um Teste de Aderência com a Calculadora TI-84 Plus.	281
CAPÍTULO 14: Aplicações da Distribuição F	285
Conhecendo a Distribuição F.	285
Definindo uma variável aleatória F.	288
Medindo os momentos da distribuição F.	288

Testando Hipóteses sobre a Igualdade de Duas Variâncias Populacionais.	291
Hipótese nula: Variâncias iguais	291
Hipótese alternativa: Variâncias desiguais	292
Estatística de teste	292
Valor(es) crítico(s)	293
Decisão sobre a igualdade de duas variâncias populacionais	295
Testando Hipóteses sobre Duas Variâncias Populacionais com a Calculadora TI-84 Plus.	296

PARTE 4: TÉCNICAS MAIS AVANÇADAS: ANÁLISE DE REGRESSÃO E MODELAGEM EM PLANILHAS 299

CAPÍTULO 15: **Análise de Regressão Simples** 301

Premissa Fundamental: As Variáveis Têm uma Relação Linear.	302
Definindo uma relação linear	303
Usando gráficos de dispersão para identificar relações lineares	304
Definindo a Equação de Regressão Populacional	308
Estimando a Equação de Regressão Populacional	309
Testando a Equação de Regressão Estimada.	316
Usando o coeficiente de determinação (R^2).	316
Calculando o coeficiente de determinação	318
Teste t	319
Usando um Software de Estatística	325
Premissas da Regressão Linear Simples	326
Conduzindo a Análise de Regressão Simples com a Calculadora TI-84 Plus	327

CAPÍTULO 16: **Principais Técnicas Estatísticas no Excel** ... 331

Implementando Funções no Excel	332
Conhecendo as Principais Funções Estatísticas do Excel ..	333
Medidas de tendência central	333
Medidas de dispersão	335
Medidas de associação	336
Distribuições de probabilidade discreta	338
Distribuições de probabilidade contínua	341
Intervalos de confiança	345
Análise de regressão	347

Aprofundando-se com as Ferramentas de Análise	349
Calculando covariância e correlação	349
Calculando as estatísticas descritivas	351
Análise de regressão	353
Teste de hipóteses	354
PARTE 5: A PARTE DOS DEZ.....	357
CAPÍTULO 17: Dez Erros Comuns na Análise Estatística ..	359
Projetando Gráficos Enganosos	360
Tirando a Conclusão Errada de um Intervalo de Confiança	361
Interpretando Incorretamente os Resultados de um Teste de Hipótese	362
Confiando Demais no Coeficiente de Determinação (R^2)...	362
Presumindo a Normalidade	363
Pensar que Correlação Implica Causalidade	363
Tirando Conclusões de uma Equação de Regressão Quando os Dados Não Seguem as Suposições	364
Análise de Regressão para Previsões Sobre Valores Fora do Intervalo dos Dados da Amostra	365
Confiando Demais nas Previsões	365
Usando a Distribuição Errada	366
CAPÍTULO 18: (Quase) Dez Fórmulas Principais para Estatística Aplicada a Negócios.....	367
Medidas de Resumo de uma População ou Amostra	368
Probabilidade	369
Distribuições de Probabilidade Discretas	370
Distribuições de Probabilidade Contínuas	371
Distribuições Amostrais	371
Intervalos de Confiança para a Média Populacional	372
Testando Hipóteses sobre Médias Populacionais	373
Testando Hipóteses sobre Variâncias Populacionais	376
Usando a Análise de Regressão	377
ÍNDICE	379

Introdução

Você sempre teve pavor de estatística? Você é praticamente todo mundo! As equações são extremamente intimidadoras e a terminologia parece tão... chata.

Porque, então, a estatística é tão importante? Todas as disciplinas de negócios podem ser analisadas com princípios estatísticos. A estatística torna possível analisar problemas do mundo real com dados reais, para que possamos entender se nossa estratégia de marketing está realmente funcionando, quanto uma empresa deve cobrar por seus produtos, ou qualquer outra das inúmeras questões práticas dos negócios.

Sem uma estrutura formal para analisar esses tipos de situações, seria impossível ter qualquer confiança em nossos resultados. É aí que entra a ciência da estatística. Longe de ser uma coleção opressora de equações, é uma estrutura lógica para analisar problemas práticos de negócios com dados reais.

Este livro foi elaborado para mostrar como aplicar a estatística a situações práticas passo a passo, para que, ao final, você saiba tanto sobre estatística quanto pessoas com muito mais formação nessa área!

Sobre Este Livro

Todos os cursos de administração exigem pelo menos algumas disciplinas de estatística, e há uma boa razão para isso! Todas as disciplinas de administração são empíricas por natureza, o que significa que precisam analisar dados reais para serem bem-sucedidas. O objetivo deste livro é:

- » Apresentar os princípios nos quais se baseia a análise estatística
- » Fornecer muitos exemplos práticos desses princípios para que você possa dominá-los
- » Melhorar sua compreensão das circunstâncias em que cada técnica estatística deve ser usada

Como um título da série *Para leigos*, este livro está organizado em módulos; você pode pular partes e aprender sobre várias técnicas estatísticas na ordem que lhe for mais conveniente. Nos casos em que o conteúdo de um capítulo se baseia em leituras anteriores, você é orientado a voltar ao

material original. Ao longo do livro, há muitas dicas e lembretes úteis para que você aproveite ao máximo cada capítulo. Explico cada equação em detalhes, e todos os termos-chave são explicados em profundidade.

Nesta segunda edição atualizada, mostro como usar as calculadoras Texas Instruments TI-84 Plus e TI-84 Plus CE para obter resultados de forma rápida e fácil para praticamente todos os problemas que você encontrará no livro. Também adicionei um novo capítulo que mostra como é fácil resolver problemas estatísticos usando o Microsoft Excel.

Este livro não pode torná-lo especialista em estatística, mas fornece uma maneira de melhorar seus conhecimentos muito rapidamente para que possa usar a estatística em situações práticas imediatamente.

Penso que...

Faço as seguintes suposições sobre você, leitor deste livro:

- » Você precisa usar as técnicas deste livro em um ambiente prático e tem pouca ou nenhuma experiência prévia com estatística.

OU

- » Você é um estudante que se sente sobrecarregado com um curso tradicional de estatística e sente a necessidade de mais conhecimentos básicos. Você pode se beneficiar ao ver mais exemplos do material; a estatística é uma ciência que pode ser aprendida por meio da prática!

OU

- » Você está simplesmente interessado em melhorar seus conhecimentos nessa área.

Em todos esses casos, você está extremamente motivado e pode se dedicar ao aprendizado da estatística tanto quanto for necessário. Parabéns! Sua recompensa por ler este livro será uma maior compreensão da estatística aplicada a negócios.

Ícones usados neste livro

Os ícones a seguir foram criados para ajudá-lo a usar este livro de forma rápida e fácil. Fique atento a eles.



LEMBRE-SE

O ícone Lembre-se indica informações que são especialmente importantes para serem lembradas para fins de exame.



DICA

O ícone Dica apresenta informações como um acrônimo de memória ou algum outro auxílio para compreender ou memorizar o material.



CUIDADO!

Quando vir este ícone, preste uma atenção especial. As informações a seguir podem ser um pouco difíceis, confusas ou prejudiciais.



PAPO DE
ESPECIALISTA

O ícone Papo de Especialista é usado para indicar informações detalhadas; para algumas pessoas, pode não ser necessário ler ou compreender essas informações.

Além Deste Livro

Além do material informativo, inteligente e (se assim posso dizer) bem escrito que você está lendo agora, este produto também vem com ajuda e informações online acessíveis. Não importa o quanto você conheça estatística ao final deste livro, um pouco de informação extra sempre é útil. Confira a Folha de Cola online deste livro para saber mais sobre como descrever populações e amostras, variáveis aleatórias, distribuições de probabilidade, testes de hipóteses e muito mais. Você pode acessar a Folha de Cola Online no site da editora Alta Books. Procure pelo título do livro. Faça o download da folha completa, bem como de erratas e possíveis arquivos de apoio.

De Lá para Cá, Daqui para Lá

Quando você ficar mais hábil em análise estatística, talvez queira experimentar um pacote estatístico completo, como o SPSS ou o SAS. Eles eliminarão grande parte da carga computacional, liberando você para se concentrar na análise dos resultados. Você também pode se interessar em obter mais formação nessa área. Por exemplo, talvez queira fazer um curso de pós-graduação, como um MBA (Mestrado em Administração de Empresas). Essa é uma qualificação extremamente importante que pode abrir muitas portas no mundo dos negócios. Você precisará de seus conhecimentos estatísticos para obter esse diploma, pois eles são amplamente utilizados em todo o currículo.

Se não estiver pronto para a pós-graduação, pode simplesmente explorar alguns cursos de estatística de nível universitário em uma universidade próxima. O mais importante é continuar usando seus conhecimentos estatísticos, pois você só se tornará proficiente em usá-los com uma prática constante.

1

Introdução da Estatística Aplicada a Negócios

NESTA PARTE ...

Use histogramas para fornecer uma representação visual da distribuição dos elementos em um conjunto de dados. Um histograma pode mostrar quais valores ocorrem com mais frequência, os valores menores e maiores, e como esses valores estão distribuídos.

Crie gráficos que reflitam dados não numéricos, como cores, sabores, nomes de marcas e assim por diante. Os gráficos são usados quando as medidas numéricas são difíceis ou impossíveis de calcular.

Identifique o centro de um conjunto de dados usando a média (a média aritmética), a mediana (o meio) e a moda (o valor que ocorre com mais frequência). Elas são conhecidas como medidas de tendência central.

Use fórmulas para calcular a covariância e a correlação para amostras e populações; um gráfico de dispersão é usado para mostrar a relação (se houver) entre duas variáveis.

NESTE CAPÍTULO

- » Analisando as principais propriedades dos dados
- » Compreendendo o papel da probabilidade nos negócios
- » Distribuições amostrais
- » Tirando conclusões com base nos resultados

Capítulo 1

Arte e Ciência na Estatística Aplicada a Negócios

A análise estatística é amplamente utilizada em todas as disciplinas de negócios. Por exemplo, pesquisadores de marketing analisam os padrões de gastos dos consumidores para planejar adequadamente novas campanhas publicitárias. As organizações utilizam consultoria de gestão para determinar a eficiência com que os recursos estão sendo utilizados. Os fabricantes utilizam métodos de controle de qualidade para garantir a consistência dos produtos que estão produzindo. Essas aplicações comerciais e muitas outras são fortemente baseadas na análise estatística.

As instituições financeiras utilizam estatísticas para uma ampla variedade de aplicações. Por exemplo, um fundo de pensões pode utilizar estatísticas para identificar os tipos de títulos que deve manter em sua carteira de investimentos. Um fundo de hedge pode utilizar estatísticas para

identificar oportunidades de negociação lucrativas. Um banco de investimento pode prever o estado futuro da economia para determinar quais novos ativos deve manter em sua própria carteira.

Embora a estatística seja uma disciplina quantitativa, o objetivo final da análise estatística é explicar eventos reais. Isso significa que, além da aplicação rigorosa de métodos estatísticos, sempre há muito espaço para julgamento. Como resultado, você pode pensar na análise estatística como ciência e arte; a arte vem da escolha da técnica estatística apropriada para determinada situação e da interpretação correta dos resultados.

Neste capítulo, apresento uma breve introdução aos conceitos abordados ao longo do livro. Apresento várias técnicas importantes que ajudam a medir e analisar as propriedades estatísticas de variáveis do mundo real, como preços de ações, taxas de juros, lucros corporativos etc.

Representando as Principais Propriedades dos Dados

A palavra *dados* refere-se a uma coleção de valores quantitativos (numéricos) ou qualitativos (não numéricos). Os dados quantitativos podem consistir em preços, lucros, vendas ou qualquer variável que possa ser medida em uma escala numérica. Os dados qualitativos podem consistir em cores, nomes de marcas, localizações geográficas e assim por diante. A maioria dos dados encontrados em aplicações comerciais é quantitativa.



PAPO DE ESPECIALISTA

A palavra *dados* é, na verdade, o plural de datum; datum refere-se a um único valor, enquanto dados se refere a um conjunto de valores.

Você pode analisar dados com técnicas gráficas ou medidas numéricas. Exploro ambas as opções nas seções a seguir.

Analizando dados com gráficos

Os gráficos são uma representação visual de um conjunto de dados, facilitando a visualização de padrões e outros detalhes. A escolha do tipo de gráfico a ser usado depende do tipo de dados que você está tentando analisar. Aqui estão alguns dos tipos mais comuns de gráficos usados em estatísticas comerciais:

- » **Histogramas:** um histograma mostra a distribuição dos dados entre diferentes intervalos ou categorias, usando uma série de barras verticais.
- » **Gráficos de linha:** um gráfico de linha mostra como uma variável muda ao longo do tempo.
- » **Gráficos de pizza:** um gráfico de pizza mostra como os dados são distribuídos entre diferentes categorias, ilustrados como uma série de fatias tiradas de uma pizza.
- » **Gráficos de dispersão (diagramas de dispersão):** um gráfico de dispersão mostra a relação entre duas variáveis como uma série de pontos. O padrão dos pontos indica o grau de relação entre as duas variáveis.

Histogramas

Você pode usar um histograma com dados quantitativos ou qualitativos. Ele foi projetado para mostrar como uma variável está distribuída entre diferentes categorias. Por exemplo, suponha que uma empresa de marketing faça uma pesquisa com 100 consumidores para determinar sua cor favorita. As respostas são

Vermelho:	23
Azul:	44
Amarelo:	12
Verde:	21

Os resultados podem ser ilustrados com um histograma, com cada cor em uma única categoria. As alturas das barras indicam o número de respostas para cada cor, facilitando a identificação das cores mais populares (veja Figura 1-1).

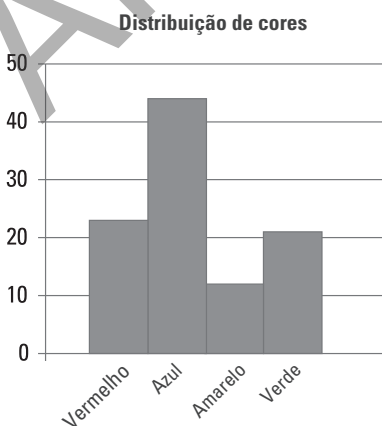


FIGURA 1-1:
Histograma
de cores
preferidas.

Com base no histograma, é possível perceber rapidamente que uma cor é a escolha mais popular, enquanto outra é menos.

Gráficos de linha

É possível usar um gráfico de linha com dados quantitativos. Ele mostra os valores de uma variável durante determinado intervalo de tempo. Por exemplo, a Figura 1-2 mostra o preço diário do ouro entre 1º de agosto de 2023 e 29 de setembro de 2023.

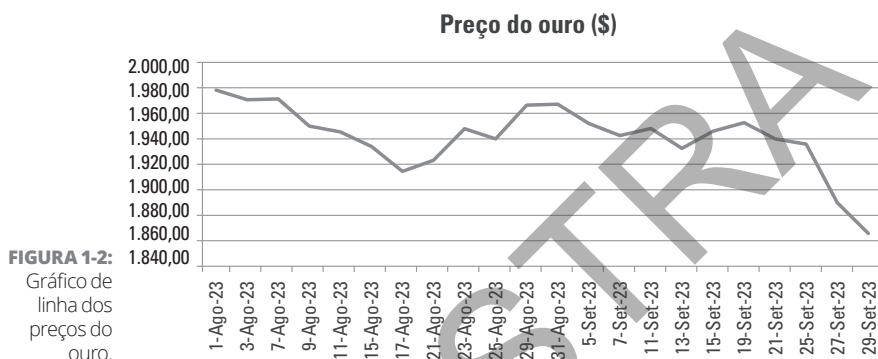


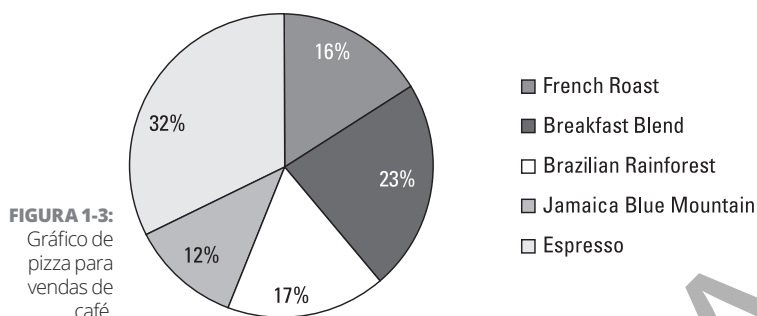
FIGURA 1-2:
Gráfico de
linha dos
preços do
ouro.

Com um gráfico de linhas, é fácil ver tendências ou padrões em um conjunto de dados. Esses gráficos podem ser usados por investidores para identificar quais ativos provavelmente terão valorização no futuro com base em seu desempenho passado.

Gráficos de pizza

Use um gráfico de pizza com dados quantitativos ou qualitativos para mostrar a distribuição dos dados entre diferentes categorias. Por exemplo, suponha que uma rede de cafeterias queira analisar suas vendas por estilo de café. Os estilos que a rede vende são French Roast, Breakfast Blend, Brazilian Rainforest, Jamaica Blue Mountain e Espresso. A Figura 1-3 mostra a proporção das vendas para cada estilo.

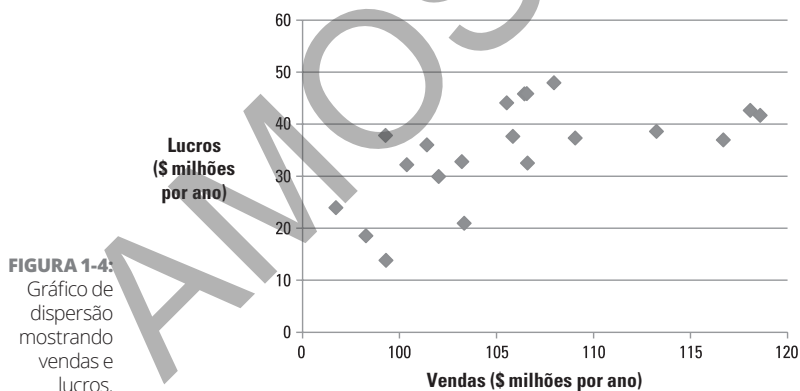
Distribuição de vendas por estilo



O gráfico mostra que o Espresso é o estilo mais vendido da rede, enquanto o Jamaica Blue Mountain representa a menor porcentagem das vendas da rede.

Gráficos de dispersão

Um gráfico de dispersão é projetado para mostrar a relação entre duas variáveis quantitativas. Por exemplo, a Figura 1-4 mostra a relação entre as vendas e os lucros de uma empresa nos últimos 20 anos.



Cada ponto no gráfico de dispersão representa o lucro e as vendas de um único ano. O padrão dos pontos mostra que níveis mais altos de vendas tendem a ser acompanhados por níveis mais altos de lucros e vice-versa. Isso é chamado de relação positiva entre duas variáveis.

Definindo propriedades e relações com medidas numéricas

Medida numérica é um valor que descreve uma propriedade fundamental de um conjunto de dados. Por exemplo, para determinar se os residentes de uma cidade tendem a ser mais velhos do que os residentes de outra cidade, você pode calcular e comparar a idade média dos residentes de cada cidade. Algumas das propriedades mais importantes de interesse em um conjunto de dados são o *centro* dos dados e a *dispersão* entre as observações.

Encontrando o centro dos dados

Para identificar o centro de um conjunto de dados, você usa medidas conhecidas como *medidas de tendência central*; as mais importantes são a média, a mediana e a moda.

A *média* representa o valor médio em um conjunto de dados, enquanto a *mediana* representa o ponto médio. A mediana é um valor que separa os dados em duas metades; metade dos elementos no conjunto de dados é *menor* ou *igual* à mediana, e a outra metade é *maior* ou *igual* à mediana. A *moda* é o valor que ocorre com mais frequência no conjunto de dados.

A média é a medida de tendência central mais amplamente utilizada, mas pode fornecer resultados enganosos se os dados contiverem valores grandes ou pequenos incomuns, conhecidos como *valores atípicos*. Neste caso, a mediana fornece uma medida mais representativa do centro dos dados. Por exemplo, o rendimento mediano das famílias é normalmente reportado pelas agências governamentais em vez do rendimento médio das famílias. Isto porque o rendimento médio das famílias é inflacionado pela presença de um pequeno número de famílias extremamente ricas. Como resultado, acredita-se que a renda familiar mediana seja uma medida melhor de como os padrões de vida estão mudando ao longo do tempo.

A moda pode ser usada para dados quantitativos ou qualitativos. Por exemplo, ela pode ser usada para determinar o número mais comum de anos de escolaridade entre os funcionários de uma empresa. Também pode ser usada para determinar o sabor mais popular vendido por um fabricante de refrigerantes.

Medindo a dispersão dos dados

As *medidas de dispersão* identificam o quanto um conjunto de dados está disperso em relação ao centro. Isso fornece uma maneira de determinar se os membros de um conjunto de dados tendem a estar muito próximos uns dos outros ou se tendem a estar amplamente dispersos. Algumas das medidas de dispersão mais importantes são

- » Variância
- » Desvio-padrão
- » Percentis
- » Quartis
- » Intervalo interquartil (IQR)

A *variância* é uma medida da diferença média ao quadrado entre os elementos de um conjunto de dados e a média. Quanto maior a variância, mais “dispersos” estão os dados. A variância é frequentemente utilizada como medida de risco em aplicações comerciais; por exemplo, pode ser utilizada para mostrar o grau de incerteza que existe em relação ao retorno de uma ação.

O *desvio-padrão* é a raiz quadrada da variância e é mais comumente usado do que a variância (porque a variância é expressa em unidades quadradas). Por exemplo, a variância de uma série de preços de gás é medida em dólares ao quadrado, o que é difícil de interpretar. O desvio-padrão correspondente é medido em dólares, o que é muito mais intuitivo e claro.

Os *percentis* dividem um conjunto de dados em 100 partes iguais, cada uma consistindo em 1% do total. Por exemplo, se a pontuação de um aluno em um exame padronizado está no 80º percentil, então o aluno teve uma pontuação igual ou melhor que 80% dos outros alunos que fizeram o exame. Um *quartil* é um tipo especial de percentil; ele divide um conjunto de dados em quatro partes iguais, cada uma consistindo em 25% do total. O primeiro quartil é o 25º percentil de um conjunto de dados, o segundo quartil é o 50º percentil e o terceiro quartil é o 75º percentil. O *intervalo interquartil* identifica os 50% centrais das observações em um conjunto de dados; ele é igual à diferença entre o terceiro e o primeiro quartis.

Determinando a relação entre duas variáveis

Para algumas aplicações, é necessário compreender a relação entre duas variáveis. Por exemplo, se um investidor deseja compreender o risco de uma carteira de ações, é essencial medir adequadamente o grau de proximidade entre os retornos das ações. É possível determinar a relação entre duas variáveis com duas medidas de associação: covariância e correlação.

A *covariância* é usada para medir a tendência de duas variáveis subirem acima de suas médias ou caírem abaixo de suas médias ao mesmo tempo. Por exemplo, suponha que uma empresa de bioengenharia descubra que o aumento dos gastos com pesquisa e desenvolvimento normalmente leva a um aumento no desenvolvimento de novas patentes. Nesse caso, os gastos com P&D e as novas patentes teriam uma covariância positiva. Se a mesma empresa descobrir que o aumento dos custos de mão de obra normalmente reduz os lucros corporativos, então os custos de mão de obra e os lucros teriam uma covariância negativa. Se a empresa descobrir que os lucros não estão relacionados à temperatura média diária, então essas duas variáveis terão uma covariância muito próxima de zero.

A *correlação* é uma medida intimamente relacionada. Ela é definida como um valor entre -1 e 1 , portanto, interpretar a correlação é mais fácil do que a covariância. Por exemplo, uma correlação $0,9$ entre duas variáveis indicaria uma relação positiva muito forte, enquanto uma correlação $0,2$ indicaria uma relação bastante fraca, mas positiva. Uma correlação $-0,8$ indicaria uma relação negativa muito forte; uma correlação $-0,3$ indicaria uma relação negativa fraca. Uma correlação 0 mostraria que duas variáveis não têm relação linear entre si.

Probabilidade: A Base de Toda Análise Estatística

A *teoria da probabilidade* fornece uma estrutura matemática para medir a incerteza. Essa área é importante para as aplicações comerciais, uma vez que grande parte da análise estatística está fortemente relacionada à teoria da probabilidade. Compreender a teoria da probabilidade fornece insights fundamentais sobre todos os métodos estatísticos usados neste livro.

A probabilidade baseia-se fortemente na noção de conjuntos. Conjunto é uma coleção de objetos. Esses objetos podem ser números, cores, sabores e assim por diante. Este capítulo foca os conjuntos de números que podem representar preços, taxas de retorno e assim por diante. Várias operações matemáticas podem ser aplicadas aos conjuntos — união, interseção e complemento, por exemplo.

A união de dois conjuntos é um novo conjunto que contém todos os elementos dos dois conjuntos originais. A interseção de dois conjuntos é um conjunto que contém apenas os elementos contidos nos *dois* conjuntos originais (se houver). O complemento de um conjunto é um conjunto que contém os elementos que não estão no conjunto original. Por exemplo, o complemento do conjunto de cartas pretas em um baralho padrão é o conjunto que contém todas as cartas vermelhas.

A teoria da probabilidade é baseada em um modelo de como resultados aleatórios são gerados, conhecido como *experimento aleatório*. Os resultados são gerados de tal forma que todos os resultados possíveis são conhecidos antecipadamente, mas o resultado real não é conhecido. As regras a seguir ajudam a determinar a probabilidade de ocorrência de resultados específicos:

- » Regra da adição
- » Regra da multiplicação
- » Regra do complemento

Você usa a regra da adição para determinar a probabilidade da união de dois conjuntos. A regra da multiplicação é usada para determinar a probabilidade da interseção de dois conjuntos. A regra do complemento é usada para identificar a probabilidade de que o resultado de um experimento aleatório não seja um elemento de um conjunto especificado.

Variáveis aleatórias

Uma *variável aleatória* atribui valores numéricos aos resultados de um experimento aleatório. Por exemplo, quando você joga uma moeda duas vezes, está realizando um experimento aleatório porque

- » Todos os resultados possíveis são conhecidos antecipadamente.
- » O resultado real não é conhecido antecipadamente.

O experimento consiste em duas tentativas. Em cada tentativa, o resultado deve ser “cara” ou “coroa”.

Suponha que uma variável aleatória X seja definida como o número de “caras” que aparecem durante este experimento. X atribui valores aos resultados do experimento da seguinte forma:

Resultado	X
{TT}	0
{HT, TH}	1
{HH}	2

onde:

T representa coroa em um único lançamento

H representa cara em um único lançamento

TT representa duas coroas consecutivas

HT representa cara seguida de coroa

TH representa coroa seguida de cara

HH representa duas caras consecutivas

X atribui um valor 0 ao resultado TT porque nenhuma cara apareceu. X atribui um valor 1 tanto a HT quanto a TH porque uma cara apareceu em cada caso. Da mesma forma, X atribui um valor 2 a HH porque duas caras apareceram.

Distribuições de probabilidade

Uma *distribuição de probabilidade* é uma fórmula ou uma tabela usada para atribuir probabilidades a cada valor possível de uma variável aleatória X . Uma distribuição de probabilidade pode ser *discreta*, o que significa que X pode assumir um número finito (contável) de valores, ou *contínua*, caso em que X pode assumir um número infinito (incontável) de valores diferentes.

Para o experimento de lançamento de moeda da seção anterior, a distribuição de probabilidade de X pode ser uma tabela simples que mostra a probabilidade de cada valor possível de X , escrita como $P(X)$:

X	P(X)
0	0,25
1	0,50
2	0,25

A probabilidade de $X = 0$ (que não apareça nenhuma cara) é igual a 0,25, porque essa experiência tem quatro resultados igualmente prováveis: HH, HT, TH e TT, e apenas em um dos casos não haverá nenhuma cara. As outras probabilidades são calculadas de forma semelhante.

Distribuições de probabilidade discretas

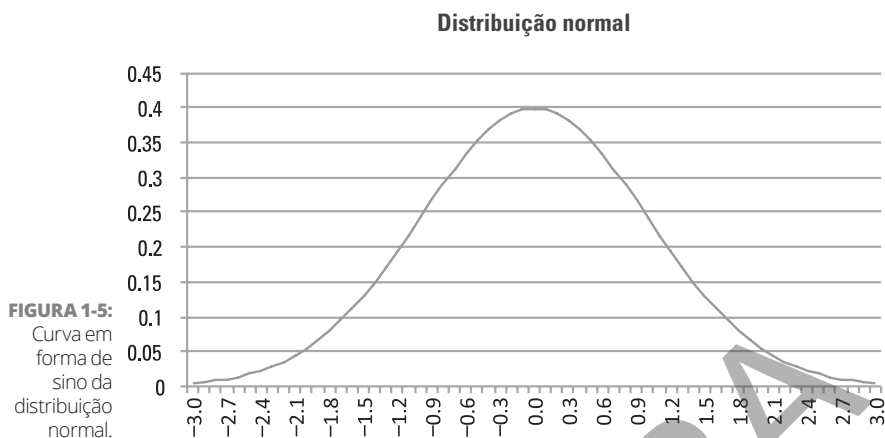
Várias distribuições de probabilidade discretas especializadas são úteis para aplicações específicas. Para aplicações comerciais, duas distribuições discretas utilizadas com frequência são:

- » Binomial
- » Poisson

Você usa a *distribuição binomial* para calcular probabilidades para um processo em que apenas um dos dois resultados possíveis pode ocorrer em cada tentativa. Você pode usar a *distribuição de Poisson* para medir a probabilidade de que determinado número de eventos ocorra durante determinado período de tempo.

Distribuições de probabilidade contínuas

Muitas distribuições contínuas podem ser usadas para aplicações comerciais; uma das mais utilizadas é a distribuição normal. A *distribuição normal* é útil para inúmeras aplicações em muitas disciplinas. Em aplicações comerciais, variáveis como retornos de ações são frequentemente consideradas como seguindo a distribuição normal. A distribuição normal é caracterizada por uma *curva em forma de sino*, e as áreas sob essa curva representam probabilidades. A curva em forma de sino é mostrada na Figura 1-5.



A distribuição normal tem muitas propriedades estatísticas convenientes que a tornam uma escolha popular para a modelagem estatística. Uma dessas propriedades é conhecida como *simetria*, a ideia de que as probabilidades de valores abaixo da média são correspondidas pelas probabilidades de valores que estão igualmente acima da média.

Usando Técnicas de Amostragem e Distribuições Amostrais

A amostragem é usada para determinar as propriedades estimadas de uma população a partir de dados amostrais. *População* é uma coleção de dados que alguém tem interesse em estudar. *Amostra* é uma seleção de dados escolhidos aleatoriamente de uma população. Por exemplo, se uma universidade estiver interessada em analisar a distribuição das médias de notas (GPAs) de seus alunos de MBA, a população de interesse seria as GPAs de todos os alunos de MBA da universidade; uma amostra consistiria nas GPAs de um conjunto de alunos de MBA escolhidos aleatoriamente.

Várias abordagens podem ser usadas para escolher as amostras; uma amostra é um subconjunto da população subjacente.

Estatística é a medida resumida de uma amostra, enquanto *parâmetro* é a medida resumida de uma população. As propriedades de uma estatística podem ser determinadas com uma distribuição amostral — um tipo especial de distribuição de probabilidade que descreve as propriedades de uma estatística.

O *teorema do limite central* (CLT) fornece as condições sob as quais a média de uma amostra segue a distribuição normal:

- » A população subjacente é normalmente distribuída.
- » O tamanho da amostra é “grande” (pelo menos 30).

Inferência Estatística: Tirando Conclusões a Partir dos Dados

A *inferência estatística* refere-se ao processo de tirar conclusões sobre uma população a partir de amostras escolhidas aleatoriamente. Nas seções a seguir, discuto duas técnicas utilizadas para a inferência estatística: intervalos de confiança e testes de hipóteses.

Intervalos de confiança

Intervalo de confiança é um conjunto de valores que se espera conter o valor de um parâmetro populacional com um nível de confiança especificado (como 90%, 95%, 99% etc.). Por exemplo, você pode construir um intervalo de confiança para a média populacional seguindo estas etapas:

1. Estime o valor da média populacional calculando a média de uma amostra escolhida aleatoriamente (conhecida como média amostral).
2. Calcule o limite inferior do intervalo de confiança subtraindo uma margem de erro da média amostral.
3. Calcule o limite superior do intervalo de confiança adicionando a mesma margem de erro à média amostral.

A margem de erro depende do tamanho da amostra usada para construir o intervalo de confiança, se o desvio-padrão da população é conhecido e o nível de confiança escolhido.

O intervalo resultante é conhecido como intervalo de confiança. Um intervalo de confiança é construído com um nível específico de probabilidade. Por exemplo, suponha que você extraia uma amostra de ações de uma carteira e construa um intervalo de confiança de 95% para o retorno médio das ações em toda a carteira.

(limite inferior, limite superior) = (0,02, 0,08)

Os retornos de toda a carteira são a população de interesse. O retorno médio em cada amostra extraída é uma *estimativa* da média populacional. A média amostral será ligeiramente diferente cada vez que uma nova amostra for extraída, assim como o intervalo de confiança. Se esse processo for repetido 100 vezes, 95 dos intervalos de confiança resultantes conterão a verdadeira média populacional.

Teste de hipóteses

O *teste de hipóteses* é um procedimento que utiliza dados amostrais para tirar conclusões sobre as características da população subjacente.

O procedimento começa com uma afirmação, conhecida como *hipótese nula*. A hipótese nula é considerada verdadeira, a menos que sejam encontradas evidências suficientes contra ela. Uma *hipótese alternativa* — o resultado aceito se a hipótese nula for rejeitada — também é afirmada.

Você calcula uma estatística de teste e a compara com um *valor crítico* (ou valores) para determinar se a hipótese nula deve ser rejeitada. A estatística de teste específica e o(s) valor(es) crítico(s) dependem de qual parâmetro populacional está sendo testado, do tamanho da amostra utilizada e de outros fatores.

Se a estatística de teste é muito extrema (por exemplo, muito grande em comparação com o(s) valor(es) crítico(s)), a hipótese nula é rejeitada em favor da hipótese alternativa; caso contrário, a hipótese nula não é rejeitada.



PAPO DE
ESPECIALISTA

Se a hipótese nula não for rejeitada, isso não significa necessariamente que ela seja verdadeira; significa apenas que não há evidências suficientes para justificar sua rejeição.

O teste de hipóteses é um procedimento geral e pode ser usado para tirar conclusões sobre muitas características de uma população, como sua média, variância, desvio-padrão e assim por diante.

Análise de regressão simples

A *análise de regressão* usa dados amostrais para estimar a força e a direção da relação entre duas ou mais variáveis. A análise de regressão simples estima a relação entre uma variável dependente (Y) e uma única variável independente (X).

Por exemplo, suponha que você esteja interessado em analisar a relação entre os retornos anuais do Índice Standard & Poor's (S&P) 500 e os retornos anuais das ações da Apple. Você pode supor que os retornos das ações da Apple estão relacionados aos retornos do S&P 500 porque o índice é um reflexo da força geral da economia. Os retornos das ações da Apple podem ser tratados como a variável dependente (Y) e os retornos do S&P 500 podem ser tratados como a variável independente (X). Você pode usar a análise de regressão para medir a relação numérica entre o S&P 500 e as ações da Apple.

A análise de regressão simples baseia-se na suposição de que existe uma relação linear entre X e Y . Uma relação linear assume a seguinte forma:

$$Y = mX + b$$

Y é a variável dependente, X é a variável independente, m é a inclinação e b é a interceptação. A inclinação indica quanto Y muda devido a uma mudança específica em X ; a interceptação indica qual seria o valor de Y se X tivesse um valor zero.

O objetivo da análise de regressão é encontrar uma linha que melhor se ajuste ou explique os dados. A linha de regressão da população é escrita da seguinte forma:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

Nesta equação, Y_i é a variável dependente, X_i é a variável independente, β_0 é a interceptação, β_1 é a inclinação e ϵ_i é um termo de erro.

Uma linha de regressão de amostra, estimada a partir dos dados, é escrita da seguinte forma:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

Aqui, $\hat{Y}_i$ é o valor estimado de Y_i , $\hat{\beta}_0$ é o valor estimado de β_0 , $\hat{\beta}_1$ é o valor estimado de β_1 e X_i é a variável independente.

A linha de regressão da amostra mostra a relação estimada entre Y e X ; você pode usar essa relação para determinar como Y responde a determinada mudança em X . Também pode usá-la para prever valores futuros de Y com base em valores presumidos de X .

Após estimar a linha de regressão da amostra, os resultados são submetidos a uma série de testes para determinar se a equação é válida. Se a equação não é válida, você rejeita os resultados e tenta um novo modelo.