Revisão e Notas sobre Estatística II

Pedro Rupf Pereira Viana

16 de janeiro de 2026

1 Introdução e Objetivos

Trata-se de uma revisão teórica e expositiva sobre Intervalos de Confiança, Testes de Hipóteses, Distribuição T de Student, Distribuição Binomial, Distribuição de Poisson, Qui-Quadrado, ANOVA, Métricas de Erros, bem como suas aplicações na área de Análise e Ciência de Dados.

2 Intervalos de Confiança

Os intervalos de confiança constituem um dos instrumentos mais importantes e amplamente utilizados na estatística inferencial contemporânea. Eles permitem que o pesquisador passe de uma estimativa pontual (um único valor calculado a partir da amostra) para uma estimativa intervalar, reconhecendo explicitamente a incerteza inerente ao processo de amostragem e oferecendo uma medida quantitativa de precisão da inferência.

Por exemplo: Imagine que você quer saber qual é a altura média dos brasileiros adultos. Você não tem como medir os 160 milhões de brasileiros adultos que existem, então você pega uma amostra aleatória de 1.000 pessoas, mede todo mundo e calcula que a média dessa amostra é 1,71 m.

Ótimo... mas será que essa média de 1,71 m é exatamente a média da população inteira? Provavelmente não. Amostras variam. Se você pegasse outra amostra de 1.000 pessoas amanhã, talvez desse 1,708 m ou 1,713 m. Esse fenômeno chama-se variabilidade amostral. Então, como a gente "chega perto" da verdade sem medir todo mundo? É aí que entra o intervalo de confiança.

De forma simples e intuitiva, um intervalo de confiança é uma faixa de valores, calculada a partir dos dados amostrais, que provavelmente contém o verdadeiro parâmetro populacional desconhecido. O nível de confiança associado (geralmente 90%, 95% ou 99%) indica a proporção de vezes que intervalos construídos pelo mesmo procedimento conteriam o parâmetro real se o processo de amostragem fosse repetido infinitas vezes sob as mesmas condições.

2.1 Interpretação correta do intervalo de confiança

É essencial distinguir a interpretação correta da interpretação frequentista clássica de algumas compreensões equivocadas bastante comuns. Um intervalo de 95% de confiança para a média populacional μ não significa que "há 95% de probabilidade de μ estar dentro do intervalo calculado", pois, na perspectiva frequentista, o parâmetro μ é um valor fixo (ainda que desconhecido) e não uma variável aleatória com distribuição de probabilidade.

A interpretação rigorosa é a seguinte: o método utilizado para construir o intervalo tem a propriedade de longo prazo de, em 95% das amostras possíveis, gerar intervalos que contêm o verdadeiro valor de μ. Em outras palavras, antes de coletar os dados, temos 95% de confiança de que o procedimento produzirá um intervalo correto; após obter o intervalo específico, ele ou contém μ ou não contém — mas continuamos confiando no método que o gerou.

Definição Formal: Seja θ um parâmetro populacional desconhecido (média μ, proporção π, diferença de médias $\mu_1 - \mu_2$, razão de chances, coeficiente de regressão etc.). Um intervalo de confiança de $100(1-\alpha)\%$ para θ é um par de estatísticas (L, U), calculadas

a partir da amostra aleatória, tais que:

$$P(L < \theta < U) = 1 - \alpha \tag{2.1}$$

para todo θ no espaço paramétrico, onde L é o limite inferior e U o limite superior.
A probabilidade é tomada antes da observação dos dados (pré-experimental); após observar a amostra, o intervalo ou contém θ ou não.

2.1.1 Caso clássico: média populacional com variância conhecida

Seja $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ uma amostra aleatória com σ conhecido.
A média amostral segue:

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Padronizando:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

Seja $z_{\alpha/2}$ o quantil tal que $P(Z > z_{\alpha/2}) = \alpha/2$ (ex.: $z_{0{,}025} = 1{,}96$ para 95%).
Então:

$$P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha$$

Reorganizando:

$$P\left(\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Portanto, o intervalo de confiança $100(1-\alpha)\%$ para μ é:

$$\left[\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

2.2 Fatores que influenciam a amplitude do intervalo

A largura do intervalo de confiança reflete o grau de incerteza da estimativa e depende basicamente de três elementos:

- Variabilidade dos dados (desvio-padrão amostral ou populacional estimado): quanto maior a dispersão natural da característica na população, maior será a incerteza e, consequentemente, mais largo o intervalo.
- Tamanho da amostra: há uma relação inversa com a raiz quadrada do tamanho amostral. Dobrar o tamanho da amostra reduz a largura aproximadamente pela metade (mais precisamente, pela raiz de $2 \approx 1{,}41$).
- Nível de confiança escolhido: níveis mais altos (ex.: 99% em vez de 95%) exigem intervalos mais largos para garantir a cobertura desejada do parâmetro.

Esses três fatores permitem ao pesquisador planejar estudos com precisão desejada: definir previamente o tamanho amostral necessário para obter intervalos suficientemente estreitos para os objetivos da pesquisa.

3 Testes de Hipóteses

Os testes de hipóteses representam um pilar central da estatística inferencial, permitindo que pesquisadores e analistas avaliem evidências empíricas para tomar decisões informadas sobre populações com base em amostras. Desenvolvidos principalmente no início do século XX, esses testes fornecem um *framework* sistemático para confrontar suposições teóricas com dados observados, promovendo uma abordagem rigorosa e replicável na ciência. Neste texto, exploraremos os princípios básicos dos testes de hipóteses, sua interpretação, exemplos práticos em diversas áreas e considerações éticas e metodológicas, sem aprofundar em derivações matemáticas complexas.

3.1 Conceitos Básicos e Estrutura Geral

Em essência, um teste de hipóteses começa com a formulação de duas proposições opostas: a hipótese nula (H_0) e a hipótese alternativa (H_1 ou H_a). A hipótese nula geralmente representa o *status quo* ou a ausência de efeito (por exemplo, quando não há diferença entre os grupos ou quando o parâmetro é igual a um valor específico). Já a hipótese alternativa expressa a mudança ou o efeito que o pesquisador suspeita existir, como há uma diferença, ou o parâmetro é maior que um valor.

O processo envolve coletar uma amostra da população de interesse, calcular uma estatística de teste (i.e. uma média, proporção ou diferença) e compará-la com o que seria esperado sob a hipótese nula. Se os dados observados forem altamente improváveis sob H_0, rejeitamos a nula em favor da alternativa; caso contrário, não a rejeitamos (no entanto, isto não equivale a aceitá-la como uma hipótese verdadeira). Isto é, matematicamente falando, seja θ um parâmetro populacional de interesse (por exemplo, média μ, proporção π, diferença $\mu_1 - \mu_2$ etc.). Um teste de hipóteses envolve duas afirmações mutuamente exclusivas e exaustivas:

- Hipótese nula (H_0): geralmente uma afirmação de "ausência de efeito" ou igualdade a um valor específico.

  Exemplo: H_0: $\mu = \mu_0$, H_0: $\mu_1 - \mu_2 = 0$, H_0: $\pi = 0,5$.

- Hipótese alternativa (H_1 ou H_a): a afirmação que o pesquisador deseja evidenciar. Pode ser:

  - Bilateral: H_1: $\mu \neq \mu_0$
  - Unilateral à direita: H_1: $\mu > \mu_0$
  - Unilateral à esquerda: H_1: $\mu < \mu_0$

Um elemento chave é o nível de significância (α), frequentemente definido em 5% ou 1%, que representa o risco aceitável de rejeitar H_0 quando ela é verdadeira (também conhecido como erro Tipo I). Há também o erro Tipo II (β), que ocorre quando falhamos em rejeitar H_0 apesar dela ser falsa, relacionado ao poder do teste (1 - β), que indica a capacidade de detectar um efeito real.

A partir de uma amostra aleatória $X_1, \cdots, X_n$, calcula-se uma estatística de teste $T = T(X_1, \cdots, X_n)$ cuja distribuição sob H_0 é conhecida (exata ou assintoticamente). A decisão é tomada comparando T com uma região crítica C definida previamente:

- Se $T \in C \rightarrow$ rejeita-se H_0 (resultado "estatisticamente significativo").

- Se $T \notin C \longrightarrow$ não se rejeita H_0.

O nível de significância α, (geralmente 0,05 ou 0,01), é a probabilidade de erro Tipo I:

$$\alpha = P(\text{rejeitar } H_0 \mid H_0 \text{ verdadeira}) = P(T \in C \mid H_0) \tag{3.1}$$

3.2 Interpretação dos Resultados: Valor-p e Suas Implicações

O valor-p é uma medida central nos testes de hipóteses, representando a probabilidade de obter resultados tão ou mais extremos que os observados, assumindo que H_0 seja verdadeira. Um valor-p baixo (menor que α) sugere que os dados são inconsistentes com a nula, levando à rejeição; um valor-p alto indica compatibilidade, resultando em não rejeição.

É **crucial** interpretar o valor-p com cautela. Ele não mede a probabilidade de H_0 ser verdadeira, nem a magnitude do efeito, **apenas a compatibilidade dos dados com a probabilidade nula**. Críticas recentes, como as da American Statistical Association (ASA) em 2016, enfatizam que o valor-p isolado pode levar a interpretações equivocadas, como confundir significância estatística com importância prática. Por isso, recomenda-se complementar os testes com estimativas de efeito (como tamanhos de efeito) e intervalos de confiança, que oferecem uma visão mais nuançada da precisão e relevância dos achados.

O valor-p é definido como a probabilidade, sob H_0 verdadeira, de obter uma estatística de teste tão ou mais extrema que a observada, na direção da alternativa:

- Para teste bilateral: $p = 2xP(T \geq |t_{obs}|H_0)$ (quando a distribuição é simétrica).
- Para teste unilateral à direita: $p = P(T \geq t_{obs}|H_0)$.

3.3 Tipos de Testes e Suas Aplicações

Existem diversos tipos de testes de hipóteses, adaptados a diferentes contextos e tipos de dados. Para comparações de médias, por exemplo, usa-se o teste T para amostras pequenas ou independentes, enquanto o teste z é aplicado em amostras grandes com variância conhecida. Em análises de proporções, o teste qui-quadrado (χ^2) ou o teste exato de Fisher são comuns para avaliar associações em tabelas de contingência.

Nas ciências da saúde, testes de hipóteses são amplamente usados em ensaios clínicos. Considere um estudo avaliando se um novo analgésico reduz a dor mais que um placebo: H_0 seria "não há diferença na redução da dor", e H_1 "o analgésico é superior". Se o valor-p for menor que 0,05, rejeita-se H_0, apoiando a aprovação do medicamento, mas sempre considerando o contexto clínico, como efeitos colaterais.

Em ciências sociais, como psicologia ou educação, testes são empregados para investigar diferenças entre grupos. Por exemplo, um pesquisador pode testar se um programa de treinamento melhora o desempenho cognitivo em idosos: H_0 "não há melhoria", H_1 "há melhoria". Resultados significativos podem informar políticas públicas, mas é essencial controlar por variáveis confundidoras para evitar conclusões espúrias.

Na economia e administração, testes de hipóteses avaliam hipóteses sobre mercados ou comportamentos. Um analista pode testar se uma campanha publicitária aumenta as vendas: H_0 "não há aumento", H_1 "há aumento". Aqui, o poder do teste é crítico, especialmente em amostras pequenas, para evitar falsos negativos que levem a decisões erradas.

3.4 Exemplos Clássicos de Testes

3.4.1 Teste z para média (σ conhecida):

Considerando H_0: $\mu = \mu_0$ versus H_1: $\mu \neq \mu_0$, calcula-se z como sendo:

$$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1) \quad \text{sob } H_0 \tag{3.2}$$

onde rejeita-se H_0 se $|z| > z_{\alpha/2}$, onde $z_{\alpha/2}$ é o quantil da normal padrão.

3.4.2 Teste T de Student para média (σ desconhecida):

Considerando H_0: $\mu = \mu_0$ versus H_1: $\mu \neq \mu_0$, calcula-se z como sendo:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \quad \text{sob } H_0 \tag{3.3}$$

onde rejeita-se $|t| > z_{\alpha/2}$.

3.4.3 Teste T para diferença entre duas médias independentes (variâncias iguais):

Considerando H_0: $\mu_1 - \mu_2 = 0$:

$$T = \frac{(\bar{X}_1 - \bar{X}_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{3.4}$$

onde s_p^2 é a variância combinada. A distribuição sob H_0 é t com $n_1 + n_2 - 2$ graus de liberdade.

3.4.4 Teste z para Proporção:

Considerando H_0: $\pi = \pi_0$ (isto é, amostras grandes):

$$z = \frac{\hat{p} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} \tag{3.5}$$

3.4.5 Teste qui-quadrado de independência:

Para tabelas de contingência rxc, comparando frequências observadas O_{ij} com esperadas E_{ij} sob independência:

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(r-1)(c-1)} \quad \text{sob } H_0 \tag{3.6}$$

3.5 Vantagens e Limitações dos Testes de Hipóteses

As vantagens dos testes de hipóteses incluem sua objetividade e capacidade de quantificar incerteza, facilitando a comunicação de resultados em artigos acadêmicos e relatórios. Eles promovem a falsificabilidade e permitem meta-análises para sintetizar evidências de múltiplos estudos.

No entanto as limitações são evidentes. O dicotomismo entre **rejeitar / não rejeitar** pode incentivar o *p-hacking* (isto é, uma manipulação de dados para obter significância), levando a crises de reprodutibilidade em campos como a psicologia. Além disso, testes não capturam a probabilidade bayesiana de hipóteses, o que tem impulsionado abordagens alternativas como a inferência bayesiana, que incorpora conhecimentos prévios.

Ética também é um aspecto importante: testes devem ser pré-especificados em protocolos de pesquisa para evitar viés, e resultados não significativos (valores-p altos) devem ser reportados para combater o viés de publicação, onde apenas achados "positivos"são divulgados.

4 Distribuição T de Student

A distribuição T de Student é uma das ferramentas mais fundamentais e amplamente utilizadas na estatística inferencial moderna, especialmente em situações em que trabalhamos com amostras pequenas e a variância populacional é desconhecida. Introduzida no início do século XX, ela permite realizar inferências sobre médias populacionais com maior realismo do que a distribuição normal padrão, reconhecendo a incerteza adicional introduzida pela estimativa da variância a partir dos próprios dados amostrais.

A distribuição t foi desenvolvida em 1908 por William Sealy Gosset, um químico e estatístico inglês que trabalhava na cervejaria Guinness em Dublin. Na época, a empresa proibia a publicação de pesquisas por seus funcionários para proteger segredos industriais, motivo pelo qual Gosset publicou seu trabalho sob o pseudônimo **Student**. Seu objetivo prático era melhorar o controle de qualidade na produção de cerveja, analisando amostras pequenas de cevada e lúpulo, onde a variância populacional não era conhecida e não podia ser assumida como fixa.

O artigo original, intitulado *The probable error of a mean*, marcou um avanço significativo: mostrou que, quando a variância é estimada a partir da amostra, a distribuição da média padronizada não segue exatamente a normal padrão (curva em forma de sino de Gauss), mas uma curva semelhante, porém com caudas mais pesadas. Essa característica reflete maior incerteza nas estimativas com amostras pequenas.

Suponha que $X_1, \cdots, X_n$ seja uma amostra aleatória independente de uma população normal com média μ e variância σ^2:

$$X_i \sim \mathcal{N}(\mu, \sigma^2), \quad i = 1, \ldots, n. \tag{4.1}$$

A média amostral é:

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \tag{4.2}$$

e a variância amostral não viesada é:

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} \tag{4.3}$$

Sabe-se que:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \quad \text{(qui-quadrado com } n-1 \text{ graus de liberdade)} \tag{4.4}$$

e que $\bar{X}$ e S^2 são independentes (propriedade das distribuições normais).

A estatística pivotal resultante é:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}}. \tag{4.5}$$

O numerador segue $\mathcal{N}(0,1)$ e o denominador é a raiz quadrada de uma variável qui-quadrado dividida por seus graus de liberdade. Portanto:

$$T \sim T_{n-1} \tag{4.6}$$

onde T_ν denota a distribuição T de Student com ν graus de liberdade.

4.1 Função Densidade de Probabilidade

A distribuição T depende de um único parâmetro: os **graus de liberdade**, geralmente denotados por ν. Os graus de liberdade estão diretamente relacionados ao tamanho da amostra: para uma amostra de tamanho n, ao estimar a variância, perdemos um grau de liberdade, resultando em $\nu = n - 1$.

Sendo assim, a densidade de probabilidade da distribuição T, com ν graus de liberdade é dada por:

$$f(t \mid \nu) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}, \quad t \in \mathbb{R} \tag{4.7}$$

onde $\Gamma(\cdot)$ é a função gama. Essa expressão mostra explicitamente a dependência dos graus de liberdade e a forma simétrica em torno de zero.

Tendo as principais propriedades definidas abaixo:

- **Simetria**: $f(-t \mid \nu) = f(t \mid \nu)$, portanto média $= 0$ (quando $\nu > 1$).
- **Variância**: $\mathrm{Var}(t) = \nu/(\nu-2)$ para $\nu > 2$ (maior que 1, refletindo caudas mais pesadas que a normal).
- **Momentos**: o quarto momento existe apenas para $\nu > 4$; curtose excessiva positiva, especialmente para ν pequeno.
- **Convergência**: quando $\nu \to \infty$, $t_\nu \to \mathcal{N}(0,1)$ em distribuição (teorema central do limite aplicado à qui-quadrado).

Podemos ver à seguir, na Figura 1, um exemplo de distribuição de densidade de probabilidade, considerando alguns graus de liberadade:

Densidades da distribuição t de Student para diferentes graus de liberdade

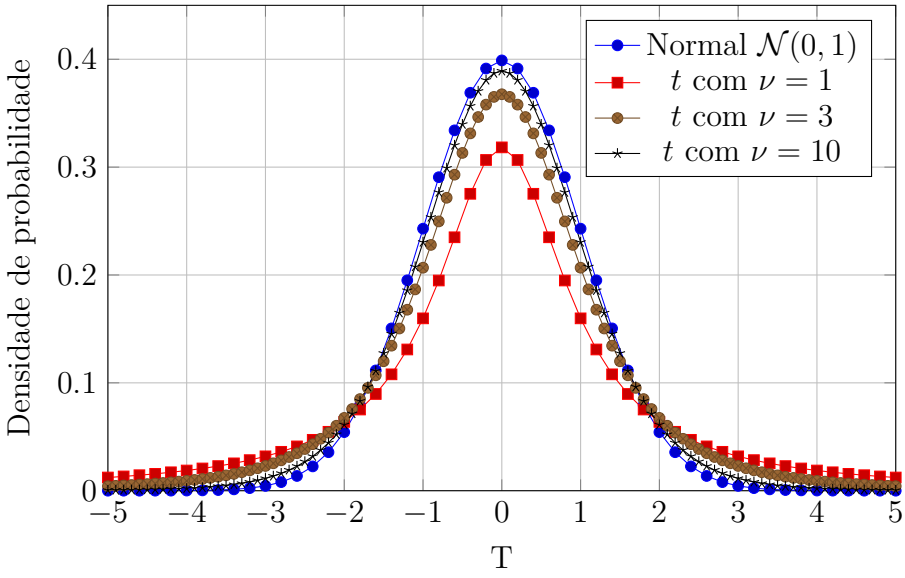

Figura 1: Comparação entre densidades da distribuição T de Student e a normal padrão. Observa-se que, à medida que os graus de liberdade aumentam, a distribuição T aproxima-se da normal.

4.2 Aplicações Principais

4.2.1 Intervalo de confiança para a média μ (σ desconhecida)

O intervalo $100(1-\alpha)\%$ é:

$$\bar{X} \pm t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}}.$$

4.2.2 Teste de hipótese para uma média

$H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ Estatística de teste:

$$t_{\text{obs}} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$

Rejeita-se H_0 ao nível α se $|t_{\text{obs}}| > t_{\alpha/2,n-1}$.

4.2.3 Teste t para duas amostras independentes (variâncias iguais)

$H_0 : \mu_1 - \mu_2 = 0$ Estatística:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

com

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$$

e $\nu = n_1 + n_2 - 2$ g.l.

4.2.4 Coeficientes em regressão linear

Para o modelo $Y = \beta_0 + \beta_1 X + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, o teste de significância de β_1 utiliza:

$$t = \frac{\hat{\beta}_1}{\mathrm{SE}(\hat{\beta}_1)} \sim t_{n-2}.$$

4.3 Robustez e Limitações

A distribuição T é exata apenas sob normalidade dos dados. Contudo, apresenta certa robustez para desvios moderados de normalidade, especialmente quando ν é grande. Em presença de assimetria forte ou *outliers*, métodos não paramétricos (tais como o de Wilcoxon e bootstrap), ou o teste de Welch (que ajusta os graus de liberdade) são preferíveis.

5 Distribuição Binomial

A distribuição binomial é uma das distribuições de probabilidade discretas mais fundamentais e amplamente empregadas na estatística e na probabilidade, servindo como base para modelar fenômenos que envolvem contagens de sucessos em um número fixo de tentativas independentes. Surgida no contexto do cálculo de probabilidades no século XVII, ela oferece uma estrutura elegante para analisar situações do mundo real onde os resultados são dicotômicos (isto é, podem ser classificados como *sucesso* ou *fracasso*).

5.1 Origem Histórica e Contexto Conceitual

A distribuição binomial tem raízes no trabalho de matemáticos como Blaise Pascal e Pierre de Fermat, que no século XVII investigaram problemas de jogos de azar, como o lançamento de dados ou moedas. No entanto, foi Jacob Bernoulli, em sua obra póstuma *Ars Conjectandi* (1713), quem formalizou o conceito ao estudar a probabilidade de obter um número específico de sucessos em repetições independentes de um experimento binário. Bernoulli demonstrou que, com um número suficiente de tentativas, a proporção de sucessos se aproxima da probabilidade verdadeira, sendo um precursor do teorema do limite central.

O termo **binomial** deriva do teorema binomial de Isaac Newton, que descreve a expansão de expressões como $(p + q)^n$, onde p é a probabilidade de sucesso, $q = 1 - p$ é a de fracasso, e n é o número de tentativas. Essa expansão reflete diretamente as probabilidades associadas a cada possível número de sucessos, ilustrando como a distribuição binomial captura a essência de processos repetitivos e independentes.

Em termos conceituais, imagine lançar uma moeda honesta (onde $p = 0,5$ para cara) dez vezes: a distribuição binomial modela a probabilidade de obter exatamente k caras (sucessos), para k variando de 0 a 10. Essa simplicidade torna a distribuição acessível para iniciantes, mas sua versatilidade a torna indispensável em pesquisas avançadas.

Em termos matemáticos, Seja $X \sim \mathrm{Bin}(n, p)$ o número de sucessos em n tentativas independentes, onde cada tentativa tem $P(\text{sucesso}) = p$ e $P(\text{fracasso}) = q = 1 - p$. A função de massa de probabilidade (fmp) é:

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \ldots, n \tag{5.1}$$

onde $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ é o coeficiente binomial. Essa expressão surge da expansão do teorema binomial $(p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$.

A função geradora de momentos (f.g.m.) é:

$$G_X(t) = (q + pt)^n \tag{5.2}$$

5.2 Características Principais da Distribuição Binomial

A distribuição binomial é definida por dois parâmetros principais: n, o número fixo de tentativas independentes, e p, a probabilidade constante de sucesso em cada tentativa. Cada tentativa deve satisfazer quatro condições chave: independência entre os eventos, número fixo de repetições, apenas dois resultados possíveis e probabilidade constante de sucesso.

Entre suas propriedades notáveis:

- **Discreta e finita:** Os valores possíveis para o número de sucessos k vão de 0 a n, resultando em uma distribuição em forma de barra que pode ser simétrica (quando $p = 0,5$) ou assimétrica (quando p está próximo de 0 ou 1).
- **Média e variância intuitivas:** A média esperada de sucessos é n_p, refletindo o que se espera "em média"(ex.: em 100 lançamentos de moeda, espera-se 50 caras). A variância, $n_p(1-p)$, indica a dispersão: maior quando p está perto de 0,5, capturando maior imprevisibilidade.
- **Forma da distribuição:** Com n pequeno, a distribuição pode ser irregular; com n grande e p moderado, ela se aproxima de uma curva em sino, facilitando aproximações por distribuições contínuas como a normal.

Essas características tornam a binomial ideal para modelar contagens, diferentemente de distribuições como a Poisson (para eventos raros sem limite superior) ou a normal (para dados contínuos).

5.3 Propriedades

- Média (esperança): $E[X] = np$
- Variância: $\text{Var}(X) = npq = np(1-p)$
- Desvio-padrão: $\sigma_X = \sqrt{np(1-p)}$
- Assimétrica: O modo é $\lfloor (n+1)p \rfloor$. A distribuição é simétrica quando $p = 0.5$; assimétrica à direita se $p < 0.5$; à esquerda se $p > 0.5$.
- Função cumulativa: $F(k) = P(X \leq k) = \sum_{i=0}^{k} \binom{n}{i} p^i q^{n-i}$

Para n grande, pela aproximação normal (teorema central do limite):

$$X \approx \mathcal{N}(np, npq) \tag{5.3}$$

com correção de continuidade: $P(a \leq X \leq b) \approx \Phi\left(\frac{b+0.5-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-0.5-np}{\sqrt{npq}}\right)$. Aproximação de Poisson: quando $n \to \infty$ e $p \to 0$ com $np = \lambda$ fixo, $X \approx \text{Poisson}(\lambda)$.

5.4 Aplicações Práticas da Distribuição Binomial

A distribuição binomial encontra aplicações em uma ampla gama de disciplinas, demonstrando sua relevância na análise de dados empíricos.

Nas ciências biológicas e médicas, ela é usada para modelar a probabilidade de resultados em ensaios clínicos ou estudos genéticos. Por exemplo, em um teste de vacina com 20 voluntários, onde a probabilidade de imunização é 0,8, a binomial calcula a chance de exatamente 16 sucessos (imunizados), auxiliando na avaliação de eficácia e no planejamento de amostras.

Em pesquisas sociais e de mercado, a distribuição auxilia na análise de opiniões ou comportamentos binários. Considere uma pesquisa com 500 eleitores, onde $p = 0,45$ é a probabilidade de apoio a um candidato: a binomial estima a probabilidade de obter entre 200 e 250 apoios, ajudando a prever resultados eleitorais com margens de erro.

Na engenharia e controle de qualidade, ela modela defeitos em processos de produção. Em uma fábrica que produz 100 itens com taxa de defeito de 0,02, a binomial avalia a probabilidade de zero defeitos (sucesso total) ou mais de cinco, orientando decisões sobre inspeções.

Em finanças e seguros, a distribuição binomial é aplicada em modelos de opções (como o modelo binomial de precificação de Cox-Ross-Rubinstein), simulando flutuações de preços de ativos em passos discretos, onde cada "subida"ou "descida"é um evento binário.

Além disso, em ecologia, ela modela a sobrevivência de espécies em habitats fragmentados, e em epidemiologia, a propagação de doenças em populações pequenas, como a probabilidade de k infecções em n contatos.

5.5 Vantagens, Limitações e Extensões

As vantagens da distribuição binomial incluem sua simplicidade computacional (facilmente calculável com softwares como R ou Excel), interpretabilidade intuitiva e robustez em cenários binários reais. Ela promove uma compreensão probabilística de incertezas, alinhando-se ao método científico ao permitir testes de hipóteses sobre proporções.

No entanto, limitações existem: assume independência estrita, o que nem sempre ocorre (ex.: em epidemias, contágios não são independentes); exige n fixo e p constante, inadequado para processos dinâmicos; e com n muito grande, cálculos manuais tornam-se impraticáveis, demandando aproximações normais ou de Poisson.

Extensões incluem a distribuição binomial negativa (para número de tentativas até k sucessos) e a multinomial (para mais de dois resultados), expandindo sua utilidade para cenários complexos.

6 Distribuição de Poisson

A distribuição de Poisson é uma das distribuições de probabilidade discretas mais importantes na estatística aplicada, especialmente projetada para modelar a contagem de eventos raros ou infrequentes que ocorrem em um intervalo fixo de tempo, espaço ou outra unidade contínua. Introduzida no início do século XIX pelo matemático francês Siméon Denis Poisson, ela oferece uma ferramenta poderosa para descrever fenômenos onde os eventos acontecem de forma independente e com uma taxa média constante, sem limite superior teórico para o número de ocorrências. Neste texto, exploraremos os fundamentos conceituais da distribuição de Poisson, sua origem histórica, propriedades intuitivas,

exemplos práticos em diferentes áreas do conhecimento e suas relações com outras distribuições, mantendo uma abordagem acessível e priorizando a compreensão intuitiva sobre formalismos matemáticos excessivos.

Do ponto de vista matemático, seja X o número de eventos em um intervalo fixo. Dizemos que $X \sim \text{Poisson}(\lambda)$ se:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots \tag{6.1}$$

A função geradora de momentos é

$$G_X(t) = e^{\lambda(t-1)} \tag{6.2}$$

e a função geradora de probabilidade é $G_X(s) = e^{\lambda(s-1)}$.

6.1 Origem Histórica

A distribuição de Poisson surgiu em 1837, no trabalho de Siméon Denis Poisson intitulado *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*. Poisson investigava a probabilidade de um número específico de condenações errôneas em julgamentos, mas o exemplo mais famoso associado à distribuição é o número de soldados prussianos mortos por coice de cavalo em diferentes corpos do exército ao longo de um ano — dados analisados posteriormente por Ladislaus Bortkiewicz em 1898, que demonstraram um ajuste impressionante à distribuição proposta por Poisson.

A intuição central é simples: imagine eventos que ocorrem aleatoriamente em um continuum (tempo, área, volume, comprimento), de forma que:

- Os eventos são independentes (a ocorrência de um não afeta a de outro).
- A taxa média de ocorrência é constante (denotada por λ).
- A probabilidade de mais de um evento em um intervalo muito pequeno é desprezível.

Nessas condições, o número de eventos em um intervalo fixo segue uma distribuição de Poisson com parâmetro λ, que representa tanto a média quanto a variância esperada do número de ocorrências.

Um exemplo clássico é a chegada de clientes a uma agência bancária: se, em média, chegam 3 clientes por hora ($\lambda = 3$), a distribuição de Poisson permite calcular a probabilidade de chegar exatamente 0, 1, 2, 3 ou mais clientes em uma hora específica.

6.2 Propriedades Principais

- **Média e variância**: $\mathbb{E}[X] = \lambda$, $\text{Var}(X) = \lambda$ (equidispersão — propriedade característica).
- **Desvio padrão**: $\sigma_X = \sqrt{\lambda}$.
- **Modo**: $\lfloor \lambda \rfloor$ (ou $\lfloor \lambda \rfloor$ e $\lfloor \lambda \rfloor - 1$ quando λ é inteiro).
- **Aditividade**: Se $X_i \sim \text{Poisson}(\lambda_i)$ independentes, então $\sum X_i \sim \text{Poisson}(\sum \lambda_i)$.
- **Função cumulativa**: $F(k) = \sum_{i=0}^{k} \frac{e^{-\lambda}\lambda^i}{i!}$.

6.3 Aproximações e Relações com Outras Distribuições

- **Limite binomial**: $X \sim \text{Bin}(n, p)$ com n grande, p pequeno e $np = \lambda$ fixo $\implies X \xrightarrow{d} \text{Poisson}(\lambda)$.
- **Aproximação normal**: Para $\lambda \geq 10$ (ou preferencialmente $\lambda \geq 20$), $X \approx \mathcal{N}(\lambda, \lambda)$. Com correção de continuidade: $P(a \leq X \leq b) \approx P(a - 0.5 < Y < b + 0.5)$, $Y \sim \mathcal{N}(\lambda, \lambda)$.

6.4 Aplicações Práticas

A distribuição de Poisson é amplamente utilizada em diversas áreas, demonstrando sua versatilidade na modelagem de contagens.

Nas ciências da saúde e epidemiologia, ela modela o número de casos de doenças raras em uma população durante um período (ex.: número de casos de uma doença genética em uma cidade por ano), o número de mutações em uma sequência de DNA ou a chegada de pacientes a um pronto-socorro em horários específicos.

Em engenharia e confiabilidade, é usada para contar falhas de equipamentos (ex.: número de quebras de uma máquina em um mês), defeitos em um metro de tecido ou acidentes em uma rodovia por dia.

No setor de serviços e gestão de filas, modela chegadas de chamadas a um call center, pedidos em um site de e-commerce ou clientes em uma fila de supermercado, servindo de base para teorias de filas (ex.: modelo M/M/1).

Em ecologia e meio ambiente, conta o número de espécies em uma área amostral, de árvores em uma parcela florestal ou de partículas radioativas detectadas por um contador Geiger.

Em finanças e seguros, modela o número de sinistros em uma carteira de apólices, de reclamações por dia ou de transações fraudulentas.

Um exemplo ilustrativo: uma central de emergências recebe, em média, 5 chamadas por hora ($\lambda = 5$). A distribuição de Poisson permite calcular a probabilidade de receber zero chamadas (período ocioso), mais de 10 (sobrecarga) ou exatamente 5 (equilíbrio), auxiliando no dimensionamento de equipes.

6.5 Vantagens, Limitações e Extensões

As vantagens da distribuição de Poisson incluem sua simplicidade (apenas um parâmetro), interpretabilidade direta (λ é taxa média) e boa adequação a dados de contagem reais. Ela promove uma visão probabilística de fenômenos aleatórios, auxiliando na previsão e planejamento.

Contudo, possui limitações: assume variância igual à média (equidispersão), o que nem sempre ocorre em dados reais, frequentemente observa-se sobredispersão (variância $>$ média) ou subdispersão (variância $<$ média). Além disso, exige independência e taxa constante, condições violadas em fenômenos com aglomeração ou tendências temporais.

Para contornar essas limitações, existem extensões como a distribuição binomial negativa (para sobredispersão) e modelos de regressão de Poisson ou *quasi-Poisson*, amplamente usados em estatística aplicada.

7 Distribuição Qui-Quadrado

A distribuição qui-quadrado (χ^2) é uma das distribuições de probabilidade contínuas mais importantes na estatística inferencial moderna, especialmente em testes de hipóteses envolvendo variâncias, tabelas de contingência e adequação de modelos. Caracterizada por ser sempre não negativa e assimétrica à direita, ela surge naturalmente quando se soma o quadrado de variáveis normais padronizadas independentes, oferecendo uma ferramenta essencial para avaliar discrepâncias entre dados observados e expectativas teóricas.

Formalmente, Sejam $Z_1, Z_2, \ldots, Z_\nu$ variáveis aleatórias independentes, cada uma com distribuição normal padrão $\mathcal{N}(0,1)$. Então a variável aleatória:

$$\chi^2 = Z_1^2 + Z_2^2 + \cdots + Z_\nu^2 \tag{7.1}$$

segue a distribuição qui-quadrado com ν graus de liberdade, denotada por $\chi^2 \sim \chi^2_\nu$.

A função densidade de probabilidade é dada por:

$$f(x \mid \nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, \quad x > 0 \tag{7.2}$$

onde $\Gamma(\cdot)$ é a função gama.

7.1 Origem Histórica

A distribuição qui-quadrado foi formalizada no início do século XX, com contribuições decisivas de Karl Pearson (que a utilizou em 1900 para desenvolver o teste de aderência) e Ronald Fisher (que a integrou ao framework de testes de hipóteses e análise de variância). Pearson buscava uma medida objetiva para avaliar se uma distribuição observada se ajustava a uma distribuição teórica esperada, enquanto Fisher expandiu seu uso para inferências sobre variâncias e associações categóricas.

A intuição básica é simples: imagine várias variáveis aleatórias independentes que seguem uma distribuição normal padrão (média zero e variância um). Se quadrarmos cada uma delas e somarmos os resultados, o valor obtido segue uma distribuição qui-quadrado. Como o quadrado elimina valores negativos e enfatiza desvios grandes, a distribuição resultante é sempre positiva, começa em zero e tem cauda longa à direita; isto é, quanto maior o número de termos somados (graus de liberdade), mais ela se aproxima de uma forma simétrica semelhante à normal.

7.2 Características Principais

A distribuição qui-quadrado depende de um único parâmetro: os *graus de liberdade*, denotados geralmente por ν. Esse parâmetro determina completamente a forma da curva:

- Para poucos graus de liberdade (i.e. $\nu = 1$ ou 2), a distribuição é fortemente assimétrica à direita, com alta probabilidade perto de zero e cauda longa.
- À medida que ν aumenta, a assimetria diminui e a distribuição se torna mais simétrica.
- Quando ν é grande (geralmente acima de 30), o valor de qui-quadrado se aproxima bastante da distribuição normal (com média ν e variância 2ν).

Outras propriedades intuitivas incluem:

- Média igual a ν: o valor esperado da variável qui-quadrado coincide com o número de graus de liberdade.
- Variância igual a 2ν: a dispersão cresce linearmente com ν.
- Sempre positiva: não há probabilidade para valores negativos, refletindo o fato de ser soma de quadrados.
- Aditividade: a soma de variáveis qui-quadrado independentes (com $\nu_1, \nu_2, \ldots, \nu_n$ graus de liberdade) também segue qui-quadrado com $\nu = \nu_1, \nu_2, \ldots, \nu_n$ graus de liberdade.

Essas características fazem do qui-quadrado uma distribuição **pivotal** em muitos procedimentos inferenciais, pois sua forma não depende de parâmetros desconhecidos da população.

7.3 Propriedades Principais

- **Domínio**: $x \geq 0$ (sempre não-negativa).
- **Média**: $\mathbb{E}[\chi^2_\nu] = \nu$.
- **Variância**: $\text{Var}(\chi^2_\nu) = 2\nu$.
- **Assimetria**: $\gamma_1 = \sqrt{8/\nu}$ (diminui com ν).
- **Curtose excessiva**: $\gamma_2 = 12/\nu$ (também diminui com ν).
- **Convergência**: Quando $\nu \to \infty$, pela lei dos grandes números e teorema central do limite,
$$\frac{\chi^2_\nu - \nu}{\sqrt{2\nu}} \xrightarrow{d} \mathcal{N}(0,1).$$
- **Aditividade**: Se $X_i \sim \chi^2_{\nu_i}$ independentes, então $\sum X_i \sim \chi^2_{\sum \nu_i}$.

7.4 Aplicações Principais

A distribuição qui-quadrado é o alicerce de vários testes estatísticos clássicos, todos baseados na comparação entre o que foi observado e o que seria esperado sob uma hipótese nula.

1. **Teste de aderência (goodness-of-fit):** Avalia se uma distribuição observada em dados categóricos se ajusta a uma distribuição teórica esperada. Por exemplo, em um dado de seis faces lançado 600 vezes, espera-se aproximadamente 100 ocorrências por face. O teste calcula quanto as frequências observadas desviam das esperadas e, se o desvio for grande demais, rejeita-se a hipótese de que o dado é honesto.

2. **Teste de independência:** Um dos usos mais comuns, onde verifica se duas variáveis categóricas são independentes em uma tabela de contingência (linhas x colunas). Exemplo clássico: uma pesquisa investiga se gênero e preferência política são associados. Calcula-se as frequências esperadas assumindo independência e compara-se com as observadas. Um valor qui-quadrado elevado indica associação estatisticamente significativa.

3. **Teste de homogeneidade:** Similar ao de independência, mas compara distribuições entre grupos diferentes. Exemplo: avaliar se a proporção de aprovação em diferentes cursos universitários é a mesma.

4. **Inferência sobre variância populacional:** Quando se tem uma amostra de uma população normal, a estatística $(n-1)S^2/\sigma^2$ segue qui-quadrado com $n-1$ graus de liberdade. Isso permite construir intervalos de confiança e testes de hipóteses para a variância σ^2 (embora menos usado hoje devido à sensibilidade à não-normalidade).

5. **Outros contextos:** Aparece em análise de variância (ANOVA), regressão linear (teste de significância global), modelos log-lineares e como base para distribuições derivadas (ex.: T de Student, F de Snedecor).

8 Análise de Variância (ANOVA)

A Análise de Variância (ANOVA, do inglês *Analysis of Variance*) constitui um dos métodos estatísticos mais amplamente utilizados nas ciências experimentais e observacionais para comparar médias de três ou mais grupos independentes. Desenvolvida principalmente por Ronald Fisher na década de 1920, no contexto da experimentação agrícola, a ANOVA oferece um framework sistemático para determinar se as diferenças observadas entre médias de grupos são suficientemente grandes para serem atribuídas a fatores controlados, em vez de mera variabilidade aleatória.

8.1 Origem Histórica

Fisher introduziu a ANOVA em seu clássico livro *Statistical Methods for Research Workers (1925)* e no contexto dos experimentos em blocos e delineamentos fatoriais na Rothamsted Experimental Station. O objetivo inicial era responder a uma pergunta simples, porém poderosa: "As diferentes variedades de trigo realmente produzem rendimentos médios diferentes, ou as variações observadas são apenas flutuações esperadas devido ao acaso?"

A intuição central da ANOVA é dividir a variabilidade total presente nos dados em duas partes principais:

- Variabilidade **entre grupos** (ou entre tratamentos): Reflete as diferenças sistemáticas causadas pelo fator de interesse (i.e. tipo de fertilizante, método de ensino, dose de medicamento).

- Variabilidade **dentro dos grupos** (ou erro/resíduo): Captura a variação natural ou aleatória que ocorre mesmo entre unidades submetidas ao mesmo tratamento.

Se a variabilidade entre grupos for muito maior do que a esperada apenas pelo erro aleatório (ou seja, muito maior do que a variabilidade dentro dos grupos), conclui-se que existe evidência estatística de que os tratamentos exercem efeito diferencial sobre a variável resposta.

8.2 Tipos Principais de ANOVA e Suas Aplicações

A ANOVA pode ser classificada conforme o número de fatores e o delineamento experimental:

1. **ANOVA de um fator (One-Way ANOVA):** Compara médias de três ou mais grupos independentes formados por um único fator. Exemplos:
   - Comparar o tempo médio de recuperação de pacientes submetidos a três diferentes tipos de fisioterapia.
   - Avaliar se o rendimento médio de plantas difere entre quatro níveis de irrigação.
   - Verificar se a satisfação média de alunos varia entre cinco modalidades de ensino (presencial, híbrido, EAD síncrono, EAD assíncrono, misto).
2. **ANOVA de dois fatores (Two-Way ANOVA):** Analisa simultaneamente o efeito de dois fatores e sua possível interação. Exemplos:
   - Efeito do tipo de solo (fator A) e do nível de adubação (fator B) no crescimento de uma cultura, verificando também se o efeito da adubação depende do tipo de solo (interação).
   - Efeito do gênero (fator A) e do nível de escolaridade (fator B) na pontuação média em um teste de raciocínio.
3. **ANOVA de medidas repetidas (Repeated Measures ANOVA):** Usada quando a mesma unidade experimental é medida em múltiplos momentos ou condições (delineamento dentro de sujeitos). Exemplos:
   - Medir a pressão arterial de pacientes em quatro momentos diferentes após administração de um medicamento.
   - Avaliar o desempenho de atletas em testes de força antes, durante e após um programa de treinamento.
4. **ANOVA fatorial e delineamentos mais complexos:** Incluem ANOVA de três ou mais fatores, ANOVA em blocos, ANOVA com medidas repetidas mistas, entre outros, permitindo investigar interações de ordem superior.

Considere k grupos independentes, com n_i observações cada (delineamento balanceado: $n_1 = n_2 = \cdots = n_k = n$). O modelo linear aditivo é:

$$X_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, \ldots, k;\ j = 1, \ldots, n$$

onde:

- μ é a média geral da população,
- τ_i é o efeito fixo do i-ésimo tratamento (com $\sum \tau_i = 0$ em alguns parametrizamentos),
- $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ são erros independentes e identicamente distribuídos.

Hipóteses:

- $H_0 : \tau_1 = \tau_2 = \cdots = \tau_k = 0$ (ou equivalentemente: $\mu_1 = \mu_2 = \cdots = \mu_k$)

- H_1: pelo menos um $\tau_i \neq 0$

Outro importante é acerca da Decomposição da Variância Total. A ANOVA baseia-se na identidade fundamental de decomposição da soma de quadrados:

$$\text{SST} = \text{SSA} + \text{SSE}$$

onde:

- SST (Soma de Quadrados Total) = $\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{..})^2$
- SSA (Soma de Quadrados entre grupos) = $n\sum_{i=1}^{k}(\bar{X}_{i.} - \bar{X}_{..})^2$
- SSE (Soma de Quadrados dentro dos grupos / erro) = $\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{i.})^2$

Graus de liberdade:

- $\text{gl}_{\text{total}} = N - 1 \quad (N = kn)$
- $\text{gl}_{\text{entre}} = k - 1$
- $\text{gl}_{\text{erro}} = N - k = k(n-1)$

Quadrados médios:

- QM_{entre} (MSA) = SSA / $(k-1)$
- QM_{erro} (MSE) = SSE / $(N-k)$

8.3 Suposições e Cuidados Metodológicos

Para que as conclusões da ANOVA sejam confiáveis, algumas suposições devem ser razoavelmente atendidas:

- Independência das observações dentro e entre grupos.
- Normalidade aproximada dos resíduos (dentro de cada grupo).
- Homocedasticidade (variâncias populacionais iguais entre grupos), onde foram verificadas por testes como os de Levene ou Bartlett.

Quando essas suposições são violadas moderadamente, a ANOVA é relativamente robusta em amostras balanceadas e de tamanho moderado a grande. Em casos de desvios graves, alternativas incluem:

- Transformações de dados (log, raiz quadrada, Box-Cox).
- Testes não paramétricos (Kruskal-Wallis para um fator; Friedman para medidas repetidas).
- Modelos lineares generalizados ou modelos mistos.

8.4 Exemplos Práticos em Diferentes Áreas

- **Medicina:** Comparar a redução média de glicemia em jejum após três diferentes esquemas dietéticos em pacientes com diabetes tipo 2.
- **Psicologia/Educação:** Avaliar se o nível de ansiedade média difere entre alunos expostos a três tipos de avaliação (prova tradicional, prova oral, portfólio).
- **Engenharia:** Testar se a resistência média à tração de concreto varia entre quatro proporções diferentes de cimento.
- **Agronomia:** Comparar o rendimento médio de soja em cinco cultivares diferentes.
- **Marketing:** Verificar se a intenção média de compra difere entre quatro embalagens distintas de um produto.

9 Métricas de Erros

A avaliação do desempenho de modelos preditivos constitui uma etapa essencial no processo de modelagem estatística e de aprendizado de máquina. Quando o objetivo é prever uma variável contínua (regressão), as métricas de erro quantificam o quanto as previsões do modelo se afastam dos valores reais observados. Essas métricas não apenas permitem comparar diferentes modelos, mas também orientam a escolha do algoritmo mais adequado ao problema, revelam limitações do modelo e comunicam a qualidade da previsão a tomadores de decisão não técnicos.

9.1 Origem Histórica

As métricas de erro têm raízes na estatística clássica do século XX, especialmente no trabalho de Fisher, Gauss e na teoria da estimação por mínimos quadrados. Com o avanço do aprendizado de máquina nas últimas décadas, elas ganharam nova relevância, sendo adaptadas para grandes volumes de dados e aplicações práticas em áreas como finanças, saúde, energia, logística e ciências sociais.

A ideia central é simples: toda previsão gera um erro (ou resíduo), definido como a diferença entre o valor real (y) e o valor previsto ($\hat{y}$). Seja y_i o valor real da i-ésima observação e $\hat{y}_i$ o valor previsto pelo modelo. Define-se o erro (ou resíduo) como:

$$e_i = y_i - \hat{y}_i, \quad i = 1, \ldots, n \tag{9.1}$$

As métricas de erro agregam os e_i de diferentes formas, destacando aspectos distintos do desempenho. do modelo:penho do modelo:

- Magnitude média dos erros
- Magnitude relativa
- Sensibilidade a erros grandes (outliers)
- Interpretabilidade em unidades originais ou percentuais

9.2 Principais Métricas de Erro Absoluto

1. **Erro Absoluto Médio (Mean Absolute Error – MAE):** Calcula a média dos valores absolutos dos erros.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i|$$

   Propriedades:

   - Mantém as unidades originais da variável resposta.
   - Robusto a outliers (não eleva erros grandes).
   - Interpretação direta: média do erro absoluto.

2. **Erro Quadrático Médio (Mean Squared Error – MSE):** Calcula a média dos quadrados dos erros.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}e_i^2$$

   Propriedades:

   - Penaliza fortemente erros grandes (efeito quadrático).
   - Unidades quadradas (ex.: reais^2), dificultando interpretação direta.
   - Base da otimização por mínimos quadrados.

3. **Raiz do Erro Quadrático Médio (Root Mean Squared Error – RMSE):** É simplesmente a raiz quadrada do MSE.

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

   Propriedades:

   - Recupera as unidades originais.
   - Mantém a penalização quadrática a erros grandes.
   - Métrica mais comum em literatura e competições de machine learning.

9.3 Métricas Relativas e Percentuais

1. **Erro Percentual Absoluto Médio (Mean Absolute Percentage Error – MAPE):** Expressa o erro médio como porcentagem do valor real.

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{e_i}{y_i}\right|$$

   Propriedades:

   - Escala relativa (percentual), independente da unidade.
   - Interpretação: erro médio percentual.

2. **Erro Percentual Absoluto Mediano (Median Absolute Percentage Error – MdAPE):** Usa a mediana em vez da média dos erros percentuais. Mais robusto a outliers e valores extremos que distorcem o MAPE.

$$\text{MdAPE} = \text{mediana}_{i=1,\ldots,n} \left\{ \left| \frac{y_i - \hat{y}_i}{y_i} \right| \right\} \times 100$$

3. **Erro Percentual Simétrico Absoluto Médio (Symmetric MAPE – sMAPE):** Corrige a assimetria do MAPE usando a média entre valor real e previsto no denominador. Recomendado quando há valores próximos de zero ou quando se deseja simetria entre superestimação e subestimação.

$$\text{sMAPE} = \frac{200}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|}$$

   Vantagem: simetria entre superestimação e subestimação; mais robusto quando valores próximos de zero.

9.4 Outras Métricas Relevantes

1. **Coeficiente de Determinação (R^2 e R^2 ajustado):** Embora não seja uma métrica de erro propriamente dita, mede a proporção da variância da variável resposta explicada pelo modelo. Se $R^2 = 1$ indica ajuste perfeito; se $R^2 = 0$ indica que o modelo não explica nada além da média. R^2 ajustado penaliza a inclusão de variáveis irrelevantes, sendo preferível em regressão múltipla.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\text{SSE}}{\text{SST}}$$

   onde $\bar{y}$ é a média dos valores observados, SSE é a soma de quadrados dos erros e SST é a soma de quadrados total.

   Interpretação: proporção da variância explicada pelo modelo. $R^2 \in (-\infty, 1]$; valores próximos de 1 indicam bom ajuste.

   No cenário de R^2 ajustado, têm-se que:

$$R^2_{\text{adj}} = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

   onde p é o número de preditores. Penaliza a inclusão de variáveis irrelevantes.

2. **Erro Máximo (Max Error):** Identifica o maior erro absoluto cometido. Útil em aplicações críticas (ex.: segurança, controle de processos) onde um erro muito grande pode ser catastrófico.

$$\text{Max Error} = \max_{i=1,\ldots,n} |y_i - \hat{y}_i|$$

   Útil para identificar o pior caso e aplicações críticas.

3. **Quantis de Erro:** Muito útil em planejamento de capacidade, estoques e previsão de demanda.

9.5 Escolha da Métrica Adequada ao Contexto

A escolha da métrica depende do objetivo do problema e do custo dos erros:

- Quando erros grandes são muito mais custosos ⇒ RMSE ou Max Error
- Quando se deseja interpretabilidade em unidades originais e robustez ⇒ MAE
- Quando se compara desempenho entre variáveis de escalas diferentes ⇒ MAPE ou sMAPE
- Quando o foco é explicação da variância ⇒ R^2
- Em previsão financeira ou de vendas ⇒ combinação de MAE + MAPE + RMSE
- Em aplicações de segurança ou monitoramento ⇒ percentis altos do erro + Max Error

9.6 Limitações e Cuidados na Interpretação

Nenhuma métrica é perfeita. Algumas armadilhas comuns incluem:

- Comparar modelos em conjuntos de dados diferentes usando apenas uma métrica.
- Ignorar o efeito de outliers em métricas quadráticas.
- Usar MAPE em séries com valores próximos de zero.
- Confundir alto R^2 com bom poder preditivo fora da amostra (overfitting).

Para isso, algumas boas práticas são recomendadas:

- Reportar múltiplas métricas (MAE, RMSE, MAPE, R^2).
- Avaliar desempenho em conjunto de validação/teste (não apenas treinamento).
- Usar validação cruzada para estimativas mais robustas.
- Complementar com gráficos diagnósticos (resíduos vs. previstos, *Q-Q plot*, histograma de erros).

10 Aplicações em Análise e Ciência de Dados

Na área de Análise e Ciência de Dados, conceitos estatísticos fundamentais servem como pilares para extrair insights acionáveis de dados, validar hipóteses e construir modelos preditivos robustos. Os tópicos discutidos até o momento são amplamente aplicados em cenários reais, como análise de dados de usuários, previsão de demanda, segmentação de clientes e otimização de processos. Neste texto, revisamos brevemente cada conceito e apresentamos exemplos práticos em Ciência de Dados, com foco em implementações computacionais usando Python. As bibliotecas como NumPy, SciPy, StatsModels e Scikit-learn facilitam essas análises, permitindo escalabilidade para grandes conjuntos de dados.

10.1 Intervalos de Confiança

Os intervalos de confiança (IC) fornecem uma faixa de valores plausíveis para um parâmetro populacional, quantificando a incerteza em estimativas baseadas em amostras. Em Ciência de Dados, eles são essenciais para relatar resultados com precisão, evitando interpretações pontuais que ignoram a variabilidade amostral.

Como exemplo de aplicação, considere um cenário em que, na análise de dados de um e-commerce, um analista quer estimar a receita média por usuário (ARPU) em uma campanha de marketing. Usando uma amostra de 1.000 transações, calcula-se o IC de 95% para a média populacional, ajudando a decidir se o ARPU atende às metas de negócio. Isso é útil em relatórios dashboards, onde a incerteza é comunicada visualmente. Usando SciPy para calcular o IC para a média de uma amostra de dados normais (simulando receitas):

```
import numpy as np
from scipy import stats

# Dados simulados: receitas de 1000 usuarios
np.random.seed(42)
receitas = np.random.normal(loc=50, scale=10, size=1000)  # Media 50,
    desvio 10

# Calculo do IC de 95% para a media
media = np.mean(receitas)
sem = stats.sem(receitas)  # Erro padrao da media
ic = stats.t.interval(confidence=0.95, df=len(receitas)-1, loc=media,
    scale=sem)

print(f"Media amostral: {media:.2f}")
print(f"IC 95%: ({ic[0]:.2f}, {ic[1]:.2f})")
```

Listing 1: Cálculo do IC para a média de uma amostra de dados normais

Saída esperada: Média amostral $\approx$ 50,10; IC 95% $\approx$ (49,50, 50,70). Isso indica que, com 95% de confiança, o ARPU verdadeiro está nessa faixa.

10.2 Testes de Hipóteses

Os testes de hipóteses avaliam se há evidência estatística suficiente para rejeitar uma suposição nula sobre os dados, comparando uma estatística observada com sua distribuição sob a nula.

Em Ciência de Dados, testes de hipóteses são cruciais em experimentos A/B, como testar se uma nova interface de site aumenta a taxa de conversão. A hipótese nula assume que não há diferença entre as versões A e B; se rejeitada, implementa-se a versão superior. Usando SciPy para um teste T de duas amostras independentes em dados de conversão (grupo controle vs. tratamento):

```python
import numpy as np
from scipy import stats

# Dados simulados: taxas de conversao (controle e tratamento)
np.random.seed(42)
controle = np.random.normal(0.10, 0.02, 500)  # Media 10%
tratamento = np.random.normal(0.12, 0.02, 500)  # Media 12%

# Teste T independente
t_stat, p_val = stats.ttest_ind(controle, tratamento)

print(f"Estatistica T: {t_stat:.2f}")
print(f"P-valor: {p_val:.4f}")

if p_val < 0.05:
    print("Rejeita H0: Ha diferenca significativa na conversao.")
else:
    print("Nao rejeita H0: Sem diferenca significativa.")
```

Listing 2: Teste T de duas amostras independentes

Saída esperada: T ≈ -7,5, p-valor ≈ 0,0000, rejeitando H_0.

10.3 Teste T de Student

A distribuição T de Student é usada para inferências sobre médias quando a variância populacional é desconhecida e a amostra é pequena, ajustando para maior incerteza com caudas mais pesadas que a normal.

Como exemplo em Ciência de Dados, aplica-se em análise de métricas de desempenho, como comparar o tempo médio de carregamento de páginas antes e após uma otimização, com amostras limitadas de logs de usuários.

```python
import numpy as np
from scipy import stats

# Dados simulados: tempos antes e apos otimizacao (10 amostras pareadas)
np.random.seed(42)
antes = np.random.normal(5, 1, 10)  # Media 5s
depois = antes - np.random.normal(0.5, 0.2, 10)  # Reducao media de 0.5s

# Teste T pareado
t_stat, p_val = stats.ttest_rel(antes, depois)

print(f"Estatistica T: {t_stat:.2f}")
print(f"P-valor: {p_val:.4f}")
```

Listing 3: Tempo de carregamento de dados para teste T de Student

Saída esperada: T ≈ 9,0, p-valor ≈ 0,0000, indicando redução significativa no tempo.

10.4 Distribuição Binomial

A distribuição binomial modela o número de sucessos em n tentativas independentes com probabilidade p constante.

Como exemplo de aplicação em Ciência de Dados, vemos abaixo um cenário onde é feita a modelagem de eventos binários como cliques em anúncios (i.e. prever a probabilidade de k conversões em 100 visualizações):

```python
from scipy import stats

# Parametros: n = 100 visualizacoes, p = 0.05 probabilidade de clique
n, p = 100, 0.05

# Probabilidade de exatamente 5 cliques
prob_exata = stats.binom.pmf(k=5, n=n, p=p)

# Probabilidade cumulativa de pelo menos 10 cliques
prob_cum = 1 - stats.binom.cdf(k=9, n=n, p=p)

print(f"P(exatamente 5 cliques): {prob_exata:.4f}")
print(f"P(pelo menos 10 cliques): {prob_cum:.4f}")
```

Listing 4: Eventos de cliques analisados por Distribuição Binomial

Saída esperada: T ≈ 9,0, p-valor ≈ 0,0000, indicando redução significativa no tempo.

10.5 Distriuição de Poisson

A distribuição de Poisson modela contagens de eventos raros em intervalos fixos, com média λ. Como exemplo de Aplicação em Ciência de Dados, podemos realizar a contagem de eventos, tais como acessos à um servidor por hora ou falhas em uma rede, ajudando em monitoramento e previsão de picos:

```python
from scipy import stats

# Parametro: lambda = 3 acessos por hora
lambda_ = 3

# Probabilidade de exatamente 2 acessos
prob_exata = stats.poisson.pmf(k=2, mu=lambda_)

# Probabilidade cumulativa de mais de 5 acessos
prob_cum = 1 - stats.poisson.cdf(k=5, mu=lambda_)

print(f"P(Exatamente 2 acessos): {prob_exata:.4f}")
print(f"P(Mais de 5 acessos): {prob_cum:.4f}")
```

Listing 5: Acesso a servidores usando Distribuição de Poisson

Saída esperada: $P(2) \approx 0,224$, $P(\geq 5) \approx 0,083$.

10.6 Qui-Quadrado

A distribuição qui-quadrado é usada em testes de aderência, independência e homogeneidade para dados categóricos. Como exemplo de aplicação, em Ciência de Dados, testa associações em dados de usuários, como se gênero influencia preferência por produtos em um e-commerce:

```python
import numpy as np
from scipy import stats

# Dados simulados: tabela 2x3 (genero x preferencia de produto)
tabela = np.array([[150, 100, 50],   # Masculino
                   [120, 130, 50]])  # Feminino

chi2_stat, p_val, dof, expected = stats.chi2_contingency(tabela)

print(f"Estatistica Qui-Quad: {chi2_stat:.2f}")
print(f"P-valor: {p_val:.4f}")
print(f"Graus de liberdade: {dof}")
```

Listing 6: Testes de aderência usando Qui-Quadrado

Saída esperada: $\chi^2 \approx 5,8$, P-valor $\approx 0,055$.

10.7 ANOVA

A ANOVA compara médias entre múltiplos grupos, decompondo variância total em entre e dentro de grupos. Como exemplo de aplicação, em Ciência de Dados, podemos comparar métricas de engajamento entre diferentes segmentos de usuários (ex.: regiões geográficas):

```python
import numpy as np
import statsmodels.api as sm
from statsmodels.formula.api import ols

# Dados simulados: engajamento por regiao (3 grupos)
regiao1 = np.random.normal(20, 5, 30)
regiao2 = np.random.normal(25, 5, 30)
regiao3 = np.random.normal(22, 5, 30)
data = np.concatenate([regiao1, regiao2, regiao3])
grupos = ['R1']*30 + ['R2']*30 + ['R3']*30

# ANOVA
model = ols('data ~ grupos', data=dict(data=data, grupos=grupos)).fit()
anova_table = sm.stats.anova_lm(model, typ=2)

print(anova_table)
```

Listing 7: Usando StatsModels para one-way ANOVA

Saída esperada: F-stat $\approx 10,2$, P-valor $\approx 0,0001$.

10.8 Métricas de Erros

Métricas de erros avaliam a precisão de modelos de regressão, quantificando discrepâncias entre previsto e real. Como exemplo de aplicação, em Ciência de Dados, podemos avaliar modelos de previsão de vendas, comparando RMSE vs. MAE para decidir o melhor:

```python
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error,
    r2_score

# Dados simulados: vendas vs. marketing spend
X = np.random.rand(100, 1) * 100  # Spend
y = 50 + 2 * X.squeeze() + np.random.normal(0, 10, 100)  # Vendas

# Modelo
model = LinearRegression().fit(X, y)
y_pred = model.predict(X)

# Metricas
mae = mean_absolute_error(y, y_pred)
rmse = np.sqrt(mean_squared_error(y, y_pred))
r2 = r2_score(y, y_pred)

print(f"MAE: {mae:.2f}")
print(f"RMSE: {rmse:.2f}")
print(f"R-Quad: {r2:.2f}")
```

Listing 8: Métricas em um modelo de regressão linear simples

Saída esperada: MAE $\approx 7,8$, RMSE $\approx 9,8$ e $R^2 \approx 0,96$.