



Estatística que Decide

Visão Geral do Curso

Estatística que Decide

Métrica Certa

Aprenda a identificar e trabalhar com métricas que realmente refletem o impacto do negócio

Incerteza Explícita

Domine intervalos de confiança, poder estatístico e gestão de erros

Economia

Traduza análises estatísticas em impacto financeiro mensurável:
 $\text{benefício} \times \text{custo} \times \text{risco}$

Agenda do Curso

01

Métricas & Incerteza

Fundamentos para entender e comunicar resultados com precisão estatística

02

A/B com Economia

Como transformar testes estatísticos em decisões financeiras inteligentes

03

Previsão & Cenários

Técnicas práticas para projetar o futuro e planejar diferentes possibilidades

04

Threshold & ROI

Encontre o ponto ótimo de decisão que maximiza retorno sobre investimento

05

Checklist + One-pager

Ferramentas práticas para estruturar e comunicar suas recomendações

Os Três Pilares da Decisão Estatística

Toda decisão baseada em dados precisa equilibrar estes três elementos fundamentais. Ignorar qualquer um deles pode levar a conclusões equivocadas ou investimentos mal direcionados.



Métrica Certa

Escolha métricas diretamente ligadas ao impacto no negócio. Uma métrica norte clara é a base de qualquer análise significativa. Sem ela, estamos apenas gerando números sem propósito.



Incerteza Explícita

Quantifique e comunique incertezas através de intervalos de confiança, cálculos de poder estatístico e análise rigorosa de erros tipo I e II. Decisões ignoram incerteza por sua conta e risco.



Economia

Traduza resultados estatísticos em termos financeiros: benefício esperado multiplicado pelo custo de implementação e ajustado pelo risco. É a linguagem que os negócios entendem.

📌 **Lembre-se:** Sem métrica norte, não há decisão possível. É o ponto de partida obrigatório.

Resumos Estatísticos que Convencem

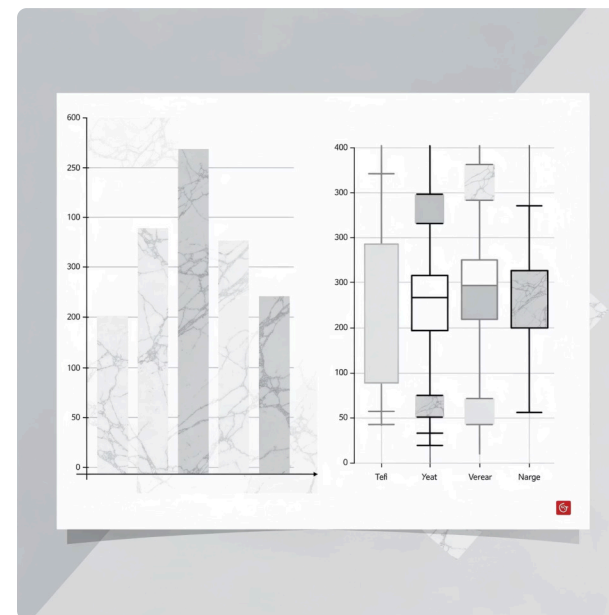
A arte de resumir dados está em escolher as estatísticas certas para cada situação. Não basta calcular – é preciso entender o que cada medida revela sobre seus dados e como isso influencia decisões.

Medidas de Tendência Central e Dispersão

- **Média:** útil quando dados são simétricos e sem outliers extremos
- **Mediana:** preferível quando há assimetria ou valores atípicos
- **IQR (intervalo interquartil):** mostra a dispersão dos 50% centrais dos dados
- **Identificação de outliers:** valores além de $1.5 \times \text{IQR}$ podem distorcer análises

Métricas de Negócio

- **Proporções:** taxas de conversão, churn, retenção
- **Segmentação estratégica:** SMB, Mid-Market, Enterprise têm comportamentos distintos



Visualizações Essenciais

Escolha o gráfico que melhor conta sua história:

- **Histograma:** distribuição de frequências
- **Boxplot:** mediana, quartis e outliers de forma compacta
- **Linhas com IC:** tendências ao longo do tempo com incerteza visível

Intervalos de Confiança: Decida na Faixa

Um dos erros mais comuns em análise de dados é tomar decisões baseadas em pontos únicos. A realidade é que toda estimativa carrega incerteza, e intervalos de confiança tornam essa incerteza explícita e acionável.

Intervalo de Confiança para Média

Quando você está estimando o valor médio de uma métrica contínua (receita média, tempo médio, etc.):

$$IC = \bar{x} \pm t(\alpha/2, n-1) \times (s/\sqrt{n})$$

Onde **\bar{x}** é a média amostral, **t** é o valor crítico da distribuição t , **s** é o desvio padrão amostral e **n** é o tamanho da amostra.

Intervalo de Confiança para Proporção

Quando você está trabalhando com taxas ou percentuais (conversão, churn, aprovação):

$$IC = \hat{p} \pm z(\alpha/2) \times \sqrt{(\hat{p}(1-\hat{p}))/n}$$

Onde **\hat{p}** é a proporção amostral, **z** é o valor crítico da normal padrão e **n** é o tamanho da amostra.

Mensagem Central: Sempre decida considerando a **faixa de valores plausíveis**, nunca apenas o ponto estimado. Um IC de 95% significa que, se repetíssemos o experimento muitas vezes, 95% dos intervalos conteriam o verdadeiro valor do parâmetro.

Testes A/B: Hipóteses e Poder Estatístico

Testes A/B são o padrão-ouro para validar mudanças em produtos digitais. Mas projetar um teste eficaz vai muito além de simplesmente dividir usuários em dois grupos. É preciso entender a lógica de hipóteses e garantir poder estatístico suficiente.

Framework de Hipóteses

- **H₀ (hipótese nula):** não há diferença entre controle e tratamento
- **H₁ (hipótese alternativa):** existe diferença significativa
- **Erro Tipo I (α):** rejeitar H₀ quando é verdadeira (falso positivo) – geralmente 5%
- **Erro Tipo II (β):** não rejeitar H₀ quando é falsa (falso negativo)
- **Poder (1-β):** probabilidade de detectar efeito quando ele existe – desejável ≥80%

Teste Z para Proporções

Usado para comparar taxas de conversão, cliques, etc.:

$$z = (p_1 - p_2) / \sqrt{(\bar{p}(1-\bar{p}))(1/n_1 + 1/n_2)}$$



Cálculo de Tamanho Amostral

Para garantir poder adequado, planeje o tamanho mínimo da amostra antes do teste:

$$n \approx 2(z(\alpha/2) + z(\beta))^2 \times \bar{p}(1-\bar{p}) / \delta^2$$

Onde δ é a diferença mínima detectável entre as proporções.

📄 No script fornecido, utilize a função `ab_z_test` para automatizar estes cálculos.

A/B Testing com Visão Econômica

Significância estatística não paga as contas. Um resultado "estatisticamente significativo" pode ser economicamente irrelevante. O verdadeiro valor de um teste A/B está em traduzir uplift estatístico em impacto financeiro real.



1. Uplift + Incerteza

Calcule o uplift percentual entre variantes, sempre acompanhado de IC e p-valor. Não pare na significância – entenda a magnitude do efeito.



2. Converter em R\$

Traduza o uplift em impacto financeiro: (Benefício - Custo). Inclua custos de implementação, manutenção e possíveis riscos associados.



3. Decisão Estruturada

Go: uplift significativo e ROI positivo robusto

No-Go: sem significância ou ROI negativo

Estender: tendência promissora mas incerteza alta – colete mais dados

Exemplo Prático

Imagine um teste onde a variante B aumenta conversão de 3.2% para 3.8% ($p < 0.01$, IC 95%: [3.5%, 4.1%]). Com 100.000 visitantes/mês e ticket médio de R\$ 250:

- Benefício mensal = 600 conversões extras \times R\$ 250 = **R\$ 150.000**
- Custo de implementação = R\$ 20.000 (dev + QA)
- ROI primeiro mês = **650%**

Decisão: Go imediato – mesmo no cenário conservador (limite inferior do IC), o retorno é fortemente positivo.



☐ **Princípio-chave:** Significância estatística sem impacto financeiro não justifica ação. Sempre feche o loop entre análise e economia.

Previsão Simples: Fundamentos Práticos

Previsões não precisam ser complexas para serem úteis. Métodos simples e interpretáveis frequentemente superam modelos sofisticados, especialmente quando você precisa explicar suas premissas para stakeholders não-técnicos.

Média Móvel

Suaviza flutuações de curto prazo para revelar tendências:

$$MA(t) = (y(t) + y(t-1) + \dots + y(t-k+1)) / k$$

Útil para identificar padrões, mas não gera previsões diretas. Ideal para visualização inicial de séries temporais.

Suavização Exponencial Simples (SES)

Pondera observações recentes mais fortemente, adaptando-se gradualmente a mudanças:

$$\hat{y}(t+1) = \alpha y(t) + (1-\alpha) \times \hat{y}(t)$$

Onde α (entre 0 e 1) controla a velocidade de adaptação. Valores maiores dão mais peso ao passado recente. Parte da família ETS (Error, Trend, Seasonality).

Métrica de Avaliação: MAPE

Mean Absolute Percentage Error mede erro em termos percentuais:

$$MAPE = (1/n) \times \sum |y - \hat{y}| / y \times 100\%$$

Interpretação direta: MAPE de 8% significa erro médio de 8% nas previsões. Cuidado com valores próximos de zero no denominador.

Cenário Base

Projeção usando parâmetros estimados direto dos dados históricos, sem ajustes. Representa a expectativa mais provável.

Cenário Otimista

Projeção considerando limite superior do IC para parâmetros ou assumindo melhoria nas tendências observadas. Útil para planejamento de capacidade.

Cenário Conservador

Projeção usando limite inferior do IC ou assumindo piora nas condições. Essencial para análise de risco e planejamento financeiro seguro.

Demonstração: SES em Ação

Nesta seção da aula, vamos executar o script ao vivo para ver a Suavização Exponencial Simples (SES) funcionando com dados reais. O gráfico abaixo será gerado durante a apresentação.

| Dados Históricos | Previsão 1-Step | Comparação Visual |
|--|--|---|
| Série temporal de vendas semanais ou outra métrica de negócio relevante carregada de <code>weekly_sales.csv</code> | Modelo SES treinado prevê o próximo período usando apenas informação até o ponto atual | Linha observada vs. linha prevista permite avaliar qualidade do ajuste e identificar pontos de maior erro |

Durante a execução, preste atenção em:

- **Aderência do modelo:** quão próximas as linhas ficam ao longo do tempo
- **Lag na resposta:** SES demora alguns períodos para "alcançar" mudanças bruscas
- **MAPE resultante:** qual o erro percentual médio e se é aceitável para o caso de uso
- **Padrões não capturados:** sazonalidade ou tendências fortes podem exigir métodos mais sofisticados

"O melhor modelo é aquele que você consegue explicar e que entrega valor no prazo necessário."

Threshold Ótimo e ROI

Em problemas de classificação (ex: aprovar crédito, identificar churn, priorizar leads), a escolha do threshold não é arbitrária. O ponto de corte ideal depende diretamente dos custos e benefícios de cada decisão.

Scores ≠ Decisão

Um modelo de machine learning retorna probabilidades ou scores, mas quem decide o threshold de ação é o negócio, não o modelo. Um threshold de 0.5 é apenas conveniente, raramente ótimo.

A Equação do Lucro

Para cada threshold t , calcule:

$$\text{Lucro}(t) = \text{TP}(t) \times B - N_{\text{ações}}(t) \times C$$

Onde:

- **TP(t):** verdadeiros positivos no threshold t
- **B:** benefício por acerto
- **N_ações(t):** total de ações tomadas
- **C:** custo por ação

Curva Lucro × Threshold

Plote o lucro esperado para cada valor de threshold entre 0 e 1. O pico dessa curva indica o threshold que maximiza retorno.

Exemplo: em um modelo de detecção de fraude:

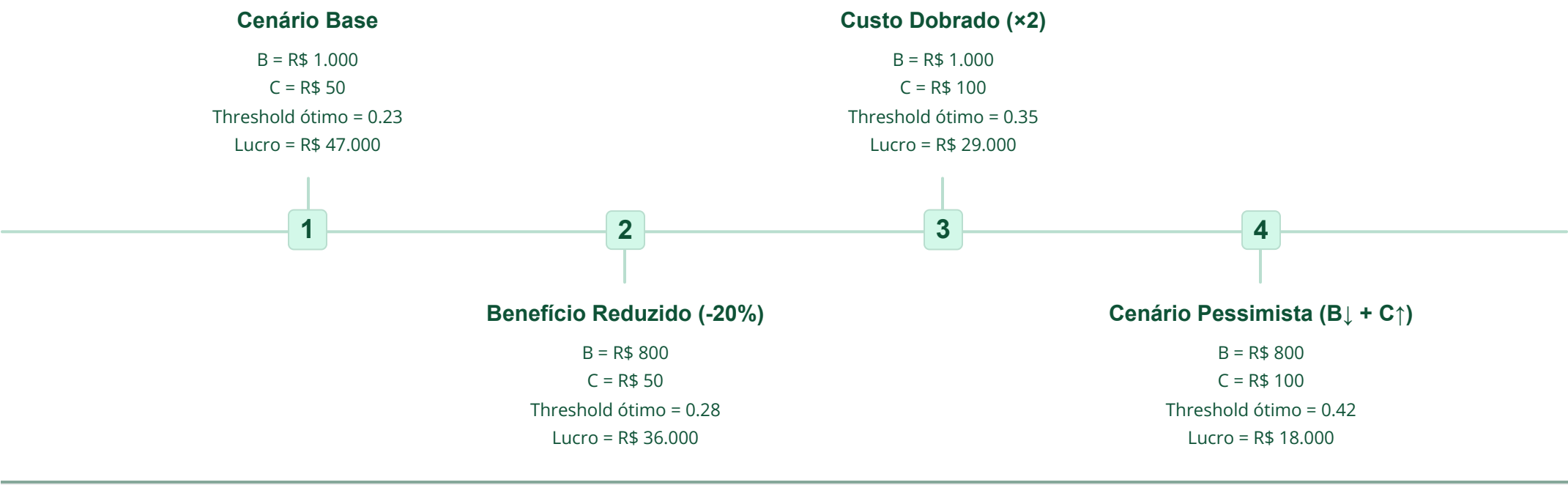
- $B = \text{R\$ } 1.000$ (perda evitada)
- $C = \text{R\$ } 50$ (custo de investigação)
- Threshold ótimo ≈ 0.23
- Lucro máximo = $\text{R\$ } 47.000/\text{mês}$

Se usássemos threshold padrão de 0.5, lucro cairia para $\text{R\$ } 31.000/\text{mês}$.

📄 No script, as funções `roi_by_threshold` e `plot_profit_curve` automatizam este cálculo. Sempre visualize a curva – o formato dela revela sensibilidade e robustez da decisão.

Análise de Sensibilidade

Suas premissas sobre benefício e custo são apenas estimativas. Análise de sensibilidade responde: "E se eu estiver errado? A decisão ainda se sustenta?" É a diferença entre recomendações frágeis e robustas.



Perguntas Críticas para Responder

- O threshold ótimo muda drasticamente com pequenas variações em B ou C?
- O lucro permanece positivo mesmo no pior cenário?
- Existe um range de thresholds "seguros" que funcionam bem em múltiplos cenários?
- A decisão Go/No-Go é robusta ou depende fortemente das premissas?
- Quais parâmetros têm maior impacto e merecem refinamento adicional?
- Vale a pena coletar mais dados para reduzir incerteza antes de agir?

Decisões robustas funcionam em múltiplos cenários. Se sua recomendação depende de premissas muito específicas, comunique isso claramente e proponha mitigações.

One-Pager: Estrutura da Decisão

Um one-pager eficaz é o cartão de visita da sua análise. Ele precisa ser autocontido, visual e orientado à ação. Stakeholders ocupados devem entender o problema, a evidência e a recomendação em 2 minutos.

| | | |
|--|---|--|
| 01 | 02 | 03 |
| Problema & Métrica Norte | Evidência Visual | Impacto Econômico |
| Declare explicitamente o problema de negócio e qual métrica única vai medir sucesso. Exemplo: "Reduzir churn de clientes Enterprise. Métrica norte: taxa de churn mensal." | Um gráfico central mostrando o resultado-chave, sempre acompanhado de IC, p-valor ou erro de previsão. A visualização deve contar a história sozinha. | Traduza em R\$ e apresente cenários (base, otimista, conservador). Exemplo: "Impacto esperado: R\$ 150k/mês (base), range: R\$ 120k-180k." |
| 04 | 05 | |
| Decisão & Métrica de Guarda | Riscos & Próximos Passos | |
| Recomendação clara (Go/No-Go/Estender) e qual métrica você vai monitorar para validar resultados pós-implementação. Exemplo: "Go. Monitorar churn semanal por 8 semanas." | Seja transparente sobre limitações, premissas críticas e plano de ação detalhado com prazos. Exemplo: "Risco: modelo assume estabilidade sazonal. Próximo passo: validação A/B em 2 semanas." | |

Checklist da Decisão Baseada em Dados

Antes de apresentar sua recomendação, passe por este checklist. Cada item não marcado é um ponto de questionamento potencial. Use isso como ferramenta de revisão final.

1

Métrica Norte

- ✓ Métrica alinhada com objetivo de negócio?
- ✓ Mensurável e disponível em frequência adequada?
- ✓ Sensível o suficiente para detectar mudanças relevantes?

2

Incerteza Quantificada

- ✓ Intervalos de confiança calculados e reportados?
- ✓ Poder estatístico adequado ($\geq 80\%$) em testes?
- ✓ Erros tipo I e II considerados na decisão?

3

Economia & Sensibilidade

- ✓ Resultado traduzido em impacto financeiro (R\$)?
- ✓ Custos de implementação incluídos?
- ✓ Análise de sensibilidade para premissas críticas?
- ✓ ROI positivo em cenários conservadores?

4

Recomendação & Passos

- ✓ Decisão clara (Go/No-Go/Estender)?
- ✓ Próximos passos com prazos definidos?
- ✓ Métrica de guarda para monitoramento pós-decisão?
- ✓ Critérios de sucesso/falha explícitos?

5

Premissas & Riscos

- ✓ Premissas críticas documentadas?
- ✓ Limitações dos dados ou modelo transparentes?
- ✓ Riscos identificados com planos de mitigação?
- ✓ Contexto de negócio incorporado na análise?

Recursos e Scripts Python

Todo o conteúdo deste curso é acompanhado de código Python prático e datasets reais. Os scripts foram projetados para serem reutilizáveis e adaptáveis aos seus próprios projetos.

Script Principal

`estatistica_que_decide.py`

Biblioteca unificada contendo todas as funções apresentadas no curso:

- Cálculo de intervalos de confiança
- Testes A/B (z-test, cálculo de poder)
- Previsão com SES e média móvel
- Otimização de threshold e curvas de ROI
- Funções auxiliares de visualização

Código documentado com docstrings e exemplos de uso em cada função.

Datasets

- `ab_counts.csv` – resultados de experimentos A/B
- `weekly_sales.csv` – série temporal de vendas
- `scores_labels.csv` – scores de modelo e labels reais

Todos os datasets estão em formato CSV limpo e pronto para análise, simulando situações reais de negócio.



Abordagem Didática

Durante a aula, executaremos **um gráfico por bloco conceitual**, permitindo que você veja os conceitos ganhando vida em tempo real.



Reprodutibilidade

Todos os scripts usam seeds fixos para garantir que você obtenha os mesmos resultados ao replicar os exemplos.



Extensibilidade

Código modular facilita adaptação para seus próprios casos de uso. Basta substituir os dados de entrada mantendo o formato esperado.