

e-book



Pós-Graduação



em



Business _____



Intelligence _____



& Analytics. _____

Como alcançar
resultados de
forma inteligente.



[09]

jornada do curso

MÓDULO 01_O PRINCÍPIO DE TUDO

MÓDULO 02_DADOS, INFORMAÇÕES E INTELIGÊNCIA

MÓDULO 03_OS DADOS CONTAM HISTÓRIAS



AULA 07_CONECTAR DADOS A RESULTADOS: OLHAR PARA TRÁS

**AULA 08_CONECTAR DADOS A RESULTADOS: OLHAR
PARA FRENTE**

**AULA 09_ DADOS COMO HIPÓTESES NA NOVA ECONOMIA:
ENTRE PREVISÕES E RESULTADOS DE SUCESSO**

AULA 10_CASE #01

AULA 11_CHECKPOINT

MÓDULO 04_APRESENTAR OS DADOS E AS DECISÕES





o novo vai ou não



MÓDULO 03:
OS DADOS CONTAM HISTÓRIAS

**DADOS COMO HIPÓTESES NA NOVA
ECONOMIA: ENTRE PREVISÕES
E RESULTADOS DE SUCESSO**



PÓS-GRADUAÇÃO EM
BUSINESS INTELLIGENCE & ANALYTICS



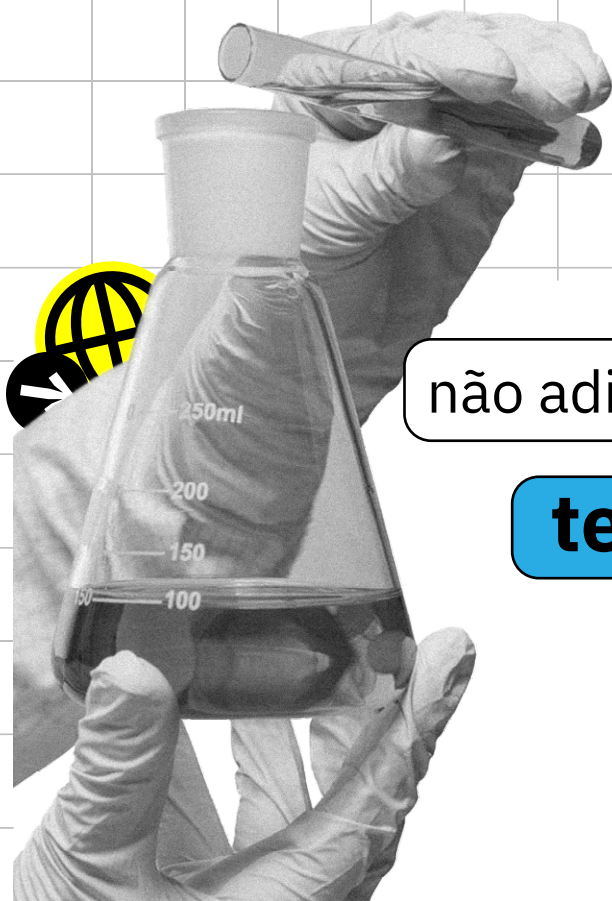
sumário



Clique, selecione ou escreva! Este material possui recursos interativos para enriquecer sua experiência.



[09]



não adivinhe nada

teste!



M

missão da aula

09

Transformar suposições em perguntas e, a partir delas, definir **hipóteses** que podem ser confirmadas por meio de **testes A/B**.

I

S

S

Ã



O



[09]

interativo! 

análises preditivas

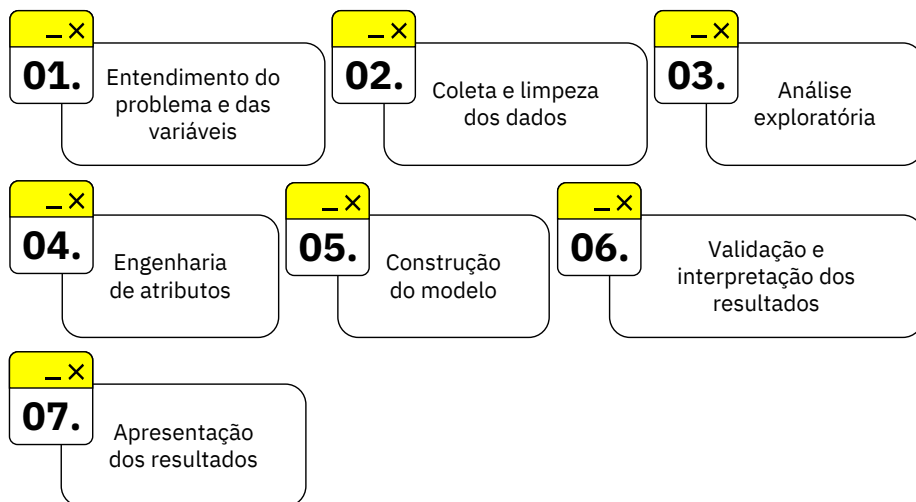


análises preditivas

breve revisão

- > É uma simplificação da realidade, baseada em probabilidade.
- > Identifica tendências com base no passado para projetar situações futuras.
- > Detecta riscos e oportunidades com antecedência para uma tomada de decisão mais assertiva e consciente.

passo a passo





1.01.01.01.01.01.0

entendimento do problema e das variáveis

- > O que você quer saber sobre o **futuro** com base no passado?
- > O que você quer entender e **prever**?
- > Quais são **as variáveis** disponíveis para análise?

2.02.02.02.02.02.0

coleta e limpeza dos dados

- > Localizar as informações corretas: podem ser acessadas de qualquer fonte.
- > Remover dados estranhos e *outliers*.
- > Organizar e ajustar os dados para que ganhem significado e possam ser utilizados na tomada de decisões.



análise exploratória

- **Visualização de dados** – histograma, dispersão etc.

engenharia de atributos

- Exemplos de transformações: **razões, binning, scaling, normalization e one-hot encoding.**



4.04.04.04.04.04.0

engenharia de atributos

One-hot encoding:

> Converte os valores categóricos em valores numéricos.

Animais		Gato	Cachorro	Peixe
Gato		1	0	0
Cachorro		0	1	0
Peixe		0	0	1
Tartaruga		0	0	0
Gato		1	0	0
Cachorro		0	1	0

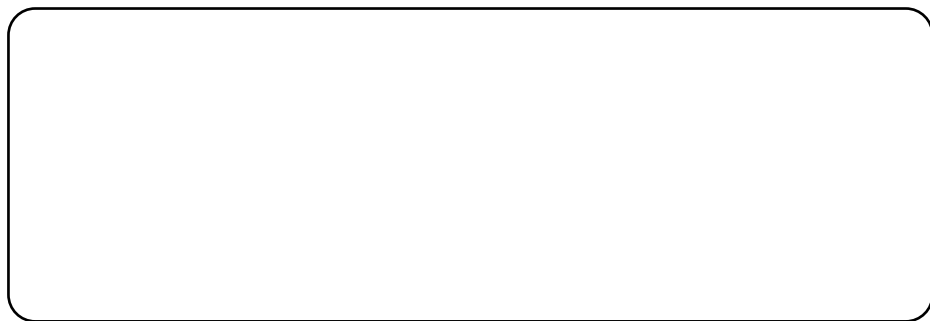
Sempre que tivermos **N** categorias, criaremos **N-1** colunas.



mão na massa

Empresa automobilística

Na tabela de carros da última aula (**carros.xlsx**), havia várias colunas categóricas.



Aplicaremos a técnica One Hot Encoding na coluna **carbody** (tipo de carro) para ver o funcionamento na prática:

_passe o mouse sobre a tabela					— X	
carbody	→	convertible	hatchback	sedan	wagon	
convertible		1	0	0	0	
hatchback		0	1	0	0	
sedan		0	0	1	0	
wagon		0	0	0	1	
hardtop		0	0	0	0	



5.05.05.05.05.0

construção do modelo

> O **modelo** escolhido vai depender do **tipo de problema** tratado.

> **Predição:**

- > Árvores de decisão;
- > Regressão linear;
- > Redes neurais;
- > DBSCAN.

> **Classificação:**

- > Naive Bayes;
- > Kneighbors;
- > Random Forest;
- > Regressão Logística.

6.06.06.06.06.0

validação e interpretação dos resultados

> Existem várias métricas para avaliar o desempenho preditivo de um algoritmo de Machine Learning.

> Cada algoritmo deve ser avaliado com a **métrica apropriada** para ele.

- > R^2 .
- > Matriz de confusão.
- > AUC.



R²

- > É também conhecido como **coeficiente de determinação** ou **coeficiente de determinação múltipla**. Ele diz o quão próximo os dados estão da linha de regressão.
- > É a porcentagem da variação da variável resposta que é explicada pelo modelo:
 - > 0% – indica que o modelo não explica nada.
 - > 100% – indica que o modelo explica toda a variabilidade dos dados-resposta.
 - > Quando maior o R², melhor o modelo se ajusta aos seus dados.

matriz de confusão

- > É uma tabela que mostra as frequências de classificação para cada classe do modelo.

- X

Valor previsto \ Valor real	Positivo	Negativo
Positivo	Verdadeiro Positivo (TP)	Falso Positivo (FP)
Negativo	Falso Negativo (FN)	Verdadeiro Negativo (TN)

- > A partir da matriz, podemos calcular alguns conceitos importantes, como: **acurácia, recall, precisão e f-score**.



AUC

- > Área under the ROC curve.
- > É uma métrica invariante em escala, pois trabalha com precisão das classificações ao invés de seus valores absolutos.
- > Um modelo cujas previsões estão 100% erradas tem uma AUC de 0, enquanto um modelo cujas previsões são 100% corretas tem uma AUC de 1.

7.07.07.07.07.07.0

apresentação dos resultados

- Deve ser totalmente baseada em dados, além de exibir os dados mais relevantes.
- É importante conhecer a sua audiência.



mão na massa

Empresa automobilística

Ainda no mesmo arquivo (**carros.xlsx**), **aplique** a engenharia de atributos nas colunas que julgar necessário.

Após esse ajuste, podemos incluir novas variáveis no modelo para avaliar.

A qualidade desse novo modelo é melhor ou pior do que os outros que fizemos?



A **análise preditiva** nos permite identificar tendências com base no passado, para **projetar** situações futuras e **tomar uma decisão**.

– x

Mas e quando você já tem opções e precisa decidir entre elas?

– x

Como **comparar** soluções para descobrir qual delas promove uma **melhor performance**?

– x

Como decidir qual é a melhor?



[09]

interativo! 

teste A/B



teste A/B

- > É uma técnica usada para comparar duas versões diferentes de algo e determinar qual delas tem o melhor desempenho.
- > Ajuda a conhecer o conteúdo que atrai mais visitantes e melhorar a experiência do usuário – design, preferências de cores etc.

Exemplo Frontlineshop



A **Frontlineshop**, loja de moda online, testou mudar a cor de fundo de um banner de seu *e-commerce*:

- > Na versão A, o fundo era preto.
- > Na versão B, o fundo era branco.

Resultado

- > A versão A **gerou 30% a mais de cliques**, comparada com a versão B.
- > Algumas possibilidades para esse resultado são: o contraste com a cor preta ao fundo, pois favoreceu o produto, e a possível conexão entre a cor e a sofisticação, o que pode ter atraído mais clientes.



teste A/B

Exemplo Shell Box



A empresa tinha como objetivo central conhecer os seus clientes com mais profundidade para aumentar as comunicação por meio de segmentações. Foi identificada uma necessidade de melhoria na taxa de abertura de *e-mails*. Assim, um teste A/B de **assunto do e-mail** foi preparado com o objetivo de melhorar essa métrica. Um ponto importante é que apenas o assunto do *e-mail* mudou. O restante manteve a mesma estrutura.

Resultado

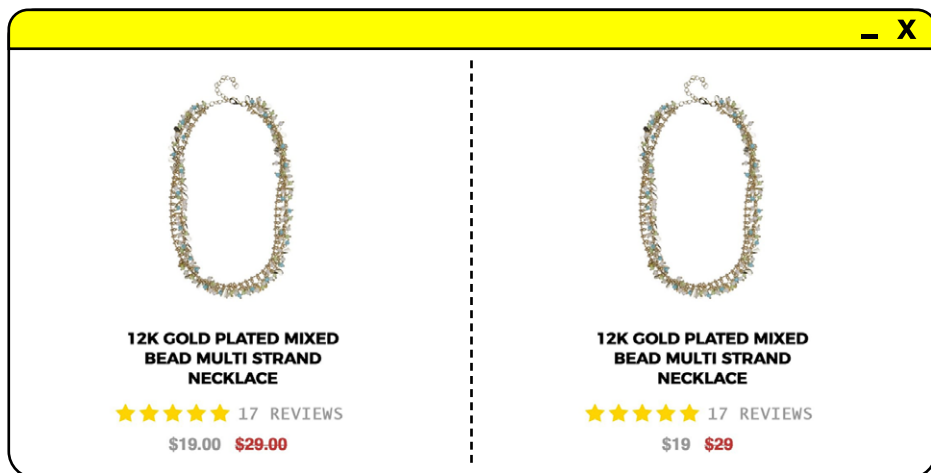
Em 15 dias de implementação do teste, houve um aumento de **192%** na taxa de abertura, representando um aumento de **20x** de respostas nas pesquisas com os clientes.



[09]

teste A/B

Exemplo de loja de joias



Uma loja virtual testou retirar os centavos dos preços. Eles acreditavam que, removendo essa informação, os clientes teriam a percepção de que os preços eram mais baixos.

Resultado

A versão sem os centavos teve um crescimento de **10%** de adições de produtos em carrinhos e uma receita por visitante **47%** maior.



teste A/B

Vantagens

- > Reduz a taxa de rejeição;
- > Melhora a experiência do usuário;
- > Aumenta a interação com a plataforma;
- > Promove melhoria contínua.

O que testar?

- > Layout;
- > Formulários;
- > Chatbot;
- > Call To Action (CTA);
- > Mudança de estilo;
- > Criar novas rotinas.

Erros comuns

- > **Usar multivariáveis em testes**, ou seja, testar várias coisas em um mesmo teste. Ao fazer isso, é difícil chegar a uma conclusão sobre qual das variáveis é a responsável pelo resultado.
- > **Não fazer testes simultâneos**. Pode existir alguma sazonalidade e os resultados de algum grupo podem ficar inflados, não condizendo com os hábitos normais das pessoas usuárias.



teste A/B

Erros comuns

- > **Comparar resultados sem usar estatística.** Não devemos comparar apenas o resultado da métrica de sucesso sem avaliar a significância estatística do resultado. Para isso, utilizamos o teste de hipóteses.

Funciona?

- > Essa é a hora em que o **neuromarketing** entra em ação;
- > Todas essas pequenas mudanças podem ser **estudadas e explicadas**;
- > O teste A/B faz a **parte matemático-estatística** para ver se há sentido em **mudar de A para B**.



não adivinhe nada

teste!



[09]

interativo! 

teste A/B & teste de hipóteses



teste A/B

Como fazer?

▼ Formule uma hipótese

- > Identifique o problema ou a métrica que você gostaria de melhorar.
- > Exemplos: taxa de cliques, taxa de abandono, número de vendas, disposição do layout.

▼ Crie um desafiante

- > **Controle:** o controle é a versão atual que você quer testar.
- > **Desafiante:** o desafiante é uma versão diferente do controle com todas as alterações que você deseja fazer.

O desafiante vai ser testado **contra** o controle para determinar se as mudanças impactam o engajamento ou as conversões.

▼ Divida seu público

- > É necessário definir qual o público que irá receber a versão controle e a versão desafiante.
- > A mesma pessoa usuária não deve receber diferentes versões.

▼ Analise os dados

- > **Teste de hipóteses:** serve para avaliar a significância estatística do resultado.



teste de hipóteses

Ao tentarmos tomar decisões, formulamos suposições. Essas suposições, que podem, ou não, ser verdadeiras, são denominadas como **hipóteses estatísticas**.

Logo, o **teste de hipóteses** é uma **técnica estatística** que nos permite rejeitar – ou não – uma hipótese estatística por meio da evidência fornecida pela amostra.

Imagine que somos donos de uma loja virtual e queremos testar um *site* com cores diferentes:

Qual *site* tem uma melhor taxa de conversão?





teste de hipóteses

Criamos as hipóteses que serão testadas:

H_0 : não há diferença significativa nas vendas entre o *site* A e *site* B.

H_1 : há uma diferença significativa nas vendas entre os *sites* A e B.

- > Qual hipótese **prevalece**? Precisamos analisar o **p-valor** e o α .

p-valor

- > Probabilidade de o experimento ser o perdedor. Quanto menor, melhor.
- > É calculado pelo teste-T.

nível de significância – α

- > É a probabilidade de **não** observar os mesmos resultados, caso o mesmo teste seja feito repetidas vezes.
- > O α padrão é de 5% ou 0,05.

Após o teste-T, verificamos que:

se $p\text{-valor} < \alpha$

- > Rejeitamos H_0 .
- > Observamos uma diferença significativa entre as duas versões.

se $p\text{-valor} \geq \alpha$:

- > Aceitamos H_0 .
- > Vemos que não há diferença significativa entre as duas versões.



mão na massa

Imagine que somos donos de uma loja virtual e queremos testar um *site* com cores diferentes:

- > Qual *site* tem uma melhor taxa de conversão?
- > Para isso, acesse o arquivo **loja_virtual.xlsx**, disponível para *download* no botão abaixo.



Passo a passo para o teste-T no Excel:

- > Habilitar o suplemento: ferramentas de análise;
- > Selecionar a aba: DADOS;
- > Selecionar a opção: análise de dados.
- > Ferramenta de análise:
Teste-T: duas amostras com variâncias iguais.
- > Preencher com as informações da tabela.

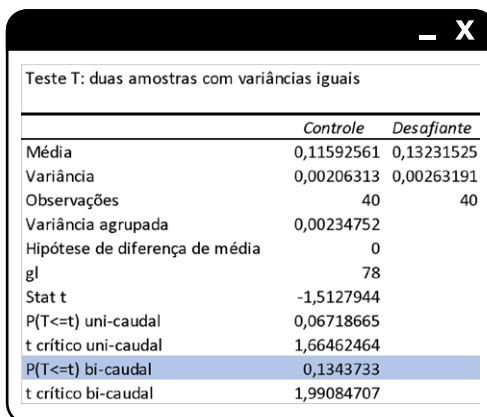
p-valor = 0,13
 $\alpha = 0,05$

Como **p-valor $\geq \alpha$:**

- > Aceitamos H_0 .
- > Observamos que não há diferença significativa.

Portanto:

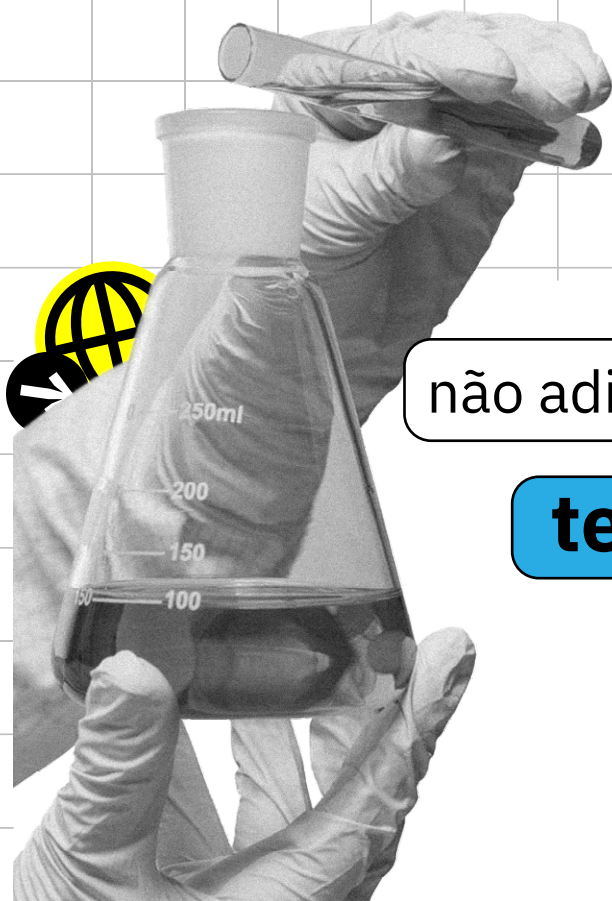
Não há diferença entre as páginas.



Teste T: duas amostras com variâncias iguais		
	Controle	Desafiante
Média	0,11592561	0,13231525
Variância	0,00206313	0,00263191
Observações	40	40
Variância agrupada	0,00234752	
Hipótese de diferença de média	0	
gl	78	
Stat t	-1,5127944	
P(T<=t) uni-caudal	0,06718665	
t crítico uni-caudal	1,66462464	
P(T<=t) bi-caudal	0,1343733	
t crítico bi-caudal	1,99084707	



[09]



não adivinhe nada

teste!



mão na massa

teste A/B

A loja Alfa mudou a identidade visual de seu *e-commerce* e fez um teste de vendas por 23 dias. Agora, precisamos **avaliar a performance** e decidir qual será a versão oficial: manter a original ou adotar a nova proposta?

- Para isso, acesse o arquivo **loja_Alfa.xlsx**, disponível para *download* no botão abaixo.



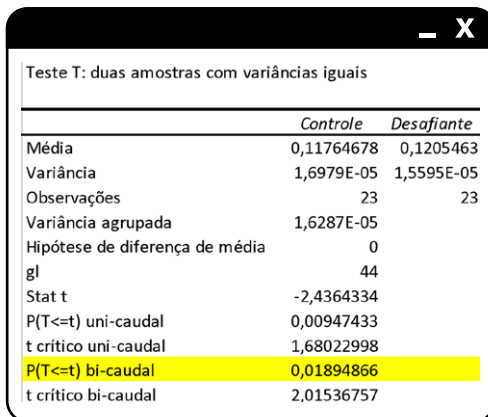
p-valor = 0,0189
 $\alpha = 0,05$

Como **p-valor < α** :

- Rejeitamos H_0 .
- Observamos que há diferença significativa.

Portanto:

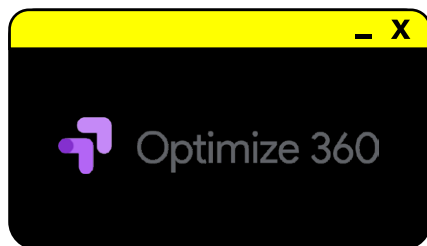
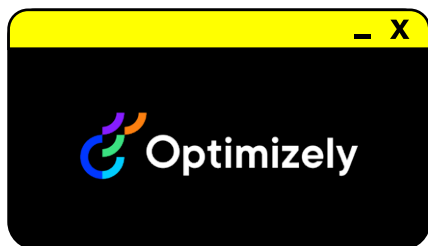
A página **desafiante** deve trazer melhores resultados para a loja Gama.



	Controle	Desafiante
Média	0,11764678	0,1205463
Variância	1,6979E-05	1,5595E-05
Observações	23	23
Variância agrupada	1,6287E-05	
Hipótese de diferença de média	0	
gl	44	
Stat t	-2,4364334	
P(T<=t) uni-caudal	0,00947433	
t crítico uni-caudal	1,68022998	
P(T<=t) bi-caudal	0,01894866	
t crítico bi-caudal	2,01536757	



ferramentas de teste A/B





[09]

interativo! 



**CAIXA DE
DÚVIDAS**





desafio conquer #09

(1/2) Somos analistas na empresa **Go First** e precisamos **estimar o preço de passagens aéreas**. Para isso, podemos utilizar preços internos, mas também tivemos acesso aos preços praticados pela concorrência.

- > Precisamos, então, construir um modelo para calcular o melhor preço para os voos no arquivo: **desafio_Conquer_09.xlsx**.
- > **Qual seria o melhor preço para as passagens indicadas no arquivo?**

(2/2) Depois de definirmos o preço, precisamos **avaliar qual é a melhor forma de vender as passagens**. Para isso, a equipe de *marketing* implementou 2 *sites* diferentes e testou por 20 dias a venda de passagens.

- > **Existe diferença nas vendas entre os *sites*?**
- > **Essa diferença, caso exista, é estatisticamente significativa?**





[09]

interativo! 

Conquer **notes**





livros

NEUROMARKETING

Darren Bridger

ESTATÍSTICA APLICADA A ADMINISTRAÇÃO, CONTABILIDADE E ECONOMIA COM EXCEL E SPSS

Patrícia Belfiore

vídeos

THE METHOD THAT CAN “PROVE” ALMOST ANYTHING

James A. Smith

+
+
+
quero