

TRATAMENTO ESTATÍSTICO PARA A PESQUISA ACADÊMICA

Mariana Curi

mcuri@icmc.usp.br

<http://icmc.usp.br/pessoas/mcuri>



[mariana-curi-55a45991](https://www.linkedin.com/in/mariana-curi-55a45991)



[Mariana Curi](https://www.youtube.com/channel/UC...)

TRÊS SITUAÇÕES PRÁTICAS

1. Teste AB
2. Comparação de Classificadores
3. Teste Adaptativo Informatizado (IA)

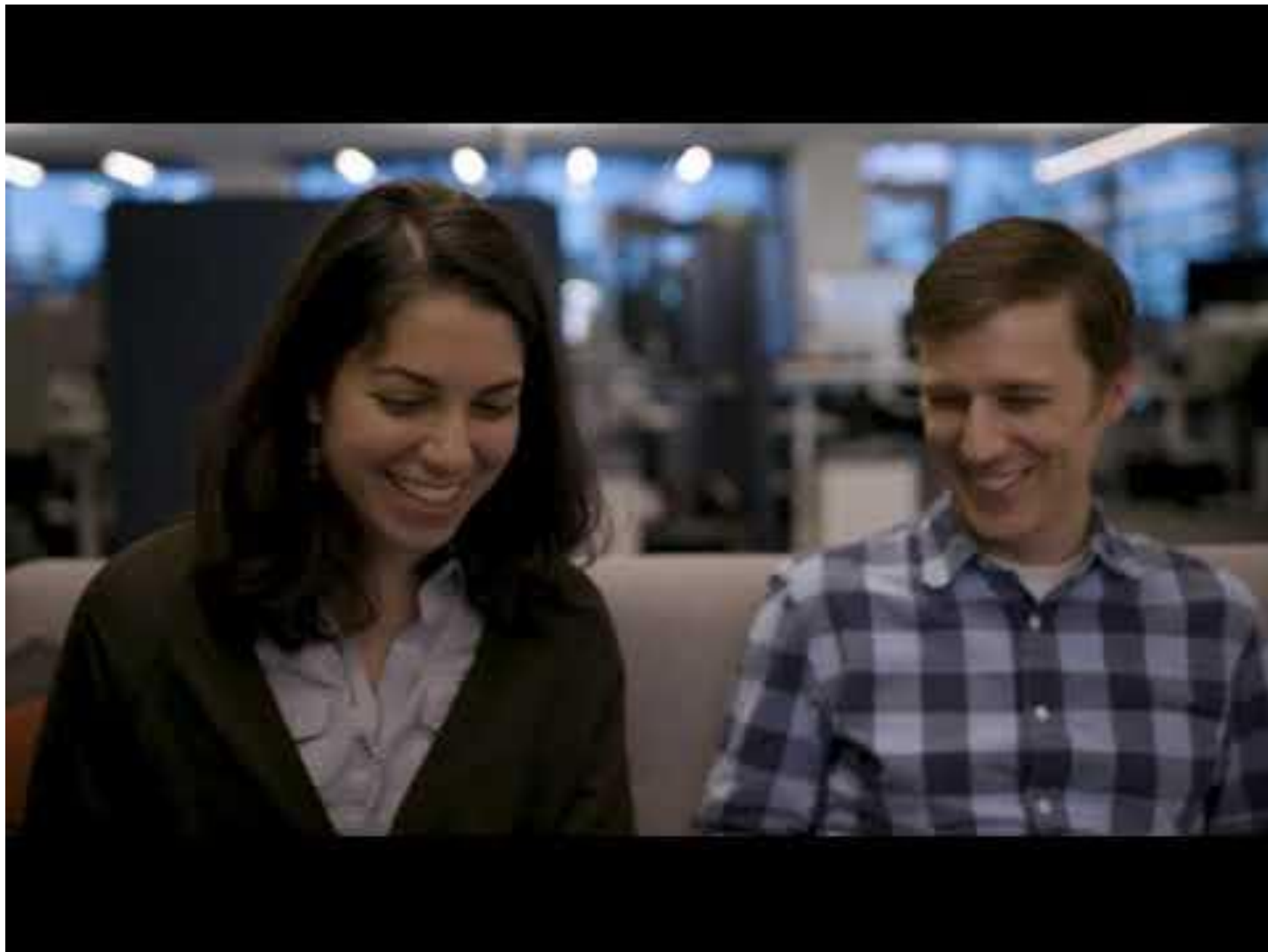
1. TESTE AB

A descrição da Wikipedia sobre o teste A/B é a seguinte:

(Fonte: https://pt.wikipedia.org/wiki/Teste_A/B)

"Teste A/B é um método de teste de **design** através do qual comparam-se elementos aleatórios com duas variantes, A e B, em que estes são o controle e o tratamento de uma experiência controlada, com o objetivo de melhorar a percentagem de aprovação. Estas experiências são muito utilizadas em desenvolvimento web e de marketing, e até mesmo em formas tradicionais de publicidade. Teste A/B também se designa por experiência aleatória controlada, experiência online controlada e teste de divisão. Em web design, o teste A/B é utilizado para identificar alterações nas páginas web que podem provocar mudanças positivas ou negativas no interesse dos utilizadores. Como o nome já diz, duas versões são comparadas, as quais são idênticas exceto por uma variante que pode impactar o comportamento do utilizador. A versão A pode ser a versão utilizada atualmente (controle), enquanto a Versão B é a modificada (tratamento)."

Tempo 0:55



Criar e gerenciar
contasPublicar e distribuir
conteúdo

Anuncie

Vender no
Facebook e no...Monetizar seu
conteúdo ou...

Criar um anúncio



Central de Ajuda para Empresas

cursosextensao.usp.br

Sobre os testes A/B

Tópicos comuns do suporte

Tipos de testes A/B disponíveis no Facebook

18.885 visualizações

Como selecionar uma variável para seu teste A/B

4.330 visualizações

Como editar ou cancelar um teste no recurso Experimentos

988 visualizações

O teste A/B permite alterar **variáveis**, como criativo do anúncio, público ou posicionamento, para determinar qual estratégia tem melhor desempenho e melhora campanhas futuras. Por exemplo, é possível **supor** que uma estratégia de público personalizado tenha melhor desempenho do que uma estratégia de público baseado em interesse para a sua empresa. Um teste A/B permite comparar rapidamente as duas estratégias para ver qual apresenta melhor desempenho.

Depois de escolher a variável que você deseja testar, dividiremos seu orçamento para comparar de forma uniforme e randomizada a exposição de cada versão do seu criativo, público ou posicionamento. Então, o teste A/B pode mensurar o desempenho de cada estratégia com base no **custo por resultado** ou no **custo por aumento de conversão com um grupo controle**.

Recomendamos o teste A/B quando você estiver tentando mensurar alterações à sua publicidade ou para comparar rapidamente duas estratégias. Você deve usar o teste A/B para aprender novas estratégias em vez de realizar testes informais, como ativar e desativar conjunto de anúncios ou campanhas manualmente, já que isso pode gerar resultados não confiáveis e veiculação ineficiente. O teste A/B ajuda a garantir que seus públicos sejam divididos de maneira uniforme e sejam estatisticamente comparáveis, enquanto testes informais podem acarretar a sobreposição de públicos.



O Google tem o compromisso de promover a igualdade racial para as comunidades negras. [Saiba como.](#)

Firestore > Documentos > Engajamento



Teste A/B do Firebase



Envie comentários

Com a tecnologia do [Google Optimize](#), o Teste A/B do Firebase ajuda você a otimizar a experiência do app, facilitando a execução, a análise e o dimensionamento de experimentos de produtos e marketing. Além disso, é possível testar campanhas de engajamento ou alterações nos recursos da IU do app para ver se elas estão funcionando conforme o esperado em relação às suas métricas principais (como receita e retenção) antes que elas sejam implementadas completamente.



 **Firestore A/B Testing**

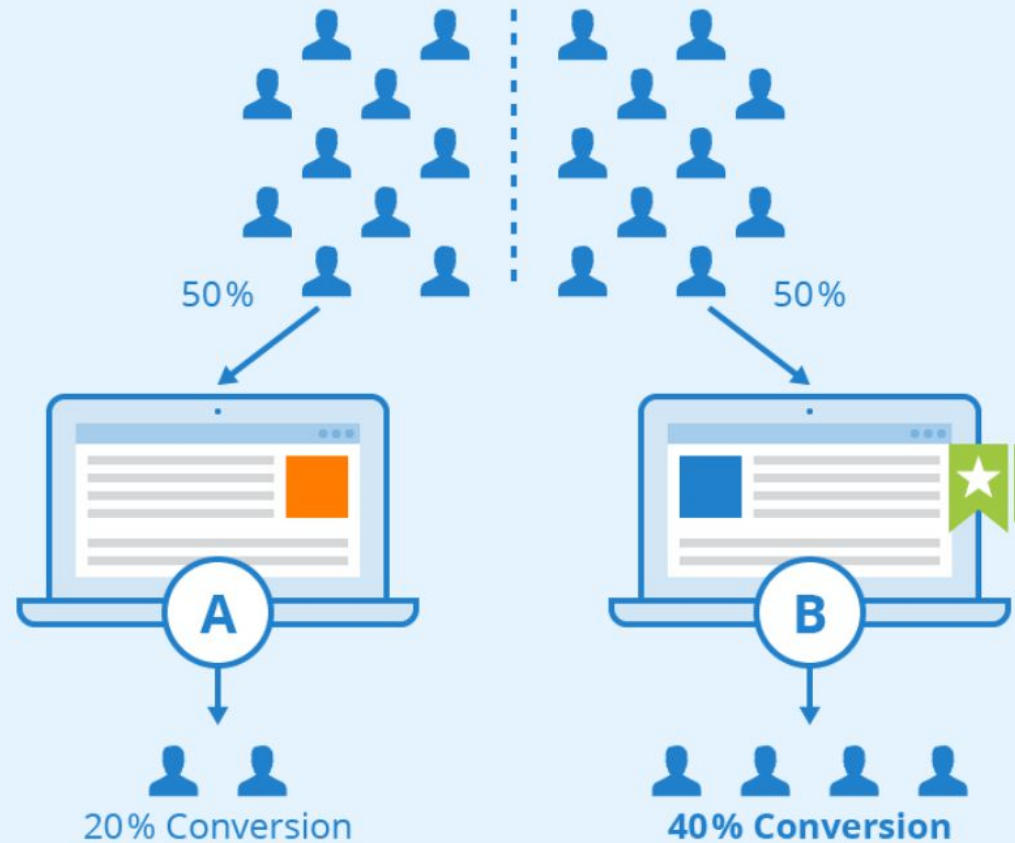
- Desenho do experimento
- População de usuários
- Predição de quem irá gastar com o APP
- Escolher o que se quer otimizar: retenção, métrica do usuário, etc
- Randomização em grupos
- Inferência bayesiana
- Os resultados obtidos refletem diferenças entre A e B ou achados aleatórios (by chance)?

Why you should try the Bayesian approach of A/B testing

The intuitive way of A/B testing. The advantages of the Bayesian approach and how to do it.



Michael Armanious Aug 29, 2020 · 7 min read ★



Diferenças *by chance*?

Suponha que um teste A/B será realizado para comparar duas versões diferentes de páginas *web* (aa e bb). Contou-se o número de visitantes em cada página e se clicaram ou não em algum dos *links* desejados, com os seguintes resultados:

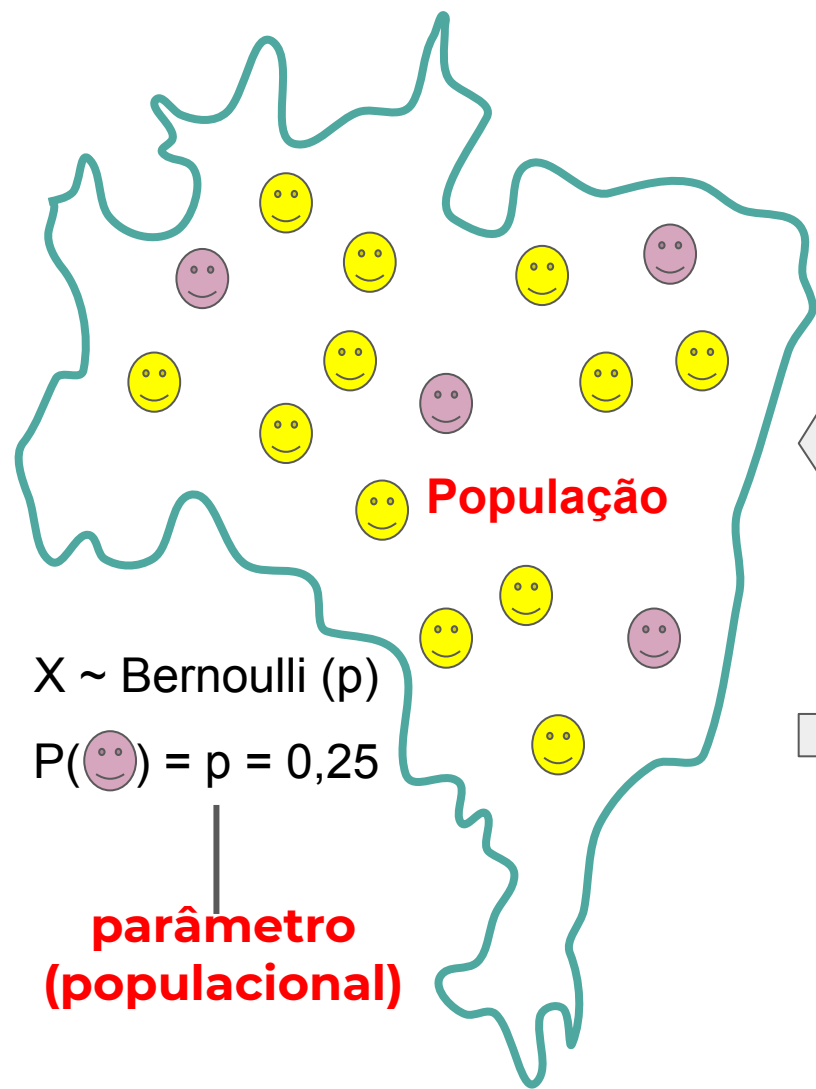
não clicou clicou

aa 4514 486 (9,7%)

bb 4473 527 (10,5%)

Para a comparação dos dois tipos de design em relação à resposta dos visitantes (clicks nos *links*), pode-se realizar um teste de hipóteses. O teste de hipóteses que compara se a proporção de visitantes que clica no *link* é a mesma entre aqueles que acessam a página com design aa e bb resultou num p-valor de 0,185.

- ❖ População x Amostra
- ❖ Parâmetro x Estimador
- ❖ Variável aleatória
- ❖ Distribuição de probabilidade
- ❖ Distribuição amostral
- ❖ Inferência clássica x bayesiana
- ❖ Testes de hipóteses
- ❖ p-valor

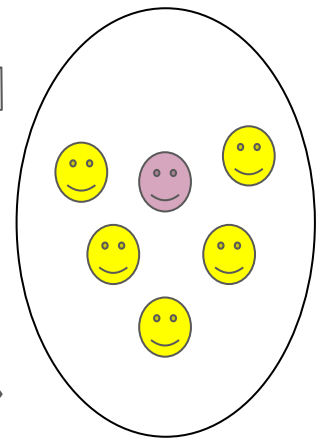


Clássica Bayesiana

Inferência estatística

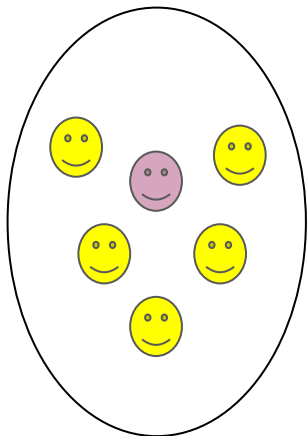


Amostra



$$\left. \begin{array}{l} X_1 = 0 \\ X_2 = 0 \\ X_3 = 1 \\ X_4 = 0 \\ X_5 = 0 \\ X_6 = 0 \end{array} \right\} \hat{p} = \sum_{i=1}^n X_i/n$$

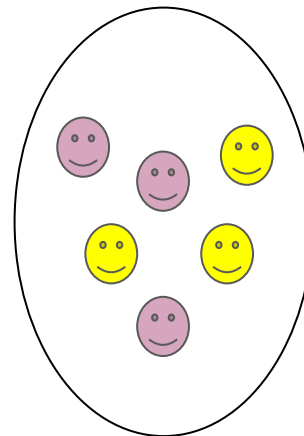
$\hat{p} = 0,167$
estimador (amostral)



Amostra 1

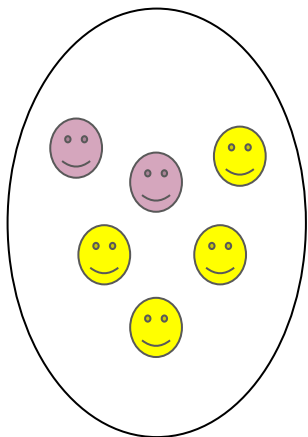
$$\hat{p}_1 = 0,167$$

estimador pontual



Amostra 3

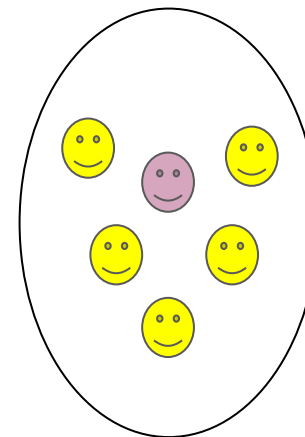
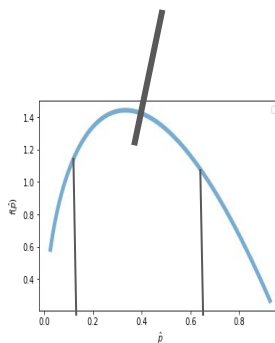
$$\hat{p}_3 = 0,5$$



Amostra 2

$$\hat{p}_2 = 0,333$$

estimador intervalar



Amostra 4

$$\hat{p}_4 = 0,167$$

distribuição amostral:

distribuição de probabilidade do estimador

Diferenças *by chance*?

Suponha que um teste A/B será realizado para comparar duas versões diferentes de páginas *web* (aa e bb). Contou-se o número de visitantes em cada página e se clicaram ou não em algum dos *links* desejados, com os seguintes resultados:

não clicou clicou

aa 4514 486 (9,7%)

bb 4473 527 (10,5%)

Para a comparação dos dois tipos de design em relação à resposta dos visitantes (clicks nos *links*), pode-se realizar um teste de hipóteses. O teste de hipóteses que compara se a proporção de visitantes que clica no *link* é a mesma entre aqueles que acessam a página com design aa e bb resultou num p-valor de 0,185.



```
import numpy as np
from scipy.stats import chi2_contingency
obs = np.array([[4514, 486], [4473, 527]])
chi2_contingency(obs)
```

```
↳ (1.7575018692680038, 0.18493641552090323, 1, array([[4493.5, 506.5],
           [4493.5, 506.5]]))
```

❖ Cálculo do tamanho amostral

❖ Amostragem

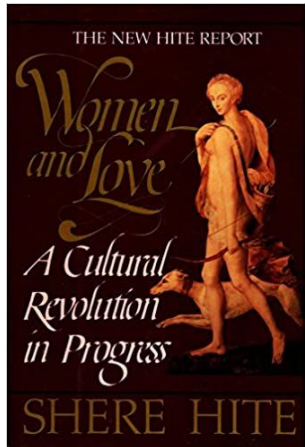
- ⌞ plano amostral: define como se dá a seleção da amostra
- ⌞ define-se a probabilidade de cada unidade ser selecionada (em amostragens probabilísticas)
- ⌞ **ignorar o plano amostral pode subestimar (ou superestimar) a variância do estimador**
- ⌞ Amostra Aleatória Simples: cada unidade da população tem mesma probabilidade de ser selecionada para a amostra, sem estratificação e em um único estágio com seleção aleatória.
- ⌞ Outras: estratificada, sistemática, múltiplos estágios, etc
- ⌞ amostragens não probabilísticas: de conveniência, voluntários

Amostragem de conveniência

- 84% insatisfeitas em seus relacionamentos
- 70% das casadas por 5+ anos têm relações fora do casamento
- 95% reportaram assédio emocional dos parceiros

- 100.000 questionários distribuídos, mas 4,5% retornaram
- 127 questões (longo!)

This "Hite Report" is not a report on average women. Its methodology is flawed. The author tells us she distributed 100,000 questionnaires to various organizations, church groups and a "wide range" of others. But wait. Even if every organization distributed every questionnaire to every woman and every woman answered it, Ms. Hite could only talk about women in these organizations, unless these members "represent" all women. How many women join feminist organizations? Or any organizations, for that matter? To accept this study as "science" would be wrong. But eventually the statisticians will come out with their brooms to sweep up the mess. Meanwhile, fishy statistics don't necessarily

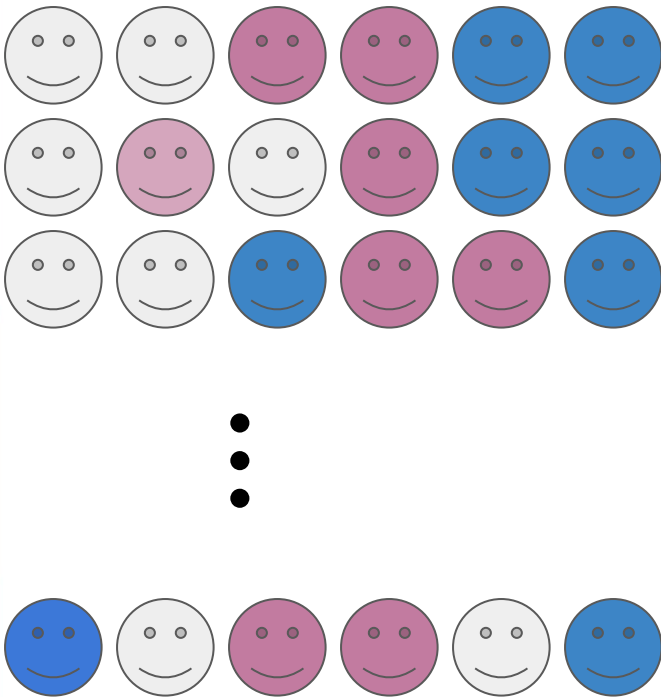


The New York Times

By Arlie Russell Hochschild
Nov. 15, 1987

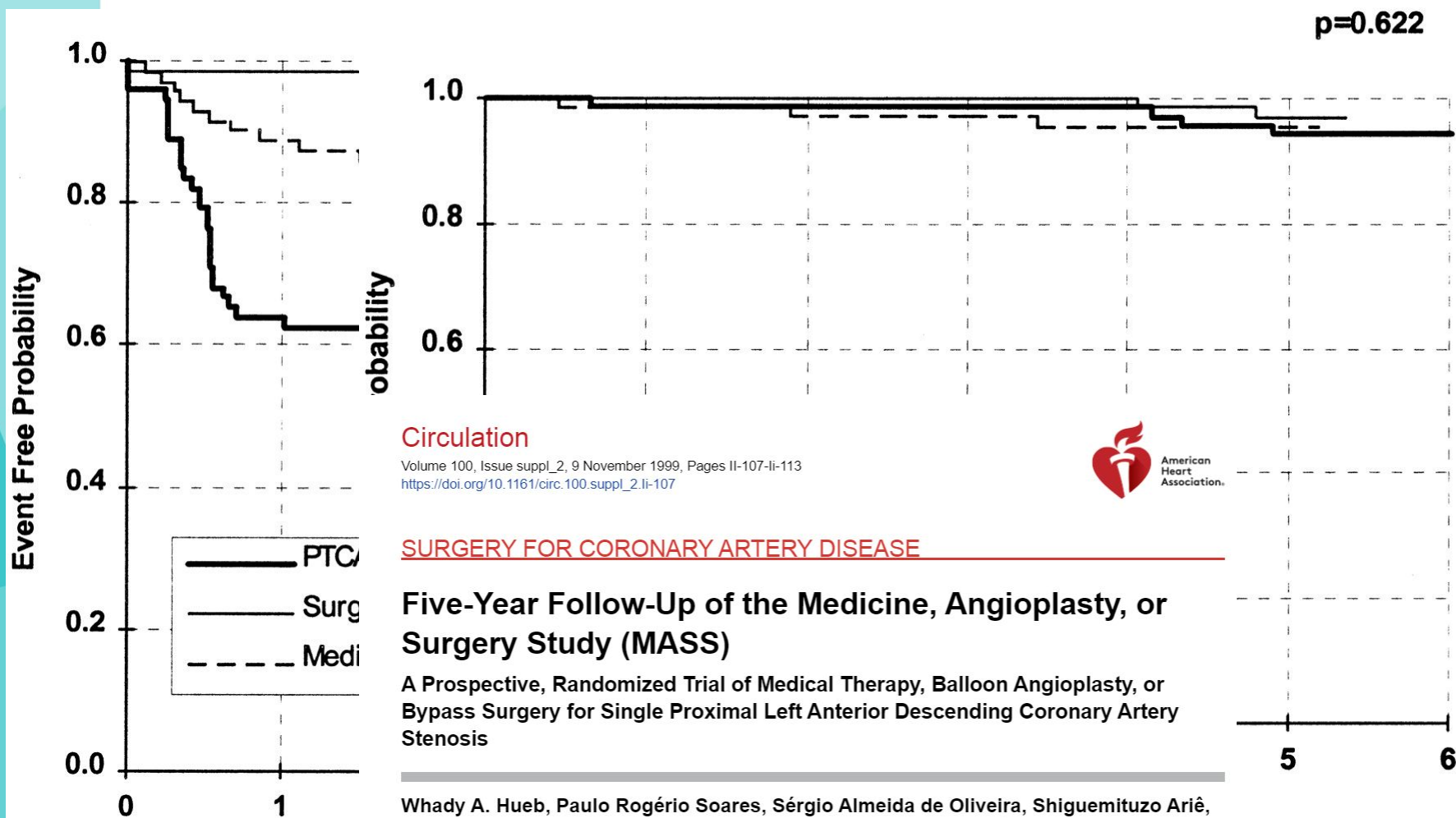
- ❖ médicos, enfermeiros estatísticos, psicólogos entre outros
- ❖ avaliar os resultados terapêuticos da DAC
- ❖ pacientes: angina estável, função ventricular preservada independentemente do número e grau de comprometimento arterial
- ❖ estudo randomizado
- ❖ após a intervenção, os pacientes tiveram seguimento clínico trimestral
- ❖ eventos primários: morte cardíaca, infarto do miocárdio ou angina refratária
- ❖ 20.769 pacientes recrutados, 611 elegíveis, randomizados para cirurgia (n= 203), angioplastia (n= 205), ou tratamento clínico (n= 203)

Amostragem em blocos



- ❖ aleatorização por blocos de 6
- ❖ equilibra o n^o de participantes nos grupos
- ❖ $P(\text{bloco})=1/45$
- ❖ análise dos dados deve considerar efeito de bloco

MASS (Medicine, Angioplasty, Surgery Study)



Circulation

Volume 100, Issue suppl_2, 9 November 1999, Pages II-107-II-113
https://doi.org/10.1161/circ.100.suppl_2.ii-107



SURGERY FOR CORONARY ARTERY DISEASE

Five-Year Follow-Up of the Medicine, Angioplasty, or Surgery Study (MASS)

A Prospective, Randomized Trial of Medical Therapy, Balloon Angioplasty, or Bypass Surgery for Single Proximal Left Anterior Descending Coronary Artery Stenosis

Whady A. Hueb, Paulo Rogério Soares, Sérgio Almeida de Oliveira, Shiguemitsu Ariê, Rita Helena A. Cardoso, Dalia Ballas Wajsbrot, Luiz A. M. Cesar, Adib D. Jatene, and José Antonio F. Ramires

❖ Inferência Bayesiana

The Power of Bayesian A/B

Testir



Michael
Jan 23, 2

Bayesian A/B testing — a

pr
sin

Why you should try the Bayesian approach of A/B testing



Bl:

The intuitive way of A/B testing. The advantages of the Bayesian approach and how to do it.

Fórmula da Probabilidade Total

Se $(B_1, B_2, B_3, \dots, B_k)$ uma partição de Ω e $A \subseteq \Omega$, então:

$$P(A) = P(B_1) \cdot P(A|B_1) + P(B_2) \cdot P(A|B_2) + \dots + P(B_k) \cdot P(A|B_k)$$

Fórmula da Bayes

$$P(B_i | A) = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^k P(A|B_i)P(B_i)}$$



Redes Bayesianas

Exemplo

Dois eventos podem fazer a grama ficar molhada: um regador ativo ou a chuva. A chuva tem um efeito direto no uso do regador (pois quando chove, o regador geralmente está inativo). Essa situação pode ser modelada com uma rede bayesiana. Cada variável tem dois valores possíveis: V (Verdadeiro) ou F (Falso). A [Distribuição de probabilidade conjunta](#) é:

$$P(G, R, C) = P(G|R, C) \cdot P(R|C) \cdot P(C)$$

onde G = grama molhada (verdadeiro/falso), R = regador ligado (verdadeiro/falso) e C = chovendo (verdadeiro falso).

O modelo pode responder questões sobre a presença de uma causa dada a presença de um efeito (também chamado de probabilidade inversa), como por exemplo "Qual é a probabilidade que esteja chovendo, dado que a grama está molhada?", usando a fórmula da [probabilidade condicional](#):

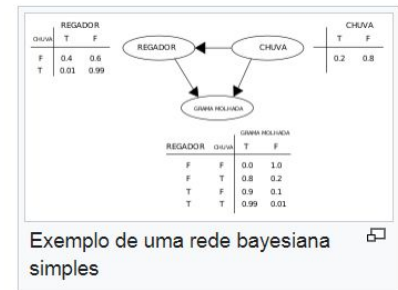
$$P(C = V|G = V) = \frac{P(G = V, C = V)}{P(G = V)} = \frac{\sum_{R \in \{V, F\}} P(G = V, R, C = V)}{\sum_{R, C \in \{V, F\}} P(G = V, R, C)}$$

Usando a expansão para a probabilidade conjunta $P(G, R, C)$ e as probabilidades condicionais das tabelas mostradas no diagrama, pode-se obter o valor de cada termo no numerador e no denominador. Por exemplo,

$$\begin{aligned} P(G = V, R = V, C = V) &= P(G = V|R = V, C = V)P(R = V|C = V)P(C = V) \\ &= 0.99 \times 0.01 \times 0.2 \\ &= 0.00198. \end{aligned}$$

Então os resultados numéricos são:

$$P(C = V|G = V) = \frac{0.00198_{VVV} + 0.1584_{VFV}}{0.00198_{VVV} + 0.288_{VVF} + 0.1584_{VFV} + 0.0_{VFF}} = \frac{0.16038}{0.44838} \approx 35.77\%.$$



2. COMPARAÇÃO DE CLASSIFICADORES

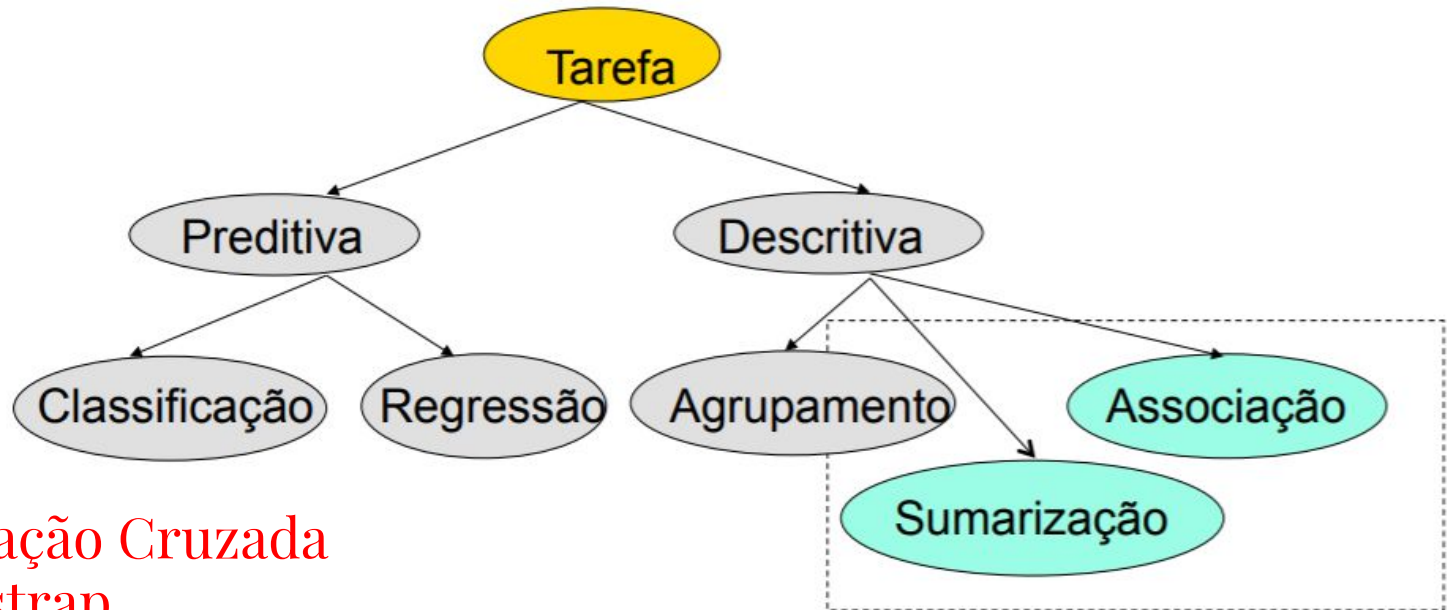
Aprendizado de Máquina

André C. P. L. F. de Carvalho
Centro de Pesquisa AMDA
ICMC-USP



Slide 24 da apresentação do André

Tarefas de aprendizado



Validação Cruzada
Bootstrap

Modelo de Regressão Logística
Modelo de Regressão Linear Múltipla, Ridge, Lasso
Análise de Agrupamento
Análise de Componentes Principais
Análise de Correlação

Slide 118 da apresentação do André

Avaliação de Desempenho de Classificador

Exemplo

- Matriz de confusão para 200 exemplos divididos em 2 classes

Classe verdadeira	Classe predita	
	p	n
P	70	30
N	40	60



Classe verdadeira	Classe predita	
	p	n
P	VP	FN
N	FP	VN

André Ponce de Leon F de Carvalho

118

Matriz de Confusão

F		Classificação (C)		
		positivo	negativo	TOTAL
Real (R)	positivo	138 (VP)	111 (FN)	249
	negativo	60 (FP)	2106 (VN)	2166
	TOTAL	198	2217	2415

ou recall

$$\text{sensibilidade} = P(C + | R+) = \frac{VP}{VP+FN}$$

$$\text{valor preditivo -} = P(R - | C-) = \frac{VN}{FN+VN}$$

$$\text{especificidade} = P(C - | R-) = \frac{VN}{FP+VN}$$

$$\text{valor preditivo +} = P(R + | C+) = \frac{VP}{FP+VP}$$

$$\text{acurácia} = P(\text{acerto}) = \frac{VP+VN}{VP+FN+FP+VN}$$

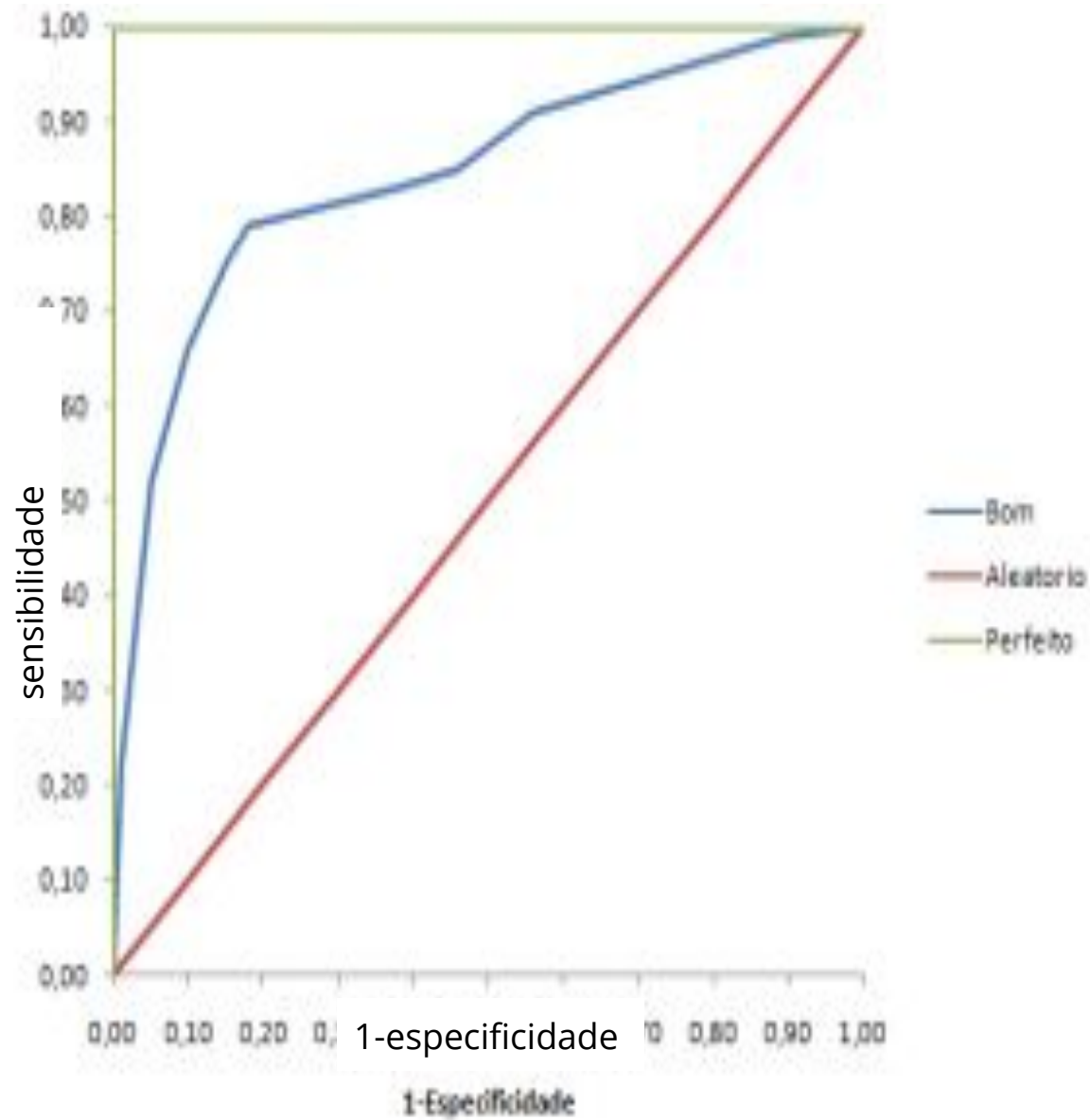
$$F_1 \text{ score} = 2 \frac{\text{precisão} * \text{recall}}{\text{precisão} + \text{recall}}$$

ou precisão

Diferença de nomenclatura

		Verdade			
População total		Condição positiva	Condição negativa	Prevalência $= \frac{\Sigma \text{Condição positiva}}{\Sigma \text{População total}}$	Acurácia (ACC) = $\frac{\Sigma \text{Verdadeiro positivo} + \Sigma \text{Verdadeiro negativo}}{\Sigma \text{População total}}$
Predito	Condição positiva prevista	Verdadeiro positivo	Falso positivo, Erro do tipo I	Valor preditivo positivo (PPV), Precisão = $\frac{\Sigma \text{Verdadeiro Positivo}}{\Sigma \text{Condição positiva prevista}}$	Taxa de falsa descoberta (FDR) = $\frac{\Sigma \text{Falso positivo}}{\Sigma \text{Condição positiva prevista}}$
	Condição negativa prevista	Falso negativo, Erro do tipo II	Verdadeiro negativo	Taxa de Falsa Omissão (FOR) = $\frac{\Sigma \text{Falso negativo}}{\Sigma \text{Condição negativa prevista}}$	Valor preditivo negativo (NPV) = $\frac{\Sigma \text{Verdadeiro negativo}}{\Sigma \text{Condição negativa prevista}}$
		Taxa de Verdadeiro Positivo (TPR), Revocação, Sensibilidade, probabilidade de detecção, Potência $= \frac{\Sigma \text{Verdadeiro positivo}}{\Sigma \text{Condição positiva}}$	Taxa de Falso Positivo (FPR), Fall-out, probabilidade de alarme falso $= \frac{\Sigma \text{Falso positivo}}{\Sigma \text{Condição negativa}}$	Teste da razão de verossimilhança positiva (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Razão de possibilidades de diagnóstico (DOR) $= \frac{\text{LR+}}{\text{LR-}}$
		Taxa de Falso Negativo (FNR), Taxa de perda = $\frac{\Sigma \text{Falso negativo}}{\Sigma \text{Condição positiva}}$	Especificidade (SPC), Seletividade, Taxa de Verdadeiro Negativo (TNR) $= \frac{\Sigma \text{Verdadeiro negativo}}{\Sigma \text{Condição negativa}}$	Teste da razão de verossimilhança negativa (LR-) = $\frac{\text{FNR}}{\text{TNR}}$	

Curva ROC e AUC



3. TESTE ADAPTATIVO COMPUTADORIZADO

Prova *linear*

1. AAAAAAAAAA

2. BBBBBBBBBB

3. CCCCCCCCCC

4. DDDDDDDDDD

5. EEEEEEEEEEE

6. FFFFFFFFFFFF

...

30. ΦΦΦΦΦΦΦΦ

Prova Adaptativa

5. EEEEEEEEEEE

7. μμμμμμμμμμμμμμ

TEORIA DE RESPOSTA AO ITEM (TRI)

Nota ENEM: <https://www.facebook.com/watch/?v=1719594845036710>

15. OOOOOOOOO

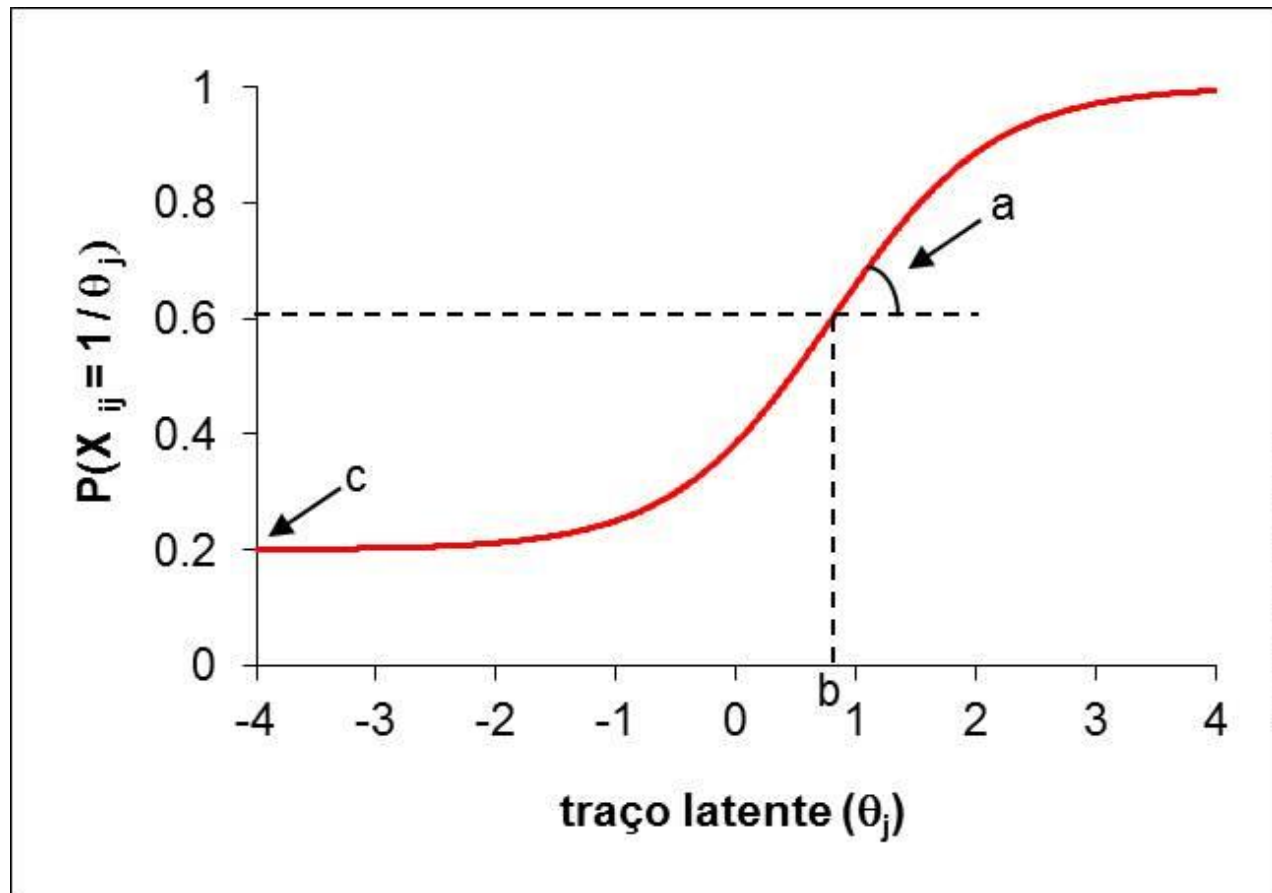
28. θθθθθθθθθθ

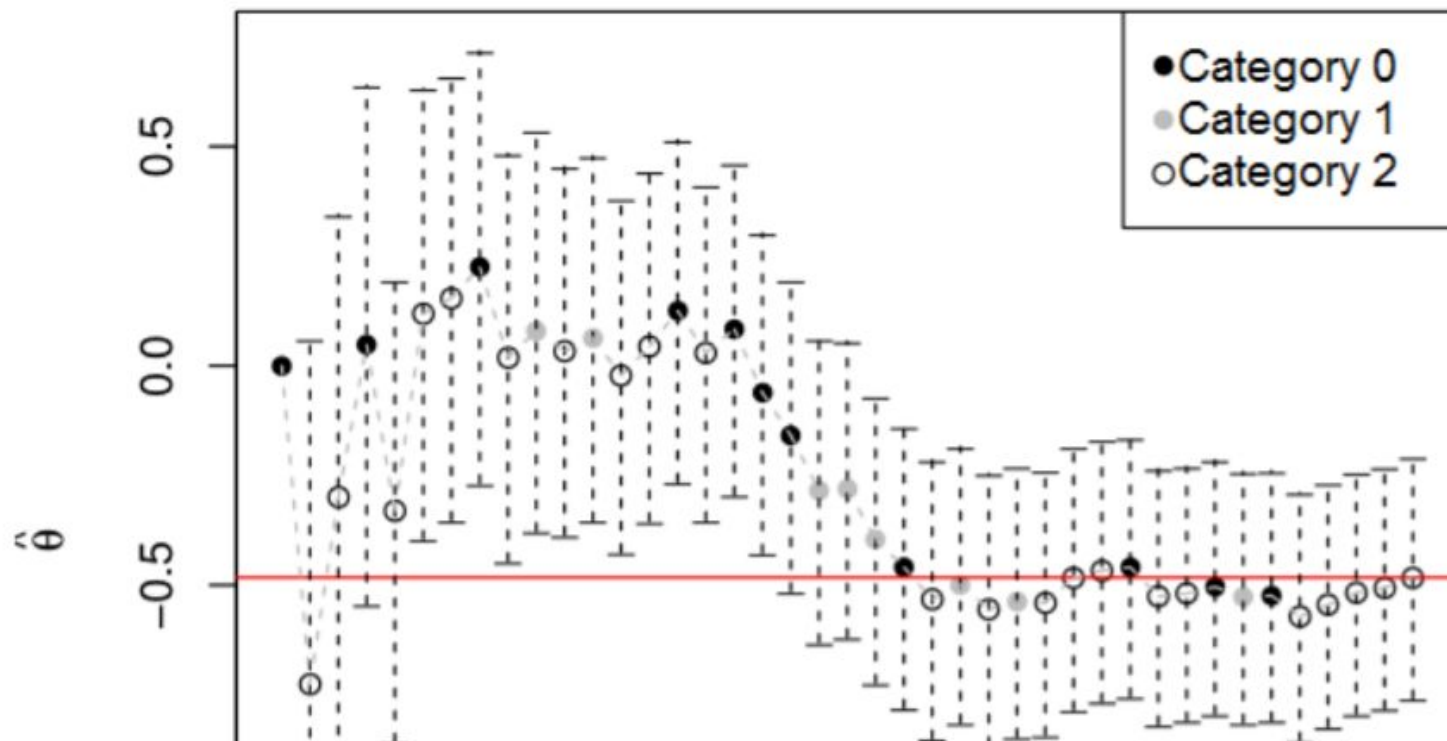
1. AAAAAAAAAA

16. LLLLLLLLLLLLLL

Modelo de 3 parâmetros para itens dicotômicos TRI

$$P(X_{ij} = 1 | \theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta_j - b_i)}}$$





Tendências em Matemática Aplicada e Computacional, **20**, N. 2 (2019), 381-401
 © 2019 Sociedade Brasileira de Matemática Aplicada e Computacional
www.scielo.br/tema
 doi: 10.5540/tema.2019.020.02.0381

Academic English Proficiency Assessment Using a Computerized Adaptive Test

M. CÚRI^{1*} and V. SILVA¹²

Interpretable Variational Autoencoders for Cognitive Models

Curi, M., Converse, G.A., Hajewski, J., Oliveira, S.

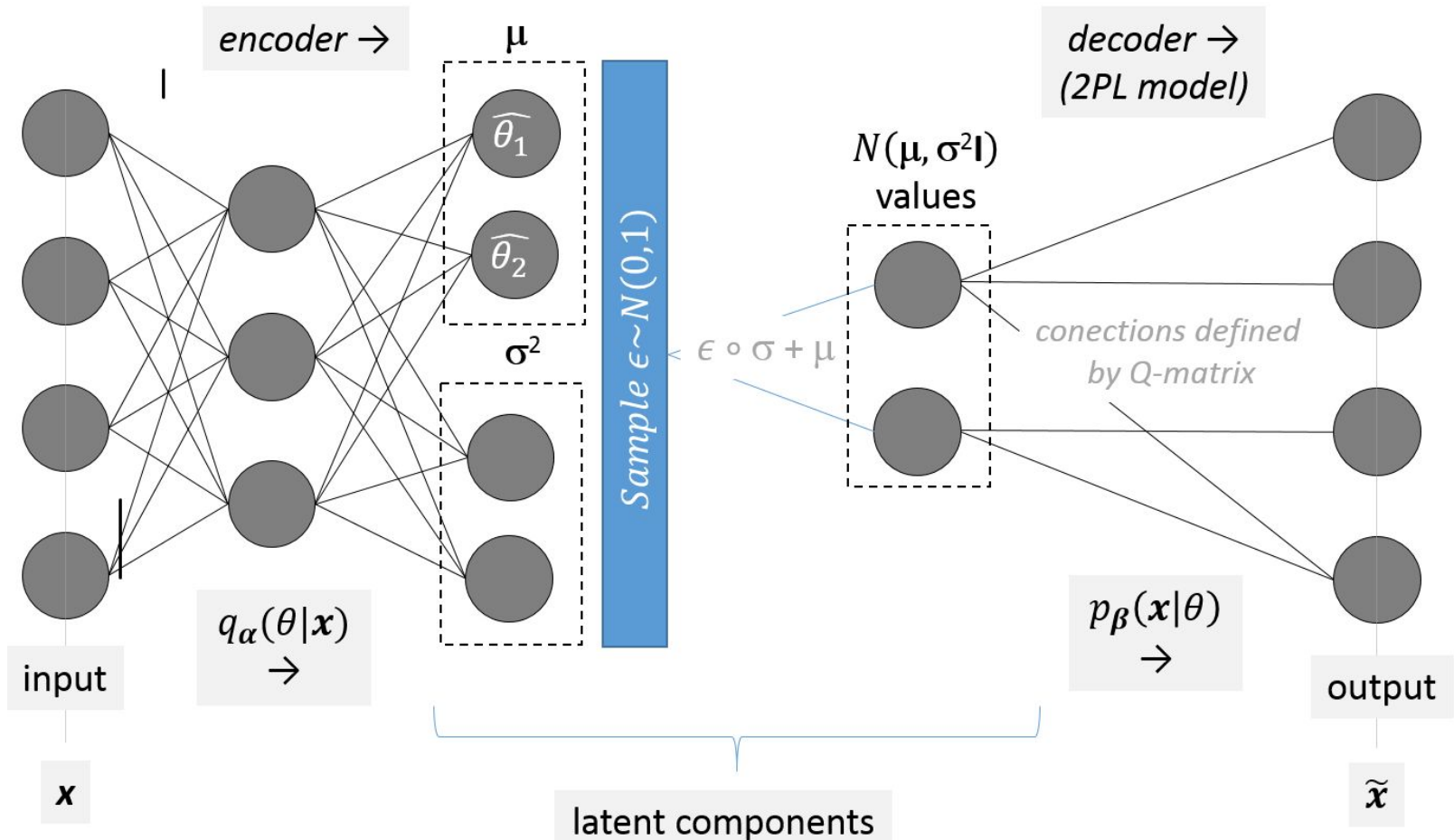
Proceedings of the International Joint Conference on Neural Networks, 2019, 2019-July, 8852333

[View abstract](#) [Related documents](#)

Autoencoders for educational assessment

Converse, G., Curi, M., Oliveira, S.

Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and



Neural Networks and Statistical Models

Proceedings of the Nineteenth Annual SAS Users Group International Conference, April, 1994

Warren S. Sarle, SAS Institute Inc., Cary, NC, USA

Abstract

There has been much publicity about the ability of artificial neural networks to learn and generalize. In fact, the most commonly used artificial neural networks, called multilayer perceptrons, are nothing more than nonlinear regression and discriminant models that can be implemented with standard statistical software. This paper explains what neural networks are, translates neural network jargon into statistical jargon, and shows the relationships between neural networks and statistical models such as generalized linear models, maximum redundancy analysis, projection pursuit, and cluster analysis.

Introduction

Neural networks are a wide class of flexible nonlinear regression and discriminant models, data reduction models, and nonlinear dynamical systems. They consist of an often large number of "neurons," i.e. simple linear or nonlinear computing elements, interconnected in often complex ways and often organized into layers.

Artificial neural networks are used in three main ways:

Jargon

Although many NN models are similar or identical to well-known statistical models, the terminology in the NN literature is quite different from that in statistics. For example, in the NN literature:

- variables are called *features*
- independent variables are called *inputs*
- predicted values are called *outputs*
- dependent variables are called *targets* or *training values*
- residuals are called *errors*
- estimation is called *training, learning, adaptation, or self-organization*.
- an estimation criterion is called an *error function, cost function, or Lyapunov function*
- observations are called *patterns* or *training pairs*
- parameter estimates are called (*synaptic*) *weights*
- interactions are called *higher-order neurons*
- transformations are called *functional links*
- regression and discriminant analysis are called *supervised learning* or *heteroassociation*
- data reduction is called *unsupervised learning, encoding, or autoassociation*
- cluster analysis is called *competitive learning* or *adaptive vector quantization*
- interpolation and extrapolation are called *generalization*

MULTIDISCIPLINARIDADE

