

FIAP

NBA

MBA em DATA SCIENCE & ARTIFICIAL INTELLIGENCE

APPLIED STATISTICS



Dra. Regina Tomie Ivata Bernal

Cientista de Dados na área da Saúde

Formação Acadêmica:

Estatístico - UFSCar

Mestre em Saúde Pública – FSP/USP

Doutor em Ciências – Epidemiologia - FSP/USP

Atividades Profissionais:

Professora de pós-graduação na FIAP

Consultora externa da SVS/MS

Cientista de Dados em Saúde

profregina.bernal@fiap.com.br
reginabernal@terra.com.br



AMOSTRAGEM



Exemplo: Quarterly Earnings per Johnson & Johnson Share

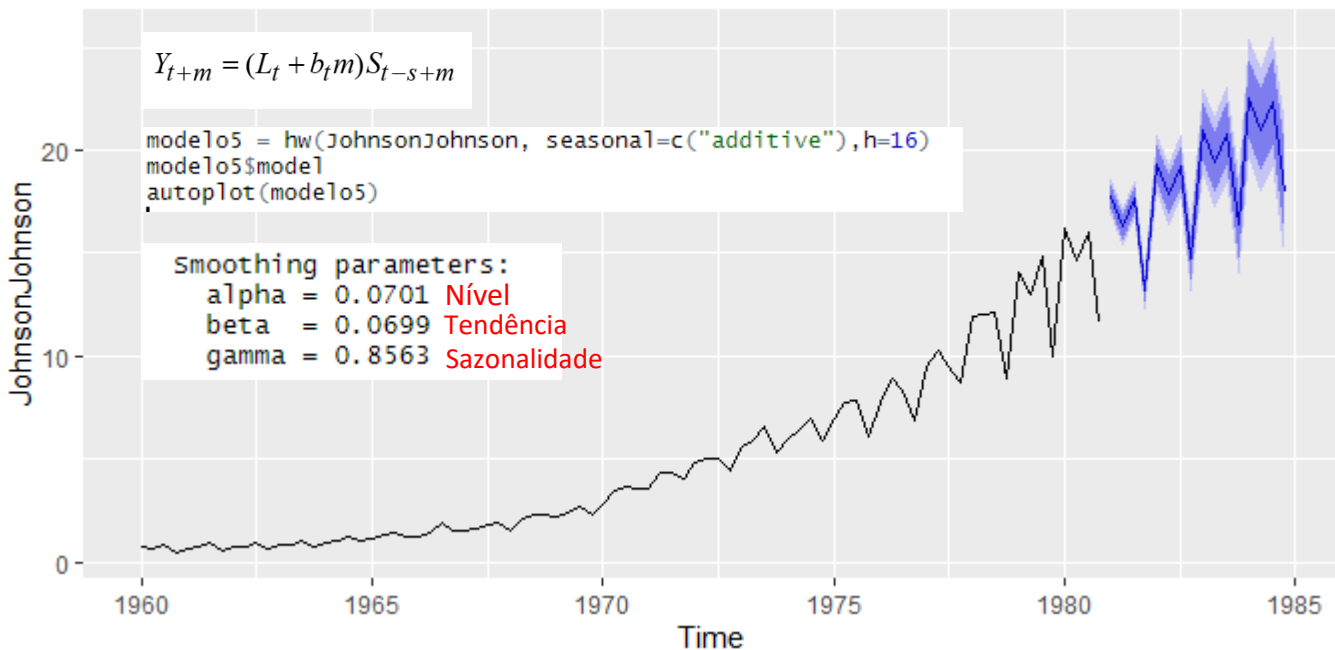
Forecasts from Holt-Winters' additive method

$$Y_{t+m} = (L_t + b_t m) S_{t-s+m}$$

```
modelo5 = hw(johnsonjohnson, seasonal=c("additive"),h=16)
modelo5$model
autoplot(modelo5)
```

Smoothing parameters:

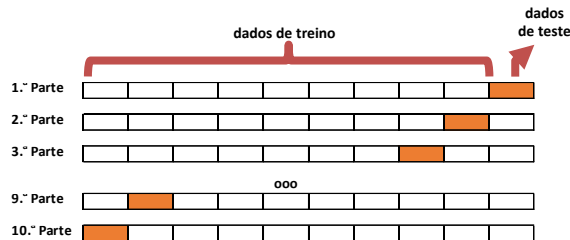
- alpha = 0.0701 **Nível**
- beta = 0.0699 **Tendência**
- gamma = 0.8563 **Sazonalidade**



APLICAÇÃO DA AMOSTRAGEM

Validação Cruzada

- Dividir os dados em partes iguais e utilizar:
 - uma fração delas para treinar o algoritmo com um hiperparâmetro;
 - outra parte testar a sua predição



Seleção do hiperparâmetro com melhor performance → definição do algoritmo com esse hiperparâmetro nos dados de treino.

Fazer o mesmo para todos os algoritmos.

A única forma de saber qual o algoritmo de melhor performance é testando todos.

APLICAÇÃO DE IA

RETAIL OCTOBER 10, 2018 / 8:04 PM / UPDATED 3 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ



SAN FRANCISCO (Reuters) - Amazon.com Inc's [AMZN.O](#) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

...

That is because Amazon's computer models were trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the tech industry.

Fonte: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>

EXEMPLO

APLICAÇÃO DE MACHINE LEARNING

Ethical Implications Of Bias In Machine Learning

Adrienne Yapo
Bentley University
175 Forest St.
Waltham, MA 02452
adrienne.yapo@gmail.com

Joseph Weiss
Bentley University
175 Forest St.
Waltham, MA 02452
jweiss@bentley.edu

1.2. Machine learning algorithm bias

Although machine learning algorithms can produce numerous benefits to individuals, consumers, businesses, investors, the government, and society at large, recent research has uncovered many instances of bias in machine learning algorithms that have troubling implications and deleterious consequences.

1.3. Machine learning in the criminal justice system

Yet perhaps the most troubling incidents of bias in machine learning to date are unfolding in the criminal justice system. Consider the following statement from then U.S. Attorney General Eric Holder on the Sentencing Reform and Corrections Act of 2015:

Table 1: Disproportionate incarceration rates

Source: *Propublica analysis from Broward County, Florida*

Prediction Fails Differently for Black Defendants

	White	African American
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

"Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes."

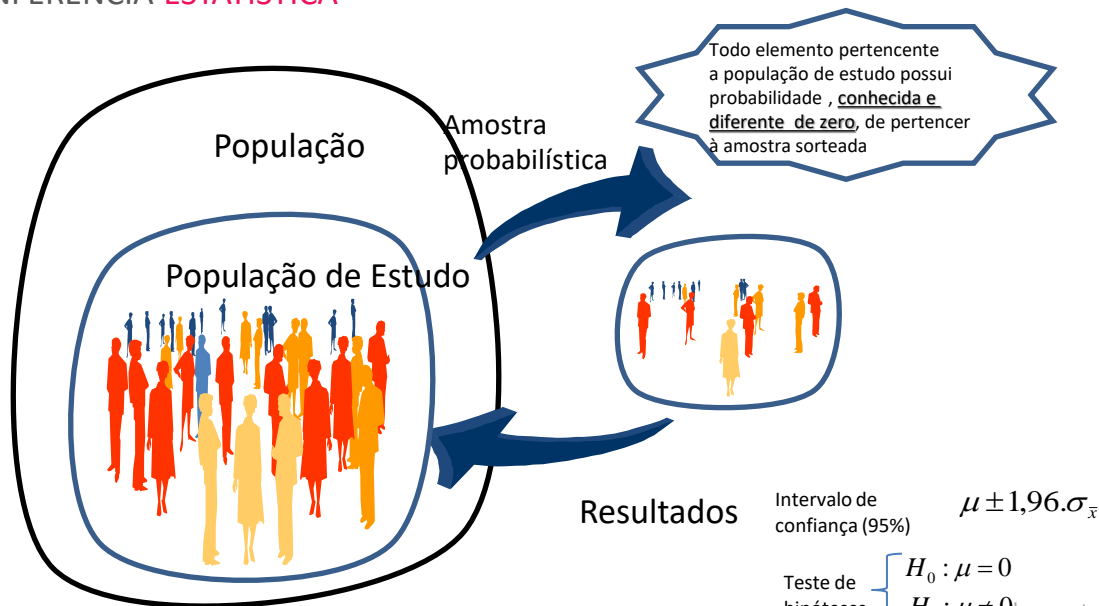
Fonte: <https://scholarspace.manoa.hawaii.edu/bitstream/10125/50557/paper0670.pdf>



INFERÊNCIA ESTATÍSTICA



INFERÊNCIA ESTATÍSTICA



INFERÊNCIA ESTATÍSTICA

• Amostragem

O que é necessário garantir?

- Que a amostra seja representativa da população

A amostra deve possuir as mesmas características básicas da população, no que diz respeito às variáveis que desejamos pesquisar.

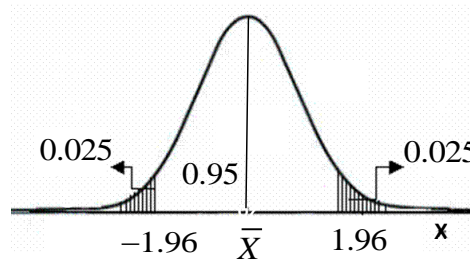
INFERÊNCIA ESTATÍSTICA

Usos de amostragem

- ✓ Pesquisa eleitoral
- ✓ Pesquisa com clientes
- ✓ Desenvolvimento de modelos estatísticos
 - ✓ Amostra de desenvolvimento
 - ✓ Amostra de validação

Intervalo de Confiança

Intervalos de confiança $P[(\bar{x} - 1.96.dp(\bar{x})) \leq \bar{X} \leq \bar{x} + 1.96.dp(\bar{x})] = 0.95$



Distribuição da média amostral segundo o modelo normal com parâmetros ($\bar{x}; dp(\bar{x})$)

O uso da distribuição normal como modelo para a distribuição da média amostral possibilita esperar que 95% das estimativas sejam diferentes do valor populacional por no máximo 1.96 desvios padrão.

Intervalo de Confiança

A figura 8.5 mostra 100 intervalos de confiança (95%) para a média de glicose no sangue de uma população. Os intervalos com estimativas menores que 34(g/l) ou maiores que 36(g/l) também não contém o valor 35.

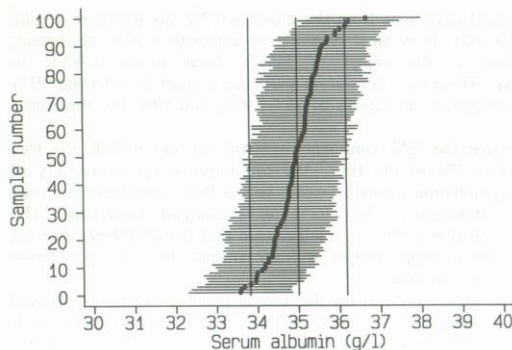
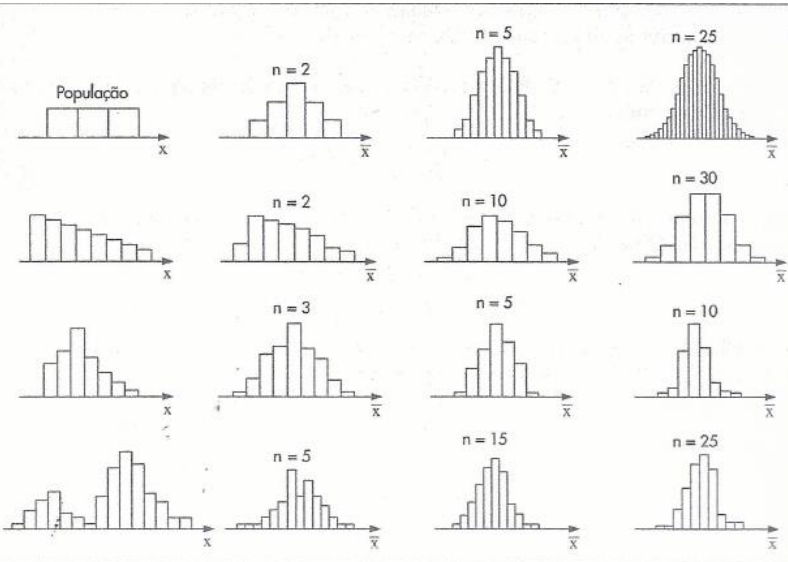


Figure 8.5 Confidence intervals from Figure 8.4 ordered by the magnitude of the mean of the random sample.

Teorema do Limite Central



Para amostras aleatórias simples (X_1, \dots, X_n) retiradas de uma população com média μ e variância σ^2 finita, a distribuição da média amostral \bar{X} aproxima-se, para n grande, de distribuição normal com média μ e variância $\frac{\sigma^2}{n}$.

Fonte: BUSSAB, W.O.; MORETTIN, P. A., Estatística Básica, 5a. ed., São Paulo: Saraiva, 2006. Página 273.

Medidas de Dispersão

Desvio Padrão X Erro Padrão (Std Error)

- Desvio Padrão: variabilidade das observações em relação à média de uma amostra.
- Erro Padrão (Std Error) : variabilidade entre as amostras

$$Erropadrao = \frac{desviopadrão}{\sqrt{n}}$$

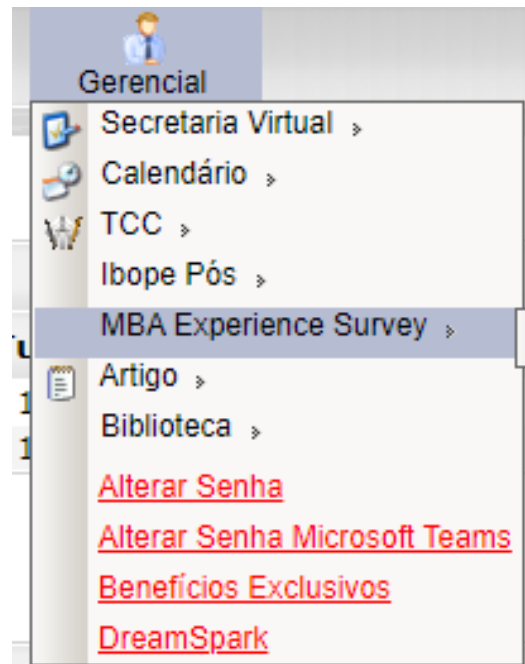
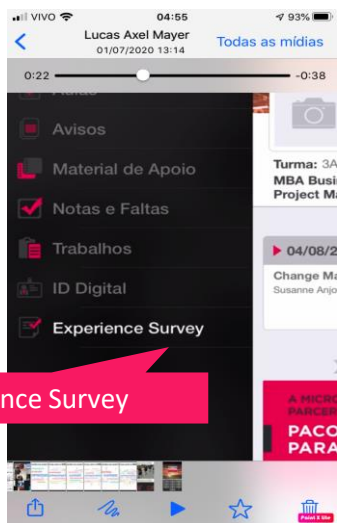
A grande finalidade do
conhecimento não é conhecer,
mas agir.

T. Huxley

O que você achou da aula de hoje?

Pelo aplicativo da FIAP

(Entrar no FIAPP, e no menu clicar em Experience Survey)



OBRIGADA



/ Regina T. I. Bernal

FIAP

Copyright © 2023 | Professora Dra. Regina Tomie Ivata Bernal
Todos os direitos reservados. Reprodução ou divulgação total ou parcial deste documento, é expressamente
proibido sem consentimento formal, por escrito, do professor/autor.

FIAP