

Estatística na Ciência

Alexandre Galvão Patriota

Departamento de Estatística
Instituto de Matemática e Estatística
Universidade de São Paulo

Webinar IVEPESP

21 de julho de 2020

Centro de Estatística Aplicada



Fundado em 1974, o CEA tem como objetivo fundamental assessoria estatística.

- Diretor Científico: Prof. Antonio Carlos Pedroso
- Vice-diretor: Prof. Julio Singer
- Sítio: www.ime.usp.br/cea/

Mais de 800 projetos desde 1975:

- 84% USP, 16% Externo.
- 28% DO, 37% ME, 26% PUB, 9% outros.

Em 2019, 19 docentes do MAE orientaram 23 alunos em 21 projetos oriundos dos seguintes órgãos:

- Escola de Educação Física e Esportes
- Escola de Enfermagem
- Escola Politécnica
- Faculdade de Direito
- Faculdade de Economia e Administração
- Faculdade de Medicina
- Faculdade de Odontologia
- Instituto da Criança da Faculdade de Medicina
- Instituto de Energia e Ambiente
- Instituto de Geociências
- Instituto Oceanográfico
- Instituto de Reabilitação Lucy Montoro
- Museu de Arqueologia e Etnologia – USP
- Universidade Federal de São Paulo

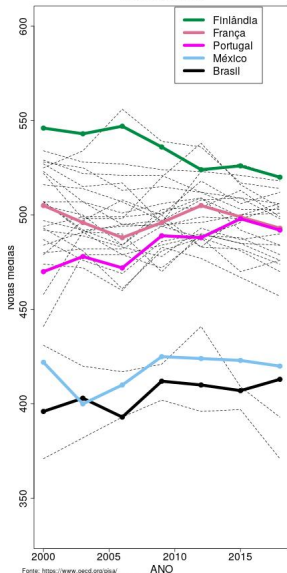
Algumas aplicações da Estatística

- Proficiência de alunos do Ensino Básico.
- Desmatamento.
- Óbitos por Covid-19.
- Eficácia de medicamentos e vacinas.
- Relações entre variáveis.
- Existência de novas partículas.

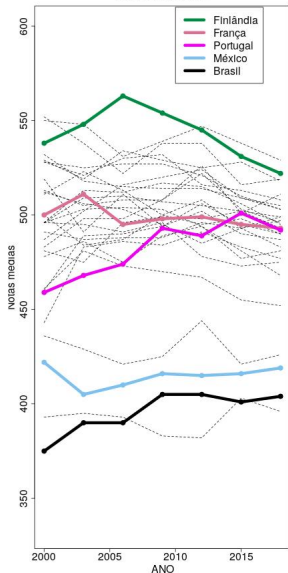
Proficiência de alunos do Ensino Básico

Alunos de 15 anos (Exame PISA)

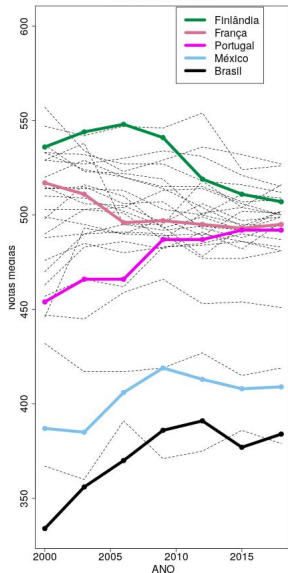
PISA em Leitura



PISA em Ciências



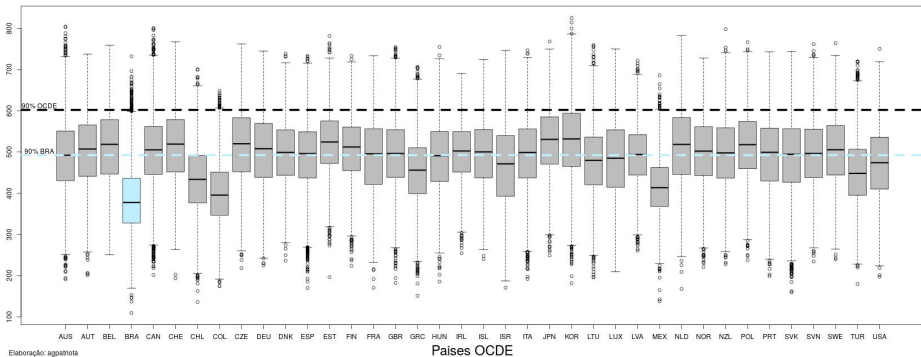
PISA em Matemática



Fonte: <https://www.oecd.org/pisa/>
Elaboração: ime@ime.usp.br

Matemática (exame PISA 2018)

Média das notas no exame Pisa (Matemática)

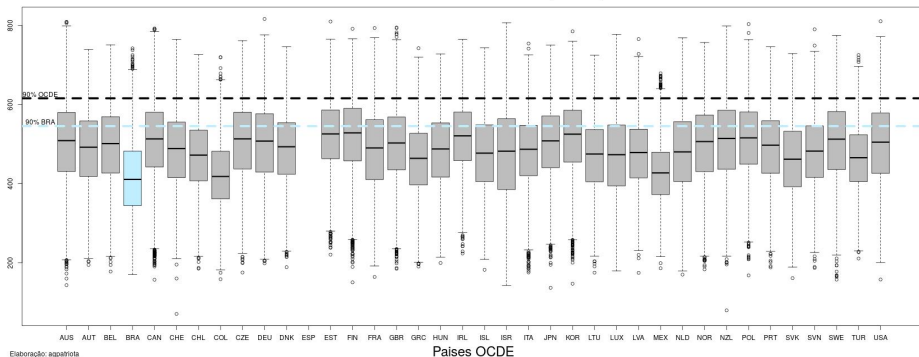


Elaboração: agosto/2019

Países OCDE

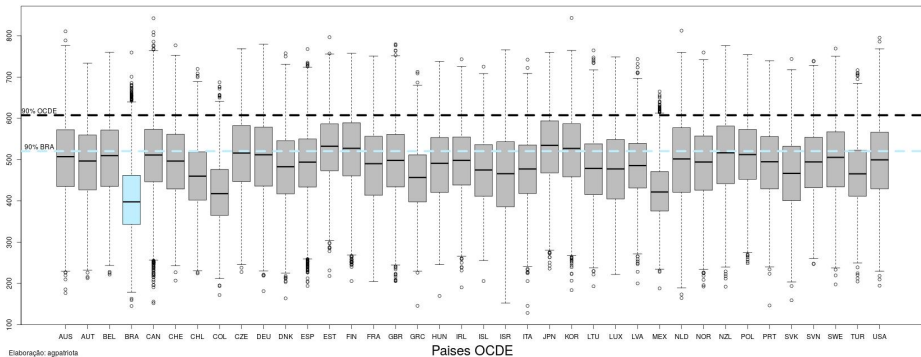
Leitura (exame PISA 2018)

Média das notas no exame Pisa (Leitura)



Ciências (exame PISA 2018)

Média das notas no exame PISA (Ciências)

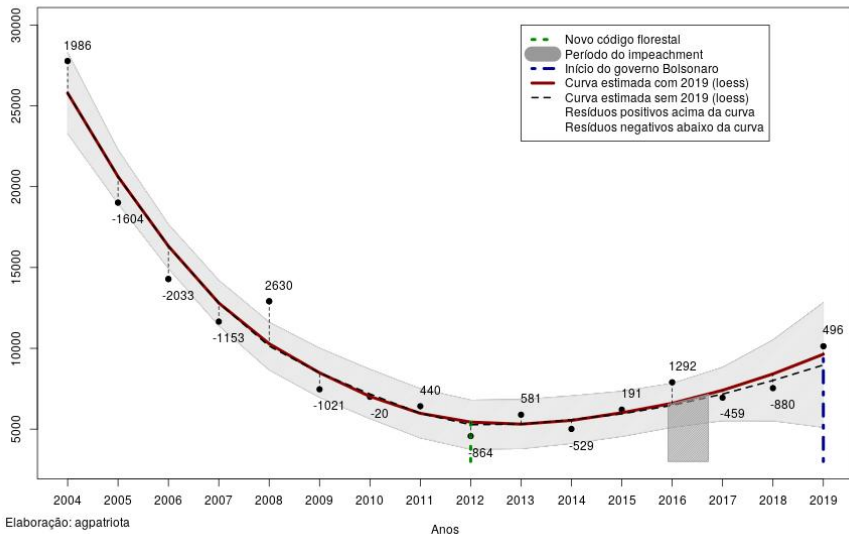


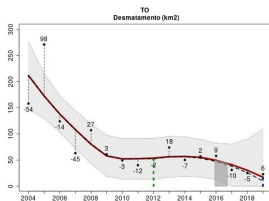
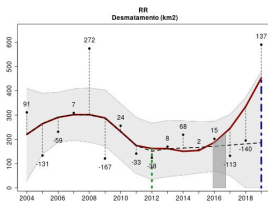
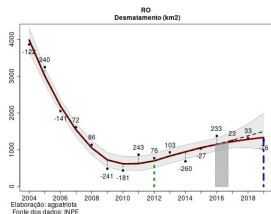
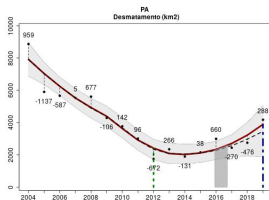
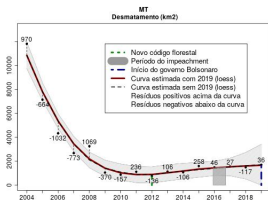
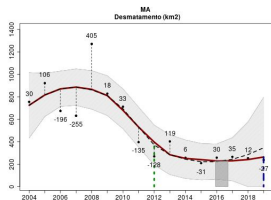
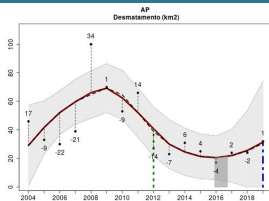
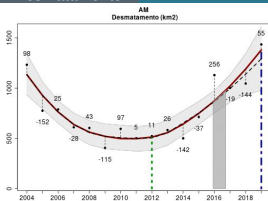
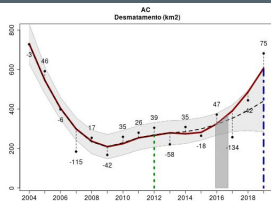
Desmatamento

Desmatamento na Amazônia legal

Desmatamento da Amazônia Legal (km²)

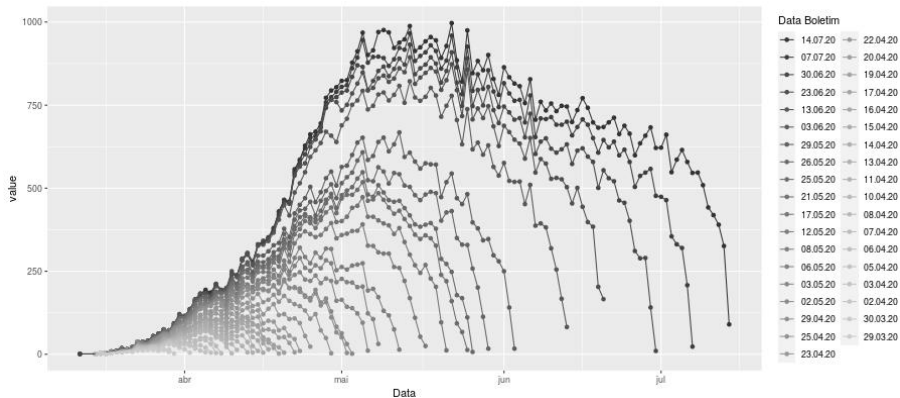
Fonte dos dados: INPE





Óbitos por Covid segundo a data de ocorrência

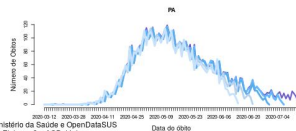
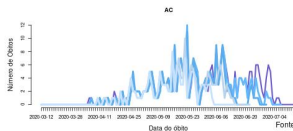
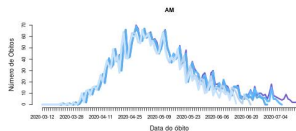
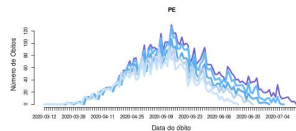
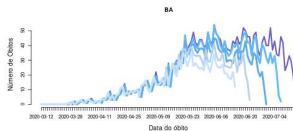
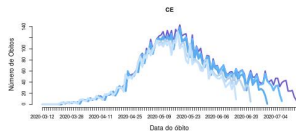
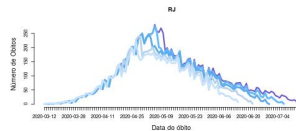
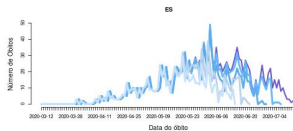
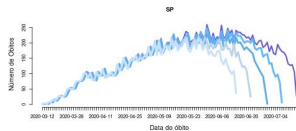
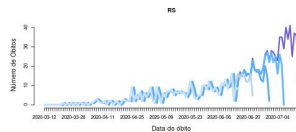
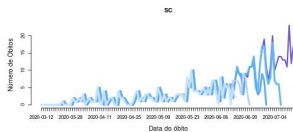
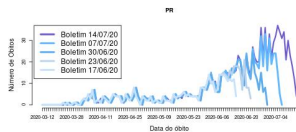
Óbitos por Covid (data de ocorrência)



Ver vídeo:

www.ime.usp.br/~patriota/ObitosBoletins.mp4

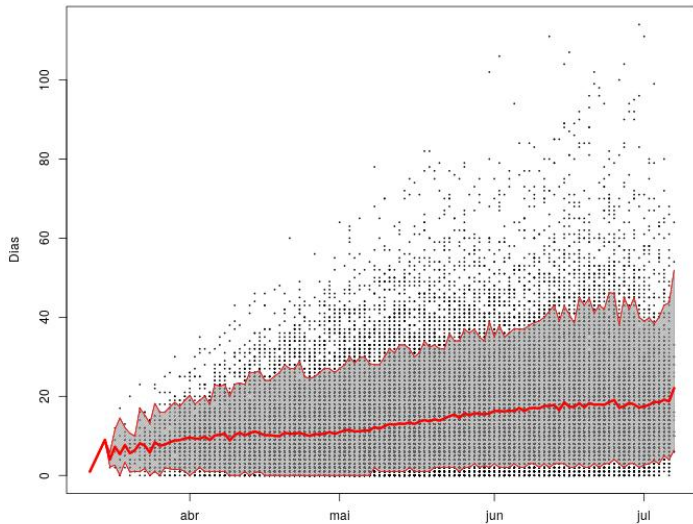
Óbitos por Covid (data de ocorrência)



Fonte: Ministério da Saúde e OpenDatASUS
Elaboração: AGPatriota

Tempo até o óbito

Tempo decorrido dos primeiros sintomas ao óbito



Alguns objetivos desta palestra

Alguns objetivos desta palestra

- ✓ Apresentar alguns exemplos.

Próximos tópicos:

- Apresentar (alguns dos) os protagonistas principais na história do método científico e da estatística.
- Discutir como a estatística, a ciência e o método científico estão relacionados.
- Por que a estatística é fundamental para a ciência moderna?

Método Científico uma breve excursão

Aristóteles (384 a.C.)

A **lógica**, a **teoria de conjuntos** e o **método científico** moderno têm a primeira formulação teórica com Aristóteles:

- *Analytica Priora*,
- *Analytica Posteriora*,
- *Topica*.

Entretanto,

- Aristóteles não tratou de proposições hipotéticas¹
- Considerava a ciência como um sistema de proposições obtidas por induções e deduções lógicas.

¹Bobzien (2002). The Development of Modus Ponens in Antiquity: From Aristotle to the 2nd Century AD, *Phronesis*, 47

Theofrastos e os peripatéticos (372 a.C.)

Theofrastos e seus colegas peripatéticos desenvolveram a **lógica condicional** empregada na ciência moderna.

- *Modus Tollens* (tolher):

Se $a \rightarrow b$ e $\neg b$, então $\neg a$

- *Modus Ponens* (pôr):

Se $a \rightarrow b$ e a , então b

Mais tarde, Ronald Fisher usa o *Modus Tollens* para testar hipóteses estatísticas.

Alhazen (965 d.C)

Alhazen foi um matemático e físico árabe.

Muitos creditam a Alhazen o desenvolvimento do **método científico moderno** (UNESCO, 2015)².

Ele desenvolveu:

- experimentos controlados para testar hipóteses,
- repetição de experimentos para verificação independente.

Influenciou fortemente Roger Bacon.

²<http://www.light2015.org/Home/ScienceStories/1000-Years-of-Arabic-Optics.html>

Roger Bacon (1220)

Roger Bacon foi um padre franciscano que se dedicou ao estudo da filosofia, lógica, aritmética, linguística, astronomia, óptica e método científico.

Escreveu o Opus Majus (parte VI):

- Hipótese-Experimento-Verificação independente.
- Os princípios devem ser testados experimentalmente.

É creditado como o pai do empirismo (Kupfer, 1974)³ e algumas vezes como pai do método científico.

³Kupfer, J. (1974). The Father of Empiricism: Roger not Francis, Vivarium, 12

-384**Aristóteles**

Lógica, conjuntos e Escrutínio racional.

$$\neg(p \implies \neg p), \neg(\neg p \implies p).$$

-372**Theofrastos**

Lógica condicional: (" $p \rightarrow q, \neg q \implies \neg p$ ").

-300**Euclides**

Axiomatização. Geometria, teoria dos números e álgebra.

965**Alhazen**

Método científico: Testes controlados, Repetições.

1220**Roger Bacon**

Opus Maius parte VI

Hipótese-Experimento-Verificação independente

Os princípios devem ser testados

experimentalmente.

Método Científico Moderno

Método Científico Moderno

Um esboço do método científico moderno (Alhazen e Roger Bacon) segue:

- Definir as **hipóteses** de interesse.
- Desenhar o **experimento**.
- Coletar os dados.
- Testar as hipóteses contra os dados.
- Verificação independente dos resultados (**repetição** do experimento).

Cada etapa é traduzida para a linguagem estatística.

Artigo científico e a Ciência

- Um **artigo científico** comunica os resultados obtidos por meio do método científico.
- A **ciência** é composta pelos resultados científicos e suas respectivas teorias.
- A ciência se **autocorrige** com o tempo por meio das verificações independentes.

Apesar de suas limitações, o método científico é o que temos de mais confiável na produção de conhecimento.

Estatística uma breve excursão

1654 ···●

Fermat e Pascal

Princípios da Probabilidade
Probabilidade Condicional (informal).

1733 ···●

Thomas Bayes

Inferência Bayesiana
Probabilidade inversa.

1884 ···●

Charles Sanders Pierce

Verossimilhança, Confiança, Axiomatização da probabilidade, Planejamento de experimentos.

1894 ···●

Karl Pearson

Método de momentos, frequentismo, Estatística Matemática.

1912 ···●

Ronald Fisher

Máxima verossimilhança, Suficiência, Testes de hipóteses, valor-p (propriedades e caso geral).

Como o método científico e a estatística estão relacionados?

Hipóteses

A hipótese científica S é traduzida para a hipótese estatística H_0 :

$$S : \text{“O medicamento não tem efeito”} \longleftrightarrow H_0 : \beta = 0,$$

em que β é o efeito médio do medicamento.

Atenção:

- **Não existe** uma única representação da hipótese científica.
- Em vez de efeitos médios poderiam ser quantis.

Experimentos

As características mensuráveis de um experimento são traduzidas em termos de modelos estatísticos:

$$\text{“Experimento científico”} \longrightarrow (\Omega, \mathcal{A}, \mathcal{P})$$

em que:

- Ω é o conjunto de possíveis resultados experimentais,
- \mathcal{A} é a classe de eventos experimentais em que se tem interesse,
- \mathcal{P} é uma família de probabilidades que podem descrever os resultados experimentais.

Atenção: O modelo nem sempre é adequado. Verificações são necessárias.

Coleta de dados → amostragem

Tecnologia de amostragem é uma disciplina que estuda como se obtêm os dados:

- Amostragem aleatória simples.
- Amostragem por estratificação.
- Amostragem por conglomerados.
- Amostragem por conveniência.
- Amostragem por quotas.

Objetivo: obter uma amostra livre de vieses de confirmação e de seleção e que tenha propriedades probabilísticas boas.

Testar a hipótese

A hipótese H_0 é testada da seguinte forma:

- Observam-se os dados x do Experimento.
- Define-se uma estatística T_{H_0} que reflita a discordância entre H_0 e os dados.
- Calcula-se o valor-p:

$$p(H_0, x) = P(T_{H_0} \geq t_{obs}),$$

em que P é a medida de probabilidade mais favorável a H_0 .

O valor-p **NÃO** é uma probabilidade condicional.

Modus Tollens na estatística

Fisher: “Se $p(H_0; x) < \alpha$, então um evento raro ocorreu ou H_0 é falsa”, para α pequeno:

Seja R_{H_0} um evento raro sob H_0 .

$$\text{Se } \underbrace{[H_0 \wedge (x \notin R_{H_0})]}_A \rightarrow \underbrace{p(H_0; x) \geq \alpha}_B \wedge \underbrace{(p(H_0; x) < \alpha)}_{\neg B},$$

$$\text{então } \underbrace{[(\neg H_0) \vee (x \in R_{H_0})]}_{\neg A}.$$

Meus principais trabalhos relacionados:

1. Patriota, 2013, Fuzzy Sets and Systems, 233.
2. Patriota, 2017, Educational and Psychological Measurement, 77.
3. Patriota, 2018, Scientometrics, 116.
4. Bickel e Patriota, 2019, Bernoulli, 25.

Estatística como linguagem

A estatística como linguagem

Como vimos, a estatística é uma linguagem formal para lidar com a incerteza.

Karl Pearson disse “*Statistics is the grammar of science*”.

Sem a estatística, não seria possível tratar a incerteza nos dados e a imprecisão dos modelos.

Limitações da estatística

Se pelo menos uma tradução estiver equivocada, os resultados estatísticos terão pouca ou nenhuma validade.

Deve-se verificar também

- a sensibilidade do modelo adotado contra pequenas perturbações,
- se o tamanho amostral é suficiente para fazer aproximações,
- se há informação redundante nas variáveis.

Palavras finais

- A estatística pode ser usada para ajudar a compreender os fatos ou para corroborar uma ideologia.
- Em áreas mais politizadas, obtêm-se conclusões opostas usando os mesmos dados (índices diferentes).
- A estatística pode ser usada para entender que tipo de manipulação foi feita nos resultados (desde que se tenha os dados).

Obrigado!