

SME0320 – Estatística I – Eng. Aeronáutica (2023)

Thomas Peron

thomas.peron@usp.br

tkdmperon.github.io

  @thomas_peron

Sala 3-250 B

Objetivos do curso

Ensinar as ideias básicas da Estatística, seus alcances e limitações. Estabelecer uma linguagem comum entre o Engenheiro e o Estatístico. Exemplificar através das técnicas mais comuns de Estatística.

(Jupiter Web, 2023)

Objetivos do curso

Ensinar as ideias básicas da Estatística, seus alcances e limitações. Estabelecer uma linguagem comum entre o Engenheiro e o Estatístico. Exemplificar através das técnicas mais comuns de Estatística.

(Jupiter Web, 2023)

Conjunto de ferramentas para trabalhar com dados e incertezas.

Critérios de avaliação

10/10/2022: Prova 1 (P1)

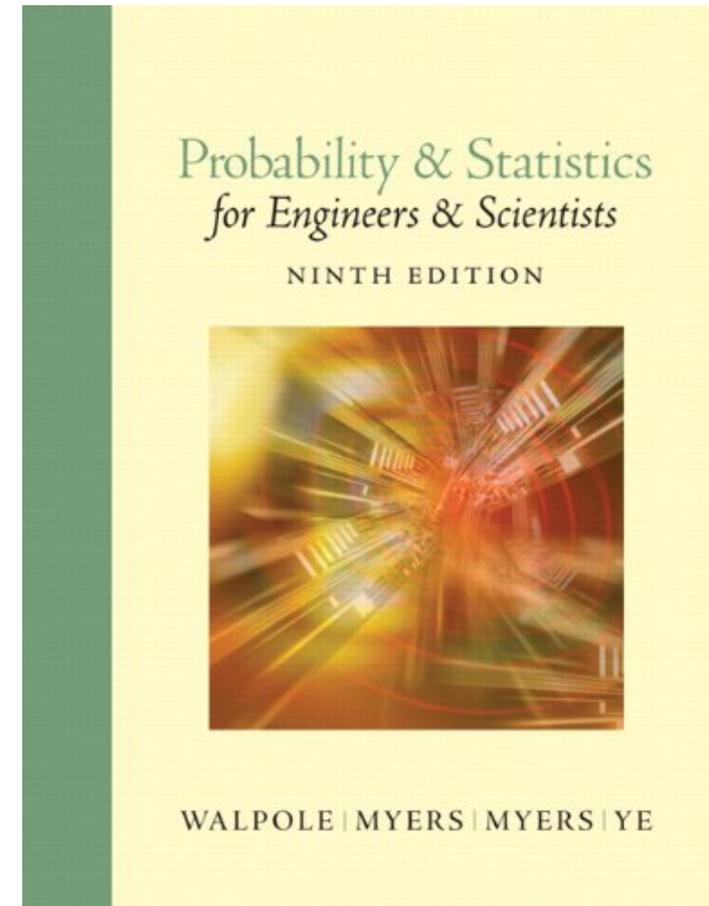
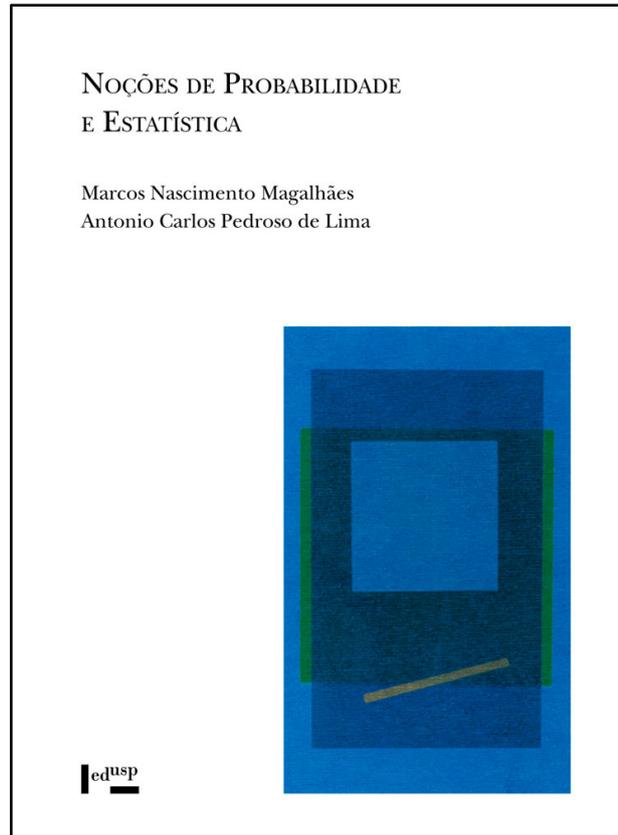
29/11/2022: Prova 2 (P2)

$$\text{Média final} = 0.8 \left[\frac{P1 + P2}{2} \right] + 0.2 \times \text{Média listas}$$

Não haverá prova substitutiva.

Data da recuperação: Uma semana antes do início do semestre.
Critérios para a Recuperação: consulte o JúpiterWeb.

Bibliografia





Pandas



Sobre as listas de exercícios e estudos

Listas semanais ou quinzenais baseadas no que foi apresentado em aula.

Verifiquem o e-mail e o site da disciplina periodicamente!

Haverá exercícios teóricos e práticos.

Mantenham o ritmo constante na resolução dos exercícios.

Dúvidas: enviem e-mail com a sigla da disciplina (SME0320) no assunto.

Uma aula antes da prova será dedicada a sanar dúvidas do conteúdo e listas.

Regras para as aulas

USP exige ao menos 70% de presença nas aulas.

Posso fazer chamada a qualquer momento durante a aula.

Questionamentos sobre o assunto da aula podem ser feitos a qualquer momento.

Algumas regras para as provas

Material permitido durante as provas:

O óbvio: lápis, caneta, lapiseira, borracha, régua.

Calculadora **própria**.

Resposta final à caneta.

Não é permitido o uso de celular.

Por que estudar estatística?

"estudo encontra"     

Videos Imagens Shopping Noticias Livros Maps Voos Finanças

Aproximadamente 204.000 resultados (0,28 segundos)

 **cnnbrasil.com.br**
<https://www.cnnbrasil.com.br> > Saúde

Estudo encontra novas pistas sobre por que pessoas com 80 ...
15 de jul. de 2023 — **Estudo encontra** novas pistas sobre por que pessoas com 80 anos têm cérebros melhores. Os chamados "SuperAgers" apresentaram mais massa ...

 **globo.com**
<https://g1.globo.com> > amazonas > noticia > 2023/07/17

Estudo encontra plástico no intestino de 6 espécies de peixes ...
17 de jul. de 2023 — **Estudo encontra** plástico no intestino de 6 espécies de peixes no Amazonas. Entre os peixes analisados pela pesquisa estão o tambaqui, o jaraqui ...

 **ebc.com.br**
<https://agenciabrasil.ebc.com.br> > geral > noticia > estu...

Estudo encontra dez novas espécies de leguminosas
26 de jan. de 2023 — **Estudo encontra** dez novas espécies de leguminosas. Grande parte delas está ameaçada de extinção, alertam pesquisadores.

 **abril.com.br**
<https://veja.abril.com.br> > comportamento > estudo-en...

Estudo encontra evidência mais antiga de uso do fogo para ...
19 de mai. de 2023 — **Estudo encontra** evidência mais antiga de uso do fogo para cozinhar. Lareiras pré-históricas encontradas perto de Madri datam de cerca de ...

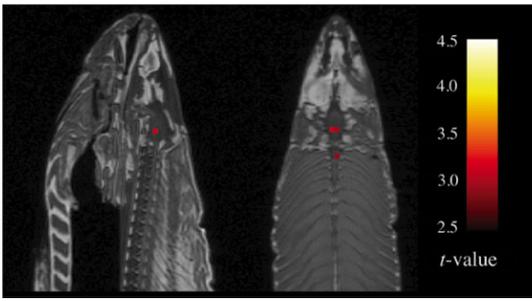
 **usp.br**
<https://jornal.usp.br> > Atualidades

Estudo encontra associação entre tipo sanguíneo e risco de ...
25 de set. de 2022 — **Estudo encontra** associação entre tipo sanguíneo e risco de AVC precoce. Ainda não é possível afirmar que pessoas de determinado grupo ...

Scanning Dead Salmon in fMRI Machine Highlights Risk of Red Herrings

Neuroscientist Craig Bennett purchased a whole Atlantic salmon, took it to a lab at Dartmouth, and put it into an fMRI machine used to study the brain. The beautiful fish was to be the lab's test object as they worked out some new methods. So, as the fish sat in the scanner, they showed it [...]

SAVE



Neuroscientist Craig Bennett purchased a whole Atlantic salmon, took it to a lab at Dartmouth, and put it into an fMRI machine used to study the brain. The beautiful fish was to be the lab's test object as they worked out some new methods.

P-hacking

8 línguas

Artigo Discussão

Ler Editar Ver histórico Ferram

Origem: Wikipédia, a enciclopédia livre.

P-hacking é um termo amplo usado em [pesquisa científica](#) para descrever vários tipos de manipulação comumente empregados na [análise de dados](#) que levam a resultados estatisticamente significativos mas equivocados ou enganosos.

Os testes convencionais de [significância](#) são baseados na probabilidade de que um resultado particular foi obtido puramente por sorte, isso é, não há uma relação real entre os resultados. Dessa forma, eles necessariamente aceitam risco de conclusões equivocadas. Esse nível de risco é chamado de *significância* (α). Quando vastas quantidades de testes são executados ao mesmo tempo, alguns desses produzem falsos resultados desse tipo, portanto 5% das hipóteses escolhidas ao acaso são estatisticamente significantes a um valor α de 5%. Quando hipóteses suficientes são testadas, é quase certo que algumas delas serão estatisticamente significante ainda que ilusórias, visto que praticamente todo conjunto de dados com qualquer grau de aleatoriedade pode conter correlação espúrias. Se não tomarem cuidado, pesquisadores que usam técnicas de [mineração de dados](#) podem ser facilmente enganado esses resultados.

Estatística

Estatísticos · Notação · Publicações de estatística · Por Categoria

Estatística – não apenas na imprensa

Hidrologia: Holanda, Séc. X, o mar avança sobre as cidades. Barragens e diques devem ser suficientemente altos para a maioria das enchentes, mas ao mesmo tempo o custo não deve ser caro.

Seguros: Dado o seu histórico de direção, informação do veículo – qual a franquia apropriada?

Testes clínicos: Uma droga é testada em 100 pacientes; 56 foram curados e 44 não mostraram melhoras. Podemos dizer que a droga é eficiente?



Estatística – não apenas na imprensa

O que é comum aos exemplos anteriores?

Aleatoriedade!

Perguntas associadas a esses problemas:

- . Noção de média (“franquia **apropriada** (...)”).
- . Quantificação de chance (“a **maioria** das enchentes (...)”).
- . Significância, variabilidade, etc.

Probabilidade

Probabilidade – disciplina que estuda fenômenos aleatórios.

Em algumas situações, o processo físico é conhecido: dados, cartas, roletas, moedas (não “viciadas”), ...



Exemplos

Jogando 1 dado:

- Maria ganha R\$ 1 se # for ≤ 3 .
- João ganha R\$ 2 se # for ≤ 2 .

Quem você preferiria ser, João ou Maria?

Jogando 2 dados:

- Escolha um número de 2 a 12.
- Ganha R\$ 100 se a soma dos dados for o número escolhido.

Qual número escolher?

$$100 \times \Pr\{X + Y = Z\} = ?$$

Processo físico é bem conhecido: 1/6 de chance de cada lado, e dados são independentes. Podemos deduzir a probabilidade e o valor esperado. Isto é teoria de probabilidades.

Estatística e modelagem

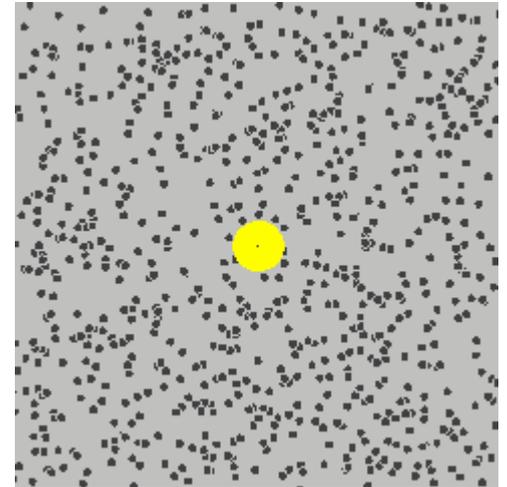
E problemas mais complicados? Precisamos estimar parâmetros dos dados. **Isto é estatística.**

Aleatoriedades do mundo real: (pessoa aleatória, dado viciado, erros na medida, ...)

Problema determinístico, porém complexo:
modelagem estatística!

Problema complicado “=” Processo simples + ruído aleatório.

Movimento Browniano



$$\frac{dx}{dt} = f(x) + \zeta(t)$$

↘ ruído

Estatística e modelagem

E problemas mais complicados? Precisamos estimar parâmetros dos dados. **Isto é estatística.**

Aleatoriedades do mundo real: (pessoa aleatória, dado viciado, erros na medida, ...)

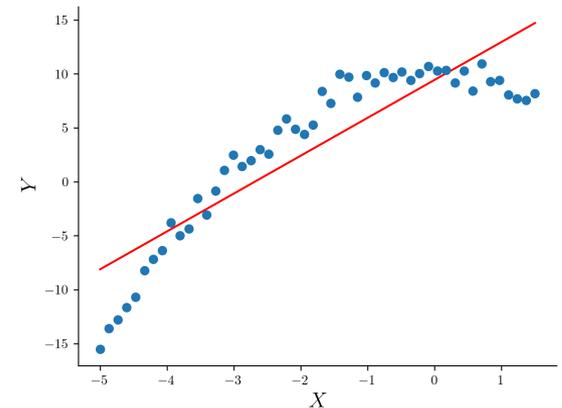
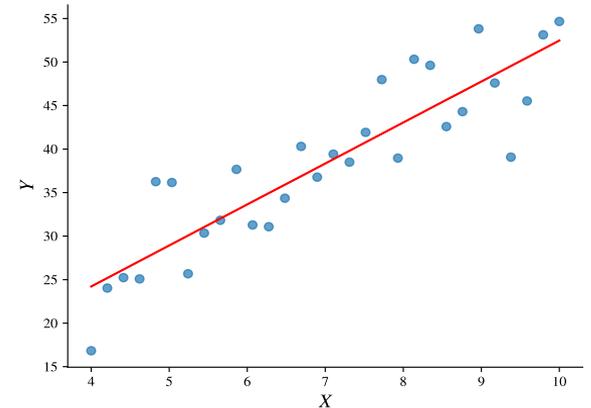
Problema determinístico, porém complexo:
modelagem estatística!

Problema complicado “=” Processo simples + ruído aleatório.

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

→ ruído, variável aleatória.

(boa) Modelagem: escolha de um processo simples e da distribuição do ruído (escolhas plausíveis!)



Estatística e modelagem

E problemas mais complicados? Precisamos estimar parâmetros dos dados. **Isto é estatística.**

Aleatoriedades do mundo real: (pessoa aleatória, dado viciado, erros na medida, ...)

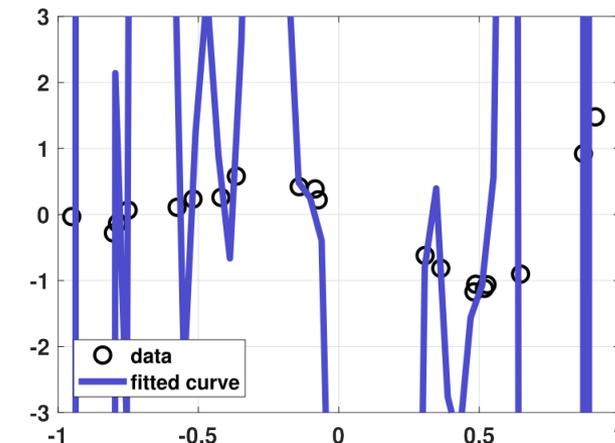
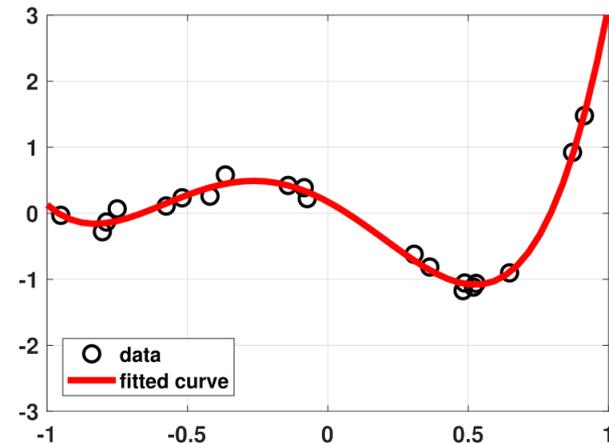
Problema determinístico, porém complexo:
modelagem estatística!

Problema complicado “=” Processo simples + ruído aleatório.

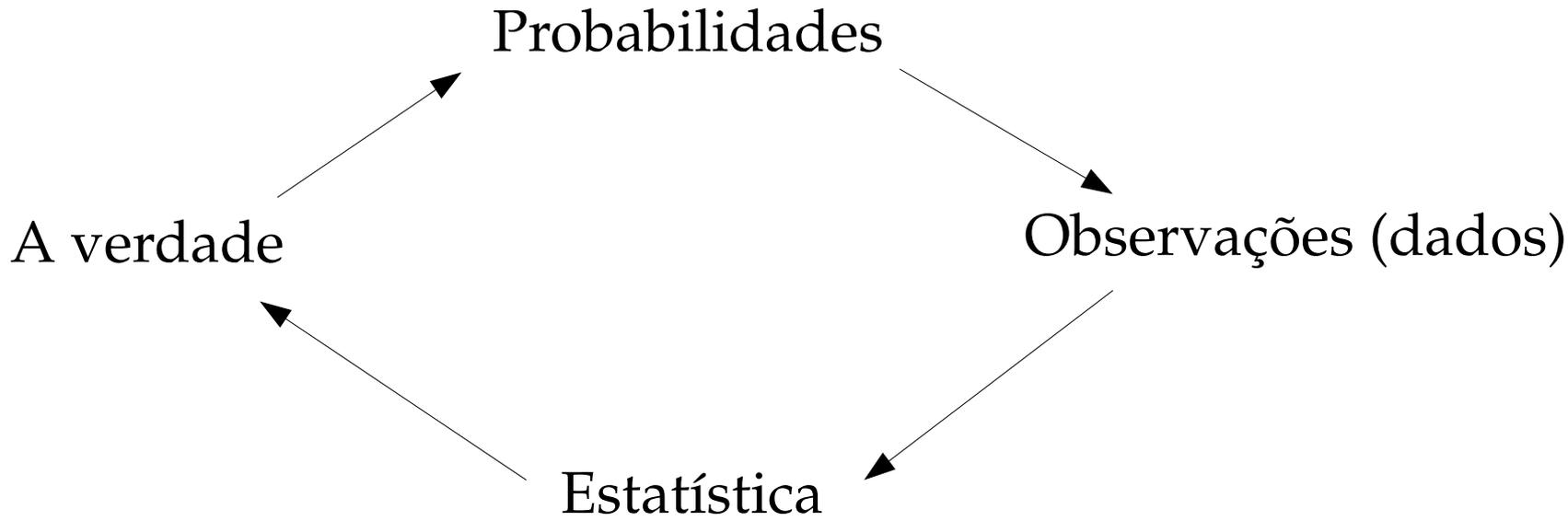
$$Y = \beta_0 + \beta_1 X + \varepsilon$$

→ ruído, variável aleatória.

(boa) Modelagem: escolha de um processo simples e da distribuição do ruído (escolhas plausíveis!)



Estatística vs. probabilidades



Estatística vs. probabilidades

Probabilidades: Estudos anteriores mostraram que a droga X foi 80% eficaz. Portanto, esperamos que, num estudo com 100 pacientes, na média 80 deles serão curados, e 65 serão com 99.99% de chance.

p

Estatística: Observamos que 78 pacientes de um total de 100 foram curados. Concluimos que, com 95% de confiança, outros estudos verificarão entre 69.88% e 86.11% dos pacientes serão curados.

\hat{p}

Correlação vs. causalidade

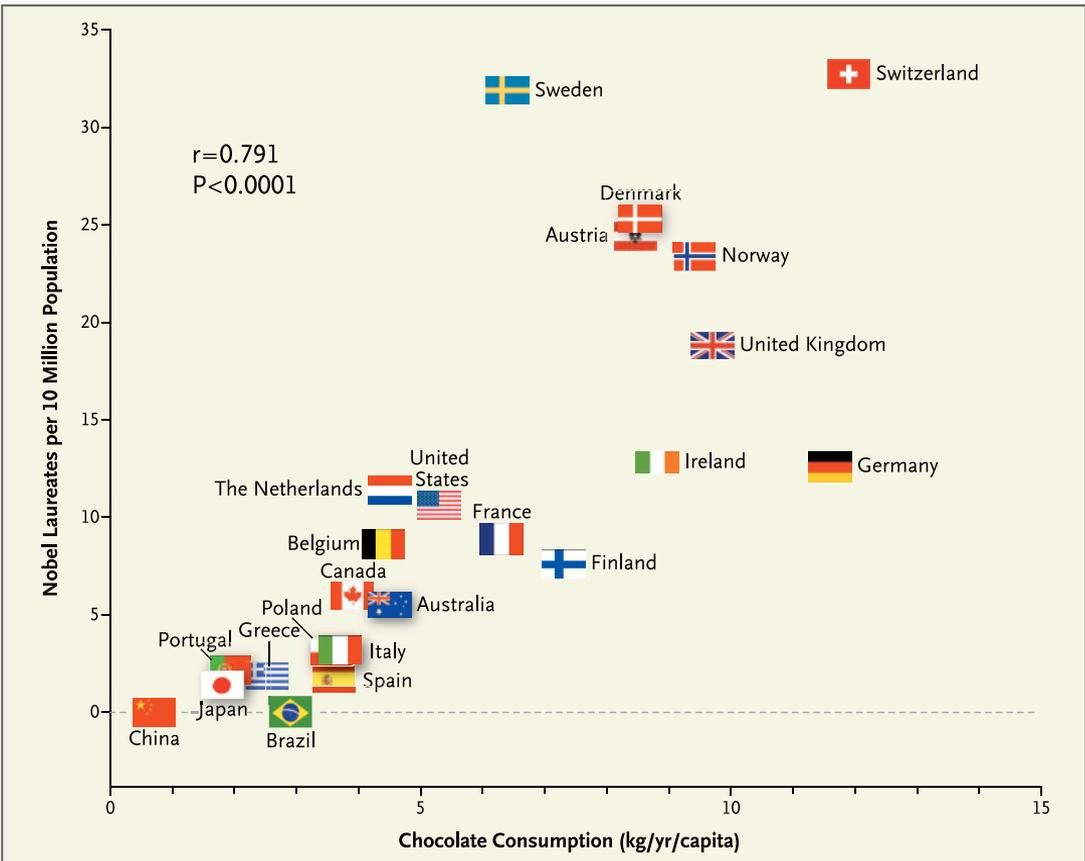
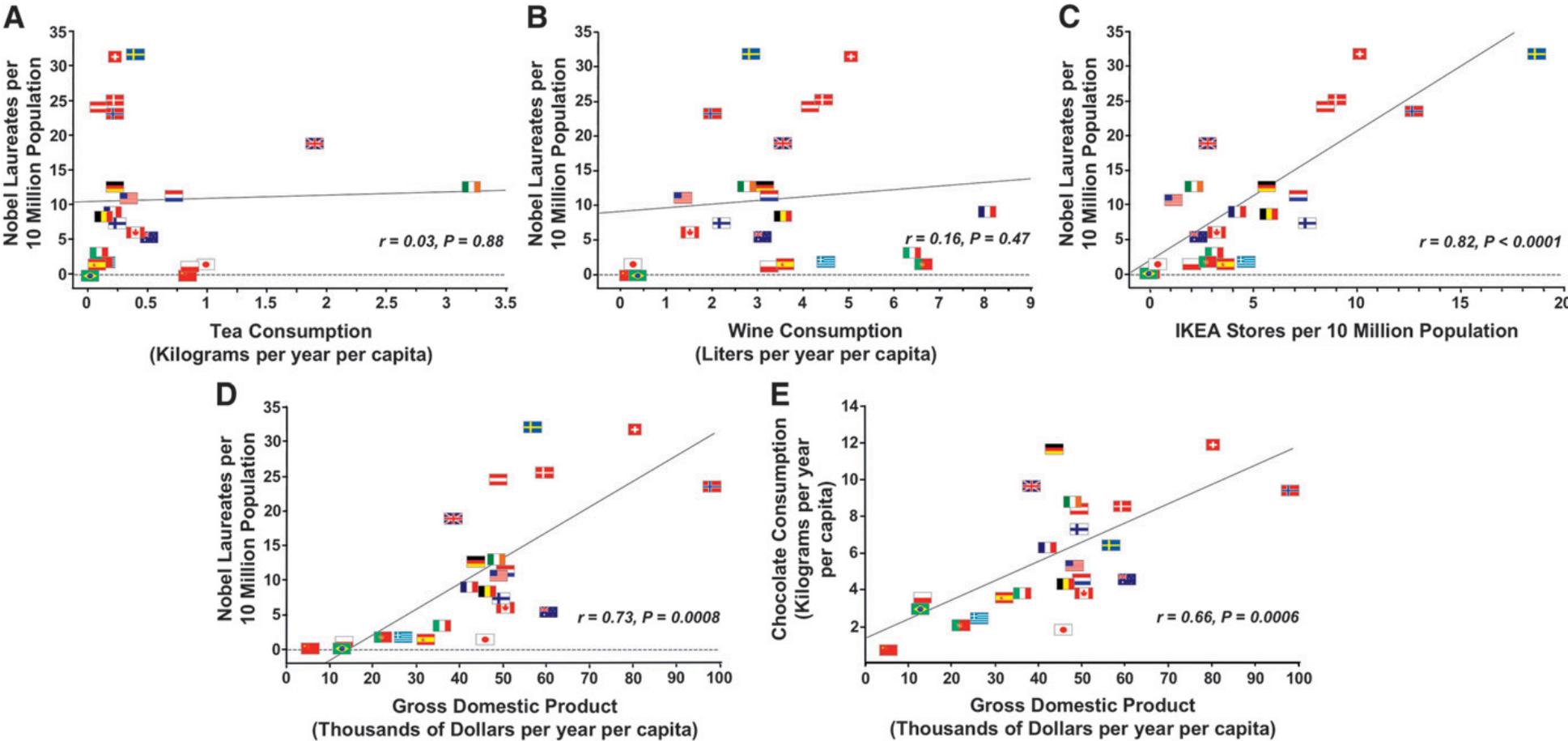
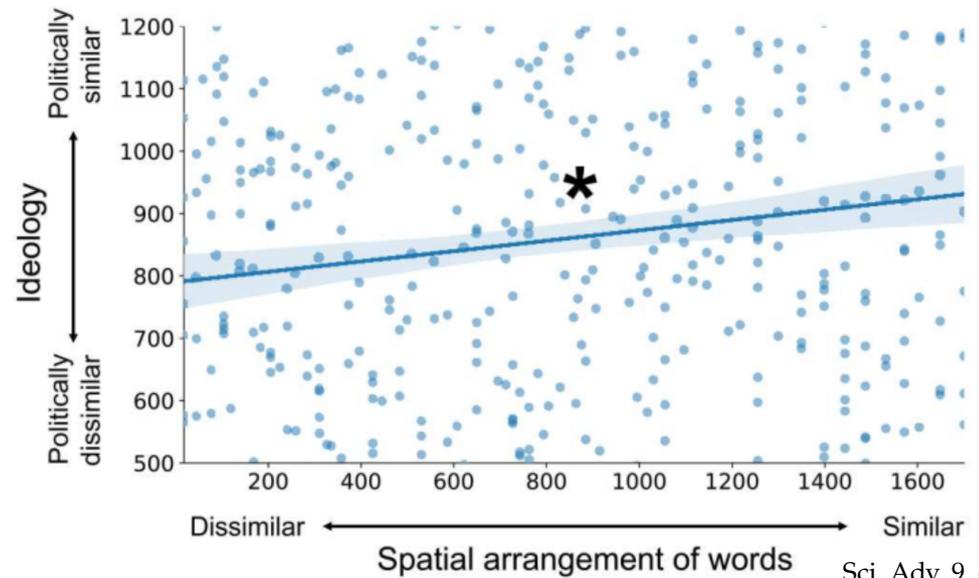


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Correlação vs. causalidade



Como avaliar um modelo estatístico?



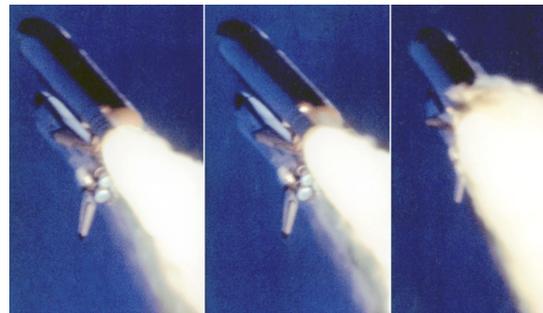
Sci. Adv. 9, eabq5920 (2023)

Aplicações

Análise de dados: Planejamento de experimentos, verificação de hipóteses/teorias, investigação de acidentes, etc.

Inteligência artificial: Previsão de demandas, controle de tráfego aéreo, ...

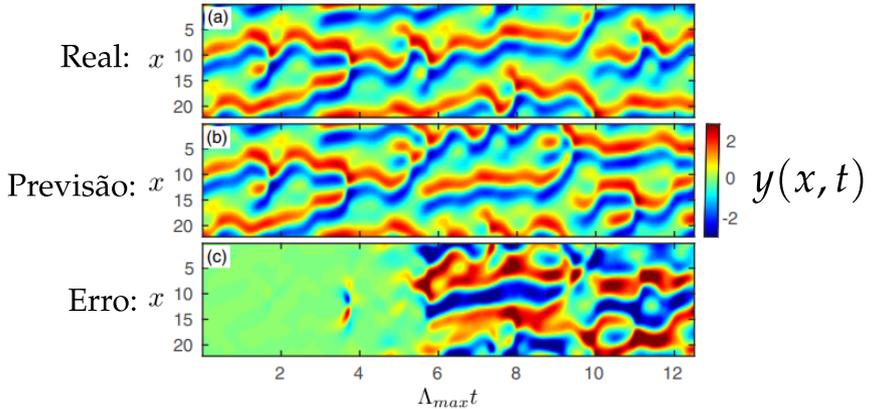
Previsão de fenômenos físicos.



Challenger Shuttle, 1986



<https://www.flightradar24.com/>



Pathak *et al.* Physical Review Letters (2020).

Machine Learning for Fluid Mechanics

Annual Review of Fluid Mechanics
 Vol. 52:477-508 (Volume publication date January 2020)
 First published as a Review in Advance on September 12, 2019
<https://doi.org/10.1146/annurev-fluid-010719-060214>

Steven L. Brunton,¹ Bernd R. Noack,^{2,3} and Petros Koumoutsakos⁴

Programa resumido



1. Análise exploratória (Intro. Python, *scatter plots*, histogramas, box plots, etc.).

2. Introdução a teoria de probabilidades.

3. Variáveis aleatórias discretas e contínuas, uni- e n -dimensionais.

4. Distribuições de probabilidades.

5. Funções de variáveis aleatórias.

6. Esperança matemática.

7. Momentos.

8. Covariância e correlação.

9. Teorema do limite central.

10. Regressão linear.

$$\Pr\{A|B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}}$$

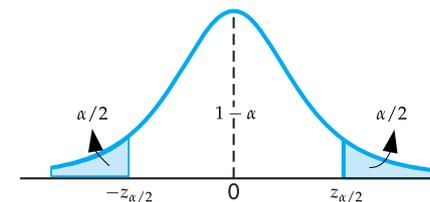
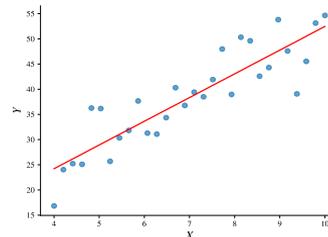
$$X : \Omega \rightarrow \mathbb{R}$$

$$\Pr\{X = x\} = P(X = x) = P_X(x)$$

$$f = f(X), f = f(X, Y)$$

$$\mathbb{E}[X^m], m \in \mathbb{N}.$$

$$\text{cov}(X, Y), \text{corr}(X, Y)$$



Próxima aula – Breve introdução ao *Python*



Por que *Python*?

1. Óbvio: automatização dos cálculos.

Próxima aula – Breve introdução ao *Python*



Por que *Python*?

1. Óbvio: automatização dos cálculos.

2. Microsoft Excel:
Não faça isso.

Comment | [Open Access](#) | [Published: 23 August 2016](#)

Gene name errors are widespread in the scientific literature

[Mark Ziemann](#), [Yotam Eren](#) & [Assam El-Osta](#) 

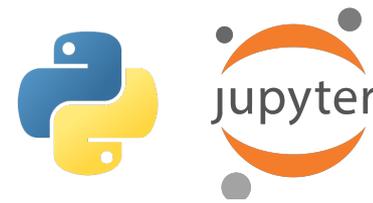
[Genome Biology](#) **17**, Article number: 177 (2016) | [Cite this article](#)

149k Accesses | **85** Citations | **3014** Altmetric | [Metrics](#)

Abstract

The spreadsheet software Microsoft Excel, when used with default settings, is known to convert gene names to dates and floating-point numbers. A programmatic scan of leading genomics journals reveals that approximately one-fifth of papers with supplementary Excel gene lists contain erroneous gene name conversions.

Próxima aula – Breve introdução ao *Python*



Por que *Python*?

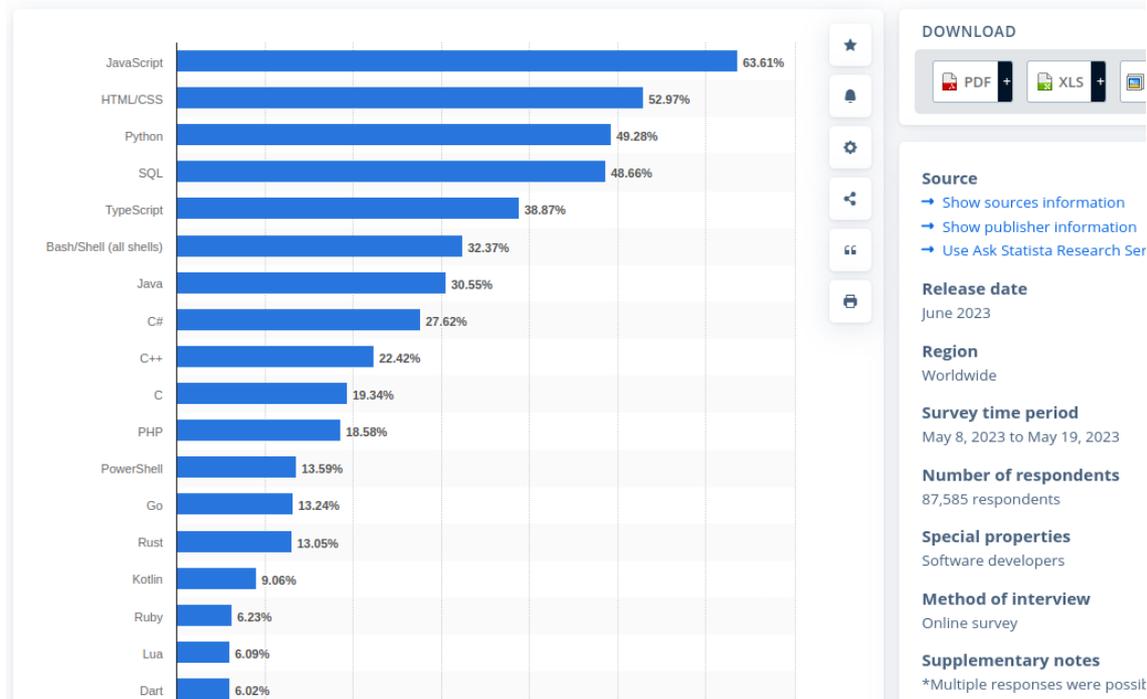
1. Óbvio: automatização dos cálculos.

2. Microsoft Excel:
Não faça isso.

3. Software livre:
altamente documentado
pela comunidade.

Technology & Telecommunications › Software

Most used programming languages among developers worldwide as of 2023



statista.com

Próxima aula – Breve introdução ao *Python*



Por que *Python*?

1. Óbvio: automatização dos cálculos.

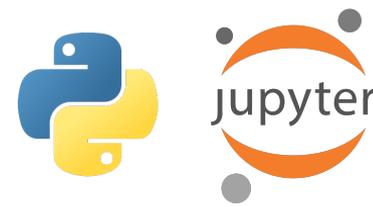
2. Microsoft Excel:
Não faça isso.

3. Software livre:
altamente documentado
pela comunidade.

4. Vários módulos.

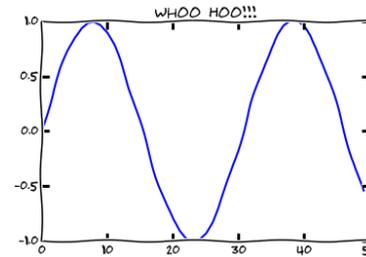


Próxima aula – Breve introdução ao *Python*



Por que *Python*?

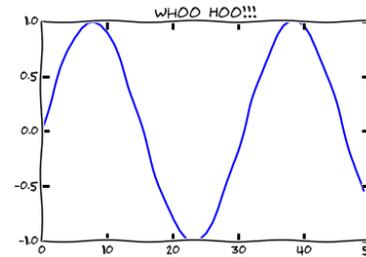
1. Óbvio: automatização dos cálculos.
2. Microsoft Excel: Não faça isso.
3. Software livre: altamente documentado pela comunidade.
4. Vários módulos.



Próxima aula – Breve introdução ao *Python*

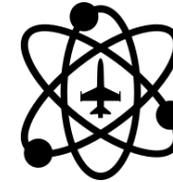
Por que *Python*?

1. Óbvio: automatização dos cálculos.
2. Microsoft Excel: Não faça isso.
3. Software livre: altamente documentado pela comunidade.
4. Vários módulos.



pypi package 0.2.5 build passing

Aircraft Design Recipes in Python



scikit-aero
Aeronautical engineering
calculations in Python

Próxima aula – Breve introdução ao *Python*



Por que *Python*?

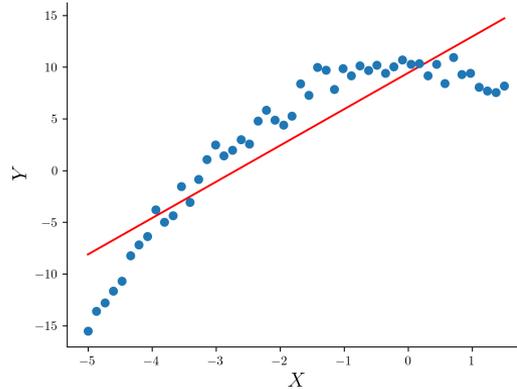
1. Óbvio: automatização dos cálculos.
2. Microsoft Excel:
Não faça isso.
3. Software livre:
altamente documentado
pela comunidade.
4. Vários módulos.
5. Linguagem de
Programação.

Próxima aula – Breve introdução ao *Python*



Por que *Python*?

1. Óbvio: automatização dos cálculos.
2. Microsoft Excel: Não faça isso.
3. Software livre: altamente documentado pela comunidade.
4. Vários módulos.
5. Linguagem de Programação.



Próxima aula – Breve introdução ao *Python*



Por que *Python*?

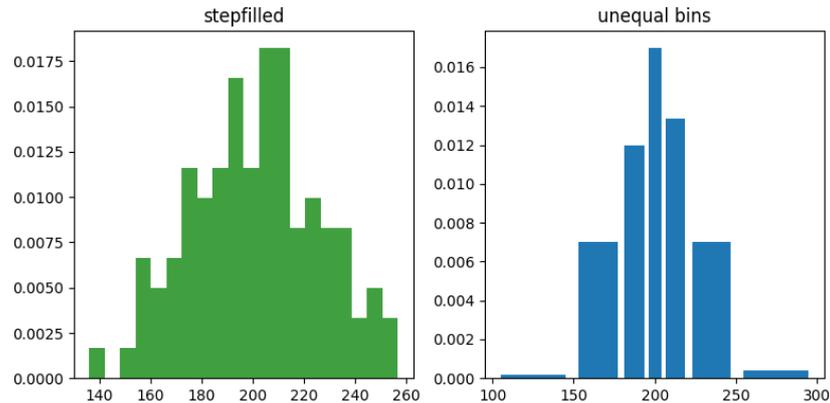
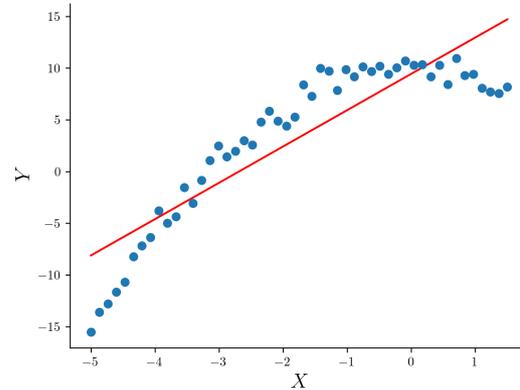
1. Óbvio: automatização dos cálculos.

2. Microsoft Excel:
Não faça isso.

3. Software livre:
altamente documentado
pela comunidade.

4. Vários módulos.

5. Linguagem de
Programação.

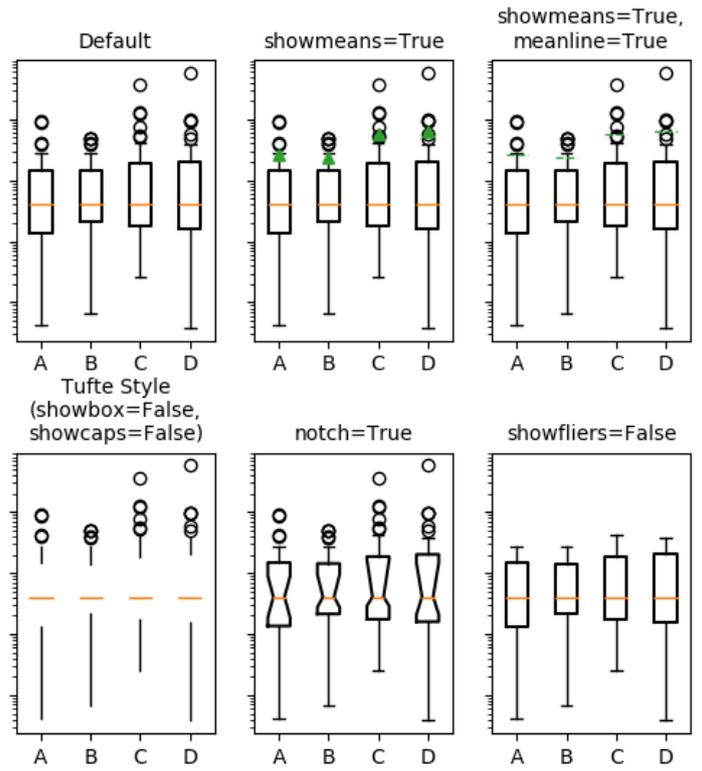
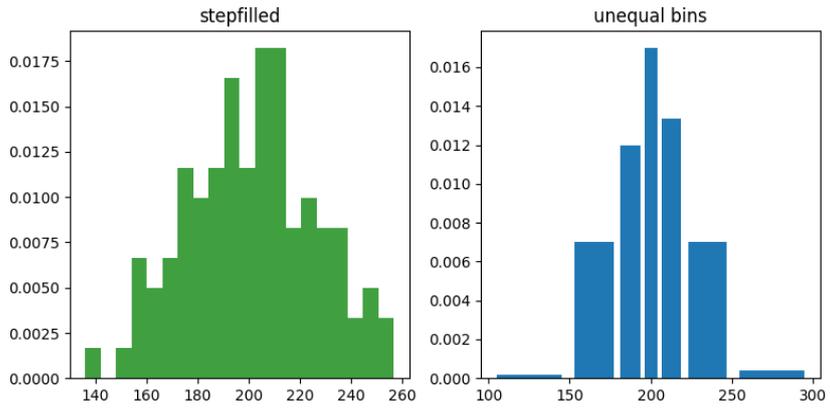
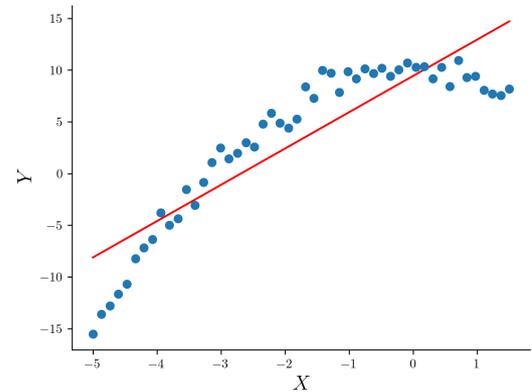


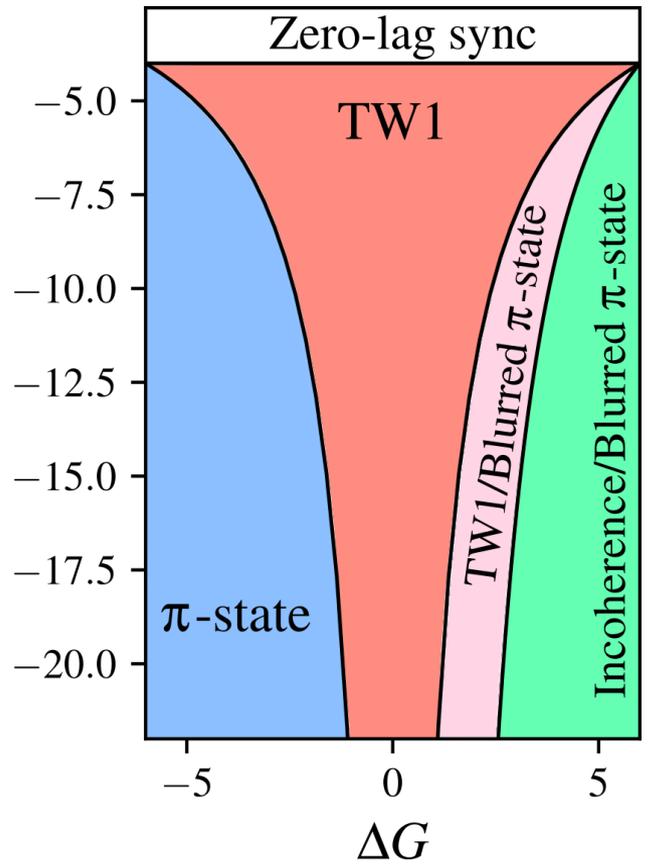
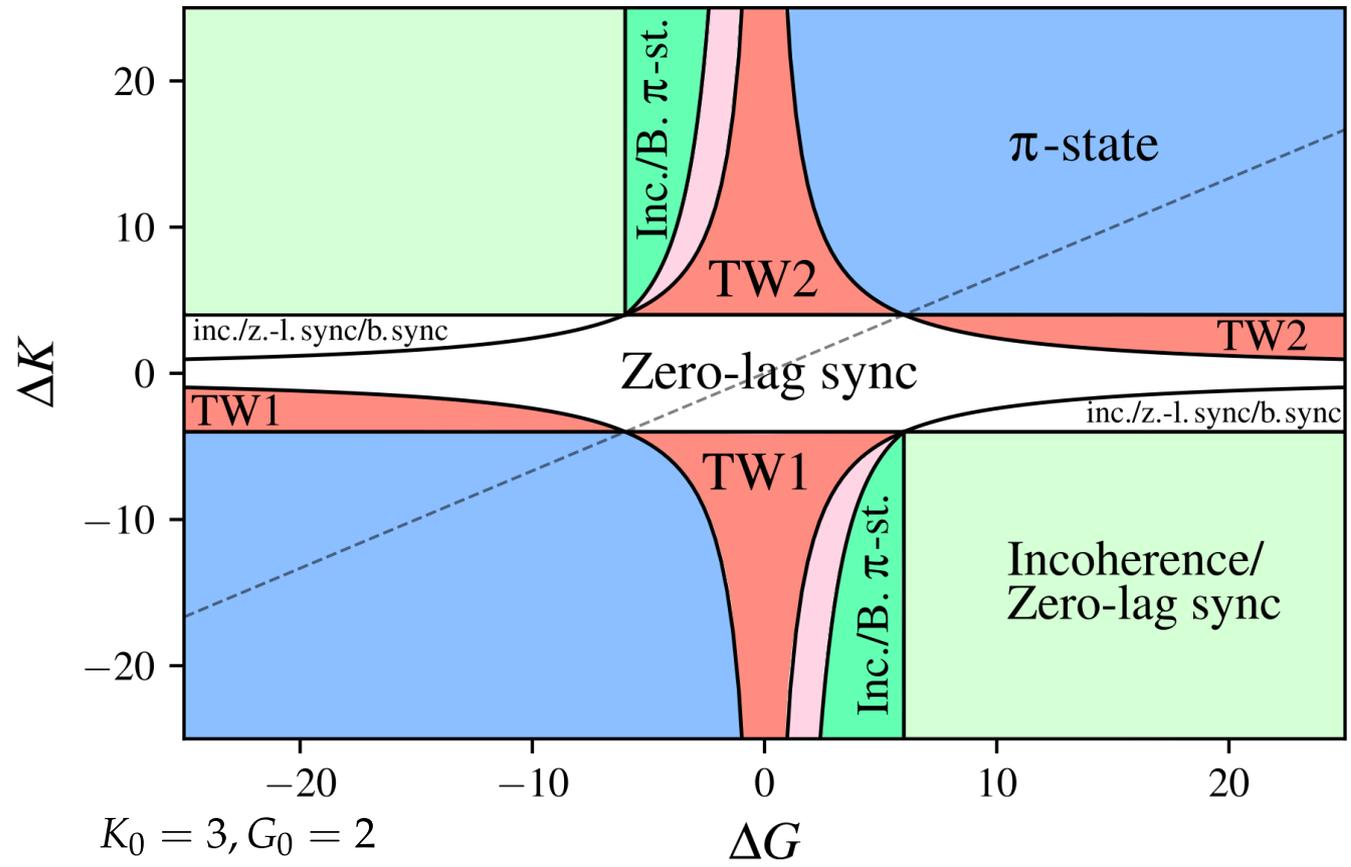
Próxima aula – Breve introdução ao *Python*



Por que *Python*?

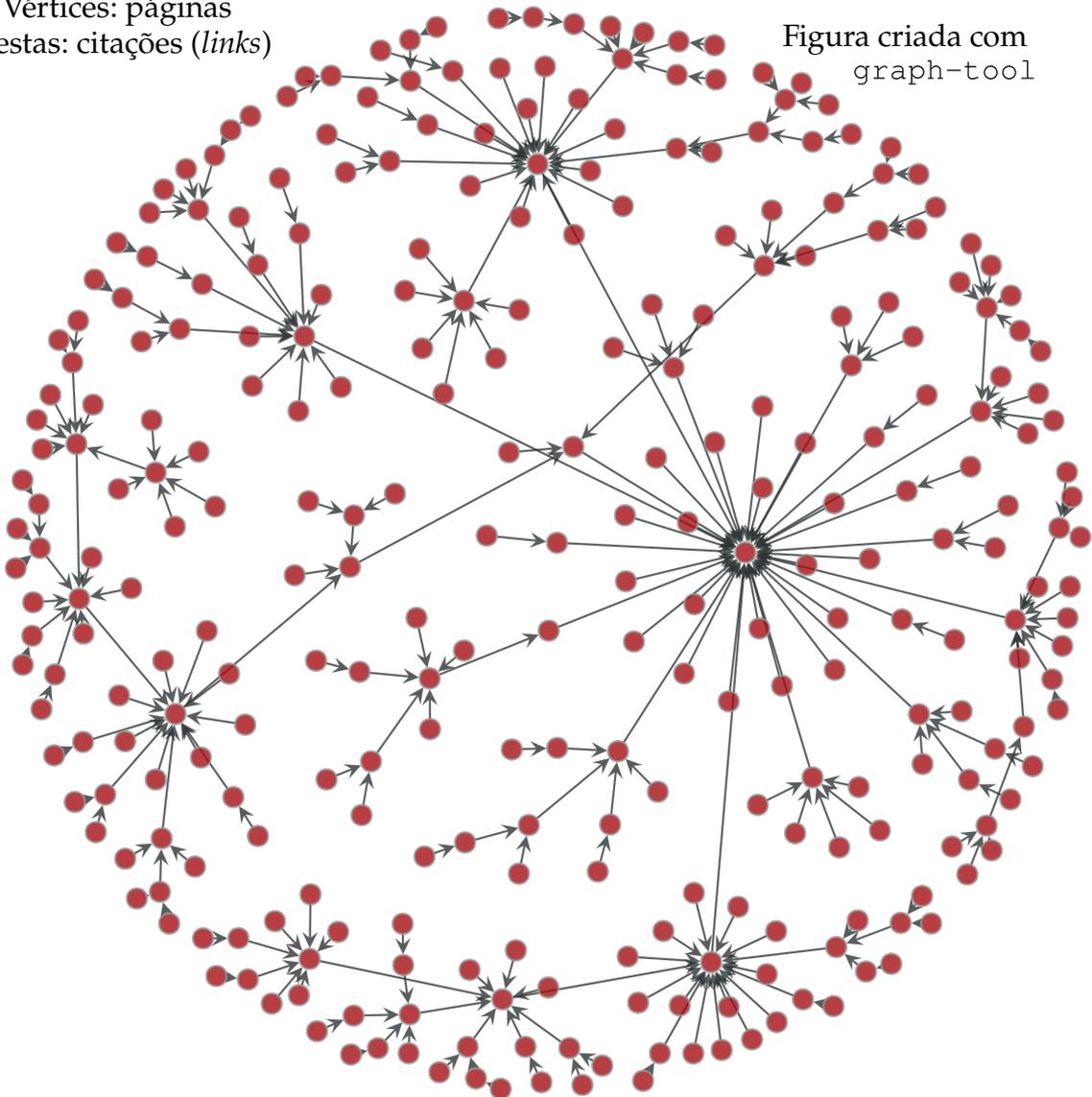
- 1. Óbvio: automatização dos cálculos.
- 2. Microsoft Excel: Não faça isso.
- 3. Software livre: altamente documentado pela comunidade.
- 4. Vários módulos.
- 5. Linguagem de Programação.

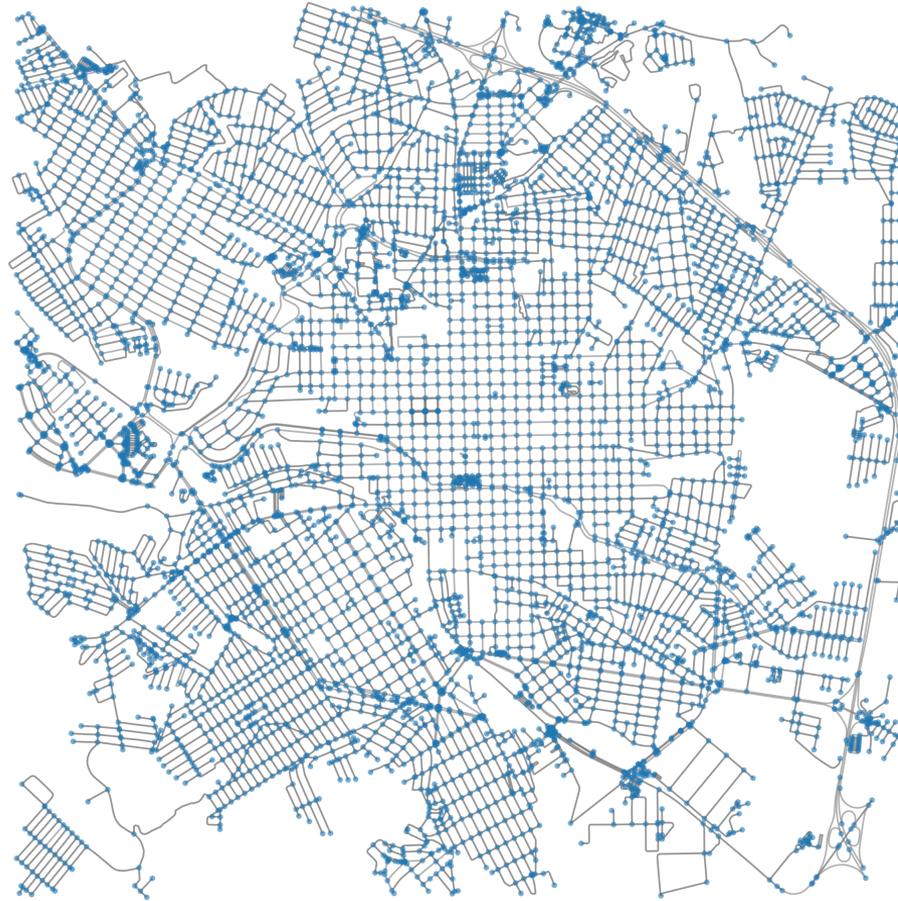




Vértices: páginas
Arestas: citações (*links*)

Figura criada com
graph-tool





Feito com osmnx.

