

Um passeio pela
ciências estatísticas...
...e pela vida de um pesquisador da área.

Hedibert Freitas Lopes

www.hedibert.org

hedibertfl@insper.edu.br

hedibert@gmail.com

Maio de 2024

Mudamos para o **Rio de Janeiro** - Dez/1970



Fiz o **ensino fundamental** na escola Raul Pederneiras, na **Cidade Alta**, Zona Norte do Rio de Janeiro, onde morei entre os 6 e os 14 anos.



Maconha, revolver, tiros, gente morta com várias perfurações, batidas policiais estavam no nosso dia-a-dia....

Ainda assim dava para ser criança e brincar nos “morrinhos”, jogar “pé-na-lata” e “pique-ajuda”.

Curiosidade: A **Cidade Alta** foi usada em 2002 para representar a **Cidade de Deus** nos anos 60/70/80 no filme com o mesmo nome (dirigido por Fernando Meirelles e baseado no livro de Paulo Lins).

Quem mora no Méier não bobéia!

Dos 14 aos 28 anos (1982 - 1996)

Morei no grande Méier (Cachambi, Camarista, Engenho de Dentro)

Fiz o ensino médio numa escola técnica na Tijuca

Joguei basquete no infanto-juvenil do Botafogo

Graduação e mestrado em Estatística na UFRJ

Fui pesquisador em estatística e econometria do IPEA

Professor auxiliar de estatística da UFF (1982 – 1996)

Professor assistente e adjunto de estatística da UFRJ (1996 - 2003)



A Estatística me tornou visível!

O interesse pela matemática foi se formando durante os anos do ensino médio: lidar com matrizes era uma fixação 😊

Passei no vestibular para matemática na Universidade Santa Úrsula.
Privada e “meia-bomba”!? Tô fora!!!

Fiz um ano de “pré-vestibular” e passei pra Matemática na UFRJ.
Gabaritei as 20 questões da prova do vestibular no final de 1986.

Matemática → Estatística

Logo no 1o semestre me encantei com a estatística através das aulas do Professor Paulo Bravo. Fiz com ele mais duas matérias: **Regressão** e **Controle de Qualidade**.

Sempre tive o privilégio de encontrar professores interessados no meu progresso.

Nota 1: Graduação tempo integral: aulas de 7h30 até 17h no Fundão.

Nota 2: Trabalhava entre 18h-24h como “digitador” no centro do Rio.

Nota 3: Minha filha nasceu no 3o semestre e eu tinha 19,5 anos.

PhD em Estatística e Teoria da Decisão

No início de 1996 fui aprovado para fazer doutorado em estatística na **Duke University**, na Carolina do Norte, EUA.

Em agosto de 1997 viajei pela primeira vez para fora do Brasil.

Curiosidade: Perdi minha conexão em NYC e fiquei 2 horas tentando falar com a telefonista de um telefone público para fazer uma ligação a cobrar para um brasileiro que ia me buscar no aeroporto de Durham, cidade sede da universidade 😊

Universidade de Chicago (2003-2013)

Depois de um período de 3 anos (Junho de 2000 – Agosto de 2003) como Professor do Departamento de Métodos Estatísticos da UFRJ, me mudei para Chicago.

Fui aceito para uma posição na “Graduate School of Business” (GSB) da Universidade de Chicago (UC)

UC é uma das 10 melhores universidades do mundo.

Chicago GSB é uma das 3 melhores BS do mundo.

Janelas de casa e do trabalho em Hyde Park

Rockefeller Memorial Chapel

Regents Park Apartments



Basquete: Botafogo, Duke & Chicago Bulls



Basquete no ensino médio



**Basquete no infanto-juvenil
Clube de Regatas Botafogo**



**Fui a vários jogos do
Chicago Bulls**

&

**Do Duke Blue Devils
também**

Samba: Grupo **Variáveis Instrumentais**

Maria



Maria: Bióloga brasileira com doutorado em bioquímica pela USP. Conheci em Chicago fazendo pós-doc em genética com uma bolsa da Pew Foundation. Nos conhecemos no dia 29/04/2006 e no dia 31/08/2006 começamos uma trajetória conjunta.

INSPER (set/2013 – até hoje)

Eu e Maria decidimos no final de 2011 voltar para o Brasil.

Em maio/2012 ela passou num concurso pra USP com 1 vaga e 50 candidatos.

Vimos imediatamente em julho/2012.

Ago-Nov/2012 fui visitante do Departamento de Estatística do IME-USP.

Passei o chapéu procurando emprego na EESP-FGV e no Insper.

O Projeto Insper parecia mais interessante e comecei lá em Set/2013.

Dados: $\{y_1, \dots, y_n\}$

Modelo: $y_i = \beta_0 + \frac{1}{\sigma} \sum_{j=1}^p \beta_j y_{ij} + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$

AR(P)

$$y = \begin{pmatrix} y_{11} \\ \vdots \\ y_{n1} \end{pmatrix}, X = \begin{pmatrix} 1 & y_{12} & \dots & y_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & y_{n2} & \dots & y_{np} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

$$y = X\beta + \epsilon, \epsilon \sim N(0, \sigma^2 I)$$

Prior: $\beta \sim N(b_0, B_0)$
 $\sigma^2 \sim IG(c_0, d_0)$
 b_0, B_0, c_0, d_0 hiperparâmetros

\Rightarrow Condições completas são "Gibbs"

$$(\beta | \sigma^2, y_1, \dots, y_n) \sim N \left[\left(B_0^{-1} + X^T X \sigma^{-2} \right)^{-1} \left(B_0^{-1} b_0 + X^T y \sigma^{-2} \right), \left(B_0 + X^T X \sigma^{-2} \right)^{-1} \right]$$

$$(\sigma^2 | y_1, \dots, y_n) \sim IG \left(c_0 + \frac{n-p}{2}, d_0 + \frac{(y - X\beta)^T (y - X\beta)}{2} \right)$$

Prior Normal-Gamma para β_j

$$E(\beta | \gamma, \lambda) = 0$$

$$V(\beta | \gamma, \lambda) = 2\sigma\lambda^2$$

$$\beta \sim NG(\lambda, \sigma)$$

$$p(\beta | \lambda, \gamma) = \frac{1}{\sqrt{\pi}^p 2^{\gamma-1/2} \lambda^{\gamma+1/2} \Gamma(\gamma)} |\beta|^{\gamma-1/2} K_{\gamma-1/2} \left(\frac{|\beta|}{\lambda} \right)$$

$\beta \in \mathbb{R}^p, \gamma, \lambda > 0$

Função Bessel modificada do tipo ν

Ass Exponencial Dupla

$$\beta \sim NG(\lambda, 1) \equiv \text{Laplace}(\lambda)$$

ALGORITMO METROPOLIS-HASTINGS

Quero amostrar de $p(\theta)$

Só sei amostrar de $q(\theta | \theta^{(i-1)})$

$$\alpha = \min \left\{ 1, \frac{p(\theta^*)/q(\theta^* | \theta^{(i-1)})}{p(\theta^{(i-1)})/q(\theta^{(i-1)} | \theta^*)} \right\}$$

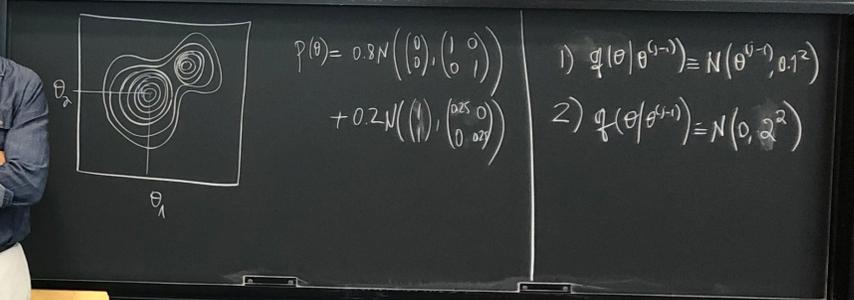
$\theta^* \sim q(\theta | \theta^{(i-1)}) \Rightarrow \theta^{(i)} = \theta^*$ com probabilidade α

$\phi_1 \sim \phi_2 \sim \dots$ com probabilidade $1-\alpha$

$\theta^{(i)} = \theta^{(i-1)}$

$\Rightarrow \theta^{(1)}, \dots, \theta^{(M)}$

$\sim P(\theta)$



Com Paulo Furquim, Carlos Melo, Naércio Menezes e outros amigos musicais do Insper



Coordenador do Núcleo de Ciências de Dados e Decisão do INSPER

Bolsista de Produtividade em Pesquisa do CNPq nível 1C

80+ artigos científicos publicados em revistas internacionais de alto impacto:

- Journal of the American Statistical Association
- Journal of Computational and Graphical Statistics
- Journal of Business and Economic Statistics
- American Economic Review

Meus artigos somam 10 mil citações no Google Scholar

Livro: *MCMC: Stochastic Simulation for Bayesian Inference* (4575 citações)

Em 2022 foi eleito ISBA Fellow durante
o *ISBA 2022 World Meeting*

Único brasileiro com essa honoraria

ISBA é a Sociedade Internacional
para Análise Bayesiana

Na foto meu orientador de doutorado
Mike West



Doutores em Estatística e Economia que orientei entre 2017 e 2024

1. Renata Tavanielli

D.Sc. in Statistics (Expected defense for Dec-2024)
Thesis topic: Cholesky-based dynamic copula modeling
Institution: University of Sao Paulo

2. Rafael Campello de Alcantara

Post-doc at UT Austin
D.Sc. in Business Economics (Fev-2024)
Thesis topic: Cutoff-aware BART for estimating Heterogeneous Treatment Effects in Regression Discontinuity Designs
Institution: Insper

3. Igor Ferreira Batista Martins

Post-doc in Econometrics at Örebro, Sweden
D.Sc. in Business Economics (Feb-2024)
Thesis topic: Essays in Bayesian Financial Econometrics
Institution: Insper

4. Bruno do Prado Costa Levy

Itaú Asset Management
D.Sc. in Business Economics (2021)
Thesis topic: Dynamic ordering learning in multivariate forecasting
Institution: Insper

5. Henrique Bolfarine

Professor UT Austin
D.Sc. in Statistics (2021)
Thesis topic: Decoupling Shrinkage and Selection in Gaussian Linear Factor Analysis
Institution: University of Sao Paulo

6. Helton Graziadei

Professor at UFPR
D.Sc. in Statistics (2020)
Thesis topic: Some Bayesian generalizations of the integer-valued autoregressive model
Institution: University of Sao Paulo

7. Paloma Waisman Uribe

Data Science Manager at Nubank
D.Sc. in Statistics (2017)
Thesis topic: Dynamic sparsity on time-varying Cholesky-based covariance matrices
Institution: University of Sao Paulo

Doutores em Estatística, Economia e Engenharia Elétrica que orientei entre 2003 e 2014

8. Samir Warty

Ph.D. in Econometrics and Statistics (2014)

Thesis topic: Sequential Bayesian Learning for Stochastic Volatility with Variance-Gamma Jumps in Returns

Institution: The University of Chicago Booth School of Business

9. Paolo Bonomolo

Ph.D. in Economics (2012)

Thesis topic: Does Inflation Walk on Unstable Paths? A Rational Sunspots Approach

Institution: Università degli studi di Pavia, Italy.

10. Maria Paula Rios Arango

Ph.D. in Econometrics and Statistics (2012)

Thesis title: Essays on Application of Particle Learning in Financial Econometrics

Institution: The University of Chicago Booth School of Business

11. Bruno Lund

D.Sc. in Economics (2009)

Thesis title: Essays in Macroeconomy and Dynamic Term-Structure Models

Institution: Graduate School of Economics, Getulio Vargas Foundation

12. Fernando Ferraz do Nascimento

Assistant Professor of Statistics, Federal University of Piauí

D.Sc. in Statistics (2009)

Thesis title: Bayesian nonparametric approach to extreme value analysis

Institution: Institute of Mathematics, Federal University of Rio de Janeiro, Brazil.

13. Esther Salazar

Researcher FDA, Washington, D.C.

D.Sc. in Statistics (2008)

Thesis title: Spatial Dynamic Factor Analysis

Institution: Institute of Mathematics, Federal University of Rio de Janeiro, Brazil.

14. Ralph dos Santos Silva

Assistant Professor, Department of Statistical Methods, Federal University of Rio de Janeiro

D.Sc. in Statistics (2006)

Thesis title: Skewed Bayesian Models

Institution: Institute of Mathematics, Federal University of Rio de Janeiro, Brazil.

15. Carlos Antonio Abanto-Valle

Assistant Professor of Statistics, Department of Statistical Methods, Federal University of Rio de Janeiro

D.Sc. in Statistics (2005)

Thesis title: Stochastic simulation methods in nonlinear dynamic models: applications in stochastic volatility models

Institution: Institute of Mathematics, Federal University of Rio de Janeiro, Brazil.

16. Cibele Behrens

D.Sc. in Operations Research (2004)

Thesis title: Bayesian Analysis of Extreme Events with Threshold Estimation

Institution: COPPE, Federal University of Rio de Janeiro, Brazil.

17. Edison Americo Huarsaya Tito

Smart Decision Support System Software Development and Consulting

D.Sc. in Electrical Engineer (2003)

Thesis Title: Evolutionary Inference in Adaptive models

Institution: Department of Electrical Engineering, PUC/RJ, Brazil.

Vamos ver um pouco de probabilidade e estatística?

- 1) **Monty Hall:** 1 carro e 2 cabras atrás de 3 portas
- 2) **Problema do aniversário:** Aniversariantes do mesmo dia e mês
- 3) **Falso-positivo & falso-negativo:** COVID-19
- 4) **Quantas hospitalizações a vacinação contra a COVID-19 já preveniu em São Paulo?** Artigo publicado em 27/Agosto/2022.
- 5) **Classificando filmes:** Ajuste linear e polinomial.

Monty Hall

```
premio = c("carro", "cabra", "cabra")
nrep = 10000
draws = matrix(0, nrep, 3)
for (rep in 1:nrep)
  draws[rep,] = sample(premio, size=3, replace=FALSE)

# Suponhamos que você escolha sempre a porta 1

# Ganhar o carro ao decidir não trocar de porta
prob.carro.porta.um = sum(draws[,1]=="carro")/nrep
prob.carro.porta.um

# Ganhar o carro ao decidir trocar de porta
# É o complementar da probabilidade acima
1-prob.carro.porta.um
```

20 simulações

```
[1,] "cabra" "cabra" "carro"
[2,] "carro" "cabra" "cabra"
[3,] "cabra" "carro" "cabra"
[4,] "cabra" "carro" "cabra"
[5,] "cabra" "carro" "cabra"
[6,] "carro" "cabra" "cabra"
[7,] "carro" "cabra" "cabra"
[8,] "cabra" "carro" "cabra"
[9,] "carro" "cabra" "cabra"
[10,] "cabra" "carro" "cabra"
[11,] "carro" "cabra" "cabra"
[12,] "cabra" "cabra" "carro"
[13,] "carro" "cabra" "cabra"
[14,] "cabra" "cabra" "carro"
[15,] "cabra" "cabra" "carro"
[16,] "carro" "cabra" "cabra"
[17,] "carro" "cabra" "cabra"
[18,] "cabra" "carro" "cabra"
[19,] "cabra" "carro" "cabra"
[20,] "cabra" "carro" "cabra"
```

Aniversário

Dia do mês de maio	Aniversariantes
18	4
19	3
20	3
21	5
22	3
23	3
24	1
25	7
26	5
27	1
28	2
29	3
30	4
31	5

Tenho cerca de 1.500 “amigos” no FB.

Com 100% de chance, num universo de 1500 pessoas, pelo menos duas fazem aniversário no mesmo dia.

Não precisamos de probabilidade pra saber disso!

Pergunta: E se ao invés de 1.500 pessoas, estivéssemos numa sala de aula de um curso de graduação do INSPER com 40 alunos (Estatística I ou Introdução à Economia, por exemplo), qual a chance de pelo menos duas pessoas fazerem aniversário no mesmo dia e no mesmo mês?

- (a) 0%
- (b) 10%
- (c) 50%
- (d) 90%
- (e) 100%

Evento A: “Pelo menos 2 pessoas aniversariando no mesmo dia e no mesmo mês do ano”.

Evento A^c: “Nenhuma das n pessoas aniversariam no mesmo dia e no mesmo mês do ano”.

Então o que estamos procurando é $\Pr(A) = 1 - \Pr(A^c)$.

$$\begin{aligned}\Pr(A^c) &= 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{n-1}{365}\right) \\ &= \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n} \\ &= \frac{365!}{365^n (365 - n)!} = \frac{n! \cdot \binom{365}{n}}{365^n}\end{aligned}$$

Usando a fórmula da página anterior

$$\Pr(A) = 1 - \frac{n! \cdot \binom{365}{n}}{365^n}$$

n	Pr(A)
10	11.7%
20	41.1%
23	50.7%
30	70.6%
40	89.1%
50	97.0%
60	99.4%
70	99.9%
75	99.97%

Usando expansão em séries de Taylor

$$\Pr(A) \approx 1 - e^{-\frac{n^2}{730}}$$

n	Pr(A)
10	12.8%
20	42.2%
23	51.6%
30	70.9%
40	88.8%
50	96.7%
60	99.3%
70	99.9%
75	99.96%

Falso-positivo & falso-negativo

A **sensibilidade** de um teste diagnóstico corresponde ao percentual de resultados positivos dentre as pessoas que tem uma determinada doença ou condição clínica.

A **especificidade** é a capacidade do mesmo teste ser negativo nos indivíduos que não apresentam a doença que está sendo investigada.

Sensibilidade = $\Pr(\text{Teste positivo} \mid \text{Doente})$

Especificidade = $\Pr(\text{Teste negativo} \mid \text{Não doente})$

Falso-positivo = $\Pr(\text{Teste positivo} \mid \text{Não doente}) = 1 - \text{Especificidade}$

Falso-negativo = $\Pr(\text{Teste negativo} \mid \text{Doente}) = 1 - \text{Sensibilidade}$

Falso-positivo & falso-negativo

RT-PCR

- Teste molecular que é baseado na pesquisa de material genético do vírus.
- É o exame padrão-ouro.
- O resultado falso-negativo pode estar relacionado à coleta precoce ou tardia
- O resultado falso-positivo é raro e costuma ter relação com a contaminação da amostra ou no processo de análise no laboratório.
- Os testes usados pela rede de laboratórios de saúde pública apresentam alta sensibilidade (aproximadamente 86%) e alta especificidade (acima de 95%).

Traduzindo para probabilidades:

Falso-positivo: $\Pr(\text{Teste positivo} \mid \text{Não doente}) = 1 - 0.95 = 0.05$

Falso-negativo: $\Pr(\text{Teste negativo} \mid \text{Doente}) = 1 - 0.86 = 0.14$

Falso-positivo & falso-negativo

$$\Pr(\text{Teste positivo} \mid \text{Não doente}) = 1 - 0.95 = 0.05$$

$$\Pr(\text{Teste negativo} \mid \text{Doente}) = 1 - 0.86 = 0.14$$

Eu acho que estou com COVID-19 (ou seja, doente)

Vou ao posto de saúde e me submeto a um teste RT-PCR.

O teste RT-PCR retorna positivo.

Qual a chance de eu estar realmente com COVID-19?

Traduzindo a pergunta com probabilidade: $\Pr(\text{Doente} \mid \text{Teste positivo}) = ?$

Note que $\Pr(\text{Doente} \mid \text{Teste positivo})$ e $\Pr(\text{Teste positivo} \mid \text{Doente})$ são **TOTALMENTE** diferentes!!!

Prevalência de COVID-19: 1% → $\Pr(\text{Doente}) = 0.01 = 1\%$

Portanto, $\Pr(\text{Teste positivo}) = 0.86 \cdot 0.01 + 0.05 \cdot 0.99 = 0.0581 = 5.81\%$

Teorema de Bayes: $\Pr(\text{Doente} \mid \text{Teste positivo}) = \frac{\Pr(\text{Teste positivo} \mid \text{Doente}) \cdot \Pr(\text{Doente})}{\Pr(\text{Teste Positivo})}$
 $= \frac{0.86 \cdot 0.01}{0.0581} = 0.1480 = 14.8\%$.

Conclusão: A chance de eu estar com COVID-19 uma vez que o teste RT-PCR retornou positivo cresceu de 1% (prevalência na população) para 14.8%, ou seja, cresceu quase 15 vezes!

Falso-positivo & falso-negativo

$$\Pr(\text{Teste positivo} \mid \text{Não doente}) = 1 - 0.95 = 0.05$$

$$\Pr(\text{Teste negativo} \mid \text{Doente}) = 1 - 0.86 = 0.14$$

Eu acho que estou com COVID-19 (ou seja, doente)

Vou ao posto de saúde e me submeto a um teste RT-PCR.

O teste RT-PCR retorna negativo.

Qual a chance de eu estar realmente com COVID-19?

Traduzindo a pergunta com probabilidade: $\Pr(\text{Doente} \mid \text{Teste negativo}) = ?$

Note que $\Pr(\text{Doente} \mid \text{Teste negativo})$ e $\Pr(\text{Teste negativo} \mid \text{Doente})$ são **TOTALMENTE** diferentes!!!

Prevalência de COVID-19: 1% → $\Pr(\text{Doente}) = 0.01 = 1\%$

Portanto, $\Pr(\text{Teste negativo}) = 0.14 \cdot 0.01 + 0.95 \cdot 0.99 = 0.9419 = 94.2\%$

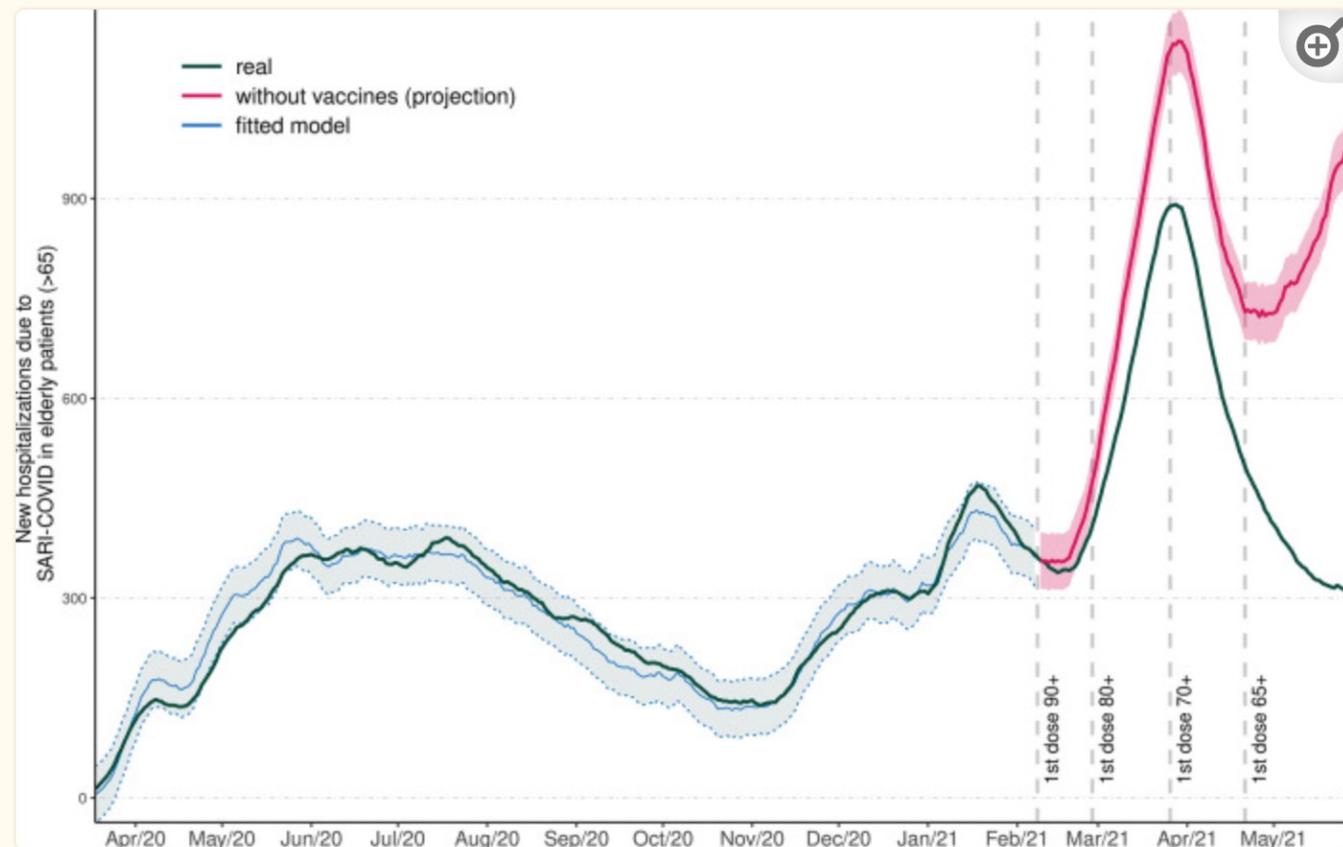
Teorema de Bayes: $\Pr(\text{Doente} \mid \text{Teste positivo}) = \frac{\Pr(\text{Teste positivo} \mid \text{Doente}) \cdot \Pr(\text{Doente})}{\Pr(\text{Teste Positivo})}$
 $= \frac{0.14 \cdot 0.01}{0.9419} = 0.0015 = 0.15\%$.

Conclusão: A chance de eu estar com COVID-19 uma vez que o teste RT-PCR retornou negativo decresceu de 1% (prevalência na população) para 0.15%, ou seja, decresceu quase 7 vezes!

Quantas hospitalizações a vacinação contra a COVID-19 já preveniu em São Paulo?

Rafael Izbicki, Leonardo S. Bastos, Meyer Izbicki, Hedibert F. Lopes, and Tiago Mendonça dos Santos

Clinics (Sao Paulo). 2021; 76: e3250. Publicado online 27/Agosto/2022. doi: 10.6061/clinics/2021/e3250



[Figure 1](#)

Number of hospitalizations due to SARI-COVID in patients aged >65 years (dark green), fitted pre-vaccination model (blue), and estimated counterfactual curve for the setting without vaccines (red).

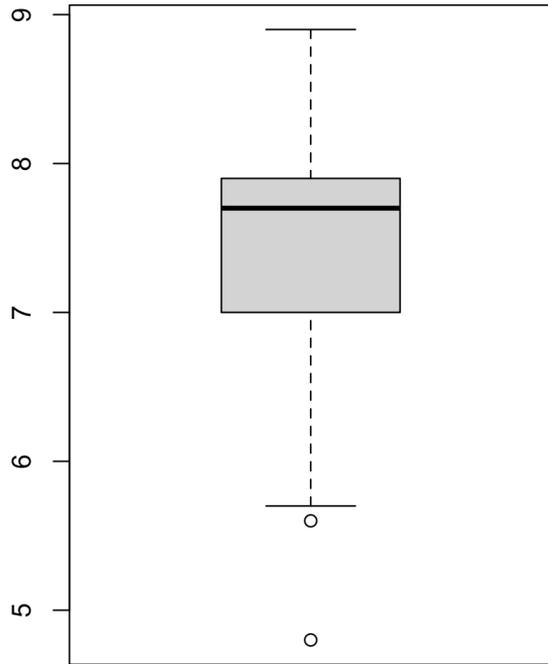
Usando as áreas entre as curvas, observamos que aproximadamente 24.364 hospitalizações foram evitadas em São Paulo devido à vacinação antes de 28 de maio de 2021. Considerando que a taxa de mortalidade estimada de pacientes hospitalizados com mais de 65 anos com SARI-COVID-19 é de aproximadamente 45%, cerca de 10.964 mortes podem ter sido prevenidas pela vacinação durante esse período. Além disso, usando estimativas dos custos médios de hospitalização por COVID-19, aproximadamente 297 milhões de dólares podem ter sido economizados, o que é suficiente para comprar quase 30 milhões de doses adicionais de vacina (a 10 dólares cada).

Em conclusão, fornecemos evidências de que as vacinas são uma forma eficaz de reduzir a mortalidade por COVID-19, além de economizar recursos financeiros substanciais durante a pandemia.

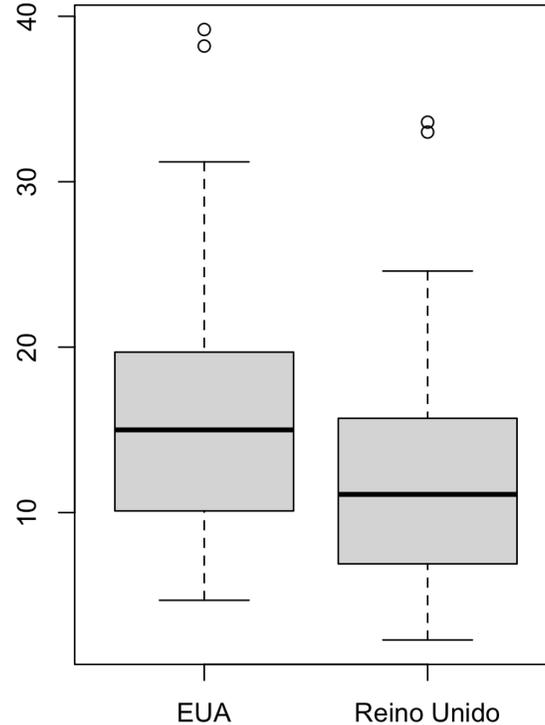
Classificando filmes (IMDB)

Correlação amostral

	Média IMDB	Nota 10 USA (%)	Nota 10 UK (%)
Média IMDB	1.000	0.795	0.805
Nota 10 USA (%)	0.795	1.000	0.987
Nota 10 UK (%)	0.805	0.987	1.000

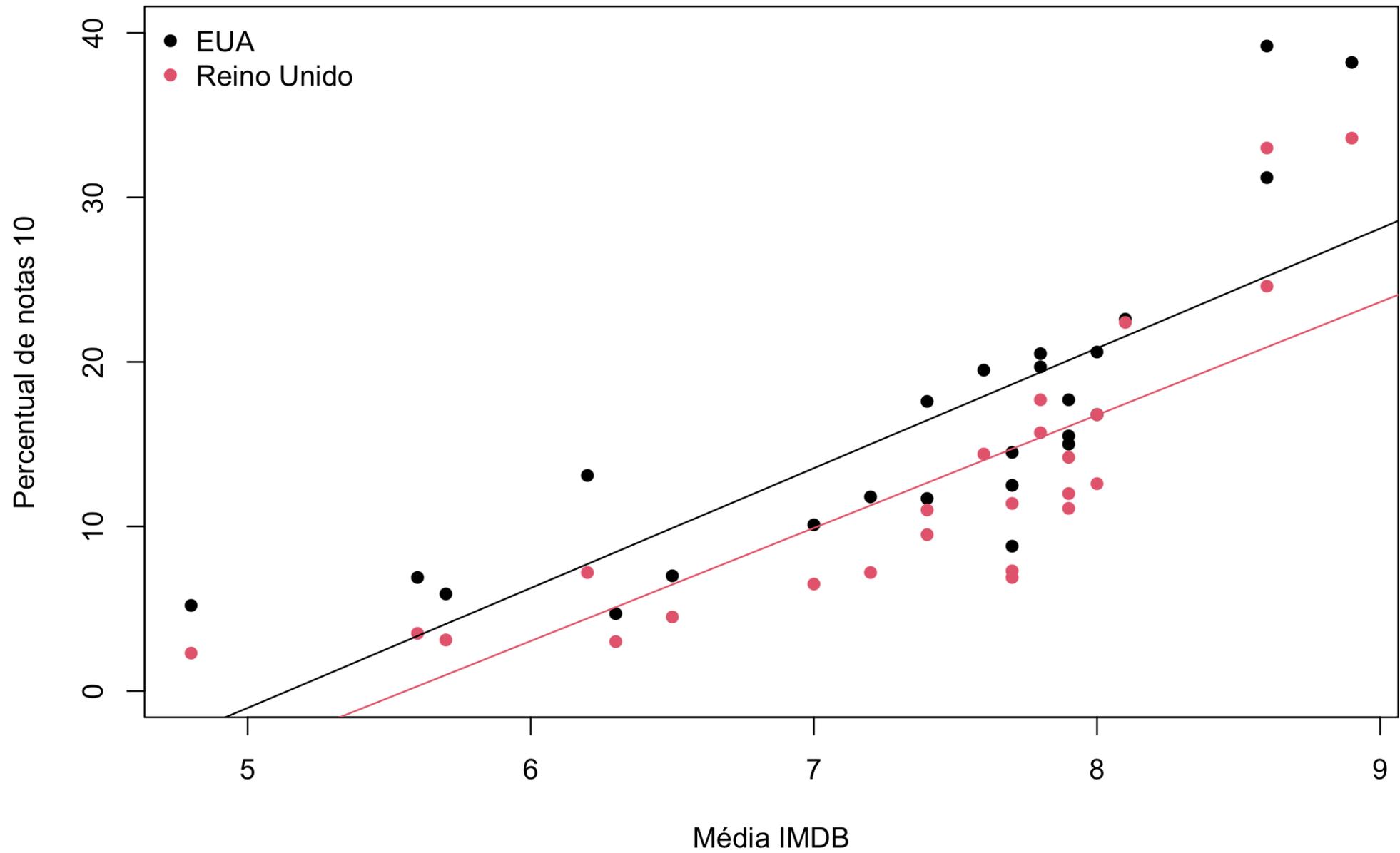


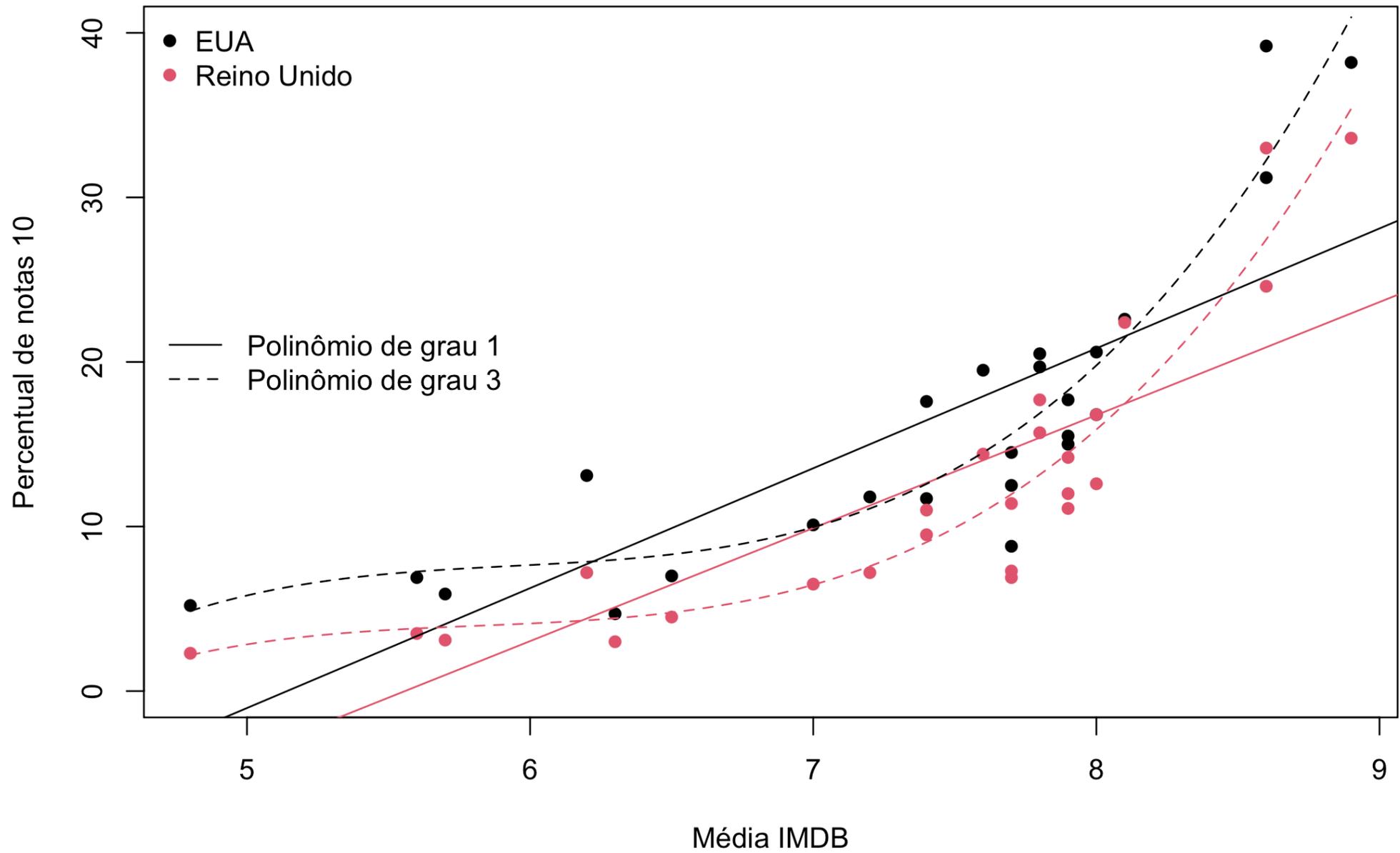
Médias IMDB



Percentual de notas 10 (máxima)

	Média IMDB	Nota 10 USA (%)	Nota 10 UK (%)
[1,]	4.8	5.2	2.3
[2,]	5.6	6.9	3.5
[3,]	5.7	5.9	3.1
[4,]	6.2	13.1	7.2
[5,]	6.3	4.7	3.0
[6,]	6.5	7.0	4.5
[7,]	7.0	10.1	6.5
[8,]	7.2	11.8	7.2
[9,]	7.4	11.7	9.5
[10,]	7.4	17.6	11.0
[11,]	7.6	19.5	14.4
[12,]	7.7	8.8	6.9
[13,]	7.7	12.5	7.3
[14,]	7.7	14.5	11.4
[15,]	7.8	20.5	17.7
[16,]	7.8	19.7	15.7
[17,]	7.9	15.5	12.0
[18,]	7.9	15.0	11.1
[19,]	7.9	17.7	14.2
[20,]	8.0	20.6	16.8
[21,]	8.0	16.8	12.6
[22,]	8.1	22.6	22.4
[23,]	8.6	39.2	33.0
[24,]	8.6	31.2	24.6
[25,]	8.9	38.2	33.6





Fenótipo Bayesiano também 😊



1. A Constrained **BART** Model for Identifying Heterogeneous Treatment Effects in RDD
2. When it counts: Econometric identification of the basic **factor model** based on GLT structures
3. Parsimonious Bayesian **factor analysis** when the number of factors is unknown
4. Dynamic mixed frequency **pooled copula** International Journal of Forecasting
5. Time series momentum predictability via dynamic binary classification
6. **Dynamic portfolio allocation** in high dimensions using sparse risk factors
7. Dynamic ordering learning in multivariate forecasting
8. **Dynamic sparsity** on dynamic regression models
9. Tree-Based Bayesian Treatment Effect Analysis
10. Decoupling shrinkage and selection in Gaussian linear **factor analysis**
11. **Deep Learning** Models For Inflation Forecasting
12. Parsimony inducing priors for **large scale state-space models**
13. Bayesian generalizations of the **integer-valued autoregressive** model
14. The illusion of the illusion of **sparsity**
15. How many hospitalizations has the **COVID-19** vaccination already prevented in Sao Paulo?
16. **Spatial Prediction** of Sea Level Trends
17. **Prior sensitivity analysis** in a semi-parametric integer-valued time series model
18. Scalable semi-parametric inference for the means of **heavy-tailed** distributions
19. Bayesian semi-parametric **Markov switching** stochastic volatility mode
20. Walk on the wild side: **Multiplicative sunspots** and temporarily unstable path
21. **Efficient sampling** for Gaussian linear regression with arbitrary priors
22. Particle learning for Bayesian **semi-parametric stochastic volatility** model
23. Dynamic models
24. **Bayesian hypothesis testing**: Redux
25. On the long run volatility of stocks: **time-varying predictive systems**
26. Bayesian factor model **shrinkage** for linear IV regression with many instruments
27. Sequential Bayesian learning for stochastic volatility with **variance-gamma jumps** in returns
28. Efficient Bayesian inference for **multivariate factor SV models**
29. Cholesky **realized stochastic volatility** model
30. Particle learning for **fat-tailed** distributions
31. **Time-varying extreme pattern** with dynamic models
32. **Bayesian instrumental variables**: likelihoods and priors
33. **Treatment effects**: a Bayesian perspective, Econometric Reviews
34. Modern Bayesian Factor Analysis

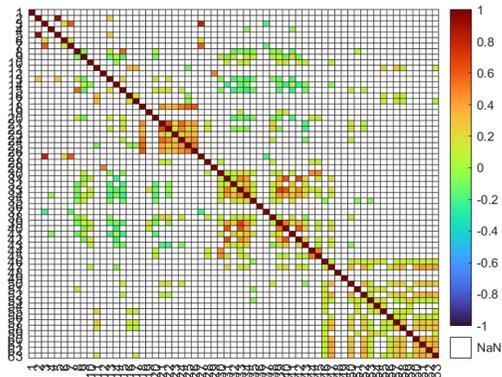


Figure 7: NYSE data; estimated marginal correlation matrix $E(\Omega^*|y)$, where $\Omega_{i\ell}^* = \text{Corr}(\Lambda_{i1}f_{1t})(y_{i\ell} - \Lambda_{i1}f_{1t})$.

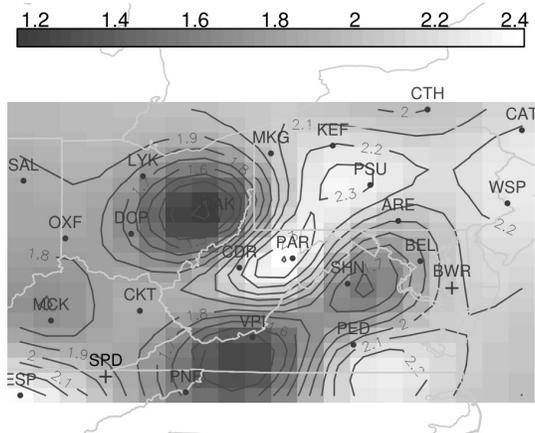
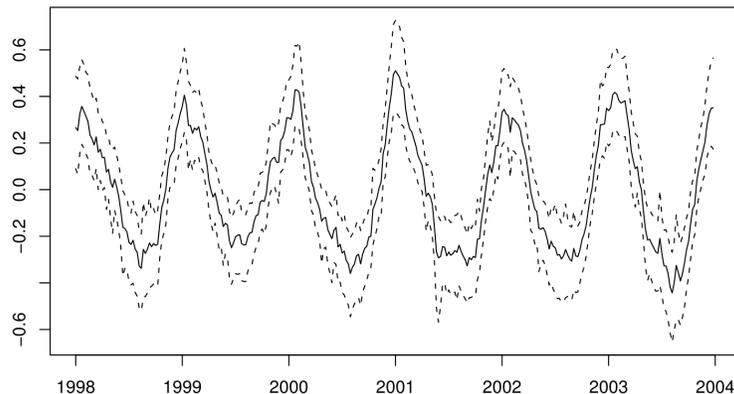
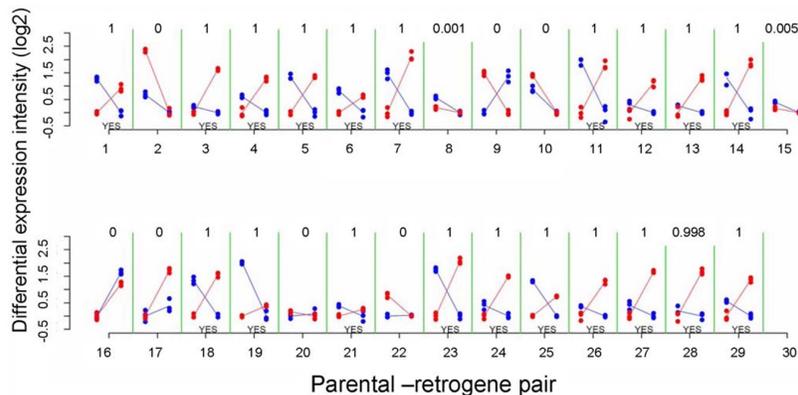
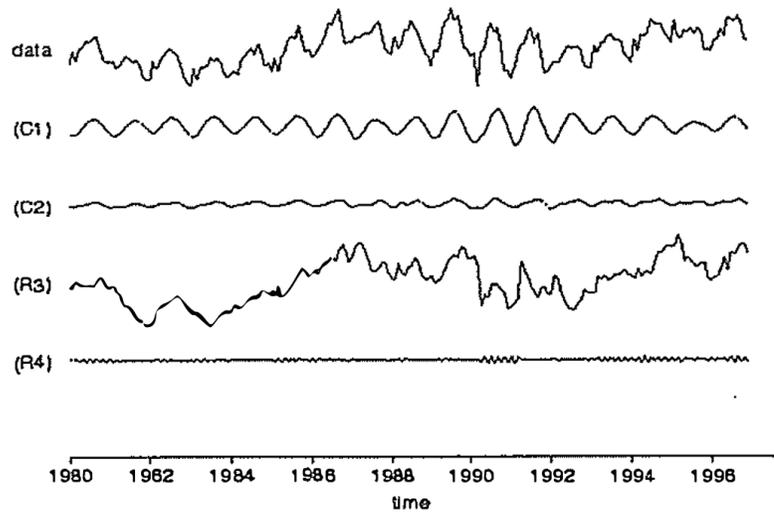


Figure 2: Google Flu Trends estimated ILI percentages (dashed line) and CDC ILI Surveillance percentages (solid line) for the United States, from June 2003 until September 2009. Separate plots correspond to separate influenza years, with each new influenza season starting in autumn, and ending in spring. Note that CDC did not use to produce ILI reports during summers before 2009, and thus no solid line appears during summer months prior to 2009.

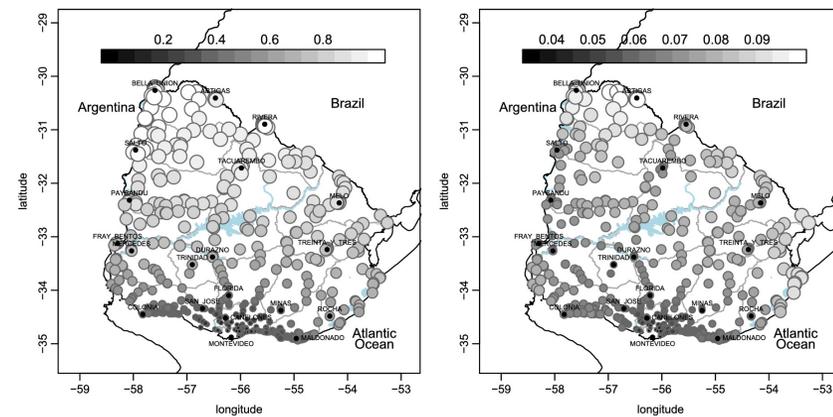
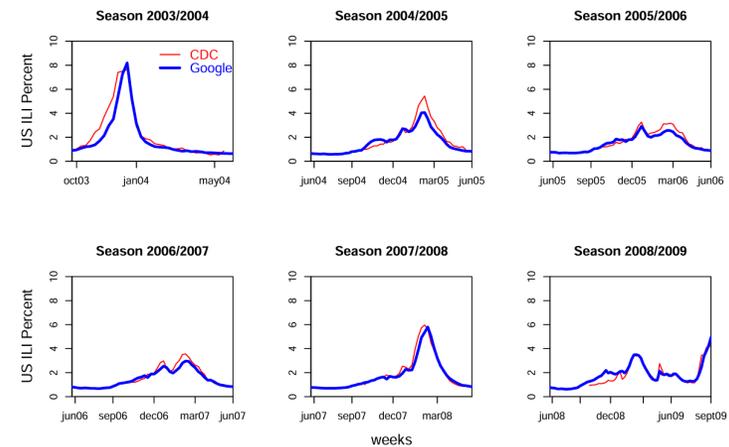
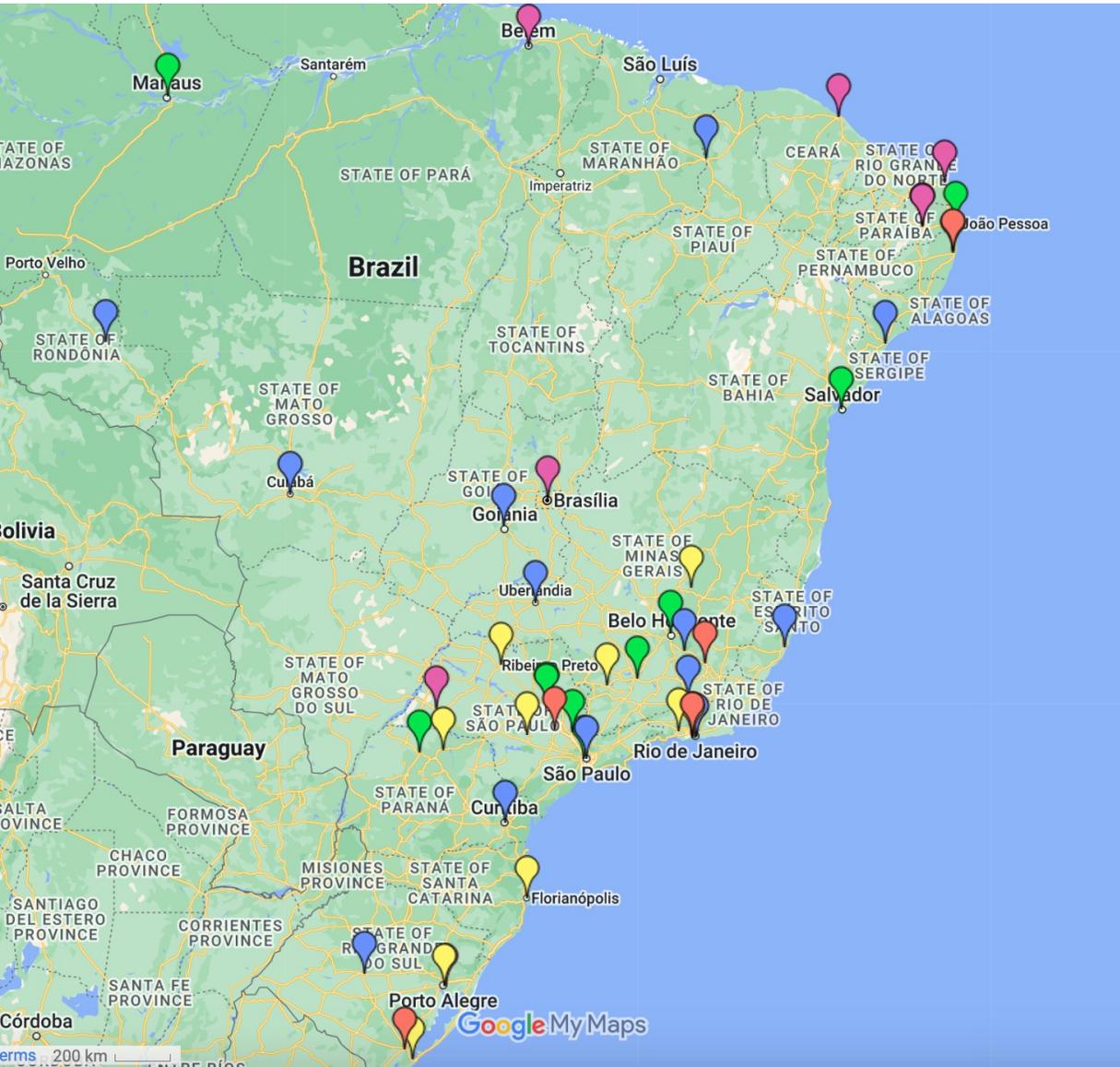


FIG. 2. Posterior mean of θ_i and standard deviations (second column) for observed and unobserved cities under the SHFM when $\phi = 5$.

Onde estudar?



Vagas Bacharelado em Estatística

autorizadas pelo MEC: 1.795 vagas

REGIÃO SUDESTE – 787 vagas

REGIÃO NORDESTE – 490 vagas

REGIÃO CENTRO-OESTE – 190 vagas

REGIÃO SUL – 176 vagas

REGIÃO NORTE – 148 vagas

Melhores universidades (subjetivo)

UFRJ

USP

UNICAMP

UFMG

UFSCAR

USP São Carlos

UFPE

UFPR

UFRGS

UFBA

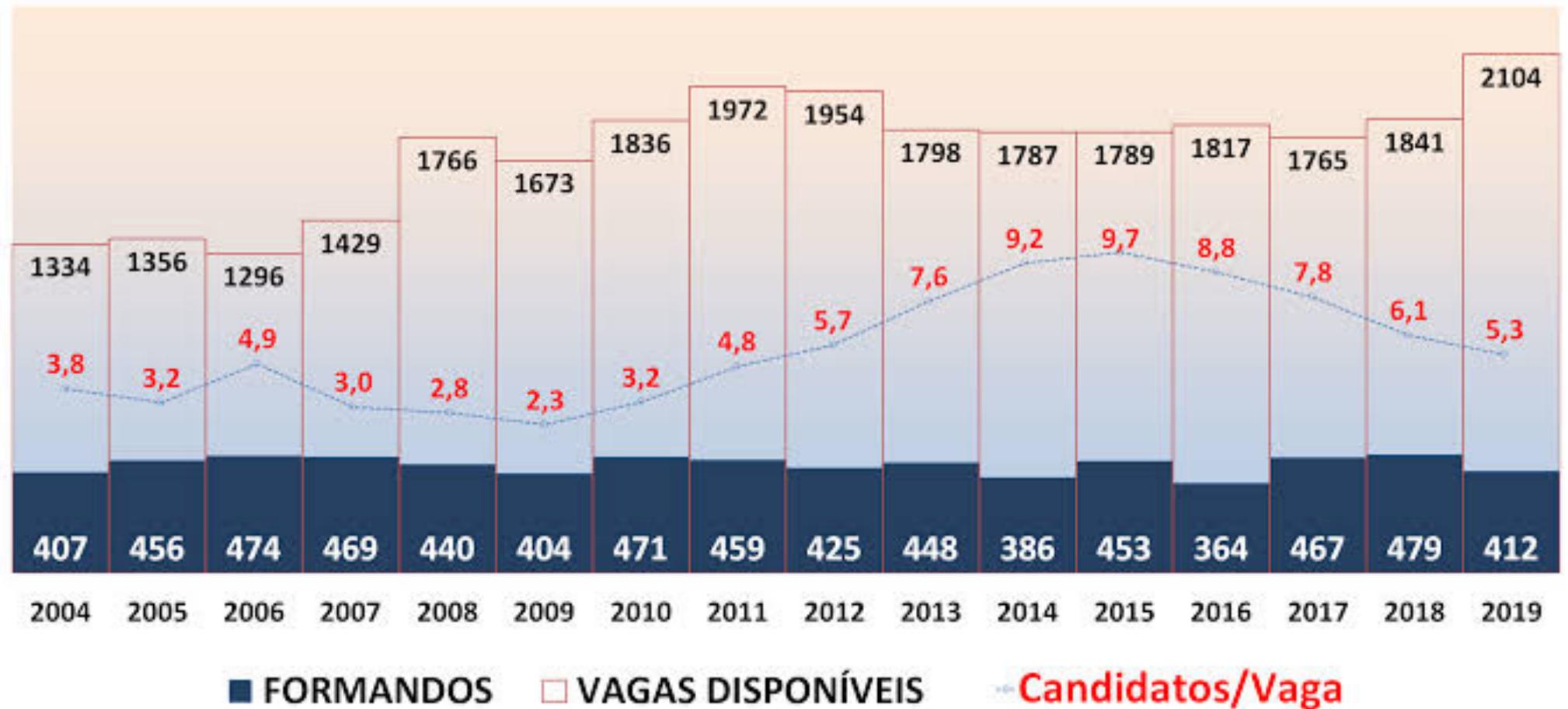
UFRN

ENCE

Fonte: <https://www.conre3.org.br/>

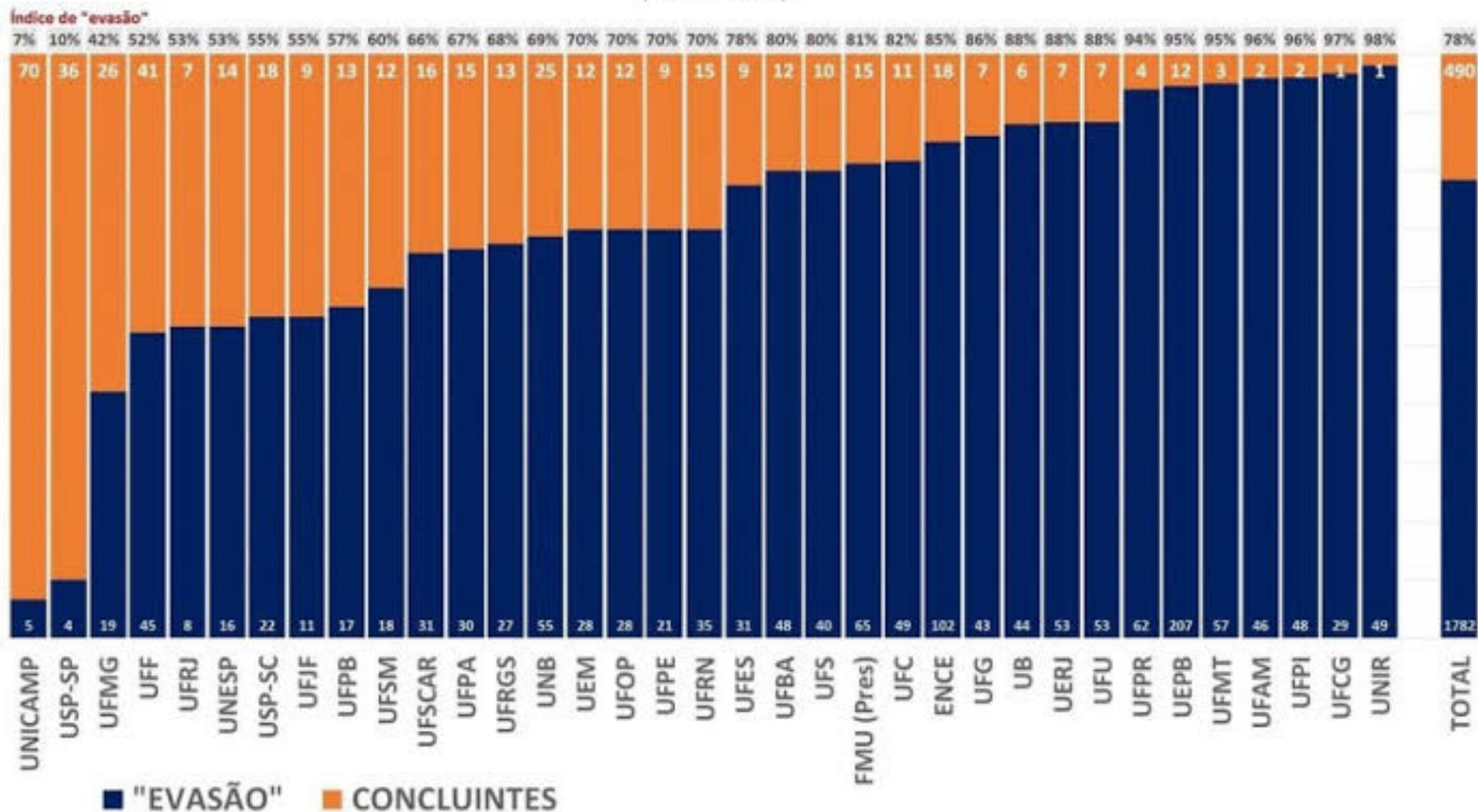
Bacharelado em Estatística

BACHARELADO EM ESTATÍSTICA
Concluintes, total de vagas oferecidas e Relação Candidatos/Vaga
(INEP)



Evasão

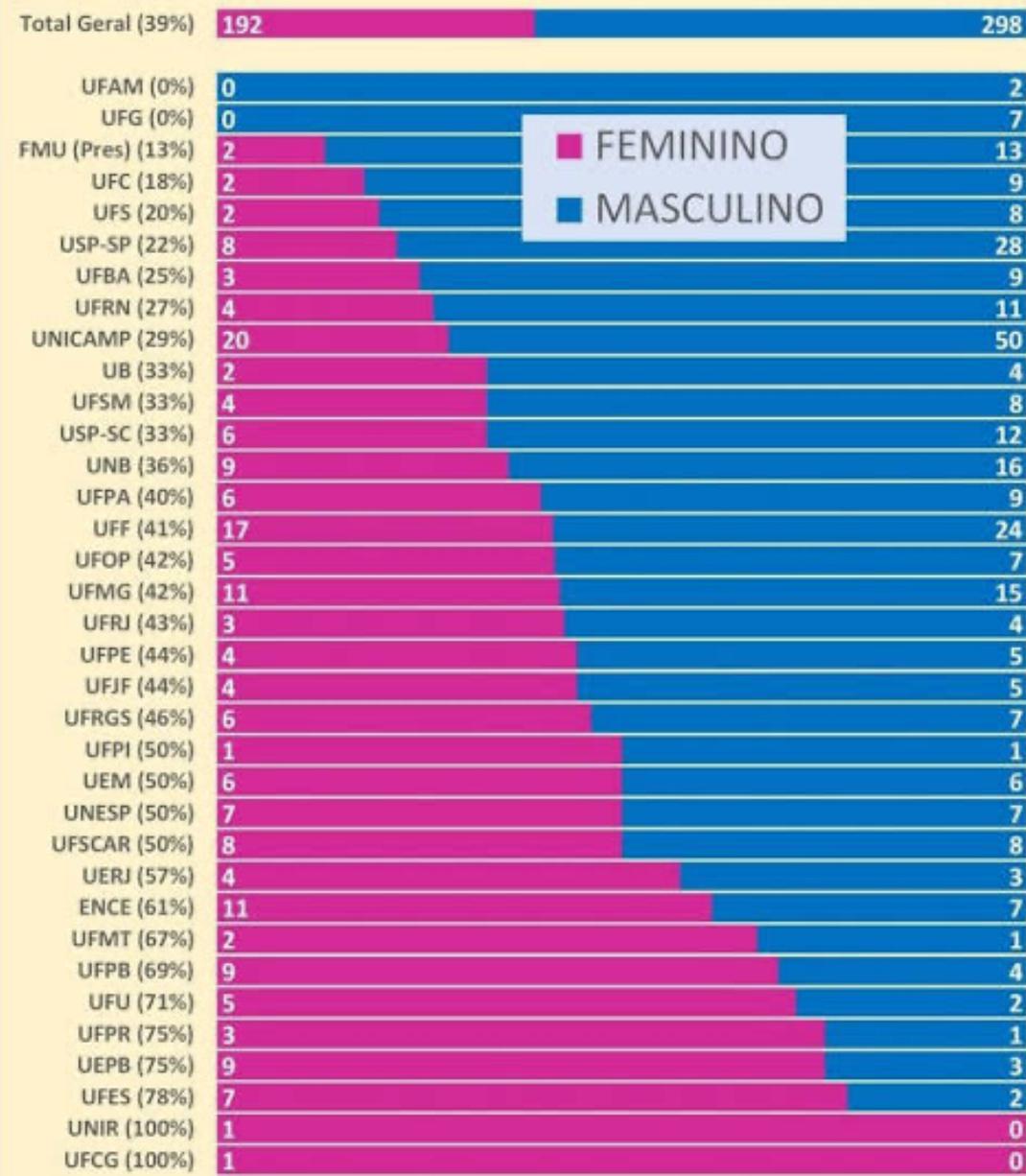
Relação entre VAGAS NOVAS e CONCLUINTES por universidade
(INEP 2021)



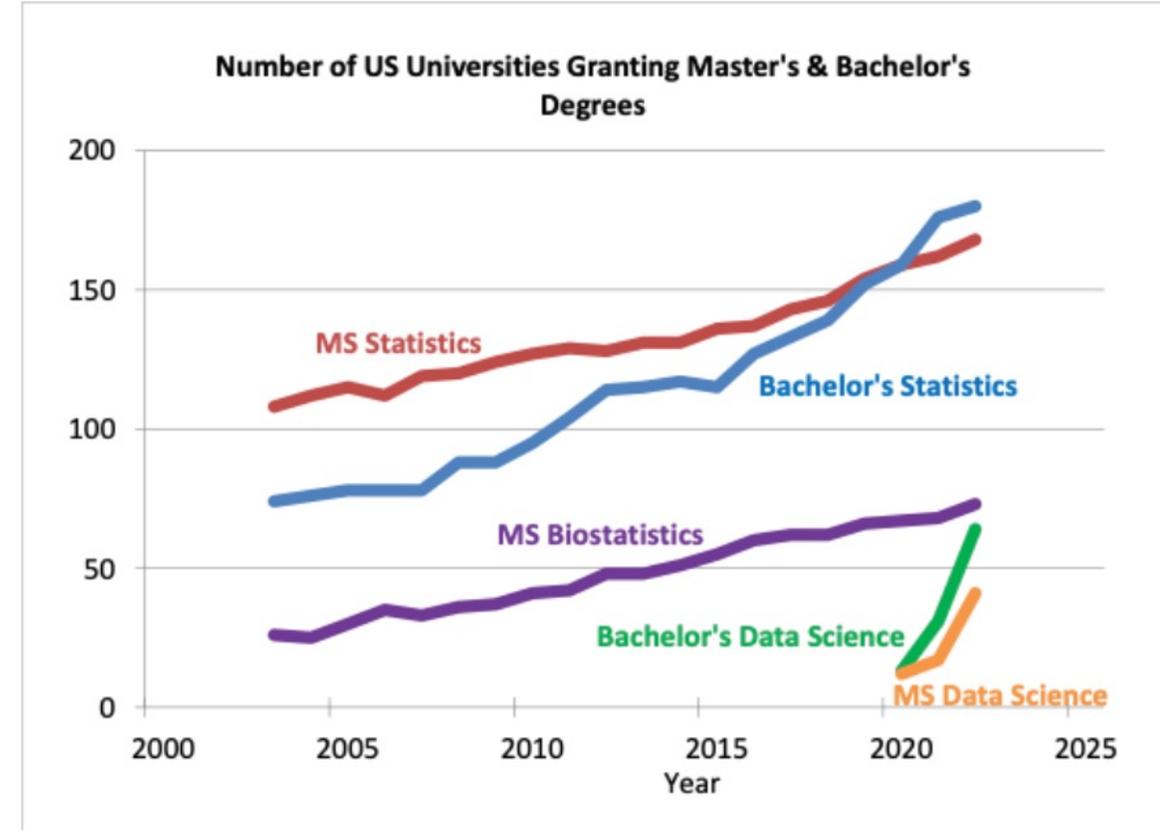
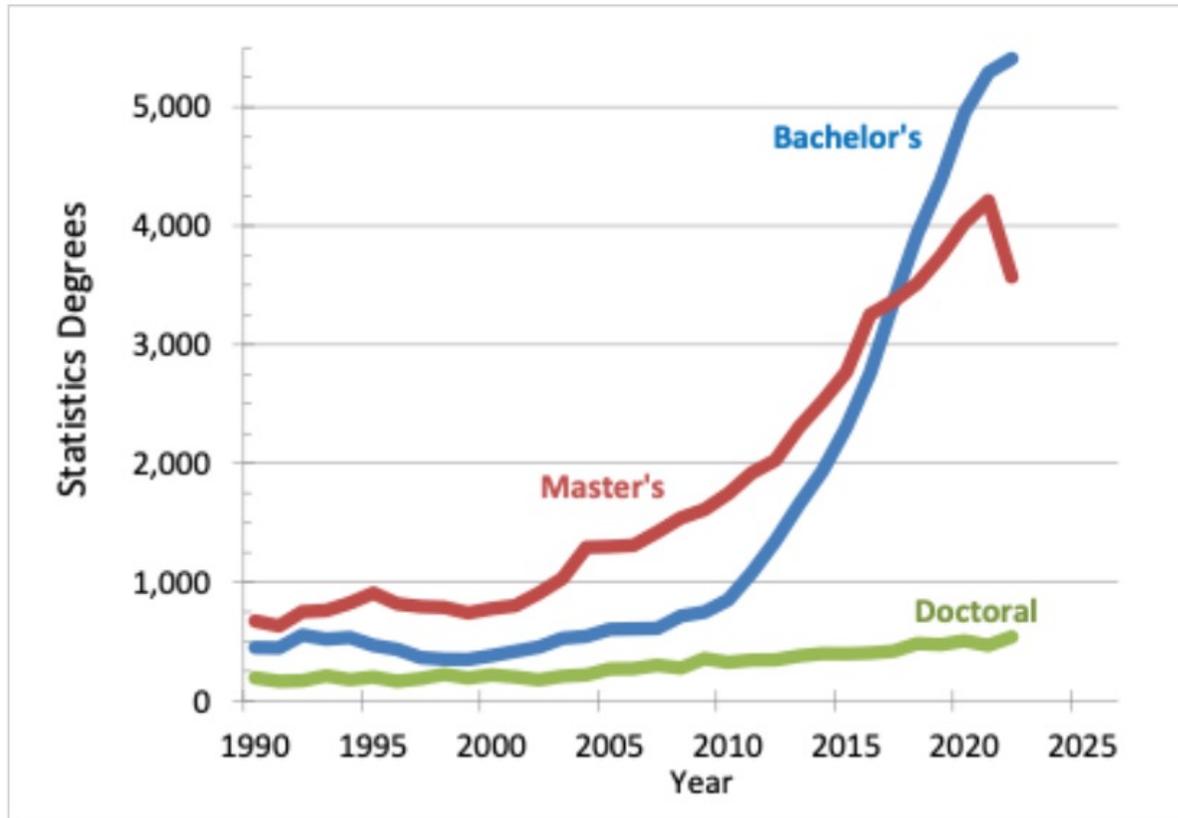
Concluintes por gênero

Bacharelado em Estatística

Concluintes por gênero - 2021 (% feminino)



Estatística & Ciência de Dados



Mais informações

A **ASA** foi criada em 1839 (19.000 membros)

A **ABE** foi criada em 1984

Simpósios Nacionais de Probabilidade e Estatística (**SINAPE**) iniciaram em 1974

Dia do estatístico - 29 de maio (dia da fundação do IBGE 88 anos atrás)

20 de outubro é o World Statistics Day (desde 2010)

Alguns encontros de Estatística no Brasil:

- SINAPE
- Escola de Regressão
- Escola de Séries Temporais e Econometria (ESTE)
- Encontro Brasileiro de Estatística Bayesiana (EBEB)

Considerações finais

- Dificuldades serão sempre **(muitas vezes!)** maiores para aqueles que viemos das massas menos abastadas.
- Perseverar (**e muito!**) é preciso.
- Ninguém avança sozinho.
- Sempre existem pessoas & oportunidades ao longo do caminho.
- Chegando ao top, você perceberá a mediocridade que lá habita.
- Nunca se deprecie por usar de ações afirmativas.
- Não se venda barato!

*Dois homens e uma mulher
Arnaldo, Carlinhos e Zé
Um dia já fui chipanzé
Agora eu ando só com o pé*

*Pé em Deus
E fé na Taba
Pé em Deus
E fé na Taba*

Tribalistas (Marisa Monte)