

Abordagem Quantitativa na Pesquisa Científica

PPG em Ciência da Informação - Metodologia da Pesquisa Científica

Marcus Nunes

16 de Novembro de 2021

Departamento de Estatística - UFRN

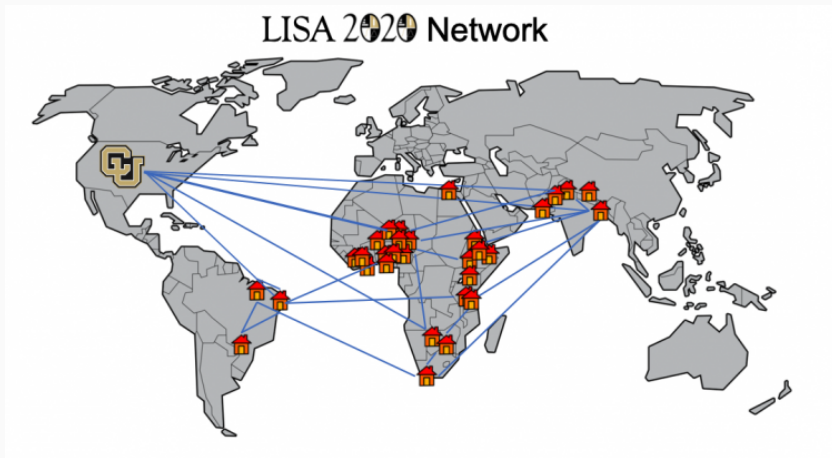
Roteiro

1. Apresentação
2. Motivação
3. Cuidados com a Coleta de Dados
4. Questionários
5. Software
6. Conclusões

Apresentação

Apresentação

- Marcus Nunes, Professor Adjunto no Departamento de Estatística
- PhD em Estatística pela Pennsylvania State University (2013)
- Mestre em Matemática Pura (2008) e Bacharel em Matemática Aplicada (2006) pela UFRGS
- Interessado em Aprendizado de Máquina, Educação Estatística e Projetos de Colaboração Estatística
- Site pessoal: marcusnunes.me
- Email: marcus@marcusnunes.me



<http://lea.estadistica.ccet.ufrn.br>

Motivação

- Fala introdutória, para não deixar ninguém de fora
- Amostragem, questionários e análise de dados
- Mostrar algumas possibilidades da análise quantitativa de dados

- Dados são produzidos a toda hora e em todo lugar
- Por que não usar estes dados para responder perguntas científicas?
- Identificar tendências
- Auxiliar em tomadas de decisão
- Generalizar resultados

Cuidados com a Coleta de Dados

Cuidados com a Coleta de Dados

- A qualidade da análise realizada depende diretamente da qualidade dos dados coletados
- Dados que não foram coletados corretamente não irão gerar boas análises
- É preciso entender bem os dados e saber das suas limitações
- Muitas vezes, a análise já surge sem futuro, pois os dados não foram coletados da maneira devida

Cuidados com a Coleta de Dados

- **População:** é o conjunto de todos os elementos ou resultados sob investigação
- **Amostra:** qualquer subconjunto da população de interesse

Cuidados com a Coleta de Dados

- As observações contidas em uma amostra são mais ou menos informativas quanto mais conhecermos a população que nos interessa
- Por exemplo, ao analisarmos a glicose de um paciente a partir de uma amostra de sangue retirada do seu dedo, teremos uma boa ideia a respeito da quantidade de glicose no corpo todo
- Por outro lado, se estivermos interessados na opinião da população sobre um projeto governamental, escolher uma amostra de 200 indivíduos moradores da região beneficiada pelo projeto vai gerar um viés de seleção

Cuidados com a Coleta de Dados

- A American Psychology Association cunhou o termo WEIRD há alguns anos no artigo *Are your findings “WEIRD”?*
- A palavra weird significa esquisito
- A sigla WEIRD significa
 - Western
 - Educated
 - Industrialized
 - Rich
 - Democratic

Cuidados com a Coleta de Dados

- **Estudo Experimental:** grupos específicos para os sujeitos do estudo são designados pelo pesquisador
- **Estudo Observacional:** não é possível definir grupos específicos para os sujeitos do estudo
- **Meta-Análise:** revisão de trabalhos publicados por outros pesquisadores
- **Estudo de Caso:** análise profunda de um ou um número pequeno de sujeitos

- Utilizado quando é possível dividir aleatoriamente os sujeitos do estudo em grupos
- Permite controlar (quase) todos os aspectos que se queiram do estudo
- Permite estabelecer relações de causa e efeito

Estudo Observacional

- Utilizados quando não é possível, por quaisquer razões, dividir os objetos de estudo em grupos
- Não permite controlar todos os aspectos do estudo
- Não permite estabelecer relações de causa e efeito

- Coleção de técnicas estatísticas para combinar estudos diferentes
- Encontra respostas para resultados pequenos e moderados ao levar em conta um grande número de estudos independentes
- Demanda técnicas especiais para ser aplicada

- Utilizado quando é necessário abordar um assunto com poucos sujeitos disponíveis
- Muito utilizado em medicina e ciências sociais
- Importante para gerar hipóteses e servir como estudo piloto
- Não demandam tratamento quantitativo

Cuidados com a Coleta de Dados

- Decidir se o estudo experimental, observacional ou meta-análise
- Tomar uma amostra representativa
- Tomar uma amostra grande o suficiente

Mas como Selecionar uma Amostra?

- Amostragem Probabilística
 - Amostragem Aleatória Simples (a.a.s.)
 - Amostragem Sistemática
 - Amostragem Estratificada
 - Amostragem por Conglomerados
- Amostragem Não-Probabilística
 - Amostragem por Cotas (proporcional)
 - Amostragem por Julgamento
 - Amostragem por Conveniência (acidental)

Questionários

- Uma pesquisa realizada nos Estados Unidos fez duas perguntas a dois grupos distintos de adolescentes:
 - “O que é mais fácil para alguém da sua idade comprar: cigarro, cerveja ou maconha?”
 - “O que é mais fácil para alguém da sua idade obter: cigarro, cerveja ou maconha?”

Resposta	Comprar (%)	Obter (%)
Cigarro	35	39
Cerveja	18	27
Maconha	34	19
Todos são iguais	4	5
Não sabe/Não Respondeu	9	10

Questionários

Há diversos problemas que podem ser encontrados quando fazemos perguntas em um questionário

- Vício deliberado
- Vício sem intenção
- Desejo de agradar
- Perguntas para os desinformados
- Complexidade desnecessária
- Ordem das perguntas
- Confidencialidade e anonimidade

Software

- Há muitas opções de software para análise de dados
- Nenhum é essencialmente melhor ou pior do que os outros
- É possível dividi-lo em dois grandes grupos:
 - *Point and click* (Excel, Google Sheets, SPSS, Statistica)
 - Linguagens de programação (R, Python, Julia, SAS)

- Software do tipo *point and click* permite que o usuário rapidamente obtenha resultados em suas análises
- Em geral, não é preciso ler nenhum manual ou tutorial para realizar as análises mais elementares
- Dependendo do programa escolhido, as análises disponíveis são muito elementares
- Não permite criar facilmente um roteiro de análise de dados reprodutível

- Minha escolha pessoal é por uma linguagem de programação (R em particular)
- A curva de aprendizagem é mais íngreme
- Por serem de código aberto, outros pesquisadores desenvolvem e programam novas técnicas analíticas
- A documentação da análise é feita naturalmente

- A seguir vou fazer uma pequena demonstração do que é possível fazer com o R
- Eu desejo encontrar a evolução no número de citações acumuladas dos docentes da linha de pesquisa Informação e Ciência do PPGCIN
- Para isso, irei utilizar as citações disponíveis no Google Scholar

```
library(scholar)
library(tidyverse)
theme_set(theme_bw())
```

```
# lista de docentes
```

```
docentes
```

Citações dos Docentes do PPGCIN

```
##          id      docente
## 1 7EYSYFAAAAAJ Docente_01
## 2 bJ948eIAAAAAJ Docente_02
## 3 EQxDJUyAAAAJ Docente_03
## 4 gqmfapMAAAAAJ Docente_04
## 5 hH8j73wAAAAJ Docente_05
## 6 IphFWwQAAAAJ Docente_06
## 7 kt-5psIAAAAAJ Docente_07
## 8 On_e41EAAAAJ Docente_08
## 9 YAYLYesAAAAJ Docente_09
```

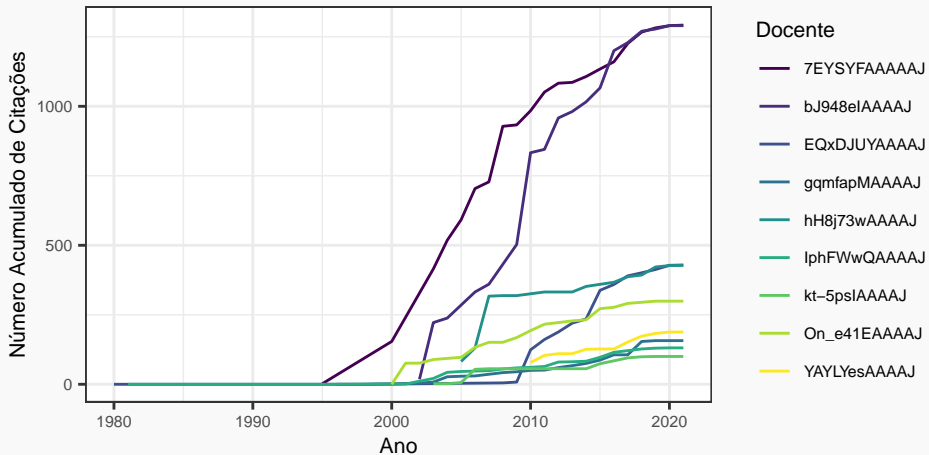


```
# citacoes  
  
citacoes <- compare_scholars(docentes$id)  
  
citacoes <-  
  citacoes %>%  
  group_by(id) %>%  
  mutate(acc_cites = cumsum(cites))
```

```
# grafico
```

```
ggplot(citacoes, aes(x = year, y = acc_cites, colour = id)) +  
  geom_line() +  
  labs(x = "Ano",  
        y = "Número Acumulado de Citações",  
        colour = "Docente") +  
  scale_colour_viridis_d()
```

Citações dos Docentes do PPGCIN



- Angelini e Nunes (202?)
- Quais fatores influenciam a produção acadêmica dos professores da UFRN?
- Analisamos dados entre 2014 e 2017

Média	Desvio Padrão
6.92	10.23

Produção Total

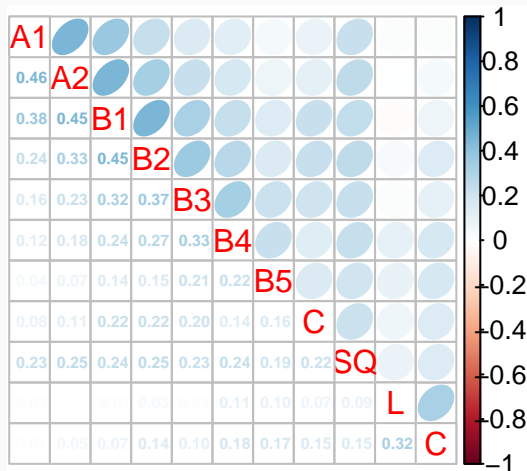
Número de Artigos	Qtde de Professores	Percentual
0	327	18.31
1	227	12.71
2	187	10.47
3	168	9.41
4	126	7.05
5	92	5.15
6	81	4.54
7	59	3.30
8	64	3.58
9	38	2.13
10+	417	23.35

Média	Desvio Padrão
0.7	1.9

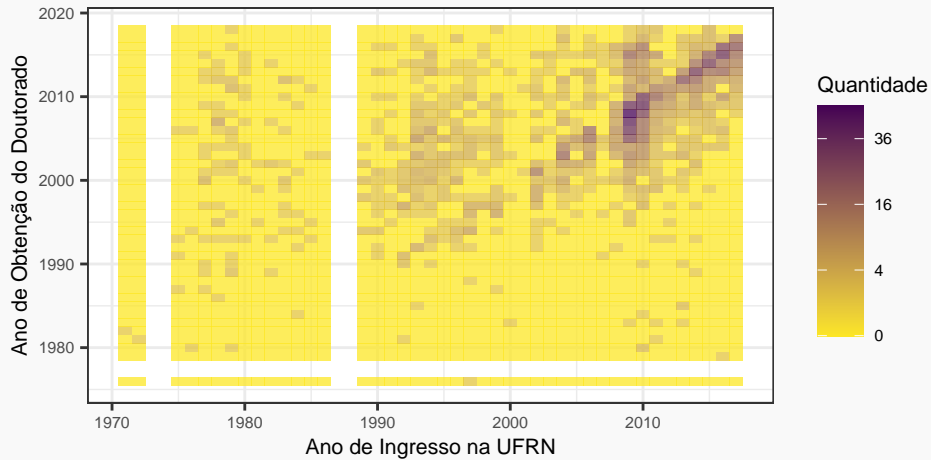
Produção A1

Número de Artigos	Qtde de Professores	Percentual
0	1311	73.40
1	239	13.38
2	84	4.70
3	53	2.97
4	25	1.40
5	28	1.57
6	13	0.73
7	7	0.39
8	5	0.28
9	4	0.22
10+	17	0.95

Correlações entre os Tipos de Produção



Heatmap



Variáveis Utilizadas

- UNG: horas de ensino anuais na graduação
- GRA: horas de ensino anuais na pós-graduação
- SDOC: orientações de doutorado concluídas
- SMAS: orientações de mestrado concluídas
- SIC: orientações de iniciação científica concluídas
- SMON: número de TCCs orientados
- PFU: número de projetos financiados
- PNF: horas dedicadas à pesquisa
- OUT: número de eventos
- BEF: número de artigos publicados antes de obter o doutorado
- YDOC: ano de obtenção do doutorado
- YUFRN: ano de ingresso na UFRN
- GEN: gênero

- Dados discretos, provenientes de contagens
- O natural seria escolher distribuições como Poisson ou Binomial Negativa para o ajuste do modelo
- Entretanto, a análise exploratória nos sugere um excesso de zeros
- Isto nos levou a considerar o ajuste de um modelo hurdle

Modelagem - Produção Total Não-Nula

Preditores	Coefficientes	Média	Desvio.Padrão	p.valor
(Intercept)	5.0415	NA	NA	0.0000
UNG	1.0013	13.9999	6.3341	0.9600
GRA	1.0763	3.4784	4.5456	0.0207
SDOC	1.3162	0.6988	1.7804	0.0000
SMAS	1.2484	2.2447	3.1184	0.0000
SIC	1.1551	2.0666	3.6978	0.0000
SMON	1.0558	3.3645	5.9741	0.0186
PFU	1.1478	3.6534	5.6673	0.0000
PNF	1.0776	21.7013	52.0439	0.0110
OUT	1.0114	5.8018	8.7833	0.6360
BEF	1.4226	3.2156	5.7848	0.0000
YDOC	0.9108	2007.2374	6.7722	0.0083
YUFRN	1.1558	2005.8970	9.7173	0.0000
GENMasculino	1.0695	NA	NA	0.1806

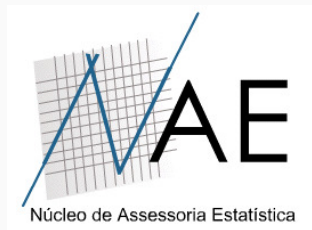
Onde Procurar Ajuda

Onde Procurar Ajuda

- Empresas de consultoria estatística
- Profissionais autônomos
- Estudantes universitários



- LEA - Laboratório de Estatística Aplicada da UFRN -
<http://lea.estadistica.ccet.ufrn.br>



- NAE - Núcleo de Assessoria Estatística da UFRGS - <http://mat.ufrgs.br/~nae/>

Conclusões

- A análise quantitativa começa antes da coleta de dados
- O planejamento da sua pesquisa vai economizar muita dor de cabeça futura
- A análise dos dados é, geralmente, a parte mais rápida
- Na dúvida, peça ajuda

Contato

- Marcus Nunes
- Site pessoal: `marcusnunes.me`
- Email: `marcus@marcusnunes.me`