

Tópicos atuais em análise quantitativa de dados para pesquisas em ensino e aprendizagem de línguas

Seminário Internacional de Pesquisas em Ensino e Aprendizagem de Línguas:
abordagens quantitativa e mista

Ronaldo Lima Jr.

`ronaldo.junior@unb.br`

`ronaldolimajr.github.io`

Universidade de Brasília / CNPq

1. Gerenciamento de Dados
2. Compartilhamento de Dados
3. Menos ênfase em valor de p

Gerenciamento de Dados

Gerenciamento de dados

- Nomeação de pastas e arquivos
- Organização de pastas
- Versionamento

- Nomeação de pastas e arquivos
- Organização de pastas
- Versionamento

- Por quê?
 - Organização pessoal
 - Encontrar arquivos
 - Manutenção de dados
 - Reanálises



projeto.ai



projeto_final.ai



projeto_final2.ai



projeto_final2_
alterado.ai



projeto_final2_
alterado_amém.ai



projeto_final2_
alterado_amém
_fenopai_quevai.ai

O nome do arquivo deve¹

1. Ser óbvio
2. Ser escrito corretamente
3. Ser legível
4. Ser versátil e abrangente
5. Ser sintaticamente correto
6. Não acarretar em problemas de codificação

¹<https://www.visuarea.com.br/tutoriais/como-nomear-arquivos-corretamente>

1. Ser óbvio

- Não ser criativo
- Pensar em buscar o arquivo posteriormente
- Pensar em mudar o arquivo de pasta e já ter outro com o mesmo nome
- Pensar no compartilhamento de arquivo

2. Ser escrito corretamente

- Para buscar o arquivo, não pode ter "alanise" para "analise"
- Cuidado com abreviações (vai lembrar do critério ao buscar?)

3. Ser legível

- Espaços causam problemas de codificação

3. Ser legível

- Espaços causam problemas de codificação
- Caracteres especiais causam problemas de codificação

3. Ser legível

- Espaços causam problemas de codificação
- Caracteres especiais causam problemas de codificação
- Tudo junto não é legível

3. Ser legível

- Espaços causam problemas de codificação
- Caracteres especiais causam problemas de codificação
- Tudo junto não é legível
- Opções:

3. Ser legível

- Espaços causam problemas de codificação
- Caracteres especiais causam problemas de codificação
- Tudo junto não é legível
- Opções:
 - *Camel case* – [seminarioAbordagemMistaSlides.pdf](#)

3. Ser legível

- Espaços causam problemas de codificação
- Caracteres especiais causam problemas de codificação
- Tudo junto não é legível
- Opções:
 - *Camel case* – `seminarioAbordagemMistaSlides.pdf`
 - *Snake case* – `seminario_abordagem_mista_slides.pdf`

3. Ser legível

- Espaços causam problemas de codificação
- Caracteres especiais causam problemas de codificação
- Tudo junto não é legível
- Opções:
 - *Camel case* – `seminarioAbordagemMistaSlides.pdf`
 - *Snake case* – `seminario_abordagem_mista_slides.pdf`
 - Hífenes – `seminario-abordagem-mista-slides.pdf`

3. Ser legível

- Espaços causam problemas de codificação
- Caracteres especiais causam problemas de codificação
- Tudo junto não é legível
- Opções:
 - *Camel case* – `seminarioAbordagemMistaSlides.pdf`
 - *Snake case* – `seminario_abordagem_mista_slides.pdf`
 - Hífenes – `seminario-abordagem-mista-slides.pdf`
 - **Combinação desses casos – `2023-seminarioAbordagemMista-slides.pdf`**

4. Ser versátil e abrangente

- Vários arquivos com o mesmo nome/mesma inicial, mudando algum detalhe (data, versão, pessoa, local, etc.)

5. Ser sintaticamente correto

Pensar em como os arquivos/pastas ficarão organizados se posicionados alfabeticamente.

5. Ser sintaticamente correto

Pensar em como os arquivos/pastas ficarão organizados se posicionados alfabeticamente.

artigo-abralin-simposio08-2023.docx

artigo-abralin-simposio10-2023.docx

slides-abralin-simposio08-2023.pdf

slides-abralin-simposio10-2023.pdf

5. Ser sintaticamente correto

Pensar em como os arquivos/pastas ficarão organizados se posicionados alfabeticamente.

artigo-abralin-simposio08-2023.docx

artigo-abralin-simposio10-2023.docx

slides-abralin-simposio08-2023.pdf

slides-abralin-simposio10-2023.pdf

abralin-simposio08-artigo-2023.docx

abralin-simposio08-slides-2023.pdf

abralin-simposio10-artigo-2023.docx

abralin-simposio10-slides-2023.pdf

2023-abralin-simposio08-artigo.docx

2023-abralin-simposio08-slides.pdf

2023-abralin-simposio10-artigo.docx

2023-abralin-simposio10-slides.pdf

5. Ser sintaticamente correto

Pensar em como os arquivos/pastas ficarão organizados se posicionados alfabeticamente.

artigo-abralin-simposio08-2023.docx

artigo-abralin-simposio10-2023.docx

slides-abralin-simposio08-2023.pdf

slides-abralin-simposio10-2023.pdf

abralin-simposio08-artigo-2023.docx

abralin-simposio08-slides-2023.pdf

abralin-simposio10-artigo-2023.docx

abralin-simposio10-slides-2023.pdf

2023-abralin-simposio08-artigo.docx

2023-abralin-simposio08-slides.pdf

2023-abralin-simposio10-artigo.docx

2023-abralin-simposio10-slides.pdf

- Outras opções?

6. Não acarretar em problemas de codificação

- Enviar como ANSI e ser recebido como UTF-8 e vice-versa
- Ser baixados com % entre os espaços

→ Evitar espaços e caracteres especiais

→ Bônus

- Pasta ou Arquivo?

- Manualmente/Localmente
- Automaticamente (ex.: google docs)
- Intencionalmente na nuvem com ferramenta apropriada (ex.: git e github)

Mão na massa:

- (Re)Organizar suas pastas e (renomear) seus arquivos
- pelo menos aqueles em que está trabalhando atualmente

- Estudar/Aprender como funciona o versionamento do google docs
- Estudar/Aprender como funciona git, github e OSF

Compartilhamento de Dados

Crises de replicabilidade

- reprodutibilidade, replicação, crítica → avanço científico
- *“with great results come great responsibility”*

Crises de replicabilidade

- reprodutibilidade, replicação, crítica → avanço científico
- “*with great results come great responsibility*”
- Exemplo 1 [McElreath 2020]: 2015, periódico de alto impacto, 1170 crianças, associação negativa entre religiosidade e generosidade
 - países (categóricos) foram lidos como variável contínua
 - Canadá (país #2) foi interpretado como duas vezes “mais país” do que os EUA (país #1)
 - artigo removido, final feliz porque os dados foram compartilhados

Importância da ciência aberta

- Exemplo 2 [McElreath 2020]:
 - 2010: “Growth in a time of debt” by Reinhart & Rogoff
 - 2013: “Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff” by Herndon, Ash & Pollin



	B	C	I	J	K	L	M
2			Real GDP growth				
3			Debt/GDP				
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	30 or less
26			3.7	3.0	3.5	1.7	5.5
27	Minimum		1.6	0.3	1.3	-1.8	0.8
28	Maximum		5.4	4.9	10.2	3.6	13.3
29							
30	US	1946-2009	n.a.	3.4	3.3	-2.0	n.a.
31	UK	1946-2009	n.a.	2.4	2.5	2.4	n.a.
32	Sweden	1946-2009	3.6	2.9	2.7	n.a.	6.3
33	Spain	1946-2009	1.5	3.4	4.2	n.a.	9.9
34	Portugal	1952-2009	4.8	2.5	0.3	n.a.	7.9
35	New Zealand	1948-2009	2.5	2.9	3.9	-7.9	2.6
36	Netherlands	1956-2009	4.1	2.7	1.1	n.a.	6.4
37	Norway	1947-2009	3.4	5.1	n.a.	n.a.	5.4
38	Japan	1946-2009	7.0	4.0	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.0	5.6
40	Ireland	1948-2009	4.4	4.5	4.0	2.4	2.9
41	Greece	1970-2009	4.0	0.3	2.7	2.9	13.3
42	Germany	1946-2009	3.9	0.9	n.a.	n.a.	3.2
43	France	1949-2009	4.9	2.7	3.0	n.a.	5.2
44	Finland	1946-2009	3.8	2.4	5.5	n.a.	7.0
45	Denmark	1950-2009	3.5	1.7	2.4	n.a.	5.6

Novamente, final feliz porque os dados foram compartilhados (quando requisitados), mas...

E se os dados não tivessem sido compartilhados?

E os dados que nunca foram compartilhados, e erros nunca detectados?

Importância da ciência aberta

- Quando pesquisadores são contactados para compartilhar seus dados/arquivos:
 - não respondem
 - não têm os dados / todos os arquivos / os detalhes das análises
 - os arquivos estão desorganizados; não se lembram mais quais são os mais recentes, de fato utilizados na pesquisa

Importância da ciência aberta

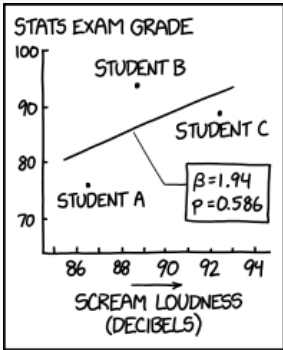
- Quando pesquisadores são contactados para compartilhar seus dados/arquivos:
 - não respondem
 - não têm os dados / todos os arquivos / os detalhes das análises
 - os arquivos estão desorganizados; não se lembram mais quais são os mais recentes, de fato utilizados na pesquisa
- O padrão deveria ser:
 - manter arquivos organizados (nomes dos arquivos, sistema de pastas, versionamento)
 - manter arquivos on-line, como back-up e para torná-los públicos



Mão na massa:

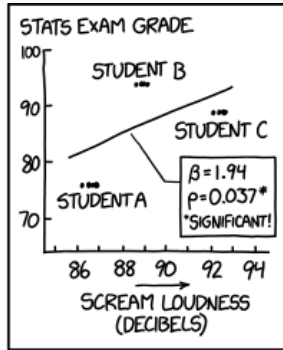
- Após organizar seus dados, pensar em como mantê-los públicos
 - github
 - OSF

Menos ênfase em valor de p



DARN, NOT SIGNIFICANT.

WE NEED MORE DATA.
HAVE THEM EACH TRY
YELLING INTO THE MIC
A FEW MORE TIMES.



PERFECT!

ARE YOU SURE
WE'RE DOING
SLOPE HYPOTHESIS
TESTING RIGHT?



Observações sobre valores de p

- Definição do valor de p (probabilidade dos dados vs probabilidade das hipóteses)

Observações sobre valores de p

- Definição do valor de p (probabilidade dos dados vs probabilidade das hipóteses)
- Problemas/limitações
 - O valor de p não diz nada sobre a hipótese de trabalho (alternativa)
 - O valor de p não diz nada sobre o tamanho do efeito
 - Limiar arbitrário
 - Decisão categórica sobre os dados (porém, comumente utilizada com gradência – de acordo com a intenção do pesquisador?)
 - É possível obter um valor de p baixo com baixo poder estatístico e/ou com pequeno tamanho de efeito
 - *p-hacking*

Observações sobre valores de p

- Definição do valor de p (probabilidade dos dados vs probabilidade das hipóteses)
- Problemas/limitações
 - O valor de p não diz nada sobre a hipótese de trabalho (alternativa)
 - O valor de p não diz nada sobre o tamanho do efeito
 - Limiar arbitrário
 - Decisão categórica sobre os dados (porém, comumente utilizada com gradência – de acordo com a intenção do pesquisador?)
 - É possível obter um valor de p baixo com baixo poder estatístico e/ou com pequeno tamanho de efeito
 - *p-hacking*
- Modelos com efeitos mistos não geram valores de p (Bates), e diferentes aproximações geram valores diferentes (e.g. `lmerTest` vs `sjPlot`)

Observações sobre valores de p

```
1 | lmerTest::fit2 = lmer(sum ~ recording + (recording|speaker))  
2 | summary(fit2)
```

Predictors	Estimates	p
Intercept	1.16	<0.001
recording	0.14	0.068

1. Satterthwaite's method
2. t-statistics and the normal distribution function
3. conditional F-test with Kenward-Roger approximation

```
1 | sjPlot::tab_model(fit2)
```

Predictors	Estimates	p
Intercept	1.16	<0.001
recording	0.14	0.048

```
1 | sjPlot::tab_model(fit2, p.val = "kr")
```

Predictors	Estimates	p
Intercept	1.16	<0.001
recording	0.14	0.071

Diferentes análises estatísticas podem levar a conclusões categoricamente distintas

Ronaldo Manguera Lima Jr, Guilherme Duarte Garcia

V. 20, N. 1 (2021)

Enviado: dez 28, 2020

Publicado: ago 5, 2021

DOI 10.25189/rabralin.v20i1.1790



<https://revista.abralin.org/index.php/abralin/article/view/1790>

Alternativas:

- Reportar e analisar tamanhos de efeito e intervalos
- Apresentar e utilizar na interpretação a visualização dos dados (descritivos e previstos)
- Utilizar estatística bayesiana

Mão na massa:

Menos ênfase em valor de p

- Estudar:
 - (i) tamanho de efeito, intervalos (de confiança, erro-padrão) e poder estatístico
 - (ii) modelos de regressão
 - Livro Guilherme Garcia
 - Cursos Ronaldo Lima Jr. e Guilherme Garcia
 - (I) visualização de dados (ggplot)
 - Guia da Royal Statistical Society
 - Curso (2023) Ronaldo Lima Jr. e Guilherme Garcia
 - (I) estatística bayesiana
 - Curso Richard McElreath – YouTube

Perguntas?