

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

A importância da análise quantitativa de dados em Linguística

Ronaldo Lima Jr.

`ronaldojr@letras.ufc.br`

`ronaldolimajr.github.io`

Universidade Federal do Ceará

1. Análise Quantitativa de Dados em Linguística
2. Objetivos da Análise Quantitativa de Dados
3. Ciência Aberta

Análise Quantitativa de Dados em Linguística

- Séc. XX → Estruturalismo e Gerativismo
 - Objeto da Linguística como estrutura invariável, sistema de oposições, inato, discreto e categórico
 - Levou a introspecção e dados/exemplos autocunhados

“Large groups of people make up all their utterances out of the same stock of lexical forms and grammatical constructions. A linguistic observer therefore can describe the speech-habits of a community without resorting to statistics” (Bloomfield 1935: 37)

“I think we are forced to conclude that grammar is autonomous and independent of meaning, and that probabilistic models give no particular insight into some of the basic problems of syntactic structure” (Chomsky 1957: 17)

- Abordagens baseadas no uso, variação e mudança
- Dados de corpora, questionários, experimentos, correlatos físicos, construtos da psicologia, etc.

- psicolinguística, linguística de corpus, fonética, sociolinguística, linguística aplicada, aquisição de L2, linguística forense, aprendizagem de máquinas, etc.
 - e mesmo subáreas “tradicionais” que não utilizavam (sintaxe, fonologia, etc.)

Possíveis barreiras / motivos de resistência

- *“Não sou bom de matemática / não sou bom com números”*
- *“Fiz Letras justamente pra não precisar de matemática”*
- *“Minha pesquisa é qualitativa”*
- Curva de aprendizagem
- Falta de tempo
- Desmotivação
- Na minha (sub)área não é necessário
- Etc...

Alguns motivos para se capacitar em análise quantitativa

- Conseguir ler/consumir pesquisa
- A pesquisa qualitativa facilmente se esbarra em dados quantitativos (questionários, quantidade de observações, etc.)
- O raciocínio quantitativo ajuda na pesquisa como um todo:
 - como operacionalizar (medir/testar/observar) os dados
 - identificar as variáveis
 - pensar em causalidade
 - pensar em variáveis de confusão (*confounding variables*)
 - pensar em hipóteses e em falseabilidade (Karl Popper)
 - motivação/idade e aprendizagem de línguas

Objetivos da Análise Quantitativa de Dados

Objetivos da análise quantitativa de dados

- Descrever
- Explicar
- Prever

Problema → Pergunta(s) → Hipótese(s)
→ Verificar/Observar/Testar → Inferência/Conclusão

- Ex. 1: encontrar o melhor preço para um produto
- Ex. 2: motivos de lentidão em um novo percurso

“Pesquisa mostra que correr pode ser prejudicial a saúde”

- identificar/caracterizar o problema
- estudar a literatura (teoria, construtos, métodos, resultados, variáveis, lacunas, etc.)
- observar o fenômeno e raciocínio dedutivo por parte do pesquisador
- compilar variáveis passíveis de influência
- estipular hipóteses
 - falsiáveis e testáveis
 - identificar/vislumbrar covariáveis (*confounding variables*)
- pensar em como operacionalizar as variáveis (observar, medir, contar)

→ Tudo isso **ANTES** de coletar os dados!

- *HARKing*
- *p-hacking*
- crises de replicabilidade
- ciência aberta
- Pré-registro
- OSF, Github
 - Palestra do Brian Nosek no Interab12

Ciência Aberta

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



2	B	C	I	J	K	L	M
3			Real GDP growth 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
46	Canada	1951-2009	1.9	3.6	4.1	n.a.	2.2
47	Belgium	1947-2009	n.a.	4.2	3.1	2.6	n.a.
48	Austria	1948-2009	5.2	3.3	-3.8	n.a.	5.7
49	Australia	1951-2009	3.2	4.9	4.0	n.a.	5.9
50							
51			4.1	2.8	2.8	=AVERAGE(L30:L44)	

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
- 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



Perguntas?