

Aula 1: Introdução

João Paulo Lazzarini Cyrino

Agosto de 2020

O trabalho com métodos estatísticos em Linguística

Neste curso vamos aprender o básico para trabalhar com estatística em linguística, especialmente para as áreas de Tipologia Linguística e Gramática (morfologia e sintaxe).

A estatística está presente na linguística já há algum tempo, inclusive trazendo suas contribuições para a área. Na sociolinguística, por exemplo, desenvolveram o programa Varbrul (ver Sankoff and Cedergren 1976), que é considerado uma das primeiras implementações de *regressão logística* mais acessíveis ao público científico não especializado em estatística computacional. Para além da sociolinguística, áreas como fonética e psicolinguística estão constantemente utilizando métodos estatísticos.

Em tipologia linguística, o uso de estatística foi se tornando mais comum ao longo do tempo e especialmente impulsionado pelas contribuições de Johanna Nichols (Nichols 1992) e Balthasar Bickel (Bickel 2011), que apontaram para a necessidade de se investigar universais probabilísticos. Atualmente, a estatística é bastante comum na área.

Já nos estudos de gramática tem havido maior presença da estatística mais recentemente. Mas, ainda assim, métodos quantitativos ainda são raros na área e sofrem alguma rejeição.

Por que devo utilizar estatística em meu trabalho sobre (ex. negação pós-verbal no PB)?

Mais do que somente utilizar a estatística no seu trabalho, aprender sobre intervalos de confiança e testes de hipótese (por exemplo) certamente fará você pensar melhor em como organizar seus dados, o que observar neles e quantos dados levantar. Enfim, a estatística estuda justamente formas de organizar e tirar conclusões sobre dados. Se você trabalha com dados, sempre será benéfico saber um pouco de estatística.

Se você está fazendo um trabalho descritivo, a estatística é muito importante. Você depende de coletar dados de uma língua e essa coleta deve seguir um método para que seus dados correspondam a uma amostra a partir da qual possamos tirar conclusões sobre como a língua funciona com respeito ao fenômeno estudado. A estatística dá as ferramentas para isso.

Se você está fazendo um trabalho tipológico, a estatística vai ajudar a determinar o número de línguas que deve ser investigado e o método de escolha (amostragem) dessas línguas. Além disso, com a correta amostragem, é possível fazer boas inferências sobre o comportamento das línguas quanto às diversas propriedades estudadas.

Trabalhos de teoria da gramática em si podem não ser tão beneficiados pela estatística caso não adotem uma visão estocástica da gramática. Mas, teorias são propostas a partir de observações/descrições. Se estas são feitas a partir de bons métodos, maiores as chances de se desenvolver melhores teorias. Além disso, há todo o campo de *teorias estocásticas de gramática* (Clark 2005). Aqui sim, a estatística tem papel principal.

O que vou aprender aqui?

Este é um curso bastante básico, mas acredito que pode ser útil para desenvolver algumas capacidades importantes para descrição e tipologia linguística. O programa do curso está organizado de forma a priorizar

as coisas mais urgentes o quanto possível, e desenvolver outras capacidades mais ao fim. Não haverá um grande enfoque na teoria matemática, mas é importante ter disposição para entender algumas das abstrações.

1. Introdução (esta aula)
2. Instalando e conhecendo a linguagem R
3. Estatística descritiva em R e pacote Tidyverse
4. Intervalos de Confiança e Amostras
5. Introdução aos Testes de Hipótese com Teste Qui-Quadrado
6. Simulando Intervalos de Confiança e o Pacote Infer
7. Outros testes de Hipótese Envolvendo Proporções
8. Testes de Hipótese Envolvendo Médias e Correlação Linear
9. Regressão Logística

No entanto, mais importante que esse programa, é entender as **capacidades** adquiridas aqui:

- Planejar um estudo, definindo bem minhas hipóteses.
- Coletar dados de forma que possam ser aproveitados no meu estudo.
- Organizar os dados em planilhas
- Saber o que observar nos dados
- Definir boas metodologias de estudo
- Criticar metodologias em outros estudos e replicá-los com melhor metodologia
- ...

Como se pode observar, o estudo de métodos estatísticos pode trazer um salto qualitativo em sua pesquisa e também te tornar um linguista mais produtivo. O esforço vale a pena!

Referências

- Bickel, Balthasar. 2011. “Statistical Modeling of Language Universals.” *Linguistic Typology* 15 (2): 401–14.
- Clark, Brady. 2005. “On Stochastic Grammar.” *Language* 81 (1): 207–17.
- Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. University of Chicago Press.
- Sankoff, David, and Henrietta J Cedergren. 1976. “The Dimensionality of Grammatical Variation.” *Language*, 163–78.