

## A ANÁLISE DAS REGRAS VARIÁVEIS

### **META**

Apresentar tutorial de investigação quantitativa

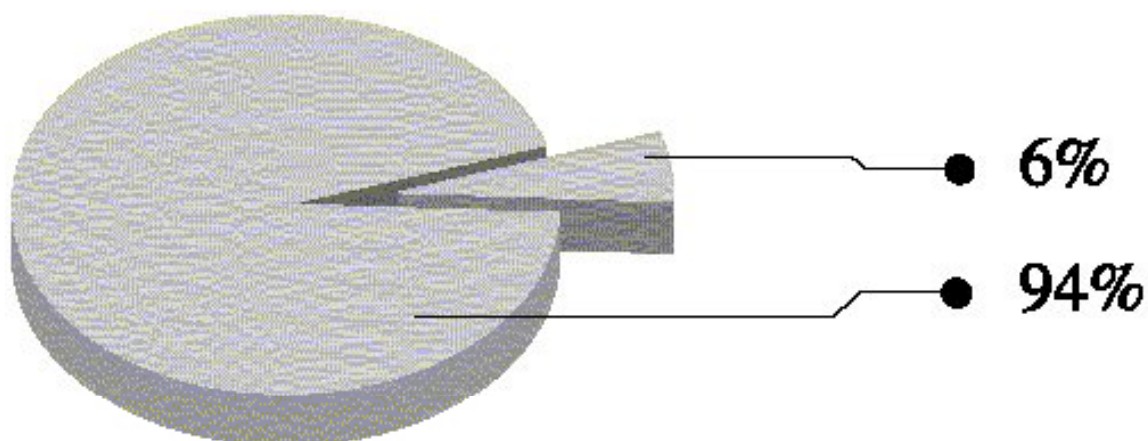
### **OBJETIVOS**

Ao final desta aula, o aluno deverá:

compreender e realizar investigações sociolinguísticas utilizando modelos quantitativos e análise estatística, a partir do estudo do rotacismo na fala de Ribeirópolis/SE

### **PRÉ-REQUISITOS**

Ter gravado e transcrito uma entrevista sociolinguística.



(Fontes: <http://www.scielo.org.ve>)

## INTRODUÇÃO

Nesta aula, vamos ver um tutorial para a pesquisa Sociolinguística Variacionista, ou, como Tarallo (1985) bem define, Sociolinguística Quantitativa, pois os resultados das investigações neste campo são baseados em frequências. Você já deve ter observado, nas leituras sugeridas, que frequentemente nos deparamos com tabelas e gráficos, com dados de frequências e percentuais. Veremos como são obtidos estes números, de modo que você, ao final desta aula, entenderá melhor o que significam as frequências e os percentuais e o que eles indicam quando se trata de um fenômeno de variação e mudança.

The screenshot shows the GoldVarb X software interface. The main window displays a table with the following data:

Group		1	0	Total	%
-----					
1 (2)					
a	N	16	4	20	33
	%	80	20		
n	N	4	16	20	33
	%	20	80		
d	N	10	10	20	33
	%	50	50		
Total	N	30	30	60	
	%	50	50		
-----					

Summary statistics shown above the table:

- Number of cells: 6
- Application value(s): 10
- Total no. of factors: 5

The style menu on the right includes options: Plain,  Bold, *Italic*, Underline, Outline, Shadow, Condense, and Extend. A font size list is also visible, with 12 selected.

Imagem da tela do programa GoldVarb X, utilizado na análise dos dados de coletas sociolinguísticas.  
(Fontes: <http://albuquerque.bioinformatics.uottawa.ca>)

## COLETA E CODIFICAÇÃO DE DADOS

Depois de termos escolhido o fenômeno variável que iremos estudar, de termos feitos pesquisas intuitivas e bibliográficas a fim de que se depreendam fatores linguísticos e sociais que podem influenciá-lo, e de constituir a amostra que irá subsidiar a análise, você precisará partir para a parte prática do trabalho que envolve a matemática. Afinal, a Sociolinguística trabalha com frequências de uso, que delinham padrões de variação e mudança na comunidade.

Então, mãos à obra!

Vamos ilustrar com um exemplo: o rotacismo em Ribeirópolis (SANTANA et alii, 2008). Você sabe o que é rotacismo? Você deve ter ouvido falar deste fenômeno na disciplina de Fonologia da língua portuguesa. Mas se não ouviu falar do fenômeno, certamente já ouviu o fenômeno em si: quando alguém diz “pranta” ou “fror”, em suma, nos contextos em que ocorre a troca do /l/ por /r/. A passagem /l/ > /r/ (planta > pranta) recebe o nome técnico de rotacismo. A neutralização do /r/ e /l/ no padrão silábico CCV (C significa “consoante” e V significa “vogal”) pode ser um indicador de variedades rurais e urbanas, um indicador regional e pode, ainda, ser um problema fono-articulatório. (BORTONI-RICARDO, 2004, p. 83-84).

As consoantes /l/ e /r/ são, do ponto de vista articulatorio, muito próximas e podem, por isso, intercambiar-se ou fundir-se na história das línguas. Na literatura linguística, são designadas como consoantes líquidas, o que revela o parentesco com as vogais. Apesar da semelhança do ponto de vista articulatorio e fonotático, as consoantes /l/ e /r/ constituem fonemas distintos no português, uma vez que podem contrastar em ambiente idêntico e, aplicando-se o princípio da comutação em pares mínimos, observa-se que a troca de /l/ por /r/ implica em mudança de significado. Apesar disto, no ambiente CVV apontado acima, ambas as consoantes podem funcionar como o mesmo valor de verdade, no mesmo contexto, configurando, assim, uma regra variável.

Vamos investigar, portanto, a neutralização do fonema /l/ por /r/ na comunidade de Ribeirópolis/SE. A coleta de dados foi feita em turmas da 4ª série (atual 5º ano) de escolas do município, na zona rural e urbana.

Segundo Marcos Bagno (2007), para os falantes urbanos escolarizados, pronúncias como broco, ingrês, chicrete, pranta, etc. são “feias”, “erradas”. Essa avaliação se prende essencialmente ao fato de essas pronúncias caracterizarem falantes socialmente desprestigiados (analfabetos, pobres, moradores de zona rural, etc). (BAGNO, 2007, p. 73-74). Assim, dado que o fenômeno é estigmatizado, a hipótese que norteia a análise é que haja maior produtividade da variante estigmatizada nas escolas da zona rural.

Para fins de ilustração, vamos controlar, além da variável dependente,

uma variável linguística e uma variável social. A variável linguística escolhida foi o item lexical utilizado como tema na coleta; a variável social foi a escola onde foi realizada a coleta.

Vamos, agora, estabelecer códigos para cada um dos fatores das variáveis. Estes códigos nos auxiliarão na análise quantitativa.

Variável dependente:

R – rotacismo

L – variante-padrão

Variáveis independentes:

Variável social – escolas: G, J, A, C

A escola G, Escola Municipal José Gonçalves dos Santos, localizada na zona rural do município, foi construída no ano de 1970, possui quatro professores e 43 alunos, divididos em cinco turmas da pré-escola a quarta série do Ensino Fundamental. A escola J (Colégio Municipal Josué Passos) pertence à zona urbana, no centro da cidade. É uma das escolas mais antigas da cidade com 34 anos, composta também de discentes da zona rural e de outras cidades vizinhas. A escola A corresponde à Escola Municipal Maria Alaíde Meneses, localizada no subúrbio da cidade, numa comunidade bastante carente economicamente. E a escola C, Centro de Educação Básica Auxiliadora Paes Mendonça, é uma escola pública mantida por uma empresa privada, situada na zona rural de Ribeirópolis, possui nove professores e 235 alunos.

Variável linguística – item lexical:

P – Placa

B – Bloco

T – Planta

S – Blusa

F – Flor

G – Globo

Eis nosso “envelope da variação”. Cada dado coletado na amostra será cotejado a cada um destes fatores, a fim de que possamos, depois, tabulá-los e chegar às generalizações estatísticas. Vamos analisar um exemplo:

A palavra [bluzə] (“blusa”), dita por um aluno da escola Escola Municipal José Gonçalves dos Santos:

LGS

Onde: L = realização padrão /l/, G = Escola Municipal José Gonçalves dos Santos e S = item lexical “blusa”

Ou então a palavra [brɔku] (“bloco”) dita por um aluno da Escola Municipal Maria Alaíde Meneses:

RAB

Onde: R = realização não padrão /r/, A = Escola Municipal Maria Alaíde Meneses e B = item lexical “bloco”.

Devemos repetir este mesmo procedimento, com todos os dados. Esta será a entrada para o pacote estatístico GoldVarb X (2001) ou Varbrul (1988), um pacote estatístico específico para a análise da variação linguística, que opera com a noção de regras variáveis. Não vamos entrar em detalhes sobre o funcionamento deste software: tutoriais podem ser encontrados em Guy e Zilles (2007) ou em <http://courses.essex.ac.uk/lg/lg654/GoldVarb2001forPCmanual.htm> (em inglês). O importante nesta aula é entender como se dá o procedimento.

## CÁLCULOS ESTATÍSTICOS

Após a codificação, os dados são submetidos à análise quantitativa, que retorna a distribuição de cada fator de cada variável em termos de frequências e percentuais.

Para o rotacismo, foram encontradas 893 ocorrências de contextos CCV em que poderia haver neutralização de /l/ e /r/. Destas 893 ocorrências, 832 foram da forma padrão /l/ e 61 da forma não padrão /r/. Dizemos, então, que o percentual de aplicação da regra foi de 93% (das 893 ocorrências, 832 são de aplicação da regra, o que totaliza 93%).

O que significa ter 93% de aplicação? Podemos dizer que é a aplicação uma regra produtiva, pois o resultado está muito próximo de 100%, do total. Ou seja, embora tenhamos a impressão de que o rotacismo é recorrente, quando analisamos empírica e cientificamente, constatamos que esta variante corresponde a apenas 7% das ocorrências.

Vamos, agora, observar os resultados referentes ao item lexical onde pode ocorrer a neutralização de /l/ e /r/.

Tabela 5: Distribuição das ocorrências de rotacismo na fala de Ribeirópolis em função do item lexical

Item lexical	Aplicação/total	Percentual	Peso Relativo
Placa	145/160	91%	0,28
Bloco	106/120	88%	0,22
Planta	127/128	99%	0,83
Blusa	127/147	86%	0,20
Flor	175/176	99%	0,88
Globo	152/162	94%	0,39
Total	832/893	93%	—

Fonte: SANTANA et alii, 2008, p. 151

Vejam os resultados. A última linha da tabela recupera o valor de referência para a amostra como um todo, ou seja, 93% de aplicação da regra. Das 162 ocorrências do item lexical “globo”, 152 foram com a variante /l/ (e, portanto, 10 com a variante /r/), totalizando 94% de aplicação da regra, percentual muito próximo do de “placa”, com 91%. Já os itens lexicais “blusa” e “bloco” são os itens onde o percentual de aplicação da regra é menor (e, conseqüentemente, é onde o percentual da variante estigmatizada /r/ é mais recorrente). Já em “planta” e em “flor”, a aplicação da regra é quase categórica (a aplicação categórica é 100%).

Você deve ter notado a terceira coluna de valores, chamada de peso relativo (ou probabilidade). O modelo estatístico adotado pela Sociolinguística Variacionista trabalha com a regressão logística com o cálculo de desvio da média ponderada. A regressão logística é um modelo utilizado para investigar, dentro de um conjunto de possíveis variáveis independentes contínuas ou discretas, quais estão associadas à ocorrência da variável dependente discreta. Nas análises sociolinguísticas, o desvio da média ponderada é chamado de **peso relativo**. Bem, para quem escolheu Letras porque não gostava de matemática, as coisas vão ficando um pouco complicadas, mas nem tanto.

**Em *Modelos quantitativos e tratamento estatístico*, (NARO, 2003) você encontrará uma descrição detalhada do modelo quantitativo adotado pela Sociolinguística Variacionista**

Quando lançamos uma moeda para o alto, a chance de sair cara ou coroa é a mesma: ou sai cara ou sai coroa (não sei de nenhum caso em que a moeda parou de pé...). Em termos de probabilidade, temos duas possibilidades em um lance: 1 dividido por 2 = 0,50 ou 50% (50% + 50% = 100%). Se lançarmos duas vezes a moeda, as chances continuam sendo de 50%, se lançarmos 20 vezes, as chances ainda continuam sendo as mesmas. Ou seja, lançar uma moeda pode dar cara ou coroa, por acaso.

Você já ouviu dizer que o pão sempre cai com a margarina para baixo? Cientistas dizem que é porque, como há mais peso num lado do pão, o da margarina, este é atraído para baixo por causa da gravidade. Se jogarmos uma vez um pão para o alto, a chance de o lado da margarina ficar para baixo é levemente maior do que o contrário. E se jogarmos 20 vezes para o alto, talvez em 2/3 das vezes o pão caia com a margarina para baixo. A probabilidade não é a mesma da moeda, perto do acaso: existe um fator atrator. O cálculo do desvio da média ponderada, na análise estatística da Sociolinguística Variacionista, tenta equilibrar os percentuais, indicando a probabilidade de ocorrência ou não do fenômeno. O ponto de equilíbrio – ou ação do acaso – é o 0,50. Perto deste valor, o fator analisado não exerce

nenhuma influência na variação, é obra de outro fator ou do acaso. Quanto mais perto do 1, maior a probabilidade de ocorrência do fenômeno, e quanto mais perto de 0, menor probabilidade de ocorrência do fenômeno.

Voltando à análise do rotacismo, os pesos relativos de 0,88 e 0,83, respectivamente para “flor” e “planta”, dizem que a probabilidade de aplicação da regra (realização da variante /l/) é muito alta, pois estes valores estão se aproximando de 1. Já para “blusa” e “bloco”, a probabilidade de aplicação da regra tende a ser reduzida, já que os pesos relativos, 0,20 e 0,22, respectivamente, estão tendendo para o 0. Ou seja, estes itens lexicais favorecem a ocorrência do rotacismo.

Vamos ver agora os resultados das escolas:

Tabela 6: Distribuição das ocorrências de rotacismo na fala de Ribeirópolis em função da escola

Escolas	Aplicação/Total	Percentual	Peso Relativo
G	69/72	96%	0,64
A	86/99	87%	0,32
J	352/371	95%	0,59
C	325/351	93%	0,42
Total	832/893	93%	—

Fonte: SANTANA et alii, 2008, p. 152

Os resultados quanto à escola apontam para duas direções: as escolas que tendem a aplicar a regra – Escola Municipal José Gonçalves dos Santos (G) e Colégio Municipal Josué Passos (J), cujos pesos relativos são superiores a 0,50, indicando a tendência de uso da variante /l/; e as escolas Escola Municipal Maria Aláide Meneses (A) e Centro de Educação Básica Auxiliadora Paes Mendonça (C), cujos pesos relativos abaixo de 0,50 indicam a probabilidade de não aplicação da regra, ou seja, favorecimento do uso da variante /r/.

Agora que já vimos o que significam os números que aparecem nas análises sociolinguísticas, vamos passar para a interpretação dos resultados. O que os números indicam?

## INTERPRETAÇÃO DOS RESULTADOS

No início da seção 1, dissemos que a hipótese norteadora da investigação do rotacismo em Ribeirópolis era de que haveria maior produtividade da variante estigmatizada nas escolas da zona rural. O que os números nos dizem?

Com os dados apresentados na tabela 6 e o perfil das escolas, é possível perceber que a hipótese não foi de todo corroborada, a exemplo da

escola Escola Municipal José Gonçalves dos Santos que, apesar de seu perfil potencialmente favorecer o uso da variante estigmatizada, os dados mostram a tendência ao uso da variante padrão. Ou seja, mesmo os alunos tendo contato com a variante não padrão em sua casa, a atuação do professor na escola pode barrar a variante estigmatizada. Já no caso da escola Escola Municipal Maria Alaíde Meneses, ao contrário da escola Escola Municipal José Gonçalves dos Santos, a hipótese foi confirmada, pois os pesos relativos indicam o desfavorecimento da aplicação da regra (ou seja, o favorecimento da forma não padrão /r/). Em suma, a escola parece ter uma influência significativa na ocorrência do fenômeno, na medida em que barra a forma não-padrão.

Quanto ao item lexical, a produtividade quase categórica da variante padrão deu-se em “flor” e “planta”, possivelmente porque são as palavras mais trabalhadas nas séries iniciais. Com o contato com a norma, o aluno adéqua sua variedade ao imposto como padrão na escola.

A metodologia quantitativa da Sociolinguística Variacionista constitui uma ferramenta segura para o estudo de qualquer fenômeno variável, nos diversos níveis de análise e em diferentes comunidades. A metodologia quantitativa reduz os dados e possibilita a comparação, permitindo as generalizações.

## CONCLUSÃO

Após esta aula, percentuais, frequências e pesos relativos não devem ser mais nenhum bicho de sete cabeças para você. Vimos como estes números são obtidos e de que modo eles contribuem para a investigação. Seguindo os passos desta aula, você não só está apto a realizar as suas próprias análises quantitativas, dos fenômenos que você desejar investigar, como também está instrumentalizado para compreender o raciocínio metodológico evocado nas análises sociolinguísticas.





## RESUMO

Nesta aula, vimos como são obtidos os números que sempre encontramos em textos da Sociolinguística, ou seja, como obtemos as frequências e percentuais e o que estes números indicam em um fenômeno de variação e mudança. Para tanto, analisamos um estudo feito na cidade de Ribeirópolis/SE sobre o fenômeno do rotacismo (passagem do /l/ > /r/) em tal cidade (SANTANA, et alii, 2008). Vimos que, após a codificação, os dados são submetidos à análise quantitativa, que retorna a distribuição de cada fator de cada variável em termos de frequências e percentuais. Depois da codificação dos dados, passamos para a interpretação das frequências e percentuais. As interpretações dos dados são feitas pelo pesquisador, pois as análises estatísticas são apenas ferramentas para a manipulação dos dados. Com esta aula, você instrumentalizado para fazer sua própria investigação sociolinguística e compreender outras investigações sociolinguísticas já realizadas.

## ATIVIDADES

1. Analise as tabelas abaixo. Foram considerados os contextos de variação /t/ ~ /t [ f ]/ e /d/ ~ /d [ ʒ ]/ antecidos de semivogal (em palavras como muito, doido, prefeito, peito) (MOTA, 2008).



Tabela 1: Palatalização de /t/ e /d/ em contextos antecidos de semivogal em Aracaju quanto à faixa de escolarização

Capital	Escolaridade	Aplicação/total	Percentual	Peso relativo
Aracaju	Fundamental	80/168	48%	0,81
	Universitário	4/148	3%	0,12

Tabela 2: Palatalização de /t/ e /d/ em contextos antecidos de semivogal em Aracaju quanto ao sexo/gênero dos informantes

Capital	Gênero	Aplicação/total	Percentual	Peso relativo
Aracaju	Masculino	40/136	29%	0,61
	Feminino	44/180	24%	0,54

Descreva os resultados obtidos. Os dados quantitativos, correlacionados às hipóteses sociais (volte à aula 5 e veja as hipóteses sociais clássicas) apontam para o prestígio ou para o estigma do fenômeno?

2. Leia o texto Sobre a concordância de número no português falado do Brasil (SCHERRE; NARO, 1998) seguindo o roteiro de leitura abaixo, a fim de consolidar os conceitos aprendidos até então, e articulá-los aos dados quantitativos.

i) Qual é a regra de concordância de número prescrita pela gramática normativa? E qual é a regra da concordância de número apresentada por Scherre e Naro (1998)?

ii) Quais são os objetivos da investigação de Scherre e Naro (1998)?

iii) Quais são as variáveis linguísticas analisadas? Quais as suas variantes?

iv) Qual a amostra utilizada pelos autores para a análise? Apresente a estratificação social da amostra.

v) Qual o total de dados analisados? Como ele se distribui pelas variáveis?

vi) Qual o caráter empírico da pesquisa: qualitativo ou quantitativo? Justifique.

vii) O que é saliência fônica?

viii) O que os resultados estatísticos apontam sobre a relação entre concordância de número e saliência fônica? Justifique com dados numéricos.

ix) Qual a relação que pode ser estabelecida entre a saliência fônica e a escolaridade dos indivíduos no que se refere à concordância de número, segundo o estudo de Scherre e Naro (1998)?

x) Como se dá a relação entre concordância e saliência fônica dentro do sintagma nominal? Justifique com dados numéricos.

xi) Explique o que é a variável posição.

xii) O que os resultados estatísticos apontam sobre a relação estabelecida entre a posição e a concordância de número? Justifique com dados numéricos.

xiii) Qual a influência do fator sexo no fenômeno de concordância de número?

xiv) Qual a influência do fator escolaridade no fenômeno de concordância de número?

xv) Qual a influência do fator faixa etária no fenômeno de concordância de número?

xvi) Qual o perfil do falante que mais tende à aplicação da regra (variante explícita) da concordância de número?

xvii) Como se dá o fenômeno de concordância de número na escrita?

xviii) A que conclusões chegam Scherre e Naro (1998) sobre o fenômeno da concordância de número no português?

## COMENTÁRIO SOBRE AS ATIVIDADES

Quem acompanhou as aulas com atenção e realizou todas as atividades, não terá nenhuma dificuldade em ler o texto de Marta Scherre e Anthony Naro sobre a concordância de número do português do Brasil. Como este texto foi feito para um congresso no exterior, é muito didático em termos de explicação de regras do funcionamento da língua. Você verá as evidências quantitativas que respaldam a argumentação dos autores. E entenderá, com muita clareza, o funcionamento da regra de concordância no português falado no Brasil.

Quanto à variação de /t/ ~ /t [ʃ]/ e /d/ ~ /d [ʒ]/ antecidos de semivogal, você deve observar a tendência de restrição do grupo de falantes com maior escolaridade e das mulheres, o que indica um fenômeno estigmatizado.

## PRÓXIMA AULA

Depois de conhecermos o campo da Sociolinguística Variacionista, veremos, na próxima aula, como esta pode ajudar no dia a dia da sala de aula, em A importância do tratamento da variação no ensino de Língua Portuguesa.



## AUTOAVALIAÇÃO

Após esta aula, sou capaz de empreender uma análise variacionista e interpretar os dados quantitativos obtidos em termos de direcionais de variação e mudança?



## REFERÊNCIAS

BAGNO, Marcos. **Nada na Língua é por acaso**: por uma pedagogia da variação linguística. São Paulo: Parábola, 2007.

BORTONI-RICARDO, Stella Maris. **Educação em língua materna**: a sociolinguística na sala de aula. São Paulo: Parábola, 2004.

GUY, Gregory; ZILLES, Ana Maria Stahl. **Sociolinguística quantitativa**: instrumental de análise. São Paulo: Parábola, 2007.

MOTA, Jacyra. **Como fala o nordestino**: a variação fônica no Atlas Linguístico do Brasil. In: **Anais do I Simpósio Mundial de estudos de Língua Portuguesa**. 2008. Disponível em: <<http://www.fflch.usp.br/eventos/simelp/new/pdf/slp22/03.pdf>>

NARO, Anthony, Julius. Modelos quantitativos e tratamento estatístico. In:

- MOLLICA, Maria Cecília; BRAGA, Maria Luiza. **Introdução à sociolinguística**: o tratamento da variação. Rio de Janeiro: Contexto, 2003, p.15-26.
- SANTANA, Alcilânia Rezende; DANTAS, Juciane Jesus; SANTANA, Maria José Barreto; GUEDES, Maria Lima; FREITAG, Raquel Meister Ko. O tratamento do rotacismo nas séries iniciais da rede municipal de ensino de Ribeirópolis. In: SILVA, Leilane Ramos da; FREITAG, Raquel Meister Ko. (Org.). **Linguagem e representação discursiva**. João Pessoa: Ed. UFPB, 2008, p. 147-154.
- SCHERRE, Maria Marta Pereira; NARO, Anthony Julius. Sobre a concordância de número no português falado do Brasil. In RUFFINO, Giovanni (org.) **Dialettologia, geolinguística, sociolinguística (Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza)**. Centro di Studi Filologici e Linguistici Siciliani, Università di Palermo. Tübingen: Max Niemeyer Verlag, 1998, p.509-523.
- TARALLO, Fernando. **A pesquisa sociolinguística**. São Paulo: Ática, 1985.