

# Análise Quantitativa de Dados em Linguística

## Correlação

---

**Ronaldo Lima Jr.**

`ronaldojr@letras.ufc.br`

`ronaldolimajr.github.io`

Universidade Federal do Ceará

1. Correlação
2. Cuidado!

# Correlação

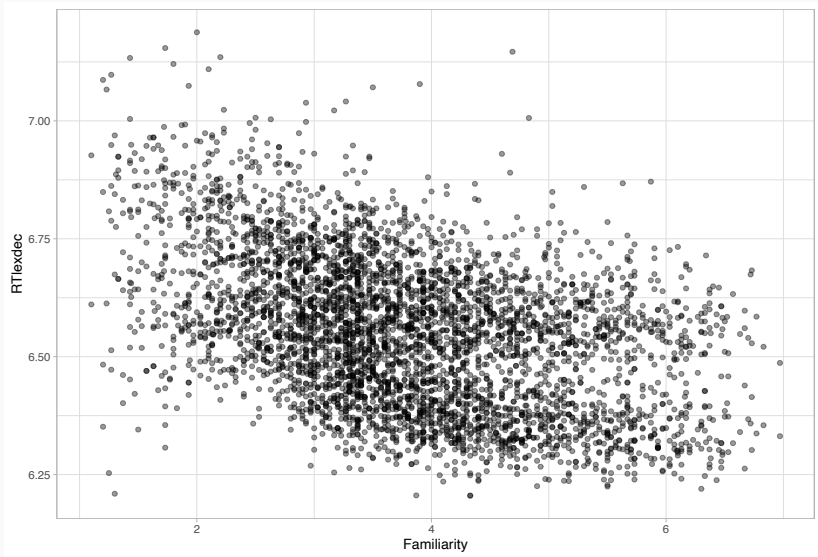
---

# Correlação

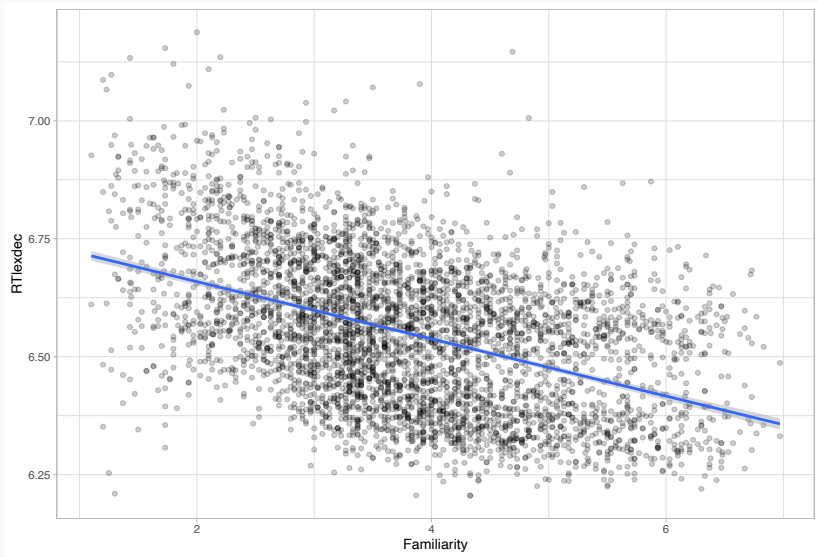
- Variável preditora e variável resposta contínuas
- Verificar o quanto as duas variáveis co-ocorrem
- **Exemplo:** Nos dados “english”, há relação entre tempo de reação e familiaridade do participante com a palavra?
  - $H_1$  Palavras mais familiares levam a tempos de reação mais baixos
  - $H_0$  Não há relação

|   | RTlexdec | Word        | Familiarity | AgeSubject |
|---|----------|-------------|-------------|------------|
|   | <dbl>    | <fct>       | <dbl>       | <fct>      |
| 1 |          |             |             |            |
| 2 |          |             |             |            |
| 3 | 1        | 6.54 doe    | 2.37        | young      |
| 4 | 2        | 6.40 whore  | 4.43        | young      |
| 5 | 3        | 6.30 stress | 5.6         | young      |
| 6 | 4        | 6.42 pork   | 3.87        | young      |
| 7 | 5        | 6.45 plug   | 3.93        | young      |
| 8 | 6        | 6.53 prop   | 3.27        | young      |
| 9 | 7        | 6.37 dawn   | 3.73        | young      |

# Correlação



# Correlação – com uma linha de tendência

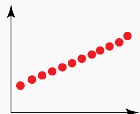


# Teste de correlação

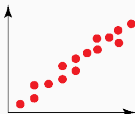
```
1 cor.test(english$RTlexdec, english$Familiarity)
2
3      Pearson's product-moment correlation
4
5 data:  english$RTlexdec and english$Familiarity
6 t = -33.522, df = 4566, p-value < 2.2e-16
7 alternative hypothesis: true correlation is not equal to 0
8 95 percent confidence interval:
9  -0.4673865 -0.4208330
10 sample estimates:
11      cor
12 -0.4444097
```

- correlação positiva, linha sobe; correlação negativa, linha desce
- $r = 1 \rightarrow$  correlação perfeita, todos os pontos co-ocorrem
- $r = 0.75 \rightarrow$  correlação forte
- $r = 0.5 \rightarrow$  correlação moderada
- $r = 0.25 \rightarrow$  correlação fraca
- $r = 0 \rightarrow$  sem correlação alguma, nenhum ponto co-ocorre

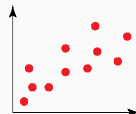
# Correlação



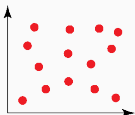
Perfect  
Positive  
Correlation



Strong  
Positive  
Correlation



Weak  
Positive  
Correlation



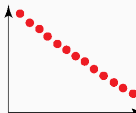
No  
Correlation



Weak  
Negative



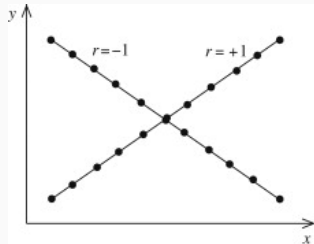
Strong  
Negative



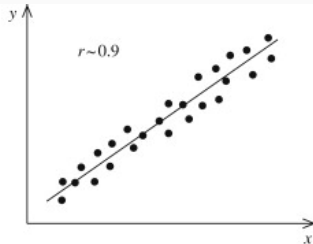
Perfect  
Negative



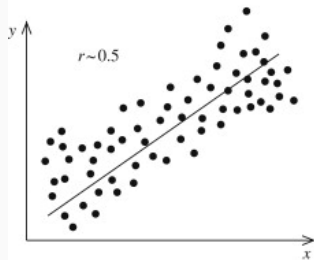
# Correlação



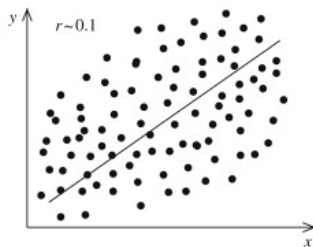
(a)



(b)

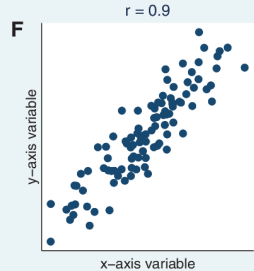
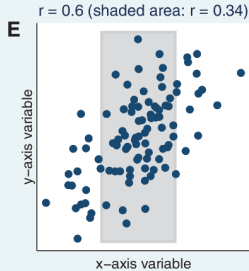
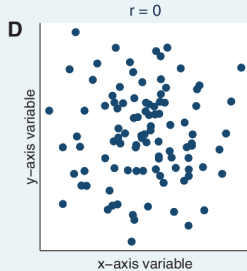
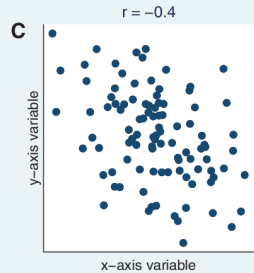
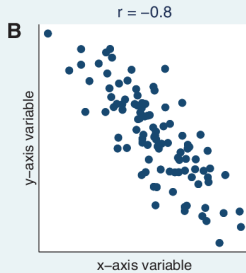
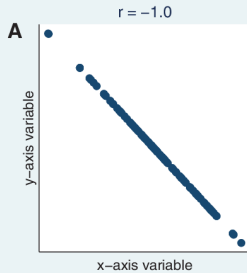


(c)



(d)

# Correlação



# Teste de correlação paramétrico e não paramétrico

O teste paramétrico de correlação (correlação de *Pearson*) pressupõe uma distribuição normal dos dados

```
1 shapiro.test(english$RTlexdec)
2 W = 0.98595, p-value < 2.2e-16

3 shapiro.test(english$Familiarity)
4 W = 0.98981, p-value < 2.2e-16
```

- Nossos dados não seguem uma distribuição normal
- teste não paramétrico (correlação de *Spearman*)

# Teste de correlação paramétrico e não paramétrico

```
1 cor.test(english$RTlexdec, english$Familiarity, method = "spearman")
2
3      Spearman's rank correlation rho
4
5 data:  english$RTlexdec and english$Familiarity
6 S = 2.2778e+10, p-value < 2.2e-16
7 alternative hypothesis: true rho is not equal to 0
8 sample estimates:
9      rho
10 -0.4338039
```

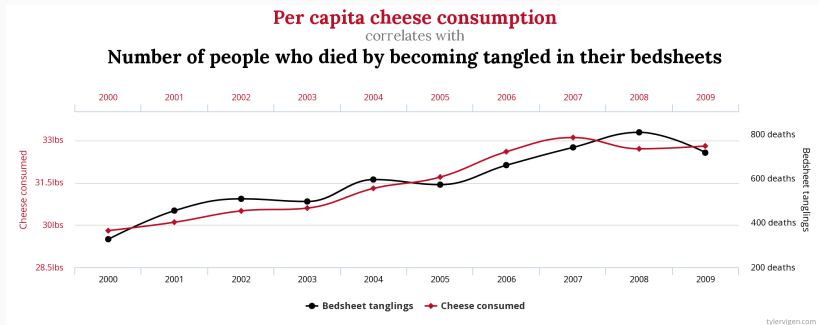
**Cuidado!**

---

# Cuidados com correlações

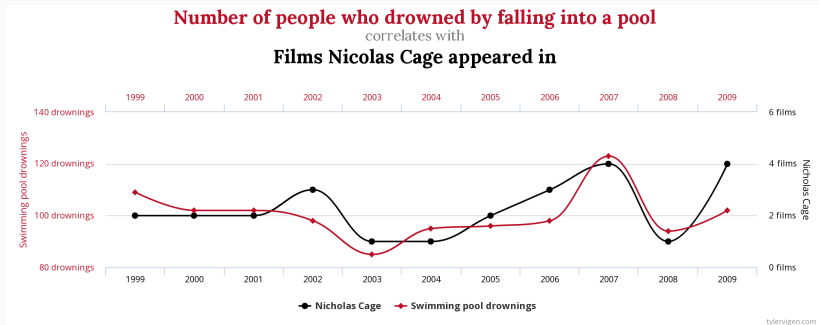
- Correlações por si só são pouco informativas
  - Apenas verificam se há uma relação entre as variáveis
- Queremos também prever valores não observados: modelos de regressão
- Co-ocorrência de duas variáveis pode não ser significativo
- Correlação não indica causação/motivação
- *Correlation does not mean causation*

# Cuidados com correlações



<http://www.tylervigen.com/spurious-correlations>

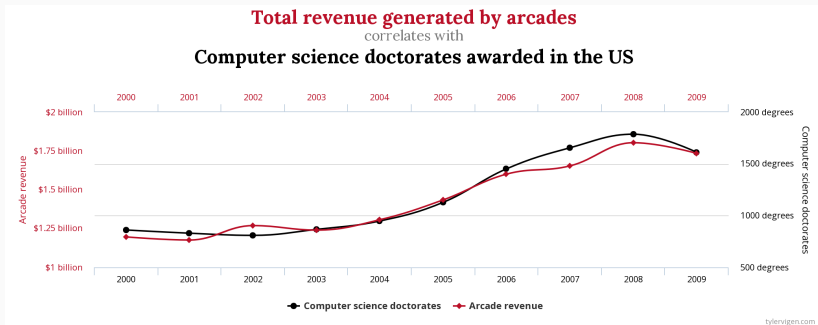
# Cuidados com correlações



<http://www.tylervigen.com/spurious-correlations>

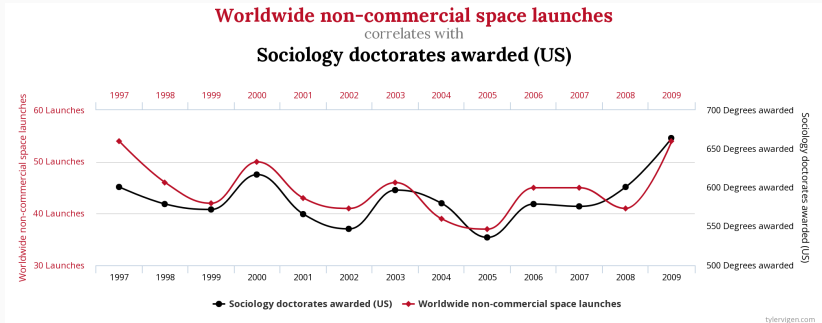


# Cuidados com correlações



<http://www.tylervigen.com/spurious-correlations>

# Cuidados com correlações



<http://www.tylervigen.com/spurious-correlations>

# Cuidados com correlações

Repita comigo:

- **correlação** não é sinônimo de **motivação**

One more time:

- **Correlation** *does not mean* **causation**

# Cuidados com correlações

- A e B podem co-ocorrer por diversos motivos:
  - Pura coincidência
  - A causa B
  - B causa A
  - C causa A e B

→ Ao observar uma correlação, como saber a causa da correlação?

Os exemplos anteriores são absurdos, mas nas pesquisas aparecem de maneira sutil

- Ex.: motivação – aprendizagem de L2
  - Há uma relação de causalidade?
  - Qual seria a direção da causalidade?
  - Há variáveis de confusão (*confounding variables*)?

**Perguntas?**