

Os métodos da psicolinguística: limites da inferência estatística e a questão da (ir)replicabilidade

Rafael Luis Beraldo

<rberaldo@unicamp.br> ou <rafael.beraldo@impatech.edu.br>

UFMG | LET406 – Linguística como ciência e aspectos biológicos da linguagem

2 de outubro de 2025

IEL/Unicamp e IMPA Tech



FALE
FACULDADE
DE LETRAS

U F M G

Sobre estes slides

- Estes slides foram apresentados em uma aula da disciplina **LET406 – Tópicos em Letras D: Linguística como ciência e aspectos biológicos da linguagem** ministrada no dia 2 de outubro de 2025, na Faculdade de Letras da UFMG.
- Os slides estão disponíveis sob a licença Creative Commons Atribuição 4.0.

Sobre mim

- Doutorado em Linguística pela Unicamp metodologia e replicação;
- Professor no IMPA Tech e pós-doc no IEL/Unicamp sob Lívia Oushiro;
- Atualmente, ministrando curso sobre irreplicabilidade e linguística:
<https://rberaldo.quarto.pub/replicabilidade-linguistica/>

A psicolinguística e seus métodos

Exemplo de Kantowitz et al. (2009)

- Tarefa: **defender uma mesa na biblioteca** usando apenas métodos não verbais.



Foto por Jason McDermott no Flickr

Exemplo de Kantowitz et al. (2009)

- Tarefa: **defender uma mesa na biblioteca** usando apenas métodos não verbais.
- “Método” (ou Variável Independente): **espalhar seus livros e pertences pela mesa** e aguardar até que alguém se sente.



Foto por Jason McDermott no Flickr

Exemplo de Kantowitz et al. (2009)

- Tarefa: **defender uma mesa na biblioteca** usando apenas métodos não verbais.
- “Método” (ou Variável Independente): **espalhar seus livros e pertences pela mesa** e aguardar até que alguém se sente.
- Medida (ou Variável Dependente): **tempo** que leva para alguém se sentar.



Foto por Jason McDermott no Flickr

Exemplo de Kantowitz et al. (2009)

- Tarefa: **defender uma mesa na biblioteca** usando apenas métodos não verbais.
- “Método” (ou Variável Independente): **espalhar seus livros e pertences pela mesa** e aguardar até que alguém se sente.
- Medida (ou Variável Dependente): **tempo** que leva para alguém se sentar.

Por que isso **não é** um experimento?



Foto por Jason McDermott no Flickr

Definição de experimento

A tarefa descrita anteriormente não é um experimento porque não satisfaz os requisitos básicos de um experimento:

An experiment occurs when the environment is systematically manipulated so that the causal effect of this manipulation on some behavior can be observed.

(Kantowitz et al., 2009, p. 52)

Definição de experimento

A tarefa descrita anteriormente não é um experimento porque não satisfaz os requisitos básicos de um experimento:

*An experiment occurs when the environment is **systematically manipulated** so that the causal effect of this manipulation on some behavior can be observed.*

(Kantowitz et al., 2009, p. 52)

Não há a manipulação sistemática do ambiente, operacionalizada como **pelo menos dois níveis de uma variável independente**.

Mais sobre experimentos

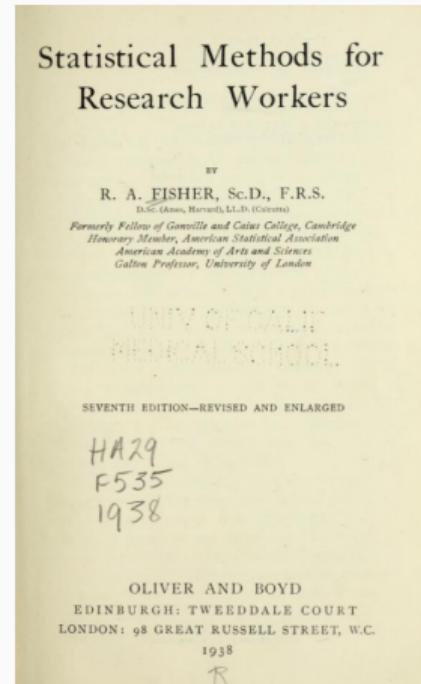
Sobre os experimentos, Kantowitz et al. (2009) ainda afirmam que eles:

- Permitem o **controle das variáveis**;
- Permitem o estabelecimento de **relações causais**;
- São **mais econômicos** que métodos observacionais;
- São conduzidos para comparar teorias (“*critical experiments*”, p. 54) ou gerar hipóteses (“*what-if experiments*”, p. 54).

Ronald Fisher e a estatística frequentista

Fisher é o autor de influentes obras de estatística, que impactaram as ciências experimentais:

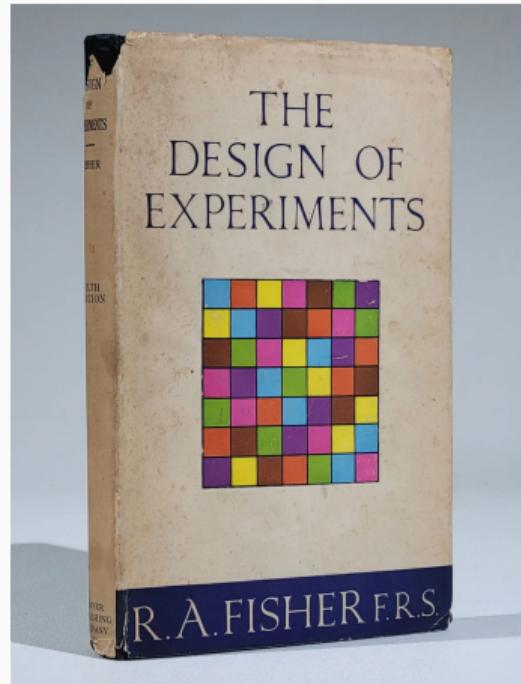
- *Statistical Methods for Research Workers* (1925)



Ronald Fisher e a estatística frequentista

Fisher é o autor de influentes obras de estatística, que impactaram as ciências experimentais:

- *Statistical Methods for Research Workers* (1925)
- *The Design of Experiments* (1935)



O tratamento formal da indução

Fisher (1955) defende que o pensamento indutivo pode atingir **o mesmo estatuto formal** que a dedução tem na lógica:

Logicians, in introducing the terms “inductive reasoning” and “inductive inference” evidently imply that they are speaking of processes of the mind falling to some extent outside those of which a full account can be given in terms of the traditional deductive reasoning of formal logic.

O tratamento formal da indução

Fisher (1955) defende que o pensamento indutivo pode atingir **o mesmo estatuto formal** que a dedução tem na lógica:

Logicians, in introducing the terms “inductive reasoning” and “inductive inference” evidently imply that they are speaking of processes of the mind falling to some extent outside those of which a full account can be given in terms of the traditional deductive reasoning of formal logic. Deductive reasoning in particular supplies no essentially new knowledge, but merely reveals or unfolds the implications of the axiomatic basis adopted. Ideally, perhaps, it should be carried out mechanically.

O tratamento formal da indução

Fisher (1955) defende que o pensamento indutivo pode atingir **o mesmo estatuto formal** que a dedução tem na lógica:

*Logicians, in introducing the terms “inductive reasoning” and “inductive inference” evidently imply that they are speaking of processes of the mind falling to some extent outside those of which a full account can be given in terms of the traditional deductive reasoning of formal logic. Deductive reasoning in particular supplies no essentially new knowledge, but merely reveals or unfolds the implications of the axiomatic basis adopted. Ideally, perhaps, it should be carried out mechanically. **It is the function of inductive reasoning to be used, in conjunction with observational data, to add new elements to our theoretical knowledge.***

O tratamento formal da indução

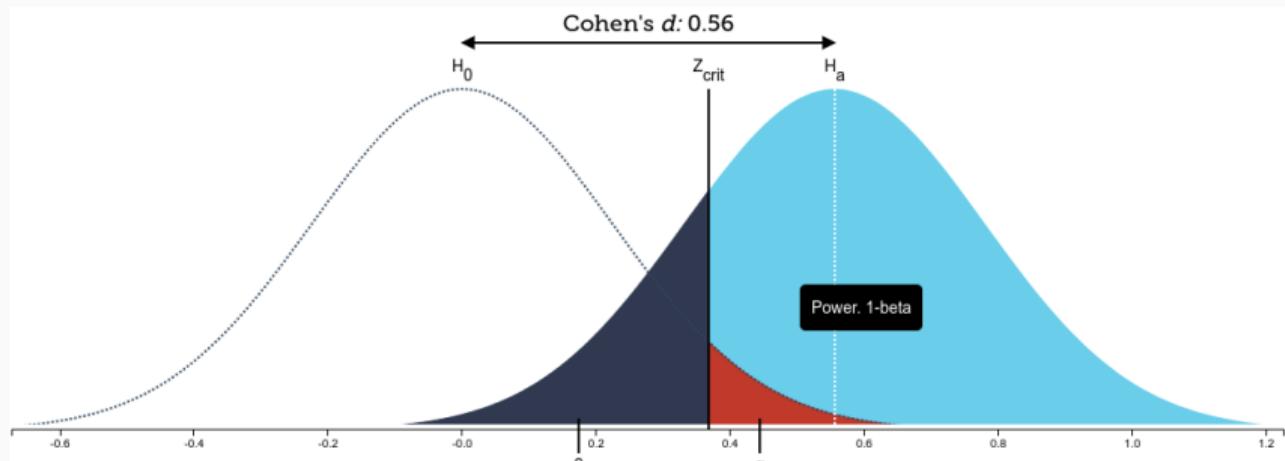
Fisher (1955) defende que o pensamento indutivo pode atingir **o mesmo estatuto formal** que a dedução tem na lógica:

Logicians, in introducing the terms “inductive reasoning” and “inductive inference” evidently imply that they are speaking of processes of the mind falling to some extent outside those of which a full account can be given in terms of the traditional deductive reasoning of formal logic. Deductive reasoning in particular supplies no essentially new knowledge, but merely reveals or unfolds the implications of the axiomatic basis adopted. Ideally, perhaps, it should be carried out mechanically. It is the function of inductive reasoning to be used, in conjunction with observational data, to add new elements to our theoretical knowledge. That such a process existed, and was possible to normal minds, has been understood for centuries; it is only with the recent development of statistical science that an analytic account can now be given, about as satisfying and complete, at least, as that given traditionally of the deductive processes.

(Fisher, 1955, p. 74, grifo nosso)

O funcionamento da estatística frequentista

O que faz a **estatística frequentista**? Ela quantifica a probabilidade de uma **distribuição observada** – os dados obtidos – **contra uma distribuição teórica**, que é dada.



Vejamos a demonstração em <https://rpsychologist.com/d3/nhst/>.

Os limites da inferência estatística frequentista

Críticas ao NHST (Null Hypothesis Statistical Testing)

As críticas ao NHST têm se acumulado desde os anos 1930; trago aqui apenas Cohen (1994):

*What's wrong with NHST? Well, among many other things, **it does not tell us what we want to know**, and we so much want to know what we want to know that, out of desperation, we nevertheless believe that it does! What we want to know is “Given these data, what is the probability that H_0 is true?” But as most of us know, what it tells us is*

Críticas ao NHST (Null Hypothesis Statistical Testing)

As críticas ao NHST têm se acumulado desde os anos 1930; trago aqui apenas Cohen (1994):

*What's wrong with NHST? Well, among many other things, **it does not tell us what we want to know**, and we so much want to know what we want to know that, out of desperation, we nevertheless believe that it does! What we want to know is “Given these data, what is the probability that H_0 is true?” But as most of us know, what it tells us is **“Given that H_0 is true, what is the probability of these (or more extreme) data?”***

(Cohen, 1994, p. 997)

O viés de publicação

Rosenthal (1979): o famoso “**problema do engavetamento**”; há um **viés de publicação** que favorece estudos “positivos”, introduzindo distorções e até mesmo falsos positivos na literatura.

Both behavioral researchers and statisticians have long suspected that the studies published in the behavioral sciences are a biased sample of the studies that are actually carried out (Bakan, 1967; McNemar, 1960; Smart, 1964; Sterling, 1959). The extreme view of this problem, the “file drawer problem,” is that the journals are filled with the 5% of the studies that show Type I errors, while the file drawers back at the lab are filled with the 95% of the studies that show nonsignificant (e.g., $p > .05$) results.



Por @mc_hankins

O viés de publicação

Rosenthal (1979): o famoso “**problema do engavetamento**”; há um **viés de publicação** que favorece estudos “positivos”, introduzindo distorções e até mesmo falsos positivos na literatura.

*Both behavioral researchers and statisticians have long suspected that the studies published in the behavioral sciences are a biased sample of the studies that are actually carried out (Bakan, 1967; McNemar, 1960; Smart, 1964; Sterling, 1959). The extreme view of this problem, the “**file drawer problem**,” is that the journals are filled with the 5% of the studies that show Type I errors, while the file drawers back at the lab are filled with the 95% of the studies that show nonsignificant (e.g., $p > .05$) results.*



Por @mc_hankins

O viés de publicação

Rosenthal (1979): o famoso “**problema do engavetamento**”; há um **viés de publicação** que favorece estudos “positivos”, introduzindo distorções e até mesmo falsos positivos na literatura.

*Both behavioral researchers and statisticians have long suspected that the studies published in the behavioral sciences are a biased sample of the studies that are actually carried out (Bakan, 1967; McNemar, 1960; Smart, 1964; Sterling, 1959). The extreme view of this problem, the “**file drawer problem**,” is that the journals are filled with the 5% of the studies that show Type I errors, while the file drawers back at the lab are filled with the 95% of the studies that show nonsignificant (e.g., $p > .05$) results.*



Por @mc_hankins

A maior parte da pesquisa publicada é falsa

Ioannidis (2005) publica o seu (agora) influente artigo *Why Most Published Research Findings Are False*.

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the

O argumento de Ioannidis (2005) ilustrado

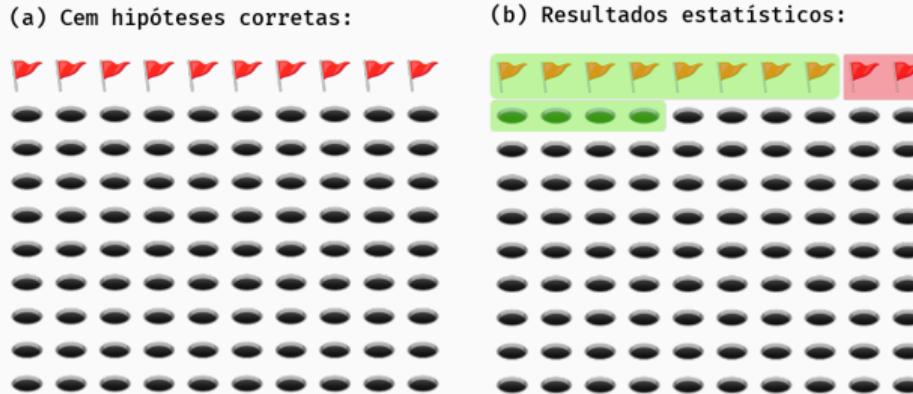


Ilustração baseada em interpretação de Pashler e Harris (2012). O modelo de Ioannidis (2005) adiciona outras considerações, tais como o baixo poder estatístico, os vieses dos pesquisadores e os múltiplos testes por laboratórios/grupos de pesquisa diferentes. (Fonte: Beraldo, 2024)

O argumento de Ioannidis (2005) ilustrado

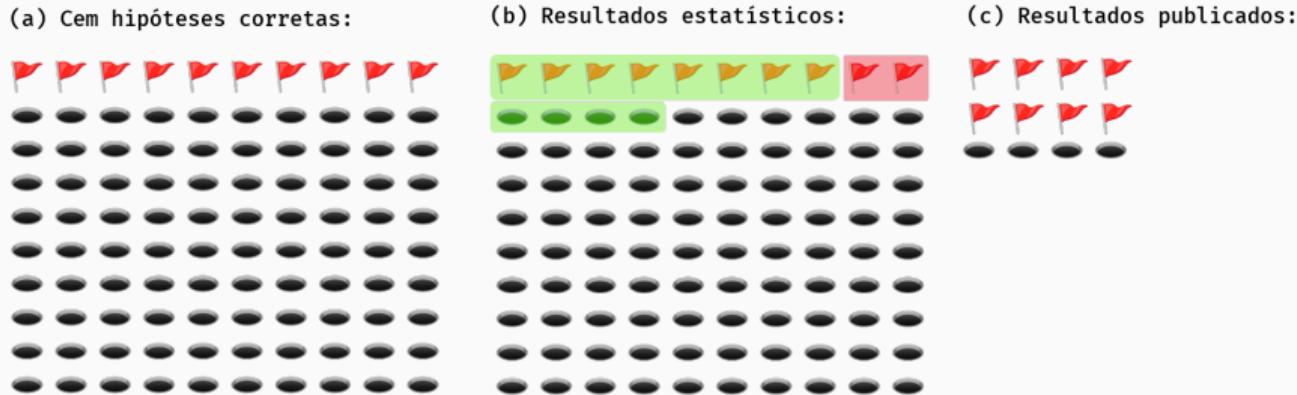


Ilustração baseada em interpretação de Pashler e Harris (2012). O modelo de Ioannidis (2005) adiciona outras considerações, tais como o baixo poder estatístico, os vieses dos pesquisadores e os múltiplos testes por laboratórios/grupos de pesquisa diferentes. (Fonte: Beraldo, 2024)

O conceito de replicação

O papel da “reprodução” de uma observação para Popper:

*We must clearly distinguish between **falsifiability and falsification**. [...]*

We say that a theory is falsified only if we have accepted basic statements which contradict it [...] This condition is necessary, but not sufficient; for we have seen that non-reproducible single occurrences are of no significance to science. Thus a few stray basic statements contradicting a theory will hardly induce us to reject it as falsified. We shall take it as falsified only if we discover a reproducible effect which refutes the theory.

(Popper, 2002 [1959], p. 66)

A “reprodução” e a falseabilidade

O papel da “reprodução” de uma observação para Popper:

*We must clearly distinguish between **falsifiability and falsification**. [...] We say that a theory is falsified only if we have accepted basic statements which contradict it [...] This condition is necessary, but not sufficient; for we have seen that **non-reproducible single occurrences are of no significance to science**. Thus a few stray basic statements contradicting a theory will hardly induce us to reject it as falsified. We shall take it as falsified only if we discover a reproducible effect which refutes the theory.*

(Popper, 2002 [1959], p. 66)

A “reprodução” e a falseabilidade

O papel da “reprodução” de uma observação para Popper:

*We must clearly distinguish between **falsifiability and falsification**. [...]*
*We say that a theory is falsified only if we have accepted basic statements which contradict it [...] This condition is necessary, but not sufficient; for we have seen that **non-reproducible single occurrences are of no significance to science**. Thus a few stray basic statements contradicting a theory will hardly induce us to reject it as falsified. **We shall take it as falsified only if we discover a reproducible effect which refutes the theory.***

(Popper, 2002 [1959], p. 66)

Uma definição de “replicação”

Nosek e Errington (2020) propõem que **a replicação seja fruto de um “acordo”**.

Uma definição de “replicação”

Nosek e Errington (2020) propõem que a replicação seja fruto de um “acordo”. Em outras palavras, **quais são as condições necessárias para que uma replicação seja informativa?**

Uma definição de “replicação”

Nosek e Errington (2020) propõem que a replicação seja fruto de um “acordo”. Em outras palavras, **quais são as condições necessárias para que uma replicação seja informativa?**

Definição de replicação Uma replicação é um estudo cujos resultados, quaisquer que sejam, seriam considerados como evidência para diagnosticar uma alegação de uma pesquisa anterior.

(Nosek e Errington, 2020, p. 2, tradução nossa)

Critérios para uma replicação

Nosek e Errington (2020) definem **dois critérios simétricos** para classificar replicações:

1. Resultados **consistentes** com uma afirmação prévia *aumentariam a confiança nessa afirmação*, e;

Critérios para uma replicação

Nosek e Errington (2020) definem **dois critérios simétricos** para classificar replicações:

1. Resultados **consistentes** com uma afirmação prévia *aumentariam* a confiança nessa afirmação, e;
2. Resultados **inconsistentes** com uma afirmação prévia *diminuiriam* a confiança nessa afirmação.

(Nosek e Errington, 2020, p. 2, tradução nossa)

O p do problema

Discussão sobre a **relevância da replicação na psicolinguística** em Vasishth e Nicenboim (2016):

This misconception is perhaps encouraged by the language we use to describe our success in getting a p-value; the effect is ‘reliable’ or ‘real’ or ‘generalizable’. The word ‘significant’ also contributes to giving the feeling that something of importance has been found. [...]

In fact, no matter how low our p-value, we will have a 0.05 probability of having mistakenly rejected the null when the null is in fact true. Thus, a p-value (regardless of whether it is low or on the border of 0.05) alone should not convince us that an effect is ‘real’ or ‘generalizable’; successful replications of the effect are much more convincing. We discuss this point further in the next section.

(Vasishth e Nicenboim, 2016, p. 355)

O p do problema

Discussão sobre a **relevância da replicação na psicolinguística** em Vasishth e Nicenboim (2016):

This misconception is perhaps encouraged by the language we use to describe our success in getting a p-value; the effect is ‘reliable’ or ‘real’ or ‘generalizable’. The word ‘significant’ also contributes to giving the feeling that something of importance has been found. [...]

In fact, no matter how low our p-value, we will have a 0.05 probability of having mistakenly rejected the null when the null is in fact true. Thus, a p-value (regardless of whether it is low or on the border of 0.05) alone should not convince us that an effect is ‘real’ or ‘generalizable’; successful replications of the effect are much more convincing. We discuss this point further in the next section.

(Vasishth e Nicenboim, 2016, p. 355)

O p do problema

Discussão sobre a **relevância da replicação na psicolinguística** em Vasishth e Nicenboim (2016):

This misconception is perhaps encouraged by the language we use to describe our success in getting a p-value; the effect is ‘reliable’ or ‘real’ or ‘generalizable’. The word ‘significant’ also contributes to giving the feeling that something of importance has been found. [...]

In fact, no matter how low our p-value, we will have a 0.05 probability of having mistakenly rejected the null when the null is in fact true. Thus, a p-value (regardless of whether it is low or on the border of 0.05) alone should not convince us that an effect is ‘real’ or ‘generalizable’; successful replications of the effect are much more convincing. We discuss this point further in the next section.

(Vasishth e Nicenboim, 2016, p. 355)

O *p* do problema

Discussão sobre a **relevância da replicação na psicolinguística** em Vasishth e Nicenboim (2016):

This misconception is perhaps encouraged by the language we use to describe our success in getting a p-value; the effect is ‘reliable’ or ‘real’ or ‘generalizable’. The word ‘significant’ also contributes to giving the feeling that something of importance has been found. [...]

In fact, no matter how low our p-value, we will have a 0.05 probability of having mistakenly rejected the null when the null is in fact true. Thus, a p-value (regardless of whether it is low or on the border of 0.05) alone should not convince us that an effect is ‘real’ or ‘generalizable’; successful replications of the effect are much more convincing. We discuss this point further in the next section.

(Vasishth e Nicenboim, 2016, p. 355)

Uma replicação (falha?) de psicolinguística

Verbos meteorológicos na gramática prescritiva

Os **verbos meteorológicos** do português brasileiro (PB) são impessoais e, portanto, **não devem apresentar concordância de número**:

- (1) *As cidades nevaram muito no inverno.

Uso no mundo real

No entanto, a concordância de número em verbos meteorológicos havia sido atestada anedoticamente, entre outros casos, **dentro de relativas cortadoras**.

- (2) não são todos os invernos q nevam nessas cidades (Costa, 2013, p. 184)

Uso no mundo real

No entanto, a concordância de número em verbos meteorológicos havia sido atestada anedoticamente, entre outros casos, **dentro de relativas cortadoras**.

- (2) não são todos os invernos q nevam nessas cidades (Costa, 2013, p. 184)

Esses casos seriam lapsos de linguagem, hipercorreções ou parte da gramática do PB?

Resultados anteriores de Costa (2011, 2013)

- Um estudo piloto de produção ($n = 27$, Costa, 2011) sugeriu que **a gramática do PB permite a concordância de número em verbos meteorológicos**; e ainda que as relativas padrão e as relativas cortadoras dentro de PPs desfavoreciam a concordância de número.

Resultados anteriores de Costa (2011, 2013)

- Um estudo piloto de produção ($n = 27$, Costa, 2011) sugeriu que a gramática do PB permite a concordância de número em verbos meteorológicos; e ainda que as relativas padrão e as relativas cortadoras dentro de PPs desfavoreciam a concordância de número.
- Um estudo de produção mais controlado ($n = 45$, Costa, 2013, Experimento 1) sugeriu que a **concordância verbal meteorológica é uma característica do PB**; entretanto, o efeito do PP não foi observado.

Experimento 2 de Costa (2013)

Um segundo experimento, dessa vez uma leitura automonitorada, foi planejado:

Condições Número do verbo (singular e plural) × tipo de antecedente (NP e PP)

Estímulos 16, quatro por condição

Amostra 32 participantes (estudantes da UFRJ)

Hipótese **Nenhuma diferença seria observada em nenhuma das condições**

Estímulos de Costa (2013)

Costa (2013) preparou 16 sentenças, quatro por condição:

Condição Estímulo

NP, SG Letícia viu **as serras** que **troveja** diariamente.

NP, PL Paulo visitou **as nações** que **nevam** excessivamente.

PP, SG Nilton ficou **nas planícies** que **venta** continuamente.

PP, PL André correu **nas praias** que **chovem** repetidamente.

A última posição (Posição 7) foi a posição crítica.

Estímulos de Costa (2013)

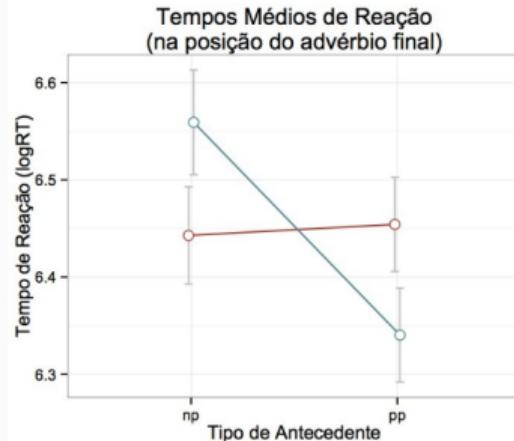
Costa (2013) preparou 16 sentenças, quatro por condição:

Condição	Estímulo
NP, SG	Letícia viu as serras que troveja diariamente .
NP, PL	Paulo visitou as nações que nevam excessivamente .
PP, SG	Nilton ficou nas planícies que venta continuamente .
PP, PL	André correu nas praias que chovem repetidamente .

A última posição (Posição 7) foi a **posição crítica**.

Resultados originais de Costa (2013)

Os resultados originais de Costa (2013, p. 69–70) foram:



Resultados de Costa (2013). As barras verticais indicam o erro padrão das médias. Azul é singular; vermelho é plural.

Table 1: Modelo de efeitos mistos ajustado à posição 7.

Efeitos fixos	Est. (logRT)	SE	valor t	valor p
Intercept	6.44279	0.08257	78.03	—
AntPP	0.01134	0.05651	0.20	0.8410
NumSG	0.11624	0.05651	2.06	0.0402
AntPP:NumSG	-0.23000	0.07992	-2.88	0.0042

Replicação 1: Características da amostra

Os métodos foram os mesmos para a Replicação 1, exceto pela **amostra expandida** e pela **coleta de dados remota**:

Total 281 participantes (estudantes da UFRN; **244 após exclusões**)

Exclusões por idioma 6 participantes (2,1%)

Atenção < 90% 15 participantes (5,3%)

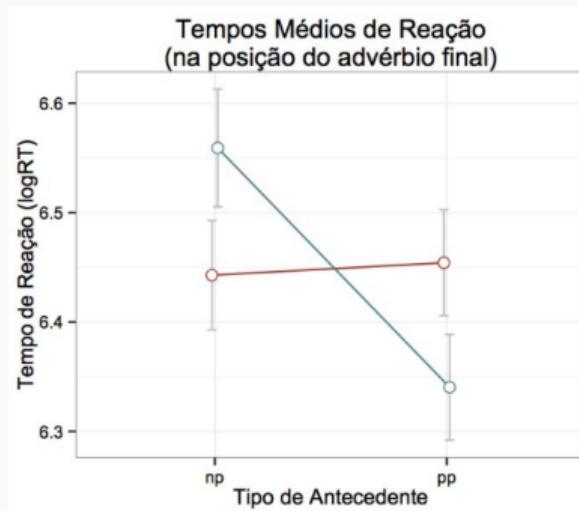
TR < 30 s 16 participantes (5,6%)

Total excluídos 37 (13,1%)

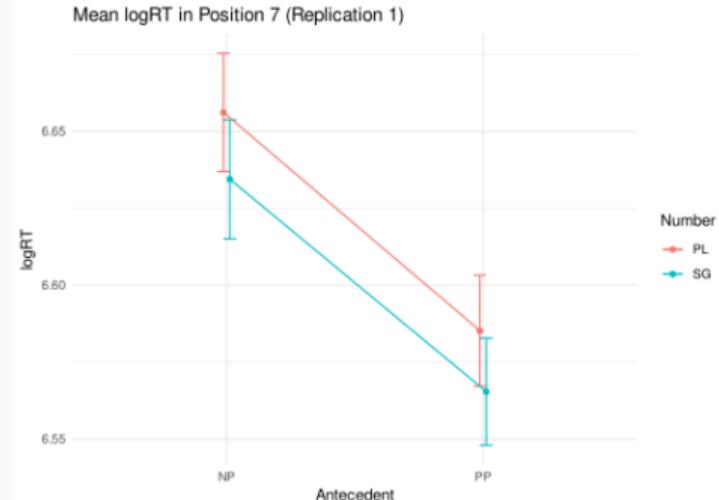
Filtragem de observações 100 ms < TR < 5000 ms (24% dos dados brutos)

A taxa de exclusão de participantes foi similar à de Sprouse (2011).

Replicação 1: Resultados



Resultados originais de Costa (2013, novamente). As barras verticais representam os erros padrão das médias. Azul é singular; vermelho é plural.



Resultados da Replicação 1. As barras verticais representam os erros padrão das médias.

Replicação 2: Modificações e características da amostra

Para controlar o efeito de integralização (Just e Carpenter, 1980), adicionarmos **duas posições de spill (P8 e P9) após a região crítica** na segunda replicação. Nossa amostra foi:

Total 261 participantes (estudantes da Unicamp; **228 após exclusões**)

Exclusões por idioma 19 participantes (7,2%)

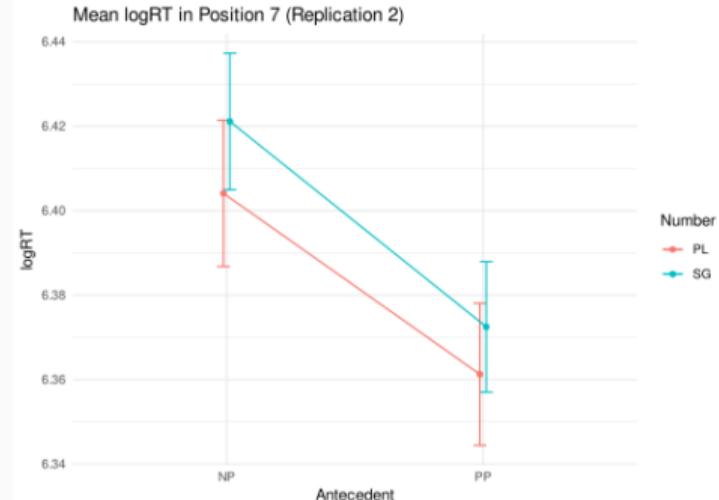
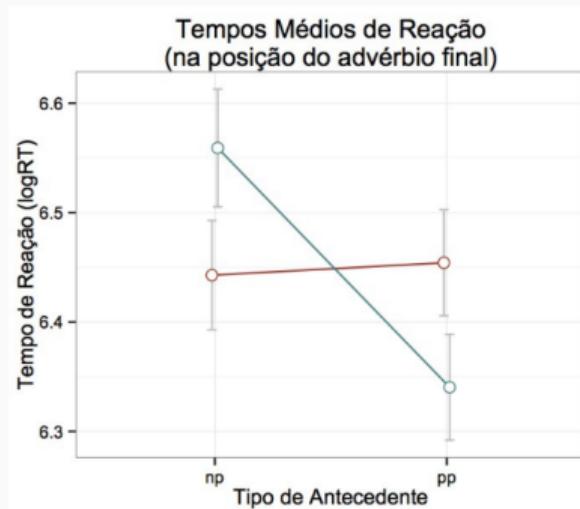
Atenção < 90% 6 participantes (2,4%)

TR < 30 s 8 participantes (3,0%)

Total excluídos 33 (12,6%)

Filtragem de observações 100 ms < TR < 5000 ms (17% dos dados brutos)

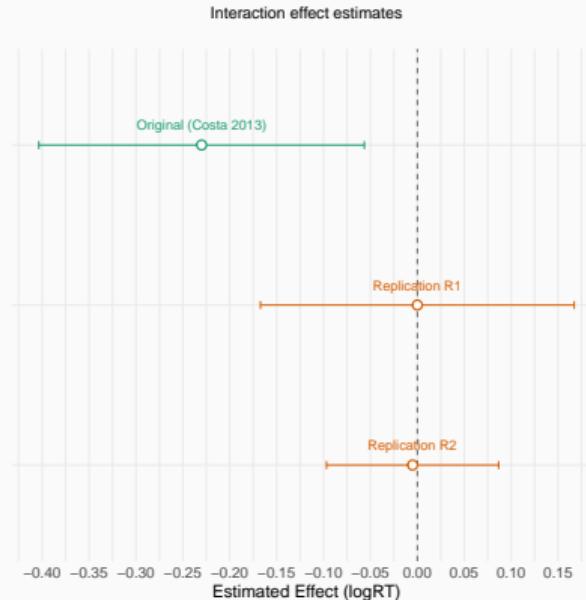
Replicação 2: Resultados



Resultados originais de Costa (2013, novamente). As barras verticais representam os erros padrão das médias. Azul é singular; vermelho é plural.

Resultados da Replicação 2. As barras verticais representam os erros padrão das médias.

Estimativa do tamanho do efeito



Estimativas do tamanho do efeito de interação NPSG × PP.SG em Costa (2013) e em nossas replicações. As barras horizontais indicam o IC de 95%. Figura adaptada de Costa (2022).

Interpretando nossos resultados

- Não conseguimos replicar os resultados do Experimento 2 de Costa (2013) – o que, de fato, está alinhado com sua hipótese original;
- Interpretamos isso como um caso de **regressão à média**: o efeito inicial desapareceu em ambas as replicações, R1 e R2.

- A psicolinguística é uma ciência, porque segue o método científico.
- Esse método – herdado em parte da psicologia –, entretanto, é fruto, disputas, problemas e inovações.
- Dentre outras causas, esses problemas derivam de equívocos no uso e interpretação das ferramentas da estatística, bem como a falta de replicações.
- Quando replicações são realizadas, há sintomas de falsos positivos, o que sugere a necessidade de inovações metodológicas (Sönning e Werner, 2021)

Conclusões da aula

- A psicolinguística é uma ciência, porque segue o método científico.
- Esse método – herdado em parte da psicologia –, entretanto, é fruto, **disputas, problemas e inovações**.
- Dentre outras causas, esses problemas derivam de equívocos no uso e interpretação das ferramentas da estatística, bem como a falta de replicações.
- Quando replicações são realizadas, há sintomas de falsos positivos, o que sugere a necessidade de inovações metodológicas (Sönning e Werner, 2021)

Conclusões da aula

- A psicolinguística é uma ciência, porque segue o método científico.
- Esse método – herdado em parte da psicologia –, entretanto, é fruto, disputas, problemas e inovações.
- Dentre outras causas, esses problemas derivam de **equívocos no uso e interpretação das ferramentas da estatística, bem como a falta de replicações.**
- Quando replicações são realizadas, há sintomas de falsos positivos, o que sugere a necessidade de inovações metodológicas (Sönning e Werner, 2021)

Conclusões da aula

- A psicolinguística é uma ciência, porque segue o método científico.
- Esse método – herdado em parte da psicologia –, entretanto, é fruto, disputas, problemas e inovações.
- Dentre outras causas, esses problemas derivam de equívocos no uso e interpretação das ferramentas da estatística, bem como a falta de replicações.
- Quando replicações são realizadas, **há sintomas de falsos positivos**, o que sugere a necessidade de inovações metodológicas (Sönning e Werner, 2021)

Referências bibliográficas

Referências i

-  BERALDO, Rafael Luis. *Irreplicabilidade na linguística experimental: causas, consequências e intervenções metodológicas*. Tese de doutorado—Campinas: Universidade Estadual de Campinas, 2024.
-  COHEN, Jacob. The earth is round ($p < .05$). *American Psychologist*, v. 50, n. 12, 1994.
-  COSTA, Igor de Oliveira. *Verbos meteorológicos no plural em orações relativas do português brasileiro: sintaxe e processamento*. Dissertação de mestrado—Rio de Janeiro: PUC-Rio, 2013.

Referências ii

-  COSTA, Igor de Oliveira. *Poder estatístico e tamanho da amostra nos estudos em psicolinguística experimental: uma abordagem compreensiva usando simulações*. Rio de Janeiro, 15 set. 2022.
-  COSTA, Igor de Oliveira; AUGUSTO, Marina R. A. Um caso de concordância com tópico: a expressão de plural em verbos meteorológicos no interior de orações relativas. In: *VI Jornada de Estudos da Linguagem. Linguagem: teoria, análise e aplicações*. Rio de Janeiro: Programa de Pós-graduação em Letras, 2011.
-  FISHER, Ronald. Statistical Methods and Scientific Induction. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, v. 17, n. 1, p. 69–78, jan. 1955.

Referências iii

-  FISHER, Ronald A. *The design of experiments*. Edinburgh: Oliver and Boyd, 1935.
-  IOANNIDIS, John P. A. Why most published research findings are false. *PLoS Medicine*, v. 2, n. 8, p. e124, 30 ago. 2005.
-  JUST, Marcel Adam; CARPENTER, Patricia A. A theory of reading: from eye fixations to comprehension. *Psychological Review*, v. 87, n. 4, p. 28, 1980.
-  KANTOWITZ, Barry H.; ROEDIGER III, Henry L.; ELMES, David G. *Experimental Psychology*. 9. ed. Belmont: Wadsworth, 2009.
-  NOSEK, Brian A.; ERRINGTON, Timothy M. What is replication? *PLOS Biology*, v. 18, n. 3, p. e3000691, 27 mar. 2020.

-  PASHLER, Harold; HARRIS, Christine R. Is the Replicability Crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, v. 7, n. 6, p. 531–536, 1 nov. 2012.
-  POPPER, Karl. *The logic of scientific discovery*. London: Routledge, 2002.
-  ROSENTHAL, Robert. The file drawer problem and tolerance for null results. *Psychological Bulletin*, v. 86, n. 3, p. 638–641, maio 1979.
-  SÖNNING, Lukas; WERNER, Valentin. The replication crisis, scientific revolutions, and linguistics. *Linguistics*, v. 59, n. 5, p. 1179–1206, 27 set. 2021.

-  VASISHTH, Shravan; NICENBOIM, Bruno. Statistical methods for linguistic research: foundational ideas – Part I. *Language and Linguistics Compass*, v. 10, n. 8, p. 349–369, ago. 2016.