

Riflessioni sull'analisi statistica di dati linguistici e psicolinguistici

Giorgio Arcara

Dipartimento di Neuroscienze, Università degli studi di Padova

La.br.i.s.
- research connection -

Le analisi statistiche

“There are three kinds of lies: lies, damned lies, and statistics.”

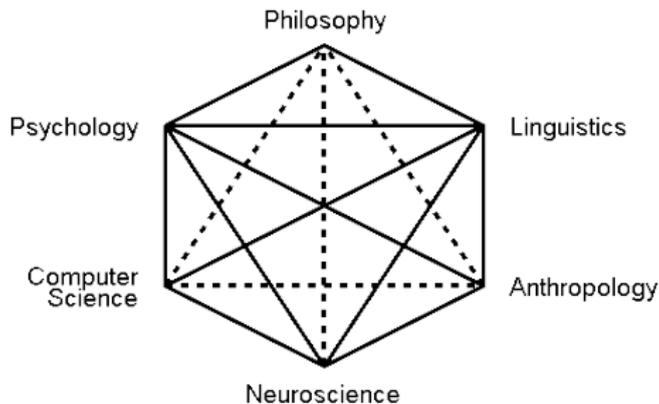
Benjamin Disraeli

“There are three kinds of lies: lies, damned lies, and statistics.”

Benjamin Disraeli

L'utilizzo delle analisi statistiche non implica buone interpretazioni dei risultati.

Il mio approccio alla statistica (e metodologia) è quello delle scienze cognitive.



- Perché sono importanti le analisi statistiche?
- Metodologia e Statistica.
- Alcune riflessioni sulle analisi statistiche in Linguistica e Psicolinguistica.

Perché sono importanti le analisi statistiche? (1)

Le scienze cognitive si basano molto sulle analisi statistiche, e si aspettano analisi statistiche dei dati (*Adger, 2015*)

Perché sono importanti le analisi statistiche? (2)

Perché sono importanti le analisi statistiche? (2)

Supponiamo ci sia una costruzione sintattica che è ritenuta accettabile dal 95% dei parlanti e inaccettabile dal 5% dei parlanti. Come dobbiamo considerare questo risultato?

Perché sono importanti le analisi statistiche? (2)

Supponiamo ci sia una costruzione sintattica che è ritenuta accettabile dal 95% dei parlanti e inaccettabile dal 5% dei parlanti. Come dobbiamo considerare questo risultato?

Le statistiche ci dicono quanto dobbiamo “credere” ai nostri dati (quanto sono generalizzabili). Da un punto di vista statistico, un risultato è interessante quando è **significativo**.

“To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of. ”

Ronald Fisher

“To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of. ”

Ronald Fisher

Statistica e Metodologia sono strettamente connesse.

Le statistiche utilizzabili dipendono dal metodo sperimentale e da come sono stati raccolti i dati.

Le statistiche utilizzabili dipendono dal metodo sperimentale e da come sono stati raccolti i dati.

Il metodo sperimentale e la raccolta dati dipendono dalle statistiche da utilizzare.

Nessuna statistica è in grado di risolvere certi problemi metodologici nella raccolta dati.

(ad esempio, numero insufficiente di stimoli, numero insufficiente di soggetti, mancanza di un gruppo di controllo, etc.)

Nessuna statistica è in grado di risolvere certi problemi metodologici nella raccolta dati.

(ad esempio, numero insufficiente di stimoli, numero insufficiente di soggetti, mancanza di un gruppo di controllo, etc.)



Come decidere la metodologia e le statistiche adatte per il proprio studio?

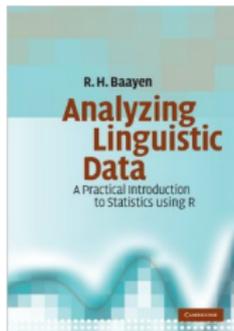
Come decidere la metodologia e le statistiche adatte per il proprio studio?

Principalmente da analisi della letteratura esistente.

Libri Utili su Statistica e Metodologia



McBurney & White (2008)
*Metodologia della ricerca in
Psicologia*. Il Mulino.



Baayen (2008) *Analyzing Linguistic
Data*. Cambridge University Press.
[download](#)



<http://cran.r-project.org/>

- gratuito e multiplatforma.
- ampliabile tramite “pacchetti”.
- costantemente aggiornato.
- grafici di alta qualità.
- è un linguaggio di programmazione.

- 1 Gli studi sul linguaggio sono quasi-esperimenti.
- 2 Correlazione e causalità.
- 3 La più comune fallacia negli studi sul linguaggio.
- 4 Numero di soggetti e di stimoli in un esperimento.
- 5 L'importanza della variabilità.
- 6 Analisi statistiche semplici e complesse.

Gli studi sul linguaggio sono quasi-esperimenti

Gli studi sul linguaggio sono quasi-esperimenti

Quasi tutti gli studi sul linguaggio sono **quasi-esperimenti**
(*McBurney & White, 2008*).

Quasi tutti gli studi sul linguaggio sono **quasi-esperimenti** (*McBurney & White, 2008*).

Non sempre possiamo avere un controllo sperimentale sulle proprietà delle parole o degli stimoli. Non possiamo decidere che una certa parola sia un verbo, o che sia breve, o frequente. Possiamo solo prenderne atto.

Gli studi sul linguaggio sono quasi-esperimenti

Quasi tutti gli studi sul linguaggio sono **quasi-esperimenti** (*McBurney & White, 2008*).

Non sempre possiamo avere un controllo sperimentale sulle proprietà delle parole o degli stimoli. Non possiamo decidere che una certa parola sia un verbo, o che sia breve, o frequente. Possiamo solo prenderne atto.

Se non possiamo *manipolare* le proprietà delle nostre unità statistiche siamo di fronte ad un **quasi-esperimento**.

Gli studi sul linguaggio sono quasi-esperimenti

Nei quasi-esperimenti c'è il costante rischio che i risultati siano riconducibili a variabili **confuse** con la nostra variabile di interesse.

Gli studi sul linguaggio sono quasi-esperimenti

Nei quasi-esperimenti c'è il costante rischio che i risultati siano riconducibili a variabili **confuse** con la nostra variabile di interesse.

La prima soluzione a questo problema è bilanciare per le principali proprietà che sappiamo essere rilevanti (es. lunghezza, frequenza).

Gli studi sul linguaggio sono quasi-esperimenti

Nei quasi-esperimenti c'è il costante rischio che i risultati siano riconducibili a variabili **confuse** con la nostra variabile di interesse.

La prima soluzione a questo problema è bilanciare per le principali proprietà che sappiamo essere rilevanti (es. lunghezza, frequenza).

Alcune analisi statistiche (es. ANCOVA, mixed-effect models) premettono di considerare molte variabili ed escludere l'influenza di variabili confuse.

Gli studi sul linguaggio sono quasi-esperimenti

Nei quasi-esperimenti non si può mai essere certi di avere escluso tutte le variabili confuse.

La ricerca scientifica spesso procede per esclusioni di spiegazioni alternative.

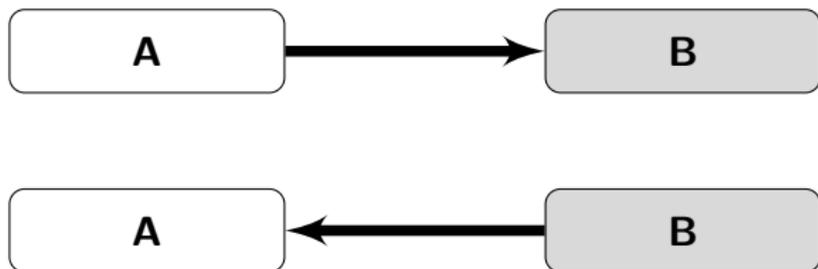
Gli studi sul linguaggio sono quasi-esperimenti

Nei quasi-esperimenti è molto rischioso effettuare inferenze di tipo causale.

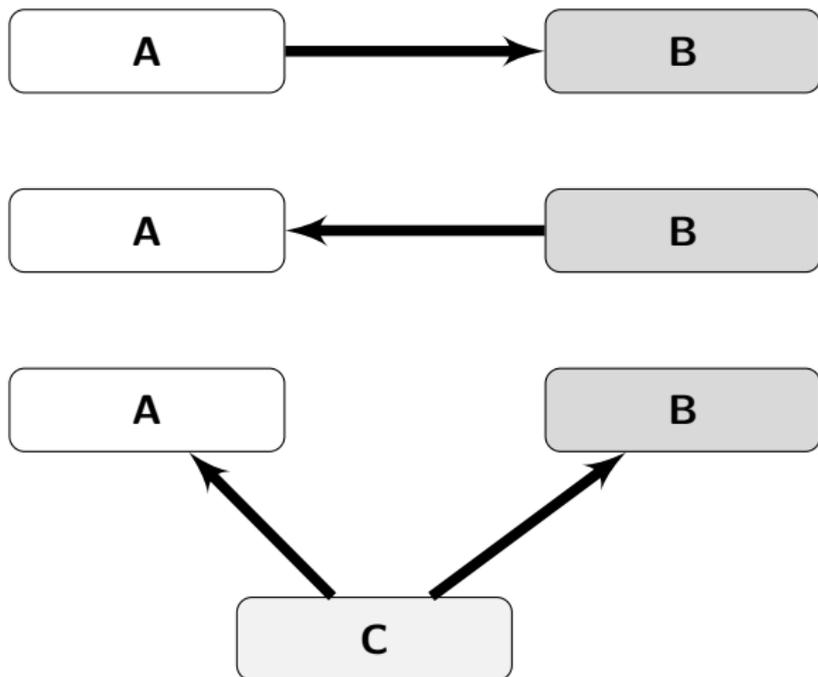
Correlazione e causalità



Correlazione e causalità



Correlazione e causalità



Le statistiche non possono essere usate quasi mai per inferire relazioni causali.

Le statistiche non possono essere usate quasi mai per inferire relazioni causali.

La causalità può essere desunta solo in esperimenti scientifici (e non in quasi-esperimenti), tramite la manipolazione delle variabili di interesse.

Un errore molto comune è pensare che confondere il nome delle statistiche con la natura del dato (soprattutto mi riferisco alla *correlazione*).

Un errore molto comune è pensare che confondere il nome delle statistiche con la natura del dato (soprattutto mi riferisco alla *correlazione*).

Il tipo di statistica utilizzata **non cambia la natura del dato** e il tipo di inferenze che possiamo fare sulla causalità.

Un errore molto comune è pensare che confondere il nome delle statistiche con la natura del dato (soprattutto mi riferisco alla *correlazione*).

Il tipo di statistica utilizzata **non cambia la natura del dato** e il tipo di inferenze che possiamo fare sulla causalità.

L'utilizzo di statistiche diverse dalla correlazione (ad esempio la *regressione lineare*) non modificano le nostre possibilità di fare inferenze su relazioni causali.

Esistono però delle analisi statistiche che **possono dare supporto** a inferenze causali.

- Modelli di equazioni strutturali (SEM).
- Granger's Causality models.
- Judea Pearl's models on causality.

Ma bisogna essere estremamente cauti con queste inferenze.

La più comune fallacia negli studi sul linguaggio

The language as-fixed-effect fallacy (*Clark, 1973*)

The language as-fixed-effect fallacy (*Clark, 1973*)

Gli item linguistici utilizzati in un esperimento sono quasi sempre una selezione tra tutti quelli possibili.

The language as-fixed-effect fallacy (*Clark, 1973*)

Gli item linguistici utilizzati in un esperimento sono quasi sempre una selezione tra tutti quelli possibili.

Questo aspetto deve essere preso in considerazione nelle analisi statistiche. Ad esempio utilizzando analisi by-subject e by-item o usando mixed-effects models (*Baayen, Davidson, & Bates, 2008*).

Numero di soggetti e di stimoli in un esperimento

Numero di soggetti e di stimoli in un esperimento

- dipende.

Numero di soggetti e di stimoli in un esperimento

- dipende.
- il maggior numero possibile.

Numero di soggetti e di stimoli in un esperimento

- dipende.
- il maggior numero possibile.
- controlla la letteratura.

Numero di soggetti e di stimoli in un esperimento

- dipende.
- il maggior numero possibile.
- controlla la letteratura.

Almeno 20-40 soggetti e almeno 30-50 stimoli per condizione sperimentale.

Ricordarsi che lo scopo è capire quanto il fenomeno è generalizzabile.

L'importanza della variabilità

Senza adeguata variabilità, molte analisi statistiche non possono essere utilizzate.

Senza adeguata variabilità, molte analisi statistiche non possono essere utilizzate.

Diversificare gli stimoli e le possibilità di risposta.

Analisi statistiche semplici e complesse

Tramite analisi semplici si può rispondere a domande semplici, tramite domande complesse si può rispondere a domande complesse.

Analisi statistiche semplici e complesse

Tramite analisi semplici si può rispondere a domande semplici, tramite domande complesse si può rispondere a domande complesse.

Cominciare con analisi semplici.
Approfondire con analisi complesse.

Tramite analisi semplici si può rispondere a domande semplici, tramite domande complesse si può rispondere a domande complesse.

Cominciare con analisi semplici.
Approfondire con analisi complesse.

Non semplificare il fenomeno per utilizzare le analisi statistiche che conosciamo.

Tramite analisi semplici si può rispondere a domande semplici, tramite domande complesse si può rispondere a domande complesse.

Cominciare con analisi semplici.
Approfondire con analisi complesse.

Non semplificare il fenomeno per utilizzare le analisi statistiche che conosciamo.

Conoscere diverse analisi statistiche permette di guardare lo stesso fenomeno da diverse angolazioni.

“The average human has one breast and one testicle. ”

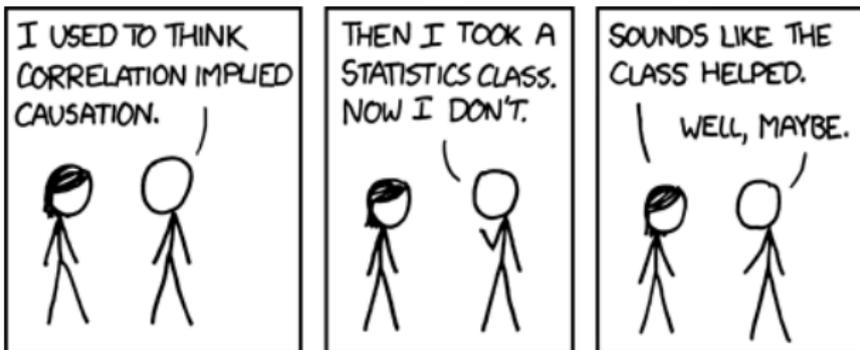
Des McHale

- Le analisi statistiche possono aiutarci nello studio scientifico di un fenomeno.

- Le analisi statistiche possono aiutarci nello studio scientifico di un fenomeno.
- Analisi statistica e metodologia sono strettamente intrecciate.

- Le analisi statistiche possono aiutarci nello studio scientifico di un fenomeno.
- Analisi statistica e metodologia sono strettamente intrecciate.
- Negli studi sul linguaggio esistono alcuni aspetti critici di tipo statistico/metodologico che è bene conoscere.

- Le analisi statistiche possono aiutarci nello studio scientifico di un fenomeno.
- Analisi statistica e metodologia sono strettamente intrecciate.
- Negli studi sul linguaggio esistono alcuni aspetti critici di tipo statistico/metodologico che è bene conoscere.
- Imparare a fare autonomamente analisi statistiche non è così complesso.



giorgio.arcara@gmail.com

Bigliografia

- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of verbal learning and verbal behavior*, 12(4), 335-359.
- Adger, D. (2015), Syntax. *WIREs Cognitive Science*, 6: 131147. doi: 10.1002/wcs.1332
- Baayen, R. H. (2007). *Analyzing Linguistic Data. A practical introduction to statistics*. Cambridge: Cambridge University Press.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory & Language*, 59, 390-412.

- McBurney, D. H., White, T. L., (2008) *Metodologia della ricerca in psicologia*. Il Mulino: Bologna.
- Judea Pearl's website:
http://bayes.cs.ucla.edu/jp_home.html
- <http://xkcd.com/>