Independence assumption

Grouping factors and generalizability

Daniela Palleschi
Humboldt-Universität zu Berlin

2024-01-12

Learning Objectives

Today we will learn about…

- the independence assumption
- types of non-independence in linguistic data
- the history of mixed models in linguistics

Resources

- this lecture covers
  - Sections 1 and 2 from Winter & Grice (2021)
  - Sections 14.1-14.3 from Ch. 14 (Mixed Models I) in Winter (2019)
  - Sections 8.1-8.2 from Ch. 8 (Mixed-effects models I: Linear Regression) in Sonderegger (2023)

Independence assumption

- we already learned about some model assumptions
  - assumption of *normality* of residuals
  - *homoscedasticity* (constant variance) of residuals
  - absence of *collinearity* of predictors
- there another, argulably more important assumption
  - assumption of *independence*

(Non-)Independence

- non-independence: any possible link or connection between groups of data points
  - e.g., two observations from the same participant will tend to be more similar than to completely independent observations
  - any case where you might expect some clustering of observations by some grouping factor
- the independence assumption assumes that our data points are *not* linked
  - i.e., the value of one observation is completely independent from another
- violations of this assumption have major implications for Type I (alpha) error
  - i.e., the chances of observing an effect where there is none (false positive)
- it also artificially inflates sample size, which affects statistical power

Repeated measures design

- the reason most (experimental) linguistic data is non-independent is the use of the repeated-measures design
  - collecting multiple data points from e.g., the same participant and for the same item
  - increases statistical power, needing fewer participants (more data points, lower variance due to control in variability between subjects)
  - saves resources (fewer subjects)

Other sources of non-independence

- non-independence is prevalent in other fields of linguistics, e.g.,
  - corpus studies: text, author, language, dialect, register
  - phonetic experiments: speaker, listener, exact repetitions
  - socio-phonetics: dialect/geographical proximity, register, speaker

Pseudoreplication

> Pseudoreplication refers to the treatment of dependent observations as independent data points, which causes an overabundance of erroneously significant results.

— Winter (2011), p. 2137

- analysing nonindependent data as if they were independent
- essentially, violating the independence assumption
  - very (*very*) common in older publications
- can also result in Type M (magnitude) and S (sign) error
- is one contributor (out of many) to the so-called replication crisis

Problem: Generalizability

- beyond spurious results, how researchers interpret the implications of their findings is problematic

> Unfortunately, outside of a few domains such as psycholinguistics, it remains rare to see psychologists model stimuli as random effects – despite the fact that most inferences researchers draw are clearly meant to generalize over populations of stimuli.

— (yarkoni_generalizability_2022?), p. 4

- if we don’t include grouping factors in our models, our findings are not generalisable beyond our sample
  - it could be that our findings are due to a few participants or experimental items who deviate from the rest
- we need to take this by-grouping factor variation into account, but how?

Solution 1: Averaging

- e.g., repeated measures ANOVA
  - seperate models for by-participant and by-item variance (with averaging) interpreted together
- PRO: takes both by-participant and -item variance into account
- CONs: not flexible or approrpriate for complex designs, and:
  - loses information regarding the variation across the grouped observations
  - lowers N
    - e.g., if we average over participants, we'd have 1 only data point per participant!
  - therefore loses statistical power (Type II error)
  - inflates Type I error (chance of a false positive)
- in sum: not optimal

Solution 2: Single observations

- run an experiment without repeated measures
  - but this lowers statistical power
  - and drastically reduces generalizability
- e.g., we could present 60 participants with a single item
  - or we could present 1 participant with 60 trials
  - but these findings also can't be generalised beyond that one item or one participant…
- in sum: not optimal

Solution 3: Linear mixed models

- best available solution: use repeated-measures design and mixed models
- a.k.a. mixed (effects) models/LM(E)Ms, multi-level models, hierarchical models
- “mixed” because they contain:
  - **fixed effects**: usually predictors; describe systematic variation in our data that we wish to explain
  - **random effects**: unsystematic variation that are due to random sampling
- random effects take dependence between observations into account
  - contain varying intercepts and slopes per level of a **grouping factor**
- fixed effects estimates are usually qualititatively unchanged
  - what is affected in the measures of *variance*

History of mixed-effects models

1973: Language-as-fixed-effect-fallacy

- none of these ideas are new to linguistics
- Clark (1973):
  - without including dependencies between repeated observations from the same **linguistic items** in our models, we cannot generalise our findings beyond our stimuli
  - our results are relevant only for the subset of the population from which we sampled

> The remedies for the language-as-fixed-effect fallacy are for the most part obvious. They include doing the right statistics, choosing the appropriate experimental design, and selecting a random or representative sample of language.

— Clark (1973), p. 347

ANOVAs: aggregation

- repeated measures ANOVAs were commonly used to take dependence between observations into account (and are still common in come fields today)
  - require aggregation (i.e., averaging) over items *or* subjects, not both simultaneously
  - drastically reduces our number of observations
  - loss of information in the variance of observed data
  - i.e., a loss of power (Type II error) and inflated Type I error (false positive)

2008: Baayen et al. (2008) and `lme4`

- enter mixed models with *crossed* random effects
- Journal of Memory and Language, Special Issue: Emerging Data Analysis
  - Baayen et al. (2008): introduction of `lme4` package for linear mixed models
  - Jaeger (2008): overview of generalised linear mixed models
- in addition, Baayen (2008) was published, a textbook for analysing linguistic data with R with an emphasis on LMMs with `lme4`

Learning Objectives 🏁

Today we learned about…

- the independence assumption ✅
- types of non-independence in linguistic data ✅
- the history of mixed models in linguistics ✅

Task

Discuss the following questions.

1. What is the independence assumption?
2. What happens when the independence assumption is violated?
3. What is the language-as-fixed-effect-fallacy?
4. What other sources of variance might be present in language research?
5. Why are repeated measures ANOVAs sub-optimal?

References

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

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. https://doi.org/10.1016/S0022-5371(73)80014-3

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Sonderegger, M. (2023). *Regression Modeling for Linguistic Data*.

Winter, B. (2011). *PSEUDOREPLICATION IN PHONETIC RESEARCH*.

Winter, B. (2019). Statistics for Linguists: An Introduction Using R. In *Statistics for Linguists: An Introduction Using R*. Routledge. https://doi.org/10.4324/9781315165547

Winter, B., & Grice, M. (2021). Independence and generalizability in linguistics. *Linguistics*, *59*(5), 1251–1277. https://doi.org/10.1515/ling-2019-0049