Understanding Mixed Models Through Simulations

A primer on power simulation

Julian Quandt

Wu Vienna University of Economics and Business

Slides and Code @ https://short.wu.ac.at/gz76

A few words before we start

- First, I know the slides contain a lot of text – this is mainly so you can look at them later and still understand them without me talking through them. You don't have to read everything as I am also going to explain it.

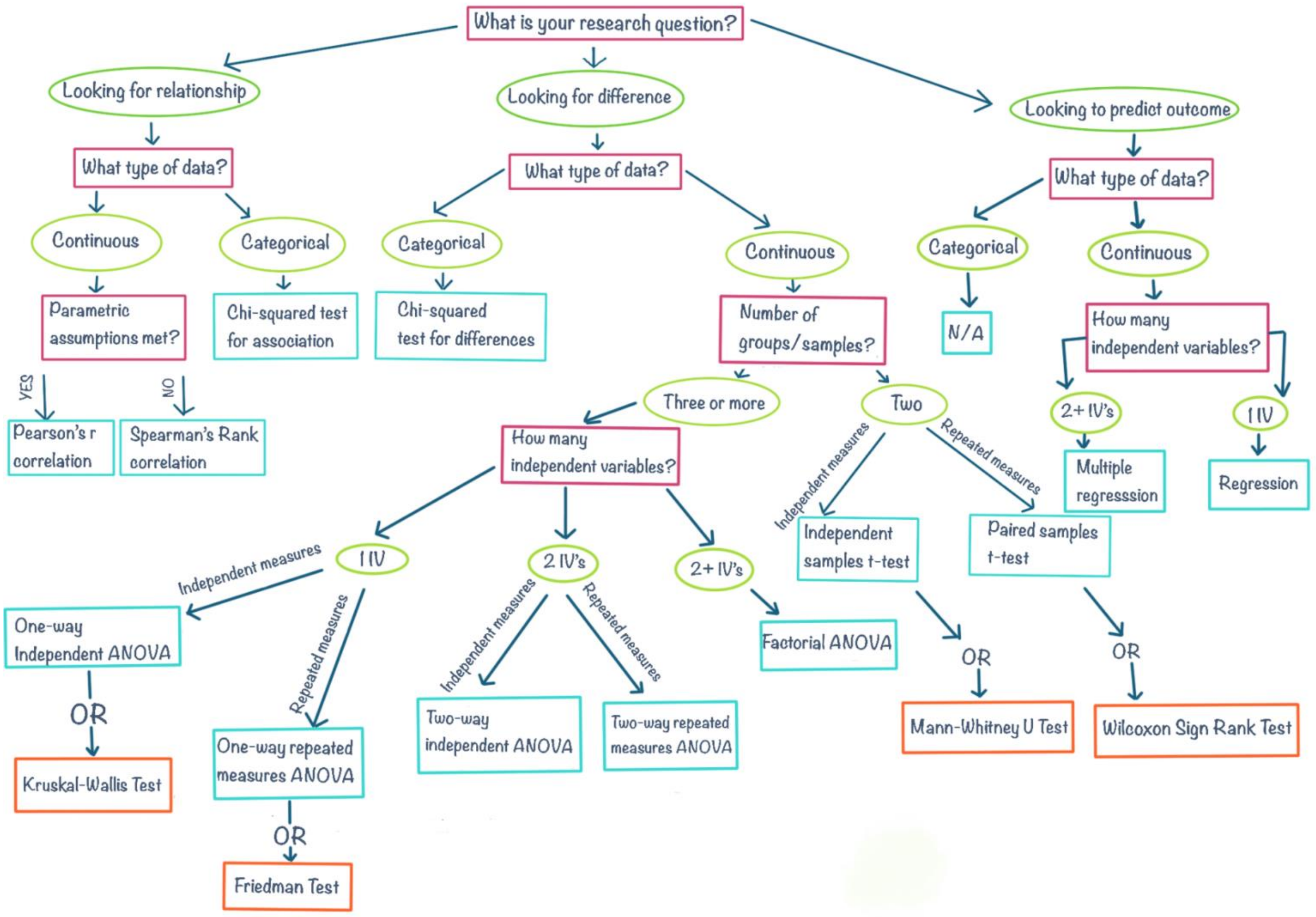
- Full explanation of power simulation for mixed models beyond the scope of this talk
  - I wrote a rather long 4-part tutorial on power simulation, that contains the examples from his talk + many other details, so if you enjoy today's talk, consider having a look
  - The focus of the tutorials is understanding simulation bottom-up, with using as little R packages as possible. IMO this gives a deeper understanding than using packages / functions on a high level of abstraction. That said, if your focus is on running power simulation with as little effort as possible, have a look at faux or simR instead - both are great and easier to use than simulating from scratch.

- Today's talk: How can simulations in R help us to understand mixed-effects models?
  - Why are simulations a useful tool
  - How to simulate mixed models from scratch
  - Understanding convergence warnings through simulations
  - If we have time: simulating random correlations, simulating binomial responses, residual variance decomposition, simulating level-2 nested data.

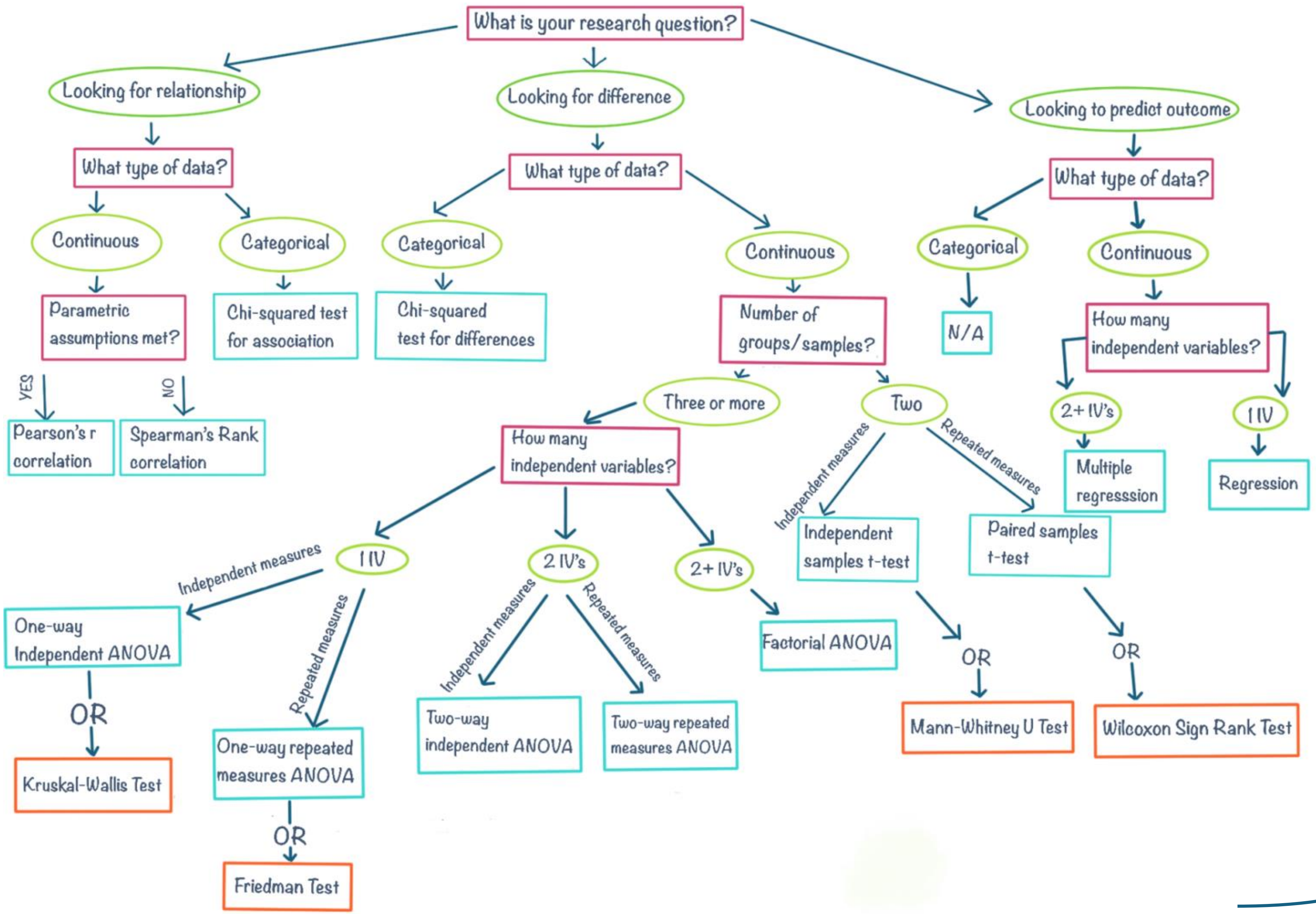
Part I

Prelude: Why simulations are useful

The garden of statistical analyses



Statistical analyses as linear models



$$Y = \beta_0 + \beta_1 * X + \in$$

[Check out this great overview] by Kristoffer Lindeløv

Two approaches to power analysis

- Imagine simple research question: ***How does people's opinion about environmental preservation change after seeing one of two video clips (positive or negative).***

→ Usually we would use a t-test or Ancova here.

- The alternative ($Y = \beta_0 + \beta_1 * X + \in$):
  - Y = **Dependent variable** (let's say change on a 7-point Likert Scale from before to after seeing the video clip)
    - → This is what we estimate
  - β_0 = **Intercept** (here the change from pre- to post-opinion for the negative clip group)
    - → let's assume the opinion of the people seeing a negative clip changes by 0.5 points (i.e. they see it as a little bit more important to preserve the environment after seeing the clip), so β_0 = 0.5
  - β_1 = **Condition Effect** (additional opinion change for the positive clip group)
    - → let's assume the positive clip works somewhat better, so people change their opinion around 0.2 points more (total of 0.7) than the people seeing the negative clip
  - X = **Condition** (0 if the person saw a negative clip, 1 if the person saw a positive clip)
  - $\in$ = **Expected noise** in this relationship (residual variance aka. how many unobserved + random factors influence this relationship, too)
    - → in this case, the model here is very simple and we do not measure any other factors that could help us to predict the change in opinions, so lets assume that the noise in the observed change is around 1 point).
  - → Y = 0.5+ 0.2*X+Norm(0,1)

Let’s try this in R

Summary of the simulation approach

- We can use the linear model to simulate data
- We come up with parameter values that we expect
- We make a design matrix, that contains all the information that we need to calculate the dependent variable.
- We add the dependent variable to the design matrix, and run the analysis on the simulated scores
- → If we repeat this like 1000 times and save the p-value for each simulation, we get the power of the test

Part II

Simulating Mixed-Effects Models

Extending the linear model with random effects

$$Y = \beta_0 + \beta_1 * X + \in$$

DV

Overall mean

Treatment-specific
adjustment
To overall mean

Unexplained
Variance /
Prediction Error

- In a mixed-model case, besides these *fixed effects* ($\beta_0$and β_1) we also assume that there will be *random* effects that are specific to participants (or stimuli in the experiment) that cause variation between participants.
- These factors are different from the unexplained variance $\in$, in that we know they are related to participants, but we do not know what exactly underlies them.

Extending the linear model with random effects

$$Y = \beta_0 + \color{red}{u_0} + \beta_1 * X + \in$$

DV

Overall mean

Treatment-specific adjustment To overall mean

Unexplained Variance / Prediction Error

Participant-specific adjustment to overall mean
Random Intercept

- In other words, the value of the DV for a person is not only a function of the overall mean and the treatment but also dependent on some *random* factors that we can attribute to specific participants.
- We do not know (or mostly also do not care) what causes these variations in participants but it is useful to model this by separating the variation in participants (the $\color{red}{u_0}$ term) from truly unknown variation in the data (the $\in$ term).

Extending the linear model with random effects

- Example: **Do people on average prefer rock or pop music?**
  - We let participants listen to multiple songs from Spotify's *best-of-rock* and *best-of-pop* playlists.
  - For each song, they will indicate on a scale from 0-100 how much they like the song (this is our DV, i.e Y in the model)
  - We will make a statistical model that predicts the liking Y for each song by means of
    - The overall liking of songs across both genres in the population (β_0)
    - The effect that the genre has on liking in the population (β_1)
    - For each participant, how much they *personally* like music compared to the population mean (u_0)
    - Variation caused by unknown factors (∈)

  - We make the following assumptions:
    - On average, people like music quite a bit, so we will set β_0 = 75
    - People will like rock music somewhat more than pop music, so β_1 = 5
    - Some people will like music more, others less, so u_0 = Norm(n_participants, 0, 7)
    - There will also be unknown factors influencing the likings, so ∈ = Norm(n_observations, 0, 10)

  - → We will simulate this just as we did before by using a linear model to populate a design matrix and then fit the model on the simulated data

Let’s try this in R

Interim Summary

- Mixed Models are extensions of the linear model
- By adding a random intercept term, we can tell the model that some variation is not truly random, but can be attributed to the specific differences between people
- This only works if we have multiple observations per person. Otherwise, the model will not be able to tell the specific variation between people (u_0) from variation that is to other random factors ($\in$).
- If you get a “singular fit” warning, a common reason for this is because the model does either not have enough observations to decompose variation terms or because the variation terms are close to the boundary (e.g. a random intercept close to 0).
- Random intercepts are useful, as they acknowledge that we do not test the entire population of participants but only a subsample. By estimating the variance in the population based on our sample of participants, we can make better predictions of what we would expect for people that we did not include in our sample.

Extending the linear model with random effects

- So far, we assume that no matter how much a person likes music in general, they will always like rock somewhat more than pop. However, it is likely that this differs among people.
  - There might be people who like music a lot in general (high u_0) but whether a specific song is pop or rock matters less to them.
  - There might also be people, for whom it matters more than 5 points, i.e. they also might like music on average but that is only because they LOVE rock songs, but strongly dislike pop songs.

Extending the linear model with random effects

$$Y = \beta_0 + {\color{red}u_0} + (\beta_1 + {\color{red}u_1}) * X + \in$$

Music ratings (0-100)

Population music liking

Population genre effect

Unexplained Variance / Prediction Error

Participant-specific adjustment to music liking
Random Intercept

Participant specific adjustment to genre effect (how much a specific participant prefers rock over pop music)
Random slope for participants over genre

Let’s try this in R

Interim Summary

- By adding random slopes to a model, we can account for our treatment / or experimental conditions, affecting different people differently.
- This only works if we have multiple observations *per participant per condition*. If this is the case, it is always advisable to at least try and add this random effect to the model, as otherwise we might increase the risk of false positive conclusions (Type I errors) while also potentially sacrificing statistical power of our test.
    - For more information see this paper and these standard operating procedures on fitting mixed models

Extending the linear model with random effects

- So far, we assume that each song is equally likely to be liked or disliked. However, it is more likely that:
  - Some songs will be universally liked more (the “hits” in each genre) while some songs will be liked less.
  - → we can add these assumptions as random effects to the model, too.

Extending the linear model with random effects

$$Y = \beta_0 + u_0 + w_0 + (\beta_1 + u_1) * X + \in$$

Music ratings (0-100)

Population music liking

Population genre effect

Unexplained Variance / Prediction Error

Participant-specific adjustment to music liking
Random Intercept for participants

Participant specific adjustment to genre effect (how much a specific participant prefers rock over pop music)
Random slope for participants over genre

Song-specific adjustment to music liking
Random Intercept for songs

Let’s try this in R

Interim Summary

- Random effects cannot only be estimated across people, but also stimuli.
- This can be an important addition to a model, as we usually do not use the entire set of possible stimuli in the model. Here, for instance, we do not have people listen to all possible rock and pop songs, so it is useful to have an estimate for how much songs differ in terms of how much they are liked. This helps us make better predictions for songs that we did not include in the sample.

Extending the linear model with random effects

- What about a random slope over songs?
  - Adding a random slope for genre over songs does not make sense here: It would mean that we would expect a certain song to differ in its propensity to be liked as a pop song compared to being liked as a rock song. Hence, this does not make sense as a song is either rock or pop but cannot "switch" genre from rock to pop or vice versa.

- → Hence, given the current variables in our model, the model is a *maximal random effects structure* model, meaning that it contains all possible random effects that we can put in there.
  - Well there is a caveat: Random effects terms can also be correlated:
    - For instance, has a high overall liking of music *across both genres* ($\beta_0 + u_0$ close to 100), this would indicate that they cannot dislike either music genre much or else they would not be able to have an average liking of close to 100 *across* genres. If they liked one music type (e.g. pop) with only 60 points, to have an average liking of e.g. 95, they would have to like rock music at 130, because (60+130)/2 = 95. However, 130 is not a possible value on our scale and therefore this can't be the case.
    - As a consequence, any person with high average liking across genres, is unlikely to have large values of u_1 . In other terms, across people the values of u_0 and u_1 will be negatively correlated.

- → I assume we are probably (close to) out of time at this point, if we still have time, we could discuss one of four things:
  - Finish the maximal model for real by adding random effects correlations (👍)
  - The elephant in the room: how to come up with parameter values? (❤️)
  - Simulating clustered data, such as students nested in classes (🎉)
  - Simulating different response data such as binomial [0,1] responses (😲)

Conclusion

- Simulating data based on the linear model gives great control over all aspects that we suspect to be present in our data
- It can help with getting a very precise idea of what experimental data will look like and can help us to identify blind-spots in our thinking and test out what might happen if certain things turn out different than expected, e.g. what happens to our estimates if all participants behave the same; u_0 ~ 0, or what happens if they all behave very differently u_0 ~ large. We can test how these affect the precision of our estimates in a certain sample size, by running the simulation multiple times and observing the differences in our estimates and p-values.
- Simulations help us think in terms of the *data-generating process*, i.e. what mechanisms in the experiment will cause which numbers to vary – in my opinion this is easier to do than thinking in terms of standardized effect sizes such as Cohen’s d, as these are usually unintuitive to express.
- We can use our domain knowledge and values from previous similar studies to inform our choices
- This simulation approach works for all linear models (gaussian, binomial etc.) and can in principle be extended to non-linear models as well (non-linear SEM / cognitive models such as DDM).