**Exercise: Constructing a multiverse**

Stefan Vermeent

Stanford March 2024

For this assignment, choose one of your own empirical research projects. This could be a current project (ideally one for which you have either a preregistration or a Methods section written down), but it could also be a finished project. The goal is to uncover the multiverse of arbitrary decisions. We will also identify non-arbitrary decisions better kept out of the multiverse.

1. If the project consists of multiple studies, focus on one. List as many decisions necessary to prepare the raw data for the main analysis (e.g., data exclusions, variable computation, variables to include, model specifications).
   1. List the type of decision (see Box 1).
   2. Determine whether each decision can be considered **arbitrary,** **non-arbitrary**, or whether it is unclear if a decision is arbitrary or not.

For example:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Decision** | **Type** | **N-A, A, or ?** |
| 1 | Remove outliers at > 3.2 SD | Changing the underlying data | A |

**Box 1. Different types of decisions**We can broadly classify analytic decisions into three types.   
  
First, decisions may **change the underlying data**. For example, we may include/exclude outliers, or include/exclude participants who missed a certain number of attention checks. For these decisions, the question of whether they are arbitrary or non-arbitrary depends on the extend to which we expect them to influence the quality of our data. For example, are the outlier values plausible in the target population, or are they likely to indicate invalid responses?  
  
Second, decisions may **change variables** included in the model. For example, the study might include two or more questionnaires that aim to measure the same construct. For these decisions, it is important to consider if our theory can in any way distinguish between the measures. Do they aim to tap the exact same construct, or do they possible tap different dimensions of the same construct? Can we reasonably assume that the reliabilities of the measures are comparable? If the measures do not tap the same construct, or if their reliability differs, the effect sizes in the multiverse are not comparable.  
  
Third, decisions may **change the causal model**. For example, we may compare effect sizes in models including or excluding a particular covariate. It is generally discouraged to add such decisions to the multiverse (Del Giudice et al., 2021; Simonsohn et al., 2018; Rohrer, 2021). As noted by Simonsohn et al (2018), analyses “with and without a certain set of covariates are not different answers to the same question, they are different answers to different questions.” (p. 10).  
  
Relevant reading:  
Del Giudice, M., & Gangestad, S. W. (2021). A Traveler’s Guide to the Multiverse: Promises, Pitfalls, and a

Framework for the Evaluation of Analytic Decisions. *Advances in Methods and Practices in*

*Psychological Science*. <https://doi.org/10.1177/2515245920954925>

Rohrer, J. (2021). Mülltiverse analysis. Retrieved from

<https://www.the100.ci/2021/03/07/mulltiverse-analysis/>

Simonsohn, U., Simmons, J.P., & Nelson, L.D. (2020). Specification curve analysis. *Nature Human*

*Behaviour, 4*(11), 1208-1214. <https://doi.org/10.1038/s41562-020-0912-z>

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Decision** | **Type** | **N-A, A, or ?** |
| 1 |  |  |  |
| 2 |  |  |  |
| 3 |  |  |  |
| 4 |  |  |  |
| 5 |  |  |  |
| 6 |  |  |  |
| 7 |  |  |  |
| 8 |  |  |  |
| 9 |  |  |  |
| 10 |  |  |  |
| 11 |  |  |  |
| 12 |  |  |  |
| 13 |  |  |  |
| 14 |  |  |  |
| 15 |  |  |  |

1. For each **arbitrary decision** and decision you are unsure about, listed above, specify all alternative decisions that may be considered equally justifiable.

For example:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Decision** | **Alternatives** | **#** |
| 1 | Exclude outliers at > 3.2 SD | Exclude at > 3.2 SD  Exclude at > 2.5 SD  Include all outliers | 3 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Decision** | **Alternatives** | **#** |
| 1 |  |  |  |
| 2 |  |  |  |
| 3 |  |  |  |
| 4 |  |  |  |
| 5 |  |  |  |
| 6 |  |  |  |
| 7 |  |  |  |
| 8 |  |  |  |
| 9 |  |  |  |
| 10 |  |  |  |
| 11 |  |  |  |
| 12 |  |  |  |

1. For each **non-arbitrary decision** listed above, specify the type(s) of nonequivalence that is likely to result from adding alternatives to the model following the terms used by Del Giudice et al. (2021; see box 2 below).

For example:

|  |  |  |
| --- | --- | --- |
|  | **Decision** | **Type(s) of nonequivalence** |
| 1 | Exclude outliers at > 3.2 SD |  |

**Box 2. Different types of outcome nonequivalence**Not all decisions are created equal. In many cases, one decision is preferable over all alternatives. Adding such non-arbitrary decisions to the multiverse can have a big impact on our inferences. Even when adding a single non-arbitrary decision with two alternatives, half the multiverse contains the unjustified alternative. This way, realistic effect sizes can quickly become overshadowed by invalid, biased effect sizes. Del Giudice et al (2021) distinguish between three types of nonequivalence that could result from adding non-arbitrary decisions to the multiverse:  
  
**1. Measurement nonequivalence.** One decision provides an objectively better measure of the construct. One example is including a composite measure as well as its indicators to the multiverse. As Del Giudice et al (2021) note, composite measures are usually more valid and reliable than their individual components (assuming that each component is itself a valid measure of the construct).  
  
**2. Effect nonequivalence.** A key assumption of multiverse analysis is that the interpretation of the effect is the same across the entire multiverse. There are different ways in which different decisions can lead to non-comparable effects. For example, adding two-and three-way interaction specifications leads to effects that can not be directly compared. In addition—as mentioned in box 1—using different sets of covariates is likely to change the causal logic of the model, making the effects incomparable. For example, if the association tested with the model is confounded by variable X, then leaving variable X out of the model yields a more biased estimate than when including it as a covariate.  
  
**3. Power/precision nonequivalence.** Some decisions might require excluding more participants than anticipated. In such cases, some data sets in the multiverse might be underpowered. Note that this can happen for decisions that are otherwise arbitrary.

Relevant reading

Del Giudice, M., & Gangestad, S. W. (2021). A Traveler’s Guide to the Multiverse: Promises, Pitfalls, and a Framework for the Evaluation of Analytic Decisions. *Advances in Methods and Practices in Psychological Science, 4*(1), 1-15. <https://doi.org/10.1177/2515245920954925>

|  |  |  |
| --- | --- | --- |
|  | **Decision** | **Type(s) of nonequivalence** |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |
| 6 |  |  |
| 7 |  |  |
| 8 |  |  |
| 9 |  |  |
| 10 |  |  |
| 11 |  |  |
| 12 |  |  |

1. What is the size of the multiverse that you constructed under (2)? In other words, How many combinations of arbitrary decisions are possible?

……………………………………………………………

1. Can you think of tests using the outcomes of the multiverse analysis to assess the impact of specific decisions in terms of reliability or validity?   
     
     
   ………………………………………………………………………………………………………………………………  
     
   ………………………………………………………………………………………………………………………………  
     
   ………………………………………………………………………………………………………………………………  
     
   ………………………………………………………………………………………………………………………………  
     
   ………………………………………………………………………………………………………………………………  
     
   ………………………………………………………………………………………………………………………………