Table X: Characteristics of simulated datasets.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Scenario A** | **Scenario B** | **Scenario C** |
| Study Design | Within-subject  Treatment study | Between-subject  Observational Design | Single Subject AB Experimental Design |
| Outcome Type | Continuous | Ordinal | Binary |
| Sample Sizes | n = 20; n = 60 | n = 74; n = 148 | n = 1 |
| Between-subject variability | SD = 0.25; SD = 0.75 | SD = 0.25; SD = 0.75 | SD = 0.25; SD = 0.75 |
| Within-subject Variabilitiy | SD = 0.25; SD = 0.75 | Not applicable | SD = 0.25; SD = 0.75 |
| Hierarchical Data Structure | Three trials clustered by participant and time (pre/post) | Four trials clustered by participant | 30 treated and 30 untreated items clustered by session |
| Data Reduction Approach for Non-Hierarchical Dataset | Average of three trials for a participant at each time point | Maximum (worst) score across four trials for each participant | Proportion of correct responses for each condition (treated/untreated) and session |

*Caption*: Note that between-subject variability indicates item-level variability for scenario C, whereas scenarios A and B indicate variability between participants.

- Define scope; change tone

- Future directions: one should systematically examine how different designs/data types/sample sizes affect the accuracy/validity/utility of synthetic data

- Goal is to illustrate, but not vet

Paper 1

* Explore its utility in different studies
* Future directions: need to understand impact of hierarchical (e.g. cog study)
* See tutorial to get started
* Potential tool to solve a field-wide problem
* Specific utility: Original model CI overlap with synthetic CI; and direction of p-value (sig or not sig)
* General utility: Visualize the raw data (compare), then pMSE
* Hierarchical (cognition) data
  + Subject, group, sex, condition
  + AT will try out syn.cart(), then maybe send original author our attempt
* Add single subject design (Will Hula study with Bayesian model – and this is hierarchical)
* Don’t need to synthesize ‘generated’ variables (e.g., BMI); need to know nuances of your data

Paper 2

* X