**Ideas for assignments (or indeed Masters thesis topics)**

* Generic things from other sources
  + More extensive simulation of something from class
  + Replicate (part of) a published simulation you have read
* Impact of violating other assumptions
  + t-test:
    - Continuous data instead of normal (which is ordered and bounded: the {faux} package has conversion fucntions to generate such data)
    - Non-normal data (some other distribution than normal or skew-normal eg to mimic reaction times, or bounded data)
    - Violation of random sampling (e.g., preselection), which can be related to the concept of regression to the mean.
* Simulate impact of other statisical choices
  + Illustrate how effect coding vs other forms of coding change regression beta estimates (but not marginal means)
* Demonstrate the difficulty with estimating and interpreting interaction effects (e.g., several of Julia Rohrer’s published articles)
* Show some common rule of thumb or practice and when they are good/bad
  + E.g., some reserachers argue we should choose which covariates to include in a regression based on what is signficiant in a bivariate regression’s p value or effect size. This gets into causal modelling and might be more complex.
* Simulate compound effect of conditional rules between tests (if assumptions test failed, run X, otherwise Y)
  + False positive rate of a given common analytic workflow for a given task or specific type of analytic flexibility
    - E.g., like the workflow for analyzing data from the Implicit Relational Assessment Procedure, which uses a 4X2 RM-ANOVA plus follow up tests, with flexbility in covariates and preprocessing.
* *p*-hacking
  + Replicate (some of) Stefan and Schönbrodt’s “a compendium of p-hacking strategies”
* Extend that compendium to data tampering and fraud
  + Condition switching
  + Participant duplication
  + Alternation of data (eg how few cells do you have to offset or by how much to get significance)
* Measurement hacking / schmeasurement
  + The impact of violating the assumptions of Cronbach’s alpha on alpha estimates (spoiler: Cronbach’s alpha doesn’t measure what you think it does)
  + Replicate Kopalle and Lehman (1996)
  + Estimate the impact of dropping items from a multi-item scale based on “cronbach’s alpha if item removed” on the in sample cronbach’s alpha.
  + Multiverse of EFA (probably too complex)
* Replicate “Why most of psychology is statistically unfalsifiable”
* Simulate the influence of rounding on the estimation of effect sizes for meta-analysis. Ie articles only report the M and SD to two decimal places usually: how accurate are the Cohen’s d values calcualted from these rounded estimates compared to the real in-sample Cohen’s d?
* Siulation of selection effects
  + Matthew effect
  + Regression to the mean
  + Trade-offs between efficacy and attrition / Efficacy paradox / Average Treatement Effect vs Average Treatment on the Treated effect
    - <https://bsky.app/profile/quentinandre.bsky.social/post/3lk4nqotlss2k>
* Scientific ponzi scheme probabalistic extension
* “Everything is a linear model” <https://lindeloev.github.io/tests-as-linear/> – demonstrate the equality of one or two different statitical methods vs. using a regression
* Robustness ML vs OLS regression models to data missingness
* Power analysis for first digit or last digit analysis
  + Geyer & Williamson 2004 Detecting Fraud in Data Sets Using Benford's Law, Durtschi,
  + Hillison, & Pacini (2004) The Effective Use of Benford’s Law to Assist in Detecting Fraud in Accounting Data
  + Nigrini (1996, 2012)
  + Fewster (2009)
  + Leemis, Schmeiser, & Evans (2000)
  + R packages benford.analysis or BenfordTests
  + Extend simulations to last digit analysis, comparisons of methods
* Other ideas you come up with - but run them by me for feasibility.