# Draft Proposal

Huifeng Wu

August 3, 2020

## Objectives

Over the summer, I am taking a large number of courses from the university. However, I have a sense of inconsistency about my productivity when dealing with coursework, and I feel left behind in comparison with my cohorts. For example, it takes me 3 or 4 hours to finish an assignment, whereas I could a spend a whole day on another one that is similar and I am very upset about that.

Therefore, I start to think about what might affect my productivity. Based on my experience, I generally feel distracted by people surrounding me, not drinking coffee and spending too long on my phone. In order to find out how to better manage study in general, I plan out a factorial experiment to investigate the effects of multi-person environment, drinking coffee, phone usage on productivity.

## Methods

For experiment, I decide to have 3 factors at two levels each, so there will be $2^3$=8 combinations to test in total. Considering potential variance at each measurement, I will replicate the proposed experiment twice.

The response variable productivity is measured using a score out of 10, where a bigger number denotes more satisfaction from an observation, Accordingly, the aforementioned factors will be defined as follow:

1. Another person proceeding his/her own activity in the same room (Yes/No)
2. Drinking coffee (Yes/No)
3. Putting cell phone away (Yes/No)

In particular, I plan on inviting 4 friends of mine to participate. The 4 participants will be randomly assigned to two combinations, under which they are asked to study for 3 hours for each. After that, they will give a productivity score. To better facilitate the participants' need, the 2 combinations will be tested on separate days. Then, we repeat the procedure on another two days and finish up the experiment with 16 measurements eventually. Please note that the other person in the first factor will proceed normal activities of daily living (ADLs), which may include but not limited to sleeping, eating, reading etc. The person will stay quiet as much as possible over the course.

To reduce bias from having the same person to test same conditions repeatedly, I will randomize our participants of 4 to ensure the reliability of further casual inference, so they may be assigned to any condition. As people's work may differ by content and difficulty, the productivity score is given on a scale of how much time the participant focus on his/her work, rather than the amount of work completed. Furthermore, it is believed that the productivity score is given individually, and that does not vary with others'.

I also assume the response variable productivity is identically independent. Each participant is assumed to have similar working habits excerpt for the control, and their inherent attributes do not impact the outcome significantly. Such differences are simply negligible here.

## Statistical Analysis Plan

At first, we should observe some plots of productivity with respect to each factor. This will briefly illustrate their relationships, even though the factors are merely two levels. In addition, it is essential to have summary statistics, such as mean, standard deviation, maximum and minimum for the outcome variable. As for data visualization, we have box plots featuring each combination and a scatter plot summarizing all data points.

We can fit a linear model with the collected data. Predictors can be the factors alone and their interactions, and we can study the main effects and interaction effects with cube and interaction plots. The estimates should be consistent between the fitted linear model and the plots. Furthermore, the model that fits the data better may be selected by using AIC/BIC backward elimination. The mean square error term and R-squared values of the final model can be presented.

With the proposed replicated runs, we can estimate the error variance of each measurement and the effect estimates. Since we can obtain the estimates of their effects from R, their 95% confidence intervals can be inferred as well. We will interpret the results in terms of statistical significance as well as the size of their effects, and see if that matches the prior hypothesis.