

# FIN 373 Homework 4

due: 9/21/21

Instructions: Please submit solutions on canvas. Only a knitted pdf of an Rmarkdown file will be accepted.

**Problem 1:** To understand why the poor are constrained in their ability to save for investments in preventative health products, two researchers designed a field experiment in rural Kenya in which they randomly varied access to four innovative saving technologies. By observing the impact of these various technologies on asset accumulation, and by examining which types of people benefit most from them, the researchers were able to identify the key barriers to saving.<sup>1</sup>

They worked with 113 ROSCAs (Rotating Savings and Credit Associations). A ROSCA is a group of individuals who come together and make regular cyclical contributions to a fund (called the “pot”), which is then given as a lump sum to one member in each cycle. In their experiment, Dupas and Robinson randomly assigned 113 ROSCAs to one of the five study arms. In this exercise, we will focus on three study arms (one control and two treatment arms). The data file, `rosca.csv` is extracted from their original data, excluding individuals who have received multiple treatments for the sake of simplicity.

Individuals in all study arms were encouraged to save for health and were asked to set a health goal for themselves at the beginning of the study. In the first treatment group (*Safe Box*), respondents were given a box locked with a padlock, and the key to the padlock was provided to the participants. They were asked to record what health product they were saving for and its cost. This treatment is designed to estimate the effect of having a safe and designated storage technology for preventative health savings.

In the second treatment group (*Locked Box*), respondents were given a locked box, but not the key to the padlock. The respondents were instructed to call the program officer once they had reached their saving goal, and the program officer would then meet the participant and open the *Locked Box* at the shop where the product is purchased. Compared to the safe box, the locked box offered stronger commitment through earmarking (the money saved could only be used for the prespecified purpose).

Participants are interviewed again 6 months and 12 months later. In this exercise, our outcome of interest is the amount (in Kenyan shilling) spent on preventative health products after 12 months.

Descriptions of the relevant variables in the data file `rosca.csv` are:

| <i>Variable</i>             | <i>Description</i>  |
|-----------------------------|---|
| <code>bg_female</code>      | 1 if female, and 0 otherwise                                    |
| <code>bg_married</code>     | 1 if married, and 0 otherwise                                   |
| <code>bg_b1_age</code>      | age at baseline   |
| <code>encouragement</code>  | 1 if encouragement only (control group), and 0 otherwise        |
| <code>safe_box</code>       | 1 if safe box treatment, and 0 otherwise                        |
| <code>locked_box</code>     | 1 if lock box treatment, and 0 otherwise                        |
| <code>fol2_amtinvest</code> | Amount invested in health products                              |
| <code>has_followup2</code>  | 1 if appears in 2nd followup (after 12 months), and 0 otherwise |

---

<sup>1</sup>This exercise is based on: Dupas, Pascaline and Jonathan Robinson. 2013. Why Don't the Poor Save More? Evidence from Health Savings Experiments. *American Economic Review*, Vol. 103, No. 4, pp. 1138-1171.

- Load the data set as a `data.frame` and create a single factor variable `treatment` that takes the value `control` if receiving only encouragement, `safebox` if receiving a safe box, and `lockbox` if receiving a locked box. How many individuals are in the control group? How many individuals are in each of the treatment arms?
- Subset the data so that it contains only participants who were interviewed in 12 months during the second followup. We will use this subset for the subsequent analyses. How many participants are left in each group of this subset? Does the drop-out rate differ across the treatment conditions? What does this result suggest about the validity of this study?
- Does receiving a safe box increase the amount invested in health products? We focus on the outcome measured 12 months from baseline during the second follow-up. Compare the mean of amount (in Kenyan shilling) invested in health products `fol2_amtinvest` between each of the treatment arms and the control group. Briefly interpret the result.
- Examine the balance of pre-treatment variables, gender (`bg_female`), age (`bg_b1_age`) and marital status (`bg_married`). Are participants in the two treatment groups different from those in the control group? What does the result of this analysis suggest in terms of the internal validity of the finding presented in the previous question?
- Does receiving a safe box or a locked box have different effects on the investment of *married* versus *unmarried women*? Compare the mean investment in health products among married women across three groups. Then compare the mean investment in health products among unmarried women across three groups. Briefly interpret the result. How does this analysis address the internal validity issue discussed in Question 4?

**Problem 2:** The data `predimed.csv` is from a randomized control trial of diet variation among participants at risk for a cardiovascular event. The type of diet was randomly assigned, and a key outcome variable is `whether or not the person had a cardiac event`, denoted `event` in the data. There are two variants of the Mediterranean diet, one supplemented with nuts and the other supplemented with extra virgin olive oil (VOO).<sup>2</sup> We will use this data to estimate a variety of probabilities and assess the effectiveness of a Mediterranean diet.

- Estimate the following probabilities:  $P(\text{event})$ ,  $P(\text{any MedDiet})$ ,  $P(\text{event}, \text{any MedDiet})$ , and  $P(\text{event} \mid \text{any MedDiet})$ .
- Estimate  $P(\text{event} \mid \text{Control})$ . Using this result and the answer from the previous question, assess whether the Mediterranean diet has an effect on the chance of a cardiac event?
- What additional information not given would be useful to characterize uncertainty in the effect estimate above. (*Hint*: Remember that this is a randomized control trial). With that information, describe the steps for characterizing this uncertainty.
- What are the effects on cardiac event likelihood of the Mediterranean diet on **Female** and **Male** subpopulations relative to the control diet?

---

<sup>2</sup>The original study is [Primary Prevention of Cardiovascular Disease with a Mediterranean Diet](#). *The New England Journal of Medicine*, 2013, 368:1279-1290.