# HW2: Health Savings Experiments

06 September 2023

To understand why those who are low income are constrained in their ability to save for investments in preventative health, 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 key barriers to saving.

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, the researchers 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 specified purpose selected by the participant) but less control of funds by the participant.

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

Descriptions of the relevant variables in the data file `rosca2.csv` are:

| Name | Description |
|---|---|
| `bg_female` | `1` if female, and `0` otherwise. This is a pre-treatment variable. |
| `bg_married` | `1` if married, and `0` otherwise. This is a pre-treatment variable. |
| `bg_b1_age` | Age at baseline. This is a pre-treatment variable. |
| `encouragement` | `1` if participant received encouragement only (control group), and `0` otherwise |
| `safe_box` | `1` if participant received safe box treatment, and `0` otherwise |
| `locked_box` | `1` if participant received lock box treatment, and `0` otherwise |
| `fol2_amtinvest` | Amount invested in health products at time of the second follow up |
| `has_followup2` | `1` if participant appears in second followup (after 12 months), and `0` otherwise |

## Question 0

Run the following the code chunk below to load the data set and create a new variable `treatment` that takes the value `control` if receiving only encouragement, `safebox` if receiving a safe box, and `lockbox` if receiving a locked box. We then designate that R should treat this new variable as a factor variable.

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(readr)
rosca <- read_csv("data2/rosca.csv")
```

```
## New names:
## Rows: 423 Columns: 9
## -- Column specification
## -------------------------------------------------------- Delimiter: "," dbl
## (9): ...1, bg_female, bg_married, bg_b1_age, encouragement, safe_box, lo...
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```r
rosca <- rosca %>%
  mutate(treatment = case_when(
    encouragement == 1 ~ 'control',
    safe_box == 1 ~ 'safebox',
    locked_box == 1 ~ 'lockbox'
  ))

rosca$treatment <- as.factor(rosca$treatment)
```

## Question 1 (10 points)

a) What does a row in the data set correspond to? How many individuals are in the control group? How many individuals are in each of the treatment arms? Use a table to show the counts. What proportion of the study participants are in each treatment arm and the control group?

b) For the remainder of this question, consider just the Locked Box treatment relative to the Encouragement Only alternative: What is the specific causal question the researchers aimed to answer? What are the potential outcomes? What is the hypothesized treatment effect?

c) For the Locked Box intervention group, what is their average missing counterfactual? How do your recommend estimating it?

## Answer 1 Code

```
nrow(rosca)
```

```
## [1] 423
```

```
ncol(rosca)
```

```
## [1] 10
```

```
summary(rosca)
```

```
##       ...1         bg_female        bg_married       bg_b1_age
##  Min.   :  1.0   Min.   :0.0000   Min.   :0.0000   Min.   :17.00
##  1st Qu.:106.5   1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:30.00
##  Median :212.0   Median :1.0000   Median :1.0000   Median :38.00
##  Mean   :212.0   Mean   :0.7447   Mean   :0.7541   Mean   :39.61
##  3rd Qu.:317.5   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:48.00
##  Max.   :423.0   Max.   :1.0000   Max.   :1.0000   Max.   :88.00
##
##  encouragement       safe_box        locked_box    fol2_amtinvest
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :   0.0
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:  40.0
##  Median :0.0000   Median :0.0000   Median :0.000   Median : 100.0
##  Mean   :0.2624   Mean   :0.2766   Mean   :0.461   Mean   : 322.2
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.: 490.0
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   :5700.0
##                                                    NA's   :30
##  has_followup2      treatment
##  Min.   :0.0000   control:111
##  1st Qu.:1.0000   lockbox:195
##  Median :1.0000   safebox:117
##  Mean   :0.9291
##  3rd Qu.:1.0000
##  Max.   :1.0000
##
```

```r
table(rosca$treatment)
```

```
##
## control lockbox safebox
##     111     195     117
```

```r
proportions(table(rosca$treatment))
```

```
##
##   control   lockbox   safebox
## 0.2624113 0.4609929 0.2765957
```

**Answer 1 Text**

**a. Each row in the data set corresponds to the profile with information regarding age, marital status,type of treatment(control,safebox and lockbox), amount invested etc of each individual that took part in the rosca study.**

**The control group has : 111, lockbox: 195 , safebox: 117.**

**The control: 26.24 % , lockbox : 46.09%, safebox : 27.65%.**

**treatment : lockbox and usual : encouragement**

**SCQ :what is the impact of technology like the Lockbox treatment relative to the usual(encouragement) on asset accumulation and saving capacity for low income individuals in rural Kenya ?**

**potential outcome #1 : What would the outcome on the saving capacity at the end of 12 months be if y/n experiences the lockbox treatment?**

**2. what would the outcome on the saving capacity at the end of 12 months be if y/n experiences the Usual treatment(encouragement)?**

**3. The hypothesised treatment effect would be to compare if the outcome on the amount of money saved with the treatment effect would be higher than the individual that experiences usual encouragement.**

**MCF :what the outcome on the saving capacity at the end of 12 months would be if y/n experiences the usual encouragement conditions instead of the lockbox condition and all else remains the same.**

**we can calculate the average MCF for the people assigned to the lockbox condition with the factual outcome of the people that experience the usual (encouragement) condition. The factual outcome : What the outcome on the saving capacity is at the end of 12 months when y/n experiences the usual(encouragement) condition.**

## Question 2 (9 points)

a) What are the drop-out rates (those for whom 12 month outcomes are missing, you may use the `has_followup2` variable) across the two treatment and one control conditions? *Hint: you can add, subtract, multiply, and divide tables in R or use Two-Way Tables*

b) Does the nature of either of the treatments (Safebox or Lockbox) suggest to you we should expect drop-out to be higher or lower in one group? What impact might drop-out have on the balance on baseline covariates we expect to obtain with randomization?

c) Subset (or filter) the data (we suggest giving the subset data a new object name) so that it contains only participants who were interviewed at 12 months during the second followup. We will use this subset for the rest of this HW which means it is what is called a complete case analysis. How many participants are left in each group (control group and two treatment arms) after removing those who dropped out of the study?

## Answer 2 Code

```
##using 2-way table
table(rosca$treatment ,rosca$has_followup2)
```

```
##
##             0   1
##   control   9 102
##   lockbox  11 184
##   safebox  10 107
```

```
proportions(table(rosca$treatment,rosca$has_followup2),margin = 1)
```

```
##
##                    0         1
##   control 0.08108108 0.91891892
##   lockbox 0.05641026 0.94358974
##   safebox 0.08547009 0.91452991
```

```
rosca_rmv <- rosca %>% filter(rosca$has_followup2== 1)
table(rosca_rmv$treatment)
```

```
##
## control lockbox safebox
##     102     184     107
```

```
table(rosca$treatment)
```

```
##
## control lockbox safebox
##     111     195     117
```

```r
table(rosca$has_followup2)
```

```
##
##   0   1
##  30 393
```

```r
table(rosca$treatment)
```

```
##
## control lockbox safebox
##     111     195     117
```

```r
table(rosca$treatment=="control")
```

```
##
## FALSE  TRUE
##   312   111
```

```r
table(rosca$has_followup2 == "1")
```

```
##
## FALSE  TRUE
##    30   393
```

```r
rosca_control <- rosca[rosca$treatment=="control",]
table(rosca_control$has_followup2=="1")
```

```
##
## FALSE  TRUE
##     9   102
```

```r
proportions(table(rosca_control$has_followup2))
```

```
##
##          0          1
## 0.08108108 0.91891892
```

```r
rosca_con_12 <- rosca_control %>% filter(rosca_control$has_followup2 == 1)
nrow(rosca_con_12)
```

```
## [1] 102
```

```r
table(rosca$treatment=="lockbox")
```

```
##
## FALSE  TRUE
##   228   195
```

```r
table(rosca$has_followup2 == "1")
```

```
##
## FALSE  TRUE
##    30   393
```

```r
rosca_lock <- rosca[rosca$treatment=="lockbox",]
table(rosca_lock$has_followup2)
```

```
##
##   0   1
##  11 184
```

```r
rosca_lock_12 <- rosca_lock %>% filter(rosca_lock$has_followup2 == 1)
nrow(rosca_lock_12)
```

```
## [1] 184
```

```r
table(rosca$treatment=="safebox")
```

```
##
## FALSE  TRUE
##   306   117
```

```r
table(rosca$has_followup2 == "1")
```

```
##
## FALSE  TRUE
##    30   393
```

```r
rosca_safe <- rosca[rosca$treatment=="safebox",]
table(rosca_safe$has_followup2)
```

```
##
##   0   1
##  10 107
```

```r
proportions(table(rosca_safe$has_followup2))
```

```
##
##          0          1
## 0.08547009 0.91452991
```

```r
rosca_safe_12 <- rosca_safe %>% filter(rosca_safe$has_followup2 == 1)
nrow(rosca_safe_12)
```

```
## [1] 107
```

**Answer 2 Text**

**a. The drop out rates for those in the control group is : 8.108 % .91.89 % is the rate for the ones that didn't drop out at the 12 month cycle.**

**The drop out rates for those in the safebox treatment group is 8.5% and 91.45% for the one's that didn't drop out at the 12 month cycle.**

**the drop out rates for those in the lockbox treatment group is 5.64% and 94.35% for the one's that didn't drop out at the 12 month cycle.**

**b. the lockbox and safebox numbers for dropouts are : 5.6% and 8.5% respectively. This would mean that the lockbox treatment proved to be more effective in saving rather than the safebox, but this difference isn't significantly large. However, the safebox treatment seems to be the treatment that would have more dropouts as compared to the lockbox as the description of the treatment states that they would have control of their savings and also will be given the keys to the lock .**

**the dropout rates would have an impact on the estimate of the average MCF we must determine for the treatment groups(safe and lock) using the usual group.The dropout rates would have an impact on the study especially if the covariates are biased after the dropouts since the randomisation does not stay randomised once there are many dropouts.**

**c. the participant numbers in each control group after the 12 month period : 102 for control, 184 for lockbox and 107 for safebox treatments.**

## Question 3 (10 points)

In this question we will investigate whether receiving a Safebox or Lockbox increases the average amount invested in health products relative to encouragement only. We focus on the outcome measured 12 months from baseline during the second follow-up `fol2_amtinvest`.

a) First, describe the distribution of this outcome over *all* study participants based on the results of the summary() and hist() commands.

b) Then find the average amount invested (in Kenyan shillings) on health products for those in the two treatment groups and the control condition.

c) Then calculate the differences in the the mean of amounts invested in health products between each of the treatment arms and the control group. Briefly interpret the results.

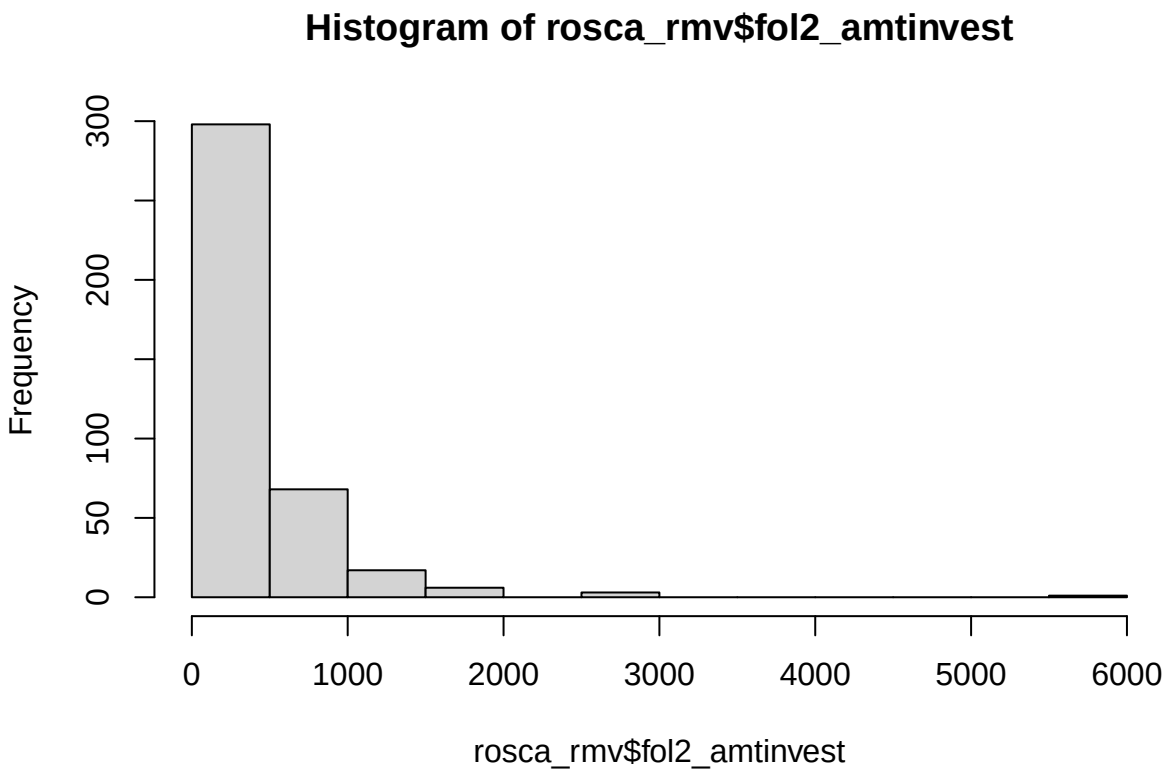
d) State any limitation of using averages (rather than a different summary statistic) to estimate impacts given what you know about the distribution of the outcome measure.

## Answer 3 Code

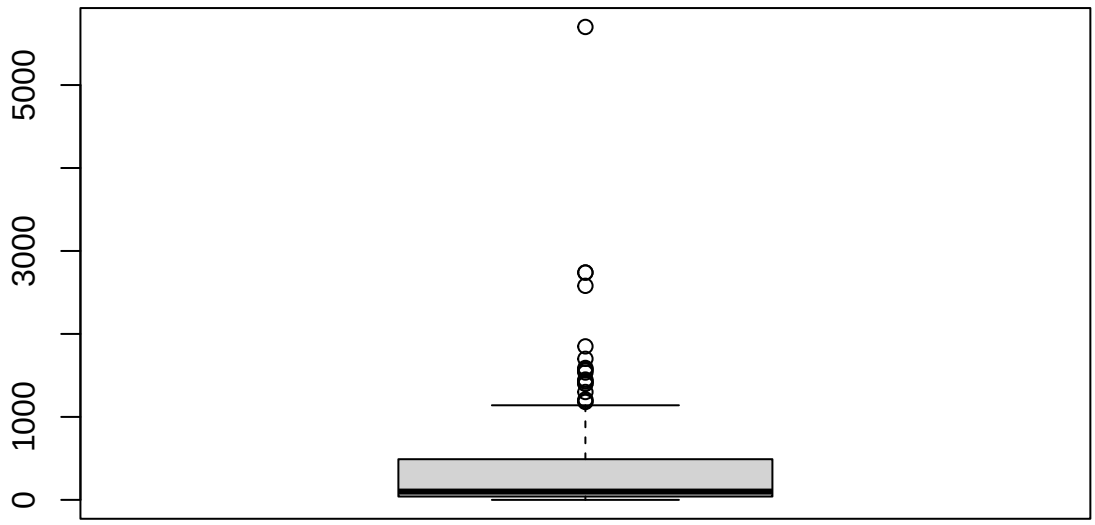
```
summary(rosca_rmv$fol2_amtinvest)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0    40.0   100.0   322.2   490.0  5700.0
```

```
hist(rosca_rmv$fol2_amtinvest)
```



```
boxplot(rosca_rmv$fol2_amtinvest)
```




```
summary(rosca_con_12$fol2_amtinvest)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0    40.0   105.0   257.8   425.0  1399.0
```

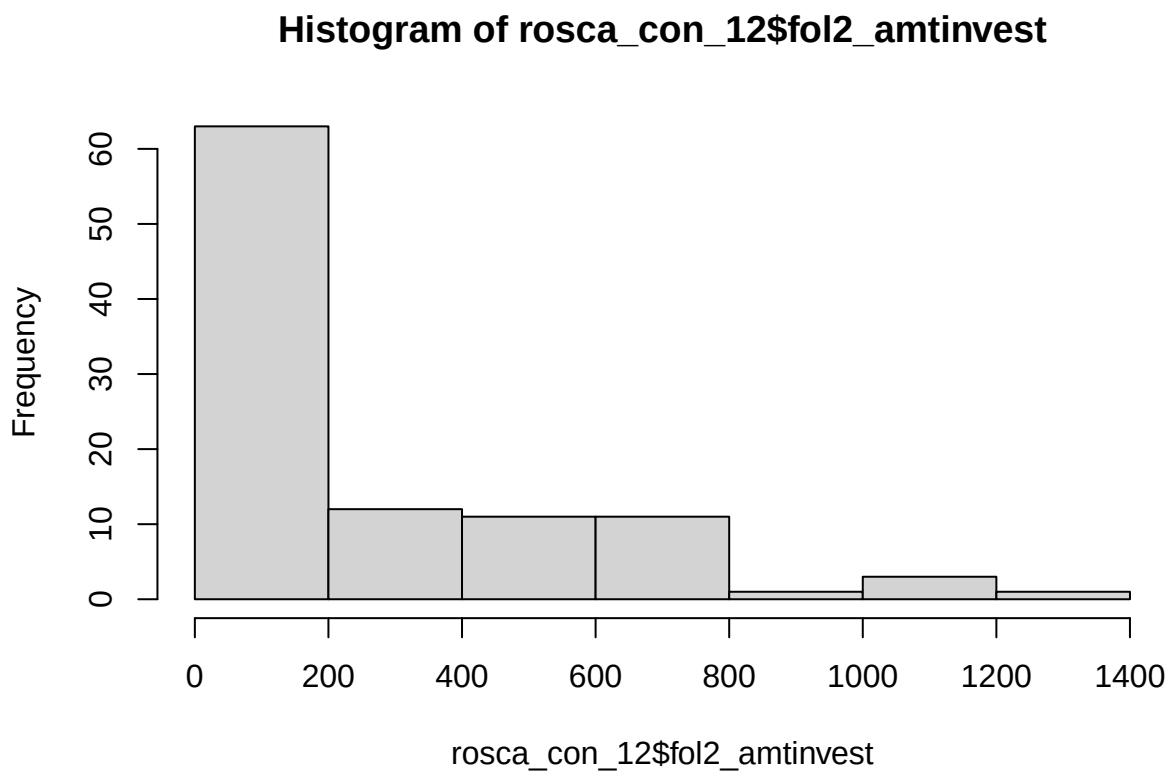
```
summary(rosca_lock_12$fol2_amtinvest)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   38.75   80.00  307.83  470.00 2740.00
```

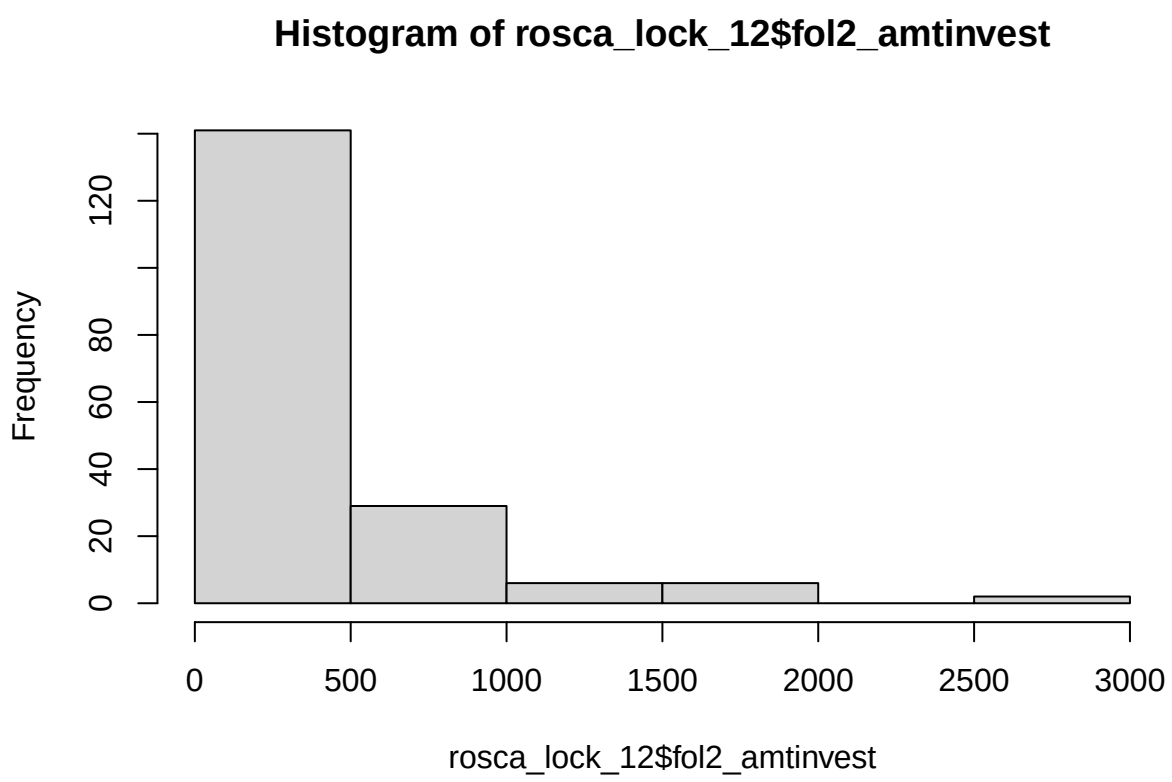
```
summary(rosca_safe_12$fol2_amtinvest)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0    60.0   150.0   408.2   580.0  5700.0
```

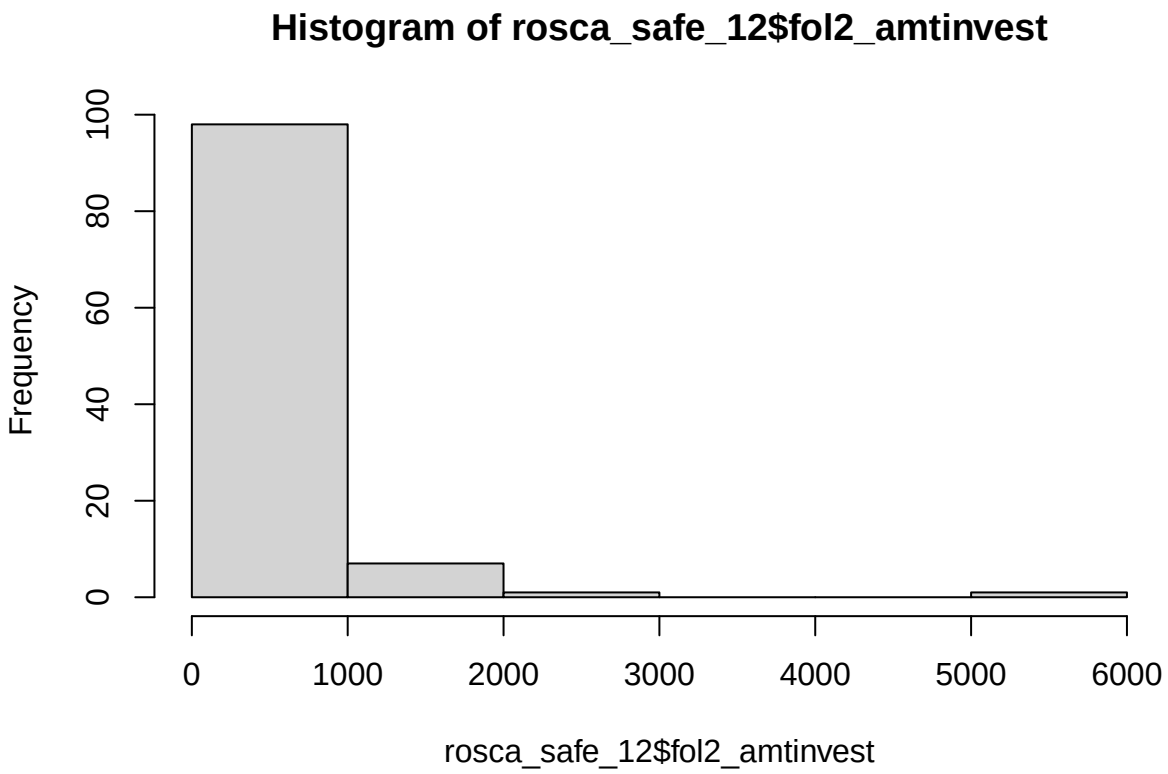
```
hist(rosca_con_12$fol2_amtinvest)
```




```
hist(rosca_lock_12$fol2_amtinvest)
```




```
hist(rosca_safe_12$fol2_amtinvest)
```




```
mean_con <- mean(rosca_con_12$fol2_amtinvest)

mean_safe <- mean(rosca_safe_12$fol2_amtinvest)

mean_lock <- mean(rosca_lock_12$fol2_amtinvest)

mean_safe- mean_con
```

```
## [1] 150.3816
```

```
mean_lock -mean_con
```

```
## [1] 49.99275
```

**Answer 3 Text**

**a. the hist function displays the highest frequency for the saving amount lying between 0-500.the summary displays that the mean investment after 12 months is approximately 322.2. The maximum amount saved is 5700 which lies at the tail end of the graph towards the right .the spread: people saved between 0-5700 shillings. .the middle 50 % of amounts 40 - 490 were saved- this means that quarter people saved 40 orless and 490 or more shillings. the median is 100: half of the people saved 100 shillings or less and half saved 100 or more.**

**there is a right skew for the values we observe. The distribution has a strong righ skew and a singal modal category near zero . there is a spike at the value 0 and there are gaps between the outliers.**

**a. For the control group : we can observe that the group skewed towards the right and 0-200 dollars.**

**for the lockbox group : we can observe that the graph is skewed towards the right of the plot and having saved amounts between 0-500 dollars were the highest to be saved at the end of 12 months.**

**for the safebox group : we can observe that the graph us skewed towards the right of the plot and having amounts between 0-1000 are saved my most people in the group.**

**b. 257.8 is the average amount spent for the control group. 408.2 is the average amount spent for the safebox treatment.307.83 is the average amount spent for the control group.**

**c. for the mean_safe - mean_con = 150.38 and the mean_lock - mean_con =49.99275.**

**the mean_safe:408.215 and the mean_lock : 307.82 and mean_con :257.833.**

**d. the average and mean are not always the best method to represent the data as there could be outliers and it takes the high and low values which can skew the data towards one direction or the other.**

## Question 4 (12 points)

Examine the distribution of the pre-treatment (or baseline) variables - gender (`bg_female`), age (`bg_b1_age`) and marital status (`bg_married`) across the three groups of interest (two treatment groups and control group). Please provide summary statistics and/or figures and interpretations for each of the three pre-treatment variables separately. For dichotomous or categorical variables, the proportions are all that are needed. For continuous variables please provide at least one measure of central tendency, at least one measure of spread, and a figure to show the shape (single or multi-modal, skew). It is not required, but consider recoding the non-numeric variables (gender and marital status) as text rather than numeric variables for ease of interpretation of the results.

a) Gender: describe the distribution of gender in each treatment/control arm

b) Age: describe the distribution of age in each treatment/control arm

c) Marital Status: describe the distribution of marital status in each treatment/control arm

d) What do the results in parts a, b, and c suggest about whether the control group is providing an unbiased estimate of the average MCF for each of the two treatment groups? What two circumstances would need to occur for a baseline covariate to indicate bias in the estimate of the MCF (and hence a biased estimate of the treatment effect)?

## Answer 4 Code

```
# gender


table(rosca_con_12$bg_female)
```

```
##
##  0  1
## 28 74
```

```
rosca_con_12$bg_female.f <- ifelse(rosca_con_12$bg_female== 1,"Female","Male")
proportions(table(rosca_con_12$bg_female.f))
```

```
##
##    Female      Male
## 0.7254902 0.2745098
```

```
table(rosca_safe_12$bg_female)
```

```
##
##  0  1
## 22 85
```

```
rosca_safe_12$bg_female.f1 <- ifelse(rosca_safe_12$bg_female== 1, "Female","Male")
proportions(table(rosca_safe_12$bg_female.f1))
```

```
##
##    Female      Male
## 0.7943925 0.2056075
```

```
table(rosca_lock_12$bg_female)
```

```
##
##   0   1
##  49 135
```

```
rosca_lock_12$bg_female.f2 <- ifelse(rosca_lock_12$bg_female== 1, "Female","Male")
proportions(table(rosca_lock_12$bg_female.f2))
```

```
##
##    Female      Male
## 0.7336957 0.2663043
```

```
# age

table(rosca_con_12$bg_b1_age)
```

```
##
## 18 19 22 24 25 26 27 28 29 30 31 32 33 34 35 36 38 39 40 41 42 43 44 45 46 47
##  1  1  1  5  2  1  4  2  1  5  3  5  2  4  4  1  6  1  5  2  2  5  1  2  2  2
## 48 49 50 52 53 54 55 56 58 62 63 65 66 68 69 70 76 78
##  5  1  4  1  1  2  2  1  1  1  1  4  1  3  1  1  1  1
```

```
mean(rosca_con_12$bg_b1_age)
```

```
## [1] 41.87255
```

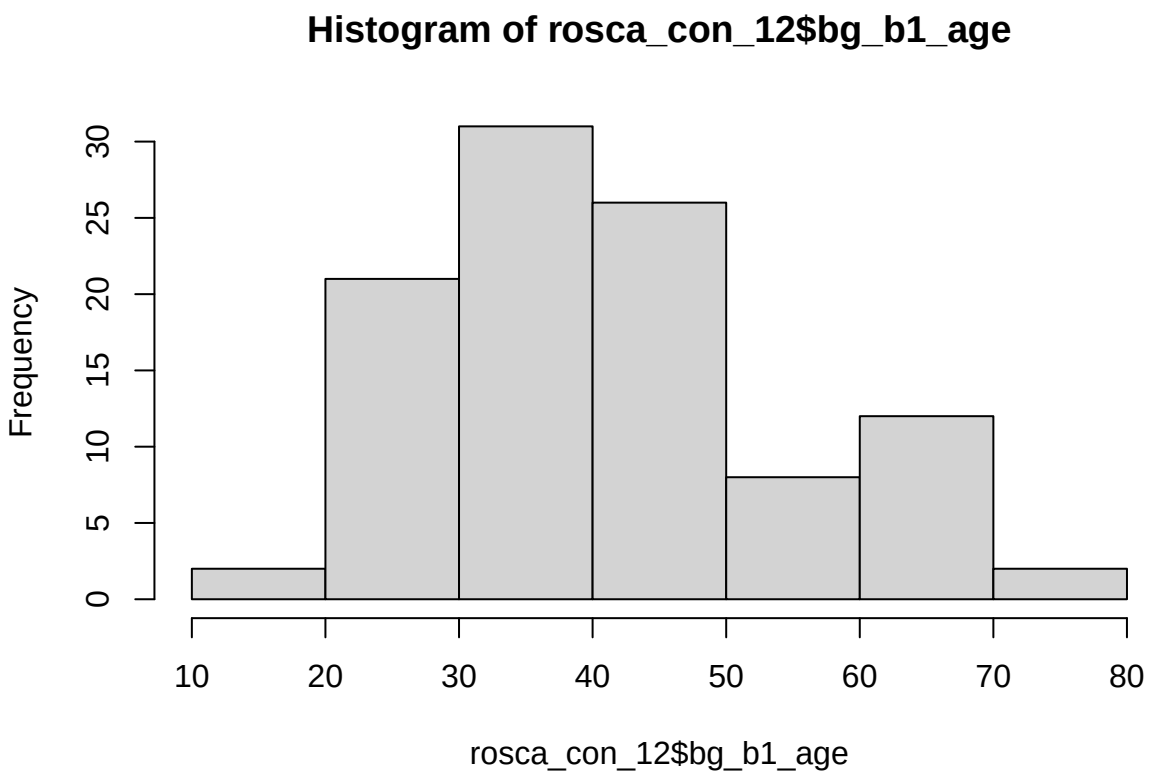
```
range(rosca_con_12$bg_b1_age)
```

```
## [1] 18 78
```

```
median(rosca_con_12$bg_b1_age)
```

```
## [1] 40
```

```
hist(rosca_con_12$bg_b1_age)
```




```
table(rosca_safe_12$bg_b1_age)
```

```
##
## 18 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 40 42 44 45 46 48 49
##  1  1  2  1  1  4  5  3  5  3 10  1  2  1  4  7  4  3  6  5  3  3  7  1  3  3
## 50 51 52 53 55 56 60 63 64 68 70 88
##  2  1  1  2  4  2  1  1  1  1  1  1
```

```
mean(rosca_safe_12$bg_b1_age)
```

```
## [1] 38.54206
```

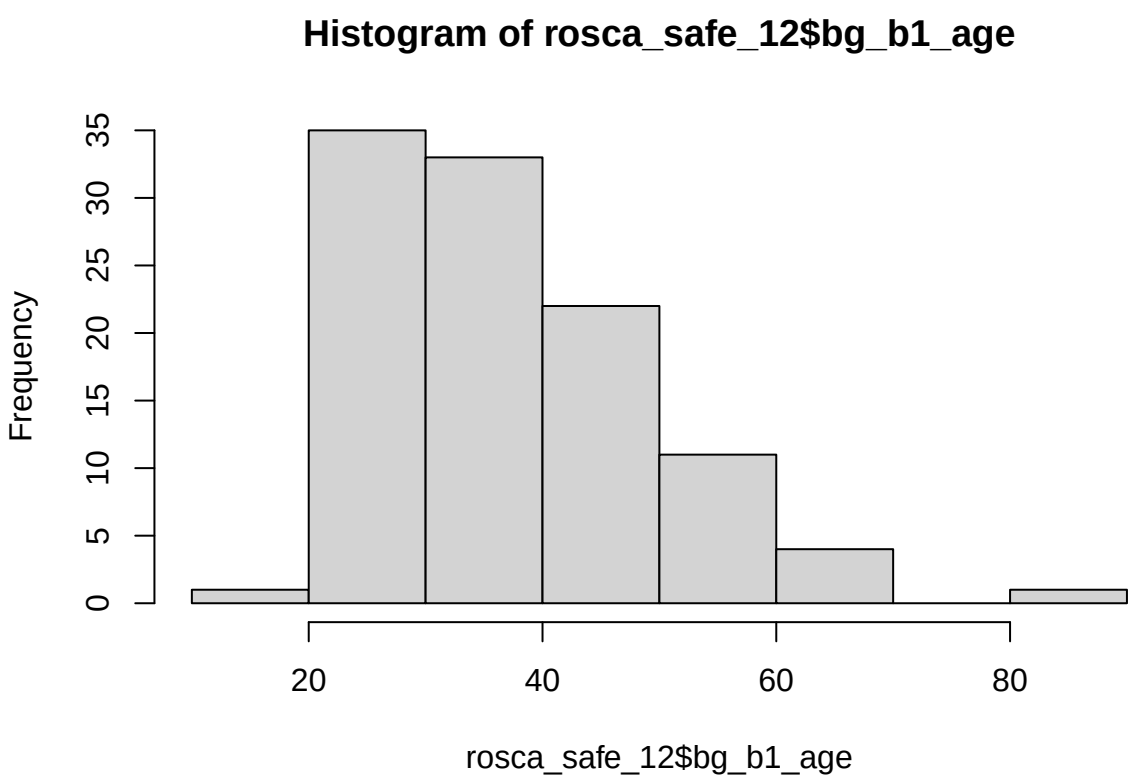
```
range(rosca_safe_12$bg_b1_age)
```

```
## [1] 18 88
```

```
median(rosca_safe_12$bg_b1_age)
```

```
## [1] 36
```

```
hist(rosca_safe_12$bg_b1_age)
```




```
table(rosca_lock_12$bg_b1_age)
```

```
##
## 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 38 39 40 41 42 43 44 45
##  2  6  5  3  5  6  5  4  3  4  4  9  3  5  4  5 11  2  3  1 12  1  7  3  1 10
## 46 47 48 49 50 51 52 53 55 56 57 58 60 62 63 65 66 67 68 70 80
##  5  4  6  2  7  3  5  4  5  2  1  1  1  3  1  1  1  1  2  4  1
```

```
mean(rosca_lock_12$bg_b1_age)
```

```
## [1] 39.58152
```

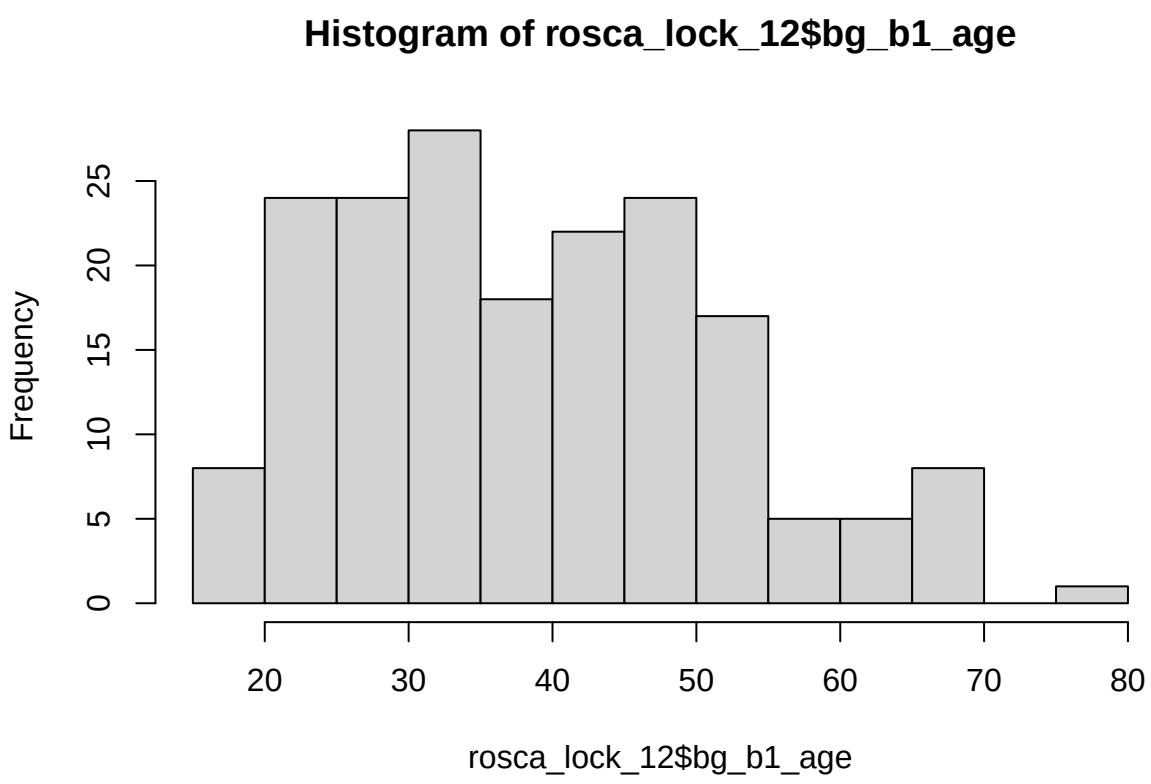
```
range(rosca_lock_12$bg_b1_age)
```

```
## [1] 19 80
```

```
median(rosca_lock_12$bg_b1_age)
```

```
## [1] 40
```

```
hist(rosca_lock_12$bg_b1_age)
```




```
# marital status

rosca_con_12$bg_married
```

```
##   [1] 1 0 1 1 1 1 1 1 1 1 0 0 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 0 1 0
##  [38] 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 0 1 1
##  [75] 1 1 1 0 1 0 0 0 0 1 0 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1
```

```
table(rosca_con_12$bg_married)
```

```
##
##  0  1
## 26 76
```

```
rosca_con_12$bg_marriedcon <- ifelse(rosca_con_12$bg_married== 1,"Married","Not Married")
proportions(table(rosca_con_12$bg_marriedcon))
```

```
##
##     Married Not Married
##    0.745098    0.254902
```

```
rosca_safe_12$bg_married
```

```
##   [1] 1 1 1 1 0 0 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 1 1 1 0 1
##  [38] 1 1 1 1 1 1 0 1 1 0 1 0 0 0 0 1 1 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 0 0 1
##  [75] 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1
```

```
table(rosca_safe_12$bg_married)
```

```
##
##  0  1
## 29 78
```

```
rosca_safe_12$bg_marriedsafe <- ifelse(rosca_safe_12$bg_married== 1,"Married","Not Married")
proportions(table(rosca_safe_12$bg_marriedsafe))
```

```
##
##     Married Not Married
##    0.728972    0.271028
```

```
rosca_lock_12$bg_married
```

```
##   [1] 1 1 0 0 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [38] 1 1 1 1 0 1 1 1 1 1 1 1 0 0 1 0 0 0 0 0 0 1 1 0 1 1 1 0 1 0 1 1 0 0 1 1 0
##  [75] 1 1 0 1 1 1 1 1 0 1 1 0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 0 1 1
## [112] 1 1 1 1 1 1 1 0 1 1 1 1 0 1 1 0 0 1 1 1 1 1 0 0 1 1 0 1 1 1 1 1 0 1 1 1 1
## [149] 0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1
```

```
table(rosca_lock_12$bg_married)
```

```
##
##   0   1
##  44 140
```

```
rosca_lock_12$bg_marriedlock <- ifelse(rosca_lock_12$bg_married== 1,"Married","Not Married")
proportions(table(rosca_lock_12$bg_marriedlock))
```

```
##
##     Married Not Married
##   0.7608696   0.2391304
```

### Answer 4 Text

**a In the control group : 73.36% were female and 26.63% were male.In the lockbox group 72.54% were female and 27.45% were male. In the safe box it was 79.43 % were female and 20.56 % were male.**

**b control group mean and range and median :41.87 , 18 - 78 and 40. Safe box group mean, range and median: 38.54 ,18-88 and 36. lockbox group mean ,range, median: 39.58 , 19-80 , 40.**

**c control group : 74.5 % are married and 25.49 % are not married. Safebox group : 72.89% are married and 27.10%. Lockbox group: 76% are married and 23.91% are not married.**

**d The gender proportions between the control group to the lockbox treatment group (treatment group 1)are approximately the same. Therefore, it would provide an unbiased estimate of the treatment group 1.the proportions between the control group and the safebox treatment group are slightly different but it is not significant enough with approximately a 6% difference in the number of women and a 6 % in the number of men between both the control and safebox (treatment group 2). The bias would be little to nothing with a significant change observed above.**

**1.for the baseline covariates to indicate bias : we must check for differences in the observed basline covariates between those receiving the alternative and the intervention these are known as confounders or lurking variables - especially those related to the outcome.If these confounders are having noticable differences then this would lead to a confounding bias.**

**2.If there is a factor or baseline covariate that is also associated to the preditive outcome it can cause the bias in estimate of the MCF.**

### Question 5 (10 points)

In this question we will investigate whether receiving a Safebox or a Lockbox had different effects on the investment of *married* versus *unmarried women*. [We recommend that you filter on gender and marital status to obtain two data sets, one with all married women the other with all unmarried women.]

a) Compare the mean investment in health products among married women across the three treatment conditions (two treatments and the control). This involves calculating the three means and then calculating the two treatment effects (Safebox and Lockbox, each relative to Control). Briefly interpret what you find.

b) Then compare the mean investment in health products among unmarried women in the three treatment conditions. Briefly interpret what you find.

c) Comment on any differences in treatment effects for the two groups of interest (married relative to unmarried women).

d) How does this analysis address any potential bias issues discussed in Question 4?

### Answer 5 Code

```r
# Married women


rosca_gender <- rosca_rmv[rosca_rmv$bg_female == "1",]
view(rosca_gender)
rosca_gender_married <- rosca_gender[rosca_gender$bg_married == 1,]
rosca_married_safe <- rosca_gender_married[rosca_gender_married$safe_box==1,]
mean(rosca_married_safe$fol2_amtinvest,na.rm=TRUE)
```

```
## [1] 557.1356
```

```r
rosca_married_lock <- rosca_gender_married[rosca_gender_married$locked_box==1,]
mean(rosca_married_lock$fol2_amtinvest,na.rm=TRUE)
```

```
## [1] 332.433
```

```r
rosca_married_con <- rosca_gender_married[rosca_gender_married$encouragement==1,]
mean(rosca_married_con$fol2_amtinvest,na.rm=TRUE)
```

```
## [1] 239.66
```

```r
# Unmarried women
rosca_gender_unmarried <- rosca_gender[rosca_gender$bg_married == 0,]
rosca_unmarried_safe <- rosca_gender_unmarried[rosca_gender_unmarried$safe_box==1,]
mean(rosca_unmarried_safe$fol2_amtinvest,na.rm=TRUE)
```

```
## [1] 264.0385
```

```r
rosca_gender_unmarried <- rosca_gender[rosca_gender$bg_married == 0,]
rosca_unmarried_lock <- rosca_gender_unmarried[rosca_gender_unmarried$locked_box==1,]
mean(rosca_unmarried_lock$fol2_amtinvest,na.rm=TRUE)
```

```
## [1] 220.4737
```

```r
rosca_gender_unmarried <- rosca_gender[rosca_gender$bg_married == 0,]
rosca_unmarried_con <- rosca_gender_unmarried[rosca_gender_unmarried$encouragement==1,]
mean(rosca_unmarried_con$fol2_amtinvest,na.rm=TRUE)
```

```
## [1] 218.5417
```

**Answer 5 Text**

**a. The three means for married women between the safebox, lockbox and control group are as follows: 557.13, 332.433 and 239.66. The difference in mean for the safebox and the control group is : 317.47. The difference in means for the lockbox and the control group is : 92.773. Therefore the highest difference we can observe is between the safebox and control group. We can infer that the highest mean for saving amongst married women in health products was seen in the safebox intervention.**

**b. The three means for unmarried women for the safebox, lockbox and control group are as follows : 264.03, 220.47, 218.54.The difference in means for the safebox and control group is 45.49. The difference in mean for the lockbox and the control group is : 1.93. We can observe a higher difference between the safebox group and control group. We can infer that the saving between all groups is quite similar with slight increase.**

**c. There are significant differences between the treatment effects on married and unmarried women. Within married women group the treatments varied and skewed widely from the control group results. difference in means :Safebox : 557.13-264.03 = 293.1 lockbox = 111.96 and control : 21.12 . Whereas, the data for unmarried women that the treatments didn't differ significantly and remained the same as compared to the control group results. the safebox and lock box treatments had different results on married women as compared to not married women.**

**d. In the 4th question we observed that the covariates are not significantly going to bias the mcf, however with the new inference at hand : we can comment on the biasing effect that the data which includes gender and marital status has on the mcf. The groups measured are married and not married women and since the difference between the two groups are significantly different it can have a significant impact on the outcome. The confounders in this case are related to the outcome which could create a bias.**