HW4_Markdown

2024-02-22

Manay Divatia

md46245

1

a

```
rosca = read.csv("~/Downloads/rosca.csv")
rosca$treatment <- ifelse(rosca$encouragement == 1, "control",
                          ifelse(rosca$safe_box == 1, "safebox", "lockbox"))
nrow(rosca[rosca$treatment == "control", ])
nrow(rosca[rosca$treatment == "safebox", ])
nrow(rosca[rosca$treatment == "lockbox", ])
```

There are 111 individuals in the control, 117 for the safebox treatment, and 195 for the lockbox treatment.

b

```
nrow(rosca)
rosca_followup <- rosca[rosca$has_followup2 == 1, ]
nrow(rosca_followup)
nrow(rosca_followup[rosca_followup$treatment == "control", ])
nrow(rosca_followup[rosca_followup$treatment == "safebox", ])
nrow(rosca_followup[rosca_followup$treatment == "lockbox", ])
```

There are 393 left in the data which is a lot considering we started with 423. The most people (11) drop from the lockbox treatment which could skew the data later on because those people were less likely anyway to invest money.

c

```
mean(rosca_followup[rosca_followup$treatment == "control", ]$fol2_amtinvest)
mean(rosca_followup[rosca_followup$treatment == "safebox", ]$fol2_amtinvest)
mean(rosca_followup[rosca_followup$treatment == "lockbox", ]$fol2_amtinvest)
```

The safebox had by far the most invested with 408. This was followed by the lockbox which had 308 and the control which had 258. This means that both the treatments seemed to have a positive effect on the amount invested in the end.

d

```
function_d <- function(treatment) {
  print(mean(rosca_followup[rosca_followup$treatment == treatment,
                            ]$bg_female))
  print(mean(rosca_followup[rosca_followup$treatment == treatment,
                            ]$bg_married))
  print(mean(rosca_followup[rosca_followup$treatment == treatment,
                            ]$bg_b1_age))
}
function_d("control")
function_d("safebox")
function_d("lockbox")
```

The safebox treatment group had the most females and the least married people. These could affect the results. However, for the rest of the data, the treatments and control seemed to be similar in terms of percentage female, age, and marriage status.

e

```
mean(rosca_followup[rosca_followup$treatment == "control" &
                      rosca_followup$bg_married == 1, ]$fol2_amtinvest)
mean(rosca_followup[rosca_followup$treatment == "safebox" &
                      rosca_followup$bg_married == 1, ]$fol2_amtinvest)
mean(rosca_followup[rosca_followup$treatment == "lockbox" &
                      rosca_followup$bg_married == 1, ]$fol2_amtinvest)
mean(rosca_followup[rosca_followup$treatment == "control" &
                      rosca_followup$bg_married == 0, ]$fol2_amtinvest)
mean(rosca_followup[rosca_followup$treatment == "safebox" &
                      rosca_followup$bg_married == 0, ]$fol2_amtinvest)
mean(rosca_followup[rosca_followup$treatment == "lockbox" &
                      rosca_followup$bg_married == 0, ]$fol2_amtinvest)
```

In all three treatments, when the individual was married, the amount invested was less. This was the biggest changei n the safebox treatment which went from an average of 470 to 241. Clearly, marriage status played a large role. It would be interesting to see how gender and age also affected amount invested.

2

a

```
predimed = read.csv("~/Downloads/predimed.csv")
total_rows = nrow(predimed)
nrow(predimed[predimed$event == "Yes", ]) / total_rows
nrow(predimed[predimed$group == "MedDiet + Nuts" |
                predimed$group == "MedDiet + VOO" , ]) / total_rows
nrow(predimed[(predimed$group == "MedDiet + Nuts" | predimed$group ==
                 "MedDiet + VOO") & predimed$event == "Yes", ]) / total_rows
any_meddiet <- predimed[predimed$group == "MedDiet + Nuts" | predimed$group
                          == "MedDiet + VOO" , ]
nrow(any_meddiet[any_meddiet$event == "Yes", ]) / nrow(any_meddiet)
```

P(event) is the chance of the event happening which was 3.98%. P(any MedDiet) was the chance of the individual being on any of the mediterranean diets which was 67.7%. P(event, any MedDiet) is the chance of the cardiac event happening and the individual being on any of the mediterranean diets which was 2.45%.

P(event | any MedDiet) is the chance of the cardiac event happening given that the individual is on any of the mediterranean diets which was 3.62%.

b

```
control_predimed <- predimed[predimed$group == "Control", ]
nrow(control_predimed[control_predimed$event == "Yes", ]) /
  nrow(control_predimed)
```

P(event | control) is the chance of a cardiac event given that the individual is not given the diet. The chance was 4.7%. From the previous question, we see a change from 3.62 to 4.7% which seems like a big difference to consider that the mediterranean diet decreases the chance of a cardiac event. However, it is important to note that there could be other factors involved.

c

Additional information like eating habits of the individual, prior health concerns, family history, and location could all affect our results and should be tracked so that we can consider them.

d

```
nrow(any_meddiet[any_meddiet$sex == "Male" & any_meddiet$event == "Yes", ]) / nrow(any_meddiet[any_medd
nrow(any_meddiet[any_meddiet$sex == "Female" & any_meddiet$event == "Yes", ]) / nrow(any_meddiet[any_me
nrow(predimed[predimed$sex == "Male" & predimed$event == "Yes" & predimed$group
              == "Control", ]) / nrow(predimed[predimed$sex == "Male" & predimed$group=="Control", ])
nrow(predimed[predimed$sex == "Female" & predimed$event == "Yes" &
                predimed$group == "Control",]) /
  nrow(predimed[predimed$sex == "Female" & predimed$group == "Control", ])
```

In both cases of control and mediterranean diet, it looks like the likelihood of a cardiac event is lower for women than it is for men which could be something to consider when looking at general trends in our data. Additionally, by splitting it up, it looks like the chance of a cardiac event is higher for both males and females when given the control.