

## ECON 204: Youth Unemployment Interventions Lab (R Version)

You are going to be analyzing (some of the) data from the paper “A Firm of One’s Own: Experimental Evidence on Credit Constraints and Occupational Choice” by Andrew Brudevold-Newman, Maddalena Honorati, Gerald Ipapa, Pamela Jakiela, and Owen Ozier, published in the *Review of Economics and Statistics*. To get started, download `ECON-204-youth-lab-template.Ro`. Save it on your computer, replacing the “template” part of the file name with your own last name. Delete the word “TEMPLATE” from the comment at the top of the file, and add your name on the “YOUR NAME” line. Modify the code so that it runs from start to finish, loading the data from github, and generates your finished figure.

0. Before you get started on the steps in the lab, familiarize yourself with the data. What values does the variable `treatment` take on?

1. The code

```
means <- mydata %>%
  group_by(treatment) %>%
  summarize(mean_selfemp = mean(selfemployment),
            mean_income = mean(income),
            sd_selfemp = sd(selfemployment),
            n = n(),
            se_selfemp = sd_selfemp / sqrt(n))
```

creates a treatment-level data frame containing means and standard errors of different variables. Look at the information contained in this data frame. Then extend the code so that you also store the standard error of the mean of the variable `income` in the new data frame as well.

2. The code below generates a bar graph showing the mean rate of self-employment in the Control group and the Franchise treatment arm.

```
ggplot(data = subset(means, treatment != "Grant"),
       aes(x = treatment, y = mean_selfemp, fill = treatment)) +
  geom_col(alpha = 0.8) +
  xlab(" ") +
  scale_fill_manual(values = c(
    "Control" = "tomato3",
    "Franchise" = "#0072B2"
  )) +
  theme(
    axis.title.y = element_text(size = 16),
    axis.text.x = element_text(size = 12),
    axis.text.y = element_text(size = 12),
    legend.position="none"
  )
```

Go through the code and make sure that you understand each part of the command. Then modify the code so that:

- (a) The outcome is income rather than self-employment
  - (b) The range of the y-axis and the axis labels are adjusted accordingly
  - (c) The graph includes the grant treatment in addition to the franchise treatment and the control group
  - (d) The y-axis has an appropriate title
3. Extend the code so that you define two new variables, **upper** and **lower**, indicating the upper and lower limits of the 95 percent confidence interval for the mean of **income**.
  4. Add the confidence intervals to your graph as an additional geom. Instead of bars, you will plot these using **ggplot**'s **geom\_errorbar** geometry. Adjust the y-axis labels as needed. Save your graph as a pdf so that you can upload it to gradescope together with your script.