

# In-Class Lab 11

ECON 425 (Justin Heflin, West Virginia University)

April 26, 2023

The purpose of this lab is to practice using R for using dummy variables in R. The lab may be completed as a group. To receive credit, upload your .R script to the appropriate place on eCampus (“In-Class Labs” folder).

## For starters

Open a new R script (named ICL11\_XYZ.R, where XYZ are your initials)

## Clean out/“Sweep” R Studio

Click the broom in the Environment panel (top-right), it is directly below the Tutorial button. Also, in the bottom-right panel, click the Plots button and then click the broom in that panel. This should help with loading things into R.

## R Packages

```
install.packages("magrittr")  
install.packages("broom")
```

```
library(ggplot2)  
library(wooldridge)  
library(modelsummary)
```

## Load the data

We’ll use a new data set on extramarital affairs, called `affairs` from the Wooldridge R Package.

```
affair_data <- as.data.frame(affairs)
```

## Variable Names and brief description:

- male: =1 if male
- age: in years
- yrs marr: years married
- kids =1 if have kids

- ratemarr: 5 = vry hap marr, 4 = hap than avg, 3 = avg, 2 = smewht unhap, 1 = vry unhap
- affair: =1 if had at least one affair
- naffairs: number of affairs within last year

Check out what's in the data by typing

```
summary(affair_data)
```

You'll notice that there are a number of variables that only take on 0/1 values: `male`, `kids`, `affair`, `hapavg`, `vryrel`, etc. There are also variables that take on a few different values: `relig`, `occup`, and `ratemarr`.

You can create a histogram of a factor variable in `ggplot()` as follows:

```
ggplot(affair_data, aes(x=ratemarr)) + geom_bar()
```

## Multiple regression with factor variables

Let's run a regression with `naffairs` as the dependent variable and `male`, `yrs marr`, `kids`, and `ratemarr` as the covariates.

```
regression <- lm(naffairs ~ male + yrs marr + kids + ratemarr, data=affair_data)
summary(regression)
```

## Linear Probability Model

Let's run the same regression as before, but this time use `affair` as the dependent variable. What happens when you run the following code?

```
regression_2 <- lm(affair ~ male + yrs marr + kids + ratemarr, data=affair_data)
summary(regression_2)
```

Finally, let's run a more flexible model where we allow the effect of fathers and mothers to be different. The way to do this in `lm()` is as follows:

```
regression_3 <- lm(affair ~ male*kids + yrs marr + ratemarr, data=affair_data)
summary(regression_3)
```