# Overview of First Half of ECON 4050

Your Name

2024-10-21

Always understand the data before using it by doing something like this:

```
summary(mtcars)
```

```
##       mpg             cyl             disp             hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##       drat             wt             qsec             vs
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
##        am              gear            carb
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
##  Median :0.0000   Median :4.000   Median :2.000
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

```
view(mtcars)
```

## The Linear Regression

1. Create a scatter plot with mpg on the y-axis and cyl on the x-axis, which is the relationship between miles per gallon and cylinders for a given car.

2. Regress mpg on cyl (dependent on independent) and interpret each coefficient.

3. Add the variable `am` to the previous model and rerun the regression. The `am` variable represents the type of transmission (0 = automatic, 1 = manual). What is the interpretation of each coefficient.

4. Is this relationship causal? If not, what is a potential omitted variable that may be important to consider when examining the relationship between mpg and cyl?

5. Create a new object called `mpgextra` and add a variable that is the log of mpg. Why would we want to take the logaritm of a variable? Rerun the regression from question 2 but using the log of mpg as the outcome and interpret the coefficients.

## Sampling, Confidence Intervals, and Hypothesis Testing

6. If you run the same regression on different samples, do you expect to get the exact same coefficient estimates or different ones? Why?

7. When a sample size gets larger, what happens to the sampling variation? Regardless of how the underlying population distribution looks like, when sample means are based on larger and larger sample sizes, the sampling distribution of these sample means becomes both more and more like what type of distribution?

8. How does the width of the confidence interval change as the confidence level increases? Why?

9. How does the width of the confidence interval change as the sample size increases? Why?

10. Define the null and alternative hypothesis. What does it mean to have a very small p-value?

11. In economics, what action should you take with regards to the null and alternative hypothesis if you obtain a p-value of 0.007? How about if it was instead 0.7?

12. What elements do you need to construct a confidence interval? What us the formula for a 95% confidence interval and what does it mean?