**Econ Lab 1 Instructions**

**App Downloads**

An app-developing company is interested in finding how income and age affect demand for one of its apps. You and your team are a consultant for this company. You suggest to the company to collect some data first. The company does not have any data, so you suggest that they get data from the online platforms on which the app is sold. For example, Apple App Store or Android Play can provide data on the age of those who downloaded the app. But what about income? These platforms do not collect data on users’ income.

You come up with a different solution: The online platforms collect data on users’ locations. What if you use location as a proxy for income? The American Community Survey (ACS) of the U.S. Census Bureau provides data on the average income in each zip code. What if you use the average income in each zip code as a proxy for users’ income? It is not the same as users’ income, but better than nothing.

In addition, you need a measure for app demand. Since you have data on zip code income from ACS, you decide to use the number of app downloads in each zip code as a measure of demand for the app. Now, that two of your variables are at the zip code level, you decide to use the average age in each zip code as the measure of age.

After about a week of work, you and your team have collected and prepared the data for the analysis. The data set is in a CSV file called “AppDownloads.csv” on Canvas. Download this dataset and save it in a folder.

1. Load the data into a data frame called “Data”.
2. How many observations are in this data frame? How many variables?
3. View the data frame. What are the variables names?
4. Get the summary statistics of the variables in this data frame (using the summary function.) Copy and paste your results in your assignment. Before doing any analysis, you have to check for abnormalities in your data. Is there anything abnormal in the data? (Hint: you may not find anything abnormal, but this is an important step in any data analysis.)
5. Create the histogram of each variable by using the function hist. Copy and paste your graphs in your assignment. Hint: the simple usage of this function is as follows: hist(dataframe$varname). The goal of creating these histograms is to look for abnormalities.
6. Run a regression of zip\_Downloads on zip\_Income. Write down the regression equation using the results. Interpret the coefficient of zip\_Income.
7. Define a dummy variable called Young, which is equal to one if zip\_Age is less than 45 and zero otherwise.
8. Add Young to the regression in step 6. Write down the regression equation. Plug in your estimated coefficients in your regression equation. Interpret each coefficient.
9. Define four dummy variables representing four age groups: 1) Below35, which is equal to one when Age is below 35 and, zero otherwise, 2) Bet35to45, which is equal to one when Age is between 35 and 45, and zero otherwise, 3) Bet45to55, which is equal to one when Age is between 45 and 55, and zero otherwise, and 4) Above55, which is equal to one when Age is 55 and more, and zero otherwise.
10. Imagine we run a regression of zip\_Downloads on zip\_Income and all four dummy variables introduced in step 11. Before you run this regression, do you think it is possible to include all four dummy variables in this regression? Why?
11. Run this regression equation with three dummy variables of your choice (among the four). Write down the regression equation using the results.
12. Interpret the coefficients of the dummy variables in step 11.