

Econ 3040: Instructions for Course Project

Ryan T. Godwin

Find and download a dataset

You may not use the same dataset as any other student. Doing so will automatically open up your project to scrutiny regarding academic misconduct. **All projects must be accomplished independently.** If you are concerned about your dataset, you may email me to confirm that your dataset is not in use by another student. **The project is due by April 16.**

You may use any data source. I suggest using [kaggle](#). kaggle is a source for free, fun, and easy to use datasets.

The dataset that you select must have at least three variables, but more is better. The y variable, and at least one of the x variables, should be continuous (not an integer). You will use the dataset to investigate a causal relationship that you are interested in. That is, think of how one of the x variables in your dataset might cause the y variable. Part of your grade will depend on whether there is a sensible relationship between x and y . Your ultimate goals are to:

- a) Test to see if your x variable of interest can have a causal effect on the y variable.
- b) Estimate by how much a change in your variable of interest leads to a change in y .
- c) Choose the right regression (population) model in order to accomplish goals (a) and (b). This includes testing to see if there are any *non-linear effects*, any possible *interactions*, and eliminating variables that are *insignificant*.

Components of your project

Your project needs to have the following sections:

Introduction: Describe the model and what you are trying to estimate.

Data Description: Briefly describe the source of the data, the variables in the data set, provide summary statistics and a scatterplot for the y variable and x variable(s) of interest.

Results: Estimate several regression models, and report the results in a table. Eliminate variables from the model that are *insignificant*. You should attempt to capture any *non-linear effects*. You should use at least one *interaction term*. Use F -tests (and/or t -tests) and *adjusted R-square* to select your preferred model.

Conclusion: Use your preferred model to test to see if the x variable(s) of interest does not cause y (goal (a)). Use your preferred model to interpret the causal effect of x on y .

References: Cite your dataset, R, and any other references that you use.

Dataset and R Code: You must submit the dataset and R Code that you use as separate files.