

PL 9239: Assignment 3

Deadline: Wednesday March 30th

Please write up all your answers in either a R script or a Word document if you prefer – this largely depends on how you want to do the graph in the last exercise.

- Irrespective of submitting R code or a word document, this document has to contain all code. The R code should be running from top to bottom in one go once the working directory is adapted to the respective machine.
- Any answers that you are adding in your own words should also go into the document.
- Please make sure that you structure your script well so that it is easy to read.

Bivariate Relationships

At the NGO :MOVE: you are having an exciting week ahead of you. You have two data sets that you want to analyse about traffic in Cardiff.

Summer (1 Point)

In a first step you want to understand whether the variable regarding summer is related to other variables.

- Measure the association between the summer variable with all other categorical variables. How strong is each respective association?
- Now also study the statistical significance of these associations.
- Taking the statistical uncertainty into account, what do you conclude for each of the substantive effects?

Family Income in 2021 (1 Point)

You now want to understand how the family income in 2021 is associated to other variables.

- Measure the association between the family income in 2021 and the family income a year before, the likelihood to own a bike, the age and the commuting distance. How strong is each respective association?
- Now also study the statistical significance of these associations.
- Taking the statistical uncertainty into account, what do you conclude for each of the substantive effects?

Multivariate Relationships

Continue with the same data. You are still interested in the family income in 2021.

Bivariate Regression (2 points)

- Run a bivariate regression where you regress the family income in 2021 on the family income in 2020.
- Interpret the two key parameters. What do they tell you regarding the substantive effect?
- What can you say regarding the statistical uncertainty?

Regression With a Dummy (3 points)

- Generate four dummy variables from the data that indicates the preferred means for transport: one for walking, using the car, taking the bus and taking the bike respectively.

- Regress the family income in 2021 on the family income in 2020 and the dummies for walking, using the car, taking the bus and taking the bike in one regression. You can interpret the dummies as respective differences from those who scoot.
- What do you conclude substantively?
- What do you conclude regarding the statistical uncertainty?

Analysis of a Data Set (3 points)

The second data set you will be working with is similar to week 6 in the lab. You are analysing what makes people spend money for bike gear.

- The goal of this exercise is to draw a graph that connects the different variables with one another. For example, for explanatory variables X1 and X2 and the outcome variable Y, you could draw something like $X1 \rightarrow X2 \rightarrow Y$. Feel free to either type up the relationship you are finding. You can of course also drawing the graph in a figure and including it in a word document.
- Report the strength of the association between the variables in your graph and interpret the statistical uncertainty.
- Report all R code that shows how you are getting to your conclusion.

For this exercise you can go *all in* and show what you have learned. Plot and describe data to get a first idea how the data looks like. Do bivariate analyses with plots and measures of association to understand how they are related. Finally, use statistical control to be able to identify confounders.

Enjoy!