

Lab 10 Activity

For this lab activity we will be working with the `MplsDemo` dataset. This is data about 84 neighborhoods of Minneapolis and the data comes from the [2015 American Community Survey](#). Run the following code to load the data as `dat`.

```
dat <- carData::MplsDemo
```

Here is a description of the columns in the data:

Variable	Description
neighborhood	Name of the neighborhood
population	Total population
white	Fraction of the population estimated to be black
black	Fraction of the population estimated to be white
foreignBorn	Fraction of the population estimated to be foreign born
hhIncome	Estimated median household income
poverty	Estimated fraction earning less than twice the poverty level
collegeGrad	Estimated fraction with a college degree

1. For this question, we will practice calculating *partial* correlations and interpreting them. For each partial correlation, first calculate it “manually” using the appropriate regression residuals and then check your results with the `pcor()` function.

- Calculate the Pearson correlation between `poverty` and `black`. Then calculate the partial correlation between `poverty` and `black` accounting for `collegeGrad`. What happens to the relationship between `poverty` and `black` *after* accounting for `collegeGrad`? What do you think this means in practice?
- Calculate the Pearson correlation between `hhIncome` and `collegeGrad`. Then calculate the partial correlation between `hhIncome` and `collegeGrad` accounting for both `foreignBorn` *and* `population`. What happens to the relationship between `hhIncome` and `collegeGrad` *after* accounting for `foreignBorn` *and* `population`?

2. For this question, we will practice calculating *semi-partial* correlations and interpreting them. For each semi-partial correlation, first calculate it “manually” using the appropriate regression residuals and then check your results with the `spcor()` function.

- Calculate the Pearson correlation between `hhIncome` and `white`. Then calculate the semi-partial correlation between `hhIncome` and `white` after accounting for `collegeGrad` for `hhIncome` only. What happens to the relationship between `hhIncome` and `white`? What does this imply about the usefulness of `white` as a predictor of `hhIncome` once we know about `collegeGrad`?

3. Run a multiple regression with `collegeGrad`, `white`, `population`, and `foreignBorn` predicting `hhIncome`. We will look at some regression diagnostics, so remember to name the rows of the dataset with the `neighborhood` variable such that the neighborhood name shows up when printing diagnostics or creating plots.

- Does this regression seem to do a good job at predicting `hhIncome`?
- Use the `influencePlot()` function to create a plot of some regression diagnostics by neighborhood. are there any neighborhoods that stand out? (**NOTE:** the `influencePlot()` function will also print the 5 data points that are overall the most extreme; this does not necessarily mean that those points are problematic or warrant further attention.)

4. Calculate DFBETAS and *only show* the DFBETAS for the neighborhoods you identified as potentially extreme. Would the regression results change by that much if these points were removed? DO you think you should be concerned about these points? (**NOTE:** DFBETAS can be interpreted in standard deviation units)