**PA 528 Public Program Evaluation**

Professor Jing Wang

**Problem Set 3: Regression** (9%)

Getting started

For this problem set, you’ll practice running and interpreting regression models using data about penguins in Antarctica.

You’ll be doing all your R work in R Markdown. You should use an RStudio Project to keep your files well organized (either on your computer or on RStudio.cloud). Either create a new project for this exercise only, or make a project for all your work in this class.

You’ll need to download a R Markdown file named “problem-set-3.Rmd” with a template for this problem set. Go to course blackboard — Assignments — Problem Set 3. Download the .Rmd file below from your blackboard and include it in your project:

* + Problem-set-3.Rmd

You’ll also need to download a CSV file (which could also be found on Blackboard under the same folder) and put it somewhere on your computer—preferably in a folder named data in your project folder:

* penguins.csv

In the end, the structure of your project directory should look something like this:

your-project-name\

problem-set-3.Rmd

your-project-name.Rproj

data\

penguins.csv

You’ll need to make sure you have these packages installed on your computer: tidyverse and modelsummary. If you try to load one of those packages with library(tidyverse) or library(modelsummary), etc., and R gives an error that the package is missing, use the “Packages” panel in RStudio to install it.

Remember that you can run an entire chunk by clicking on the green play arrow in the top right corner of the chunk. You can also run lines of code line-by-line if you place your cursor on some R code and press ⌘ + enter (for macOS users) or ctrl + enter (for Windows users).

Make sure you run each chunk sequentially. If you run a chunk in the middle of the document without running previous ones, it might not work, since previous chunks might do things that later chunks depend on.

Instructions

For this problem set, we’re less interested in causal relationships and more interested in the mechanics of manipulating data and running regressions in R. We’ll start caring about identification and causal models in the next problem set. Because of this, don’t put too much causal weight into the interpretations of these different models—this is an actual case of correlation not implying causation.

1. Download all the needed files and open it in RStudio.
2. Complete the tasks given in the R Markdown file. Fill out code in the empty chunks provided (you can definitely copy, paste, and adapt from other code in the document —*don’t try to write everything from scratch!*), and replace text in ALL CAPS with your own. (i.e. You’ll see a bunch of TYPE YOUR ANSWER HEREs. Type your answers there.). Again, you don’t need to type your answers in all caps.

Turning everything in

When you’re all done, click on the “Knit” button at the top of the editing window and create an HTML or Word version (or PDF if you’ve [installed **tinytex**](https://evalsp21.classes.andrewheiss.com/resource/install/#install-tinytex)) of your document. Upload that file to Blackboard.

