Homework 4

Due date: Friday, October 7

Complete the following exercises and submit your assignment via gradescope (linked on the course webpage).

Problems to start after class Sept 30

Q1

Chapter 2 exercise E.13. In parts (b)-(d), only report the best transformation you found.

In your answers to parts (b)-(d), please include your fitted model equation, a scatterplot with the transformed model superimposed on the **original scale** of the variables, and residual plots justifying that you improved the fit.

Note: Look at the handout from class to help you specify the appropriate formula for your `gf_lm()` layer in the scatterplot.

Q2

For the models you selected in parts (b)-(d) of Q1 go back and provide an interpretation of the slope in the context of the problem. Be sure to read the "Interpretation of transformed models" section at the end of the activity from Friday before doing this.

Q3

Consider the following data on penguin heart rate as a function of duration of dive (in minutes).

```
penguins <- read.csv("http://aloy.rbind.io/data/penguins.csv")
```

(a) Plot heart rate against duration. What problems do you see in fitting the simple linear regression model?

(b) Plot the standardized residuals against fitted values, what problem do you see? Are they they same problems you saw previously?

(c) Fit the simple linear regression model of `log(heart.rate)` against `duration`. Report the fitted regression equation.

(d) Interpret the estimated slope in the context of the problem.

(e) Calculate a 90% confidence interval for the slope and interpret it in the context of the problem.

(f) Does the model seem more appropriate? That is, do the assumptions/conditions for the model appear to be met? Explain.

(g) Create a scatterplot with the transformed model superimposed on the **original scale** of the variables.

Problems to start after class Oct 3

Q4

Researchers have a data set consisting of the isotopic composition structural bone carbonate and the isotopic composition of the coexisting calcite cements in 18 bone samples from a specimen of the dinosaur *Tyrannosaraus rex*.

```
trex <- read.csv("https://aloy.rbind.io/data/trexbones.csv")
```

(a) Create a scatterplot of `Calcite` (y) against `Carbonate` (x) and describe the relationship.

(b) Model 1: Using all 18 data cases, fit the regression of Calcite on Carbonate. Report the estimated slope, the standard error for the slope, the and R^2 value. Add the fitted regression line to the scatterplot.

(c) Do you think any points are influential in model 1? Clearly justify your answer.

(d) Model 2: now fit a model removing the first observation (i.e., the observation with the smallest carbonate value). Again, report the estimated slope, the standard error for the slope, the and R^2 value. Add the fitted regression line to the scatterplot using a different `color` and `linetype` to distinguish it from model 1. (There should now be two lines on your scatterplot.)

> **Note**
>
> To add this regression line, you can pass a different data set into the `gf_lm()` command. For example,
>
> ```
> # fill in the blank with the row number(s) to delete
> gf_point(Calcite ~ Carbonate, data = trex) |>
>   gf_lm() |>
>   gf_lm(Calcite ~ Carbonate, data = dplyr::slice(trex, -c(___)),
>         linetype = 2, color = "orange")
> ```

(e) Do you think any points are influential in model 2? Clearly justify your answer.

(f) Model 3: finally, fit a model removing the first and second observations. Again, report the estimated slope, the standard error for the slope, the and R^2 value. Add the fitted regression line to the scatterplot using a different `color` and `linetype` to distinguish it from models 1 and 2. (There should now be three lines on your scatterplot.)

(g) Do you think any points are influential in model 3? Clearly justify your answer.

(h) Compare the R^2 values, slope estimates, and standard errors for the three models. Why is there such a big difference?

(i) Why might pairs (or groups) of influential observations not be found with the usual case influence statistics?