Homework 6 – Stat 230 – Fall 2022

Due date: Friday, October 21

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

Problems to start after class Oct 14

Q1

Chapter 2 exercise E.16

> Polynomial regression in R
>
> You can include polynomial terms in your regression model by calculating the higher-order terms within your regression equation. For example, if your data set has columns `y` and `x`, then you can fit the model $\mu(y|x) = \beta_0 + \beta_1 x + \beta_2 x^2$ using the following code:
>
> ```
> lm(y ~ x + I(x^2), data)
> ```
>
> The key is to put your polynomial terms inside an `I()`.

Q2

The `milk` data set contains comparative primate milk composition data taken from Table 2 of Hinde and Milligan (2011) *Evolutionary Anthropology* 20:9-23.

(a) Create a scatterplot matrix displaying the kilocalories per gram of milk (`kcal.per.g`), the percent fat (`perc.fat`), and percent lactose (`perc.lactose`). Describe the associations you see.

(b) Fit a multiple linear regression model to predict the kilocalories per gram of milk (`kcal.per.g`) using the percent fat (`perc.fat`) and percent lactose (`perc.lactose`) measurements and report table of coefficients produced by `tidy()` in the {broom} package.

(c) You should have found that only one of the slopes was statistically discernibly different from 0 based on the individual t-tests. Your friend, who hasn't taken Stat 230 is surprised by this, since they were convinced that both the percent fat and the percent lactose would be important predictors based on the scatterplot matrix. Explain to your friend why this isn't surprising.

(d) Caclulate the variance inflation factor for each predictor. Are there any indications of multicollinearity?

Q3

Explain why multicollinearity it not a problem for researchers who are trying to develop a model that accurately predict their response.