**This document contains a bunch of linear modeling questions which can be used as needed.**

**Questions 1-6.**

Recent surveys have shown that Pima Indian women have increased incidences of Diabetes. To identify possible underlying factors of Diabetes in this population, researchers took various measurements of these women. Data from this study can be found in the file "pima.csv." Variables in this dataset include,

* Number of Pregnancies (Num\_Preg), a categorical variable
* Glucose levels (units unknown)
* Blood pressure (BP) in mmHg units
* Skin thickness (mm)
* Body mass index (BMI)
* Age, in years
* Diabetic (yes or no).

1. Conduct a linear model analysis with the response "Glucose" and predictor "Diabetic." In your analysis, be sure to state the type of linear model you are building and why.
2. Conduct a linear model analysis with the response variable "BP" and predictor variable "Num\_Preg." In your analysis, be sure to state the type of linear model you are building and why.
3. Conduct a linear model analysis with the response variable "BMI" and predictor variables "Diabetic" and "Num\_Preg." In your analysis, be sure to state the type of linear model you are building and why.
4. Conduct a linear model analysis with the response variable "Skin\_Thickness" and predictor variables "Num\_Preg" and "Diabetic." In your analysis, be sure to state the type of linear model you are building and why.
5. Conduct a linear model analysis with the response variable "BP" and predictor variable "BMI." In your analysis, be sure to state the type of linear model you are building and why.
6. Conduct a linear model analysis with the response variable "BMI" and predictor variables "Glucose" and "Diabetic." In your analysis, be sure to state the type of linear model you are building and why.

**Questions 7-11.**

In the early 1930’s, evolutionary biology and statistician Sir Ronald Fisher and one of his students collected data from three different species of iris flowers. They measured various physical attributes (all recorded in cm) of roughly 50 flowers from each species of iris in a single pasture on Quebec. All flowers were collected on the same day and measured by the student to reduce bias. Data are listed in the file “iris.csv”.

1. Conduct a linear model analysis with the response variable "Sepal.Width" and predictor variable "Species." In your analysis, be sure to state the type of linear model you are building and why.
2. Conduct a linear model analysis with the response variable "Sepal.Length" and predictor variable "Species." In your analysis, be sure to state the type of linear model you are building and why.
3. Conduct a linear model analysis with the response variable "Sepal.Length" and predictor variable "Sepal.Width" for *Setosa* irises only. In your analysis, be sure to state the type of linear model you are building and why.
4. Conduct a linear model analysis with the response variable "Petal.Length" and predictor variable "Petal.Width" for *Versicolor* irises only. In your analysis, be sure to state the type of linear model you are building and why.
5. Conduct a linear model analysis with the response variable "Sepal.Length" and predictor variables "Sepal.Width" and "Species." In your analysis, be sure to state the type of linear model you are building and why.

**Questions 12-16.**

The data set for the following questions comes from measurements taken on various attributes of different wines grown in a single region in Italy from three different cultivar types. This data set can be found in the file "wine.csv". Variables (assume units unknown) include,

* Cultivar (categorical variable)
* Alcohol content
* Malic Acid content
* Ash content
* Magnesium content
* Total Phenol content
* Flavanoid content
* Non-flavanoid phenol content
* Color (a numeric variable).

1. Conduct a linear model analysis with the response variable "Magnesium " and predictor variable "Cultivar." In your analysis, be sure to state the type of linear model you are building and why.
2. Conduct a linear model analysis with the response variable "Color" and predictor variables "Alcohol" and "Cultivar." In your analysis, be sure to state the type of linear model you are building and why.
3. Conduct a linear model analysis with the response variable "Flavanoids" and predictor variables "Cultivar." In your analysis, be sure to state the type of linear model you are building and why.
4. Multiple regression with phenol response, flavonoids and nonflavphenols predictors
5. Regression with magnesium response and ash predictor (log transformation might be good)