Stat 425 Hw 3

Question 1 - Diagnostics Overview (5 points)

Using the `teengamb` data, fit a model with `gamble` as the response and the other variables as predictors. Perform regression diagnostics on this model to answer the following questions. Display any plots that are relevant to each part.

a. Check the constant variance assumption for the errors. Comment on the appropriate plot and state your conclusion

b. Check the normality assumption for the errors. Briefly comment

c. Check for and identify any large leverage points.

d. Check for and identify any outliers. Use any reasonable cutoff value

e. Check for and identify any influential points.

Question 2 – What is QQNorm? (5 points)

a. (1 pt) First, type `set.seed = 1`. Then, generate a random standard normal sample of length 100. Store this sample in a vector (name it anything you like)

b. (4 pts) Using R, but without using the `qqnorm()` function or any other similar functions, write a simple algorithm to assess the normality of this sample by creating your own qqnorm plot.

Hint: You may use the algorithm described in Faraway 7.8 (Assessing Normality)

Hint: **Some** of the following functions *may* be useful:

```
sort(), order(), if(), while(), for(), rnorm(), qnorm(), pnorm(), dnorm(), plot(), windows()
```

Question 3 – WLS (4 points)

By comparing minions adult height to their parents' adult height, Gru wanted to observe their inheritance pattern. Gru selected one minion of each of the following heights: 15, 16, 17, 18, 19, 20, and 21cm. The minions had many children each. The average heights of their offspring are measured in the following table. The sample sizes are unknown, but the sample standard deviations are also given.

below. In the table, only the mean diameters of the offspring seed are given along with respective standard deviations; sample sizes are unknown.

```
parent <- 15:21
offspring <- c(15.89, 16.16, 16.12, 16.39, 16.36, 17.06, 17.25)
sd = c(1.764, 1.595, 1.655, 2.036, 1.895, 1.937, 1.987)
Gru <- data.frame(parent, offspring, sd)
print(Gru)
```

```
##   parent offspring    sd
## 1     15     15.89 1.764
## 2     16     16.16 1.595
## 3     17     16.12 1.655
## 4     18     16.39 2.036
## 5     19     16.36 1.895
## 6     20     17.06 1.937
## 7     21     17.25 1.987
```

a. (0.5 pt) Draw a scatterplot of X = parent height vs Y = offspring height

b. (0.5 pt) Fit an OLS model

c. (0.5 pt) Fit a WLS model (assuming that the sds are accurate)

d. (1.5 pts) Add both fitted lines to the scatterplot. Create a legend to show which is the OLS line and which is the WLS line.

e. (1 pt) Gru thinks that perfect inheritance would correspond to $\beta_1 = 0.5$ since there are 2 parents. Test the hypothesis that $H_0: \beta_1 = 0.5$ vs $H_1: \beta_1 < 0.5$. State the test statistic, its distribution under H_0, and the *p-value*

Note: pay special attention to the formulation of the alternative hypothesis. Also, we were previously testing if the βs were equal to 0. How is this different?

Question 4 – GLS (3 points)

```
library(faraway)
```

```
## Warning: package 'faraway' was built under R version 3.5.3
```

```
data(strongx)
strongx
```

```
##    momentum energy crossx sd
## 1         4  0.345    367 17
## 2         6  0.287    311  9
## 3         8  0.251    295  9
## 4        10  0.225    268  7
## 5        12  0.207    253  7
## 6        15  0.186    239  6
## 7        20  0.161    220  6
## 8        30  0.132    213  6
## 9        75  0.084    193  5
## 10      150  0.060    192  5
```

In class, we looked at a WLS example of the `strongx` dataset in the `faraway` package. We also said that WLS was a specific form of GLS. We can solve for $\tilde{\beta}$ using the method of GLS as well.

a. (0.5 pt) What are the matrices Σ and Σ^{-1}? (print both matrices). You may find the `diag()` command to be useful.

b. (2.5 pt) Use GLS to solve for $\tilde{\beta}$ using either method shown in class. Show your steps/code.

Question 5 – Golf (3 points)

Load the data file `lpga2009.csv` and review the data documentation file for more details about the variables.

a. (0.5 pts) Fit a SLR model to predict prize with percentile in tournaments. Print and show the diagnostics plot (`plot(lm)`) in a 2 x 2 grid.

b. (0.5 pts) Comment on the normality of the residuals

c. (0.5 pts) Are there any outliers? If so, identify them by observation number.

d. (0.5 pts) Are there any influential points? If so, identify them by observation number.

e. (1 pt) For this model, perform and interpret results of the Breusch-Pagan test of nonconstant error variance using $\alpha = .05$. What is your conclusion about the error variance?