MTH 325, Lab #2, Spring 2023 Name _____

Instructions: Follow along with the tutorial portion of the lab. Replicate the code examples in R on your own, along with the demonstration. Then use those examples as a model to answer the questions/perform the tasks that follow. Copy and paste the results of your code to answer questions where directed. Submit your response file and the code used (both for the tutorial and part two). Your code file and your lab response file should each include your name inside.

Linear Regression Diagnostics
In the last lab, we looked at generating our linear regression models, and now we want make predictions with our model, and analyze the residuals to assess model assumptions.

```
10  fit <- lm(mpg ~ qsec, data = mtcars)
11  summary(fit)
12
13  mtcars$predicted <- predict(fit)   # Save the predicted values
14  mtcars$residuals <- residuals(fit)
```

The predict() function applied to our model generates values on the regression line at each of the input x values. Here, we've saved it to a new column in our dataset for easy plotting. Likewise, the function residuals() calculates the residuals of our model, and here we've also save it to our data set.

To create our residual plot, we use the same plotting functions uses in the last lab, and just select the residuals as the y-value. This will be very useful when we want to plot the residuals against multiple models.

```
16  ggplot(data=mtcars, aes(x=qsec, y=residuals))+geom_point()+labs(title='Residual Plot')
17
```

These don't look too bad, but we can also look at the qq-plot.

```
26  plot(fit, which=2, col=c("red"))
27
```

This doesn't make the errors look very normal. We can look at the boxplot and/or the histogram to look more closely at the outliers.

```
28  ggplot(mtcars,aes(x=residuals))+geom_histogram(bins=8)
29  ggplot(mtcars,aes(x=residuals))+geom_boxplot()
30
```

This data suggests that we might want remove the Toyota Corolla and the Fiat 128.

After removing the outliers, redo the model with the reduced dataset and replot.

```r
mtcars2 <- mtcars[-c(18,20),]
fit <- lm(mpg ~ qsec+0, data = mtcars2)
summary(fit)
mtcars2$predicted <- predict(fit)   # Save the predicted values
mtcars2$residuals <- residuals(fit)
plot(fit, which=2, col=c("red"))
```

There are a number of additional diagnostic plots you can try out.

```r
plot(fit, which=1, col=c("blue"))
plot(fit, which=3, col=c("blue"))
plot(fit, which=4, col=c("blue"))
plot(fit, which=5, col=c("blue"))
plot(fit, which=6, col=c("blue"))
```

Tasks

1. Use the built-in Orange data set. Continue your analysis of the data set. Revisit your model from last lab. Conduct a thorough residual analysis. Include graphs. Identify outliers and assess whether they need to be removed and justify your reasoning. Explain your process and interpret the final model. Include all graphs and output of the model. Construct a prediction interval for age 750 based on your model.

2. Use the built-in trees data set. Continue your analysis of the data set. Revisit your model from last lab. Conduct a thorough residual analysis. Include graphs. Identify outliers and assess whether they need to be removed and justify your reasoning. Explain your process and interpret the final model. Include all graphs and output of the model(s). Construct a prediction interval for Girth is 13.0 and Height is 75.

References:

1. Discovering Statistics Using R. Andy Field, Jeremy Miles, Zoe Field. (2012)
2. http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf
3. https://book.stat420.org/applied_statistics.pdf
4. https://scholarworks.montana.edu/xmlui/bitstream/handle/1/2999/Greenwood_Book_Version3_CC_optimized.pdf?sequence=7&isAllowed=y
5. https://www.rstudio.com/resources/cheatsheets/
6. http://www.sthda.com/english/wiki/r-built-in-data-sets
7. https://rpubs.com/iabrady/residual-analysis
8. http://r-statistics.co/Outlier-Treatment-With-R.html
9. https://statdoe.com/step-by-step-scatterplot-for-one-factor-in-r/