title: "Homework 2"

subtitle: "4375 Machine Learning with Dr. Mazidi"

author: "Tyler Echols"

date: "6/6/2022"

output:

html\_document:

df\_print: paged

pdf\_document: default

---

This homework gives practice in using linear regression in two parts:

\* Part 1 Simple Linear Regression (one predictor)

\* Part 2 Multiple Linear Regression (many predictors)

You will need to install package ISLR at the console, not in your script.

## Step 3 (No code. Write your answers in white space)

\* Write the equation for the model, y = wx + b, filling in the parameters w, b and variable names x, y

\* Is there a strong relationship between horsepower and mpg?

\* Is it a positive or negative correlation?

\* Comment on the RSE, R^2, and F-statistic, and how each indicates the strength of the model

\* Comment on the RMSE and whether it indicates that a good model was created

## Step 4: Examine the model graphically

\* Comment on the predicted value given the graph you created

## Step 5: Evaluate on the test data

\* Test on the test data using the predict function

\* Find the correlation between the predicted values and the mpg values in the test data

\* Print the correlation

\* Calculate the mse on the test results

\* Print the mse

\* Compare this to the mse for the training data

\* Comment on the correlation and the mse in terms of whether the model was able to generalize well to the test data

Your commentary here:

```{r}

# your code here

```

## Step 6: Plot the residuals

\* Plot the linear model in a 2x2 arrangement

\* Do you see evidence of non-linearity from the residuals?

Your commentary here:

```{r}

# your code here

```

## Step 7: Create a second model

\* Create a second linear model with log(mpg) predicted by horsepower

\* Run summary() on this second model

\* Compare the summary statistic R^2 of the two models

Your commentary here:

```{r}

# your code here

```

## Step 8: Evaluate the second model graphically

\* Plot log(train\$mpg)~train\$horsepower

\* Draw a blue abline()

\* Comment on how well the line fits the data compared to model 1 above

Your commentary here:

```{r}

# your code here

```

## Step 9: Predict and evaluate on the second model

\* Predict on the test data using lm2

\* Find the correlation of the predictions and log() of test mpg, remembering to compare pred with log(test$mpg)

\* Output this correlation

\* Compare this correlation with the correlation you got for model

\* Calculate and output the MSE for the test data on lm2, and compare to model 1. Hint: Compute the residuals and mse like this:

```

residuals <- pred - log(test$mpg)

mse <- mean(residuals^2)

```

Your commentary here:

```{r}

# your code here

```

## Step 10: Plot the residuals of the second model

\* Plot the second linear model in a 2x2 arrangement

\* How does it compare to the first set of graphs?

Your commentary here:

```{r}

# your code here

```

# Problem 2: Multiple Linear Regression

## Step 1: Data exploration

\* Produce a scatterplot matrix of correlations which includes all the variables in the data set using the command “pairs(Auto)”

\* List any possible correlations that you observe, listing positive and negative correlations separately, with at least 3 in each category.

Your commentary here:

```{r}

# your code here

```

## Step 2: Data visualization

\* Display the matrix of correlations between the variables using function cor(), excluding the “name” variable since is it qualitative

\* Write the two strongest positive correlations and their values below. Write the two strongest negative correlations and their values as well.

Your commentary here:

```{r}

# your code here

```

## Step 3: Build a third linear model

\* Convert the origin variable to a factor

\* Use the lm() function to perform multiple linear regression with mpg as the response and all other variables except name as predictors

\* Use the summary() function to print the results

\* Which predictors appear to have a statistically significant relationship to the response?

Your commentary here:

```{r}

# your code here

```

## Step 4: Plot the residuals of the third model

\* Use the plot() function to produce diagnostic plots of the linear regression fit

\* Comment on any problems you see with the fit

\* Are there any leverage points?

\* Display a row from the data set that seems to be a leverage point.

Your commentary here:

```{r}

# your code here

```

## Step 5: Create and evaluate a fourth model

\* Use the \* and + symbols to fit linear regression models with interaction effects, choosing whatever variables you think might get better results than your model in step 3 above

\* Compare the summaries of the two models, particularly R^2

\* Run anova() on the two models to see if your second model outperformed the previous one, and comment below on the results

Your commentary here:

```{r}

# your code here

```