Homework 1

IN YOUR CODE, PLEASE USE THE SAME VARIABLE NAMES MENTIONED IN THIS DOCUMENT OR THE AUTO GRADER WILL NOT GRANT YOU THE GRADE. THE VARIABLE NAMES ARE HIGHLIGHTED

For your first homework, you need to use the Auto dataset (loaded for you in the provided script) to do a simple linear regression where you predict mpg from horsepower. You also need to do a multiple linear regression where you predict mpg from horsepower, weight, cylinders, and displacement.

Simple Linear Regression

Create a different data frame for the simple linear regression.
Call it hpdf

You should not use the lm() function to calculate beta_0 and beta_1.
Calculate hp_beta_0 and hp_beta_1 manually.

Once, you are done. You can use the lm() function to verify your answers.
Store the results of the lm() function in a variable called hp.fit

```
hp.fit <- lm(#You code here)
```

Create a summary of the hp.fit model and store the results into a variable called hp.fit.s

```
hp.fit.s <- summary(hp.fit)
```

Verifying the model statistically

From the hp.fit.s variable, extract the t-values of the intercept and slope
I will help you with this one
You can get the t-values through the coefficients matrix

```
hp.fit.s$coefficients
              Estimate  Std. Error   t value      Pr(>|t|)
```

```
(Intercept) 39.9358610 0.717498656  55.65984 1.220362e-187
horsepower  -0.1578447 0.006445501 -24.48914  7.031989e-81
```

If you want the t value for the intercept, do the following

```
> hp.fit.s$coefficients["(Intercept)", "t value"]
[1] 55.65984
```

Store this value into a variable called hp.fit.t.b0

```
hp.fit.t.b0 <- hp.fit.s$coefficients["(Intercept)", "t value"]
```

Do the same for beta_1 and store the value into a variable called hp.fit.t.b1

Now, we want to verify these values are statistically significant. Let us do a Boolean test with both variables. If they fall outside the range of -2 and 2, then they are statistically significant.
Notice that -2 and 2 might be the optimal range and you need to verify this range from the t-distribution table, but for the sake of this assignment, this range have to work.

Store the Boolean test results in the following variable for both betas
hp.fit.t.b0_test
hp.fit.t.b1_test

Multiple Linear Regression

Again, you need to calculate this manually. Follow Week 2 Lab code and do the same.
The following are the names of the variables you need to use

lm.matrix : this is the matrix that will include the predictors. Don't forget to add a column of 1s to it.
betas : this is a vector that will contain the results of the matrix operation.

Now that you calculated the coefficients, use the lm() function to verify your answers.
store your model into a variable called lm.fit

Statistical test

You will also need to verify the t-values of each predictor
Follow the same steps above to extract the t-values from your model

You need the following variables
lm.fit.s : this is the summary of your model
lm.fit.t.b0, lm.fit.t.b1, lm.fit.t.b2, lm.fit.t.b3, lm.fit.t.b4 : variables to store the t-values of each coefficient in.
lm.fit.t.b0_test, lm.fit.t.b1_test, lm.fit.t.b2_test, lm.fit.t.b3_test, lm.fit.t.b4_test : variable to store the Boolean test results in.

Question

Should we remove some of the predicators from the model? Why?

Table of variables and Rubric

Task	Variable Names	Grade
Simple Linear Regression Variables		
Create hpdf data frame	hpdf	1
Calculate hp_beta_0 manually	hp_beta_0	1
Calculate hp_beta_1 manually	hp_beta_1	1
Store lm() results	hp.fit	1
Create summary of hp.fit	hp.fit.s	1
Extract and store t-value of intercept	hp.fit.t.b0	1
Extract and store t-value of slope	hp.fit.t.b1	1
Boolean test for hp.fit.t.b0	hp.fit.t.b0_test	1
Boolean test for hp.fit.t.b1	hp.fit.t.b1_test	1
Multiple Linear Regression Variables		
Calculate lm.matrix	lm.matrix	1
Calculate betas	betas	1
Store lm() results	lm.fit	1
Create summary of lm.fit	lm.fit.s	1
Extract t-values for coefficients in lm.fit	lm.fit.t.b0 lm.fit.t.b1 lm.fit.t.b2 lm.fit.t.b3 lm.fit.t.b4	1 1 1 1 1
Boolean tests for lm.fit t-values	lm.fit.t.b0_test lm.fit.t.b1_test	1 1

Task	Variable Names	Grade
	lm.fit.t.b2_test lm.fit.t.b3_test lm.fit.t.b4_test	1 1 1
Question to answer		2
	Total Points	25