Matthew L. Pergolski

IST 772

Dr. Block

11/25/2021

**Homework 8**

**Beginning Statement**

“I produced the material below with no assistance [direct quote from IST 772 class syllabus].” Note: Homework questions from the book may have been copied/pasted into the document for both the student and viewer’s convenience.

The homework for week 8 is exercises 1-8 on pages 181-182.

**Homework Question 1**

**Question:**

1. The data sets package in R contains a small data set called mtcars that contains n = 32

observations of the characteristics of different automobiles. Create a new data frame

from part of this data set using this command: myCars <- data.frame(mtcars[,1:6]).

**Answer/Student Response:**

The following was generated and observed:

Graphical user interface

Description automatically generated

**Homework Question 2**

**Question:**

2. Create and interpret a bivariate correlation matrix using cor(myCars) keeping in mind

the idea that you will be trying to predict the mpg variable. Which other variable might

be the single best predictor of mpg?

**Answer/Student Response:**

The following was generated and observed:

Text

Description automatically generated

Other than the already mentioned variable of 'wt' (at -0.8676594), we also see a strong correlation with 'cyl', coming in at a value of -0.8521620. Another one to mention is 'disp' with a value of -0.8475514.

**Homework Question 3**

**Question:**

3. Run a multiple regression analysis on the myCars data with Im(), using mpg as the

dependent variable and wt (weight) and hp (horsepower) as the predictors. Make

sure to say whether or not the overall R-squared was significant. If it was significant,

report the value and say in your own words whether it seems like a strong result or

not. Review the significance tests on the coefficients (B-weights). For each one that

was significant, report its value and say in your own words whether it seems like a

strong result or not.

**Answer/Student Response:**

The following as generated and observed:

Graphical user interface, text

Description automatically generated with medium confidence

Graphical user interface

Description automatically generated

Text

Description automatically generated

The associated overall p-value with the summary function of the linear model comes out to be p-value: 9.109e-12, which suggests the results are significant. The R-squared value is 0.8268 and adjusted R-squared value is 0.8148, which suggests there is a sizable/strong relationship between the variables. The 'wt' variable results are significant at 1.12e-06 \*\*\* while the hp variable is also significant at 0.00145 \*\*. Both of these variables suggest significance due to their p-values being below the 0.05 alpha standard (and both show a strong relationship to 'mpg', with 'wt' showing a stronger relationship than 'hp'.

**Homework Question 4**

**Question:**

4. Using the results of the analysis from Exercise 2, construct a prediction equation for

mpg using all three of the coefficients from the analysis (the intercept along with the

two B-weights). Pretend that an automobile designer has asked you to predict the

mpg for a car with 110 horsepower and a weight of 3 tons. Show your calculation and

the resulting value of mpg.

**Answer/Student Response:**

The following was generated and observed:

Graphical user interface, text, application

Description automatically generated

A picture containing graphical user interface

Description automatically generated

Text

Description automatically generated with medium confidence

To summarize the calculations above, the updated mpg value after the changing variable values equates to ~22.1.

**Homework Question 5**

**Question:**

5. Run a multiple regression analysis on the myCars data with ImBF(), using mpg as the

dependent variable and wt (weight) and hp (horsepower) as the predictors. Interpret

the resulting Bayes factor in terms of the odds in favor of the alternative hypothesis. If

you did Exercise 2, do these results strengthen or weaken your conclusions?

**Answer/Student Response:**

The following was generated and observed:

Graphical user interface, application

Description automatically generated

**Graphical user interface, text, application

Description automatically generated**

The Bayes Factor with the lmBF command indicates very strong evidence for the alternative hypothesis, coming in at 788547604 to 1. This strengthens the findings presented by the 'frequentist' perspective in Exercise 2.

**Homework Question 6**

**Question:**

6. Run ImBF() with the same model as for Exercise 4, but with the options posterior=TRUE

and iterations=10000. Interpret the resulting information about the coefficients.

**Answer/Student Response:**

The following was generated and observed:

Graphical user interface, application

Description automatically generated

Text

Description automatically generated

In looking at 'wt' and 'hp', we do not observe any values that cross 0 within the 95% HDI; this adds to our confidence in relation to the alternative hypothesis.

**Homework Question 7**

**Question:**

7. Run install.packages() and library() for the "car" package. The car package is "com-

panion to applied regression" rather than more data about automobiles. Read the

help file for the vif() procedure and then look up more information online about how to

interpret the results. Then write down in your own words a "rule of thumb" for interpret-

ing vif.

**Answer/Student Response:**

The following was generated and observed:

Text

Description automatically generated

This function alculates variance-inflation and generalized variance-inflation factors (VIFs and GVIFs) for linear, generalized linear, and other regression models. It measures the extent of multicollinearity (generally speaking, the lower this value, the better).

**Homework Question 8**

**Question:**

8. Run vif() on the results of the model from Exercise 2. Interpret the results. Then run a

model that predicts mpg from all five of the predictors in myCars. Run vif() on those

results and interpret what you find.

**Answer/Student Response:**

The following was generated and observed:

Text

Description automatically generated

Text

Description automatically generated

Table

Description automatically generated with low confidence

Graphical user interface, text

Description automatically generated

For multicollinearity, the lower the value, (generally) the more ideal our lm function will be; the vif(myCars.lm) line of code indicates a rather low value of ~1.7 for each variable; however, when all variables are taken into account, we see higher values, with 'cyl' (~7.8) and 'disp' (~10.4) clocking in at the highest. If these variables were to be removed from the model, we would likley see 'better' results.