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:

```
myCars <- data.frame(mtcars[,1:6])
myCars
```

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:

```
cor(myCars)

# Comments
  # 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.
```

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:

```
myCars.lm <- lm(mpg ~ wt + hp, data = myCars)
myCars.lm
summary(myCars.lm)
```

```
>   myCars.lm <- lm(mpg ~ wt + hp, data = myCars)
>   myCars.lm

Call:
lm(formula = mpg ~ wt + hp, data = myCars)

Coefficients:
(Intercept)           wt           hp
   37.22727     -3.87783     -0.03177
```

```
>   summary(myCars.lm)

Call:
lm(formula = mpg ~ wt + hp, data = myCars)

Residuals:
   Min     1Q Median     3Q    Max
-3.941 -1.600 -0.182  1.050  5.854

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 37.22727    1.59879  23.285  < 2e-16 ***
wt          -3.87783    0.63273  -6.129 1.12e-06 ***
hp          -0.03177    0.00903  -3.519  0.00145 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.593 on 29 degrees of freedom
Multiple R-squared:  0.8268,    Adjusted R-squared:  0.8148
F-statistic: 69.21 on 2 and 29 DF,  p-value: 9.109e-12

>
```

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:

```
# coefficients for variables
lm.c <- coefficients(myCars.lm)
lm.c

# update metrics
hp <- 110
weight <- 3

# prediction equation with updated variable values
prediction.equation <- lm.c[1] + (lm.c[2] * weight) + (lm.c[3] * hp)
prediction.equation
```

```
>   lm.c
(Intercept)          wt          hp
37.22727012 -3.87783074 -0.03177295
>
```

```
>   prediction.equation
(Intercept)
   22.09875
```

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:

```r
# Bayesian method
lmBF(mpg ~ wt + hp, data = myCars)

# Summary command
summary(lmBF(mpg ~ wt + hp, data = myCars))

# STR command
str(lmBF(mpg ~ wt + hp, data = myCars))
```

```
>   lmBF(mpg ~ wt + hp, data = myCars)
Bayes factor analysis
--------------
[1] wt + hp : 788547604 ±0%

Against denominator:
  Intercept only
---
Bayes factor type: BFlinearModel, JZS
```

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:

```r
# Bayesian method
lmBF(mpg ~ wt + hp, data = myCars, posterior = TRUE, iterations = 10000)

# STR command
str(lmBF(mpg ~ wt + hp, data = myCars, posterior = TRUE, iterations = 10000))

# Summary command
summary(lmBF(mpg ~ wt + hp, data = myCars, posterior = TRUE, iterations = 10000))
```

```
>   summary(lmBF(mpg ~ wt + hp, data = myCars, posterior = TRUE, iterations = 10000))
0
|----|----|----|----|----|----|----|----|----|----|
**************************************************|

Iterations = 1:10000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

       Mean        SD  Naive SE Time-series SE
mu   20.094  0.484601 4.846e-03      4.945e-03
wt   -3.785  0.673210 6.732e-03      6.948e-03
hp   -0.031  0.009498 9.498e-05      9.678e-05
sig2  7.451  2.290613 2.291e-02      2.778e-02
g     4.083 18.862535 1.886e-01      1.886e-01

2. Quantiles for each variable:

        2.5%      25%      50%      75%    97.5%
mu   19.1512 19.77686 20.09371 20.41232 21.05850
wt   -5.0941 -4.22151 -3.79270 -3.34836 -2.47748
hp   -0.0494 -0.03738 -0.03116 -0.02456 -0.01212
sig2  4.3603  5.94446  7.07209  8.54068 12.67478
g     0.3677  0.94608  1.68289  3.39866 19.25631

>
```

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 "companion 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 interpreting vif.

Answer/Student Response:
The following was generated and observed:

```
#install.packages("car")
library("car")

?vif
```

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:

```
# vif command for lm displayed in problem # 2
vif(myCars.lm)

# vif command for new lm containing all variables
myCars.lm.all <- lm(mpg ~ ., data = myCars)
myCars.lm.all
summary(myCars.lm.all)

vif(myCars.lm.all)
```

```
>   vif(myCars.lm)
      wt       hp
1.766625 1.766625
```

```
>   summary(myCars.lm.all)

Call:
lm(formula = mpg ~ ., data = myCars)

Residuals:
    Min      1Q  Median      3Q     Max
-3.7014 -1.6850 -0.4226  1.1681  5.7263

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 36.00836    7.57144   4.756  6.4e-05 ***
cyl         -1.10749    0.71588  -1.547  0.13394
disp         0.01236    0.01190   1.039  0.30845
hp          -0.02402    0.01328  -1.809  0.08208 .
drat         0.95221    1.39085   0.685  0.49964
wt          -3.67329    1.05900  -3.469  0.00184 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

Residual standard error: 2.538 on 26 degrees of freedom
Multiple R-squared:  0.8513,    Adjusted R-squared:  0.8
F-statistic: 29.77 on 5 and 26 DF,  p-value: 5.618e-10

>
```

```
>   vif(myCars.lm.all)
      cyl      disp        hp      drat        wt
 7.869010 10.463957  3.990380  2.662298  5.168795
```

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.