HW8_Tully

IST 772 HW8

11/26/21

Assignment:

The homework for this week are exercises:

- 1-7 (Stanton,p.181)
- 8 (Stanton,p.182)

Exercise 1 (Stanton,p.181):

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]).

```
# View the column names of the original data set
colnames(mtcars)
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

```
# Truncateing the data set
myCars <- data.frame(mtcars[,1:6])
# View the column names of the created data set
colnames(myCars)
```

```
## [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"
```

> We truncated the data set and created a new data set named myCars by removing the following columns: "qsec", "vs", "am", "gear", and "carb".

Exercise 2 (Stanton,p.181):

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?

```
# creating a bivariate correlation matrix
cor(myCars)
```

```
##             mpg        cyl       disp         hp       drat         wt
## mpg   1.0000000 -0.8521620 -0.8475514 -0.7761684  0.6811719 -0.8676594
## cyl  -0.8521620  1.0000000  0.9020329  0.8324475 -0.6999381  0.7824958
## disp -0.8475514  0.9020329  1.0000000  0.7909486 -0.7102139  0.8879799
## hp   -0.7761684  0.8324475  0.7909486  1.0000000 -0.4487591  0.6587479
## drat  0.6811719 -0.6999381 -0.7102139 -0.4487591  1.0000000 -0.7124406
## wt   -0.8676594  0.7824958  0.8879799  0.6587479 -0.7124406  1.0000000
```

The other variable that might be the single best predictor of mpg is wt because it has a high negative correlation at -0.8677.

Exercise 3 (Stanton,p.181):

Run a multiple regression analysis on the myCars data with lm(), 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.

```r
# multiple regression analysis
myCarslm <- lm(mpg ~ wt + hp, data =myCars)
summary(myCarslm)
```

```
##
## 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
```

In this multiple regression analysis the R-squared at 0.8268 is significant. The F(2,29)=69.21, p < .001, meaning that we reject the null hypothesis. The wt variable has a very strong negative relationship on mpg with an estimate of -3.9 and it is significant with a Pr(>|t|) value of 1.12e-06 (<.001). hp variable also has a strong negative relationship on mpg with an estimate of -0.03 and it is significant with a Pr(>|t|) value of 0.001.

Exercise 4 (Stanton,p.181):

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.

```
# prediction equation for mpg
37.22727 + (3 * -3.87783) + (100 * -0.03177)
```

```
## [1] 22.41678
```

> This is the prediction equation for mpg for a car with 110 horsepower and a weight of 3 tons. mpg = intercept + (x1 * wt) + (x2 * hp); mpg = 37.22727 + (3 * -3.87783) + (100 * -0.03177); mpg = 22.41678

Exercise 5 (Stanton,p.181):

Run a multiple regression analysis on the myCars data with lmBF(), 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?

```
#
mpgOutMCMC <- lmBF(mpg ~ wt + hp,  data=myCars)
summary(mpgOutMCMC)
```

```
## Bayes factor analysis
## --------------
## [1] wt + hp : 788547604 ±0%
##
## Against denominator:
##   Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS
```

```
library(effectsize)
bf = 788547604
interpret_bf(
  bf,
  rules = "raftery1995",
  log = FALSE,
  include_value = FALSE,
  protect_ratio = TRUE,
  exact = TRUE
)
```

```
## [1] "very strong evidence in favour of"
## (Rules: raftery1995)
```

> This multiple regression analysis resulted in a very strong evidence in favor or the desired hypothesis. We would be rejecting the null hypothesis. In comparison to the earlier excercise this result would strengthen our conclusion.

Exercise 6 (Stanton,p.181):

Run lmBF() with the same model as for Exercise 4, but with the options posterior=TRUE and iterations=10000. Interpret the resulting information about the coefficients.

```
#
mpgOutMCMC <- lmBF(mpg ~ wt + hp,  data=myCars, posterior=TRUE, iterations=10000)
summary(mpgOutMCMC)
```

```
##
## 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   20.08735  0.484985 4.850e-03      4.850e-03
## wt   -3.80347  0.663505 6.635e-03      6.757e-03
## hp   -0.03094  0.009474 9.474e-05      9.474e-05
## sig2  7.45640  2.134577 2.135e-02      2.579e-02
## g     3.94594 10.473100 1.047e-01      1.047e-01
##
## 2. Quantiles for each variable:
##
##         2.5%      25%      50%      75%    97.5%
## mu   19.1269 19.76373 20.08416 20.40308 21.06338
## wt   -5.0916 -4.23916 -3.81312 -3.36245 -2.49148
## hp   -0.0498 -0.03713 -0.03101 -0.02473 -0.01238
## sig2  4.3787  5.95084  7.10846  8.53284 12.68250
## g     0.3656  0.94731  1.74149  3.58458 19.84098
```

> While Interpreting the resulting information of the coefficients wt and hp we can see they both have a negative relationship to mpg. Since 95% HDI do not pass through zero it demonstrates that there is significance in rejecting the null hypothesis.

Exercise 7 (Stanton,p.181):

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.

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

```
## Loading required package: carData
```

```r
# interpret_vif
interpret_vif(c(3,7,15))
```

```
## [1] "low"      "moderate" "high"
## (Rules: default)
```

> The rule of thumb for the diagnostic for multicollinearity and interpreting the VIF is that if it is under five then it is low. If it is five but less than 10 then it is moderate. If it is ten or higher then it is high multicollinearity.

Exercise 8 (Stanton,p.182):

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.

```r
#
library(effectsize)
# run a  model that predicts mpg from hp and wt of the predictors in myCars.
mpg_vif <- vif(lm(mpg ~ hp + wt, data = myCars))
mpg_vif
```

```
##       hp       wt
## 1.766625 1.766625
```

```r
# interpret_vif
interpret_vif(mpg_vif)
```

```
##    hp    wt
## "low" "low"
## (Rules: default)
```

```r
# run a  model that predicts mpg from all five of the predictors in myCars.
mpg_vif <- vif(lm(mpg ~ cyl + disp + hp + drat + wt, data = myCars))
mpg_vif
```

```
##       cyl      disp        hp      drat        wt
##  7.869010 10.463957  3.990380  2.662298  5.168795
```

```r
# interpret_vif
interpret_vif(mpg_vif)
```

```
##        cyl       disp         hp       drat         wt
## "moderate"     "high"      "low"      "low" "moderate"
## (Rules: default)
```

> The first test resulted in low multicollinearity with hp and wt. The second test with all 5 variables shows low multicollinearity with hp and drat. It has a moderate multicollinearity in cyl and wt. It has a high multicollinearity in disp.

Reference(s):

Stanton, Jeffrey M.. Reasoning with Data. Guilford Publications. Kindle Edition.