Homework 8
Syracuse University
IST 772
Summer 2021

```
# load packages
library(BayesFactor)
library(BEST)
library(car)
```

Question 1

```
# create a subset dataframe from mtcars
myCars <- data.frame(mtcars[, 1:6])
```

Question 2

```
# create and interpret a bivariate correlation matrix
cor(myCars, method = "pearson")
```

```
##             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
```

```
# cyl, disp, hp, and wt all have negative correlations with mpg. There
# might be multicollinearity between these variables because they are
# also correlated with eachother. The drat variable, on the other hand,
# has a positive correlation with mpg.
```

Question 3

```
# run a multiple regression on myCars
carsOut <- lm(formula = mpg ~ wt + hp,
              data = mtcars)

# summarize the model results
summary(carsOut)
```

```
##
## Call:
## lm(formula = mpg ~ wt + hp, data = mtcars)
##
## 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
```

```
# yes the overall R-squared is significant because the p-value is much less
than
# any of the standard alphas, 0.05, 0.01, 0.001.

# the value of R-squared is 0.8268. This is a strong result, indicating that
# 82.68% of the variance in miles per gallon is explained by the weight and
# the horse power of the car.

# coefficient for weight = -3.87783. Yes this is significant because the
# p value is < .001. This is a strong result because it is significant in
# explaining variance in the miles per gallon.

# coefficient for horsepower = -0.03177. Yes this is significant because the
# p value is < 0.01. This is a strong result becuase it is significant in
# explaining variance in the miles per gallon.
```

Question 4

```
# construct a prediction equation for mpg using the result from exercise 3
# mpg = 37.22727 - 3.8778 * weight - 0.03177 * horsepower
mpg <- function(x1, x2) {37.22727 + (-3.8778 * x1) + (-0.03177 * x2)}

# predict the mpg for a car with 110 weight and 3 horse power
mpg(x1 = 3, x2 = 110)
```

```
## [1] 22.09917
```

Question 5

```
# run a multiple regression analysis on mycars with lmBF()
# carsOutBayes <- lmBF(formula = mpg ~ wt + hp,
#                      data = mtcars)

# carsOutBayes
```

Question 6

```
# run a multiple regression analysis on mycars with lmBF()
# carsOutBayes <- lmBF(formula = mpg ~ wt + hp,
#                      data = mtcars,
#                      posterior = TRUE,
#                      iterations = 10000)

# carsOutBayes
```

Question 7

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

# read help on vif
help(vif)

## starting httpd help server ... done

# the vif is a way to look at how much one of the variables is contributing
# to the error in a model. Using this could help narrow down a set of
features
# that are optimal for the model. if the vif is very high for a particular
# variable it may indicate that there is multicollinearity and so either
# that variable will need to be removed altogether or combined with the
variable
# that it is related with to make one combined variable.
```

Question 8

```
# run vif on the results of the model from exercise 2
vif(carsOut)

##       wt       hp
## 1.766625 1.766625

# run a vif on the results of a model that uses all predictor variables
vif(lm(formula = mpg ~ .,
       data = mtcars))

##       cyl      disp        hp      drat        wt      qsec        vs
am
## 15.373833 21.620241  9.832037  3.374620 15.164887  7.527958  4.965873
4.648487
##      gear      carb
##  5.357452  7.908747

# a number of variables have a vif of greater than 5 which means that they
# are causing variance inflation. The variables that have a vif of greater
# than 5 should either be removed or combined with another variable.
```