Victoria Haley

Homework 8

1. R code: myCars <- data.frame(mtcars[,1:6])
   Output:

	mpg	cyl	disp	hp	drat	wt
Mazda RX4	21.0	6	160.0	110	3.90	2.620
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875
Datsun 710	22.8	4	108.0	93	3.85	2.320
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440
Valiant	18.1	6	225.0	105	2.76	3.460
Duster 360	14.3	8	360.0	245	3.21	3.570
Merc 240D	24.4	4	146.7	62	3.69	3.190
Merc 230	22.8	4	140.8	95	3.92	3.150
Merc 280	19.2	6	167.6	123	3.92	3.440
Merc 280C	17.8	6	167.6	123	3.92	3.440
Merc 450SE	16.4	8	275.8	180	3.07	4.070
Merc 450SL	17.3	8	275.8	180	3.07	3.730
Merc 450SLC	15.2	8	275.8	180	3.07	3.780

Showing 1 to 15 of 32 entries, 6 total columns

2. Output:

```
> 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
>
```

‘wt’ might be the single best predictor of mpg as it has the strongest correlation (-0.86) of the variables.

3. Output:

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

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 overall R-squared (0.83) was significant. With an F-statistic of $F(2,29)$ $p<0.05$, this seems like a strong result.

Both B-weights were significant and seem like strong results, with ‘wt’ being 1.12e-06, and ‘hp’ being 0.00145.

From these results, we can say that as weight increases, fuel efficiency decreases by about 4 MPG and as horsepower increases, fuel efficiency decreases by about 0.03 MPG.

4. Equation: MPG = Intercept + (-3.878 * wt) + (-0.032 * hp)
   Prediction for a car with 110 horsepower and weight of 3 tons:
   37.227 + (-3.878 * 3) + (-0.032 * 110) = *22.073 MPG*

5. Output:

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

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

The resulting Bayes factor is in support of the alternative hypothesis that weight and horsepower do influence fuel efficiency. These results strengthen the results from exercise 2 above.

6. R code:

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

output:

```
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   20.08554 4.905e-01 4.905e-03      0.0048162
wt   -3.78086 6.644e-01 6.644e-03      0.0069027
hp   -0.03094 9.454e-03 9.454e-05      0.0000973
sig2  7.49016 2.255e+00 2.255e-02      0.0274834
g     5.42913 1.019e+02 1.019e+00      1.0190421

2. Quantiles for each variable:

         2.5%      25%      50%      75%    97.5%
mu   19.11353 19.76179 20.08851 20.40985 21.03895
wt   -5.06250 -4.21744 -3.78441 -3.33940 -2.45684
hp   -0.04946 -0.03727 -0.03092 -0.02458 -0.01259
sig2  4.36551  5.98015  7.10697  8.55589 12.77966
g     0.35251  0.92429  1.70816  3.41497 20.58301
```

The B-weights of both predictors are very similar to what was produced by the lm() function above in exercise 2. ‘wt’ has a mean of -3.78, with the highest density intervals ranging from -5.06 to -2.46. ‘hp’ has a mean of -0.031, with the highest density intervals ranging from -0.05 to -0.01.

7. ? vif(): Calculates variance-inflation and generalized variance-inflation factors (VIFs and GVIFs) for linear, generalized linear, and other regression models.
   my rule of thumb: as VIF goes up, so does the correlation.

8. Output for model above:

```
> vif(regOut)
      wt       hp 
1.766625 1.766625 
```

   Interpretation: Since the vif is close to 1, there is not much multicollinearity between ‘wt’ and ‘hp’.

   Output for all 5 variables:

```
> regOutAll <- lm(mpg ~ cyl + disp + hp + drat + wt, data=myCars)
> vif(regOutAll)
      cyl      disp        hp      drat        wt 
 7.869010 10.463957  3.990380  2.662298  5.168795 
> 
```

   Interpretation:
   Seeing that the vif of all variables is greater than 1, with a range of 2.7 (drat) to 10.5 (disp), there is multicollinearity between all variables. This is to be expected though as each predictor will have an influence on the criterion, and each predictor may have correlation with other predictors.