Stat 021 Homework 6

Suzanne Thornton

Due: Friday, Nov. 8, 12:00pm

Instructions: A **pdf** version of your homework must be submitted to Gradescope by **noon** on the due date. The course passcode is **MPKJ4Z**. If you are having trouble getting your *.Rmd* file to compile, you need to get help with this **before** the due date.

You are allowed to hand in **only one** late homework assignment throughout the semester. If you need to hand in this particular assignment late, you must let me know via email by noon on the due date.

You are encouraged to study with your peers to help complete the homework assignments but no copying is allowed. If I see that two or more homework assignments are copied, all students involved will receive a grade of 0 on that assignment and will forfeit (perhaps retroactively) the opportunity to hand in a late homework.

Q0) To help you with some programming tools you will need on your final project, please complete the R *swirl* tutorial on writing functions. You can access this tutorial by typing the following commands in the R console window:

```
install.packages("swirl")
library("swirl")
swirl()
```

Then, the tutuorial will ask what to call you so enter your name and next type

```
main()
```

Make the following sequence of selections:

- 1: R Programming: The basics of programming in R
- 1: R Programming
- 9: Functions

Please complete this tutorial up until you get to the part about binary operators (this occurs at about 94% of the way through). Although you don't get points for this problem, it will dramatically help you with your final project and the material in this tutorial is fair game for future homework assignments.

Q1) Read the data uploaded to Moodle called "mileage.csv". This data describles the gasoline mileage performance for 32 automobiles. Use this data to answer the following questions. Before fitting any models make sure the data is being correctly read into R.

a) Build a linear regression model relating gasoline mileage, y to engine displacement x_1 and the type of transmission, x_2. (Note that transmission type is a binary categorical variable.) Does the type of transmission significantly affect the mileage performance? Justify your answer. (4 points)

b) Modify the model developed in part a to include an interaction between engine displacement and the type of transmission. What is the average effect on gasoline mileage when the engine is automatic? What is the average effect on gasoline mileage when the engine is manual? (4 points)

c) Build a linear regression model relating gasoline mileage, y, to vehicle weight x_3 and the type of transmission x_2 . Does the type of transmission significantly affect the mileage performance? Justify your answer. (4 points)

d) Modify the model developed in part a to include an interaction between vehicle weight and the type of transmission. What is the average effect on gasoline mileage when the transmission is automatic? What is the average effect on gasoline mileage when the transmission is manual? (4 points)

e) Based off of the results for parts (a)-(d), what terms do you think should be included in the final regression model and why? (4 points)

‘‘‘

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
gasperform <- read_csv("mileage.csv", skip = 2, col_names = FALSE)
```

```
## Parsed with column specification:
## cols(
##   X1 = col_character(),
##   X2 = col_double(),
##   X3 = col_double(),
##   X4 = col_double(),
##   X5 = col_character()
## )
```

```
gasperform
```

```
## # A tibble: 32 x 5
##    X1            X2    X3    X4 X5
##    <chr>      <dbl> <dbl> <dbl> <chr>
##  1 Apollo      18.9 350    3910 A
##  2 Omega       17   350    2860 A
##  3 Nova        20   250    3510 A
##  4 Monarch     18.2 351    3890 A
##  5 Duster      20.1 225    3365 M
##  6 JensonConv  11.2 440    4215 A
##  7 Skyhawk     22.1 231    3020 A
##  8 Monza       21.5 262    3180 A
##  9 Scirocco    34.7  89.7  1905 M
## 10 CorollaSR5  30.4  96.9  2320 M
## # ... with 22 more rows
```

```
colnames(gasperform) <- c("Car", "mpg", "displacement", "weight", "transmission_type" )
head(gasperform)
```

```
## # A tibble: 6 x 5
##   Car          mpg displacement weight transmission_type
##   <chr>      <dbl>        <dbl>  <dbl> <chr>
## 1 Apollo      18.9          350   3910 A
## 2 Omega       17            350   2860 A
## 3 Nova        20            250   3510 A
```

```
## 4 Monarch     18.2          351   3890 A
## 5 Duster      20.1          225   3365 M
## 6 JensonConv  11.2          440   4215 A
```

```
gasperform_standard <- gasperform %>% mutate_at(vars("mpg", "displacement", "weight"), funs(scale))
```

```
## Warning: funs() is soft deprecated as of dplyr 0.8.0
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once per session.
```

```
gasperform2 <- gasperform_standard %>% mutate(transmission_type_cat = transmission_type) %>% select(mpg
ANOVA_gasperform2 <- lm(mpg ~ transmission_type_cat + displacement, data = gasperform2)
summary(ANOVA_gasperform2)
```

```
##
## Call:
## lm(formula = mpg ~ transmission_type_cat + displacement, data = gasperform2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.09448 -0.29564  0.02061  0.28342  1.07349
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -0.02303    0.13204  -0.174    0.863
## transmission_type_catM   0.08187    0.35256   0.232    0.818
## displacement            -0.84738    0.16105  -5.262 1.23e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4933 on 29 degrees of freedom
## Multiple R-squared:  0.7724, Adjusted R-squared:  0.7567
## F-statistic:  49.2 on 2 and 29 DF,  p-value: 4.779e-10
```

```
#The type of transmission does not severely affect the mileage performance.
#There is a 0.08187 mpg difference in transmissions types A and M.
#This difference is far less than one SE from the mean.

gasperform2_model <- lm(mpg ~ transmission_type_cat + displacement + transmission_type_cat*displacement
summary(gasperform2_model)
```

```
##
## Call:
```

```
## lm(formula = mpg ~ transmission_type_cat + displacement + transmission_type_cat *
##     displacement, data = gasperform2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.99255 -0.20037  0.02234  0.24272  0.73992
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        -0.1242     0.1120  -1.110 0.276566
## transmission_type_catM             -1.5459     0.5144  -3.006 0.005540
## displacement                       -0.6505     0.1423  -4.572 8.94e-05
## transmission_type_catM:displacement  -1.5127     0.3944  -3.835 0.000653
##
## (Intercept)
## transmission_type_catM             **
## displacement                       ***
## transmission_type_catM:displacement ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4065 on 28 degrees of freedom
## Multiple R-squared:  0.8508, Adjusted R-squared:  0.8348
## F-statistic: 53.21 on 3 and 28 DF,  p-value: 1.092e-11
```

```
#When the engine is automatic, there is a 5 standard deviation difference in gasoline mileage.
#When the engine is manual, the difference is about 3 standard deviations

gasperform3 <- gasperform_standard %>% mutate(transmission_type_cat = transmission_type) %>% select(mpg
ANOVA_gasperform3 <- lm(mpg ~ transmission_type_cat + weight, data = gasperform3)
summary(ANOVA_gasperform3)
```

```
##
## Call:
## lm(formula = mpg ~ transmission_type_cat + weight, data = gasperform3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9828 -0.3575  0.0480  0.3546  1.1190
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -0.1654     0.1287  -1.285 0.208969
## transmission_type_catM   0.5881     0.3132   1.877 0.070552 .
## weight                  -0.6389     0.1431  -4.465 0.000112 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5309 on 29 degrees of freedom
## Multiple R-squared:  0.7364, Adjusted R-squared:  0.7182
## F-statistic:  40.5 on 2 and 29 DF,  p-value: 4.025e-09
```

```r
#The type of transmission does not sinificantly affect mileage performance.
#The estimated difference was still below 2 standard errors from the mean.


gasperform3_model <- lm(mpg ~ transmission_type_cat + weight + transmission_type_cat*weight, data = gasp
summary(gasperform3_model)
```

```
##
## Call:
## lm(formula = mpg ~ transmission_type_cat + weight + transmission_type_cat *
##     weight, data = gasperform3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.86311 -0.29206  0.05883  0.22432  0.77915
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -0.2480     0.1051  -2.360 0.025490 *
## transmission_type_catM        -0.7997     0.4188  -1.909 0.066525 .
## weight                        -0.4590     0.1227  -3.743 0.000834 ***
## transmission_type_catM:weight  -1.4331     0.3462  -4.140 0.000289 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4255 on 28 degrees of freedom
## Multiple R-squared:  0.8365, Adjusted R-squared:  0.8189
## F-statistic: 47.73 on 3 and 28 DF,  p-value: 3.908e-11
```

```r
#The type of transmission was not a significant enough affect on mileage performance.

#Both weight and displacement cause significant effects on mileage performance.
#They should both be included in the final regression model.
```