

# Stat 021 Homework 6

Bing Xin Tu

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

**Q0)** To help you with some programming tools you will need on your final project, please complete the R *swirl* tutorial on writing functions.

Swirl tutorial completed.

**Q1)** Read the data uploaded to Moodle called “mileage.csv”. This data describes 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.

```
car_data <- read_csv("~/Google Drive Swat/Swat docs/Stat 21/Homework/mileage.csv",
                     col_names = c("Car", "mpg", "displacement", "weight", "transmission_type"),
                     cols(col_character(), col_double(), col_double(), col_double(), col_factor()))
car_data = as_tibble(car_data)
car_data = car_data[c(-1,-2),]

#standardizes data and fct_infreq makes transmission type A the default
car_data_standard <- car_data %>%
  mutate_at(vars("displacement", "weight"), funs(scale)) %>%
  mutate(transmission_type = fct_infreq(transmission_type))

head(car_data_standard)
```

```
## # A tibble: 6 x 5
##   Car      mpg displacement[,1] weight[,1] transmission_type
##   <chr>    <dbl>          <dbl>      <dbl> <fct>
## 1 Apollo    18.9           0.558      0.371 A
## 2 Omega     17            0.558     -0.728 A
## 3 Nova     20           -0.297     -0.0476 A
## 4 Monarch   18.2           0.566      0.350 A
## 5 Duster    20.1          -0.510     -0.199 M
## 6 JensonConv 11.2           1.33       0.691 A
```

- 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)

```
MLR_car_standard <- lm(mpg ~ displacement + transmission_type, data = car_data_standard)

summary(MLR_car_standard)
```

```
##
## Call:
## lm(formula = mpg ~ displacement + transmission_type, data = car_data_standard)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.9153 -1.8679  0.1302  1.7907  6.7826
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)      20.0776      0.8343  24.065 < 2e-16 ***
## displacement    -5.3540      1.0176  -5.262 1.23e-05 ***
## transmission_typeM  0.5173      2.2276   0.232  0.818
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.117 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
```

Based on the regression output, we can see that the type of transmission does not significantly affect the mileage performance. Based on the summary of the regression output, we can see that the p-value for transmission is at 0.818. If we are using an alpha value of 0.05, then we fail to reject the null, meaning the transmission type of the car does not significantly affect its mileage.

- 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)

```
#avg effect on gasoline mileage including interaction terms
MLR_car_interaction <- lm(mpg ~ displacement +
                          transmission_type +
                          displacement*transmission_type, data = car_data_standard)

summary(MLR_car_interaction)

##
## Call:
## lm(formula = mpg ~ displacement + transmission_type + displacement *
##     transmission_type, data = car_data_standard)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.2712 -1.2660  0.1412  1.5336  4.6750
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   19.4381     0.7074  27.478 < 2e-16 ***
## displacement                   -4.1102     0.8990  -4.572 8.94e-05 ***
## transmission_typeM              -9.7677     3.2498  -3.006 0.005540 **
## displacement:transmission_typeM -9.5579     2.4922  -3.835 0.000653 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.568 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

#car_data_standardA <- car_data %>%
#     mutate_at(vars("displacement", "weight"), funs(scale)) %>%
#     mutate(transmission_fctA = fct_rev(transmission_type))

#MLR_car_interactionA <- lm(mpg ~ displacement +
#     transmission_fctA +
#     displacement*transmission_fctA, data = car_data_standardA)
```

```
#summary(MLR_car_interactionA)
```

The average effect on gasoline mileage when the engine is automatic is  $19.44 - 4.11x_1$ . This means that we expect gasoline mileage to change on average of  $19.44 - 4.11x_1$  miles when the transmission type of the car is automatic.

The average effect on gasoline mileage when the engine is manual is  $9.67 - 13.67x_1$ . This means that we expect gasoline mileage to change on average of  $9.67 - 13.67x_1$  miles when the transmission type of the car is manual.

- 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)

```
MLR_weight_trans <- lm(mpg ~ weight +
                        transmission_type, data = car_data_standard)

summary(MLR_weight_trans)

##
## Call:
## lm(formula = mpg ~ weight + transmission_type, data = car_data_standard)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.2095 -2.2586  0.3033  2.2403  7.0699
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    19.1781     0.8133   23.581  < 2e-16 ***
## weight         -4.0369     0.9041   -4.465  0.000112 ***
## transmission_typeM  3.7158     1.9792    1.877  0.070552 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.354 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
```

Based on the regression output, we can see that the type of transmission does not significantly affect the mileage performance. Based on the summary of the regression output, we can see that the p-value for transmission is at 0.07. If we are using an alpha value of 0.05, then we fail to reject the null, meaning the transmission type of the car does not significantly affect its mileage.

- d) Modify the model developed in part c 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)

```
#avg effect on gasoline mileage including interaction terms
MLR_weight_trans_int <- lm(mpg ~ weight +
                           transmission_type +
                           weight*transmission_type, data = car_data_standard)

summary(MLR_weight_trans_int)
```

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

```
## lm(formula = mpg ~ weight + transmission_type + weight * transmission_type,
##     data = car_data_standard)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.4534 -1.8453  0.3717  1.4173  4.9229
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      18.6563     0.6640  28.098 < 2e-16 ***
## weight           -2.9003     0.7749  -3.743 0.000834 ***
## transmission_typeM -5.0526     2.6464  -1.909 0.066525 .
## weight:transmission_typeM -9.0550     2.1874  -4.140 0.000289 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.689 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
```

The average effect on gasoline mileage when the engine is automatic is  $18.66 - 2.09x_1$ . This means that we expect gasoline mileage to change on average of  $18.66 - 2.09x_1$  miles when the transmission type of the car is automatic.

The average effect on gasoline mileage when the engine is manual is  $13.61 - 11.96x_1$ . This means that we expect gasoline mileage to change on average of  $13.61 - 11.96x_1$  miles when the transmission type of the car is manual.

- 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)

Based off of the results for parts (a)-(d), I think displacement and weight should be included as quantitative variables in the model. This is because for all four regressions, the outputs of these two variables have shown to have a statistically significant effect on mileage. In addition, while transmission type does not seem to have a significant effect on mileage in three of the four models ran, I would still consider adding this categorical binary variable as its interaction with other variables have significant effects. Thus, it is imperative to include the weight and transmission interaction with transmission type in our final model because both interaction terms have yielded statistically significant effects on mileage, seen in models b and d.