Course 2 Section 3.5 - PREDICTION

Jiaying Wu

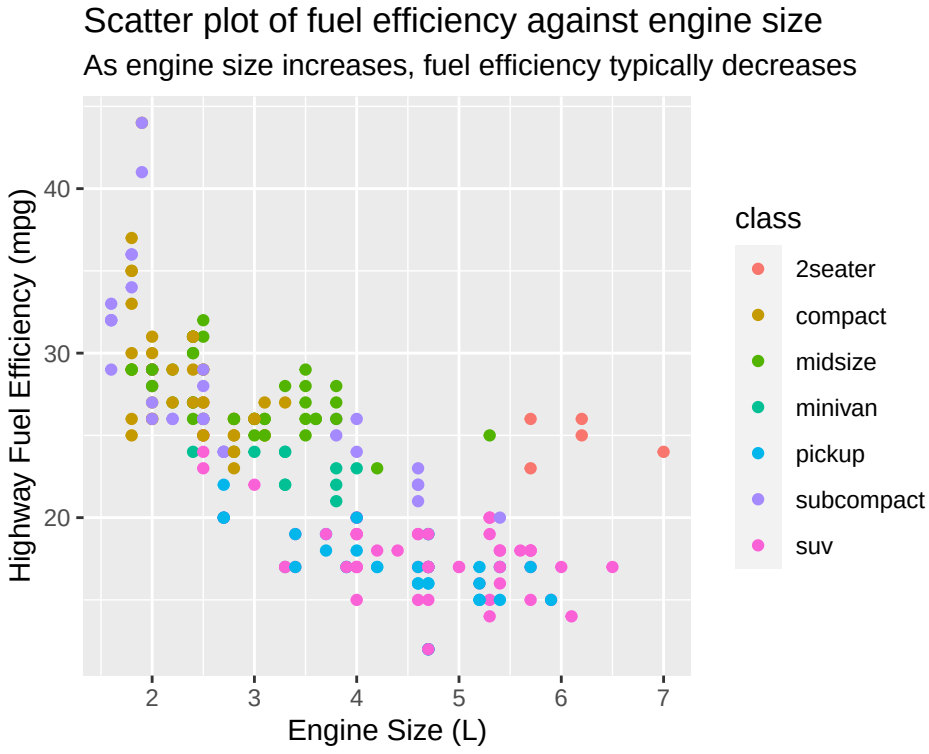
16/10/2020

```
#load library
library(tidyverse)
library(broom)
```

Give it a go!

Q1.Produce a scatter plot of fuel efficiency against engine size with each dot coloured by vehicle class.

```
mpg %>%
  ggplot(aes(x = displ, y = hwy, colour=class)) + # hwy - hwy fuel efficiency & displ - engine size
  geom_point() +
  labs(x = "Engine Size (L)",
       y = "Highway Fuel Efficiency (mpg)",
       title = "Scatter plot of fuel efficiency against engine size",
       subtitle = "As engine size increases, fuel efficiency typically decreases")
```



Q2.Estimate a multiple linear regression model of highway fuel efficiency using engine size and vehicle class as independent variables.

```
multi_lin_fit <- lm(formula = hwy ~ displ + class, data = mpg)
```

Q3.Use the the summary() or tidy() function to return the estimated coefficients and report the fitted model in equation form.

```
summary(multi_lin_fit)
```

```
##
## Call:
## lm(formula = hwy ~ displ + class, data = mpg)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.572 -1.569 -0.245  1.355 14.724
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      38.9533     1.7976  21.669  < 2e-16 ***
## displ            -2.2976     0.2132 -10.778  < 2e-16 ***
## classcompact     -5.3122     1.5283  -3.476 0.000610 ***
## classmidsize     -4.9471     1.4722  -3.360 0.000914 ***
## classminivan     -8.7986     1.5939  -5.520 9.26e-08 ***
## classpickup     -11.9232     1.3687  -8.711 6.46e-16 ***
## classsubcompact  -4.6988     1.5097  -3.112 0.002095 **
## classsuv        -10.5851     1.3268  -7.978 7.43e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.745 on 226 degrees of freedom
## Multiple R-squared:  0.7939, Adjusted R-squared:  0.7875
## F-statistic: 124.3 on 7 and 226 DF,  p-value: < 2.2e-16
```

```
tidy(multi_lin_fit)
```

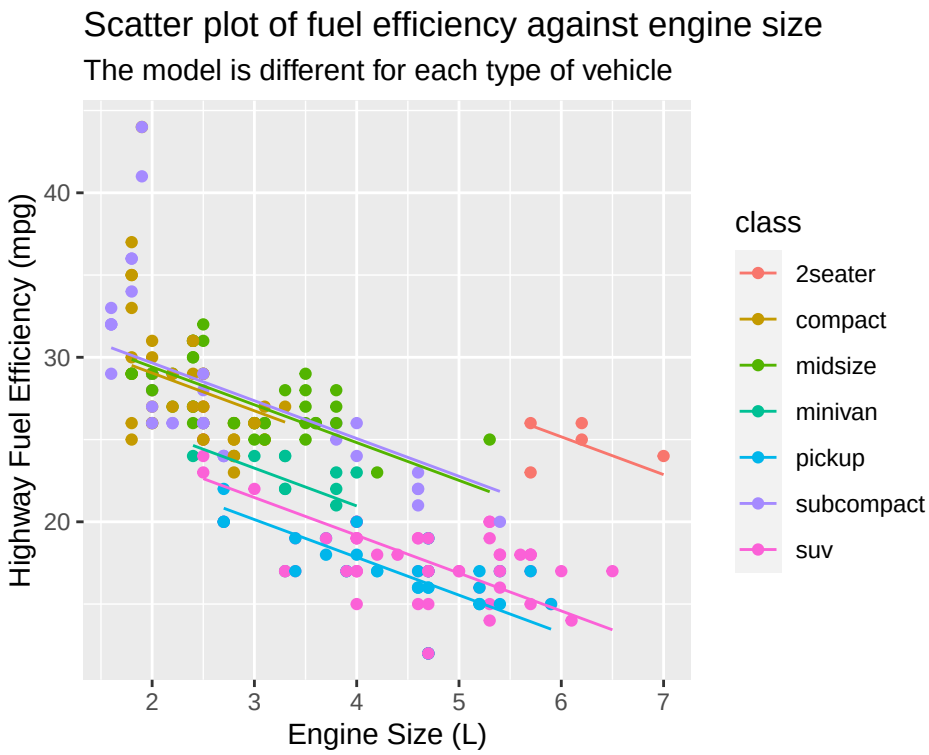
```
## # A tibble: 8 x 5
##   term            estimate std.error statistic  p.value
##   <chr>              <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)        39.0      1.80      21.7  4.36e-57
## 2 displ              -2.30     0.213    -10.8  3.99e-22
## 3 classcompact       -5.31     1.53      -3.48 6.10e- 4
## 4 classmidsize       -4.95     1.47      -3.36 9.14e- 4
## 5 classminivan       -8.80     1.59      -5.52 9.26e- 8
## 6 classpickup       -11.9      1.37      -8.71 6.46e-16
## 7 classsubcompact    -4.70     1.51      -3.11 2.09e- 3
## 8 classsuv          -10.6      1.33      -7.98 7.43e-14
```

$\hat{hwy} = 38.953 - 2.2976displ - 5.3122classcompact - 4.9471classmidsize - 8.7986classminivan - 11.9232classpickup - 4.6988classsubcompact - 10.5851classsuv$

Q4.Plot the fitted model over the scatter plot.

```
mpg_fit <- augment(multi_lin_fit)

mpg_fit %>%
  ggplot() +
  geom_point(aes(x = displ, y = hwy, colour=class)) +
  geom_line(aes(x = displ, y=.fitted, colour=class)) +
  labs(x = "Engine Size (L)",
       y = "Highway Fuel Efficiency (mpg)",
       title = "Scatter plot of fuel efficiency against engine size",
       subtitle = "The model is different for each type of vehicle")
```



Q5.How much of the variation in highway fuel efficiency can be explained by engine size and vehicle class (the summary() function returns a statistic that will help you answer this question)?

The R-squared is 0.7939, the engine size and vehicle class help to explain 79.39% of the variation in highway fuel efficiency.

Q6.Interpret the estimated slope coefficient attached to the engine size variable.

displ, coef = -2.3. with class fixed, engine size increases by 1L, the hwy decreases 2.3 miles per gallon.

Q7.Predict the highway fuel efficiency of a pickup with a 2L engine (hint: let pickup indicator variable be equal to 1 and all other the class indicator variables be equal to 0).

hwy = 39 - 2.3*2 - 11.9 = 22.5

Q8.Why might predictions of the highway fuel efficiency of a Tesla Model 3 using this regression model be unreliable?

There are no data for this car model (electricity engine) in `mpg`.