Econometrics project

Spring 2023

Dr, Rania Mamdouh

Dr, Yousra Hassan

Name:	**Code:**
Haneen Ahmed Elgabry	**5200199**
Nour Mohamed Eletreby	**5200917**
Sherine Reda Fatehy	**5200386**

Introduction

This data used in this project is collected about cars the size of the data was 805 observations, 9 variables.

This analysis is conducted on a random sample of size 250 in which the dependent variable: " Price of used cars “and the independent variables:

1- Mileage: The number of miles the car has been driven (Unit of measurement: miles)

2- Make: manufacturer of the car such as Saturn, Pontiac, and Chevrolet

3-Type: body type such as sedan, coupe, etc.

4-Cylinder: engine’s capacity low, moderate, and high

5-Doors: number of doors

6-Cruise: indicator variable representing whether the car has cruise control (1 = cruise)

7-Sound: indicator variable representing whether the car has upgraded speakers (1=upgraded)

8-Leather: indicator variable representing whether the car has leather seats (1 = leather)

We must convert the categorical variables into dummy variables. Therefore, we will use K-1 dummies as follows:

For (make): We will create 5 dummies and the base category is Buick.

For (Type): We will create 4 dummies and the base category is Convertible.

For (Cylinder): We will create 2 dummies and the base category is High.

The data we have about the price of concerned cars and these factors we will use it in studying the effect of these factors on the car price through answering some questions in the following pages.

Analysis

In the analysis part we are going to answer the following questions in order to see the effect of each independent variable on the price.

Question (1): Does the mean of the price of the car differ according to the make of the car?

The model: $Y = \beta 0 + \beta 1 D1 + \beta 2 D2 + \beta 3 D3 + \beta 4 D4 + \beta 5 D5 + +Ui$

$E(Price) = \beta 0 + \beta 1 Cadillac + \beta 2 Chevrolet + \beta 3 Pontiac + \beta 4 SAAB + \beta 5 Saturn$

R-output:

```
Call:
lm(formula = price ~ Make, data = car_data)

Residuals:
Min      1Q  Median      3Q     Max 
-3.5057 -0.3943  0.0825  0.5280  2.0109 

Coefficients:
Estimate Std. Error t value Pr(>|t|)    
(Intercept)     7.0146     0.1825  38.436  < 2e-16 ***
MakeCadillac   -1.9303     0.2453  -7.869 1.16e-13 ***
MakeChevrolet   1.1170     0.2051   5.446 1.26e-07 ***
MakePontiac     0.5237     0.2267   2.310   0.0217 *  
MakeSAAB       -1.0609     0.2419  -4.385 1.73e-05 ***
MakeSaturn      1.4924     0.2737   5.452 1.22e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9125 on 244 degrees of freedom
Multiple R-squared:  0.6051,   Adjusted R-squared:  0.597 
F-statistic: 74.78 on 5 and 244 DF,  p-value: < 2.2e-16
```

The estimated model:

$E\ (Price)$ = 7.0146-1.9303cadillac+1.1170chevrolet+0.5237pontiac-1.0609SAAB+1.4924saturn

The hypotheses:

$H0: \beta 1 = \beta 2 = \cdots \ldots . = \beta 5 = 0\ (insignificant)$

$H1: at\ least\ one\ of\ \beta i \neq 0\ (significant)$

Interpretation:

The p-values of all the differential intercepts are less than 0.05 which means that we will reject $H0$ and they are all significant.

$E\ (Y|X = \text{Cadillac}) = \beta 0 + \beta 1$

$E\ (Y|X = \text{Chevrolet}) = \beta 0 + \beta 2$

$E\ (Y|X = \text{Pontiac}) = \beta 0 + \beta 3$

$E\ (Y|X = \text{SAAB}) = \beta 0 + \beta 4$

$E\ (Y|X = \text{Saturn}) = \beta 0 + \beta 5$

$E\ (Y|X = Buick) = \beta 0$

They are statistically different.

$\beta 0$: The mean price to make Buick car is 7.0146 thousand $.

$\beta 1$: The mean price to make the Cadillac car is less than the mean price to make the Buick car by 1.9303 thousand $.

$\beta 2$: The mean price to make the Chevrolet car is greater than the mean price to make the Buick car by 1.117 thousand $.

$\beta 3$: The mean price to make the Pontiac car is less than the mean price to make the Buick car by 0.5237 thousand $.

$\beta 4$: the mean price to make the SAAB car is less than the mean price to make the Buick car by 1.0609 thousand $.

$\beta 5$: the mean price to make the Saturn car is less than the mean price to make the Buick car by 1.4924 thousand $.

Answer: Yes, we concluded that the mean of the car price differs according to the make of the car.

<u>Question (2):</u> Does the mean of the price of the car differ according to the make of the car, for any car type, and any Cylinder size?

The model: $Y = \beta 0 + \beta 1 D1 + \beta 2 D2 + \beta 3 D3 + \beta 4 D4 + \beta 5 D5 + \alpha 1 D6 + \alpha 2 D7 + \alpha 3 D8 + \alpha 4 D9 + \gamma 1 D10 + \gamma 2 D11 + Ui$

$E(Price) = \beta 0 + \beta 1 Cadillac + \beta 2 Chevrolet + \beta 3 Pontiac + \beta 4 SAAB + \beta 5 Saturn + \alpha 1 Coupe + \alpha 2 Hatchback + \alpha 3 Sedan + \alpha 4 Wagon + \gamma 1 low + \gamma 2 moderat$.

R-output:

```
Call:
lm(formula = price ~ Make + Type + Cylinder, data = car_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.20034 -0.31320  0.00493  0.34568  1.18182 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)        4.9964     0.1796  27.824  < 2e-16 ***
MakeCadillac      -1.0249     0.1563  -6.559 3.35e-10 ***
MakeChevrolet      0.3437     0.1225   2.806  0.00543 ** 
MakePontiac        0.3188     0.1254   2.542  0.01165 *  
MakeSAAB          -2.1668     0.1558 -13.907  < 2e-16 ***
```

```
MakeSaturn          0.3324     0.1593   2.087  0.03797 *
TypeCoupe           0.8782     0.1569   5.596 6.01e-08 ***
TypeHatchback       1.5117     0.1886   8.014 4.96e-14 ***
TypeSedan           0.9658     0.1409   6.856 6.03e-11 ***
TypeWagon           0.3979     0.1795   2.216  0.02763 *
Cylinderlow         2.5353     0.1353  18.737  < 2e-16 ***
Cylindermoderate    1.0523     0.1267   8.306 7.53e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4775 on 238 degrees of freedom
Multiple R-squared:  0.8945,   Adjusted R-squared:  0.8897
F-statistic: 183.5 on 11 and 238 DF,  p-value: < 2.2e-16
```

The estimated model:

$E(Price)$ = 4.9964 – 1.0249 $Cadillac$ + 0.3437 $Chevrolet$ + 0.3188 $Pontiac$ − 2.1668 $SAAB$ + 0.3324 $Saturn$ + 0.8782$Coupe$ + 1.5117 $Hatchback$ + 0.9658 $Sedan$ + 0.3979 $Wagon$ + 2.5353 low + 1.0523 $moderate$

Hypotheses:

$H0$: $\beta 1 = \beta 2 = \cdots \ldots. = \beta 5 = 0$

$H1$: $at\ least\ one\ of\ \beta \neq 0$

Interpretation:

-There is not statistically difference between the mean price in (Make Saturn and Make Pontiac) and (make Buick) (as p-value >0.05) which means statistically that the mean price of cars of (Make Saturn and Make Pontiac) and make Buick are the same.

- While the mean price of cars which (Make Cadillac, Make SAAB and Make Chevrolet) are statistically different from the mean price of cars which (make Buick) by -1.0249, -2.1668, 0.3437 thousand of dollars respectively all other variables constant. (As p-value<0.05).

- There is not statistically difference between the mean price in (Type Wagon) and (Type Convertible) (as p-value>0.05) which means statistically that the mean price of cars of (Type Wagon) and (Type Convertible) are the same.

- While the mean price of cars which (Type Hatchback, Type Coupe, Type Sedan) are statistically different from the mean price of teachers in (Type convertible) by 1.5117, 0.8782, 0.9658 thousand of dollars respectively all other variables constant (as p-value<0.05).

-The mean price of cars which (Cylinder low, Cylinder moderate) are statistically different from the mean price of cars in (Cylinder high) by 2.5353, 1.0523 thousand of dollars respectively all other variables constant (as p-value<0.05).

Answer: we can conclude that the mean of the price of the car differ according to the make of the car, for any car type, and any Cylinder size.

Question (3): Is there an interaction between the make of the car and its cylinder size?

The model:

$$Y = \beta 0 + \beta 1D1 + \beta 2D2 + \beta 3D3 + \beta 4D4 + \beta 5D5 + \alpha 1D6 + \alpha 2D7 + \gamma 1D1D6 + \gamma 2D1D7 + \gamma 3D2D6 + \gamma 4D2D7 + \gamma 5D3D6 + \gamma 6D3D7 + \gamma 7D4D6 + \gamma 8D4D7 + \gamma 9D5D6 + \gamma 10D5D7 + Ui$$

$$E(Price) = \beta 0 + \beta 1Cadillac + \beta 2Chevrolet + \beta 3Pontiac + \beta 4SAAP + \beta 5Saturn + \alpha 1Low + \alpha 2Moderate + \gamma 1Cadillac.Low + \gamma 2Cadillac.Moderate + \gamma 3Chevrolet.Low + \gamma 4Chevrolet.Moderate + \gamma 5Pontiac.Low + \gamma 6Pontiac.Moderate + \gamma 7SAAP.Low + \gamma 8SAAP.Moderate + \gamma 9Saturn.Low + \gamma 10Saturn.Moderate$$

R output:

```
Call:
lm(formula = price ~ Make + Cylinder + Make * Cylinder, data = car_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.1798 -0.3126  0.0052  0.3275  1.1608 

Coefficients: (6 not defined because of singularities)
                              Estimate Std. Error t value Pr(>|t|)    
(Intercept)                     5.9000     0.2476  23.830  < 2e-16 ***
MakeCadillac                   -0.9204     0.2663  -3.456 0.000649 ***
MakeChevrolet                  -0.8775     0.2983  -2.942 0.003584 ** 
MakePontiac                     0.3968     0.1306   3.039 0.002638 ** 
MakeSAAB                       -2.3095     0.2910  -7.935 8.22e-14 ***
MakeSaturn                      0.4935     0.2533   1.948 0.052555 .  
Cylinderlow                     2.3632     0.3487   6.777 9.56e-11 ***
Cylindermoderate                1.1146     0.2290   4.867 2.07e-06 ***
MakeCadillac:Cylinderlow            NA         NA      NA       NA    
MakeChevrolet:Cylinderlow       1.5960     0.3916   4.075 6.26e-05 ***
MakePontiac:Cylinderlow        -0.4335     0.3052  -1.420 0.156773    
MakeSAAB:Cylinderlow                NA         NA      NA       NA    
MakeSaturn:Cylinderlow              NA         NA      NA       NA    
MakeCadillac:Cylindermoderate  -0.7087     0.2995  -2.366 0.018784 *  
MakeChevrolet:Cylindermoderate  1.3570     0.2946   4.606 6.70e-06 ***
MakePontiac:Cylindermoderate        NA         NA      NA       NA    
MakeSAAB:Cylindermoderate           NA         NA      NA       NA    
MakeSaturn:Cylindermoderate         NA         NA      NA       NA    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4704 on 238 degrees of freedom
Multiple R-squared:  0.8976,	Adjusted R-squared:  0.8929 
F-statistic: 189.7 on 11 and 238 DF,  p-value: < 2.2e-16
```

The estimated model:

$$E(Price) = 5.9 - 0.9204Cadillac - 0.8775Chevrolet + 0.3968pontiac - 2.3095SAAP + 0.4935Saturn + 2.3632Low + 1.1146Moderate - 0.7087Cadillac.Moderate + 1.596Chevrolet.Low + 1.357Chevrolet.Moderate - 0.4335Pontiac.Low$$

The hypotheses:

$H0: \gamma 1 = \gamma 2 = \cdots \ldots = \gamma 10 = 0$

$H1: at\ least\ one\ of\ \gamma \neq 0$

Interpretation:

-The p-values of the interaction coefficients are less than 0.05 (They are significant) except for interaction term Pontiac.Low

-We can notice that there is no observations for interaction term (Cadillac.Low),(SAAP.Low) , (Saturn.Low) , (Pontiac.Moderate),(SAAP.Moderate),(Saturn.Moderate) and this represented on R output by NA , So we excluded them from the analysis.

Answer: Yes, there is an interaction between the make and its cylinder size. (The effect of cylinder size is not constant among different types of make of car).

<u>Question (4):</u> On average does the make of the car affect the price of the car holding the number of miles the car has been driven and number of doors constant, and what is the effect of increasing the number of miles the car has been driven and number of doors on the price of car?

Model:

$Y = \beta 0 + \beta 1 D1 + \beta 2 D2 + \beta 3 D3 + \beta 4 D4 + \beta 5 D5 + \alpha 1 X1 + \alpha 2 X2 + Ui$

$E(Price) = \beta 0 + \beta 1 Cadillac + \beta 2 Chevrolet + \beta 3 Pontiac + \beta 4 SAAP + \beta 5 Saturn + \alpha 1 \text{Mileage} + \alpha 2 \text{Doors}$

R-output:

```
Call:
lm(formula = price ~ Make + Mileage + Doors, data = car_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.80247 -0.44658  0.02195  0.47898  1.85938 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)    5.490e+00  3.583e-01  15.321  < 2e-16 ***
MakeCadillac  -1.866e+00  2.357e-01  -7.917 8.77e-14 ***
MakeChevrolet  1.275e+00  2.008e-01   6.350 1.06e-09 ***
MakePontiac    6.489e-01  2.196e-01   2.954  0.00344 ** 
MakeSAAB      -9.884e-01  2.342e-01  -4.221 3.45e-05 ***
MakeSaturn     1.645e+00  2.654e-01   6.198 2.44e-09 ***
Mileage        1.854e-05  6.985e-06   2.654  0.00848 ** 
Doors          2.915e-01  6.789e-02   4.294 2.54e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8742 on 242 degrees of freedom
Multiple R-squared:  0.6405,	Adjusted R-squared:  0.6301 
F-statistic:  61.6 on 7 and 242 DF,  p-value: < 2.2e-16
```

The estimated model:

$E(Price)$ = 5.940e + 00 −1.866e + 00$Cadillac$ + 1.275e + 00$Chevrolet$ +6.489e − 01$Pontiac$ −9.884e - 01$SAAP$ + 1.645e + 00$Saturn$ +1.854e − 01Mileage +2.915e − 05Doors.

Hypothesis:

$H0: \beta 1 = \beta 2 = \cdots .... = \beta 5 = 0$

$H1: at\ least\ one\ of\ \beta \neq 0$

Interpretation:

-The p-values of coefficients ($\beta 1$, $\beta 2$, $\beta 3$, $\beta 4$, $\beta 5$) are less than 0.05 (They are significant) we can say that on average the make of the car has a significant role in affecting the price of the car making other factors that affects the price of the car as constant.

-The mean price of cars which (MakeCadillac, MakeChevrolet, MakeSAAB, MakePontiac, MakeSaturn) are statistically different from the mean price of cars which (makeBuick) by (−1.866e + 00), (1.275e + 00) , (−9.884e - 01), (6.489e − 01),(1.645e + 00) thousands of dollars respectively and by making all other variables constant .

-If we increase the number of doors the car has by 1 door the price of the car will increase by 2.915e − 05thousands of dollars on average, holding other variables that affect the price as constant.

-If we increase the number of miles the car has been driven by 1 mile using one liters of Oil the price of the car will increase by 1.854e − 01thousands of dollars on average, holding other variables that affect the price as constant.

Answer: we can conclude that the price of the car increases when number of miles the car has been driven increases but with less increase than when the number of doors increases.

Question 5: The effect of the number of miles and the price of the car differ according to the type of the car?

The model:

$Y = \beta 0 + \beta 1D1 + \beta 2D2 + \beta 3D3 + \beta 4D4 + \alpha 1X + \gamma 1X.\ D1 + \gamma 2X.\ D2 + \gamma 3X.\ D3 + \gamma 4X.\ D4 + Uit$

$E(Price) = \beta 0 + \beta 1\text{Hatchback} + \beta 2\text{Sedan} + \beta 3\text{Coupe} + \beta 4\text{Wagon} + \alpha 1\text{Mile} + \gamma 1\text{Mile. Hatchback} + \gamma 2\text{Mile. Sedan} + \gamma 3\text{Mile. Coupe} + \gamma 4\text{Mile. Wagon}$

R-output:

```
Call:
lm(formula = price ~ Mileage + Type + Mileage * Type, data = car_data)

Residuals:
    Min      1Q  Median      3Q     Max
```

```
-2.8162 -0.9237  0.1252  0.7621  2.7111

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)    
(Intercept)            4.346e+00  7.107e-01   6.114 3.90e-09 ***
Mileage                2.800e-05  2.712e-05   1.032  0.30292    
TypeCoupe              2.942e+00  9.022e-01   3.261  0.00127 ** 
TypeHatchback          4.485e+00  1.039e+00   4.315 2.34e-05 ***
TypeSedan              2.292e+00  7.598e-01   3.017  0.00283 ** 
TypeWagon              3.404e+00  1.252e+00   2.720  0.00701 ** 
Mileage:TypeCoupe      6.699e-06  3.689e-05   0.182  0.85606    
Mileage:TypeHatchback -2.798e-05  4.622e-05  -0.605  0.54546    
Mileage:TypeSedan     -2.673e-08  2.988e-05  -0.001  0.99929    
Mileage:TypeWagon     -6.474e-05  5.254e-05  -1.232  0.21903    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.217 on 240 degrees of freedom
Multiple R-squared:  0.3092,	Adjusted R-squared:  0.2833 
F-statistic: 11.94 on 9 and 240 DF,  p-value: 1.597e-15
```

The estimated model:

$E(Price)$ = 4.346 + 4.485Hatchback + 2.292 Sedan + 2.942Coupe + 3.404Wagon + 2.800e-05Mile -2.798e-05Mile.Hatchback -2.673e-08Mile. Sedan + 6.699e-06Mile. Coupe – 6.474e-05Mile. Wagon

The hypotheses:

$H0: \gamma 1 = \gamma 2 = \gamma 3 = \gamma 4 = 0$

$H1: at\ least\ one\ of\ \gamma i \neq 0$

Interpretation:

-The p-values of all the interaction coefficients are greater than 0.05, They are insignificant, Therefor Don’t reject $H0$ with confidence interval 95%.

- (The effect of no. of miles is constant among different types of cars).

Answer: We can conclude from this that there is an interaction between the number of the miles and the type of the car, for example, the effect of the number of miles is the same for any type (the effect of miles is constant among different types of car).

<u>Question 6:</u> What are the effects of each of cruise sound, and leather on the price of car is there any interaction between them?

The model:

$Y = \beta 0 + \beta 1D1 + \beta 2D2 + \beta 3D3 + \alpha 1D1D2 + \alpha 2D1D3 + \alpha 3D2D3 + Ui$

$E(Price)$ = $\beta 0$ + $\beta 1$Cruise + $\beta 2$Sound + $\beta 3$Leather + $\alpha 1$Cruise. Sound +$\alpha 2$Cruise. leather + $\alpha 3$Sound. Leather

R-output:

```
Call:
lm(formula = price ~ Cruise + Sound + Leather + Cruise * Sound +
    Cruise * Leather + Sound * Leather, data = car_data)

Residuals:
    Min      1Q  Median      3Q     Max
-2.9663 -0.8426 -0.1115  0.7524  3.1614

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)      9.0045     0.3168  28.421  < 2e-16 ***
Cruise          -1.3573     0.3733  -3.636 0.000338 ***
Sound           -0.4468     0.4708  -0.949 0.343497
Leather         -0.6910     0.4596  -1.504 0.133972
Cruise:Sound     0.3328     0.4532   0.734 0.463443
Cruise:Leather  -0.7573     0.4711  -1.608 0.109235
Sound:Leather    0.7004     0.3565   1.965 0.050597 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.177 on 243 degrees of freedom
Multiple R-squared:  0.3461,   Adjusted R-squared:   0.33
F-statistic: 21.44 on 6 and 243 DF,  p-value: < 2.2e-16
```

The estimated model:

$E(Price)$ = 9.0045 – 1.3573Cruise – 0.4468Sound – 0.6910Leather – 0.7573 Cruise. Leather + 0.3328 Cruise. Sound + 0.7004 Sound. Leather

The hypotheses:

***Part* 1:**

$H0: \alpha 1 = \alpha 2 = \alpha 3 = 0$

$H1: at\ least\ one\ of\ \alpha i \neq 0$

***Part* 2:**

$H0: \beta 1 = \beta 2 = \beta 3 = 0$

$H1: at\ least\ one\ of\ \beta i \neq 0$

Interpretation:

Part 1:

P-values of the interaction coefficients “$\alpha 1$ and $\alpha 2$ and a3” are insignificant which means that (p-value > 0.05). Then, don’t reject $H0$ with confidence level 95%.

Part 2:

-P-value of $\beta 1$ is < 0.05 “it’s significant” with confidence interval 95% This means that, on average, a one-unit increase in the Cruise feature is associated with a 1.3573 thousand dollar decrease in the predicted price of the car, while holding all other independent variables constant.

- P-value of $\beta 2$ and $\beta 3$ is > 0.05 “it’s insignificant” with confidence interval 95%, then it can be said that there is no evidence to suggest that the Sound and Leather features have a significant impact on the price of the car in this model. That is, the coefficients for Sound (b2) and Leather (b3) may not be reliably different from zero.

- therefore, we reject H0 with confidence level 95%. As B1 is statistically different from 0 unlike B2 and B3 .

Question (7): Does the regression price- number of miles change once the number of miles achieves 15 thousand miles?

The model:

$Y = \beta 0 + \beta 1X + \beta 2(X - X*)D + Ui$

$E(Price) = \beta 0 + \beta 1\text{Mile} + \beta 2(\text{Mile} - 15000)\text{D}$

Where, D = 1 when Mile ≥ 15000 D = 0 when Mile < 15000

R-output:

```
Call:
lm(formula = price ~ Mileage + Mileage_15, data = car_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.2566 -1.2239  0.2034  1.1133  2.8502 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 6.995e+00  2.507e-01  27.907   <2e-16 ***
Mileage     1.171e-05  1.752e-05   0.668    0.505    
Mileage_15  5.291e-02  3.331e-01   0.159    0.874    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.439 on 247 degrees of freedom
Multiple R-squared:  0.006003, Adjusted R-squared:  -0.002046 
F-statistic: 0.7458 on 2 and 247 DF,  p-value: 0.4754
```

The estimated model:

$E(Price) = 6.778 + 0.00002199\text{Mile} + 0.000008422\ (\text{Mile} – 15000)\ \text{D}$

The hypotheses:

$H0: \beta 2 = 0$

$H1: \beta 2 \neq 0$

Interpretation:

From the previous R output, we observed that the p-value of $\beta 2$ is greater than 0.05 which means that we won't reject $H0$ and it's insignificant with confidence interval 95%.

Answer: We can conclude that there's no "Break Point". i.e., the regression model price-number of miles doesn't change once the number of miles achieves 15 thousand miles. It's the same before and after 15000 thousand miles "no need to define a piecewise regression model we can define a simple model directly."

Conclusion

Based on the data and analysis provided in the report, several conclusions can be drawn. Firstly, the mean price of a car varies based on the make of the car, and this is true for any car type and cylinder size. Additionally, there is an interaction between the make of the car and its cylinder size, meaning that the effect of the cylinder size on the price of the car is not constant among different makes of cars.

The study also found that the price of the car increases as the number of miles driven increases, but this increase is less significant than when the number of doors increases. There is also an interaction between the number of miles driven and the body type of the car, indicating that the effect of the number of miles on the price of the car is constant among different body types of cars.

Finally, the study found that there is no "break point" in the relationship between the price of the car and the number of miles driven. This means that the regression model for price and number of miles driven does not change once the number of miles achieved 15,000 miles. Therefore, there is no need to define a piecewise regression model, and a simple model can be used directly.

In summary, the study provides insights into the factors that affect the price of a car, including the make of the car, its cylinder size, the number of miles driven, and the type of the car. The findings of the study can be used to inform pricing strategies for buyers and sellers in the car market and can also serve as a basis for future research on the topic.

Appendix

```
> data <- read.csv("data car project econo.csv")
> sample_data <- data[sample(1:nrow(data), size = 250, replace = FALSE), ]
> new_file <- "C:\\Users\\20106\\Downloads\\eco project.csv"
> write.csv(sample_data, file = new_file, row.names = FALSE)
> car_data <- read.csv("C:/Users/20106/Downloads/eco project/eco project.csv"
)
> model1 <- lm(price ~ Make, data = car_data)
> summary(model1)

Call:
lm(formula = price ~ Make, data = car_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5057 -0.3943  0.0825  0.5280  2.0109 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)     7.0146     0.1825  38.436  < 2e-16 ***
MakeCadillac   -1.9303     0.2453  -7.869 1.16e-13 ***
MakeChevrolet   1.1170     0.2051   5.446 1.26e-07 ***
MakePontiac     0.5237     0.2267   2.310   0.0217 *  
MakeSAAB       -1.0609     0.2419  -4.385 1.73e-05 ***
MakeSaturn      1.4924     0.2737   5.452 1.22e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9125 on 244 degrees of freedom
Multiple R-squared:  0.6051,	Adjusted R-squared:  0.597 
F-statistic: 74.78 on 5 and 244 DF,  p-value: < 2.2e-16

> model2 <- lm(price ~ Make + Type + Cylinder, data = car_data)
> summary(model2)

Call:
lm(formula = price ~ Make + Type + Cylinder, data = car_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.20034 -0.31320  0.00493  0.34568  1.18182 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)        4.9964     0.1796  27.824  < 2e-16 ***
MakeCadillac      -1.0249     0.1563  -6.559 3.35e-10 ***
MakeChevrolet      0.3437     0.1225   2.806  0.00543 ** 
MakePontiac        0.3188     0.1254   2.542  0.01165 *  
MakeSAAB          -2.1668     0.1558 -13.907  < 2e-16 ***
MakeSaturn         0.3324     0.1593   2.087  0.03797 *  
TypeCoupe          0.8782     0.1569   5.596 6.01e-08 ***
TypeHatchback      1.5117     0.1886   8.014 4.96e-14 ***
TypeSedan          0.9658     0.1409   6.856 6.03e-11 ***
TypeWagon          0.3979     0.1795   2.216  0.02763 *  
Cylinderlow        2.5353     0.1353  18.737  < 2e-16 ***
Cylindermoderate   1.0523     0.1267   8.306 7.53e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4775 on 238 degrees of freedom
Multiple R-squared:  0.8945,	Adjusted R-squared:  0.8897 
F-statistic: 183.5 on 11 and 238 DF,  p-value: < 2.2e-16
```

```
> table(Make,Cylinder)
           Cylinder
Make        high low moderate
  Buick        0   0       25
  Cadillac    23   0        8
  Chevrolet    8  54       33
  Pontiac      5  14       27
  SAAB         0  33        0
  Saturn       0  16        4
> model3 <- lm(price ~ Make +Cylinder+Make* Cylinder, data = car_data)
> summary(model3)

Call:
lm(formula = price ~ Make + Cylinder + Make * Cylinder, data = car_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.1798 -0.3126  0.0052  0.3275  1.1608 

Coefficients: (6 not defined because of singularities)
                               Estimate Std. Error t value Pr(>|t|)    
(Intercept)                      5.9000     0.2476  23.830  < 2e-16 ***
MakeCadillac                    -0.9204     0.2663  -3.456 0.000649 ***
MakeChevrolet                   -0.8775     0.2983  -2.942 0.003584 ** 
MakePontiac                      0.3968     0.1306   3.039 0.002638 ** 
MakeSAAB                        -2.3095     0.2910  -7.935 8.22e-14 ***
MakeSaturn                       0.4935     0.2533   1.948 0.052555 .  
Cylinderlow                      2.3632     0.3487   6.777 9.56e-11 ***
Cylindermoderate                 1.1146     0.2290   4.867 2.07e-06 ***
MakeCadillac:Cylinderlow             NA         NA      NA       NA    
MakeChevrolet:Cylinderlow        1.5960     0.3916   4.075 6.26e-05 ***
MakePontiac:Cylinderlow         -0.4335     0.3052  -1.420 0.156773    
MakeSAAB:Cylinderlow                 NA         NA      NA       NA    
MakeSaturn:Cylinderlow               NA         NA      NA       NA    
MakeCadillac:Cylindermoderate   -0.7087     0.2995  -2.366 0.018784 *  
MakeChevrolet:Cylindermoderate   1.3570     0.2946   4.606 6.70e-06 ***
MakePontiac:Cylindermoderate         NA         NA      NA       NA    
MakeSAAB:Cylindermoderate            NA         NA      NA       NA    
MakeSaturn:Cylindermoderate          NA         NA      NA       NA    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4704 on 238 degrees of freedom
Multiple R-squared:  0.8976,	Adjusted R-squared:  0.8929 
F-statistic: 189.7 on 11 and 238 DF,  p-value: < 2.2e-16

> model4 <- lm(price ~ Make + Mileage + Doors, data = car_data)
> summary(model4)

Call:
lm(formula = price ~ Make + Mileage + Doors, data = car_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.80247 -0.44658  0.02195  0.47898  1.85938 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)    5.490e+00  3.583e-01  15.321  < 2e-16 ***
MakeCadillac  -1.866e+00  2.357e-01  -7.917 8.77e-14 ***
MakeChevrolet  1.275e+00  2.008e-01   6.350 1.06e-09 ***
MakePontiac    6.489e-01  2.196e-01   2.954  0.00344 ** 
MakeSAAB      -9.884e-01  2.342e-01  -4.221 3.45e-05 ***
MakeSaturn     1.645e+00  2.654e-01   6.198 2.44e-09 ***
```

```
Mileage        1.854e-05  6.985e-06   2.654  0.00848 **
Doors          2.915e-01  6.789e-02   4.294 2.54e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8742 on 242 degrees of freedom
Multiple R-squared:  0.6405,	Adjusted R-squared:  0.6301 
F-statistic:  61.6 on 7 and 242 DF,  p-value: < 2.2e-16

> model5 <- lm(price ~ Mileage + Type + Mileage*Type, data = car_data)
> summary(model5)

Call:
lm(formula = price ~ Mileage + Type + Mileage * Type, data = car_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.8162 -0.9237  0.1252  0.7621  2.7111 

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)    
(Intercept)            4.346e+00  7.107e-01   6.114 3.90e-09 ***
Mileage                2.800e-05  2.712e-05   1.032  0.30292    
TypeCoupe              2.942e+00  9.022e-01   3.261  0.00127 ** 
TypeHatchback          4.485e+00  1.039e+00   4.315 2.34e-05 ***
TypeSedan              2.292e+00  7.598e-01   3.017  0.00283 ** 
TypeWagon              3.404e+00  1.252e+00   2.720  0.00701 ** 
Mileage:TypeCoupe      6.699e-06  3.689e-05   0.182  0.85606    
Mileage:TypeHatchback -2.798e-05  4.622e-05  -0.605  0.54546    
Mileage:TypeSedan     -2.673e-08  2.988e-05  -0.001  0.99929    
Mileage:TypeWagon     -6.474e-05  5.254e-05  -1.232  0.21903    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.217 on 240 degrees of freedom
Multiple R-squared:  0.3092,	Adjusted R-squared:  0.2833 
F-statistic: 11.94 on 9 and 240 DF,  p-value: 1.597e-15

> model6 <- lm(price ~ Cruise + Sound + Leather + Cruise*Sound + Cruise*Leath
er + Sound*Leather, data = car_data)
> summary(model6)

Call:
lm(formula = price ~ Cruise + Sound + Leather + Cruise * Sound + 
    Cruise * Leather + Sound * Leather, data = car_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.9663 -0.8426 -0.1115  0.7524  3.1614 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)      9.0045     0.3168  28.421  < 2e-16 ***
Cruise          -1.3573     0.3733  -3.636 0.000338 ***
Sound           -0.4468     0.4708  -0.949 0.343497    
Leather         -0.6910     0.4596  -1.504 0.133972    
Cruise:Sound     0.3328     0.4532   0.734 0.463443    
Cruise:Leather  -0.7573     0.4711  -1.608 0.109235    
Sound:Leather    0.7004     0.3565   1.965 0.050597 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.177 on 243 degrees of freedom
Multiple R-squared:  0.3461,	Adjusted R-squared:   0.33 
```

```
F-statistic: 21.44 on 6 and 243 DF,  p-value: < 2.2e-16

> car_data$Mileage_15 <- ifelse(car_data$Mileage >= 15000, 1, 0)
> model7 <- lm(price ~ Mileage + Mileage_15, data = car_data)
> summary(model7)

Call:
lm(formula = price ~ Mileage + Mileage_15, data = car_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.2566 -1.2239  0.2034  1.1133  2.8502 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 6.995e+00  2.507e-01  27.907   <2e-16 ***
Mileage     1.171e-05  1.752e-05   0.668    0.505    
Mileage_15  5.291e-02  3.331e-01   0.159    0.874    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.439 on 247 degrees of freedom
Multiple R-squared:  0.006003,	Adjusted R-squared:  -0.002046 
F-statistic: 0.7458 on 2 and 247 DF,  p-value: 0.4754


library(writexl)
> write.xlsx(car_data, file = "C:/Users/20106/Downloads/eco project/car_data.
xlsx")
```