**PART 1**

Load the dataset named Carseats (in the ISLR package) into R.

**Background Information**

Carseats is a simulated dataset in the ISLR package with sales of child car seats at 400 different stores. The 11 variables are:

str(Carseats)

**Sales**: Unit sales (in thousands) at each location

**CompPrice**: Price charged by competitor at each location

**Income**: Community income level (in thousands of dollars)

**Advertising**: Local advertising budget for company at each location (in thousands of dollars)

**Population**: Population size in region (in thousands)

**Price**: Price company charges for car seats at each site

**ShelveLoc**: A factor with levels Bad, Good and Medium indicating the quality of the shelving location

for the car seats at each site

**Age**: Average age of the local population

**Education**: Education level at each location

**Urban**: A factor with levels No and Yes to indicate whether the store is in an urban or rural location

**US**: A factor with levels No and Yes to indicate whether the store is in the US or not

Create a new dataframe (call it C1) that is a copy of Carseats. Create two indicator (dummy) variables:

* + - Bad\_Shelf = 1 if ShelveLoc = “Bad”, 0 otherwise
    - Good\_Shelf = 1 if ShelveLoc = “Good”, 0 otherwise

Also, create two interaction variables:

* + - Price\_Bad\_Shelf = Price\* Bad\_Shelf
    - Price\_Good\_Shelf = Price\* Good\_Shelf

Question 1

Please estimate a linear regression model (using the lm function) with Sales as the dependent variable and Price as the independent variable. What is this model’s R-squared value?

0.198

0.196

0.100

0.050

Question 2

Use the model estimated in Question 1. What is this model’s adjusted R-squared value?

0.198

0.196

0.100

0.050

Question 3

Use the model estimated in Question 1. What is the estimated coefficient of Price?

-0.053073

13.641915

0.632812

0.005354

Question 4

Use the model estimated in Question 1. What is the t- value of the coefficient of Price?

-0.053073

0.632812

21.558

-9.912

Question 5

Use the model estimated in Question 1. Is the estimated coefficient for price statistically speaking different from 0 (at a significance level of 5%)?

Yes, the p-value is less than 0.05

No, the p-value is greater than 0.05

We don’t have enough info to answer this question

No, the coefficient is too close to zero (negligible)

Question 6

Please estimate a linear regression model (using the lm function) with Sales as the dependent variable and Price, Bad\_Shelf, and Good\_Shelf as independent variables. What is this model’s coefficient for the variable "Price"?

-0.053073

-1.862022

-0.056698

3.033825

Question 7

Use the model estimated in Question 6. What is this model’s coefficient for the variable “Bad\_Shelf”?

-0.053073

-1.862022

-0.056698

3.033825

Question 8

Use the model estimated in Question 6. What is this model’s coefficient for the variable “Good\_Shelf”?

-0.053073

-1.862022

-0.056698

3.033825

Question 9

Use the model estimated in Question 6. What is the average value of Sales when Price = 0 for ShelveLoc = “Medium”?

3.033825

16.8976

12.0018

13.86382

Question 10

Use the model estimated in Question 6. What is the average value of Sales when Price = 0 for ShelveLoc = “Bad”?

3.033825

16.8976

12.0018

13.86382

Question 11

Use the model estimated in Question 6. What is the average value of Sales when Price = 0 for ShelveLoc = “Good”?

3.033825

16.8976

12.0018

13.86382

Question 12

Please estimate a linear regression model (using the lm function) with Sales as the dependent variable and Price and ShelveLoc as independent variables. What is this model’s R-square value?

0.5391

0.5426

0.5620

0.5578

Question 13

For the model in Question 12, what is the average value of Sales when Price = 0 for ShelveLoc = “Bad”?

3.033825

16.8976

12.0018

13.86382

Question 14

For the model in Question 12, what is the average value of Sales when Price = 0 for ShelveLoc = “Medium”?

3.033825

16.8976

12.0018

13.86382

Question 15

For the model in Question 12, what is the average value of Sales when Price = 0 for ShelveLoc = “Good”?

3.033825

16.8976

12.0018

13.86382

Question 16

For the model in Question 12, what is the estimated Coefficient of Price?

-0.053073

-1.862022

-0.056698

3.033825

Question 17

Please estimate a linear regression model (using the lm function) with Sales as the dependent variable and Price, Bad\_Shelf, Good\_Shelf , Price\_Bad\_Shelf, and Price\_Good\_Shelf as independent variables. What is this model’s base case?

Bad\_Shelf

Good\_Shelf

Medium\_Shelf

Average of the three.

Question 18

For the model in Question 17, when price = 0, is the average sales for products in bad shelf significantly different compared to the base case?

Yes

No

Maybe

Not enough info.

Question 19

For the model in Question 17, for a unit change in price, does the average sales differ significantly between products in good shelf compared to the base case?

Yes

No

Maybe

Not enough info.

**PART 2**

Download the [EDSAL.csv](https://prod-edxapp.edx-cdn.org/assets/courseware/v1/670b2876f877485d5dc8e47bd5b72ec8/asset-v1:GTx+MGT6203x+1T2018+type@asset+block/EDSAL__1_.csv) file and upload it to a dataframe (in R). The three variables are Education, Experience and Salary.

Run 4 linear regressions using the lm function in R. (note – you have to use the natural log; the default 'log()' function in R)

* + - Linlin: Use Salary as the dependent variable and Experience as the independent variable.
    - Linlog: Use Salary as the dependent variable and log(Experience) as the independent variable.
    - Loglin: Use log(Salary) as the dependent variable and Experience as the independent variable.
    - Loglog: Use log(Salary) as the dependent variable and log(Experience) as the independent variable.

Question 20

Which of the 4 fitted models has the lowest R-square value?

Loglog

Loglin

Linlin

Linlog

Question 21

In all the models run in Part 2, which of the 4 fitted models has the highest R-square value?

Loglog

Loglin

Linlin

Linlog

Question 22

What is the interpretation of the slope coefficient for the Linlin model estimated in Part 2?

Increasing Experience by 1 unit leads to 3.0959 units increase in Salary

Increasing Experience by 1% leads to 0.01\*3.0959 units increase in Salary

Increasing Experience by 1 unit leads to (3.0959\*100) % increase in Salary

Increasing Experience by 1% leads to 3.0959% increase in Salary

Question 23

What is the interpretation of the slope coefficient for the Linlog model estimated in Part 2?

Increasing Experience by 1 unit leads to 34.985 units increase in Salary

Increasing Experience by 1% leads to 0.01\*34.985 units increase in Salary

Increasing Experience by 1 unit leads to (34.985 \*100)% increase in Salary

Increasing Experience by 1% leads to 34.985% increase in Salary

Question 24

What is the interpretation of the slope coefficient for the Loglin model estimated in Part 2?

Increasing Experience by 1 unit leads to 0.037087 units increase in Salary

Increasing Experience by 1% leads to 0.01\*0.037087 units increase in Salary

Increasing Experience by 1 unit leads to (0.037087 \*100) % increase in Salary

Increasing Experience by 1% leads to 0.037087% increase in Salary

Question 25

What is the interpretation of the slope coefficient for the Loglog model estimated in Part 2?

Increasing Experience by 1 unit leads to 0.45949 units increase in Salary

Increasing Experience by 1% leads to 0.01\*0.45949 units increase in Salary

Increasing Experience by 1 unit leads to (e^0.45949-1) \*100% increase in Salary

Increasing Experience by 1% leads to 0.45949% increase in Salary