**Multiple Regression Using R**

* Collecting the data
* Capturing the data in R
* Checking for linearity
* Applying the multiple linear regression model
* Making a prediction

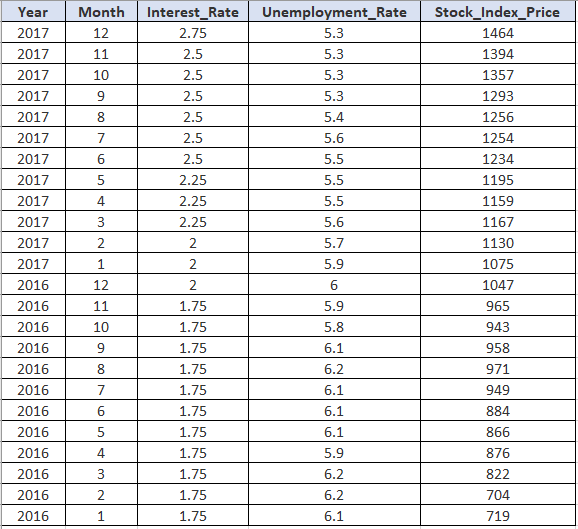
**Steps to apply the multiple linear regression in R**

**Step 1: Collect the data**

So let’s start with a simple example where the goal is to predict the stock\_index\_price (the dependent variable) of a fictitious economy based on two independent/input variables:

* Interest\_Rate
* Unemployment\_Rate

Here is the data to be used for our example:



### Step 2: Capture the data in R

Next, you’ll need to capture the above data in R. The following code can be used to accomplish this task:

Year <- c(2017,2017,2017,2017,2017,2017,2017,2017,2017,2017,2017,2017,2016,2016,2016,2016,2016,2016,2016,2016,2016,2016,2016,2016)

Month <- c(12, 11,10,9,8,7,6,5,4,3,2,1,12,11,10,9,8,7,6,5,4,3,2,1)

Interest\_Rate <- c(2.75,2.5,2.5,2.5,2.5,2.5,2.5,2.25,2.25,2.25,2,2,2,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75)

Unemployment\_Rate <- c(5.3,5.3,5.3,5.3,5.4,5.6,5.5,5.5,5.5,5.6,5.7,5.9,6,5.9,5.8,6.1,6.2,6.1,6.1,6.1,5.9,6.2,6.2,6.1)

Stock\_Index\_Price <- c(1464,1394,1357,1293,1256,1254,1234,1195,1159,1167,1130,1075,1047,965,943,958,971,949,884,866,876,822,704,719)

**Step 3: Check for linearity**

Before you apply linear regression models, you’ll need to verify that several assumptions are met. Most notably, you’ll need to make sure that a linear relationship exists between the dependent variable and the independent variable/s.

A quick way to check for linearity is by using scatter plots.

For our example, we’ll check that a linear relationship exists between:

* The Stock\_Index\_Price (dependent variable) and the Interest\_Rate (independent variable); and
* The Stock\_Index\_Price (dependent variable) and the Unemployment\_Rate (independent variable)

Here is the code that can be used in R to plot the relationship between the Stock\_Index\_Price and the Interest\_Rate:

Year <- c(2017,2017,2017,2017,2017,2017,2017,2017,2017,2017,2017,2017,2016,2016,2016,2016,2016,2016,2016,2016,2016,2016,2016,2016)

Month <- c(12, 11,10,9,8,7,6,5,4,3,2,1,12,11,10,9,8,7,6,5,4,3,2,1)

Interest\_Rate <- c(2.75,2.5,2.5,2.5,2.5,2.5,2.5,2.25,2.25,2.25,2,2,2,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75)

Unemployment\_Rate <- c(5.3,5.3,5.3,5.3,5.4,5.6,5.5,5.5,5.5,5.6,5.7,5.9,6,5.9,5.8,6.1,6.2,6.1,6.1,6.1,5.9,6.2,6.2,6.1)

Stock\_Index\_Price <- c(1464,1394,1357,1293,1256,1254,1234,1195,1159,1167,1130,1075,1047,965,943,958,971,949,884,866,876,822,704,719)

plot(x=Interest\_Rate, y=Stock\_Index\_Price)

And for the second case, you can use the code below in order to plot the relationship between the Stock\_Index\_Price and the Unemployment\_Rate:

Year <- c(2017,2017,2017,2017,2017,2017,2017,2017,2017,2017,2017,2017,2016,2016,2016,2016,2016,2016,2016,2016,2016,2016,2016,2016)

Month <- c(12, 11,10,9,8,7,6,5,4,3,2,1,12,11,10,9,8,7,6,5,4,3,2,1)

Interest\_Rate <- c(2.75,2.5,2.5,2.5,2.5,2.5,2.5,2.25,2.25,2.25,2,2,2,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75)

Unemployment\_Rate <- c(5.3,5.3,5.3,5.3,5.4,5.6,5.5,5.5,5.5,5.6,5.7,5.9,6,5.9,5.8,6.1,6.2,6.1,6.1,6.1,5.9,6.2,6.2,6.1)

Stock\_Index\_Price <- c(1464,1394,1357,1293,1256,1254,1234,1195,1159,1167,1130,1075,1047,965,943,958,971,949,884,866,876,822,704,719)

plot(x=Unemployment\_Rate, y=Stock\_Index\_Price)

### Step 4: Apply the multiple linear regression in R

You may now use the following template to perform the multiple linear regression in R:

model <- lm(Dependent variable ~ First independent Variable + Second independent variable + ...)

summary(model)

Using the template for our example:

Year <- c(2017,2017,2017,2017,2017,2017,2017,2017,2017,2017,2017,2017,2016,2016,2016,2016,2016,2016,2016,2016,2016,2016,2016,2016)

Month <- c(12, 11,10,9,8,7,6,5,4,3,2,1,12,11,10,9,8,7,6,5,4,3,2,1)

Interest\_Rate <- c(2.75,2.5,2.5,2.5,2.5,2.5,2.5,2.25,2.25,2.25,2,2,2,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75)

Unemployment\_Rate <- c(5.3,5.3,5.3,5.3,5.4,5.6,5.5,5.5,5.5,5.6,5.7,5.9,6,5.9,5.8,6.1,6.2,6.1,6.1,6.1,5.9,6.2,6.2,6.1)

Stock\_Index\_Price <- c(1464,1394,1357,1293,1256,1254,1234,1195,1159,1167,1130,1075,1047,965,943,958,971,949,884,866,876,822,704,719)

model <- lm(Stock\_Index\_Price ~ Interest\_Rate + Unemployment\_Rate)

summary(model)

You can use the coefficients in the summary in order to build the multiple linear regression equation as follows:

Stock\_Index\_Price = (Intercept) + (Interest\_Rate coef)\*X1  (Unemployment\_Rate coef)\*X2

And once you plug the numbers from the summary:

Stock\_Index\_Price = (1798.4) + (345.5)\*X1 + (-250.1)\*X2

In the next section, we’ll see how to use this equation to make predictions.

**Step 5: Make a prediction**

Now let’s make a prediction based on the equation above.

For example, imagine that you want to predict the stock index price after you collected the following data:

* Interest Rate = 1.5 (i.e., X1= 1.5)
* Unemployment Rate = 5.8 (i.e., X2= 5.8)

And if you plug that data into the regression equation you’ll get:

Stock\_Index\_Price = (1798.4) + (345.5)\*(1.5) + (-250.1)\*(5.8) = 866.07

The predicted value for the Stock\_Index\_Price is therefore 866.07.

Some additional stats to consider in the summary:

1. **Adjusted R-squared** reflects the fit of the model, where a higher value generally indicates a better fit
2. **Intercept coefficient** is the Y-intercept
3. **Interest\_Rate coefficient** is the change in Y due to a change of one unit in the interest rate (everything else held constant)
4. **Unemployment\_Rate coefficient** is the change in Y due to a change of one unit in the unemployment rate (everything else held constant)
5. **Std. Error**reflects the level of accuracy of the coefficients
6. **Pr(>|t|)** is the *p-value*. A p-value of less than 0.05 is considered to be statistically significant