

**Name:** Ankit Sharma

**Email:** [kumarankitx022@gmail.com](mailto:kumarankitx022@gmail.com)

**Mob no.:** +91-7677241423

**Week:** 03

## **ASSIGNMENT – 06**

### **TOPIC: MULTIPLE LINEAR REGRESSION (With Backward Elimination)**

Q.1. Automatic Backward Elimination.

**Ans. Backward elimination** is a feature selection technique while building a machine learning model. It is used to remove those features that do not have a significant effect on the dependent variable or prediction of output.

#### **STEPS FOR BACKWARD ELIMINATION:**

**Step-1:** Firstly, We need to select a significance level to stay in the model. (SL=0.05)

**Step-2:** Fit the complete model with all possible predictors/independent variables.

**Step-3:** Choose the predictor which has the highest P-value, such that.

- a. If P-value > SL, go to step 4.
- b. Else Finish, and Our model is ready.

**Step-4:** Remove that predictor.

**Step-5:** Rebuild and fit the model with the remaining variables.

In [5]:

```

# Creating automated backward elimination function with p-values
import statsmodels.api as sm
def backwardElimination(x, SL):
    numVars = len(x[0])
    temp = np.zeros((50,6)).astype(int)
    for i in range(0, numVars):
        regressor_OLS = sm.OLS(Y, x).fit()
        maxVar = max(regressor_OLS.pvalues).astype(float)
        if maxVar > SL:
            for j in range(0, numVars - i):
                if (regressor_OLS.pvalues[j].astype(float) == maxVar):
                    temp[:,j] = x[:, j]
                    x = np.delete(x, j, 1)
                    tmp_regressor = sm.OLS(Y, x).fit()
            regressor_OLS.summary()
    return x

# Applying the backward elimination

X = np.append(arr = np.ones((50, 1)).astype(int), values = X, axis = 1)
SL = 0.05
X_opt = X[:,[0,1,2,3,4,5]]
X_opt = np.array(X_opt, dtype=float)

X_Res = backwardElimination(X_opt, SL)
regressor_OLS = sm.OLS(endog = Y, exog = X_Res).fit()
regressor_OLS.summary()

```

Out[5]:

OLS Regression Results

<b>Dep. Variable:</b>	y	<b>R-squared:</b>	0.947
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.945
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	849.8
<b>Date:</b>	Tue, 19 May 2020	<b>Prob (F-statistic):</b>	3.50e-32
<b>Time:</b>	16:42:53	<b>Log-Likelihood:</b>	-527.44
<b>No. Observations:</b>	50	<b>AIC:</b>	1059.
<b>Df Residuals:</b>	48	<b>BIC:</b>	1063.
<b>Df Model:</b>	1		
<b>Covariance Type:</b>	nonrobust		

  

	coef	std err	t	P> t	[0.025	0.975]
const	4.903e+04	2537.897	19.320	0.000	4.39e+04	5.41e+04
x1	0.8543	0.029	29.151	0.000	0.795	0.913

  

<b>Omnibus:</b>	13.727	<b>Durbin-Watson:</b>	1.116
<b>Prob(Omnibus):</b>	0.001	<b>Jarque-Bera (JB):</b>	18.536
<b>Skew:</b>	-0.911	<b>Prob(JB):</b>	9.44e-05
<b>Kurtosis:</b>	5.361	<b>Cond. No.</b>	1.65e+05