

Simple Linear Regression

- In R linear regression uses the `lm()` function to create a regression model given some formula, in the form of **$Y \sim X + X2 + X3 + X4$** (X is variables)
- `summary()` functions to look the model and it's parameters such as formula, coefficients, standard error, residual, multiple/adjusted R-Square..etc to analyze regression model
- `predict()` function used to make a prediction on new data, and we can derived formula for prediction $y = b_0 + b_1x_1 + b_2x_2 + b_3x_3..etc$

In Dataset, 12 months of monthly spend and sales

- Month & Spend are independent variable
- Sales is dependent variable, so we need to predict the Sales using Spend & Month

[Download Dataset](#)

Month	Spend	Sales
1	1000	9914
2	4000	40487
3	5000	54324
4	4500	50044
5	3000	34719
6	4000	42551
7	9000	94871
8	11000	118914
9	15000	158484
10	12000	131348
11	7000	78504
12	3000	36284

```

In [ ]: data <- read.csv("marketing-spend.csv")

In [ ]: # Check the correlation between variable
cor(data$Spend,data$Sales) # Default method is pearson correlation

In [ ]: # Check the correlation between variable, We can change correlation method based on ou
# Methods: "pearson", "kendall", "spearman"
cor(data$Spend,data$Sales, method = "spearman")

In [ ]: # Display the Scatter plot between Spend vs Sales
plot(data$Spend,data$Sales,
      main="Spend vs Sales")

In [ ]: # Prepare Simple Regression Model

# Parameters for lm() function
# Dependent variable
# Independent Variable
# Data Source

model_1 <- lm(Sales~Spend, data)

In [ ]: # Quick Formula & Coefficients
model_1

In [ ]: #Coefficients
# B0 : 1383.47
# B1 : 10.62

model_1$coefficients

In [ ]: # Details Summary of Model
summary(model_1)

```

- Summary Display the Formula for the Regression Model
- Descriptive Summary for Residuals (Actual_Sales - Predicted_Sales)
- Coefficients for the Model
- Significance code for the Variables
- Residual Standard Error
- R-Squares : R-squared (R²) is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model
- Example : An R-squared of 100% means that all movements of a dependent variables are completely explained by independent variable(s) you are interested in. if the R-squared of a model is 0.50, then approximately half of the observed variation can be explained by the model's inputs.

```
In [ ]: # Predict on new data
Spend <- c(1200, 2500)
new_data <- data.frame(Spend)
```

```
In [ ]: predict(model_1, new_data)
```

Preparing Formula for quick prediction

```
In [ ]: b0 <- model_1$coefficients['(Intercept)']
b1 <- model_1$coefficients['Spend']
```

```
In [ ]: #  $Y = b_0 + b_1 * spend$ 
spend = 10000
b0 + (b1 * spend)
```