



# Linear Regression

ISMAIL SEVER

# Table of Content

- Description of Analysis
- Summary of Analysis
- Increasing the Accuracy
- R script

# Description of Analysis

- Linear regression is used to **predict** the value of an **outcome variable Y** based on one or more input **predictor variables X**.
- We have a dataset called `ols_stock`. We will import this dataset into our RStudio and build a linear model to check whether the predictor “`dividend`” is significantly associated with outcome variable `stock_return_scaled`.
- **Null Hypothesis:**  $\beta=0$ , co-efficient  $\beta$  of the predictor is zero and not statistically significant
- **Alternative Hypothesis:**  $\beta \neq 0$ , co-efficient  $\beta$  of the predictor is not equal to zero and is statistically significant

# Summary of Analysis

```
Call:
lm(formula = stock_return_scaled ~ dividend, data = ols_stock)

Residuals:
    Min       1Q   Median       3Q      Max
-174.38  -71.47  -36.62   26.19   779.78

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   204.76     29.29   6.991 8.43e-09 ***
dividend     -12.97     41.85  -0.310  0.758
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 146.4 on 47 degrees of freedom
Multiple R-squared:  0.002041, Adjusted R-squared:  -0.01919
F-statistic: 0.09611 on 1 and 47 DF,  p-value: 0.7579
```

The equation:

$$(\text{stock\_return\_scaled}) = 204.76 - 12.97(\text{dividend})$$

## Part I (Error)

The median is -36.62. The difference between median and other values are not huge, so it is not a normal distribution.

## Part II (Model Outcome)

Since p-value is 0.758 which is bigger than our significance value 0.05, we accept the **null hypothesis** that the predictor variable is not significantly associated with outcome variable.

## Part III (Model Performance)

R-squared value is 0.002041 which is very small. This means the model doesn't fit very well the data. Higher is the R-squared value, better is the model.

# Increasing the Accuracy

I would like to include another two predictors such as “**marketcap**” and “**stock\_return**” to increase the accuracy of my model, because I believe that there is a statistically significant relationship between the predictors “marketcap” and “stock\_return” and the outcome variable “stock\_return\_scaled”. Moreover, the model that we built will fit very well the data.

# R Script

```
#Build Linear Model
```

```
simple.fit<-lm(stock_return_scaled~dividend, data=ols_stock)
```

```
LinearModel<-simple.fit
```

```
#Summary of Key Statistics of the Model
```

```
summary(LinearModel)
```



Thank You