/02 Linear Regression with One Regressor

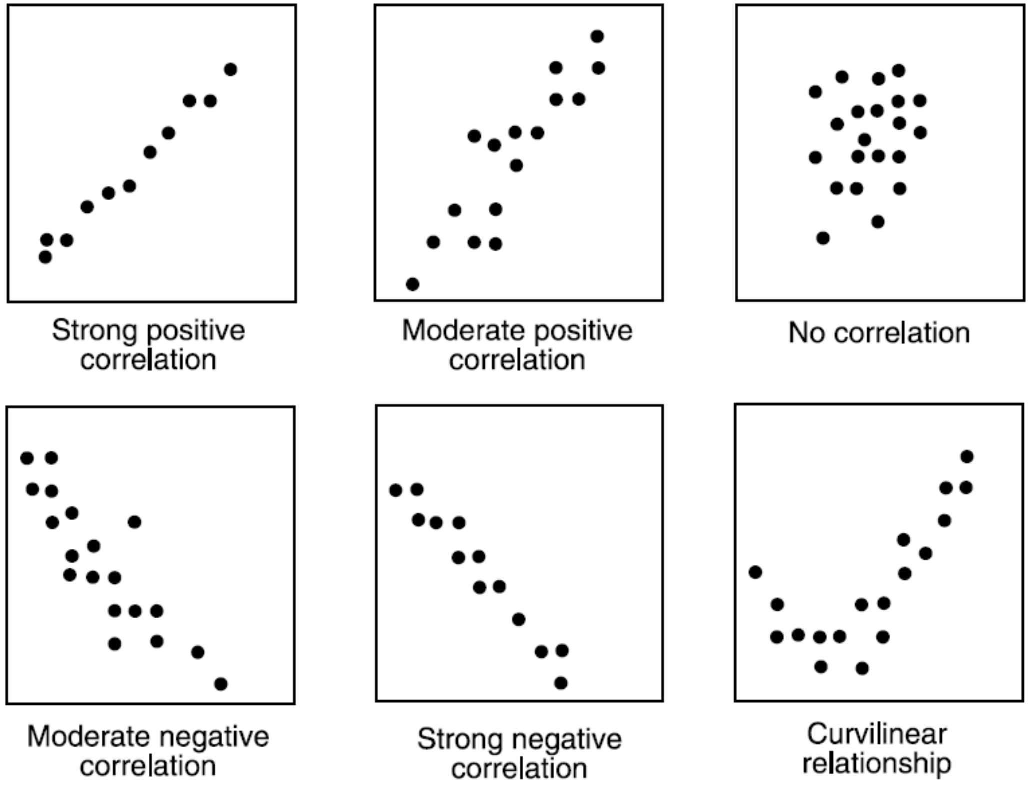
Linear Regression with One Regressor

➢ REGRESSION ANALYSIS:

Lets us estimate the slope of the population regression line.

➢The slope of the population regression line is the expected effect on *Y of a unit change in X.*

➢Ultimately our aim is to estimate the causal effect on *Y of* a unit change in *X – for that we prepare a scatter plot, which indicates the nature of relationship between the dependent and independent variable.*



Linear Regression with One Regressor

- **POPULATION REGRESSION FUNCTION:**

$$E(Y_i \mid X_i) = B_0 \times B_1 \times (X_i)$$

Slope coefficient

Explained variable

Intercept coefficient

Explanatory variable

Linear Regression with One Regressor

- **POPULATION REGRESSION FUNCTION:**

The Error Term

- The difference between each Y and its corresponding conditional expectation (i.e., the line that fits the data) is the error term or noise component denoted ε_1.

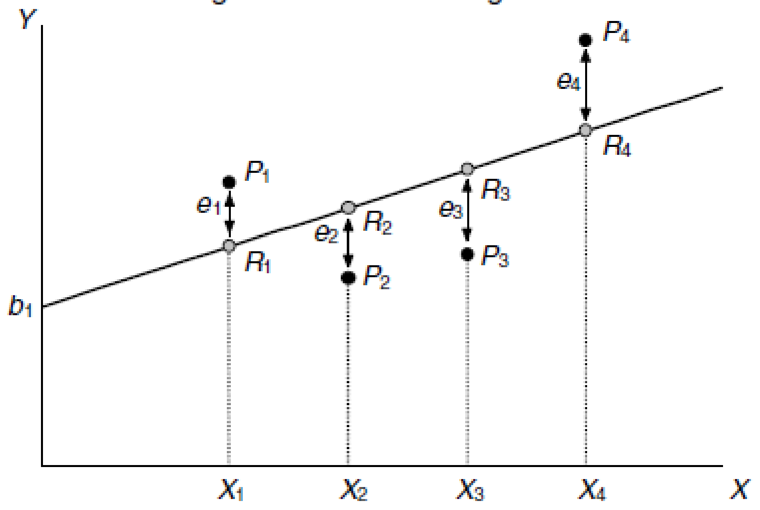
$$\varepsilon_i = Y_i - E(Y_i \mid X_i)$$

- The error term represents effect from independent variables not included in the model.

- The population regression function can also be expressed as:

$$Y_i = B_0 + B_1 \times X_i + \varepsilon_i$$

Fitted regression line showing residuals



Linear Regression with One Regressor

- **SAMPLE REGRESSION FUNCTION:**

  - The sample regression function is an equation that represents a relationship between the Y and X variable(s) that is based on the information in a sample of the population.

$$Y_i = b_0 + b_1 \times X_i + e_i$$

Linear Regression with One Regressor

➢ PROPERTIES OF REGRESSION:

➢ **Linear regression** makes several key assumptions:

- Linear relationship
- Multivariate normality
- No or little multicollinearity
- No auto-correlation
- Homoscedasticity

➢ If the relationship between the dependent variable and an independent variable is non-linear, then an analyst would do that transformation first and then enter the transformed value into the linear equation.

Linear Regression with One Regressor

➢ ORDINARY LEAST SQUARES REGRESSION:

Ordinary least squares (OLS) estimation is a process that estimates the population parameters B_i with corresponding values for b_i that minimize the squared residuals (i. e., error terms).

The OLS sample coefficients are those that:

$$\text{minimize} \sum e_i^2 = \sum [Y_i - (b_0 + b_1 x X_i)]^2$$

The estimated slope coefficient (b_1) : $b_1 = \frac{\text{cov}(X, Y)}{\text{Var}(X)}$

& the intercept term (b_0) : $b_0 = \overline{Y} - b_1 \overline{X}$

Linear Regression with One Regressor

➢ Assumptions Underlying Linear Regression:

❖The three key assumptions are:

- The expected value of the error term, conditional on the independent variable, is zero ($E(\varepsilon_i|X_i) = 0$).
- All (X,Y) observations are independent and identically distributed (i.i.d.).
- There is no large outliers will be observed in the data.

❖Additional assumptions include:

- A linear relationship exists between the dependent and independent variable.
- The model is correctly specified in that is includes the appropriate independent variable and does not omit variables.
- The independent variable is uncorrelated with the error terms.
- The variance of ε_i is constants for all X_i: $Var(\varepsilon_i|X_i) = \sigma^2$
- No serial correlation of the error terms exits [i.e., Corr $(\varepsilon_i,\varepsilon_{i+j}) = 0$ for j=1,2,3…].
- The error term is normally distributed.

Linear Regression with One Regressor

Properties of OLS Estimators:

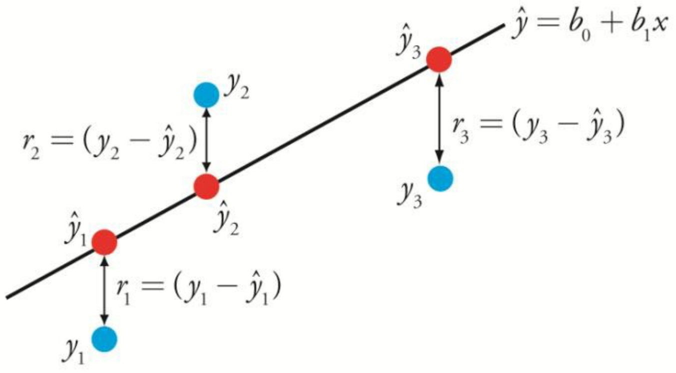
Three key properties are:

- Unbiased
- Consistent
- Efficient

Linear Regression with One Regressor

OLS REGRESSION RESULTS:

The sum of squared residuals(SSR), is the sum of squares that results from placing a given intercept and slope coefficient in to the equation and computing the residuals, squaring the residuals and summing them.
It is an indicator of how well the sample regression function explains the data.

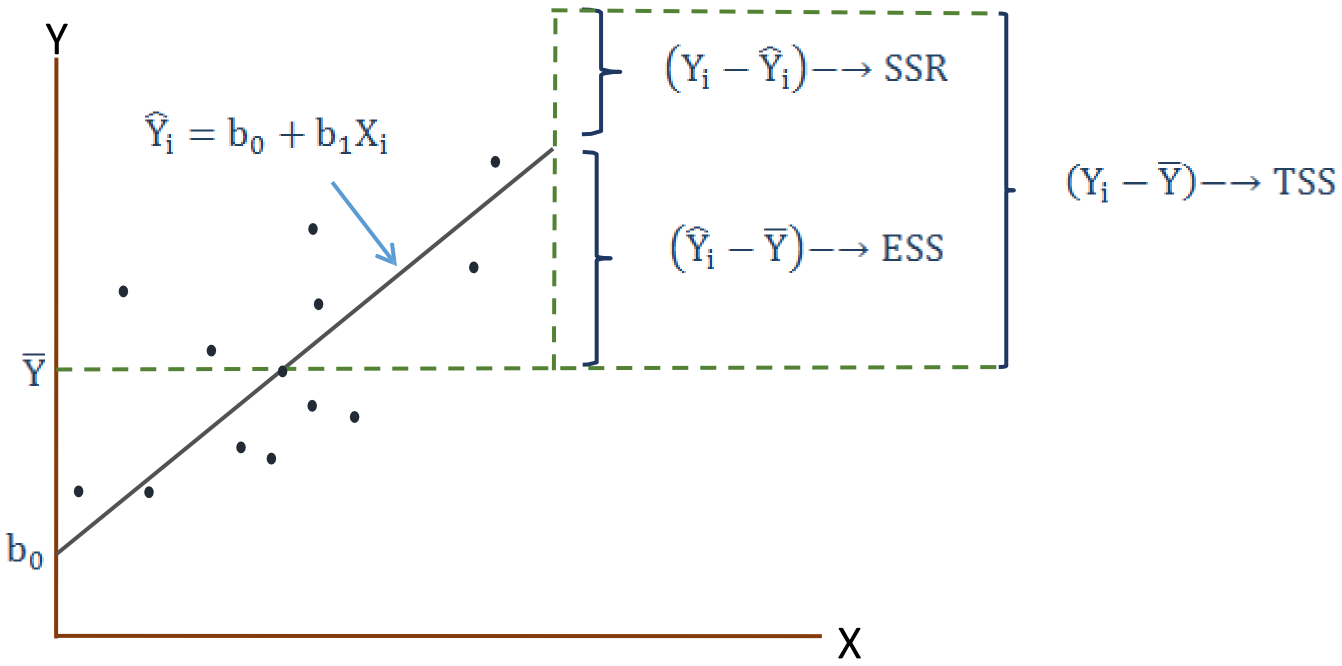


Example

Linear Regression with One Regressor

➢ OLS REGRESSION RESULTS:

The **coefficient of determination**, represented by R^2, is a measure of the "goodness of fit" of the regression. It is interpreted as a percentage of variation in the dependent variable explained by the interdependent variable.



TSS ESS SSR

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y} - \bar{Y})^2 + \sum (Y_i - \hat{Y})^2$$

The coefficient of determination can be calculated as follows:

$$R^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$$

$$R^2 = 1 - \frac{SSR}{TSS} = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2}$$

Note: Sum of squared residuals (SSR) is also known as the sum of squared errors (SSR). In the same regard, total sum of squares (TSS) is also known as sum of squares total (SST), and explained sum of squares (ESS) is also known as regression sum of squares (RSS).

Linear Regression with One Regressor

- **OLS REGRESSION RESULTS:**

In a simple two-variable regression, the square root of R^2 is the correlation coefficient(r) between X_i and Y_i .
if the relationship is positive, then: r = $\sqrt{R^2}$

<u>Difference between correlation coefficient & coefficient of determination:</u>

- the correlation coefficient indicates the sign of the relationship, whereas the coefficient of determination does not.
- the coefficient of determination can apply to an equation with several independent variables, and it implies a causation or explanatory power, while the correlation coefficient only applies to two variables and does not imply causation between the variables.

Linear Regression with One Regressor

- **OLS REGRESSION RESULTS:**

The Standard Error of the Regression

- The standard error of the regression (SER) measures the degree of variability of the actual Y-values relative to the estimated Y-values from a regression equation.
- The SER gauges the “fit” of the regression line.
- The SER is the standard deviation of the error terms in regression.
- SER is also referred to as the standard error of the residual, or the standard error of estimate (SEE).

Thanks
Leading Derivative & Risk Advisors

成蹊风险研究

riskmacro.com