$$y = \beta_0 + \beta_1 x + \varepsilon$$

y → dependent Variable

x → Independent Variable

β_0 = intercept (value of y when $x=0$)

β_1 = slope (change in y for unit change in x)

Assumptions

1. $y|x$ is normally distributed
   ↳ y given x is normally distributed
2. observations need to be independent.
3. $E(\varepsilon) = 0$
4. ε and x (independent Variable) is not correlated.

Goodness of Fit

R^2 (R-Square) (or Coefficient of Determination.

$$R^2 = \frac{SSR}{SST} = \frac{\text{Sum of Squared Residuals}}{\text{Sum of Squared total}}$$

= total variation explained by the model.

$0 \le R^2 \le 1$ (or $0 \le R^2 \le 100\%$

$$wage = \beta_0 + \beta_1(educ) + \varepsilon$$

wage = annual salary

educ = # of years of education.

β_1 = effect of education on wage while all other factors are held constant

ε — includes all other factors that affect wage

$$wage = \beta_0 + \beta_1(educ) + \beta_2(experience) + \beta_3(skills) + \varepsilon$$

β_0 — intercept (wage when all other $x = 0$)

β_2 = effect of experience on wage while all other factors are fixed/held constant

β_3 = effect of skills on wage when all others are held constant

In MLR (Multiple linear regression) All x's (i.e independent variables) should not be correlated.

$$wage = \beta_0 + \beta_1(\overbrace{\underline{education}}^{x_1}) + \beta_2(\underbrace{\underline{experience}}_{x_2}) + \beta_3\underbrace{\boxed{(experience + education)}}_{x_3}$$

$$x_3 = x_2 + x_1$$

Consumption and income

$$Consumption = \beta_0 + \beta_1(income) + \beta_2(income)^2 + \varepsilon$$

$$x_1 = income$$

$$x_2 = income^2$$

$$\frac{\Delta\, cons}{\Delta\, inc} = \beta_1 + 2\beta_2(income)$$

Generalized form of MLR.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

β_0 – intercept (Value of Y when $x_i = 0$)

$\beta_1 \cdots \beta_k$ – Slope parameters for $x_1 \cdots x_k$

ε – error term (accounts for all the factors that are not included in the model).

Assumptions

1. y given x_i is normally distributed.
2. Observation are independent (random sample)

3. No perfect Collinearity

- None of the independent variables are constant
- No linear relationship among any of the independent variables.

Sales(Y)	Advertising (X)
5	3
10	3
15	3
20	3
25	3

Not

4. $E(\varepsilon \mid x_1\ x_2 \cdots x_k) = 0$

5. Homoskedasticity

$$Var(\varepsilon \mid x_1\ x_2 \cdots x_k) = \sigma^2$$

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots$

$\beta_k x_k$

$H_0 : \beta_1 = 0 \longrightarrow$

If x_1 has an effect on y

If $p < 0.05$ reject H_0

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$

x_1 and x_4 are highly correlated

$$\text{College GPA} = \beta_0 + \beta_1 (\text{hsGPA}) + \beta_2 (\text{Skipped classes}) + \beta_3 (\text{business major}) + \beta_4 (\text{clubs}) + \varepsilon,$$

$\beta_3 = 0.094$

business major $= 1$ if in business school
$= 0$ otherwise.

→ Compared to an engg major a business major will have a 0.094 times effect on Col GPA.