**Regression and Correlation**

**Regression:** It is the dependence of one variable (dependent variable) on one or more other variables (independent variables).

**Simple regression:** The dependence of a variable on single variable is called simple regression.

**Multiple regression:** The dependence of a variable on more than one variable is called simple regression.

**Regression line Y on X**

$$Y=a+bx$$

**Where, Y= dependent variable, X= independent variable, a= intercept, b=slope (rate of change)**

$$b=\frac{n\sum\_{}^{}XY-\sum\_{}^{}X\sum\_{}^{}Y}{n\sum\_{}^{}X^{2}-\left(\sum\_{}^{}X\right)^{2}}$$

$$a=\overbar{y}-b\overbar{x}$$

**Correlation:** It is a linear relationship (association) between two variables. It is denoted by r, ρ

**Positive Correlation:** If both variables tend to increase or decrease together, the correlation is said to be positive correlation.

**Negative correlation:** If one variable tend to increase as the other variable decreases the correlation is said to be negative correlation.

$$r=\frac{n\sum\_{}^{}XY-\sum\_{}^{}X\sum\_{}^{}Y}{\sqrt{\left[n\sum\_{}^{}X^{2}-\left(\sum\_{}^{}X\right)^{2}\right]\left[n\sum\_{}^{}Y^{2}-\left(\sum\_{}^{}y\right)^{2}\right]}}$$

**Range:** $-1\leq r\leq +1$

**Example:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Statistics (x)** | **Economics(Y)** | **XY** | $$X^{2}$$ | $$Y^{2}$$ |
| **64** | **57** | **3648** | **4096** | **3249** |
| **71** | **59** | **4189** | **5041** | **3481** |
| **53** | **49** |  |  |  |
| **67** | **62** |  |  |  |
| **55** | **51** |  |  |  |
| **58** | **50** |  |  |  |
| **77** | **55** |  |  |  |
| **57** | **48** |  |  |  |
| **56** | **52** |  |  |  |
| **51** | **42** |  |  |  |
| **76** | **61** |  |  |  |
| **68** | **57** |  |  |  |
| **Sum= 753** | **Sum=643** | **Sum=40830** | **Sum=48139** | **Sum=34843** |

$$Y=a+bx$$

$$b=\frac{n\sum\_{}^{}XY-\sum\_{}^{}X\sum\_{}^{}Y}{n\sum\_{}^{}X^{2}-\left(\sum\_{}^{}X\right)^{2}}$$

$b=\frac{12(40830)-(753)(643)}{12\left(48139\right)-(753)^{2}}$**= 0.54**

$$a=\overbar{y}-b\overbar{x}$$

$$a=53.58-\left(0.54\right)\left(62.75\right)=19.54$$

$$Y=a+bx$$

$$Y=19.54+0.54x$$

$$r=\frac{n\sum\_{}^{}XY-\sum\_{}^{}X\sum\_{}^{}Y}{\sqrt{\left[n\sum\_{}^{}X^{2}-\left(\sum\_{}^{}X\right)^{2}\right]\left[n\sum\_{}^{}Y^{2}-\left(\sum\_{}^{}y\right)^{2}\right]}}$$