CORRELATION COEFFICIENT

Measure of closeness of linear relationship between 2 variables—*ρ* for population; *r* for sample

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

or

$$r = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sqrt{\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right]\left[\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right]}} = \frac{L_{xy}}{\sqrt{L_{xx} * L_{yy}}}$$

Quantitative measure of the dependence between 2 variables—best if both variables are normally distributed

Example: A Patient Satisfaction Test has been in use in hospitals. We want to evaluate a new Short Form of the test. Both were administered to a sample of patients. The following results were obtained:

Patient	Short Form (X)	Standard (Y)
1	40	45
2	55	61
3	60	59
4	65	82
5	70	80
6	73	76
7	80	90
8	85	102
9	90	98
10	97	100
11	100	114
12	105	111
13	111	115

Summary Statistics:

$$\sum X_i = 1031$$
$$\sum X_i^2 = 87{,}159$$
$$\sum Y_i = 1133$$
$$\sum Y_i^2 = 104{,}777$$
$$\sum X_i Y_i = 95{,}383$$

$$r = \frac{95{,}383 - \frac{1031*1133}{13}}{\sqrt{\left(87{,}159 - \frac{1031^2}{13}\right)\left(104{,}777 - \frac{1133^2}{13}\right)}}$$
$$= \frac{5527.3846}{\sqrt{5392.7693*6031.6923}}$$
$$= \frac{5527.3846}{5703.2907}$$
$$= 0.969$$