

Sample correlation coefficient, r

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{([n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2])}}$$

For the data, take values of men to be x and values of women to be y

SEX LIFE EXPECTANCY

Men(x)	59.7	72.9	41.9	46.2	50.3	43.2
Women(y)	63.8	77.8	44.5	48.3	54.0	43.5

Table of calculations, n=6 (count no. of columns)

X	Y	XY	X^2	Y^2
59.7	63.8	3808.86	3564.09	4070.44
72.9	77.8	5671.62	5314.41	6052.84
41.9	44.5	1864.55	1755.61	1980.25
46.3	48.3	2231.46	2134.44	2332.89
50.3	54.0	2716.20	2530.09	2916.00
43.2	43.5	1879.20	1866.24	1892.25
$\sum X=314.2$	$\sum Y=331.9$	$\sum XY=18171.89$	$\sum X^2=17164.88$	$\sum Y^2=19244.67$

$$r = \frac{6(18171.89) - (314.2)(331.9)}{\sqrt{([6(17164.88) - 314.2^2][6(19244.67) - 331.9^2])}}$$

$$r = \frac{4748.36}{\sqrt{22662918.13}}$$

$$r = \frac{4748.36}{4760.56}$$

$$r = 0.997 (+ve)$$

The correlation coefficient is Pearson's coefficient of relation, usually denoted by r.

Interpretation of Pearson's correlation coefficient; The sign of the correlation determines whether the correlation is positive or negative. In our case, $r=0.997(+ve)$ hence the correlation is positive. The magnitude of the correlation coefficient determines the strength of the correlation, 0.997 is close to one hence a relatively strong correlation. The r value should not exceed 1 hence our calculation is within the range of Pearson's coefficient of correlation.

RE-CALCULATION BY CALCULATOR

$$\sum X = 314.2 \quad \sum Y = 331.9 \quad \sum XY = 18171.89 \quad \sum X^2 = 17164.88 \quad \sum Y^2 = 19244.67$$

$$r = \frac{6(18171.89) - (314.2)(331.9)}{\sqrt{([6(17164.88) - 314.2^2][6(19244.67) - 331.9^2])}}$$

$$r = \frac{4748.36}{\sqrt{22662918.13}}$$

$$r = \frac{4748.36}{4760.56}$$

$$r = 0.997 (+ve)$$

1.The correlation coefficient is Pearson's correlation coefficient, usually denoted by r.

2.

x	Y
41.9	44.5
43.2	43.5
46.2	48.3
50.3	54.0

The difference in the value of x on y is not uniform and there is no uniform equation relating values of x to y hence the relation is non-linear.

3.Regression line Y on X: $(Y - \bar{Y}) = b_{yx}(X - \bar{X}) \dots (i)$

Regression line X on Y: $(X - \bar{X}) = b_{xy}(Y - \bar{Y}) \dots (ii)$

NOW FOR OUR DATA:

SEX

LIFE EXPECTANCY

Men(x)	59.7	72.9	41.9	46.2	50.3	43.2
Women(y)	63.8	77.8	44.5	48.3	54.0	43.5

Where

$$n=6 \quad \sum x = 314.2 \quad \sum y = 331.9 \quad \sum x^2 = 17,164.88 \quad \sum y^2 = 19,244.67 \quad \sum XY = 18171.89$$

SOLUTION

$$\bar{X} = \frac{1}{n} \sum X = \frac{314.2}{6} = 52.36667 \quad \bar{Y} = \frac{1}{n} \sum Y = \frac{331.9}{6} = 55.31667$$

$$Cov(X, Y) = \frac{1}{n} \sum XY - \bar{X} \bar{Y} = \frac{18171.89}{6} - (52.36667 \times 55.31667)$$

$$= 3028.648333 - 2896.749803$$

$$Cov(X, Y) = 131.8985$$

$$Var(X) = \frac{1}{n} \sum X^2 - \bar{X}^2 = \frac{17164.88}{6} - 52.36667^2 = 2860.81333 - 2742.268127$$

$$Var X = 118.5452$$

$$Var(Y) = \frac{1}{n} \sum Y^2 - \bar{Y}^2 = \frac{19244.67}{6} - 55.31667^2 = 3207.445 - 3059.93398$$

$$Var Y = 147.5110$$

$$b_{yx} = \frac{Cov(X, Y)}{Var(X)} = \frac{131.8985}{118.5452} = 1.1126$$

$$b_{xy} = \frac{Cov(X, Y)}{Var(Y)} = \frac{131.8985}{147.5110} = 0.8942$$

Now substituting values of $\bar{X}$, $\bar{Y}$, b_{xy} , b_{yx} in the regression equations we get,

$$(Y - 55.31667) = 1.1126(X - 52.36667) \dots Y \text{ on } X \text{ (i)}$$

$$(X - 52.36667) = 0.8942(Y - 55.31667) \dots X \text{ on } Y \text{ (ii)}$$

4. If $x=67.6$, $y=?$

In this case, we use the regression equation Y on X

$$(Y - 55.31667) = 1.1126(67.6 - 52.36667)$$

$$Y = 55.31667 + 16.9486$$

$$Y = 72.2627$$

$$Y = 72.3$$

5. women=y men=x $\therefore y=?$ if $x=60$

In this case, we use regression equation Y on X

$$(Y - 55.31667) = 1.1126(60 - 52.36667)$$

$$Y = 55.31667 + 8.4928$$

$$\text{Women's L.E} = 63.8$$

6. women=y men=x $\therefore x=?$ if $y=65$

In this case, we use the regression equation X on Y

$$(X - 52.36667) = 0.8942(65 - 55.31667)$$

$$X = 52.36667 + 8.6588$$

$$X = 61.02547$$

$$\textit{Men's } L.E = 61.0$$