Example 10·1. *Calculate the correlation coefficient for the following heights (in inches) of fathers (X) and their sons (Y) :*

X :	65	66	67	67	68	69	70	72
Y :	67	68	65	68	72	72	69	71

$$r(X,Y) = \frac{\left(\frac{1}{n}\sum x_i y_i - \bar{x}\,\bar{y}\right)}{\sqrt{\left(\frac{1}{n}\sum x_i^2 - \bar{x}^2\right)\left(\frac{1}{n}\sum y_i^2 - \bar{y}^2\right)}}$$

$$r(X,Y) = \rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$$

$$= \frac{E(XY) - E(X)E(Y)}{\sqrt{\{E(X^2)-(E(X))^2\}\{E(Y^2)-(E(Y))^2\}}}$$

$$E(XY) = \sum_y \sum_x xy\, P(X=x, Y=y)$$

$f(x,y)$

$$= \sum_{i=1}^{n} x_i y_i P(X=x_i, Y=y_i)$$

CALCULATIONS FOR CORRELATION COEFFICIENT

X	Y	X^2	Y^2	XY
65	67	4225	4489	4355
66	68	4356	4624	4488
67	65	4489	4225	4355
67	68	4489	4624	4556
68	72	4624	5184	4896
69	72	4761	5184	4968
70	69	4900	4761	4830
72	71	5184	5041	5112
Total 544	552	37028	38132	37560

$x=1$, $x=0$, $y=0,1,2$

$$E(XY) = \sum_{i=1}^{n} (x_i y_i)\left(\frac{1}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$$

$$E(X) = \sum x\, P(X=x)$$

$$= \sum x_i P(X=x_i) = \sum x_i\left(\frac{1}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

$$r(X,Y) = \frac{E(XY) - E(X)E(Y)}{\sqrt{\{E(X^2)-(E(X))^2\}\{E(Y^2)-(E(Y))^2\}}}$$

$$= \frac{\frac{1}{n}\sum x_i y_i - \bar{x}\,\bar{y}}{\sqrt{\left\{\frac{1}{n}\sum x_i^2 - (\bar{x})^2\right\}\left\{\frac{1}{n}\sum y_i^2 - (\bar{y})^2\right\}}}$$

$$E(X^2) = \sum x^2 P(X=x)$$

$$= \sum x_i^2 P(X=x_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

$$P(Y=y_i) = \frac{1}{n}$$

r

CALCULATIONS FOR CORRELATION COEFFICIENT

X	Y	X^2	Y^2	XY
65	67	4225	4489	4355
66	68	4356	4624	4488
67	65	4489	4225	4355
67	68	4489	4624	4556
68	72	4624	5184	4896
69	72	4761	5184	4968
70	69	4900	4761	4830
72	71	5184	5041	5112
Total 544	552	37028	38132	37560

$$\bar{x} = \frac{544}{8} = 68$$

$(37560) \quad (68)(69)$

Total	544	552	37028	38132	37560

$$r(X,Y) = \frac{\frac{1}{8}(37560) - (68)(69)}{\sqrt{\left[\frac{1}{8}(37028) - (68)^2\right]\left[\frac{1}{8}(38132) - (69)^2\right]}} = 0.60$$

Poll Que: Which of the following options is correct?

(a) $r(X,Y) = \dfrac{\frac{1}{n}\left(\sum x_i y_i - \bar{x}\bar{y}\right)}{\sqrt{\frac{1}{n}\left(\sum x_i^2 - \bar{x}\right)\frac{1}{n}\left(\sum y_i^2 - \bar{y}\right)}}$

(b) $r(X,Y) = \dfrac{\frac{1}{n}\left(\sum x_i y_i - \bar{x}\bar{y}\right)}{\sqrt{\frac{1}{n}\left(\sum x_i^2 - \bar{x}^2\right)\frac{1}{n}\left(\sum y_i^2 - \bar{y}^2\right)}}$

(c) $r(X,Y) = \dfrac{\left(\frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}\right)}{\sqrt{\left(\frac{1}{n}\sum x_i^2 - \bar{x}^2\right)\left(\frac{1}{n}\sum y_i^2 - \bar{y}^2\right)}}$

(d) *None of these.*

$\frac{1}{n}\left(\sum x_i y_i - n\bar{x}\bar{y}\right)$

Theorem: *Correlation coefficient is independent of change of origin and scale.*

If $U = \frac{X-a}{h}$, $V = \frac{Y-b}{k}$

then $r(U,V) = r(X,Y)$

$X = \frac{X-0}{1}$, $Y = \frac{Y-0}{1}$, $(0,0)$

(a,b)

$(0,0)$

$X = a + hU$ $\quad$ $Y = b + kV$

$E(X) = a + hE(U)$ $\quad$ $E(Y) = b + kE(V)$

$\{X - E(X)\}^2 = h^2\{U - E(U)\}^2$ $\quad$ $\{Y - E(Y)\}^2 = k^2\{V - E(V)\}^2$

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

$$= E[ab + akV + bhU + hkUV] - [ab + akE(V) + bhE(U) + hkE(U)E(V)]$$

$$= \cancel{ab} + \cancel{akE(V)} + \cancel{bhE(U)} + hkE(UV) - [\cancel{ab} + \cancel{akE(V)} + \cancel{bhE(U)} + hkE(U)E(V)]$$

$$= hk[E(UV) - E(U)E(V)] = hk\,Cov.(U,V)$$

$$Cov.(U,V) = E(UV) - E(U)E(V)$$

$U = \frac{X-a}{h}$ $\quad$ $V = \frac{Y-b}{k}$

$$= E\left[\frac{XY - bX - aY + ab}{hk}\right] - \frac{1}{h}[E(X) - a]\frac{1}{k}[E(Y) - b]$$

$$= \frac{1}{}\left[E(XY) - \cancel{bE(X)} - \cancel{aE(Y)} + \cancel{ab}\right] - \frac{1}{}\left[E(X)E(Y) - \cancel{bE(X)}\right.$$

$$= \frac{1}{hk}\left[E(XY) - \cancel{bE(X)} - \cancel{aE(Y)} + \cancel{ab}\right] - \frac{1}{hk}\left[E(X)E(Y) - \cancel{bE(X)} - \cancel{aE(Y)} + \cancel{ab}\right]$$

$$= \frac{1}{hk}\left[E(XY) - E(X)E(Y)\right] = \frac{1}{hk}\,Cov(X,Y)$$

$$V(X) = E\left[\{X - E(X)\}^2\right] = E\left[h^2\{U - E(U)\}^2\right] = h^2 E\left[\{U - E(U)\}^2\right]$$

$$V(X) = h^2 V(U) \qquad V(Y) = k^2 V(V)$$

$$V(U) = E\left[\{U - E(U)\}^2\right] = E\left[\frac{1}{h^2}\{X - E(X)\}^2\right] = \frac{1}{h^2} E\left[\{X - E(X)\}^2\right]$$

$$= \frac{1}{h^2} V(X)$$

$$\sqrt{h^2} = |h| \quad \sqrt{k^2} = |k| \quad h>0,\ k>0$$

$$r(X,Y) = \frac{Cov.(X,Y)}{\sqrt{Var.(X)\,Var.(Y)}} = \frac{hk\,Cov.(U,V)}{\sqrt{h^2 Var.(U)\,k^2 Var.(V)}} = \frac{Cov.(U,V)}{\sqrt{Var.(U)\,Var.(V)}}$$

$$= r(U,V) \qquad Cov.(X,Y) = hk\,Cov.(U,V)$$

$$U = \frac{X-a}{h} \qquad V = \frac{Y-b}{k}$$

indep. of change of origin

dep. of " " scale.

$$V(X) = h^2 V(U)$$

(SHORT-CUT METHOD)

X	Y	$U = X - 68$	$V = Y - 69$	U^2	V^2	UV
65	67	−3	−2	9	4	6
66	68	−2	−1	4	1	2
67	65	−1	−4	1	16	4
67	68	−1	−1	1	1	1
68	72	0	3	0	9	0
69	72	1	3	1	9	3
70	69	2	0	4	0	0
72	71	4	2	16	4	8
Total		0	0	36	44	24

$a = x = 68$ $\quad b = y = 69$

$$\frac{E(UV) - E(U)E(V)}{\sqrt{Var.(U)\,Var.(V)}}$$

$$r(X,Y) = r(U,V) = \frac{\frac{1}{n}\sum u_i v_i - \bar{u}\bar{v}}{\sqrt{\left[\frac{1}{n}\sum u_i^2 - (\bar{u})^2\right]\left[\frac{1}{n}\sum v_i^2 - (\bar{v})^2\right]}}$$

$$= \frac{1}{n}(24) - (0)(0)$$

$$= \frac{\frac{1}{8}(24) - (0)(0)}{\sqrt{\left\{\frac{1}{8}(36) - 0^2\right\}\left\{\frac{1}{8}(44) - 0^2\right\}}} = 0.60.$$

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} \qquad \bar{x} = \frac{\sum x_i}{n}$$

$$= n\bar{x} - n\bar{x} = 0.$$

Corollary. *If X and Y are random variables and a, b, c, d are any numbers provided only that $a \neq 0$, $c \neq 0$, then*

$$r(aX + b, cY + d) = \frac{ac}{|ac|} r(X, Y)$$

If $ac > 0$, *i.e.*, if a and c are of same signs, then $ac/|ac| = +1$

If $ac < 0$, *i.e.*, if a and c are of opposite signs, then $ac/|ac| = -1$.

$|ac| = ac$

$\frac{ac}{|ac|} = \frac{ac}{-ac} = -1$

$r(3X+4, -4Y-7) = r(X,Y)$

$r(-3X+4, 4Y-7) = -r(X,Y)$

$aX + b = U \Rightarrow aE(X) + b = E(U)$

$cY + d = V \Rightarrow aE(Y) + d = E(V)$

$\{U - E(U)\}^2 = a^2\{X - E(X)\}^2$

$\{V - E(V)\}^2 = b^2\{Y - E(Y)\}^2.$

$$r(aX+b, cY+d) = r(U,V)$$

$$= \frac{E(UV) - E(U)E(V)}{\sqrt{\{E(U^2) - (E(U))^2\}\{E(V^2) - (E(V))^2\}}}$$

$$\text{Cov.}(U,V) = E(UV) - E(U)E(V)$$

$$= E[acXY + adX + bcY + bd] - [acE(X)E(Y) + adE(X) + bcE(Y) + bd]$$

$$= ac[E(XY) - E(X)E(Y)] = ac\,\text{Cov.}(X,Y)$$

$\text{Var.}(U) = a^2\text{Var.}(X) \qquad \text{Var.}(V) = c^2\text{Var.}(Y)$

$$r(U,V) = r(aX+b, cY+d) = \frac{\text{Cov.}(U,V)}{\sqrt{\text{Var.}(U)\text{Var.}(V)}} = \frac{ac\,\text{Cov.}(X,Y)}{\sqrt{a^2\text{Var.}(X)c^2\text{Var}(Y)}}$$

$$= \frac{ac\,Cov(X,Y)}{\sqrt{a^2 Var(X)\,c^2 Var(Y)}}$$

$$\sqrt{Var(U)Var(V)}$$

$\sqrt{a^2} = |a|$

$\sqrt{c^2} = |c|$

$\sqrt{9} = \sqrt{(-3)^2} = -3$ ✗

$\sqrt{(-3)^2} = |-3| = 3$ ✓

$\sqrt{(-3)^2} = [(-3)^2]^{1/2} = -3$ ✗

$$= \frac{ac}{|ac|} \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{ac}{|ac|} \rho(X,Y)$$