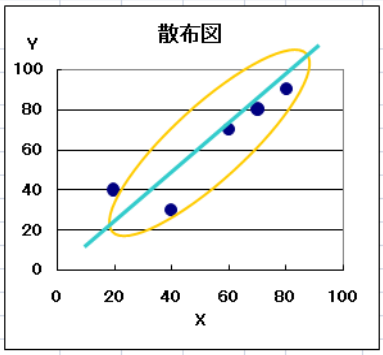
2変量の関係

データ						平均
X	20	40	60	80	70	54.0
Y	40	30	70	90	80	62.0

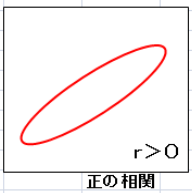
2変量の傾向を知るには ⇒ 散布図 (scatter diagram)



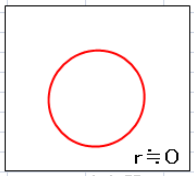
相関係数(r)の性質

$-1 \leqq r \leqq 1 \quad \Leftrightarrow \quad |r| \leqq 1$

完全な正の相関



正の相関



無相関

負の相関

完全な負の相関

相関係数 (correlation coefficient)

データ						平均
X	20	40	60	80	70	54.0
Y	40	30	70	90	80	62.0

$$r = \frac{\Sigma (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\Sigma (x_i - \overline{x})^2 \cdot \Sigma (y_i - \overline{y})^2}} = \frac{\frac{1}{n}\Sigma x_i y_i - \overline{x}\,\overline{y}}{\sqrt{(\frac{1}{n}\Sigma x_i^2 - \overline{x}^2)\cdot(\frac{1}{n}\Sigma y_i^2 - \overline{y}^2)}}$$

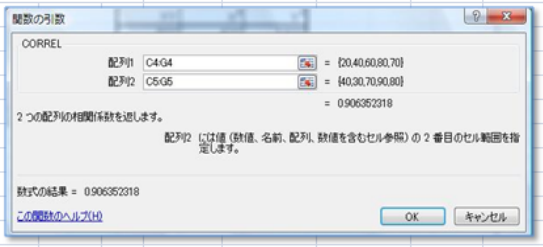
≒ 0.906

X	Y	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$
20	40	-34	-22	748	1156	484
40	30	-14	-32	448	196	1024
60	70	6	8	48	36	64
80	90	26	28	728	676	784
70	80	16	18	288	256	324
			Σ	2260	2320	2680

XY	X^2	Y^2
800	400	1600
1200	1600	900
4200	3600	4900
7200	6400	8100
5600	4900	6400
19000	16900	21900

r＝ 2260／√2320×2680
＝ 0.906

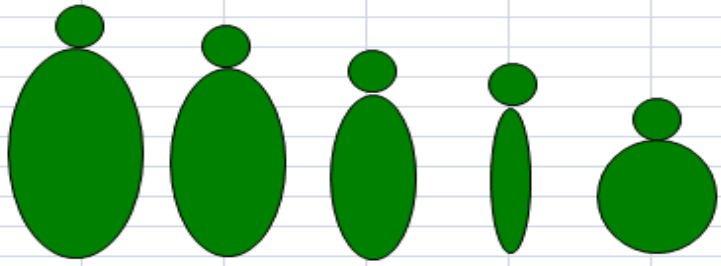
r＝ (M20/5-I4*I5)/sqrt((N20/5-I4*I4)*(O20/5-I5*I5))
＝ 0.906



順位相関（ranked correlation coefficient）

たとえば

5人の身長と体重に関して，体重の重い順に並べた



身長	1	2	3	4	5
体重	1	2	3	5	4

↑

個々の数値でなく，順番(順位)を示す値の比較でも2変量の関係は分かる

代表的な順位相関係数（タイスコアが無い場合）

データ

X	20	40	60	80	70
Y	40	30	70	90	80

Ⅰ．*Spearman's ranked correlation coefficient*

$$r_s = 1 - \frac{6\Sigma(u_i - v_i)^2}{n(n^2 - 1)}$$

u	1	2	3	5	4	
v	2	1	3	5	4	計
$(u-v)^2$	1	1	0	0	0	2

rs＝ 0.9

Ⅱ．*Kendall's ranked correlation coefficient*

$$\tau = \frac{S}{0.5 \times n(n-1)}$$

u	1	2	3	4	5	
v	2	1	3	4	5	
大	3	3	2	1	0	
小	1	0	0	0	0	(S)
大－小	2	3	2	1	0	8

τ＝ 0.8