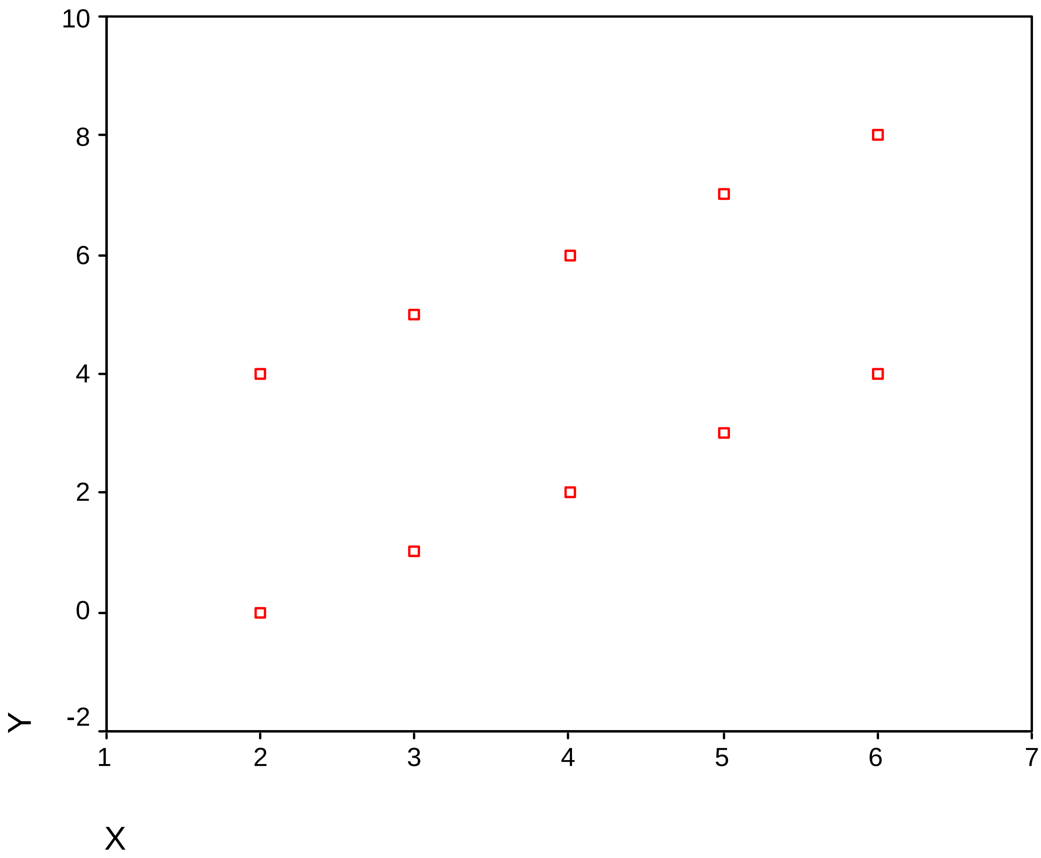
Tutorial 2

Exercise 1:

Table 1:

X	2	2	3	3	4	4	5	5	6	6
Y	0	4	1	5	2	6	3	7	4	8

Scatter plot:



Mean:

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

Thus

$$\bar{x} = \frac{2+2+3+3+4+4+5+5+6+6}{10} = 4$$

$$\bar{y} = \frac{0+4+1+5+2+6+3+7+4+8}{10} = 4$$

Variances:

$$\mathrm{var}(x) = s_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

$$s_x^2 = \frac{4+4+1+1+1+1+4+4}{9} = \frac{20}{9} = 2.22$$

$$s_y^2 = \frac{16+9+1+4+4+1+9+16}{9} = \frac{60}{9} = 6.67$$

Standard deviations:

$$sd(x) = s_x = \sqrt{\mathrm{var}(x)}$$

$$s_x = \sqrt{2.22} = 1.49$$

$$s_y = \sqrt{6.67} = 2.58$$

Covariances:

$$\mathrm{cov}(x, y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

$$\mathrm{cov}(x, y) = \frac{8+3+-1+-1+3+8}{9} = \frac{20}{9} = 2.22$$

Correlations:

$$r_{xy} = \frac{\operatorname{cov}(x,y)}{s_x s_y} = \frac{2.22}{1.49 * 2.58} = 0.577$$

Exercise 2a

Linear transformations

$v_i = 2x_i + 3$

$w_i = y_i - 8$

Linear combinations

$z_i = v_i + w_i = 2x_i + y_i - 5$

<u>Means</u>

$\bar{v}_i = a\bar{x}_i + b = 2*4+3 = 11$

$\bar{w}_i = \bar{y}_i - b = 1*4 - 8 = -4$

$\bar{z}_i = 2\bar{x}_i + \bar{y} - 5 = 2*4+1*4 - 5 = 7$

<u>Variances</u>

Calculate using table:

Table var(v) = cov (v,v) $v = 2x_i + 3$

	$2x_i$
$2x_i$	$(2)*(2)\,\mathrm{cov}(x,x) = 4\,\mathrm{var}(x) = 8.89$

Table var(w) $w = y_i - 8$

	$1y_i$
$1y_i$	$(1)*(1)\,\mathrm{cov}(y,y) = 1\,\mathrm{var}(y) = 6.67$

Table var(z) $z = 2x_i + y_i - 5$

	$2x_i$	$1y_i$
$2x_i$	8.89	$2\,\mathrm{cov}(x,y) = 4.44$
$1y_i$	$2\,\mathrm{cov}(x,y) = 4.44$	6.67

$\mathrm{var}(z) = 8.89 + 4.44 + 4.44 + 6.67 = 24.44$

Standard deviations

$$s_v = \sqrt{8.89} = 2.98$$
$$s_w = \sqrt{6.67} = 2.58$$
$$s_z = \sqrt{24.44} = 4.94$$

Exercise 2b

Covariances

cov(v,w)

$v_i = ax_i + b$

and

$w_i = cy_i + d$

$\text{cov}(v, w) = ac\,\text{cov}(x, y) = (2)*(1)*(2.22) = 4.44$

Calculate with help of tables.

Table cov(v,w)

	$1y_i$
$2x_i$	$2\,\text{cov}(x, y) = 4.44$

Table cov(v,z)

	$2x_i$	$1y_i$
$2x_i$	$4\,\text{cov}(x, x) = 8.89$	$2\,\text{cov}(x, y) = 4.44$

$\text{cov}(v, z) = 8.89 + 4.44 = 13.33$

Table cov(w,z)

	$2x_i$	$1y_i$
$1y_i$	$2\operatorname{cov}(x,y)=4.44$	$1\operatorname{cov}(y,y)=6.67$

$$\operatorname{cov}(w,z)=4.44+6.67=11.11$$

Correlations

Correlation between v end w

If ac >0 , than $r_{vw} = r_{xy}$
If ac<0 , than $r_{vw} = -r_{xy}$

With linear transformations correlation remains the same in absolute values.

$$r_{xy}=\frac{\operatorname{cov}(x,y)}{s_x s_y}$$

$$r_{vw}=\frac{4.44}{2.98*2.58}=0.577$$

$$r_{vz}=\frac{13.33}{2.98*4.94}=0.905$$

$$r_{wz}=\frac{11.11}{2.58*4.94}=0.870$$

Total

	X	Y	V	W	Z
Means	4	4	11	-4	7
Variances	2.22	6.67	8.89	6.67	24.44

Covariance matrix

Covariance	V	W	Z
V	8.89		
W	4.44	6.67	
Z	13.33	11.11	24.44

Correlation matrix

Correlation	V	W	Z
V	1		
W	0.577	1	
Z	0.905	0.870	1

Exercise 3a

$$z_i = \frac{(x_i - \overline{x})}{s_x}$$

Table 1:

X	2	2	3	3	4	4	5	5	6	6
Y	0	4	1	5	2	6	3	7	4	8

$$x = 2 \Rightarrow z = \frac{(2-4)}{1.49} = -1.342$$

$$x = 3 \Rightarrow z = \frac{(3-4)}{1.49} = -0.671$$

$$x = 4 \Rightarrow z = \frac{(4-4)}{1.49} = 0$$

$$x = 5 \Rightarrow z = \frac{(5-4)}{1.49} = 0.671$$

$$x = 6 \Rightarrow z = \frac{(6-4)}{1.49} = 1.342$$

$$y = 0 \Rightarrow z = \frac{(0 - 4)}{2.58} = -1.550$$

$$y = 1 \Rightarrow z = \frac{(1 - 4)}{2.58} = -1.162$$

$$y = 2 \Rightarrow z = \frac{(2 - 4)}{2.58} = -0.775$$

$$y = 3 \Rightarrow z = \frac{(3 - 4)}{2.58} = -0.387$$

$$y = 4 \Rightarrow z = \frac{(4 - 4)}{2.58} = 0$$

$$y = 5 \Rightarrow z = \frac{(5 - 4)}{2.58} = 0.387$$

$$y = 6 \Rightarrow z = \frac{(6 - 4)}{2.58} = 0.775$$

$$y = 7 \Rightarrow z = \frac{(7 - 4)}{2.58} = 1.162$$

$$y = 8 \Rightarrow z = \frac{(8 - 4)}{2.58} = 1.550$$

Exercise 3b

$$r_{xy} = \frac{\mathrm{cov}(x, y)}{s_x s_y}$$

For z-scores standard deviations are always equal to 1, thus for z-scores:

$$r_{z_x z_y} = \frac{\mathrm{cov}(z_x, z_y)}{1*1}$$

$$\mathrm{cov}(z_x, z_y) = \sum_{i=1}^{n} \frac{(z_{x_i} - z_{\bar{x}})(z_{y_i} - z_{\bar{y}})}{n-1}$$

For z-scores means are always 0, thus for z-scores:

$$r_{z_x z_y} = \sum_{i=1}^{n} \frac{z_{x_i} z_{y_i}}{n-1}$$

Table 1:

	A	B	C	D	E	F	G	H	I	J
X	2	2	3	3	4	4	5	5	6	6
Y	0	4	1	5	2	6	3	7	4	8

Calculate with z-scores from 3a.

A: -1.342 * -1.549 = 2.079
B: 0 = 0
C: -0.671 * -1.162 = 0.780
D: -0.671 * 0.387 = -0.260
E: 0 = 0
F: 0 = 0
G: 0.671 * -0.387 = -0.260
H: 0.671 * 1.162 = 0.780
I: 0 = 0
J: 1.342 * 1.549 = $\underline{2.079}$
5.198 +

$$r_{z_x z_y} = \frac{5.198}{9} = 0.577 = r_{xy}!!$$

Exercise 4

a) $U = ax + b$ (when a > 0)
b) $U = -ax + b$ (when a > 0)
c) $U = b$ (when a = 0)
d) $U = (x - 4)^2$

Examples:

a) U = 2x+1

X	2	2	3	3	4	4	5	5	6	6
U	5	5	7	7	9	9	11	11	13	13

b) U = -2x+1

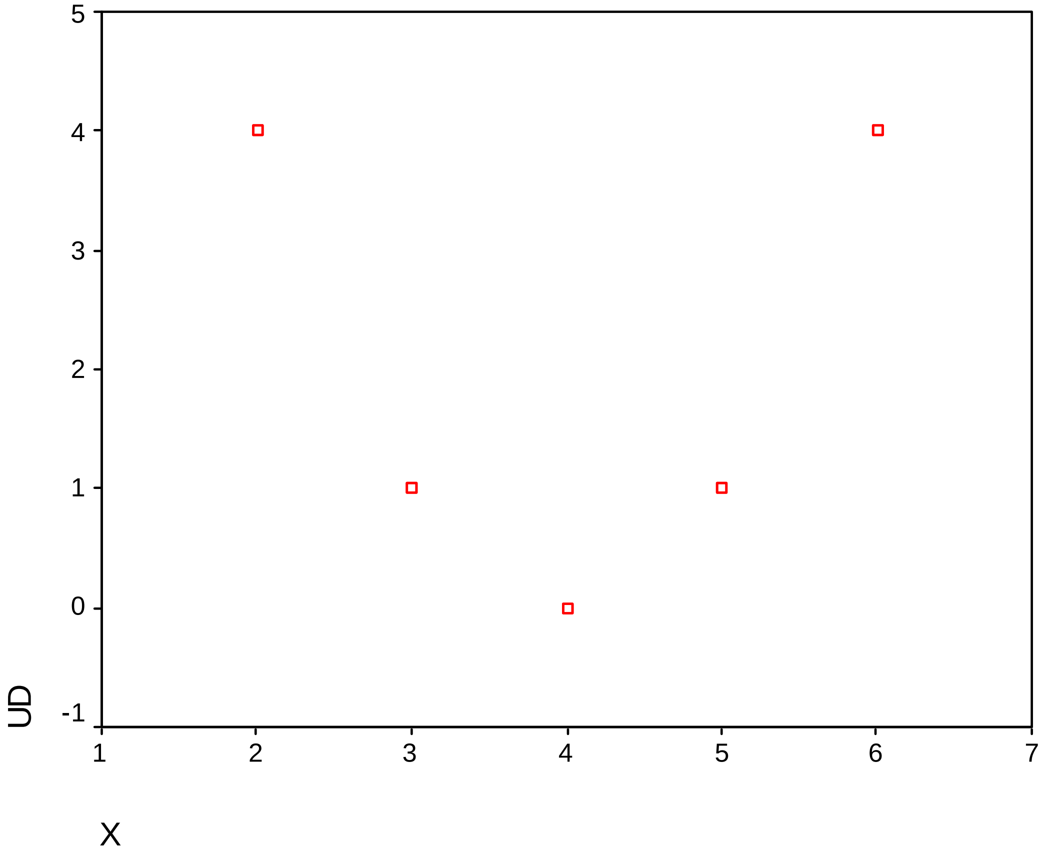
X	2	2	3	3	4	4	5	5	6	6
U	-3	-3	-5	-5	-7	-7	-9	-9	-11	-11

c) U = 9 + E

X	2	2	3	3	4	4	5	5	6	6
U	8	10	8	10	10	8	10	8	8	10

d) U = $(x - 4)^2$

X	2	2	3	3	4	4	5	5	6	6
U	4	4	1	1	0	0	1	1	4	4



Exercise 5

Null hypothesis: correlation between X and Y in population is equal to zero:

H_0: $\rho_{xy} = 0$
H_a: $\rho_{xy} \neq 0$

Test statistic:

$$t = r_{xy}\sqrt{\frac{n-2}{1-r_{xy}^2}} = 0.577\sqrt{\frac{8}{1-0.577^2}} = 1.998$$

Two-tailed t-test:
$t_{cv}(df = n-2) = t_{cv}(8)_{\alpha=0.05} = \pm\ 2.306$

Conclusion: H_0 not rejected.

Seee Hinkle, Wiersma and Jurs Chapter.10 pp. 235-238

<u>Spss ouput</u>

Correlations

		X	Y
X	Pearson Correlation	1,000	,577
	Sig. (2-tailed)	,	,081
	N	10	10
Y	Pearson Correlation	,577	1,000
	Sig. (2-tailed)	,081	,
	N	10	10

Exercise 6

Proportion of explained (predicted) variance =

$R^2 xy$

$R^2 xy = 0.577^2 = 0.33 = 33\%$

Exercise 7

a)

	A	B	C	D	E	F	G	H	I	J	K
X	2	2	3	3	4	4	5	5	6	6	10000
Y	0	4	1	5	2	6	3	7	4	8	-10000

R is -1.00

b)

	A	B	C	D	E	F	G	H	I	J	K
X	2	2	3	3	4	4	5	5	6	6	10000
Y	0	4	1	5	2	6	3	7	4	8	10000

R is 1.00

Exercise 8

Table 1 when x>3

X	4	4	5	5	6	6
Y	2	6	3	7	4	8

$$\bar{x} = \frac{\sum_{i-1}^{n} x_i}{n}$$

$$\bar{x} = 5$$
$$\bar{y} = 5$$

$$\operatorname{var}(x) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}$$

$$\operatorname{var}(x) = 0.8$$
$$\operatorname{var}(y) = 5.6$$

$$sd(x) = \sqrt{\mathrm{var}(x)}$$

$$sd(x) = 0.89$$
$$sd(y) = 2.37$$

$$\mathrm{cov}(x, y) = \sum_{i=1}^{n} \frac{(x_i - \overline{x})(y_i - \overline{y})}{n - 1}$$

$$\mathrm{cov}(x, y) = 0.8$$

$$r_{xy} = \frac{\mathrm{cov}(x, y)}{s_x s_y}$$

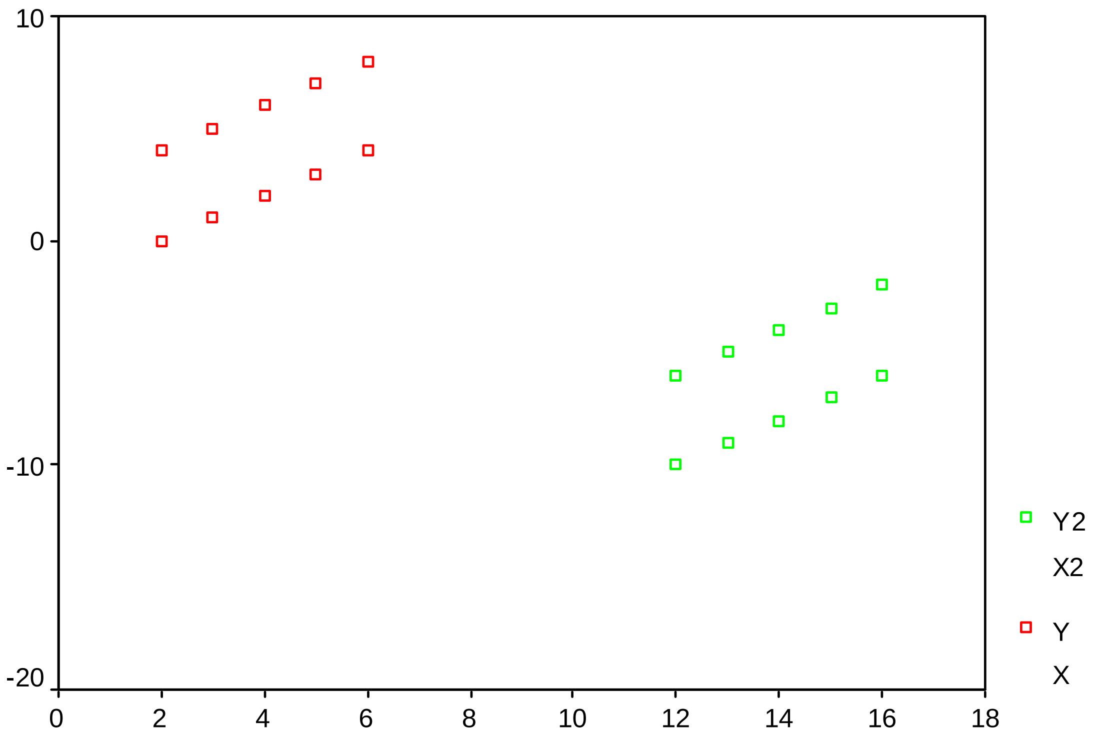
$$r_{xy} = 0.378$$

Exercise 9a.

X2 = X+10
Y2 = Y-10

Table:

X	2	2	3	3	4	4	5	5	6	6
Y	0	4	1	5	2	6	3	7	4	8
X2	12	12	13	13	14	14	15	15	16	16
Y2	-10	-6	-9	-5	-8	-4	-7	-3	-6	-2



X2 = X +10

Y2 = Y -10

$r_{x2y2} = r_{xy}$

Correlations

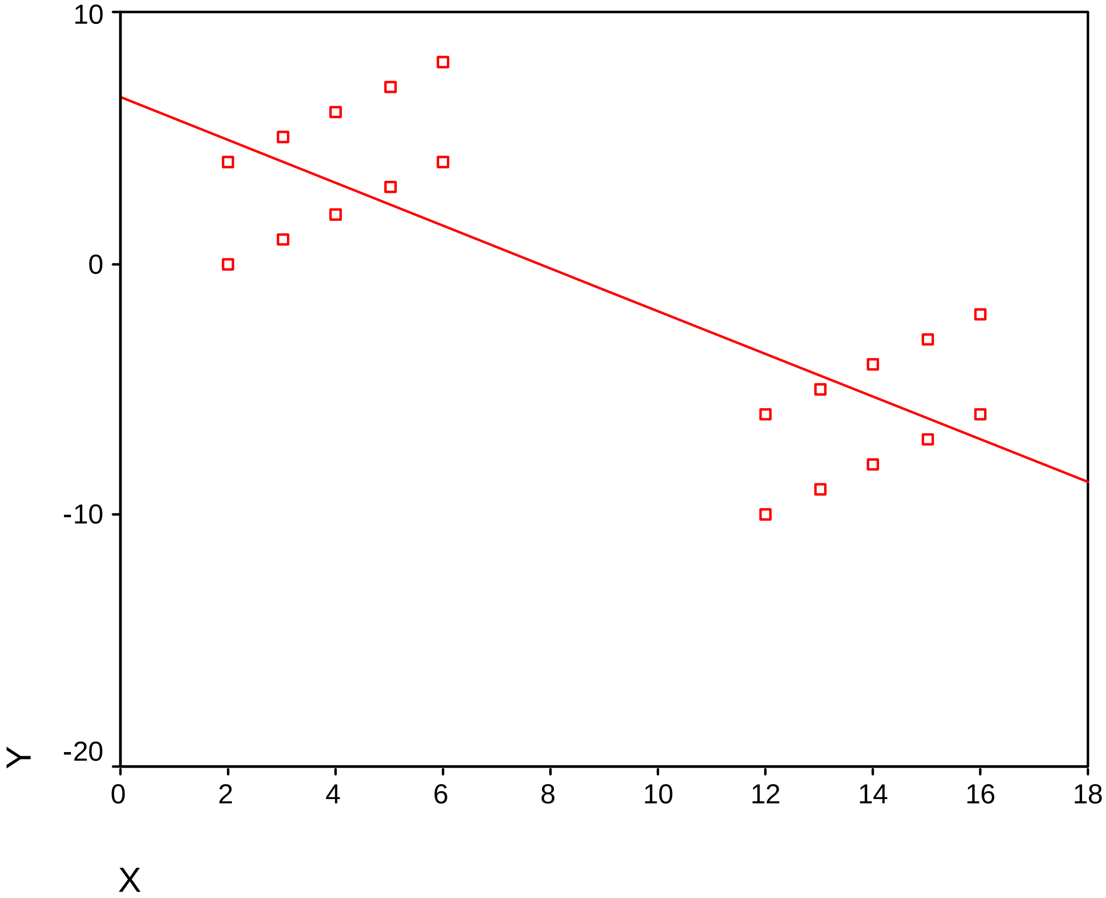
		X2	Y2
X2	Pearson Correlation	1	,577
	Sig. (2-tailed)	.	,081
	N	10	10
Y2	Pearson Correlation	,577	1
	Sig. (2-tailed)	,081	.
	N	10	10

r_{xy} in group I +II is negative

Correlations

		X	Y
X	Pearson Correlation	1	-,688**
	Sig. (2-tailed)	.	,001
	N	20	20
Y	Pearson Correlation	-,688**	1
	Sig. (2-tailed)	,001	.
	N	20	20

**. Correlation is significant at the 0.01 level


10
0
-10
-20
Y
0
2
4
6
8
10
12
14
16
18
X

Exercise 9b.

X2 = X
Y2 = -Y

Table:

X	2	2	3	3	4	4	5	5	6	6
Y	0	4	1	5	2	6	3	7	4	8
X2	2	2	3	3	4	4	5	5	6	6
Y2	-0	-4	-1	-5	-2	-6	-3	-7	-4	-8

$r_{x2y2} = - r_{xy}$

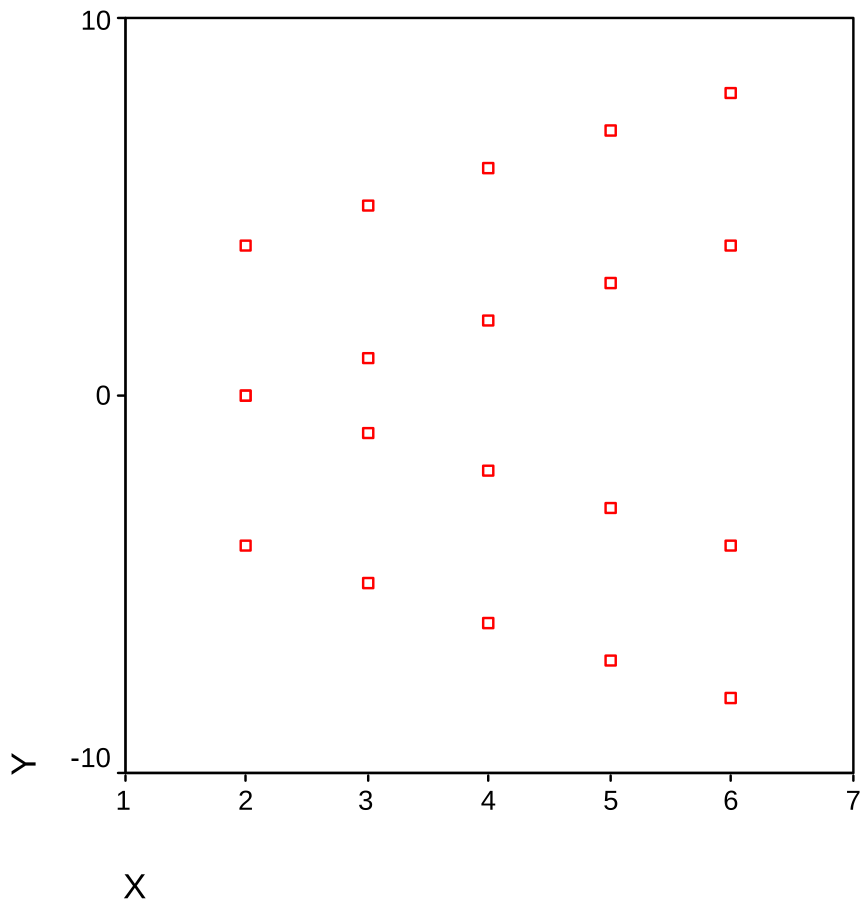
Correlations

		X3	Y3
X3	Pearson Correlation	1	-,577
	Sig. (2-tailed)	.	,081
	N	10	10
Y3	Pearson Correlation	-,577	1
	Sig. (2-tailed)	,081	.
	N	10	10

r_{xy} in group I +II is 0

Correlations

		X	Y
X	Pearson Correlation	1	,000
	Sig. (2-tailed)	.	1,000
	N	20	20
Y	Pearson Correlation	,000	1
	Sig. (2-tailed)	1,000	.
	N	20	20



Exercise 10

	X=0	X=1	Total

Y=0	A	B	12
Y=1	C	D	28
Total	36	4	40

Correlation r for 2 dichotomous variables is equal to Φ:

$$\phi = \frac{(ad - bc)}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

Note:

with minimum correlation, we mean:

lower limit (largest neg. value)
lower limit as ad-bc takes maximal negative value.

and with maximum correlation:

higher limit (highest pos. value)
higher limit as ad-bc takes maximum positive value

Lower limit:

	X=0	X=1	Total
Y=0	8 (a)	4 (b)	12
Y=1	28 (c)	0 (d)	28
Total	36	4	40

$$\phi = \frac{(ad - bc)}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

$$\phi = \frac{(0 - 112)}{\sqrt{(12)(28)(36)(4)}} = \frac{-112}{\sqrt{48384}} = -0.51$$

Higher limit:

	X=0	X=1	Total
Y=0	12 (a)	0 (b)	12
Y=1	24 (c)	4 (d)	28
Total	36	4	40

$$\phi = \frac{(ad - bc)}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

$$\phi = \frac{(48 - 0)}{\sqrt{(12)(28)(36)(4)}} = \frac{48}{\sqrt{48384}} = 0.21$$

In this way we can see that the correlation lays between: $-0.51 \leq \Phi \leq 0.21$