Application of Statistics in Educational Research I

MPU1034

CORRELATION*

Prof. Dr. Mohd Salleh Abu
Dr. Hamidreza Kashefi

main source: Vernoy & Vernoy (1997)

Some Commonly Used Jargons…

- Linear and Non-linear Relationship
- Correlation
- Scatterplot
- Correlation Coefficient
- Pearson (Product-Moment *r*)
- Spearman Rank (Spearman rho ρ)
- Kendall's Fax (Kendall tau τ)
- Phi (ϕ)
- Point-Biserial (φ)

The General Idea About Correlation

Two variables may correlate to each other in 3 possible ways:

- ✓ Positive Relationship:

  Both variables vary in the same direction - as one goes up, the other goes up - Eg. salary and years of education are positively correlated because the people who make the highest salaries tend to be the ones who have gone to school the longest

- ✓ Negative Relationship:

  Two variables vary in the opposite direction - as one goes up, the other goes down. For example, the number of daily hassles and the amount of immunoglobulin *A* in a person's system are negatively correlated because as the number of hassles goes up, the amount of immunoglobulin A tends to go down.

- ✓ Zero Relationship:

  Two variables vary with freely without any kind of rules or relationship. For example, the relationship between personality fluctuations and the movement of distant stars has a zero correlation. Contrary to strong beliefs in astrology held by a surprising number of people, bona fide studies have found that there is absolutely no correlation between the position of stars at the time you are born and the true nature of your personality

Degree of Correlation: How Strongly Are Variables Correlated?

The degree of correlation between two variables can be established using two methods:

- ☑ Scatterplot – a graph with plotted values for two variables being compared
- ☑ Correlation Coefficient

Scatterplot

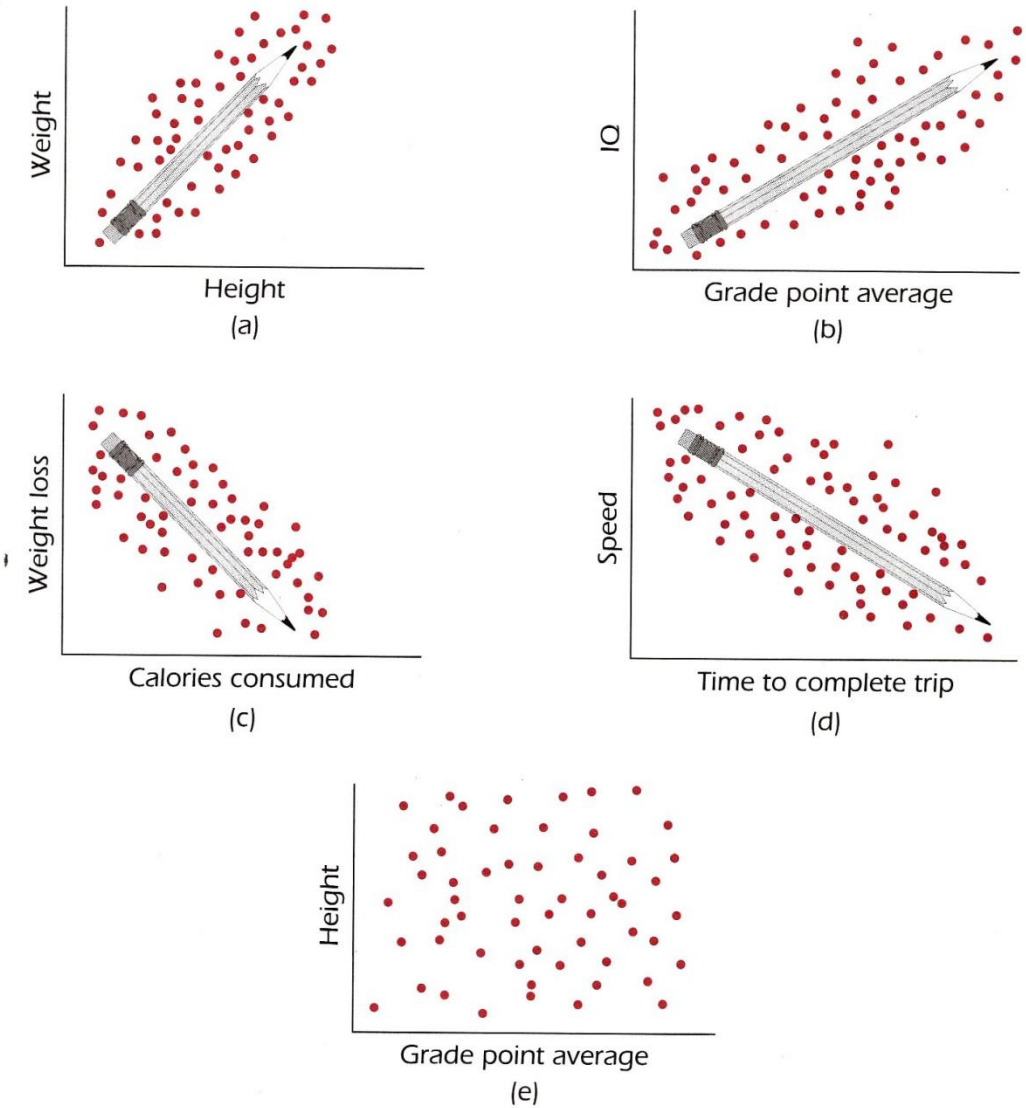


Figure 8.1 Five scatterplots indicating (a) a positive correlation between weight and height; (b) a positive correlation between IQ and grade point average; (c) a negative correlation between weight loss and calories consumed; (d) a negative correlation between speed and time to complete a trip; and (e) a zero correlation between height and grade point average

Correlation Coefficient

CORRELATION

- ✓ Correlation characterizes the existence of a relationship between variables
- ✓ Correlation coefficient is a quantitative indicator of the degree of relationship between variables
- ✓ There are several measures of correlations depending to the nature of variables being investigated. The most commonly used measure of correlation is called Pearson product moment correlation (denoted by ***r***)
- ✓ Each measure of correlation uses different kind of rules and modeling

✓ A brief description of the measures of correlations are described below:

Scales of Variables Being Compared	Applicable Measures	Symbols of Correlations
Interval-Interval (or 'stronger' forms)	Pearson Product Moment	r
Ordinal-Ordinal	Spearman Rank (Spearman rho)	ρ
	Kendall's tau	τ
Nominal-Nominal	Phi	ϕ
Interval-Nominal	Point-Biserial	φ

Computing The Correlation Coefficients

Pearson Product Moment (r)

The Pearson correlation ***r*** can be computed using the formula:

$$r = \frac{(n \cdot \Sigma XY) - (\Sigma X \cdot \Sigma Y)}{\sqrt{[(n \cdot \Sigma X^2) - (\Sigma X)^2] \cdot [(n \cdot \Sigma Y^2) - (\Sigma Y)^2]}}$$

Eg 1.

Compute the correlation between the men's Heights (in inches) and Weights (in pounds) for the following data:

Man	Height (X)	Weight (Y)
M.P.	72	190
T.D.	66	135
C.Q.	69	155
C.Y.	72	165
D.P.	71	155

Note that the paired data are in the form of interval-interval. Thus, the most suitable correlation measure would be Pearson product moment.

Table 8.6 Computation of the Correlation Coefficient Using the Raw-Score Formula for Men's Heights and Weights

Man	Height X	X^2	Weight Y	Y^2	XY
M. P.	72	5,184	190	36,100	13,680
T. D.	66	4,356	135	18,225	8,910
C. Q.	69	4,761	155	24,025	10,695
C. Y.	72	5,184	165	27,225	11,880
D. P.	71	5,041	155	24,025	11,005
	$\Sigma X = 350$	$\Sigma X^2 = 24,526$	$\Sigma Y = 800$	$\Sigma Y^2 = 129,600$	$\Sigma XY = 56,170$

Computation of ***r*** :

$$r = \frac{(n \cdot \Sigma XY) - (\Sigma X \cdot \Sigma Y)}{\sqrt{[(n \cdot \Sigma X^2) - (\Sigma X)^2] \cdot [(n \cdot \Sigma Y^2) - (\Sigma Y)^2]}}$$

$$= \frac{(5 \cdot 56{,}170) - (350 \cdot 800)}{\sqrt{[(5 \cdot 24{,}526) - 350^2] \cdot [(5 \cdot 129{,}600) - 800^2]}}$$

$$= \frac{280{,}850 - 280{,}000}{\sqrt{(122{,}630 - 122{,}500) \cdot (648{,}000 - 640{,}000)}}$$

$$= \frac{850}{\sqrt{130 \cdot 8{,}000}} = \frac{850}{\sqrt{1{,}040{,}000}} = \frac{850}{1{,}019.804} = .83$$

Eg 2.

Table 8.7 Computation of the Correlation Coefficient Using the Raw-Score Formula for Number of Hassles and Immunoglobulin A Counts for Ten Subjects

Subject	Daily hassles X	X^2	Immunoglobulin A count Y	Y^2	XY
C. D.	12	144	1.72	2.958	20.64
P. M.	15	225	1.15	1.322	17.25
R. R.	20	400	0.15	0.023	3.00
M. D.	12	144	2.00	4.000	24.00
J. L.	17	289	0.25	0.063	4.25
M. F.	16	256	1.33	1.769	21.28
J. P.	10	100	2.10	4.410	21.00
H. S.	12	144	1.75	3.063	21.00
M. H.	15	225	1.35	1.823	20.25
M. V.	19	361	0.20	0.040	3.80
	$\Sigma X = 148$	$\Sigma X^2 = 2288$	$\Sigma Y = 12.00$	$\Sigma Y^2 = 19.470$	$\Sigma XY = 156.47$

$$r = \frac{(n \cdot \Sigma XY) - (\Sigma X \cdot \Sigma Y)}{\sqrt{[(n \cdot \Sigma X^2) - (\Sigma X)^2] \cdot [(n \cdot \Sigma Y^2) - (\Sigma Y)^2]}}$$

$$= \frac{(10 \cdot 156.47) \cdot (148 \cdot 12)}{\sqrt{[(10 \cdot 2288) - 148^2] \cdot [(10 \cdot 19.470) - 12^2]}}$$

$$= \frac{1564.7 - 1776}{\sqrt{(22{,}880 - 21{,}904) \cdot (194.70 - 144)}}$$

$$= \frac{-211.3}{\sqrt{976 \cdot 50.7}} = \frac{-211.3}{\sqrt{49{,}483.2}} = \frac{-211.3}{222.448} = -.95$$

- There is another way to calculate Pearson coefficient. In the following example you can see how the Pearson coefficient can be calculated by the new set of formula.

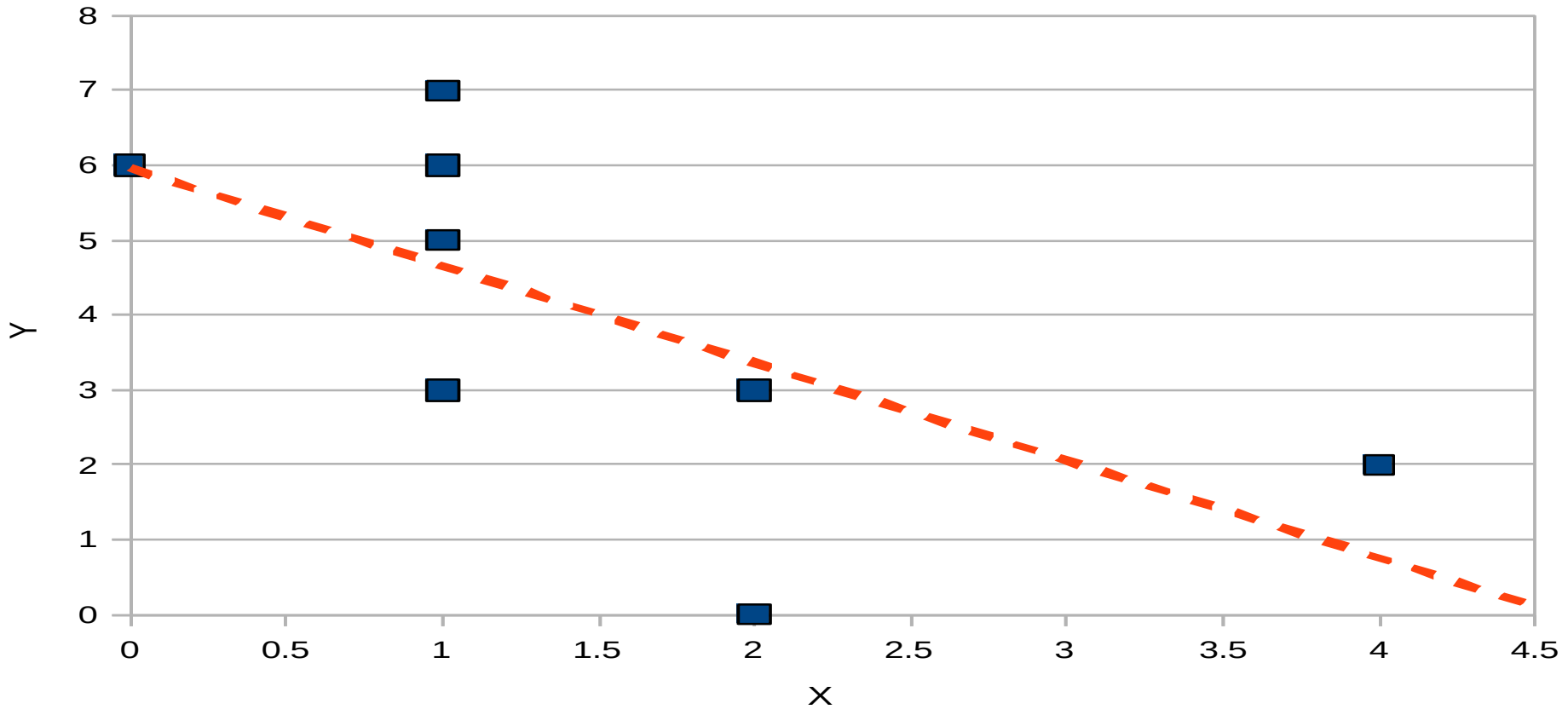
source: Gravetter (2012)

For the following scores,

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

a) Sketch a scatter plot and estimate the Pearson Correlation.
b) Compute the Pearson correlation

a) Sketch a scatter plot and estimate the Pearson Correlation.



The scatter plot shows points moderately scattered around a line slopping down to the right

STEP 2 : USE FORMULA TO FIND THE VALUE OF r

We need to find *Mx*, *My*, *SSx*, *SSy*, *SP*

$$M_X = \frac{\sum X}{n} \qquad M_Y = \frac{\sum Y}{n}$$

$$SS_x = \sum (X - M_x)^2$$

$$SP = \sum (X - M_x)(Y - M_y)$$

$$SS_y = \sum (X - M_y)^2$$

$$r = \frac{SP}{\sqrt{(SSx)(SSy)}}$$

b) Compute the Pearson correlation

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

Formula Pearson correlation:

$$r = \frac{SP}{\sqrt{(SSx)(SSy)}}$$

$$M_x = \frac{\sum x}{n}$$

$$M_y = \frac{\sum y}{n}$$

$$M_x = \frac{1+4+1+1+2+0+2+1}{8}$$

$$= 1.5$$

$$M_y = \frac{7+2+3+6+0+6+3+5}{8}$$

$$= 4$$

Scores		Deviations		Squared Deviations		Products
X	y	$(x - M_X)$	$(y - M_y)$	$(x - M_x)^2$	$(y - M_y)^2$	$(x - M_X)(y - M_y)$
1	7	1 – 1.5 = **- 0.5**	7 – 4 = **3**	(- 0.5)² = **0.25**	(3)² = **9**	(-0.5)(3) = **- 1.5**
4	2	4 – 1.5 = **2.5**	2 – 4 = **-2**	(2.5)² = **6.25**	(-2)² = **4**	(2.5)(-2) = **-5**
1	3	1 – 1.5 = **- 0.5**	3 – 4 = **- 1**	(- 0.5)² = **0.25**	(- 1)² = **1**	(- 0.5)(-1) = **0.5**
1	6	1 – 1.5 = **- 0.5**	6 – 4 = **2**	(- 0.5)² = **0.25**	(2)² = **4**	(- 0.5)(2) = **-1**
2	0	2 – 1.5 = **0.5**	0 – 4 = **- 4**	(0.5)² = **0.25**	(- 4)² = **16**	(0.5)(-4) = **-2**
0	6	0 – 1.5 = **- 1.5**	6 – 4 = **2**	(- 1.5)² = **2.25**	(2)² = **4**	(-1.5)(2) = **-3**
2	3	2 – 1.5 = **0.5**	3 – 4 = **- 1**	(0.5)² = **0.25**	(- 1)² = **1**	(0.5)(-1) = **- 0.5**
1	5	1 – 0.5 = **- 0.5**	5 – 4 = **1**	(- 0.5)² = **0.25**	(1)² = **1**	(-0.5)(1) = **- 0.5**
				SS_x = 10	**SS_y = 40**	**SP = - 13**

$SS_X = 10 \quad SS_Y = 40 \quad SP = -13$

$$r = \frac{SP}{\sqrt{(SSx)(SSy)}}$$

$$r = \frac{-13}{\sqrt{(10)(40)}}$$

$$= -0.65$$

Eg 3.

Table 8.8 Correlation Coefficient for Shoe Size and Number of Friends for Five Students

Student	Shoe size X	X^2	Number of friends Y	Y^2	XY
B. D.	12	144	3	9	36
J. C.	10	100	1	1	10
B. W.	9	81	2	4	18
K. T.	7	49	1	1	7
K. H.	6	36	3	9	18
	$\Sigma X = 44$	$\Sigma X^2 = 410$	$\Sigma Y = 10$	$\Sigma Y^2 = 24$	$\Sigma XY = 89$

$$r = \frac{(n \cdot \Sigma XY) - (\Sigma X \cdot \Sigma Y)}{\sqrt{[(n \cdot \Sigma X^2) - (\Sigma X)^2] \cdot [(n \cdot \Sigma Y^2) - (\Sigma Y)^2]}}$$

$$= \frac{(5 \cdot 89) - (44 \cdot 10)}{\sqrt{[(5 \cdot 410) - 44^2] \cdot [(5 \cdot 24) - 10^2]}}$$

$$= \frac{445 - 440}{\sqrt{(2050 - 1936) \cdot (120 - 100)}} = \frac{5}{\sqrt{114 \cdot 20}} = \frac{5}{\sqrt{2280}}$$

$$= \frac{5}{\sqrt{47.749}} = .105$$

Spearman Rank (Spearman rho ρ)

- ✓ Spearman rank (sometimes called Spearman rho) is used to find the correlation of variables containing data presented in the *ranked order*.
- ✓ The Spearman correlation is denoted by ρ
- ✓ Speaman rank is suitable for small sample size (30 or less)
- ✓ The computation of ρ is given by:

$$\rho = 1 - \frac{6\sum d^2}{N\left(N^2 - 1\right)}$$

where

d is the difference of each of the paired-rank

N is the number of pairs of ranks

Eg 1.

A study was conducted to find the relationship between the appraisal of an employee's personal traits (in ranking form) as evaluated by two Heads of Department (namely HD1 and HD2). The data are as follows:

Traits Being Evaluated	HD1 rank	HD2 rank
Communication Skills	1	7
Teamworking	8	2
Problem Solving	7	6
Self Esteem	6	4
Adaptability	5	8
Independent	4	5
Personal Integrity	3	1
Ethics	2	3

Find the Spearman rank ρ? Give you comment on the ρ obtained.

Traits Being Evaluated	HD1 rank (X)	HD2 rank (Y)	d = Y - X	d^2
Communication Skills	1	7	6	36
Teamworking	8	2	-6	36
Problem Solving	7	6	-1	1
Self Esteem	6	4	-2	4
Adaptability	5	8	3	9
Independent	4	5	1	1
Personal Integrity	3	1	-2	4
Ethics	2	3	1	1

$N = 8$ $\sum d^2 = 92$

Spearman rank(ρ)

$$= 1 - \frac{6\sum d^2}{N(N^2-1)}$$

$$= 1 - \frac{6(92)}{8(8^2-1)}$$

$$= 1 - \frac{552}{8(63)}$$

$$= 1 - 1.095$$

$$= -0.095$$

Conclusion:

The correlation between the the two appraisal is almost negligible i.e. there is no correlation between appraisal evaluated by the two Heads of Department

When is a correlation 'strong enough'? (Guilford's rough guide only)

< 0.2	slight; almost negligible relationship
0.2 – 0.4	low correlation; definite but small relationship
0.4 – 0.7	moderate correlation; substantial relationship
0.7 – 0.9	high correlation; marked relationship
> 0.9	very high correlation; very dependable relationship

Words of Caution

- Examine your data distribution (i.e. using scatterplot) before you do anything with the correlation and make sure you know the do's and don'ts with correlation coefficients!

- Correlation coefficient is just an index of relationship which tells nothing about the 'cause and effect' of the relationship! The correlation coefficient tells you nothing about the reasons of the relationship, if any.

- Limit yourself to linear relationship if you don't have adequate statistical background!

- Never compare correlation coefficients soley at their face values! Any comparative investigation MUST be accompanied with careful checking on scatterplot, sample homogeniety and sample size.

Check this….

A researcher is studying the factors that are likely to affect Reading Ability (RA) among a particular group of young school children. He employs a comparative correlational approach (using Pearson product moment) to establish the following table:

TABLE A Correlation Coefficients Between RA and Factors Investigated (ordered by the degree of contribution)

Factors Investigated	Correlation Coefficients
Mastery of Vocabulary	0.91
Exposure to Early Reading	0.83
Spatial Ability	0.77
Scale of Motivation from Parents	0.75
Reasoning Ability	0.58

Critise the method of analysis employed by this researcher.

The Importance of Scatterplot

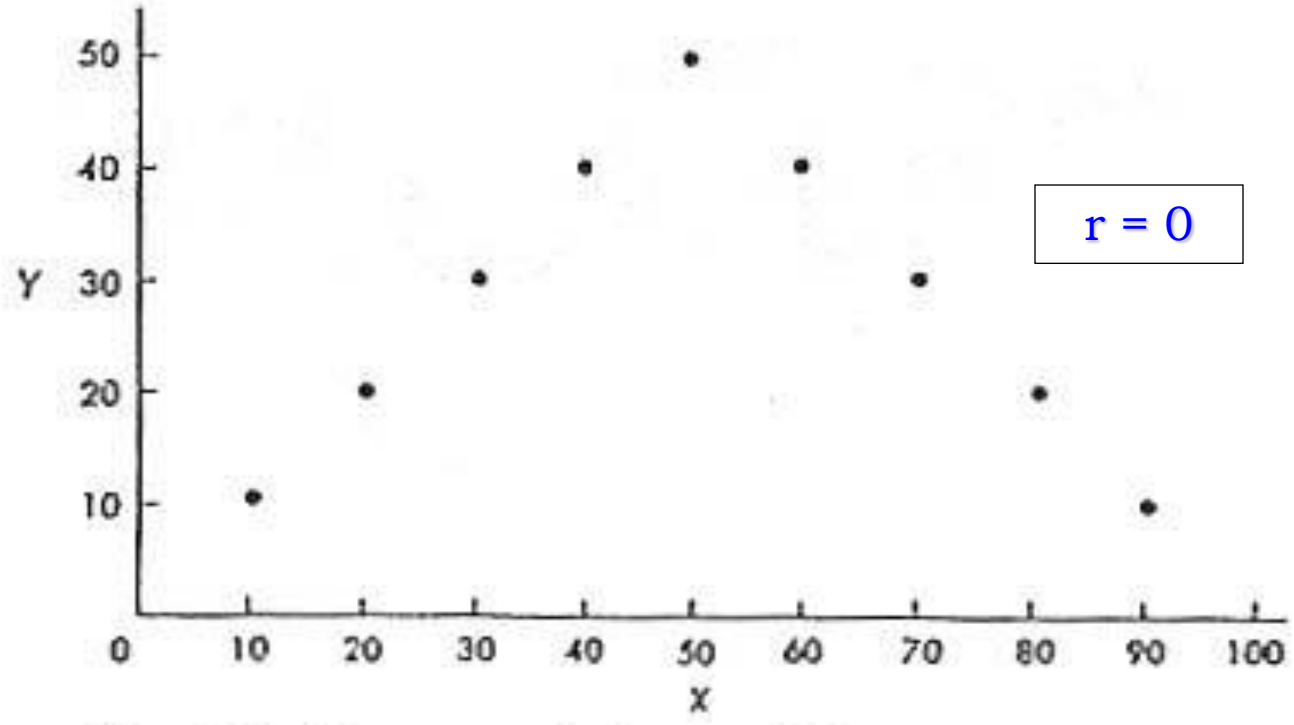
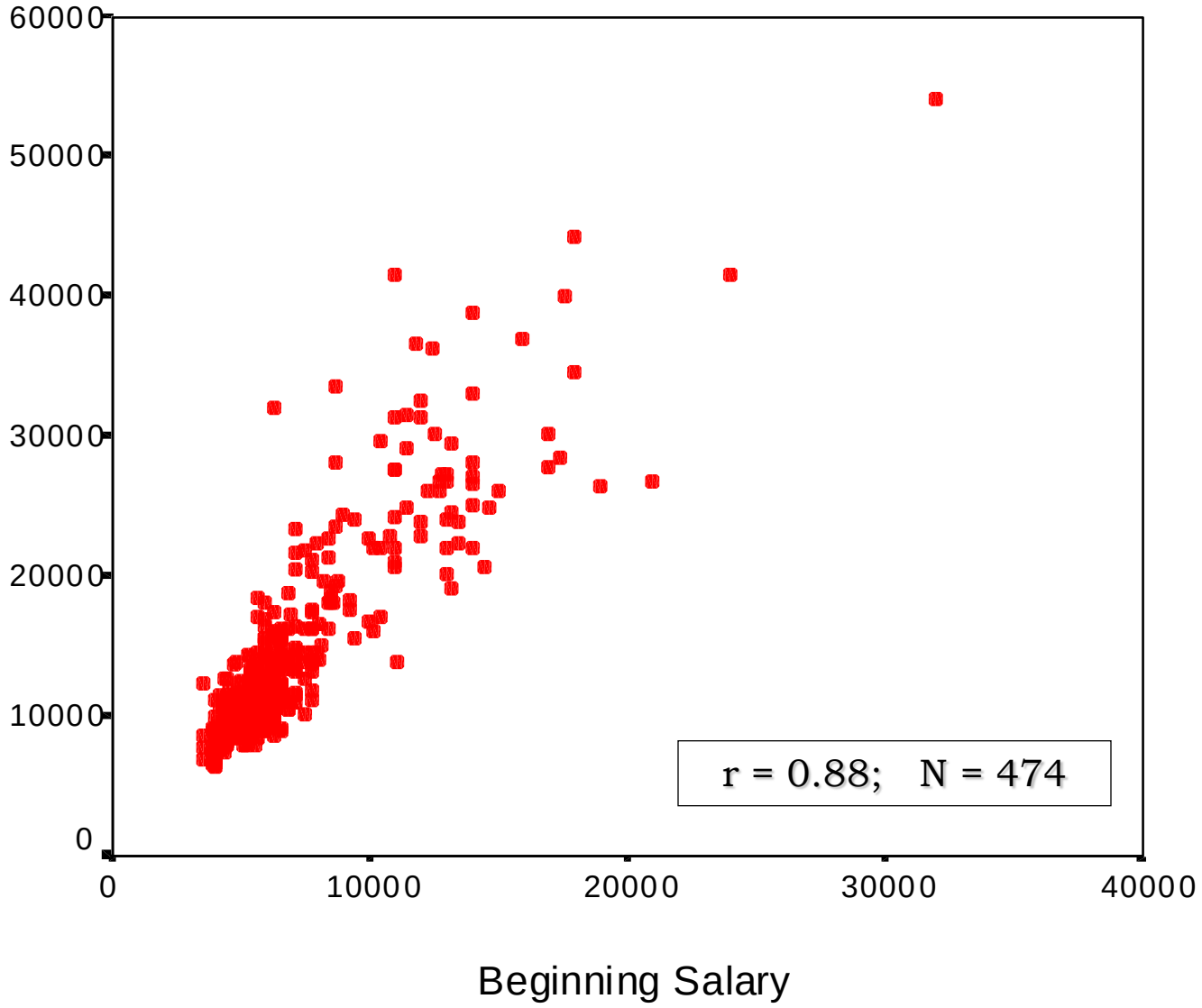


Figure 2–6 A scattergram showing two variables that are perfectly related but with a zero Pearson product moment *r*.

Performing Correlation Using SPSS

Scatterplot



Pearson Eg.1

Pearson Correlation	0.8335
Sig. (2-tailed)	0.0795
N	5

Pearson Eg.2

Pearson Correlation	-0.9499
Sig. (2-tailed)	0.0000
N	10

Pearson Eg.3

Pearson Correlation	0.1047
Sig. (2-tailed)	0.8669
N	5

Spearman Eg.1

Correlation Coefficient	-0.0952
Sig. (2-tailed)	0.8225
N	8