

CORRELATION AND SCATTER PLOTS

corr-sg(3)

Correlation coefficient (Pearson r) is a numerical representation of the relationship between two sets of data; the direction and the degree of the relationship.

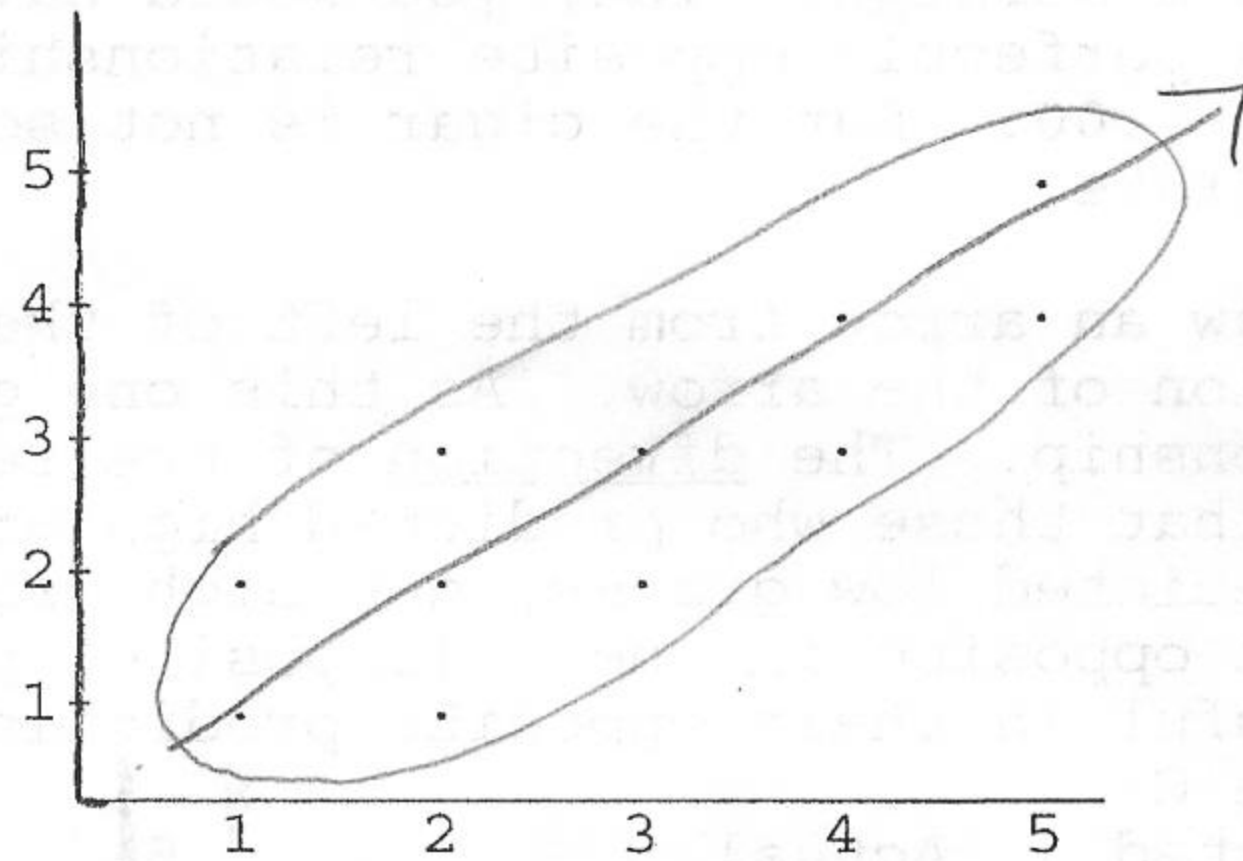
NOTE: this relationship does not imply a causal relationship

A scatter plot shows pictorially the relationship between two sets of data (variable X and variable Y), direction and degree. Given two sets of scores for a group of subjects, plot the points on the graph.

Examples: A class of education majors is thought to have knowledge about their own learning ability and grades. On the first day of class, and after reviewing the syllabus, the students are asked to predict the grade they expect to get in the course. Here are the results of the first class:

Example A:

Predicted Grade (X)	Actual Grade (Y)
5	5
5	4
4	3
4	4
3	3
3	2
2	1
2	3
2	2
1	2
1	1



Once you have plotted the points, draw the best cigar around the points that you can. The thickness of the cigar indicates the degree of the relationship. The thinner the cigar, the stronger the relationship. If all the points fell on a straight line, you would have a very thin cigar, and a perfect relationship.

Now draw an arrow from the left of the graph (0,0) to the right, and note the direction of the arrow. As this one goes up, you have a positive relationship. The direction of the relationship is positive. This means that those who predicted high grades, got high grades; and those who predicted low grades, got low grades. Although this group was not perfect in their predictions, they were moderately successful.

If all the points fell on a straight line, and the arrow from left to right went up, the correlation coefficient (r) would be +1.00

The correlation coefficient (Pearson r) ranges from -1.00 to +1.00. The closer to ± 1.00 , the stronger the relationship. The closer to 0, the weaker the relationship. if $r = 0.00$, there is no relationship (also called zero relationship or zero correlation).