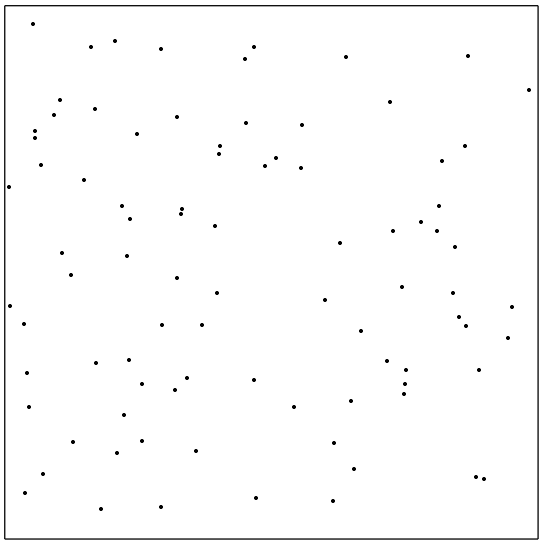
COUNTING DOTS

• 54 students were asked to count the dots in the square below without making any marks on their sheet, and then to count them again. (There are 87 dots).



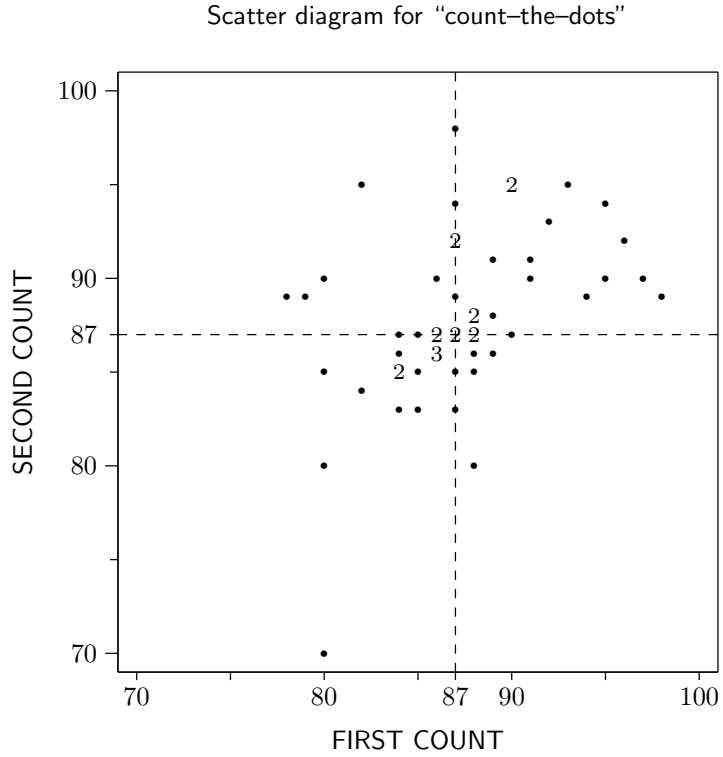
How do you think the results turned out?

• Would the first or second count tend to be more accurate? Why?

• Would there be any association between the first and second counts? Why?

• What's a good way to show how the results came out?

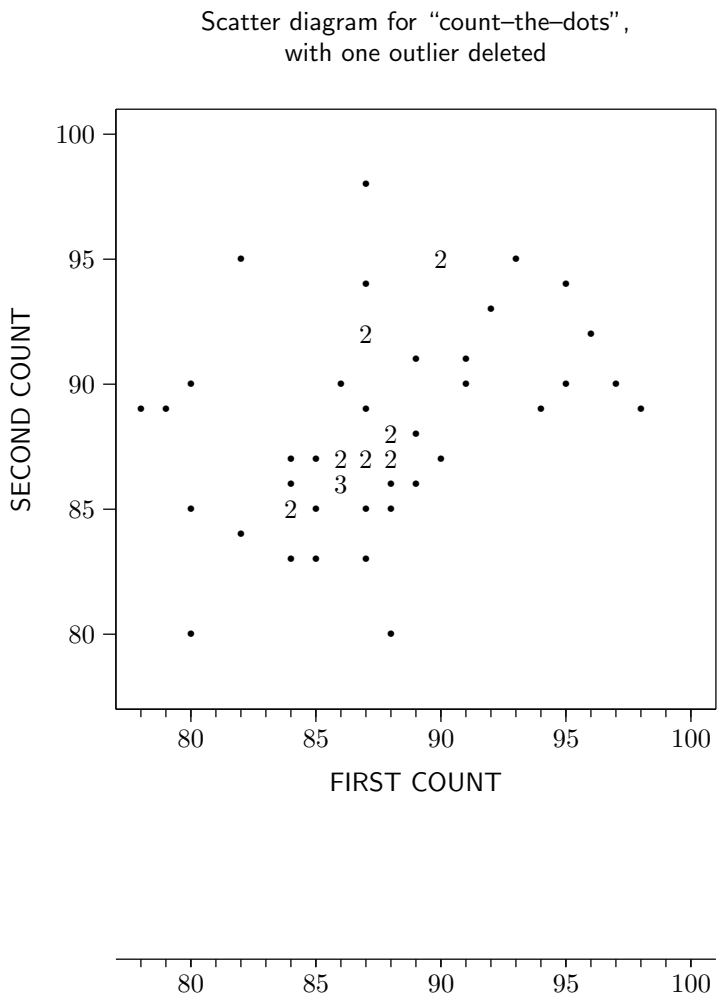
• The relationship between two variables can be represented visually by a SCATTER DIAGRAM.



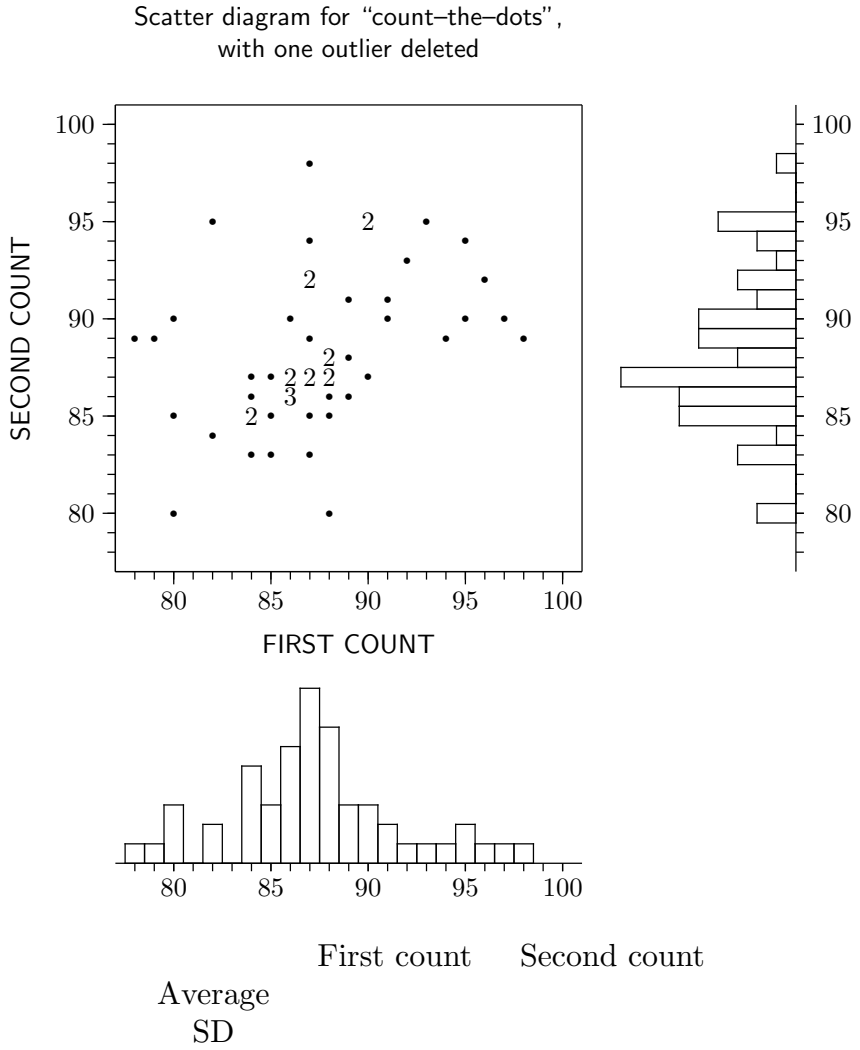
• How many students got the correct count the first time? _____ The second time? _____ Both times? _____.

• One point in the scatter diagram seems to be quite unusual. Which one? What are its coordinates? Is there an explanation?

- How can we find the distribution of the first count?



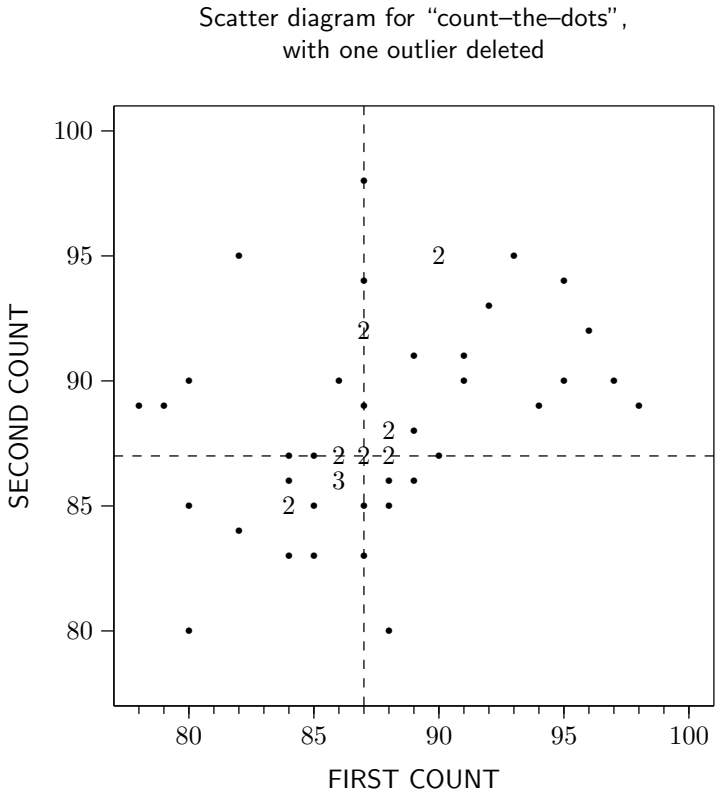
- Was the first or second count more accurate?



	First count	Second count
Average		
SD		

- How unusual is a first count of 70?

• Is there any association between the first and second counts?



Points per quadrant

	Left	Right
Upper		
Lower		

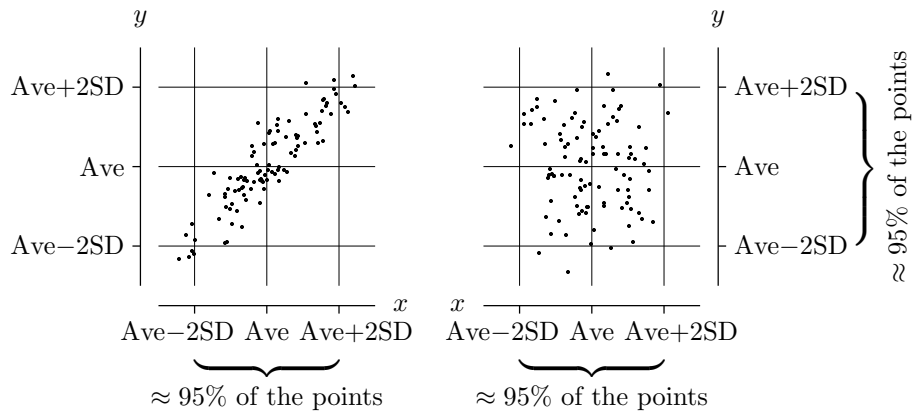
• How can we measure the strength of the association between the first and second counts?

SUMMARIZING A SCATTER DIAGRAM

• A scatter diagram can be summarized by means of five statistics:

- The average and SD of the x-values
- The average and SD of the y-values
- The correlation coefficient r

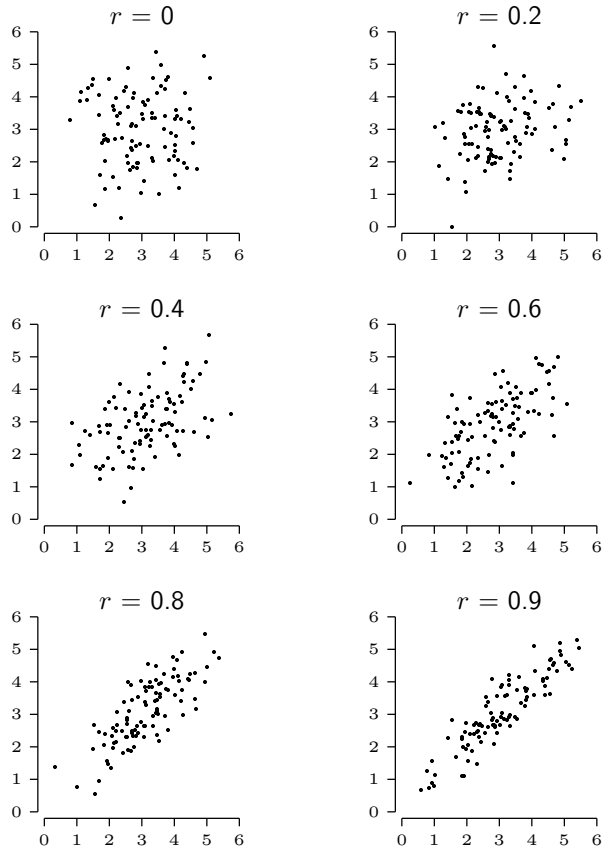
The averages and the SDs specify the location and spread of the cloud of points, horizontally and vertically. The correlation coefficient measures the amount of linear association, i.e., clustering about a line.



Strong association	Weak association
Tight clustering	Loose clustering
Knowing x helps in predicting y	Knowing x isn't much help in predicting y
Correlation near 1	Correlation near 0

EXAMPLES OF POSITIVE CORRELATIONS

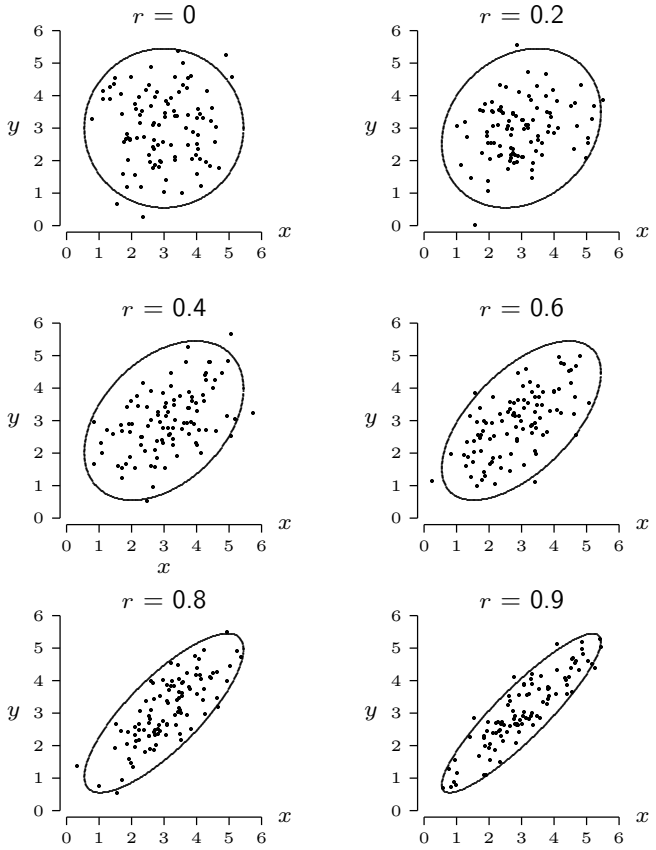
• In each diagram there are 100 points, with an average of 3 and and SD of 1, both horizontally and vertically. The correlation coefficient r measures the amount of clustering about a line.



r always lies between -1 and 1. What values of r are common in the social sciences? The biological sciences? The physical sciences?

EXAMPLES, CONTINUED

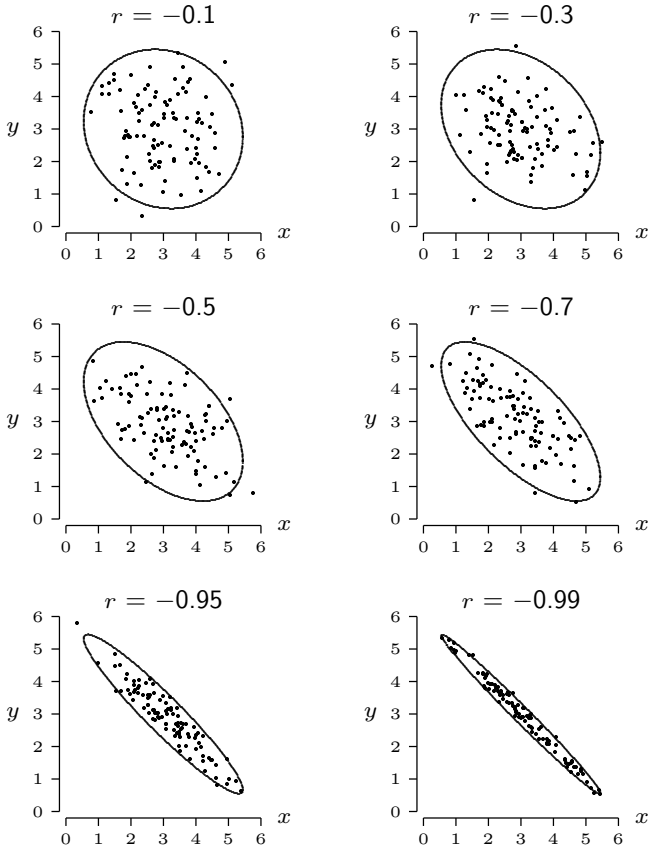
• Here are the same scatter diagrams, with ovals showing the main portion of the clouds of points.



Positive association (as x increases, y tends to increase also) is indicated by a sign in the correlation coefficient r. As r gets closer to 1, the points cluster more tightly around a line.

EXAMPLES OF NEGATIVE CORRELATIONS

• Here are some scatter plots with negative correlation coefficients. Again the average is 3 and the SD is 1, both horizontally and vertically.

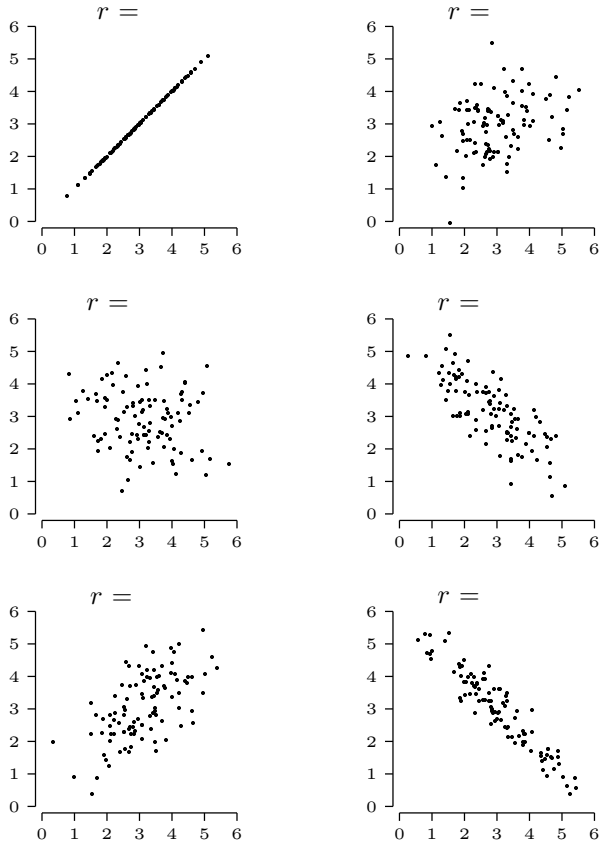


Negative association (as x increases, y tends to decrease) is indicated by a sign in the correlation coefficient r. As r gets closer to -1, the points cluster more tightly around a line.

EXERCISES

• In scrambled order, the correlation coefficients in the following diagrams are

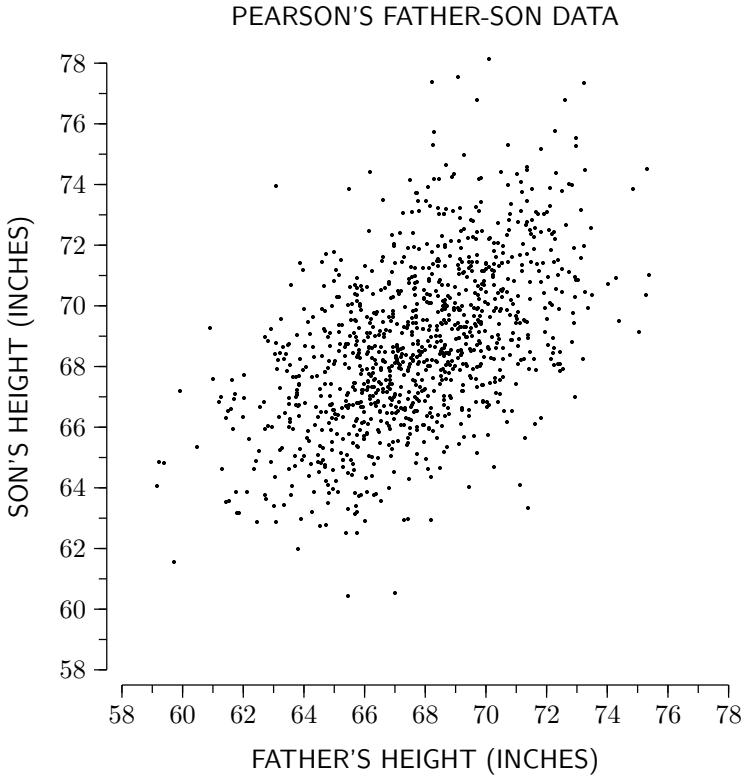
$$-0.93,\ -0.75,\ -0.20,\ 0.27,\ 0.63,\ \text{and}\ 1.0$$



Match the diagrams with the correlation coefficients.

• The diagrams on pages 8 and 9 can be used to guess r by eye, in cases where the spread of the scatter diagram from side-to-side is about the same as the spread up-and-down.

• The following scatter diagram shows the heights of 1,078 fathers and their full-grown sons. There is one dot for each father-son pair. The SD of height is 2.7 inches, both for the fathers and for the sons.

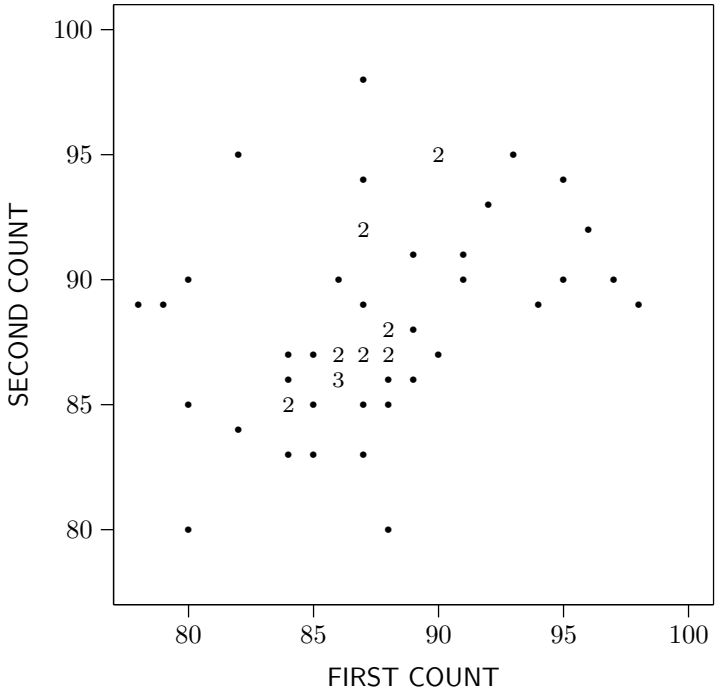


• Is the correlation between heights of fathers and heights of sons around −0.25, 0, 0.25, 0.5, or 0.75?

• If you took only the father-son pairs where the father was taller than six feet, would the correlation between the heights be around −0.25, 0, 0.25, 0.5, or 0.75?

• Is the correlation between the heights of husbands and wives in the U.S. around −0.7, −0.3, 0, 0.3, or 0.7? Explain.

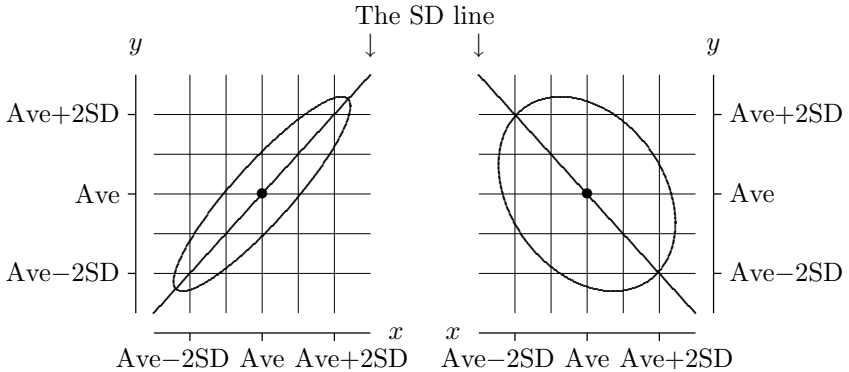
• Count-the-dots, revisited.



The horizontal and vertical SDs are about the same. Is the correlation coefficient about 0.1, 0.4, or 0.7? Explain.

THE SD LINE

• The points in a scatter diagram cluster (tightly or loosely) about the so-called SD line, which goes through all the points which are an equal number of SDs away from the average, for both variables.



- The SD line goes through the point of averages (marked by the "•" in the diagrams); this is the point with coordinates (_____, _____).
- When the correlation coefficient is positive, the line rises one vertical SD for each run of one horizontal SD.
- When the correlation coefficient is negative, the line falls one vertical SD for each run of one horizontal SD.

• One study on male college students found their average height to be 69 inches, with an SD of 3 inches. Their average weight was 140 pounds, with an SD of 20 pounds. And the correlation was 0.60. If one of these men is 72 inches tall, how heavy would he have to be to fall on the SD line?

COMPUTING THE CORRELATION COEFFICIENT

• To compute a correlation coefficient, you convert each variable to standard units and take the average of the products:

r = average of (x in standard units) × (y in standard units).

• A TA gives a quiz with 10 questions to 5 students. The results are:

Number right	Number wrong
10	0
8	2
7	3
6	4
4	6

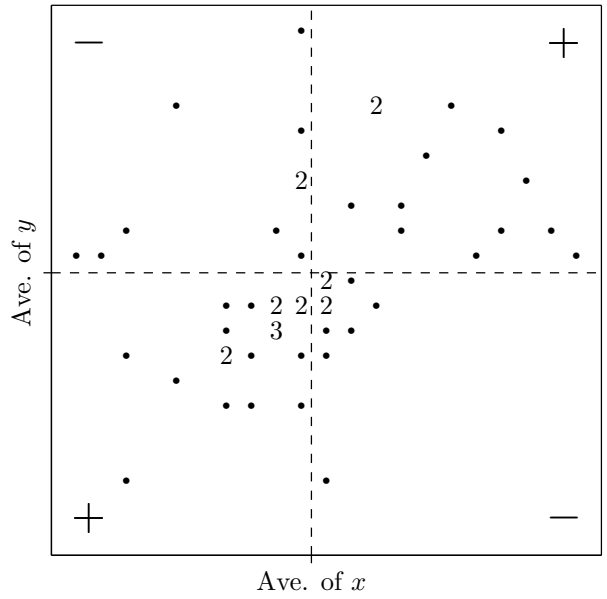
What is the correlation coefficient?

x	deviations	std units	y	deviations	std units	product
10			0			
8	1	1/2	2	−1	−1/2	−1/4
7	0	0	3	0	0	0
6	−1	−1/2	4	1	1/2	−1/4
4	−3	−3/2	6	3	3/2	−9/4
Ave	SD		Ave	SD		

$$r = \text{average product} = \frac{1}{-}\left(\frac{-1+0-1-9}{4}\right) = \text{—} =$$

• Is there an easier way to reach this conclusion?

• Why does r work as a measure of correlation? Consider the computation of r for the "count-the-dots" data:



In the upper-right and lower-left quadrants, the products

(x in standard units) × (y in standard units)

are _____, while in the remaining two quadrants the products are _____. The average all the products is the correlation coefficient. r is positive in this case because the points in the two positive quadrants predominate.