Section 12.6

Scatter Plots, Correlation, and Regression Lines.

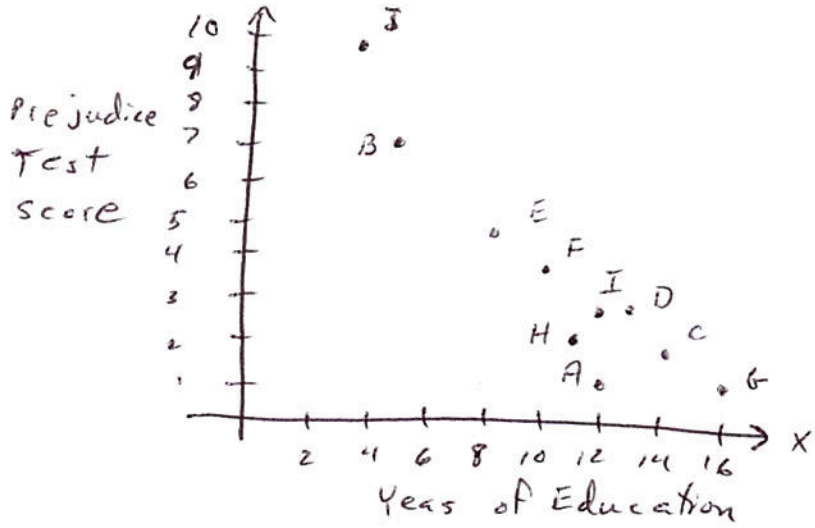
Scatter Plots

Table 12.15 pg 726

Person	A	B	C	D	E	F	G	H	I	J
Yeas of Ed (x)	12	15	14	13	8	10	16	11	12	4
Score on Prej Test (y)	1	7	2	3	5	4	1	2	3	10

High Prej test score (1-10) indicates prejudice.

Two data items per person



Scatter Plot of Data From Table 12.15

say

Are the quantities related?
Could we draw a line across the plot so that the data items are basically on the line?

If there is a clear relationship between two quantities, then we say that they are <u>correlated</u>.

Our scatter plot shows that more education corresponds to lower scores on prejudice test.

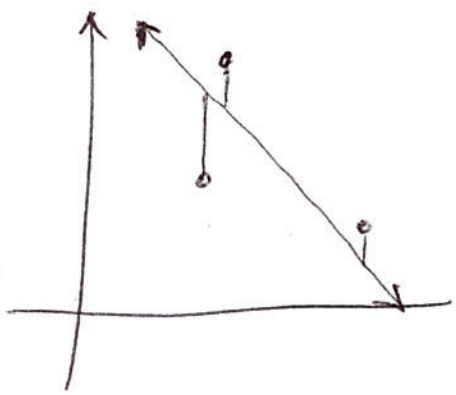
write

<u>Correlation</u> is used to determine if there is a relationship between two variables and, if so, the strength and direction of that relationship.

Correlation does not mean causation.

Regression Lines and Correlation Coefficients.

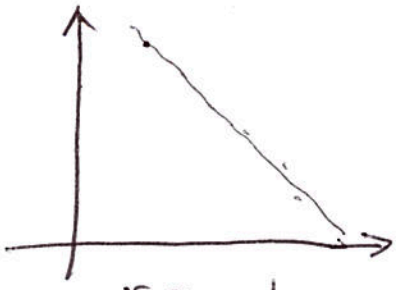
A line that best fits the data in a scatter plot is called a <u>regression line</u>.



<u>Best</u> means the line that minimizes the sum of the distances from the line to the data points, measured vertically.

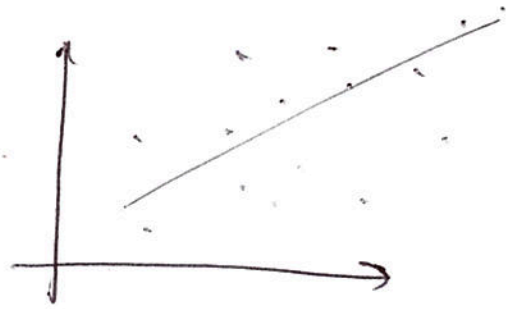
The <u>correlation coefficient</u>, designated by r, describes the strength and direction of a relationship between variables whose data points on or near a line.

$r \approx 1$ strong positive correlation



$r \approx -1$

strong negative correlation.



$r \approx 0.3$

weak positive correlation.

Figure 12.28 pg 727.

$|r|$ near 1 means strong correlation

$r > 0$ means as one quantity grows, the other grows.

$r < 0$ means as one quantity grows, the other shrinks.

Sec 12.6

The correlation coefficient does not tell the whole story.

If you only have two items in your sample, then $|r| = 1$.

The larger the sample size, the smaller $|r|$ needs to be to indicate a correlation in the population.

Table 12.16 pg 731.

$\alpha = 0.05$ column

sample size of 7. If $|r| > 0.754$, then the data in the population are correlated, with a 5% that we are wrong.

Example:

In a survey of 572 randomly selected service members it was determined that there was a -0.84 correlation coeff for years deployed and job satisfaction while the correlation coefficient for salary and job satisfaction was 0.75. which variable, years deployed or salary is a better indication of job satisfaction among service members?