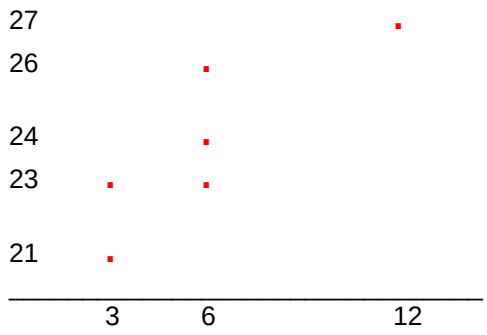
r = Correlation = a measure of linear association in bivariate data

= the average product of standard values

A researcher records the lifetimes of 6 products with various warranty periods:

x (warranty)	**3**	**3**	**6**	**6**	**6**	**12**	avr =	SD =
y (lifetime)	**21**	**23**	**23**	**24**	**26**	**27**	avr =	SD =



x	std val x	y	std val y	prod of SV's
3		21		
3		23		
6		23		
6		24		
6		26		
12		27		
				sum =

r =

Properties of r

(1) $-1 \leq r \leq +1$
(extreme cases: r = -1, 0, 1)

(2) r is a pure number (no units)

(3) r is scale invariant

(4) r can be affected by outliers

(5) high (+) or (-) r is not proof of cause & effect

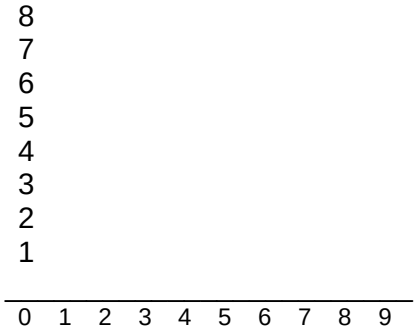
(6) r measures linear association.
(there are other forms of association)

(7) x,y assignments to variables
do not affect r

(8) r applies only to numerical variables (not categorical variables)

Ecological correlation

x	1	2	3	4	5	6	7	8	9
y	0	2	1	5	4	3	7	6	8



The heights (in inches) of 200 women were recorded at age 9 and at adulthood.

age 9 mean height = 55 SD(age 9) = SD(x) = 1.6
mean adult height = 66 SD(adult) = SD(y) = 2.4 r = .54

Individual prediction

(1) Susan is 59 inches tall as a 9 year old.
Predict her adult height.

(2) Jane is 53 inches tall as a 9 year old.
Predict her adult height.

(3) Helen is 55 inches tall as a 9 year old.
Predict her adult height.

The regression line used to predict y from x has **slope = rSD(y)/SD(x)**
and passes through the **point of averages**

In Broome County,

avr low temp = 53.68 SD(low) = SD(x) = 11.2
avr high temp = 70.79 SD(high) = SD(y) = 10.4 r = .765

Regression equation to predict high temp(y) from low temp (x): **y = mx + b**

- find m **m = rSD(y)/SD(x)** =
- replace m y =
- replace x by avr x and y by avr y
- solve for b b =
- write out the equation **y =**

(4) Find the regression line used to predict adult height from height at 9 years.

The participants in a walk-a-thon kept track of their distances and money raised. The following statistics were recorded:

average distance walked = 10.2 miles	SD = 2.8 miles	
average amount of money raised = $52	SD = $12	r = .84

Both distance walked and amount of money raised were normally distributed.

a) Jim walked 15 miles. Estimate the amount of money he raised.

b) John walked 5 miles. Estimate the amount of money he raised.

c) Susan was in the 90^{th} percentile for distance walked. Estimate her percentile for money raised.

d) Sara was in the 40^{th} percentile for distance walked. Estimate her percentile for money raised.

e) Write the **equation of the regression line** used to predict amount of money raised from distance walked. (Chapter 12)

At ACME P&V, worker salary (in thousands of dollars) and seniority (in years of service) are highly correlated. Here are the summary statistics:

mean seniority = 15 SD(seniority) = 3.6

mean salary = 42 SD(salary) = 5.4 r = .88

(1) Find the equation of the regression line used to predict salary from seniority.

Predict the salary of Sally Davis (25 years) and Roger Greene (5 years).

(2) Find the equation of the regression line used to predict seniority from salary.

Predict the seniority level of Mary Johnson ($55,200/year).

In problems (3) & (4), assume salaries & seniority levels are approx. normally distributed.

(3) Bob Smith's seniority level is in the 80th percentile. What is the predicted percentile of his salary?

(4) Barb Smith's seniority level is in the 36th percentile. What is the predicted percentile of her salary?