9.2 Correlation

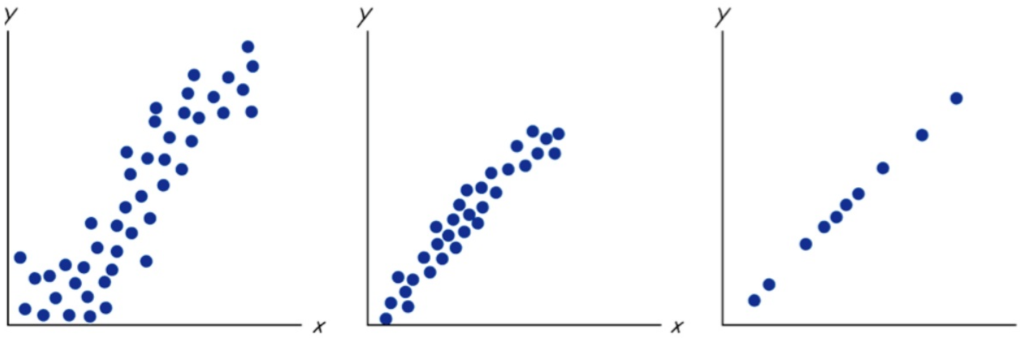
We are going to look at pairs of data to see if a relationship exists. If one does exist, we will make an equation that can be used to make predictions.

A correlation is a relationship between two variables.

We're only going to look at linear equations today.

You can look at a scatterplot to see if a correlation exists. You can also use it to determine the type of correlation. Is it strong/weak? Positive/negative?

Positive correlation

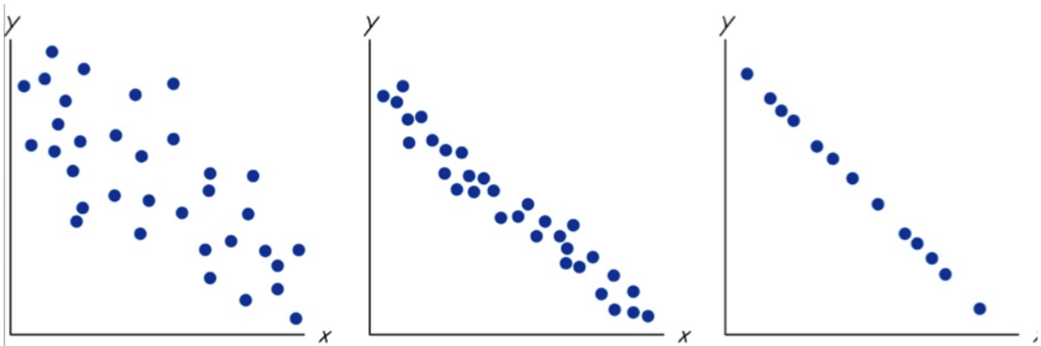


(a) Positive correlation between x and y

(b) Strong positive correlation between x and y

(c) Perfect positive correlation between x and y

Negative Correlation

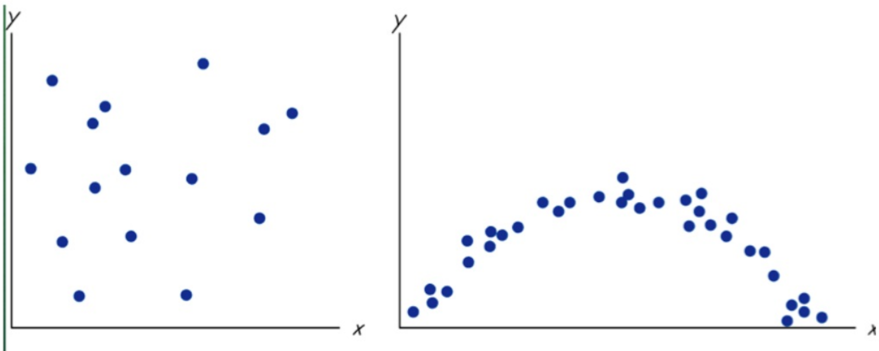


(d) Negative correlation between x and y

(e) Strong negative correlation between x and y

(f) Perfect negative correlation between x and y

No linear correlation

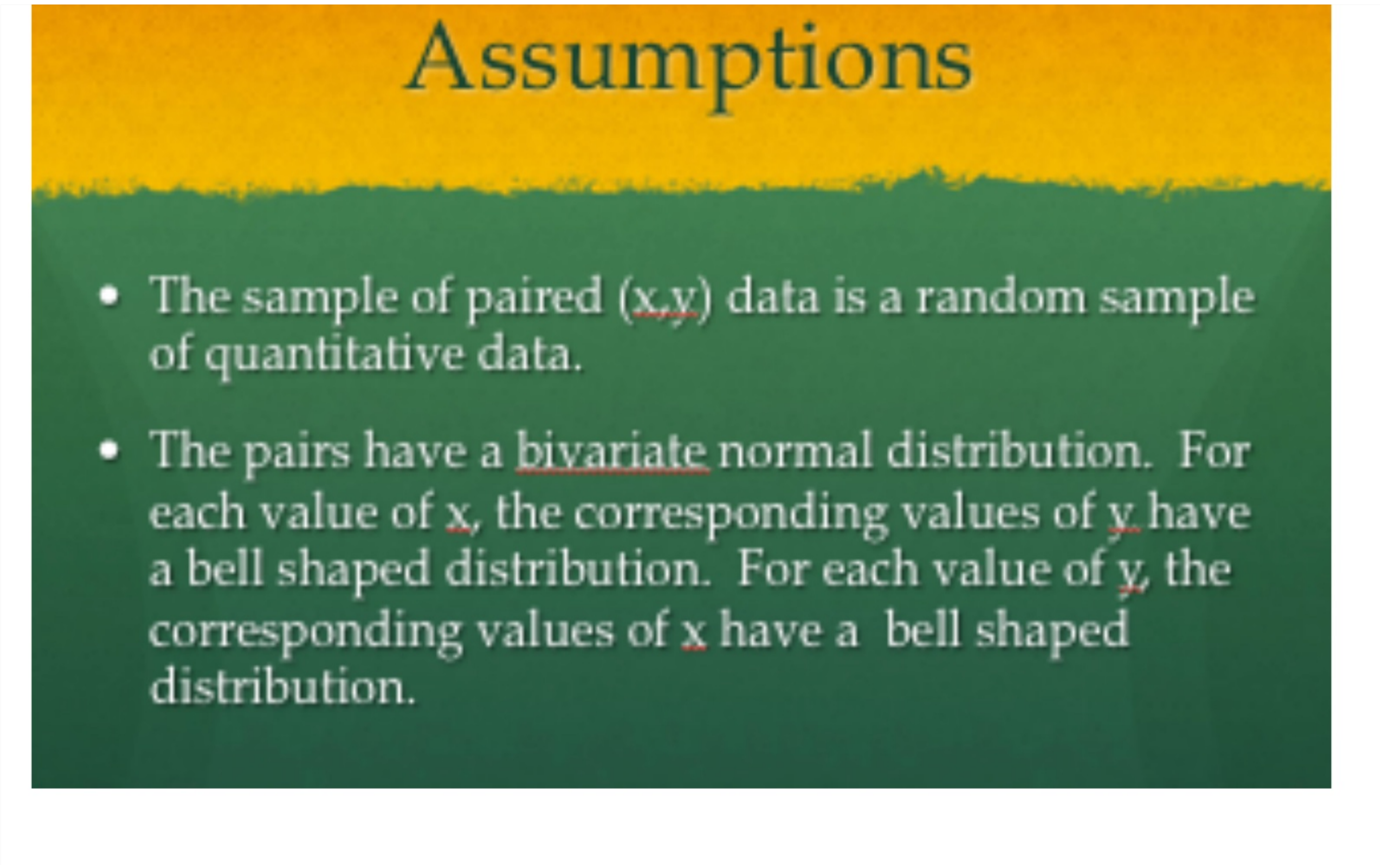


(g) No correlation between x and y

(h) Nonlinear relationship between x and y

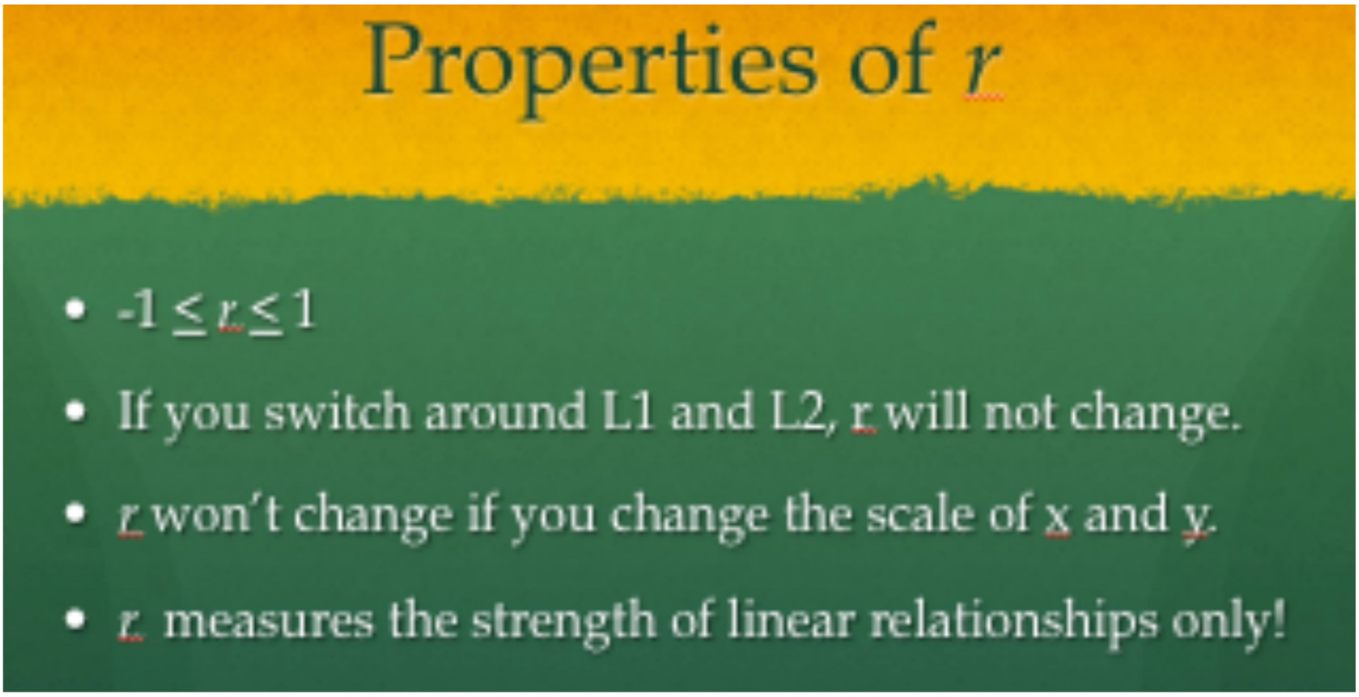
r measures the strength of the linear relationship between the paired x and y quantative data in a sample

The value of r will determine if there is significant linear correlation between paired data.


Assumptions
The sample of paired (x,y) data is a random sample of quantitative data.
The pairs have a bivariate normal distribution. For each value of x, the corresponding values of y have a bell shaped distribution. For each value of y, the corresponding values of x have a bell shaped distribution.

Interpreting *r*

- If the absolute value of *r* exceeds the value in Table A-6 where n= sample size, then a significant correlation exists. Otherwise, there isn't enough evidence to support the conclusion of a significant linear correlation.
- Ex. If α=0.05 and n=10 and no linear correlation between x & y, there is a 5% chance that the absolute value of *r* will exceed 0.632. (This value came from the table.)


Properties of r
-1 ≤ r ≤ 1
If you switch around L1 and L2, r will not change.
r won't change if you change the scale of x and y.
r measures the strength of linear relationships only!

If $|r| > r_{cv}$, then there is significant linear correlation.

p.742

1) n = 10; 95%; r = 0.623

$r_{cv} = .632$ $\quad |.623| \overset{?}{>} .632$ X

not sign. linear corr.

2) n = 15; 99%; r = -0.842

$r_{cv} = .641$ $\quad |-.842| \overset{?}{>} .641$ ✓

There is sign. linear correlation

Proportion of Variation

r^2

- The value of r^2 is the proportion of the variation in y that is explained by the linear relationship between x & y.
- Ex. If r=.755, then r^2 = *.570025*. This means that 57% of the data that can be explained by the linear relationship between the pairs of data. It also means that 43% of the variations in the data cannot be explained by the linear relationship.
- I'll show you a better example in a few minutes.

Example: The paired data below consits of test scores for 6 randomly selected students and the number of hours they studied for a test.

$n=6$

Hours	5	10	4	6	10	9
Score	64	86	69	86	59	87

a) Find the value of r. Is there a signficant linear correlation between the number of hours studying for the test and the test score?

$r=.224$ $\quad r_{cv}=.811$

$.224 < .811$
not sign. for linear corr.

b) What proportion of the test scores can be explained by the linear relationship? (What is the proportion of variation?

$r^2=.05=5\%$

5% of the data can be explained by the linear relationship; 95% is from other factors.

p. 510: 1-4 All
10 - 14 All