We've already been introduced to paired data. In this section, we are going to look at how well matched the pairs are.

Our first step is to graph the data. One set is graphed on the x-axis; the other on the y-axis.

The result is called a Scatter Diagram, or a Scatter Plot.

Once the points are plotted, we want to determine how closely they form a straight line. This is called linear correlation. The closer the points are to a line, the more correlated they are. The line can be slanted positively or negatively.

Correlation is a measure of how related a pair of data is. Some examples:

1. Height and weight are positively correlated because, generally, a taller person is heavier.

2. Age and height for children are positively correlated because, generally, an older child is taller.

3. Age and height for seniors are negatively correlated because, generally, on older person is shorter (due to loss of calcium in the spinal column).

4. Distance and size are negatively correlated because an object appears smaller as the distance increases.

The two type of correlation are:

Positive: as one increases (decreases), the other also increases (decreases).

Negative: as one increases (decreases), the other decreases (increases).

Linear Correlation Coefficient

Just because we can find a line, doesn't mean that the pairs of points are well correlated.

To determine how well correlated they are, we calculate the linear correlation coefficient.

The linear correlation coefficient is given by:

$$r = \frac{SS_{xy}}{\sqrt{SS_x SS_y}}$$

where:

$$SS_{xy} = n\sum(xy) - (\sum x)(\sum y)$$
$$SS_x = n\sum(x^2) - (\sum x)^2$$
$$SS_y = n\sum(y^2) - (\sum y)^2$$

To interpret the result, use Table A-5 to get the critical value for $|r|$.

Coefficient of Determination

The coefficient of determination tells us how much of the variation in the data is explained by the least-squares line.

The formula is given by:

$$r^2$$

Testing the Correlation Coefficient

1. $H_0:$ $\rho = 0$
2. $H_a:$ $\rho < 0$ (Left-tail) $\rho > 0$ (Right-tail) $\rho \neq 0$ (Two-tail)
3. $df = n - 2$, t-test, $n \geq 3$
4. $$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$