Regression

Correlation Coefficient

So you have the following data :

x	y
1	100
2	150
5	300
8	290
10	500
15	450

You plug these into your favorite statistics program and learn that the correlation coefficient is

$$r = 0.897080517 \ldots$$

Now what?

p-value

Turns out that one thing you can do is find a "p-value", which is the probability that random chance alone can account for a value of r that far from 0.

The details look like this :

$$\boldsymbol{H_0 = \text{null hypothesis} = \text{" r is zero ".}}$$

$$\boldsymbol{N = \text{number of data points} = 6 \text{ in this example}}$$

$$\boldsymbol{d.f. = \text{degrees of freedom} = N - 2}$$

$$\boldsymbol{t = \frac{r}{\sqrt{(1 - r^2)/(N - 2)}} = \text{Students} - t \text{ statistic}}$$

Then you do a "two-tailed test" to find the probability of getting a value of t that extreme by random chance alone.

in Excel

To do this with Excel, the function calls look like this :

```
r = CORREL ( Xarray, Yarray )

t = r / sqrt ((1 - r^2) / (N - 2))
```

```
p - value =  = TDIST ( t, N - 2, 2)
```

So there, mnnh.

Jim Mahoney, April 2005