BASIC STATISTICS

THE PEARSON r — Dr. A. A. Walsh

The Pearson "Product Moment" Correlation Coefficient, or Pearson r for short, allows you to determine if two sets of observations, on a single randomly selected sample, are co-related (correlated) with each other such that high and low values on one variable are associated with high and low values on the other or vice versa. The Pearson r calculates the degree of linear association (correlation) between the variables on a scale from -1.00 to +1.00. To use the Pearson r, the variables being measured must not depart significantly from normality, they must be measured at the interval or ratio level, the variation in both variables must be similar (i.e., homoscedastic), and their relationship must presumably be linear (curvilinear relationships can not be assessed by the Pearson r).

There are many ways of calculating r. For example, the Pearson r may be calculated from z-scores, from raw scores combined with means and standard deviations, and from raw scores directly. There are at least 6 commonly used formulas. THE RAW SCORE COMPUTATIONAL FORMULA IS SHOWN BELOW. Ho (null) holds that r = 0 (zero... no correlation).

$$r = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{\left[\sqrt{n\Sigma X^2 - (\Sigma X)^2}\right]\left[\sqrt{n\Sigma Y^2 - (\Sigma Y)^2}\right]}$$

n = number of pairs of scores

ΣXY = sum of the products of the paired scores

ΣX = sum of scores on one variable

ΣY = sum of scores on the other variable

ΣX^2 = sum of the squared scores on the X variable

ΣY^2 = sum of the squared scores on the Y variable

There are a number of methods for determining the significance of the Pearson r. One method employs a standard table(below). The Degrees of Freedom (df) for this test are equal to N - 2, where N equals the number of pairs of observations. If our calculated r is equal to or greater than the tabled value at a particular probability level then r is said to be significant at that level, *i.e.*, we can Reject Ho (null), ... if r is less, we must Accept Ho.

Finally, the COEFFICIENT OF DETERMINATION (r^2) can be calculated by squaring r and multiplying by 100. This value tells us the proportion of variability among the Y scores(the criterion) that can be accounted for by variability among the X scores(the predictor). Said another way, it tells us the amount of variation in Y explained by its linear connection to X. When r = 1.00 or -1.00, *i.e.*, when you have a perfect correlation, the **COEFFICIENT = 100%.**

Table E Critical values of r for the Pearson correlation coefficient.

df	p=.05	p = .01	df	p=.05	p = .01
1	.997	.9999	21	.413	.526
2	.950	.990	22	.404	.515
3	.878	.959	23	.396	.505
4	.811	.917	24	.388	.496
5	.754	.874	25	.381	.487
6	.707	.834	26	.374	.479
7	.666	.798	27	.367	.471
8	.632	.765	28	.361	.463
9	.602	.735	29	.355	.456
10	.576	.708	30	.349	.449
11	.553	.684	35	.325	.418
12	.532	.661	40	.304	.393
13	.514	.641	45	.288	.372
14	.497	.623	50	.273	.354
15	.482	.606	60	.250	.325
16	.468	.590	70	.232	.302
17	.456	.575	80	.217	.283
18	.444	.561	90	.205	.267
19	.433	.549	100	.195	.254
20	.423	.537			

If your calculated r is greater than table r, reject *Ho.* If your value of degrees of freedom is not listed, use the tabled r for the next smaller value of degrees of freedom.

Source: Table VI of Fisher and Yates: *Statistical Tables for Biological, Agricultural and Medical Research,* 6th edition, 1974, published by Longman Group Ltd., London; (previously published by Oliver and Boyd Ltd., Edinburgh).