Testing the Significance of a Correlation Coefficient

(W. J. Wilson, March 2015)

In the population that you care about there is a particular relationship between two variables x & y. If you characterized this relationship as linear, you could calculate an exact correlation coefficient that accurately represents the relationship for this population. That would be called rho (ρ) (a Greek letter because this is a population parameter).

Unfortunately, you cannot access ρ directly, because you cannot measure every member of the population. Instead you draw a sample from the population and calculate r, a sample statistic that you hope represents ρ. Based on the size of the r that you obtain, how likely is it that the underlying population can be characterized by a non-zero correlation coefficient?

Null Hypothesis: $\rho = 0$
Alternate Hypothesis: $\rho \neq 0$
$\alpha = .05$

What is the probability of getting an r value as large or larger than the one you obtained if the null hypothesis is true?

Consult a table. Table B.6 in the text has the critical values for r. These critical values are based on the same principle we used to evaluate the z-test: there is a pile of all possible z scores. Is our z near the middle or way out in one of the extreme tails? For r, there are *infinitely many piles of possible r values* (based on n, the number of pairs of observations). Is our r near the middle of its pile, or is it far out in one of the tails?

Degrees of Freedom: for r, df = n – 2. For any set of paired data points, any two pairs will always lie on a straight line (giving an r = 1.00). Therefore, only the pairs past those two affect the value of r: r is free to vary only for n – 2 data points.

If our r ≥ critical-r, then $p < \alpha$, and we reject the null hypothesis that $\rho = 0$, and instead conclude that a relationship exists in the underlying population.