Introduction to Power

Prerequisites
Significance Testing, Type I and Type II Errors, Misconceptions

Suppose you work for a foundation whose mission is to support researchers in mathematics education and your role is to evaluate grant proposals and decide which ones to fund. You receive a proposal to evaluate a new method of teaching high-school algebra. The research plan is to compare the achievement of students taught by the new method with those taught by the traditional method. The proposal contains good theoretical arguments why the new method should be superior and the proposed methodology is sound. In addition to these positive elements, there is one important question still to be answered: Does the experiment have a high probability of providing strong evidence that the new method is better than the standard method even if, in fact, the new method is actually better? It is possible, for example, that the proposed sample size is so small that even a fairly large *population* difference would be difficult to detect. That is, if the sample size is small, then even a fairly large difference in sample means might not be significant. If the difference is not significant, then no strong conclusions can be drawn about the population means. It is not justified to conclude that the *null hypothesis* that the *population* means are equal is true just because the difference is not significant. Of course, it is not justified to conclude that this null hypothesis is false. Therefore, when an effect is not *significant*, the result is inconclusive. You may prefer that your foundation's money be used to fund a project that has a higher probability of being able to make a strong conclusion.

Power is defined as the probability of correctly rejecting a false null hypothesis. In terms of our example, it is the probability that given there is a difference between the population means of the new method and the standard method, the sample means will be significantly different. The probability of failing to reject a false null hypothesis is often referred to as β. Therefore power can be defined as:

$$\text{power} = 1 - \beta.$$

It is very important to consider power while designing an experiment. You should avoid spending a lot of time and/or money on an experiment that has little chance of finding a *significant* effect.