Factors Affecting Power

Learning Objectives

1. State factors affecting power
2. State what the effect of each of the factors is

Several factors affect the power of a statistical test. Some of the factors are under the control of the experimenter, whereas others are not. The following example will be used to illustrate the various factors.

Suppose a math achievement test were known to be normally distributed with a mean of 75 and a *standard deviation* of σ. A researcher is interested in whether a new method of teaching results in a higher mean. Assume that although the experimenter does not know it, the population mean μ for the new method is larger than 75. The researcher plans to sample N subjects and do a one-tailed test of whether the sample mean is significantly higher than 75. In this section, we consider factors that affect the probability that the researcher will correctly reject the false *null hypothesis* that the *population* mean is 75 or lower. In other words, factors that affect power.

SAMPLE SIZE

Figure 1 shows that the larger the sample size, the higher the power. Since sample size is typically under an experimenter's control, increasing sample size is one way to increase power. However, it is sometimes difficult and/or expensive to use a large sample size.

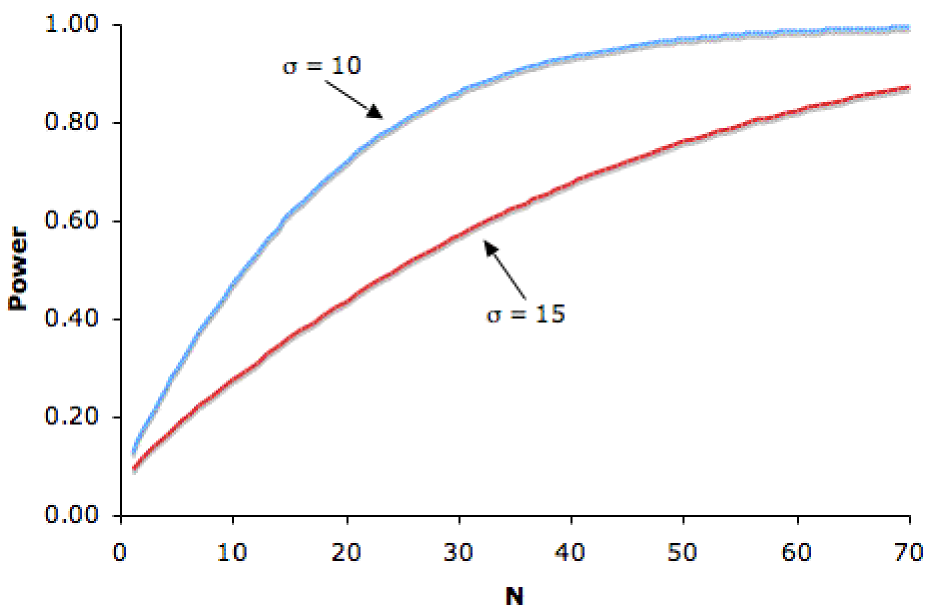


Figure 1. The relationship between sample size and power for H_0: $\mu = 75$, real $\mu = 80$, one-tailed $\alpha = 0.05$, for σ's of 10 and 15.

STANDARD DEVIATION

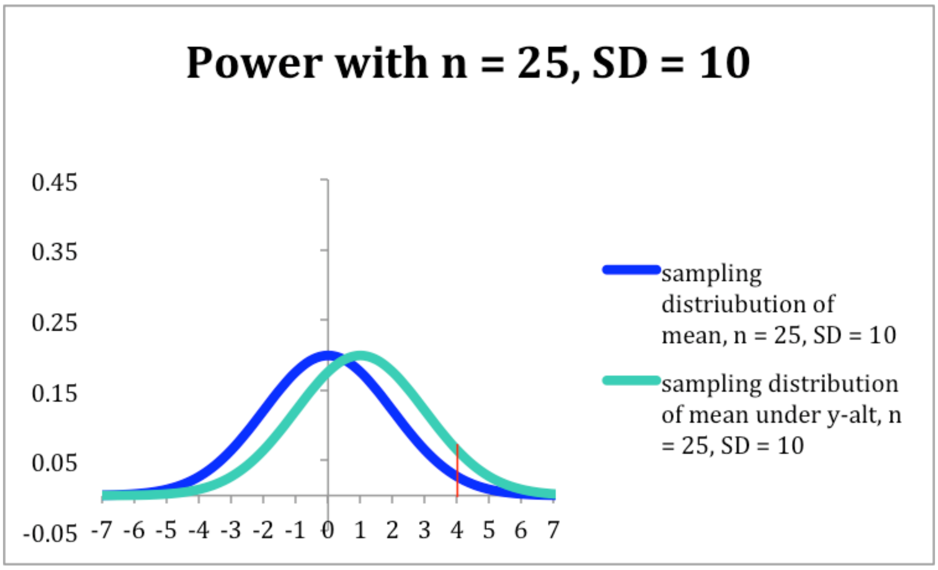
Figure 1 also shows that power is higher when the standard deviation is small than when it is large. For all values of N, power is higher for the standard deviation of 10 than for the standard deviation of 15 (except, of course, when N = 0). Experimenters can sometimes control the standard deviation by sampling from a homogeneous population of subjects, by reducing random measurement error, and/or by making sure the experimental procedures are applied very consistently.

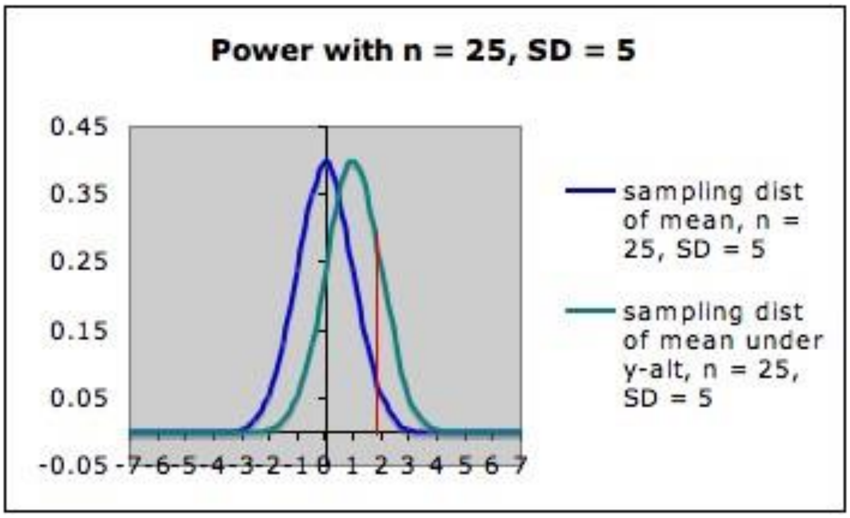
Variance

Power also depends on variance: smaller variance yields higher power.

Example: The pictures below each show the sampling distribution for the mean under the null hypothesis $\mu = 0$ (blue -- on the left in each picture) together with the sampling distribution under the alternate hypothesis $\mu = 1$ (green -- on the right in each picture), both with sample size 25, but *for different standard deviations of the underlying distributions*. (Different standard deviations might arise from using two different measuring instruments, or from considering two different populations.)

- In the first picture, the standard deviation is 10; in the second picture, it is 5.
- Note that *both graphs are in the same scale*. In both pictures, the blue curve is centered at 0 (corresponding to the null hypothesis) and the green curve is centered at 1 (corresponding to the alternate hypothesis).
- In each picture, the red line is the cut-off for rejection with alpha = 0.05 (for a one-tailed test) -- that is, in each picture, the area under the *blue* curve to the right of the red line is 0.05.
- In each picture, the area under the *green* curve to the right of the red line is the power of the test against the alternate depicted. Note that this area is *larger* in the second picture (the one with smaller standard deviation) than in the first picture.





DIFFERENCE BETWEEN HYPOTHESIZED AND TRUE MEAN

Naturally, the larger the effect size, the more likely it is that an experiment would find a significant effect. Figure 2 shows the effect of increasing the difference between the mean specified by the null hypothesis (75) and the population mean μ for standard deviations of 10 and 15.

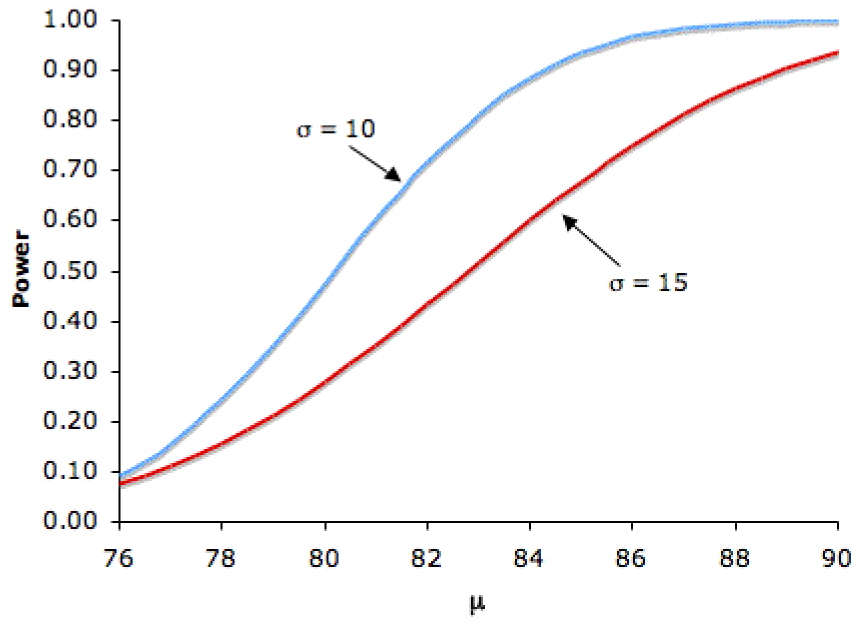


Figure 2. The relationship between μ and power for H_0: μ = 75, one-tailed α = 0.05, for σ's of 10 and 15.

SIGNIFICANCE LEVEL

There is a trade-off between the *significance level* and power: the more stringent (lower) the significance level, the lower the power. Figure 3 shows that power is lower for the 0.01 level than it is for the 0.05 level. Naturally, the stronger the evidence needed to reject the null hypothesis, the lower the chance that the null hypothesis will be rejected.

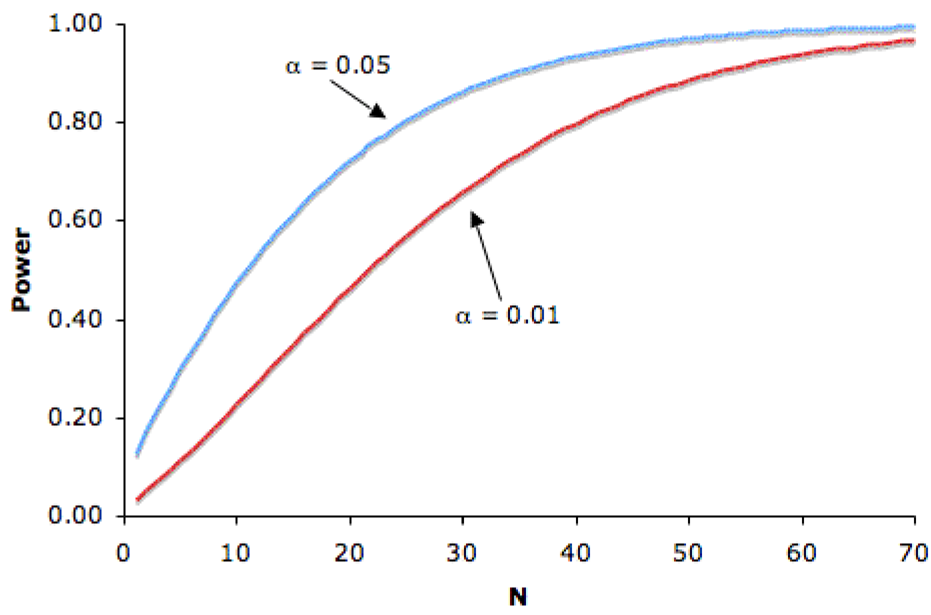


Figure 3. The relationship between significance level and power with one-tailed tests: μ = 75, real μ = 80, and σ = 10.

ONE- VERSUS TWO-TAILED TESTS

Power is higher with a *one-tailed* test than with a *two-tailed* test as long as the hypothesized direction is correct. A one-tailed test at the 0.05 level has the same power as a two-tailed test at the 0.10 level. A one-tailed test, in effect, raises the significance level.

References

https://web.ma.utexas.edu/users/mks/statmistakes/FactorsInfluencingPower.html

https://stattrek.com/hypothesis-test/power-of-test.aspx