

Statistical Power Analysis

Power is defined as $1 - b$ where b is the probability of committing a Type II error. Since a Type II error consists of deciding that the tested null hypothesis is not false when in fact it is false, $1 - b$ or power is the probability of correctly failing to reject the tested null hypothesis.

Power and b depend on the value of a (the probability of a Type I error, finding a tested null hypothesis to be false when it is not) chosen by the experimenter, the variability of the observed data, the sample size, and the specific alternative hypothesis against which the null hypothesis is being contrasted.

The concept of power is easy to illustrate graphically, but its precise calculation is more difficult. Suppose you had performed an experiment in which the density of *Peromyscus leucopus* had been estimated with individuals of *Peromyscus maniculatus* present or absent and obtained the following results:

Density (number/hectare) of *P. leucopus* *P. maniculatus*

| Present | Absent |
|--------------------|---------------------|
| 2.1 | 3.5 |
| 1.9 | 2.0 |
| 1.5 | 5.4 |
| 2.3 | 4.2 |
| 3.0 | 6.4 |
| 2.5 | 7.2 |
| 2.4 | 1.9 |
| 3.5 | 1.2 |
| 2.2 | 2.8 |
| 2.4 | 5.9 |
| | 2.4 |
| | 1.5 |
| | 1.2 |
| | 1.1 |
| | 1.7 |
| ----- | |
| $\bar{x}_p = 2.38$ | $\bar{x}_a = 3.226$ |
| $s_p = 0.555$ | $s_a = 2.086$ |
| $n_p = 10$ | $n_a = 15$ |

Assuming that these are independent estimates of the density of *P. leucopus* we could apply the *t* - test with variances assumed to be unknown and unequal to determine if *P. leucopus* density is altered by the presence or absence of *P. maniculatus*.

$$t = \frac{\bar{x}_a - \bar{x}_b}{\sqrt{\frac{s_p^2}{n_p} + \frac{s_a^2}{n_a}}}$$

If the null hypothesis of no difference in density regardless of the presence or absence of *P. maniculatus* is true, then this quotient will follow Student's - *t* distribution with $(n_a + n_p) - 2$ degrees of freedom (note that we should actually use Welch's correction here for degrees of freedom since variances are unequal, but I will not do so for computational brevity).

$$t = \frac{3.226 - 2.38}{\sqrt{\frac{4.351}{15} + \frac{0.308}{10}}}$$

$$t = 1.49$$

$$df = (10 + 15) - 2$$

$$df = 23$$

The probability of obtaining a *t* value of 1.49 or greater if the null hypothesis is true is $0.10 < p < 0.05$, $df = 23$ then we would fail to reject the null hypothesis because the critical value of *t* is 1.71. (Critical value of *t* from *t* - tables with $df = 23$).

The issue now becomes one of confidence in our decision. Did we perform a strong or "powerful" test of this null hypothesis or did we reach the decision to not reject the null hypothesis because of an inadequate database? Should we view our results as an indication that our experiment was unable to detect an effect that must be present, or as sound evidence that there truly is no effect of *P. maniculatus* on *P. leucopus* density? To answer these questions we must know the probability that we have incorrectly failed to reject the tested null hypothesis, and therefore the probability that the decision not to reject is correct, the power.

Under the tested null hypothesis the value of *t* is "centrally" distributed about its expected value $E(t) = 0$. As *df*'s increase, the shape of the distribution of *t* approaches normality $[N(0,1)]$. If we assume for the test specified above that the alternative hypothesis of interest is that *P. leucopus* density should increase in the absence of *P. maniculatus* then the region of rejection can be depicted as in Figure 1.

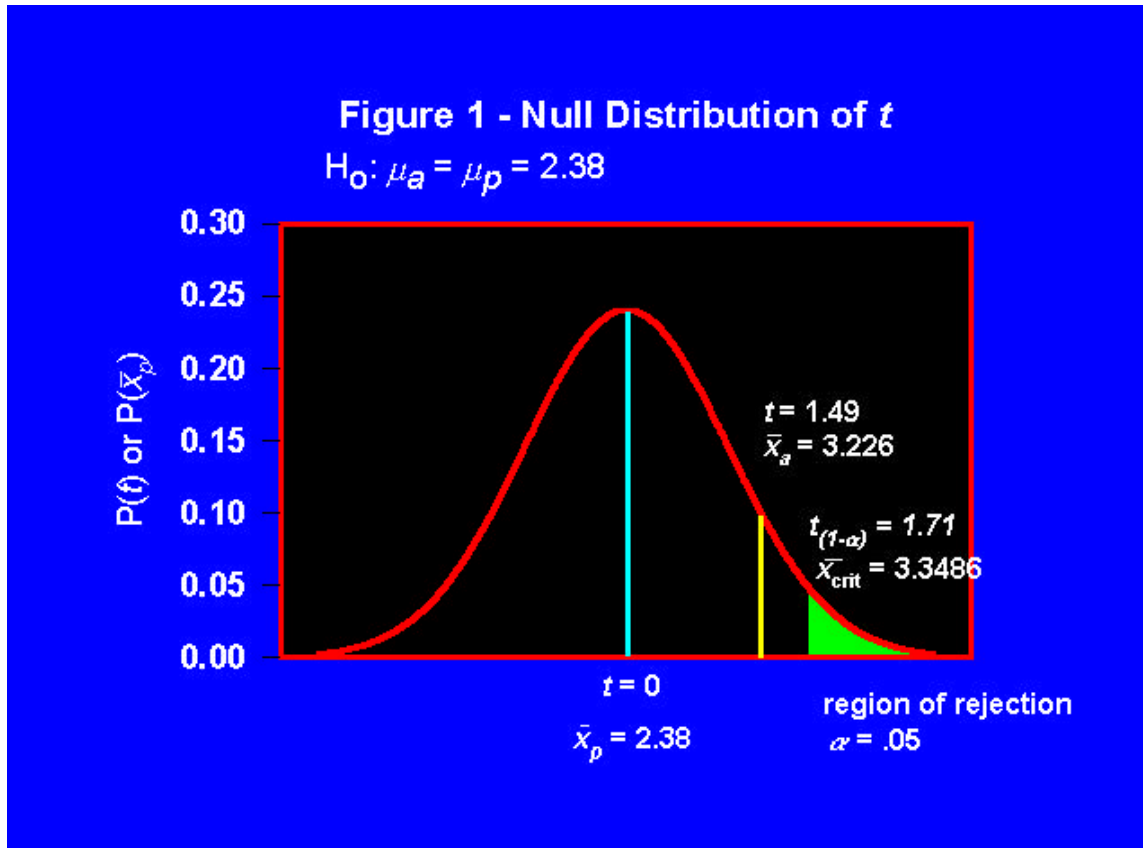


Figure 1. The curve depicts both the distribution of the observations and the distribution of t under the null hypothesis that the two population means m_a and m_p are equal. For sake of calculation we have fixed the before removal population mean and contend that after removal the under the null hypothesis no change should occur ($m_a = m_p = 2.38$). Under the null hypothesis the expected value of t is 0. For an α of 0.05 the minimum value necessary to reject the null hypothesis is 1.71. In terms of the value of \bar{x}_a , this t value translates into an \bar{x}_a of 3.3486. Since the computed t value from the actual data is 1.49 and it is associated with an \bar{x}_a value of 3.266 we decide not to reject the null hypothesis.

But, if the tested null hypothesis is actually false, some other hypothesis must be true. In this example it might be logical to expect a 35%, 50% or even 100% increase in the density of *P. leucopus* when *P. maniculatus* is absent. I chose 35% as the smallest alternative hypothesis of interest because the observed increase in density was 35%.

Now, if the correct hypothesis is that $\bar{x}_a = 1.35\bar{x}_p$ then the expected value of t under this alternative hypothesis is equal to the value we calculated previously (1.49) and the distribution of t under this alternative is about the value 1.49. The term "centrally" distributed is usually reserved for the distribution of a test statistic under the null

hypothesis and "non-central" is used to describe the distribution of a test statistic under a specific alternative hypothesis. Figure 2 shows the distribution of t under the null and alternative hypothesis. b is the area under the curve depicting the non-central t distribution (the one associated with the alternative hypothesis) falling within the region of rejection defined in Figure 1 in terms of the central distribution (the one associated with the null hypothesis). Power, again, is just $1 - b$

What remains is how to obtain a more or less precise estimate of this area. If we think of these data in terms of the original units of the observations, then the expected value of \bar{x}_a associated with the alternative hypothesis is 3.226 and the value of \bar{x}_a necessary to just enter the region of rejection under the null hypothesis is 3.3486. This is obtained by back calculating from the equation for t substituting ($\alpha = 0.05$, $df = 23$) $t = 1.71$. The question can be rephrased and an approximate estimate of obtained in this way. If a distribution has $\bar{x} = 3.226$ and $s = 1.66$ (this is obtained by pooling the variance estimates from the 2 samples populations, $s^2 = ((s_a^2(n_a - 1) + s_p^2(n_p - 1)) / (n_a + n_p - 2))$), then using the z transformation $z = (x - \bar{x}) / (s / \sqrt{n})$ we can determine the distance of a particular value, say the value of $\bar{x}_{crit} = 3.3486$ associated with the region of rejection from the value $\bar{x}_a = 3.226$ assumed under our alternative hypothesis. This z value is $z = (3.3486 - 3.226) / (1.66 / 5)$, which is 0.369. The area to the left of a z value of 0.369 under the normal curve is 64.4%. Therefore our first approximate estimate of $b = 64.4\%$ and of power is 35.6%, not a very comforting result. The chance of making a Type II error is very high indeed so the test has little power. Notice that we recast the problem in terms of the sample estimates of the distributions of the underlying (and assumed to be normal) populations to get this quick estimate of b but it is only a crude estimate.

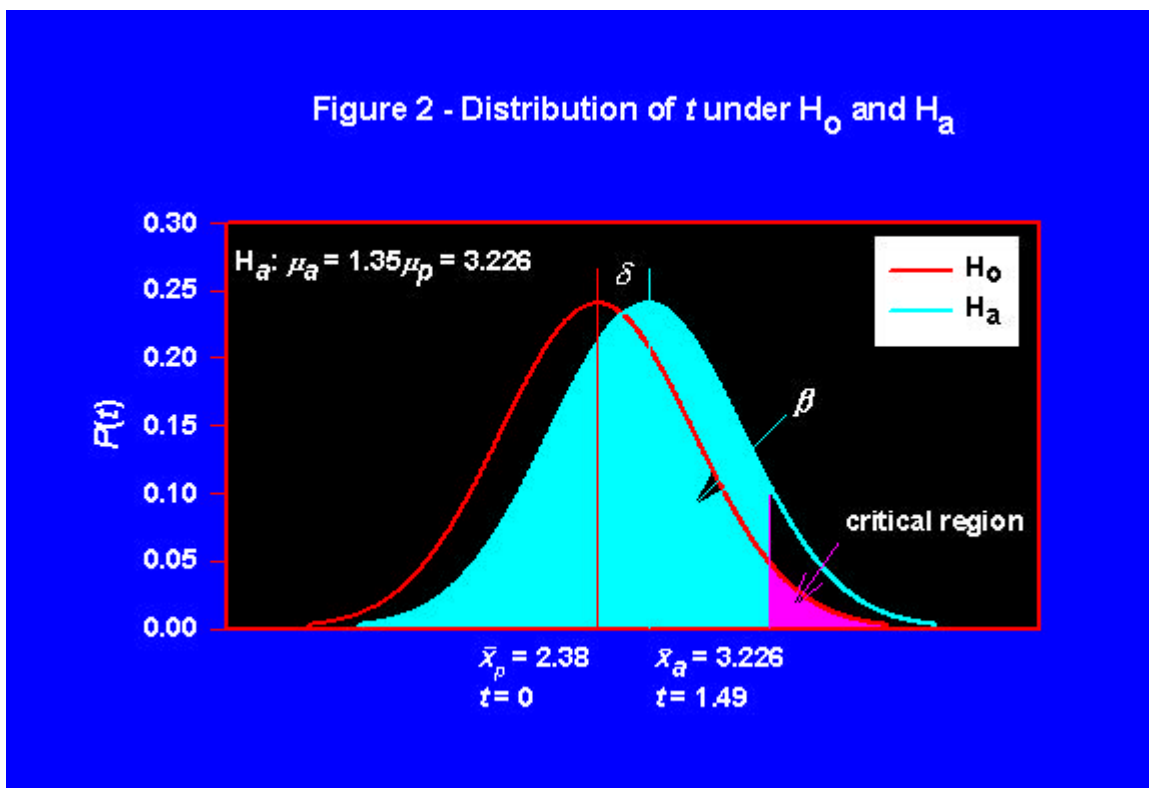


Figure 2. The curve on the left depicts the distribution of t and the observations under the null hypothesis, with the pink area indicating the region of rejection for an $\alpha = 0.05$ level test. The curve to the right illustrates the distribution of observations from the *P. maniculatus* absent sample and t under the alternative hypothesis that $m_a = 1.35m_p$ and therefore that $E(t) = 1.49$. The light blue region under this curve is b , the area lying within the region of non-rejection under the null hypothesis.

To summarize this procedure:

1. determine \bar{x}_{crit} associated with $t_{(1-\alpha)}$ (the critical region)
2. determine \bar{x}_a associated with H_a
3. Calculate z value using \bar{x}_a as the mean, \bar{x}_{crit} as x and s pooled as calculated above.
4. From tables of the standard normal distribution determine the area under this curve to the left or below z .
5. This value is b and power is $1 - b$

Two other procedures can be used to obtain more precise estimates of and power. These rely on estimating what is called the non-centrality parameter, d . Cohen (1977) also calls this or a related quantity the "effect size". It is a measure of the deviation between the value of the test statistic expected under the null and the alternative hypothesis. For the example above where the alternative hypothesis is that $m_a = 1.35m_p$. If we assume that m_p is 2.38, $m_a = 3.226$ and therefore $(m_a - m_p) = 0.846$.

$$\begin{aligned}
 d &= \frac{(m_a - m_p)}{s \sqrt{\frac{1}{n_a} + \frac{1}{n_p}}} \\
 &= \frac{0.846}{1.66 \sqrt{\frac{1}{10} + \frac{1}{15}}} \\
 &= 1.248
 \end{aligned}$$

If one has tables of non-central t distributions then one uses this additional parameter to read out the value of b for particular df 's and a). If tables are not available then one can use the formula:

$$z_b = \frac{t_{(1-a,df)} - d}{\sqrt{1 + [t_{(1-a,df)}^2 / 2df]}}$$

and refer z_b to the normal distribution to determine the area to the left of z_b and therefore the Type II error rate and power.

For the example above we could also have computed power for the alternative hypothesis of $m_a = 1.5m_p$, $m_a = 1.75m_p$, $m_a = 2.0m_p$, etc. These values can also be computed for different a values and for different sample sizes. This permits us to determine whether we should allow a higher a in order to decrease b and whether repeating the experiment with an increased sample size will give us the power we desire.

For example:

| | | |
|-----------------------|-------------|--------------|
| $H_a : m_a = 1.35m_p$ | $d = 1.248$ | $b = 64.4\%$ |
| $H_a : m_a = 1.5m_p$ | $d = 1.756$ | $b = 48.4\%$ |
| $H_a : m_a = 1.75m_p$ | $d = 2.630$ | $b = 18.7\%$ |
| $H_a : m_a = 2.0m_p$ | $d = 3.510$ | $b = 4.0\%$ |

(using the normal approximation of the non-central t distribution).

For the first alternative hypothesis listed above had been 0.2, 0.1, or 0.01 rather than 0.5 our estimates of b would have been 35.2%, 52.8%, 87.9%, respectively. Figure 3 shows in the form of a "power curve" the b values for various sample sizes.

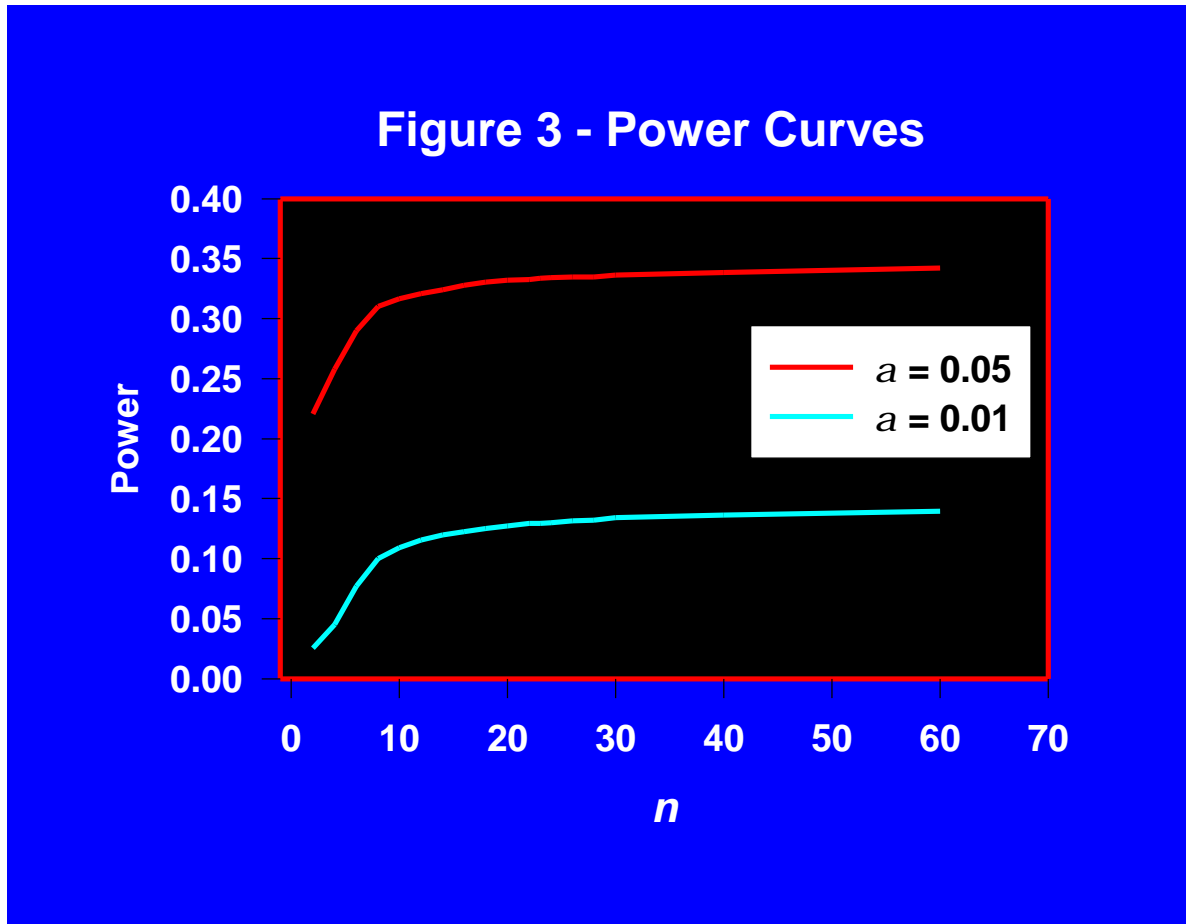


Figure 3. Power curves from the *Peromyscus* example for various α - levels and sample sizes. The curves give the power of the t - test with the estimated s pooled for specific combinations of α and with an alternative hypothesis of $m_u = 1.35m_p$. Additional curves could be plotted for other alternative hypothesis, α - levels, and s pooled estimates.

Literature Cited

Cohen, J. 1977. Statistical power analysis for the behavioral sciences. Academic Press; New York.