More Power Stuff

What is the statistical power of a hypothesis test? Statistical power is the probability of rejecting the null conditional on the null being false. In mathematical terms it is

$$P\left(\text{reject } H_0 \,|\, H_0 \text{ is false}\right)$$

Note the above probability cannot generally be computed. It cannot generally be computed because we have not specified the actual distribution that sample results (i.e. $\bar{x},\ \bar{p},\ \bar{x}_1 - \bar{x}_2,$ or $\bar{p}_1 - \bar{p}_2$) is drawn from. In order to calculate the power we need to affix values to the parameters of the distribution that the sample results are actually drawn from. To affix these values we may use results (estimates) from prior studies. For example, in Chill Pill we used estimates of the population variances from a smaller trial. We also used an estimate of the effect of another therapy to give us the mean of the distribution. Once we have mean and variance of the distribution that the sample results are drawn from the power can be computed straightforwardly as

$$P\left(\text{reject } H_0 \,|\, E(\ ) = A \text{ and } Var(\ ) = B\right)$$

where the expected value and variance operators, $E(\ )$ and $Var(\ )$, may be applied to $\bar{x},\ \bar{p},\ \bar{x}_1 - \bar{x}_2,$ or $\bar{p}_1 - \bar{p}_2$ (depending on whether the test relates to $\mu,\ p,\ \mu_1 - \mu_2,$ or $p_1 - p_2$), A is any value of the mean of the distribution of the sample results that satisfies the alternative hypothesis, and B is any non-negative constant.

How is power related to a Type II error? Conditional on the null being false two things can happen: you could reject the null or fail to reject the null (commit a Type II error). Because power is the probability of rejecting the null, conditional on the null being false, the probability of a Type II error may be computed as 1-power.

Steps for Computing Statistical Power
1. Figure out what values of the Z-stat (or t-stat) that will lead to rejection of the null. For a one tailed test this critical Z value is going to be Z_α (right tailed) or minus Z_α (left-tailed). For a two-tailed test the critical Z value is going to be $\pm Z_{\alpha/2}$.

2. Using this information on the values of the Z-stat that will lead to rejection of the null, figure out what values of the sample result $\left(\bar{x},\ \bar{p},\ \bar{x}_1 - \bar{x}_2,\text{ or } \bar{p}_1 - \bar{p}_2\right)$ will lead to rejection of the null. *For these computations you use the distribution of the sample result under the null.*

3. Calculate the probability of drawing a sample result $\left(\bar{x},\ \bar{p},\ \bar{x}_1 - \bar{x}_2,\text{ or } \bar{p}_1 - \bar{p}_2\right)$ that will lead to rejection of the null. *For these computations you use the actual distribution of the sample result.*

An Easier Power Example

Compute the power of the following hypothesis test assuming that $E(x_i) = 33$, $Var(x_i) = 100$, and a sample size of 100.

$$H_0 : \mu \leq 30$$
$$H_a : \mu > 30 \qquad \alpha = 0.05$$

We would like to compute

$$P\left(\text{reject } H_0 \mid E(\bar{x}) = 33 \text{ and } Var(\bar{x}) = 1\right)$$

The null is not correct if
$E(\bar{x}) = \mu = 33$

1. Note that

   reject $H_0 \Rightarrow$ compute a $p - value < 0.05 \Rightarrow$ compute a $Z - stat > Z_{0.05} = 1.645$ [1]

2. We can use this information to calculate the values of $\bar{x}$ that will lead to the rejection of the null. Basically we know that the null will be rejected as long as

   $$Z - stat = \frac{\bar{x} - 30}{1} > 1.645 \Rightarrow \bar{x} > 30 + 1.645 = 31.55$$

   Note that I am using the distribution of x-bar under the null to compute the values off x-bar that will lead the rejection of the null

---

[1] This is a right tailed test. If it were a left tailed test we would need $Z - stat < -Z_\alpha$ and if it were a two tailed test we would need $|Z - stat| > Z_{\alpha/2}$.

3. To complete the power calculation all we need to do is calculate the probability of drawing an $\bar{x}$ that is larger than 31.55. Because we know $\bar{x} \sim N\left(33, \dfrac{100}{100}\right)$ the computation is straightforward.

$$Z_{31.55} = \frac{31.55 - 33}{1} = -1.45 \Rightarrow P\left(\bar{x} > 31.55\right) = P\left(Z > -1.45\right) \approx 0.93$$

Note that I am using the actual distribution of x-bar to compute the probability off drawing a value of x-bar that will lead the rejection of the null

Thus, the power of the test is 0.93.

A Somewhat Harder Example

Consider a Moe and Thorton trial of the Chill Pill where Control Group I is the relevant control group and where the sample sizes are 144 (for convenience). Compute the power of the following hypothesis test

$$H_0 : \mu_T - \mu_C \le 0$$
$$H_a : \mu_T - \mu_C > 0 \qquad \alpha = 0.01$$

assuming population variances equal to the sample variances from the prior (smaller) trial and that the true effect of the Chill Pill is to increase test scores by 3 points.

_____

We would like to compute

$$P\left(\text{reject } H_0 \mid E\left(\bar{x}_T - \bar{x}_C \right) = \mu_T - \mu_C = 3 \text{ and } Var\left(\bar{x}_T - \bar{x}_C \right) = \frac{12.1^2 + 14.1^2}{144} \right)$$

1. Note that

   reject $H_0 \Rightarrow$ compute a $p - value < 0.01 \Rightarrow$ compute a $Z - stat > Z_{0.01} = 2.33$

2. We can use this information to calculate the values of $\Delta \bar{x} = \bar{x}_T - \bar{x}_C$ that will lead to the rejection of the null. Basically we know that the null will be rejected as long as

$$Z - stat = \frac{\Delta \bar{x} - 0}{\sqrt{\dfrac{12.1^2 + 14.1^2}{144}}} > 2.33 \Rightarrow \Delta \bar{x} > 3.61$$

3. To complete the power calculation all we need to do is calculate the probability of drawing an $\Delta \bar{x}$ that is larger than 3.61. Because we know
   $\Delta \bar{x} \sim N\left(3, \dfrac{12.1^2 + 14.1^2}{144} \right)$ the computation is straightforward.

$$Z_{3.61} = \frac{3.61 - 3}{\sqrt{\dfrac{12.1^2 + 14.1^2}{144}}} = 0.30 \Rightarrow P\left(\Delta \bar{x} > 3.61 \right) = P\left(Z > 0.39 \right) \approx 0.35$$

One Last Harder Power Example (with more guesswork needed)

I've chatted a bit about Moving to Opportunity (MTO) a few times during the semester. Basically, MTO was a randomized mobility experiment. In the experiment public housing residents in Boston that were on a very long waiting list were randomly assigned to one of three groups. The experimental (treatment) group received a voucher that they could use to move (immediately) to a low-poverty neighborhood (and only a low poverty area). Another group (the section 8 control group) received an un-targeted housing voucher while the remaining group remained on the long section 8 waiting list.

Assuming that the true effect of the treatment is to decrease the incidence of asthma attacks requiring treatment among children by 5 percentage points (0.05), calculate the power of the following test.

$$H_0 : p_T - p = 0$$
$$H_a : p_T - p_C \neq 0$$
$$\alpha = 0.01$$

You may assume that asthma rates in a comparable population are 0.12. Also assume sample sizes of 200.

_____

We would like to compute

$$P\left(\text{reject } H_0 \mid E\left(\overline{p}_T - \overline{p}_C\right) = p_T - p_C = -0.5 \text{ and } Var\left(\overline{p}_T - \overline{p}_C\right) = \frac{(0.07)(0.93) + (0.12)(0.88)}{200} \right)$$

> It is not clear what numbers to use here – thus picking numbers involves a little guess work. What we do know is that whatever numbers are used the difference between p_T and p_C should be exactly minus 0.05. I chose $p_T = 0.07$ and $p_C = 0.12$, but other choices are perfectly acceptable.

1. Note that

   reject $H_0 \Rightarrow$ compute a $p-value < 0.005 \Rightarrow$ compute $|Z - stat| > Z_{0.005} = 2.575$

2. We can use this information to calculate the values of $\Delta \bar{p} = \bar{p}_T - \bar{p}_C$ that will lead to the rejection of the null. Basically we know that the null will be rejected as long as

$$|Z - stat| = \left| \frac{\Delta \bar{p} - 0}{\sqrt{\dfrac{2 \cdot 0.095 \cdot 0.905}{200}}} \right| = \frac{|\Delta \bar{p}|}{\sqrt{\dfrac{2 \cdot 0.095 \cdot 0.905}{200}}} < 2.575 \Rightarrow |\Delta \bar{p}| > 0.076$$

> When we do a hypothesis where the null relates the difference of two population proportions to zero we need to assume the population proportions are equal. We have been dealing with this by using a weighted average of the sample proportions (p-tilda). As we don't yet have any sample information for this test – we have to make up some reasonable numbers. By picking 0.095 I'm basically guessing that the midpoint of the p-bar treatment and p-bar control (i.e. p-tilta) will be 0.095.

3. To complete the power calculation all we need to do is calculate the probability of drawing an $|\Delta \bar{p}|$ that is larger than 0.076. Because we know (are assuming) that

$$\Delta \bar{p} \sim N\left(5, \frac{0.07 \cdot 0.93 + 0.12 \cdot 0.88}{200} \right)$$ the computation is straightforward.

$$Z_{-0.076} = \frac{-0.076 - \bar{} 0.05}{\sqrt{\dfrac{0.07 \cdot (0.93) + 0.12 \cdot (0.88)}{200}}} = -0.89 \Rightarrow$$

$$P\left(\Delta \bar{p} < -0.076 \right) = P\left(Z < -0.89 \right) \approx 0.19$$

$$Z_{0.076} = \frac{0.076 - 0.05}{\sqrt{\dfrac{0.07 \cdot (0.93) + 0.12 \cdot (0.88)}{200}}} = 4.31 \Rightarrow P\left(\Delta \overline{p} > 0.076\right) = P\left(Z > 4.31\right) \approx 0$$

Thus the power is equal to

$$P\left(\left|\Delta \overline{p}\right| > 0.76\right) = P\left(\Delta \overline{p} > 0.76\right) + P\left(\Delta \overline{p} < -0.76\right) \approx 0.19$$

If the true difference in asthma rates between a treated population (a hypothetical population) and the control population is 0.05, the above test will lead to the correct rejection of the null 19 percent of the time assuming sample sizes of 200. In other words – we would only be able to detect a statistically significant of the treatment (at significance level 0.01) in 19 percent of the sample draws (sample experiments). If you were designing this study, and you wanted a higher probability of being able to detect a statistically significant effect of the treatment the incidence of serious asthma attacks, you would want a larger sample.