Power and Sample Size

Fixed α Testing

- Recall that in this approach, a small significance level α is provided, often .05.
- If $pvalue \leq \alpha$ we conclude that the results are statistically significant at the α level.
- If $pvalue > \alpha$ we conclude that the results are not statistically signficant at the α level.

Errors in Fixed α Testing

- There are two kinds of errors which can be made.
- Type I: we find statistical significance when H_0 is true.
- Type II: we fail to find statistical significance when H_0 is false.

Conclusion	H_0 true	H_0 false
significant	type I	
not significant		type II

- The probability of a type I error, the significance level, $P(type\ I) = \alpha$, is given.
- A type I error is considered much worse than a type II error, so α is typically small.
- The probability of a type II error is denoted $P(type\ II) = \beta$.
- The power is the probability of correctly finding statistically significant evidence against H_0

$$Power = 1 - \beta = P(reject\ H_0\ when\ H_0\ false).$$

- We can increase power (reduce β) by increasing the sample size.
- Formulae for calculating power, and for determining the sample size required to achieve desired power are given below. You can check your calculations with the applet available at

  http://www.stat.ubc.ca/ rollin/stats/ssize/

- We examine three cases
  1. paired comparisons of two means
  2. independent comparisons of two means
  3. comparing two binomial probabilities
- In each case we will need to specifiy how different the null and alternative hypothesis are.
- We will also need to assume that we know the variance of the population we are sampling from.
  - Information about the variance is often available from previous studies.
- The following table lists commonly used standard normal deviate values (z_α and z_β) used in power and sample size calculation. Note: a correction to the table: For z_α, the second row (Direction of testing being Two-sided), $\alpha = 0.025$ should be $\alpha = 0.02$ so that $Z_\alpha = 2.326$).

	Direction of testing	**α or β**	**Value**
Z_α	Two-sided	α=0.05	Z_α=1.960
	Two-sided	α=0.025	Z_α=2.326
	Two-sided	α=0.01	Z_α=2.576
	One-sided	α=0.05	Z_α=1.645
	One-sided	α=0.025	Z_α=1.960
	One-sided	α=0.01	Z_α=2.326
Z_β		β=0.20	Z_β=0.840
		β=0.10	Z_β=1.282

Power analysis for paired sample t-test

- The hypotheses are
  $H_0 : \mu_d = 0$ versus $H_a : \mu_d = \delta > 0$,

  where μ_d is the mean of the differences of the paired observations, and δ is the mean difference when H_a is true.

- The test statistic is
$$T = \frac{\bar{d}}{s_d/\sqrt{n}}$$
  where $\bar{d}$ and s_d are the sample mean and standard deviation of the differences of the paired observations.

- If we assume the variance σ_d^2 of the differences is known, the test statistic is
$$T = \frac{\bar{d}}{\sigma_d/\sqrt{n}}.$$

- We declare statistical significance at level α if $P \leq \alpha$, or equivalently if $T \geq z_\alpha$ where z_α is the upper αth quantile of the standard normal distribution.

- We can calculate the power, or probability of finding significance, as
$$\begin{aligned} Power &= P(P \leq \alpha) \\ &= P(T \geq z_\alpha) \\ &= P(\bar{d} \geq z_\alpha \sigma_d/\sqrt{n}). \end{aligned}$$

- Now we standardize both sizes of the inequality by subtracting the mean difference under H_a and dividing by the standard error $\sigma_d/\sqrt{n})$
$$\begin{aligned} Power &= P(\frac{\bar{d}-\delta}{\sigma_d/\sqrt{n}} \geq z_\alpha - \frac{\delta}{\sigma_d/\sqrt{n}}) \\ &= P(Z \geq z_\alpha - \frac{\delta}{\sigma_d/\sqrt{n}}) \end{aligned}$$

- So
$$Power = 1 - \Phi(z_\alpha - \frac{\delta}{\sigma_d/\sqrt{n}}). \quad (1)$$

- The symbol Φ refers to the cumulative probability distribution function for the normal distribution.

- We use the normal rather than the t distribution because we have assumed we know σ_d.

- These probabilities can be obtained from the normal tables.

- The power increases with δ, the difference between the null and alternative values, from α when $\delta = 0$ to one when δ is large.

- Power also increases with the number of pairs n.
- Power decreases with the standard deviation of the differences σ_d, although we don't usually have any control over this quantity.
- Given a desired value for power, we can determine the sample size required by rearranging the formula (1) above

$$\begin{aligned}
1-\beta &= 1-\Phi(z_\alpha - \frac{\delta}{\sigma_d/\sqrt{n}}) \\
\beta &= \Phi(z_\alpha - \frac{\delta}{\sigma_d/\sqrt{n}}) \\
\Phi^{-1}(\beta) &= z_\alpha - \frac{\delta}{\sigma_d/\sqrt{n}} \\
-z_\beta &= z_\alpha - \frac{\delta}{\sigma_d/\sqrt{n}} \\
\sqrt{n} &= \frac{\sigma_d(z_\alpha + z_\beta)}{\delta} \\
n &= \frac{\sigma_d^2(z_\alpha + z_\beta)^2}{\delta^2}. \qquad (2)
\end{aligned}$$

- We should round the answer up to the nearest integer to be sure of achieving the required power.
- The required sample size increases with σ_d, z_α and z_β, but decreases with δ.
- For a two-sided alternative, the formulae (1) and (2) are used with $z_{\alpha/2}$ in place of z_α.

Example: A researcher wishes to be able to detect a mean difference of .2 units using paired data and a one-sided alternative with $\alpha = .05$. She believes the standard deviation of the differences is $\sigma_d = .5$.

1. What would be the power if she uses 36 pairs?
   - We use formula (1) with $n = 36$, $z_{.05} = 1.645$ and $\delta = .2$.

$$\begin{aligned}
Power &= 1-\Phi(1.645 - \frac{.2}{.5/\sqrt{36}}) \\
&= 1-\Phi(1.645-2.4) \\
&= 1-\Phi(-.755) \\
&= 1-.2251 = .7749
\end{aligned}$$

2. How many pairs are required so that there is a 90% power of detecting a difference of .2 units using a one-sided alternative with significance level $\alpha = .05$? Suppose $\sigma_d = .5$.
   - Using formula (2) with $\delta = .2$, $\alpha = .05$, $\beta = .1$

$$n = \frac{.25(1.645+1.282)^2}{.04} = 53.47$$

   - We should use $n = 54$ pairs (rounding up).

Power analysis for two group independent sample t-test
Comparing means of two normal populations of equal variances, based on two independent samples.

- To test the hypotheses
  $H_0 : \mu_1 - \mu_2 = 0$

  versus

  $H_a : \mu_1 - \mu_2 = \delta > 0$

  using independent samples, the test statistic is

$$T = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

- The formula for power, when the variance is assumed known, is

$$Power = 1 - \Phi \left(z_\alpha - \frac{\delta}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right).$$

- If the sample sizes are equal in the two samples, this is

$$Power = 1 - \Phi \left(z_\alpha - \frac{\delta}{\sigma \sqrt{2/n}} \right). \qquad (3)$$

- Once again, the power increases with δ, n (or n_1 and n_2), but decreases with σ.
- To determine the sample size n required in each sample to achieve a given power, we use the equation

$$n = \frac{2\sigma^2 (z_\alpha + z_\beta)^2}{\delta^2}. \qquad (4)$$

- As before, for a two-sided alternative, we use $z_{\alpha/2}$ in place of z_α.

Example: A researcher wishes to detect a mean difference of 1.5 using a two-sided level $\alpha = .05$ test, and believes the standard deviation of measurements in each population is $\sigma = 2$.

1. What is the power if she uses $n_1 = n_2 = 15$?

   - Use $z_{\alpha/2} = 1.96$ in equation (3), so

$$\begin{aligned} Power &= 1 - \Phi(1.96 - \frac{1.5}{2\sqrt{2/15}}) \\ &= 1 - \Phi(-.094) \\ &= 1 - .462 = .537. \end{aligned}$$

2. What sample size is required to acheive 80% power?

   - Use $z_{\alpha/2}$ for a two-sided alternative in equation (4).
   - $z_\beta = .84$

$$\begin{aligned} n &= \frac{2(2^2)(1.96 + .84)^2}{(1.5)^2} \\ &= 27.92 \end{aligned}$$

   so we should take 28 in each group.

Power analysis for comparing two proportions, independent samples

- The test for
$$H_0 : p_1 - p_2 = 0$$
versus
$$H_a : p_1 - p_2 = \delta$$
for $\delta > 0$ uses the test statistic
$$Z = \frac{\hat{p}_1 - \hat{p}_2}{SE(\hat{p}_1 - \hat{p}_2)}.$$
Assuming equal sample sizes $n_1 = n_2 = n$ in the two groups, and using standard error
$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{p_1(1-p_1) + p_2(1-p_2)}\sqrt{\frac{1}{n}}$$

- the power of the test is
$$Power = 1 - \Phi\left(z_\alpha - \frac{\delta}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}\sqrt{\frac{1}{n}}}\right) \quad (5)$$

- and sample size per group is
$$n = \frac{(p_1(1-p_1) + p_2(1-p_2))(z_\alpha + z_\beta)^2}{\delta^2}. \quad (6)$$

- Usually you don't have values of p_1 and p_2 in advance, in which case you can use the worst choice $p_1 = p_2 = .5$, which gives the largest possible value for $p_1(1-p_1) + p_2(1-p_2) = \frac{1}{2}$, in which case the sample size per group is
$$n = \frac{(z_\alpha + z_\beta)^2}{2\delta^2}. \quad (7)$$
and the power is
$$Power = 1 - \Phi\left(z_\alpha - \frac{\delta}{\sqrt{1/(2n)}}\right). \quad (8)$$

- This choice will give a conservative answer, i.e. the true power will be at least as large as the value calculated.
- As before, we substitute $z_{\alpha/2}$ for z_α for a two-sided alternative.
- Once again, we round the sample size up to ensure desired power.

Example: A researcher wants to determine the difference in two probabilities to within .1 using a one-sided test at level $\alpha = .05$.

1. What is the power if the sample size in each group is 100?

   - Since the sample proportions were not given, consider the worst choice $p_1 = p_2 = .5$ and use formula (5) with $z_\alpha = 1.645$, $\delta = .1$ and $n = 100$ gives

$$\begin{aligned} Power &= 1 - \Phi(1.645 - \frac{.1}{\sqrt{.5(1-.5) + .5(1-.5)}\sqrt{1/100}}) \\ &= 1 - \Phi(1.645 - 1.414) \\ &= 1 - \Phi(.231) = .4088. \end{aligned}$$

2. What sample size is required to achieve a power of .75?

   - Again using the worst choice $p_1 = p_2 = .5$ and use equation (6) with $z_\alpha = 1.645$, $z_\beta = .6745$ and $\delta = .1$ gives

$$\begin{aligned} n &= \frac{(.5(1-.5) + .5(1-.5))(1.645 + .6745)^2}{(.1)^2} \\ &= \frac{2.69004}{.01} \\ &= 269.004 \end{aligned}$$

   - So $n = 270$ are required in each group.
   - Typically, large samples are required to estimate probabilities, or differences in probabilities.
   - Note: Both questions 1 and 2 could have used formula (8) and formula (7), respectively, and would give the same answers.

3. Suppose that the researcher knew that $p_1 = .3$, $p_2 = .4$. So $\delta = |p_1 - p_2|$ is still $.1$. To have power =.75 for a one sided test at level $\alpha = .05$, use equation (6) with $z_\alpha = 1.645$ and $z_\beta = .6745$. The required sample size per group is:

$$n = \frac{(.3(.7) + .4(.6))(1.645 + .6745)^2}{.1^2} = 242.07$$

   so round up to 243 subjects per group.

   - So sample size is smaller when p_1 and p_2 are distant from 0.5.

In summary, statistical power nearly always depends on the three factors:

- the statistical significance criterion used in the test, i.e., **the α level**
- the magnitude of the effect of interest in the population, i.e., **the effect size**
- **the sample size** used to detect the effect