community project

encouraging academics to share statistics support resources

stcp-rothwell-samplesize

The following resources are associated:
Medical Statistics, Introduction to Hypothesis testing, Types of Trials

Power-Based Sample Size Calculations

Sample size calculations are needed when designing a study or a trial. They vary depending on the type of outcome you are investigating. There are two types which will be discussed here, continuous and binary outcomes. Continuous outcomes are numerical scale variables, such as blood pressure measurements after a new treatment or weight lost on a diet. Binary variables are categorical, such as yes/no, dead/alive, responded/not responded. Sample size calculations also depend on the type of trial being designed, a parallel group trial or a cross-over trial (see Types of Trials sheet).

Sample size calculations for t-tests

In order to calculate a sample size for an independent t-test, several values are needed.
- d is the expected difference between the means of the groups and represents the clinically meaningful difference you are trying to find.
- σ^2 is the estimated population variance (assuming both groups have the same). This needs to be estimated from either a pilot study or previous research.
- The required **power** and the **significance level** of the study.
  Here is a table which explains where these values come from.

	Null hypothesis is actually	
Decide to	**False**	**True**
Reject Null Hypothesis	Correct decision Power of test $(1 - \beta)$	Type I Error (α) = significance level (0.05)
Not Reject Null Hypothesis	Type II Error (β)	Correct Decision

The power is usually set at 80% ($1 - \beta = 0.8$) or 90% ($1 - \beta = 0.9$). The aim is to have the power as high as possible so the probability of rejecting the null hypothesis when the null is actually false (detecting a significant result) is high. The significance level is also known as the Type I error, which is conventionally set at 5%, $\alpha = 0.05$. This is set to be as low as possible and this is also the value which we will compare the test results with (testing if p<0.05).

Medical Statistics - Power based sample size calculations

Once the levels of power and significance are decided, Z scores, $Z_{1-\beta}$ and $Z_{1-\frac{\alpha}{2}}$ are calculated
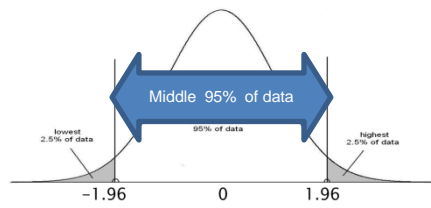
using the standard normal distribution. For example, if the significance level is 0.05, we wish to know the Z scores within which 95% of values lie.

$$Z_{1-\frac{\alpha}{2}} = Z_{1-\frac{0.05}{2}} = Z_{0.975} = 1.96$$

The most commonly used Z scores for power are

$$For\ 80\%\ power, Z_{1-\beta} = Z_{1-0.8} = Z_{0.2} = 0.84$$
$$For\ 90\%\ power, Z_{1-\beta} = Z_{1-0.9} = Z_{0.1} = 1.28$$

Parallel Group Trials

The null and alternate hypotheses are as follows:

$H_0: \mu_A = \mu_B$ (The two treatments are the same)

$H_1: \mu_A \neq \mu_B$ (The two treatments are different)

The sample size for each arm of the trial is given by:

$$n_A = \frac{2\left(Z_{1-\beta}+Z_{1-\frac{\alpha}{2}}\right)^2 \sigma^2}{d^2} \tag{1}$$

where d is the expected difference between the two treatments, $Z_{1-\beta}$ and $Z_{1-\frac{\alpha}{2}}$ are the Normal values for power and significance, and σ^2 is the estimated population variance.

Quick Results

For 90% power, equal allocation, 2-sided 5% significance level, the sample size for each arm can be approximated by:

$$n_A = \frac{21\sigma^2}{d^2} \tag{2}$$

Example

Let us consider a trial of blood pressure medication which is expected to reduce systolic blood pressure by 10mmHg. The allocation the investigator has chosen is equal between the treatment and control groups, and the expected standard deviation in the population which the trial is aimed at is 50mmHg.

Using equation (1), we get

$$n_A = \frac{(r+1)\left(Z_{1-\beta}+Z_{1-\frac{\alpha}{2}}\right)^2 \sigma^2}{rd^2} = \frac{2\times(1.282+1.96)^2\times50^2}{10^2} = 525.53 \approx 526\ per\ group$$

Note 1: Any sample sizes which result in a decimal shall be rounded up to account for the extra participant needed.

The quick result (2) yields:

$$n_A = \frac{21\sigma^2}{d^2} = \frac{21\times50^2}{10^2} = 525\ per\ group$$

Cross over Trials

In order for us to be able to estimate a sample size for a cross-over trial, we need to first estimate the within-subject standard deviation σ_w. This can be estimated from previous studies or knowledge.

$$n = \frac{2(Z_{1-\beta} + Z_{1-\alpha/2})^2 \sigma_w^2}{d^2} \tag{3}$$

where n is the total sample size, σ_w is the within-subject standard deviation, d is the expected difference between the two treatments, $Z_{1-\beta}$ and $Z_{1-\frac{\alpha}{2}}$ are the Normal values for power and significance.

Quick Results
For 90% power, equal allocation, 2-sided 5% significance level, the total sample size can be approximated by:

$$n = \frac{21\sigma_w^2}{d^2} \tag{4}$$

Example

Let us consider two different inhalers for asthma sufferers, one being the standard and the other being a new inhaler. The two treatments will be trialled in a cross over design, with all patients receiving both treatments. The expected difference (d_s) in the number of asthma incidences is 2 (units) with an expected within-subject population standard deviation (σ_w) of 4 (units). The patients are randomised to either have the standard inhaler first, or the new inhaler first.

Using the formula (4)

$$n = \frac{2(Z_{1-\beta} + Z_{1-\alpha})^2 \sigma_w^2}{d^2} = \frac{2 \times (1.282 + 1.96)^2 \times 4^2}{2^2} = 84.08 \approx 85 \; in \; total$$

Binary (Categorical) Outcome

Parallel Group Trials
The parallel group superiority trial is investigating whether two treatments or interventions are different in terms of the proportion of patients with a particular outcome.
Let p_A and p_B be the proportion of adverse events in groups A and B respectively.
The two hypotheses of interest would be

$H_0: p_A = p_B$ (There is no difference between the two treatment effects).
$H_1: p_A \neq p_B$ (There is a difference between the two treatment effects).
Let us illustrate the results of a trial for two different sun protection creams. The outcome of interest is whether the participants burned or not, given they followed trial protocol.

	Outcome		
Treatment	**Burned**	**Not burned**	**Total**
Cream A	p_A	$1 - p_A$	n_A
Cream B	p_B	$1 - p_B$	n_B
Overall Response	$\bar{p}$	$1 - \bar{p}$	$n = n_A + n_B$

Let p_A be the proportion of responses in group A and p_B be the response in group B, with n_A and n_B be the total number of patients in groups A and B respectively such that $n = n_A + n_B$ is the total number of patients in the study and $\bar{p} = \frac{n_A p_A + n_A p_B}{n_A + n_B}$ is the average response across the treatments.
The most common sample size formulae for this type of data is

$$n_A = \frac{\left(Z_{1-\beta}+Z_{1-\frac{\alpha}{2}}\right)^2 (p_A(1-p_A)+p_B(1-p_B))}{(p_A-p_B)^2} \qquad (5)$$

Where p_A and p_B are the proportion of responses in groups A and B respectively, and $Z_{1-\beta}$ and $Z_{1-\frac{\alpha}{2}}$ are the Normal values for power and significance as in the continuous case.

Quick Results

Here are two conservative quick estimates for the sample size in each group. For 90% power, equal allocation, 2-sided 5% significance level, the sample size for each arm can be approximated by:

$$n_A = \frac{5.25}{(p_A-p_B)^2} \qquad (6)$$

Whilst for an 80% power and 5% significance level, the quick sample size calculation is

$$n_A = \frac{4}{(p_A-p_B)^2} \qquad (7)$$

Example

An investigator wishes to compare two treatments for nausea, one being placebo and the other being a new experimental drug. The absolute risk of nausea on placebo is predicted to be 50% and it is thought that the new treatment would be worth using if it reduced the absolute risk of nausea to 30%, meaning that the treatment effect would have an absolute risk reduction of 20%. The trial will have 90% power and a two-sided significance level of 5%.

Using equation (5), we get

$$n_A = \frac{\left(Z_{1-\beta} + Z_{1-\frac{\alpha}{2}}\right)^2 \left(\pi_A(1-\pi_A) + \pi_B(1-\pi_B)\right)}{(\pi_A - \pi_B)^2}$$

$$= \frac{(1.282 + 1.96)^2 \times (0.30 \times 0.70 + 0.50 \times 0.50)}{(0.30 - 0.50)^2} = 120.87 \approx 121 \; per \; group$$

Note 1: Any sample sizes which result in a decimal shall be rounded up to account for the extra participant needed.

Using the quick formula (6) for 90% power and 5% type I error rate we get

$$n_A = \frac{5.25}{(\pi_A - \pi_B)^2} = \frac{5.25}{(0.3 - 0.5)^2} = 131.25 \approx 132 \; per \; group$$

which is clearly more conservative than the more accurate formulae from Equation (6).

Cross over Trials

Cross over trials with a binary outcome require a more complicated sample size calculation which is not discussed here. There are a number of medical statistics books available for further information.

Note: There is a sample size app available for iPhone/iPad and Android devices, called **SampSize**.

https://play.google.com/store/apps/details?id=uk.co.epigenesys.sampsize&hl=en_GB

https://itunes.apple.com/gb/app/epi-infotm-companion-for-iphone/id762428884?mt=8

There is also a helpful online calculator which is more flexible:

http://sampsize.sourceforge.net/iface/