Sample Size Estimation
Sample size planning

„How large a sample should I take?“
Clinical trials should have sufficient statistical power to detect difference between groups considered to be of **clinical interest**.

Therefore calculation of sample size with provision for adequate levels of significance and power is an essential part of planning.
Sample size

Depending on:

1) variability in the target population
2) desired precision in the estimate
3) desired confidence in the estimate
4) Feasibility and costs
The calculation of sample size is only approximate and depends on the accuracy and relevance of the estimates we use for it.
Sample size calculation

• Parameters used in sample size calculation are estimates with uncertainty.

• They are based on very small pilot studies or.

• They are based on previous studies (literature):
  • Population may be different.
  • Publication bias - overly optimistic.
  • Different inclusion and exclusion criteria.

• Mathematical models approximation.
\( \alpha \) and Confidence Level

- \( \alpha \): The significance level of a test: the probability of rejecting the null hypothesis when it is true (or the probability of making a Type I error - put an useless medicine into the market!).

- Confidence level: The probability that an estimate of a population parameter is within certain specified limits of the true value; commonly denoted by “1-\( \alpha \)”. 
Power and $\beta$

- Power: The probability of correctly rejecting the null hypothesis when it is false; commonly denoted by “1-\(\beta\).

- $\beta$: The probability of failing to reject the null hypothesis when it is false (or the probability of making a Type II error - keep a good medicine away from patients!).
A measure of how close an estimate is to the true value of a population parameter. It may be expressed in absolute terms or relative to the estimate.
Estimating Population Mean

Sample size required for estimating population mean:

- The objective in interval estimation is to obtain narrow intervals with high reliability
- The width of the interval is determined by the magnitude of the quantity
We may wish to estimate the mean FEV1 in a population of young men.

- Literature: $s = 0.67$ litre

- We set the standard error (SE = 0.1 litre) or confidence limits and choose the sample size to achieve this.

$$SE = 0.67 / \sqrt{n}, \quad n = 0.67^2 / SE^2 = 0.67^2 / 0.1^2 = 45$$

<table>
<thead>
<tr>
<th>n</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE</td>
<td>0.212</td>
<td>0.150</td>
<td>0.095</td>
<td>0.067</td>
<td>0.047</td>
<td>0.030</td>
</tr>
<tr>
<td>95% CI</td>
<td>+/- 0.42</td>
<td>+/- 0.29</td>
<td>+/- 0.19</td>
<td>+/- 0.13</td>
<td>+/- 0.09</td>
<td>+/- 0.06</td>
</tr>
</tbody>
</table>
Estimating a population proportion

- We want to estimate the true immunization coverage in a community of school children.

- Previous studies tell us that immunization coverage should be somewhere around 80%.

- Precision (absolute): we would like the result to be within 4% of the true value.

- Confidence level: conventional 95% = 1 - \(\alpha\); therefore \(\alpha = 0.05\) and \(z_{(1-\alpha/2)} = 1.96\) : value of the standard normal distribution corresponding to a significance level of \(\alpha\) (1.96 for a 2-sided test at the 0.05 level).
Example 2

• The standard error depends on the very quantity which we wish to estimate

\[ SE = \sqrt{\frac{p(1-p)}{n}} \]

• \( p = \) guess for the expected proportion in the population = 0.80

• \( d = \) absolute precision = 0.04

• \( z_{(1-\alpha/2)} = 1.96 \) : value of the standard distribution corresponding to a significance level of \( \alpha \) (1.96 for a 2-sided test at the 0.05 level)
Example 2 - Sample Size

\[ n = \frac{z^2 p(1 - p)}{d^2} \]

\[ = \frac{(1.96)^2 (0.8)(0.2)}{0.04^2} \]

\[ = 384 \]
Sample size for significance tests

- How small a **treatment difference** is important to detect and with what degree of **certainty**? ($\delta$, $\alpha$ and $\beta$)

- How many patients will drop out?
What is $\delta$

- $\delta$ is the minimum difference between groups that is judged to be clinically important
  - Minimal effect which has clinical relevance in the management of patients
  - The anticipated effect of the new treatment (larger)
The Choice of $\alpha$ and $\beta$

- common approach:

$$\alpha = 0.05$$

$$\beta = 0.05 - 0.2$$
Comparison of two means

- $H_0$: $\delta = \mu_C - \mu_I = 0$
- $H_A$: $\delta = \mu_C - \mu_I \neq 0$

- If the variance is known

$$z = \frac{(\bar{x}_c - \bar{x}_I)}{\sigma \sqrt{\frac{1}{n_C} + \frac{1}{n_I}}} = \frac{\delta}{SE(\delta)}$$

- If $|z| > z_{\alpha}$

$H_0$ will be rejected at the $\alpha$ level of significance.
*Sample size*

- A total sample $2n$ would be needed to detect a true difference $\delta$ between $\mu_I$ and $\mu_C$ with power $(1-\beta)$ and significant level $\alpha$ by formula:

$$2n = \frac{4\left(z_\alpha + z_{1-\beta}\right)^2 \sigma^2}{\delta^2}$$
An investigator wish to estimate the sample size necessary to detect a 10 mg/dl difference in cholesterol level in a diet intervention group compared to the control group.

The variance from other data is estimated to be 50 mg/dl.

For a two sided 5% significance level, \( z_\alpha = 1.96 \), and for 90% power, \( z_{1-\beta} = 1.282 \).

\[
2n = 4(1.96 + 1.282)^2(50)^2/10^2 = 1050
\]
Example 4

- An investigator interested in the mean levels of change might want to test whether diet intervention lowers serum cholesterol from baseline levels when compared with a control.

- $H_0: \Delta_c - \Delta_i = 0$

- $H_A: \Delta_c - \Delta_i \neq 0$

- $\sigma = 20\text{mg/dl}, \delta = 10\text{mg/dl}$

$$2n = 4(1.96 + 1.282)^2(20)^2/10^2 = 170$$
• A sample size of 85 in each group will have 90% power to detect a difference in means of 10mg/dl assuming that the common standard deviation is 20mg/dl using a two group t-test with a 0.05 two-sided significant level.
# Comparison of two means

<table>
<thead>
<tr>
<th>Problem</th>
<th>$\mu_1$</th>
<th>$\mu_2$</th>
<th>s</th>
<th>$\alpha$</th>
<th>Power 1-$\beta$</th>
<th>n per group</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4.5</td>
<td>4.25</td>
<td>0.5</td>
<td>0.05</td>
<td>0.9</td>
<td>86</td>
</tr>
<tr>
<td>B</td>
<td>4.5</td>
<td>4.25</td>
<td>1</td>
<td>0.05</td>
<td>0.9</td>
<td>338</td>
</tr>
<tr>
<td>C</td>
<td>4.5</td>
<td>4.35</td>
<td>0.5</td>
<td>0.05</td>
<td>0.9</td>
<td>235</td>
</tr>
<tr>
<td>D</td>
<td>4.5</td>
<td>4.25</td>
<td>0.5</td>
<td>0.05</td>
<td>0.8</td>
<td>63</td>
</tr>
</tbody>
</table>
Values of $f(\alpha, P)$

<table>
<thead>
<tr>
<th>$\alpha$ (Type I error)</th>
<th>0.1</th>
<th>0.05</th>
<th>0.02</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.8</td>
<td>13.0</td>
<td>15.8</td>
<td>17.8</td>
</tr>
<tr>
<td></td>
<td>8.6</td>
<td>10.5</td>
<td>13.0</td>
<td>14.9</td>
</tr>
<tr>
<td></td>
<td>6.2</td>
<td>7.9</td>
<td>10.0</td>
<td>11.7</td>
</tr>
<tr>
<td></td>
<td>2.7</td>
<td>3.8</td>
<td>5.4</td>
<td>6.6</td>
</tr>
</tbody>
</table>

$\left(z_\alpha + z_{1-\beta}\right)^2 = f(\alpha, P)$
Comparing two proportions

- The standard error of the difference between the sample proportions is:

\[
SE(\delta) = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}
\]

\[
(p_1 - p_2)^2 = f(\alpha, P) \left( \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} \right)
\]
Comparing two proportions

- When the sample sizes are equal, i.e. $n_1 = n_2 = n$:

\[
n = \left( \frac{f(\alpha, P)(p_1(1 - p_1)p_2(1 - p_2))}{(p_1 - p_2)^2} \right)
\]
### Comparison of two proportions

<table>
<thead>
<tr>
<th>Problem</th>
<th>$p_1$</th>
<th>$p_2$</th>
<th>$\alpha$</th>
<th>Power 1-$\beta$</th>
<th>$n$ per group</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>20 %</td>
<td>10 %</td>
<td>0.05</td>
<td>0.8</td>
<td>199</td>
</tr>
<tr>
<td>B</td>
<td>20 %</td>
<td>10 %</td>
<td>0.05</td>
<td>0.95</td>
<td>329</td>
</tr>
<tr>
<td>C</td>
<td>20 %</td>
<td>10 %</td>
<td>0.05*</td>
<td>0.8</td>
<td>157</td>
</tr>
<tr>
<td>D</td>
<td>55 %</td>
<td>45 %</td>
<td>0.05</td>
<td>0.8</td>
<td>392</td>
</tr>
</tbody>
</table>

* One-sided test
Power

- Power Depends on 4 Elements:
  - The real difference between two treatments, \( \delta \)
    - Big \( \delta \) \( \Rightarrow \) big power
  - The variation among individuals, \( \sigma \)
    - Small \( \sigma \) \( \Rightarrow \) big power
  - The sample size, \( n \)
    - Large \( n \) \( \Rightarrow \) big power
  - Type I error, \( \alpha \)
    - Large \( \alpha \) \( \Rightarrow \) big power
Books and Software

- Statistical Tables for the Design of Clinical Trials
  David Machin and Michael Campbell, Blackwell Science Ltd 1987

- PASS 2000

- nQuery Advisor 7.0

- Internet sites: i.e. http://home.clara.net/sisa/