SAMPLE SIZE CALCULATION

Dr. Abubakir Majeed
Assist. Prof. of Community Medicine

Concept of sample size

Statistical analyses:

- making inferences about a population while taking account of sampling error

E.g., the proportion of men who currently smoke is 37.9% (32.6% - 43.2%)

- Because of sampling error, the true proportion is very unlikely to be exactly 37.9%

- However it is likely to be contained in the 95% confidence interval of 32.6% to 43.2%

Concept of sample size

Making interpretation of results is difficult or even useless;

- Where confidence intervals are very wide
- Where seemingly large differences between groups are not statistically significant

Confidence intervals and significance tests;

- An important influence is the sample size, n (recall the formulae)

Concept of sample size

- At analysis stage, when CI and statistical tests carried out, it is too late to change the sample size
- When we plan an investigation, we must decide how many people or other units need to be included in order to answer the study objectives
- If the number is too small, we may fail to detect important effects or may estimate effects too imprecisely and will be unable to draw meaningful conclusions
- If the study is too large, we waste resources and may compromise reliability by over-stretching the resources available

Sample size calculation

- The focus is on;
  - A single estimate of a desired outcome for the whole of a sample (proportion or mean)
  - Two groups are to be compared (difference between proportions or means)

Important points to note

- Sample size determinations give only rough estimates of the numbers needed
  - To distinguishing between 200 and 300 subjects
  - But not between 200 and 217
- Large numbers cannot compensate for poor sampling strategy or bias from any other source
- Sample size formulae and computer programs, assume the sampling strategy is simple random sampling
  - Adjustments must be made to the sizes when a different sampling strategy is used
  - A different strategy increases the numbers required

Approaches to sample size calculation

Two approaches:

- Based on estimation of a proportion, mean, difference, relative risk etc with a **certain degree of precision**

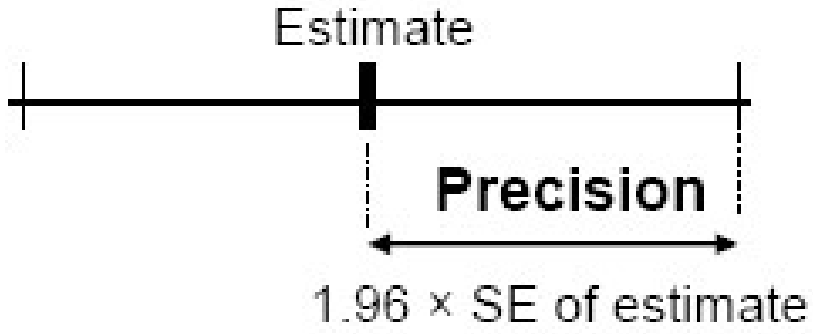
  E.g., to estimate the proportion of children aged 12-23 months who are vaccinated to within 10%. We can consider this as **"how wide will the confidence interval be?“**

- Based on **testing a hypothesis**

  E.g. to compare the duration of exclusive breastfeeding between an intervention and a control group. We can consider this as **"what are the chances of making a wrong conclusion from the significance test?"**

I . Estimating an effect with a certain degree of precision

A 95% confidence interval:



Example

Precision for a proportion = $1.96 \times \sqrt{\frac{p(1-p)}{n}}$

rearranging the formula: n = $3.84 \times \frac{p(1-p)}{precision^2}$

Estimating with a certain degree of precision

- We have to guess the value of p
  We must do this using common sense, previous studies, etc

- There is no set rule for what level of precision you should select for a study - this will depend on the purpose of the study, resources available, etc

  To have an idea, the following table shows the sample sizes required for a study which aims to measure a proportion (e.g. smoking prevalence) for varying levels of precision

Precision	Sample size
20%	24
10%	96
5%	384
1%	9604

Notation

e = precision

σ = standard deviation

p = proportion or percentage

n = sample size

(The width of a confidence interval = 2e)

One Sample

1. Proportion

$$n > \frac{3.84p(1-p)}{e^2}$$

Percentage

$$n > \frac{3.84p(100-p)}{e^2}$$

One Sample – Proportion
Example

- A researcher wishes to assess the prevalence of obesity in a community and wants the estimate to be within 2% of the true value
- Previous studies have shown that prevalence of obesity is 10%
- Solution: e = 2% and p = 10%

$$n = \frac{3.84 \times p(1-p)}{(\text{precision})2}$$

$$n = \frac{3.84 \times 0.1(1-0.1)}{(0.02)^2} = 864$$

- Minimum sample size required is 864 patients

One Sample – Proportion

Absolute precision

The prevalence of obesity in a population is between 20% and 30%. We want the confidence interval for the sample proportion to be 5% above or below

The precision has been formulated as the value ±5% or ±2.5%, i.e. in absolute terms

Assumed prevalence (%)	Target precision required (±)	Lower confidence limit (%)	Sample size
30%	±5%	25%	323
	±2.5%	27.5%	1291
20%	±5%	15%	246
	±2.5%	17.5%	984

Relative precision

Assumed prevalence (%)	Relative Precision	Actial precision (±) to be used for sample size	Lower confidence limit (%)	Sample size
30%	20%	±6%	24%	224
	10%	±3%	27%	896
20%	20%	±4%	16%	384
	10%	±2%	18%	1536

One Sample

- 2. Mean

$$n > \frac{3.84\sigma^2}{e^2}$$

One Sample - Mean

Example

- A researcher wishes to assess the mean level of serum HDL-c (mmol/L) in non-insulin dependent diabetics and wants the estimate to be within **0.25**mmol/L of the true value

- Previous studies have shown that serum HDL-c measurements have a standard deviation of **0.3**mmol/L.

- Solution: e = 0.25 and σ = 0.3

$$n > \frac{3.84\sigma^2}{e^2} \qquad n > \frac{3.84 * 0.3^2}{0.25^2} = 5.53$$

- Minimum sample size required is 6 patients

Two samples

3. Difference between two proportions

$$n > \frac{3.84 \times [p_1(1-p_1) + p_2(1-p_2)]}{e^2}$$

- required in each sample

Difference between two proportions

Example

- Suppose we want to estimate the percentage change in mortality rate due to pneumonia when patients are treated with a new drug and we want the estimate to be accurate to within **2%**. Mortality in the first two weeks after pneumonia is estimated to be **15%** and it is estimated that the new drug could reduce this to **10%**

- Solution:
- p1=15%, p2=10% and e=2%

$$n > \frac{3.84 \times [15(100-15)+10(100-10)]}{2^2} = 2{,}088$$

- We would need a total sample of 4,176 (2,088 in each group).

Two samples

4. Difference between two means

$$n > 2 \times \frac{3.84\sigma^2}{e^2}$$

- required in each sample

Difference between two means

Example

- A researcher wishes to estimate the difference in mean level of serum HDL-c (mmol/L) between non-insulin dependent diabetics and normal controls. He wants the estimate of the difference to be within **0.25** mmol/L of the true mean difference.
- SD from previous studies is **0.3**mmol/L
- Solution: e = 0.25 and σ = 0.3

$$n > 2 \times \frac{3.84\sigma^2}{e^2}$$

$$n > 2 \times \frac{3.84 * 0.3^2}{0.25^2} = 11.06$$

- Minimum sample size is 12 patients in each group.

II. Sample size calculation based on testing a hypothesis - comparing two groups

- To compare a proportion between two groups - an intervention and a control group, we can test whether the two proportions are significantly different

- In carrying out a significance test to test a hypothesis, there are basically two types of error we can make

For testing the null hypothesis H0: D=0

		TRUE SITUATION
TEST RESULT	D = 0	D <> 0
Don't reject H0 (not significant)	Probability= 1-α	Type II error Prob. = β
Reject H0 (signif., P<α)	Type I error Prob. = α	Probability = 1-β = Power

Testing a hypothesis

- Ideally, we would like to minimise both types of error
- The power of the study (equal to 1-Type II error) - the larger the power of the study, the smaller the Type II error
  - **The power is the probability of getting a statistically significant result with the selected sample if a true difference exists**
- The objective is to choose a sample size such that if there is a clinically important difference between the groups, we have a good chance of finding a statistically significant difference between them
- If, at the analysis stage, our particular sample selected yields results that are statistically significant, then most people feel satisfied with the interpretation to be made
- But, if the results are not statistically significant then this could have arisen because either

  (a) there is no true difference between the groups being compared or

  (b) there is a true difference but our particular sample did not show a difference because the study had low power to do so
- If a study have a high level of power and the results are not statistically significant then we can be more certain that this is because no true difference exists

To choose an adequate sample size, we must specify:

- The Type I error or significance level we want our study to have; *commonly 5%, also called 95% confidence*

- The Type II error or the POWER we want our study to have *commonly a power of 80-90% (Type II error of 10-20%)*

- The "base level" in one of the group, e.g. the % of controls in a case-control study which are exposed to the risk factor or the % of people who have disease among those who are unexposed to the risk factor in a cohort study *typically estimated from a pilot study or existing data*

- The minimum level of effect which we wish our study to detect (e.g. the minimum odds ratio in a case-control study, the minimum difference in the % diseased in a cohort study) *must decide what is an effect of clinical or public health importance*

Notation

- **σ** = standard deviation
- **p** = proportion or percentage
- **n** = sample size
- **d** = smallest difference of clinical or scientific importance
- **F** = value that depends on the *significance level* and the *power* of your study

‘F’ values

	Power			
	80%	90%	95%	99%
Significance level				
0.1	6.18	8.56	10.82	15.77
0.05	7.85	10.51	12.99	18.37
0.025	9.51	12.41	15.10	20.86
0.010	11.68	14.88	17.81	24.03

These values, we shall call ‘F’, are equal to: $\left(Z_{\frac{\alpha}{2}} + Z_{\beta}\right)^2$

where:
α = probability(type I error)
β = probability(type II error)

Example:
If α = 5% and β = 10%, then
$F=(1.96+1.2816)^2 = 10.51$

One sample

Mean

$$n > \frac{F\sigma^2}{d^2}$$

Example

- Suppose some outcome is measured twice on each patient, once after treatment A and once after treatment B, and that the order of the treatments is randomized. A true difference in outcome of 5 units is considered clinically important. The standard deviation of repeated measurements from previous studies is known to be 11.31. We want to set our significance level at 1% and we want to ensure that the difference would not be missed in 99% of trials.

- Solution:
  d=5, F=24.03 and σ=11.31

$$n > \frac{24.03 * 11.31^2}{5^2} = 122.95$$

- We would need to recruit 123 patients.

One Sample

Proportion $$n > \frac{Fp(1-p)}{d^2}$$

Percentage $$n > \frac{Fp(100-p)}{d^2}$$

Example

- A health authority wants to know if the prevalence of children under 2 years with wheeze in their district is different from the national average of 30%. A random sample is to be selected and they wish to identify a difference of 10% or higher. They think that a power of 95% and a significance level of 5% will be appropriate.

$$n > \frac{Fp(1-p)}{d^2}$$

	Power			
Sig. lev	80%	90%	95%	99%
0.1	6.18	8.56	10.82	15.77
0.05	7.85	10.51	12.99	18.37
0.025	9.51	12.41	15.10	20.86
0.01	11.68	14.88	17.81	24.03

Two samples

Difference between two means

$$n > \frac{2F\sigma^2}{d^2}$$

- required in each sample

Example

- A study is to be carried out in a rural area of East Africa to ascertain whether giving food supplementation during pregnancy increases birth weight. Women attending the antenatal clinic are to be randomly assigned to either receive or not to receive food supplementation. The difference between mean birth weights regarded as clinically important is 0.25Kg. From previous studies the standard deviation of birth weight is found to be 0.4Kg. The significance level is to be 1% and the power is to be 95%.

- Solution:
- d = 0.25, F = 17.81 and σ = 0.4

$$n > \frac{2*17.81*0.4^2}{0.25^2} = 91.19$$

- The study will require 92 women in each group.

Two samples

Difference between two proportions

- **or between two percentages**

$$n > \frac{F[p_1(1-p_1)+p_2(1-p_2)]}{d^2}$$

- required in each sample

$$n > \frac{F[p_1(100-p_1)+p_2(100-p_2)]}{d^2}$$

Example

- The MRC CRASH Trial was a large placebo controlled trial among adults with head injury and impaired consciousness, of the effects of a 48-hour infusion of corticosteroids on death and neurological disability. In a previous trial with similar inclusion criteria the overall risk of death among controls was **15%.** A **2%** absolute risk reduction was regarded as clinically important. The significance level was **1%** and the power was **95%.**

- Solution:
- d=2%, p1=15%, p2=13%, F=17.81

$$n > \frac{17.81[15(100-15)+13(100-13)]}{2^2} = 10,712.72$$

- The CRASH Trial required 10,713 patients in each group.

Further notes

- Large study sizes are worse than useless if our samples are not representative
- Increase the calculated sample size to allow for non-contact, non-response and other factors that tend to reduce the final sample size
- The usual calculations assume simple random sampling. Adjustments need to be made if other methods such as cluster sampling or stratified sampling are used. These may imply double the sample size, or even more
- Sample sizes should be increased if adjustment from confounding factors is anticipated. Typically, increases of 20-25% should be considered
- The calculations have assumed a very large or infinite population size from which the sample is drawn, so the sample is only a small fraction of the total population. Adjustments need to be made if this fraction becomes sizeable (above 10%; in fact, in Epi-Info one can allow for this in the section on precision and confidence intervals, but not in the section on power)
- Computer packages make it quick and easy to calculate trial sample sizes for different scenarios
- We must strike a balance between what is desirable and what is logistically feasible

Working backwards

- It is a useful possibility to consider
  - Suppose that the sample size is determined by costs, time, availability of patients etc
  - Then we can take this sample size and see what precision our estimate will have or what power our study will have for detecting a significant difference
  - We may discover that the estimate would be so imprecise or the power so low that the study is not worth carrying out in the first place