Significance of mean – clinical application

Standard Deviation

Is a measure of how widely values are dispersed from the mean, or how closely they are flocked around the mean.

Formula:

$$SD = \sqrt{\frac{\Sigma d^2}{n-1}}$$

where d is difference between individual value and the mean

Example

Calculate the SD for the following figures

5, 10 1 7 17 27 99 34 78 81 3 7 17 34 56 43

31 22 56 34 89 67 35 13 48 98 24 47 12 24 5

Mean

Mode

Median

Sd

Se

STANDARD ERROR

- SE = SD/sqrt of “n”
- Chance variation from sample to sample or from sample to population
- In any sampling distribution;
- $\mu \pm 1$ SE limits include 68% of obs
- $\mu \pm 1.96$ SE limits include 95% of obs
- $\mu \pm 2.58$ SE limits include 99% of obs
- This sampling distribution forms the basis of tests of significance

Confidence interval

- The confidence interval estimates the boundaries likely to include (desired target) proportions (often 95%) of future similar measurements made from that statistical population.
- A range likely to include true population values
- Statisticians can calculate a range (interval) in which we can be fairly sure (confident) that the “true value” lies
- The size of a CI is related to the sample size of the study. Larger studies usually have a narrower CI.

CI application

- The average systolic BP before treatment in study A, of a group of 100 hypertensive patients, was 170 mmHg. After treatment with the new drug the mean BP dropped by 20 mmHg.
- If the 95% CI is 15–25, this means we can be 95% confident that the true effect of treatment is to lower the BP by 15–25 mmHg.
- In study B 50 patients were treated with the same drug, also reducing their mean BP by 20 mmHg, but with a wider 95% CI of -5 to +45.
- This CI includes zero (no change).
- This means there is more than a 5% chance that there was no true change in BP, and that the drug was actually ineffective.

Importance of Sample Size calculation

- Scientific reasons
- Ethical reasons
- Economic reasons

Scientific Reasons

- In a trial with **negative** results and a *sufficient sample size*, the result is concrete
- In a trial with **negative** results and *insufficient power (insufficient sample size)*, may mistakenly conclude that the treatment under study made no difference

Ethical Reasons

- An *undersized* study can expose subjects to potentially harmful treatments without the capability to advance knowledge
- An *oversized* study has the potential to expose an unnecessarily large number of subjects to potentially harmful treatments

Economic Reasons

- *Undersized study* is a waste of resources due to its inability to yield useful results
- *Oversized study* may result in statistically significant result with doubtful clinical importance leading to waste of resources

Which variables should be included in sample size calculation?

- The sample size calculation should relate to the study's primary outcome variable.
- If the study has secondary outcome variables which are also considered important (as is often the case), the sample size should also be sufficient for the analyses of these variables too.

Consistency with study aims

- If the aim is to demonstrate that a new drug is superior to an existing one then it is important that the sample size is sufficient to detect a clinically important difference between the two treatments.
  - However, sometimes the aim is to demonstrate that two drugs are equally effective. This type of trial is called an equivalence trial or a 'negative' trial.
  - The sample size required to demonstrate equivalence will be larger than that required to demonstrate a difference.
- The sample size calculation should also be consistent with the study's proposed method of analysis, since both the sample size and the analysis depend on the design of the study.

Some commonly used terms:

- Random error
- Systematic error (bias)
- Precision (Reliability)
- Null hypothesis
- Alternative hypothesis
- Type I error (α)
- Type II error (β)
- Hypothesis Testing
- Power of the study (1-β)

Random error

- It describes the role of chance
- Sources of random error include:
  - sampling variability
  - subject to subject differences
  - measurement errors
- It can be controlled and reduced to acceptably low levels by:
  - Increasing the sample size
  - Repeating the experiment

Precision (Reliability)

- Degree to which a variable has the same value when measured several times
- It is a measure of consistency
- It is a function of :
  - random error

    (the greater the error, the less precise the measurement)
  - sample size
  - confidence interval required &

A larger sample size would give precise estimates

Systematic error (bias)

- It describes deviations that are not a consequence of chance alone
- Several factors including **patient selection criteria** might contribute to it.
- These factors may not be amenable to measurement
- Removed or reduced by good design and conduct of the experiment
- A strong bias can yield an estimate very far from the true value

Accuracy (Validity)

- It indicates the degree to which the variable actually represents what it is supposed to represent
- It is a function of systematic error
- The greater the error the less accurate the variable

Null hypothesis

- Null hypothesis is a hypothesis which states that there is no difference among groups or that there is no association between the predictor and the outcome variables

- This hypothesis needs to be tested

Type I error

- Rejecting a null hypothesis actually true in the population
- probability of erroneously finding a disease exposure association, when none exists in reality.

Type II error

- Fails to reject a null hypothesis that is actually false in the population
- probability of erroneously not finding disease exposure association, when it exists in reality.

Hypothesis Testing

Fact	Inference: No Difference	Inference: Difference
No Difference		Type I
Difference	Type II	

Alternative hypothesis

- It assumes that there is a difference among the groups or there exists an association between the predictor and outcome variable

- There are two types of alternative hypothesis:
  - one-tailed (one-sided) hypothesis &
  - two-tailed (two-sided) hypothesis

- One-tailed hypothesis specifies the difference (or effect or association) in one direction only.

- Two-tailed hypothesis specifies the difference (or effect or association) in either direction.

Hypothesis testing:

One-sided or two-sided?

- Ho: $\mu_1 = \mu_2$ (equal means)
- Ha:

  $\mu_1 \neq \mu_2$ (two-sided test), or

  $\mu_1 > \mu_2$ (one-sided test), or

  $\mu_1 < \mu_2$ (one-sided test)
- Examples:
  - Two -sided: new drug vs. old drug. May help or hurt. Safer to test both possibilities.
  - One-sided: Standard Tx +/- Yoga. Very unlikely to hurt. Possible benefit. More power with one-sided test.

Differences Small and Big

- In any parameter if the difference between values is high, it is easy to detect, the test to detect such differences need not be very sophisticated.
- To know the difference between the weight of rat and a man we don't need much equipment (we really don't need any equipment)

Small differences

- To detect the difference between the weights of two rats of the same size, we require a balance.
- If the difference is in grams a simple balance will be adequate
- If the difference is in milligrams we need a sensitive balance.

Difference and Power

- We need sophisticated tools to detect very small differences
- For large differences, a less sophisticated tool is good enough
- In statistics, the sophistication is termed as power
- A trial with high power is required to detect small differences between treatments.

Power of a test

- Power is the probability that we reject the null hypothesis given a particular alternative hypothesis is true.(difference detected)
- Power = 1 - β = 1- Pr(Type 2 error)

- β = Pr (missing the difference)
- power = Pr (detecting the difference)

Importance of Sample Size calculation

- Scientific reasons
- Ethical reasons
- Economic reasons

Scientific Reasons

- In a trial with **negative** results and a *sufficient sample size*, the result is concrete
- In a trial with **negative** results and *insufficient power (insufficient sample size)*, may mistakenly conclude that the treatment under study made no difference

Ethical Reasons

- An *undersized* study can expose subjects to potentially harmful treatments without the capability to advance knowledge
- An *oversized* study has the potential to expose an unnecessarily large number of subjects to potentially harmful treatments

Economic Reasons

- *Undersized study* is a waste of resources due to its inability to yield useful results
- *Oversized study* may result in statistically significant result with doubtful clinical importance leading to waste of resources

Which variables should be included in sample size calculation?

- The sample size calculation should relate to the study's primary outcome variable.
- If the study has secondary outcome variables which are also considered important (as is often the case), the sample size should also be sufficient for the analyses of these variables too.

Consistency with study aims

- If the aim is to demonstrate that a new drug is superior to an existing one then it is important that the sample size is sufficient to detect a clinically important difference between the two treatments.
  - However, sometimes the aim is to demonstrate that two drugs are equally effective. This type of trial is called an equivalence trial or a 'negative' trial.
  - The sample size required to demonstrate equivalence will be larger than that required to demonstrate a difference.
- The sample size calculation should also be consistent with the study's proposed method of analysis, since both the sample size and the analysis depend on the design of the study.

Asking the sample size question

- What sample size do I need to have so that I have adequate power to detect a particular difference?
- I only have N subjects available. What power will I have to detect a particular difference with that sample size?

Preparing to calculate sample size or power

- What is the study question?
- What is the principal outcome measure?
- What statistical test will be used?
- Will the test be one- or two-tailed?
- What α level? (usually 0.05)
- For some designs, other parameters must be specified.
  - e.g., For comparing means, what is the standard deviation?

Free Software Sources for Sample Size

- **UnifyPow** (SAS-based): download from www.bio.ri.ccf.org/power.html
- **STPLAN**: free via anonymous ftp from http://odin.mdacc.tmc.edu/anonftp/
- Free web-base software
  - http://www.statistics.com/
  - http://ebook.stat.ucla.edu/calculators/powercalc/

General Rules of Thumb

- Non RCTs generally require a much larger sample to allow adjustment for confounding factors in the analysis
- Equivalence studies need a larger sample size than studies aimed to demonstrate a difference
- For moderate to large effect size (0.5<effect size<0.8), 30 subjects per group
- For comparison between 3 or more groups, to detect a moderate effect size of 0.5 with 80% power, will require 14 subjects/group