Significance of mean – clinical application

Module12 Topic 3

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. (4)
- 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(6)
- The size of a CI is related to the sample size of the study. Larger studies usually have a narrower CI.

Confidence Interval example

- We may be interested in blood pressure (BP) reduction with antihypertensive treatment.
- From a sample of treated patients we can work out the mean change in BP.
- However, this will only be the mean for our particular sample. If we took another group of patients we would not expect to get exactly the same value, because chance can also affect the change in BP.
- The CI gives the range in which the true value (i.e. the mean change in BP if we treated an infinite number of patients) is likely to be.

Confidence Interval 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.

Designing a Study:
~~Sample Size Calculation~~

- The study design (case/control, cohort study, RCT, etc.) is the first decision, but sample size is a close second.
- Sometimes sample size will drive the design.
  - To test a possible association between breast implants and scleroderma, should you:
    1. study women with/without breast implants to observe the proportions with scleroderma, or
    2. study women with/without scleroderma to observe breast implants?
  - Both are rare, but scleroderma is much more rare than breast implants. Design (1) above would require a huge sample size. Design (2) is feasible.

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 (Cardiac Studies)

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.

Allowing for response rates and other losses to the sample

- The sample size calculation should relate to the final, achieved sample.
- Therefore, the initial numbers approached in the study may need to be increased in accordance with the expected response rate, loss to follow up, lack of compliance, and any other predicted reasons for loss of subjects.
- The link between the initial numbers approached and the final achieved sample size should be made explicit.

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

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

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

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 not erroneously finding disease exposure association, when it exists in reality.

Hypothesis Testing

	Inference	
Fact	No Difference	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.
- Power = 1 - β = 1- Pr(Type 2 error)

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

Sample size calculations are ~~approximate~~

- Often based on roughly estimated parameters.
- Usually based on mathematical models that only approximate the truth.
- Often based on a simplification of the study question.
- Changes may occur in the expected treatment effect or in the target population between design and data collection.
- **So, be conservative when estimating sample size**.

Asking the sample size question

- What sample size do I need to 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?

Preparing to calculate

- The hard one: How small a difference is it important to detect (Δ or effect size)?
  - i.e., What difference would you not want to miss?
- What degree of certainty (power) do you want to detect that difference?
  - (usually 0.80 - 0.95)

Sample size formulas

- For every statistical test, a sample size formula can be derived.
- Example: testing for a difference between two means, the formula is

$$\text{N in each group} = \frac{2\,\sigma^2\,(z_{\alpha/2} + z_{\beta})^2}{\Delta^2}$$

where σ = the common standard deviation of each group,

Δ = the smallest difference you want to detect with high power,

α = significance level, and β = 1- Power.

What is Z score/number(8)

- Z_{α}, *Z* is a constant (set by convention according to the accepted α error and whether it is a one-sided or two-sided effect) :

α-error	5%	1%	0.1%
2-sided	1.96	2.5758	3.2905
1-sided	1.65	2.33	

- For Z1-β,Z is a constant set by convention according to power of the study :

Power	80%	85%	90%	95%
Value	0.8416	1.0364	1.2816	1.6449

Two Sample t-test

For the hypothesis: $H_0: \mu_1 = \mu_2$ vs. $H_1: \mu_1 \neq \mu_2$

For a two tailed t-test, the formula is:

$$N = n_1 + n_2 = \frac{4\sigma^2 (z_{1-\alpha/2} + z_{1-\beta})^2}{(d = \mu_1 - \mu_2)^2}$$

Sample Size for Testing Two tailed t-test

$H_0: \mu_1 = \mu_2$ vs. $H_1: \mu_1 \neq \mu_2$

How large a sample would be needed for comparing two approaches to cholesterol lowering using $\alpha = 0.05$, to detect a difference of d = 20 mg/dl or more with

Power = $1 - \beta = 0.90$

The formula is:

$$N = n_1 + n_2 = \frac{4\sigma^2 (z_{1-\alpha/2} + z_{1-\beta})^2}{(d = \mu_1 - \mu_2)^2}$$

Note: Textbooks do not always clearly indicate whether the formula they provide is for one group only or for both groups combined.

When σ = 30 mg/dl, β = 0.10, α = 0.05; $z_{1-\alpha/2}$ = 1.96

Power = 1- β ; $z_{1-\beta}$ = 1.282 , d = 20mg/dl

$$N = n_1 + n_2 = \frac{4(30)^2(1.96+1.282)^2}{(20)^2}$$

$$= \frac{4\times 900\times(3.242)^2}{400} = \frac{37,838.03}{400}$$

$$N = 94.6$$

Hence about 50 for each group

Another simple formula(9)

- The formula to calculate the sample size for a mean (or point) estimate is:
- N = (SD/SE)squared

where N = the required sample size,

SD = the standard deviation, and

SE = the standard error of the mean

- The standard deviation could be estimated either by looking at some previous study or by carrying out a pilot study.
- **If you want a 95% confidence interval, then divide the maximum acceptable MRE(margin for random error) by 1.96 to calculate the SE.**
- **If instead you want a 99% confidence interval, then divide the maximum acceptable MRE by 2.56 to calculate the SE.**

Application(7)

		Clinically Significant	
		Yes	No
Statistically significant	Yes	Typically assume the groups, outcomes, or treatments are different	Consider that the sample size may be too large
	No	Consider that the sample size may be too small	Typically assume the groups, outcomes, or treatments are not different

Commercial Software for Sample Size

- **PASS** (phone: 1-801-546-0445):
  http://www.ncss.com/
  e-mail:ncss@ix.netcom.com

- **nQuery Advisor** (phone 1-800-262-1171)
  http://www.statsolusa.com/
  e-mail:info@statsolusa.com

- **Sample Power** by SPSS (phone: 312-329-2400)
  http://www.spss.com/

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/

Pitfalls to avoid (1)

- "The throughput of the clinic is around 50 patients a year, of whom 10% may refuse to take part in the study. Therefore over the 2 years of the study, the sample size will be 90 patients. "
- Although most studies need to balance feasibility with study power, the sample size should not be decided on the number of available patients alone.
- Where the number of available patients is a known limiting factor, sample size calculations should still be provided, to indicate either
  - The power which the study will have to detect the desired difference of clinical importance, or
  - The difference which will be detected when the desired power is applied.

Pitfalls to avoid (2)

- "Sample sizes are not provided because there is no prior information on which to base them."
- Where prior information on standard deviations is unavailable, sample size calculations can be given in very general terms, i.e. by giving the size of difference that may be detected in terms of a number of standard deviations.

Pitfalls to avoid (3)

- "A previous study in this area recruited 150 subjects and found highly significant results (p=0.014), and therefore a similar sample size should be sufficient here."
- Previous studies may have been 'lucky' to find significant results, due to random sampling variation.
- Calculations of sample size specific to the present, proposed study should be provided, including details of
  - power
  - significance level
  - primary outcome variable
  - effect size of clinical importance for this variable
  - standard deviation (if a continuous variable)
  - sample size in each group (if comparing groups)

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
- Use *sensitivity analysis* to create a sample size table for different power, significance, or effect size and then sit and ponder over it for the optimal sample size

References

1. Aggarwal YP, Statistical Methods, concepts, application,computation.Sterlingpublishers(1986)
2. Varalakshmi V et al.Statistics – Higher Secondary, First Year. Tamilnadu Textbook Corporation 2005
3. Langley R,Practical Stats. The Chauser press 1968
4. Barkan H, Annals of Cardiac Anaesthesia (2015)18:74
5. Dunn OJ, Clark VA, Basic stats, a primer for biomed sciences. 4th Ed, 2009,John Weiley and sons inc
6. Harris M, Taylor G. Medical Statistics made easy. Ist Ed,Martin Dunitz,2003
7. Krousel wood M, clinicials guide to stats part 1 The Ochsner Journal (2006) 6:2,68
8. KadamP ,Sample size calculation . International J of Ayurveda (2010)1:55-57
9. Fox N, Hunn A,MathewsN,Sampling and sample size calculation . The NIHR RDS EM / YH(2009)

Assessing the Need for a Transformation

- Several rules of thumb have been suggested .
- Two of these rules should be used only for ratio data.
- If the standard deviation divided by the mean is < 1/4,
- For example, In systolic blood pressure data the sd is 32.4 and the mean is 137.3, so the coefficient of variation is 32.4/137.3 = 0.24. This would be an indication that it is questionable if a transformation is needed.
- An alternative criterion is if the ratio of the largest to the smallest number is < 2, a transformation may not be helpful. e.g. the ratio is 230/87 = 2.6, so perhaps a transformation is helpful, but it again seems borderline
- When,, a log transformation results in near normal data we know that our original data follows a skewed distribution called a *lognormal distribution.*
- Transformation makes data difficult to interpret and compare
- Often, researchers perform their analyses both with and without transformation and see if it affects the final results appreciably.
- If it does not, the transformation may not be used in the final