Basic Statistical Methodology in Clinical Medical Research

HKCOG Research Course

May 22, 2010

Daniel Y.T. Fong
School of Nursing
The University of Hong Kong

Learning Objectives

- Be able to choose an appropriate descriptive statistic
- Be able to know what to report in estimation
- Be able to understand p-value in hypothesis tests
- Be able to calculate a sample size

Are you Ready?

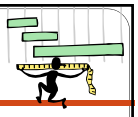
- Descriptive Statistics
- Inferential Statistics
- Sample Size Calculation

What is Statistics?

- Is Variation Good?
- Statistics is
  - about managing "variability",
  - about quantifying uncertainty and the strength of evidence of an experiment

(Evidence based)

Data Types



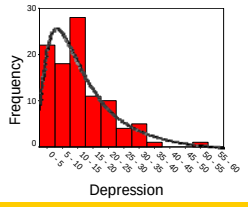
Quantitative (takes numerical values)	**Qualitative/Categorical** (takes coded numerical values)
• **Discrete** (whole numbers) e.g. number of accidents, household size	• **Ordinal** (ranking order exists) e.g. Poor/Average/Good
• **Continuous** (takes decimal places) e.g. height, weight	• **Nominal** (no ranking order) e.g. gender, race

Descriptive Statistics

Why Descriptive Statistics?

Depression levels of 100 cancer patients

6	2	13	17	22	9	19	7	10	5
0	11	7	19	24	16	3	7	13	4
12	29	3	4	33	1	2	6	13	3
25	30	13	25	16	30	12	10	14	2
20	30	4	2	6	12	31	10	3	3
8	24	8	8	4	8	26	12	12	15
2	8	8	20	15	6	14	21	3	8
10	11	10	23	10	14	13	35	22	17
4	10	4	0	20	53	19	5	12	8
11	20	4	13	17	12	11	15	10	2



- Presenting the raw data is often infeasible
- Presenting the distribution is still be overwhelming
- Good to have only a few numbers to summarize a dataset

Good Measure of Location?

Fridge-Like Condition on Buses in Hong Kong

Source: Metro Daily, September 2006.

The minimum on-bus temperature in each bus company is reported in the following table.

Route	Temperature on bus*
CityBus 8X	15.1°C
KMB 57M and 60M	19.2°C and 18.4°C
NWFB 23	21.5°C
Exhaust gas temperature:	without air-conditioning 31.2°C to 32.8°C with air-conditioning 32.1°C to 36.1°C.

*outdoor temperature was 31.2°C

Measures of Location

	Advantages	Disadvantages
Median • Middle value • 中位數	1. Robust", i.e. not affected by aberrant values	1. Does not use all the data 2. Not easy to manipulate mathematically
Mean • Average value • 平均數	1. Is the "expected" value 2. Uses all the data 3. Easy to calculate	1. Not robust to aberrant values 2. Can be difficult to interpret due to aberrant values
Mode • Most popular value • 眾數	1. Can be useful for discrete and categorical measurements	1. Not useful for continuous data 2. May not be unique 3. Does not use all the data

Why Not Just Mean or Median?

Sample 1: 6, 2, 13, 17, 22

Sample 2: 10, 8, 13, 14, 15

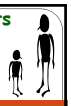
Mean	Median
• Average value	• Middle value
12	13

Are they different?

- Sample 1 has a higher variability than Sample 2
- The mean or the median does not tell variability

Measures of Dispersion

for measurements at least ordinal



Advantages	Disadvantages
Range :- Maximum - minimum	
1. The simplest measure of dispersion	1. Sensitive to aberrant values 2. Does not use all the data
Interquartile range :- 3rd quartile – 1st quartile	
1. Robust to aberrant values 2. Included 50% of the data	1. Does not use all the data 2. Not easy to manipulate mathematically
SD :- Average distance of each score from the mean	
1. For normal distribution, mean and SD describe the entire distribution 2. Easy to manipulate mathematically	1. Sensitive to aberrant values

Deciding the Marriage

Location Measures

Median
- Middle value

Mean
- Average value

Dispersion Measures

Range :-
Maximum - minimum

Interquartile range :-
3rd quartile – 1st quartile

SD :-
Average distance of each score from the mean

Q & A

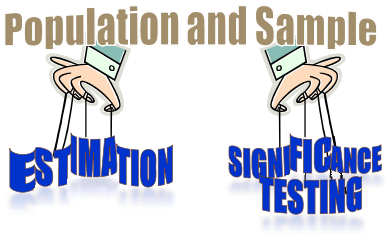
1. Among the descriptive statistics

- ✗ Mean is always an appropriate measure of location
- ✓ Median is always an appropriate measure of location
- ✓ Mean can be preferable to median
- ✗ Interquartile range covers 95% of data

✓ True or False ? ✗

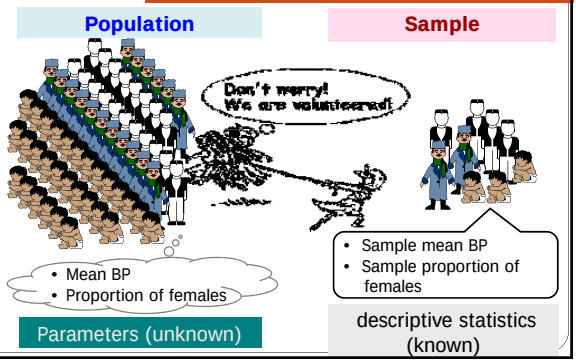
Inferential Statistics

The Three Basic Elements



- The fundamental concept from which the statistical concept is based

Population and Sample



Parameters (unknown)

descriptive statistics (known)

Ready for Analysis?

Estimation

What is the blood pressure, on average, of an African-Chinese after taking calcium for 12 weeks?

- Decide a sample size !
- Let's take it as 5 !

Working Through...

Estimation

Assume:

- there were only 10 African-Chinese in this world, and
- they took calcium for 12 weeks.
- Their BPs were

101, 97, 112, 109, 99, 121, 108, 113, 92, 98

Population

What is the blood pressure, on average, of an African-Chinese after taking calcium for 12 weeks?

Done ?

Estimation

Sample

97, 121, 113, 98, 101 → **Average = 106**

Another Sample

112, 108, 97, 101, 113 → Average = 106.2

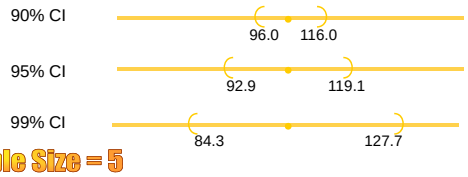
Third Sample

99, 109, 92, 98, 121 → Average = 103.8

Quantification of Precision - Confidence Interval

Estimation

- An interval with certain confidence to include the true population mean
- E.g. a 95% confidence interval has 95% chance to include the true mean



Sample Size = 5

The width reflects the precision

Sample Size = 9 ?

Q & A

2. For quantifying the precision in using a sample mean

- ✗ Small sample size would result in a shorter confidence interval
- ✓ A confidence interval should always be reported during estimation
- ✗ A 95% CI has limits between which, in the long run, 95% of observations fall

✓ True or False ? ✗

Hypothesis Testing

- A choice between
  Null hypothesis and Alternative hypothesis
- Do the data support the Alternative Hypothesis (H_A) more than the Null Hypothesis (H_0)?

Is the New Drug Effective?

Significance testing

- Mean reduction in BP by the new drug is 5.3mmHg more than the old drug
- P-value for testing H_0: same reduction is 0.003 (< 0.05)

↑ ↑
????????? **Significant**

Samples

Placebo drug n=10 Mean reduction = -0.3mmHg

New drug n=11 Mean reduction = 5mmHg

Statistical Errors

Significance testing

- Our statistical conclusions can be wrong though the chance may be low

		Experiment: Significant	Experiment: Insignificant
Actual	New≈Placebo	Patients' risk	✓
	New≠Placebo	✓	Sponsor's risk

- Type I error **/ False Positive Error**
- Type II error **/ False Negative Error**

Definition of p-value

Significance testing

It is the observed chance of committing a False Positive Error

- P-value is a probability between 0 and 1
- The larger is the p-value, the less likely to reject the null hypothesis (significant), Why?

Level of Zidovudine (AZT) in the Blood

Significance testing

- Two groups of AIDS patients underwent the same drug treatment
- AZT measured at 6 hours after treatment

Samples n=5 normal fat absorption; n=9 fat malabsorption

Sample mean difference: 0.34 − 0.2 = 0.14

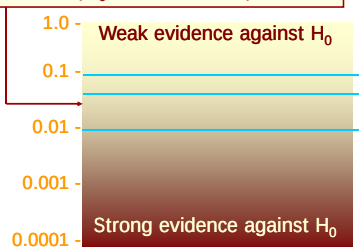
p-value (H_0: no difference) is 0.038

Significant ?

Statistically Significant?

Significance testing

p-value (H_0: no difference) is 0.038



Must be specified before data analysis in confirmatory trials

Clinically Significant?

Significance testing

- Not relevant when the difference is not statistically significant
- When it is statistically significant, the p-value does not help
- Relies on confidence interval

A 95% CI for the difference of AZT in blood = (0.08, 0.20)

Normal fat absorption - Mal-absorption

Clinically Significant ?

Uniform Requirements for Manuscripts Submitted to Biomedical Journals

Significance testing

….When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as **confidence intervals**). **Avoid relying solely on statistical hypothesis testing**, such as the use of p-values, which fails to convey important quantitative information …..

~ Adopted by BMJ, JAMA, etc.

Q & A

3. In a small RCT of a new treatment in acute myocardial infarction, the mortality in the treated group was half that in the control group, but the difference was not significant. We can conclude that

- ✗ the treatment is useless
- ✗ there is no point in continuing to develop the treatment
- ✓ we should carry out a new trial of much greater size

✓ True or False ? ✗

Sample Size Calculation

How?

- Design
- Main objective
- Main outcome(s)
- Method of analysis
- Level of significance

Determine your Plan of the Main Analysis

Look for additionally required information and method of calculating the Sample Size

- In a study protocol, the sample size calculation method should be **consistent** to the analysis plan

Two Types of Analysis

1. Parameter Estimation
   - e.g. prevalence of hypertension, mortality rate, quality of life level, etc.
2. Hypothesis/Significance Testing
   - e.g. blood pressure changed after 12-week of treatment, a test treatment differs from the standard treatment, etc.

A Motivating Study

Parameter Estimation

- To calculate the sample size for **estimating** the blood pressure of African Chinese after taking calcium for 12 weeks.

Hint

How would you do the analysis in order to achieve this objective?

Analysis Plan

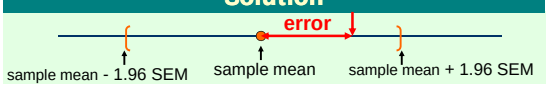
A 95% CI for the mean BP at 12-week will be constructed.
i.e. (sample mean - 1.96 SEM, sample mean + 1.96 SEM)

To calculate the sample size for estimating the blood pressure of African Chinese after taking calcium for 12 weeks.

The Motivating Study

Parameter Estimation

Solution



- In order to have a maximum error of **e**, we must have

$$1.96\,SEM \le e \Rightarrow 1.96\frac{SD}{\sqrt{n}} \le e \Rightarrow n \ge \left[\frac{1.96\,SD}{e}\right]^2$$

From clinicians: **e** = 10 and **SD** = 15 (Average = 106)

n = 8.64 and **take it as 9 !**

Key Information to be Determined

Parameter Estimation

1. Confidence level
   - Often taken as 95%
2. Maximum tolerable error
3. Dispersion
   - For estimation of mean, it is the anticipated SD
   - For estimation of proportion, it is the anticipated proportion
     (can be taken conservatively as 0.5 if no idea at all)

Effect Size

Hypothesis Testing

Effect size is the smallest clinically significant magnitude

- Its determination is a medical and scientific judgment, not a statistical decision
- The smaller the effect size sought by a study, the larger the sample size required

Statistical Method?

Hypothesis Testing

- Testing about two means:
  - Student's t-test
  - Wilcoxon rank sum test/Mann-Whitney U
- Testing about two proportions:
  - χ^2-test
  - Fisher's exact test
  - Logistic regression

Anticipated results with standard treatment (for RCTs)?

Hypothesis Testing

- Very often, there isn't much information about the new/test treatment.
- It is fair to expect the availability of more complete information about the standard treatment.
  - e.g. mean and SD for continuous outcomes, proportion for binary outcomes, etc.
- Make good use of published results

Key Information to be Determined – Not exclusive

Hypothesis Testing

- Primary objective (research hypothesis)
  - e.g. superiority or equivalence/non-inferiority?
- Primary outcome
  - e.g. continuous or categorical?
- Statistical methods
  - must be consistent with the statistical analysis plan
- For controlled studies,
  - Anticipated results with standard treatment
  - Smallest clinically important difference to detect and how certain (power)

Efficacy of borage oil

~ Takwale et al. (2003) BMJ.

Hypothesis Testing

Planning the Main Analysis

Main objective	To study the efficacy of borage oil in subjects with atopic eczema	From Clinician
Design	Randomised, placebo-controlled, parallel groups	From Clinician
Main outcome	% change of total sign score at 12 weeks	From Clinician

A two independent sample t-test will be used to compare the % changes of total sign score between subjects who received borage oil and subjects who received placebo. A p-value < 5% will conclude statistical significance.

From Statistician

Efficacy of borage oil (2)

Hypothesis Testing

Calculating the Sample Size

Smallest meaningful difference (Effect Size, e)	20%	From Clinician
SD in a group	35%	From pilot data
Power (1-β)	80% (generally accepted as the minimal)	Pretty standard

Based on t-test

$$n = \frac{2(SD)^2}{e^2} \times f(\alpha,\beta) + \frac{z^2_{1-\alpha/2}}{4} \quad where \ f(\alpha,\beta) = (z_{1-\alpha/2} + z_{1-\beta})^2$$

$z_{1-\alpha/2} = 1.96$

$z_{1-\beta} = 0.84$

n = 48.98 and **take it as 49 / group**

From Statistician

Catches

- Check for aberrant values when choosing a descriptive statistic
- Always report a confidence interval in estimation
- P-value is the observed chance of committing a false positive error
- Plan for the primary analysis before calculating a sample size

References

- Campbell MJ and Machin D. Medical Statistics: A Commonsense Approach. Third edition, John Wiley & Sons: New York, 1999.
- Altman DG. Practical Statistics for Medical Research. Chapman & Hall: New York, 1991
- Machin D, Campbell M, Fayers P, and Pinol A. Samples size tables for clinical studies. 2nd ed. Blackwell Science: United Kingdom, 1997.