Hypothesis Testing

Dr Trevor Bryant

Learning Outcomes

Following this session you should be able to:

- Understand the concept and general procedure of hypothesis testing
- Understand the concept and interpretation of P values
- Explain the relationship between CI (point estimate ± 1.96 x S.E) & Hypothesis Testing
- Describe Type I & Type II Errors

Hypothesis testing - milestones

- Develop the research question
- Develop the research hypothesis
- State it as a statistical hypothesis
- Test the hypothesis
- Was it a good idea?
- Next question(s)

The Four Elements of a Research Question

- Cells, Patient or Population
  - What or Who is the question about?
- Intervention or Exposure*
  - What is being done or what is happening to the cells, patients or population?
- Outcome(s)
  - How does the intervention affect the cells, patients or population?
- Comparison(s)
  - What could be done instead of the intervention

**Intervention is intentional whereas an exposure is incidental*

Defining a Research Hypothesis

'A well-defined hypothesis crystallizes the research question and influences the statistical tests that will be used in analyzing the results'

http://intra.som.umass.edu/nakosteen/Topics/Developing%20the%20research%20design.doc [Accessed 17 Feb 2009]

You cannot prove a hypothesis

- Falisifiability
  - (Karl Popper, 1902-1994)
- Scientific laws cannot be shown to be True or False
- They are held as Provisionally True
- 'All Swans are White'
  - (David Hume,1711-1776)

What is a Hypothesis?

- A tentative statement that proposes a possible explanation to some phenomenon or event
- A useful hypothesis is a testable statement which may include a prediction
- Any procedure you follow without a hypothesis is not an experiment

Formalized Hypothesis

- **IF** and **THEN**
- **Specify a tentative relationship**
- **IF** skin cancer is **related** to ultraviolet light, **THEN** *people with a high exposure to UV light will have a higher frequency of skin cancer*

  Dependent variable

  Independent variable

Disproving a hypothesis

- Collect evidence
- If evidence supports current hypothesis
  Hold hypothesis to be *Provisionally True*
- If evidence does not support hypothesis
  Reject hypothesis and develop new one
- Statistical testing uses Null Hypothesis
  - No difference unless unlikely event (p)
  - Alternative hypothesis – a difference?
  - Swans

Statistical Hypothesis testing - Overview

- Define the problem
- State null hypothesis (H_0)
- State alternative hypothesis (H_1)
- Collect a sample of data to gather evidence
- Calculate a test statistic

$$\text{Test statistic} = \frac{\text{observed value - hypothesised value}}{\text{standard error of observed value}}$$

- Relate test statistic to known distribution to obtain P value
- Interpret P value

Defining the problem

- The *null hypothesis* assumes **No Effect**
  - H_0 : There is no treatment effect in the population of interest
- The *alternative hypothesis* opposite of null hypothesis
  - H_1 : There is a treatment effect in the population of interest

Note: These are specified before collecting the data, they relate to the population not the sample and usually no direction is specified for the effect

Calculating the test statistic

The test statistic summarises the data from the sample in a single number. It's size indicates the amount of evidence gathered for either hypothesis

- The choice of test statistic will depend on the type of data collected and the hypotheses of interest
- 'Large' test statistic - more evidence for H_1
- Values of the test statistic are standardized and can 'compare to published tables' – calculated

How do we choose the test statistic?

- What is the measurement of interest?
  Means, proportions, etc
- What is the distribution of the measurement
  Normal or skewed
- How many groups of patients are being studied?
  1, 2, 3 or more
- Are they independent groups?
  or paired

Interpretation of the P value

The P value is the probability of getting a test statistic as large as, or larger than, the one obtained in the sample if the null hypothesis were true

It is the probability that our results occurred by chance

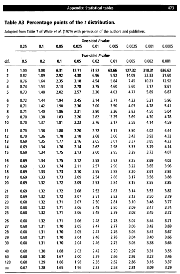
Interpretation of the P value (2)

- By convention, **P values of <.05** are often accepted as "statistically significant" in the medical literature
- It is an **arbitrary cut-off**
- A cut-off of P <.05 means that in about 5 out of 100 (1 in 20) experiments, a result would appear significant just by chance ("Type I error")
- We can use other P values for example 0.01

Interpretation of the P value (3)

- **Large P value** (usually > 0.05)
- Likely to have got results by chance if H_0 was true
  - *Accept null hypothesis*
  - *Result is non-significant*
- **Small P value** (usually < 0.05)
- Unlikely to have got results by chance if H_0 was true
  - *Reject null hypothesis – accept alternative hypothesis*
  - *Result is significant*

Where do P >0.05 P >0.01 P >0.001 fit in?



Appendix: Statistical tables 473

Table A3 Percentage points of the *t* distribution.

Adapted from Table 7 of White *et al.* (1979) with permission of the authors and publishers.

	One-sided *P*-value								
	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
	Two-sided *P*-value								
d.f.	0.5	0.2	0.1	0.05	0.02	0.01	0.005	0.002	0.001
1	1.00	3.08	6.31	12.71	31.82	63.66	127.32	318.31	636.62
2	0.82	1.89	2.92	4.30	6.96	9.92	14.09	22.33	31.60
3	0.76	1.64	2.35	3.18	4.54	5.84	7.45	10.21	12.92
4	0.74	1.53	2.13	2.78	3.75	4.60	5.60	7.17	8.61
5	0.73	1.48	2.02	2.57	3.36	4.03	4.77	5.89	6.87
6	0.72	1.44	1.94	2.45	3.14	3.71	4.32	5.21	5.96
7	0.71	1.42	1.90	2.36	3.00	3.50	4.03	4.78	5.41
8	0.71	1.40	1.86	2.31	2.90	3.36	3.83	4.50	5.04
9	0.70	1.38	1.83	2.26	2.82	3.25	3.69	4.30	4.78
10	0.70	1.37	1.81	2.23	2.76	3.17	3.58	4.14	4.59
11	0.70	1.36	1.80	2.20	2.72	3.11	3.50	4.02	4.44
12	0.70	1.36	1.78	2.18	2.68	3.06	3.43	3.93	4.32
13	0.69	1.35	1.77	2.16	2.65	3.01	3.37	3.85	4.22
14	0.69	1.34	1.76	2.14	2.62	2.98	3.33	3.79	4.14
15	0.69	1.34	1.75	2.13	2.60	2.95	3.29	3.73	4.07
16	0.69	1.34	1.75	2.12	2.58	2.92	3.25	3.69	4.02
17	0.69	1.33	1.74	2.11	2.57	2.90	3.22	3.65	3.96
18	0.69	1.33	1.73	2.10	2.55	2.88	3.20	3.61	3.92
19	0.69	1.33	1.73	2.09	2.54	2.86	3.17	3.58	3.88
20	0.69	1.32	1.72	2.09	2.53	2.84	3.15	3.55	3.85
21	0.69	1.32	1.72	2.08	2.52	2.83	3.14	3.53	3.82
22	0.69	1.32	1.72	2.07	2.51	2.82	3.12	3.50	3.79
23	0.68	1.32	1.71	2.07	2.50	2.81	3.10	3.48	3.77
24	0.68	1.32	1.71	2.06	2.49	2.80	3.09	3.47	3.74
25	0.68	1.32	1.71	2.06	2.48	2.79	3.08	3.45	3.72
26	0.68	1.32	1.71	2.06	2.48	2.78	3.07	3.44	3.71
27	0.68	1.31	1.70	2.05	2.47	2.77	3.06	3.42	3.69
28	0.68	1.31	1.70	2.05	2.47	2.76	3.05	3.41	3.67
29	0.68	1.31	1.70	2.04	2.46	2.76	3.04	3.40	3.66
30	0.68	1.31	1.70	2.04	2.46	2.75	3.03	3.38	3.65
40	0.68	1.30	1.68	2.02	2.42	2.70	2.97	3.31	3.55
60	0.68	1.30	1.67	2.00	2.39	2.66	2.92	3.23	3.46
120	0.68	1.29	1.66	1.98	2.36	2.62	2.86	3.16	3.37
∞	0.67	1.28	1.65	1.96	2.33	2.58	2.81	3.09	3.29

Example of a hypothesis test

Example of a hypothesis test

Randomised controlled trial of cranberry-lingonberry juice and Lactobacillus GG drink for the prevention of urinary tract infections in women. Kontiokari et al. BMJ (2001) 322: 1571-3

150 women were randomised to three groups (cranberry-lingonberry juice, lactobacillus drink or control group).

At six months, 8/50 (16%) women in the cranberry group, 19/50 (38%) in the lactobacillus group, and 18/50 (36%) in the control group had had at least one recurrence.

Question: Is there any EFFECT of cranberry to prevent infection?

Example of a hypothesis test

What is the Hypothesis?

If women drink cranberry-lingoberry juice **then** there will be a reduction in the recurrence of urinary tract infection

Statistical Hypothesis

Null H_0: There are no differences in recurrence rates among women in the population who drink cranberry-lingoberry juice, lactobacillus drink or neither of these

Alternative H_1: There is a difference in the recurrence rates between these three groups in the population

Example of a hypothesis test

- Which test should be used?

  Chi-squared test

- What is the test statistic?

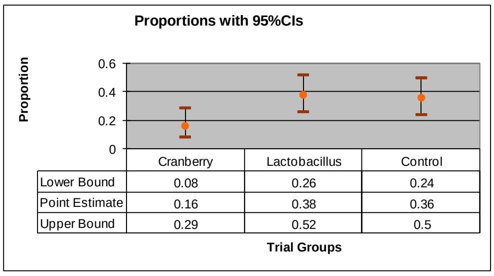
  X^2 7.05, P = 0.03

- How to interpret the result?

  Reject null hypothesis

  There is a significant difference in recurrence rates between these three groups (based on 5% significance)

Example of a hypothesis test



	Cranberry	Lactobacillus	Control
Lower Bound	0.08	0.26	0.24
Point Estimate	0.16	0.38	0.36
Upper Bound	0.29	0.52	0.5

5 minute break

Errors in Hypothesis testing

Jury's verdict	True state of Defendant	
	Defendant really is **Guilty**	Defendant really is **Innocent**
Guilty	✓ Correct Decision	✗
Not guilty	✗	✓ Correct Decision

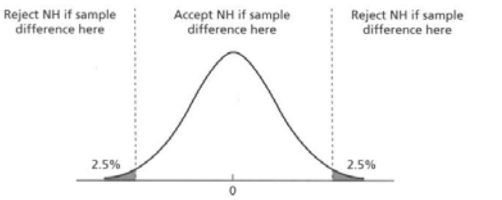
Types of Error in hypothesis testing

Statistical Decision	True state of null hypothesis - Reality	
	Null hypothesis is True	Null Hypothesis is False
Accept	H_0 accepted correctly	Type II error (β)
Reject	Type I error (α)	H_0 rejected correctly

Type I error

- The probability that we reject null hypothesis when it is true
- 'False positive'
- Rejected H_0 because the results occurred by chance
- Conclude that there is a significant effect, even though no true effect exists
- Probabilities of Type 1 error called – alpha (α) Determined in advance, typically 5%

Type 1 Error – Null Hypothesis is True



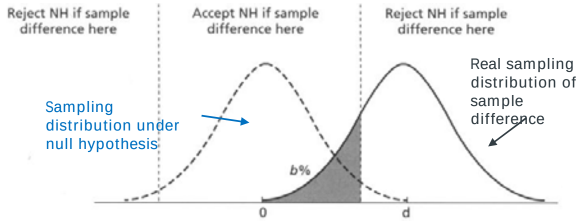
Shaded areas gives the probability that the Null hypothesis is wrong rejected

Adapted from Kirkwood & Sterne 2nd Ed

Type II error

- The probability that we accept null hypothesis when it is false
- 'False Negative'
- Accept H_0 even though it is not true
- Conclude that there is no significant effect, even though a true difference exists
- Probabilities of Type II error called – beta (β)

Type II Error – Null Hypothesis is False



Shaded area is the probability (b%) that the null hypothesis fails to be rejected

Adapted from Kirkwood & Sterne 2nd Ed

Type II error rate

- Type II error rate depends on :
  - the size of the study
  - the variability of the measurement
- The implications of making either a type I or type II error will depend on the context of the study

The Power of the Study

The *power* of the study is the probability of correctly detecting a true effect

Or the probability of correctly rejecting the null hypothesis

Power = 100% - Type II error rate = (1 – β) x 100

The Power of the Study (2)

- The power will be low if there are only a few observations

  taking a larger sample will improve the power

- The power will be low if there is variability amongst the observations

  reducing variability will improve power

- Ideally we would like a power of 100% but this is not feasible

  usually accept a power of 80%

Things to consider

We can never be 100% certain that the correct decision has been reached when carrying out a hypothesis test

An hypothesis test cannot prove that a null hypothesis is true or false. It only gives an indication of the strength of evidence

References:

- Altman, D.G. *Practical Statistics for Medical Research*. Chapman and Hall 1991. Chapter 8
- Kirkwood B.R. & Sterne J.A.C. *Essential Medical Statistics*. 2nd Edition. Oxford: Blackwell Science Ltd 2003. Chapter 8
- Machin D. and Campbell M.J. *The Design of Studies for Medical Research*, John Wiley and Sons 2005 Chapter1

Questions