statstutor

community project

encouraging academics to share statistics support resources

Statistical Methods 14 Sample Size Calculations

Based on materials provided by Coventry University and Loughborough University under a National HE STEM Programme Practice Transfer Adopters grant

www.statstutor.ac.uk

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Type II errors and sample size calculations

We shall consider:

- Things that can go wrong in statistical testing (recap from Workshop 8)
- Effect sizes
- Statistical power
- Power calculations

www.**stats**tutor.ac.uk

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Things that can go wrong

	H_0 really true	H_0 really false
H_0 rejected	Type I error	Correct decision
H_0 accepted	Correct decision	Type II error

- Type I error is equivalent to **convicting the innocent**
- Type II error is equivalent to **acquitting the guilty**
- Reducing the chance of a Type I error by changing the significance threshold increases the chance of a Type II error
- The best solution is to **increase the sample size**
- The **power** of a test is **1 – Probability(Type II error)**

www.**stats**tutor.ac.uk

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Effect sizes

- Different definitions
- A common one is Cohen's *d*:
  - Given two samples X_1 and X_2, $d = |X\downarrow 1 - X\downarrow 2 / s|$
  - Where *s* is the standard deviation of the combined sample X_1 and X_2
  - Cohen (1988) classifies *d* as follows:

Effect size	**Interpretation**
0.2 to 0.3	Small
About 0.5	Medium
0.8 and above	Large

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Example

For the Pulse data set:

- X_1 = Initial pulse for non/not regular smokers
- X_2 = Initial pulse for regular smokers
- $\bar{X}_1 = 71.94$
- $\bar{X}_2 = 76.00$
- $s = 10.75$
- $d = 0.378$
- So the effect size of smoking on pulse for the samples is “small to medium”

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Statistical power

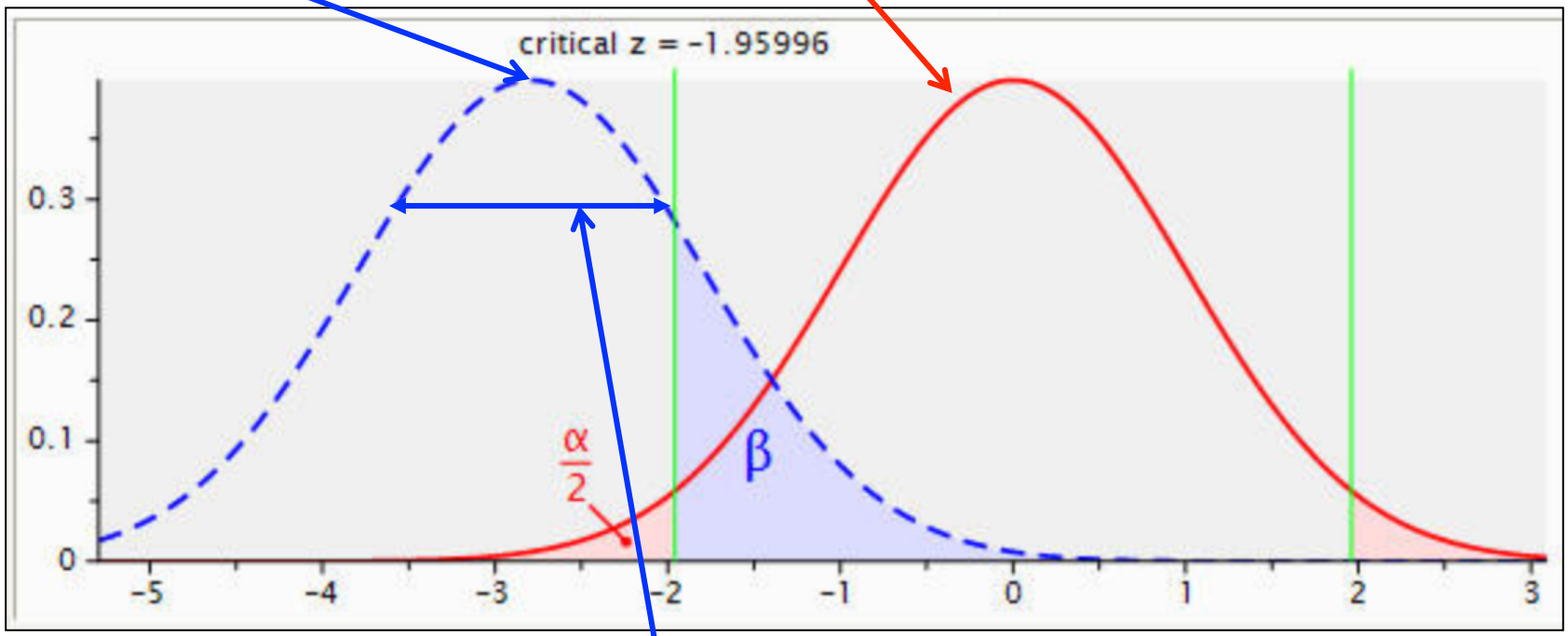
- Just as there is a standard level of statistical significance (α) of 0.05 to reduce the possibility of Type I errors there is also a recommended minimum value for the power of statistical test of **0.8**
- This means the cut-off value for the probability of a Type II error (β) is **0.2**
- Given α and an estimate of the effect size (*d*) the minimum sample size(s) required to achieve a given power level (1-β) can be calculated
- **Note**: The effect size must be estimated **before** the experiment, e.g. from prior research, and not based on the data

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Example – normal distribution

- Standard normal distribution (mean 0, standard deviation 1)
- Estimate of second sample parameter mean based on X_2



- $\alpha = 0.05$, $\beta = 0.2$
- Standard deviation of estimate is proportional to $1/\sqrt{n_2}$

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Sample sizes for t-tests

Paired samples t-test, $\alpha = 0.05$, $\beta = 0.8$:

Effect	Small	Medium	Large
Effect size	0.2	0.5	0.8
Minimum total sample size	199	34	15

Independent samples t-test, $\alpha = 0.05$, $\beta = 0.8$, equal sample sizes:

Effect	Small	Medium	Large
Effect size	0.2	0.5	0.8
Minimum sample size per group	392	64	26

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Application – Pulse data

- d = 0.378 (from the data)
- If we had assumed a small effect size we would have needed 392 in each group with equally sized groups
- If we had assumed a medium effect size we would have needed 64 in each group with equally sized groups
- As there were 64 in Group 1 and 27 in Group 2 we could have used http://www.biomath.info/power/ttest.htm to estimate the group sizes with this ratio for an effect size of 0.5:
  - Minimum Group 1 size = 105
  - Minimum Group 2 size = 45
- Clearly there were insufficient sample sizes to reduce the risk of a Type II error to a satisfactory level

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Bibliography

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. New York: Lawrence Erlbaum.

Columbia University Medical Center (n. d.) *What is your study design?* Available at: http://www.biomath.info/power/index.htm [Accessed 8/01/14].

Fox, N., Hunn, A. & Mathers, N. (2009) *Sampling and sample size calculation*. [pdf]Available at: http://rds-eastmidlands.nihr.ac.uk/resources/doc_download/9-sampling-and-sample-size-calculation.html [Accessed 8/01/14].

Heinrich-Heine-Universität Düsseldorf (2013) *G*Power 3*. Available at: http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/ [Accessed 8/01/14].

statstutor (2006) *Statistics: An introduction to sample size calculations*. [pdf] Available at: http://www.statstutor.ac.uk/resources/uploaded/sample-size.pdf [Accessed 8/01/14].

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield