Statistics

Jo Sweetland
Research Occupational Therapist

First a test…

… testing your knowledge on

statistics!

…please be honest with yourself

Statistics

- Statistics give us a common language to share information about numbers
- To cover some key concepts about statistics which we use in everyday clinical research
  - Probability
  - Inferential statistics
  - Power

What are statistics for?

- Providing information about your data that helps to understand what you have found - ‘descriptive’ statistics
- Drawing conclusions which go beyond what you see in your data alone – ‘inferential’ statistics
  - What does our sample tell us about the population
  - Did our treatment make a difference?
  - Depends on the probability theory

Probability
two ways to think about it

The probability of an event, say the outcome of a coin toss, could be thought of as:

The chance of a single event
(toss one coin 50% chance of head)
OR
The proportion of many events
(toss infinite coins, 50% will be heads)

It is the same thing and is known as the *frequentist* view

Probability

- Definition: a measurement of the likelihood of an event happening.
- Calculating probability involves three steps
  - E.g. coin toss
  - Simplifying assumptions
    - P(heads)=P(tails); no edges
  - Enumerating all possible outcomes
    - heads/tails=2 outcomes
  - Calculate probability by counting events of a certain kind as a proportion of possible outcomes
    - P(heads)=1 out of 2 = ½ or 50% or 0.5

Basic laws for combining Probability

- The additive law
  - The probability of either of two or more mutually exclusive events occurring is equal to the sum of their individual probabilities
  - E.g. toss a coin – can be heads or tails but NOT both **P(head OR tail)= .5 + .5 =1**
- The multiplicative law
  - The probability of two or more independent events occurring together = P(A) x P(B) x P(C) etc
  - E.g. toss two coins – probability of two heads **P(head&head)= .5 x .5 = .25**

Probability
an example

- Three drug treatments for severe depression
  - Drug A effective for 60%
  - Drug B effective for 75%
  - Drug C effective for 43%
- Assume independence
- What proportion of people would benefit from drug treatment?

Probability
an example

- Is it 60% x 75% x 43% = 20%?
  - Less than any one treatment
  - This 20% represents those who would improve from each and every drug
- We would want those who would improve from *some combination* of the three
- Solution:

those who improve at all = everyone – those who don't improve from any drug

= 40% x 25% x 57% = 6%

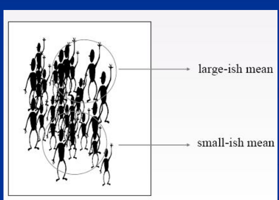
- So answer = 100 – 6 = 94%

Inferential statistics:
main concepts

- Populations are too big to consider everyone, so we randomly sample
- Sampling is necessary, but it introduces variation
  - Different samples will produce different results
  - Systematic and non-systematic
    - E.g. height – men tend to be systematically taller than women but lots of random variability
- Variation is what we study
- The difference between characteristics of the sample and the (theoretical) population is called 'sampling error'
- Statistics = sets of tools for helping us make decisions about the impact of sampling error on measurements

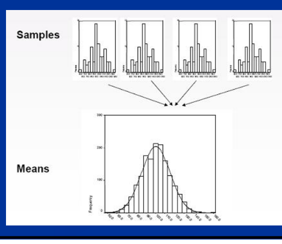
Sampling

Sampling is an inherently *probabilistic* process



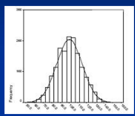
Sampling distributions

- Take lots of small samples from the same large population
- Calculate the mean each time and plot them



Normal distribution

- Sample means are "normally distributed"



- This happens regardless of the population so is a powerful tool
- Commonest value = population mean
- Spread of means gets less as sample size increases
  - Smoothing the effect of extreme values

Standard deviation

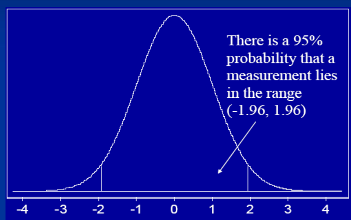
- A standard deviation is used to measure the amount of variability or spread among the numbers in a data set. It is a standard amount of deviation from the mean.
- Used to describe where most data should fall, in a relative sense, compared to the average. E.g. in many cases, about 95% of the data will lie within two standard deviations of the mean (the empirical rule).

Empirical rule

- As long as there is a normal distribution the following rules applies:
  - About 68% of the values lie within one standard deviation of the mean
  - About 95% of values lie within 2 standard deviation of the mean
  - About 99.7% of values lie within 3 standard deviation of the mean

Normal distribution

- Most of the data are centred around the average in a big lump, the farther out you move on either side the fewer the data points.
- Most of the data to lie within two standard deviations of the mean.
- Normal distribution is symmetric because of this the mean and the median are equal and both occur in the middle of the distribution.

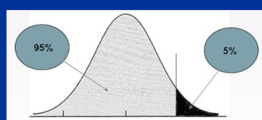


Central Limit Theorem

- The central limit theorem tells us that, no matter what the shape of the distribution of observations in the population, the sampling distribution of statistics derived from the observations will tend to 'Normal' as the size of the sample increases.
- This theorem gives you the ability to measure how much your sample will vary, without having to take any other sample means to compare it with. It basically says that your sample mean has a normal distribution, no matter what the distribution of the original data looks like.

Rejection region

If we can describe our population in terms of the likelihood of certain numbers occurring, we can make inferences about the numbers that actually do come up



Probability = area under curve between intervals
Shaded area = rejection region = area in which only 1 in 20 scores would fall

Null Hypothesis (H_0)

'a straw man for us to knock down'

- H_0: 'the sample we got was from the general population'
- H_A: 'the sample was from a different population'
- We calculate the probability it was from H_0 population
- If <5%, we're prepared to accept that the sample was NOT from the general population, but from some other population
- This cut-off is denoted as alpha, α. Sometimes we choose a smaller value e.g. 1% or even 1/10th%
- So a null hypothesis is a hypothesis set up to be nullified or refuted in order to support an alternative hypothesis

Type I error

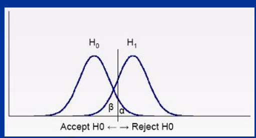
- We will get it wrong 5% of the time
- One in twenty (5%) is considered a reasonable risk - more than one in twenty is not
- ***Type I error = the probability of rejecting the null hypothesis when it is in fact true***

  ("Cheating" – saying you found something when you didn't)
- False positive
- The greater the Type I error the more spurious the findings and study be meaningless
- However if you do more than one test the overall probability of a false positive will be greater than .05

Type II error and power

- Type II error = flip-side of Type I error
- ***Probability of accepting the null hypothesis when it is actually false***

  ("gutting!" not finding something that was really there)
- False negative
- If you have a 10% chance of missing an effect when it is there, then you obviously have a 90% chance of finding it – 90% power
- Power = (1- prob of type II error)



What affects power?

- Distances between distributions – e.g. the mean difference, effect size
- Spread of distributions
- The rejection line (alpha: = .05, .01, .001)

excel example of power.xls

Doing a power calculation

- Usually done to estimate sample size
  - Decide alpha (usually 5%)
  - Decide power (often 80% but ideally more)
  - Ask a statistician to help!

Our randomised control trial

Evaluation of an Early Intervention Model of Occupational Rehabilitation

- A randomised control trial
- A comprehensive evaluation of an early intervention (proactive) vocational rehabilitation service primarily focusing on work related outcomes, cost analysis, general health and well being outcomes.

‘Powering’ our study

Our sample size:

"It is considered **clinically important** to detect at least a difference in scores on the Psychological MSimpact sub-scale (the primary outcome) of 10 points. Using an estimated standard deviation of 23 points the study will require 112 patients per group to detect a 10 point difference with 90% power and a significance level of 5%. In order to allow for up to 30% dropout over the 5 year follow-up period, the target sample size is inflated to 146 per group. This sample size calculation assumes the primary analysis will be a 2 sample t-test and that assumptions of Normality are appropriate for the primary outcome."

- [reference Machin D, Campbell M, Fayer P, Pinol A. Sample size tables for clinical studies Blackwell Science 1997]"

Reference List

- Rowntree, D. Statistics without Tears – an introduction for non-mathematicians. Penguin Books 2000
- Rumsey, D. Statistics for Dummies. Wiley Publishing 2003
- Machin D, Campbell M, Fayer P, Pinol A. Sample size tables for clinical studies Blackwell Science 1997