Statistics and Clinical Evidence

An introduction to statistical methods for analysing clinical data

Tony Brady

Reference books

- Altman DG. *Practical Statistics for Medical Research*
- Bland JM. *An Introduction to Medical Statistics*

Clinical trials

- Methods suitable for analysing clinical trial data
- Parallel design - usually 2 groups, one group 'placebo' or control, other group 'active' or experimental
- Patients are randomly allocated to the two groups to avoid bias
- Not going to cover methods specific to crossover trials, laboratory experiments or non-randomised studies

Outcome measure

- Primary outcome of trial should be predefined in protocol
- Appropriate analysis depends on type of outcome
- Main analysis to be performed should also be predefined before outcome data is seen

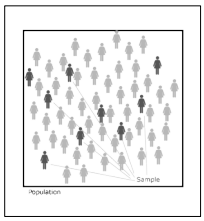
Types of outcome

- Continuous - blood pressure, length of hospital stay, number of asthma attacks
- Categorical - presence/absence of ulcer, vital status, Rankin scale: 0 (no symptoms) to 5 (severe disability) and 6 (death)
- Time to event - survival time, time to cancer progression, time to relapse

Continuous outcomes

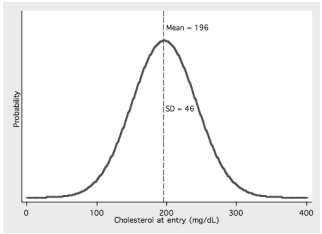
- Example: trial of placebo vs. beta-blockade to lower blood pressure in women
- Sample of 62 women in trial (30 placebo and 32 beta-blockade)
- Outcome is systolic blood pressure (continuous)

Populations and samples

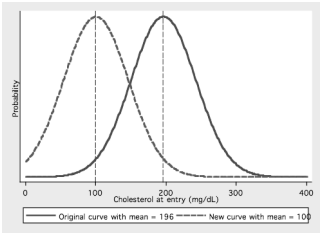


- We are interested in the population
- But we analyse a sample
- From the sample make *inferences* about the population
- If we do not sample randomly can get *bias*

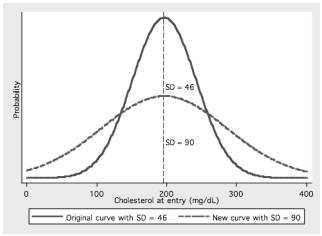
Normal distribution



Different mean



Different standard deviation

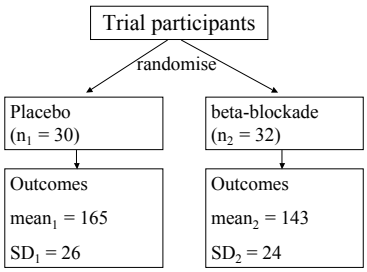


Two-sample t-test (1)

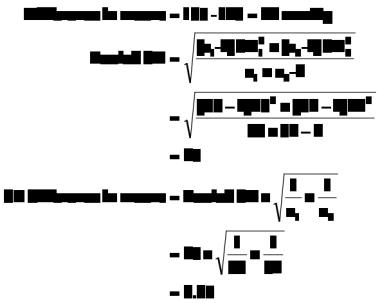
- Compare mean of X in group 1 with mean of X in group 2
- Null hypothesis is that there is no difference between the means in each group
- Example:
  - mean systolic blood pressure (SBP) in placebo group was 165mmHg, SD=26
  - mean SBP in beta-blocker group was 143mmHg, SD=24
  - Is there evidence that the beta-blocker group had different average SBP?

Simulation

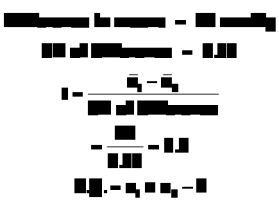
- Two groups
- Population outcome in group 1 has Normal distribution with mean = 100, SD=60
- Population outcome in group 2 has Normal distribution with mean = 200, SD=60
- Take samples of size N=100 from each group (so N=200 in total)



Two sample t-test (2)



Two sample t-test (3)



From statistical tables, t=3.5 on 60 *d.f.* gives P<0.001

There is strong evidence of a difference in SBP between the two groups

More extreme difference = bigger t = smaller P value

P value reflects the strength of evidence against the null hypothesis (smaller P = stronger evidence)

Assumptions

- The outcome is approximately normally distributed
- The SD in the two groups is approximately equal
- Non-parametric equivalent to two sample t-test is Mann-Whitney test

Note: always use "2 sided" P-values. "1 sided" are rarely justified

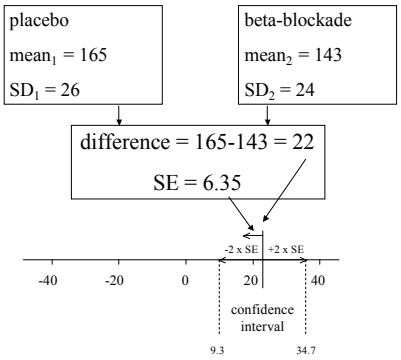
Confidence intervals

- Useful to know estimated size of treatment effect and associated uncertainty
- Difference in mean outcome between the two groups is treatment effect
- Confidence interval (CI) is calculated using SE and critical values of t-distribution:
- 95% CI: effect ± ($t_{0.975}$ x SE)

A value you look up from the t-distribution (usually around 2.0)

Confidence intervals (2)

- Example
  - mean SBP was 22mmHg lower in beta-blocker group compared to placebo group
  - Standard error was 6.35, $t_{0.975}$ on 60 *d.f.*=2.00
  - 95% CI: 22 ± (2.00 x 6.35)
  - which is 9.3mmHg to 34.7mmHg
  - "We are 95% certain that the true treatment effect lies in the range 9.3 to 34.7mmHg"



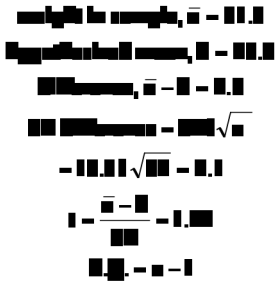
One sample t-test

- Used to compare the mean of a random variable with a fixed value
- e.g. are the trial participants representative of women in general population?
- Compare average age & weight in sample with population averages

One sample t-test (1)

- Compare mean of X with hypothesized value
- Example:
  - mean weight in sample, 51.3 kg, SD=16.3
  - mean weight in population 48.7 kg
  - Are the women in our sample heavier, on average, than women generally?

One sample t-test (2)



From statistical tables, t=1.24 on 61 *d.f.* gives P=0.22

There is little evidence that the women in our sample differ from women generally in weight

Assumptions

- The outcome is approximately normally distributed
- Non-parametric equivalent to one sample t-test is sign test

Paired t-test

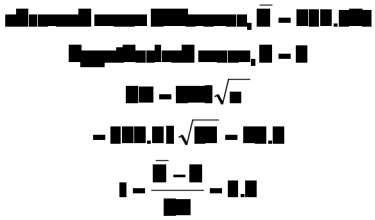
- Two-sample t-test assumes groups are independent
- This won't be true if comparison is within the same patient
- Example:
  - does dietary intake change over the course of the beta-blockade trial?
  - Intake measured at baseline and end of the trial in each woman

Dietary intake (kJ)

Subject	Baseline	End	Difference
1	5260	6410	-1150
2	5610	5520	90
3	7810	8230	-420
4	4570	4335	235
5	6190	5220	970
.	.	.	.
Mean	5486.2	5251.7	234.5
SD	1158.3	1296.7	322.4

If dietary intake was unchanged we expect the mean difference to be 0

We apply the one sample t-test where the hypothesized mean is 0



From statistical tables, t=5.7 on 61 *d.f.* gives P<0.001

There is strong evidence that the women in our sample reduced their dietary intake during the trial

More than 2 groups

- How do you analyse 3 arm trial (A vs. B vs. C)?
- Inefficient to use two sample t-tests for A vs. B and A vs. C
- Solution is to use one way analysis of variance (ANOVA)
- ANOVA with only 2 groups is identical to two sample t-test

Categorical outcomes

Trial of infra-red stimulation (IRS) vs. placebo for pain caused by cervical osteoarthrosis

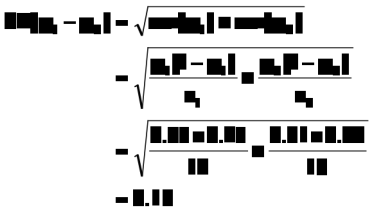
		IRS	Placebo	Total
Improvement in pain	Yes	9	4	13
	No	3	9	12
	Total	12	13	25

IRS $p_1 = 9/12 = 0.75$

Placebo $p_2 = 4/13 = 0.31$ $p_1 - p_2 = 0.44$

CI for difference in proportions

- Confidence interval (CI) is calculated using SE and critical values of Normal distribution:
- 95% CI: effect ± $Z_{0.975}$ x SE
- $Z_{0.975}$ =1.96 (doesn't depend on *d.f.*)



CI for difference in proportion

- $p_1 - p_2 = 0.44$
- SE($p_1 - p_2$)= 0.18
- 95% CI is 0.44 ± (1.96 x 0.18)
- which is 0.09 to 0.79

"We are 95% certain that the true difference in proportion lies between 0.09 and 0.79"

Chi-squared test

- Used to test for differences in proportion between 2 or more groups
- Calculate expected frequencies under the null hypothesis (that the outcome variable and group variable are unrelated)
- Not valid if <80% of expected frequencies <5, in this case use Fisher's exact test

Observed, O:

		IRS	Placebo	Total
Improvement in pain	Yes	9	4	13
	No	3	9	12
	Total	12	13	25

Under null hypothesis:

P(pain=yes) = 13/25 = 0.52

P(IRS) = 12/25 = 0.48

Expected number:

(pain=yes) & IRS

= 0.52 x 0.48 x 25 = 6.24

Expected, E:

		IRS	Placebo	Total
Improvement in pain	Yes	6.24	6.76	13
	No	5.76	6.24	12
	Total	12	13	25

$(O - E)^2/E$:

		IRS	Placebo
Improvement in pain	Yes	1.22	1.13
	No	1.32	1.22

Chi-squared statistic, X^2 is sum of $(O-E)^2/E$ = 4.89

on (no. rows-1) x (no. columns-1) *d.f.*

= (2-1) x (2-1) = 1 *d.f.*

P=0.03 from tables

P=0.05 from Fisher's exact test

Measures of treatment effect

p_1=0.75, p_2=0.31

- Difference in proportion, p_1-p_2 = 0.44
- Relative risk, p_1/p_2 = 2.4
- Odds ratio, $[p_1/(1-p_1)] / [p_2/(1-p_2)]$ = 6.7
- Number needed to treat (NNT) = $1/(p_1-p_2)$ = 2.3
  - "Need to treat 2.3 patients to prevent 1 pain"

Odds ratio

- Odds ratio tends to be larger than relative risk unless p_1 and p_2 are small (<0.1)
- Can be estimated using logistic regression
- Logistic regression can be extended to incorporate baseline factors such as age, sex etc
- *Adjusted* treatment effect odds ratio can be estimated from this model

Summary

Aim	Test
Compare mean with hypothesized value	One sample t-test
Compare means in two independent groups	Two sample t-test
Compare means in >2 independent groups	ANOVA
Compare means before/after in same group	Paired t-test (=one sample t-test on differences)
Compare proportions in 2 or more independent groups	Chi-squared test