Introduction to Statistics for Clinical Trials

Comparisons of means

Martin Bland

Professor of Health Statistics

University of York

http://www-users.york.ac.uk/~mb55/

Methods for comparison of means

- Large sample Normal method (z method)
  (any distribution)
- Two sample t method (unpaired t, two group t method)
  (Normal distribution, uniform variance)
- Paired t method
  (Normal distribution for differences,
  differences independent of magnitude)

Large sample Normal method (z method)

This method can be used to compare two means for any large samples. Data may be from any distribution.

Example: interventions for depression delivered using the internet.

Recruited 525 people with symptoms of depression identified in a survey.

Randomly allocated to websites:

BluePages, information about depression (n = 166),

MoodGYM, cognitive behaviour therapy (n = 182),

Control, attention placebo (n = 178).

Christensen H, Griffiths KM, Jorm AF. (2004) Delivering interventions for depression by using the internet: randomised controlled trial. *British Medical Journal* **328**, 265-268.

Two means, large sample method

	number	Fall in scores mean	SD	SE of mean
MoodGYM	182	4.2	9.1	0.67
Controls	178	1.0	8.4	0.63

Confidence interval for difference between two means, MoodGYM minus Control.

Difference = 3.2

Standard error for the difference = $\sqrt{(0.67^2 + 0.63^2)} = 0.92$

N.B. This only works when the groups are independent.

Two means, large sample method

	number	Fall in scores mean	SD	SE of mean
MoodGYM	182	4.2	9.1	0.67
Controls	178	1.0	8.4	0.63

Confidence interval for difference between two means, MoodGYM minus Control.

Difference = 3.2

Standard error for the difference = $\sqrt{(0.67^2 + 0.63^2)} = 0.92$

Because sample is large:

- ❖ The means follow a Normal distribution, so the difference will as well.
- ❖ The standard error provides a good estimate of the standard deviation of this Normal distribution.

Two means, large sample method

	number	Fall in scores mean	SD	SE of mean
MoodGYM	182	4.2	9.1	0.67
Controls	178	1.0	8.4	0.63

Confidence interval for difference between two means, MoodGYM minus Control.

Difference = 3.2

Standard error for the difference = $\sqrt{(0.67^2 + 0.63^2)} = 0.92$

95% CI = $3.2 - 1.96 \times 0.92$ to $3.2 + 1.96 \times 0.92$
= 1. 40 to 5.00

Two means, large sample method

	Fall in scores			
	number	mean	SD	SE of mean
MoodGYM	182	4.2	9.1	0.67
Controls	178	1.0	8.4	0.63

Test of null hypothesis means are equal, MoodGYM versus Control.

Difference = 3.2, standard error for the difference = 0.92.

If null hypothesis true, difference/standard error will be from a Standard Normal distribution.

difference/standard error = 3.2/0.92 = 3.48.

From Normal distribution P = 0.0005.

Large sample Normal or z test for two means.

Large sample Normal method (z method)

Assumptions:

- The observations and groups are independent. This means that we should not have, for example, a group of 100 observations where there are 10 subjects with 10 observations on each. We should not have links between observations in the two groups, such as a matched study where each subject in one group is matched, e.g. by age and sex, with a subject in the other group.
- The samples are large enough for the standard errors to be well estimated. My rule of thumb is at least 50 in each group.

Two sample t method

This is also called the unpaired t method or test and the two group t method, Student's two sample t test.

Example: Blood glucose (mmol/L) measured in a comparison of two infant feed formulae for small neonates

```
      Formula 1                    Formula 2
1.2  3.5  3.9  4.2  4.8       2.8  3.3  4.0  4.6
2.8  3.7  3.9  4.3  5.7       2.9  3.4  4.0  4.7
3.1  3.8  4.0  4.4  5.9       3.1  3.6  4.2  5.0
3.1  3.9  4.0  4.4            3.1  3.7  4.3  5.1
3.2  3.9  4.2  4.7            3.1  3.9  4.5  6.8

n = 23                        n = 20
mean = 3.94                   mean = 4.01
SD = 0.95                     SD = 0.96
```

Two sample t method

Example: Blood glucose (mmol/L) measured in a comparison of two infant feed formulae for small neonates

```
      Formula 1                    Formula 2
1.2  3.5  3.9  4.2  4.8      2.8  3.3  4.0  4.6
2.8  3.7  3.9  4.3  5.7      2.9  3.4  4.0  4.7
3.1  3.8  4.0  4.4  5.9      3.1  3.6  4.2  5.0
3.1  3.9  4.0  4.4           3.1  3.7  4.3  5.1
3.2  3.9  4.2  4.7           3.1  3.9  4.5  6.8

n = 23                       n = 20
mean = 3.94                  mean = 4.01
SD = 0.95                    SD = 0.96
```

We cannot use the large sample Normal method because the samples are too small.

The standard error will not be sufficiently well estimated.

Two sample t method

We cannot use the large sample Normal method because the samples are too small.

The standard error will not be sufficiently well estimated.

The distribution of the standard error estimate depends on the distribution of the observations themselves.

We must make two assumptions about the data:

1. the observations come from Normal distributions,
2. the distributions in the two populations have the same variance. (N.B. The populations, not the samples from them, have the same variance.)

Two sample t method

If the distributions in the two populations have the same variance, we need only one estimate of variance. We call this the common or pooled variance estimate.

The degrees of freedom are number of observations minus 2.

We use this common estimate of variance to estimate the standard error of the difference between the means.

Blood glucose example:

Common variance = 0.89, SD = 0.95,
df = 23 + 20 – 2 = 41.

SE of difference = 0.29.

Difference (Formula 1 – Formula 2) = 4.01 – 3.94 = 0.07.

Two sample t method

Blood glucose example:

Common variance = 0.89, SD = 0.95, df = 41.

SE of difference = 0.29.

Difference (Formula 1 – Formula 2) = 4.01 – 3.94 = 0.07.

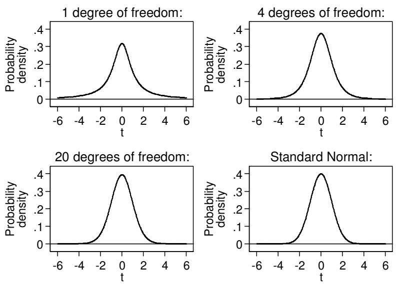
95% confidence interval for difference:

0.07 – ? × 0.29 to 0.07 + ? × 0.29

? comes not from the Normal distribution but the t distribution with 41 degrees of freedom.

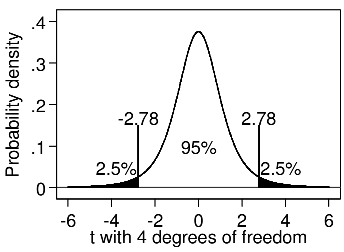
Two sample t method

The t distribution family has one parameter, the degrees of freedom.



As the degrees of freedom increases, the t distribution becomes closer to the Standard Normal distribution.

Two tailed probability point of the t Distribution



Two tailed probability points of the t Distribution

D.f.	Probability				D.f.	Probability			
	0.10	0.05	0.01	0.001		0.10	0.05	0.01	0.001
	(10%)	(5%)	(1%)	(0.1%)		(10%)	(5%)	(1%)	(0.1%)
1	6.31	12.70	63.66	636.62	16	1.75	2.12	2.92	4.02
2	2.92	4.30	9.93	31.60	17	1.74	2.11	2.90	3.97
3	2.35	3.18	5.84	12.92	18	1.73	2.10	2.88	3.92
4	2.13	2.78	4.60	8.61	19	1.73	2.09	2.86	3.88
5	2.02	2.57	4.03	6.87	20	1.73	2.09	2.85	3.85
6	1.94	2.45	3.71	5.96	21	1.72	2.08	2.83	3.82
7	1.90	2.36	3.50	5.41	22	1.72	2.07	2.82	3.79
8	1.86	2.31	3.36	5.04	23	1.71	2.07	2.81	3.77
9	1.83	2.26	3.25	4.78	24	1.71	2.06	2.80	3.75
10	1.81	2.23	3.17	4.59	25	1.71	2.06	2.79	3.73
11	1.80	2.20	3.11	4.44	30	1.70	2.04	2.75	3.65
12	1.78	2.18	3.06	4.32	40	1.68	2.02	2.70	3.55
13	1.77	2.16	3.01	4.22	60	1.67	2.00	2.66	3.46
14	1.76	2.15	2.98	4.14	120	1.66	1.98	2.62	3.37
15	1.75	2.13	2.95	4.07	∞	1.65	1.96	2.58	3.29

D.f. = Degrees of freedom
∞ = infinity, same as the Standard Normal Distribution

Two sample t method

Blood glucose example:

Common variance = 0.89, SD = 0.95, df = 41.

SE of difference = 0.29.

Difference (Formula 1 – Formula 2) = 4.01 – 3.94 = 0.07.

95% confidence interval for difference:

0.07 – ? × 0.29 to 0.07 + ? × 0.29

? comes not from the Normal distribution but the t distribution with 41 degrees of freedom.

Two tailed probability points of the t Distribution

D.f.	Probability				D.f.	Probability			
	0.10	0.05	0.01	0.001		0.10	0.05	0.01	0.001
	(10%)	(5%)	(1%)	(0.1%)		(10%)	(5%)	(1%)	(0.1%)
1	6.31	12.70	63.66	636.62	16	1.75	2.12	2.92	4.02
2	2.92	4.30	9.93	31.60	17	1.74	2.11	2.90	3.97
3	2.35	3.18	5.84	12.92	18	1.73	2.10	2.88	3.92
4	2.13	2.78	4.60	8.61	19	1.73	2.09	2.86	3.88
5	2.02	2.57	4.03	6.87	20	1.73	2.09	2.85	3.85
6	1.94	2.45	3.71	5.96	21	1.72	2.08	2.83	3.82
7	1.90	2.36	3.50	5.41	22	1.72	2.07	2.82	3.79
8	1.86	2.31	3.36	5.04	23	1.71	2.07	2.81	3.77
9	1.83	2.26	3.25	4.78	24	1.71	2.06	2.80	3.75
10	1.81	2.23	3.17	4.59	25	1.71	2.06	2.79	3.73
11	1.80	2.20	3.11	4.44	30	1.70	2.04	2.75	3.65
12	1.78	2.18	3.06	4.32	40	1.68	2.02	2.70	3.55
13	1.77	2.16	3.01	4.22	60	1.67	2.00	2.66	3.46
14	1.76	2.15	2.98	4.14	120	1.66	1.98	2.62	3.37
15	1.75	2.13	2.95	4.07	∞	1.65	1.96	2.58	3.29

D.f. = Degrees of freedom
∞ = infinity, same as the Standard Normal Distribution

Two sample t method

Blood glucose example:

Common variance = 0.89, SD = 0.95, df = 41.

SE of difference = 0.29.

Difference (Formula 1 – Formula 2) = 4.01 – 3.94 = 0.07.

95% confidence interval for difference:

0.07 – ? × 0.29 to 0.07 + ? × 0.29

? comes not from the Normal distribution but the t distribution with 41 degrees of freedom. t = 2.02.

0.07 – 2.02 × 0.29 to 0.07 + 2.02 × 0.29

= –0.52 to +0.66 mmol/L.

Two sample t method

Blood glucose example:

Common variance = 0.89, SD = 0.95, df = 41.

SE of difference = 0.29.

Difference (Formula 1 – Formula 2) = 4.01 – 3.94 = 0.07.

Test of significance, null hypothesis that in the population the difference between means = 0:

(difference – 0)/SE = 0.07/0.29 = 0.23.

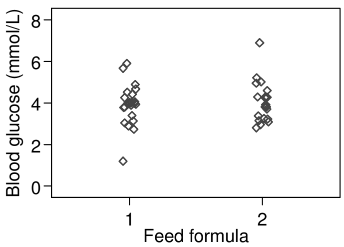
If the null hypothesis were true, this would be an observation from the t distribution with 41 degrees of freedom.

The probability of such an extreme value is greater than 0.1. More accurately, P = 0.8229 = 0.8.

Two sample t method

Assumptions of two sample t method

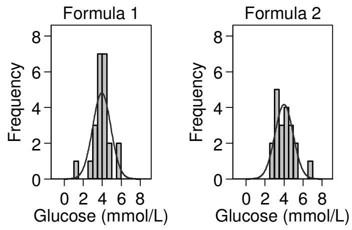
1. Distribution of blood glucose expenditure follows a Normal distribution in each population.
2. Variances are the same in each population.



Two sample t method

Assumptions of two sample t method

1. Distribution of blood glucose expenditure follows a Normal distribution in each population.
2. 2. Variances are the same in each population.



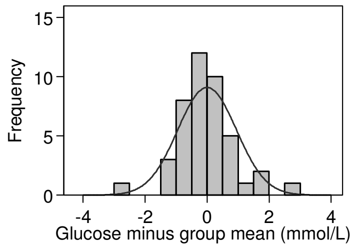
Not enough observa-tions to judge.

Two sample t method

Assumptions of two sample t method

Could combine the two graphs by subtracting the group mean from each observation to give **residuals**.

Residuals have mean = 0.



Looks a bit skew, but few observa-tions.

A better way: Normal plot.

Normal plot

Order the observations:

2.8 2.9 3.1 3.1 3.1 3.3 3.4 3.6 3.7 3.9
4.0 4.0 4.2 4.3 4.5 4.6 4.7 5.0 5.1 6.8

What would we expect the first observation of a Normal sample with 20 observations to be?

-1.67

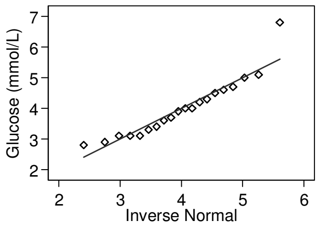
We can do this for what we expect the 2nd, 3rd, 4th, etc., to be. We then multiply by the standard deviation of the glucose and add the mean.

$-1.67 \times 0.96 + 4.01 = 2.41$

We do this for all 20 values and get what we would expect to get if glucose followed a Normal distribution.

Normal plot

If we plot blood glucose against the expected Normal value, we should get a straight line if the blood glucose follows a Normal distribution.

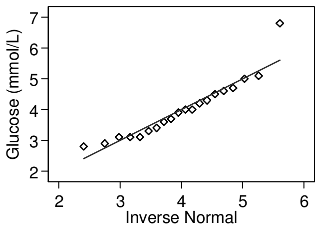


Distributions close to the Normal produce a straight line.

Skew distributions produce a clear bend or curve.

Normal plot

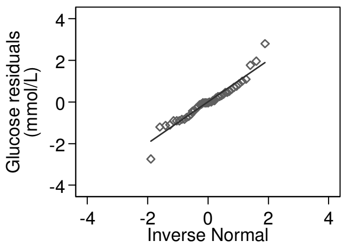
If we plot blood glucose against the expected Normal value, we should get a straight line if the blood glucose follows a Normal distribution.



This looks fairly straight except for one outlier.

Normal plot

We can do this for all the data, using the residuals.



This looks fairly straight except for a couple of outliers.

Normal plot

There are several ways of drawing Normal plots.

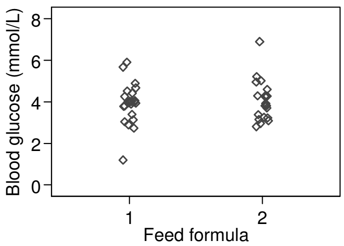
Some programs, such as SPSS, put the expected Normal values on the vertical axis and the observed data on the horizontal axis. A downward curve then indicates positive skewness, an upward curve negative skewness.

Some programs use the Standard Normal expected values rather than those for a Normal distribution with the same mean and standard deviation as the data.

Some offer a Normal probability plot rather than a Normal quantile plot, but these look very similar and are interpreted in the same way.

Two sample t method

Standard deviations very similar, 0.95 and 0.96.



Data fit the assumptions approximately but not exactly.

Two sample t method

Effect of deviations from assumptions

Methods using the t distribution depend on some strong assumptions about the distributions from which the data come.

In general for two equal sized samples the t method is very resistant to deviations from Normality, though as the samples become less equal in size the approximation becomes less good.

The most likely effect of skewness is that we lose power.

P values are too large and confidence intervals too wide.

We can usually correct skewness by a transformation.

Two sample t method

Effect of deviations from assumptions

If we cannot assume uniform variance, the effect is usually small if the two populations are from a Normal Distribution.

Unequal variance is often associated with skewness in the data, in which case a transformation designed to correct one fault often tends to correct the other as well.

If distributions are Normal, can use the Satterthwaite correction to the degrees of freedom.

Two sample t method

Unequal variances: Satterthwaite correction to the degrees of freedom.

If variances are unequal, we cannot estimate a common variance.

Instead we use the large sample form of the standard error of the difference between means. We replace the t value for confidence intervals by t with fewer degrees of freedom.

Degrees of freedom depend on the relative sizes of the variances. The larger variance dominates and if one is much larger than the other the degrees of freedom for that group are the only degrees of freedom.

Two sample t method

Unequal variances: Satterthwaite correction to the degrees of freedom.

For the blood glucose example:

Degrees of freedom: 41 (= 23 + 20 – 2)

Satterthwaite's degrees of freedom: 40.1

We round this down to 40 to use the t table.

For this example, analyses are almost identical.

N.B. Satterthwaite’s method is an approximation for use in unusual circumstances. The equal variance method is the standard t test.

Paired t method

MAGS score before and after treatment with topical placental extract in 9 patients with non-healing wounds (Shukla *et al.*, 2004)

Subject	Before	After	Difference
1	20	32	12
2	31	47	16
3	34	43	9
4	39	43	4
5	43	55	12
6	45	52	7
7	49	61	12
8	51	55	4
9	63	71	8
Mean			9.33
Standard deviation			4.03
Standard error of mean			1.34

Paired t method

For a small sample, we must assume that the differences themselves follow a Normal distribution.

95% confidence interval:

mean difference – $t_{0.05}$ standard errors to
mean difference + $t_{0.05}$ standard errors.

where $t_{0.05}$ is the two-sided 5% point of the t distribution with degrees of freedom = number of observations minus one.

Test of significance: refer mean difference / standard error to the t distribution with degrees of freedom = number of observations minus one.

Paired t method

Example: Increase in MAGS score

Mean difference = 9.33, SE = 1.34 litres/min.

9 differences, hence 9 – 1 = 8 degrees of freedom.

Two tailed probability points of the t Distribution

D.f.	Probability 0.10 (10%)	0.05 (5%)	0.01 (1%)	0.001 (0.1%)	D.f.	Probability 0.10 (10%)	0.05 (5%)	0.01 (1%)	0.001 (0.1%)
1	6.31	12.70	63.66	636.62	16	1.75	2.12	2.92	4.02
2	2.92	4.30	9.93	31.60	17	1.74	2.11	2.90	3.97
3	2.35	3.18	5.84	12.92	18	1.73	2.10	2.88	3.92
4	2.13	2.78	4.60	8.61	19	1.73	2.09	2.86	3.88
5	2.02	2.57	4.03	6.87	20	1.73	2.09	2.85	3.85
6	1.94	2.45	3.71	5.96	21	1.72	2.08	2.83	3.82
7	1.90	2.36	3.50	5.41	22	1.72	2.07	2.82	3.79
8	1.86	2.31	3.36	5.04	23	1.71	2.07	2.81	3.77
9	1.83	2.26	3.25	4.78	24	1.71	2.06	2.80	3.75
10	1.81	2.23	3.17	4.59	25	1.71	2.06	2.79	3.73
11	1.80	2.20	3.11	4.44	30	1.70	2.04	2.75	3.65
12	1.78	2.18	3.06	4.32	40	1.68	2.02	2.70	3.55
13	1.77	2.16	3.01	4.22	60	1.67	2.00	2.66	3.46
14	1.76	2.15	2.98	4.14	120	1.66	1.98	2.62	3.37
15	1.75	2.13	2.95	4.07	∞	1.65	1.96	2.58	3.29

D.f. = Degrees of freedom

∞ = infinity, same as the Standard Normal Distribution

Paired t method

Example: Increase in MAGS score

Mean difference = 9.33, SE = 1.34 litres/min.

9 differences, hence 9 – 1 = 8 degrees of freedom.

Using the 8 d.f. row, we get $t_{0.05} = 2.31$.

The 95% confidence interval:

$9.33 - 2.31 \times 1.34$ to $9.33 + 2.31 \times 1.34$
= 6.2 to 12.4.

Test of significance:

Mean/SE = 9.33/1.34 = 6.96

From t table, P<0.001. From computer program, P=0.0001.

Paired t method

Assumptions:

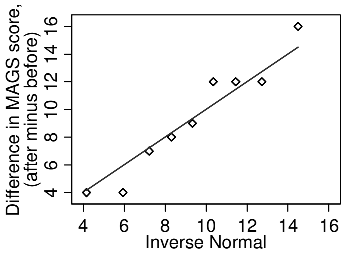
- The observations are independent.
- The differences follow a Normal distribution.
- The mean and standard deviation of differences are constant, i.e. not related to the size of the variable.

Paired t method

Assumptions of the paired t method

The differences follow a Normal distribution.

Check with histogram, Normal plot.



Paired t method

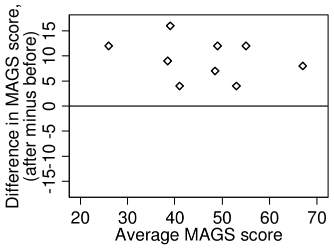
Assumptions of the paired t method

The differences follow a Normal distribution.

Check with histogram, Normal plot.

The mean and SD of the differences are constant, i.e. unrelated to magnitude.

Check with plot of difference against average.



Paired t method

Deviations from assumptions

1. The differences follow a Normal distribution.

   Need at least 100 observations to ignore non-Normal. However, differences tend to have a symmetrical distribution, so this assumption is usually met.

2. The mean and SD of the differences are constant, i.e. unrelated to magnitude.

   Essential, can be dealt with by transformation.

Paired t method

Deviations from assumptions

1. The differences follow a Normal distribution.

   Not as robust as the two sample t method. Need at least 100 observations to ignore non-Normal. However, differences tend to have a symmetrical distribution, so this assumption is usually met.

2. The mean and SD of the differences are constant, i.e. unrelated to magnitude.

   Essential, can be dealt with by transformation.