Comparing 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)
- One-way analysis of variance (anova)
  (For more than two means, not included in course)
  (Normal distribution, uniform variance)

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: study of respiratory symptoms in school children.

Do children reported by their parents to have respiratory symptoms had worse lung function than children who were not reported to have symptoms, and, if so, by how much?

Cough during the day or at night

	Number	Mean PEFR litre/min	SD litre/min
cough	92	294.8	57.1
no cough	1643	313.6	55.2

Large sample Normal method (z method)

	Number	Mean PEFR litre/min	SD litre/min	SE litre/min
cough	92	294.8	57.1	5.95
no cough	1643	313.6	55.2	1.36

Standard error of difference = $\sqrt{5.95^2 + 1.36^2} = 6.11$ litre/min.

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

Difference = 294.8 – 313.6 = –18.8 litre/min.

Because sample is large:

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

Large sample Normal method (z method)

Difference = –18.8 litre/min
Standard error of difference = 6.11 litre/min

Sample difference is from a Normal distribution with standard deviation = 6.11.

95% confidence interval for the difference:
–18.8 – 1.96 × 6.11 to –18.8 + 1.96 × 6.11
i.e. –30.8 to –6.8 litre/min.

Test of significance for the difference:
Null hypothesis value for the difference = 0
We are (–18.8 – 0)/6.11 = –3.1 standard deviations from the expected mean difference if null hypothesis true.
For a Normal distribution, P=0.002.

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:

24 hour energy expenditure (MJ) in groups of lean and obese women

Lean		Obese	
6.13	8.08	8.79	9.97
7.05	8.09	9.19	11.51
7.48	8.11	9.21	11.85
7.48	8.40	9.68	12.79
7.53	10.15	9.69	
7.58	10.88		
7.90			

	Lean	Obese
Number	13	9
Mean	8.066	10.298
SD	1.238	1.398

Two sample t method

24 hour energy expenditure (MJ) in groups of lean and obese women

Lean		Obese	
6.13	8.08	8.79	9.97
7.05	8.09	9.19	11.51
7.48	8.11	9.21	11.85
7.48	8.40	9.68	12.79
7.53	10.15	9.69	
7.58	10.88		
7.90			

	Lean	Obese
Number	13	9
Mean	8.066	10.298
SD	1.238	1.398

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.

Energy expenditure example:

Common variance = 1.701, SD = 1.304,
df = 13 + 9 – 2 = 20.

SE of difference = 0.566.

Difference (obese – lean) = 10.298 – 8.066 = 2.232.

Two sample t method

Energy expenditure example:

Common variance = 1.701, SD = 1.304, df = 20.

SE of difference = 0.566.

Difference (obese – lean) = 10.298 – 8.066 = 2.232.

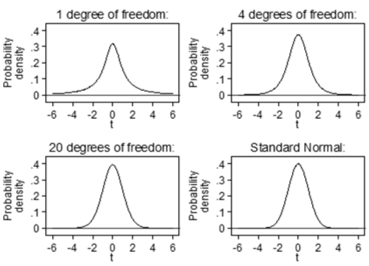
95% confidence interval for difference:

2.232 – ? × 0.566 to 2.232 + ? × 0.566

? comes not from the Normal distribution but the t distribution with 20 degrees of freedom.

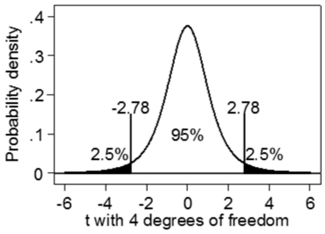
Two sample t method

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 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

Two sample t method

Energy expenditure example:

Common variance = 1.701, SD = 1.304, df = 20.

SE of difference = 0.566.

Difference (obese – lean) = 10.298 – 8.066 = 2.232.

95% confidence interval for difference:

2.232 – ? × 0.566 to 2.232 + ? × 0.566

? comes not from the Normal distribution but the t distribution with 20 degrees of freedom.

Two tailed probability points of the t Distribution

D.f.	0.10 (10%)	0.05 (5%)	0.01 (1%)	0.001 (0.1%)	D.f.	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

Two sample t method

Energy expenditure example:

Common variance = 1.701, SD = 1.304, df = 20.

SE of difference = 0.566.

Difference (obese – lean) = 10.298 – 8.066 = 2.232.

95% confidence interval for difference:

2.232 – ? × 0.566 to 2.232 + ? × 0.566

? comes not from the Normal distribution but the t distribution with 20 degrees of freedom.

2.232 – 2.09 × 0.566 to 2.232 + 2.09 × 0.566

= 1.05 to 3.42 MJ.

Two sample t method

Energy expenditure example:

Common variance = 1.701, SD = 1.304, df = 20.

SE of difference = 0.566.

Difference (obese – lean) = 10.298 – 8.066 = 2.232.

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

(difference – 0)/SE = 2.232/0.566 = 3.943.

If the null hypothesis were true, this would be an observation from the t distribution with 20 degrees of freedom.

Two sample t method

Energy expenditure example:

Common variance = 1.701, SD = 1.304, df = 20.

SE of difference = 0.566.

Difference (obese – lean) = 10.298 – 8.066 = 2.232.

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

(difference – 0)/SE = 2.232/0.566 = 3.943.

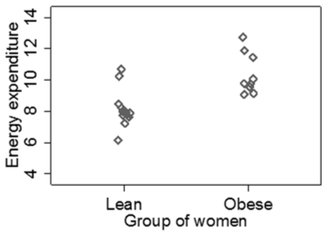
If the null hypothesis were true, this would be an observation from the t distribution with 20 degrees of freedom.

The probability of such an extreme value is less than 0.001. More accurately, P = 0.0008.

Two sample t method

Assumptions of two sample t method

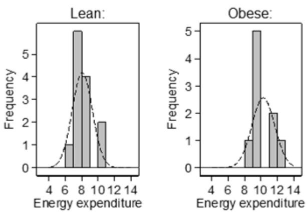
1. Distribution of energy expenditure follows a Normal distribution in each population.
2. Variances are the same in each population.



Two sample t method

Assumptions of two sample t method

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



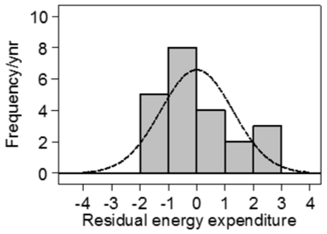
Not enough observa-tions to judge.

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.

8.79 9.19 9.21 9.68 9.69 9.97 11.51 11.85 12.79

The average values for the 1st, 2nd, 3rd, etc. observations of a Standard Normal sample with 9 observations are:

–1.28 –0.84 –0.52 –0.25 0.00 0.25 0.52 0.84 1.28

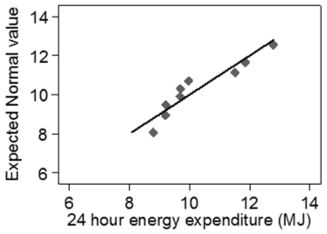
For a Normal sample with the same mean and standard deviation as energy expenditure, we multiply by standard deviation = 1.398 and add mean = 10.298:

8.51 9.12 9.57 9.95 10.30 10.65 11.02 11.47 12.09

Normal plot

E: 8.79 9.19 9.21 9.68 9.69 9.97 11.51 11.85 12.79
N: 8.51 9.12 9.57 9.95 10.30 10.65 11.02 11.47 12.09

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



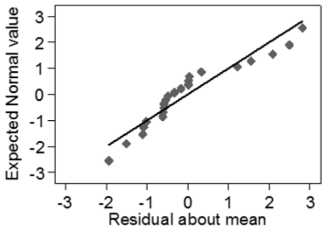
Normal plot

There are several variations on the Normal plot. Sometimes we use a Standard Normal distribution rather than one with mean and standard deviation equal to those of the variable.

Sometimes we plot the expected Normal value on the vertical axis and the observed value on the horizontal (SPSS does this).

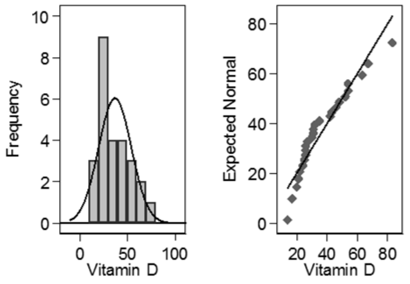
Normal plot

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



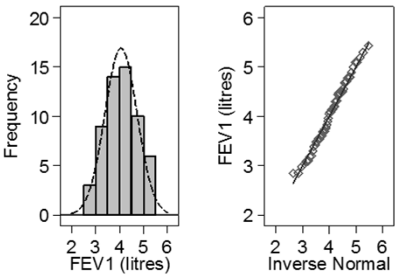
Normal plot

Skew distributions produce a clear bend or curve.



Normal plot

Distributions close to the Normal produce a straight line.



Two sample t method

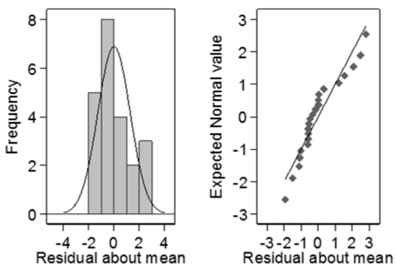
Example: energy expenditure

	Lean	Obese
Number	13	9
Mean	8.066	10.298
SD	1.238	1.398

Standard deviations very similar.

Two sample t method

Example: energy expenditure



Distribution rather skew.

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 energy expenditure example:

Degrees of freedom: 20 (= 13 + 9 – 2)

Satterthwaite's degrees of freedom: 15.9187

We round this down to 15 to use the t table.

Equal variances: 95% CI = 1.05 to 3.42 MJ, P=0.0008.
Unequal variances: 95% CI = 1.00 to 3.46 MJ. P=0.0014.

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

PEFR (litre/min) measured by Wright meter and mini meter, female subjects

Subject	Wright PEFR	Mini PEFR	Difference
1	490	525	-35
2	397	415	-18
3	512	508	4
4	401	444	-43
5	470	500	-30
6	415	460	-45
7	431	390	41
8	429	432	-3
9	420	420	0
10	275	227	48
11	165	268	-103
12	421	443	-22
Mean			-17.2
Standard deviation			40.3
Standard error of mean			11.6

Paired t method

For a large sample, mean difference will be an observation from a Normal distribution,
mean = population mean,
standard deviation = standard error of sample mean.

95% confidence interval:

mean difference – 1.96 standard errors to
mean difference + 1.96 standard errors.

Test of significance: refer mean difference / standard error to the standard Normal distribution.

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: PEFR meters

Mean difference = –17.1, SE = 11.6 litres/min.

12 differences, hence 12 – 1 = 11 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: PEFR meters

Mean difference = –17.1, SE = 11.6 litres/min.

12 differences, hence 12 – 1 = 11 degrees of freedom.

Using the 11 d.f. row, we get $t_{0.05}$ = 2.20.

The 95% confidence interval:

–17.2 – 2.20 × 11.6 to –17.2 + 2.20 × 11.6
= –42.7 to +8.3 litre/min.

Test of significance:

Mean/SE = –17.2 /11.6 = –1.48

From t table, P>0.10. From computer program, P=0.17.

Paired t method

Assumptions of the paired t method

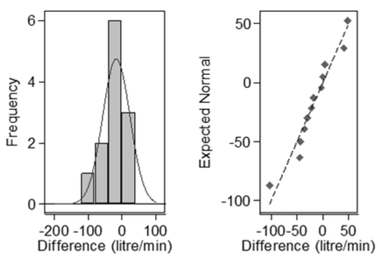
1. Observations are independent.
2. The differences follow a Normal distribution.
3. The mean and SD of the differences are constant, i.e. unrelated to magnitude.

Paired t method

Assumptions of the paired t method

2. The differences follow a Normal distribution.

Check with histogram, Normal plot.

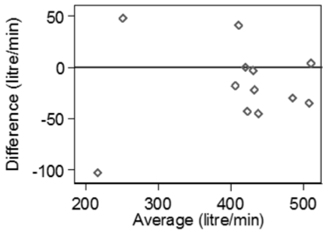


Paired t method

Assumptions of the paired t method

3. 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. Lack of independence.

   Test not valid. Needs advanced method.

2. 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.

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

   Essential, can be dealt with by transformation.

Deviations from assumptions of t methods

Data do not follow a Normal distribution or do not have uniform variance.

Transformations

We analyse a mathematical function of the data.

E.g. Logarithm, square root, reciprocal (one over).

Significance tests: works fine.

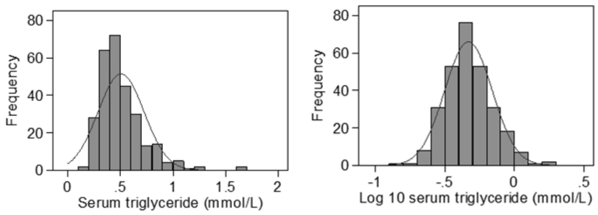
Confidence intervals: difficult to interpret.

Deviations from assumptions of t methods

Data do not follow a Normal distribution or do not have uniform variance.

Transformations

E.g. Serum triglyceride, logarithmic transformation:



Deviations from assumptions of t methods

No transformation possible

Non-parametric methods: do not require these assumptions.

E.g. for two groups: Mann Whitney U test, also known as the Wilcoxon two sample test.

Assume observations independent, as for t method.

Assume observations can be ordered.

Approximation if observations are tied, i.e. Two cannot be separated.

Deviations from assumptions of t methods

No transformation possible

Non-parametric methods: do not require these assumptions.

E.g. for pairs: sign test.

Assume pairs of observations independent, as for t method.

Deviations from assumptions of t methods

No transformation possible

Non-parametric methods: do not require these assumptions.

E.g. for pairs: Wilcoxon matched pairs or signed rank test.

Assume pairs of observations independent, as for t method.

Assume differences between pairs are meaningful, quantitative data.

Assume distribution of differences is symmetrical.

Approximation if observations are tied, i.e. differences for two pairs cannot be separated.