

AN INTRODUCTION TO STATISTICS

Article 7. An introduction to hypothesis testing. Parametric comparison of two groups—2

P Driscoll, F Lecky

Objectives

- Dealing with unpaired (independent) parametric data
 - Discuss common mistakes using the *t* test
- In covering these objectives the following terms will be introduced:
- Unpaired *z* test
 - Unpaired *t* test
 - One and two tailed tests

In the previous article we discussed the comparison of paired (dependent) data.¹ These result when there is a relation between the groups, for example investigating the before and after effects of a drug on the same group of patients. The key measurement here is the difference between each pair. If this comes from a population that is normally distributed the mean difference can be calculated along with the standard error of the mean (SEM). The 95% confidence intervals for the true mean difference can then be derived along with the *p* value for the null hypothesis (table 1).

When there is no relation between the groups, the data are called “unpaired” or “independent”. A common example of this is the controlled trial where the effect of an intervention on one group is compared with a control group without the intervention. Here the selection of the experimental group does not tell you which people will be in the control group. They are therefore independent of one another.

It is useful to note at this stage that when you compare groups you are taking into account two variations. One is due to the difference between subjects within the same group and is called the intra-group variation. The other results from the difference between the groups and so is known as the inter-group variation. With paired data the difference in subjects is removed because each subject acts as its own control. Consequently you are simply measuring the inter-group variation. In contrast, when using unpaired data, both these variations have an effect (table 1).

Table 1 Comparison of two groups using *z* and *t*-tests

Paired test	Unpaired test
Deals with the difference between the paired values	Deals with the difference between the means of both groups
Relies on the population of this difference being normally distributed	Relies on the population of this difference being normally distributed
It is not affected by the distribution of the before and after samples	Is affected by the inter and intra group distribution

Box 1 System for statistical comparison of two groups

- State the null hypothesis and the alternative hypothesis of the study
- Select the level of significance
- Establish the critical values
- Select the groups and calculate their means and standard errors of the mean (SEM)*
- Choose and calculate the test statistic
- Compare the calculated test statistic with the critical values
- Express the chances of obtaining results at least this extreme if the null hypothesis is true

* or estimated standard error of the mean (ESEM) if using a sample size < 100.²

Unpaired *z* test

We have shown previously that a systematic approach is used to determine if the null hypothesis is valid (box 1).¹

To see how this works when dealing with unpaired data consider the following example. Dr Egbert Everard has been working for just over a year in the emergency department of Deathstar General. During this time the department has dealt with 100 patients who have ingested a new rave drug “Hothead”. He suspects these people may have an abnormal sodium concentration on presentation—how can he investigate this?

1 STATE THE NULL HYPOTHESIS AND

ALTERNATIVE HYPOTHESIS OF THE STUDY

Having considered the problem, Egbert writes down the null hypothesis as:

“There is no significant difference between the sodium concentration in the patient’s ingesting “Hothead” and those of a similar age attending the emergency department”. In other words they are part of the same population.

Accident and
Emergency
Department, Hope
Hospital, Salford
M6 8HD, UK

Correspondence to:
Mr Driscoll, Consultant in
Accident and Emergency
Medicine (pdriscoll@
hope.srht.nwest.nhs.uk)

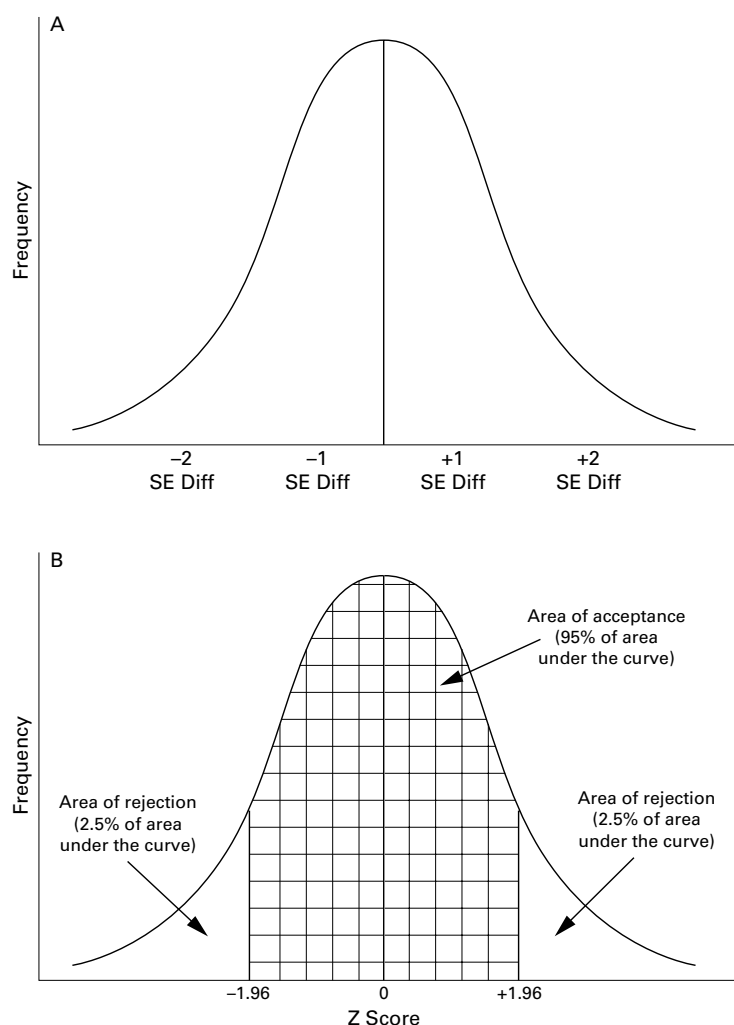


Figure 1 Normal distribution of the difference between the means of the two groups. (A) Shows the hypothetical population mean of this difference (central vertical line). The horizontal axis is divided up into standard errors of the difference of the mean (SE Diff). (B) Displays the same data as a standard normal distribution and replaces the SE Diff with z scores. The areas of acceptance (0.95 of the total area under the curve) and rejection (0.05 of the total area under the curve) of the null hypothesis are demonstrated. The area of rejection is split into two tails of the distribution curve.

This can be summarised to:

mean "Hothead" $[Na^+] =$ mean control $[Na^+]$

The alternative hypothesis is the logical opposite of this, that is:

"There is a significant difference between the sodium concentration in the patient's ingesting "Hothead" and those of a similar age attending the emergency department".

This can be summarised to:

mean "Hothead" $[Na^+] \neq$ mean control $[Na^+]$

2 SELECT THE LEVEL OF SIGNIFICANCE

If the null hypothesis is correct, and the two groups are part of the same population, then their mean sodium concentrations should be the same. Therefore the difference between them would be zero. However, there is bound to be some small variation simply due to chance. Therefore how big a difference are we going to allow before we reject the idea that the two groups are all part of the same population?

In answering this question we rely on an interesting mathematical fact that the difference of the means represents a population that has a normal distribution. In other words, if you repeated the experiment many times, and plotted all the mean differences, the scatter diagram would take the form of a normal distribution. If the null hypothesis was correct this distribution would have a mean of zero. The standard deviation of this distribution is known as the standard error of the differences between the means (SE Diff). This is equivalent to the standard error of the mean (SEM) that has been discussed in previous articles (fig 1A).¹⁻³

When comparing large groups of independent data we can determine the size of different areas under the distribution curve by using the z statistic:

$$z = [\mu_1 - \mu_2] / \text{SE Diff}$$

Where:

$\mu_1 =$ mean of group 1

$\mu_2 =$ mean of group 2

You will note that this equation is slightly different from the one used to compare paired data.¹ The numerator has changed to reflect the fact that we are interested in the difference between the means of the two groups rather than the mean difference between paired readings. Furthermore, the standard error of the difference between the means (SE Diff) has replaced the SEM to take account of the errors in estimating the means in each group:

$$\text{SE Diff} = \sqrt{[(s_1^2 / n_1) + (s_2^2 / n_2)]}$$

where:

$s_1 =$ estimation of the population's standard deviation derived from group 1 with n_1 subjects

$s_2 =$ estimation of the population's standard deviation derived from group 2 with n_2 subjects

As discussed in article 4, s is used because in clinical practice we usually do not know the value of a population's standard deviation (σ). However, provided the sample size is large enough (that is, greater than or equal to 100) the z statistic can still be derived using s as an estimation of the population's standard deviation.³

By convention, the outer 0.025 probabilities (that is, the tips of the two tails representing 2.5% of the area under the curve) are considered to be sufficiently away from the population mean as to represent values that cannot be simply attributed to chance variation (fig 1B). Consequently, if the sample mean is found to lie in either of these two tails then the null hypothesis is rejected. Conversely, if the sample mean lies within these two extremes then the null hypothesis will be accepted (fig 1B).

Following convention, Egbert picks a significance level of 0.05 for his study. He now needs to determine the sodium concentration that demarcates these two tails.

3 ESTABLISH THE CRITICAL VALUES

Using the z table Egbert finds that the critical value (z_{CRIT}) demarcating the middle 95% of the distribution curve is $z = \pm 1.96$ (fig 1B). In other words a z value of ± 1.96 separates

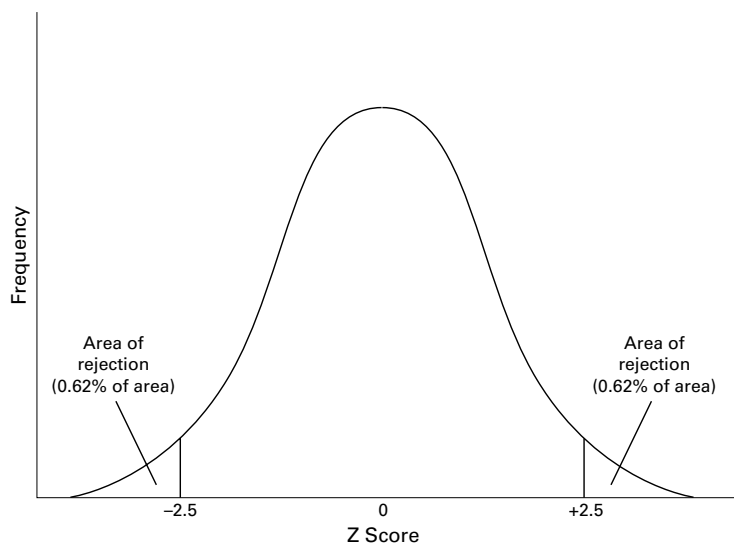


Figure 2 Standard normal distribution curve demonstrating two ways of getting a difference with a magnitude of 2.5. 0.0062 of the area under the curve lies between -2.5 to the tip of the left tail. A similar area lies between $+2.5$ to the tip of the right tail.

the middle 95% area of acceptance of the null hypothesis from two, 2.5% areas of rejection.

With the null and alternative hypotheses defined, and the critical values established (z_{CRIT}), the patients for the study can now be selected. The z statistic derived from the sample (z_{CALC}) can then be determined.

4 SELECT THE GROUPS AND CALCULATE THEIR MEANS

Egbert gathered the presenting sodium concentrations from a sample of 100 patients who have taken “Hothead”. The mean (μ_1) and estimation of the population’s standard deviation (s_1) were calculated and found to be 131 mmol/l and 10 mmol/l respectively.

Egbert had arranged with Ivor Whitecoat, senior laboratory technician at Deathstar General, to measure the sodium concentration in a control group of patients who had not taken “Hothead”. Ivor tells him that the mean sodium concentration in a 100 patients of a similar age presenting to the emergency department (μ_2) is 134 mmol/l with an s of 7 mmol/l (s_2).

5 CHOOSE AND CALCULATES THE TEST STATISTIC

As explained before, the z statistic is equal to:

$$z = [\mu_1 - \mu_2] / \text{SE Diff}$$

where:

$$\text{SE diff} = \sqrt{[(10^2 / 100) + (7^2 / 100)]} = \sqrt{[1 + 0.49]} = 1.2 \text{ (rounded down)}$$

An interesting feature can be noted from the equation above. Supposed Ivor Whitecoat used several thousand patients to work out a mean. In this case s_2^2 / n_2 would become so small as to be negligible. Consequently the SE Diff would simply be equal to the ESEM of Egbert’s group.

The difference between the two means in this study is:

$$[\mu_1 - \mu_2] = 131 - 134 = -3$$

Therefore:

$$z = -3 / 1.2 = -2.5$$

Key point 1

When comparing a sample with a large group you only need to know the ESEM of the sample.

6 COMPARE THE CALCULATED TEST STATISTIC WITH THE CRITICAL VALUES

The calculated value of -2.5 lies beyond the larger critical value of -1.96 . It therefore falls outside the area of accepting the null hypothesis.

7 EXPRESS THE CHANCES THAT THE NULL HYPOTHESIS IS IN KEEPING WITH THE DATA

The p value is the probability of getting a difference equal to or greater than that found in the experiment (that is, -3), if the null hypothesis was correct.⁴ As the z value can be negative or positive, there are two ways of getting a value with a magnitude of 2.5. Consequently the p value is represented by the area demarcated by -2.5 to the tip of the left tail plus the area demarcated by $+2.5$ to the tip of the right tail (fig 2).

From the tables of z statistics, it can be seen that the size of the tail from $+2.5$ to the right tip is $0.5 - 0.4938 = 0.0062$. The equivalent value in the other half of the distribution curve is the same. The p value is therefore doubled to give a total value of 0.0124. Consequently there is a 1.2% chance that a difference with a magnitude of 3, or larger, could have resulted if the null hypothesis was true.

WHAT IS THE ESTIMATED RANGE FOR THE TRUE DIFFERENCE BETWEEN THE MEANS?

In view of the importance of confidence intervals, Egbert also wants to determine the 95% CI for the difference. From the previous explanation, Egbert knows that 95% of all possible values of the difference between the means will lie within a range 1.96 SE Diff below his experimental difference to 1.96 SE Diff above it:

$$95\% \text{ confidence intervals} = \text{mean} \pm [1.96 \times \text{SEDiff}]$$

As described above, the difference between the means is $131 - 134 = -3$ mmol/l. Therefore the 95% confidence intervals of the difference between the two groups are:

$$1.96 \times 1.2 \text{ below } -3 \text{ mmol/l} = -5.4 \text{ (rounded up) mmol/l}$$

to

$$1.96 \times 1.2 \text{ above } -3 \text{ mmol/l} = -0.6 \text{ (rounded down) mmol/l}$$

As this range does not include zero, Egbert concludes that the data are not compatible with the null hypothesis being correct. The range is on the side of there being a lower sodium concentration in the patients taking “Hothead” but the differences are small. Therefore, rather than simply presenting a p value, it would be better if Egbert uses the 95% confidence intervals in discussing the clinical relevance of these data.

In summary, Egbert concludes that he can confidently reject the null hypothesis.

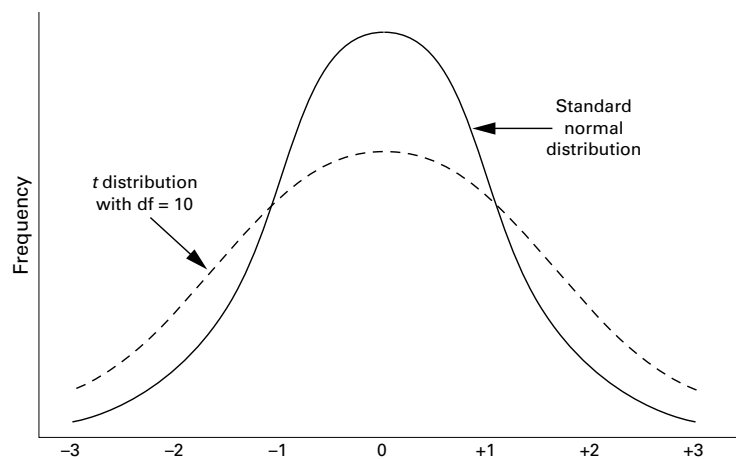


Figure 3 t distribution with 10 degrees of freedom and a standard normal distribution. As the degrees of freedom increase the t distribution becomes more like a normal distribution.

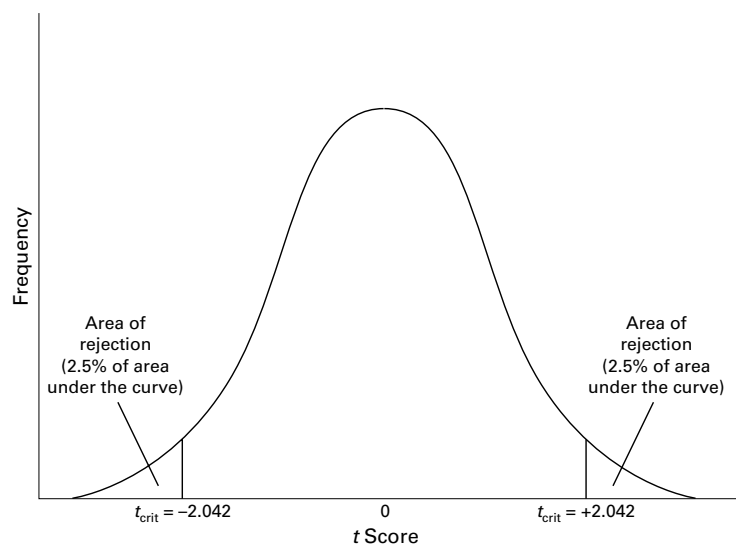


Figure 4 t distribution curve with 30 degrees of freedom demonstrating two tails. These are demarcated by t_{crit} values of -2.042 and 2.042 . The area between these values represents those values that are in keeping with the null hypothesis and amount to 0.95 of the area under the curve.

Difference between the means of the two groups = -3 mmol/l (95% confidence intervals -5.4 to -0.6 mmol/l); $p = 0.012$.

Unpaired t test

When the sample size is less than 100, the effect of intra-group variation becomes greater. In these cases the normal distribution has to be replaced by the t distribution. Unlike the normal distribution, the shape of the t distribution is dependent upon the size of the group (fig 3). It is always symmetrical but with small sample sizes the curve is flatter and has longer “tails”. However, as the sample size increases, the curve becomes normally distributed.

Irrespective of which t distribution is chosen, the same principle applies regarding set areas under the curve representing particular probabilities. Consequently the p value for the null hypothesis is derived from the test statistic. However, tables of the t distribution are used rather than the z distribution ones. Likewise, the 95% confidence intervals for the true difference between the means are the experi-

mental mean difference, \pm the appropriate t value, multiplied by the SE Diff.

To see how this works, consider the following example. Egbert by chance has a night off and discusses his findings over a romantic meal with Endora Lonely, an emergency physician at St Heartsinc. She is surprised by the topic of conversation but does wonder if the same applies to patients having taken an analogue of “Hothead” called “Brainboil”. She tackles this using the previously described systematic approach but this time considers using an unpaired t test as her study and control groups have only 16 patients in each.

1 STATE THE NULL HYPOTHESIS AND THE ALTERNATIVE HYPOTHESIS OF THE STUDY

These remain unchanged from those used in the larger study. Consequently:

“There is no significant difference between the sodium concentration in the patient’s ingesting “Brainboil” and those of a similar age attending the emergency department”. In other words they are part of the same population.

This can be summarised to:

mean “Brainboil” $[Na^+] =$ mean control $[Na^+]$

The alternative hypothesis is the logical opposite of this, that is:

“There is a significant difference between the sodium concentration in the patient’s ingesting “Brainboil” and those of a similar age attending the emergency department”.

This can be summarised to:

mean “Brainboil” $[Na^+] \neq$ mean control $[Na^+]$

2 SELECT THE LEVEL OF SIGNIFICANCE

Following convention, Endora picks a significance level of 0.05 for her study.

3 ESTABLISH THE CRITICAL VALUES

This is carried out in the same way as comparing larger groups but this time the t statistic is used because the group size is less than 100. The t table enables you to calculate the probability of lying within the middle 95% of the distribution where the null hypothesis is valid. The tables also take into account the changes in t with variation in sample size.

For mathematical reasons rather than use the absolute number in the group you use the degrees of freedom. In this type of comparison it is equal to the sum of the number in each group minus 1 (that is, $(n_1 - 1) + (n_2 - 1)$).

Consequently the degrees of freedom is $[16 - 1] + [16 - 1] = 30$.

Using the t distribution tables, Endora finds that the t value for a probability of 0.05 with 30 degrees of freedom is 2.042. This is known as the t_{crit} value. It means that for this study, a t value of ± 2.042 divides the middle 95% of the population accepting the null hypothesis from the two, 2.5% tails of rejection (fig 4).

The t statistic from the study (t_{calc}) can now be determined.

4 SELECT THE GROUPS AND CALCULATE THEIR MEANS

After carrying out her study with “Brainboil” she finds the mean presenting sodium concentrations to be 130 mmol/l ($s=10$ mmol/l). Dr Rooba Tube tells her the mean sodium concentration in a control group of patients at St Heartsinc is 135 mmol/l ($s=10$).

5 CHOOSE AND CALCULATE THE TEST STATISTIC

Unpaired *t* test

This test requires the data to have two properties:

(1) The two groups need to come from populations with normal distributions.

It is sometimes possible to determine if the distribution is skewed simply by checking the raw data or looking at a distribution curve. A more sophisticated way is to use a computer to develop a “normal plot” of the data. This programme manipulates normally distributed data so that they form a straight line. The closer the data comply with this line, the closer they are to being normally distributed. Altman gives a good description for those who are interested to getting further information about this type of manipulation.⁵ It is possible to also prove the same thing mathematically using various probability tests. However, these are of little additional benefit when populations are less than 30 and the normal probability plot gives a reasonably straight line.

(2) The groups should be from populations with the same standard deviation.

It is not possible to know the standard deviations of the populations but an estimation comes from using the *s* value calculated from each group. The closer these values are, the more likely they have similar standard deviations. As with distributions, it is possible to formally assess the difference in group variance by carrying out an *F* test (see appendix).

Endora checks her datasets for both groups and is satisfied that these pre-conditions are met. She therefore proceeds with the analysis.

Key points 2 Requirements for unpaired *t* tests

- The groups need to have normal distributions
- The groups need to have similar standard deviations

Calculate the *t* statistic

The *t* statistic for comparing two unpaired groups is:

$$t = [\mu_1 - \mu_2] / \text{SE Diff}$$

where:

μ_1 = the mean for group 1

μ_2 = the mean for group 2

When using the *t* test for comparing unpaired means the SE diff is derived from the following formula:

$$\text{SE diff} = \sqrt{[(s^2 / n_1) + (s^2 / n_2)]}$$

Where:

n_1 and n_2 are the number of subjects in the two groups.

Table 2 Extract of the table of the *t*-statistic values. The first column lists the degrees of freedom (*df*). The headings of the other columns give probabilities for *t* to lie within the two tails of the distribution.

<i>df</i>	0.1	0.05	0.01	0.001
1	6.314	12.706	63.657	636.619
5	2.015	2.571	4.032	6.869
10	1.812	2.228	3.169	4.587
15	1.753	2.131	2.947	4.073
20	1.725	2.086	2.845	3.850
30	1.697	2.042	2.750	3.646
43	1.681	2.017	2.695	3.532

s represents the pooled variance of the two groups:

$$s^2 = [(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2] / [n_1 + n_2 - 2]$$

From the data Endora calculates:

$$s^2 = [(16 - 1)100 + (16 - 1)100] / 30 = [3000] / 30 = 100$$

Therefore:

$$\text{SE diff} = \sqrt{[(100/16) + (100/16)]} = \sqrt{12.5} = 3.5$$

Therefore:

$$t = [130 - 135] / 3.5 = -1.43$$

6 COMPARE THE CALCULATED TEST STATISTIC WITH THE CRITICAL VALUES

The calculated value of -1.43 lies inside the area of accepting the null hypothesis (that is, a *t* value of ± 2.042).

7 EXPRESS THE CHANCES THAT THE NULL HYPOTHESIS IS IN KEEPING WITH THE DATA

As the *t* value can be negative or positive, there are two ways of getting a difference with a magnitude of 1.43. Consequently the *p* value is represented by the area demarcated by -1.43 to the tip of the left tail plus the area demarcated by $+1.43$ to the tip of the right tail.

To determine this *p* value, Endora needs to consult the *t* distribution table using the appropriate degrees of freedom (table 2). In this case the degrees of freedom equals two less than the total number of subjects in both groups (that is, 30). The table indicates that the size of these two tails is greater than 0.1. Consequently there is over a 10% chance that a difference of 5 mmol/l, or larger could have resulted if the null hypothesis was true.

The null hypothesis is therefore accepted and Endora can conclude that the sodium concentrations in patients taking “Brainboil” are not significantly different from the general population.

What is the estimated range for the true difference between the means?

Similar to Egbert, Endora would also like to know the 95% confidence intervals for the difference. The SE Diff is used in the usual way to determine this:

$$95\% \text{ CI} = \text{difference between the means} \pm [t_0 \times \text{SE Diff}]$$

Where t_0 is the *t* statistic appropriate for the required CI for the true difference between the means.

To find the *t* value representing the middle 95% of the distribution curve, Endora needs to look down the column representing the outer 5% (0.05 probability) of the curve because this identifies the extreme point of the middle

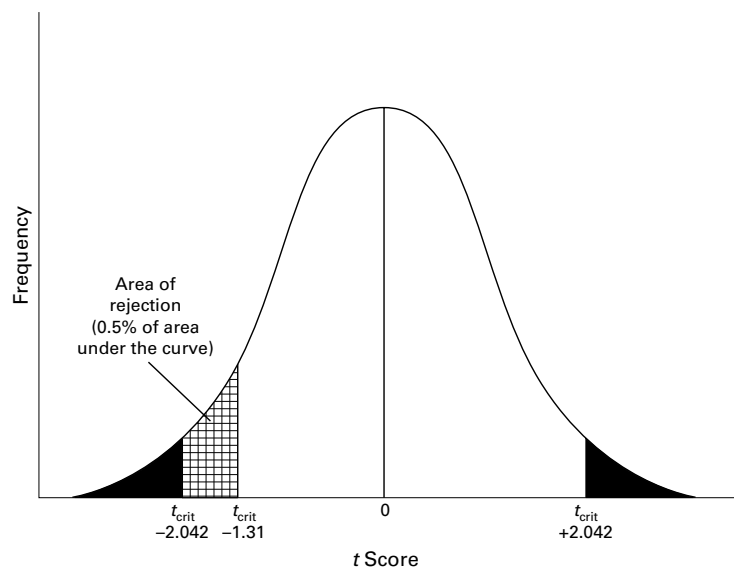


Figure 5 t distribution curve with 30 degrees of freedom demonstrating one and two areas of rejection. A t statistic to be -1.43 lies in an area of acceptance of the null hypothesis using the two tailed test [$t_{crit} = -2.042$] or rejection using the one tailed test [$t_{crit} = -1.31$].

section. The t value in the column representing $p = 0.05$ at 30 degree of freedom is 2.042. This is the number of SE Diff above and below the mean that cover the middle 95% of the distribution curve.

Therefore:

$95\% \text{ CI} = -5 \pm [2.042 \times 3.5] = -12 \text{ to } 2 \text{ mmol/l}$ (to the nearest whole numbers)

As this includes zero, the null hypothesis is valid at the 5% level. The confidence intervals are wide indicating that the study lacks precision, possibly due to the small sample size.¹

Common mistakes made in using the z and t test

ONE AND TWO TAILED TEST

When using the null hypothesis in comparing two groups, you are determining what is the probability that they are from the same population. p Values of less than 0.05 are usually taken as the point where the null hypothesis can be rejected. However, this result does not tell you how they are different. Indeed when you set up your initial study you commonly do not know which group will produce the biggest result.

To demonstrate this consider a trial comparing the analgesic effect of a new non-steroidal anti-inflammatory drug (NSAID) with ibuprofen. Before carrying out this trial you may hope that the new drug is better than the old one but you cannot be sure—it could be worse. Your study, and statistical analysis, should therefore be able to detect three possibilities:

- The new NSAID is no different than ibuprofen (that is, supporting the null hypothesis)
- The new NSAID is a better than ibuprofen (that is, rejects the null hypothesis)
- The new NSAID is worse than ibuprofen (that is, rejects the null hypothesis)

The latter two displacements are referred to as the “tails” or “sides”. Consequently the tests measuring the probability of the null hypothe-

sis being valid in one or both situations are called “one” and “two” tailed (sided) tests respectively.

When using a “one tailed test” there is only one area of rejection on the random sampling distribution of the means. Consequently the area of rejection is concentrated on one of the tails. This reduces the critical value for t and so makes the p value smaller and more impressive for any given difference between means.

For example, if Endora used a one tail test for a t statistic of -1.43 she would find it to lie in an area of rejection of the null hypothesis ($p < 0.05$, shaded area). This is because the t_{crit} for this degree of freedom is -1.31 . In contrast, as we have seen, using a two tailed test leads to acceptance of the null hypothesis ($p > 0.05$, checkered area) (fig 5).

You may therefore be tempted to use a one tailed test in an analysis. Beware though; it is rare that the direction of displacement can be predicted before the study. Furthermore, in comparison to the standard treatment a worse result by the new treatment is also clinically important. Consequently two tailed tests should be routinely used when comparing two groups. If a decision is made to carry out a one tailed test then the rationale must be clearly described. An example of such a case is in public health work when you want to determine if a product does not fall below a particular standard (for example, water purity). Here you are not concerned with how pure the water is, just as long as it is better than a preset level.

Key points 3

- The decision to use a one tailed test must depend upon the nature of the hypothesis being tested and should therefore be decided upon before the data are collected.
- As a rule of thumb, when comparing two groups always use a two tailed test for the null hypothesis.
- The concept of one tailed and two tailed tests does not apply when more than two groups are being compared.

USING A COMPUTER WITHOUT CONSIDERING THE DATA AND THE QUESTION BEING ASKED

Obviously using computer software can greatly help you in working out the long calculations described above. This leads to the temptation of doing the analysis on your own and not seeking statistical help when necessary. However, be aware in doing this that you make sure the appropriate test is carried out so that the correct calculations and distribution tables are used. Applying the wrong test will still produce an answer, but it will be meaningless!

ASSUMPTIONS OF NORMALITY AND SIMILAR VARIANCE

The t test is able to deal with all but major deviations from normality or uniform variance between the groups. The main problems occur when dealing with small data with a skewed distribution. In these cases the t statistic does

Table 3 Comparison of age and admission temperatures in two groups of patients selected for Steele *et al*'s study⁹

Characteristic	Blanket rewarming (7 patients)	Forced-air rewarming (9 patients)
Mean age (s)	54 yr. (14 yr.)	58 yr. (19 yr.)
Mean admission temperature (s)	29.8°C (1.5°C)	28.8°C (2.5°C)

not comply with the t distribution curve. A possible solution in these cases is to see if transforming the data can make them have a normal distribution and uniform variance.⁶ Data that still fail to achieve these prerequisites cannot be analysed using the t test. Instead a non-parametric method will have to be used.

MULTIPLE COMPARISONS

This article has concentrated on comparing two groups. In medical research however we may be faced with having to compare three or more groups. If this problem is tackled using two group comparisons for every possible pair of combinations then we run the risk of finding some "significant" differences simply by chance. This probability gets bigger as the number of pairs increase. Therefore for multiple comparisons, t tests should not be used. Instead a different test, known as an analysis of variance, is required.

Summary

When carrying out medical studies we often have to compare two unpaired (independent) groups. This is best carried out using a systematic approach in which the null hypothesis, the levels of significance and the critical levels are decided upon before the experiment is started. Once the data have been collected the appropriate statistic test is chosen and calculated. The likelihood of the null hypothesis being valid can then be determined along with the confidence intervals for the difference between the means of the two groups.

When comparing large samples (a 100 or greater), the z statistic can be used. However, with smaller groups the assumptions made in its calculation are no longer valid. In these circumstances the t statistic should be calculated, provided the data are normally distributed and the two groups have similar standard deviations.

Quiz

1 Why are the confidence intervals for unpaired comparisons usually greater than the paired variety?

2 What are the requirements of the data if an unpaired t test is to be carried out?

3 Allison *et al* carried out an experiment to compare the capillary leakage in trauma victims resuscitated with either hydroxyethyl starch ($n = 24$) or gelatine ($n = 21$).⁷ State the null hypothesis and alternative hypothesis of the study? Assuming a level of significance of 0.05, what would be the critical values (t_{crit}) of the study if an unpaired t test was carried out?

4 The following question is adapted from Boyd *et al* study comparing two activated charcoal preparations.⁸ They found the mean (s)

amount of charcoal drunk was 26.5 (13.3) g for Carbomix and 19.5 (13.7) g for Actidose-Aqua. The sample size for the Carbomix group was 47 and 50 for Actidose-Aqua. Assuming the prerequisites for carrying out an unpaired, two tailed t test are valid, calculate the t statistic.

5 One for you to try on your own. Steele *et al* carried out a study comparing two types of rewarming of hypothermic patients.⁹ Part of the data are adapted and shown in table 3.

Carry out an unpaired, two tailed t test comparing these datasets in the two groups. Is there a significant difference between the groups for the two variables?

Answers

1 When using paired data the confidence interval is smaller because you have removed the variability between the subjects and are solely comparing the inter-groups difference.

2 The groups need to have normal distributions and similar standard deviations.

3 The null hypothesis is:

"There is no significant difference in the capillary leakage of trauma victims resuscitated with hydroxyethyl starch or gelatine.

The alternative hypothesis is:

"There is a significant difference in the capillary leakage of trauma victims resuscitated with hydroxyethyl starch or gelatine.

To determine the critical value the degrees of freedom need to be calculated first. In this study this is equal to $[24 - 1] + [21 - 1] = 43$.

Using the t distribution tables, the t_{crit} value for a probability of 0.05 with 43 degrees of freedom is ± 2.017 if a two tailed test is being used.

4 The t statistic for comparing two unpaired groups is:

$$t = [\mu 1 - \mu 2] / SE \text{ Diff}$$

where:

$\mu 1$ = the mean for Carbomix

$\mu 2$ = the mean for Actidose-Aqua

$$SE \text{ diff} = \sqrt{[s^2 / n_1 + s^2 / n_2]}$$

and:

$$s^2 = [(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2] / [n_1 + n_2 - 2]$$

From the data of Boyd *et al*:

$$s^2 = [(46 \times 176.89) + (49 \times 187.69)] / 95 = 182.46$$

Therefore:

$$SE \text{ diff} = \sqrt{[182.46/47 + 182.46/50]} = 2.74$$

Therefore:

$$t = [26.5 - 19.5] / 2.74 = 2.6 \text{ (rounded up)}$$

The authors would like to thank Jim Wardrope and Iram Butt for their invaluable suggestions.

Appendix

The F test is the ratio of the larger variance/smaller variance. The resulting F value is then used to determine the probability that the two variances are from the same population (that is, that the null hypothesis is correct). This is done by reading the F tables using the F value and the appropriate degrees of freedom. The latter is equal to the sum of $(n_1 - 1)$ and $(n_2 - 1)$ where n_1 and n_2 are the number in both groups. If the null hypothesis is not consistent with the data (that is, the " F " statistic is significant), the t test should not be used.

1 Driscoll P, Lecky F. An introduction to hypothesis testing. Parametric comparison of two groups—1. *Emerg Med J* 2001;18:124–30.

- 2 Driscoll P, Lecky F, Crosby M. An introduction to estimation—1. Starting from z. *J Accid Emerg Med* 2000;17:409–15.
- 3 Driscoll P, Lecky F. An introduction to estimation—2. From z to t. *Emerg Med J* 2001;18:65–70.
- 4 Driscoll P, Lecky F, Crosby M. An introduction to statistical inference. *J Accid Emerg Med* 2000;17:357–63.
- 5 Altman D. Normal plot. In: *Practical statistics for medical research*. London: Chapman and Hall, 1997:133–43.
- 6 Driscoll P, Lecky F, Crosby M. *An introduction to everyday statistics—1*. *J Accid Emerg Med* 2000;17:205–11.
- 7 Allison K, Gosling P, Jones S, et al. Randomised trial of hydroxyethyl starch versus gelatine for trauma resuscitation. *J Trauma* 1999;47:1114–21.
- 8 Boyd R, Hanson J. Prospective single blinded randomised controlled trial of two orally administered activated charcoal preparations. *J Accid Emerg Med* 1999;16:24–5.
- 9 Steele M, Nelson M, Sessier D, et al. Forced air speeds rewarming in accidental hypothermia. *J Trauma* 1996;27:479–84.

Further reading

- Altman D. Theoretical distributions. In: *Practical statistics for medical research*. London: Chapman Hall, 1991:48–73.
- Bland M. *An introduction to medical statistics*. Oxford: Oxford University Press, 1987.
- Gaddis G, Gaddis M. Introduction to biostatistics: Part 4, statistical inference techniques in hypothesis testing. *Ann Emerg Med* 1990;19:820–5.
- Gardner M, Altman D. Calculating confidence intervals for means and their differences. In: *Statistics with confidence*. London: BMJ Publications, 1989:20–7.
- Glaser A. Hypothesis testing. In: *High yield biostatistics*. Baltimore: Williams and Wilkins, 1995:31–46.
- Koosis D. Difference between means. In: *Statistics—a self teaching guide*. 4th ed. New York: Wiley, 1997:127–52.
- Swincow T. The t test. In: *Statistics from square one*. London: BMJ Publications, 1983:33–42.