THE *t* TEST

An introduction

ANTONY UGONI B.Sc. (Hons).*

BRUCE F. WALKER D.C. †

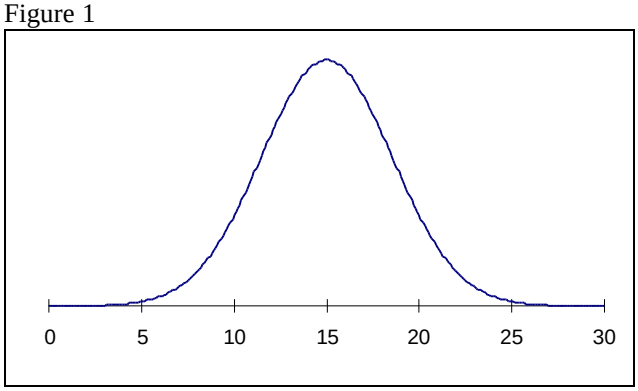
Abstract: The *t* distribution is a probability distribution similar to the Normal distribution. It is commonly used to test hypotheses involving numerical data. This paper provides an understanding of the *t* distribution and uses a musculo-skeletal example to illustrate its application.

Key Indexing Terms: *t*-distribution, *t*-test, chiropractic, osteopathy.

THE *t*-TEST

The distribution of continuous data can often be closely approximated by the normal distribution (1). For example, the distribution of Cobb angles for scoliotic females may be represented by a normal distribution with a notional mean (μ) = 15 degrees and a notional standard deviation (σ) = 5 degrees, and is shown in figure 1. By assuming that these Cobb angles follow a normal distribution, we are able to answer questions such as 'What is the probability that a randomly chosen scoliotic female has a Cobb angle less than 20°.

Figure 1



This is an example of probability at an individual level.

* DEPARTMENT OF SOCIAL AND PREVENTIVE MEDICINE.
MONASH UNIVERSITY. COMMERCIAL RD, PRAHRAN. VIC, 3181.

† PRIVATE PRACTICE.
33 WANTIRNA RD, RINGWOOD, VIC. 3134. PH 03 879 5555

Often, however, we are interested in describing a group of individuals and hope to summarise the group with one statistic. When the underlying population data is normally distributed, the best summary statistic is the mean (average).

For example, suppose we were interested in the erythrocyte sedimentation rate (ESR) of people with polymyalgia rheumatica, if we were to randomly choose 10 people with polymyalgia rheumatica, the group can be described by the average of the 10 observations.

The primary objective of most studies is to use the sample mean to estimate the population mean. A problem with this, however, is that 2 identical studies will not yield the same sample mean for (say) polymyalgia rheumatica.

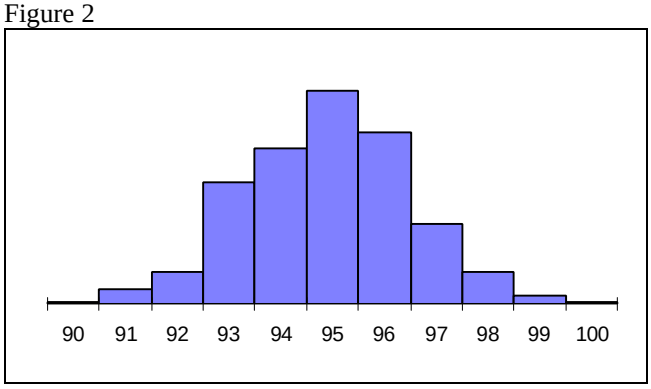
An analogy is coin tossing and estimating the probability of 'Heads'. Suppose the trial is to flip the coin 20 times and count the number of 'Heads'. If we perform the trial once, we may see 12 'Heads' and conclude the probability of 'Heads' is 0.55. If we were to repeat the trial again, we may see 9 'Heads', and conclude the probability of Heads' is 0.45. The dilemma for the researcher is this: 2 identical trials, both trying to estimate the same thing, have produced different results. Which trial has the correct result? Which trial has the incorrect result? Which trial is the researcher to place his faith in?

The questions above can never be answered in research. If we knew which test provided the correct result, why would we need to perform a trial when the correct result is known? The best we can hope for is that every trial yields the correct result with some random variability thrown in for good measure.

The same holds true for means. A random sample of 20 people with polymyalgia rheumatica may give an average of 90 mm/hr, but an identical study with another random sample of 20 people with polymyalgia rheumatica may give the different average of 100 mm/hr. If we were to conduct 1000 studies of polymyalgia rheumatica, it certainly would be likely to get 1000 different average ESR's.

An example of this is shown below in figure 2, where 1000 (simulated) studies each randomly selected 20 people and calculated the mean ESR. The histogram shows the (simulated) distribution of the mean ESR.

Figure 2

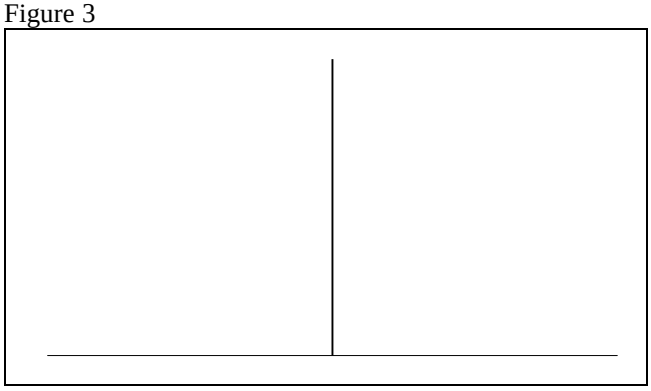


If we were to perform an infinite number of polymyalgia rheumatica studies, and we knew the population standard deviation (σ) of polymyalgia rheumatica, the Central Limit Theorem (2) tells us that the distribution of sample means will have a normal distribution.

This may be seen conceptually in the following way.

Suppose our study design was to sample <u>everybody</u> in the population with polymyalgia rheumatica and calculate the mean ESR. If we were to repeat this study design an infinite number of times, the distribution of means would look like

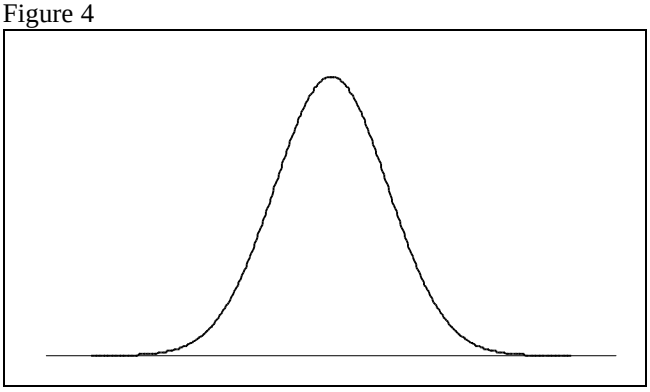
Figure 3



since the mean for polymyalgia rheumatica would be the same for every study.

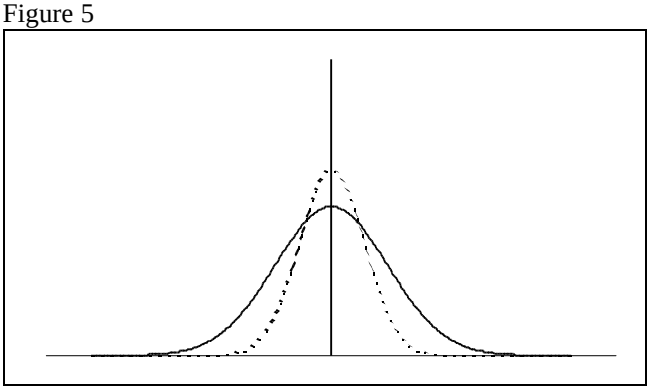
Now suppose that our study design would be to randomly sample <u>1</u> person and calculate the average ESR for polymyalgia rheumatica. Each study would simply give us the value of the ESR for each randomly chosen person. If we were to repeat this study design an infinite number of times, the distribution of means would look exactly like the distribution of the raw data. To simplify this, let us assume that the ESR is normally distributed (with mean μ and standard deviation σ). The distribution of means (and the raw data) would look like

Figure 4



We have now described the two extremes of study design. The first study samples everybody in the population of interest, and the second study samples only one person from the population of interest. The 2 distributions of means are superimposed in figure 5 below (unbroken lines). Intuitively, then, it follows that the distribution of means from a sample size of size n (n is greater than 1, but less than the population size), should be somewhere between the extremes (broken line).

Figure 5



This is the distribution of sample means all calculated from the same sample size (n). The mean of the distribution of means is equal to the population mean μ, so we know where the distribution is centred. To be able to trace the smooth bell shaped curve that is the normal distribution, we also need to know what the standard deviation is. That is, what would the standard deviation of the means be?

To answer this question, we would need to perform several studies (say 100), and estimate the standard deviation of the means by

Sample standard deviation of the mean =

$$\sqrt{\frac{\sum(\bar{x} - \bar{\bar{x}})^2}{99}}$$

where $\bar{x}$ = sample mean

$\bar{\bar{x}}$ = mean of the sample means

This is simply a manipulation of the usual formula for the sample standard deviation of observed data.

In the everyday of clinical research, however, we have enough trouble gaining funding for one study let alone 100, so performing several identical studies is certainly out of the question. Fortunately, a result exists that allows us to calculate the standard deviation of the means.

Note: The term 'Standard Deviation' is usually reserved for use when discussing the spread of data on individuals. When referring to the mean, we call the standard deviation of the means, the standard error (SE) of the means.

If we know the standard deviation (σ) of the data that the sample was drawn from, and if the sample size is of size 'n', the SE of the mean is

$$\text{SE} = \frac{s}{\sqrt{n}}$$

Thus

$$\bar{X} \sim N(m, \frac{s}{\sqrt{n}})$$

Or, in words: The sample mean has a normal distribution. The mean of these means is equal to the population mean (μ), and the standard error (standard deviation of the means) is equal to $\frac{s}{\sqrt{n}}$.

This result can be used to answer questions of the following nature:

Polymyalgia rheumatica in (population X) is normally distributed with mean (95 mm/hr) and standard deviation (20 mm/hr). A sample of size 15 is to be randomly selected and the mean ESR is to be calculated. What is the probability that the sample mean will be less than 80 mm/hr ?

It is rarely the case that we know what the population standard deviation (σ) is, and usually need to estimate it with the following

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}}$$

where s denotes the *sample* standard deviation

The SE of the mean can then be estimated by

$$\text{SE} = \frac{s}{\sqrt{n}}$$

When this is the case, the Central Limit Theorem does not apply any more. When the underlying data are normally distributed, however, the statistic

$$t^* = \frac{\bar{x} - m}{s/\sqrt{n}}$$

has a t distribution. This distribution is defined by 1 parameter, called the 'degrees of freedom'.

Need for the t distribution stems from the fact that we have had to estimate the standard deviation, throwing extra variability into the problem. We penalise ourselves for this by using the t distribution instead of the normal distribution.

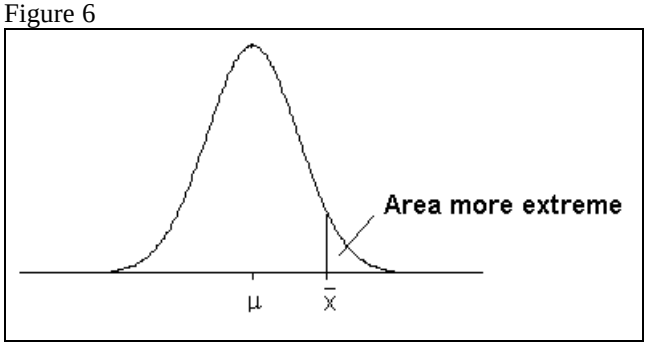
In our previous paper (1), we highlighted the point that to obtain probabilities for an observation when using the normal distribution, we need to know how many standard deviations from the mean the observation lies. When looking at means instead of observations, we are interested in the number of SE's the sample mean lies from the population mean. t^* does just that. Once t^* is known, we can then compare this to the t distribution to calculate probabilities.

ONE SAMPLE t TEST

In a previous paper (3), the concept of an hypothesis was introduced. The one sample t test is concerned with making inference regarding a population mean. For example, suppose you were interested in testing the hypothesis that the average ESR for polymyalgia rheumatica was 95 mm/hr (μ). To show this, you would need to randomly select 'n' (say 100) people with polymyalgia rheumatica. From this sample we obtain 2 statistics. The sample mean $\bar{x}$, and the sample standard deviation (s) which will yield the SE ($\frac{s}{\sqrt{n}}$)

When testing an hypothesis, we always assume the hypothesis is correct. We now want to know what the probability of our observed sample mean ($\bar{x}$) or something more extreme occurring is.

Figure 6



In mathematical notation, this can be written as

$$2P(\overline{X} > \overline{x})$$

Where $\overline{X}$ denotes the random variable (the thing which can vary from study to study) and $\overline{x}$ denotes the actual observation. The extra factor of '2', is needed to perform a 2 sided test (4).

By calculating the number of standard errors the sample mean lies from the hypothesised mean (t^*), we are able to obtain the probability $P(\overline{X} > \overline{x})$, by comparing t^* to the appropriate t distribution. Having multiplied this probability by '2', we have then calculated the 2 sided p-value (4).

Common practice is to reject the hypothesis when the p-value is less than 0.05, and not reject it when the p-value is greater than 0.05.

The concept of a p-value and its interpretation are discussed in (3) and will not be re discussed here.

TWO SAMPLE t TEST

This is more common a scenario than the one sample t test.

Usually we want to compare the means of 2 groups. For example, the mean of a treatment group and the mean of a control group for polymyalgia rheumatica. The hypothesis tested here, is the hypothesis stated in (3) ie. 'Nothing Happens', or the means in the 2 groups are equal to each other. If we denote the mean of the treatment group by μ_1 and the mean of the control group by μ_2, then the hypothesis that we want to test is

$$\mu_1 - \mu_2 = 0$$

the study design would be to take a random sample of n_1 people who have treatment, and a random sample of n_2 people who act as controls, and calculate the difference between the sample means by

$$\overline{X}_1 - \overline{X}_2$$

If we assume that the underlying distributions which the two samples were taken from are both normally distributed, then the distribution of each of the means will also be normally distributed as discussed before. It can be shown that the difference between 2 normally distributed variables will also have a normal distribution.

Since our hypothesis is that $\mu_1 - \mu_2 = 0$, we assume that the mean of this distribution is zero. All that is required is to find the SE of $\overline{X}_1 - \overline{X}_2$. This in itself can be derived using different assumptions (5), and will not be discussed here.

The reader should be able to see that we have

a sample mean	$\overline{X}_1 - \overline{X}_2$
an hypothesised mean	$\mu_1 - \mu_2 = 0$
a standard error	

The p-value can then be derived using the same method as with the one sample t test. That is, calculate the number of SE's the sample mean lies from the hypothesised mean, and compare this t statistic to the appropriate t distribution.

CONCLUSION

t distributions help us decide if a mean is different from a known standard value.

When reading the literature it is important to understand the meaning of t distributions and how they differ from other important distributions.

REFERENCES

1. Ugoni A, Walker BF. An Introduction to Probability Distributions. COMSIG Review 1995; 4(1):Pages 16-23.
2. Dawson-Saunders B., Trapp RG. Basic and Clinical Biostatistics. Connecticut: Prentice-Hall, 1990: 84-86.
3. Ugoni A. On the Subject of Hypothesis Testing. COMSIG Review 1993; 2(2): 45-48.
4. Zar J. H. Biostatistical Analysis. New Jersey: New Jersey 1984: 97-101.
5. Zar J. H. Biostatistical Analysis. New Jersey: New Jersey 1984: 126-137.