Student's *t* - test

By
Dr. Asaad Al-Yassen
Assist. Professor of dermato-epidemiology
& occupational health
Department of Community Medicine
University of Basrah

Objectives:

The objectives of our study to enable the students to know the indication and the application of t – test and how to make a conclusion whether accept or reject the null hypothesis.

Student's *t*- test

- Introduced in 1908 by **William Sealy Gosset.**
- Gossset published his mathematical work under the pseudonym **"Student".**

Definition of t – test: It's a method testing hypothesis about the mean of small sample drawn from a normally distributed population when the standard deviation for the sample is unknown.

When we have to use t -test?

- When the population standard deviation σ is not known.
- When sample size is small, $n < 30$.
- It is used only for quantitative data.
- Use to compare between two samples mean.

Types of t – test:

*** Single sample t – test:** we have only 1 group; want to test against a hypothetical mean.

* **Independent sample t – test:** we have 2 means, 2 groups, no relation between groups. e.g. When we want to compare the mean of the treatment group with the placebo group.

*** Paired t – test:** It consists of samples of matched pairs of similar units or one group of units tested twice. e.g. the difference of mean pre and post drug intervention.
i.e. either same people in both groups, or people are related, e.g. husband-wife, left hand-right hand, hospital patient and visitor.

***When they have a large sample size, the range depends on $\bar{X}$ ± 1.96 SE(x)** at 95% confidence interval, while
***When sample size (n) is small, the width or range will be more. So *t* constant must be changed.

$$\bar{X} \pm t\ SE(x)$$

$$\bar{X} \pm t \frac{sd}{\sqrt{n}}$$

From ***t* table** we can get the appropriate *t* value for various levels of **P** for a given sample size **(n – 1)**what wecalled **the degree of freedom.**
Degree of freedom: Is the number of values in a study that are free to vary.

The "t" distributions

- The shape of ***the t-distribution*** depends on the sample size ***(df)***.
- ***t*** distribution spreads less and less as the sample size gets larger. The larger the sample size (or ***df***), the closer the ***t-distribution*** is to a normal distribution.
-The t distribution approaches the standard normal distribution relatively quickly, when df=30 the two distributions are almost identical.

One - sample *t* test

- Compare one sample mean versus population mean when σ is unknown.
- The formula for *t* -test is similar in structure to the Z – test, except that in *t* – test used to estimate the standard error from the sample standard deviation instead of population standard deviation.

$$Z = \frac{|\bar{X} - \mu|}{\frac{\sigma}{\sqrt{n}}}$$

$$t_{n-1} = \frac{|\bar{X} - \mu|}{\frac{sd}{\sqrt{n}}}$$

Steps of calculation of *t* test:

1. State the null hypothesis and alternative hypothesis
2. Calculate the *t* value
3. Calculate the degree of freedom (df = n – 1)
4. Stat the critical value (s) from the ((*t* distribution table)) at 0.05 and 0.01.
5. Draw a conclusion: Compare the calculated *t* to the critical values from the *t* distribution table to determine significance.

- **If the calculated *t* value is smaller than the critical values.**

The difference between the sample mean and the population mean is likely due to sampling error or chance, so accept the null hypothesis.

P > 0.05

- **If the calculated *t* value is equal to or larger than the critical values.**

The difference between the sample mean and the population mean is not likely due to sampling error or chance, so reject the null hypothesis.

P < 0.05 or **P < 0.01**

Sample questions:

Example: The mean weight of 9 under nourished 6 years old children was 17.3 Kg. with standard deviation 2.51 Kg. If the weight of 6 years old children in the general population is normally distributed with a mean weight 20.9 Kg. Determine if the weight of this sample is significantly different from the population of 6 years old children.

Key answer to sample questions:

n = 9

$\bar{X}$ = 17.3 Kg.

Sd = 2.51 Kg.

μ = 20.9 Kg.

1.State the null hypothesis and alternative hypothesis:

The null hypothesis

There is no significant difference between the mean weight 6 years old malnourished children and the mean weight of 6 years old in the general population, and if there is any difference is due to chance or sampling error.

$$H_0 : \mu = \mu_0$$

Alternative hypothesis

H_α**:** $\mu \neq \mu_0$

2.Calculate the *t* value:

$$t_{n-1} = \frac{|\bar{X} - \mu|}{\frac{sd}{\sqrt{n}}}$$

$$t_8 = \frac{|17.3 - 20.9|}{\frac{2.51}{\sqrt{9}}} = 4.301$$

3.Calculate the degree of freedom (df)

$$df = n - 1$$

$$\mathbf{= 9 - 1 = 8}$$

4.State the critical values: From the *t* distribution table, the critical values at the 8 degrees of freedom are:

df	0.05	0.01
8	**2.31**	**3.36**

Since the calculated *t* value is larger than the critical tabulated value at 95% level, therefore, the probability to find difference by chance is less than 0.05.

(Calculated*t* value > critical tabulated value at 95% level)

So $P < 0.05$

The calculated *t* value is larger than the critical tabulated value at 99.7% level, therefore, the probability to find difference by chance is less than 0.01.

(Calculated *t* value > critical tabulated value at 99.7% level)

So $P < 0.01$

5.Conclusion: We reject the null hypothesis at both 95% and 99.7% confidence limit, which means that there is highly significant difference between mean weight of 6 years old malnourished children and the mean weight of 6 years old children in the general population.

t test for two independent samples((*t* – test for the difference between two means))

The t test must be used when:

- The population variances are unknown & summation of both sample size are less than 30.

For application of t test it is assumed that:

- The population variances are equal and
- The samples are taken from two normally or approximately normally distributed population.

The hypotheses are: $H_0 : \mu_1 - \mu_2 = 0$

$H_\alpha : \mu_1 - \mu_2 \neq 0$

$$t_{n_1 + n_2 - 2} = \frac{|\bar{X}_1 - \bar{X}_2|}{SE\,(x_1 - x_2)}$$

$$SE_{(x1-x2)} = \sqrt{\left[\frac{(n_1 - 1)\,S_1 + (n_2 - 1)\,S_2}{n_1 + n_2 - 2}\right]\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}$$

or

$$SE_{(x1-x2)} = \sqrt{S^2_p\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}$$

$$S^2_p = \frac{(n_1 - 1)\, S^2_1 + (n_2 - 1)\, S^2_2}{n_1 + n_2 - 2}$$

$$df = n_1 + n_2 - 2$$

Pooled estimate of variance:

- The pooled estimate of the population variance is the average of both sample variances, once adjusted for their degrees of freedom.

1- You can know if you have made a mistake in calculating the pooled estimate of the variance if it does not come out between the two estimates.

2- The pooled variance should be closer to the variance of the larger sample.

t test for dependent samples
paired samples t test

Dependent samples: are samples that are paired or matched in some way.

Examples of dependent sample:

1- Samples in which the ***same subjects*** are used in a pre&post situation.

2- Another type of dependent samples is ***matched samples***.

Steps

1.State the null and the alternative hypotheses.

$$H_0 : \mu_d = 0$$

$$H_a : \mu_d \neq 0$$

Calculate the difference for each of the pairs of data, ***d.***

2. Find the mean of the differences. $\bar{d}$.
3. Find the standard deviation of the differences, S_d.
4. Find the estimated standard error of the differences.
5. Find the test value, **t.**

Test statistic:

$$t = \frac{|\bar{d}|}{S_d / \sqrt{n}}$$

d.f. = n - 1

n = no. of pairs