t-Test: Types and Application

Swami Vidyamritananda

The t test is one type of inferential, parametric statistics.

The t-test tells us how significant the differences between groups are; In other words, it lets us know if those differences (measured in means/averages) could have happened by chance.

- **There are three main types of t-test:**

1. Independent Samples t-test compares the means for two groups.

2. Paired sample t-test compares means from the same group at different times (say, one year apart).

3. One sample t-test tests the mean of a single group against a known mean.

Student's t-test or Independent t-test

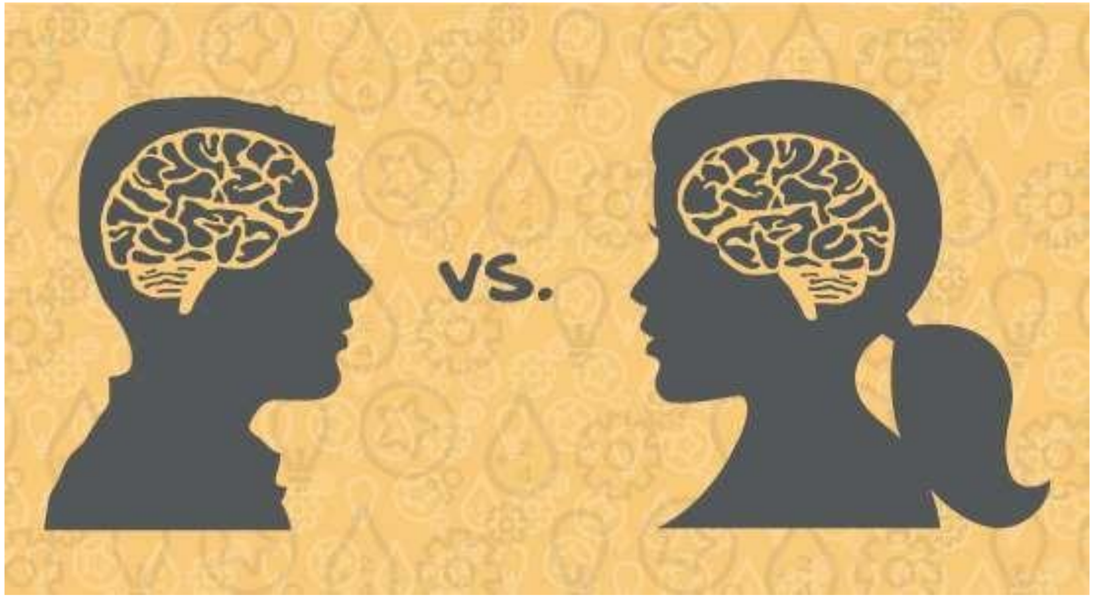
(William Sealy Gosset, 1908)

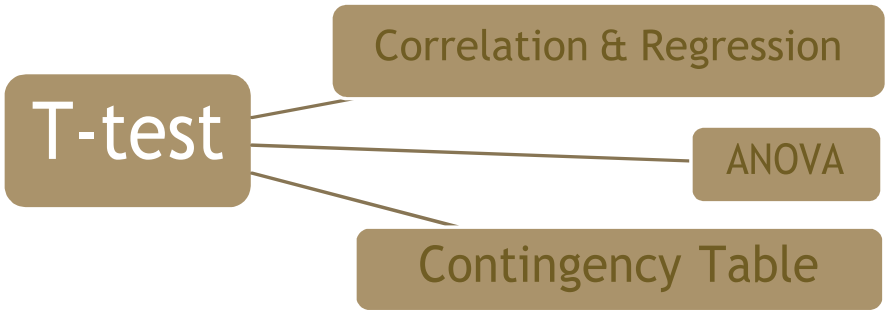
Determines whether there is a statistically significant difference between the means in two unrelated groups.

- E.g. - compare whether systolic blood pressure differs between a control and treated group, IQ between men and women, or any other two groups.

- In social science research this type of t-test is mostly used in survey type of research design.

It is also known as independent samples t-test, two sample t-tests, between samples t-test and unpaired samples t-test.





T- Test vs Correlation & regression

- The *t*-test compares one variable (perhaps blood pressure) between two groups.
- Correlation and regression is performed to see how two variables (perhaps blood pressure and heart rate) vary together.

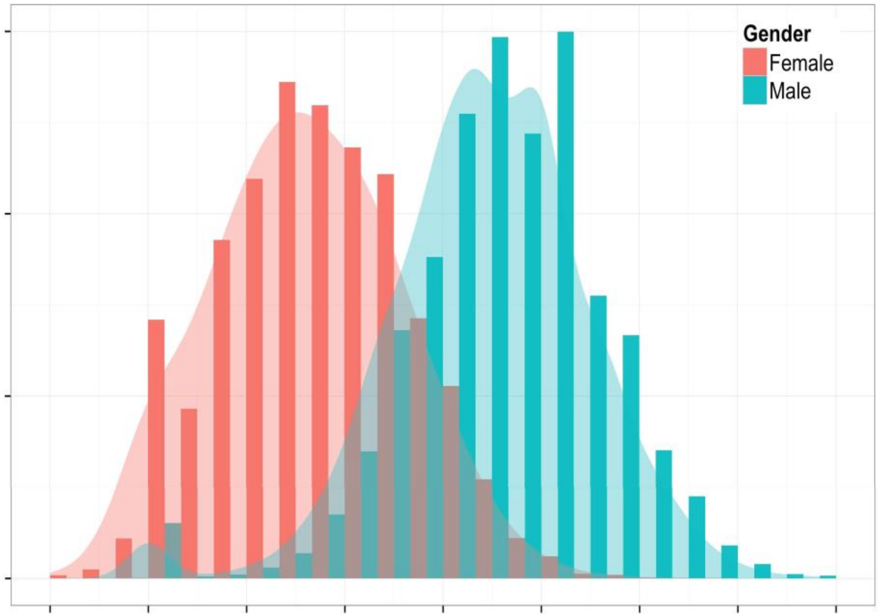
T- Test vs ANOVA

- The *t*-tests (and related nonparametric tests) compare exactly two groups.
- ANOVA (and related nonparametric tests) compare three or more groups.

T- Test vs Fishers or chi-square tests

- *t*-test is performed to compare a continuous/scalable variable (e.g., blood pressure, weight or enzyme activity).
- a contingency table is used to compare a categorical variable (e.g., pass vs. fail, viable vs. not viable).

A *t*-test is used when we are looking at a numerical variable (which is in ratio or interval /scale) - for example, height - and then comparing the averages of two separate populations or groups (categorical variables e.g., males and females).



Assumptions:

- **Two independent samples;** One independent, categorical variable that has two levels/groups. One continuous dependent variable.
- **Unrelated groups**
- **Data should be normally distributed**
- **The two samples should have the same variance**

Assumption of homogeneity of variance

The independent t-test assumes the variances of the two groups we are measuring are equal in the population. If variances are unequal, this can affect the Type I error rate. The assumption of homogeneity of variance can be tested using Levene's Test of Equality of Variances.

Null Hypothesis:

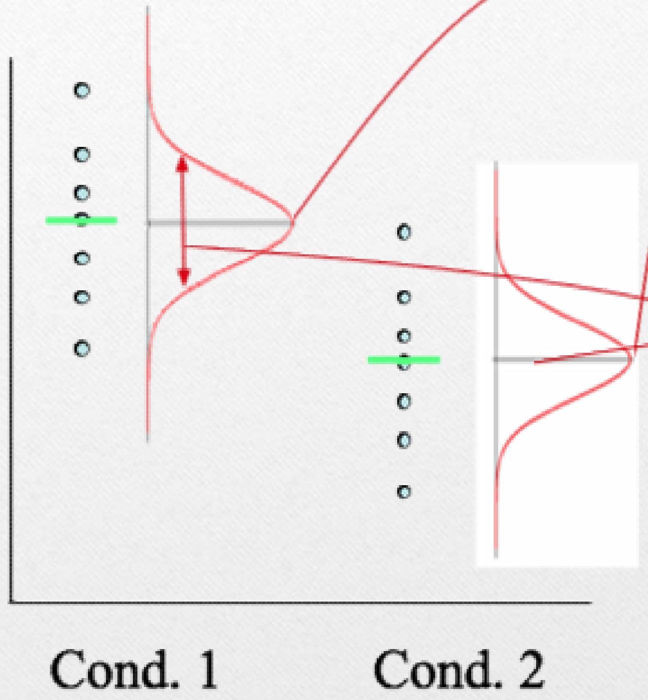
H_0: u1 - u2 = 0, where u1 is the mean of first population and u2 the mean of the second.

As above, the null hypothesis tends to be that there is no difference between the means of the two populations; or, more formally, that the difference is zero (so, for example, that there is no difference between the average heights of two populations of males and females).

Equation:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$$

The difference between the mean divided by the pooled standard error of the mean.



$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$$

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^{\,2}}{n_1} + \frac{s_2^{\,2}}{n_2}}$$

Example:

Student’s T-tests can be used in real life to compare means. For example, a drug company may want to test a new cancer drug to find out if it improves life expectancy. In an experiment, there’s always a control group (a group who are given a placebo, or “sugar pill”). The control group may show an average life expectancy of +5 years, while the group taking the new drug might have a life expectancy of +6 years. It would seem that the drug might work. But it could be due to a fluke. To test this, researchers would use a Student’s t-test to find out if the results are repeatable for an entire population.

Examples of typical questions that the independent samples t-test answers are as follows:

- **Medicine** – Has the quality of life improved for patients who took drug A as opposed to patients who took drug B?
- **Sociology** – Are men more satisfied with their jobs than women?
- **Biology** – Are foxes in one specific habitat larger than in another?
- **Economics** – Is the economic growth of developing nations larger than the economic growth of the first world?
- **Marketing** – Does customer segment A spend more on groceries than customer segment B?

Paired-samples t-test

Paired-samples t-test compares the means between two related groups on the same continuous, dependent variable.

In social science research, paired sample t-test is mostly used in Experimental research design.

Equation:

$$t = \frac{\Sigma d}{\sqrt{\frac{n\left(\Sigma d^2\right) - \left(\Sigma d\right)^2}{n-1}}}$$

Where; sum of

d = Difference

n = No. of cases

d^2 = Square of differences

Assumptions:

Assumption #1: Dependent variable should be measured on a **continuous** scale (i.e., it is measured at the **interval** or **ratio** scale). Examples of variables include revision time (measured in hours), intelligence (measured using IQ score), exam performance (measured from 0 to 100), weight (measured in kg), and so forth.

Assumption #2: Independent variable should consist of **two categorical**, "**related groups**". “Related groups" indicates that the same subjects are present in both groups. It is possible to have the same subjects in each group is because each subject has been measured on two occasions on the same dependent variable.

Assumption #3: There should be **no significant outliers** in the **differences** between the two related groups. Outliers are simply single data points within data that do not follow the usual pattern (e.g., in a study of 100 students' IQ scores, where the mean score was 108 with only a small variation between students, one student had a score of 156, which is very unusual, and may even put her in the top 1% of IQ scores globally).

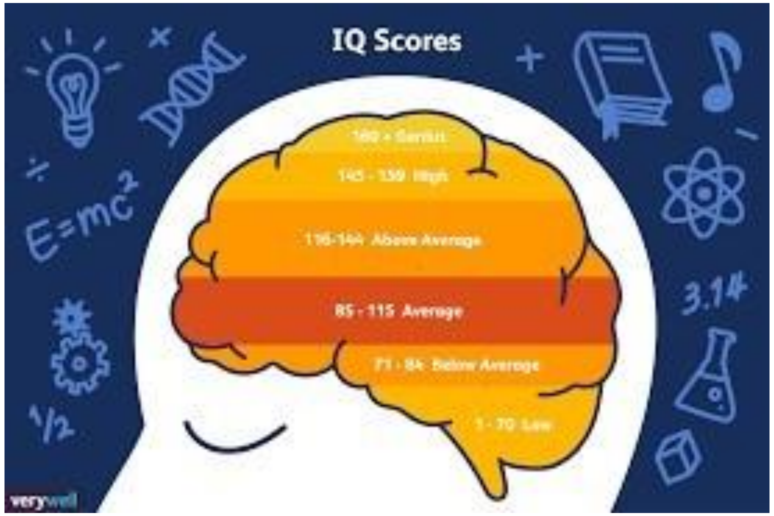
Assumption #4: The **distribution of the differences** in the **dependent variable** between the two related groups should be **approximately normally distributed**.

Example:

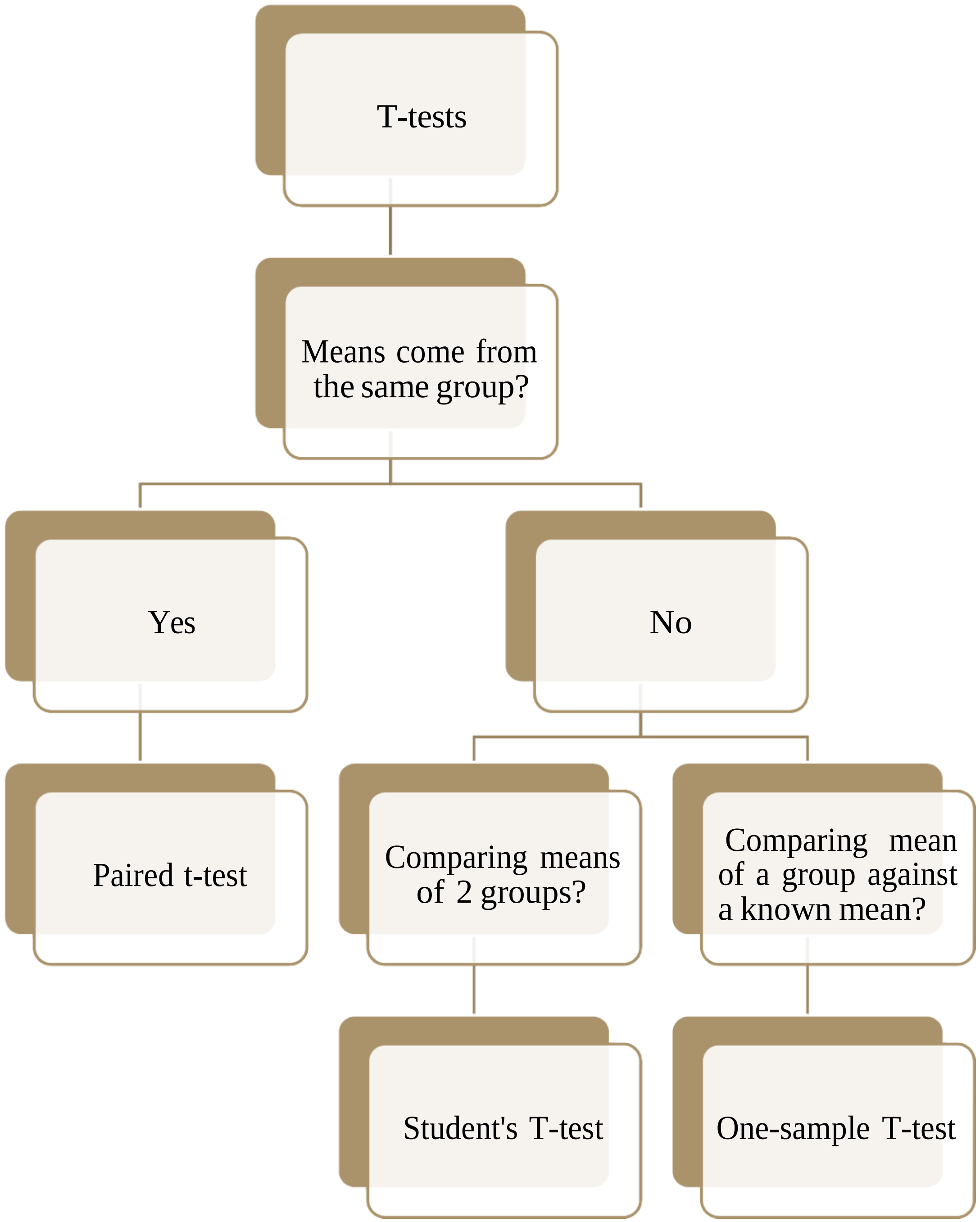
A paired *t*-test can be run on a variable that was measured twice for each sample subject to test if the mean difference in measurements is significantly different from zero. For example, consider a sample of people who were given a pre-test measuring their knowledge of a topic. Then, they were given a video presentation about the topic, and were tested again afterwards with a post-test:

Sample Subject	Pre-Test Score	Post-Test Score	Difference	D^2
1	10	18	8	16
2	14	12	-2	4
3	15	15	0	0
Total =			10	20

One sample T-test



There is only one group which is to be compared to a set value or a known population mean.


T-tests
Means come from the same group?
Yes
No
Paired t-test
Comparing means of 2 groups?
Comparing mean of a group against a known mean?
Student's T-test
One-sample T-test