## T- Test

A t-test (also known as Student's t-test) is a tool for evaluating the means of one or two populations using hypothesis testing.

A t-test may be used to evaluate whether a single group differs from a known value (a one-sample t-test), whether two groups differ from each other (an independent two-sample t-test), or whether there is a significant difference in paired measurements (a paired, or dependent sample t-test).

Q What if I have more than two groups?

You cannot use a t-test. Use a multiple comparison method. Examples are Analysis of variance (ANOVA), Tukey - Kramer pairwise comparison, Dunnett's comparison to a control, and analysis of means (ANOM).

## T-Test Assumptions

While t-tests are relatively robust to deviations from assumptions, t-test do assume that:

- The data are continous
- The sample data have been randomly sampled from a population.
- There is homogeneity of variance (i.e., the variability of data in each group is similar).
- The distribution is approximately normal.

For two-sample t-tests, we must have independent samples. If the samples are not independent, then a paired t-test may be appropriate.

| | One-Sample t-test | Two-Sample t-test | Paired t-test |
|---|---|---|---|
| Synonyms | Student's t-test | • Independent groups t-test<br>• Independent samples t-test<br>• Equal variances t-test<br>• Pooled t-test<br>• Unequal variances t-test | • Paired groups t-test<br>• Dependent samples t-test |
| Number of variables | One | Two | Two |
| Type of variable | • Continous measurement | • continous measurement<br>• Categorical or Nominal to define groups | • Continous Measurement<br>• Categorical or Nominal to define pairing within group |
| Purpose of test | Decide if the population mean is equal to a specific value or not | Decide if the population means for two different groups are equal or not. | Decide if the difference b/w paired measurements for a population is zero or not. |

| | One-sample t-test | Two-Sample t-test | Paired t-test |
|---|---|---|---|
| Example test if... | mean heart rate of a group of people is equal to 65 or not. | Mean heart rates for two groups of people are the same or not. | mean difference in heart rate for a group of people before and after exercise is zero or not. |
| Estimate of Population Mean | Sample average | Sample average for each group. | Sample average of the differences in paired measurements |
| Population Standard deviation | Unknown, use sample standard deviation | Unknown, use sample standard deviations for each group. | Unknown, use sample standard deviation of differences in paired measurements. |
| Degrees of Freedom | Number of observations in sample minus 1, or $n-1$ | Sum of observations in each sample minus 2, or: $n_1 + n_2 - 2$ | Number of paired observations in sample minus 1, or: $n-1$ |

The table above shows only the t-tests for population means. Another common t-test is for correlation coefficients. You can use this t-test to decide if the correlation coefficient is significantly different from zero.

## How to Perform a t-test

- For all of the t-tests involving means, you perform the same steps in analysis:

1. Define your null ($H_0$) and Alternative ($H_\alpha$) hypothesis before collecting your data.

2. Decide on the alpha value (or $\alpha$ value). This involves determining the risk you are willing to take of drawing the wrong conclusion.
   For example, suppose you set $\alpha = 0.05$ when comparing two independent groups. Here, you have decided on a 5% risk of concluding the unknown population means are different when they are not.

3. Check the data for errors.

4. Check the assumptions for the test.

5. Perform the test and draw your conclusion. All t-tests for means involves calculating a test ~~stas~~ statistics. You compare the test statistic to a theoretical value from the t-distribution. The theoretical value ~~of~~ involves both the ~~of~~ value and the degrees of freedom for your data.

One Sample T-test Statistics

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

$\bar{X}$ = Sample Mean
$\mu$ = Population Mean
$S$ = Standard Deviation
$n$ = Number of observations

Two Sample t test for Comparing Two means

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \Delta}{\sqrt{\frac{S_1^2}{n_1^2} + \frac{S_2^2}{n_2^2}}}$$

where, $\bar{x}_1$ and $\bar{x}_2$ are the means of the two samples

$\Delta$ is the hypothesize difference between the population mean (0 if testing for equal means).

$S_1$ and $S_2$ are the standard deviations of the two samples.

$n_1$ and $n_2$ are the sizes of the two samples.

## Paired T-test Formula

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

where d: difference per paired value.
n: number of samples

## Z-test

Z-test is a stastical tool used for the comparison or determination of the significance of several measures, particularly the mean in a sample from a normally distributed population or between two independent population

- Like t-tests, Z tests are also based on normal probability distribution.
- Z-test is the most commonly used stastical tool in research methodology, with it being used for studies where the sample size is large $(n > 30)$
- In the case of the Z-test, the variance is usually known.
- Z-test is more convenient than t-test as the critical value at each significance level in the confidence interval is the sample for all sample sizes.

- A z-score is a number indicating how many standard deviations above or below the mean of the population is.

## One Sample z-test (One tailed z-test)

- One Sample z-test is used to determine whether a particular population parameter, which is mostly mean, significantly different from an assumed value.

- It helps to estimate the relationship between the mean of the sample and the assumed mean.

- In this case, the standard normal distribution is used to calculate the critical value of the test.

- If the z-value of the sample being tested falls into the criteria for the one-sided tests, the alternative hypothesis will be accepted instead of the null hypothesis.

- A one-tailed test would be used when the study has to test whether the population parameter being tested is either lower than or higher than some hypothesized value.

- A one-sample z-test assumes that data are a random sample collected from a normally distributed population that all have the same mean and same variance.

- This hypothesis implies that the data is continous, and distribution is symmetric.

- Based on the alternative hypothesis set for a study, a one-sided z test can be either a left-sided z test or a right sided z-test.

## Two Sample z-test (two-tailed z-test)

- In this the case of two sample z-test, two normally distributed independent samples are required.

- A two-tailed z-test is performed to determine the relationship between the population parameters of the two samples.

- In the case of the two-tailed z-test, the alternative hypothesis is accepted as long as the population parameter is not equal to the assumed value.

- The two-tailed test is appropriated when we have $H_0: \mu = \mu_0$ and $H_a: \mu \neq \mu_0$ which may mean $\mu > \mu_0$ or $\mu < \mu_0$

- Thus, in a two-tailed test, there are two rejection regions, one on each tail of the curve.

## Z-test Application

- Z-test is performed in studies where the sample size is larger and the variance is known.

- It is also used to determine if there is a significant difference between the mean of two ~~idep~~ independent samples.

- The z-test can also be used to compare the population proportion to an assumed proportion or to determine the difference between the population proportion of two samples.

## Z-test Formula

→ For the normal distribution population with one sample:-

$$Z = \frac{\bar{x} + \mu}{\frac{\sigma^2}{\sqrt{n}}}$$

where $\bar{x}$ is the mean of the sample, and $\mu$ is the assumed mean, $\sigma$ is the standard deviation and $n$ is the number of observations.

→ Z-test for the difference in mean:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Where $\sigma$ is the standard deviation of the samples and $n_1$ and $n_2$ are the number of observation of two samples.

| Basis for Comparison | T-test | Z-test |
|---|---|---|
| Definition | The t-test is a test in statistics that is used for testing hypothesis regarding the mean of a small sample taken population when the standard deviation of the population is not known. | Z-test is a statistical tool used for the comparison or determination of the significance of several statistical measure, particularly the mean in a sample to from a normally distributed population or between two independent samples. |
| Sample Size | The t-test is usually performed in samples of a smaller size ($n \leq 30$) | Z-test is generally performed in samples of a larger size ($n > 30$). |
| Types of distribution Population | T-test is performed on samples distributed on the basis of t-distribution. | Z-tests is performed on samples that are normally distributed. |
| Assumptions | It is not based on the assumption that all key points on the sample are independent. | Z-test is based on the assumption that all key points on the sample are independent. |

| | T-test | Z-test |
|---|---|---|
| Variance or Standard Deviation | Variance or standard deviation is not unown in the t-test. | Variance or Standard Deviation is unown in z-test. |
| Distribution:- | The sample values are to be ~~calculated~~ recorded or calculated by the researcher. | In a normal distribution, the average is considered 0 and the variance as 1. |
| Population Parameters | In addition, to the mean, the t-test can also be used to compare partial or ~~single~~ simple correlations among two samples. | In addition to mean, z-test can also be used to compare the population proportion. |
| Convenience | t-tests are less convenient as they have separate critical values for different sample sizes. | Z-test is more convenient as it has the same critical value for different sample sizes. |

## Anova

Analysis of variance (ANOVA) is a statistical technique that is used to check if the means of two or more groups are significantly different from each other. Anova checks the impact of one or more factors by comparing the means of different samples.

Basically, Here we are testing groups to see if there's a difference between them.

examples of when we to might want to test different groups.

-> A group of psychiatric patients are trying 3 diff different therapies: Counseling, Medication and Biofeedback. You want to see if one therapy is better than the others.

-> A manufacturer has two different processes to make light bulbs. They want to know if one process is better than the other.

The Formula for ANOVA is

$$F = \frac{MST}{MSE}$$

where:-

F = ANOVA coefficient

MST = Mean Sum of Squares due to treatment.

MSE = Mean Sum of Squares due to error.

## One-way Anova

In statistics, one-way Analysis of variance (One-way Anova) is a technique that can be used to compare whether two sample's means are significantly different or not (using the F-distribution)

**Example:-** you have group of individuals randomly split into smaller groups and completing different tasks.
Studying the effects of tea on weight weight loss and form ~~tree~~ three groups green tea, black tea, and no tea.

## Assumptions

The results of a one-way ANOVA can be considered reliable as long as the following assumptions are met.

- Response variable residuals are normally distributed (or approximately normally distributed).

- variance of population are equal.

- Response for a given group are in independent and identically distributed normal random variables (not a simple random sample (SRS)).

~~If data are ordinal, a non-parametric alternative to this test~~

## Two way ANOVA

A Two way ANOVA is an extension of the One way ANOVA. With a One way, you have 1 independent variable affecting a dependendent variable. With a Two way ANOVA, there are two independents. Use a two way ANOVA when you have One measurement variable (i.e. a qualitative variable) and two nominal variables.

In other words, if your experiment has a quantitative outcome and you have two categorical explanatory variables, a two way ANOVA is appropriate.

Example:-

You might want to find out if there is an interaction between income and gender for anxiety level at job interviews

The anxiety level is the outcome, or the variable that can be measured. Gender and Income are 2 categorical variables. These ~~independent variables~~ categorical variables are also the independent variables, which are called factors in a two way ANOVA.

The factors can be split into levels.

In the above example, income level could be split into three levels low, middle and high income. Gender could be split into three levels : Male, Female and Transgender groups are all possible combinations of the factors. In this example there would be 3 X 3 = 9 treatment groups.

Assumptions :-

- The population must be close to a normal distribution.
- Samples must be independent.
- Population variance must be equal (i.e. homoscedastic)
- Groups must have equal sample sizes.