

# Health Research ToolBox: A Step by Step Guide for Beginners

---

Jubayer Hossain <https://jhossain.me/>

September 10, 2021

Lead Organizer, Introduction to Scientific Computing for Biologists

Founder,

Center for Health Innovation, Research, Action and Learning



## **Compare Sample Means (parametric)**

---

# Compare Sample Means

- Student's t-test
- Paired Student's t-test
- Analysis of Variance Test (ANOVA)
- Repeated Measures ANOVA Test

## t-test

- It compares **mean** of **two groups**
- It is a parametric statistical test.
- It's used to study if there is **statistical difference** between **two groups**

# Types of t-test

- One sample t-test
- Paired t-test(Dependent)
- Unpaired t-test(Independent)

Unpaired t-test also have 2 categories

- Student's t-test
  - Equal variance
  - Two sample t-test
- Welch t-test
  - Unequal variance
  - Unequal variance t-test

## Selection of t-test

- One sample t-test(for one sample)
- Paired t-test(for dependent samples)
- Student t-test(When sample size and variance are equal)
- Welch t-test(When sample size and variance are different)

# One Sample t-test

It compares the mean of one sample

- Known(from previous study) mean (  $\mu$  )
- Hypothetical mean(  $\mu$  )

## Student's t-test (Con..)

- The independent t-test is also called the two sample t-test, student's t-test, or unpaired t-test.
- It's an univariate test that tests for a significant difference between the mean of two unrelated groups.
- It compares the mean of two independent samples.

### Assumptions

The assumptions that the data must meet in order for the test results to be valid are:

- The independent variable (IV) is categorical with at least two levels (groups)
- The dependent variable (DV) is continuous which is measured on an interval or ratio scale
- The distribution of the two groups should follow the normal distribution
- The variances between the two groups are equal
- This can be tested using statistical tests including Levene's test, F-test, and Bartlett's test.



## Student's t-test (Con..)

If any of these assumptions are violated then another test should be used.

### Interpretation

**Question: Is there a difference in the height between men and women?**

### Hypothesis

- $H_0$ : the means of the samples are equal.
- $H_a$ : the means of the samples are unequal.

### References

<https://pythonfordatascienceorg.wordpress.com/independent-t-test-python/>

# The Hypothesis Being Tested

- Null Hypothesis ( $H_0$ ):  $\mu_1 = \mu_2$ , which translates to the mean of sample\_01 is equal to the mean of sample 02
- Alternative Hypothesis ( $H_1$ ):  $\mu_1 \neq \mu_2$ , which translates to the means of sample01 is not equal to sample 02

## Homogeneity of variance

Of these tests, the most common assessment for homogeneity of variance is Levene's test. The Levene's test uses an F-test to test the null hypothesis that the variance is equal across groups. A p value less than .05 indicates a violation of the assumption.

[https://en.wikipedia.org/wiki/Levene%27s\\_test](https://en.wikipedia.org/wiki/Levene%27s_test)

<https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.levene.html>

# Levene's test

Levene's test is an inferential statistic used to assess the equality of variances for a variable calculated for two or more groups.

## **Interpretation**

- $H_0$ : The variances are equal between two groups
- $H_a$ : The variances are not equal between two groups

## Checking normal distribution by shapiro method

- <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.shapiro.html>
- <https://stats.stackexchange.com/questions/15696/interpretation-of-shapiro-wilk-test>

## Paired t-test

It compares the mean between two related samples.(each subject is measured twice)

# The Hypothesis Being Tested

- Null Hypothesis ( $H_0$ ):  $\mu_1 = \mu_2$ , which translates to the mean of sample 01 is equal to the mean of sample 02
- Alternative hypothesis ( $H_a$ ):  $\mu_1 \neq \mu_2$ , which translates to the means of sample 01 is not equal to sample 02

## Assumption check

- The samples are independently and randomly drawn
- The distribution of the residuals between the two groups should follow the normal distribution
- The variances between the two groups are equal

## Welch's t-test (Con..)

- It compares the mean of two independent samples.
- It assumes:
  - Samples don't have equal variance
  - Sample size is not equal.

Welch's t-test Assumptions Like every test, this inferential statistic test has assumptions. The assumptions that the data must meet in order for the test results to be valid are:

- The independent variable (IV) is categorical with at least two levels (groups)
- The dependent variable (DV) is continuous which is measured on an interval or ratio scale
- The distribution of the two groups should follow the normal distribution If any of these assumptions are violated then another test should be used.

# Welch's t-test

## Interpretation

- Null hypothesis ( $H_0$ ):  $\mu_1 = \mu_2$ , which translates to the mean of sample 1 is equal to the mean of sample 2
- Alternative hypothesis ( $H_A$ ):  $\mu_1 \neq \mu_2$ , which translates to the mean of sample 1 is not equal to the mean of sample 2



# Analysis of Variance(ANOVA)

---

# ANOVA - Analysis of Variance

- Compares the means of 3(+) groups of data.
- Used to study if there is **statistical difference** between 3(+) group of data.
- Assumes the data are **normally distributed** and have **equal variances**

# One-way ANOVA

- Compares the mean of 3(+) groups of data considering one independent variable or factor.
- Within each group there should be at least three observations.

# Two-way ANOVA

- Compares the means of 3(+) groups of data considering two independent variables or factors.

## Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample are normally distributed.
- Observations in each sample have the same variance.

## Interpretation

- $H_0$ : the means of the samples are equal.
- $H_a$ : one or more of the means of the samples are unequal.

Thank You

