Hypothesis Testing Overview

Hypothesis testing provides confidence in favor of a certain hypothesis. In simpler terms, it is testing an assumption regarding a population parameter and gives an ability to check if the assumption was meaningful or not.

How to conduct a hypothesis test?

1. Have an empirically verifiable claim.
2. Design the experiment.
    a. Determine the appropriate sample size based on the desired power, alpha significance level and expected effect size.
    b. Define the null and alternative hypothesis.
    c. Define the test.
3. Conduct the experiment and collect the data.
4. Based on the data collected
    a. Calculate the test statistic and p-value
5. Interpret the results to accept or reject the null hypothesis.

Although there are hundreds of statistical hypothesis tests that you could use, there is only a small subset that you may need to use in a machine learning project. Here we are discussing parametric tests.

Parametric tests are based on assumptions about the distribution of the underlying population from which the sample was taken.

The assumptions are as follows:

1. Observations in each sample are independent and identically distributed (iid).
2. Sample means come from normal sampling distributions.
3. Observations in each sample have the same variance.

Here is a summary of the tests we have discussed. Note that there are plenty of other tests out there which are designed for special purposes such as working with variances. The following tests are used most commonly and are based off of working with mean values:

Continuous Data:

- One Sample T-Test: Sample mean vs Known or hypothesized population mean
- Two-Sample Unpaired t-Test: 1st Group Sample mean vs 2nd Group Sample Mean in two Conditions

- Paired Two-Sample t-Test: Group Sample mean vs Same Group Sample Mean in two Conditions
- ANOVA: More than two Groups

Categorical Data:

- Chi-Squared

How about one-tailed or two-tailed?

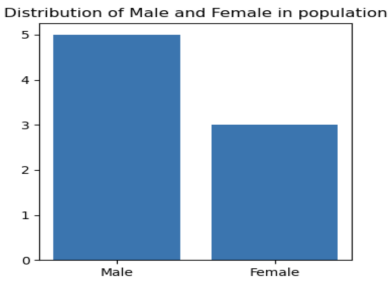
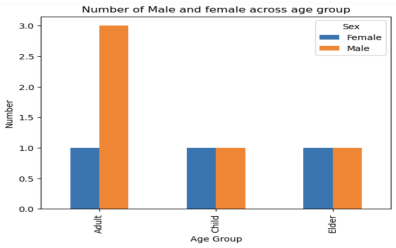
In most of the cases we would conduct above tests in a two-tailed fashion since we are aiming to reject equality of mean, variance, etc. One-tailed would only be used when we are strictly investigating whether a mean/variance is bigger in one group compared to the other.

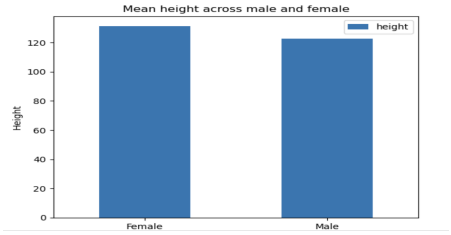
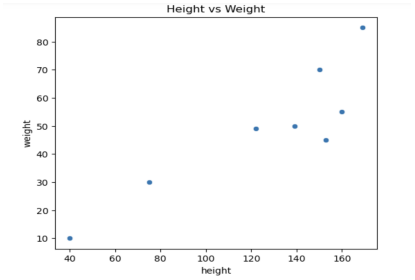
How about Pearsonr?

We would use Pearsonr when dealing with correlation to measure the linear relationship between two datasets.

Have a look at the flowchart and table below to better help you in making decisions on which test to use.

Different types of statistical tests, purpose and example along with type of data on which it is performed are tabulated below:

Sr. No	Type of data	Purpose and example	Suitable hypothesis test
1.	One categorical variable	Ex- Out of a population of male and females, 3 out of 8 people are females. But data recorded for a county of 10000 people had 68 % females. 	Proportion Z test - One sample proportion Z test
	Binary variables	Ex- comparing conversion rate for a person signing up for term deposit (1 or 0) if a sales rep provided 2 hours of additional support (1 or 0).	Two sample proportion Z test
2.	Two categorical variables	Ex - Is there any relationship between the number of male and female being in certain age group. 	Chi Squared test
3.	One numerical continuous value variable	Ex - Comparing if mean height of population is same as a known mean height of population	T test - One sample t-test
4.	One continuous numerical variable	Ex - Comparing mean height of the male and female	Two sample t-test (two groups) - Two-tailed two sample

	& one categorical variable		- t-test (compare if means of two population are equal) - One tailed two sample t-test (compare if mean of one sample is greater than the other)
		Ex - Comparing heights of more than two different species of cats.	Anova (multiple groups)
5.	Two or more numerical variables	Ex - Checking linear relationship between height and weight 	Pearson R correlation test

Flowchart: