Statistical Inference

DATA SCINCE | COM 221-ML

- Learning about what we do not observe using what we observe (data).
- With proper use of statistics, observations can be an **educated / principled approximations**. Otherwise, it's just a **wild guess**.

Three Modes of Statistical Inference:

1. Descriptive Inference
2. Predictive Inference
3. Causal Inference

STATISTICAL TESTING

- The method of hypothesis testing uses **tests of significance** to determine the likelihood that a statement (often related to the mean or variance of a given distribution) is **true**, and at what likelihood we would, as statisticians, accept the statement as true.
- Knowledge of how to appropriately use each test (and when to use which test) is equally important.

Research question: Exploring a comparison, association or relationship

Level of measurement of data: Categorical or scale (continuous)

From your research question you should formulate one or more hypotheses each leading to two possible outcomes:

- **Null Hypothesis (H_0):** Assumes no difference, association or relationship between the variables
- **Alternative Hypothesis (H_A):** Assumes a difference, association or relationship between the variables

The use of an **appropriate** statistical hypothesis test will allow you to decide between these two outcomes by examining the probability (or p) value of the test statistic.

p-value

A decision between the two hypotheses is made by viewing the **p-value**, which is the probability or **chance of a more extreme event happening under the assumption of the null hypothesis**.

- If this probability is small, H_0 is rejected in favor of H_A, which is termed a **statistically significant result.**
- Otherwise we 'fail to reject H_0', which is termed a **non-statistically significant result**.

Significance

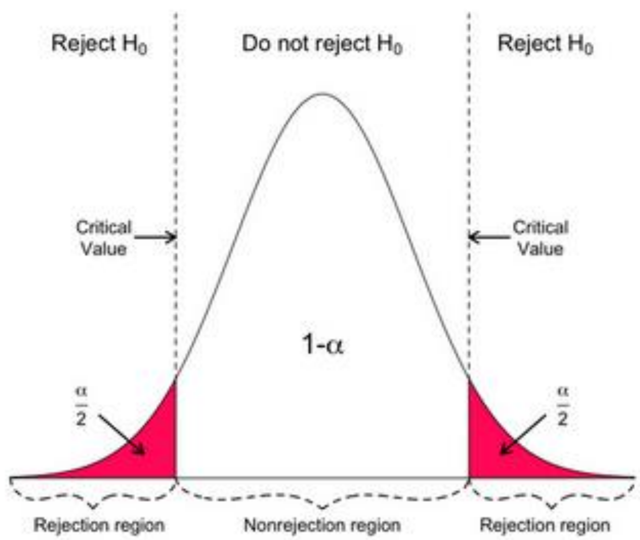
Significance, or statistical significance, describes a **decision made concerning a value stated in the null hypothesis**. When the null hypothesis is rejected, we reach significance. When the null hypothesis is retained, we fail to reach significance.

The decision to reject or retain the null hypothesis is called **significance**.

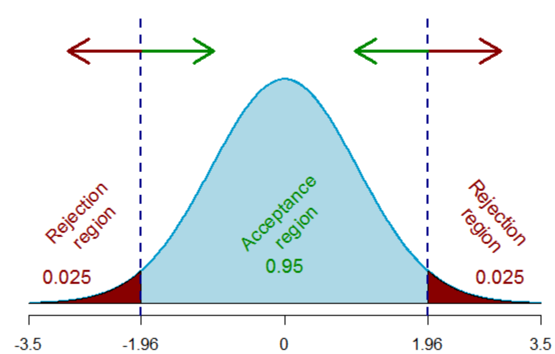
p-value of test statistic	Statistically significant?	Formal action	Informal interpretation
Greater than 0.1	No	Retain H_0	**No evidence** to reject H_0
Between 0.1 and 0.05	No	Retain H_0	**Weak evidence** to reject H_0
Between 0.05 and 0.01	Yes, at the 0.05 level	Reject H_0	**Evidence** to reject H_0
Between 0.01 and 0.001	Yes, at the 0.01 level	Reject H_0	**Strong evidence** to reject H_0
Less than 0.001	Yes, at the 0.001 level	Reject H_0	**Very strong evidence** to reject H_0

You should always **relate back the outcome of the hypothesis testing to particular variables in your study**; don't just conclude with 'reject the null hypothesis'.

You will need to discuss your results and conclusions in the context of the work of others, and the potential for further studies. Note that **statistical significance does not necessarily indicate practical usefulness of a result**.

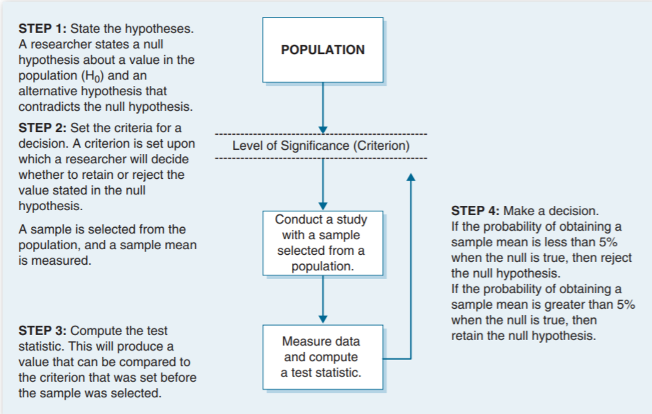


Critical Region for Two-tailed test



Steps to Hypothesis Testing

Step 1: State the hypotheses.
Step 2: Set the criteria for a decision.
Step 3: Compute the test statistic.
Step 4: Make a decision.



Common hypothesis tests

Level of measurement of data	Test of association (relationship)	Test of difference (comparison): 2 groups	Test of difference (comparison): >2 groups
Nominal	Chi-squared	Chi-squared	
Ordinal	Spearman	Mann-Whitney U (unrelated) Wilcoxon (related)	Kruskall-Wallis (unrelated) Friedman (related)
Scale (continuous)	Pearson	Independent t-test (unrelated) Paired t-test (related)	One-way ANOVA (unrelated) Repeated measures ANOVA (related)

Limitations

For your project, you may be restricted in the data you are able to collect. For example, you may not have a true 'random sample', but rather a 'convenience' sample.
Always try to get as representative a sample as possible, but you may only be able to obtain a relatively small number of cases (e.g. less than 30). This may affect the results of statistical testing.
Another influence on outcomes is underlying assumptions necessary for the validity of the tests. Sometimes, data are assumed to be representative of the population from which they have been drawn and come from a normal distribution.

Two Sample T-test for comparing two Means

The two-sample t-test (also known as the independent samples t-test) is a method used to test whether the unknown means of two groups are equal or not.
Determines whether there is statistical evidence that the associated population means are **significantly different**. The independent samples t-test is a **parametric test**.
Examples:
The independent samples *t*-test is commonly used to test the following:

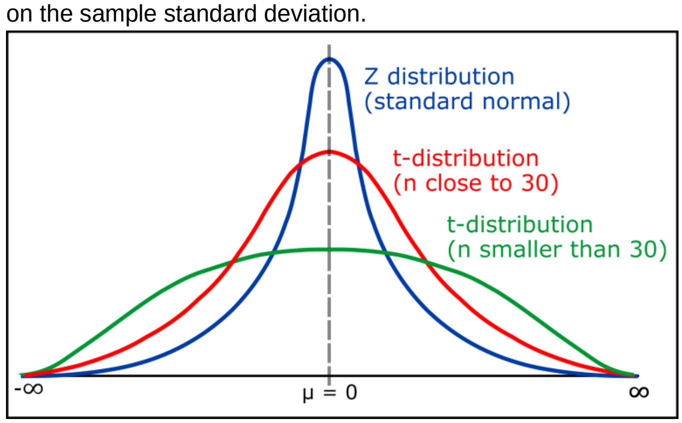
1. Statistical differences between the means of two groups
2. Statistical differences between the means of two interventions
3. Statistical differences between the means of two change scores

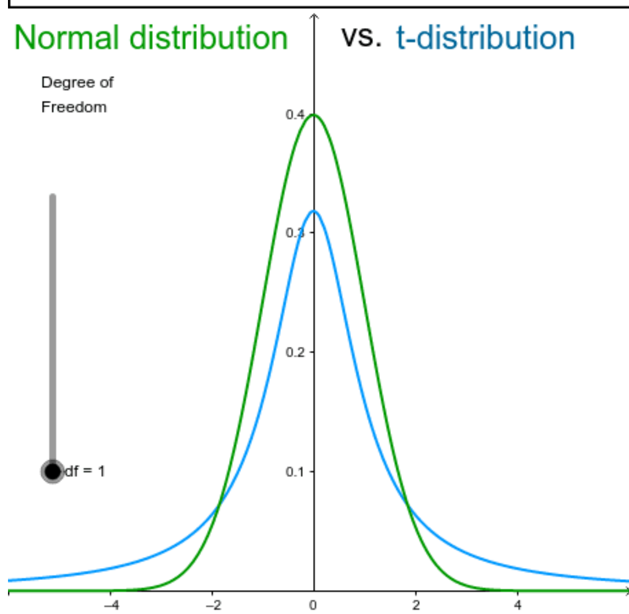
Assumptions:

1. The sampling method for each sample is simple random sampling.
2. The samples are independent.
3. There is no relationship between the subjects in each sample. This means that:
    - Subjects in the first group cannot also be in the second group
    - No subject in either group can influence subjects in the other group
    - No group can influence the other group
4. Normal distribution (approximately) of the dependent variable for each group.
5. No outliers.

T distribution

The t-distribution describes the **standardized distances of sample means** to the population mean **when the population standard deviation σ is not known**, and the observations come from a normally distributed population.
The standard normal or z-distribution assumes that you know the population standard deviation. The t-distribution is based on the sample standard deviation.





T distribution VS. Normal Distribution

1. Like the normal distribution, the t-distribution has a smooth shape, is symmetric, and has a mean of zero.

2. The normal distribution assumes that the population standard deviation is known. The t-distribution does not make this assumption.
3. The t-distribution is defined by the **degrees of freedom**. These are related to the sample size. The curves with more degrees of freedom are taller and have thinner tails.
4. **The t-distribution is most useful for small sample sizes**, when the population standard deviation is not known, or both.
5. As the sample size increases, the t-distribution becomes more similar to a normal distribution.

Hypothesis

Every hypothesis test requires the analyst to state a null hypothesis and an alternative hypothesis. The hypotheses are stated in such a way that they are mutually exclusive. That is, if one is true, the other must be false; and vice versa.

Set	Null hypothesis	Alternative hypothesis	Number of tails
1	$\mu_1 - \mu_2 = d$	$\mu_1 - \mu_2 \neq d$	2
2	$\mu_1 - \mu_2 \geq d$	$\mu_1 - \mu_2 < d$	1
3	$\mu_1 - \mu_2 \leq d$	$\mu_1 - \mu_2 > d$	1

Test Statistics t (Equal variance)

There are actually two forms of the test statistic for this test, depending on whether or not equal variances are assumed.

Equal Variances:

$$t = \frac{\text{difference of group averages}}{\text{standard error of difference}}$$

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Where

$\bar{x}_1$ = Mean of first sample

$\bar{x}_2$ = Mean of second sample

n_1 = Sample size (i.e., number of observations) of first sample

n_2 = Sample size (i.e., number of observations) of second sample

s_1 = Standard deviation of first sample

s_2 = Standard deviation of second sample

s_p = Pooled standard deviation

Test Statistics t (Unequal Variance)

When the two independent samples are assumed to be drawn from populations with unequal variances (i.e., σ12 ≠ σ22), the test statistic t is computed as:

Unequal Variances:

$$t = \frac{\text{difference of group averages}}{\text{standard error of difference}}$$

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where

$\bar{x}_1$ = Mean of first sample

$\bar{x}_2$ = Mean of second sample

n_1 = Sample size (i.e., number of observations) of first sample

n_2 = Sample size (i.e., number of observations) of second sample

s_1 = Standard deviation of first sample

s_2 = Standard deviation of second sample

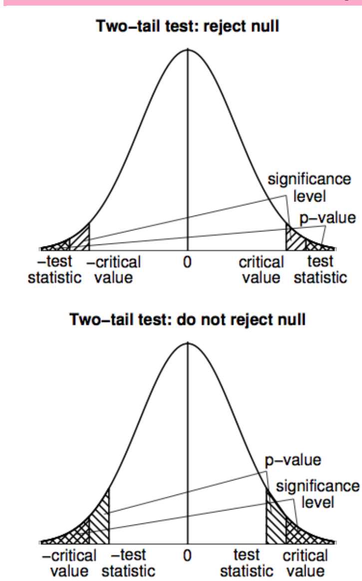
Compare T statistics with critical T value

The calculated t value is then compared to the critical t value from the t distribution table with degrees of freedom. **Degrees of freedom** are the number of independent values that a statistical analysis can estimate.

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2}$$

If the calculated t statistic > critical t value from t table, then we reject the null hypothesis.

When will I reject the null?



Mann-Whitney U Test

The Mann-Whitney U test is a **non-parametric test** that can be used in place of an unpaired t-test. Transforms and ranks values before computing the **test statistic *U***.

It is used to test the null hypothesis that two samples whether observations in one sample tend to be larger (or different) than observations in the other.

$$U' = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

Mann-Whitney U test assumptions

1. The Mann Whitney U-Test is used when the requirements of the t-test are not met.
2. The sample drawn from the population is random.
3. Independence within the samples and mutual independence is assumed.
4. Ordinal or continuous measurement scale is assumed.