Lecture: 06

Hypothesis Testing:
Different Types of Statistical Tests

Topics we will cover

Different Types of Statistical Tests

1. Z-Test
2. T-Test
3. One vs Two Sample Test
4. Paired Sample T-Test
5. Chi-squared test
6. Anova
7. Some Non-Parametrics Test
8. Post-Hoc Analysis

Statistical Tests

A statistical hypothesis test is a method of statistical inference used to decide whether the data at hand sufficiently support a particular hypothesis.

- Test statistics such as z-tests,one sample t-tests, chi-square tests, etc., are commonly used to assess hypotheses about population parameters based on sample data.

NOTE: Generally, in hypothesis tests, test statistic means to obtain all of the sample data and convert it to a single value. For example, Z-test calculates Z statistics, t-test calculates t-test statistic, and F-test calculates F values etc., are the test statistics. Test statistics need to compare to an appropriate critical value (cv) or p-value. A decision can then be made to reject or not reject the null hypothesis.

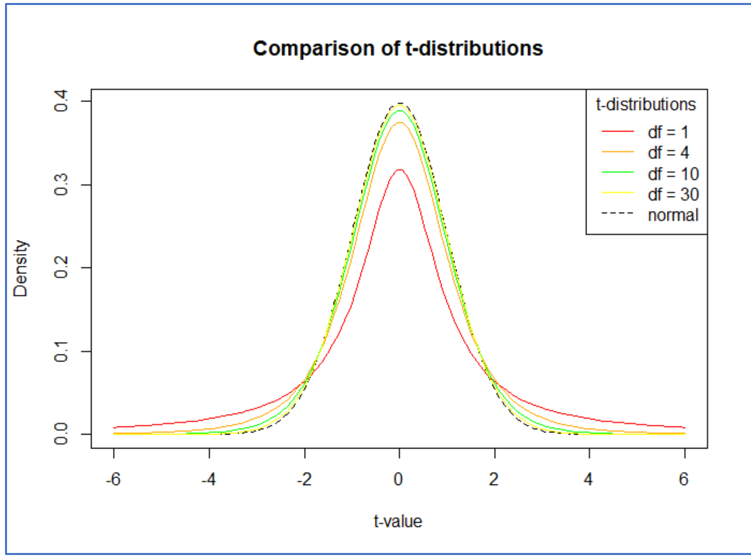
Does knowing more help us?

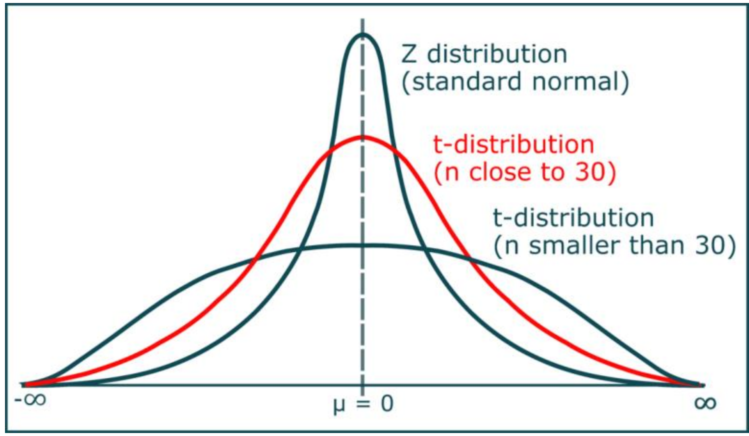
Yes! If we have an **idea of the standard deviation of the underlying population** (or even just have enough data to make an estimate), we can use a **z-test instea**d, which give more accurate results.

According to the theory, we cannot use **z-tests for sample sizes under 30 elements**.

Z – Test vs T – Test

Feature	Z-Test	T-Test
Population Standard Deviation	Known	Unknown
Sample Size	Large (typically n ≥ 30)	Small (typically n < 30) or when population standard deviation is unknown
Underlying Distribution	Normal Distribution	Student's t-Distribution
Sensitivity & Robustness	More sensitive but requires more information	More robust (safer option when population standard deviation is unknown) but less sensitive
When to Use	When you know the population standard deviation and have a large sample	When the population standard deviation is unknown, and you're working with small or moderate samples





Z Test - When to use

The Z-test compares a sample mean to a population mean. It is used when :

- population standard deviation (σ) is known (or can make a reasonable assumption about it) AND
- Either you have a large sample size (typically $n > 30$) OR data is normally distributed.

Assumption: It assumes that the sample data is normally distributed or that the sample size is large enough for the Central Limit Theorem to apply

Compute the **Z-Test Statistic** using the sample mean, *μ_1*, the population mean, *μ_0*, the number of data points in the sample, *n* and the population's standard deviation, *σ*:

$$z = \frac{\mu_1 - \mu_0}{\sigma / \sqrt{n}}$$

T-Test

T-test is a statistical test used to determine if there is a significant difference between the means of two groups.It is used when

- population standard deviation (σ) is unknown AND
- Either you have a smaller sample size (typically n < 30) OR even with large n, σ is still unknown

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

In General, the type of test statistic used depends on the number of samples being compared

- **One Sample:** when there is only one sample that needs to be compared with a given value.
- **Two Samples**, when there are two or more samples to be compared. In this case, tests can include correlation tests and tests for differences between samples.

Additionally, samples can be paired or not paired.

- Paired samples are also called **dependent samples (***observations that are related or matched in some way***)**, while not paired samples are also called **independent samples (***not related or matched***)**.

One sample T-Test

Determine if the mean of a **single sample** is significantly different from a known or hypothesized population mean.

- Commonly used when you have collected data from a single group or sample and want to compare its mean to a specific value or a hypothesized population mean.

How to run a one-sample t test:

```
import numpy as np

from scipy import stats

 stats.ttest_1samp(your_data, popmean=0.5)

>>> TtestResult(statistic=2.456308468440, pvalue=0.017628209047638, df=49)
```

Two Sample t-test

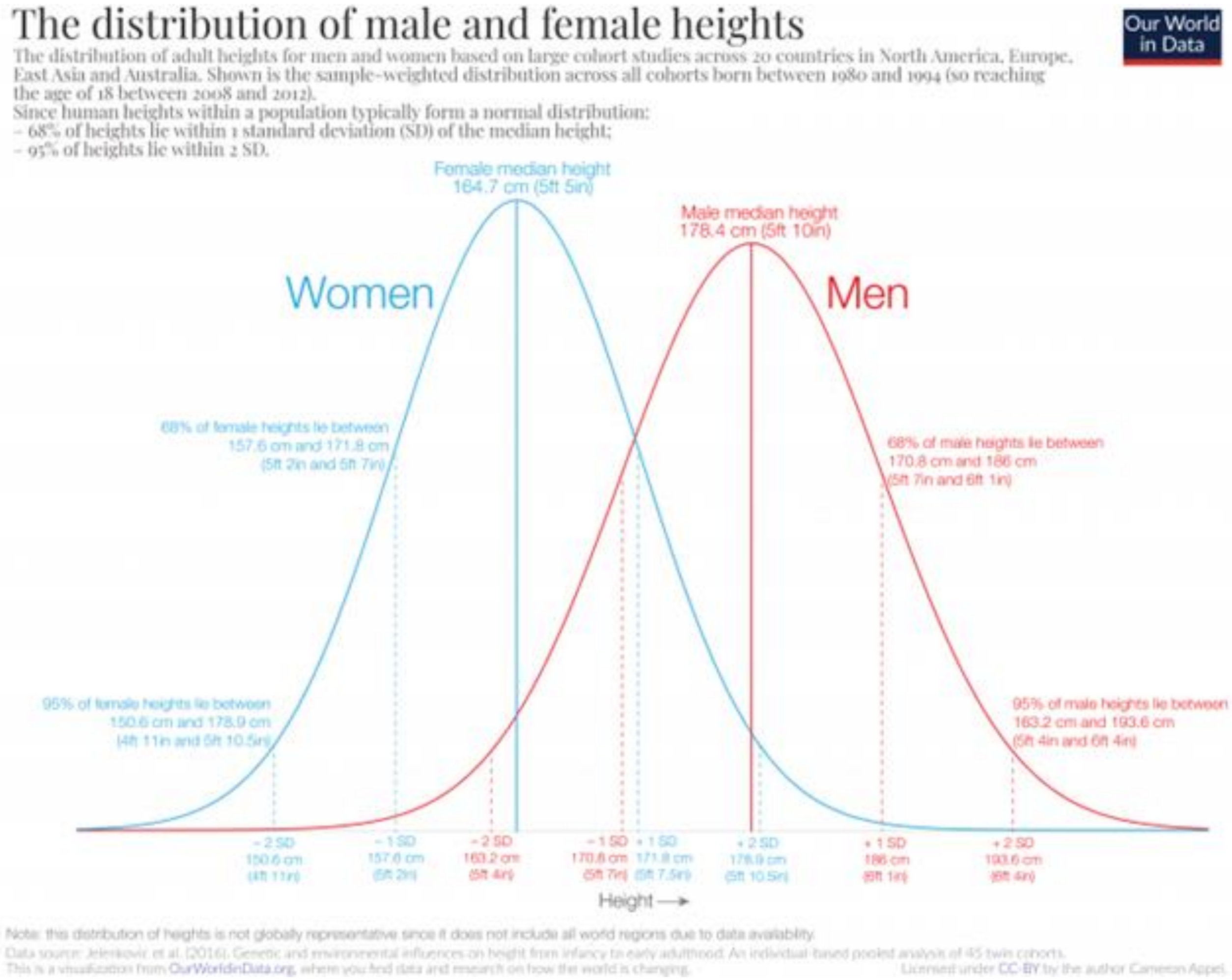
The two sample t test, (also referred to as the unpaired t test), is used to compare the means of two different samples.

Example: We have noticed most humans fall into one of two distinct categories– male or female. We would like to know if our sample of males is taller than our sample of females.

Can we just take the average of the two samples?

Two-Sample T-test

- NO! Simply taking the average of the two samples is not sufficient to determine if one group is taller than the other (does not account for variability within each group).
- Instead, we would conduct a statistical test to determine whether there is a statistically significant difference between the two groups.



Two Sample T-Test

- **Null hypothesis:** Men and women are the same height
- **Alternative hypothesis:** Men and women are different heights
- **p-value:** the probability that we would see these observations if the null hypothesis is true/correct

Q: What sort of p value would we see if men and women had different heights?

When can we use the T-test?

- We can use the test for continuous data obtained from a random sample that follows a normal distribution.

Our Assumptions: For the t-test we assume:

- The data is normally distributed
- We care about **the mean** → ex: whether there is a significant difference in the means of the two groups.

Q: What if my data isn't nearly normally distributed?

- If your **sample sizes are very small**, you might not be able to test for normality. You might need to rely on your understanding of the data.
- When you cannot safely assume normality, you can **perform a nonparametric test** that doesn't assume normality.

Paired Sample t test

The paired sample t test is used to compare the means of **two related groups** of samples.

- It is used in a situation where you have two values (i.e., a pair of values) for the same group of samples.
- Often these two values are measured from the same samples either at two different times, under two different conditions, or after a specific intervention.

Paired Sample t test: Example

A company monitors a bunch of humans, test them for intelligence, and then run one half of them through a training session to make them smarter. Afterwards, they want to know if their training worked.

This would be called a **paired t-test.**

Null Hypothesis: ?

Alternative Hypothesis: ?

Paired Sample t test: Example

A company monitors a bunch of humans, test them for intelligence, and then run one half of them through a training session to make them smarter. Afterwards, they want to know if their training worked.

This would be called a **paired t-test.**

Null Hypothesis: The training did nothing

Alternative Hypothesis: The training came from a different distribution

Paired Sample t test: Example

A company monitors a bunch of humans, test them for intelligence, and then run one half of them through a training session to make them smarter. Afterwards, they want to know if their training worked.

This would be called a **paired t-test.**

Null Hypothesis: The training did nothing

Alternative Hypothesis: The training came from a different distribution

Ques: If they get a p-value of .05. What can they conclude?

What about this?

We are monitoring **birds from two different places on the planet,** and get the following results:

Bird Type	**Location A**	**Location B**
Grackle	7	13
Pigeon	2	7
Sea pigeon	15	1
One of those big fish-beak things	13	0
Big long bird	22	0
Bat	3	4

Each bird type and location falls into distinct categories, making them categorical variables suitable for analysis using methods like the **chi-square test**.

We want to find out if two different places on Earth have the same types of birds

Chi Square Test

- The Chi-Square Test is a vital statistical test performed to analyse two **nominal variables** and see **whether or not they are related**.
- It is most useful when the data is split into areas such as gender, product type, or favourite services.
- This test is crucial for making decisions when conducting relations analysis in categorical data.
- **<u>Types of Chi-Square Tests:</u>**

1. **Chi-Square Goodness-of-Fit Test:** This test checks **whether the sample data belongs to a particular population** with a given distribution. For example, it can be used to determine **whether the results in a given trial are random** by comparing the observed values of each constructor with the theoretical values.
2. **Chi-Square Test of Independence:** It tests the hypothesis that **two categorical variables are related**. For example, a data scientist can apply this test to determine whether there is **any correlation** between customer satisfaction and gender.

Do these locations have the same underlying bird population?

Enter the Chi Square Test! A test for checking if two sets of categorical variables come from the same distribution.

Null hypothesis: ?

Alternative hypothesis: ?

Do these locations have the same underlying bird population?

Enter the Chi Square Test! A test for checking if two sets of categorical variables come from the same distribution.

Null hypothesis: The bird populations observed in Location A and Location B are the same.

Alternative hypothesis: The bird populations observed in Location A and Location B are different.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Multiple Groups

The Zoo management team decides to take 50 wild monkeys to study, but they don’t know what those monkeys would like to eat! They have five different food mixes they want to try. They split the monkeys up into five groups and feed each group a different mix, and then measure how much the monkeys grow over the next few years.

Ques: How do Zoo team know if the mixes have different effects?

Anova (Analysis of Variance) Test

ANOVA is a powerful statistical test for comparing the **means of multiple groups (three or more groups (more than two))** to determine if there are significant differences among them.

We use a **anova** test.

- Null hypothesis:
- Alternative hypothesis:

Anova Test

We use a **anova** test.

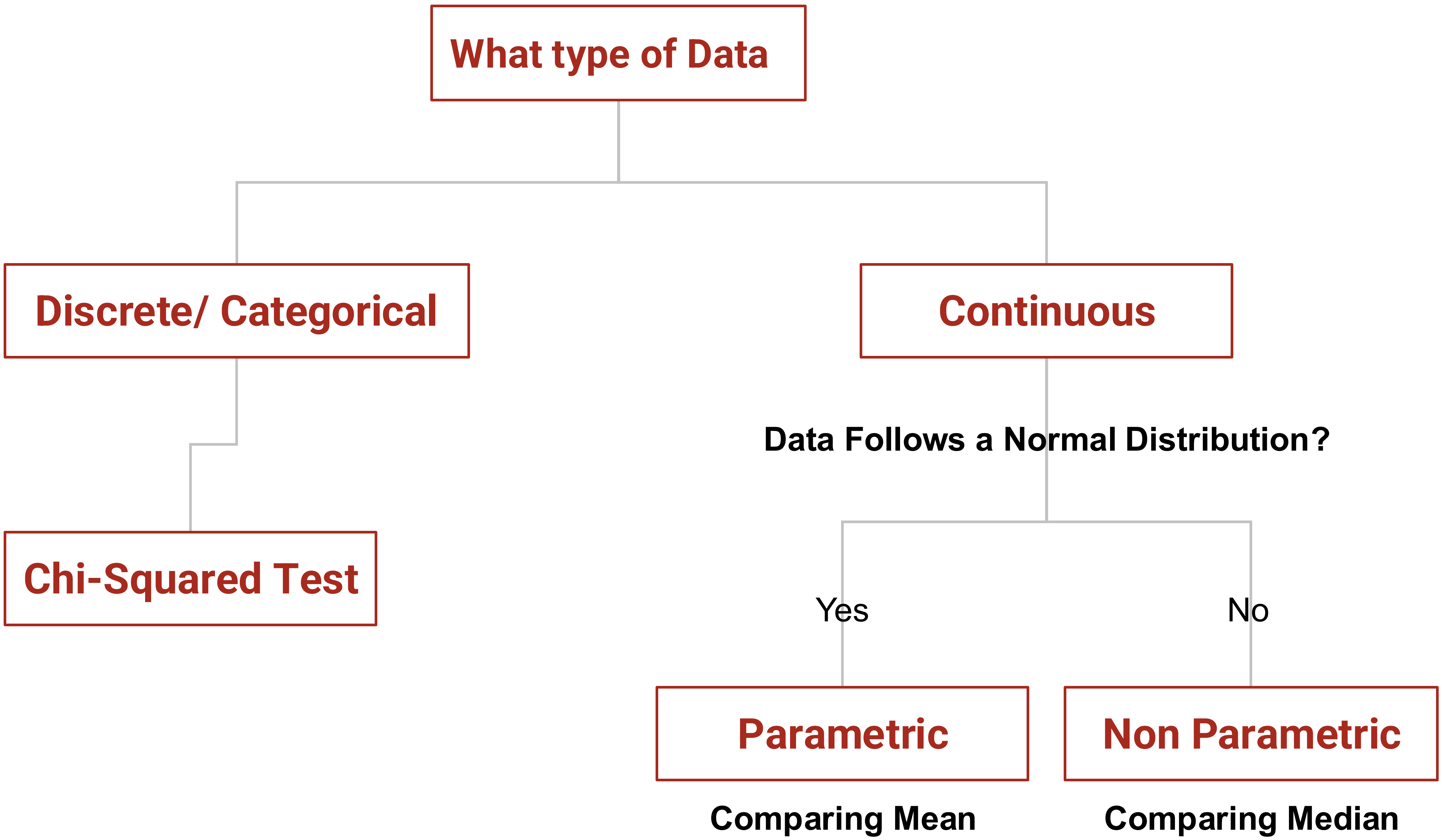
- **Null hypothesis:** There is no difference between any of the groups
- **Alternative hypothesis:** There is a difference between at least one of the groups

Notes: In t-tests and z-tests, we typically compare means of <u>two groups</u> using individual datasets or assess the mean of a single group against a known value. **ANOVA** evaluates differences in means across <u>three or more groups</u> as a whole, considering both within-group and between-group variability.

Analysis of variance, or ANOVA

- Analysis of variance, or ANOVA, is one of the most valuable statistical techniques for comparing the mean of three or more groups to test the null hypothesis that at least one group's means differ.
- **Types of ANOVA:**

1. **One-Way ANOVA:** Explores the significance of at least three unrelated groups about one independent variable.
2. **Two-Way ANOVA:** Analyse how two independent variables influence the dependent variable and determine if the two independent variables interact.
3. **MANOVA (Multivariate Analysis of Variance):** A more advanced type of ANOVA that allows for testing of multiple dependent variables at the same time.


What type of Data
Discrete/ Categorical
Continuous
Chi-Squared Test
Data Follows a Normal Distribution?
Yes
No
Parametric
Non Parametric
Comparing Mean
Comparing Median

Parametric Tests

These tests make strong assumptions about the population from which the data is drawn, including that the data follows a **normal distribution** and has **equal variances**.

Data Type: They require data to be on **an interval** or **ratio scale**.

Central Tendency: They use the **mean** to measure the central tendency of the data.

Advantages: Parametric tests are generally **more statistically powerful**, meaning they have a greater ability to detect significant differences and are less likely to result in a Type 2 error (failing to reject a false null hypothesis).

Examples: T-tests and Analysis of Variance (ANOVA).

Non-Parametric Tests

These tests make **fewer assumptions** about the data and population distribution.

Data Type: They are suitable for **ordinal data (ranked data) or nominal data**, and are appropriate when parametric test conditions are not met.

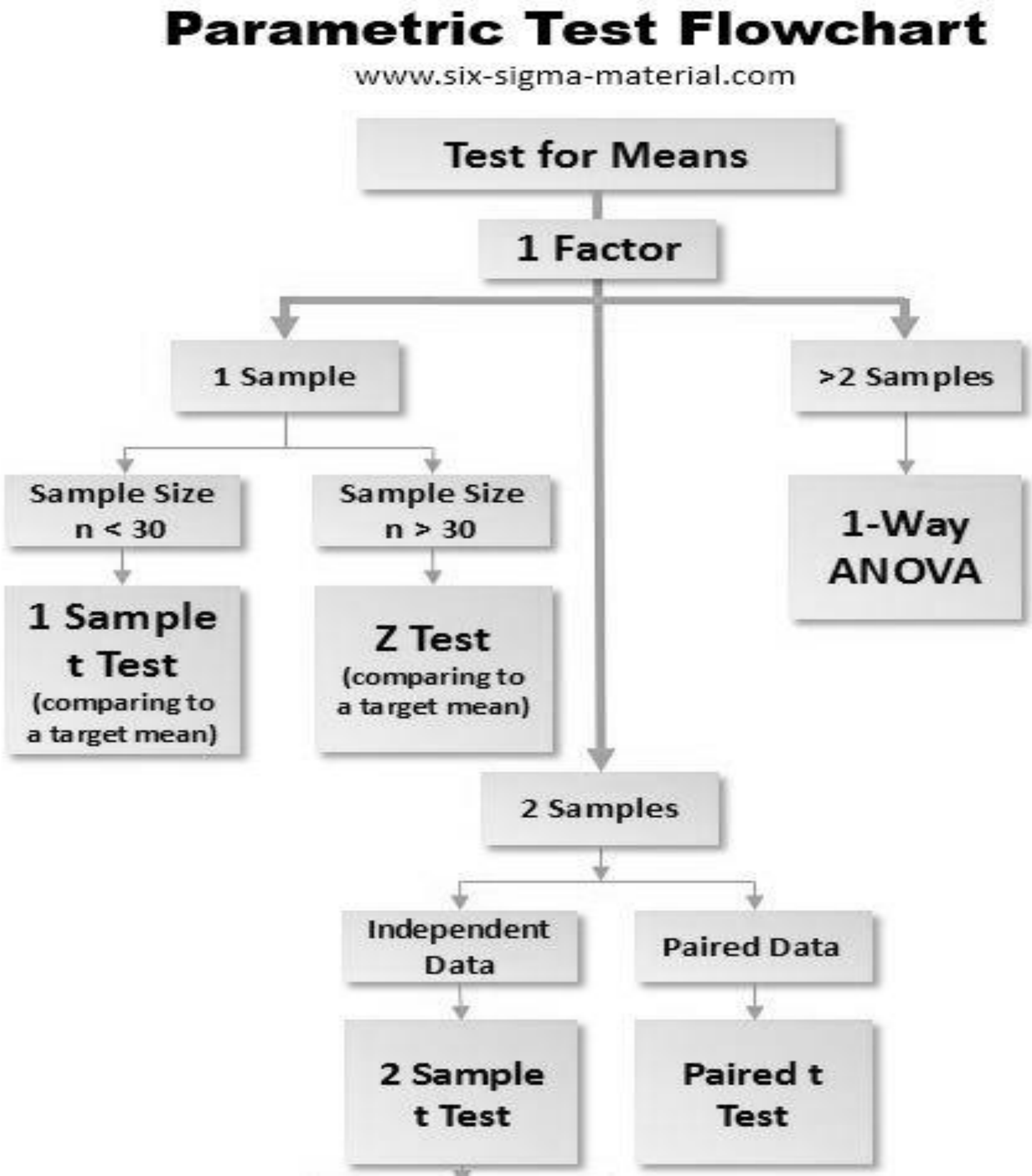
Central Tendency: They use the **median** as the measure of **central tendency**.

Advantages: They are useful for **small sample sizes, data with significant outliers, or when the assumption of normality cannot be met**.

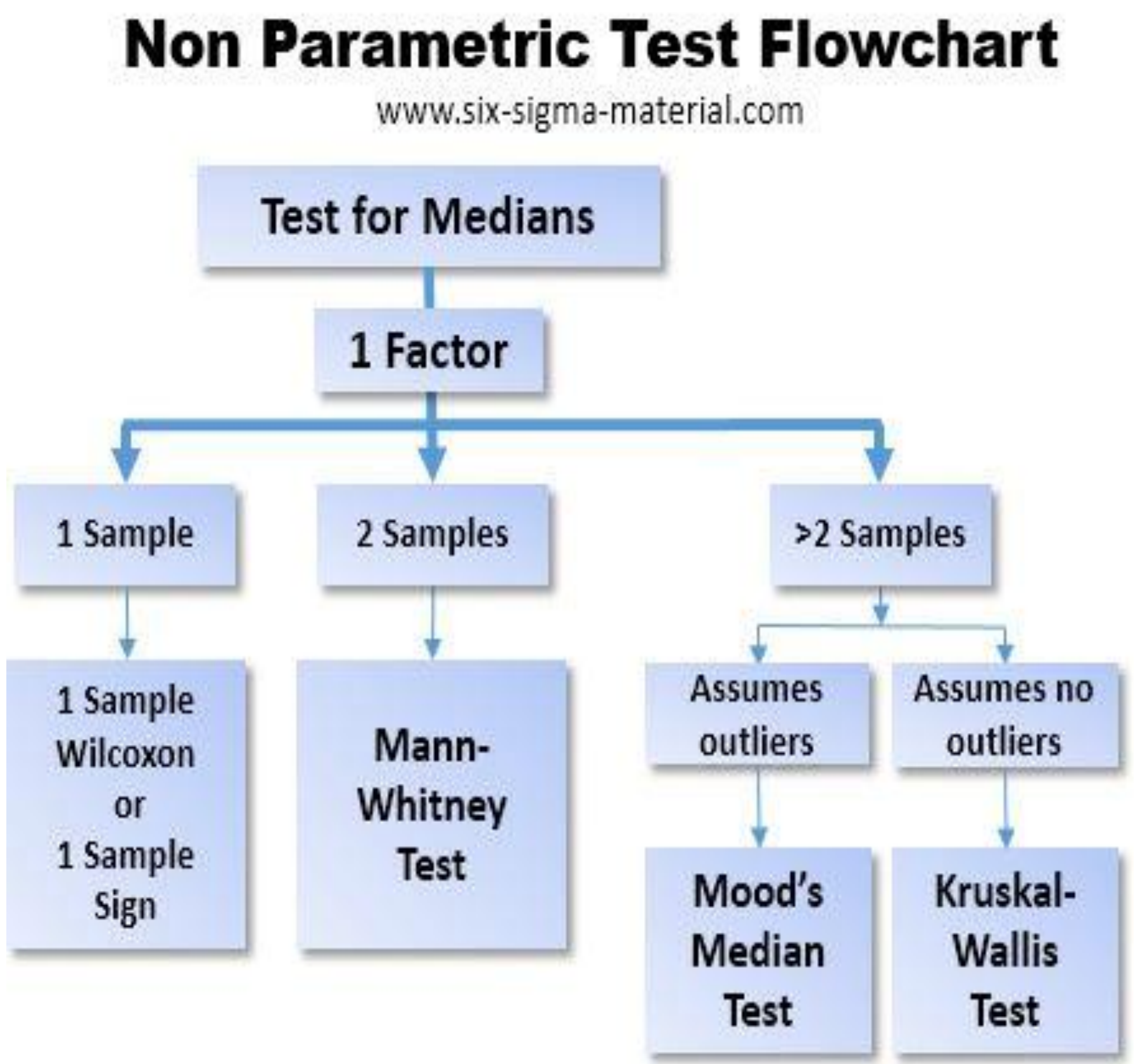
Disadvantages: They are often **less powerful** than their parametric counterparts, can be **less precise**, and are typically geared towards hypothesis testing rather than parameter estimation.

Examples: Mann-Whitney U test and Kruskal-Wallis test.

Parametric Tests (Comparing Means)



Nonparametric Tests (Comparing Medians)



Kruskal-Wallis Test

This nonparametric test is an extension of the one-way ANOVA and is used to compare the medians of more than two independent groups (three or more independent groups) when the data do not meet the assumptions of normality required for ANOVA.

Mann-Whitney U Test

This nonparametric test is used to

- compare two independent groups
- when the data do not meet the assumptions of normality required for a t-test.

Wilcoxon Signed-Rank Test

This nonparametric test is used to

- compare two related or paired groups.
- when the data do not meet the assumptions of normality required for a paired t-test.

Considerations: (How to decide an appropriate statistical test?)

- What are you curious about?
  - Mean? Standard deviation? Frequency?
- Is your data categorical or continuous?
- Do you have one or two samples?
- Is your data normally distributed?
  - If it is, you would use a **parametric** test. If it is *very* non-normal, you would use a non-parametric test
- Is your data paired? Is there a before and after?

Post Hoc Tests for ANOVA

Post Hoc Test/ Analysis

When you use ANOVA to find the main effect is significant, <u>indicate that the main effect is different between groups.</u>

However, you might want to know the **which group is different than the other group, that is called post-hoc analysis**.

Post Hoc Test/ Analysis

Definition: Post hoc analysis is the examination of data after an initial statistical test has been performed, <u>typically after rejecting the null hypothesis</u> in an analysis of variance (ANOVA) or a similar test.

- **Purpose:** It helps to **identify which specific group(s) are different from each other** when there are more than two groups, often in the context of multiple comparisons.

Post Hoc Test/ Analysis

Some common types of Post Hoc Test:

- Tukey's Honest Significant Difference (HSD)
- Bonferroni Correction
- Duncan's New Multiple Range Test

All of these tests identify which group(s) differ significantly from others

Example

Scenario: We conducted a study to compare the test scores of students from three different schools: School A, School B, and School C.

School	Sample Size	Mean Score	Standard Deviation
School A	20	85	8
School B	25	92	7
School C	18	88	9

More than 2 groups (A,B,C) → Apply ANOVA
ANOVA result: p-value < 0.05 (indicating significant overall differences among schools).

Example

Next Step: Applied Tukey's HSD post hoc test.

Comparison	p-value	Significant Difference
School A vs. School B	0.02	Yes
School A vs. School C	0.04	Yes
School B vs. School C	0.90	No

Interpretation from ANOVA: School A and School B, as well as School A and School C, have significantly different test scores ($p < 0.05$). But there is no significant difference in test scores between School B and School C.

Summary

- Hypothesis testing is a powerful tool for <u>making data-driven decisions</u>.
- It involves formulating hypotheses, collecting data, conducting statistical tests, and drawing conclusions.
- Be mindful of the significance level and the potential for Type I and Type II errors.
- Considering factors like sample size and assumptions enhances result interpretation.

Conclusion

If you are ever speculating about the properties of a population based on a sample, you **must do a hypothesis test**.

Figure out what you want to measure, look up the correct test, and perform it.

The p-value will tell you the probability that the null hypothesis is correct.

Recap:

Statistical tests let us reason about one more samples and how they relate to each other and the population.

- **z-test:** A test we can use if your data is normally distributed and we have a large number of samples
- **t-test:** A test we can use if your data is normally distributed and we have a small number of samples
- **Paired tests:** Used when we follow specific population members through time
- **One tailed vs two tailed tests:** Two tailed tests look for *any* difference in population; one tailed tests require us to pick a direction

Recap: Main Steps of Hypothesis Testing

1. **State the Null Hypothesis:** Assumption what you're trying to test
2. **State the Alternative Hypothesis:** what you believe might be true
3. **Pick a Level of Significance α:** the probability of rejecting the null hypothesis when it's actually true. Common values for α are 0.05 or 0.01.
4. **Choose a Test:** Select the right tool to check your guesses based on your data.
5. **Collect Data:** Get the information/data you need through observation or experimentation.
6. **Calculate a test statistic:** Using the collected data, calculate the appropriate test statistic.
7. **Calculate P-Value and compare with α:** Based on the comparison, decide if you have enough evidence to believe your guess is right or if you need to keep looking.
8. **Draw a Conclusion:** Based on your decision in the previous step, draw a conclusion regarding the null hypothesis. If you reject it, you accept the alternative hypothesis. If you fail to reject it, you do not have enough evidence to support the alternative hypothesis.