Statistics

The study of data: how to collect, analyze, summarize, and present it.

Types of statistics:

1. Descriptive statistics
2. Inferential statistics

Descriptive statistics

- Descriptive statistics is mainly for understanding data, simply the analysis of data that helps describe, show or summarize data in a meaningful way.
- Descriptive statistics is a branch of statistics that deals with the collection of data to give a precise information about which data is obtained.

Types of Descriptive Statistics

1) Measures of Central Tendency – Mean, Median and Mode

Mean - The mean is also known as "M" and is the most common method for finding averages. You get the mean by adding all the response values together, dividing the sum by the number of responses, or "N."

Mode - The mode is just the most frequent response value

Median - Finally, we have the median, defined as the value in the precise center of the dataset.

2) Measure of Variability (or dispersion)

- which shows how spread out the values are – Range, Variance and SD
- Dispersion refers to the spread of the values around the central tendency.

Range - Use range to determine how far apart the most extreme values are

SD - standard deviation (s) is your dataset's average amount of variability, showing you how far each score lies from the mean. Low standard deviation implies that most values are close to the mean. High standard deviation suggests that the values are more broadly spread out. The larger your standard deviation, the greater your dataset's variable. Follow these six steps:

1. List the scores and their means.
2. Find the deviation by subtracting the mean from each score.
3. Square each deviation.
4. Total up all the squared deviations.
5. Divide the sum of the squared deviations by N-1.
6. Find the result's square root.

Raw Number/Data	Deviation from Mean	Deviation Squared
4	4-7.3= -3.3	10.89
6	6-7.3= -1.3	1.69
7	7-7.3= -0.3	0.09
8	8-7.3= 0.7	0.49
8	8-7.3= 0.7	0.49
9	9-7.3=1.7	2.89
10	10-7.3= 2.7	7.29
M=7.3	Sum = 0.9	Square sums= 23.83

When you divide the sum of the squared deviations by 6 (N-1): 23.83/6, you get 3.971, and the square root of that result is 1.992. As a result, we now know that each score deviates from the mean by an average of 1.992 points.

Difference between Mean absolute deviation vs standard deviation

Variance - Variance reflects the dataset's degree spread. The greater the degree of data spread, the larger the variance relative to the mean.

Square root of variance is standard deviation

Inferential statistics

- Inferential differs from descriptive in that inferential stats makes guesses about the population whereas descriptive describes the subset.
- Inferential statistics is where you use characteristics of a sample to infer something about a population.
- Inferential statistics can be defined as a field of statistics that uses analytical tools for drawing conclusions about a population by examining random samples. The goal of inferential statistics is to generalize about a population. In inferential statistics, a statistic is taken from the sample data (e.g., the sample mean) that used to make inferences about the population parameter (e.g., the population mean).

Types of inferential statistics

1) Hypothesis Testing
2) Regression Analysis

Hypothesis Testing:

Hypothesis testing is a type of inferential statistics that is used to test assumptions and draw conclusions about the population from the available sample data. It involves setting up a null hypothesis and an alternative hypothesis followed by conducting a statistical test of significance. A conclusion is drawn based on the value of the test statistic, the critical value, and the confidence intervals.

Regression Analysis:

- Regression analysis is used to quantify how one variable will change with respect to another variable.
- Checks the effect of a unit change of the independent variable in the dependent variable

Difference between Descriptive and inferential statistics:

1) Descriptive statistics doesn’t infer any conclusions or predictions, which implies that inferential statistics do so.
2) Inferential statistics takes a random sample of data from a portion of the population and describes and makes inferences about the entire population. For instance, in asking 50 people if they liked the movie they had just seen, inferential statistics would build on that and assume that those results would hold for the rest of the moviegoing population in general.
3) Therefore, if you stood outside that movie theater and surveyed 50 people who had just seen Rocky 20: Enough Already! and 38 of them disliked it (about 76 percent), you could extrapolate that 76% of the rest of the movie-watching world will dislike it too, even though you haven’t the means, time, and opportunity to ask all those people.
4) Simply put: Descriptive statistics give you a clear picture of what your current data shows. Inferential statistics makes projections based on that data.

	Descriptive Statistics	**Inferential Statistics**
1	It gives information about raw data which describes the data in some manner.	It makes inference about population using data drawn from the population.
2	It helps in organizing, analyzing and to present data in a meaningful manner.	It allows us to compare data, make hypothesis and predictions.
3	It is used to describe a situation.	It is used to explain the chance of occurrence of an event.
4	It explains already known data and limited to a sample or population having small size.	It attempts to reach the conclusion about the population.

Types of sampling:

- **Population**: The whole group we are interested in
- **Census**: A collection of data from the whole population
- **Sample**: A collection of data from part of the population

But how do we choose what members of the population to sample?

There are 4 main methods:

1) **Random Sampling**

- The best way is to choose randomly
- Imagine slips of paper each with a person's name, put all the slips into a barrel, mix them up, then dive your hand in and choose some slips of paper.
- But this means you need a full list of the population to choose from.

  Example: You want to know the favorite colors for people at your school, but do not have the time to ask everyone.

- Somehow get a full list of students printed out, then place all pages on the ground, drop a pencil and note down the student's name.
- Repeat until you have 50 names. Now survey those 50.Your results will hopefully be nearly as good as if you had asked everyone.
- Random surveys are the best way to avoid bias.
- And your results are better when you ask more people.

  Example: nationwide opinion polls survey around 2,000 people, and the results are nearly as good (within about 1%) as asking everyone.

2) **Systematic Sampling**

- This is where we follow some system of selection like "every 10th person"

  Example: You want to know the favorite colors for people at your school, but don't have the time to ask everyone.

Solution: stand at the gate and choose "every 4th person to arrive"

3) **Stratified Sampling**

- This is where we divide the population into groups by some characteristic such as age or occupation or gender.
- Then make sure our survey includes people from each group in proportion to how many there are in the whole population.

Example: We want to survey 300 people in the USA

This is the population breakdown for the USA in 2010:

Age Range	Percent
0-4	6.50%
17-May	17.50%
18-23	9.90%
24-44	26.60%
45-64	26.40%
65+	13.00%
	100%

We want to survey 300 people, so we choose:

Age Range	Percent	People
0-4	6.50%	20
17-May	17.50%	52
18-23	9.90%	30
24-44	26.70%	80
45-64	26.40%	79
65+	13.00%	39
	100%	300

4) **Cluster Sampling**

- We break the population into many groups, then randomly choose whole groups.
- Example: we divide the town into many different zones, then randomly choose 5 zones and survey everyone in those zones.
- Cluster sampling works best when the clusters are similar in character to each other.

Example: if the town has rich and poor zones then try to create a new way of dividing the town into fairer regions.

Quiz

1. Mrs. Trahan samples her class by selecting 5 girls and 7 boys. This type of sampling is called.
2. Mrs. Trahan samples her class by selecting every third person on her class list. Which type of sampling method is this?
3. Mrs. Trahan samples her class by selecting all students sitting at group 1 and group 5 in her classroom. This sampling technique is called.
4. Mrs. Trahan samples her class by picking 10 numbers from her hat and each number is assigned to a student.
5. In order to use samples to estimate something from the population, the sample should be _________________ the population.
6. If every individual in a population has the same chance of being included in a sample, the sample is a(n) _________ sample.
7. The school librarian wants to determine how many students use the library on a regular basis. What type of sampling method would she use if she chose to randomly select 20 students in the cafeteria during each of the three lunch periods on Monday.
8. A student wants to determine the favorite professional sport of students in her high school. Which of the following samples is most likely to giver her a representative sample?

   answer choices

   A random Sample of the students in the art honor society

   A random sample of students on the basketball team

   A random sample of students on the official school enrollment roster

   A random sample of students in the library during lunch
9. To determine the most popular type of cake amount all the students at your school, ask your entire math class.

What are variables:

A variable is any characteristics, number, or quantity that can be measured or counted. A variable may also be called a data item. Age, sex, business income and expenses, country of birth, capital expenditure, class grades, eye colour and vehicle type are examples of variables.

Types of variables:

1) Quantitative
2) Qualitative

What is quantitative data?

- Quantitative data refers to any information that can be quantified — that is, numbers. If it can be counted or measured, and given a numerical value, it's quantitative in nature. Think of it as a measuring stick.
- Quantitative variables can tell you “How many," "how much," or "how often."

  Some examples of quantitative data:

- How many people attended last week's webinar?
- How much revenue did our company make last year?
- How often does a customer rage click on this app?

What is qualitative data?

- Unlike quantitative data, qualitative data is descriptive, expressed in terms of language rather than numerical values.
- Qualitative data analysis describes information and cannot be measured or counted. It refers to the words or labels used to describe certain characteristics or traits.

Difference between quantitative and qualitative data –

- Quantitative data is numbers-based, countable, or measurable. Qualitative data is interpretation-based, descriptive, and relating to language.
- Quantitative –
  - Expressed in numbers, Math/ statistical analysis
  - If your goal is to confirm or test a theory or hypothesis, opt for quantitative.
- Qualitative –
  - Expressed in words, summarize, categorize and interpret
  - If you want to understand or explore an idea, then opt for qualitative

Types of Quantitative data:

1) Discrete data is counted
2) Continuous data is measured.

Discrete data

- Example: the number of students in a class
  - We can't have half a student!
- Example: the results of rolling 2 dice

- Only has the values 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12

Continuous Data

Continuous Data can take any value (within a range)

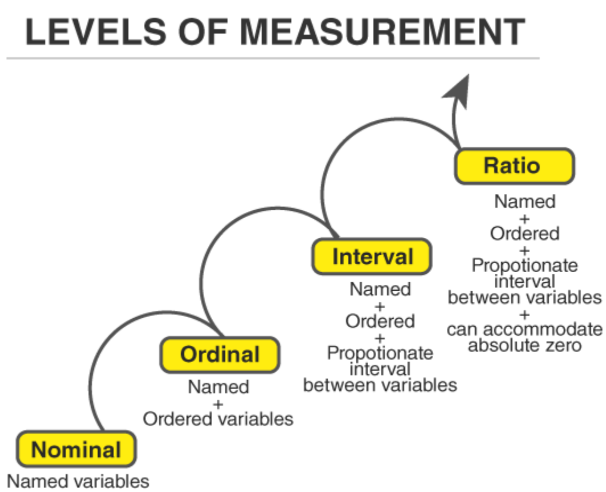
Examples:

- A person's height: could be any value (within the range of human heights), not just certain fixed heights,
- Time in a race: you could even measure it to fractions of a second,
- A dog's weight,
- The length of a leaf,

Quiz:

- 7 dogs is qualitative or quantitative?
- If cat fur is black
- 6 ft inches

Level of measurements:



1) Nominal Scale:

- Nominal variables (also called categorical variables) can be placed into categories. They do not have a numeric value and so cannot be added, subtracted, divided or multiplied. They also have no order; if they appear to have an order then you probably have ordinal variables instead.

Example:

An example of a nominal scale measurement is given below:

What is your gender?

M- Male

F- Female

Here, the variables are used as tags, and the answer to this question should be either M or F.

- Cannot be quantified. In other words, you can't perform arithmetic operations on them, like addition or subtraction, or logical operations like "equal to" or "greater than" on them.
- Cannot be assigned any order.

2) Ordinal Scale.

- The ordinal scale contains things that you can place in order. For example, hottest to coldest, lightest to heaviest, richest to poorest. Basically, if you can rank data by 1st, 2nd, 3rd place (and so on), then you have data that's on an ordinal scale.
- Ordinal scales report the ordering and ranking of data without establishing the degree of variation between them. Ordinal represents the "order." Ordinal data is known as qualitative data or categorical data. It can be grouped, named and also ranked.
- The ordinal scale contains things that you can place in order. It measures a variable in terms of magnitude, or rank.
- Ordinal scales tell us relative order but give us no information regarding differences between the categories.

Example:

- Ranking of school students – 1st, 2nd, 3rd, etc.
- Ratings in restaurants
- Evaluating the frequency of occurrences
  - Very often
  - Often
  - Not often
  - Not at all

3) Interval Scale

- An interval scale has ordered numbers with meaningful divisions. Temperature is on the interval scale: a difference of 10 degrees between 90 and 100 means the same as 10 degrees between 150 and 160. Compare that to high school ranking (which is ordinal), where the difference between 1st and 2nd might be .01 and between 10th and 11th .5. If you have meaningful divisions, you have something on the interval scale.

- It is defined as a quantitative measurement scale in which the difference between the two variables is meaningful. In other words, the variables are measured in an exact manner, not as in a relative way in which the presence of zero is arbitrary.

- An interval scale has ordered numbers with meaningful divisions, the magnitude between the consecutive intervals are equal.

-

- For example, temperature on Fahrenheit/Celsius thermometer i.e. 90° are hotter than 45° and the difference between 10° and 30° are the same as the difference between 60° degrees and 80°

- For instance, the distance between 29 and 30 degrees is the same as the distance between 99 and 100 degrees.

- However, due to the absence of absolute zero, you cannot tell by how much the temperature is higher or lower. For example, you cannot say if 40 degrees is twice hot as 20 degrees or if – 20 degrees is half as cold as -40 degrees.

4) Ratio Scale

- The ratio scale is the same as the interval scale with one major difference: zero is meaningful. For example, a height of zero is meaningful (it means you don't exist). Compare that to a temperature of zero, which while it exists, it doesn't mean anything in particular (although admittedly, in the Celsius scale it's the freezing point for water).

- which is quantitative. It is a type of variable measurement scale. It allows researchers to compare the differences or intervals. The ratio scale has a unique feature. It possesses the character of the origin or zero points.

- Along with all the other values, in a Kelvin scale, the zero point has a relevant meaning. For instance, you can tell on a Kelvin scale that 40K is twice hot as 20K.

- Also, the presence of absolute zero in a Kelvin scale means that nothing can be colder than OK. This is because on a ratio scale there can be no negative number.

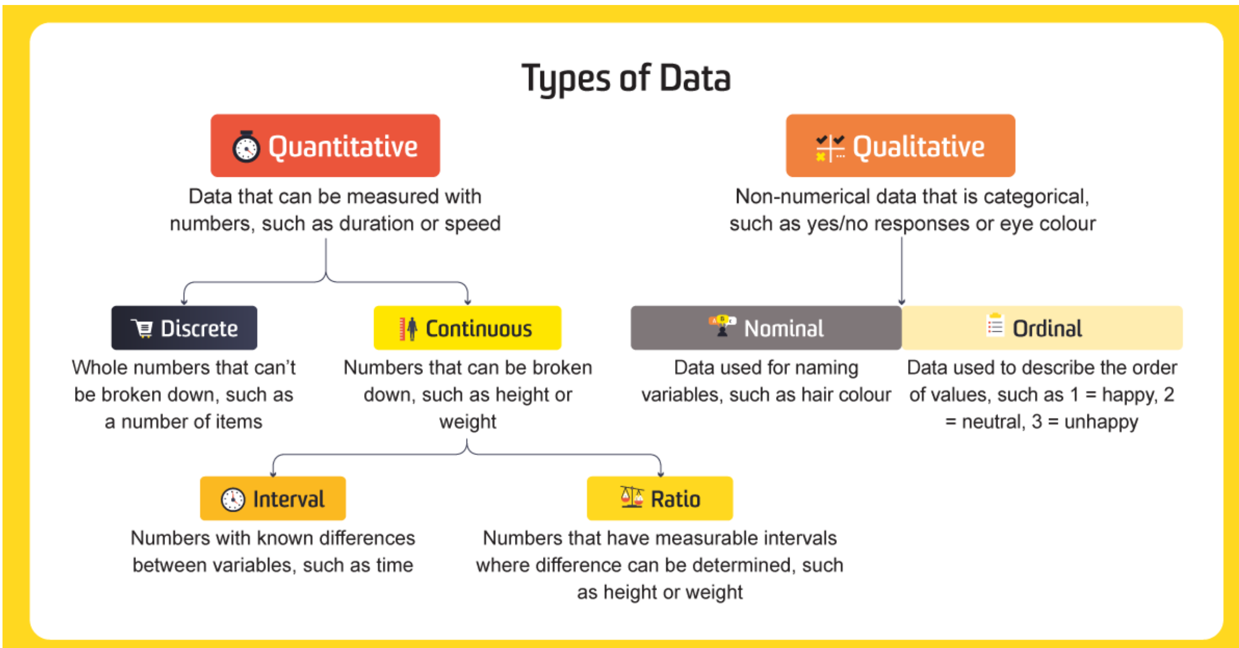
Difference between interval and ratio:

- Time and duration are two examples of interval and ratio scale respectively.
    - Time is the value of the interval scale because there is no zero. You cannot tell when time started.

- Duration is a case of ratio scale for the fact that duration has a starting point. The zero in duration has a meaningful presence. You can tell that 20 days is twice of 10 days.

- A ratio scale can measure any data that has “zero points” characteristics. A ratio scale is ideal for measuring age, weight, height, etc. In marketing research it can calculate sales, shares, the volume of the customer, etc.

- An interval scale is mostly used to gather feedback based on agreement, satisfaction level, or likelihood. It is commonly found in question-type surveys where the choice of options is so scaled that a numerical value can be allotted to them, in order to calculate.

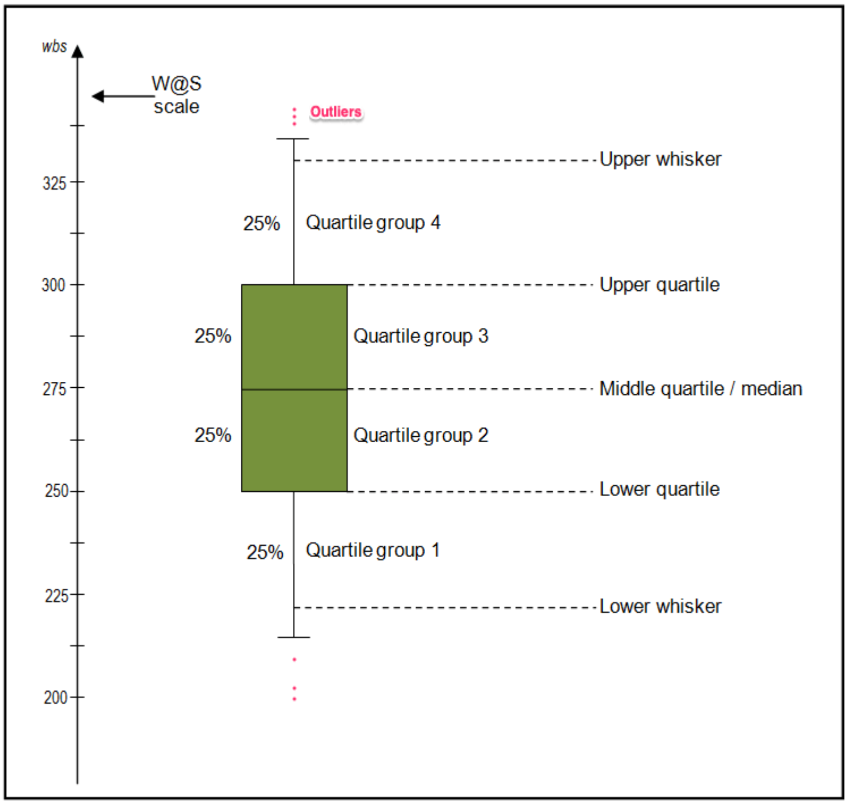


Quiz:

1. Example 1: The order in which athletes cross the finish line in a race is an example of which of the scales of measurement?
2. Example 2: Amount of calories in a pack of cheese is an example of which of the measurement scales?
3. Example 3: Children in a primary school are evaluated and classified as #0 - non-readers, #1 - beginners, #2 - grade level readers, or #3 - advanced readers. Which of the four scales of measurement is used in this evaluation?
4. “Places from where students are travelling to take a statistics course” is an example of which scale of measurement?
5. “Students’ scores on a biology test” is an example of which scale of measurement?
6. “Number of cars” is an example of which scale of measurement?
7. Calendar year is an example of what scale of measurement?
8. "Amount of calories in a small Al Marai Yogurt" is which scale of measurement?
9. shades of lipstick available in A MAC store, is which scale of measurement?
10. Your shoe size, is an example of which scale of measurement?
11. Arranging the shirt sizes as small, medium and large is an example of which scale of measurement?
12. Which level of measurement is used to measure the size of these different size M&Ms?
13. Which level of measurement is used to measure the national debt?

Quartiles:

Quartiles are values that separate the data into four equal parts.



Percentiles

Percentiles are values that separate the data into 100 equal parts.

For example, the 95th percentile separates the lowest 95% of the values from the top 5%

Percentile reflects the percentage of students who have scored less than you in any exam.

So if your percentile is 99 in an exam taken by 5000 students, it means that 99% of 5,000 students i.e. 4,950 students are below you. Hence, it also means that you are at 50th rank from the top.

There are 100 people in exam, Suppose your rank is 21 in the exam, then your percentile score is

Difference between percentiles and quartiles:

Percentiles: Range from 0 to 100.

Quartiles: Range from 0 to 4.

Note that percentiles and quartiles share the following relationship:

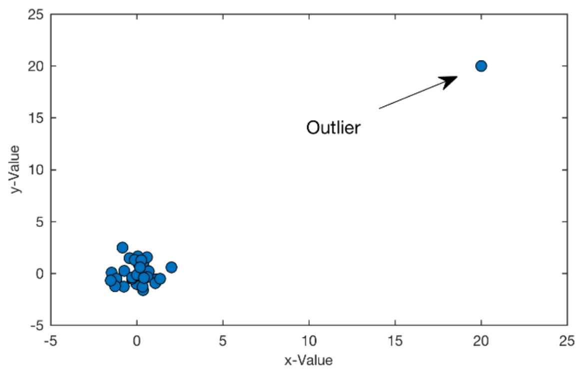
- 0 percentile = 0 quartile (also called the minimum)
- 25th percentile = 1st quartile
- 50th percentile = 2nd quartile (also called the median)
- 75th percentile = 3rd quartile
- 100th percentile = 4th quartile (also called the maximum)

- First quartile: the lowest 25% of numbers
- Second quartile: between 0% and 50% (up to the median)
- Third quartile: 0% to 75%
- Fourth quartile: the highest 25% of numbers
- Q1 is the central point between the smallest score and the median. Q1 tells us that 25% of the scores are less than x value and 75% of the class scores are greater
- Q3 is the middle value between Q2 and the highest score. Q3, the 75th percentile, reveals that 25% of the scores are greater and 75% are less than x value.

Difference between percentiles and Percentages

Outlier Detection

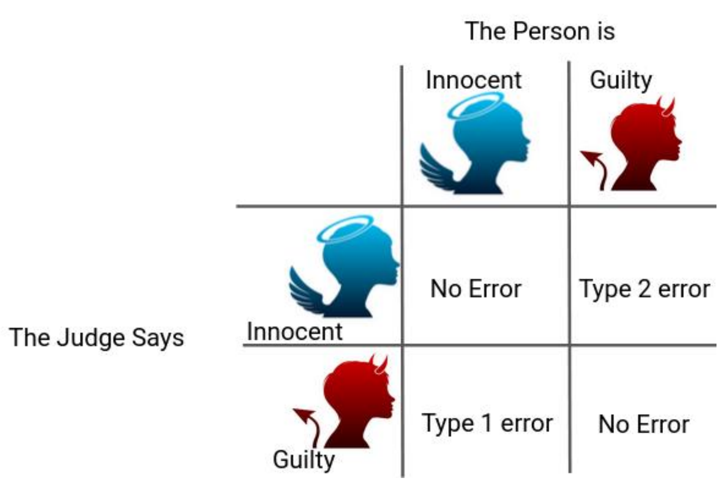
- A data set should be checked for extremely high or extremely low values. These values are called outliers.
- An outlier is an extremely high or an extremely low data value when compared with the rest of the data values.
- An outlier can strongly affect the mean and standard deviation of a variable. For example, suppose a researcher mistakenly recorded an extremely high data value. This value would then make the mean and standard deviation of the variable much larger than they really were.
- Outlier is a extreme value of data. data points are follow certain pattern , some values are not include the group the values are out of from the group is called the outlier



Hypothesis Testing:

Let's take an example to understand the concept of Hypothesis Testing. A person is on trial for a criminal offense and the judge needs to provide a verdict on his case. Now, there are four possible combinations in such a case:

- First Case: The person is innocent and the judge identifies the person as innocent
- Second Case: The person is innocent and the judge identifies the person as guilty
- Third Case: The person is guilty and the judge identifies the person as innocent
- Fourth Case: The person is guilty and the judge identifies the person as guilty



As you can clearly see, there can be two types of error in the judgment – Type 1 error, when the verdict is against the person while he was innocent and Type 2 error, when the verdict is in favor of Person while he was guilty

According to the Presumption of Innocence, the person is considered innocent until proven guilty. That means the judge must find the evidence which convinces him “beyond a reasonable doubt”. This phenomenon of **“Beyond a reasonable doubt”** can be understood as **Probability (Judge Decided Guilty | Person is Innocent) should be small.**

The basic concepts of Hypothesis Testing are actually quite analogous to this situation.

We consider **the Null Hypothesis** to be true until we find strong evidence against it. Then. we accept the **Alternate Hypothesis**. We also determine the **Significance Level (α)** which can be understood as the Probability of (Judge Decided Guilty | Person is Innocent) in the previous example. Thus, if α is smaller, it will require more evidence to reject the Null Hypothesis. Don’t worry, we’ll cover all of this using a case study later.

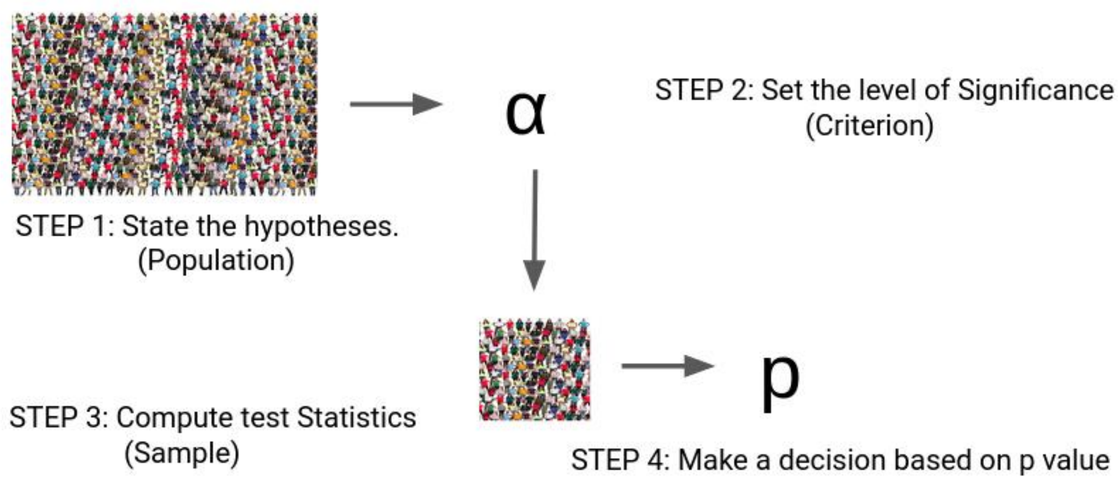
In inferential statistics

Truth about Population / Decision based on sample	Null Hypothesis (H_0)	Alternative Hypothesis
Null Hypothesis (H_0)	No error (1 - α)	Type 2 error
Alternative Hypothesis (H_1)	Type 1 error (α)	No error

Steps to Perform Hypothesis testing

There are four steps to perform Hypothesis Testing:

- Set the Hypothesis
- Set the Significance Level, Criteria for a decision
- Compute the test statistics
- Make a decision

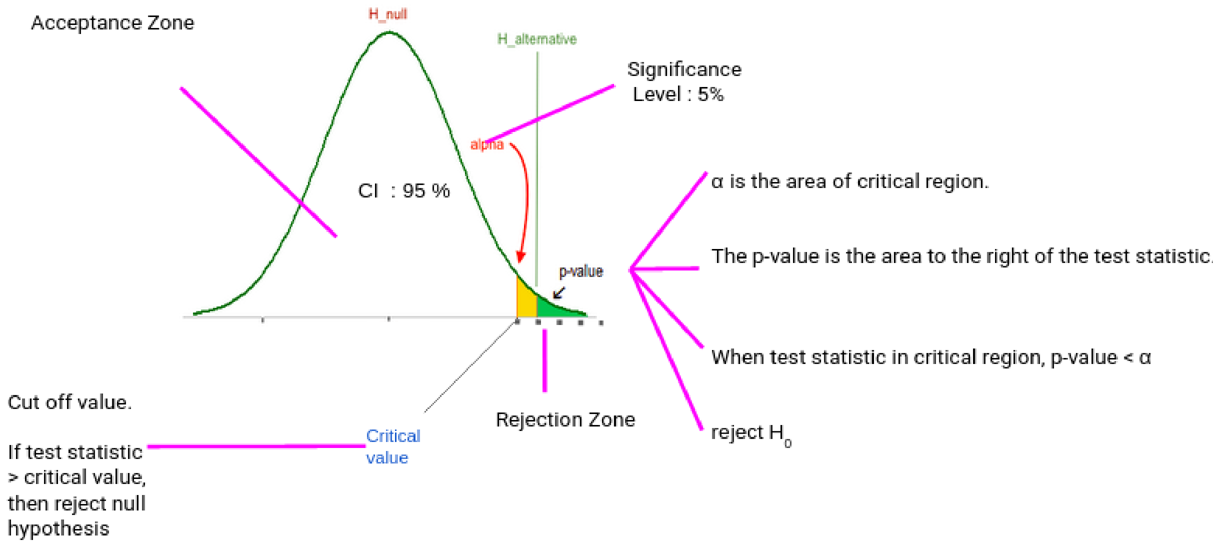


Steps 1 to 3 are quite self-explanatory but on what basis can we make a decision in step 4? What does this p-value indicate?

We can understand this p-value as the measurement of the Defense Attorney's argument. If the p-value is less than α , we reject the Null Hypothesis or if the p-value is greater than α, we fail to reject the Null Hypothesis.

Critical Value, p-value

Let's understand the logic of Hypothesis Testing with the graphical representation for Normal Distribution.



Typically, we set the Significance level at 10%, 5%, or 1%. If our test score lies in the Acceptance Zone we fail to reject the Null Hypothesis. If our test score lies in the critical zone, we reject the Null Hypothesis and accept the Alternate Hypothesis.

Critical Value is the cut off value between Acceptance Zone and Rejection Zone. We compare our test score to the critical value and if the test score is greater than the critical value, that means our test score lies in the Rejection Zone and we reject the Null Hypothesis. On the opposite side, if the test score is less than the Critical Value, that means the test score lies in the Acceptance Zone and we fail to reject the null Hypothesis.

But why do we need p-value when we can reject/accept hypotheses based on test scores and critical value?

p-value has the benefit that we **only need one value** to make a decision about the hypothesis. We don't need to compute two different values like critical value and test scores. Another benefit of using p-value is that we can test at **any desired level of significance** by comparing this directly with the significance level.

This way we don't need to compute test scores and critical value for each significance level. We can get the p-value and directly compare it with the significance level.

P value is the probability for null hypothesis to be true

P =0.01 – 10 out of 100, null hypothesis to be true

P value > level of significance, Null hypothesis is accepted

P value <level of significance, Null hypothesis is rejected

How to understand p-value in layman terms?

One of the most important concepts in Statistics is the interpretation of the p-value. I would like to share my understanding of p-value in simpler terms.

Suppose a company claims that chocolate 'X' it produces has 70 gm of nut in 1 bar of chocolate. But some customers complain that the chocolate bars have fewer nuts than 70gm. We want to test whether the company's claim about the mean weight of nuts in chocolate X is correct. Let's try to work through this problem.

We take a sample of 20 chocolate bars and find the mean amount of nuts in those bars(sample mean) to make inferences about the population mean. Difference between population and sample? There are different ways of sampling from the population. Here we take a random sample of bars available at the nearest store.

The population mean (not known to us) is the mean of nuts contained in all the bars produced by the company — which in reality is not feasible to calculate. We hypothesize a value for it which is 70gm (hypothesized mean — what the company has claimed) in our example.

Suppose we get a nut mean of 68.5 for the 20 bars we sampled. Can we conclude that the customer claims are true? Does this provide enough evidence that the bars are short of nuts or could this result just be from luck? To answer this we do Hypothesis Testing.

Let’s state the null and alternate hypotheses.

Null hypothesis: The population mean of nuts in chocolate bars is 70gm.(This is the thing we are trying to provide evidence against.)

Alternate hypothesis: The population mean of nuts in chocolate bars is less than 70 gm. (This is what we are trying to prove)

The rule says we reject the null hypothesis if the p-value is less than the significance level(alpha). Next question:

What is the level of significance and p-value?

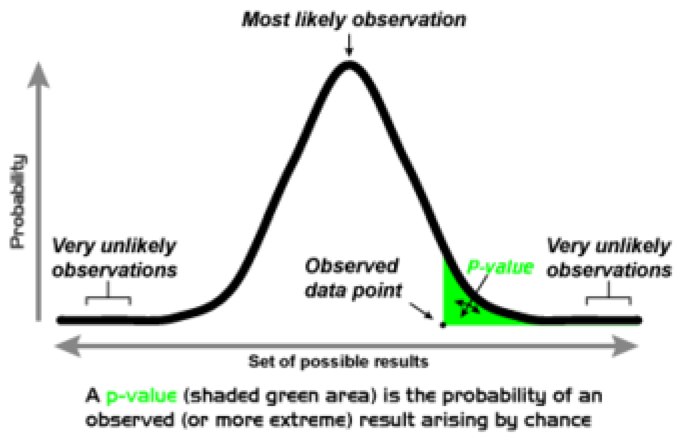
The level of significance(alpha) is the percentage of risk we are willing to take while rejecting the null hypothesis.

P-value is the probability that a random chance generated the data or something else that is equal or rarer (under the null hypothesis). We calculate the p-value for the sample statistics(which is the sample mean in our case). We can do it manually by looking at the z-table or use some statistical software to compute it.

Let’s say the true population of chocolates does have a nut mean of 70gm. This does not necessarily imply that if there exits 1 million chocolate of type X in the world, each chocolate will have a nut mean of 70 gm. There might be few chocolates that have a mean less than 70gm and some that have mean greater than 70gm. So when we sample out a few chocolates the sample mean might be less than, equal to or greater than 70gm due to sampling variation.

Coming back to the interpretation of p-value.

“The p-value is used to determine if the outcome of an experiment is statistically significant. A low p-value means that assuming the null hypothesis is true, there is a very low likelihood that this outcome was a result of luck. A high p-value means that assuming the null hypothesis is true, this outcome was very likely”



p-value — area of the shaded region under the Normal curve

Why do we reject the null hypothesis if we p-value < alpha (Let’s say 0.05)?

Suppose we get p-value as 0.04. This means under the assumption that population nut mean is 70 gm, we will get a value of sample mean (which is 68.5 in our case) 4 out of 100 times if we re-run the test under the null hypothesis. Very Unlikely! So, We might say — the result suggests that the sample probably was derived from different parent populations.

Hence we reject the null hypothesis. In this case, we have strong evidence again the null.

Getting something that rare (with chances 4/100) might be very unlikely but there are still chances that it happened by luck.

Let’s try to understand our basis of rejecting the null — why did we conclude the sample was not derived from the population under null hypothesis. Suppose you got 2 lottery tickets and randomly shuffled them. One ticket was from an event X with 3,00,000 participants and 1 prize. The other ticket was from an event Y with 30 participants with 10 prizes. You won the lottery on one ticket but you don’t know which event it belonged to. Won’t your guess be: event Y, because there are more chances you could have won when there were more prizes and fewer people. But suppose your winning ticket was from event X! Hurray!!! Luck played a role and you got it!

So when we say an event is very unlikely, there are still chances that it happened by luck and that is when we stumble on Type I error — Rejecting the null when it was true. Failure to reject a false null hypothesis is called Type II error.

One of the best examples to remember the definition of type I and type II errors. (Ref. Google Images)

Coming back to our definition as it will make more sense now. Level of significance (Alpha) — the percentage of risk we are willing to take while rejecting the null hypothesis or the Type I error rate.

One thing we all are well versed is about the fact that when your p-value is greater than 0.05 we fail to reject our null hypothesis, while when it is less than 0.05 we reject the null hypothesis. (At 5 % level of significance.)

So what is the idea behind this?

Let us take an example of an experiment which we are all familiar with. We order food from Zomato and it claims that your food will be delivered within 45 minutes.

Here our null hypothesis is an ideal scenario where the food will be delivered within 45 minutes.

Ho: The food delivery time of Zomato is equal to 45 minutes.

But that is not the case for some of us. Why? Due to external factors such as traffic jams, rains, technical error, delay from the restaurant, delay due to wrong/longer routes, etc, there might be incidents wherein many of us have experienced delivery times greater than 45 minutes.

So our Alternative Hypothesis would be:

H1: The food delivery time of Zomato is greater than 45 minutes.

So how often these external factors distort the delivery time claimed by Zomato? How many times has it gone beyond 45 minutes? Is that number significant? How are we going to decide upon its significance? Is the claim even true for about even 100 or 1000 of the population? How do we trust this claim?

We run an experiment on a sample of customers ordering from Zomato. This sample should be such that it summarizes our true population in terms of various factors such as location, age groups of delivery boys who fulfilled the order, experience of delivery boys, time of order, number of orders per customers, etc. All in all our sample should not be biased and should not contain only delivery times which are closer to 45 minutes.

Let us say we ran the experiment on a sample of 10,000 customers who had ordered from Zomato in the last year. We take a 5 % level of significance, in other words 0.05 as our cut off to accept or reject our claim. (If you want to be stricter with your experiment you can choose even 0.001(0.1%) and if you want to be lenient you can take the cut off as 0.1 (10%).)

After running the experiment (meaning statistical tests such as Z-test, t-test, ANOVA, etc) we observe one of the following scenarios:

Lets take the P-value definition which says ‘The probability under the assumption of no effect/difference of obtaining a result which is equal to or extreme than what was actually observed. Here no difference indicates the Null Hypothesis (Delivery time=45) while extreme result is Alternative Hypothesis (Delivery time>45).

P-value as 0.03. As it is less than 0.05 we reject Ho. Only 3 % of the samples(300 out of 10000 customers) had delivery times equal to 45 minutes which is far below the cut off you have set which is 5 % (Level of Significance=0.05).

It is very less likely that the observed difference of delivery time being greater than 45 mins is due to chance, therefore we have sufficient evidence from our sample to reject the claim. Your result is statistically significant.

Hence, you have stronger evidence against the null hypothesis, implying Zomato’s claim is false and therefore delivery times are greater than 45 minutes.

P-value as 0.98 As it is greater than 0.05 we fail to reject Ho. 98 % of the samples (9800 out of 10000 customers) had delivery times equal to 45 minutes which is far above the cut off you have set which is 5 % (Level of Significance=0.05).

It implies no difference in the delivery times as claimed by Zomato other than due to chance. Your result is statistically not significant. (However, it could be possible that this result would have been due to chance , one of the reason being your sampling methodolgy is not correct and you might be sampling from a biased population with delivery times equal to 45 mins. Or if your sample is not biased, then woah, Zomato’s claim is true and their services are really awesome!)

Hence, you have weak evidence against the null hypothesis, implying Zomato’s claim is true and delivery times are indeed equal to 45 minutes.

Suppose a child in the family goes to the school daily and one day, his teacher writes to his mother in the school diary that your son is very naughty and he was found fighting with another kid. This situation is quite common in schools and like any mother, this child's mother says no my son is not naughty, it must have been the other kid who provoked him. Simple...agreed?

After some days, this happens again with another kid and, again the mother of the kid in question does not accept the fault of her child.

The situation repeats itself again with our hero kid the third time with a third kid and now the mother becomes suspicious as to probably her child is really naughty.

Interestingly, this third instance in our example actually becomes what is Statistically called the Threshold of Significance (or the level of significance).

When the same child again is reported to fight with another kid, the mother has no other option but to accept that her child is really naughty and he finds ways to fight with other kids. This is actually when it is said that the evidence is "Statistically Significant". This is accepted to be significant since it has occurred beyond the level of significance.

Please remember, up to two complaints, the mother does not accept the blame but after certain repetitions, she has to accept it. Statistically, when we design an experiment, we have to fix a level of significance before commencing the experiment and then take trials and see what is the result. In many life situations and tests, a level of significance is fixed at 5%. When some results are found to be occurring more than 5% of the time, it is said that the result is significant. From here comes the concept of p-value and a p-value of less that .05 is said to be significant and accordingly results are interpreted.

Please allow me here to slightly touch base with the definition or explanation of the p-value. I hope that you are aware that the p-value is the probability of finding any result under the test conditions for a given set of data presuming that the null hypothesis were true. The p-value has to be smaller than the significance level because that gives us an indication that the probability of that result happening is very low (in most cases, less than 5%) by chance only and that is why it is most probably the real state of the world.

what is Significance value?

The significance level, also denoted as alpha ,is the probability of rejecting the null hypothesis when it is true. For example, a significance level of 0.05 indicates a 5% risk of conducting that a difference exists when there is no actual difference.

T test

- T Test (Students T Test) is a statistical significance test that is used to compare the means of two groups and determine if the difference in means is statistically significant.
- Let's suppose, you have measurements from two different groups, say, you are measuring the marks scored by students from a class you attended special coaching versus those who did not. And you want to determine if the scores of those who attending the coaching is significantly higher than those who did not. You can use the T-Test to determine this.
- Another example - If you want to test a car manufacturer's claim that their cars give a highway mileage of 20kmpl on an average. You sample 10 cars from the dealership, measure their mileage and use the T-test to determine if the manufacturer's claim is true.
- However, depending on how the groups are chosen and how the measurements were made you will have to choose a different type of T-Test for different situations.
- This test assumes that the test statistic, which is often the 'sample mean', should follow a normal distribution

t-tests are a statistical way of testing a hypothesis when:

- We do not know the population variance

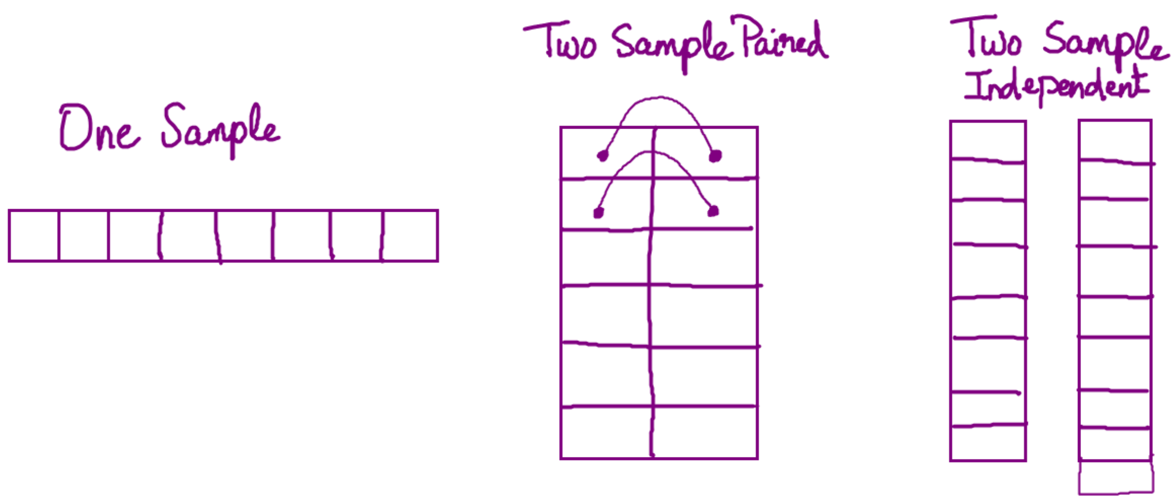
- Our sample size is small, $n < 30$

Three types of t-test

1) **One sample T-test**
   Tests the null hypothesis that the population mean is equal to a specified value `μ` based on a sample mean `x̄`.
   For example, given a sample of 15 items, you want to test if the sample mean is the same as a hypothesized mean (population). That is, essentially you want to know if the sample came from the given population or not.

   Because, there is only one sample involved and you want to compare the mean of this sample against a particular (hypothesized) value..

2) **Two Sample Independent T-Test**
   Tests the null hypothesis that two sample means $\bar{x}_1$ and $\bar{x}_2$ are equal. Compares the means of two numeric variables obtained from two independent groups.
   When to use (example):
   Test of the mean weight of mangoes from Farm A equals mean weight of mangoes from Farm B.
3) **Two Sample Dependent T-Test (aka Paired T-Test)**
   Compare the means of two numeric variables of same size where the observations from the two variables are paired. Typically, it may be from the same entity before and after a treatment, where treatment could be showing a commercial and the measured value could be opinion score about a brand.
   When to use (example):
   Given the numeric ratings of same menu items from two different restaurants (rated by same connoisseur), determine which restaurant's food tastes better



Difference between independent and paired test?

1. An Independent Samples t-test compares the means for two groups.

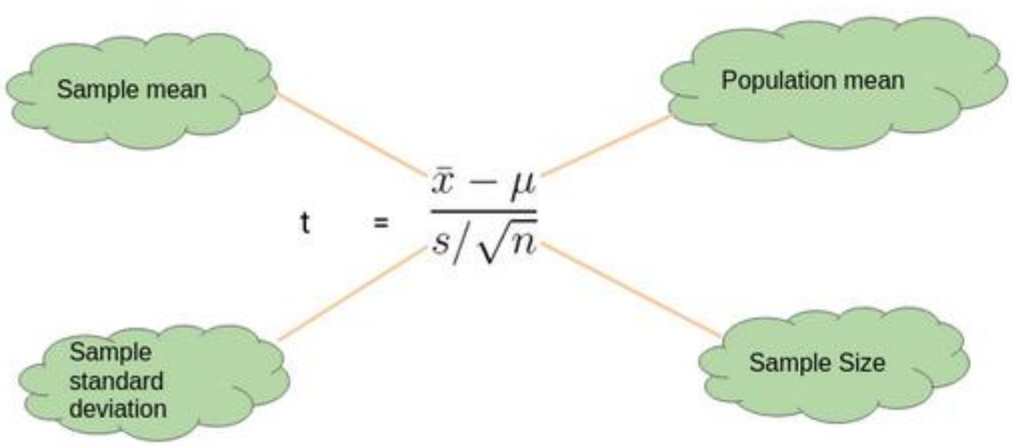
2. A Paired sample t-test compares means from the same group at different times (say, one year apart)

If the calculated t-statistic is greater than the critical t-value, the test concludes that there is a statistically significant difference between the two populations. Therefore, you reject the null hypothesis that there is no statistically significant difference between the two populations.

In any other case, there is no statistically significant difference between the two populations. The test fails to reject the null hypothesis and we accept the alternate hypothesis which says that the height of men and women are statistically different

One-Sample t-Test

We perform a One-Sample t-test when we want to **compare a sample mean with the population mean**. The difference from the Z Test is that we do **not have the information on Population Variance** here. We use the **sample standard deviation** instead of population standard deviation in this case.



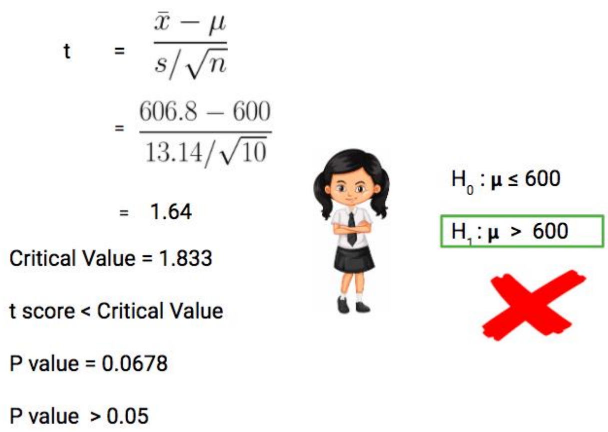
Here’s an Example to Understand a One Sample t-Test

Let’s say we want to determine if on average girls score more than 600 in the exam. We do not have the information related to variance (or standard deviation) for girls’ scores. To a perform t-test, we randomly collect the data of 10 girls with their marks and choose our α value (significance level) to be 0.05 for Hypothesis Testing.

Girls_Score
587
602
627
610
619
622
605
608
596
592

In this example:

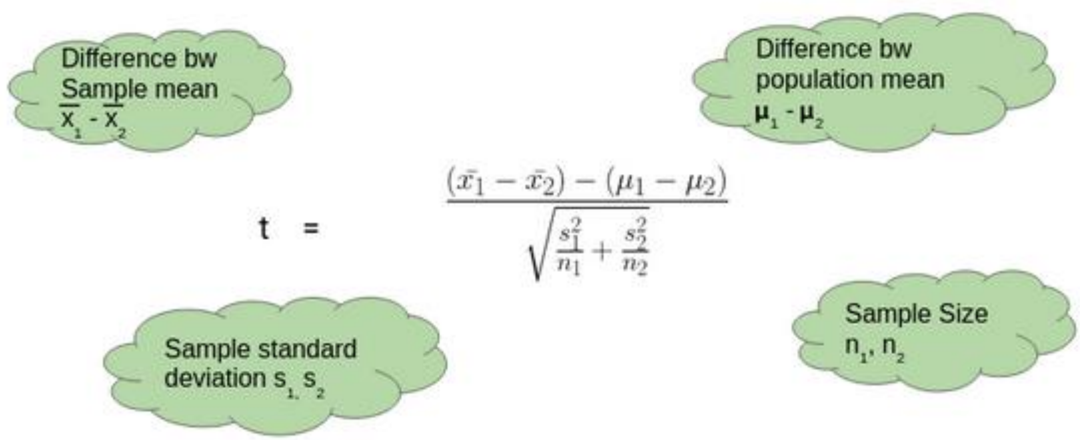
- Mean Score for Girls is 606.8
- The size of the sample is 10
- The population mean is 600
- Standard Deviation for the sample is 13.14



Our **P-value is greater than 0.05 thus we fail to reject the null hypothesis** and don't have enough evidence to support the hypothesis that on average, girls score more than 600 in the exam.

Two-Sample t-Test

We perform a Two-Sample t-test when we want to compare the mean of two samples.



Here’s an Example to Understand a Two-Sample t-Test

Here, let’s say we want to determine if on average, boys score 15 marks more than girls in the exam. We do not have the information related to variance (or standard deviation) for girls’ scores or boys’ scores. To perform a t-test. we randomly collect the data of 10 girls and boys with their marks. We choose our α value (significance level) to be 0.05 as the criteria for Hypothesis Testing.

Girls_Score
587
602
627
610
619
622
605
608
596
592

Boys_Score
626
643
647
634
630
649
625
623
617
607

In this example:

- Mean Score for Boys is 630.1
- Mean Score for Girls is 606.8
- Difference between Population Mean 15
- Standard Deviation for Boys’ score is 13.42
- Standard Deviation for Girls’ score is 13.14

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\frac{(630.1 - 606.8) - (15)}{\sqrt{\frac{(13.42)^2}{10} + \frac{(13.14)^2}{10}}}$$

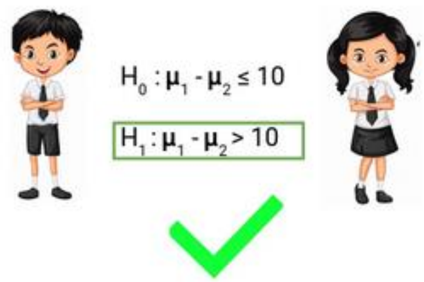
Critical Value = 1.833

t = 2.23

P value = 0.019

Critical Value > t score

P value < 0.05



Thus, **P-value is less than 0.05 so we can reject the null hypothesis** and conclude that on average boys score 15 marks more than girls in the exam.

Some more examples on One-Sample Student’s t-test:

You perform a one sample T Test when you want to test if the population mean is of a specified value μ given by a sample mean x̄.

Let me put it plainly: you have a sample of observations for which you know the mean (sample mean), and you want to test if the sample came from a population with given (hypothesized) mean.

Steps to be followed

Step 1: Define the Null and Alternate Hypothesis

H0 (Null Hypothesis): Sample Mean (x̄) = Hypothesized Population Mean(μ)

H1 (Alternate Hypothesis): Sample Mean (x̄) != Hypothesized Population Mean(μ)

Step 2: Compute the T statistic

Use the following formula to compute the t-statistic:

$$t = \frac{Z}{s} = \frac{\bar{X} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}}$$

where, $\bar{x}$=sample mean, μ=population mean, n=sample size, s=sample standard error.

The only difference with the z-statistic formula is that instead of population standard deviation, it uses the sample's standard error.

Step 3: Lookup T critical and check if computed T statistic falls in rejection region.

Based on the degrees of freedom (n-1) and the alpha level (typically 0.05), lookup the critical T value from the T-Table. If the absolute value of the computed T-statistic is greater than the critical T-value, that is, it falls in the rejection region, then we reject the null hypothesis that the population mean is of the specified value.

Alternately, you can check the output P-value. P-value is typically an output from the software as a result of the T Test. If the P-value is lesser that the significance level (typically 0.05), then reject the null hypothesis and conclude that the population mean is different from what is stated.

Example 1: A customer service company wants to know if their support agents are performing on par with industry standards.

According to a report the standard mean resolution time is 20 minutes per ticket. The sample group has a mean at 21 minutes per ticket with a standard deviation of 7 minutes.

Can you tell if the company's support performance is better than the industry standard or not?

Procedure to do One Sample T Test

Step 1: Define the Null Hypothesis (H0) and Alternate Hypothesis (H1)

Example:

H0: Sample mean ($\bar{x}$) = Hypothesized Population mean (μ)

H1: Sample mean ($\bar{x}$) != Hypothesized Population mean (μ)

The alternate hypothesis can also state that the sample mean is greater than or less than the comparison mean.

Step 2: Compute the test statistic (T)

t=Zs=X¯–μσ^n√

where s is the standard error.

Step 3: Find the T-critical from the T-Table

Use the degree of freedom and the alpha level (0.05) to find the T-critical.

Step 4: Determine if the computed test statistic falls in the rejection region.

Alternately, simply compute the P-value. If it is less than the significance level (0.05 or 0.01), reject the null hypothesis.

Example 2: We have the potato yield from 12 different farms. We know that the standard potato yield for the given variety is $\mu=20$.

x = [21.5, 24.5, 18.5, 17.2, 14.5, 23.2, 22.1, 20.5, 19.4, 18.1, 24.1, 18.5]

Test if the potato yield from these farms is significantly better than the standard yield.

Solution:

Step 1: Define the Null and Alternate Hypothesis

H0: $\overline{x} = 20$

H1: $\overline{x} > 20$

n = 12. Since this is one sample T test, the degree of freedom = n-1 = 12-1 = 11.

Let's set alpha = 0.05, to meet 95% confidence level.

Step 2: Calculate the Test Statistic (T)
1. Calculate sample mean

$$\bar{X} = \frac{x_1 + x_2 + x_3 + .. + x_n}{n}$$

$$\bar{x} = 20.175$$

1. Calculate sample standard deviation

$$\bar{\sigma} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + .. + (x_n - \bar{x})^2}{n-1}$$

$$\sigma = 3.0211$$

1. Substitute in the T Statistic formula

$$T = \frac{\bar{x} - \mu}{se} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$T = (20.175 - 20)/(3.0211/\sqrt{12}) = 0.2006$$

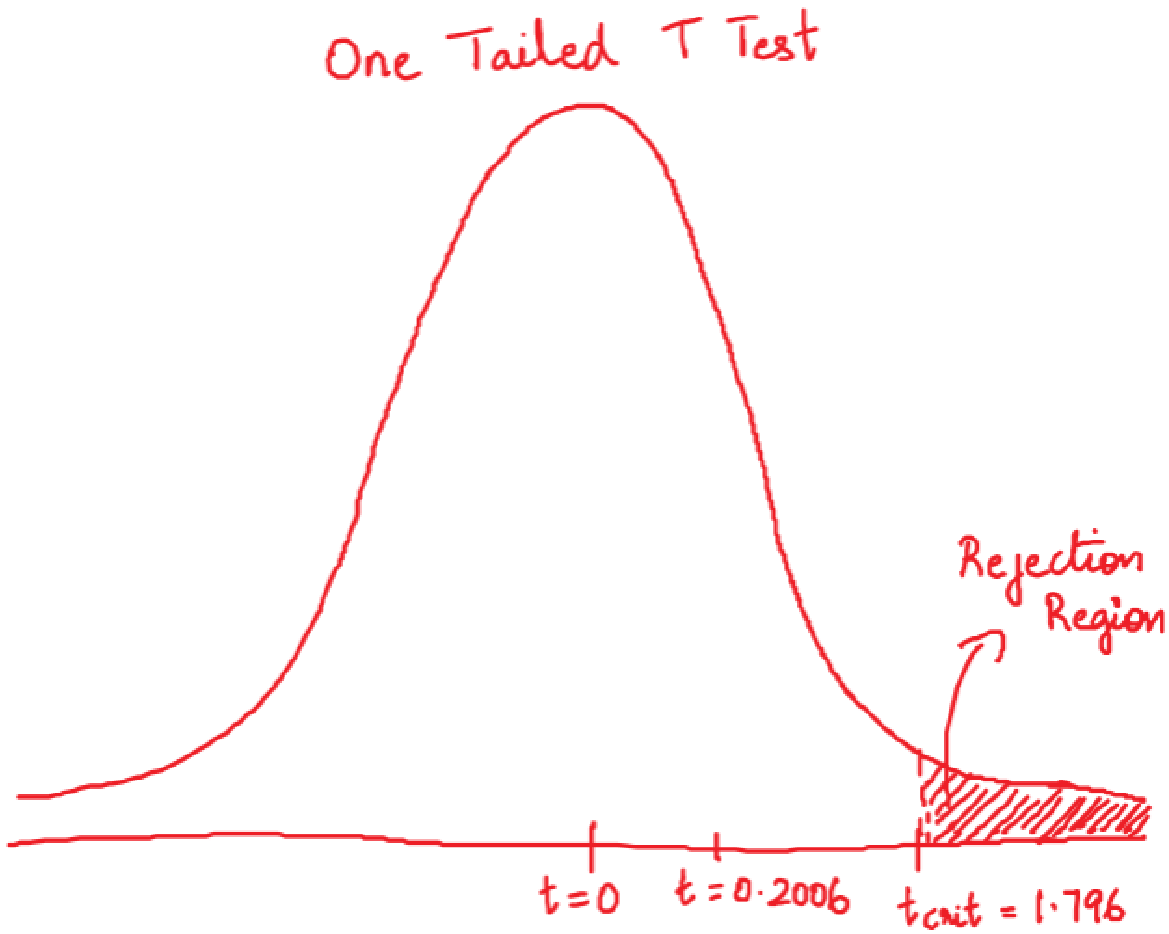
Step 3: Find the T-Critical

Confidence level = 0.95, alpha=0.05. For one tailed test, look under 0.05 column. For d.o.f = 12 – 1 = 11, T-Critical = 1.796.

Now you might wonder why 'One Tailed test' was chosen. This is because of the way you define the alternate hypothesis. Had the null hypothesis simply stated that the sample means is not equal to 20, then we would have gone for a two tailed test. More details about this topic in the next section.

Step 4: Does it fall in rejection region?

Since the computed T Statistic is less than the T-critical, it does not fall in the rejection region.



Clearly, the calculated T statistic does not fall in the rejection region. So, we do not reject the null hypothesis.

Some more examples on Two sample independent test Use Cases

Example1: Problem Statement

A global fast-food chain wants to venture into a new city by setting up a store in a popular mall in the city. It has shortlisted 2 popular locations A and B for the same and wants to choose one of them based on the number of foot falls per week.

The company has obtained the data for the same and wants to know which of the two locations to choose and if there is significant difference between the two.

Here's how the footfalls per day (in 1000's) looks like for 10 randomly chosen days:

Please refer t test input from excel file

Looks like, Location A gets 8.88k more footfalls on an average over a period of 10 days
However, the average number of footfalls is in these locations in much more and tend to vary on a day to day basis. So, the question remains: Is an additional avg footfall of 8.88k measured over 10 days, enough to say that Loc A receives different footfalls than Loc B? Is it statistically significant?

Step 0: Identify the type of T Test.

Well, here are the facts: There are two samples. Each sample contains the number of footfalls but since both are random samples the observations are not paired. That is, the number of footfalls of 116.2 in location A does not correspond to 110.0 in location B because they may not be captured on the same day. So, the observations are not paired, as a result, the T-Test to perform is the Two Sample Independent T Test.

Step 1: Frame the null and alternate hypothesis
Null Hypothesis H0: $\bar{x}$_LocA = $\bar{x}$_LocB

Alternate Hypothesis H1: $\bar{x}$_LocA != $\bar{x}$_LocB

Step 2: Degrees of freedom
Since this is a 2 sample t-Test, the degrees of freedom = 10 + 10 – 2 = 18

Ideally we should use Welch-Satterthwaite’s formula. But for simplicity of manual calculation, I am using this one for now.

Step 3: Compute the t-Statistic

Let’s start by computing the variance
Variance of Loc_A = 80.508
Variance of Loc_B = 95.584

$$t = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{8.88}{\sqrt{\frac{80.51}{10} + \frac{95.58}{10}}} = 2.1161$$

Step 4: Lookup the critical value from t-table

Since its a two-tailed test with 18 degrees of freedom. And assuming a default 95% confidence, the critical value from the t-table is 2.101

In our case, the t-statistic (2.1161) > t-critical (2.101)

This means, the t-statistic does falls in the rejection zone and so, we reject the null hypothesis and conclude that the means are in fact different.

Example2:

Your company wants to improve sales. Past sales data indicate that the average sale was $100 per transaction. After training your sales force, recent sales data (taken from a sample of 25 salesmen) indicates an average sale of $130, with a standard deviation of $15. Did the training work? Test your hypothesis at a 5% alpha level.

Step 1: Write your null hypothesis statement). The accepted hypothesis is that there is no difference in sales, so:
H0: μ = $100.

Step 2: Write your alternate hypothesis. This is the one you’re testing. You think that there *is* a difference (that the mean sales increased), so:
H1: μ > $100.

Step 3: Identify the following pieces of information you'll need to calculate the test statistic. The question should give you these items:

1. The sample mean($\bar{x}$). This is given in the question as $130.
2. The population mean(μ). Given as $100 (from past data).
3. The sample standard deviation(s) = $15.
4. Number of observations(n) = 25.

Step 4: Insert the items from above into the t score formula.

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

t = (130 – 100) / ((15 / √(25))
t = (30 / 3) = 10
This is your **calculated t-value.**

Step 5: Find the t-table value. You need two values to find this:

1. The alpha level: given as 5% in the question.
2. The degrees of freedom, which is the number of items in the sample (n) minus 1: 25 – 1 = 24.

Look up 24 degrees of freedom in the left column and 0.05 in the top row. The intersection is 1.711.This is your one-tailed critical t-value.

What this critical value means is that we would expect most values to fall under 1.711. If our calculated t-value (from Step 4) falls within this range, the null hypothesis is likely true.

Step 6: Compare Step 4 to Step 5. The value from Step 4 does not fall into the range calculated in Step 5, so we can reject the null hypothesis. The value of 10 falls into the rejection region (the left tail).
In other words, it's highly likely that the mean sale is greater.

The sales training was probably a success.

Example 3 Two sample t-test:

Sam Sleepresearcher hypothesizes that people who are allowed to sleep for only four hours will score significantly lower than people who are allowed to sleep for eight hours on a cognitive skills test. He brings sixteen participants into his sleep lab and randomly assigns them to one of two groups. In one group he has participants sleep for eight hours and in the other group he has them sleep for four. The next morning he administers the SCAT (Sam's Cognitive Ability Test) to all participants. (Scores on the SCAT range from 1-9 with high scores representing better performance).

SCAT scores
8 hours sleep group (X) 5 7 5 3 5 3 3 9
4 hours sleep group (Y) 8 1 4 6 6 4 1 2

One sample t test Use Cases

One Sample T Test Example

Paired t test example:

An instructor wants to use two exams in her classes next year. This year, she gives both exams to the students. She wants to know if the exams are equally difficult and wants to check this by looking at the differences between scores. If the mean difference between scores for students is "close enough" to zero, she will make a practical conclusion that the exams are equally difficult. Here is the data:

Student	Exam 1 Score	Exam 2 Score	Difference
Bob	63	69	6
Nina	65	65	0
Tim	56	62	6
Kate	100	91	-9
Alonzo	88	78	-10
Jose	83	87	4
Nikhil	77	79	2
Julia	92	88	-4
Tohru	90	85	-5
Michael	84	92	8
Jean	68	69	1
Indra	74	81	7
Susan	87	84	-3
Allen	64	75	11
Paul	71	84	13
Edwina	88	82	-6

If you look at the table above, you see that some of the score differences are positive and some are negative. You might think that the two exams are equally difficult. Other people might disagree. The statistical test gives a common way to make the decision, so that everyone makes the same decision on the same data.

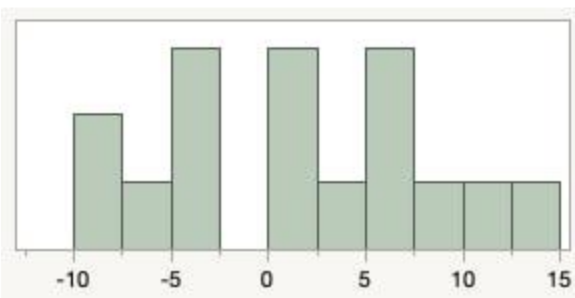
Checking the data

Let’s start by answering: Is the paired t-test an appropriate method to evaluate the difference in difficulty between the two exams?

- Subjects are independent. Each student does their own work on the two exams.
- Each of the paired measurements are obtained from the same subject. Each student takes both tests.
- The distribution of differences is normally distributed. For now, we will assume this is true. We will test this later.

We decide that we have selected a valid analysis method.

Before jumping into the analysis, we should plot the data. The figure below shows a histogram and summary statistics for the score differences.



Summary Statistics

Mean	1.3125
Std Dev	7.002083
Std Err Mean	1.7505208
Upper 95% Mean	5.0436467
Lower 95% Mean	-2.418647
N	16

From the histogram, we see that there are no very unusual points, or *outliers*. The data are roughly bell-shaped, so our idea of a normal distribution for the differences seems reasonable.

From the statistics, we see that the average, or mean, difference is 1.3. Is this “close enough” to zero for the instructor to decide that the two exams are equally difficult? Or not?

How to perform the paired *t*-test

We'll further explain the principles underlying the paired *t*-test in the Statistical Details section below, but let's first proceed through the steps from beginning to end. We start by calculating our test statistic. To accomplish this, we need the average difference, the standard deviation of the difference and the sample size. These are shown in Figure 1 above. (Note that the statistics are rounded to two decimal places below. Software will usually display more decimal places and use them in calculations.)

The average score difference is:

$$\overline{x_d} = 1.31$$

Next, we calculate the standard error for the score difference. The calculation is:

$$\text{Standard Error} = \frac{s_d}{\sqrt{n}} = \frac{7.00}{\sqrt{16}} = \frac{7.00}{4} = 1.75$$

In the formula above, n is the number of students – which is the number of differences. The standard deviation of the differences is s_d.

We now have the pieces for our test statistic. We calculate our test statistic as:

$$t = \frac{\text{Average difference}}{\text{Standard Error}} = \frac{1.31}{1.75} = 0.750$$

To make our decision, we compare the test statistic to a value from the *t*-distribution. This activity involves four steps:

1. We decide on the risk we are willing to take for declaring a difference when there is not a difference. For the exam score data, we decide that we are willing to take a 5% risk of saying that the unknown mean exam score difference is zero when in reality it is not. In statistics-

speak, we set the significance level, denoted by α, to 0.05. It's a good practice to make this decision before collecting the data and before calculating test statistics.

2. We calculate a test statistic. Our test statistic is 0.750.
3. We find the value from the *t*-distribution. Most statistics books have look-up tables for the distribution. You can also find tables online. The most likely situation is that you will use software for your analysis and will not use printed tables.

   To find this value, we need the significance level ($\alpha = 0.05$) and the *degrees of freedom*. The degrees of freedom (*df*) are based on the sample size. For the exam score data, this is:

$$Df = 16\text{-}1 = 15$$

1. The t value with $\alpha = 0.05$ and 15 degrees of freedom is 2.131.
2. We compare the value of our statistic (0.750) to the t value. Because $0.750 < 2.131$, we cannot reject our idea that the mean score difference is zero. We make a practical conclusion to consider exams as equally difficult.

Statistical details

Let’s look at the exam score data and the paired t-test using statistical terms.

Our null hypothesis is that the population mean of the differences is zero. The null hypothesis is written as:

$$H_o : \mu_{\mathrm{d}} = 0$$

The alternative hypothesis is that the population mean of the differences is not zero. This is written as:

$$H_o : \mu_{\mathrm{d}} \neq 0$$

We calculate the standard error as:

$$Standard Error = \frac{s_d}{\sqrt{n}}$$

The formula shows the sample standard deviation of the differences as s_d and the sample size as n.

The test statistic is calculated as:

$$t = \frac{\mu_{\mathrm{d}}}{\frac{s}{\sqrt{n}}}$$

We compare the test statistic to a *t* value with our chosen alpha value and the degrees of freedom for our data. In our exam score data example, we set α = 0.05. The degrees of freedom (*df*) are based on the sample size and are calculated as:

$$df = n - 1 = 16 - 1 = 15$$

Statisticians write the *t* value with α = 0.05 and 15 degrees of freedom as:

$$t_{0.05,15}$$

The t value with α = 0.05 and 15 degrees of freedom is 2.131. There are two possible results from our comparison:

- The test statistic is lower than the t value. You fail to reject the hypothesis that the mean difference is zero. The practical conclusion made by the instructor is that the two tests are equally difficult. Next year, she can use both exams and give half the students one exam and half the other exam.
- The test statistic is higher than the t value. You reject the hypothesis that the mean difference is zero. The practical conclusion made by the instructor is that the tests are not of equal difficulty. She must use the same exam for all students.

Quiz

Q1. A study was set up to look at whether there was a difference in the mean arterial blood pressure between two groups of volunteers, after 6 weeks of following one of two treatment programs. One group of volunteers were given an exercise regimen to follow for the 6 weeks and the other group were given the same exercise regimen with the addition of an experimental tablet.

Which type of t-test should be used in this situation?

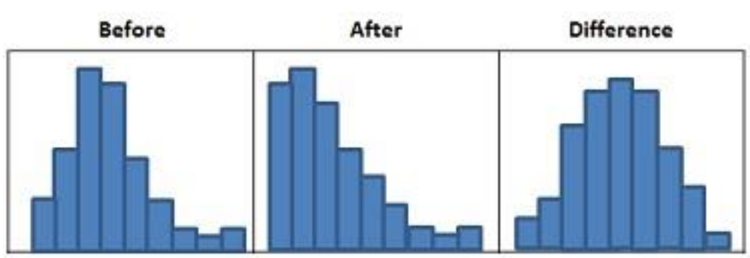
a) One sample t-test

b) independent samples t-test

c) Paired samples t-test

d) None of the t-tests would be suitable

Q2: The following Histograms display the distribution of reported alcohol consumption (units) in patients diagnosed with alcoholic liver disease before an intervention and after the intervention has been completed. A histogram of the difference (before minus after) is also presented.



If we were interested in testing to see if there had been a significant change in reported alcohol consumption then we could use which of the following t-tests and for what reason?

a) One sample t-test because the “Before” data is normally distributed

b) Independent samples t-test because the “Before” data is normally distributed

c) A paired samples t-test because the “Before” is normally distributed.

d) A paired samples t-test because the “Difference” is normally distributed.

e) No t-tests would be suitable because the “Before” data is not normally distributed

f) No t-tests would be suitable because the “After” data is not normally distributed.

Q3. When should you use the t-test? (select all that apply)

a) When the sample is large
b) When you are testing for a mean
c) When you are given the population standard deviation
d) When you are ONLY given the sample standard deviation
e) When population parameters are unknown

Q4: What t-test type compares the means for two groups?
a) Paired Samples
b) Independent
c) Dependent
d) Mixed

Q5: What is an independent t-test used for?
a) Comparing means from a sample and a population
b) Comparing means from two separate samples
c) Comparing means from two related samples
d) Comparing means from more than two samples

Q6: To use a t-test, the dependent variable must have
a) Nominal or interval data
b) Ordinal or ratio data
c) Interval or ratio data
d) Ordinal or interval data

Z score

What is the Z Test?

z tests are a statistical way of testing a hypothesis when either:

- We know the population variance, or
- We do not know the population variance but our sample size is large $n \geq 30$

If we have a sample size of less than 30 and do not know the population variance, then we must use a t-test.

In statistics, a z-score tells us how many standard deviations away a given value lies from a population mean.

We use the following formula to calculate a z-score for a given value:

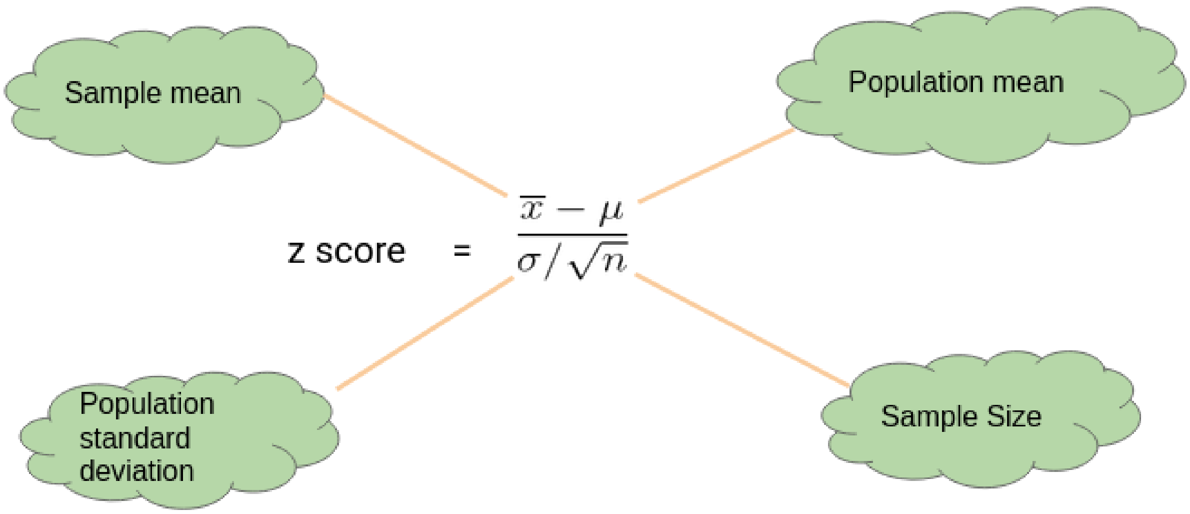
$$z = (x - \mu) / \sigma$$

where:

- **x**: Individual data value
- **μ**: Mean of population
- **σ**: Standard deviation of population

One-Sample Z test

We perform the One-Sample Z test when we want to compare **a sample mean with the population mean.**



Here's an Example to Understand a One Sample Z Test

Let's say we need to determine if girls on average score higher than 600 in the exam. We have the information that the standard deviation for girls' scores is 100. So, we collect the data of 20 girls by using random samples and record their marks. Finally, we also set our α value (significance level) to be 0.05.

Score
650
730
510
670
480
800
690
530
590
620
710
670
640
780
650
490
800
600
510
700

In this example:

- Mean Score for Girls is 641
- The size of the sample is 20
- The population mean is 600
- Standard Deviation for Population is 100

$$\text{z score} = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}}$$

$$= \frac{641 - 600}{100/\sqrt{20}}$$

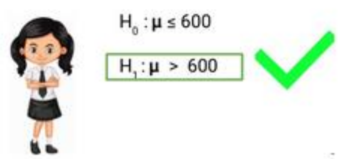
$$= 1.8336$$

$$\text{p value} = .033357.$$

Critical Value = 1.645

Z score > Critical Value

P value < 0.05



Since the P-value is less than 0.05, we can reject the null hypothesis and conclude based on our result that Girls on average scored higher than 600.

Here's an Example to Understand a Two Sample Z Test

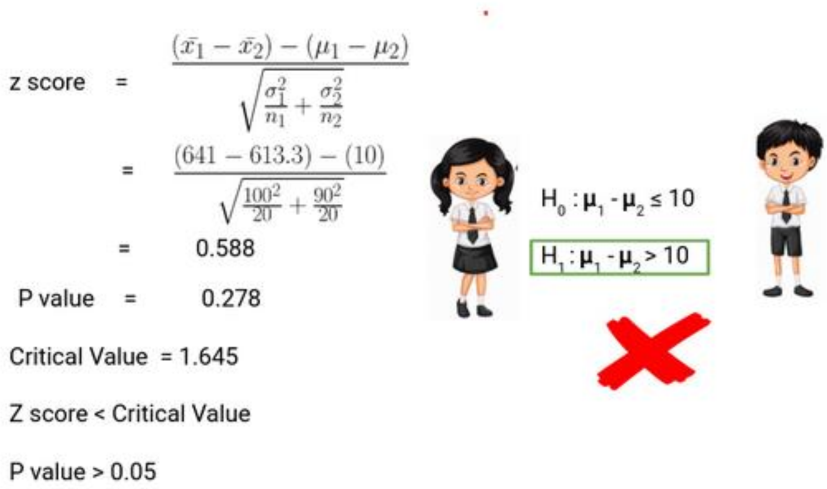
Here, let's say we want to know if Girls on average score 10 marks more than the boys. We have the information that the standard deviation for girls' Score is 100 and for boys' score is 90. Then we collect the data of 20 girls and 20 boys by using random samples and record their marks. Finally, we also set our α value (significance level) to be 0.05.

Score
650
730
510
670
480
800
690
530
590
620
710
670
640
780
650
490
800
600
510
700

Score
630
720
462
631
440
783
673
519
543
579
677
649
632
768
615
463
781
563
488
650

In this example:

- Mean Score for Girls (Sample Mean) is 641
- Mean Score for Boys (Sample Mean) is 613.3
- Standard Deviation for the Population of Girls’ is 100
- Standard deviation for the Population of Boys’ is 90
- Sample Size is 20 for both Girls and Boys
- Difference between Mean of Population is 10



Thus, we can **conclude based on the P-value that we fail to reject the Null Hypothesi**s. We don't have enough evidence to conclude that girls on average score of 10 marks more than the boys. Pretty simple, right?

Some more examples on z test

Example: Calculate and Interpret Z-Scores

Suppose the scores for a certain exam are normally distributed with a mean of 80 and a standard deviation of 4.

Example 1: Find the z-score for an exam score of 87.

We can use the following steps to calculate the z-score:

- The mean is $\mu = 80$
- The standard deviation is $\sigma = 4$
- The individual value we're interested in is $X = 87$
- Thus, $z = (X - \mu) / \sigma = (87 - 80) / 4 = 1.75$.

This tells us that an exam score of 87 lies 1.75 standard deviations above the mean.

Example 2: Find the z-score for an exam score of 75.

We can use the following steps to calculate the z-score:

- The mean is $\mu = 80$
- The standard deviation is $\sigma = 4$
- The individual value we're interested in is $X = 75$
- Thus, $z = (X - \mu) / \sigma = (75 - 80) / 4 = -1.25$.

This tells us that an exam score of 75 lies 1.25 standard deviations below the mean.

Example 3: Find the z-score for an exam score of 80.

We can use the following steps to calculate the z-score:

- The mean is $\mu = 80$
- The standard deviation is $\sigma = 4$
- The individual value we're interested in is $X = 80$
- Thus, $z = (X - \mu) / \sigma = (80 - 80) / 4 = 0$.

This tells us that an exam score of 80 is exactly equal to the mean.

Important Notes on Z Test

- Z test is a statistical test that is conducted on normally distributed data to check if there is a difference in means of two data sets.
- The sample size should be greater than 30 and the population variance must be known to perform a z test.
- The one-sample z test checks if there is a difference in the sample and population mean,
- The two sample z test checks if the means of two different groups are equal.

Example 4: A teacher claims that the mean score of students in his class is greater than 82 with a standard deviation of 20. If a sample of 81 students was selected with a mean score of 90 then check if there is enough evidence to support this claim at a 0.05 significance level.

Solution: As the sample size is 81 and population standard deviation is known, this is an example of a right-tailed one-sample z test.

$H_0 : \mu = 82$

$H_1 : \mu > 82$

From the z table the critical value at α = 1.645

$z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

$\overline{x} = 90, \mu = 82, n = 81, \sigma = 20$

$z = 3.6$

As 3.6 > 1.645 thus, the null hypothesis is rejected and it is concluded that there is enough evidence to support the teacher's claim.

Answer: Reject the null hypothesis

-

Example 5: An online medicine shop claims that the mean delivery time for medicines is less than 120 minutes with a standard deviation of 30 minutes. Is there enough evidence to support this claim at a 0.05 significance level if 49 orders were examined with a mean of 100 minutes?

Solution: As the sample size is 49 and population standard deviation is known, this is an example of a left-tailed one-sample z test.

$H_0 : \mu = 120$

$H_1 : \mu < 120$

From the z table the critical value at α = -1.645. A negative sign is used as this is a left tailed test.

$$z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$\overline{x} = 100$, $\mu = 120$, $n = 49$, $\sigma = 30$

$z = -4.66$

As -4.66 < -1.645 thus, the null hypothesis is rejected and it is concluded that there is enough evidence to support the medicine shop's claim.

- **Answer:** Reject the null hypothesis
-

Example 6: A company wants to improve the quality of products by reducing defects and monitoring the efficiency of assembly lines. In assembly line A, there were 18 defects reported out of 200 samples while in line B, 25 defects out of 600 samples were noted. Is there a difference in the procedures at a 0.05 alpha level?

Solution: This is an example of a two-tailed two proportion z test.

H_0: The two proportions are the same.

H_1: The two proportions are not the same.

As this is a two-tailed test the alpha level needs to be divided by 2 to get 0.025.

Using this, the critical value from the z table is 1.96.

$n_1 = 200$, $n_2 = 600$

$p_1 = 18 / 200 = 0.09$

$p_2 = 25 / 600 = 0.0416$

$p = (18 + 25) / (200 + 600) = 0.0537$

$$z = \frac{p_1 - p_2 - 0}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 2.62$$

As 2.62 > 1.96 thus, the null hypothesis is rejected and it is concluded that there is a significant difference between the two lines.

Answer: Reject the null hypothesis

Why Are Z-Scores Useful?

Z-scores are useful because they give us an idea of how an individual value compares to the rest of a distribution.

For example, is an exam score of 87 good? Well, that depends on the mean and standard deviation of all exam scores.

If the exam scores for the whole population are normally distributed with a mean of 90 and a standard deviation of 4, we would calculate the z-score for 87 to be:

$z = (X – \mu) / \sigma = (87 – 90) /4 = -0.75$.

Since this value is negative, it tells us that an exam score of 87 is actually below the average exam score for the population. Specifically, an exam score of 87 is 0.75 standard deviations below the mean.

In a nutshell, z-scores give us an idea of how individual values compare to the mean.

Example 1: Exam Scores

Z-scores are often used in academic settings to analyze how well a student's score compares to the mean score on a given exam.

For example, suppose the scores on a certain college entrance exam are roughly normally distributed with a mean of 82 and a standard deviation of 5.

If a certain student received a 90 on the exam, we would calculate their z-score to be:

- $z = (x – \mu) / \sigma$
- $z = (90 – 82) / 5$
- $z = 1.6$

This means that this student received a score that was 1.6 standard deviations above the mean.

z-score of 1.6 represents a value that is greater than **94.52%** of all exam scores.

Example 2: Newborn Weights

Z-scores are often used in a medical setting to analyze how a certain newborn's weight compares to the mean weight of all babies.

For example, it's well-documented that the weights of newborns are normally distributed with a mean of about 7.5 pounds and a standard deviation of 0.5 pounds.

If a certain newborn weights 7.7 pounds, we would calculate their z-score to be:

- $z = (x – \mu) / \sigma$
- $z = (7.7 – 7.5) / 0.5$
- $z = 0.4$

This means that this baby weighs 0.4 standard deviations above the mean.

z-score of 0.4 represents a weight that is greater than **65.54%** of all baby weights.

Example 3: Giraffe Heights

Z-scores are often used in a biology to assess how the height of a certain animal compares to the mean population height of that particular animal.

For example, suppose the heights of a certain species of giraffe is normally distributed with a mean of 16 feet and a standard deviation of 2 feet.

If a certain giraffe from this species is 15 feet tall, we would calculate their z-score to be:

- $z = (x – \mu) / \sigma$
- $z = (15 – 16) / 2$
- $z = -0.5$

This means that this giraffe has a height that is 0.5 standard deviations below the mean.

z-score of -0.5 represents a height that is greater than just **30.85%** of all giraffes.

Example 4: Shoe Size

Z-scores can be used to determine how a certain shoe size compares to the mean population size.

For example, it's known that shoe sizes for males in the U.S. is roughly normally distributed with a mean of size 10 and a standard deviation of 1.

If a certain man has a shoe size of 10, we would calculate their z-score to be:

- $z = (x - \mu) / \sigma$
- $z = (10 - 10) / 1$
- $z = 0$

This means that this man has a shoe size that is 0 standard deviations away from the mean.

z-score of 0 represents a shoe size that is greater than exactly **50%** of all males.

Example 5: Blood Pressure

Z-scores are often used in medical settings to assess how an individual's blood pressure compares to the mean population blood pressure.

For example, the distribution of diastolic blood pressure for men is normally distributed with a mean of about 80 and a standard deviation of 20.

If a certain man has a diastolic blood pressure of 100, we would calculate their z-score to be:

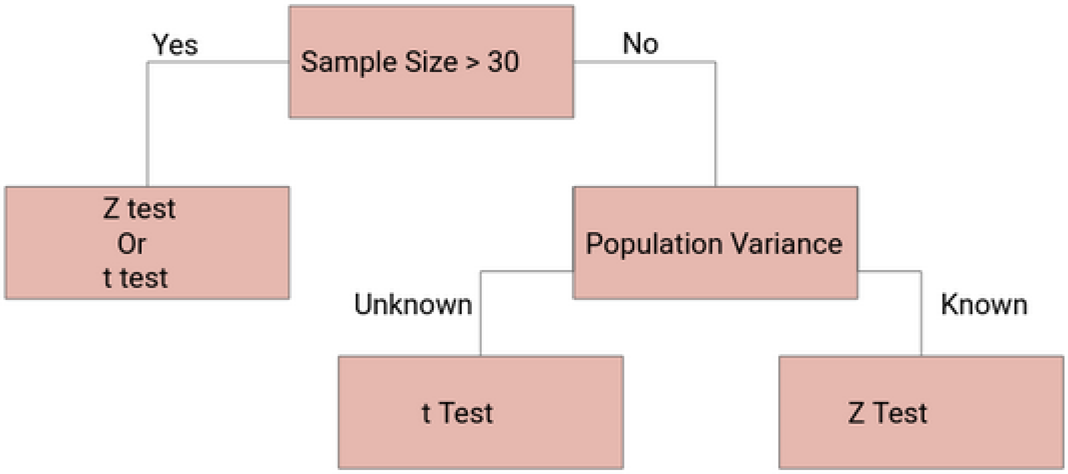
- $z = (x - \mu) / \sigma$
- $z = (100 - 80) / 20$
- $z = 1$

This means that this man has a diastolic blood pressure that is 1 standard deviation above the mean.

z-score of 1 represents a blood pressure size that is greater than **84.13%** of all males.

Deciding between Z Test and T-Test

So when we should perform the Z test and when we should perform t-Test? It's a key question we need to answer if we want to master statistics.



If the sample size is large enough, then the Z test and t-Test will conclude with the same results. For a **large sample size**, **Sample Variance will be a better estimate** of Population variance so even if population variance is unknown, we can **use the Z test using sample variance.**

Similarly, for a **Large Sample**, we have a high degree of freedom. And since **t-distribution approaches the normal distribution**, **t**he difference between the z score and t score is negligible.

Chi Square

Example 1: "Which holiday do you prefer?"

	Beach	Cruise
Men	209	280
Women	225	248

Does Gender affect Preferred Holiday?

If Gender (Man or Woman) **does** affect Preferred Holiday we say they are **dependent**.

By doing some special calculations (explained later), we come up with a "p" value:

p value is 0.132

Now, **$p < 0.05$** is the usual test for **dependence**.

In this case **p is greater than 0.05**, so we believe the variables are **independent** (ie not linked together).

In other words Men and Women probably do **not** have a different preference for Beach Holidays or Cruises.

It was just random differences which we expect when collecting data.

Example 2: "Which pet do you prefer?"

	Cat	Dog
Men	207	282
Women	231	242

By doing the calculations (shown later), we come up with:

P value is 0.043

In this case **$p < 0.05$**, so this result is thought of as being "significant" meaning we think the variables are **not** independent.

In other words, because **$0.043 < 0.05$** we think that Gender is linked to Pet Preference (Men and Women have different preferences for Cats and Dogs).

Chi-Square Test

Note: **Chi** Sounds like "Hi" but with a **K**, so it sounds like "**Ki** square"

And Chi is the greek letter X, so we can also write it X^2

Important points before we get started:

- This test only works for **categorical** data (data in categories), such as Gender {Men, Women} or color {Red, Yellow, Green, Blue} etc, but **not numerical** data such as height or weight.
- The numbers must be large enough. Each entry must be **5** or more. In our example we have values such as 209, 282, etc, so we are good to go.

Our first step is to state our **hypotheses**:

Hypothesis: A statement that might be true, which can then be tested.

The two **hypotheses** are.

- Gender and preference for cats or dogs are **independent**.
- Gender and preference for cats or dogs are **not independent**.

Lay the data out in a table:

	Cat	Dog
Men	207	282
Women	231	242

Add up rows and columns:

	Cat	Dog	
Men	207	282	489
Women	231	242	473
	438	524	962

Calculate "Expected Value" for each entry:

Multiply each row total by each column total and divide by the overall total:

	Cat	Dog	
Men	$\frac{489 \times 438}{962}$	$\frac{489 \times 524}{962}$	489
Women	$\frac{473 \times 438}{962}$	$\frac{473 \times 524}{962}$	473
	438	524	962

Which gives us:

	Cat	Dog	
Men	222.64	266.36	489
Women	215.36	257.64	473
	438	524	962

Subtract expected from observed, square it, then divide by expected:

In other words, use formula $\frac{(O-E)^2}{E}$ where

- O = **Observed** (actual) value
- E = **Expected** value

	Cat	Dog	
Men	$\frac{(207-222.64)^2}{222.64}$	$\frac{(282-266.36)^2}{266.36}$	489
Women	$\frac{(231-215.36)^2}{215.36}$	$\frac{(242-257.64)^2}{257.64}$	473
	438	524	962

Which gets us:

	Cat	Dog	
Men	1.099	0.918	489
Women	1.136	0.949	473
	438	524	962

Now add up those calculated values:

$$1.099 + 0.918 + 1.136 + 0.949 = 4.102$$

Chi-Square is 4.102

From Chi-Square to p

Degrees of Freedom

First we need a "Degree of Freedom"

$$\text{Degree of Freedom} = (\text{rows} - 1) \times (\text{columns} - 1)$$

p-value

The rest of the calculation is difficult, so either look it up in a table or use the Chi-Square Calculator.

The result is:

$$p = 0.04283$$

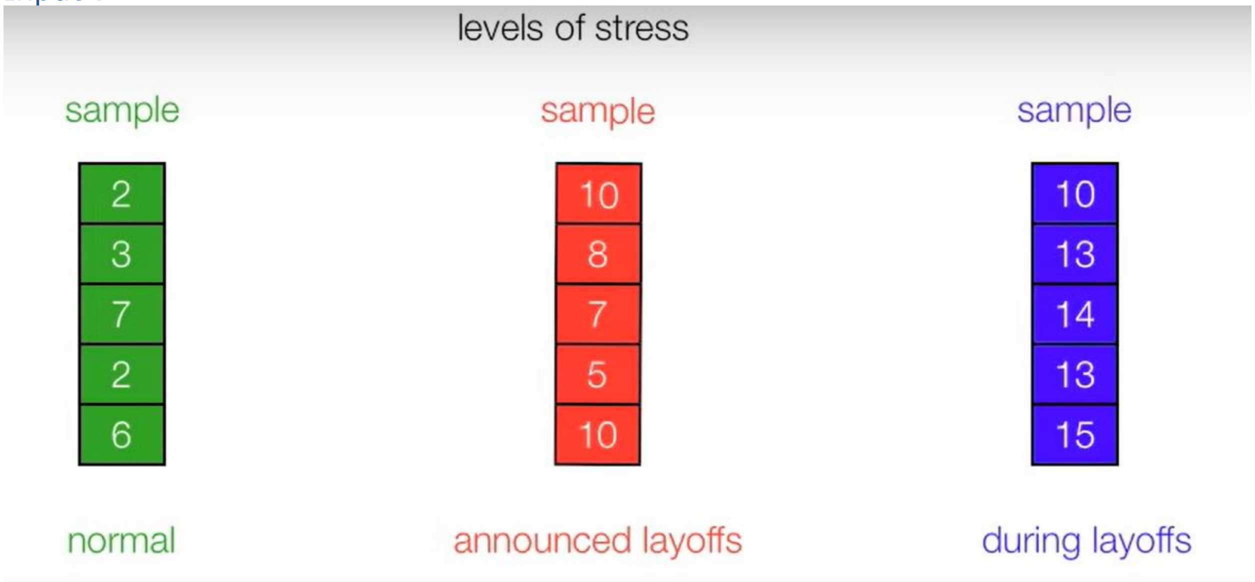
Done!

Critical value is 3.841

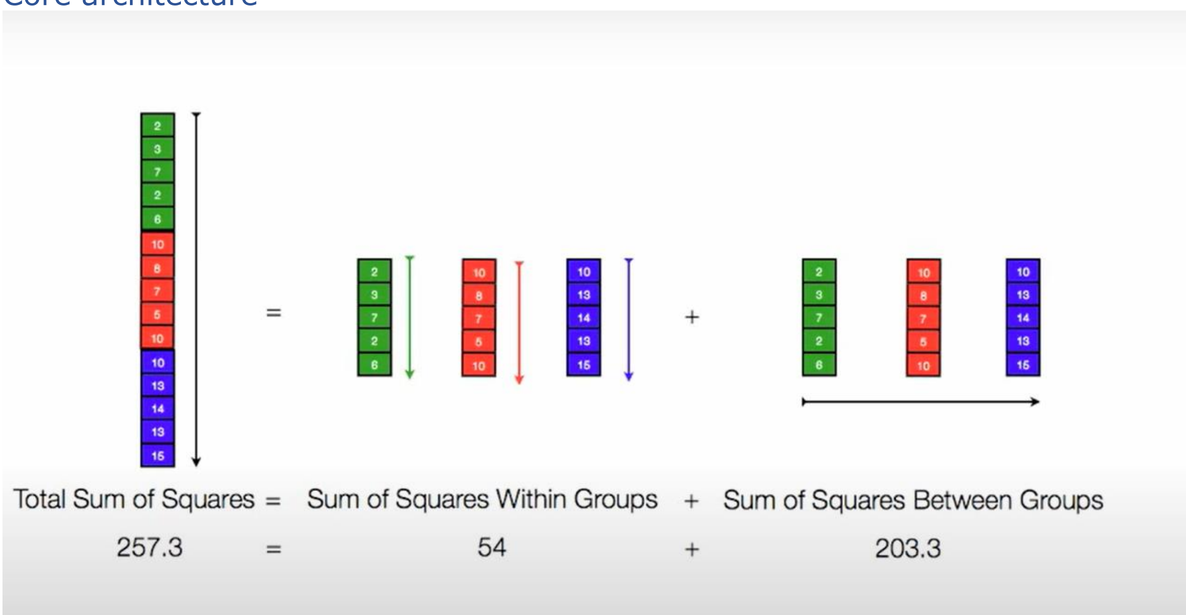
Critical value(3.841) is greater than chi square value(4.102), so we are rejecting our null hypothesis.

Anova

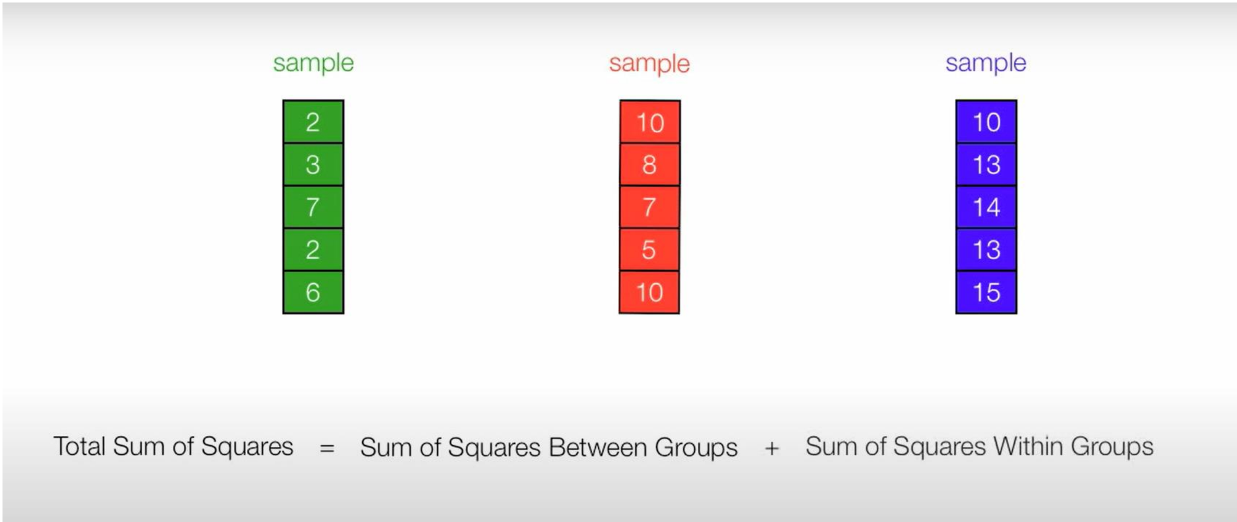
Input :



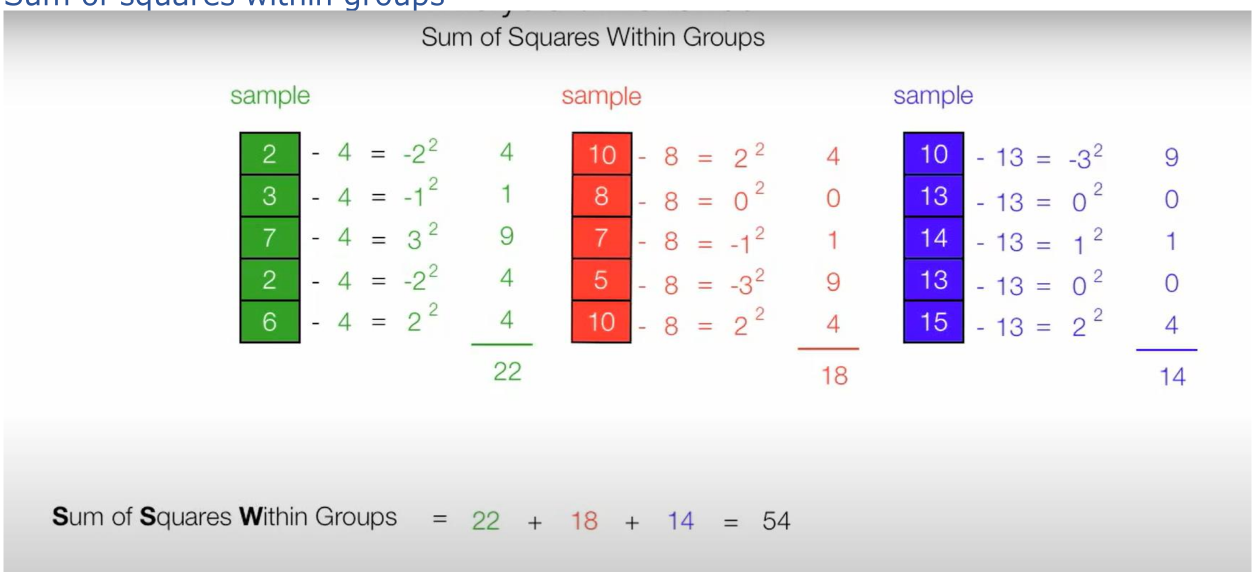
Core architecture



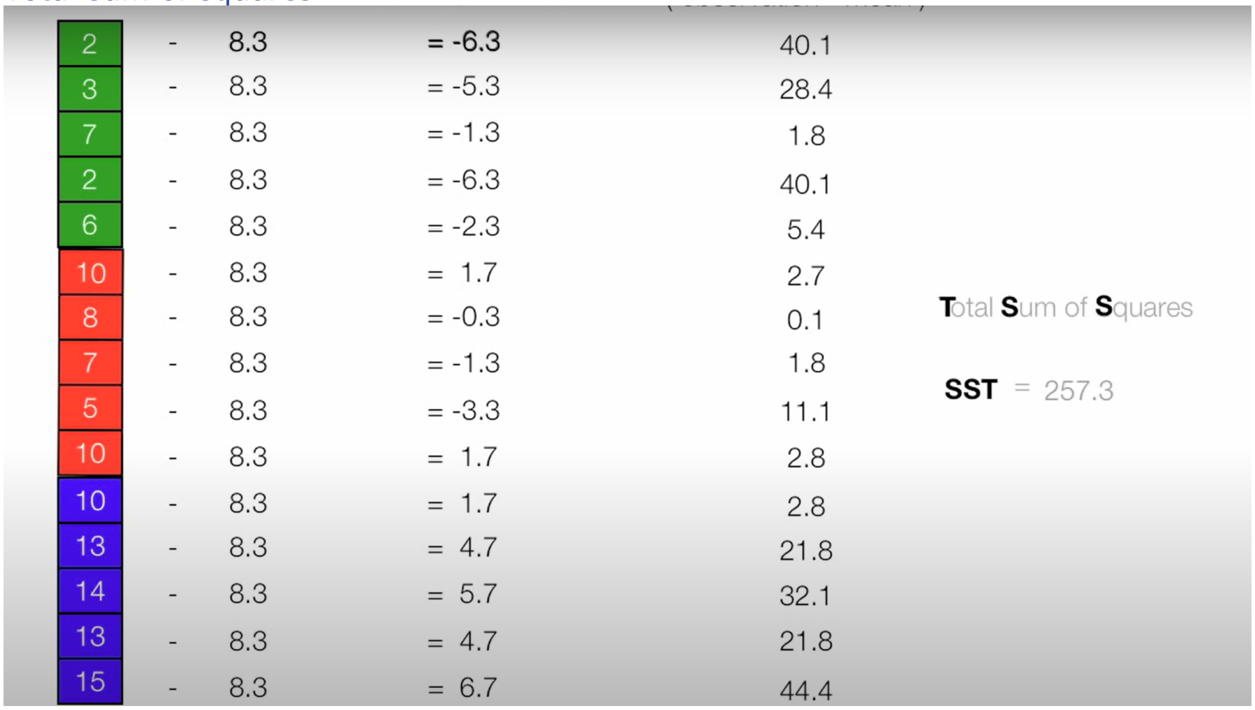
Anova formula



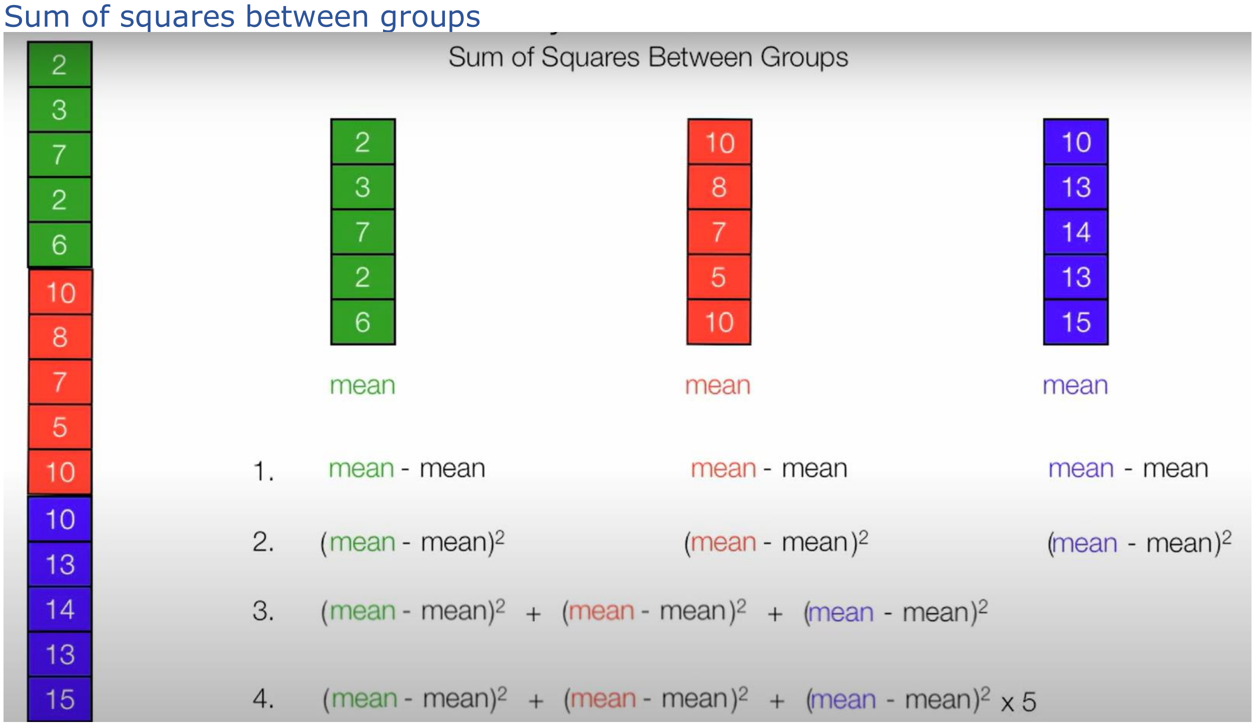
Sum of squares within groups



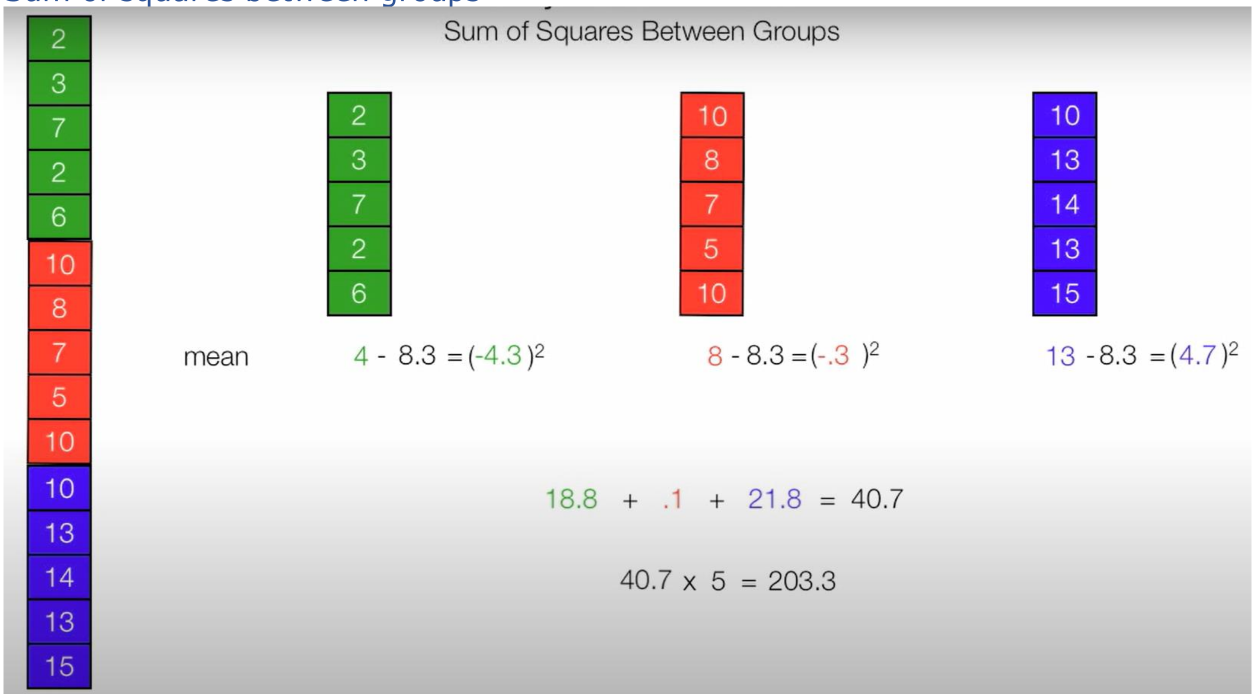
Total sum of squares



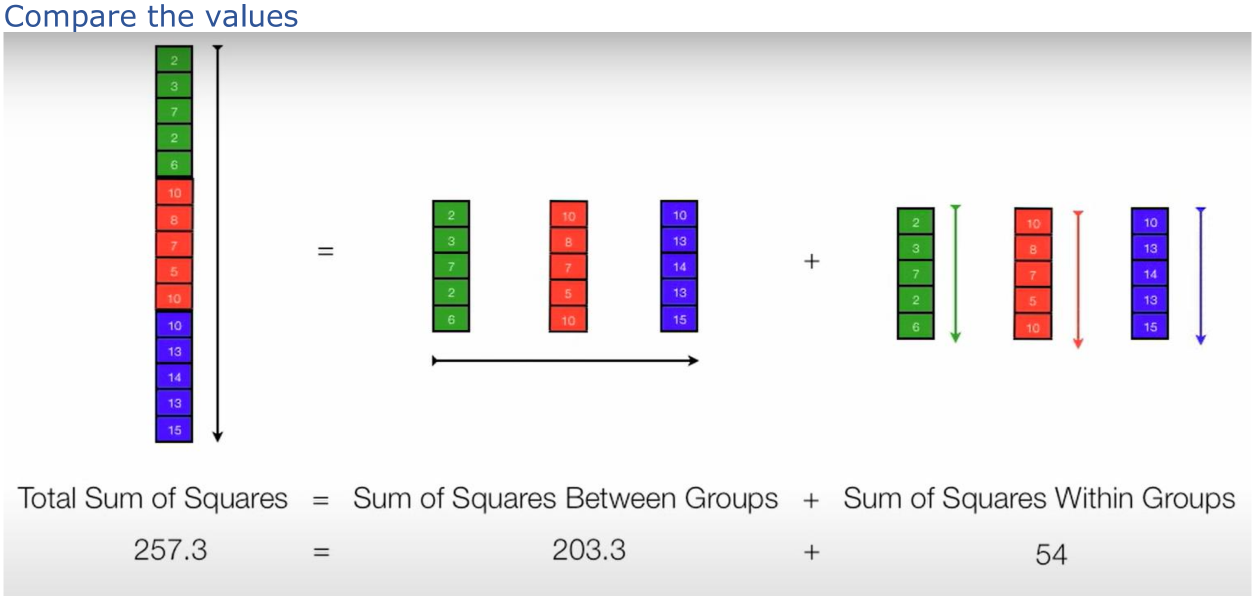
Sum of squares between groups



Sum of squares between groups



Compare the values



Final calculation

Final Calculations

$$\frac{\text{Sum of Squares Between Groups}}{\text{degrees of freedom}} = \frac{203.3}{2} = 101.667$$

$$F = \frac{101.667}{4.5} = 22.59$$

$$\frac{\text{Sum of Squares Within Groups}}{\text{degrees of freedom}} = \frac{54}{12} = 4.5$$

F ratio value

$$F(2,12) = 22.59,\ p<.05$$

Find critical value

degrees of freedom denominator

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	161.5	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	246.0	248.0	249.1	250.1
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31

Critical Value = 3.89

F = 22.59

value