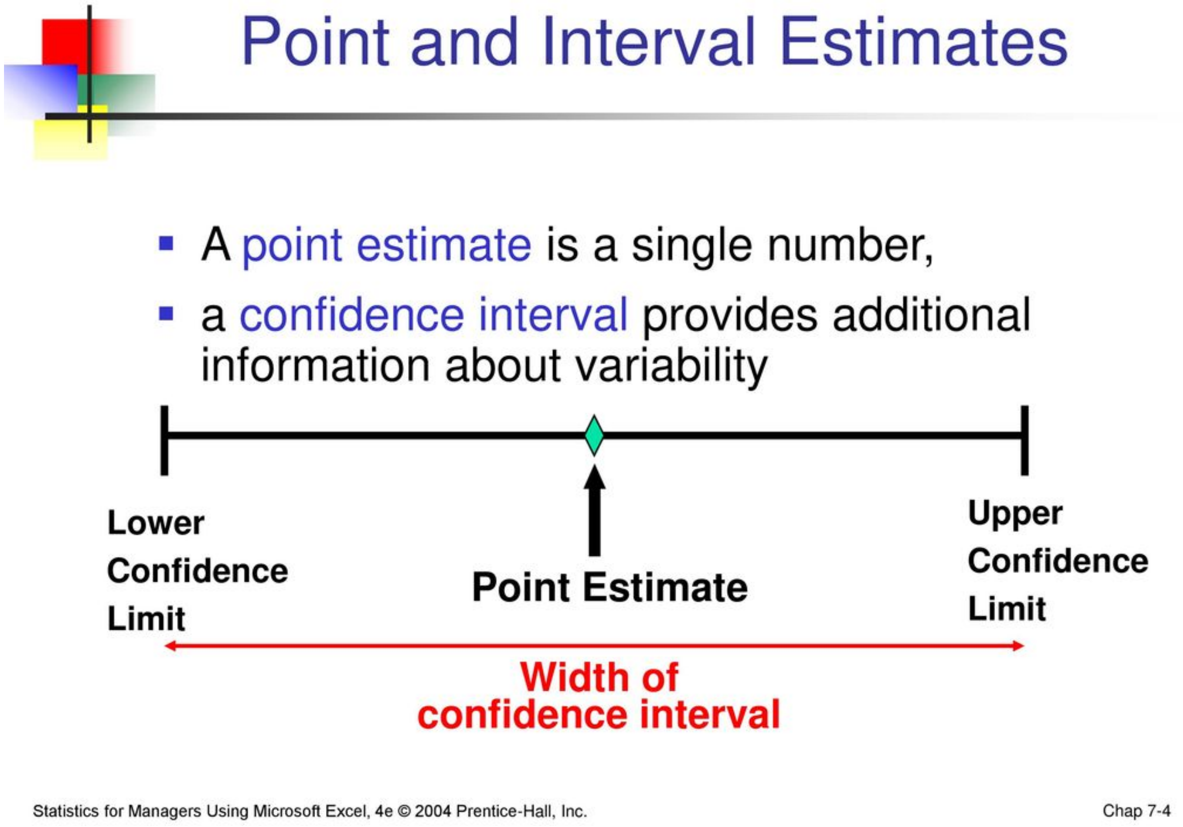
Road to Data Science in 50 Days - Day 4 ¶

Statistics for Data Science

Previously, we got introduced to inferential statistcs and different types of probability distributions. If you want to learn more about probability such as Conditional probability, bayes theorem (Which we are going to cover ahead when we will learn about a Machine Learning algorithm called Naive Bayes), head to this link which explains it with very good visualizations and examples: https://www.mathsisfun.com/data/probability-events-types.html (https://www.mathsisfun.com/data/probability-events-types.html)

Estimation

In statistics, **estimation** refers to the process by which one makes inferences about a population, based on information obtained from a sample.



Point and Interval Estimates

- A point estimate is a single number,
- a confidence interval provides additional information about variability

Lower Confidence Limit

Point Estimate

Upper Confidence Limit

Width of confidence interval

Statistics for Managers Using Microsoft Excel, 4e © 2004 Prentice-Hall, Inc.

Chap 7-4

Point Estimate vs. Interval Estimate

Statisticians use sample statistics to estimate population parameters. For example, sample means are used to estimate population means; sample proportions, to estimate population proportions.

An estimate of a population parameter may be expressed in two ways:

- Point estimate. A point estimate of a population parameter is a single value of a statistic. For example, the sample mean x is a point estimate of the population mean μ. Similarly, the sample proportion p is a point estimate of the population proportion P.

- Interval estimate. An interval estimate is defined by two numbers, between which a population parameter is said to lie. For example, a < x < b is an interval estimate of the population mean μ. It indicates that the population mean is greater than a but less than b.

Types Estimators can be described in several ways (click on the bold word for the main article on that term):

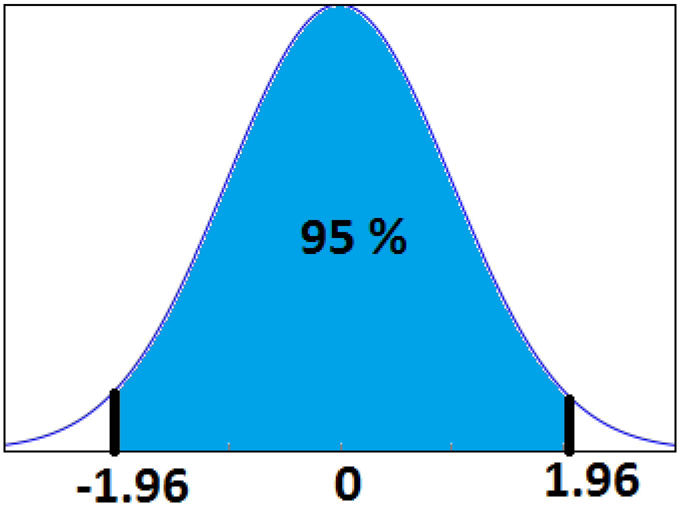
- **Biased** (https://www.statisticshowto.com/what-is-bias/): a statistic that is either an overestimate or an underestimate.

- **Efficient** (https://www.statisticshowto.com/efficient-estimator-efficiency/): a statistic with small variances (the one with the smallest possible variance is also called the "best"). Inefficient estimators can give you good results as well, but they usually requires much larger samples.
- **Invariant** (https://www.statisticshowto.com/scale-invariance/): statistics that are not easily changed by transformations, like simple data shifts.
- **Shrinkage** (https://www.statisticshowto.com/shrinkage-estimator/): a raw estimate that's improved by combining it with other information. See also: The James-Stein estimator.
- **Sufficient** (https://www.statisticshowto.com/sufficient-statistic/): a statistic that estimates the population parameter as well as if you knew all of the data in all possible samples.
- **UnBiased** (https://www.statisticshowto.com/unbiased/): an accurate statistic that neither underestimates nor overestimates.

Confidence Intervals

Statisticians use a confidence interval to express the precision and uncertainty associated with a particular sampling method. A confidence interval consists of three parts.

- A confidence level.
- A statistic.
- A margin of error.



The confidence level describes the uncertainty of a sampling method. The statistic and the margin of error define an interval estimate that describes the precision of the method. The interval estimate of a confidence interval is defined by the sample statistic + margin of error.

For example, suppose we compute an interval estimate of a population parameter. We might describe this interval estimate as a 95% confidence interval. This means that if we used the same sampling method to select different samples and compute different interval estimates, the true population parameter would fall within a range defined by the sample statistic + margin of error 95% of the time.

Confidence intervals are preferred to point estimates, because confidence intervals indicate (a) the precision of the estimate and (b) the uncertainty of the estimate.

Formally, Confidence Interval is defined as:

$$\text{C.I} = \bar{X} \pm Z_{\alpha/2}\, \sigma/\sqrt{n}$$

whereas, **X̄** = the sample mean

Za/2= Z value for desired confidence level α

σ = the population standard deviation

For an alpha value of 0.95 i.e 95% confidence interval, z=1.96.

Confidence Level

The probability part of a confidence interval is called a confidence level. The confidence level describes the likelihood that a particular sampling method will produce a confidence interval that includes the true population parameter.

Here is how to interpret a confidence level. Suppose we collected all possible samples from a given population, and computed confidence intervals for each sample. Some confidence intervals would include the true population parameter; others would not. A 95% confidence level means that 95% of the intervals contain the true population parameter; a 90% confidence level means that 90% of the intervals contain the population parameter; and so on.

Margin of Error

In a confidence interval, the range of values above and below the sample statistic is called the margin of error.

For example, suppose the local newspaper conducts an election survey and reports that the independent candidate will receive 30% of the vote. The newspaper states that the survey had a 5% margin of error and a confidence level of 95%. These findings result in the following confidence interval: We are 95% confident that the independent candidate will receive between 25% and 35% of the vote.

Note: Many public opinion surveys report interval estimates, but not confidence intervals. They provide the margin of error, but not the confidence level. To clearly interpret survey results you need to know both! We are much more likely to accept survey findings if the confidence level is high (say, 95%) than if it is low (say, 50%).

Interesting points to note about Confidence Intervals:

- Confidence Intervals can be built with difference degrees of confidence suitable to a user’s needs like 70 %, 90% etc.
- Greater the sample size, smaller the Confidence Interval, i.e more accurate determination of population mean from the sample means.
- There are different confidence intervals for different sample means. For example, a sample mean of 40 will have a difference confidence interval from a sample mean of 45.
- By 95% Confidence Interval, we do not mean that – The probability of a population mean to lie in an interval is 95%. Instead, 95% C.I means that 95% of the Interval estimates will contain the population statistic.

Many people do not have right knowledge about confidence interval and often interpret it incorrectly. So, I would like you to take your time visualizing the 4th argument and let it sink in.

Too many technical terms has been introduced. Let's test our understanding.

Which of the following statements is true.

I. When the margin of error is small, the confidence level is high.
II. When the margin of error is small, the confidence level is low.
III. A confidence interval is a type of point estimate.
IV. A population mean is an example of a point estimate.

(A) I only
(B) II only
(C) III only
(D) IV only.
(E) None of the above.

Solution: The correct answer is (E). The confidence level is not affected by the margin of error. When the margin of error is small, the confidence level can low or high or anything in between. A confidence interval is a type of interval estimate, not a type of point estimate. A population mean is not an example of a point estimate; a sample mean is an example of a point estimate.

Practical example

Calculate the 95% confidence interval for a sample mean of 40 and sample standard deviation of 40 with sample size equal to 100.

Solution:

We know, z-value for 95% C.I is 1.96. Hence, Confidence Interval (C.I) is calculated as:

C.I= [{x(bar) – (z*s/√n)},{x(bar) – (z*s/√n)}]

C.I = [{40-(1.96*40/10},{ 40+(1.96*40/10)}]

C.I = [32.16, 47.84]

What is the Standard Error?

The standard error is an estimate of the standard deviation of a statistic. This lesson shows how to compute the standard error, based on sample data.

The standard error is important because it is used to compute other measures, like confidence intervals and margins of error.

Refer to this link to learn more about the notations and formulas: https://stattrek.com/estimation/standard-error.aspx?tutorial= (https://stattrek.com/estimation/standard-error.aspx?tutorial=)

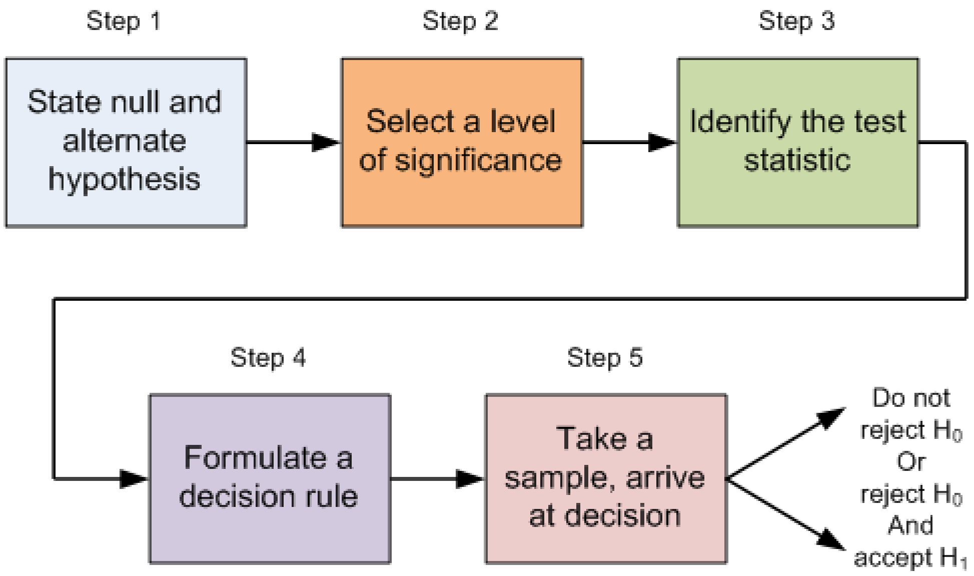
Now that we have been introduced to a lot of technical terms, we enter the territory of one of the most important and critical part of statistics for researchers and data science in general i.e. **Hypothesis Testing**

1. Hypothesis Testing

Before I get into the theoretical explanation, let us understand Hypothesis Testing by using a simple example.

Example: Class 8th has a mean score of 40 marks out of 100. The principal of the school decided that extra classes are necessary in order to improve the performance of the class. The class scored an average of 45 marks out of 100 after taking extra classes. Can we be sure whether the increase in marks is a result of extra classes or is it just random?

Five-Step Procedure for Testing a Hypothesis


Step 1
State null and alternate hypothesis
Step 2
Select a level of significance
Step 3
Identify the test statistic
Step 4
Formulate a decision rule
Step 5
Take a sample, arrive at decision
Do not reject H_0
Or
reject H_0
And
accept H_1

Hypothesis testing lets us identify that. It lets a sample statistic to be checked against a population statistic or statistic of another sample to study any intervention etc. Extra classes being the intervention in the above example.

Hypothesis testing is defined in two terms – *Null Hypothesis* and *Alternate Hypothesis*.

- **Null Hypothesis** being the sample statistic to be equal to the population statistic. For eg: The Null Hypothesis for the above example would be that the average marks after extra class are same as that before the classes.

- **Alternate Hypothesis** for this example would be that the marks after extra class are significantly different from that before the class.

Hypothesis Testing is done on different levels of confidence and makes use of z-score to calculate the probability. So for a 95% Confidence Interval, anything above the z-threshold for 95% would reject the null hypothesis.

Points to be noted:

1. We cannot accept the Null hypothesis, only reject it or fail to reject it.
2. As a practical tip, Null hypothesis is generally kept which we want to disprove. For eg: You want to prove that students performed better after taking extra classes on their exam. The Null Hypothesis, in this case, would be that the marks obtained after the classes are same as before the classes.

2. Types of Errors in Hypothesis Testing

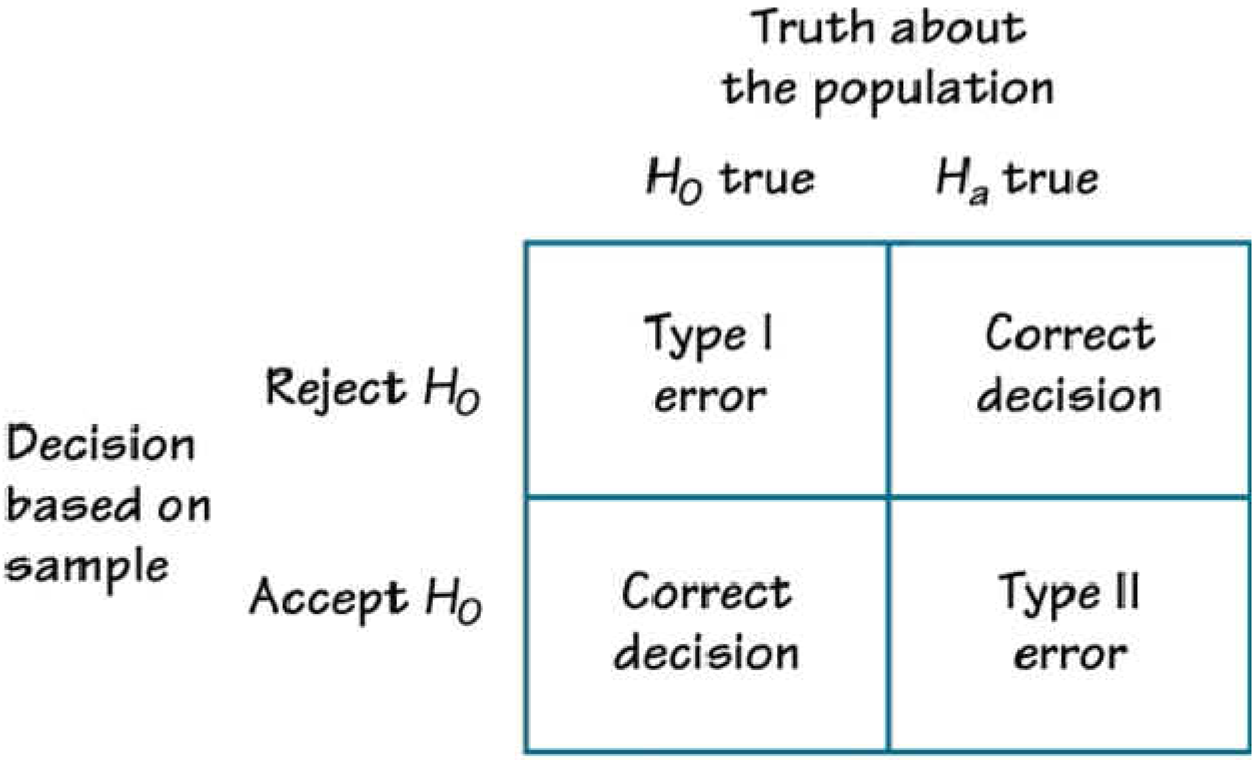
Now we have defined a basic Hypothesis Testing framework. It is important to look into some of the mistakes that are committed while performing Hypothesis Testing and try to classify those mistakes if possible.

Now, look at the Null Hypothesis definition above. What we notice at the first look is that it is a statement subjective to the tester like you and me and not a fact. That means there is a possibility that the Null Hypothesis can be true or false and we may end up committing some mistakes on the same lines.

There are two types of errors that are generally encountered while conducting Hypothesis Testing.

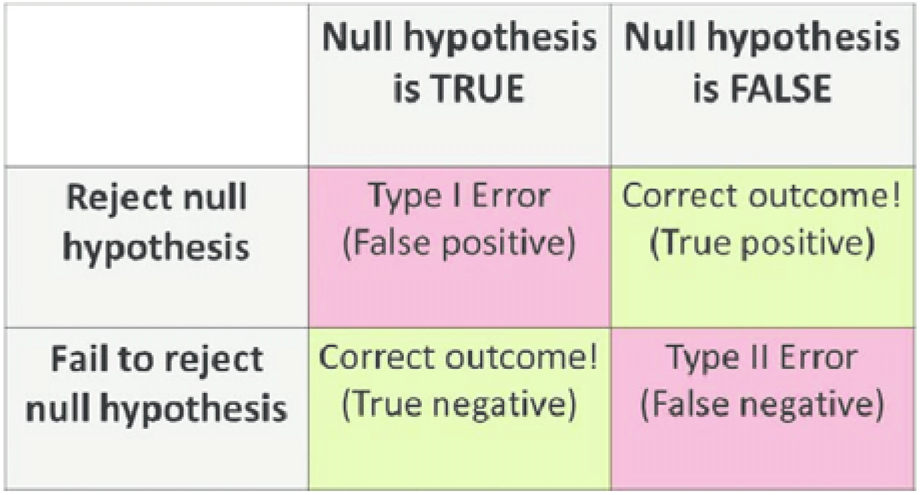
- **Type I error**: Look at the following scenario – A male human tested positive for being pregnant. Is it even possible? This surely looks like a case of False Positive. More formally, it is defined as the incorrect rejection of a True Null Hypothesis. The Null Hypothesis, in this case, would be – Male Human is not pregnant.
- **Type II error**: Look at another scenario where our Null Hypothesis is – A male human is pregnant and the test supports the Null Hypothesis. This looks like a case of False Negative. More formally it is defined as the acceptance of a false Null Hypothesis.

The below image will summarize the types of error :



		Truth about the population	
		H_0 true	H_a true
Decision based on sample	Reject H_0	Type I error	Correct decision
	Accept H_0	Correct decision	Type II error

Another very impressive examples is:



	Null hypothesis is TRUE	Null hypothesis is FALSE
Reject null hypothesis	Type I Error (False positive)	Correct outcome! (True positive)
Fail to reject null hypothesis	Correct outcome! (True negative)	Type II Error (False negative)

3. T-Tests

T-tests are very much similar to the z-scores, the only difference being that instead of the Population Standard Deviation, we now use the Sample Standard Deviation. The rest is same as before, calculating probabilities on basis of t-values.

The Sample Standard Deviation is given as:

$$S = \frac{\sqrt{\sum(x-\bar{x})^2}}{(n-1)}$$

where n-1 is the Bessel’s correction for estimating the population parameter.

Another difference between z-scores and t-values are that t-values are dependent on Degree of Freedom of a sample. Let us define what degree of freedom is for a sample.

The Degree of Freedom – It is the number of variables that have the choice of having more than one arbitrary value. For example, in a sample of size 10 with mean 10, 9 values can be arbitrary but the 1oth value is forced by the sample mean.

Points to note about the t-tests:

- Greater the difference between the sample mean and the population mean, greater the chance of rejecting the Null Hypothesis. Why? (We discussed this above.)

- Greater the sample size, greater the chance of rejection of Null Hypothesis.

4. Different types of t-tests

4.1 1-sample T-test

This is the same test as we described above. This test is used to:

- Determine whether the mean of a group differs from the specified value.
- Calculate a range of values that are likely to include the population mean.

For eg: A pizza delivery manager may perform a 1-sample t-test whether their delivery time is significantly different from that of the advertised time of 30 minutes by their competitors.

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{N}}}$$

where, X(bar) = sample mean

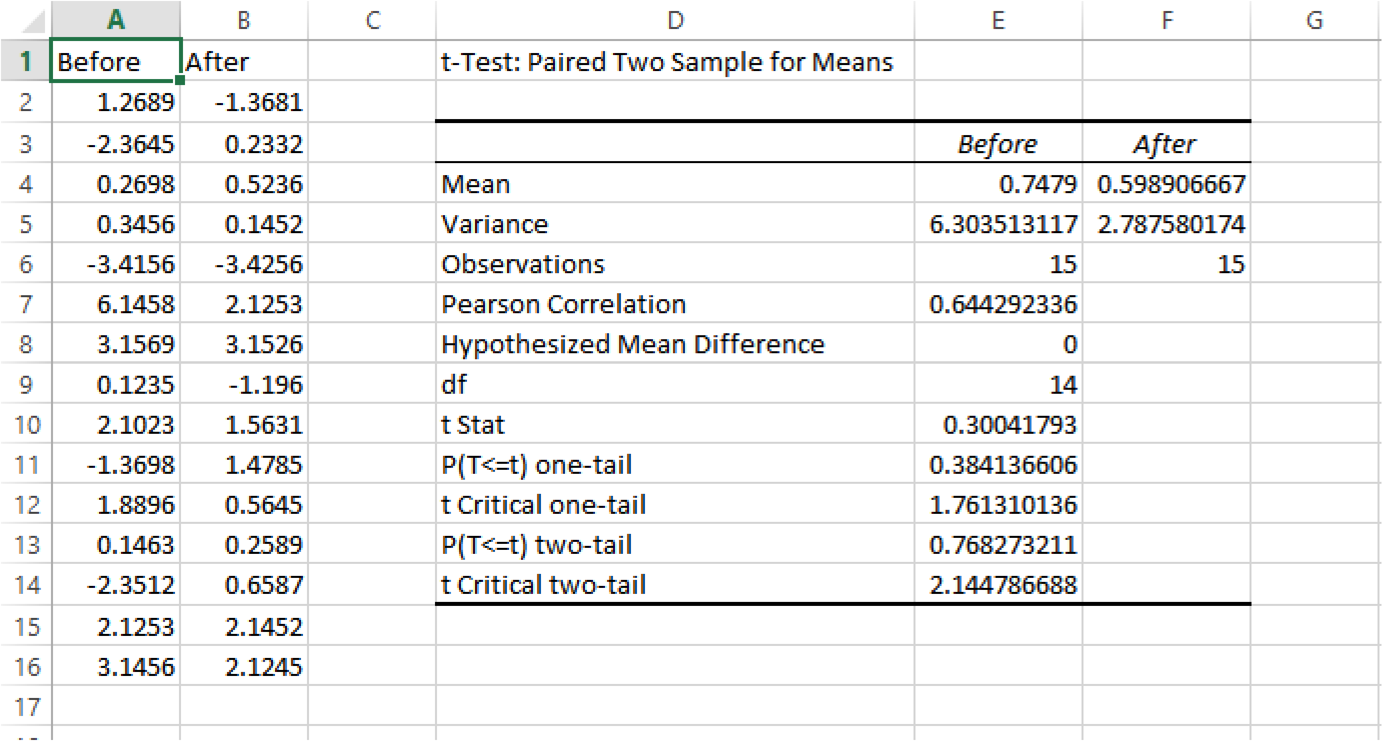
μ = population mean

s = sample standard deviation

N = sample size

4.2 Paired t-test

Paired t-test is performed to check whether there is a difference in mean after a treatment on a sample in comparison to before. It checks whether the Null hypothesis: The difference between the means is Zero, can be rejected or not.



	A	B	C	D	E	F	G
1	Before	After		t-Test: Paired Two Sample for Means			
2	1.2689	-1.3681					
3	-2.3645	0.2332			*Before*	*After*	
4	0.2698	0.5236		Mean	0.7479	0.598906667	
5	0.3456	0.1452		Variance	6.303513117	2.787580174	
6	-3.4156	-3.4256		Observations	15	15	
7	6.1458	2.1253		Pearson Correlation	0.644292336		
8	3.1569	3.1526		Hypothesized Mean Difference	0		
9	0.1235	-1.196		df	14		
10	2.1023	1.5631		t Stat	0.30041793		
11	-1.3698	1.4785		P(T<=t) one-tail	0.384136606		
12	1.8896	0.5645		t Critical one-tail	1.761310136		
13	0.1463	0.2589		P(T<=t) two-tail	0.768273211		
14	-2.3512	0.6587		t Critical two-tail	2.144786688		
15	2.1253	2.1452					
16	3.1456	2.1245					
17							

The above example suggests that the Null Hypothesis should not be rejected and that there is no significant difference in means before and after the intervention since p-value is not less than the alpha value (o.o5) and t stat is not less than t-critical.

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$

where, d (bar) = mean of the case wise difference between before and after,

Sd = standard deviation of the difference

n = sample size.

4.3 2-sample t-test

This test is used to determine:

- Determine whether the means of two independent groups differ.
- Calculate a range of values that is likely to include the difference between the population means.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

The above formula represents the 2 sample t-test and can be used in situations like to check whether two machines are producing the same output. The points to be noted for this test are:

1. The groups to be tested should be independent.
2. The groups' distribution should not be highly skewed. where, X1 (bar) = mean of the first group

S1 = represents 1st group sample standard deviation

N1 = represents the 1st group sample size.

4.4 Practical example

We will understand how to identify which t-test to be used and then proceed on to solve it. The other t-tests will follow the same argument.

Example: A population has mean weight of 68 kg. A random sample of size 25 has a mean weight of 70 with standard deviation =4. Identify whether this sample is representative of the population?

Step 0: Identifying the type of t-test

Number of samples in question = 1

Number of times the sample is in study = 1

Any intervention on sample = No

Recommended t-test = 1- sample t-test.

Had there been 2 samples, we would have opted for 2-sample t-test and if there would have been 2 observations on the same sample, we would have opted for paired t-test.

Step 1: State the Null and Alternate Hypothesis

- Null Hypothesis: The sample mean and population mean are same.
- Alternate Hypothesis: The sample mean and population mean are different.

Step 2: Calculate the appropriate test statistic

df = 25-1 =24

t= (70-68)/(4/√25) = 2.5

Now, for a 95% confidence level, t-critical (two-tail) for rejecting Null Hypothesis for 24 d.f is 2.06 . Hence, we can reject the Null Hypothesis and conclude that the two means are different.

You can use the t-test calculator here (https://www.danielsoper.com/statcalc/calculator.aspx?id=98)

5. ANOVA

ANOVA (Analysis of Variance) is used to check if at least one of two or more groups have statistically different means. Now, the question arises – Why do we need another test for checking the difference of means between independent groups? Why can we not use multiple t-tests to check for the difference in means?

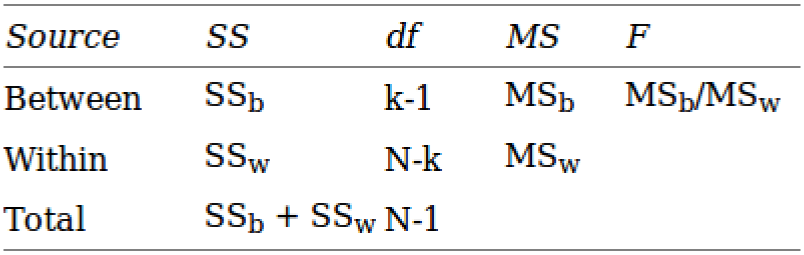
The answer is simple. Multiple t-tests will have a compound effect on the error rate of the result. Performing t-test thrice will give an error rate of ~15% which is too high, whereas ANOVA keeps it at 5% for a 95% confidence interval.

To perform an ANOVA, you must have a continuous response variable and at least one categorical factor with two or more levels. ANOVA requires data from approximately normally distributed populations with equal variances between factor levels. However, ANOVA procedures work quite well even if the normality assumption has been violated unless one or more of the distributions are highly skewed or if the variances are quite different.

ANOVA is measured using a statistic known as F-Ratio. It is defined as the ratio of Mean Square (between groups) to the Mean Square (within group).

Mean Square (between groups) = Sum of Squares (between groups) / degree of freedom (between groups)

Mean Square (within group) = Sum of Squares (within group) / degree of freedom (within group)



Source	*SS*	*df*	*MS*	*F*
Between	SS_b	k-1	MS_b	MS_b/MS_w
Within	SS_w	N-k	MS_w	
Total	$SS_b + SS_w$	N-1		

$$\sum_{j=1}^{p}\sum_{i=1}^{n_j}(X_{ij} - \bar{X}_j)^2 = SS_{w/in}$$

$$\sum_{j=1}^{p} n_j(\bar{X}_j - \bar{X})^2 = SS_{Betw}$$

Here, **p** = represents the number of groups

n = represents the number of observations in a group

Xj (bar) = represents the mean of a particular group

X (bar) = represents the mean of all the observations

Now, let us understand the degree of freedom for within group and between groups respectively.

Between groups : If there are k groups in ANOVA model, then k-1 will be independent. Hence, k-1 degree of freedom.

Within groups : If N represents the total observations in ANOVA (∑n over all groups) and k are the number of groups then, there will be k fixed points. Hence, N-k degree of freedom.

5.1 Steps to perform ANOVA

1. Hypothesis Generation
   - Null Hypothesis : Means of all the groups are same
   - Alternate Hypothesis : Mean of at least one group is different
2. Calculate within group and between groups variability
3. Calculate F-Ratio
4. Calculate probability using F-table
5. Reject/fail to Reject Null Hypothesis

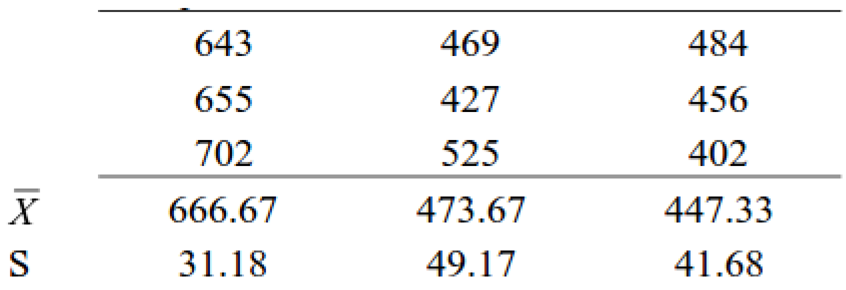
There are various other forms of ANOVA too like Two-way ANOVA, MANOVA, ANCOVA etc. but One-Way ANOVA suffices the requirements of this course.

Practical applications of ANOVA in modeling are:

1. Identifying whether a categorical variable is relevant to a continuous variable.
2. Identifying whether a treatment was effective to the model or not.

5.2 Practical Example

Suppose there are 3 chocolates in town and their sweetness is quantified by some metric (S). Data is collected on the three chocolates. You are given the task to identify whether the mean sweetness of the 3 chocolates are different. The data is given as below:



	643	469	484
	655	427	456
	702	525	402
$\overline{X}$	666.67	473.67	447.33
S	31.18	49.17	41.68

Here, first we have calculated the sample mean and sample standard deviation for you. Now we will proceed step-wise to calculate the F-Ratio (ANOVA statistic).

Step 1: Stating the Null and Alternate Hypothesis

Null Hypothesis: Mean sweetness of the three chocolates are same.

Alternate Hypothesis: Mean sweetness of at least one of the chocolates is different.

Step 2: Calculating the appropriate ANOVA statistic

In this part, we will be calculating SS(B), SS(W), SS(T) and then move on to calculate MS(B) and MS(W). The thing to note is that,

Total Sum of Squares [SS(t)] = Between Sum of Squares [SS(B)] + Within Sum of Squares [SS(W)].

So, we need to calculate any two of the three parameters using the data table and formulas given above.

As, per the formula above, we need one more statistic i.e Grand Mean denoted by X(bar) in the formula above.

X bar = (643+655+702+469+427+525+484+456+402)/9 = 529.22

SS(B)=[3*(666.67-529.22)^2]+ [3*(473.67-529.22)^2]+[3*(447.33-529.22)^2] = 86049.55

SS (W) = [(643-666.67)^2+(655-666.67)^2+(702-666.67)^2] + [(469-473.67)^2+(427-473.67)^2+(525-473.67)^2] + [(484-447.33)^2+(456-447.33)^2+(402-447.33)^2]= 10254

MS(B) = SS(B) / df(B) = 86049.55 / (3-1) = 43024.78

MS(W) = SS(W) / df(W) = 10254/(9-3) = 1709

F-Ratio = MS(B) / MS(W) = 25.17 .

Now, for a 95 % confidence level, F-critical to reject Null Hypothesis for degrees of freedom(2,6) is 5.14 but we have 25.17 as our F-Ratio.

So, we can confidently reject the Null Hypothesis and come to a conclusion that at least one of the chocolate has a mean sweetness different from the others.

You can use the F-calculator here (http://stattrek.com/online-calculator/f-distribution.aspx).

Note: ANOVA only lets us know the means for different groups are same or not. It doesn’t help us identify which mean is different.To know which group mean is different, we can use another test know as Least Significant Difference Test.

6. Chi-square Goodness of Fit Test

Sometimes, the variable under study is not a continuous variable but a categorical variable. Chi-square test is used when we have one single categorical variable from the population.

Let us understand this with help of an example. Suppose a company that manufactures chocolates, states that they manufacture 30% dairy milk, 60% temptation and 10% kit-kat. Now suppose a random sample of 100 chocolates has 50 dairy milk, 45 temptation and 5 kitkats. Does this support the claim made by the company?

Let us state our Hypothesis first.

- Null Hypothesis: The claims are True
- Alternate Hypothesis: The claims are False.

Chi-Square Test is given by:

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

where, **Oi** = sample or observed values

Ei = population values

The summation is taken over all the levels of a categorical variable.

Ei = [n * Pi] Expected value of a level (i) is equal to the product of sample size and percentage of it in the population.

Let us now calculate the Expected values of all the levels.

E (dairy milk)= 100 * 30% = 30

E (temptation) = 100 * 60% =60

E (kitkat) = 100 * 10% = 10

Calculating chi-square = [(50-30)^2/30+(45-60)^2/60+(5-10)^2/10] =19.58

Now, checking for p (chi-square >19.58) using chi-square calculator (http://stattrek.com/online-calculator/chi-square.aspx), we get p=0.0001. This is significantly lower than the alpha(0.05).

So we reject the Null Hypothesis.

7. Correlation

Correlation is used to test relationships between quantitative variables. In other words, it’s a measure of how things are related. The study of how variables are correlated is called correlation analysis.

Some examples of data that have a high correlation:

- Your caloric intake and your weight.
- Your eye color and your relatives’ eye colors.
- The amount of time your study and your GPA.

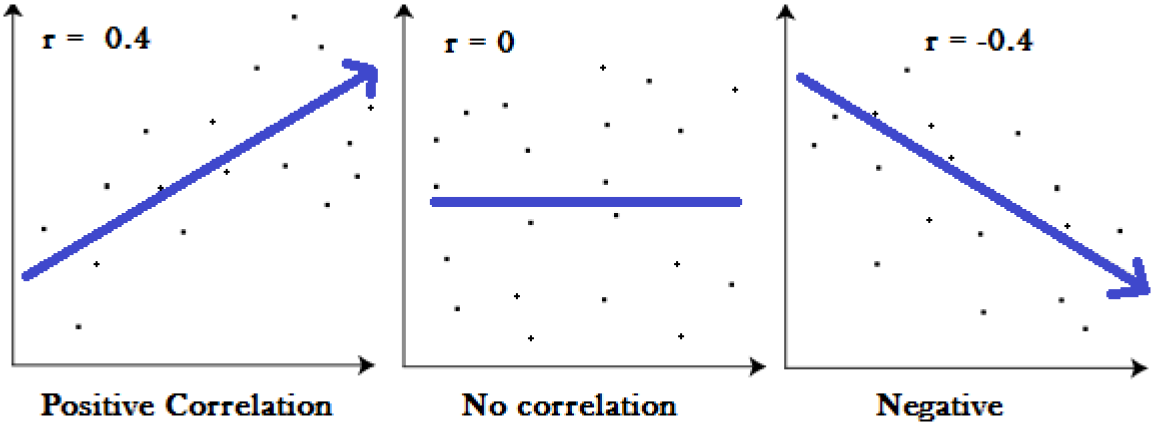
Some examples of data that have a low correlation (or none at all):

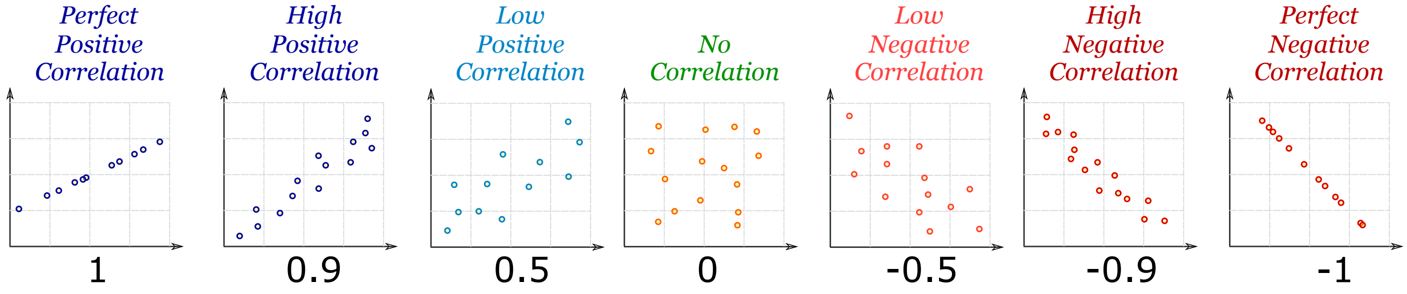
- Your sexual preference and the type of cereal you eat.
- A dog’s name and the type of dog biscuit they prefer.
- The cost of a car wash and how long it takes to buy a soda inside the station.

Correlations are useful because if you can find out what relationship variables have, you can make predictions about future behavior. Knowing what the future holds is very important in the social sciences like government and healthcare. Businesses also use these statistics for budgets and business plans.

7.1 The Correlation Coefficient

A correlation coefficient is a way to put a value to the relationship. Correlation coefficients have a value of between -1 and 1. A “0” means there is no relationship between the variables at all, while -1 or 1 means that there is a perfect negative or positive correlation (negative or positive correlation here refers to the type of graph the relationship will produce).





7.2 Types

The most common correlation coefficient is the Pearson Correlation Coefficient. It's used to test for linear relationships between data. In AP stats or elementary stats, the Pearson is likely the only one you'll be working with. However, you may come across others, depending upon the type of data you are working with. For example, Goodman and Kruskal's lambda coefficient is a fairly common coefficient. It can be symmetric, where you do not have to specify which variable is dependent, and asymmetric where the dependent variable is specified.

7.3 Correlation vs Covariance

This has been explained very well by one of my connection on Linkedin - Chayan Kathuria (https://www.linkedin.com/in/chayankathuria/)

Correlation and Covariance are often confused with each other. They are similar but tell different characteristics of the data. ⁉

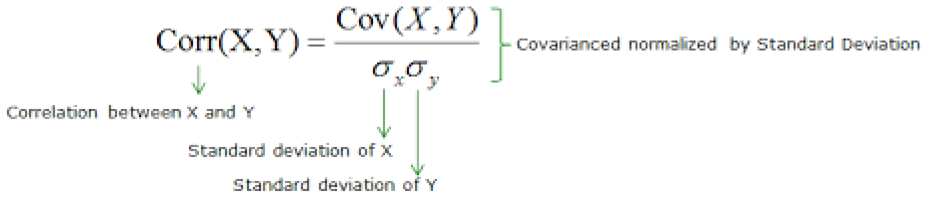
Correlation and Covariance both the terms measure the relationship and the dependency between two variables. 💡

Covariance defines the direction of the variance or the relationship between 2 variables. If Covariance is 0, it means that the variables have no variance with respect to each other. This would mean they are not related and provide completely different information about data.

If it is positive, it means that one variable increases as the other increases (directly proportional) and if it is negative, then one of the variables decreases as the other increases (inversely proportional).📈

Correlation, on the other hand, also defines the magnitude of the relation between 2 variables. Correlation is nothing but the covariance divided by the standard deviations of the variables. Hence, the value of correlation lies between -1 and 1. 📉

-1 being a complete inverse proportionality and 1 being a complete direct proportionality. And such variables won't add any more information to the data.


$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_x \sigma_y}$
Covarianced normalized by Standard Deviation
Correlation between X and Y
Standard deviation of X
Standard deviation of Y

Another very detailed video explanation on Correlation and Covariance by one of my inspiration - Josh Starmer

1. Covariance - https://www.youtube.com/watch?v=qtaqvPAeEJY (https://www.youtube.com/watch?v=qtaqvPAeEJY)
2. Correlation - https://www.youtube.com/watch?v=xZ_z8KWkhXE (https://www.youtube.com/watch?v=xZ_z8KWkhXE)

Links to Refer:

https://www.analyticsvidhya.com/blog/2017/01/comprehensive-practical-guide-inferential-statistics-data-science/ (https://www.analyticsvidhya.com/blog/2017/01/comprehensive-practical-guide-inferential-statistics-data-science/) - Comprehensive & practical inferential statistics guide for Data Science (most of the above content is from this particular link)

https://statisticsbyjim.com/basics/descriptive-inferential-statistics/ (https://statisticsbyjim.com/basics/descriptive-inferential-statistics/) - Descriptive vs Inferential Statistics

Video Tutorials:

https://www.youtube.com/watch?v=ZxK7SXURFcM (https://www.youtube.com/watch?v=ZxK7SXURFcM) - Inferential Statistics by GreyAtom

https://www.youtube.com/watch?v=qBigTkBLU6g&list=PLblh5JKOoLUK0FLuzwntyYI10UQFUhsY9 (https://www.youtube.com/watch?v=qBigTkBLU6g&list=PLblh5JKOoLUK0FLuzwntyYI10UQFUhsY9) - Statistics Fundamentals Playlist by Joshua Starmer