Central Limit Theorem:

Central Limit Theorem says whether the population follows normal distribution or not, but sample means of randomly selecting (more than or equal to 30 data points) from the population follows normal distribution.

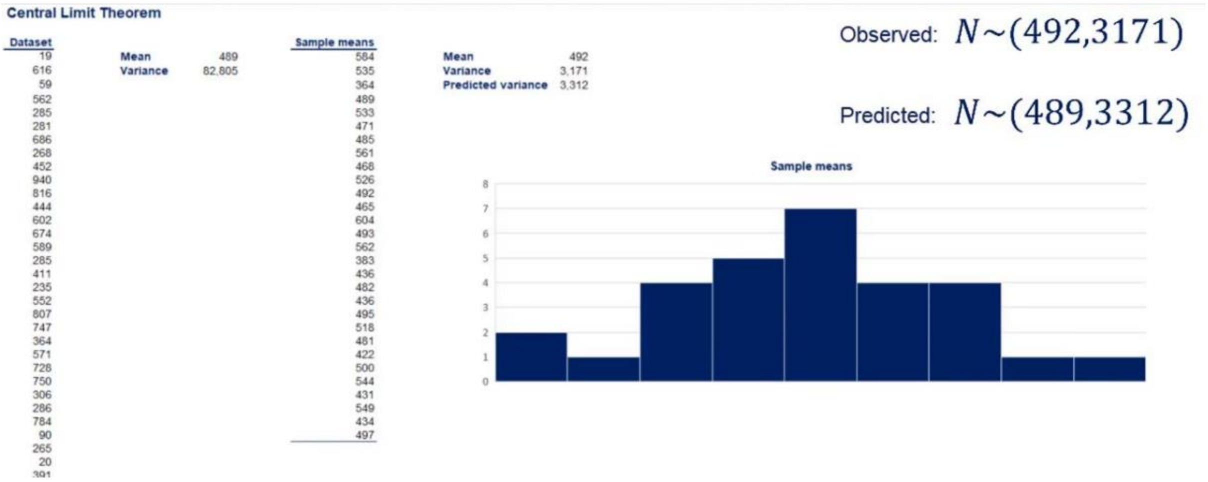
The size of the sample increases, the distribution of the mean among multiple samples will be normally distributed.

Therefore, as a sample size increases, the sample Mean and Standard Deviation will be closer in value to the population mean μ and standard deviation σ.

Why we need Central Limit Theorem:

As in the universe everything follows normal distribution. In the population/sample however the data is distributed, convert that into normal distribution by using Central Limit Theorem. If the data is normally distributed, we can easily apply all the predefined statistical functions on it.

When comparing to actual dataset and fairly large sample Means data, Means and Variances of actual dataset and Means and Variances of fairly large sample mean data are nearly same.



Central Limit Theorem

Dataset			Sample means		
19	Mean	489	584	Mean	492
616	Variance	82,805	535	Variance	3,171
59			364	Predicted variance	3,312
562			489		
285			533		
281			471		
686			485		
268			561		
452			468		
940			526		
816			492		
444			465		
602			604		
674			493		
589			562		
285			383		
411			436		
235			482		
552			436		
807			495		
747			518		
364			481		
571			422		
728			500		
750			544		
306			431		
286			549		
784			434		
90			497		
265					
20					
391					

Observed: $N \sim (492,3171)$

Predicted: $N \sim (489,3312)$

In the above example Mean of 489 and Variance of 82,805 are the population parameters. Mean of 492 and Variance (observed) of 3172 are the 25 Sample Means data.

$$N \sim \left(\mu , \frac{\sigma^2}{n}\right)$$

σ^2 is Population variance

n is sample size (number of data points)

If we calculate the predicted Variance with the above formulae,

= 82,805/25

= 3315

In the above example

- Actual dataset Mean of 489 and Predicted Variance 3312
- Sample Mean data of Mean of 492 and Predicted Variance 3171

That means Means and Variances of Actual dataset verse Sample Mean Data are nearly same, but it converts to Normal Distribution.

Standard Error: Standard Deviation of the distribution formed by the Sample Means or root of Variance of actual data divided by number of sample Means.

Population

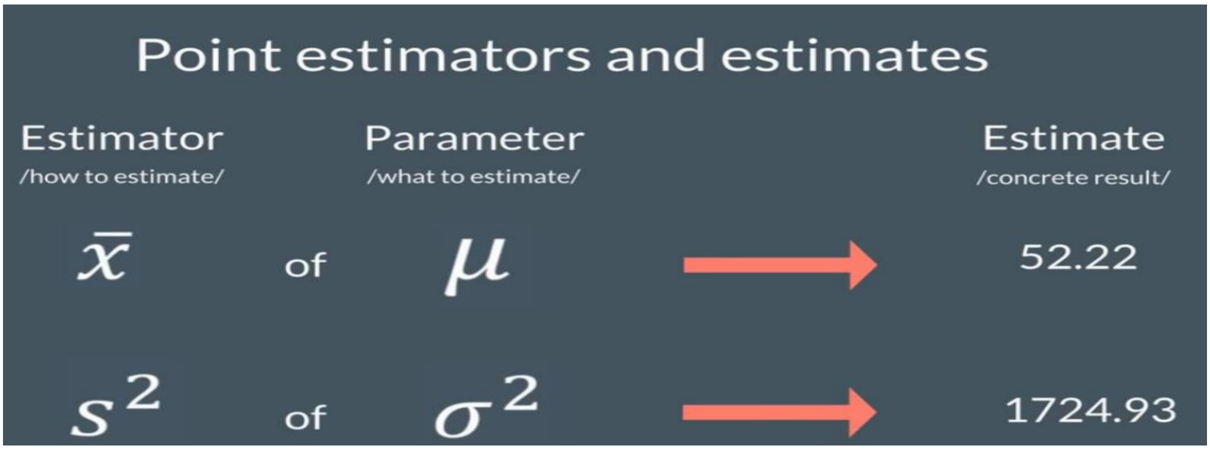
$$= \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

Sampling

$$\frac{s}{\sqrt{n}}$$

We use Standard Error in most of the Statistical Tests. With this we can estimate or approximate true Mean (Population Mean).

Standard Error decreases when Sample size increase that means bigger sample size will better approximation.



Point estimators and estimates

Estimator /how to estimate/		Parameter /what to estimate/		Estimate /concrete result/
$\bar{x}$	of	μ	$\longrightarrow$	52.22
s^2	of	σ^2	$\longrightarrow$	1724.93

The result (52.22, 1724.93) of Estimator (X bar and S square) from Population Parameter are called Estimate

Sample Mean is the Estimator of Population Mean and Sample Variables is the Estimator of Population Variances. These Estimators Estimate the results or predictions or values.

Two types of Estimates

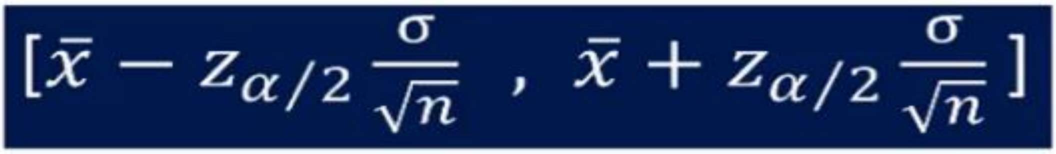
1) Point Estimates: It estimates the single number
2) Confidence Interval Estimates: It is the range or interval within which you expect the population parameter to be or estimate the population parameter within the range or interval.

Here we set the probability (confidence) percentage of the expected value within range (interval).

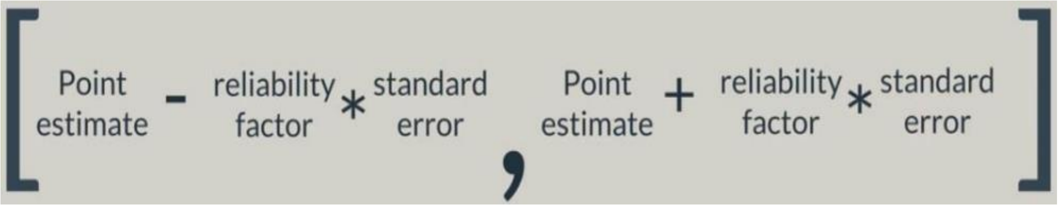
Confidence Interval are two types

1) **Population Variance known (Z distribution):**
Here we know Population Variance.
With the sample Mean and sample size we can find the confidence interval.

Formulae for one Population: (if we know Population Standard Deviation)



$$[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\ ,\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$$



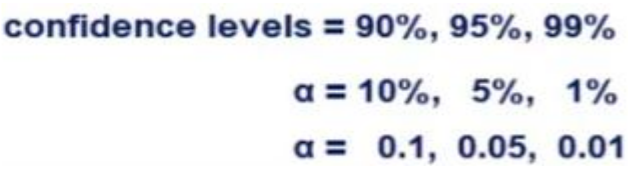
[Point estimate - reliability factor * standard error , Point estimate + reliability factor * standard error]

Confidence Level: The level at which our population parameter will be falling in that range.
Example if confidence level is 95% means, 95% of the cases population parameter will be following within that range (confidence interval) or we are 95% confidence that the population parameter (population average) will be within the range or confidence interval.

1 – α is always Confidence Level

α (Alpha) or non-confidence Level: The level at which our population parameter may not fall within that range Or probability of population parameter falling into the non-confidence interval.

Example if confidence level is 95% means, there are 5% of population parameter fall out of this range (confidence interval).



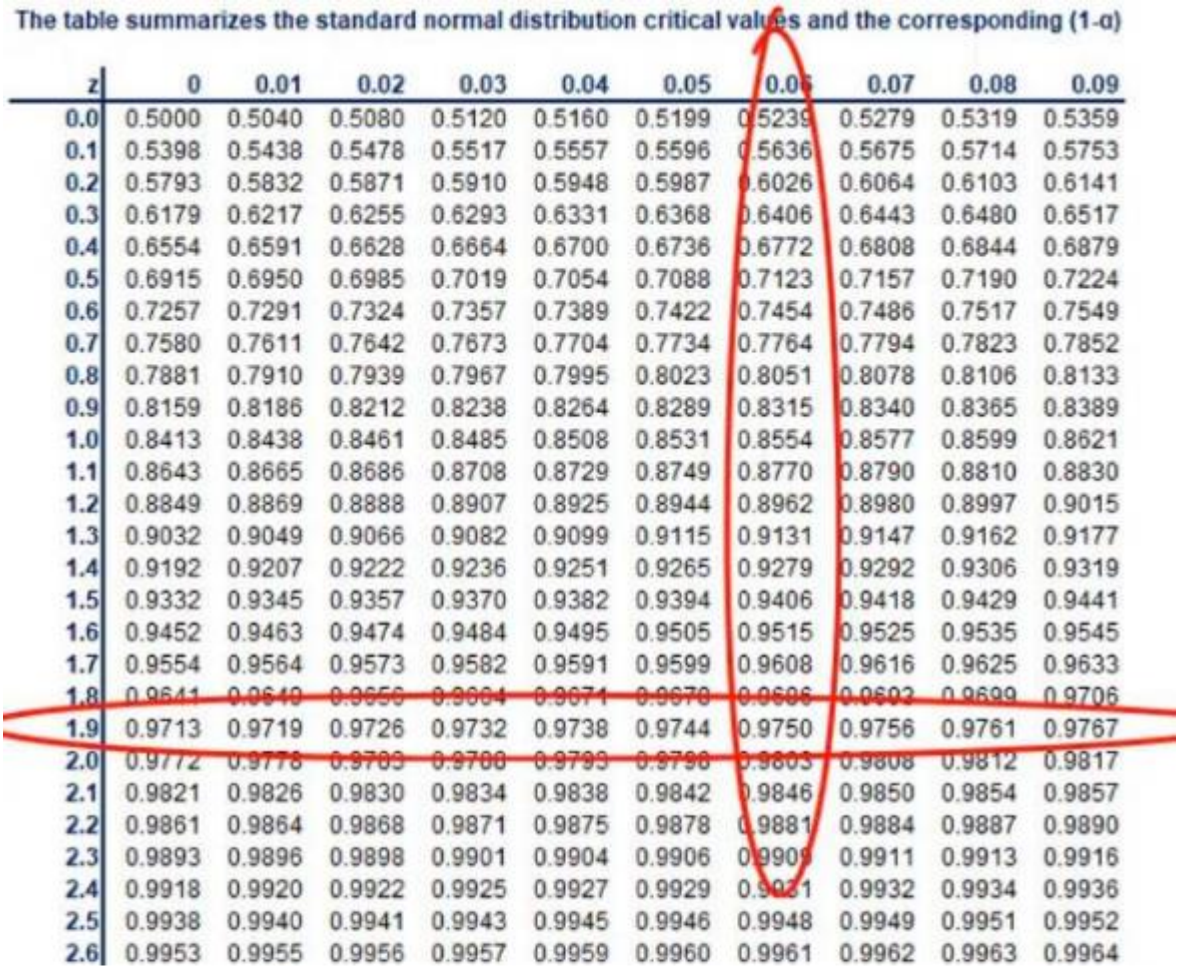
confidence levels = 90%, 95%, 99%

α = 10%, 5%, 1%

α = 0.1, 0.05, 0.01

In the above formulae if Confidence Interval 95% means α is 1 - 0.95 → 0.05 and α/2 is 0.05/2 → 0.025

If $z_{0.025}$ means, need to check the critical value for 0.975 (that is 1 – 0.025) in z distribution table



The table summarizes the standard normal distribution critical values and the corresponding (1-α)

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964

$$z_{0.025} = 1.9+0.06=1.96$$

$z_{\alpha/2}$ value for 95% confidence level is **1.96**

When comparing to confidence level from 95% to 99%, the interval range will be more from lower limit to upper limit.

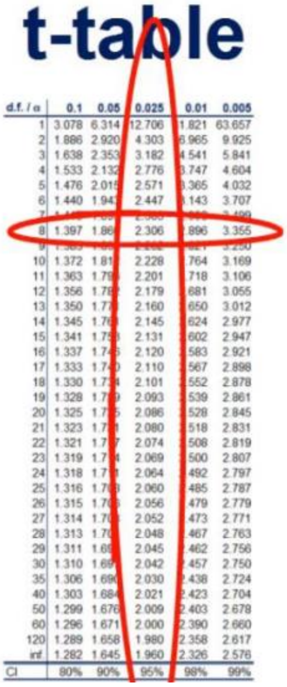
2) **Population Variance unknown (T Distribution):**
Here we don't know Population Variance or Standard Deviation. Most of the cases we don't know Population Standard Deviation hence we calculate the population mean with this. We can easily find the inferences through small samples.

Formulae for One Population:

$$\bar{x} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}$$

In the above formulae
t represents as T Distribution
n-1 is degrees of freedom (sample size -1)

In the following example consider n value as 9 sample size and α is 0.05



t-table

d.f. / α	0.1	0.05	0.025	0.01	0.005
1	3.078	6.314	12.706	1.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	[illegible]	[illegible]	[illegible]	[illegible]	3.499
8	1.397	1.860	2.306	2.896	3.355
9	[illegible]	[illegible]	[illegible]	[illegible]	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
35	1.306	1.690	2.030	2.438	2.724
40	1.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
inf	1.282	1.645	1.960	2.326	2.576
CI	80%	90%	95%	98%	99%

$$t_{n-1,\alpha/2}$$

$$t_{8,0.025} = 2.31$$

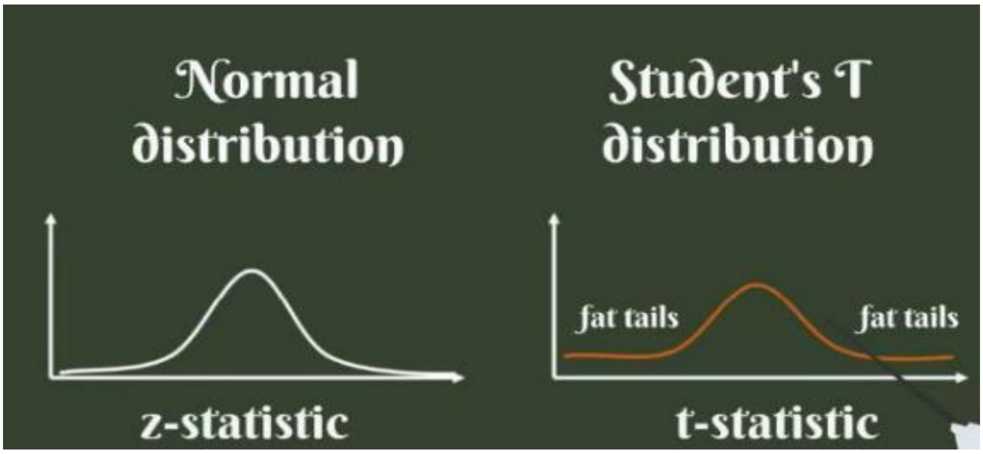
$t_{n-1,\alpha/2}$ **or Reliability Factor value is 2.31**

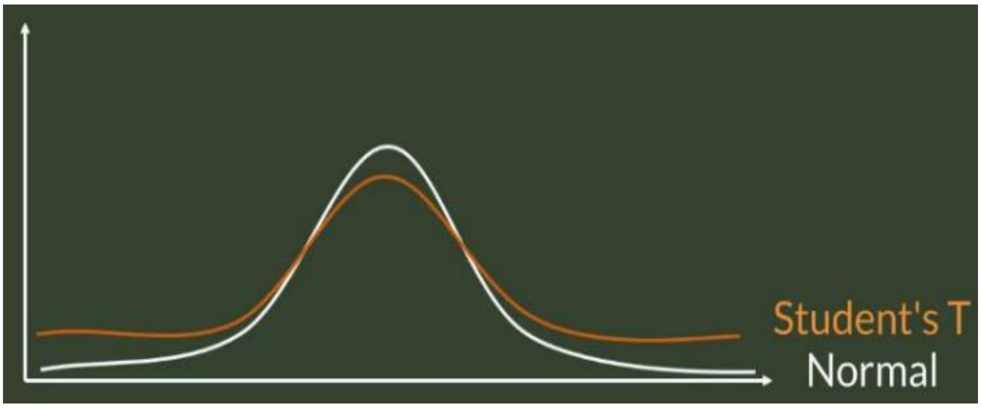
Normal Distribution Vs Student’s T Distribution:

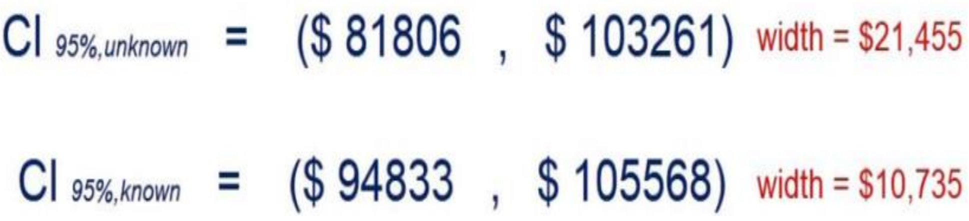
Normal Distribution also called as Standard Normal Distribution, Z Distribution, Z Statistics.

Student’s T Distribution also call as T Distribution or T Statistics.

Z Distribution	**T Distribution**
Narrow Tails	Fat Tails
More Height	Less Height
Less Width	More Width





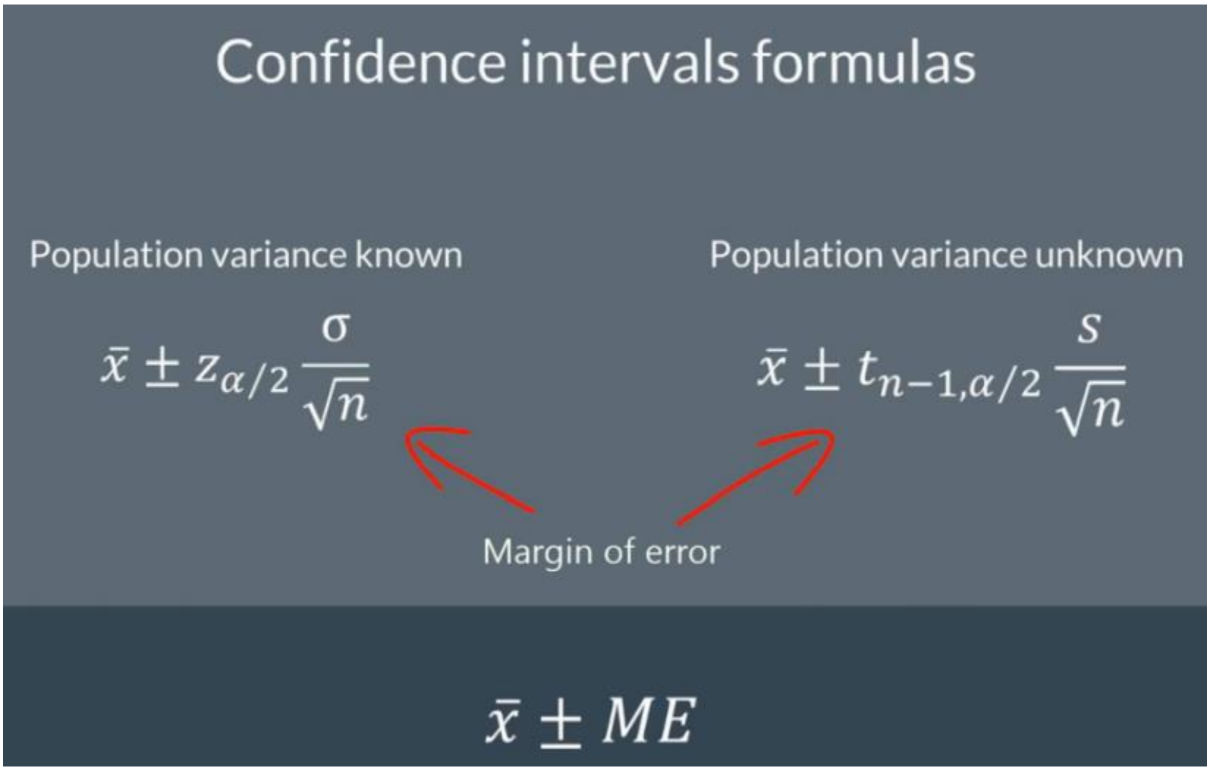


$CI_{95\%,unknown} = (\$\,81806\ ,\ \$\,103261)$ width = $21,455

$CI_{95\%,known} = (\$\,94833\ ,\ \$\,105568)$ width = $10,735

Margin of Error:

To get the confidence interval range, the value which we add or subtract to mean is called Margin of Error.



Confidence intervals formulas

Population variance known

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

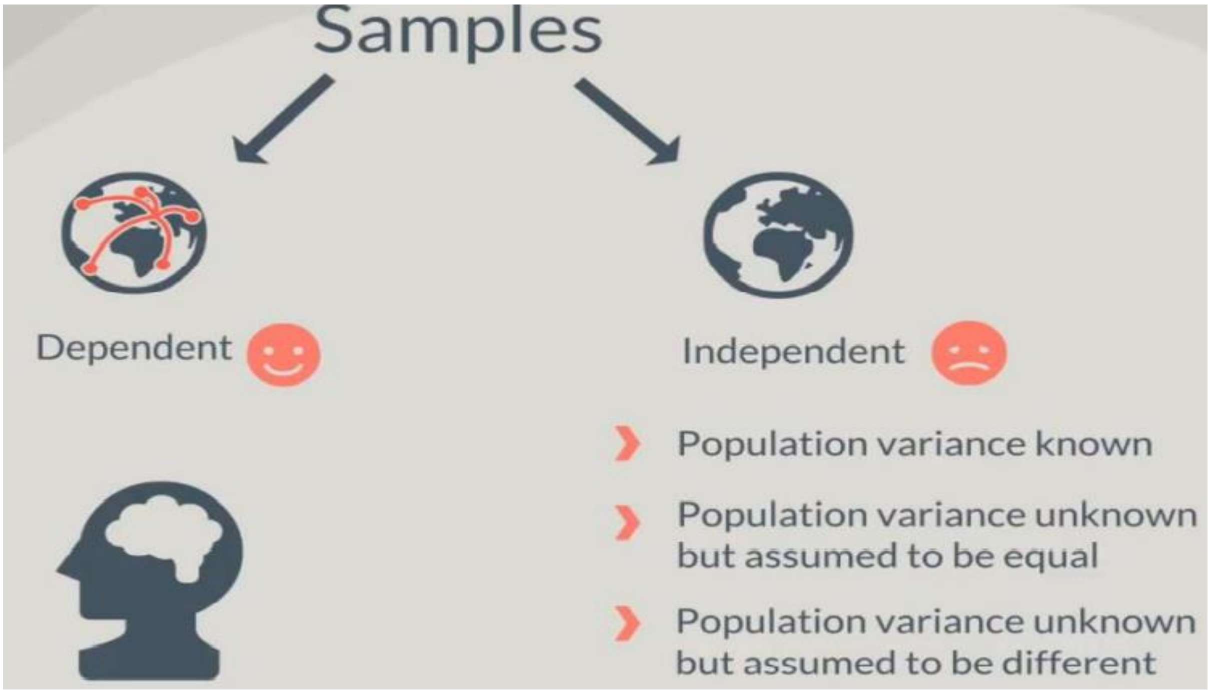
Population variance unknown

$$\bar{x} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}$$

Margin of error

$$\bar{x} \pm ME$$

Confidence Interval for Two Population:



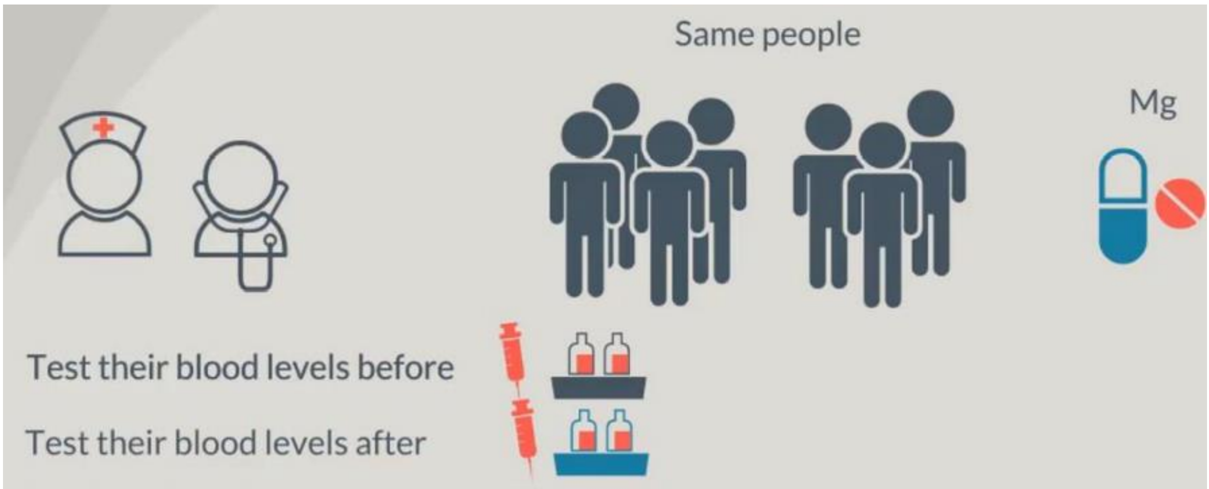
Samples

Dependent

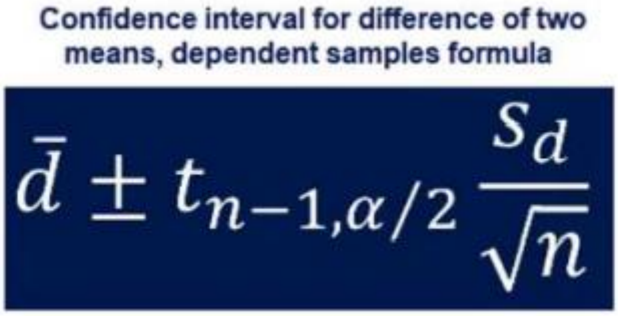
Independent

- Population variance known
- Population variance unknown but assumed to be equal
- Population variance unknown but assumed to be different

Dependent: Test the same sample before and after the action. Example we collected the blood samples from few people then gave the Magnesium medicine to them. After taking the medicine again collected blood samples from the same sample and test whether the medicine is working properly or not.

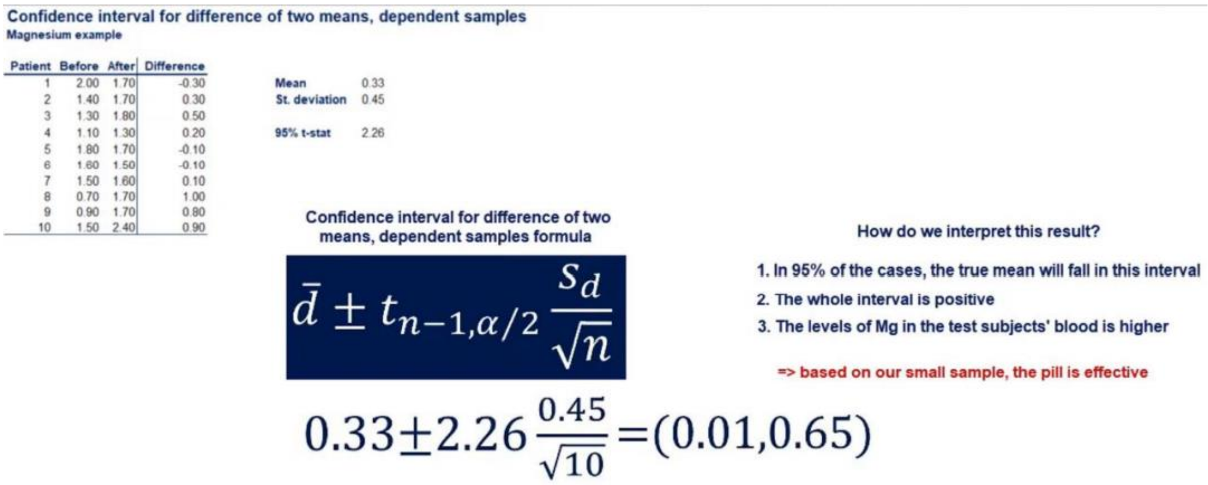


Formulae:



Confidence interval for difference of two means, dependent samples formula

$$\bar{d} \pm t_{n-1,\alpha/2} \frac{s_d}{\sqrt{n}}$$



Confidence interval for difference of two means, dependent samples

Magnesium example

Patient	Before	After	Difference
1	2.00	1.70	-0.30
2	1.40	1.70	0.30
3	1.30	1.80	0.50
4	1.10	1.30	0.20
5	1.80	1.70	-0.10
6	1.60	1.50	-0.10
7	1.50	1.60	0.10
8	0.70	1.70	1.00
9	0.90	1.70	0.80
10	1.50	2.40	0.90

Mean	0.33
St. deviation	0.45
95% t-stat	2.26

Confidence interval for difference of two means, dependent samples formula

$$\bar{d} \pm t_{n-1,\alpha/2} \frac{s_d}{\sqrt{n}}$$

$$0.33 \pm 2.26 \frac{0.45}{\sqrt{10}} = (0.01, 0.65)$$

How do we interpret this result?

1. In 95% of the cases, the true mean will fall in this interval
2. The whole interval is positive
3. The levels of Mg in the test subjects' blood is higher

=> based on our small sample, the pill is effective

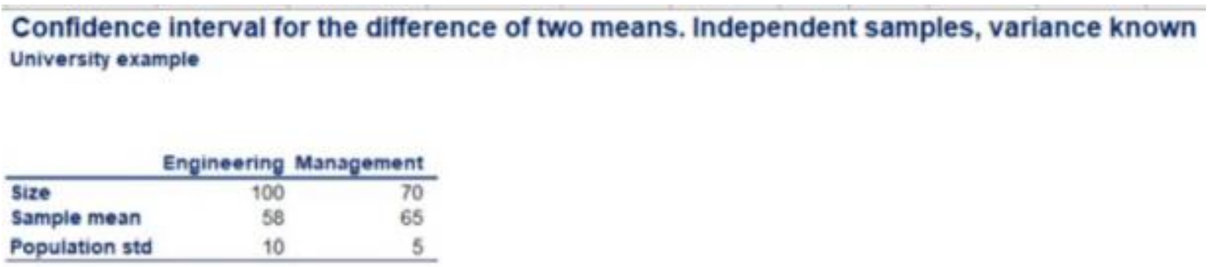
In the above example, calculated the confidence interval from Difference values (Before taking pill – After taking pill). Calculated the mean and Standard Deviation for Difference column values then apply the formulae of T Distribution (same the number of data points are less than 30). After applying the T Distribution, we got the positive values means. That means 95% of the cases, our pill(medicine) is effective.

Independent: Sample are different or independent.

Know Population Variances:

Calculate the Confidence interval between two independent variables with population Mean and Standard Deviations

Example:

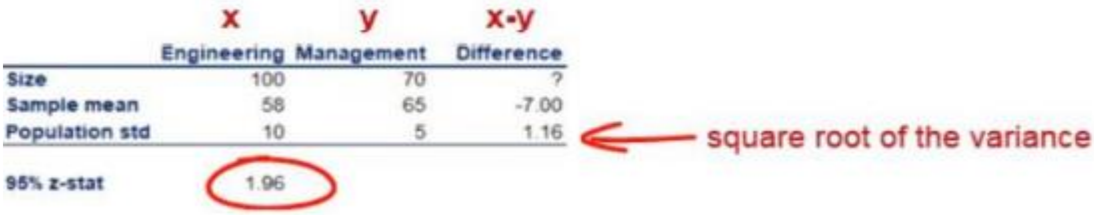


Confidence interval for the difference of two means. Independent samples, variance known

University example

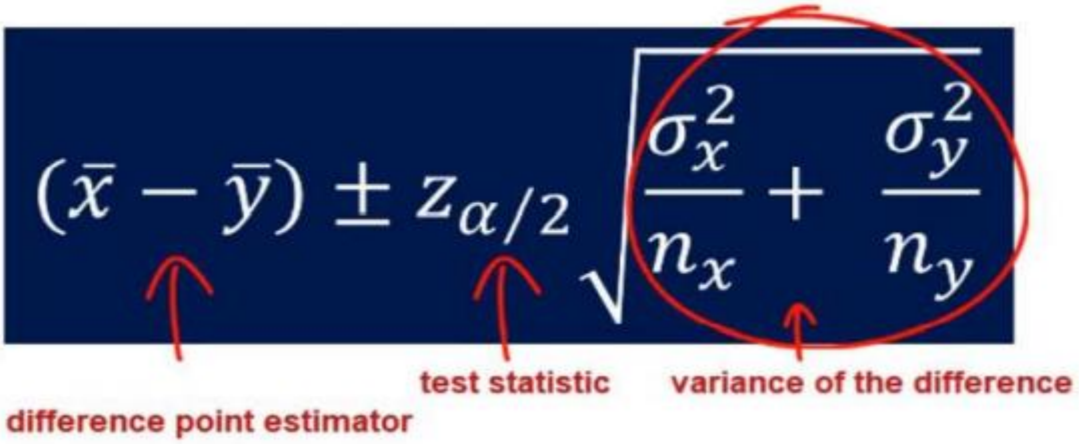
	Engineering	Management
Size	100	70
Sample mean	58	65
Population std	10	5

In the above example Engineering and Management students are independent, Sample sizes are different



	x	y	x-y
	Engineering	Management	Difference
Size	100	70	?
Sample mean	58	65	-7.00
Population std	10	5	1.16
95% z-stat	1.96		

square root of the variance

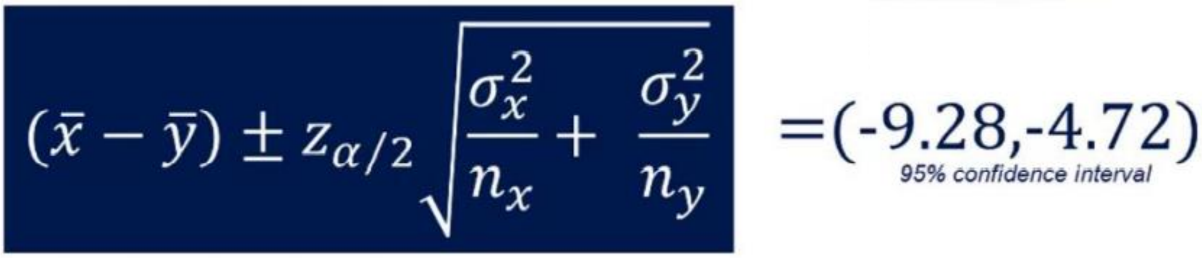


$$(\bar{x}-\bar{y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}$$

test statistic

variance of the difference

difference point estimator



$$(\bar{x}-\bar{y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}} =(-9.28,-4.72)$$

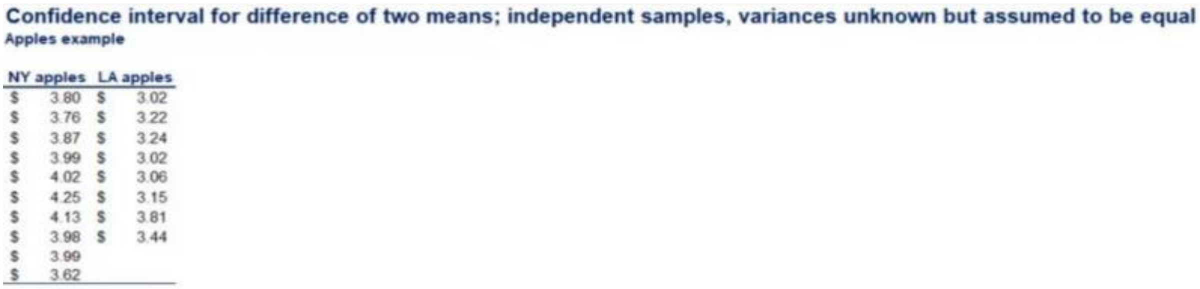
95% confidence interval

95% of the cases Management students get good marks than Engineering students. Engineering students’ average marks less than 9.28 marks to 4.72 marks. Even in the above example sample Mean difference is -7.

Unknow Population Variances but assumed to be equal:

Confidence interval for two different means of independent samples but assumed variances to be equal.

Example:

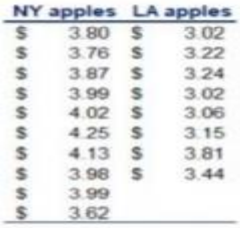


Confidence interval for difference of two means; independent samples, variances unknown but assumed to be equal

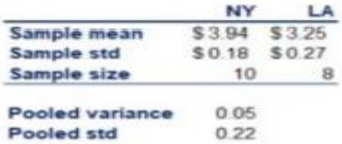
Apples example

NY apples	LA apples
$ 3.80	$ 3.02
$ 3.76	$ 3.22
$ 3.87	$ 3.24
$ 3.99	$ 3.02
$ 4.02	$ 3.06
$ 4.25	$ 3.15
$ 4.13	$ 3.81
$ 3.98	$ 3.44
$ 3.99	
$ 3.62	

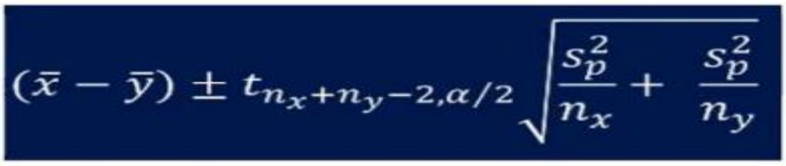
Above example showing apple prices for NY and LA. We don’t know the apple price variances for NY and LA but assumed as both price variances are equal.



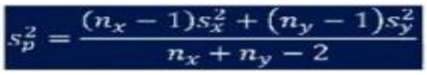
NY apples	LA apples
$ 3.80	$ 3.02
$ 3.76	$ 3.22
$ 3.87	$ 3.24
$ 3.99	$ 3.02
$ 4.02	$ 3.06
$ 4.25	$ 3.15
$ 4.13	$ 3.81
$ 3.98	$ 3.44
$ 3.99	
$ 3.62	



	NY	LA
Sample mean	$3.94	$3.25
Sample std	$0.18	$0.27
Sample size	10	8
Pooled variance	0.05	
Pooled std	0.22	

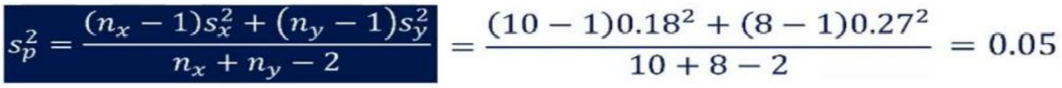


$$(\bar{x}-\bar{y}) \pm t_{n_x+n_y-2,\alpha/2}\sqrt{\frac{s_p^2}{n_x}+\frac{s_p^2}{n_y}}$$



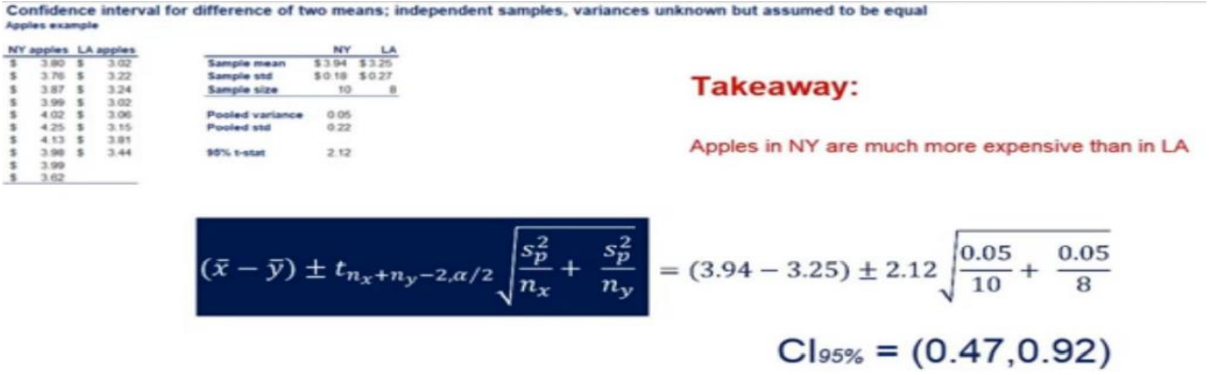
$$s_p^2 = \frac{(n_x-1)s_x^2+(n_y-1)s_y^2}{n_x+n_y-2}$$

Calculated the Mean and Standard Deviation for both the cities.



$$s_p^2 = \frac{(n_x-1)s_x^2+(n_y-1)s_y^2}{n_x+n_y-2} = \frac{(10-1)0.18^2+(8-1)0.27^2}{10+8-2} = 0.05$$

Also calculated Pooled Variance and Pooled Standard Deviation.



Confidence interval for difference of two means; independent samples, variances unknown but assumed to be equal

Apples example

NY apples	LA apples
$ 3.80	$ 3.02
$ 3.76	$ 3.22
$ 3.87	$ 3.24
$ 3.99	$ 3.02
$ 4.02	$ 3.06
$ 4.25	$ 3.15
$ 4.13	$ 3.81
$ 3.98	$ 3.44
$ 3.99	
$ 3.62	

	NY	LA
Sample mean	$3.94	$3.25
Sample std	$0.18	$0.27
Sample size	10	8
Pooled variance	0.05	
Pooled std	0.22	
95% t-stat	2.12	

Takeaway:

Apples in NY are much more expensive than in LA

$$(\bar{x}-\bar{y}) \pm t_{n_x+n_y-2,\alpha/2}\sqrt{\frac{s_p^2}{n_x}+\frac{s_p^2}{n_y}} = (3.94-3.25) \pm 2.12\sqrt{\frac{0.05}{10}+\frac{0.05}{8}}$$

$CI_{95\%} = (0.47, 0.92)$

After calculating the Confidence interval, concluded as 95% of the cases NY prices are more than LA with intervals of 0.47 to 0.92

For example, in LA apple price is $5 then in NY 95% of cases apple prices between $5.47 and $5.92.

Unknow Population Variances but assumed to be different:

Confidence interval for two different means of independent samples but assumed variances to be different.

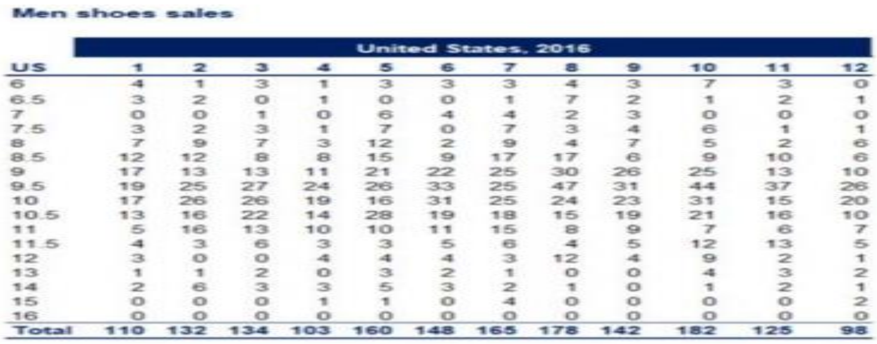
Example if we want to compare apples vs oranges with unknow variance. We can do but at present it is not in our scope

Practical Example:

One of the Shoe company facing inventory problem how many pairs of shoes need to by for each size and gender. If they buy less, all the customers may not get and if they buy more, the remaining inventory will pileup(accumulate) it will be over cost to the company.

If we find the number of shoes of each size that we are selling based on the historical data, we can order same from manufactures.

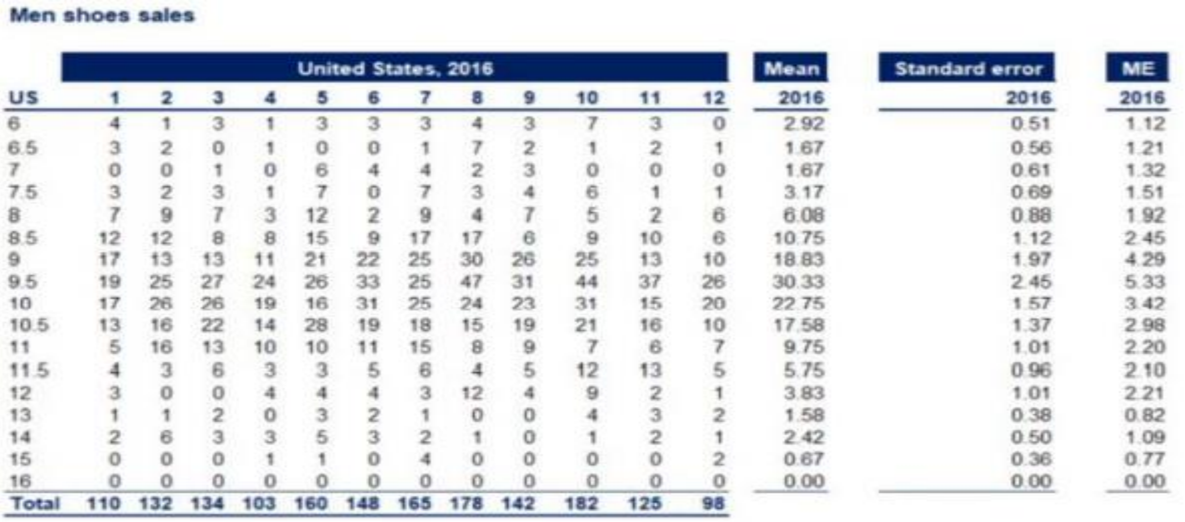
We have the following historical data of 12 months data.



Men shoes sales

	United States, 2016											
US	1	2	3	4	5	6	7	8	9	10	11	12
6	4	1	3	1	3	3	3	4	3	7	3	0
6.5	3	2	0	1	0	0	1	7	2	1	2	1
7	0	0	1	0	6	4	4	2	3	0	0	0
7.5	3	2	3	1	7	0	7	3	4	6	1	1
8	7	9	7	3	12	2	9	4	7	5	2	6
8.5	12	12	8	8	15	9	17	17	6	9	10	6
9	17	13	13	11	21	22	25	30	26	25	13	10
9.5	19	25	27	24	26	33	25	47	31	44	37	26
10	17	26	26	19	16	31	25	24	23	31	15	20
10.5	13	16	22	14	28	19	18	15	19	21	16	10
11	5	16	13	10	10	11	15	8	9	7	6	7
11.5	4	3	6	3	3	5	6	4	5	12	13	5
12	3	0	0	4	4	4	3	12	4	9	2	1
13	1	1	2	0	3	2	1	0	0	4	3	2
14	2	6	3	3	5	3	2	1	0	1	2	1
15	0	0	0	1	1	0	4	0	0	0	0	2
16	0	0	0	0	0	0	0	0	0	0	0	0
Total	110	132	134	103	160	148	165	178	142	182	125	98

Based on 12 months data calculate the Mean, Standard Deviation and ME.



Men shoes sales

	United States, 2016												Mean	Standard error	ME
US	1	2	3	4	5	6	7	8	9	10	11	12	2016	2016	2016
6	4	1	3	1	3	3	3	4	3	7	3	0	2.92	0.51	1.12
6.5	3	2	0	1	0	0	1	7	2	1	2	1	1.67	0.56	1.21
7	0	0	1	0	6	4	4	2	3	0	0	0	1.67	0.61	1.32
7.5	3	2	3	1	7	0	7	3	4	6	1	1	3.17	0.69	1.51
8	7	9	7	3	12	2	9	4	7	5	2	6	6.08	0.88	1.92
8.5	12	12	8	8	15	9	17	17	6	9	10	6	10.75	1.12	2.45
9	17	13	13	11	21	22	25	30	26	25	13	10	18.83	1.97	4.29
9.5	19	25	27	24	26	33	25	47	31	44	37	26	30.33	2.45	5.33
10	17	26	26	19	16	31	25	24	23	31	15	20	22.75	1.57	3.42
10.5	13	16	22	14	28	19	18	15	19	21	16	10	17.58	1.37	2.98
11	5	16	13	10	10	11	15	8	9	7	6	7	9.75	1.01	2.20
11.5	4	3	6	3	3	5	6	4	5	12	13	5	5.75	0.96	2.10
12	3	0	0	4	4	4	3	12	4	9	2	1	3.83	1.01	2.21
13	1	1	2	0	3	2	1	0	0	4	3	2	1.58	0.38	0.82
14	2	6	3	3	5	3	2	1	0	1	2	1	2.42	0.50	1.09
15	0	0	0	1	1	0	4	0	0	0	0	2	0.67	0.36	0.77
16	0	0	0	0	0	0	0	0	0	0	0	0	0.00	0.00	0.00
Total	110	132	134	103	160	148	165	178	142	182	125	98			

Calculated Standard Error with the following formulae

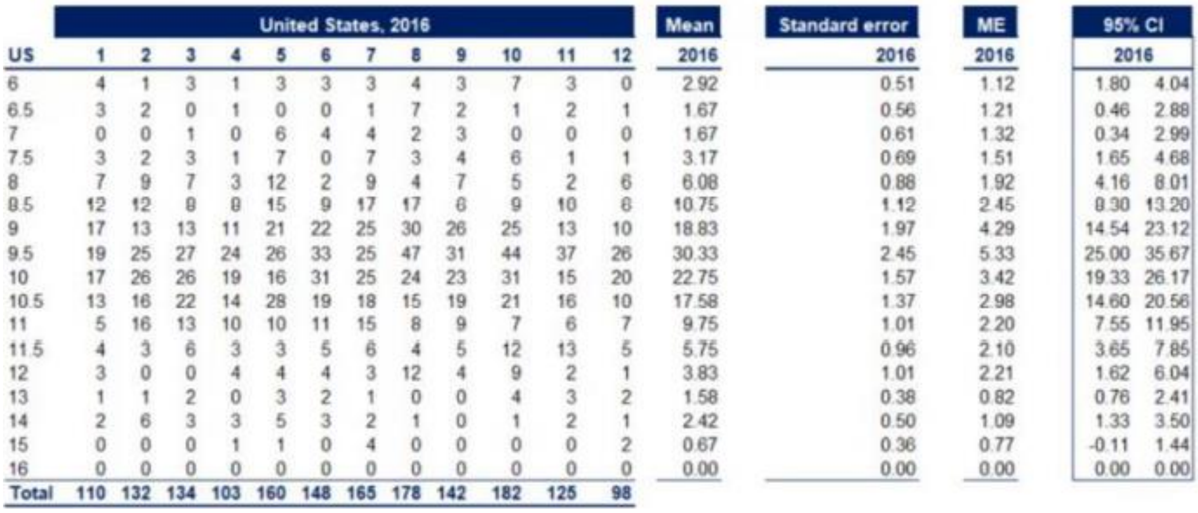
$$\frac{s}{\sqrt{n}}$$

Calculate the 95% confidence level of Margin of Error with the following formulae

$$t_{n-1,\alpha/2}\,\frac{s}{\sqrt{n}}$$

Finally apply the formulae $\bar{x} \pm t_{n-1,\alpha/2}\,\frac{s}{\sqrt{n}}$

We get the following confidence interval values for 95% confidence level.



	United States, 2016												Mean	Standard error	ME	95% CI	
US	1	2	3	4	5	6	7	8	9	10	11	12	2016	2016	2016	2016	
6	4	1	3	1	3	3	3	4	3	7	3	0	2.92	0.51	1.12	1.80	4.04
6.5	3	2	0	1	0	0	1	7	2	1	2	1	1.67	0.56	1.21	0.46	2.88
7	0	0	1	0	6	4	4	2	3	0	0	0	1.67	0.61	1.32	0.34	2.99
7.5	3	2	3	1	7	0	7	3	4	6	1	1	3.17	0.69	1.51	1.65	4.68
8	7	9	7	3	12	2	9	4	7	5	2	6	6.08	0.88	1.92	4.16	8.01
8.5	12	12	8	8	15	9	17	17	6	9	10	6	10.75	1.12	2.45	8.30	13.20
9	17	13	13	11	21	22	25	30	26	25	13	10	18.83	1.97	4.29	14.54	23.12
9.5	19	25	27	24	26	33	25	47	31	44	37	26	30.33	2.45	5.33	25.00	35.67
10	17	26	26	19	16	31	25	24	23	31	15	20	22.75	1.57	3.42	19.33	26.17
10.5	13	16	22	14	28	19	18	15	19	21	16	10	17.58	1.37	2.98	14.60	20.56
11	5	16	13	10	10	11	15	8	9	7	6	7	9.75	1.01	2.20	7.55	11.95
11.5	4	3	6	3	3	5	6	4	5	12	13	5	5.75	0.96	2.10	3.65	7.85
12	3	0	0	4	4	4	3	12	4	9	2	1	3.83	1.01	2.21	1.62	6.04
13	1	1	2	0	3	2	1	0	0	4	3	2	1.58	0.38	0.82	0.76	2.41
14	2	6	3	3	5	3	2	1	0	1	2	1	2.42	0.50	1.09	1.33	3.50
15	0	0	0	1	1	0	4	0	0	0	0	2	0.67	0.36	0.77	-0.11	1.44
16	0	0	0	0	0	0	0	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00
Total	110	132	134	103	160	148	165	178	142	182	125	98					

In the above 95% CI consider upper boundary value then round value to nearest zero. In the above example Men’s shoes 6 size can be sold 95% of cases in a month is 4 pairs. Similarly for shoe size 9.5 is 36 shoe pairs.