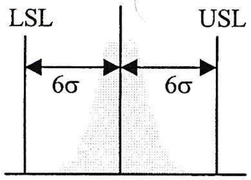
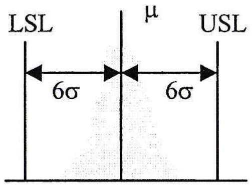
Confidence Intervals



Definition: Confidence Interval

A metric that gives an estimated range of values expected from a sampling that will include the actual value of the population, within a chosen level of certainty.



Confidence Interval for μ (mean)

The Tech Group
WE PROVIDE INNOVATIVE SOLUTIONS

Suppose we wish to know the mean of a population by taking a sample.

Our sample will only be an estimate of our mean, with:

- A degree of accuracy (or range)
- A level of confidence the estimate is within that accuracy range.

We can choose sample size, N

And one of the following:

Degree of accuracy or *level of confidence*

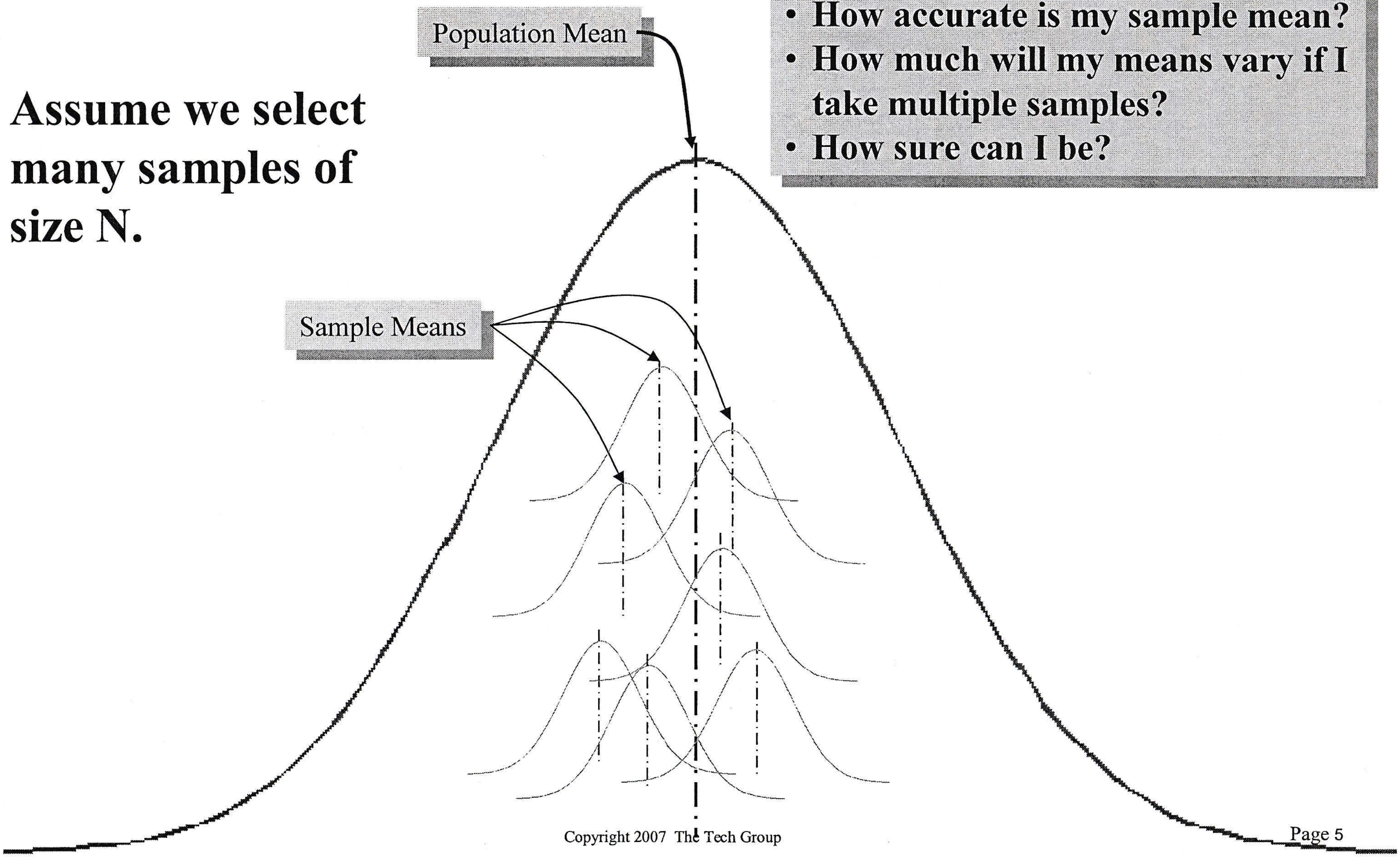
Whichever one we choose, the other one will be calculated.

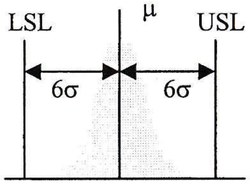
Population Mean vs. Sample Mean

The Tech Group
WE PROVIDE INNOVATIVE SOLUTIONS

Assume we select many samples of size N.

- **How accurate is my sample mean?**
- **How much will my means vary if I take multiple samples?**
- **How sure can I be?**



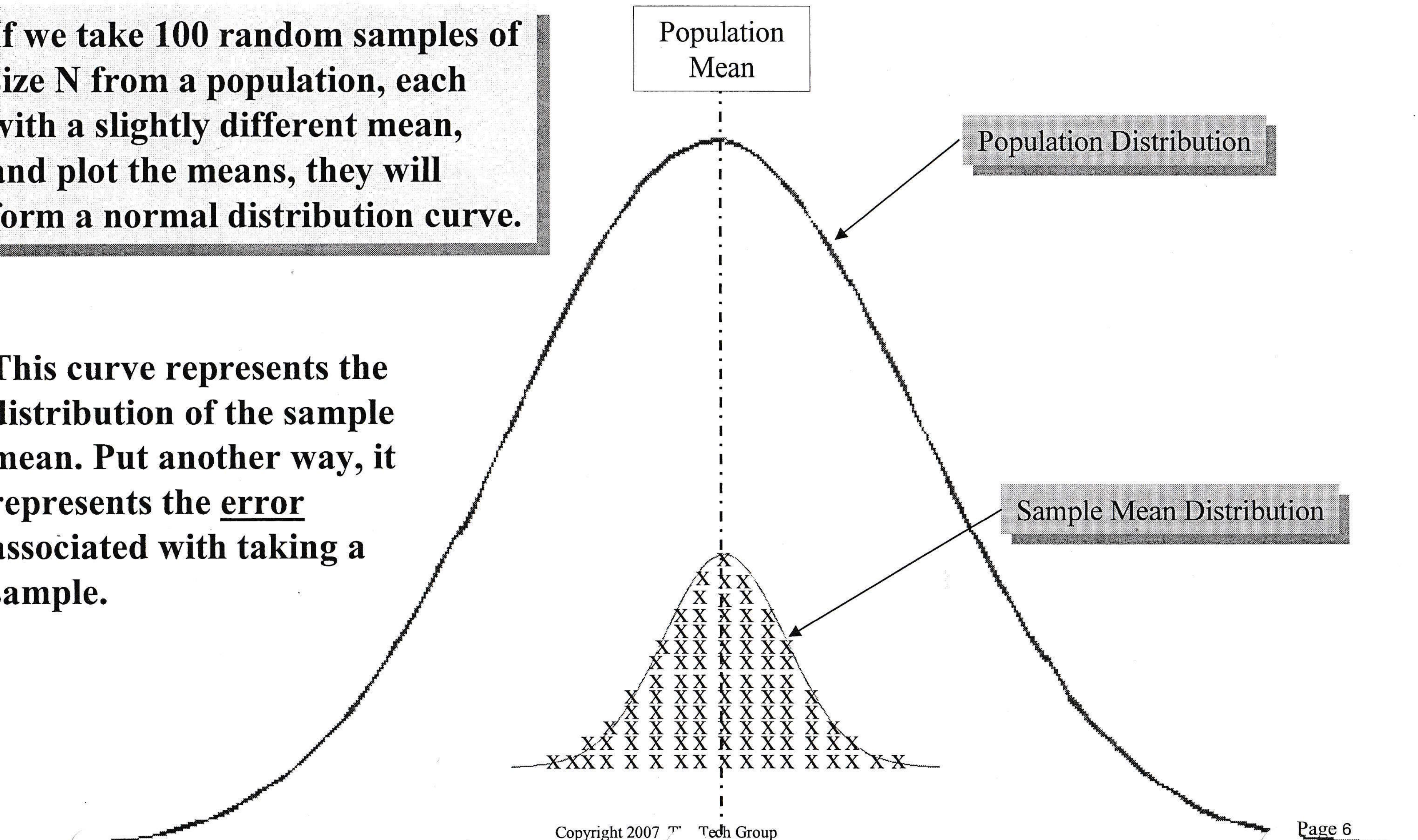


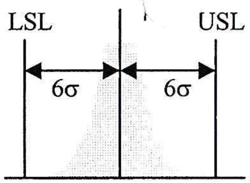
The Tech Group
WE PROVIDE INNOVATIVE SOLUTIONS

Population Mean vs. Sample Mean

If we take 100 random samples of size N from a population, each with a slightly different mean, and plot the means, they will form a normal distribution curve.

This curve represents the distribution of the sample mean. Put another way, it represents the <u>error</u> associated with taking a sample.

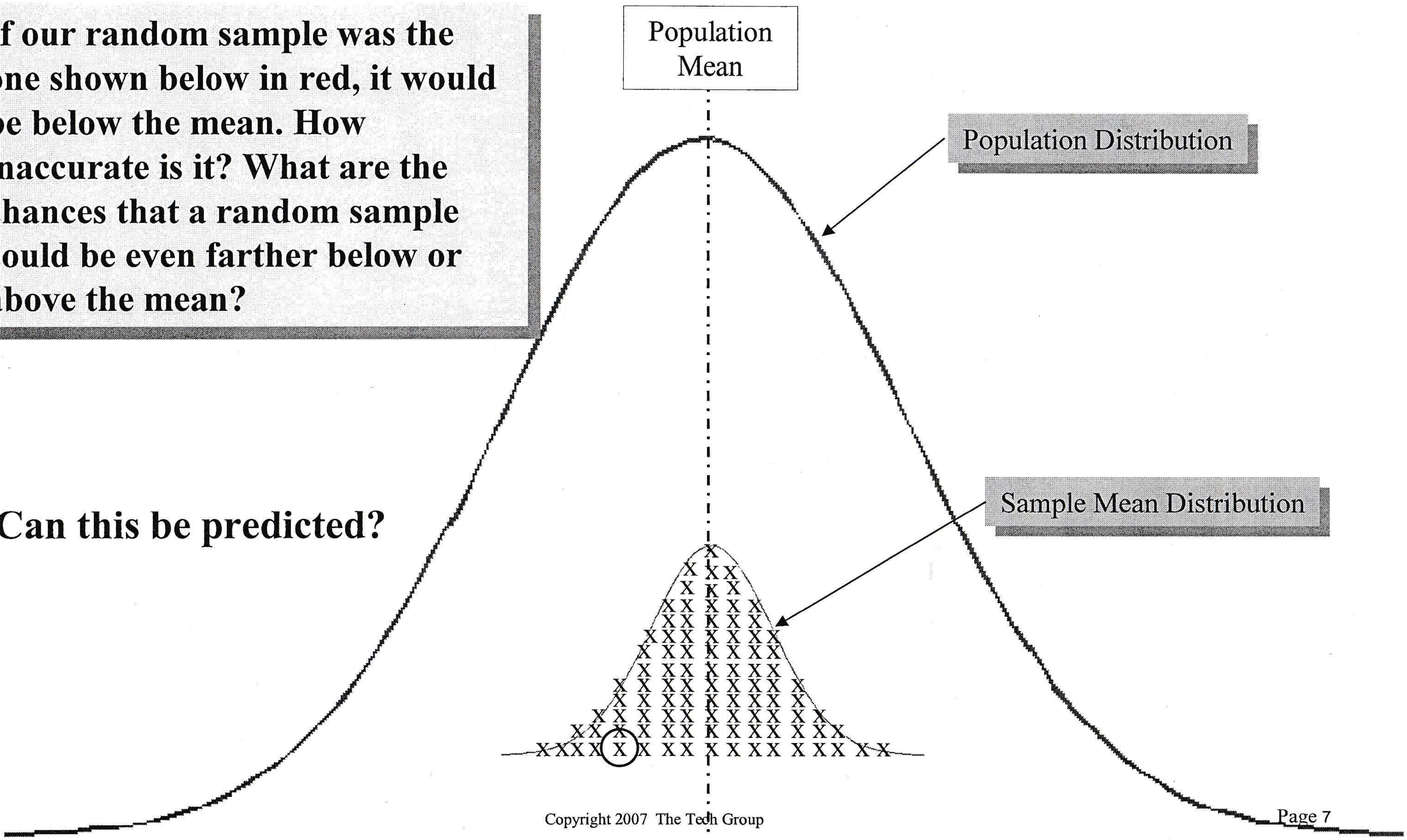


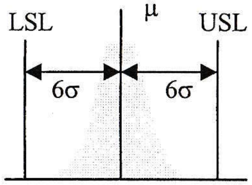


Population Mean vs. Sample Mean

If our random sample was the one shown below in red, it would be below the mean. How inaccurate is it? What are the chances that a random sample could be even farther below or above the mean?

Can this be predicted?

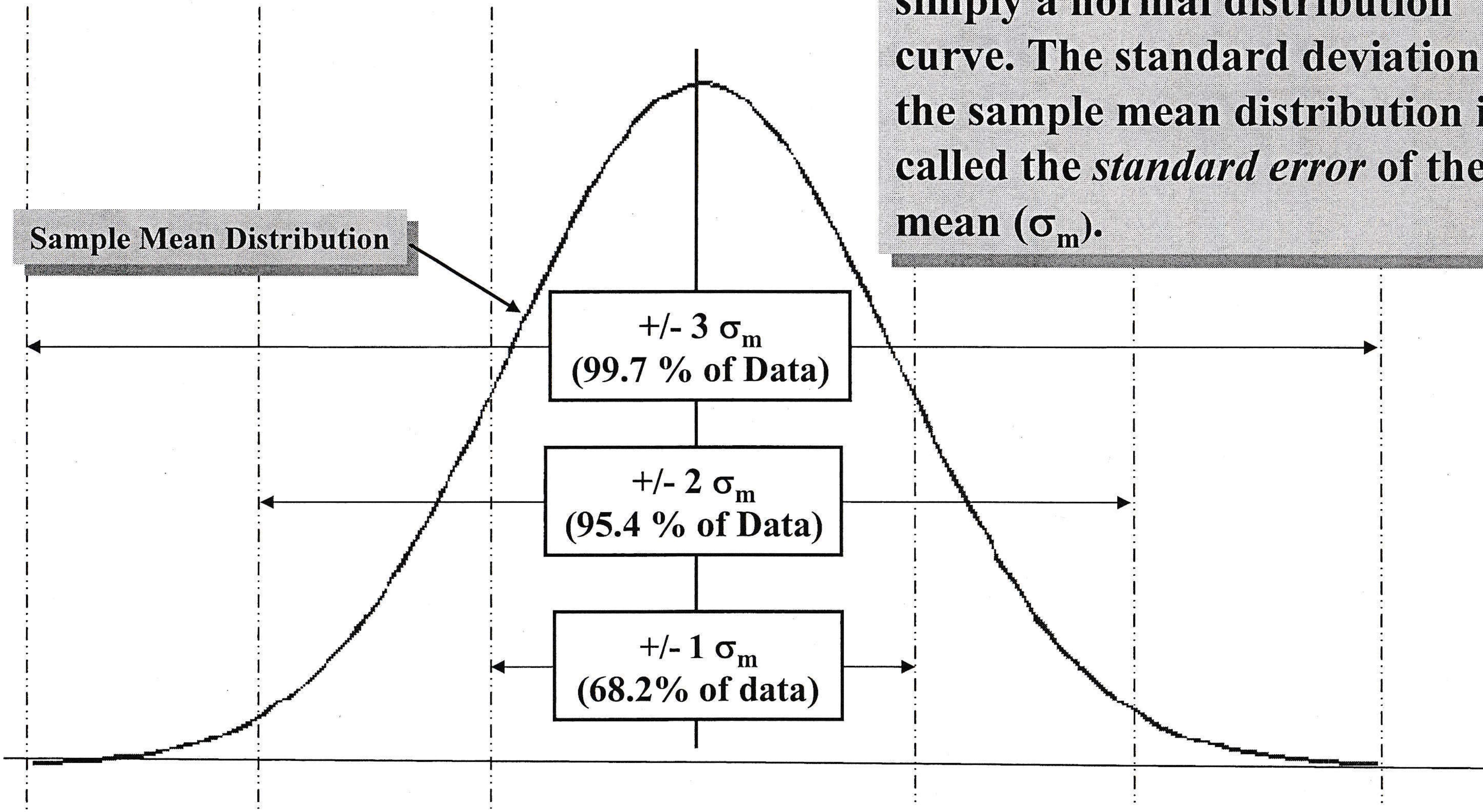




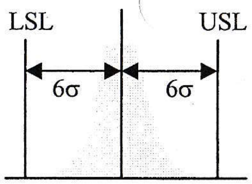
Standard Error

The Tech Group
WE PROVIDE INNOVATIVE SOLUTIONS

The sample mean distribution is simply a normal distribution curve. The standard deviation of the sample mean distribution is called the *standard error* of the mean (σ_m).



How can we determine the Standard Error?



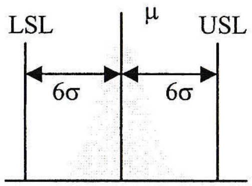
Standard Error

The *standard error* of the mean is defined as:

$$\sigma_m = \frac{\sigma}{\sqrt{N}}$$

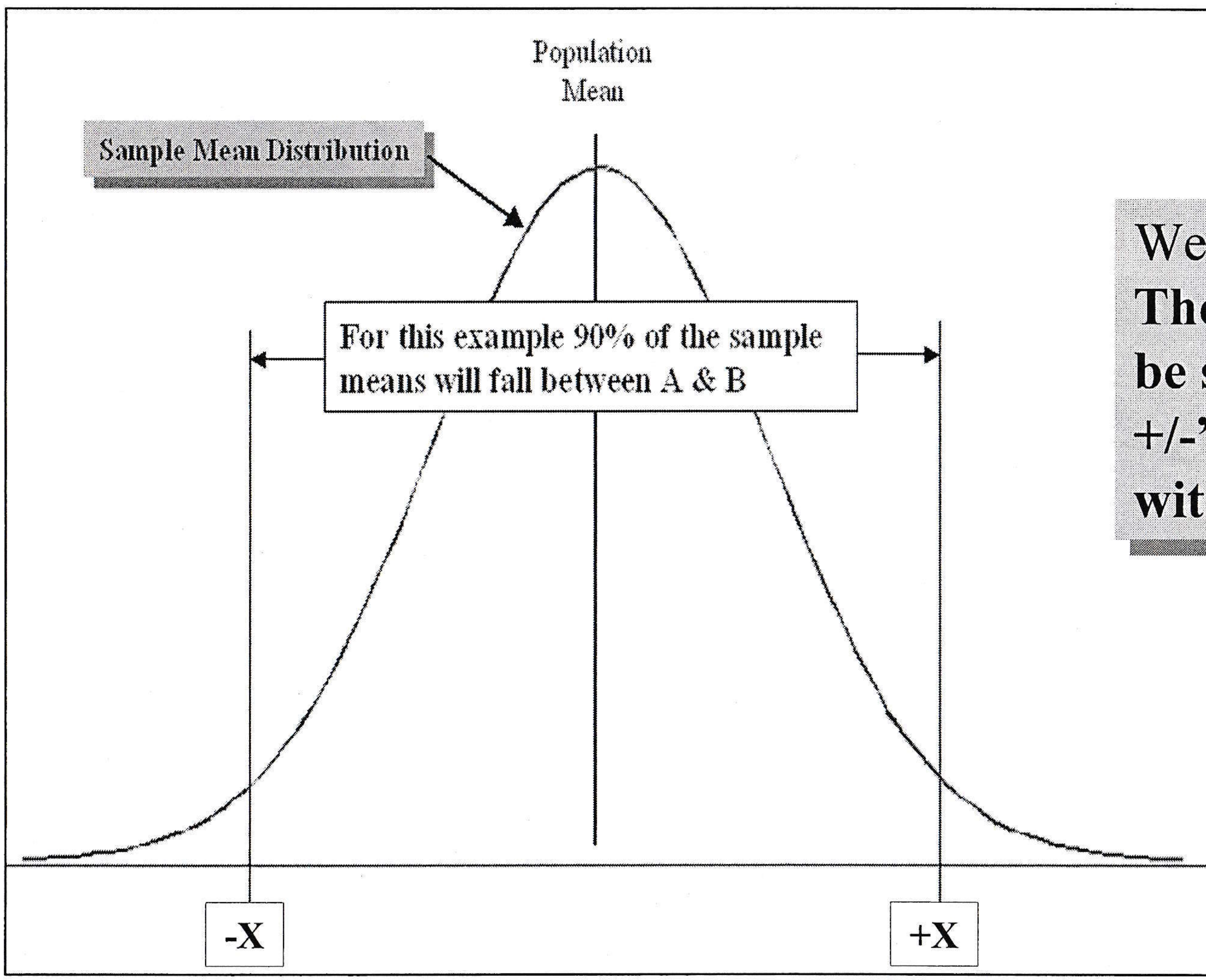
Where N is the sample size and σ is the standard deviation of the population.

What happens to the standard error as the sample size gets larger?

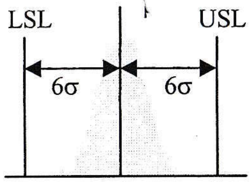


Confidence Interval

Using the concept of standard error, it is now possible to calculate the range in which you can expect the actual mean to fall based on a confidence level.



We can also say it like this:
The population mean will be somewhere between +/-"X" of the sample mean, with a 90% confidence.

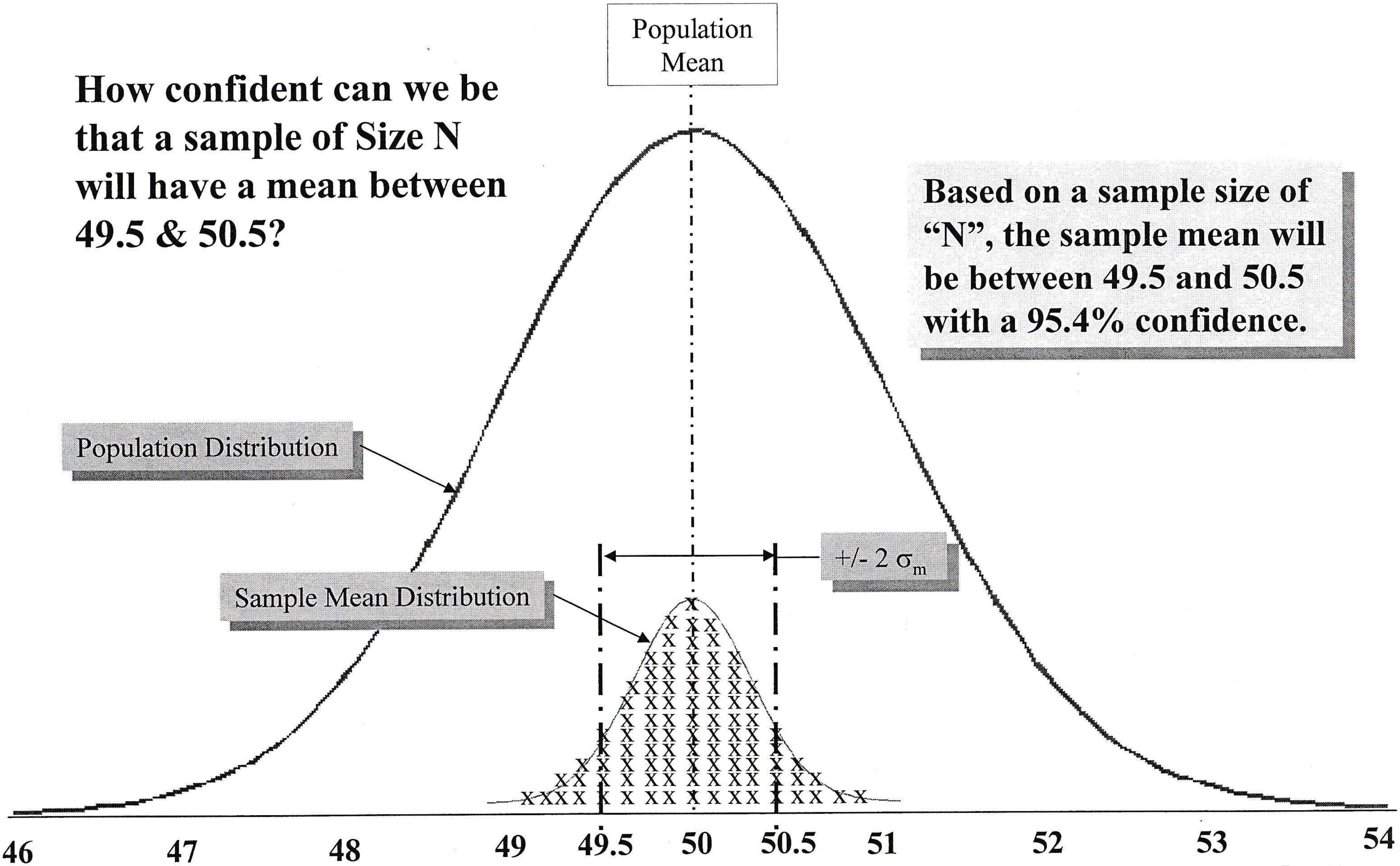


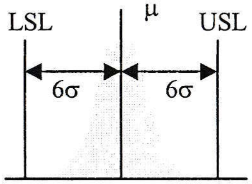
Confidence Interval Concept

The Tech Group
WE PROVIDE INNOVATIVE SOLUTIONS

How confident can we be that a sample of Size N will have a mean between 49.5 & 50.5?

Based on a sample size of "N", the sample mean will be between 49.5 and 50.5 with a 95.4% confidence.





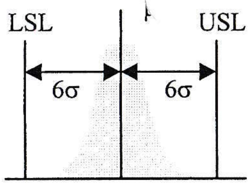
Formulas for Confidence Interval

To calculate the confidence interval (μ) we must first decide on our confidence level, and select a Z value from the Z-table.

Using this value the formula is:

$$\boxed{\mu = M \pm Z\sigma_M}$$

Where **M** is the sample mean, σ_M is the standard error of the mean and **Z** is the value from the Z-table corresponding to the our desired confidence interval (1-α).



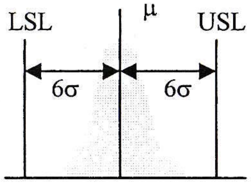
Example #1

Assume the standard deviation of SAT verbal scores in a school system is known to be 100. A researcher wishes to estimate the mean SAT score and compute a 95% confidence interval from a random sample of 10 scores.

The 10 scores are: 320, 380, 400, 420, 500, 520, 600, 660, 720 & 780. Therefore M= 530, N=10 and

$$\boxed{\sigma_M = \frac{\sigma}{\sqrt{N}} = \frac{100}{\sqrt{10}} = 31.62}$$

The value of Z, from the Z-table for the 95% confidence interval is the number of σ_M's one must go from the mean <u>in both directions</u> to contain 95% of the distribution of means.



Example #1 (con't)

The Tech Group
WE PROVIDE INNOVATIVE SOLUTIONS

The value from the Z-table corresponding to .025% of the data at each tail of the normal distribution curve is Z=1.96. We now know "M", "Z" and σ_M.

Z-Table

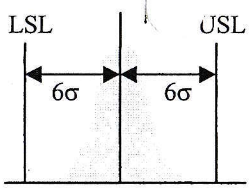
$$\sigma_M = \frac{\sigma}{\sqrt{N}} = \frac{100}{\sqrt{10}} = 31.62$$

So,

$$\mu = M \pm Z\sigma_M$$

$$\mu = M \pm Z\sigma_M = 530 \pm 1.96(31.62) = 530 \pm 61.98$$

This means that researchers can be 95% certain that the mean SAT score in the school system will be 530 +/- 61.98.



Confidence Interval When Sigma is Not Known

It is rare we know the standard deviation of a population, when we want to estimate the mean. So we must estimate both σ and μ. We know from before:

$$\mu = M \pm Z\sigma_M$$

and

$$\sigma_M = \frac{\sigma}{\sqrt{N}}$$

But if σ is not known, we must estimate our standard error based on the σ from our sample and use the t-distribution instead of the Z-distribution.

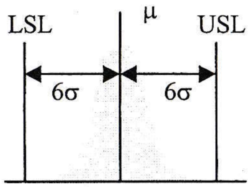
$$\mu = M \pm (t)\, s_M$$

and

$$s_M = \frac{s}{\sqrt{N}}$$

Est. sigma of the population

The value of "t" can be determined from a t-table. The degrees of freedom (DOF) for "t" is equal to the DOF for the estimate of σ_M which is equal to N-1.



Example #2

The Tech Group
WE PROVIDE INNOVATIVE SOLUTIONS

A Green Belt working on a lean project is interested in estimating the mean reading speed (words per minute) of high school graduates & computing the 95% confidence interval. A sample of 6 graduates was taken and the reading speeds were:

200, 240, 300, 410, 450, 600. For this data,

$M = 366.66$ (mean)

$s = 149.35$ (est. σ of the pop.)

$s_M = 60.97$

DOF=6-1=5

t-Table

t=2.571 (from t-table)

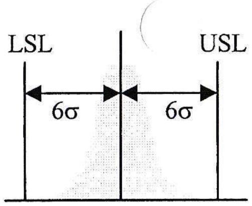
$$\boxed{s_M = \frac{s}{\sqrt{N}}}$$

$$\boxed{\mu = M \pm (t)\, s_M}$$

Therefore, the <u>confidence interval</u> = 366.66 +/- 2.571 (60.97)

= 366.66 +/- 156.75 or

Upper Limit: 523.41
Lower Limit: 209.91



Exercise

An auto manufacturer wants high mileage to be a selling feature of their new model. Through several experiments mileage results were collected and averages are listed below. What average mileage range can the manufacturer advertise with 99% confidence? (Sigma for the data set is 1.69)

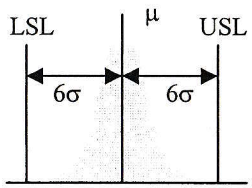
Test #	Mileage
1	61.8
2	58.6
3	62.0
4	60.9
5	60.6
6	61.1
7	59.3
8	60.0
9	61.0
10	63.7
11	57.6
12	62.7
13	63.1
14	59.3
15	60.4
Mean:	**60.8**

$$\sigma_m = \frac{\sigma}{\sqrt{N}} = \frac{1.69}{\sqrt{15}} = .436$$

What is the confidence level if we want to advertise an average mileage of 60 MPG?

$$60.8 \pm 2.575(.436)$$

$$59.68 - 61.92$$



Example #3
Confidence Intervals for a Proportion

A Manufacturing Engineer wishes to estimate the failure rate for a piece of automation equipment considered ready for production. He selects a sampling of 400 assemblies from the equipment and finds 120 of the 400 to be defective.

Compute the 95% confidence level of the proportion that is defective in the population.

Applying the same rules as before, we know the confidence interval is:

Standard error of the proportion

P = proportion in the sample

$$\boxed{= P \pm Z(\sigma_P)}$$

Where

$$\boxed{\sigma_P = \sqrt{\frac{p(1-p)}{N}}}$$

(Condition: np & n(1-p)>=5)

Example #3 (con't)

- The value of **p** is 120/400 = .30
- The estimated value of σ_p is:

$$\sigma_P = \sqrt{\frac{p(1-p)}{N}} =$$

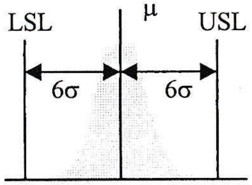
$$\sigma_P = \sqrt{\frac{.3(1-.3)}{400}} = .023$$

- From before, we know Z for a 95% confidence interval is 1.96, therefore our confidence interval is:

$$= P \pm Z(\sigma_P)$$

$$= .3 \pm 1.96\,(.023) = .3 \pm .045$$

With 95% confidence,we can expect the defect rate to be 30% +/- 4.5%.

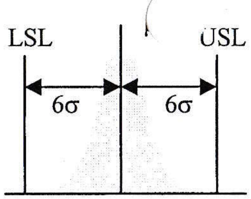


The Tech Group
WE PROVIDE INNOVATIVE SOLUTIONS

Other Applications of Confidence Intervals

Confidence intervals can be used where ever an estimate must be made using a sample from a normally distributed (or predictable) data set.

- Between two means (differences & sums)
- Between two proportions (differences & sums)
- For Variance
- For Standard Deviations
- For Cp/Cpk values

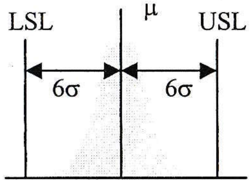


The Tech Group
WE PROVIDE INNOVATIVE SOLUTIONS

Relationships of Variables

Confidence interval width will increase or decrease if:

- Population sigma increases?
- Desired confidence level is increased?
- Number of samples (N) decreased?

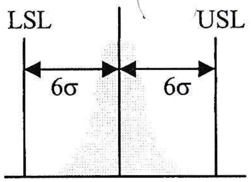


Summary

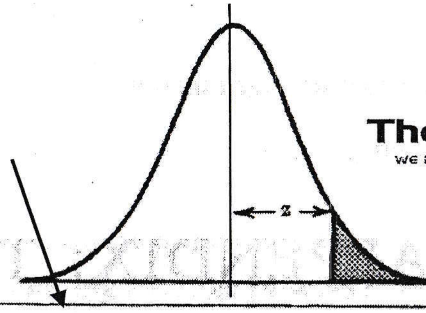
- Confidence intervals provide a powerful way to determine the precision of an estimate for a population or populations based on a sample.

- Without it you may draw incorrect conclusions.

- If the precision level is unacceptable, higher precision (smaller interval) can be attained by either increasing the sample size, or reducing the confidence level.



Z-Table

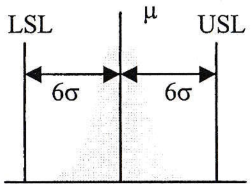


The Tech Group
WE PROVIDE INNOVATIVE SOLUTIONS

For a 95% confidence interval, $\alpha = .05$, Therefore $\alpha/2 = .025$. So we locate .025 from the table and select Z.

TABLE A. ***Tail area of unit normal distribution***

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010



The Tech Group

WE PROVIDE INNOVATIVE SOLUTIONS

t-Table

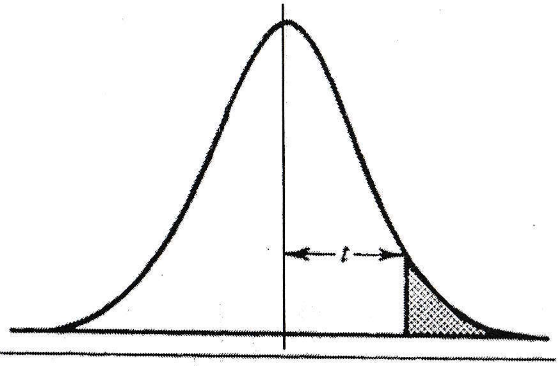


TABLE B1. Probability points of the t distribution with ν degrees of freedom

	tail area probability									
ν	**0.4**	**0.25**	**0.1**	**0.05**	**0.025**	**0.01**	**0.005**	**0.0025**	**0.001**	**0.0005**
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850

For a 95% confidence interval, α= .05, Therefore α/2= .025. So we locate .025 on the x-axis and the DOF on the y-axis & select "t" where they intersect.