Confidence Interval Estimation

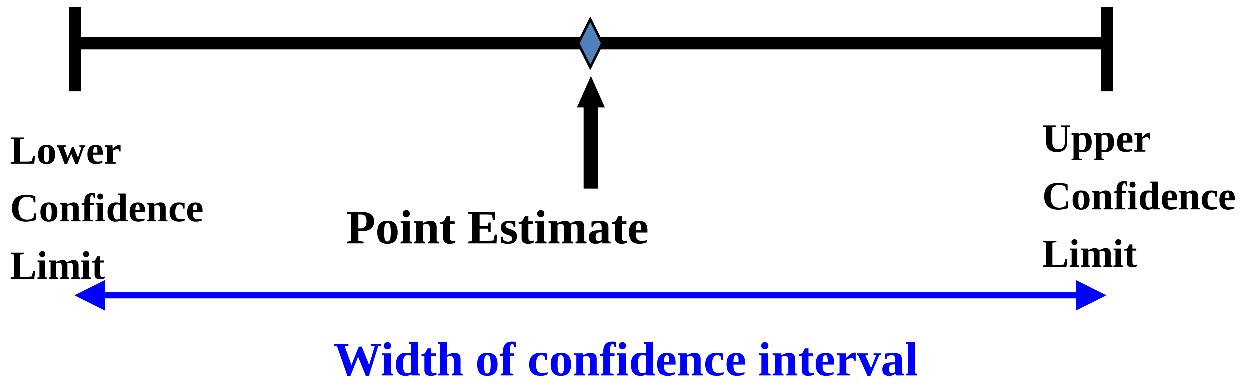
A manager of human resources in a company might want to estimate the **average number of days of work an employee misses per year because of illness.**

The cellular company wants to ascertain the average number of minutes of time used per month customer.

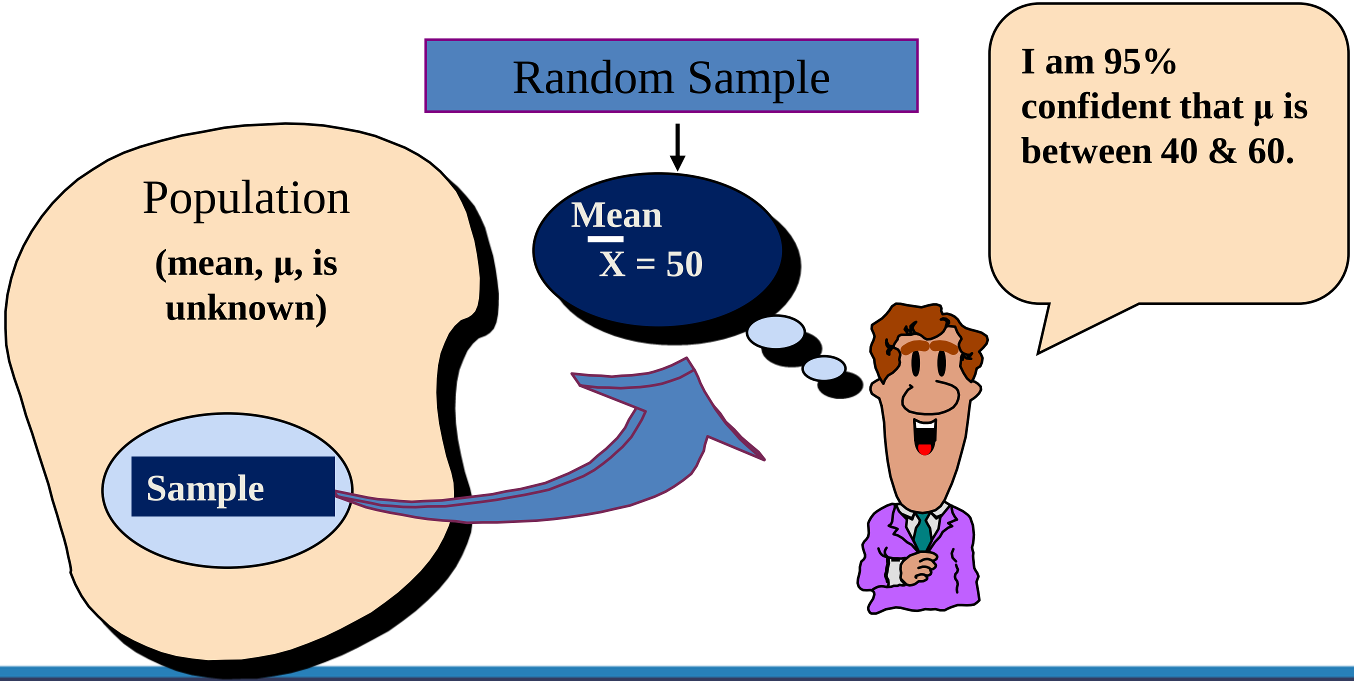
A **point estimate** is *a statistic taken from a sample that is used to estimate a population parameter.*

Point and Interval Estimates

- A point estimate is a single number,
- An **interval estimate** (confidence interval) is *a range of values within which the analyst can declare, with some confidence, the population parameter lies* .
- A confidence interval provides additional information about the variability of the estimate



Estimation Process



The general formula for all confidence intervals is:

Point Estimate ± (Critical Value)(Standard Error)

Where:

- Point Estimate is the sample statistic estimating the population parameter of interest
- Critical Value is a table value based on the sampling distribution of the point estimate and the desired **confidence level**
- Standard Error is the standard deviation of the point estimate

Point Estimates

We can estimate a Population Parameter …		with a Sample Statistic (a Point Estimate)
Mean	μ	$\bar{X}$
Proportion	π	p

The Greensboro coliseum is considering expanding its seating capacity and needs to know both the average number of people who attend events there and the variability in this number. The following are the attendances (in thousands) at nine randomly selected sporting events. Find point estimates of the mean and the variance of the population from which the sample was drawn.

8.8	14.0	21.3	7.9	12.5	20.6	16.3	14.1	13.0

Solution:

$$\sum x^2 = 2003.65 \qquad \sum x = 128.5 \qquad n = 9$$

$$\bar{x} = \frac{\bar{x}}{n} = \frac{128.5}{9} = 14.2778 \text{ thousands of people}$$

$$s^2 = \frac{1}{n-1}\left(\sum x^2 - nx^2\right) = \frac{2003.65-9(14.2778)^2}{8}$$

$$= 21.119 \text{ (1,000s of people)}^2$$

For a population with a known variance of 185, a sample of 64 individuals leads to 217 as an estimate of the mean.

a) Find the standard error of the mean.
b) Establish an interval estimate that should include the population mean 68.3 percent of the time.

Solution:

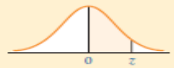
$\sigma^2 = 185$ $\sigma = \sqrt{185} = 13.60$ $n = 64$ $\bar{x} = 217$

$$a)\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{13.60}{\sqrt{64}} = 1.70$$

$$b)\ \bar{x} \pm \sigma_{\bar{x}} = 217 \pm 1.70 = (215.3, 218.7)$$

Given the following confidence levels, express the lower and upper limits of the confidence interval for these levels in terms of $\bar{x}$ and $\sigma_{\bar{x}}$.

a) 54 percent
b) 75 percent
c) 94 percent
d) 98 percent



SECOND DECIMAL PLACE IN z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998									
4.0	.49997									
4.5	.499997									
5.0	.4999997									
6.0	.499999999									

Solution:

$a)\,\bar{x} \pm 0.74\sigma_{\bar{x}}$

$b)\,\bar{x} \pm 1.15\sigma_{\bar{x}}$

$c)\,\bar{x} \pm 1.88\sigma_{\bar{x}}$

$d)\,\bar{x} \pm 2.33\sigma_{\bar{x}}$

Confidence Intervals

- How much uncertainty is associated with a point estimate of a population parameter?

- An interval estimate provides <u>more information about a population characteristic</u> than does a point estimate

- Such interval estimates are called confidence intervals

Confidence Interval Estimate

- An interval gives a range of values:
  - Takes into consideration **variation** in sample statistics from sample to sample
  - Based on **observations** from 1 sample
  - Gives information about closeness to unknown population parameters
  - Stated in terms of level of confidence
    - e.g. 95% confident, 99% confident

Confidence Interval Example

<u>Cereal fill example</u>

- Population has $\mu = 368$ and $\sigma = 15$.
- If you take a sample of size n = 25 you know
  - $368 \pm 1.96 * 15 / \sqrt{25} = (362.12, 373.88)$ contains 95% of the sample means
  - When you don't know μ, you use $\overline{X}$ to estimate μ
    - If $\overline{X} = 362.3$ the interval is $362.3 \pm 1.96 * 15 / \sqrt{25} = (356.42, 368.18)$
    - Since $356.42 \leq \mu \leq 368.18$ the interval based on this sample makes a correct statement about μ.

But what about the intervals from other possible samples of size 25?

SECOND DECIMAL PLACE IN z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998									
4.0	.49997									
4.5	.499997									
5.0	.4999997									
6.0	.499999999									

Confidence Interval Example *(continued)*

Sample #	$\overline{X}$	Lower Limit	Upper Limit	Contain μ?
1	362.30	356.42	368.18	Yes
2	369.50	363.62	375.38	Yes
3	360.00	354.12	365.88	No
4	362.12	356.24	368.00	Yes
5	373.88	368.00	379.76	Yes

Confidence Interval Example

(continued)

- In practice you only take one sample of size n
- <u>In practice you do not know μ so you do not know if the interval actually contains μ</u>
- However you do know that 95% of the intervals formed in this manner will contain μ
- Thus, based on the one sample, you actually selected you can be <u>95% confident</u> your interval will contain μ (this is a 95% confidence interval)

Note: 95% confidence is based on the fact that we used $Z = 1.96$.

Confidence Level

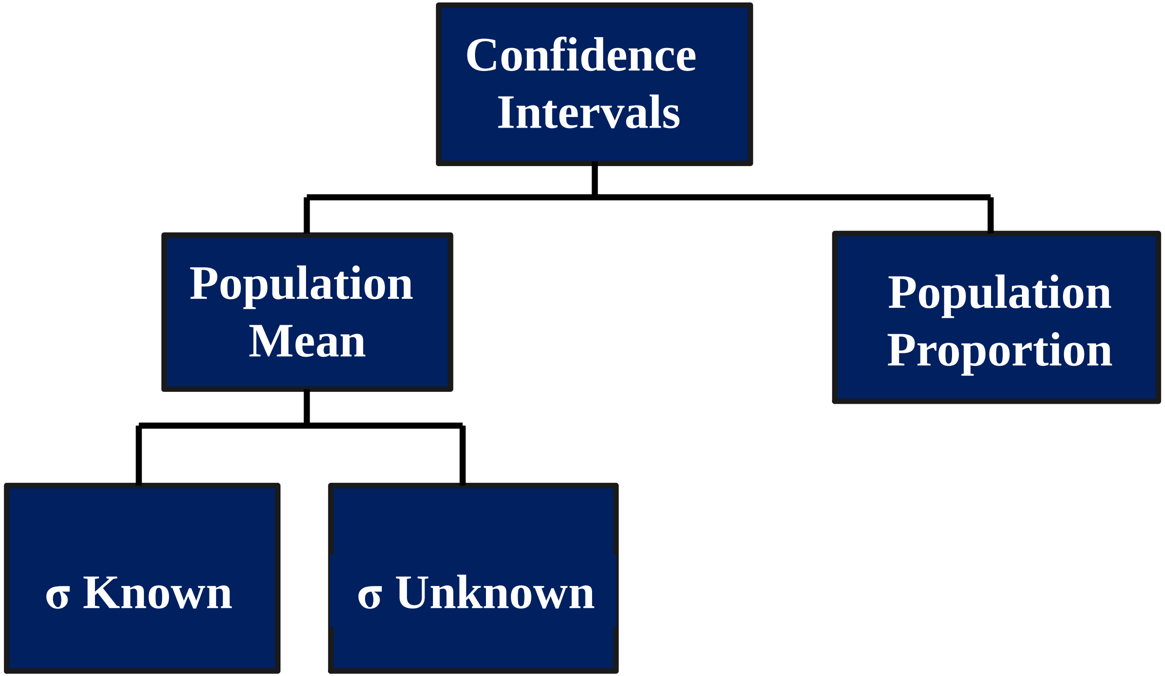
- Confidence Level
  - Confidence the **interval will contain** the unknown population parameter
  - A percentage (less than 100%)

Confidence Level, $(1-\alpha)$

(continued)

- Suppose confidence level = 95%
- Also written $(1 - \alpha) = 0.95$, (so $\alpha = 0.05$)
- A relative frequency interpretation:
  - 95% of all the **confidence intervals** that can be constructed will contain the **unknown** true parameter
- A specific interval either will contain or will not contain the true parameter
  - No probability involved in a specific interval

Confidence Intervals



Confidence Interval for μ (σ Known)

- Assumptions
  - Population standard **deviation σ is known**
  - Population is **normally distributed**
  - If population is **not normal**, use **large** sample
- Confidence interval estimate:

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

where $\bar{X}$ is the point estimate

$Z_{\alpha/2}$ is the normal distribution critical value for a probability of $\alpha/2$ in each tail

$\sigma/\sqrt{n}$ is the standard error

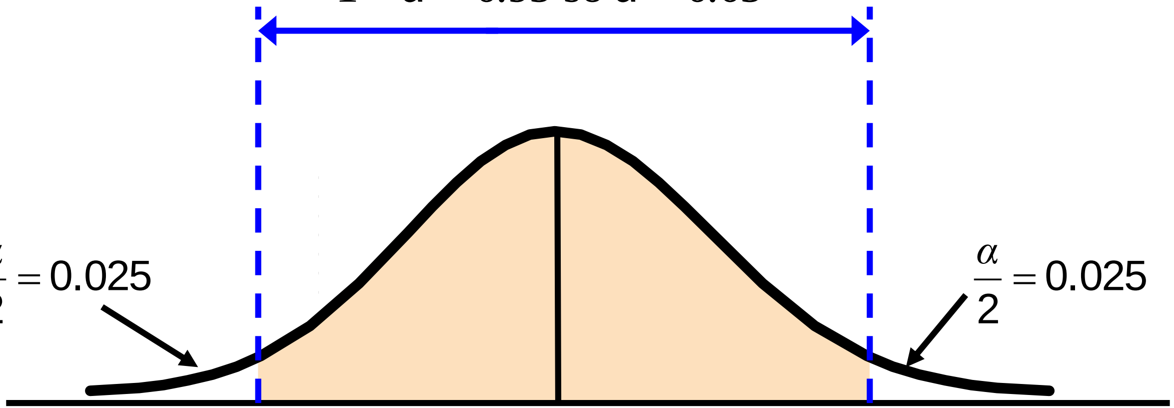
Finding the Critical Value, $Z_{\alpha/2}$

$$Z_{\alpha/2} = \pm 1.96$$

- Consider a 95% confidence interval:

$$1 - \alpha = 0.95 \text{ so } \alpha = 0.05$$

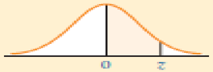
$\frac{\alpha}{2} = 0.025$



$\frac{\alpha}{2} = 0.025$

Z units: $Z_{\alpha/2}$ **= -1.96** | **0** | $Z_{\alpha/2}$ **= 1.96**

X units: **Lower Confidence Limit** | **Point Estimate** | **Upper Confidence Limit**



SECOND DECIMAL PLACE IN z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998									
4.0	.49997									
4.5	.499997									
5.0	.4999997									
6.0	.499999999									

Confidence Level	***Confidence Coefficient,*** $1-\alpha$	$Z_{\alpha/2}$ ***value***
80%	**0.80**	**1.28**
90%	**0.90**	**1.645**
95%	**0.95**	**1.96**
98%	**0.98**	**2.33**
99%	**0.99**	**2.58**
99.8%	**0.998**	**3.08**
99.9%	**0.999**	**3.27**

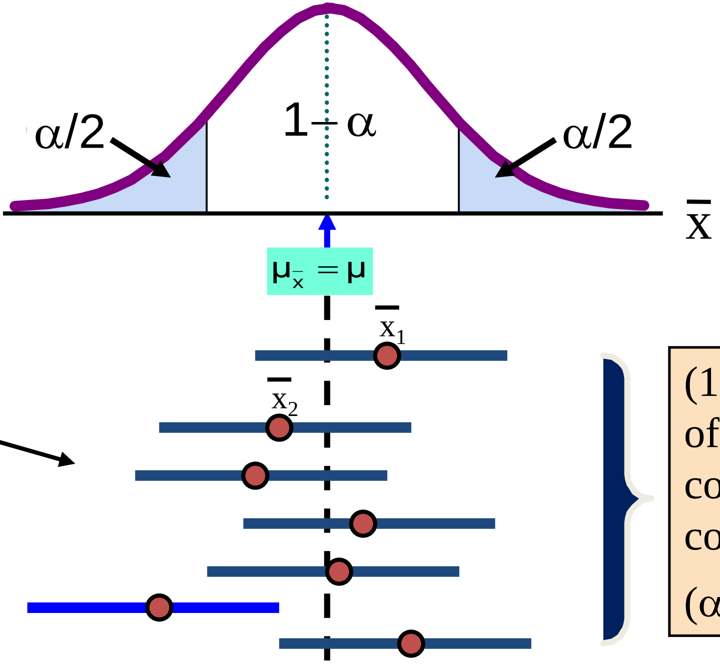
Common Levels of Confidence

- Commonly used confidence levels are 90%, 95%, and 99%

Confidence Level	*Confidence Coefficient,* $1-\alpha$	$Z_{\alpha/2}$ *value*
80%	0.80	1.28
90%	0.90	1.645
95%	0.95	1.96
98%	0.98	2.33
99%	0.99	2.58
99.8%	0.998	3.08
99.9%	0.999	3.27

Intervals and Level of Confidence

Sampling Distribution of the Mean



Confidence Intervals

Intervals extend from

$$\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

to

$$\bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$(1-\alpha)$x100% of intervals constructed contain μ;

(α)x100% do not.