EIN 5226

Confidence Intervals –Continuous Data

Chapter 16 Sections 16.1-2

Chapter 17 Sections 17.4-6

Need Z table and t table for this lecture

Karen E. Schmahl Ph.D., P.E.

FLORIDA INTERNATIONAL UNIVERSITY
COLLEGE OF ENGINEERING

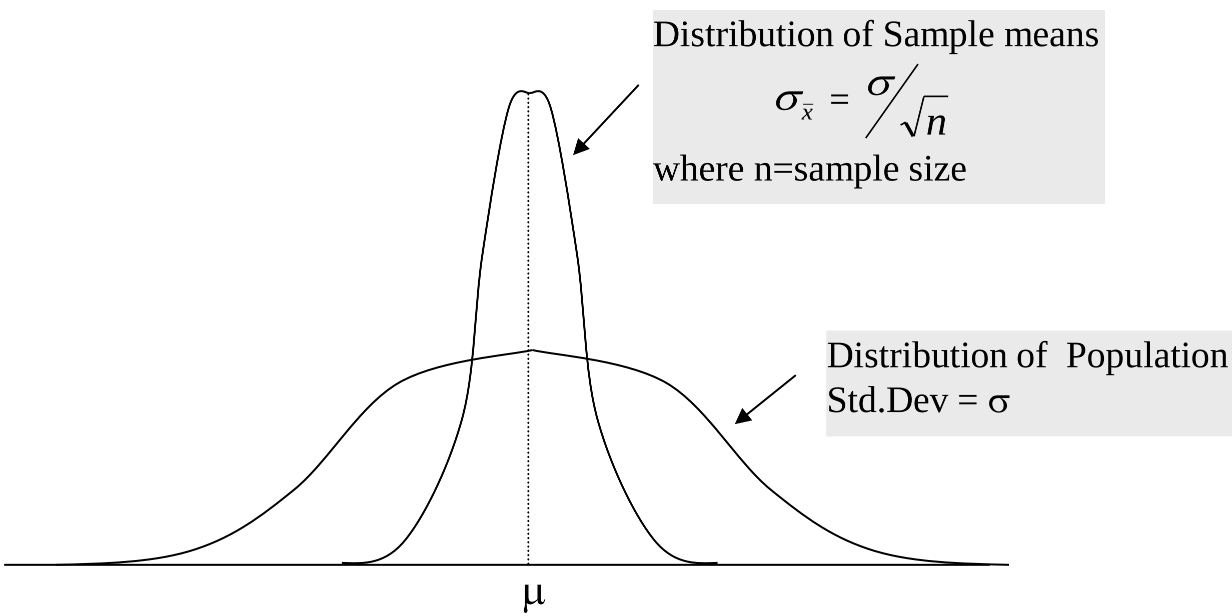
The Central Limit Theorem

If a random sample of size n is drawn from a population with mean μ and variance, σ^2,

then the sample mean, $\bar{x}$, has approximately a normal distribution with mean μ and variance σ^2 / n .

That is, the distribution function of $\dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ is approximately a standard normal.

The approximation improves as the sample size increases.

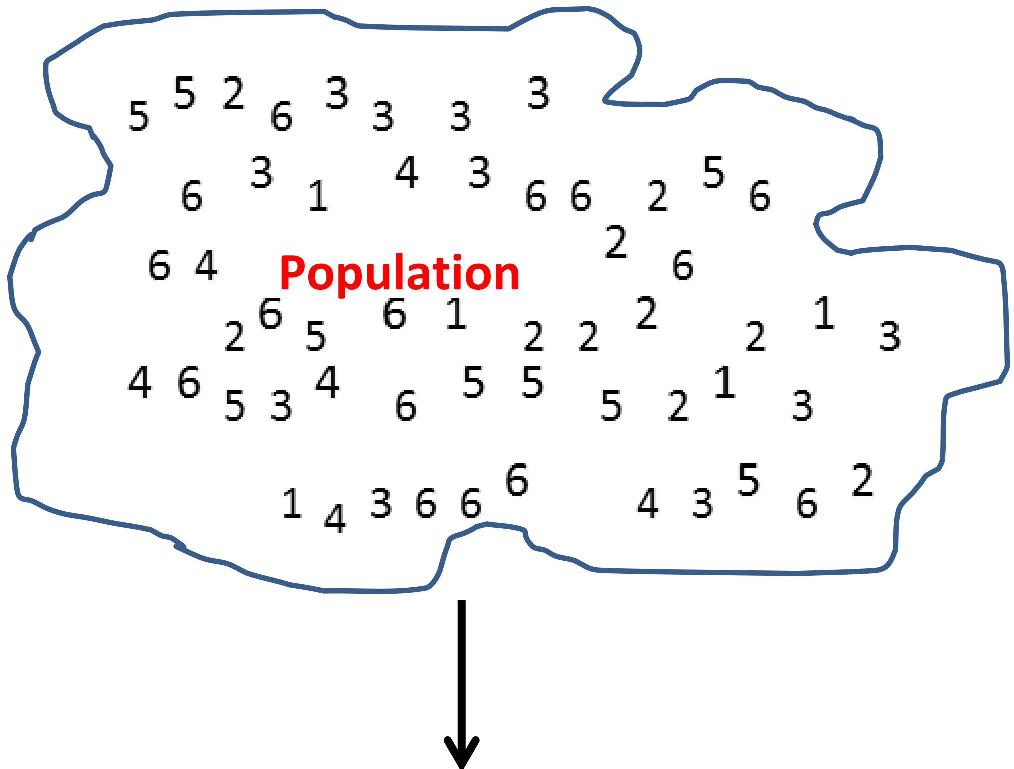


$\sigma_{\bar{x}} = \sigma / \sqrt{n}$

1. What is theoretical standard deviation, $\sigma_{\bar{x}}$, of the distribution of sample averages of size n= 8 where for the population μ= 45 and σ= 3.3?

A) 1.35 B) 0.418 C) 0.875 D) 1.17

What can a sample tell us about a population?



Sample of size n=5

3 2 4 6 5

$\bar{x} = 4.0$, s=1.58

Sample statistics vs. population parameters

Population – entire group of interest in an analysis

Parameter – descriptive number calculated from entire populations values

Sample – subset of items selected from population

Statistic – any descriptive value calculated from the sample group's observations

Sample statistics are often used to estimate parameters

Point Estimates

- From a sample, we calculate statistics and use them as **point estimates** of population parameters
  - sample mean (Normal Distribution)
  - sample proportion (Binomial Distribution)

- More useful are interval estimates, which are called **confidence intervals**.

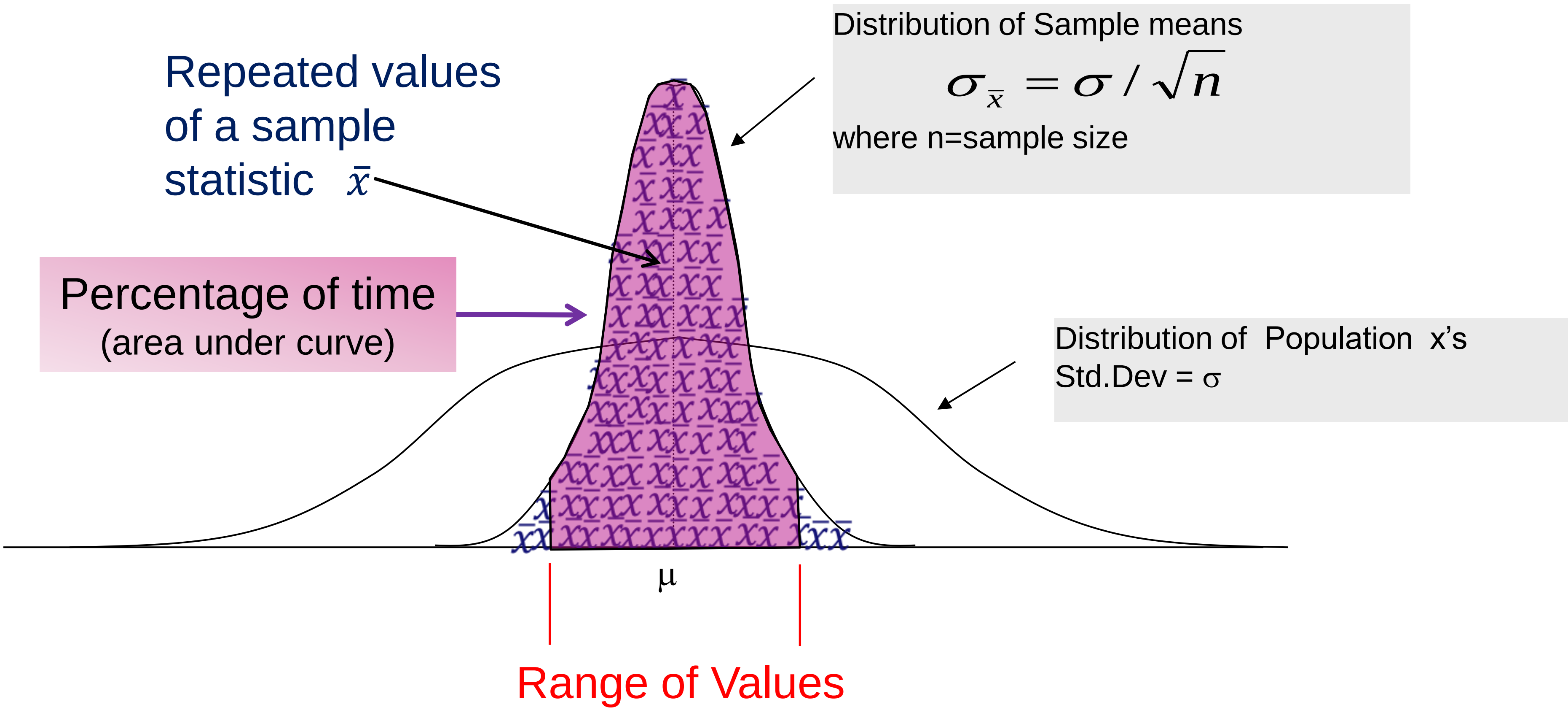
Confidence Intervals for Means

.The confidence interval

- Provides margin of error for the sample statistic to indicate how far off the true value the point estimate could be.

- Provides a range of values in which repeated values of a sample statistic are expected to fall a certain percentage of time

The Central Limit Theorem



Confidence intervals

A confidence interval is a range of values within which repeated sample statistics are predicted to fall a certain percentage of the time.

α = risk that true population parameter will fail to fall in defined interval

Confidence level = 100 (1-α)%

Example: Risk $\boldsymbol{\alpha} = .05$ or 5%

Confidence Level$= .95$ or 95%

Taking a sample of size n=36, we get a mean of 19.60 and the standard deviation of the population is know to be 2. What is the 90% confidence interval for the true mean of the population?

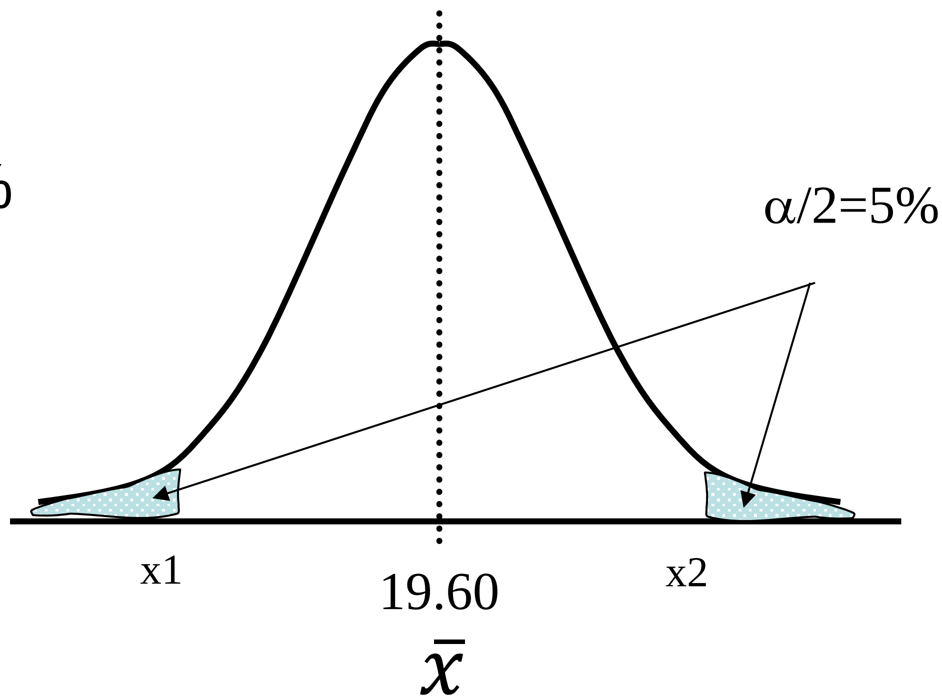
n=36

$\bar{x}$ =19.60

σ= 2

α = 10%

$$\sigma_{\bar{x}} = \sigma / \sqrt{n} = 2 / \sqrt{36} = .333$$



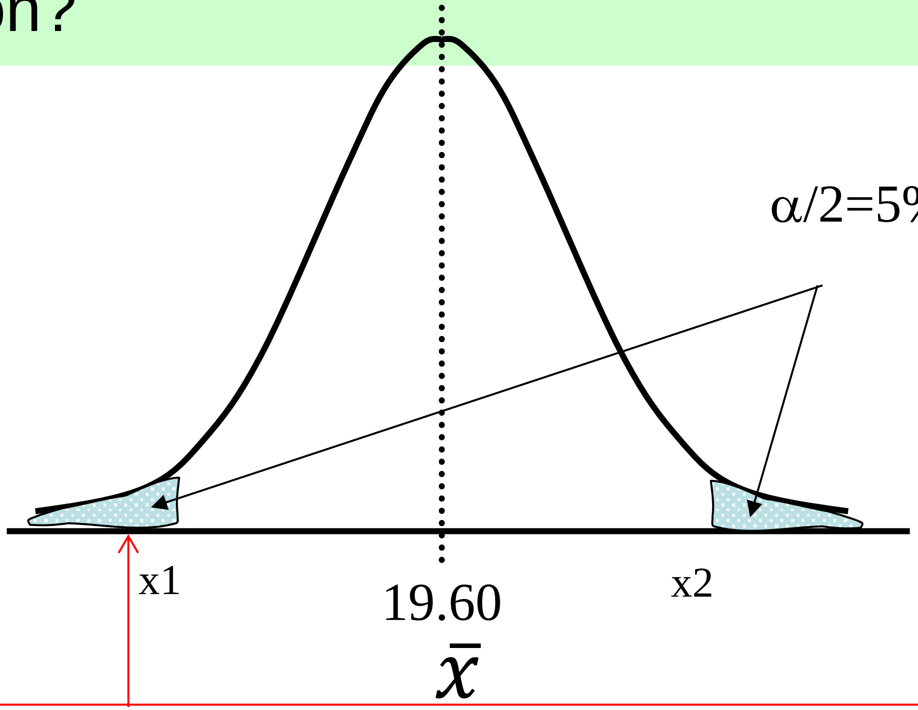
Taking a sample of size n=36, we get a mean of 19.60 and the standard deviation of the population is know to be 2. What is the 90% confidence interval for the true mean of the population?

n=36

$\bar{x}$ =19.60

σ= 2 $\sigma_{\bar{x}} = .333$

α = 10%



2. What value of Z will correspond to the probability in the lower tail?

A) -3.30 B) -2.575 C) -1.645 D) +2.578

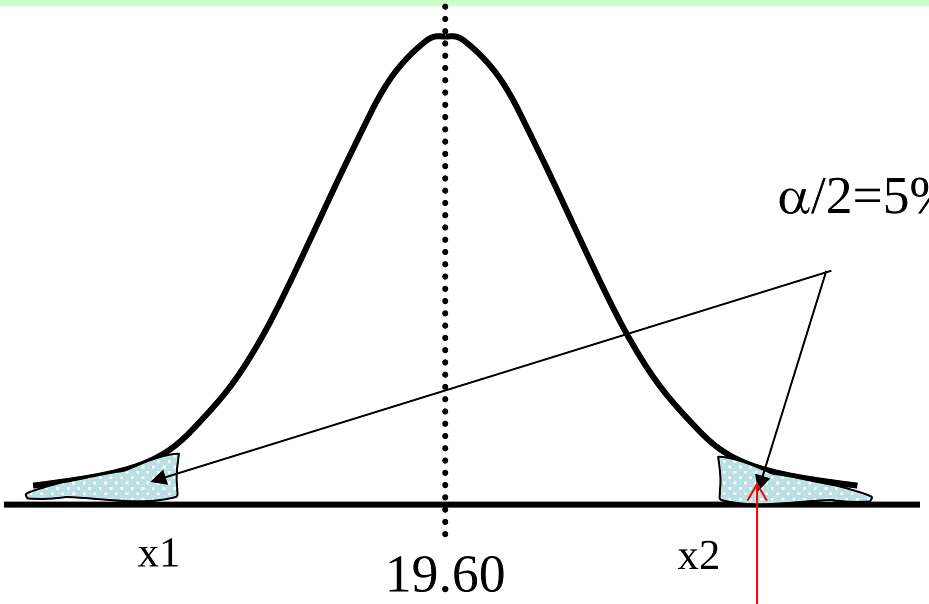
Taking a sample of size n=36, we get a mean of 19.60 and the standard deviation of the population is know to be 2. What is the 90% confidence interval for the true mean of the population?

n=36

$\bar{x}$ =19.60

σ= 2 $\sigma_{\bar{x}} = .333$

α = 10%



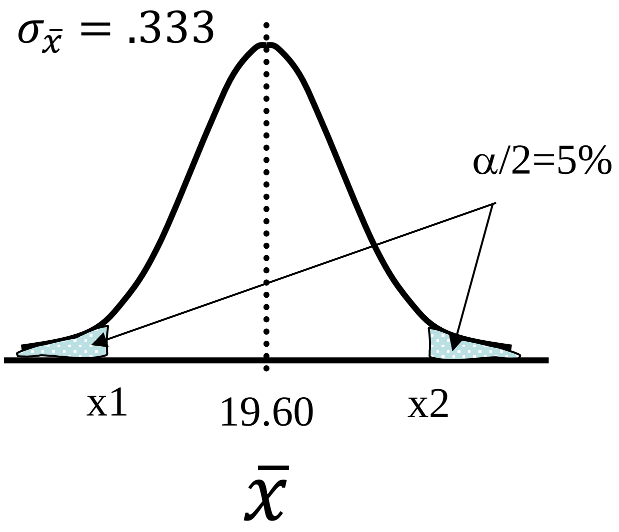
3. What value of Z will correspond to the probability in the upper tail?

A) +3.30 B) +2.575 C) +1.645 D) -2.578

Taking a sample of size n=36, we get a mean of 19.60 and the standard deviation of the population is know to be 2. What is the 90% confidence interval for the true mean of the population?

$$Z = \frac{x - \hat{u}}{\sigma/\sqrt{n}} = \frac{x - \bar{x}}{\sigma/\sqrt{n}} = \pm 1.645$$

$$x = \bar{x} + Z\frac{\sigma}{\sqrt{n}}$$



Lower limit = x_1 =

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 19.60 - 1.645\frac{2}{\sqrt{36}} = 19.05$$

Upper limit = x_2 =

$$\bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 19.60 + 1.645\frac{2}{\sqrt{36}} = 20.15$$

Therefore with 90% confidence we can say the true mean of the population is between 19.05 and 20.15

Large Sample (n≥30)
Confidence Interval for Means

<u>Needed</u>

Sample mean, $\bar{x}$

Standard deviation of distribution of sample mean, $\sigma_{\bar{x}}$

Use s=$\hat{\sigma}$ to estimate $\sigma_{\bar{x}}$ with $s/\sqrt{n}$

Confidence level (1-α) or level of risk (α)

Large Sample (n≥30) Confidence Interval for Means

Needed

Sample mean, $\bar{x}$

Standard deviation of distribution of sample mean, $\sigma_{\bar{x}}$

Use s=$\hat{\sigma}$ to estimate $\sigma_{\bar{x}}$ with ${}^{s}/_{\sqrt{n}}$

Confidence level (1-α) or level of risk (α)

Confidence interval formulas:

Lower Limit

$$\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}}$$

Upper limit

$$\bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}}$$

Practice
Large Sample, Interval for Means

$$\sigma_{\bar{x}} = {}^{s}/_{\sqrt{n}}$$

A dishwasher manufacturer studied the noise level on its new design by studying 55 units. The average decibel level was found to be 65 with a standard deviation of 12. What is the 95% confidence level for the new model's average noise level?

4. The risk level, $\boldsymbol{\alpha}$, for this problem is

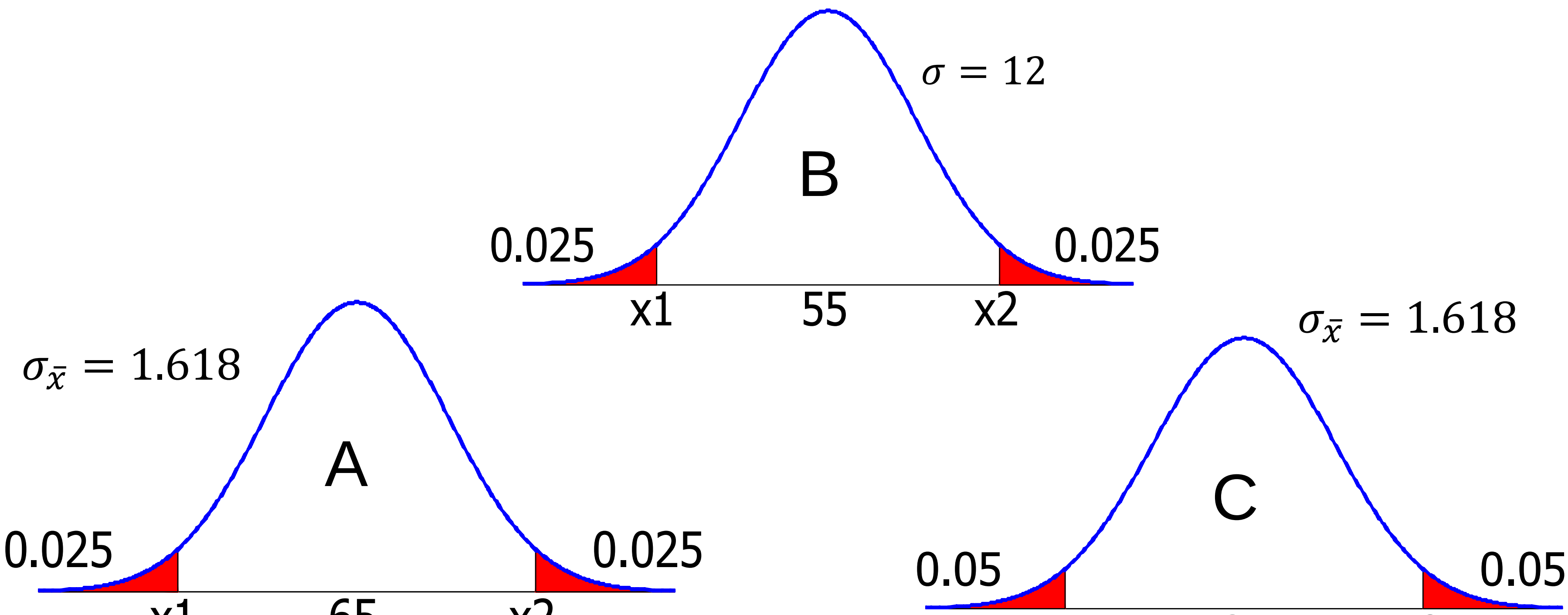
A) 10% B) 5% C) 2.5% D) 12%

5. The standard error of the mean, $\sigma_{\bar{x}}$ applicable in this problem is

A) 12 B) 7.218 C) 1.618 D) 1.839

A dishwasher manufacturer studied the noise level on its new design with 55 units. The average decibel level was found to be 65 with a standard deviation of 12. What is the 95% confidence level for the new model's average noise level?

6. Which is the correct depiction of the problem being analyzed?



Confidence interval formulas: Lower Limit Upper limit

$$\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}} \qquad \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}}$$

A dishwasher manufacturer studied the noise level on its new design by studying 55 units. The average decibel level was found to be 65 with a standard deviation of 12. What is the 95% confidence level for the new model's average noise level?

7. The Z value for this problem is

A) ± 1.645 B) ± 1.96 C) ± 2.81 D) ± 1.85

8. The lower and upper limits are

A) 61.83, 72.01 B) 60.83, 69.17

C) 52.08, 57.92 D) 61.83, 68.17

Question?

Does this 95% confidence interval actually cover the true mean of the population, μ?

- The sample observations come from the middle 95% of the population distribution and then the true mean would be in the interval.
- If the sample mean was unusually large or small, the observations in it may have come from the outer 5% of the population. In this case, the true mean will not be in the interval.
- In the long run, if we repeated these confidence intervals over and over, then 95% of the samples will have means in the middle 95% of the population. Then 95% of the confidence intervals will cover the population mean.

A dishwasher manufacturer studied the noise level on its new design by studying 55 units. The average decibel level was found to be 65 with a standard deviation of 12. Based on the sample, the 95% confidence level for the new model's average noise level was determined to be [61.8 and 68.2].

9. T / F The mean of the population is 65.

10. T / F The probability that the true mean is in the defined interval is 95%.

11. T / F If I increase the sample size and calculate a new 95% confidence interval, the probability of true being in the interval will improve.

Confidence Interval Simulation

Permissions
Permission is granted to link to any portion of the Rice Virtual Lab. The applets in the simulations/demonstrations are hereby in the public domain and can therefore be used without restriction.

Confidence interval Simulation
stat_sim\RVLS_simulations\stat_sim\conf_interval\Contents.html

Estimating Sample Sizes

Larger Sample Size

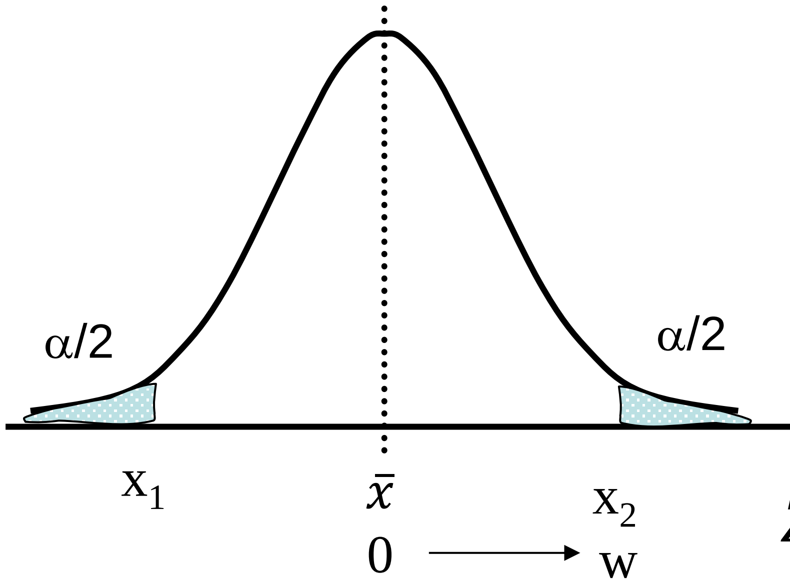
= Smaller Confidence Interval

= Greater Accuracy & Precision

Estimate sample size for mean:

$$n = \left(\frac{Z_{\alpha/2} \sigma}{w} \right)^2$$

where w = desired width (±w)



$$Z_{\alpha/2} = \frac{x - u}{\sigma/\sqrt{n}} = \frac{w - 0}{\sigma/\sqrt{n}}$$

Formula for estimating sample size: $$n = \left(\frac{Z_{\alpha/2}\,\sigma}{w}\right)^2$$

To use need: Level of risk willing to take (to determine Z)

Population standard deviation (or estimate)

w: desired ± interval

Sample Size - Means

A dishwasher manufacturer studied the noise level on its new design with 55 units. The average decibel level was found to be 65 with a standard deviation of 12.

A minor design change was made and an additional study is required. How many units must be sampled so that a 99% confidence interval specifies the mean to within 10 decibels

$Z_{(.01/2)} = Z_{(.005)} = 2.575$

$\hat{\sigma} = s = 12$

$w = \pm 10$

$$n = \left(\frac{Z_{\alpha/2}\sigma}{w}\right)^2$$

$$n = \left(\frac{(2.575)(12)}{10}\right)^2 = 9.55 >> 10$$

Sample Size - Means

$$n = \left(\frac{Z_{\alpha/2}\sigma}{w}\right)^2$$

A dishwasher manufacturer studied the noise level on its new design with 55 units. The average decibel level was found to be 65 with a standard deviation of 12.

12. The company president thinks the interval previously used is too large. How many units must be sampled so that a 95% confidence interval specifies the mean to within 5 decibels.

A) 39 B) 5 C) 22 D) 23

One-Sided Confidence Intervals

- Often situations require only an upper or a lower bound. A one sided confidence interval is used.
- With the same conditions as with the two-sided CI, the level 100(1-α)%
  - A lower confidence bound for μ is

$$\overline{X} - z_{\alpha}\sigma_{\overline{X}}.$$

  - An upper confidence bound for μ is

$$\overline{X} + z_{\alpha}\sigma_{\overline{X}}.$$

- All the risk, $\boldsymbol{\alpha}$, is included in the one tail.

One-sided Confidence Interval

40 adults took part in a simulated burning house exercise to gather data on the average time to escape after the smoke alarm sounded. The sample mean escape time was 94 seconds with a standard deviation of 14 seconds.

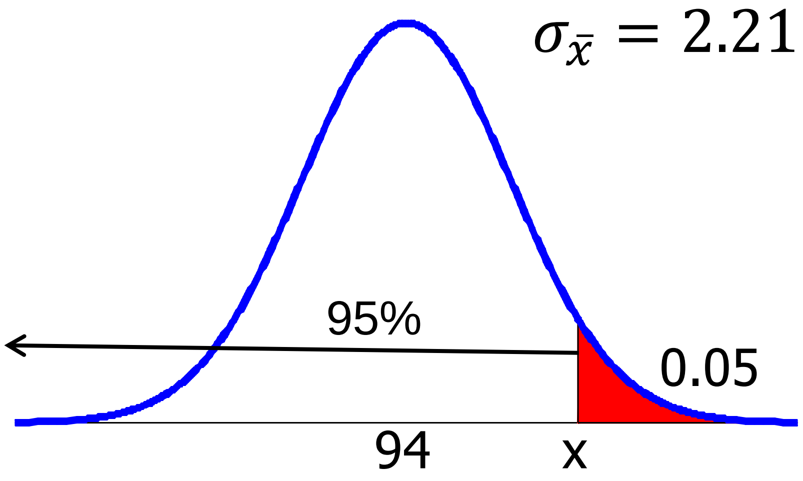
What is the 95% upper confidence level for the average time?

$$\sigma_{\bar{x}} = {}^{s}/_{\sqrt{n}} = {}^{14}/_{\sqrt{40}} = 2.21$$

$Z_{.05} = 1.645$ (from table)

$$x = \bar{x} + Z_{\alpha}\frac{s}{\sqrt{n}}$$

$$= 94 + (1.645)(2.21) = 97.64$$



40 adults took part in a simulated burning house exercise to gather data on the average time to escape after the smoke alarm sounded. The sample mean escape time was 94 seconds with a standard deviation of 14 seconds.

By what time would 95% of individuals be expected to escape from the house?

13. Which of the following statements is correct about working this problem?

A) This problem would be worked the same way as the last one.

B) The applicable standard error of the mean is

$$\sigma_{\bar{x}} = {}^{s}/_{\sqrt{n}} = {}^{14}/_{\sqrt{40}} = 2.21$$

C) You would not use the distribution of the means as the question is about individual observations .

D) The sample size requires use of a double sided approach.

Alternate Notation Confidence Interval Equations

Common notation, two sided

Lower limit

$$\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}}$$

Upper limit

$$\bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}}$$

Textbook notation, two sided

Lower limit

$$\bar{x} - U_{\alpha} \frac{s}{\sqrt{n}}$$

Upper limit

$$\bar{x} + U_{\alpha} \frac{s}{\sqrt{n}}$$

U values found in Table C

My notation: Where
$\boldsymbol{\alpha}$ = total risk in problem,
use Z table with
$Z_{\alpha/2}$ to work problems

If risk $\boldsymbol{\alpha}$=5%, and test is two tailed test, use Z table for $Z_{.05/2}$ = $Z_{.025}$= 1.96

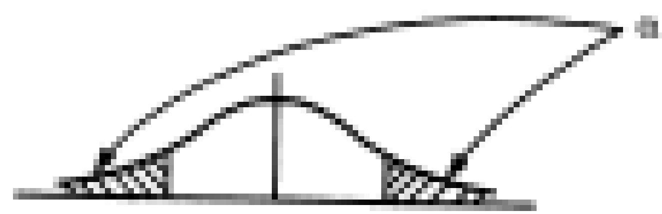
Books notation: Alpha is area under the curve – same on both sides.

If risk $\boldsymbol{\alpha}$=5%, and test is two tailed test, use Table C for U with $\boldsymbol{\alpha}$=5%,.
$U_{.05}$= 1.96

TABLE C Probability Points of the Normal Distribution: Double-Sided (Variance Known)

α only	U	α only	U
0.001	3.291	0.100	1.645
0.005	2.807	0.150	1.440
0.010	2.576	0.200	1.282
0.015	2.432	0.300	1.036
0.020	2.326	0.400	0.842
0.025	2.241	0.500	0.675
0.050	1.960	0.600	0.524

Note 1: The same information can be obtained from Table A; however, this table format is different.
Note 2: In this test the tabular value corresponds to U_α, where α is the value of probability associated with the distribution area pictorially represented as



U_α

Student’s t probability distribution

- Normal tables not as accurate for smaller sample sizes (n<30) where the standard deviation is unknown.

- Student’s t distribution considers sample size - degrees of freedom (n-1)

- Developed by W.S Gosset in 1908

Degrees of Freedom

Degrees of Freedom (df or *v*) = n-1

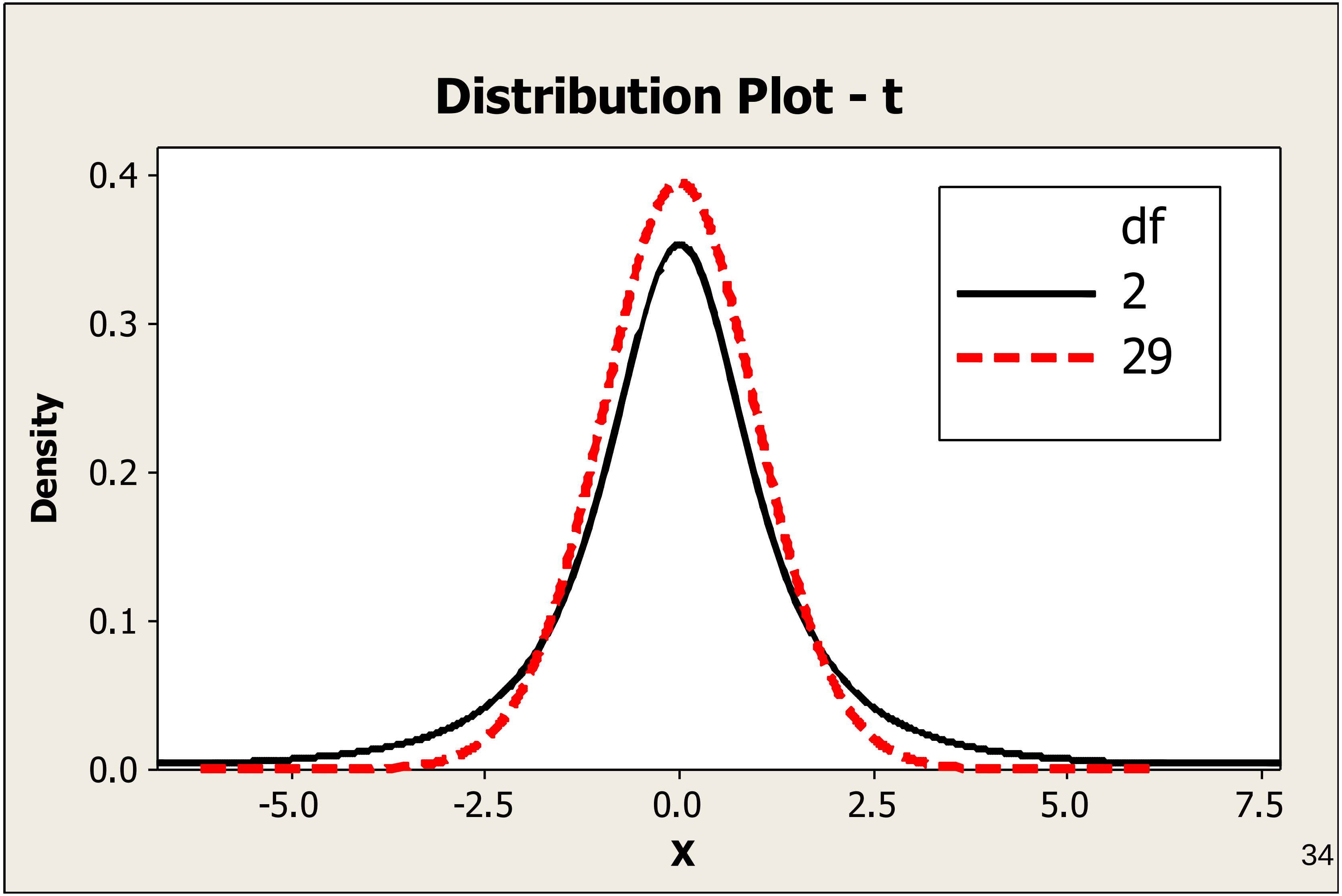
- Is equal to the number of observation values minus the number of parameters estimated in calculating the statistic in question.
- For the sample variance, the 1 is subtracted because we are using $\bar{x}$, which was calculated from the sample data.

Sample Variance $$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

More on Student’s *t*

- Where the sample size is small the distribution of the sample means follows a *t* distribution with n-1 degrees of freedom.
- The probability density of the Student’s *t* distribution is different for different degrees of freedom.
- When *n* is large, the distribution is very close to normal, so the normal curve can be used, rather than the Student’s *t*.

Student’s t probability distribution



Small Sample confidence interval for means ($n \leq 30$)

Must know:

Sample mean, $\bar{x}$

Standard deviation of the distribution of sample mean, $\sigma_{\bar{x}}$

Estimate $\sigma_{\bar{x}}$ with ${}^{s}\!/\!_{\sqrt{n}}$

Confidence Level ($1-\alpha$) / level of risk (α)

Degrees of freedom, n-1

Confidence interval formulas:

Lower Limit

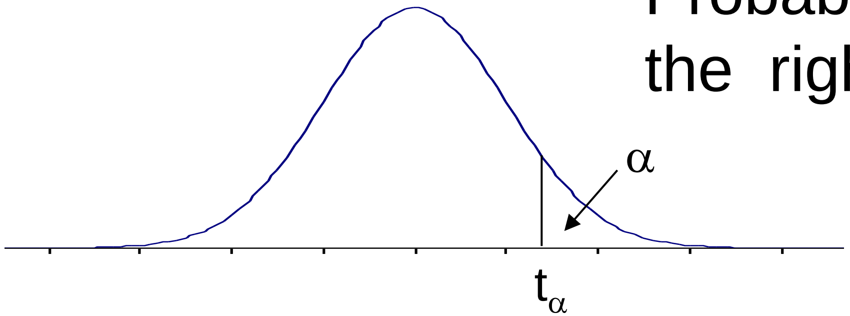
$$\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Upper limit

$$\bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Portion of Table D (page 1043)
Probability Points of the *t* Distribution:
Single sided

Probabilities are in the right side tail



Degrees of freedom
v = n-1

Probabilities across the top ***α***

v	0.40	0.30	0.20	0.10	0.05	0.025	0.01	0.005	0.001
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055	3.930
13	0.259	0.538	0.870	1.354	1.771	2.160	2.650	3.012	3.852
14	0.258	0.537	0.868	1.771	1.761	2.145	2.624	2.997	3.787
15	0.258	0.536	0.866	1.761	1.753	2.131	2.602	2.947	3.733

T values are in the body of the table

Small Sample CI

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

For a random sample of 15 bolts, the average height was found to be 0.330 inches with a standard deviation of 0.012 in. Obtain a 95% confidence interval for the true mean.

v	0.40	0.30	0.20	0.10	0.05	0.025
11	0.260	0.540	0.876	1.363	1.796	2.201
12	0.259	0.539	0.873	1.356	1.782	2.179
13	0.259	0.538	0.870	1.354	1.771	2.160
14	0.258	0.537	0.868	1.771	1.761	2.145
15	0.258	0.536	0.866	1.761	1.753	2.131

n=15

v=n-1=14

$\bar{x}$=0.33

s=0.012

$\boldsymbol{\alpha}=1\text{-}.95=0.05$

$\boldsymbol{\alpha}/2=.025$

$$t_{.05/2,14} = t_{.025,14} = 2.145$$

Small Sample CI

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

For a random sample of 15 bolts, the average height was found to be 0.330 inches with a standard deviation of 0.012 in. Obtain a 95% confidence interval for the true mean.

N=15
v=n-1=14
$\bar{x}$=0.33
s=0.012
$\boldsymbol{\alpha}$=1-.95=..05
$t_{.05/2,14} = t_{.025,14}$= 2.145

Lower confidence limit =

$$0.33 - 2.145 \frac{.012}{\sqrt{15}} = .3233$$

Upper confidence limit =

$$0.33 + 2.145 \frac{.012}{\sqrt{15}} = .3367$$

Small Sample CI

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

A part is assembled by hand. Time studies were performed on assembly of 20 parts. The mean was 5.63 minutes and standard deviation was 0.40 minutes. What is the 95% confidence interval for the true mean for the assembly job?

14. What t value is correct for this problem?

A) $t_{0.025,20}$= 2.086 B) $t_{0.025,19}$= 2.093

C) $t_{0.05,20}$= 1.725 D) $t_{0.025,19}$= 0.688

Small Sample CI

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

A part is assembled by hand. Time studies were performed on assembly of 20 parts. The mean was 5.63 minutes and standard deviation was 0.40 minutes.

15. What is the 95% confidence interval for the true mean for the assembly job?

A) [5.437, 5.822] B) [5.475, 5.785]

C) [5.443, 5.817] D) [5.374, 5.886]

Related Assignments

Please see Blackboard for related assignments