

Topic 7. Confidence Intervals for Population Means and Proportions

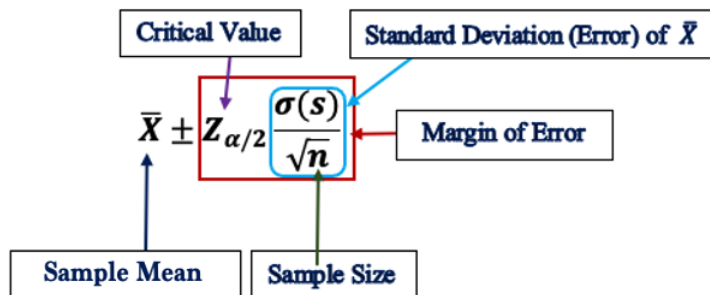
Cheng Peng

Contents

1	Introduction	1
2	Confidence Interval of Proportion	2
3	t - Confidence Interval for Mean (μ)	4
3.1	t-distribution and t-Table	4
3.2	Why t-Confidence Interval	5
3.3	t-confidence Interval by Examples	6
3.4	Normal or t-Confidence Interval: A Summary	7
4	Use of Technology	8
4.1	IntroStatsApps: Normal CI - Proportion	8
4.2	IntroStatsApps: t-CI	9
4.3	IntroStatsApps: t-Table	10
5	Practice Exercises	11

1 Introduction

We have developed the steps for constructing confidence intervals for population means based on large samples with which the central limit theorem can be used to find the sampling distribution of the sample mean.



As a review, we present the following example to highlight the importance of the steps for constructing a confidence interval.

Example 1. Market researchers use the number of sentences per advertisement as a measure of readability for magazine advertisements. The following represents a random sample of the number of sentences found in 50 advertisements. (Source: Journal of Advertising Research)

9 20 18 16 9 9 11 13 22 16 5 18 6 6 5 12 25

17 23 7 10 9 10 10 5 11 18 18 9 9 17 13 11 7
14 6 11 12 11 6 12 14 11 9 18 12 12 17 11 20

The summarized statistics are given in the following table.

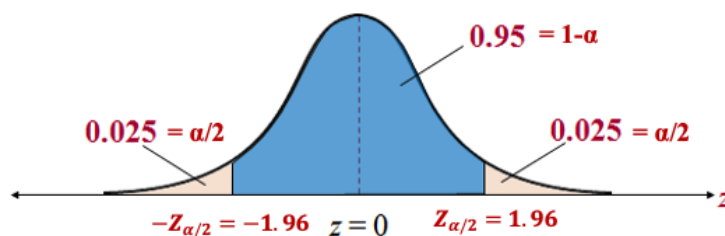
Population Parameter...	Sample Statistic
Mean: μ	$\bar{x} = 12.4$
Standard deviation: σ	$s = 5$

Based on the above data, construct a 95% confidence interval of the mean number of sentences (μ) in all magazine advertisements.

Solution We are going to use the 5-step procedure to construct the confidence interval in the following.

Step 1: We are given a confidence level of 95%.

Step 2: Since the sample size $n = 50 > 30$, using the CLT, we claim that the sample mean is approximately normally distributed. The critical value based on the standard normal distribution should be used. The critical value was found and labeled in the following figure.



Step 3: Find the margin of error E in the following

$$E = Z_{\alpha/2} \times \frac{s}{\sqrt{n}} \approx 1.96 \times \frac{5.0}{\sqrt{50}} \approx 1.4.$$

Step 4: The explicit form of the confidence interval is given by

$$(\bar{X} - E, \bar{X} + E) = (12.4 - 1.4, 12.4 + 1.4) = (11.0, 13.8).$$

Step 5: The confidence interval (11,13.8,) has a 95% chance to include the true average number of sentences in all magazine advertisements.

2 Confidence Interval of Proportion

Recall that the sample proportion (\hat{p}) is approximately normally distributed if both $n\hat{p}$ and $n(1 - \hat{p})$ are large. To be more specific, in this course, if $n\hat{p} \geq 5$ and $n(1 - \hat{p}) \geq 5$ (some textbook use $n\hat{p} \geq 10$ and $n(1 - \hat{p}) \geq 10$). we have

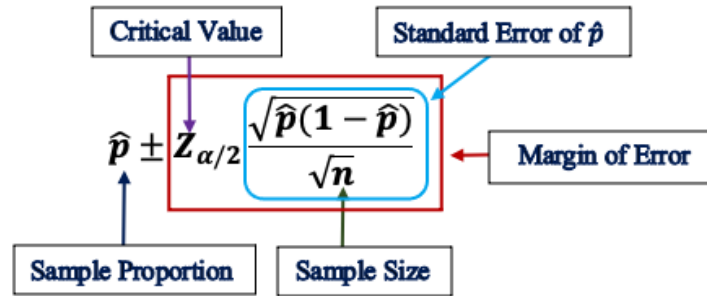
$$\hat{p} \rightarrow N\left(p, \sqrt{\frac{p(1-p)}{n}}\right),$$

which is equivalent to

$$\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \rightarrow N(0, 1).$$

The next Youtube video demonstrate the logical process of constructing the confidence interval for population proportions.

Therefore, we can use the same 5-step procedure as used previously in constructing the confidence interval of the population mean for the large sample. The only difference is that the form of the margin of error involves the sampling error of the sample proportion. The following annotated formula explains the components of the confidence interval of the proportion.



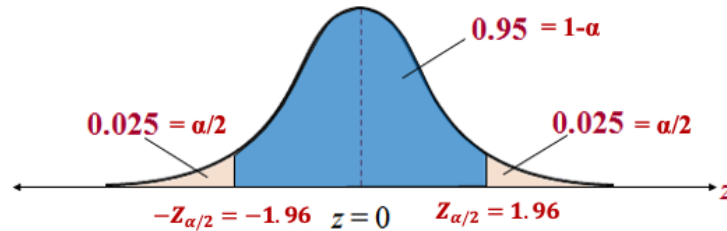
Next, we use an example to illustrate the 5-step procedure for constructing the confidence interval of proportion.

Example 2. In a survey of 1219 U.S. adults, 354 said that their favorite sport to watch is football. Construct a 95% confidence interval for the proportion of adults in the United States who say that their favorite sport to watch is football.

Solution: First of all, the sample proportion $\hat{p} = 354/1219 \approx 0.29$ and sample size $n = 1219$.

Step 1: The confidence level is $1 - \alpha = 0.95$. $\alpha/2 = 0.025$.

Step 2: Since $n\hat{p} = 354 > 5$ and $n(1 - \hat{p}) = 1219 \times 0.71 = 865 > 5$, the sampling distribution of \hat{p} is normally distributed. The critical value corresponding to 95% confidence level is $Z_{0.025} = 1.96$ (from the normal table).



Step 3: The margin of error for the confidence interval of population proportion is given by

$$E = Z_{\alpha/2} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96 \times \sqrt{\frac{0.29(1-0.29)}{1219}} \approx 0.025.$$

Step 4: The explicit form of the confidence interval is given by

$$(\hat{p} - E, \hat{p} + E) = (0.29 - 0.025, 0.29 + 0.025) = (0.365, 0.315).$$

Step 5: The confidence interval (0.265, 0.315) has a 95% chance to include the true proportion of adults who say football is their favorite sport.

3 t - Confidence Interval for Mean (μ)

We have constructed confidence intervals based on the CLT that require a large sample size. When we construct a confidence interval based on small samples, we need to make stronger assumptions about the population so that we have enough information for the interval.

3.1 t-distribution and t-Table

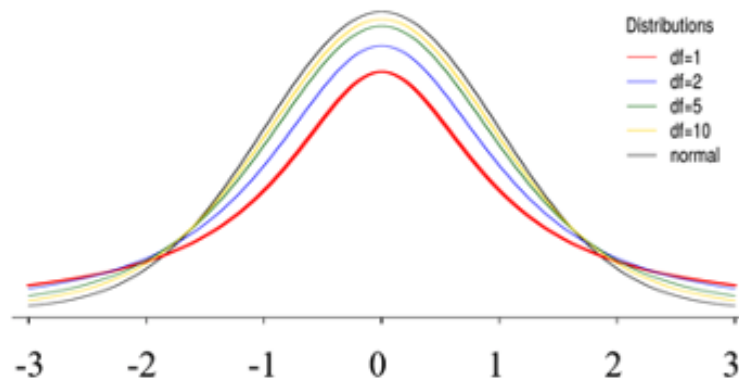
We have pointed out that

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \rightarrow N(0, 1)$$

if the sample is from a normal population regardless of the sample size. However, if the population is unknown, we need to use the sample standard deviation to estimate the population standard deviation. In this case,

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \not\rightarrow N(0, 1)$$

if the sample size is small! The correct distribution is called t-distribution with $n - 1$ degrees of freedom. The difference between standard normal and t-distributions are depicted in the following figure.

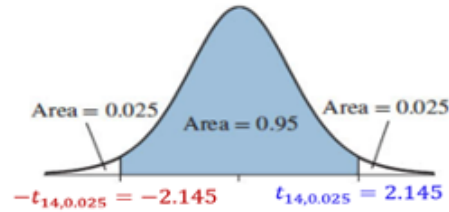


We can see from the above figure that t-distributions are also bell-shaped. As the degrees of freedom increase, the t-distributions approach the standard normal distribution! **However**, when the degree of freedom is small, the t-distribution and the standard normal distribution are VERY different.

This implies that when a small sample from a normal population with unknown population variance, we should NOT use the normal critical value to construct the confidence interval for the unknown population mean. Instead, we MUST use the t-critical value that can be found in the t-table!

Example 3. Find the critical value for a 95% confidence when the sample size is 15.

Solution since sample size $n = 15$, the degrees of freedom $df = n - 1 = 14$. We use the notation $t_{df, \alpha/2}$ to denote the critical value based on the t-distribution. The following figure shows the location of the CV on the density curve.



The critical value $CV = t_{14,0.025} = 2.145$ which is found from the t-distribution table. The following table illustrates the structure of the t-table and how to use the t-table to find the critical value.

Degrees of Freedom	Area in Right Tail									
	0.40	0.25	0.10	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	127.321	318.309	636.619
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215	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	3.923

3.2 Why t-Confidence Interval

Note that

$$TS = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \rightarrow N(0, 1)$$

If one of the following two conditions is satisfied

- Sample size is large (the conventional threshold is $n > 30$), or
- Populations is normal and population variance, σ^2 , or standard deviation, σ , is **known**. In this case, there is no restriction on the sample size.

However, if population is normal and population standard deviation σ is **unknown**, we need to use sample standard deviation, s , to estimate the **unknown** σ . In this case

$$TS = \frac{\bar{x} - \mu}{s/\sqrt{n}} \rightarrow t_{n-1}.$$

That is, the critical value required in constructing the confidence interval of population mean is based on the t-distribution. The following YouTube video explain this situation with examples.

Remark: Note that there is no restriction on the sample size in this scenario. If $n > 30$, we can use either the central limit theorem to estimate the underlying t-distribution with the standard normal distribution or the t-distribution. However, if $n < 30$, the t-critical value **must be used** in constructing the confidence interval.

The following Youtube video explains the steps in constructing the t-confidence interval with an examples.

3.3 t-confidence Interval by Examples

The steps for constructing t-confidence intervals are identical to those in the normal confidence intervals except for the critical value that is found in the t-table. We will use an example to show the steps for constructing a t-confidence interval.

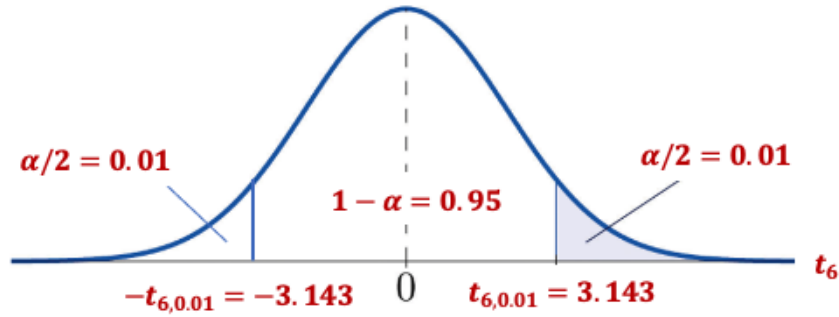
Example 4. . Estimating Car Pollution - In a sample of seven cars, each car was tested for nitrogen-oxide emissions (in grams per mile) and the following results were obtained: 0.06, 0.11, 0.16, 0.15, 0.14, 0.08, 0.15 (based on data from the Environmental Protection Agency). Assuming that this sample is representative of the cars in use. Further, the amounts of nitrogen-oxide emission for all cars are normally distributed. Construct a 98% confidence interval estimate of the mean amount of nitrogen-oxide emission for all cars

Solution: We first calculate the sample mean and sample standard deviation using the formulas introduced in the note of descriptive statistics.

$$\bar{x} = 0.1214, \quad s = 0.0389.$$

Step 1: The confidence level for the t-confidence interval is $1 - \alpha = 0.98$. The right-tailed area of the t-density curve is $\alpha/2 = 0.01$

Step 2: We need to use the t-critical value to calculate the confidence interval: $CV = t_{6,0.01} = 3.143$. The density curve and the table below.



Degrees of Freedom	0.40	0.25	0.10	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	127.321	318.309	636.619
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.277	0.765	1.638	2.353	3.182	4.841	5.841	7.453	10.215	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	4.397	4.781

Step 3: The margin of error

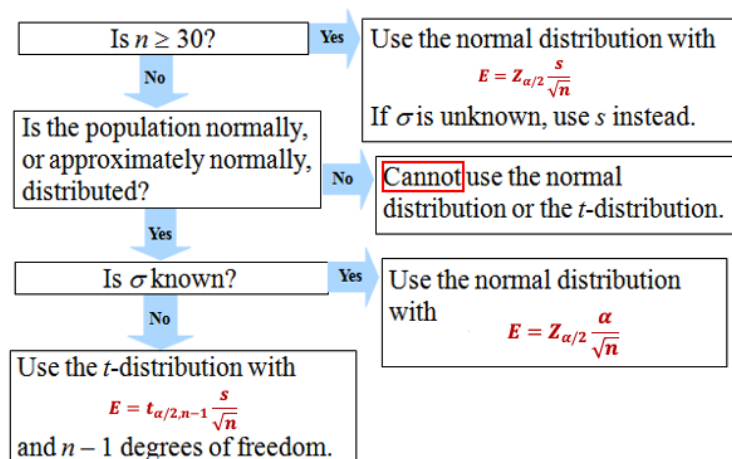
$$E = t_{n-1, \alpha/2} \times \frac{s}{\sqrt{n}} = 3.143 \times \frac{0.0398}{\sqrt{7}} \approx 0.0462.$$

Step 4: Therefore, the 98% confidence interval estimate of the population mean is $(\bar{x} - E, \bar{x} + E) = (0.1214 - 0.0462, 0.1214 + 0.0462) = (0.075, 0.168)$.

Step 5: The interval (0.075, 0.168) has a 98% chance to include the true average amount of nitrogen-oxide emission of all cars in use.

3.4 Normal or t-Confidence Interval: A Summary

When a normal or t confidence interval should be constructed is dependent on the given amount of information. The following brief flow chart summarizes the selection of the two types of confidence intervals based on different conditions.



4 Use of Technology

Two **ISLA** apps can be used to generate solutions to most of the problems. You can use these apps to check your work and make sure you correctly understand the concepts and steps for finding the confidence intervals of either population mean or proportion.

4.1 IntroStatsApps: Normal CI - Proportion

The following apps generate solutions for confidence intervals of population proportions. <https://wcupeng.shinyapps.io/NormalCI4MeanProp/>

Example: Suppose that a market research firm is hired to estimate the percent of adults living in a large city who have cell phones. Five hundred randomly selected adult residents in this city are surveyed to determine whether they have cell phones. Of the 500 people surveyed, 421 responded yes - they own cell phones. Using a 95% confidence level, compute a confidence interval estimate for the true proportion of adult residents of this city who have cell phones.

To use the app, we need to estimate the sample proportion: $421/500 = 0.842$ and $n = 500$.

ISLA: NORMAL C.I.s FOR POPULATION MEAN AND PROPORTION

1. CI for μ or p ?

☒ Population Mean (μ)
☐ Population Proportion (p)

2. Sample Statistics

Sample Proportion [in decimal form]: \hat{p}


3. Sample Size and Confidence Level

Sample Size: n

Confidence Level: α

80% 99%

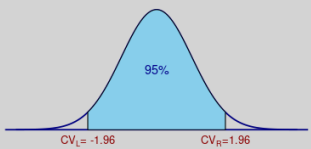
80 82 84 86 88 90 92 94 96 98 99


 Report bugs to C. Peng

Steps for Constructing C.I.

Step 1. The given confidence level.
 $\text{conf.level} = 1 - \alpha = 95\%$.

Step 2. CV on the standard normal density curve.
 $CV = Z_{\alpha/2} = 1.960$.



Step 3: Margin of Error

$$E = CV \times \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = 1.96 \times \sqrt{\frac{0.842(1 - 0.842)}{500}} = 0.032.$$

Step 4: Expression of Confidence Interval

$$(\hat{p} - E, \hat{p} + E) = (0.842 - 0.032, 0.842 + 0.032) = (0.81, 0.874).$$

Step 5: Interpretation of Confidence Interval

There is a 95% chance that the confidence interval (0.81, 0.874) contains the true population proportion.

4.2 IntroStatsApps: t-CI

The following apps generate solutions for t-confidence intervals. The direct link is at: <https://wcu-peng.shinyapps.io/Student-t-CI/>

Example: The yearly salary for mathematics assistant professors are normally distributed. A random sample of 8 math assistant professor's salaries are listed below in thousands of dollars. Estimate the population mean salary with a 99% confidence interval.

66.0, 75.8, 70.9, 73.9, 63.4, 68.5, 73.3, 65.9

To use the app, we need to calculate the sample mean and sample **standard deviation**: Using the formulas, we have $\bar{x} = 69.71$ and $s = 4.45$.

ISLA: t CONFIDENCE INTERVALS FOR POPULATION MEAN

1. Sample Statistics

Sample Mean: \bar{X}

Standard Deviation: s or σ_0

2. Sample Size and Confidence Level

Sample Size: n

Confidence Level: $1 - \alpha$


80%

90%

95%

99%

95%

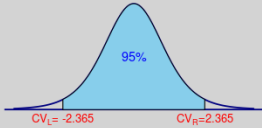


Report bugs to C. Peng

Solution: The desired confidence interval is constructed in the following steps.

Step 1. The given confidence level.
 $\text{conf.level} = 1 - \alpha = 1 - 0.05 = 95\%$

Step 2. CV on the t density curve.
 $CV = t(\alpha/2, df) = t(0.03, 7) = 2.365$



Step 3: Margin of Error

$$E = CV \times \frac{s}{\sqrt{n}} = 2.365 \times \frac{4.45}{\sqrt{8}} = 3.721.$$

Step 4: Expression of Confidence Interval

$$(\bar{X} - E, \bar{X} + E) = (69.71 - 3.721, 69.71 + 3.721) = (65.989, 73.431).$$

Step 5: Interpretation of Confidence Interval

There is a 95% chance that the confidence interval (65.989, 73.431) contains the true population mean.

4.3 IntroStatsApps: t-Table

For convenience, I also create an interactive t-table that provides the correct probability or percentile based on the provided information. The direct link is at: <https://wcu-peng.shinyapps.io/t-Table/>

The following screenshot is the solution to **Example 3**.

ISLA: T-DISTRIBUTION TABLE

Instruction

This app creates a t-table that is commonly used in introductory statistics textbooks. It is primarily used for finding critical values.

1. What Tail Probability Is Given?

- ☐ Right
☒ Left

2. Given Right Tail Probability

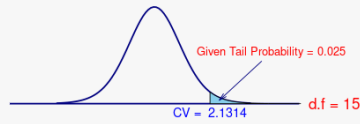
0.025

3. Degrees of Freedom ($n - 1$)

15



Report bugs to C. Peng



The numbers in the top row are tail probabilities.

d.f.	0.4000	0.2500	0.1000	0.0500	0.0250	0.0100	0.0050	0.0025	0.0010	0.0005
10	0.2602	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693	3.5814	4.1437	4.5869
11	0.2596	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058	3.4966	4.0247	4.4370
12	0.2590	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545	3.4284	3.9296	4.3178
13	0.2586	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123	3.3725	3.8520	4.2208
14	0.2582	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768	3.3257	3.7874	4.1405
15	0.2579	0.6912	1.3406	1.7531	2.1314	2.6025	2.9467	3.2860	3.7328	4.0728
16	0.2576	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208	3.2520	3.6862	4.0150
17	0.2573	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982	3.2224	3.6458	3.9651
18	0.2571	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784	3.1966	3.6105	3.9216
19	0.2569	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609	3.1737	3.5794	3.8834
20	0.2567	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453	3.1534	3.5518	3.8495

5 Practice Exercises

Please do the following problems manually and then use the **ISLA** apps to check your work.

1. A team of efficiency experts intends to use the mean of a random sample of size $n = 150$ to estimate the average mechanical aptitude of assembly-line workers in a large industry (as measured by a certain standardized test) and found that the sample mean is 19.9 minutes, and the sample standard deviation is 5.73 minutes. Construct a 95% confidence interval for the average mechanical aptitude of assembly line workers in the given industry.
2. A college admissions director wishes to estimate the mean age of all students currently enrolled. In a random sample of 20 students, the mean age is found to be 22.9 years. From past studies, the standard deviation is known to be 1.5 years, and the population is normally distributed. Construct a 90% confidence interval of the population mean age.
3. Suppose you are interested in investigating the factors that affect the prevalence of tuberculosis among intravenous drug users. In a group of 97 individuals who admit to sharing needles, 24.7% had a positive tuberculin skin test result. Construct 95% confidence intervals for the population proportions.

4. Find the confidence interval for average SAT score for women given that $n = 15$, the sample mean is 496, and the sample standard deviation is 108. Assume that the SAT scores are normally distributed.