

Chapter Seven Confidence Intervals and Sample Size



- 1. Preface
- 2. Confidence Intervals for the Mean When σ is Known
- 3. Confidence Intervals for the Mean When σ is Unknown
- 4. Confidence Intervals and Sample Size for Proportions
- 5. Confidence Intervals for Variances and Standard Deviations

Assi. Prof. Dr. Taher M. Ahmed Civil Engineering Department University of Anbar

Chapter Six

Confidence Intervals and Sample Size

- ❑ A survey by the Roper Organization found that 45% of the people who were offended by a television program would change the channel, while 15% would turn off their television sets. The survey further stated that the margin of error is 3 percentage points, and 4000 adults were interviewed. Several questions arise:
- 1. How do these estimates compare with the true population percentages?
- 2. What is meant by a margin of error of 3 percentage points?
- 3. Is the sample of 4000 large enough to represent the population of all adults who watch television in the United States?
- Inferential statistical techniques have various assumptions that must be met before valid conclusions can be obtained.
- 1. The samples must be randomly selected.

1. Preface:

- 2. The sample size must be greater than or equal to 30 or less.
- 3. The population must be normally or approximately normally distributed based on sample size.

2. Confidence Intervals for the Mean When σ is Known

<u>2.1. A point estimate</u> is a specific numerical value estimate of a parameter. The best point estimate of the population mean " μ " is the sample mean " \overline{X} ".

Example: The president of university want to estimate the average age of the student (μ). He could select a random sample of 100 students and find the average age (\overline{X}) of these students, say, 22.3 years. From the sample mean, the president could infer that the average age of all the students is 22.3 years. So \overline{X} is <u>estimator</u> for the population (μ).

A good estimator should satisfy the following criteria:

1. The estimator should be an **unbiased estimator.** That is, the expected value or the mean of the estimates obtained from samples of a given size is equal to the parameter being estimated.

2. The estimator should be consistent. For a **consistent estimator**, as sample size increases, the value of the estimator approaches the value of the parameter estimated.

3. The estimator should be a **relatively efficient estimator.** That is, of all the statistics that can be used to estimate a parameter, the relatively efficient estimator has the smallest variance

- **2.2.** An interval estimate of a parameter is an interval or a range of values used to estimate the parameter. This estimate may or may not contain the value of the parameter being estimated.
- In an interval estimate, the parameter is specified as being between two values. For example, an interval estimate for the average age of all students might be 21.9 ≤ μ ≤ 22.7, or 22.3 ± 0.4 years.

2.3. Confidence Intervals

- A confidence interval is a specific interval estimate of a parameter determined by using data obtained from a sample and by using the specific confidence level of the estimate.
- The confidence level of an interval estimate of a parameter is the probability that the interval estimate will contain the parameter, assuming that a large number of samples are selected and that the estimation process on the same parameter is repeated.



 For instance, you may wish to be 95% confident that the interval contains the true population mean.

2.3. Confidence Intervals Formula

Formula for the Confidence Interval of the Mean for a Specific α When σ is Known is: $\overline{\mathbf{X}} = \mathbf{Z}$ is $(\frac{\sigma}{2}) \leq \mu \leq \overline{\mathbf{X}} + \mathbf{Z}$ is $(\frac{\sigma}{2})$

$$\overline{X} - Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \overline{X} + Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

- For a 90% confidence interval, $z_{a/2} = 1.65$; for a 95% confidence interval, $z_{a/2} = 1.96$; and for a 99% confidence interval, $z_{a/2} = 2.58$. $\sigma/2$ is the standard error,
- The term $Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$ is called the margin of error (also called the maximum error of the estimate). For a specific value, say, $\alpha = 0.05$, 95% of the sample means will fall within this error value on either side of the population mean.
- The margin of error also called the maximum error of the estimate is the maximum likely difference between the point estimate of a parameter and the actual value of the parameter.



Assumptions for Finding a Confidence Interval for a Mean When σ Is Known

1. The sample is a random sample.

2. Either $n \ge 30$ or the population is normally distributed if n < 30.

Example 1: A researcher wishes to estimate the number of days it takes an automobile dealer to sell a Chevrolet Aveo. A sample of 50 cars had a mean time on the dealer's lot of 54 days. Assume the population standard deviation to be 6.0 days. Find the best point estimate of the population mean and the 95% confidence interval of the population mean.

Solution

The best point estimate of the mean is 54 days. For the 95% confidence interval use z = 1.96 (from table E).

$$\overline{X} - Z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \overline{X} + Z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}}\right)$$

$$54 - 1.96 \left(\frac{6.0}{\sqrt{50}}\right) < \mu < 54 + 1.96 \left(\frac{6.0}{\sqrt{50}}\right)$$

$$54 - 1.70 < \mu < 54 + 1.70$$

$$52.3 < \mu < 55.7 \quad OR \ 54 \pm 1.70$$



One can say with 95% confidence that the interval between 52.3 and 55.7 days does contain the population mean, based on a sample of 50 automobiles. **Example 2:** A survey of 30 emergency room patients found that the average waiting time for treatment was 174.3 minutes. Assuming that the population standard deviation is 46.5 minutes, find the best point estimate of the population mean and the 99% confidence of the population mean.

Solution

The best point estimate is 174.3 minutes. The 99% confidence is interval use z = 2.58 (from table E).

 $1JL + \setminus \mu \setminus 1JUL$

$$\overline{X} - Z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \overline{X} + Z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}}\right)$$

$$174.3 - 2.58 \left(\frac{46.5}{\sqrt{30}}\right) < \mu < 174.3 + 2.58 \left(\frac{46.5}{\sqrt{30}}\right)$$

$$.74.3 - 21.9 < \mu < 174.3 + 21.9$$

Example 3: The following data represent a sample of the assets (in millions of dollars) of 30 credit unions in southwestern Pennsylvania. Find the 90% confidence interval of the mean.

One	can	say	with	99%			
confid	ence	that	the	mean			
waitin	g tim	e for	eme	rgency			
room	treatr	ment	is be	tween			
152.4 and 196.2 minutes.							

	12.23	16.56	4.39
	2.89	1.24	2.17
	13.19	9.16	1.42
22	73.25	1.91	14.64
1	11.59	6.69	1.06
3	8.74	3.17	18.13
	7.92	4.78	16.85
	40.22	2.42	21.58
	5.01	1.47	12.24
1	2.27	12.77	2.76

Solution

- 1. Find the mean and standard deviation ($\overline{X} = 11.091$, $\sigma = 14.405$.
- 2. Confident intervals = 0.9 ; α = 1- 0.9 = 0.1 ; $\alpha/2$ = 0.05.
- 3. $Z_{\alpha/2} = 1.68$ from table E. $\overline{X} Z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \overline{X} + Z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}}\right)$

$$11.091 - 1.65 \left(\frac{14.405}{\sqrt{30}}\right) < \mu < 11.091 + 1.65 \left(\frac{14.4}{\sqrt{30}}\right)$$
$$11.091 - 4.339 < \mu < 11.091 + 4.339$$

3. Sample Size

- Sample size depends on: the margin of error, the population standard deviation, and the degree of confidence.
- the margin of error formula is:



 $6.752 < \mu < 15.430$



where E is the margin of error. If necessary, round the answer up to obtain a whole number. That is, if there is any fraction or decimal portion in the answer, use the next whole number for sample size n.

Example 4: A scientist wishes to estimate the average depth of a river. He wants to be 99% confident that the estimate is accurate within 2 feet. From a previous study, the standard deviation of the depths measured was 4.33 feet.



4. Confidence Intervals for the Mean When σ is Unknown

Most of the time, the value of " σ " is not known, so it must be estimated by using "S", namely, the standard deviation of the sample. When *S* is used, especially when the sample size is small, the *Student* **t** *distribution*, most often called the **t** distribution is used instead of normal distribution (Z).



The t distribution shares some characteristics of the normal distribution and differs from it in others.

Similarity: between t and normal distributions:

- 1. It is bell-shaped.
- 2. It is symmetric about the mean.
- 3. The mean, median, and mode are equal to 0 and are located at the center of the distribution.
- 4. The curve never touches the x axis.

The *t* distribution differs from the standard normal distribution in the following:

- 1. The variance is greater than 1.
- 2. The *t* distribution is actually a family of curves based on the concept of *degrees of freedom*, which is related to sample size.

3. As the sample size increases, the *t* distribution approaches the standard normal distribution.

Formula for a Specific Confidence Interval for the Mean When S is Unknown

- The values for $t_{a/2}$ are found in Table F.
- Degree of freedom d.f. = n-1

$$\overline{X} - t_{\alpha/2}\left(rac{s}{\sqrt{n}}
ight) < \mu < \overline{X} + t_{\alpha/2}\left(rac{s}{\sqrt{n}}
ight)$$

Example 5: Find the ta2 value for a 95% confidence interval when the sample size is 22.

Solution

The d.f. = 22 - 1 = 21.

Find 21 in the left column and 95% in the row labeled Confidence Intervals. The intersection where the two meet gives the value for $t_{a/2}$, which is 2.080.

Table F									
The <i>t</i> Distribution									
	Confidence Intervals	50%	80%	90%	95%	98%	99%		
d f	One tail α	0.25	0.10	0.05	0.025	0.01	0.005		
u.i.	Two tails α	0.50	0.20	0.10	0.05	0.02	0.01		
1									
2									
3					V				
21-				> >	2.080	2.518	2.831		
: (Z)∞		0.674	1.282 ^a	1.645 ^{<i>b</i>}	1.960	2.326 [°]	2.576 ^d		

Table F	The t Distribution					
	Confidence intervals	80%	90%	95%	98 %	99 %
	One tail, α	0.10	0.05	0.025	0.01	0.005
d.f.	Two tails, α	0.20	0.10	0.05	0.02	0.01
1		3,078	6.314	12,706	31,821	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		1,415	1,895	2,365	2,998	3.499
8		1.397	1.860	2.306	2,896	3.355
9		1.383	1.833	2,262	2.821	3.250
10		1.372	1.812	2,228	2.704	3.109
12		1.365	1.790	2,201	2.718	3.000
12		1.350	1.764	2.179	2,081	3.033
14		1.330	1.7/1	2.160	2.630	3.012
15		1.341	1.761	2.145	2.602	2.9/7
16		1 3 37	1.735	2.130	2.583	2.94/
17		1 333	1.740	2110	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
32		1.309	1.694	2.037	2,449	2,738
34		1.307	1.691	2.032	2,441	2,728
36		1.306	1,688	2,028	2,434	2,719
38		1.304	1.686	2.024	2,429	2,712
40		1.303	1.684	2.021	2,423	2,704
45		1.301	1.679	2,014	2,412	2,690
50		1,299	1.070	2,009	2,403	2.0/8
55		1.297	1.671	2,004	2,390	2,008
60		1.290	1.671	2.000	2.390	2,660
70		1.295	1,667	1.99/	2.380	2.034
75		1 293	1.665	1.997	2 3 77	2 643
80		1.292	1.664	1.990	2.374	2.639
90		1.291	1.662	1.987	2 368	2 632
100		1,290	1.660	1.984	2.364	2,626
500		1.283	1.648	1.965	2.334	2.586
1000		1.282	1.646	1.962	2.330	2.581
		1.0007	1 (40)	1.040	2.2.00	2 576

"This value has been rounded to 1.28 in the textbook.

^b This value has been rounded to 1.65 in the textbook. "This value has been munded to 2.33 in the textbook.

^d This value has been rounded to 2.58 in the textbook.

Source: Adapted from W. H. Beyer, Handbook of Tables for Probability and Statistics, 2nd ed., CRC Press, Boca Raton, Fla., 1986. Reprinted with permission.



> Assumptions for Finding a Confidence Interval for a Mean When S is Unknown

1. The sample is a random sample.

2. Either $n \ge 30$ or the population is normally distributed if n < 30

Example 6: Ten randomly selected people were asked how long they slept at night. The mean time was 7.1 hours, and the standard deviation was 0.78 hour. Find the 95% confidence interval of the mean time. Assume the variable is normally distributed.

Solution

Since σ is unknown and S must replace it, the t distribution (Table F) must be used for the confidence interval.

d.f. = 10 -1 = 9 The confidence interval = 95% Substituting in the formula. $\overline{X} - t_{\alpha/2} \left(\frac{s}{\sqrt{n}}\right) < \mu < \overline{X} + t_{\alpha/2} \left(\frac{s}{\sqrt{n}}\right)$ $7.1 - 2.262 \left(\frac{0.78}{\sqrt{10}}\right) < \mu < 7.1 + 2.262 \left(\frac{0.78}{\sqrt{10}}\right)$ $6.54 < \mu < 7.66$

Therefore, 95% confident that the population mean is between 6.54 and 7.66 hr. Example 7: The data represent a sample of the number of home fires started by candles for the past several years. (Data are from the National Fire Protection Association.) Find the 99% confidence interval for the mean number of home fires started by candles each year. **5460 5900 6090 6310 7160 8440 9930**

Solution

- Find the mean and standard deviation for the data. ($\overline{X} = 7041.4 \& S = 1610.3$)
- Confidence interval = 99%; d.f. = 6. From table $t_{\alpha/2}$ = 3.707.
- Substitute in the formula and solve. $\overline{X} t_{\alpha/2} \left(\frac{s}{\sqrt{n}}\right) < \mu < \overline{X} + t_{\alpha/2} \left(\frac{s}{\sqrt{n}}\right)$

 $7041.4 - 3.707 \left(\frac{1610.3}{\sqrt{7}}\right) < \mu < 7041.4 + 3.707 \left(\frac{1610.3}{\sqrt{7}}\right) \quad 4785.2 < \mu < 9297.6$

 So, at 99% confident that the population mean number of home fires started by candles each year is between 4785.2 and 9297.6.

OVERALL: As stated previously, when σ is known, $Z_{\alpha/2}$ values can be used *no matter* what the sample size is, as long as the variable is normally distributed or $n \ge 30$. When σ is unknown and $n \ge 30$, then S can be used in the formula and $t_{\alpha/2}$ values can be used. Finally, when σ is unknown and n < 30, S is used in the formula and $t_{\alpha/2}$ values are used, as long as the variable is approximately normally distributed.



5. Confidence Intervals for Variances and Standard Deviations

- In statistics, the variance and standard deviation of a variable are as important as the mean. For example, when products that fit together (such as pipes) are manufactured, it is important to keep the variations of the diameters of the products as small as possible; otherwise, they will not fit together properly and will have to be scrapped. In the manufacture of medicines, the variance and standard deviation of the medication in the pills play an important role in making sure patients receive the proper dosage. For these reasons, confidence intervals for variances and standard deviations are necessary.
- To calculate these confidence intervals, a new statistical distribution is needed. It is called the chi-square distribution (χ^2).
- The chi-square variable is similar to the t variable in that its distribution is a family
 of curves based on the number of degrees of freedom.
- The chi-square distribution is obtained from the values of



- A chi-square variable cannot be negative, and the distributions are skewed to the right. At about 100 degrees of freedom, the chi-square distribution becomes somewhat symmetric. The area under each chisquare distribution is equal to 1.00.
- Table G gives the values for the chisquare distribution.



Table G	The Chi-Squ	are Distrib	ution							
Degrees of	α									
freedom	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	_	_	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24,736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36,191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.262	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17,708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Source: Owen, Handbook of Statistical Tables, Table A.-4 "Chi-Square Distribution Table," O 1962 by Addison-Weakey Publishing Company, Inc. Copyright renewal © 1990. Reproduced by permission of Pearson Education, Inc.



Example 8: Find the values for χ^2 right and χ^2 left for a 90% confidence interval when n 25.

Solution:

```
To find \chi^2 right, \alpha = 1 - 0.90 = 0.10; \alpha/2 = 0.05.
To find \chi^2 left, \alpha = 1 - 0.95 = 0.05.
d.f. = n-1=25-1=24
```

Hence, use the 0.95 and 0.05 columns and the row corresponding to d.f. = 24.

• From table: χ^2 = 36.415 at right; and χ^2 = 13.848 at left



5.1. Confidence Interval for a Variance

a Variance
$$\frac{(n-1)S^2}{\chi^2_{right}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{left}}$$

5.2. Confidence Interval for a Standard Deviation

Assumptions:

- 1. The sample is a random sample.
- 2. The population must be normally distributed

Example 9: Find the 95% confidence interval for the variance and standard deviation of the nicotine content of cigarettes manufactured if a sample of 20 cigarettes has a standard deviation of 1.6 milligrams.

Solution

d.n. = n - 1 = 20 - 1 = 19 $\alpha = 1 - 0.95 = 0.05$ $\alpha/2 = 0.05/2 = 0.025 \text{ (right)}$ $\chi^{2}_{right} = \chi^{2}_{0.025} = 32.852$ $1 - \alpha/2 = 0.975 \text{ (left)}$ $\chi^{2}_{left} = \chi^{2}_{0.975} = 8.907$

$$\frac{(n-1)S^2}{\chi^2_{right}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{left}}$$
$$\frac{(20-1)(1.6)^2}{32.852} < \sigma^2 < \frac{(20-1)(1.6)^2}{8.907}$$
$$1.5 < \sigma^2 < 5.5$$

$$\sqrt{\frac{(n-1)S^2}{\chi^2_{right}}} < \sigma < \sqrt{\frac{(n-1)S^2}{\chi^2_{left}}}$$



 $1.2 < \sigma < 2.3$

Example 10: Find the 90% confidence interval for the variance and standard deviation for the stability test of asphalt cores in kN. The data represent a selected sample from a specific mix designed for a road. Assume the variable is normally distributed. 59 53 52 51 39 49 54 46 49 48

Solution:

0

 χ^2_{left} = 8.907

Determine the variance for the data; S² = 28.2.

1-α=0.95

• $1 - \alpha = 1 - 0.9 = 0.1$; $\alpha_{left} = 0.05$, $\alpha_{right} = 0.9 + 0.05 = 0.95$; d.n. = n - 1 = 10 - 1 = 9

 χ^2_{right} = 32.852

•α/2

• Find χ^2_{left} from Table = 3.325; χ^2_{right} from Table = 16.919

$$\frac{(n-1)S^2}{\chi^2_{right}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{left}} \quad \frac{(9)(28.2)}{16.919} < \sigma^2 < \frac{(9)(28.2)}{3.325}$$

$$\frac{15 < \sigma^2 < 76.3}{\sqrt{\frac{(n-1)S^2}{\chi^2_{right}}}} < \sigma < \sqrt{\frac{(n-1)S^2}{\chi^2_{left}}} \quad 3.87 < \sigma < 8.73$$