Estimation and Confidence Intervals

Chapter 9

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

GOALS

- Define a point estimate.
- Define level of confidence.
- Construct a confidence interval for the population mean when the population standard deviation is known.
- Construct a confidence interval for a population mean when the population standard deviation is unknown.
- Construct a confidence interval for a population proportion.
- Determine the sample size for attribute and variable sampling.

Point and Interval Estimates

- A point estimate is the statistic, computed from sample information, which is used to estimate the population parameter.
- A confidence interval estimate is a range of values constructed from sample data so that the population parameter is likely to occur within that range at a specified probability. The specified probability is called the level of confidence.

Factors Affecting Confidence Interval Estimates

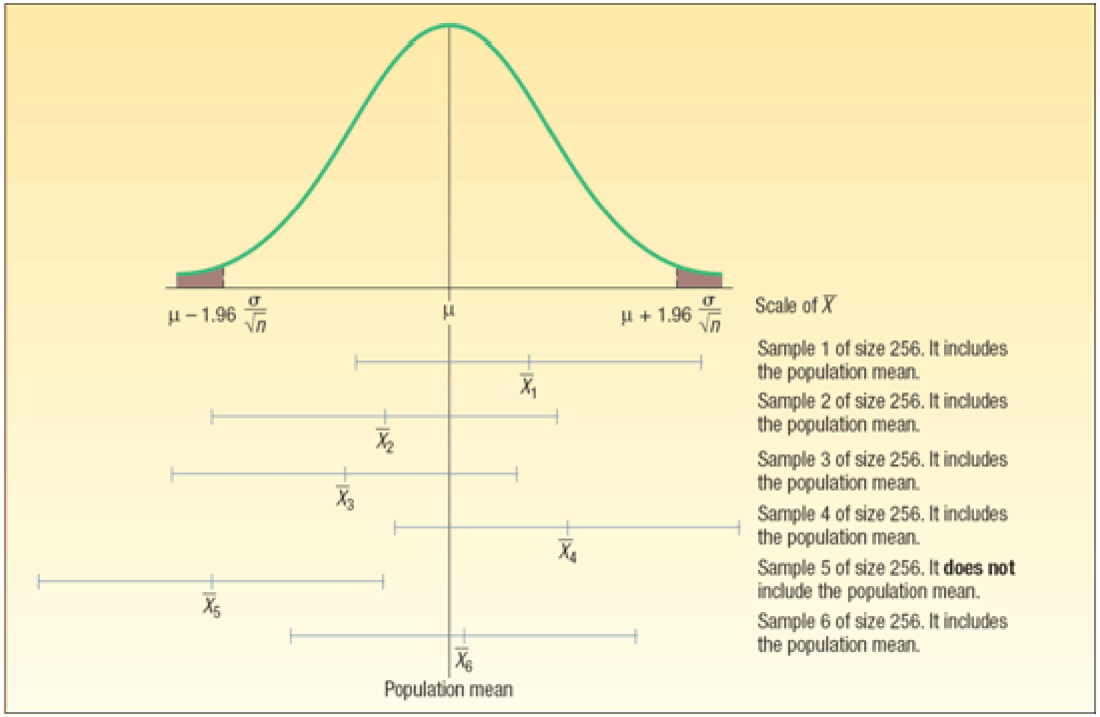
The factors that determine the width of a confidence interval are:

1. The sample size, *n*.
2. The variability in the population, usually σ estimated by *s*.
3. The desired level of confidence.

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

Interval Estimates - Interpretation

For a 95% confidence interval about 95% of the similarly constructed intervals will contain the parameter being estimated. Also 95% of the sample means for a specified sample size will lie within 1.96 standard deviations of the hypothesized population



Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

Characteristics of the t-distribution

1. It is, like the *z* distribution, a continuous distribution.
2. It is, like the *z* distribution, bell-shaped and symmetrical.
3. There is not one t distribution, but rather a family of t distributions. All *t* distributions have a mean of 0, but their standard deviations differ according to the sample size, *n.*
4. The t distribution is more spread out and flatter at the center than the standard normal distribution As the sample size increases, however, the *t* distribution approaches the standard normal distribution,

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

Comparing the z and t Distributions when *n* is small

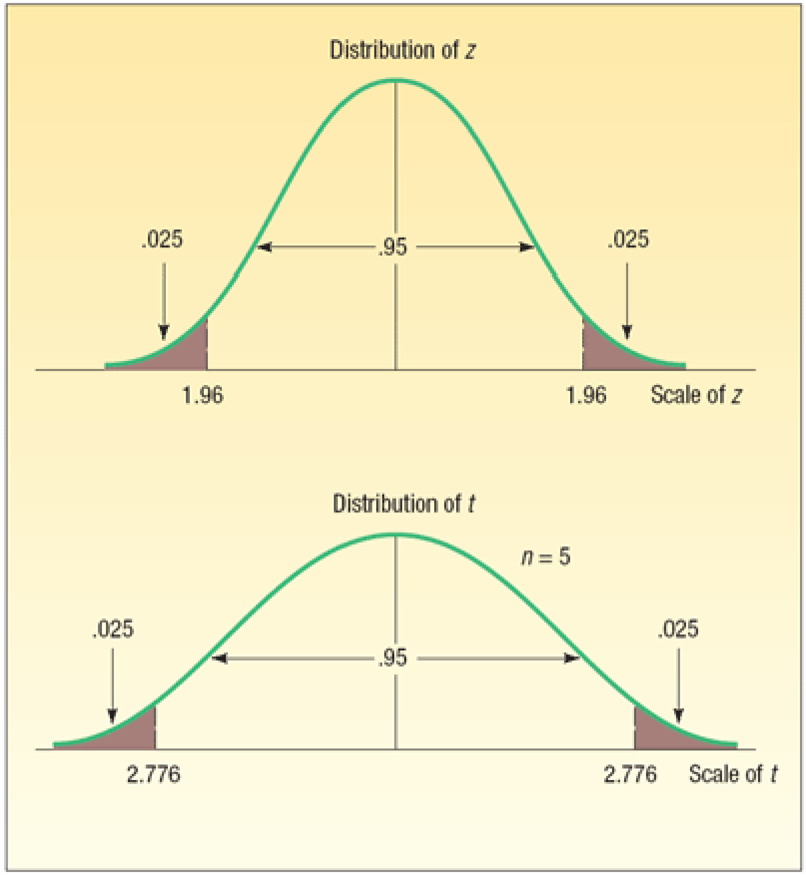


CHART 9–2 Values of z and t for the 95 Percent Level of Confidence

Confidence Interval Estimates for the Mean

<u>Use *Z*-distribution</u>

If the population standard deviation is known or the sample is greater than 30.

$$\bar{X} \pm z \frac{\sigma}{\sqrt{n}}$$

<u>Use *t*-distribution</u>

If the population standard deviation is unknown and the sample is less than 30.

$$\bar{X} \pm t \frac{s}{\sqrt{n}}$$

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

When to Use the *z* or *t* Distribution for Confidence Interval Computation

Estimation and Confidence Intervals

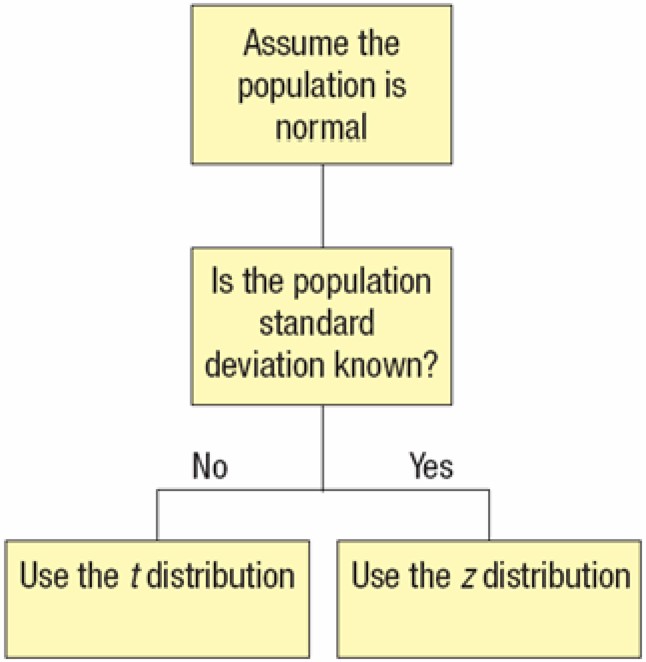


CHART 9–3 Determining When to Use the *z* Distribution or the *t* Distribution

Confidence Interval for the Mean – Example using the t-distribution

A tire manufacturer wishes to investigate the tread life of its tires. A sample of 10 tires driven 50,000 miles revealed a sample mean of 0.32 inch of tread remaining with a standard deviation of 0.09 inch. Construct a 95 percent confidence interval for the population mean. Would it be reasonable for the manufacturer to conclude that after 50,000 miles the population mean amount of tread remaining is 0.30 inches?

Given in the problem :

$$n = 10$$

$$\bar{x} = 0.32$$

$$s = 0.09$$

Compute the C.I. using the t - dist. (since σ is unknown)

$$\bar{X} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$$

Student's t-distribution Table

Compute the C.I.

using the t - dist. (since σ is unknown)

$$\bar{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

$$= \bar{X} \pm t_{.05/2,10-1}\frac{s}{\sqrt{n}}$$

$$= 0.32 \pm t_{.025,9}\frac{0.09}{\sqrt{10}}$$

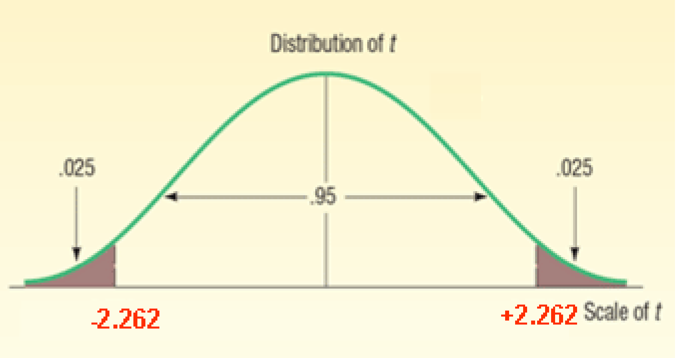
$$= 0.32 \pm 2.262\frac{0.09}{\sqrt{10}}$$

$$= 0.32 \pm 0.064$$

$$= (0.256, 0.384)$$

Conclude : the manufacturer can be reasonably sure (95% confident) that the mean remaining tread depth is between 0.256 and 0.384 inches.

	Confidence Intervals				
	80%	90%	95%	98%	99%
	Level of Significance for One-Tailed Test				
df	0.100	0.050	0.025	0.010	0.005
	Level of Significance for Two-Tailed Test				
	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	$\boxed{2.262}$	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169



Confidence Interval Estimates for the Mean – Using Minitab

The manager of the Inlet Square Mall, near Ft. Myers, Florida, wants to estimate the mean amount spent per shopping visit by customers. A sample of 20 customers reveals the following amounts spent.

$48.16	$42.22	$46.82	$51.45	$23.78	$41.86	$54.86
37.92	52.64	48.59	50.82	46.94	61.83	61.69
49.17	61.46	51.35	52.68	58.84	43.88	

Confidence Interval Estimates for the Mean – By Formula

Compute the C.I.

using the t - dist. (since σ is unknown)

$$\bar{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

$$= \bar{X} \pm t_{.05/2,20-1}\frac{s}{\sqrt{n}}$$

$$= 49.35 \pm t_{.025,19}\frac{9.01}{\sqrt{20}}$$

$$= 49.35 \pm 2.093\frac{9.01}{\sqrt{20}}$$

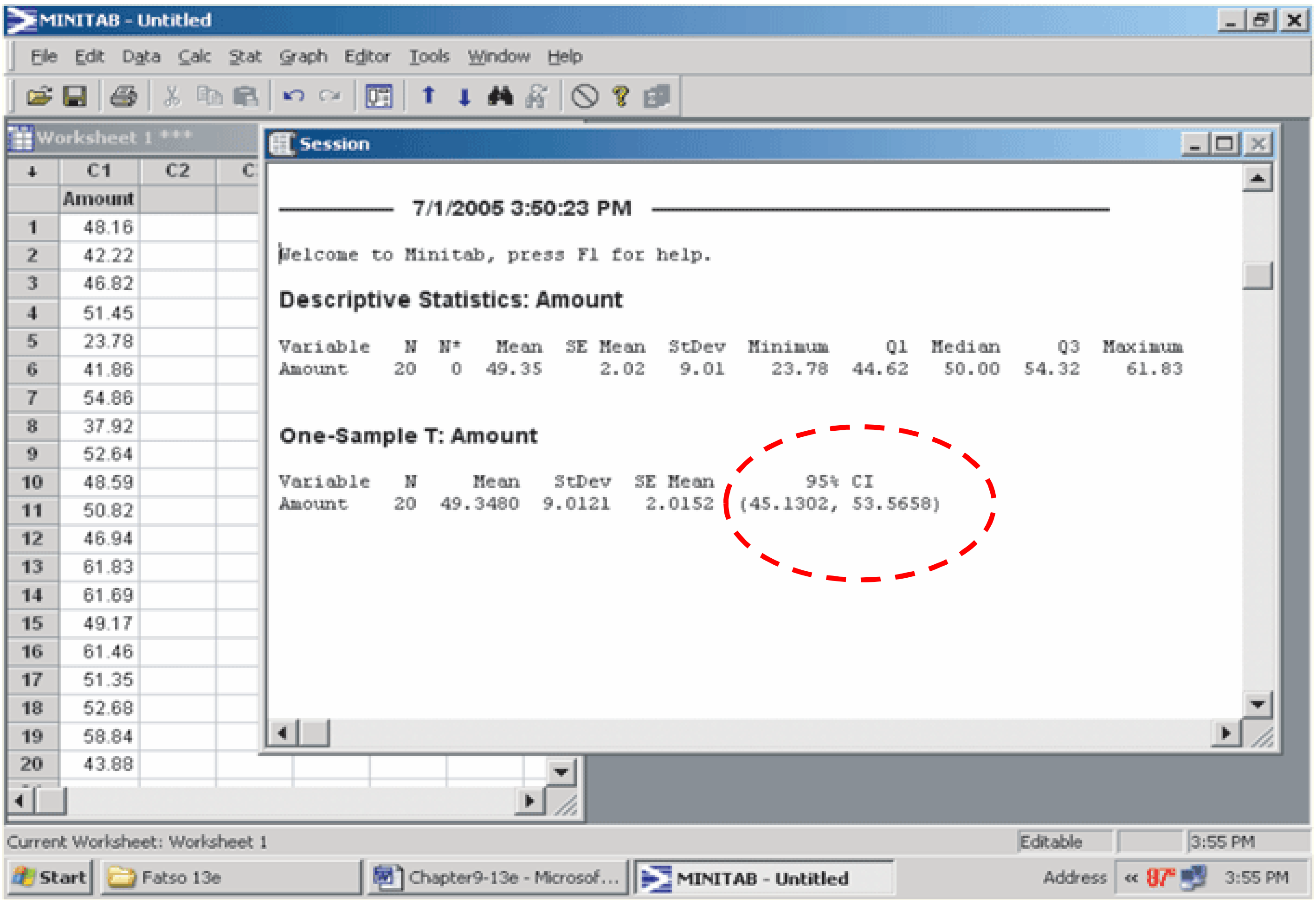
$$= 49.35 \pm 4.22$$

The endpoints of the confidence interval are \$45.13 and \$53.57

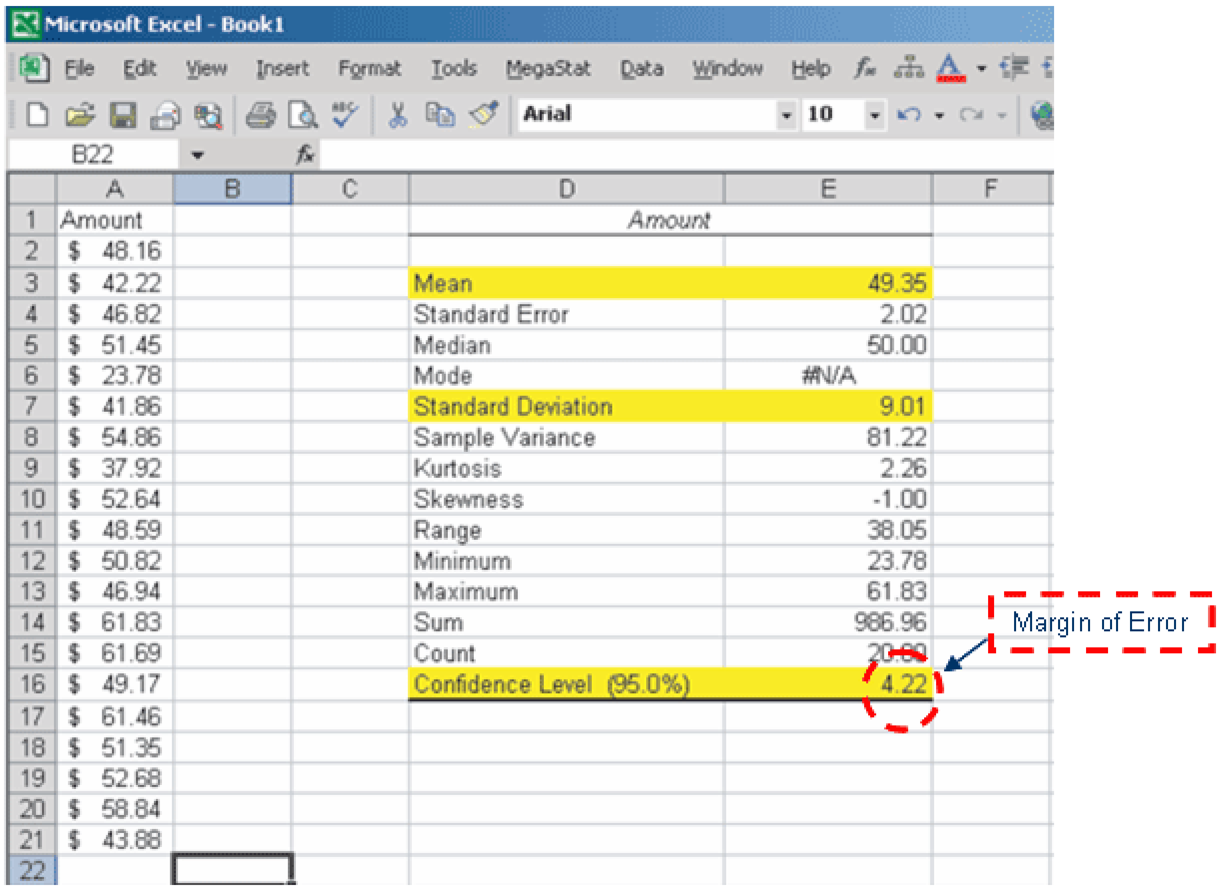
Conclude : It is reasonable that the population mean could be \$50.

The value of \$60 is not in the confidence interval. Hence, we conclude that the population mean is unlikely to be \$60.

Confidence Interval Estimates for the Mean – Using Minitab



Confidence Interval Estimates for the Mean – Using Excel



	A	B	C	D	E	F
1	Amount			Amount		
2	$ 48.16					
3	$ 42.22			Mean	49.35	
4	$ 46.82			Standard Error	2.02	
5	$ 51.45			Median	50.00	
6	$ 23.78			Mode	#N/A	
7	$ 41.86			Standard Deviation	9.01	
8	$ 54.86			Sample Variance	81.22	
9	$ 37.92			Kurtosis	2.26	
10	$ 52.64			Skewness	-1.00	
11	$ 48.59			Range	38.05	
12	$ 50.82			Minimum	23.78	
13	$ 46.94			Maximum	61.83	
14	$ 61.83			Sum	986.96	
15	$ 61.69			Count	20.00	
16	$ 49.17			Confidence Level (95.0%)	4.22	
17	$ 61.46					
18	$ 51.35					
19	$ 52.68					
20	$ 58.84					
21	$ 43.88					
22						

Using the Normal Distribution to Approximate the Binomial Distribution

To develop a confidence interval for a proportion, we need to meet the following assumptions.

1. The binomial conditions, discussed in Chapter 6, have been met. Briefly, these conditions are:
   - a. The sample data is the result of counts.
   - b. There are only two possible outcomes.
   - c. The probability of a success remains the same from one trial to the next.
   - d. The trials are independent. This means the outcome on one trial does not affect the outcome on another.
2. The values $n\pi$ and $n(1-\pi)$ should both be greater than or equal to 5. This condition allows us to invoke the central limit theorem and employ the standard normal distribution, that is, z, to complete a confidence interval.

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

Confidence Interval for a Population Proportion

The confidence interval for a population proportion is estimated by:

$$p \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

Confidence Interval for a Population Proportion- Example

The union representing the Bottle Blowers of America (BBA) is considering a proposal to merge with the Teamsters Union. According to BBA union bylaws, at least three-fourths of the union membership must approve any merger. A random sample of 2,000 current BBA members reveals 1,600 plan to vote for the merger proposal. What is the estimate of the population proportion?

Develop a 95 percent confidence interval for the population proportion. Basing your decision on this sample information, can you conclude that the necessary proportion of BBA members favor the merger? Why?

First, compute the sample proportion:

$$p = \frac{x}{n} = \frac{1,600}{2000} = 0.80$$

Compute the 95% C.I.

$$\text{C.I.} = p \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

$$= 0.80 \pm 1.96\sqrt{\frac{.80(1-.80)}{2,000}} = .80 \pm .018$$

$$= (0.782, 0.818)$$

Conclude: The merger proposal will likely pass because the interval estimate includes values greater than 75 percent of the union membership.

Finite-Population Correction Factor

- A population that has a fixed upper bound is said to be finite.
- For a finite population, where the total number of objects is *N* and the size of the sample is *n*, the following adjustment is made to the standard errors of the sample means and the proportion:
- However, if $n/N < .05$, the finite-population correction factor may be ignored.

Standard Error of the Sample Mean

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Standard Error of the Sample Proportion

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

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

Effects on FPC when n/N Changes

Finite-Population Correction Factor for Selected Samples When the Population Is 1,000

Sample Size	Fraction of Population	Correction Factor
10	.010	.9955
25	.025	.9879
50	.050	.9752
100	.100	.9492
200	.200	.8949
500	.500	.7075

Observe that FPC approaches 1 when *n/N* becomes smaller

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

Confidence Interval Formulas for Estimating Means and Proportions with Finite Population Correction

C.I. for the Mean (μ)

$$\bar{X} \pm z \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

C.I. for the Mean (μ)

$$\bar{X} \pm t \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

C.I. for the Proportion (π)

$$p \pm z \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}}$$

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

CI For Mean with FPC - Example

There are 250 families in Scandia, Pennsylvania. A random sample of 40 of these families revealed the mean annual church contribution was $450 and the standard deviation of this was $75.

Develop a 90 percent confidence interval for the population mean.

Interpret the confidence interval.

Given in Problem:

N – 250

n – 40

s - $75

Since n/N = 40/250 = 0.16, the finite population correction factor must be used.

The population standard deviation is not known therefore use the t-distribution (may use the z-dist since n>30)

Use the formula below to compute the confidence interval:

$$\bar{X} \pm t \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

CI For Mean with FPC - Example

$$\bar{X} \pm t \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$= \$450 \pm t_{.10,40-1} \frac{\$75}{\sqrt{40}} \sqrt{\frac{250-40}{250-1}}$$

$$= \$450 \pm 1.685 \frac{\$75}{\sqrt{40}} \sqrt{\frac{250-40}{250-1}}$$

$$= \$450 \pm \$19.98 \sqrt{.8434}$$

$$= \$450 \pm \$18.35$$

$$= (\$431.65, \$468.35)$$

It is likely that the population mean is more than \$431.65 but less than \$468.35. To put it another way, could the population mean be \$445? Yes, but it is not likely that it is \$425 because the value \$445 is within the confidence interval and \$425 is not within the confidence interval.

Selecting a Sample Size

There are 3 factors that determine the size of a sample, none of which has any direct relationship to the size of the population. They are:

- The degree of confidence selected.
- The maximum allowable error.
- The variation in the population.

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

Sample Size Determination for a Variable

- To find the sample size for a variable:

$$n = \left(\frac{z \cdot s}{E} \right)^2$$

where:

E - the allowable error

z - the z - value corresponding to the selected level of confidence

s - the sample deviation (from pilot sample)

Sample Size Determination for a Variable-Example

A student in public administration wants to determine the mean amount members of city councils in large cities earn per month as remuneration for being a council member. The error in estimating the mean is to be less than \$100 with a 95 percent level of confidence. The student found a report by the Department of Labor that estimated the standard deviation to be \$1,000. What is the required sample size?

Given in the problem:

- *E*, the maximum allowable error, is \$100
- The value of *z* for a 95 percent level of confidence is 1.96,
- The estimate of the standard deviation is \$1,000.

$$n = \left(\frac{z \cdot s}{E}\right)^2$$

$$= \left(\frac{1.96 \cdot \$1{,}000}{\$100}\right)^2$$

$$= (19.6)^2$$

$$= 384.16$$

$$= 385$$

Sample Size Determination for a Variable- Another Example

A consumer group would like to estimate the mean monthly electricity charge for a single family house in July within $5 using a 99 percent level of confidence. Based on similar studies the standard deviation is estimated to be $20.00. How large a sample is required?

$$n = \left(\frac{(2.58)(20)}{5} \right)^2 = 107$$

Sample Size for Proportions

- The formula for determining the sample size in the case of a proportion is:

$$n = p(1-p)\left(\frac{Z}{E}\right)^2$$

where :

p is estimate from a pilot study or some source,
otherwise, 0.50 is used

z - the z - value for the desired confidence level

E - the maximum allowable error

Another Example

The American Kennel Club wanted to estimate the proportion of children that have a dog as a pet. If the club wanted the estimate to be within 3% of the population proportion, how many children would they need to contact? Assume a 95% level of confidence and that the club estimated that 30% of the children have a dog as a pet.

$$n = (.30)(.70)\left(\frac{1.96}{.03}\right)^2 = 897$$

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

Another Example

A study needs to estimate the proportion of cities that have private refuse collectors. The investigator wants the margin of error to be within .10 of the population proportion, the desired level of confidence is 90 percent, and no estimate is available for the population proportion. What is the required sample size?

$$n = (.5)(1-.5)\left(\frac{1.65}{.10}\right)^2 = 68.0625$$

$$n = 69 \text{ cities}$$

End of Chapter 9