Dr. Saeed Alghamdi
Room 542
Statistics Department
Faculty of Sciences
King Abdulaziz University
web sit
http://saalghamdy.kau.edu.sa

Elementary Statistics

A Step by Step Approach
Eighth Edition

by
Allan G. Bluman

SLIDES PREPARED
BY

Dr. Saeed A Dobbah Alghamdi
Statistics Department
King Abdulaziz University

CHAPTER 7

Confidence Intervals
and Sample Size

Notes

- ……
- ……
- ……
- ……
- ……
- ……

Objectives

- Find the confidence interval for the mean when σ is known or $n \geq 30$.
- Determine the minimum sample size for finding a confidence interval for the mean.
- Find the confidence interval for the mean when σ is unknown and $n < 30$.
- Find the confidence interval for a proportion.
- Determine the minimum sample size for finding a confidence interval for a proportion.

Notes

- ……
- ……
- ……
- ……
- ……
- ……

Introduction

- *Estimation* is the process of estimating the value of a parameter from information obtained from a sample.

Notes

- ❑……
- ❑……
- ❑……
- ❑……
- ❑……
- ❑……

Three Properties of a Good Estimator

- 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.
- 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.
- 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.

Notes

- ❑……
- ❑……
- ❑……
- ❑……
- ❑……
- ❑……

Point and Interval Estimates

- A *point estimate* is a specific numerical value of a parameter. The best point estimate of the population mean μ is the sample mean $\bar{X}$.
- 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

Notes

- ❑……
- ❑……
- ❑……
- ❑……
- ❑……
- ❑……

Confidence Level and Confidence Interval

- The *confidence level* of an interval estimate of a parameter is the probability that the interval estimate will contain the parameter.
- 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.

Notes

❑……

❑……

❑……

❑……

❑……

❑……

Confidence Interval for the mean when σ is known

The confidence interval for the mean can be found using the standard normal distribution z if the following requirements are satisfied

1. *The sample is simple random sample.*
2. *The value of the population standard deviation σ is known.*
3. *The population is normally distributed or the sample size n is 30 or more.*

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

Notes

❑……

❑……

❑……

❑……

❑……

❑……

Confidence Interval for the mean when σ is known

Characteristics of the *z* Distribution

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

Notes

❑……

❑……

❑……

❑……

❑……

❑……

Confidence Interval for the mean when σ is known

EXAMPLE

The following data (BODYTEMP) are the body temperature of a sample of 106 people selected randomly. Assuming the population standard deviation of the body temperature is 3, find the 95% confidence interval for the body temperature average.

Since $\alpha = 1 - 0.95 = 0.05$, $Z_{\alpha/2} = 1.96$, $\bar{x} = 98.2$, $\sigma = 3$ and $n = 106$, substituting in the formula, we get

$$98.2 - 1.96 * 3/\sqrt{106} < \mu < 98.2 + 1.96 * 3/\sqrt{106}$$

$$97.63 < \mu < 98.77$$

Thus, we can be 95% confident that the population mean of body temperature is between 97.63 and 98.77 degree, based on a sample of 106 people.

Notes

❑……

❑……

❑……

❑……

❑……

❑……

Confidence Interval for the mean when σ is known

EXCEL

1. Enter the data into Excel worksheet
2. From the toolbar, select Add-Ins, **MegaStat>Descriptive Statistics…**
3. Enter the cell range in the **Input range**
4. Check both **Mean** and **Sample variance and standard deviation**. Click [OK]
5. From the toolbar, select Add-Ins, **MegaStat>Confidence Intervals/Sample Size**
6. Enter the **Mean, Std. Dev.** and **n** of the data.
7. Type in or scroll the **Confidence Level**. Click [OK]

Notes

❑……

❑……

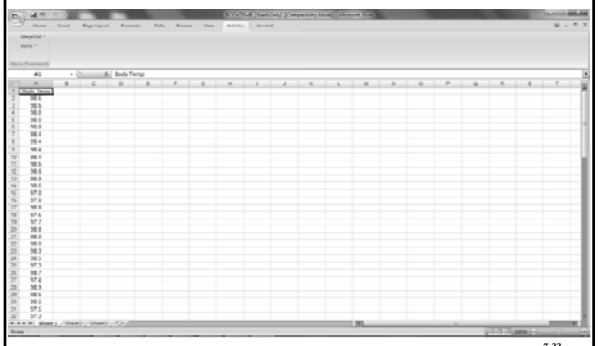
❑……

❑……

❑……

❑……

Confidence Interval for the mean when σ is known



Notes

❑……

❑……

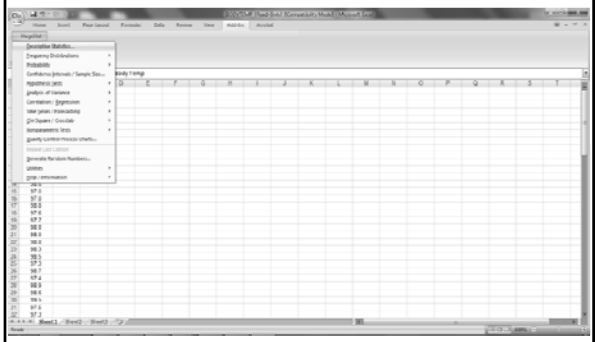
❑……

❑……

❑……

❑……

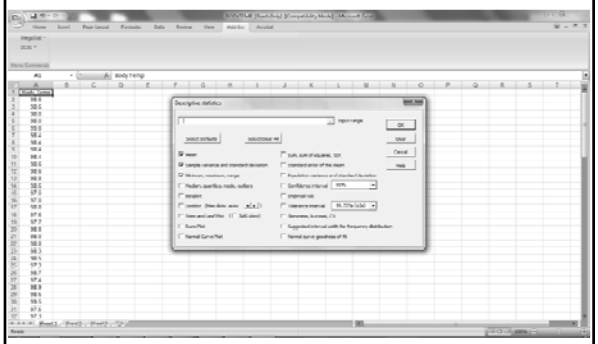
Confidence Interval for the mean when σ is known



Notes

- .....
- .....
- .....
- .....
- .....
- .....

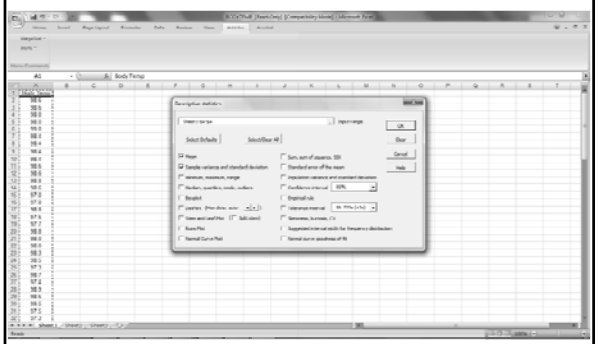
Confidence Interval for the mean when σ is known



Notes

- .....
- .....
- .....
- .....
- .....
- .....

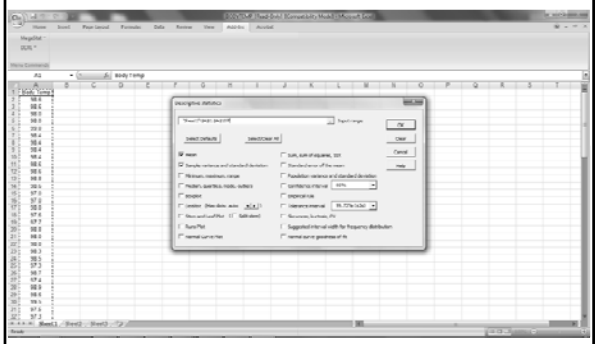
Confidence Interval for the mean when σ is known



Notes

- .....
- .....
- .....
- .....
- .....
- .....

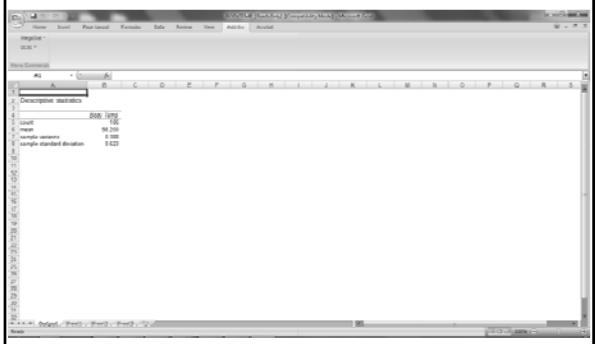
Confidence Interval for the mean when σ is known



Notes

- .....
- .....
- .....
- .....
- .....
- .....

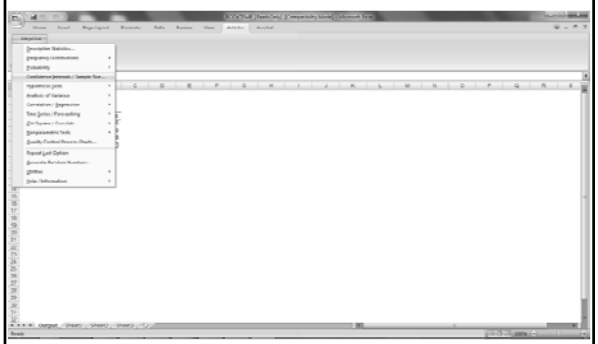
Confidence Interval for the mean when σ is known



Notes

- .....
- .....
- .....
- .....
- .....
- .....

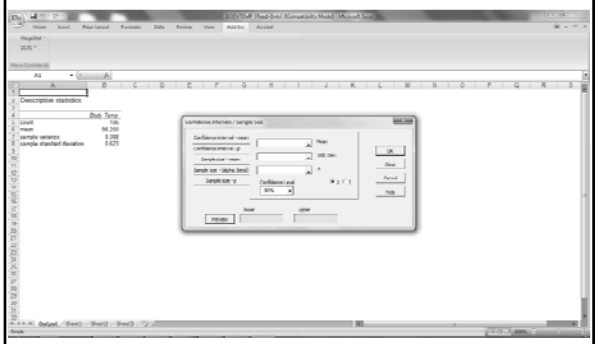
Confidence Interval for the mean when σ is known



Notes

- .....
- .....
- .....
- .....
- .....
- .....

Confidence Interval for the mean when σ is known



Notes

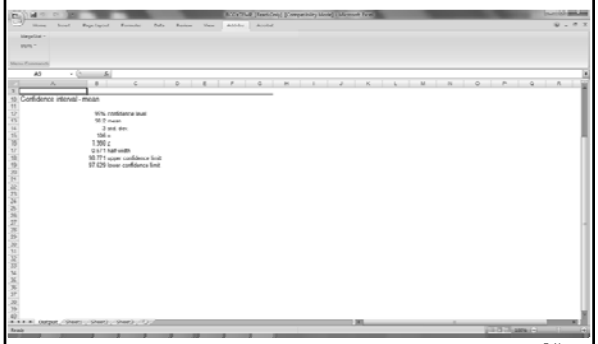
- .....
- .....
- .....
- .....
- .....
- .....

Confidence Interval for the mean when σ is known

Notes

- .....
- .....
- .....
- .....
- .....
- .....

Confidence Interval for the mean when σ is known



Notes

- .....
- .....
- .....
- .....
- .....
- .....

Confidence Interval for the mean when σ is known

Maximum Error of Estimate

The *maximum error of estimate* is the maximum likely difference between the point estimate of a parameter and the actual value of the parameter. It is represented by the term

$$E = z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Notes

❑.....

❑.....

❑.....

❑.....

❑.....

❑.....

Confidence Interval for the mean when σ is known

The Minimum Sample Size Needed for an Interval Estimate of the Population Mean is given by

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

where *E* is the maximum error of estimate. If there is any fraction or decimal portion in the answer, use the next whole number for sample size, *n*.

Notes

❑.....

❑.....

❑.....

❑.....

❑.....

❑.....

Confidence Interval for the mean when σ is known

EXAMPLE

A college president asks a statistics teacher to estimate the average age of the students at their college. How large a sample is necessary?

The statistics teacher would like to be 99% confident that the estimate should be accurate within 1 year. From a previous study, the standard deviation of the age is known to be 3 years.

Since $\alpha = 1 - 0.99 = 0.01$, $Z_{\alpha/2} = 2.58$, $\sigma = 3$ and $E = 1$, substituting in the formula, we get $n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{E}\right)^2 = \left(\frac{2.58(3)}{1}\right)^2 = 59.9 \approx 60$

Thus, to be 99% confident that the estimate is within 1 year of the true mean age, the teacher needs a sample size of at least 60 students.

Notes

❑.....

❑.....

❑.....

❑.....

❑.....

❑.....

Confidence Interval for the mean when σ is known

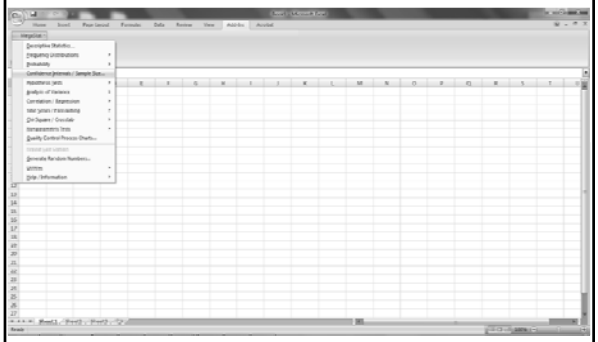
EXCEL

1. From the toolbar, select Add-Ins, **MegaStat>Confidence Intervals/Sample Size**
2. Select **Sample size - mean**
3. Enter **E** and **Std. Dev.**
4. Type in or scroll the **Confidence Level**. Click [OK]

Notes

- ❑……
- ❑……
- ❑……
- ❑……
- ❑……
- ❑……

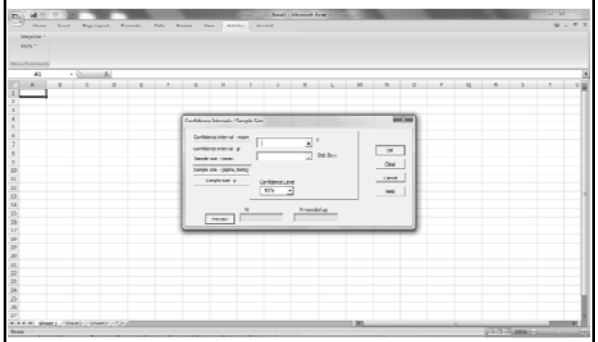
Confidence Interval for the mean when σ is known



Notes

- ❑……
- ❑……
- ❑……
- ❑……
- ❑……
- ❑……

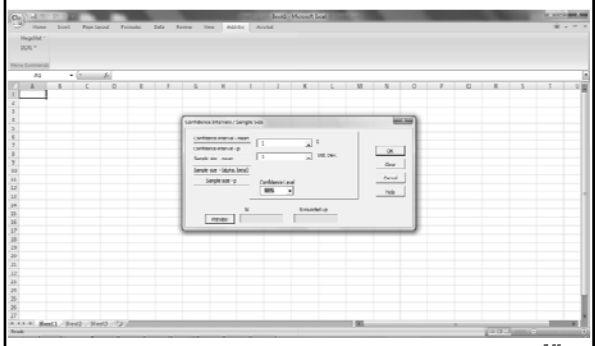
Confidence Interval for the mean when σ is known



Notes

- ❑……
- ❑……
- ❑……
- ❑……
- ❑……
- ❑……

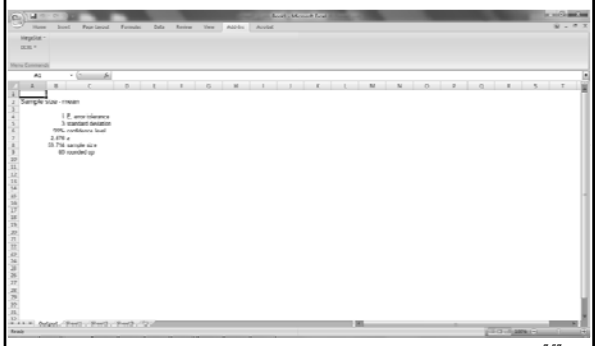
Confidence Interval for the mean when σ is known



Notes

- .....
- .....
- .....
- .....
- .....
- .....

Confidence Interval for the mean when σ is known



Notes

- .....
- .....
- .....
- .....
- .....
- .....

Confidence Interval for the mean when σ is unknown

The confidence interval for the mean can be found using the student t distribution with $n-1$ degrees of freedom if the following requirements are satisfied

1. *The sample is simple random sample.*
2. *The sample is from normally distributed population or the sample size n is 30 or more.*

$$\bar{X} - t_{(\frac{\alpha}{2},n-1)}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{X} + t_{(\frac{\alpha}{2},n-1)}\left(\frac{s}{\sqrt{n}}\right)$$

Notes

- .....
- .....
- .....
- .....
- .....
- .....

Confidence Interval for the mean when σ is unknown

Characteristics of the *t* Distribution

The *t* distribution is similar to the *z* distribution in the following

1. It is bell shaped.
2. It is symmetrical about the mean.
3. The mean, median, and mode are equal to 0.
4. The curve never touches the *x* axis.

The *t* distribution differs from the *z* distribution in the following

1. The variance of the *t* distribution is greater than 1.
2. The *t* distribution is 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 *z* distribution.

Notes

❑.....

❑.....

❑.....

❑.....

❑.....

❑.....

Confidence Interval for the mean when σ is unknown

EXAMPLE

The following data (BODYTEMP) are the body temperature of a sample of 28 people selected randomly. Assuming the population standard deviation of the body temperature is unknown, find the 95% confidence interval for the body temperature average.

Since $\alpha = 1 - 0.95 = 0.05, t_{(\alpha/2, n-1)} = 2.052$, $\bar{x} = 98.25, \sigma = 0.568$ and $n = 28$ substituting in the formula, we get

$$98.25 - 2.052 * 0.568 / \sqrt{28} < \mu < 98.25 + 2.052 * 0.568 / \sqrt{28}$$

$$98.03 < \mu < 98.47$$

Thus, we can be 95% confident that the population mean of body temperature is between 98.03 and 98.47 degree, based on a sample of 28 people.

Notes

❑.....

❑.....

❑.....

❑.....

❑.....

❑.....

Confidence Interval for the mean when σ is unknown

EXCEL

1. Enter the data into Excel worksheet
2. From the toolbar, select Add-Ins, **MegaStat>Descriptive Statistics…**
3. Enter the cell range in the **Input range**
4. Check both **Mean** and **Sample variance and standard deviation**. Click [OK]
5. From the toolbar, select Add-Ins, **MegaStat>Confidence Intervals/Sample Size**
6. Enter the **Mean, Std. Dev.** and **n** of the data.
7. Type in or scroll the **Confidence Level**.
8. Select *t* test. Click [OK]

Notes

❑.....

❑.....

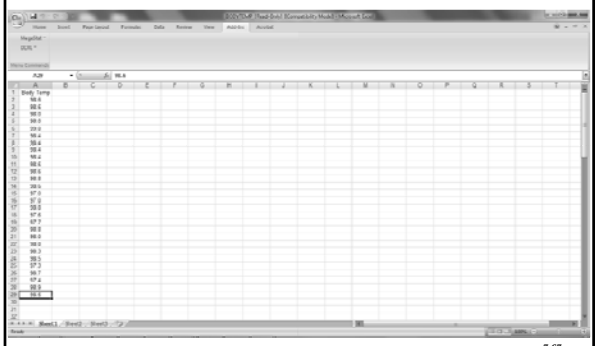
❑.....

❑.....

❑.....

❑.....

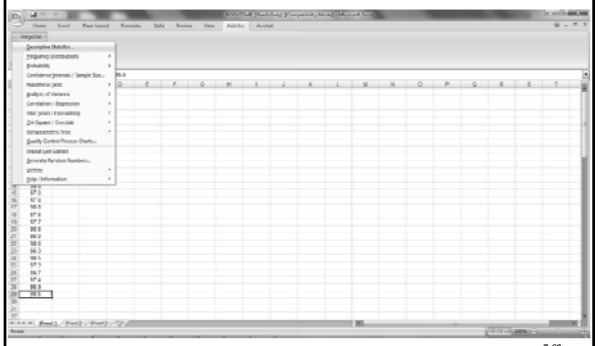
Confidence Interval for the mean when σ is unknown



Notes

- .....
- .....
- .....
- .....
- .....
- .....

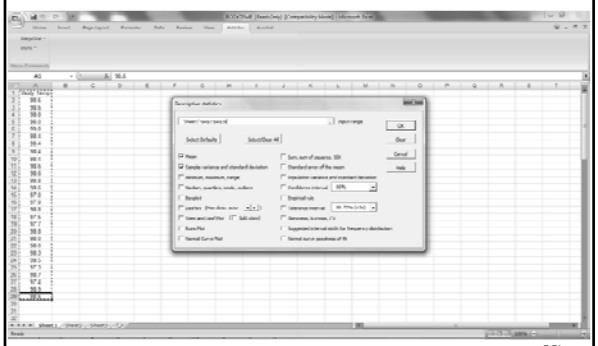
Confidence Interval for the mean when σ is unknown



Notes

- .....
- .....
- .....
- .....
- .....
- .....

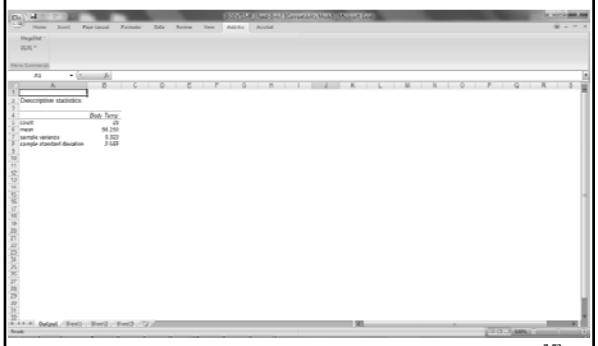
Confidence Interval for the mean when σ is unknown



Notes

- .....
- .....
- .....
- .....
- .....
- .....

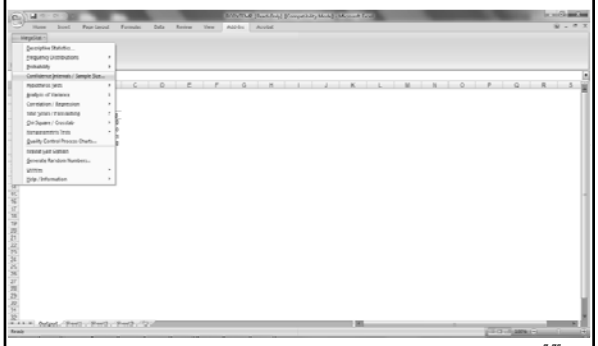
Confidence Interval for the mean when σ is unknown



Notes

- .....
- .....
- .....
- .....
- .....
- .....

Confidence Interval for the mean when σ is unknown



Notes

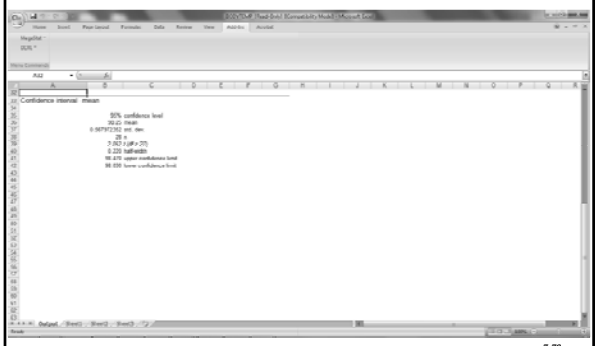
- .....
- .....
- .....
- .....
- .....
- .....

Confidence Interval for the mean when σ is unknown

Notes

- .....
- .....
- .....
- .....
- .....
- .....

Confidence Interval for the mean when σ is unknown



Notes

❑……

❑……

❑……

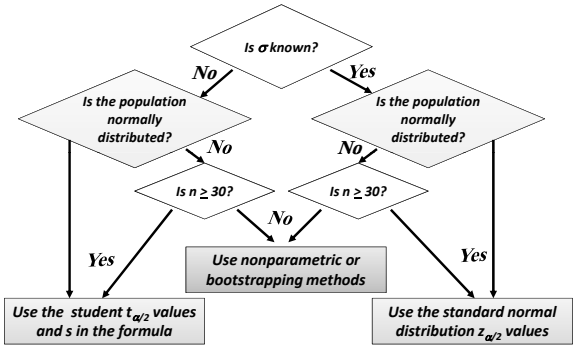
❑……

❑……

❑……

Confidence Interval for the mean

When to Use the z or t Distribution



Notes

❑……

❑……

❑……

❑……

❑……

❑……

Confidence Interval for the Proportion

Symbols Used in Proportion Notation

- p = symbol for the population proportion
- $\hat{p}$ (read p "hat") = symbol for the sample proportion
- For a sample proportion,

$$\hat{p} = \frac{X}{n} \text{ and } \hat{q} = \frac{n - X}{n} \text{ or } 1 - \hat{p}$$

where

X = number of sample units that possess the characteristics of interest

n = sample size.

Notes

❑……

❑……

❑……

❑……

❑……

❑……

Confidence Interval for the Proportion

The confidence interval for the proportion p can be found using the standard normal distribution z if the following requirements are satisfied

1. *The sample is simple random sample.*
2. *The conditions for the binomial distribution are satisfied (there are fixed number of trials, the trials are independent, there are two outcomes any trial and the probabilities remain constant for each trial)*
3. *There are at least 5 successes and 5 failures.*

$$\hat{p} - z_{a/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{a/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Notes

❑……

❑……

❑……

❑……

❑……

❑……

Confidence Interval for the Proportion

EXAMPLE

There were 500 nursing applications in a sample, including 60 from men. Find the 90% confidence interval for the true proportion of male applicants.

Since $\alpha = 1 - 0.90 = 0.10$, $z_{\alpha/2} = 1.65$, $\hat{p} = 60/500 = 0.12$, $\hat{q} = 1 - 0.12 = 0.88$ and $n = 500$, we get

$$0.12 - 1.65*\sqrt{\frac{0.12\times 0.88}{500}} < p < 0.12 + 1.65*\sqrt{\frac{0.12\times 0.88}{500}}$$

$$0.096 < p < 0.144$$

Thus, we can be 90% confident that the population proportion of male applicants is between 9.6% and 14.4%, based on a sample of 500 applicants.

Notes

❑……

❑……

❑……

❑……

❑……

❑……

Confidence Interval for the Proportion

EXCEL

1. From the toolbar, select Add-Ins, **MegaStat>Confidence Intervals/Sample Size**
2. In the dialog box, select the Confidence interval-p.
3. Enter 60 in the first box; *p* will automatically switch to *x*.
4. Enter 500 in the second box for *n*.
5. Type in or scroll the **Confidence Level**. Click [OK]

Notes

❑……

❑……

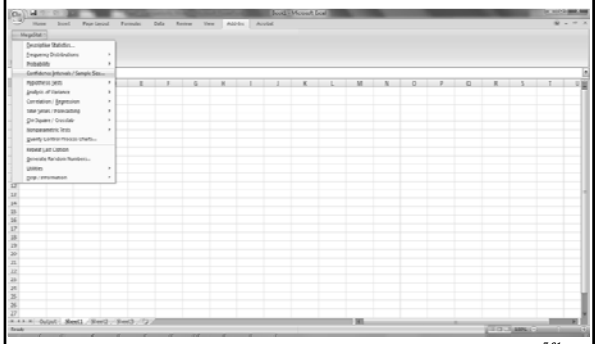
❑……

❑……

❑……

❑……

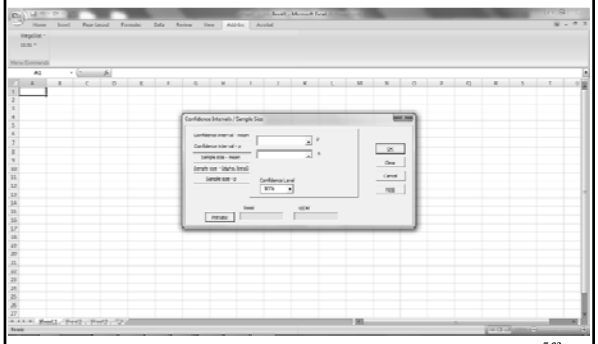
Confidence Interval for the Proportion



Notes

- ……
- ……
- ……
- ……
- ……
- ……

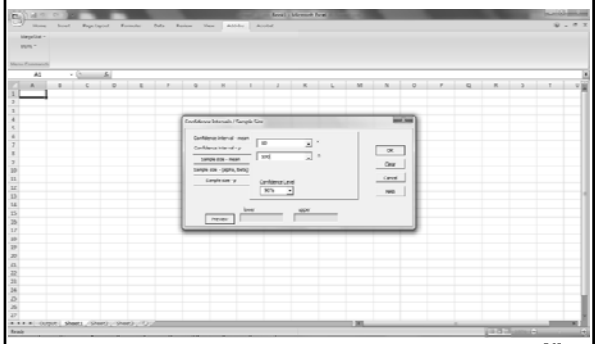
Confidence Interval for the Proportion



Notes

- ……
- ……
- ……
- ……
- ……
- ……

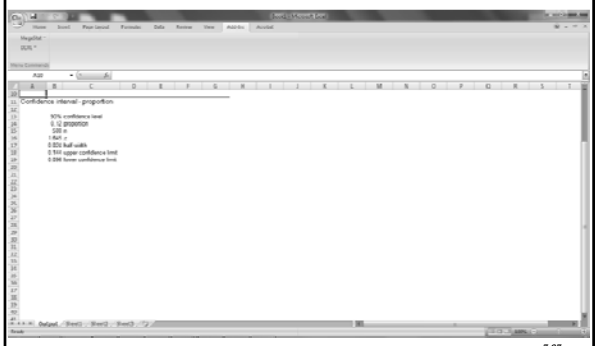
Confidence Interval for the Proportion



Notes

- ……
- ……
- ……
- ……
- ……
- ……

Confidence Interval for the Proportion



Notes

- .....
- .....
- .....
- .....
- .....
- .....

Confidence Interval for the Proportion

The minimum sample size needed for interval estimate of a population proportion is given by

$$n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2$$

where the maximum error of estimate is

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

- If necessary, round up to obtain a whole number.
- When no estimate $\hat{p}$ is known use $\hat{p} = 0.5$.

Notes

- .....
- .....
- .....
- .....
- .....
- .....

Confidence Interval for the Proportion

EXAMPLE

A researcher wishes to estimate, with 95% confidence, the proportion of people who own a home computer. A previous study shows that 40% of those interviewed had a computer at home. The researcher wishes to be accurate within 2% of the true proportion. Find the minimum sample size necessary.

Since $\alpha = 1 - 0.95 = 0.05$, $z_{\alpha/2} = 1.96$, $E = 0.02$, $\hat{p} = 0.4$ and $\hat{q} = 0.6$ then

$$n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2 = 0.4(0.6)\left(\frac{1.96}{0.02}\right)^2 = 2304.96 \approx 2305$$

Notes

- .....
- .....
- .....
- .....
- .....
- .....

Confidence Interval for the Proportion

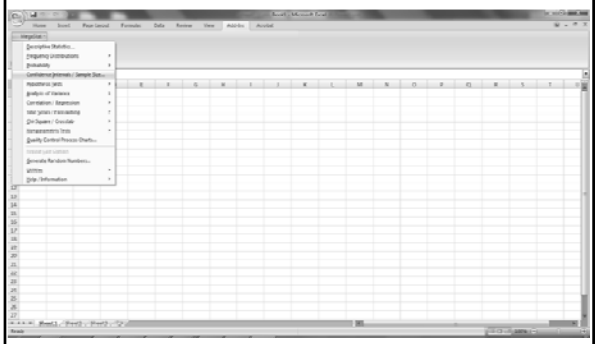
EXCEL

1. From the toolbar, select Add-Ins, **MegaStat>Confidence Intervals/Sample Size**
2. In the dialog box, select the Sample size-p.
3. Enter 0.02 in the first box for *E*.
4. Enter 0.40 in the second box for *p*.
5. Type in or scroll the **Confidence Level**. Click [OK]

Notes

- ……
- ……
- ……
- ……
- ……
- ……

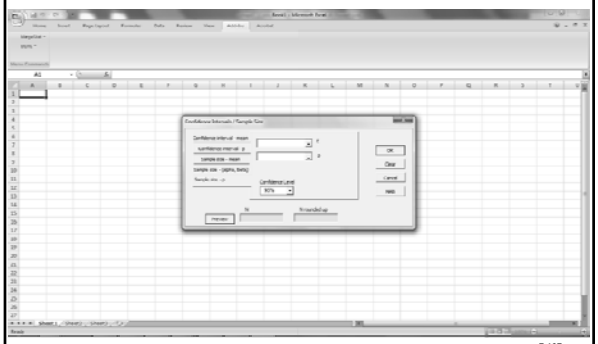
Confidence Interval for the Proportion



Notes

- ……
- ……
- ……
- ……
- ……
- ……

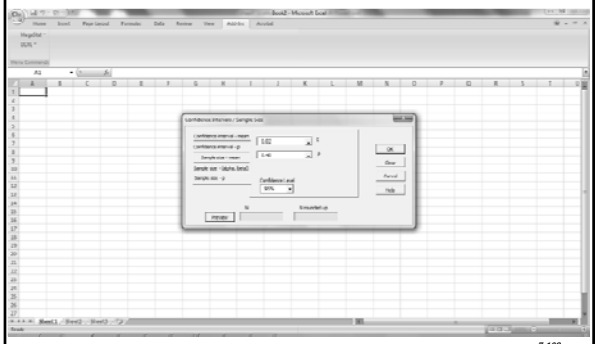
Confidence Interval for the Proportion



Notes

- ……
- ……
- ……
- ……
- ……
- ……

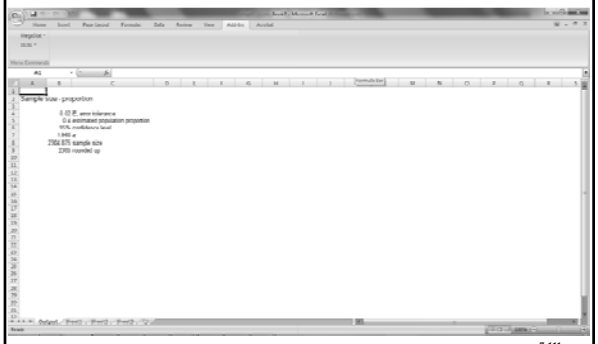
Confidence Interval for the Proportion



Notes

- ……
- ……
- ……
- ……
- ……
- ……

Confidence Interval for the Proportion



Notes

- ……
- ……
- ……
- ……
- ……
- ……

Summary

- A good estimator must be unbiased, consistent, and relatively efficient.
- There are two types of estimates of a parameter: point estimates and interval estimates.
- A *point estimate* is a single value. The problem with point estimates is that the accuracy of the estimate cannot be determined, so the *interval estimate* is preferred.
- By calculating a 95% or 99% confidence interval about the sample value, statisticians can be 95% or 99% confident that their estimate contains the true parameter.

Notes

- ……
- ……
- ……
- ……
- ……
- ……