**CHAPTER TWO**

**STATISTICAL ESTIMATIONS**

### Objectives:

After completing this chapter, you should be able to:

* Define estimation and related concepts.
* Compute point and confidence interval estimates for unknown population parameters.
* Determine sample size to estimate population mean and proportion.

1. **Introduction**

In chapter one, we have seen that the sampling distribution of the mean shows how far sample means could be from a known population mean similarly we have seen that the sampling distribution of the proportion shows how far sample proportions could be from a known population proportion.

But our objective in this chapter is to determine how far an unknown population mean could be from the mean of a simple random sample selected from the population, or how far an unknown population proportion could be from a sample proportion. That is, it tries to determine how large the error could be if a sample statistic is used to estimate a population parameter. The chapter also deals with how to construct the right sample size that will serve to estimate the unknown population parameters.

## **Definitions**

**Estimation: -** Is the process of predicting or estimating the unknown population parameter through sampling. That is, it is the process of using sample statistic so as to estimate the unknown population parameter.

**Estimator:** - is a sample statistic that is used to estimate an unknown population parameter.

**E.**g.  etc.

**Estimate:** - is a single numerical value obtained for an estimator.

**E.**g. 1, 2, 3, …

For example, the sample mean  is an estimator for the population mean.

**A parameter:** - is a characteristic of a population.

**E.**g. etc.

**A statistic:** - is a characteristic of a sample.

**E.**g.  etc

* 1. **Point estimates of the mean and proportion**

***A point estimate*** is when a statistic taken from the sample is used to estimate a population parameter. A point estimate is a single value of an estimator. Thus, the sample average  is a point estimate of the population mean.

**Example 1:**

Suppose we have the following random sample of n = 6 elements from a population whose parameter values are not known.

1. 2 4 5 7 11

The sample mean  is



Therefore the point estimate of the population mean  is 5. i.e. 

Estimates

The sample proportion of success for calling an even number from the sample is



The statistic is an estimator of the unknown population proportion. Thus, the point estimate of the population proportion is 0.33 i.e. p = 0.33

Estimates

The sample standard deviation is an estimator of the unknown population standard deviation .

, Where  = sample mean

estimates

The sample standard error of the mean  is an estimator of the unknown standard error of the mean . When the sample size is small  and  are computed as:

 and 

Estimates

**Example 2**

The results of a sample taken from a population are:

n = 9  

Find the point estimates for the following

1. The mean.
2. The standard deviation 
3. The variance 
4. The standard deviation (error) of the mean 

**Solution:**

1. The sample mean  is an estimator of . Therefore we have to calculate  to get the point estimate of .



Thus the point estimate for the population mean  is 4. 

1. The sample standard deviation  is an estimator of . Therefore we have to calculate 



Therefore, the point estimate of is 6. 

1. The sample variance  is an estimator of the population variance . Because we have obtained that  in (b), the sample variance  is



Thus, the point estimate for the population variance  is 36.



1. The sample standard error of the mean is an estimator of the standard error of the mean . So we have to calculate.



Therefore, the point estimate for 

***Check point:***

Suppose we have the following random sample of n=6 elements from a population whose parameter values are not known;

4 10 11 13 16 and 18

Compute the following point estimates;

1. The estimate of the population mean
2. The estimate of the population standard deviation
3. The estimate of the standard error of the mean.
4. The estimate of the population proportion of even numbers.
   1. **Interval Estimations of The Mean and Proportion**

### Definition

**Interval estimate**: - specifies the range of values with in which the unknown population parameter is expected to lie. For example, in the previous example of sample:

1 2 4 5 7 11

The sample mean  was 5. Therefore, 5 is our point estimate of  . On the other hand, if we state that, the range of values for  will be 5-1 = 4 and 5 + 1 = 6. Thus the interval estimate of  is between 4 and 6



The interval estimate incorporates:

1. Measure of variability i.e. standard error of the point estimate

2. Confident coefficient/level- measures how confident we are that the

Interval is correct.

For example, a confidence interval estimate of  is an interval estimate; together with a statement of how confident we are that the interval is correct.

**2.3.1 Interval estimations of the population mean **

From statistics for management I, we have



Rearranging the formula we have





Therefore, the formula for calculating an interval estimate of the mean,  is as follows



The choice of method used in constructing a confidence interval estimate for  depends up on whether the:

1. Population distribution is normal or not
2. Population standard deviation is known or unknown.
3. Sample size is large or small

Therefore, there are different formulas (methods) that we have to use for the different conditions above.

Let us now see each of them one by one.

**Note:** Confidence interval = interval estimate.

**Confidence interval estimate for  when population is normal,  is known and n is large or small**



Where  = sample mean

 = Population mean

 = Population standard deviation

 = Sample size

C = confidence coefficient (level)

 = 1-C

### Example 1

The vice president of operations for Ethiopian Telecommunication Corporation is in the process of developing a strategic management plan. He believes that the ability to estimate the length of the average phone call on the system is important. Because thousands of calls have been placed on the system, averaging all the calls is virtually impossible. He takes a random sample of 60 calls from the company records and find that the mean sample length for a call is 4.26 minutes, Past history for these types of calls has shown that the population standard deviation for call length is about 1.1 minutes. Assuming that the population is normally distributed and he wants to have a 95% confidence, help him estimate the population mean.

# Given:

* Sample mean 
* Population standard deviation 
* n = 60
* c = 0.95



If he wants to produce a single number estimate of the value of , he can use the sample mean,  = 4.26. In this case, the sample mean is a point estimate of the population mean. However he realizes that if he were to randomly sample another 60 telephone calls, this second sample mean likely would be different from the first, and hence the point estimate would change. In fact, for every sample taken, the like hood is strong that the point estimate would change. Thus estimating a population parameter with an interval estimate often is preferable. In using interval estimate, the researcher must select a desired level of confidence; some of the more common ones are 90%, 95%, 98% and 99%. And the researcher cannot simply select the highest confidence level because there are trade offs between sample size, interval width and the level of confidence.

In our example, the vice president selected a 95% confidence level. When using a 95% level of confidence he is selecting on interval centered on  with in which 95% of all sample mean values will falls as shown in the figure below.

95%



0.25



0.25

z =-1.96

0.475

z =+1.96

0.475

***Figure 2-1 Distribution of sample means about population mean for 95% confidence***

Because the distribution is symmetric and the interval is equal on each side of the population mean,  (95%) or 0.4750 of the area falls on each side of the mean. Thus the Z table fives a Z value of 1.96 for this portion of the normal curve. Thus the Z value for a 95% confidence interval is always 1.96. In other words, of all the possible  values along the horizontal axis of the diagram, 95% of them should fall with in a Z score of 1.96 from the population mean. Now he can estimate the average call length by substituting 1.96 in place of Z.



4.26 - 1.96 





This can be interpreted, as the vice president can be 95% confident that the average length of a call for the population is between 3.8 and 4.54 minutes or we have a 95% chance that the population mean will be between 3.98 and 4.54. This means if he were to randomly select 100 samples of 60 calls each and use the results of each sample to construct a 95% confidence interval, approximately 95 of the 100 intervals would contain the population mean. It also indicates that 5% of the intervals would not contain the population mean.

We get the Z value for a certain (specific) confidence level from Z table.

You will find different tables at the back of this material. One of these tables is the Z table, which contains Z-values for different confidence levels. Let us now see how to pick up the Z-values of a specific confidence level.

For example we have said that the Z-value of 95% confidence level is 1.96. To get this number from the table we divide the confidence level by two.



We then search this number from the table.

Thus we obtain this figure in the row of 1.9 and under the column of 0.06. When we combine the row and column figures i.e. 1.9+ 0.0.6 = 1.96, which is the z- value for 95% confidence level.

Let us take another example to find the Z-value of a 90% confidence level.



Let us now search 0.45 from the table and we obtain 0.4495 in the 1.6 row and 0.04 column and 0.4505 in the 1.6 row and 0.05 column.

Because both this figures has equal difference from 0.45 we have to take a number between the 0.04 and 0.05 columns. And it would be . And then when we combine the row and column figures it would give us 1.6 + 0.045 = 1.645. Therefore, the Z-value of a 90% confidence level is 1.645. We can get Z-values for other confidence levels in this manner. The following table provides the Z-values only for some of the more common levels of confidence.

Confidence level Z- value

90% 1.645

95% 1.96

98% 2.33

99% 2.575

## **Example 2**

A survey conducted by Addis Zemen newspaper found that the sample mean age of men was 44 years and the sample mean age of women was 47 years. All together 454 people from Addis Ababa were included in the reader poor, 340 women and 114 men. Assume that the population distribution is normal and the standard distribution of age for both men and women is 8 years

1. Develop a 95% confidence interval estimate for the man age of the population men who read the newspaper.
2. Develop a 95% confidence interval estimate for the mean age of the population women who read the newspaper.
3. Compare the widths of the two interval estimates from part (a) and (b). Which one has a better precision? Why?

## **Solution**

#### Given:

   

   

a) 





**Interpretation**: We can be 95% confident that the mean age of the population men who read the newspaper is between 42.5 and 45.5 years

b) 





**Interpretation**: We can be 95% confident that the mean age of the population women who read the newspaper is between 46.1 and years 47.85

c) (b) is more precise than (a) because the interval is less while the

Confidence level is the same.

Precision is affected by

1. ***Sample size***- to increase precision, increase sample size.
2. ***Confidence level***- to increase precision decreases the confidence level.

***Check point:***

The standard deviation of the amounts poured in to bottles by an automatic filing machine is 1.8 milli litters. The amounts of fill in a random sample of bottles in ml were:

418 479 482 480 477 478 481 482

Suppose the population of amount of fill is normal. Construct 90% confidence interval estimate for the mean amount in all bottles filled by the machine.

**Confidence Interval estimate for  when population is normal, is unknown and n is large**

When the population standard deviation is unknown the sample standard deviation is an acceptable estimate of and substitutes the population standard deviation and can be used in the confidence interval, along with the sample mean, to estimate the population mean if sample size is large. A value of  is generally considered a lower limit for large sample size.

 Where  = sample standard deviation



## **Example 1**

Suppose that a car rental firm in Addis Ababa wants to estimate the average number of miles traveled by each of its cars rented. A random sample of 110 cars rented reveals that the sample mean travel distance per day is 85.5 miles with a sample standard deviation of 19.3 miles compute a 99% confidence interval to estimate.

## **Solution**

#### Given:

n = 110  miles S = 19.3 miles Z = 2.57







**Interpretation:**

We state with 99% confidence that the average distance traveled by rented cars lies between 80.77 and 90.23 miles

**Example 2**

A study is being conducted in a company that has 800 engineers. A random sample of 50 of these engineers reveals that the average sample age is 34.3 years and the sample standard deviation is 8 years. Construct a 97% confidence interval to estimate the average age of all engineers in this company.

## **Solution**

#### Given:

n = 50 N = 800  S = 8 years Z = 2.17

1) Calculate. If the number of the total population is known the formula

for  is as follows:

, Where N = population number





34.3 - 2.17 (1.096) 



**Interpretation:** we have a 97% chance that the population mean will lie. between 31.92 and 36.68

**Confidence interval estimate for  when population is normal,  is unknown and n is small**

If the sample size is small, we can only construct a confidence interval estimate for  if and only if the population is normally distributed. Therefore, when the population is normal, n is small and sample standard deviation (s) is used to estimate the population standard deviation , we use a distribution called t distribution to get the interval estimate of the population mean .

The t formula for sample when  is unknown, population is normal and n is small is



## **Characteristics of t-distribution**

1. The t distribution is symmetric about its mean 0 and ranges from negative infinity to positive infinity.



t-distribution

**Figure 2.1**

2. The t-distribution is bare-shaped (uni-modal) and has approximately

the same appearance as the standard normal distribution

3. The variance of the t-distribution is calculated by

Variance =  for v > 2 where V = degrees of freedom

4. The variance of the t-distribution always exceeds 1. The t-distribution

depends on a parameter V, called degrees of freedom of the

distribution.

 Where n is sample size

5. As V increases the variance of the t-distribution approaches 1 and the

shape approaches that of the standard normal distribution.

6. Because the variance of t exceeds 1 while the variance of the standard normal distribution equals 1, the t-distribution is slightly flatter in the middle than the standard normal distribution and has a ticker tails.

8. The t-distribution is a family of distributions with a different density function corresponding to each different value of the parameter.





Just like the Z-value, t value can be obtained from the t table found at the back of the material. To get this value, that is, we have to first find. For example, to get t value for a 98% confidence level where V= 26, first calculate 



We can now look up the table. In the table the figures at the top of every column are  and at the beginning of each row are degrees of freedom, V.

Therefore when we look at the intersection of 0.01 and 26 we get 2.48. This is the t-value.

Example 1

If a random sample of 27 items produces  and s = 20.6, what is the 98% confidence interval for. Assume that x is normally distributed for the population.

Solution

Given:

n = 27  s = 20.6 ν = n - 1 = 27- 1 = 26

First find out 



When we look at the table we obtain









Interpretation:

We can state with 98% confidence that the population mean lies between 118.58 and 138.22

Example 2

A sample of 20 taxi fares in Addis Ababa shows a sample mean of 2.5 Birr and a sample standard deviation of 0.5 Birr. Assuming that the population is normally distributed, develop a 90% confidence interval estimate of the mean taxi fares in Addis.

Solution

Given:

n = 20  s = 0.5  c = 0.9

First calculate 



Now look up  from the table



We can now use the formula to estimate 







Interpretation: we state with 90% confidence that the mean taxi fares in Addis lies between Birr 2.3 and Birr 2.67.

2.3.2 Interval estimation of the population proportion

We have seen how to estimate the population mean,  under different conditions. Let us now see how to estimate the population proportion p.

Business decision makers and researchers often need to be able to estimate an unknown population proportion. For example, companies may want to estimate their proportion of the market, the proportion of their produced goods that are defective etc. The symbol p represents the population proportion of "successes" and , represents the population proportion of "failure". For example, if we want to estimate the proportion of television viewers who watched a particular program, a television program viewer is a success and p is the population proportion of viewers and q = 1-p is the proportion of the population who did not watch the program.

If a simple random sample of size n is selected from a population that has p as the proportion of successes, the sample proportion of successes is symbolized as ; and . The formulas are





you learned that the sampling distribution of p is nearly normal if *n* is "sufficiently" large, that is, if both *np* and *nq* are grater than 5. Here, because we don't know the value of *p*, we will use as the estimator of *p*, and both  and  should be greater than 15 

Therefore, when constructing a confidence interval estimate for the unknown value of *p*, we use the estimator  in place of p in the formula

 



Where 



The above is the formula that we are going to use for estimating the unknown population proportion using the sample proportion.

## **Steps to be followed**

The objective is to estimate p using a confidence interval with confidence coefficient C.

1. Select a sample random sample such that both  and  are greater

than 15. And compute  and 

2. Compute 

3. Then look up  from the table

4. Finally compute  to get the confidence interval estimate of P.

P is in the interval





### Example 1

Last July, a random sample of 400 members of the labor force in one region showed that 32 were unemployed. Construct the 95 percent confidence interval for the proportion unemployed in the region.

**Solution**

1. First we have to compute the sample proportion of unemployed

 

2. Then we have to find out 



3. Then we have to look up  from the table and we find



4. Finally we have constructed the interval using the formula







Interpretation:

We state with 95 percent confidence that the population proportion of unemployed for the region is between 0.053 and 0.107 that is, between 5.3 and 10.7 percent.

Example 2

Recently a study of 87 randomly selected companies with tele-marketing operation was completed. The study revealed that 39% of the sampled companies had used tele-marketing to assist them in order processing. Using this information estimate the population proportion of tele marketing companies who use the tele marketing operation to assist them in order processing taking a 95% confidence level.

Solution

Given:

*n* = 87  *C* = 0.95

1. Calculate  and 

 

2. Compute



3. Look up  from the table



4. Construct the interval using the formula







Interpretation: We state with 95% confidence that the population proportion of tele marketing companies who use tele-marketing operation to assist them in order processing lies between 0.29 and 0.49.

2.3.3 Interval estimation of the unknown difference between two population means 

i) Interval estimation of the unknown difference between two population means  when population is normal, population standard deviation is known and *n* is large or small

In the last section we have seen how to calculate an interval estimate for a given population mean. Sometimes, however we may be interested to know the difference between two different population means.

Therefore, the formula for the confidence interval estimate for  when the population is normal and the population standard deviation is known is as follows:



Rearranging the formula



Where  = difference between two population mean

 = Difference between two sample means

 = standard error of the difference between two means.



### Example 1

In sex discrimination case an employee raised that a large proportion of men earn more than women for comparable work. Let population I represent all men employees performing certain jobs. And population 2 represents all female employees performing comparable jobs at the corporation. Independent samples of 100 males and 100 females were taken. And the sample means are found to be 20, 600 Birr for males and 19, 700 Birr for females. The sample standard deviations are 3000 Birr for males and 2, 500 Birr for the females.

1. Construct a 95% confidence interval estimate for 
2. What do you conclude from this

Solution

a) Given:

      

Steps

1. Calculate 



2. Calculate 



3. Look up from the table



4. Construct the confidence interval using the formula



20,600-19,700-1.96(390.5) ≤≤20, 600-19,700+1.96(390.5)



Interpretation: we state with a 95% confidence that the mean salary difference between male and female employees lies between 135.6 Birr and 1665.4 Birr.

b) From the above result we can conclude that there is sex discrimination between the two sexes because the interval contains positive number. That is, 

Example 2

A farmer wants to determine if different types of foods can influence the mean number of eggs that hens lay per month. In a random sample of 100 hence that ate food 1 the average number of eggs per month was 15.2 with variance 4. In a random sample of 100 hens that ate food 2, the average number of eggs per month was 14 with variance 4. Construct a 95% interval for. What do you conclude from your result?

Solution

Given:

   

  

Steps

1. Calculate 



2. Calculate 



3. Find out the value of  from the table.



4. Construct the interval using the formula







**Interpretation:** We state with a 95% confidence that the difference in the mean number of eggs that hens, which eat different types of foods, lay lies between 0.65 and 1.75

And form the above result we can conclude that the types of food hens eat can influence the number of eggs they lay.

**ii)Confidence interval estimate for, when the population is normal,  and  are unknown and sample size is small**

If  and  are unknown when the population is normal and sample size is small, to estimate  we have to use t-distribution, just like what we have done previously to estimate  when  is unknown, with slight modification. And to use a t distribution in this case, we have to assume that the unknown population variances are equal. That is,



And in this case the formula for constructing a confidence interval for is as follows



Where: ν is the pooled degree of freedom which is calculated as:  and 

**Example 1**

Two manufacturing companies produce drill tips that are used to cut holes in steel sheets. A customer wishing to know which drill tips has the longer life purchases independent samples of  drill tips from company 1 and  drill tips from company 2. The mean lives of the drill tips are  minutes and  minutes. The sample variances are  and . Construct a 95% confidence interval for assuming that the two populations are normally distributed. What do you conclude from you results?

### Solution

**Given:**    

 

 

### Steps

1. Calculate the pooled estimator of common population variance 







2. Calculate  and 









3. Look up the value of  from the table



4. Calculate 



 = 2.13

5. Construct the confidence interval using the formula







Interpretation: we state with 95% confidence that the difference between the two population mean lies between -10.35 and -1.65.

From the above result, we can conclude that drill tips which are manufactured in company 2 have longer life than drill tips which are manufactured in company 1 because the interval for  is negative where  is the mean life of drill tips for company 1 and  is the mean life of drill tips for company 2. So drill tips of company 2 are preferred than company 1.

### Example 2

5 years old children were being studied to determine whether children whose parents are college graduate watch more or less TV than children whose parents are not college graduates. Independent random samples of 21 children were selected from each population. The sample means and variances were  hours,  hours,  and . The population variances are assumed to be equal and the populations are assumed to be normal. Calculate a 90% confidence interval for the difference between two population mean

## **Solution**

#### Given:

  C = 0.95

 

 

1. Calculate 







3. Look up  from the table



4. Calculate 



5. Construct the confidence interval using its formula







Interpretation: We can state with 90% confidence that the difference between two populations lies between -6.55 and -1.45.

2.3.4 Interval estimation of the difference between two independent proportions 

Previously we have seen how to calculate an interval estimate of a population proportion p. However, we may also be interested to know the difference between two different population proportions. Therefore, given that  are all greater than or equal to five the formula to calculate the confidence interval estimate for  is as follows



Where



### Example

Ethiopia Teachers Association wants to determine whether membership in the Teachers' ' Idir' is independent of marital status or not. A random sample of 600 teachers indicated that 200 are members of the "Idir" where as the rest are non members. Concerning marital status, 100 of the "Idir" members and 170 of the non-members are single. The rest of the samples are married.

1. Is membership in the "Idir" independent of marital status? Test at 90% confidence level.
2. Interpret the result

## **Solution**

#### Given:

 Is the number of teachers who are a member of “Idir?”

 Is the number of teachers who are non-members?

C = 0.90

1. First we have to calculate 









2. Calculate 







3. Calculate 



4. Look up  from the table



5. Construct the interval for  using its formula







**Interpretation:** We can state with 90 percent confidence that the difference between the two population proportions lays between 0.004 and 0.146 that 0.5 percent and 14.6 percent.

1. Therefore, from the above fact that  we can say that membership in the "Idir" is not independent of marital status because the proportion of single persons who are members of the "Idir"  is greater than the proportion of single persons who are non-members 

## **2.4 Determination of Sample Size**

The reason for taking a sample from a population is that it would be too costly to gather data for the whole population. But collecting sample data also costs money; and the larger the sample, the higher the cost. To hold cost down, we want to use as small a sample as possible. On the other hand, we want the sample to be large enough to provide "good" estimates of population parameters. Consequently, the questions is " How large should the sample be?" the answer depends on three factors:

1. How precise (narrow) do we want a confidence interval estimate to be.

2. How confident do we want to be that the interval estimate is correct.

3. How variable is the population being sampled.

In general, the Higher the desired precision or level of confidence, the larger (more costly) will be the sample; also for a given precision and level of confidence, the larger population variability is , the larger (more costly) will be the sample.

Let us now see how to determine sample size for estimating  and p

**2.4.1 Sample size for estimating a population mean **

The confidence interval estimate of  is



The above can be expressed as:



The above is of the form  where



Squaring both sides, we have



Solving for n, we have



To use this formula, we must guess a value for . Using the guessed value for , the formula for the sample size is needed in estimating  is



**Example 1:**

A gasoline service station shows a standard deviation of 6.25 Birr for the charges made by the credit card customers. Assume that the station's management would like to estimate the population mean gasoline bill for its credit card customers to be with in  Birr 1. For a 95% confidence level how large a sample would be necessary.

**Solution**

## **Given:**

 Birr C = 0.95





Therefore we have to calculate 









Therefore, for a 95% confidence level the sample size should be 150.

**Example 2:**

The National Travel and Tour organization would like to estimate the mean amount of money spent by a tourist to be with in 100 with 95% confidence. If the amount of money spent by tourists is considered to be normally distributed with a standard deviation of 200 Birr, what sample size would be necessary for the NTO to meet their objective in estimating this mean amount.

**Solution**

**Given:**

 c= 0.95





Therefore we have to calculate 









Therefore, there sample size should be 15.37

**2.4.2 Determination of  if  is unknown**

If is unknown, we can roughly approximate  by the following formula



**Example**

A sample is to be taken to estimate the mean salary of plumbers to with in  500, with a confidence level of 99%. A plumbers union official states that $40, 000 and $26, 000 would be unusually large and small salaries for plumbers in the union what should the sample size be?

**Solution**

**Given**

e = 500

C = 0.99

First we have to calculate 







And is calculated as follows



Then, we have to look up  from the table



Now we can find 







Therefore, sample size should be 266

**2.4.3 Determination of sample size for estimating a population proportion **

The confidence interval estimate for p is



Let us call  the error term and symbolize it by e. Thus, 

So, we can express the confidence interval as follows:



If we square both sides of



We obtain,



When we rearrange the above we get



Therefore, the formula for, the sample size needed in estimating a population proportion with confidence coefficient c, is



**Example 1**

Suppose that a production facility purchases a particular component parts in large its from a supplier. The production manager wants to estimate the proportion of defective parts received from the supplier. She believes that the proportion of defects is no more than 0.2 and wants to be with in 0.02 of the true proportion of defects with a 90% level of confidence. How large a sample should she take?

## **Solution**

#### Given:

e = 0.02 c = 0.90

P = 0.2



To use the above formula we have to first calculate the values of  and q.









We can now substitute the above figures in to the formula to obtain the sample size.







Therefore, she should take a sample of 1076

**2.4.4 Determining  when population proportions are not known**

If population proportions are not known we take 0.5 as p and 0.5 as q to determine the largest sample size.

### Example

What is the largest sample size that would be needed in estimating a population proportion to be with in 0.02 with a confidence level of 95%

Solution

Given:

e = 0.02 c = 0.95

Because population proportions are not given we take

p = 0.05 and q = 1- p = 0.05



Let us now calculate 





Therefore,  can be calculated as follows





The above result implies that the largest sample size is 2401

**2.5.5 Determining sample size for estimating **

We have seen that



Let e represent 

Therefore, 

But here we have to assume that

and

Therefore the formula for determining sample size for estimating  I is as follows



**Note:**

The sample size that is needed to estimate  is twice as the sample size needed to estimate.

### Example

A college admissions officer wants to estimate the difference in the average scores of men and women. She plans to take a random sample of men and women who had taken the test at the same time. She wants to be with in 10 points of the true difference in the mean scores of men and women and 95% confident of her results. Past test results indicate that the standard deviation of the test scores is about 105 points. How large should the sample size be?

## Solution

#### Given

  





Therefore 

This implies the sample size should be 847

##### *Solved problems*

1. A sample of data for the price changes in ten-years Treasury bonds

taken from a normal population is

4 10 2 8 4 14 10 12 8

Find the point estimates for the following

1. The mean 
2. The standard deviation 
3. The variance 
4. The standard deviation of the mean 

## Solution

a) 

Therefore, the point estimate of  is 

b) 

We calculate  and obtain 128



Therefore, the point estimate of is 4 

c) 

Therefore, the point estimate of variance 

d) 

Therefore, the point estimate of 

2. Time magazine reports information on the time required for caffin products such as coffee and soft drinks to leave the body after consumption. Assume that the 99% confidence interval estimate of the population mean time for adults is 5.6 hours to 6.4 hours. Assuming that the population distribution is normal;

a) What is the point estimate of the mean time for caffin to leave the

body after consumption.

b) If the population standard deviation is 2 hours, how large a sample

was used to provide the interval estimate.

## **Solution**

#### Given:



**a)**

5.6



6.4

The difference of the interval is

6.4 - 5.6 = 0.8

When we divide 0.8 by two we get the difference between 6.4 or 5.6

and  is calculated as:







Therefore, the point estimate of 

**b) Given:**

 hours





Let us take the minimum interval which is 5.6











3. A fast-food restaurant took a random sample of 400 customers to determine the proportion of customers who are in a confidence interval of 0.73 to 0.87 was reported.

1. Find the number of females and the sample proportion
2. Find the level of confidence of this interval

**Solution**

**a) Given**

n = 400 0.73  p  0.87

Therefore taking the highest and lowest intervals we can construct the following two equations



Solving the two equations simultaneously



Therefore, the sample proportion  = 0.8

Let us now find the number of females.

Number of females will be  multiplied by n



b) To find the level of confidence we have to first find 







Let us now do the reserve of what we have been doing to obtain Z -value, to get the confidence level. So the table shows us that the Z-value 3.50 has a value of 0.4997. That is



. Therefore the confidence level of the interval is 99.55%

4. An advertising executive thinks that the proportion of consumers who have seen his company's advertising in newspapers is somewhere between 0.65 and 0.85. The executive wants to estimate the consumer population proportion to be between 0.05 and have 98% confidence in the estimate. How large a sample should be taken?

Solution

Given:

 e = 0.05 c = 0.98



The largest sample is obtained if we take the minimum interval as P.











Summary

An estimator is a sample statistic that is used to estimate an unknown population parameter. A point estimate is a single value of an estimator. The sample mean

=x/n

Is used as the estimator of the population mean . The sample standard deviation



is used as the estimator of the population standard deviation ðx. The sample standard error of the mean ;



Is the estimator of the standard error of the mean ð. The estimator of the population proportion of success p is the sample proportion of successes;

=number of success in a sample of size n

n

a confidence interval estimate is a range of values within which we state a population parameter lies. The proportion of confidence intervals that are correct (i.e,that contain the population parameter) is called the *confidence coefficient* C; the confidence level C is expressed as a percent. The proportion of incorrect interval is symbolized as ; = 1-C and /2=1-C/2

The symbol Z/2 means the z value for which the one tail area of the standard normal curve is /2. The confidence interval( with confidence coefficient C) for the mean of a normal population whose standard deviation is ðx is



If ðx is estimated by Sx, and sample size n is greater than 30, the approximate confidence interval for is; 

If the sample is drawn from a normal population whose standard deviation is unknown, and n is less than or equal to 30,we use the student t value(rather than z) in constructing a confidence interval for . The t value depends on /2 and the number of degrees of freedom v. For the methods discussed in this chapter, v=n-1. The symbol t/2,v means the t value for a one-tail t distribution area of /2, with v degrees of freedom. The confidence interval for , with confidence coefficient C, is;



When both n and nq are greater than 15, the approximate confidence interval for a proportion of success (with confidence coefficient C) is;



The sample size needed to estimate unknown population parameter to within + e, depends up on the value of e, the confidence coefficient C, and the population variability. To compute the sample size n for estimating a population mean, first decide on the value to be used for e, and the confidence coefficient desired, C. next guess a value for ðx ,the population standard deviation. Then; 

to compute the sample size np  for estimating a population proportion, first establish the values for e and C. Next, guess the value of the proportion p. Then 

*Exercises*

1. A candy company fills a 20g package of a certain candy with individually wrapped pieces of candy. The number of pieces of candy per package varies, because the package is sold by weight. The company wants to estimate the number of pieces per package. Inspectors randomly sample 120 packages of this candy and count the number of pieces in each package. They find that the sample mean number of pieces is 18.72 with a sample variance of 0.763. What is the point estimate of the number of pieces per package? Construct a 99% confidence interval to estimate the mean number of pieces per package for the population.

2. The marketing director of a large department store wants to estimate the average number of customers who enter the store every 5 minutes. She has a research assistant who randomly selects 5-minutes intervals and counts the number of arrivals at the store. The assistant obtains the figures 58 32 41 47 56 80 45 29 32 and 78. The analyst assumes that the number of arrivals is normally distributed. Using these data the analyst computes a 95% confidence interval to estimate the mean value for all 5-minite intervals. What interval value does she get?

3. A recent survey of 1060 women shoppers revealed that, on the average they spent $550 a year on clothes. Forty two percent said that they are buying more clothes at mass merchants. About 15% revealed that they are shopping less frequently at department stores. Construct a 95% confidence interval to estimate the population proportion of women shoppers who are shopping less frequently at department stores.

4. The Ethiopian Civil Service Commission says that the average work week in Ethiopia is down to only 35 hour, largely because of a rise in part-time workers. Suppose this figure was obtained from a random sample of 20 workers and that the standard deviation of the sample was 4.3 hours. Assume that hours worked per week are normally distributed in the population. Use this sample information to develop a 98% confidence interval for the population variance of the number of hours worked per week for a worker. What is the point estimate?

5. Suppose that a production facility purchases a particular component part in large lots from a supplier. The production manager wants to estimate the proportion of defective parts received from this supplier. She believes that the proportion of defects is no more than 20 and wants to be with in 0.02 of the true proportion of defects with a 90% level of confidence. How large a sample should she take?

6. A health service provider has been reimbursed Gray cross and Gray shield for a large number of services. It appears that overpayments may have occurred on many of the reimbursements due to errors made during the billing process or owing to up coding (billing for a more complicated service than was actually provided). Find the 95% confidence interval estimate of the mean overpayment per service given the following random sample of overpayments. (Assume normality)

$7 $3 $3 $3

7. A sample of 100 stocks from the oil and Gas field services industry revealed that 40 of two stocks had negative annual return on investment the previous year. Under economic conditions similar to those of the previous year, find the 95% confidence interval estimate for the fraction of stocks that will have negative return for this industry.

8. If a normal population of overpayments by National Insurance of health service provider is known to have a population standard deviation equal to $5, how large a sample would we take in order to be 95% confident that the sample mean overpayment will not differ from the population mean overpayment by more than  80?

9. Cattle are sold by weight. A sample of six animals from one herd at American Livestock Corporation yielded the following weights (in pounds):

692 800 685 790 695 793

Find a 95% confidence interval for the mean weight per animal in the herd. Assume normality.

10. Twenty secretaries were given a spelling test. They were then given a special short course designed to improve spelling ability and were tested again at the end of the course. The differences between the first and the second scores had a mean equal to 4(that is, there was a 4-point improvement). The variance of the improvement was 16. Find a 90% confidence interval for the mean of the score improvement if this course were given to all secretaries in the company. Assume normality.