1. Introduction

Welcome to the series of E-learning modules on practical illustrations on construction of Confidence Intervals on proportions. In this module we are going cover the basic concepts of proportions, the procedure to obtain interval estimates for a single population proportion and an interval estimate for the difference between two population proportions through practical examples.

By the end of this session, you will be able to:

- Understand basic concepts of proportions
- Apply the method of Interval Estimation for the estimation of a Population proportion

• Describe the procedure to obtain Interval Estimates for the difference between two Population proportions

It is always possible to classify the whole population into various classes with respect to the attributes or qualitative characteristics under study.

We consider the cases where the population is divided into two classes only; say C and C dash with respect to an attribute.

Hence any unit in the population may be placed in class C or C dash respectively depending on whether it possess or does not possess the given attribute.

In the study of attributes we are interested in the estimate of population proportion.

2. Notations

Notations:

Let, P be equal to the Proportion of units possessing the given attribute in a population equal to X divided by N

Let Q be equal to the Proportion of units that do not possess the given attribute which equals to X dash divided by N which is equal to one minus P.

In statistical language P and Q are the proportion of successes and failures respectively in the population.

Let us consider Simple Random Sample of size n from this population. Let 'x' be the number of units in the sample possessing the given attribute.

Then, p equals to Proportion of sampled units possessing the given attribute which is equal to x divided by n and q is equal to Proportion of sampled units which do not possess the given attribute which is equal to one minus p.

A confidence interval for the population proportion P gives an estimated range of values which is likely to include an unknown population parameter.

In this case P, the population proportion, the estimated range being calculated from a given set of sample data.

Similarly we can obtain the interval estimates for the difference between two population proportions from the given sets of sample data drawn from two populations.

Hundred into (one minus alpha) percent Confidence Interval for the population proportion P is given by

[p minus Z alpha by two into square root of P into Q by n, p plus Z alpha by two into square root of P into Q by n]

Note: If P is not known then we use the estimate of P

The estimate of P is p

Therefore hundred into (one minus alpha) percent Confidence Interval for the population proportion P is given by

[p minus Z alpha by two into square root of p into q by n, p plus Z alpha by two into square root of p into q by n]

Therefore hundred into (one minus alpha) percent confidence interval for the difference of population proportions is given by

(p one minus p two) minus Z alpha by two into square root of (P one into Q one by n one, plus P two into Q two by n two), (p one minus p two) plus Z alpha by two into square root of (P one into Q one by n one plus P two into Q two by n two).

Note: If P one and P two are not known then we use the estimates of P one and P two as p one and p two.

Therefore hundred into (one minus alpha) percent confidence interval for the difference of population proportions is given by

(p one minus p two) minus Z alpha by two into square root of (p one into q one by n one plus p two into q two by n two), (p one minus p two) plus Z alpha by two into square root of (p one into q one by n one plus p two into q two by n two).

3. Illustrative Examples

Problem1:

Ina random sample of one hundred articles taken from a large batch of articles, ten are found to be defective. Obtain ninety five percent confidence interval for the true proportion of defectives in the batch.

Solution:

Let xi denote the number of defective articles, which is equal to ten.

Let n be the total number of articles, which is equal to one hundred.

Estimate of the proportion of defective articles is given by,

p is equal to x i by n which is equal to ten by hundred which is equal to zero point one.

Then q is equal to one minus p which is equal to zero point nine zero.

Since hundred into (one minus alpha) percent is equal to ninety five percent, alpha is equal to zero point zero five and alpha by two is equal to zero point zero two five. Then from the table of probabilities of Normal distribution Z alpha by two which is equal to Z zero point zero two five is equal to one point nine six.

Then hundred into (one minus alpha) percent confidence interval for the population proportion P is given by

[p minus Z alpha by two into square root of p into q by n p plus Z alpha by two into square root of p into q by n]

Which is equal to [zero point one zero minus one point nine six into square root of (zero point one zero into zero point nine zero) by one hundred, zero point one zero plus one point nine six into square root of (zero point one zero into zero point nine zero) by one hundred] Which is equal to [zero point zero four one two, zero point one five eight, eight]

Hence ninety five percent confidence interval for the true proportion of defectives is zero point zero four one two, zero point one five eight, eight.

That is we conclude that the percentage of defectives in the batch is in between **four** and **sixteen** percent.

Problem 2:

In a sample of eight hundred and fifty men from district A, three hundred and fifty are found to be smokers. In a sample of one thousand two hundred men from another district B, the number of smokers is five hundred and seventy five. Establish ninety nine percent confidence interval for the difference between the proportions of the smokers in the two districts.

Solution:

Proportion of smokers in district A is equal to p one which is equal to three hundred and fifty by eight hundred and fifty which is equal to zero point four one.

Proportion of smokers in district B is equal to p two which is equal to five hundred and seventy five by one thousand two hundred which is equal to zero point four eight.

q One is equal to one minus p one which is equal to one minus zero point four one which is equal to zero point five nine.

Q two is equal to one minus p two which is equal to one minus zero point four eight which is equal to zero point five two.

Since hundred into (one minus alpha) percent is equal to ninety nine percent, alpha is equal to zero point zero one and alpha by two is equal to zero point zero zero five. Then from the table of probabilities of Normal distribution, Z alpha by two is equal to Z zero point zero zero five which is equal to two point five seven.

Then hundred into (one minus alpha) percent confidence interval for the difference between the proportions of the smokers in the two districts is given by, (p one minus p two) minus Z alpha by two into square root of (p one into q one by n one plus p two into q two by n two), (p one minus p two) plus Z alpha by two into square root of (p one into q one by n one plus p two into q one by n one plus p two into q two by n one plus p two into q two by n two).

Which is equal to [(zero point four eight minus zero point four one) minus two point five seven into square root of (zero point four one into zero point five nine by eight hundred and fifty plus zero point four eight into zero point five two by one thousand two hundred), (zero point four eight minus zero point four one) plus two point seven into square root of (zero point four one into zero point five nine by eight hundred and fifty plus zero point four eight into zero point five nine by eight hundred and fifty plus zero point four eight into zero point five nine by eight hundred and fifty plus zero point four eight into zero point five nine by eight hundred and fifty plus zero point four eight into zero point five two by one thousand two hundred)] which is equal to [zero point zero one two six, zero point one two seven four].

Hence we can conclude that there is a difference of one to thirteen percent in the proportion of smokers in the two districts.

4. Examples (Part 1)

Problem 3:

Ten Life insurance policies in a sample of two hundred taken, out of fifty thousand were found to be insured for less than five thousand rupees. How many policies can be reasonably expected to be insured for less than five thousand rupees in the whole lot at ninety five percent confidence level?

Solution:

Let x i denote the number of policies who are insured for less than five thousand rupees which is equal to ten. Let n be the total number of polices which is equal to two hundred.

Estimate of the proportion of policies who are insured for less than five thousand rupees is given by,

p is equal to x i by n which is equal to ten by two hundred which is equal to zero point zero five.

Then q is equal to one minus p which is equal to zero point nine five.

Since hundred into (one minus alpha) percent is equal to ninety five percent, alpha is equal to zero point zero five and alpha by two is equal to zero point zero two five. Then, from the table of probabilities of Normal distribution

Z alpha by two which is equal to Z zero point zero two five is equal to one point nine six.

Then hundred into (one minus alpha) percent confidence interval for the population proportion P is given by

[p minus Z alpha by two into square root of p into q by n, p plus Z alpha by two into square root of p into q by n]

Which is equal to [zero point zero five minus one point nine six into square root of (zero point zero five into zero point nine five) by two hundred, zero point zero five plus one point nine six into square root of (zero point zero five into zero point nine five) by two hundred] which is equal to [zero point zero one nine eight, zero point zero eight zero two.

Hence ninety five percent confidence interval for the proportion of policies who are insured for less than Rupees five thousand rupees is [zero point zero one nine eight, zero point zero eight zero two]

Number policies can be reasonably expected to be insured for less than five thousand rupees in the whole lot at ninety five percent confidence level is

[Zero point zero one nine eight, zero point zero eight zero two] into fifty thousand.

Hence number policies that can be reasonably expected to be insured for less than five thousand rupees in the whole lot at ninety five percent confidence level is in between nine hundred and ninety and four thousand and ten.

5. Examples (Part 2)

Problem 4:

An investigation of the performance of two machines in a factory manufacturing large number of fans gives the following results.

In machine one, number of fans examined were two hundred and seventy five out of which seventeen defective were found.

In machine two, three hundred and fifty fans were examined and twenty two defective fans were found.

Find the ninety five percent confidence Interval for the difference in proportions of defectives.

Solution:

Proportion of defective fans produced by machine one is equal to p one which is equal to seventeen by two hundred and seventy five which is equal to zero point zero six one.

Proportion of defective fans produced by machine two is equal to p two which is equal to twenty two by three hundred and fifty which is equal to zero point zero six three

q one is equal to one minus p one is equal to one minus zero point zero six one which is equal to zero point nine three nine

q Two is equal to one minus p two is equal to one minus zero point zero six three which is equal to zero point nine three seven.

Since hundred into (one minus alpha) percent is equal to ninety five percent; alpha is equal to zero point zero five. Then from the table of probabilities of Normal distribution Z alpha by two is equal to one point nine six.

Then hundred into (one minus alpha) percent confidence interval for the for the difference in proportions of defectives is given by, (p one minus p two) minus Z alpha by two into square root of (p one into q one by n one plus p two into q two by n two) , (p one minus p two) plus Z alpha by two into square root of (p one into q one by n one plus p two into q two by n two) Which is equal to [(zero point zero six three minus zero point zero six one) minus one point nine six into square root of (zero point zero six three into zero point nine three nine by two hundred and seventy five plus zero point zero six three minus zero point six one) plus one point nine six into square root of (zero point zero six three minus zero point six one) plus one point nine six into square root of (zero point zero six three minus zero point nine three nine by two hundred and fifty), (zero point zero six three minus zero point nine three nine by two hundred and seventy five plus zero point zero six one into zero point nine three nine by two hundred and seventy five plus zero point zero six one into zero point nine three nine by two hundred and seventy five plus zero point zero six three minus zero point nine three nine by two hundred and seventy five plus zero point zero six three into zero point nine three seven by three hundred and fifty)]. This is equal to [minus zero point zero three six one two, zero point zero four zero one two]

Notice that this confidence interval includes the value of zero which implies there may be negligible difference in the difference of proportion of defective fans produced by two machines.

Problem 5

Obtain ninety five percent interval estimates for the proportion of persons who would vote for a guilty verdict in a particular sexual harassment case. And also estimate the limits for proportions among male and female jurors. Of hundred and sixty mock jurors of both sexes, hundred and five voted guilty and fifty five voted not guilty. Among the eighty female jurors fifty eight voted guilty. Among the eighty male jurors forty seven voted guilty. Solution:

Our point estimate of the population proportion is simply our sample proportion, p is equal to hundred and five by hundred and sixty, is equal to zero point six five six.

For a ninety five percent confidence interval we compute:

[p minus or plus z alpha by two into square root of p q by n].

Z alpha by two is equal to Z point zero five by two which is equal to one point nine six from the table of Normal probabilities

By substituting in the above we get, [point six five six minus or plus one point nine six into square root of point six five six into point three four four by one sixty which is equal to point six five six minus or plus point zero seven four.

Hence the interval is [zero point five eight two, zero point seven three zero]

If we look at the proportions separately for female and male jurors, we see that among the eighty female jurors fifty eight voted guilty. For a ninety five percent confidence interval we compute: p is equal to fifty eight by eighty equal to point seven two five. Point seven two five minus or plus one point nine six into square root of zero point seven two five into zero point two seven five by eighty is equal to zero point seven two five minus or plus zero point zero nine eight. Hence the interval is [zero point six two seven, point eight two three].

Among the eighty male jurors forty seven voted guilty. For a ninety five percent confidence interval we compute: p is equal to forty seven by eighty equal to zero point five eight eight. Zero point five eight eight minus or plus one point nine six into square root of point five eight eight into point four one two by eighty is equal to zero point five eight eight minus or plus zero point one zero eight.

Hence the interval is [point four eight zero, zero point six nine six]

Do notice that the confidence interval for the male jurors overlaps the confidence interval for the female jurors.

Here's a summary of our learning in this session where we have :

- Understood the concept of proportions
- Described the procedure to apply Interval Estimation technique for the estimation of a population proportion through practical examples
- Explained how to apply Interval Estimation method for the estimation of the difference between two population proportions