1. Introduction

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

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

- Explain the proportions
- Explain the interval estimate for the population proportion
- Explain the interval estimate for the difference between two population proportions
- Apply these interval estimates for the estimation of the population proportions

A qualitative characteristic which cannot be measured quantitatively is known as an attribute. For example, honesty, beauty, intelligence etc. Quite often, we come across the situations, where it may not be possible to measure the characteristic under study. However, it may be possible to classify the whole population into various classes with respect to the attributes under study.

We consider the cases, where the population is divided into two classes say C and C dash with respect to an attribute. Such a classification is termed as dichotomous classification. Hence, any unit in the population may be placed in class C or C dash respectively, depending whether it possess or does not possess the given attribute. In the study of attributes, we are interested in the estimate of population proportion.

For example:

- Proportion of defective items in a large consignment of such items
- Proportion of the literates or the bread winners in a town
- Percentage of sales accounted for by a particular product
- Proportion of viewing public for a particular program
- Proportion of trees with a diameter of ten inches or more

Suppose we don't have a mean, then how do we proceed to get the population parameters?
For example:

- Percentage of people who vote for Congress party in an election
- Proportion of the population who is in a certain category

To estimate above characteristics, we need another method.

Surveys and experiments often produce counts, which we can turn into proportions.
Count is equal to f by n which is equal to a proportion
For example hundred by six hundred is equal to zero point six zero is a proportion
Or multiply by hundred to get a percentage
Zero point six zero into hundred is equal to sixty percent

2. Notations

Let us consider that a population with N units Y one, Y two, up to YN is classified into two disjoint and exhaustive classes C and C dash respectively for a given attribute.
Let the number of individuals in the classes C and C dash be X and X dash respectively such that X plus X dash is equal to N

Then,
P is equal to the proportion of units possessing the given attribute, which is equal to X divided by N
Q is equal to the proportion of units does not possess the given attribute, which is equal 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 a 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 is equal to Proportion of sampled units possessing the given attribute 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 is calculated from a given set of sample data.

3. Result 1

We have to find hundred into (1 minus alpha) percent Confidence Interval for the population proportion. Let X be a number of units possessing an attribute (in the population). Let P be the population proportion and
Let X follows Binomial with parameters n and P. Then, expected value of X is equal to n into P and variance of X is n into P into Q, which is equal to n into P into (1 minus P)
Let us take a sample of size n from the above population.

Let p is equal to x by n be the sample proportion where, x is the number of units possessing an attribute in the sample and n is the sample size.
Expected value of p is equal to Expected value of x by n which is equal to 1 by n into expected value of x is equal to 1 by n into n into P which is equal to P
Variance of p is equal to Variance of x by n which is equal to 1 by n square into variance of x is equal to 1 by n square into n into P into Q which is equal to P into Q by n
As the sample size is large (as n tends to infinity) Binomial distribution tends to Normal distribution.
Therefore, p follows Normal distribution with parameters P and P into Q by n
Then, Z is equal to p minus P by square root of P into Q by n which follows Normal distribution with mean zero and variance1

For the given (1 minus alpha), we can always find two quantities minus Z alpha by 2 and Z alpha by 2 from Standard Normal variate tables such as

Probability of minus Z alpha by 2 less than or equal to Z less than or equal to Z alpha by 2, which is equal to 1 minus alpha

Probability of minus Z alpha by 2 less than or equal to p minus P by square root of P into Q by n less than or equal to Z alpha by 2, which is equal to 1 minus alpha

Probability of minus Z alpha by 2 into square root of P into Q by n less than or equal to p minus P less than or equal to Z alpha by 2 into square root of P into Q by n, which is equal to 1 minus alpha

Probability of p minus Z alpha by 2 into square root of P into Q by n less than or equal to P less than or equal to p plus Z alpha by 2 into square root of P into Q by n, which is equal to 1 minus alpha

Therefore, hundred into (1 minus alpha) percent Confidence Interval for the population proportion P is given by

[p minus Z alpha by 2 into square root of P into Q by n , p plus Z alpha by 2 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 (1 minus alpha) percent Confidence Interval for the population proportion P is given by

[p minus Z alpha by 2 into square root of p into q by n, p plus Z alpha by 2 into square root of p into q by n]

4. Confidence Interval for Difference

- We use the two sample proportions to construct a confidence interval for the difference in population proportions P one minus P two between two groups
- Interval is centred at the difference of the two sample proportions
- As usual, the multiple Z we use depends on the confidence level that is needed
- For example, for a ninety five percent confidence interval, Z is equal to one point nine six

Examples:

- The difference in the proportion of males and females who will vote for Congress
- The difference in the proportion of republicans and democrats who will vote for Mr. Gore
- The difference in the proportion of smokers and non-smokers in a district

5. Result 2

On the basis of independent random samples of sizes n1 and n2 from two Binomial populations, it is desirable to estimate the difference between the parameters P one and P two.
For example, if we wish to estimate the difference between the proportions of voters in two different constitutes that favour a candidate X for a particular election.

We have to find hundred into (1 minus alpha) percent confidence interval for the difference of population proportions. Let X one be the number of units possessing an attribute in the first population. Let P one be the population proportion of that population.
Let X one follows Binomial with parameters n one and P one. Then, Expected value of X one is n one into P one and Variance of X one is equal to n one into P one into Q one, which is equal to n one into P one into (1 minus P one)
Let us take a sample of size n one from the above population. Let p one is equal to x one by n one be the sample proportion, where x one is the number of units possessing an attribute in the sample of the first population and n one is the sample size.

Expected value of p one is equal to Expected value of x one by n one, which is equal to 1 by n one into expected value of x one is equal to 1 by n one into n one into P one, which is equal to P one
Variance of p one is equal to Variance of x one by n one, which is equal to 1 by n one square into variance of x one, which is equal to 1 by n one square into n one into P one into Q one which is equal to P one into Q one by n one
Let X two be a number of units possessing an attribute in the second population. Let P two be the population proportion of that population.
Let X two follows Binomial with parameters n two and P two. Then, Expected value of X two is n two into P two and Variance of X two is equal to n two into P two into Q two, which is equal to n two into P two into (1 minus P two)
Let us take a sample of size n two from the above population. Let p two is equal to x two by n two be the sample proportion, where x two is the number of units possessing an attribute in the sample of the second population and n two is the sample size.

Expected value of p two is equal to Expected value of x two by n two which is equal to 1 by n two into expected value of x two is equal to 1 by n two into n two into P two, which is equal to P two
Variance of p two is equal to Variance of x two by n two, which is equal to 1 by n two square into variance of x two is equal to 1 by n two square into n two into P two into Q two, which is equal to P two into Q two by n two
As the sample size is large (as n two tends to infinity), Binomial distribution tends to Normal distribution.
Therefore, p one follows Normal distribution with parameters P one and P one into Q one by n one
And p two follows Normal distribution with parameters P two and P two into Q two by n two
p one minus p two follows Normal distribution with parameters (P one minus P two) and (P one into Q one by n one plus P two into Q two by n two)
Then, Z is equal to (p one minus p two) minus (P one minus P two) by square root of (P one

into Q one by n one plus P two into Q two by n two)which follows Normal distribution with mean zero and variance 1

For the given (1 minus alpha), we can always find two quantities minus Z alpha by 2 and Z alpha by 2 from Standard Normal variate tables such as

Probability of minus Z alpha by 2 less than or equal to Z less than or equal to Z alpha by 2 which is equal to 1 minus alpha

Probability of minus Z alpha by 2 less than or equal to (p one minus p two) minus (P one minus P two) by square root of (P one into Q one by n one plus P two into Q two by n two) less than or equal to Z alpha by 2, which is equal to 1 minus alpha

Probability of minus Z alpha by 2 into square root of (P one into Q one by n one plus P two into Q two by n two) less than or equal to (p one minus p two) minus (P one minus P two) less than or equal to Z alpha by 2 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 1 minus alpha

Probability of (p one minus p two) minus Z alpha by 2 into square root of (P one into Q one by n one plus P two into Q two by n two) less than or equal to (P one minus P two) less than or equal to (p one minus p two) plus Z alpha by 2 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 1 minus alpha

Therefore, hundred into (1 minus alpha) percent confidence interval for the difference of population proportions is given by

(p one minus p two) minus Z alpha by 2 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 2 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 (1 minus alpha) percent confidence interval for the difference of population proportions is given by

(p one minus p two) minus Z alpha by 2 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 2 into square root of (p one into q one by n one plus p two into q two by n two)

For example:

Suppose we wish to estimate the proportion of persons who would vote for a guilty verdict in a particular sexual harassment case. We shall use the data from a study by Egbert, Moore, Wuensch, and Castellow (J*ournal of Social Behaviour and Personality*). Of hundred and sixty mock jurors of both sexes, hundred and five voted guilty and fifty five voted not guilty.

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

For a ninety five percent confidence interval we compute:
[p minus or plus z alpha by 2 into square root of pq 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 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 [point five eight two, point seven three zero]

Suppose we look at the proportions separately for female and male jurors. 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 is equal to point seven two five
Point seven two five minus or plus one point nine six into square root of point seven two five into point two seven five by eighty. Hence, the interval is [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, which is equal to point five eight eight
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
Hence, the interval is [point four eight 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
- The interval estimation of population proportion
- The interval estimation of difference between two population proportions
- These estimation techniques to the practical problems