Subject	Statistics
Semester	04
Paper no	10
Paper Name	Testing of Hypothesis
Topic no	17
Topic name	Large Sample Test for Testing Single Proportion and Equality of Two Proportions
SME	Ms Shubharekha
ID	Ms Varsha Shetty

E-learning Module On Large Sample Test for Testing Single Proportion and Equality of Two Proportions

Learning Objectives

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

- Understand the concept of proportions
- Explain Large Sample Test for the Population Proportion
- Explain Large Sample Test for the Equality of Two Population Proportions
- Explain the practical application of the tests

Introduction

In this paper we will discuss the tests of significance using

- Large samples
- Samples from a population which is divided into two mutually exclusive classes

We have seen that for large values of **n**, the sample size, almost all the distributions like the Binomial, Poisson, Negative Binomial etc. are very closely approximated by Normal distribution.

Thus in this case we apply the Normal test which is based upon the following fundamental property of the Normal probability curve

$$X \sim N(\mu, \sigma^2)$$

then $$Z = \frac{X - \mu}{\sigma} = \frac{X - E(X)}{\sqrt{V(X)}} \sim N(0,1)$$

Attribute

- A qualitative characteristic which cannot be measured quantitatively
- Examples include honesty, beauty and intelligence

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 only say A and A′ with respect to an attribute.

Such a classification is termed as **dichotomous** classification.

Hence any sampling unit in the population may be placed in class A or A′ 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 and test the validity of the estimate.

For Example

- Proportion of defective items in a large consignment of such items
- Proportion of literates in a town
- Proportion of people who view a particular program
- Proportion of trees with a diameter of nine inches or more

Sometimes we cannot use averages.

For example:

- Percentage of people who vote for a communist
- Proportion of the population who is in a certain category

In such cases we need another method to estimate above characteristics.

Notations:

Let us suppose that a population with N units Y_1, Y_2....Y_N is classified into two disjoint and exhaustive classes A and A′ respectively with respect to a given attribute.

Let the number of individuals in the classes

$$A \Longrightarrow X$$

$$A' \Longrightarrow X'$$

such that $X+X' = N$

Then

P = The Proportion of units possessing the given attribute

= X/N

Q = The Proportion of units does not possess the given attribute

= X'/N = 1-P

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

Let us consider SRS sample of size **n** from this population.
Let 'x' is the number of units in the sample possessing the given attribute.

Then
p = Proportion of sampled units possessing the given attribute
= x/n

q = Proportion of sampled units which do not possess the given attribute
= 1 – p

The presence of an attribute may be termed as a success and its absence as failure.

In this case a sample of **n** observations is identified with that of a series of **n** independent Bernoulli trials with constant probability **p** of success for each trial.

Then the probability of **x** successes in **n** trials is given by the Binomial distribution as

$$p(x) = n_{C_x} p^x q^{n-x}, x = 0,1,2,.., n$$

Test for Single Proportion:

To test a null hypothesis

$$H_0: P = P_0$$

against an alternative hypothesis

$$H_1: P \neq P_0$$

when the sample size is very large

Let X be a number of units possessing an attribute (in the population).

Let P be the population proportion

Let $X \sim B(n, P)$

Then $E(X) = nP$

$V(X) = nPQ = nP(1-P)$

Let us take a sample of size **n** from the above population.

Let $\mathbf{p = x/n}$, be the sample proportion where $\mathbf{x}$ is the number of units possessing an attribute in the sample and $\mathbf{n}$ is the sample size

$$E(p) = E(x / n)$$
$$= [1/n] E(x)$$
$$= [1 / n] n P$$
$$= P$$

$$V(p) = V(x/n)$$
$$= [1/n^2]\, V(x)$$
$$= [1/n^2]\, n\, PQ$$
$$= PQ/n$$

As the sample size is large (as **n** tends to infinity) Binomial distribution tends to Normal distribution

Therefore

$$p \sim N(P, PQ/n)$$

$$Z = \frac{p - E(p)}{\sqrt{V(p)}} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \sim N(0,1)$$

We can apply a Normal test.

Under $\mathbf{H_0: P = P_0}$ the test statistic is

$$Z = \frac{p - P_0}{\sqrt{\frac{P_0 Q_0}{n}}}$$

Since the alternative hypothesis is two sided the test procedure is to reject the null hypothesis if

$$|Z| = \left| \frac{p - P_0}{\sqrt{\frac{P_0 Q_0}{n}}} \right| \geq Z_{\alpha/2}$$

One Sided Tests

To test a null hypothesis

$$H_0: P = P_0$$

against an alternative hypothesis

$$H_1: P > P_0$$

when the sample size is very large.

The test procedure is to reject the null hypothesis if

$$Z = \frac{p - P_0}{\sqrt{\frac{P_0 Q_0}{n}}} > Z_\alpha$$

To test a null hypothesis

$$H_0: P = P_0$$

against an alternative hypothesis

$$H_0: P < P_0$$

when the sample size is very large

The test procedure is to reject the null hypothesis if

$$Z = \frac{p - P_0}{\sqrt{\frac{P_0 Q_0}{n}}} < -Z_\alpha$$

Test of Significance for Difference of Proportions

To compare two distinct populations with respect to the prevalence of a certain attribute say **'A'** among their members.

For example:

- The difference in the proportions of Democrats and Republicans who will vote for Bush
- The difference in the proportions of defectives produced by employees of shift I and shift II

To estimate the difference in two proportions and test for the significant difference in population proportions (P_1- P_2) between two groups:

- We use the two sample proportions
- Analysis is centered at the difference of the two sample proportions

On the basis of independent random samples of sizes n_1 and n_2 from two Binomial populations it is desirable to estimate the difference between the parameters P_1 and P_2

For example: We may wish to estimate the difference between the proportion of voters in two different constitutes that favor a candidate X for a particular election.

To test the null hypothesis

$$H_0: P_1 = P_2 = P$$

against an alternative hypothesis

$$H_1: P_1 \neq P_2$$

for sample of large sizes

Let X_1 be a number of units possessing an attribute in the first population

Let P_1 be the population proportion of that population

Let $X_1 \sim B(n_1, P_1)$

Then,

$$E(X_1) = n_1 P_1$$

$$V(X_1) = n_1 P_1 Q_1 = n_1 P_1 (1 - P_1)$$

Let us take a sample of size n_1 from the above population.

Let,

$$p_1 = x_1 / n_1$$

be the sample proportion where x_1 is the number of units possessing an attribute in the sample of the first population and n_1 is the sample size

$$E(p_1) = E(x_1/n_1)$$
$$= [1/n_1]\ E(x_1)$$
$$= [1/n_1]\ n_1 P_1$$
$$= P_1$$

$$V(p_1) = V(x_1/n_1)$$
$$= [1/_{n1}{}^2]\ V(x_1)$$
$$= [1/_{n1}{}^2]\ n_1 P_1 Q_1$$
$$= P_1 Q_1 /\ n_1$$

Let X_2 be a number of units possessing an attribute in the second population.

Let P_2 be the population proportion of that population.

Let

$$X_2 \sim B(n_2, P_2)$$

Then,

$$E(X_2) = n_2P_2$$

$$V(X_2) = n_2P_2Q_2 = n_2P_2(1-P_2)$$

Let us take a sample of size n_2 from the above population.

Let $p_2 = x_2 / n_2$ be the sample proportion where x_2 is the number of units possessing an attribute in the sample of the second population and n_2 be the sample size.

$$E(p_2) = E(x_2 / n_2)$$
$$= [1 / n_2]\, E(x_2)$$
$$= [1 / n_2]\, n_2 P_2$$
$$= P_2$$

$$V(p_2) = V(x_2 / n_2)$$
$$= [1 / n_2^2]\, V(x_2)$$
$$= [1 / n_2^2]\, n_2 P_2 Q_2$$
$$= P_2 Q_2 / n_2$$

As the sample size is large (as **n** tends to infinity) Binomial distribution tends to Normal distribution

Therefore

$$p_1 \sim N\,(P_1,\ P_1Q_1 / n_1)$$

$$p_2 \sim N\,(P_2,\ P_2Q_2 / n_2)$$

$$p_1 - p_2 \sim N(P_1 - P_2,\ \ P_1Q_1/n_1 + P_2Q_2/n_2)$$

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}} \sim N\,(0,1)$$

Under $H_0: P_1 = P_2 = P$;

$$Z = \frac{(p_1 - p_2)}{\sqrt{PQ\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} \sim N(0,1)$$

Here the alternative hypothesis is 2 sided.
The test procedure is to reject the null hypothesis if

$$\left| \frac{(p_1 - p_2)}{\sqrt{PQ\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} \right| \geq Z_{\alpha/2}$$

Note:

In general we do not have any information as to the proportion of A's in the populations from which the samples have been taken.

Under $H_0: P_1 = P_2 = P$ (say) an unbiased estimate of the population proportion P based on the samples is given by

$$\widehat{P} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

The estimate is unbiased since

$$E(\hat{P}) = \frac{E(n_1 p_1 + n_2 p_2)}{n_1 + n_2} = \frac{n_1 E(p_1) + n_2 E(p_2)}{n_1 + n_2} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = P$$

because $P_1 = P_2 = P$ under H_0

One Sided Tests:

To test the null hypothesis

$$H_0: P_1 = P_2 = P$$

against an alternative hypothesis

$$H_1: P_1 > P_2$$

for sample of large sizes.

The test procedure is to reject the null hypothesis if

$$\frac{(p_1 - p_2)}{\sqrt{PQ\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} > Z_\alpha$$

To test the null hypothesis

$$H_0: P_1 = P_2 = P$$

against an alternative hypothesis

$$H_1: P_1 < P_2$$

for sample of large sizes.

The test procedure is to reject the null hypothesis if

$$\frac{(p_1 - p_2)}{\sqrt{PQ\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} < -Z_\alpha$$

For example:



In a locality survey of 100 adults over 40 years old a total of 15 people indicated that they participated in fitness activities at least twice a week.

Do these data indicate that the participant rate of adults over 40 years of age is significantly less than the 20% figure?

Here we are required to test the null hypothesis

H_0: P = 0.2 against H_1: P < 0.2

Sample proportion, p = x/n

=15 / 100

=0.15

P_0 =0.2

q_0 = 1 - 0.2 = 0.8

and the test statistic is

$$Z = \frac{p - P_0}{\sqrt{\frac{P_0 Q_0}{n}}} = -1.25$$

Suppose $\alpha = 0.05$ then $Z_{0.05} = 1.64$, from the standard Normal probabilities table.

We reject the null hypothesis if $Z < -Z_{\alpha}$

But in this case $Z = -1.25 > -1.64$

Hence we don't have enough evidence to reject the null hypothesis.

There is insufficient evidence to conclude that the percentage of adults over age 40 who participate in fitness activities twice a week is less than 20%

Statistical Significance and Practical Importance

It is important to understand the difference between results that are significant and the results that are practically important.

In statistical language the word significant does not necessarily mean important, but only that the results could not have occurred by chance.

From the test procedure we will learn to determine whether the results are statistically significant but when we use these procedures in a practical situation however we must also make sure the results are practically important.