TEST OF HYPOTHESIS

Defination of Hypothesis

A hypothesis is a tentative statement about the relationship between two or more variables. It is a specific, testable prediction about what you except to happen in a study.

Diagramming hypothesis:-

Diagramming hypotheses is a useful technique to help clarify your thinking.

Usually, a hypothesis takes the form 'X causes Y' or 'X is related to Y'

$$X \longrightarrow Y$$

$$X \longleftrightarrow Y$$

For example, the first hypothesis sated that above could be represented by a diagram as follows-

Financial resources $\xrightarrow{+ve}$ Adoption of new technology

The two variables or concepts are linked by an arrow going from one concept to the other. The arrow indicates that one variable (financial resources) does something to the other variable (adoption of new technology).

The plus sign indicates that the relationship is seen as positive, that is more of the one will lead to more of the other. Not all concepts ~~are~~ have a positive relationship.

Suppose, involving more than two concepts or variables for example:

Financial resources $\xrightarrow{+ve}$ Adoption of New Technology

Distance from Market $\xrightarrow{-ve}$ Adoption of New Technology

In this case two concepts, financial resources and distance from market, are related as independent concepts to the dependent concept, adoption of technology.

Here one of the independent concepts is positively related and other negatively related to the dependent concept.

There are endless possibilities. Most research proposals or projects deal only with one small area of the diagram. But it is often useful to make a diagram of more than you plan to study in order to show where your research fits into the larger frame of things and help you to indentify factors which may have to taken into account.

How is a hypothesis used in the scientific method

In the scientific method, in social science research involves the following steps:-

1. Forming a questions
2. Performing background research.
3. Creating hypothesis
4. Designing an experiment
5. Collecting data
6. Analyzing the results.
7. Drawing conclusions
8. Communicating the results.

The hypothesis is what the researchers predict the relationship between two or more variables, but it involves more than a guess. Most of the time, the hypothesis begins with a question which is then explored through background research. It is only at this point the researcher beign to develop a testable hypothesis.

Types of Statistical Hypotheses:

There are two types of statistical hypotheses:

1. Null Hypothesis:

The null hypothesis, denoted by H_0, is usually the hypothesis that sample observations result purely from change. The null hypothesis, always states that the treatment has no effect (no change, no difference). According to the null hypothesis, the population mean after treatment is the same as it was before treatment.

How to set up a Null hypothesis:

The following points may be followed in mind in setting the null hypothesis:-

(i) If we want to ~~so~~ test the significance of the difference between a statistic and parameter or between two sample statistics then set up the null hypothesis H_0 that the difference is not significant.

(ii) If we want to test any statement about the population, we set up the null hypothesis that it is true. For example, If the population mean has specified value μ_0, then $H_0: \mu = \mu_0$

(iii) To test the significance of any statistic 't' we compute the standardised variate

$$Z = \frac{t - E(t)}{S.E(t)} \quad [\text{for large sample test}]$$

For any statistic t, the values of $E(t)$ and $S.E(t)$ are invariably in terms of the population parameters. For example for statistic $t = \bar{x}$, $E(\bar{x}) = \mu$ and $S.E(\bar{x}) = \sigma/\sqrt{n}$ and for the statistic $t = p$ (sample proportion), $E(p) = P$ and $S.E(p) = \sqrt{PQ/n}$

Alternative hypothesis:

For any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis and it is denoted by H_1. It is very important to explicitly state the alternative hypothesis in respect of any null hypothesis H_0 because the acceptance or rejection of H_0 is meaningful only if it is being tested against a rival hypothesis. For example, if we want to test the null hypothesis that the population has specified mean (μ_0) i.e

$$H_0: \mu = \mu_0$$

Then the alternative hypothesis could be:

i) $H_1: \mu \neq \mu_0$ (i.e $\mu > \mu_0$ or $\mu < \mu_0$)

ii) $H_1: \mu > \mu_0$

iii) $H_1: \mu < \mu_0$.

The alternative hypothesis in

(i) is known as a two-tailed alternatives

(ii) is known as right-tailed alternatives

(iii) is known as left tailed alternatives.

The null hypothesis consists of only a single parameter value and is usually simple while alternative hypothesis is usually composite.

Scanned by CamScanner

Types of Error in Testing of hypothesis:

Type I error: A type I error occours when the researcher rejects a null hypothesis when it is true. The probability of committing a Type I error is called the significance level. This probability is also called alpha (α).

Type-II error: A Type-II error occurs when the researcher fails to reject a null hypothesis That is false. The probability of committing a Type-II error is called Beta (β). The probability of not committing a Type-II error is called the Power of the test.

In any test procedure, the four possible multually disjoint and exhustive decissions are:

i) Reject Ho when actually it is not true. i.e. when Ho is false.

ii) Accept Ho when it is true.

iii) Reject Ho when it is true.

iv) Accept Ho when it is false.

The decisions in (i) and (ii) are correct decissions While decissions in (iii) and (iv) are wrong decissions These decissions may be expressed in the following ~~dict~~ dichotomous table:

	Decision from Sample: Reject Ho	Asccept Ho
Ho (True)	Wrong (Type-I)	Correct
Ho (False) (H_1 True)	Correct	Wrong (Type-II error)

Then we can write,

$$P[\text{Reject } H_0 \text{ when it is true}] = P[\text{Type I Error}] = \alpha$$

$$P[\text{Accept } H_0 \text{ when it is wrong}] = P[\text{Type-II Error}] = \beta$$

Then α and β are also called the sizes of type-I and Type-II error respectively.

Level of Significance :

The maximum size of the type-I error, which we are prepared to risk is known as the level of significance.

$$P[\text{Rejecting } H_0 \text{ when } H_0 \text{ is true}] = \alpha.$$

Commonly used levels of significance in practice are 5% (0·05) and 1% (0·01). If we adpot 5% level of significance, it implies that 5 samples out of 100 we are likely to reject a correct H_0.
In other words this implies that we are 95% confident that our decission to reject H_0 correct. Level of significance is always fixed in advance before collecting the sample information.

Thus we can say that, 'α' is the level of significance and $(1-\alpha)$ is the degree of confidence which is called the confidence coefficent.

Critical Region :

Suppose we take several samples of the same size from the given population and compute some statistic t (say $\bar{x}$, p etc), for each of these samples. Let $t_1, t_2 \cdots t_k$ be the values of the statistic for these samples. Each of these values may be used to test some null hypothesis H_0. The statistics which lead to the rejection of H_0 give us a region called Critical Region (C) or Rejection Region (R) while those which lead to the acceptance of H_0 gives us a region called Acceptance Region (A).

Critical Values or Significant Value :

The value of test statistic which separates the critical (or rejection) region and the acceptance region is called critical value or significant value.

It depends on :

i) The level of significance used

ii) The alternative hypothesis, whether it is two-tailed or single tail test.

Now

$$Z = \frac{t - E(t)}{S.E(t)} \sim N(0,1)$$

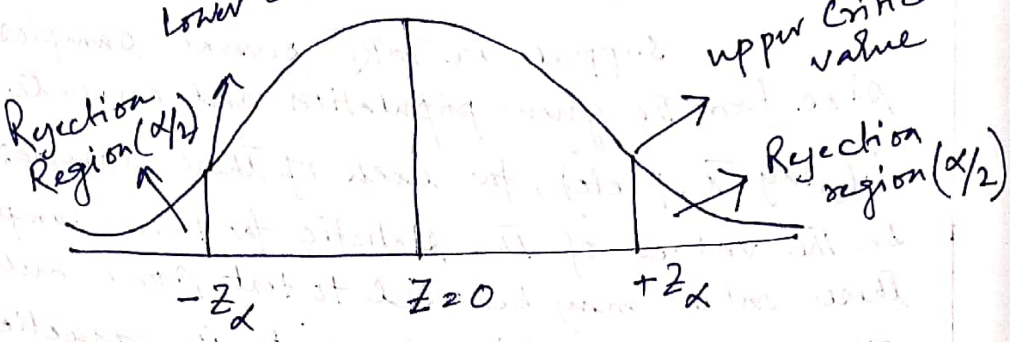
asymptotically, as $n \to \infty$, The value of Z is called test statistics under null hypothesis.

The critical value of the test statistic at level of significance (α) for two tailed test is given by Z_α where Z_α is determined by the equation

$$P(|Z| > Z_\alpha) = \alpha.$$

i.e Z_α is the value ~~of~~ so that total area of the critical region on both tails is α. Since normal probability curve is a symmetrical curve —

Two-Tailed Test (Level of Significance 'α')



from the normal distribution —

$$P(Z > Z_\alpha) + P(Z < -Z_\alpha) = \alpha.$$

$$\Rightarrow P(Z > Z_\alpha) + P(Z > Z_\alpha) = \alpha.$$

(By symmetry for the normal distribution)

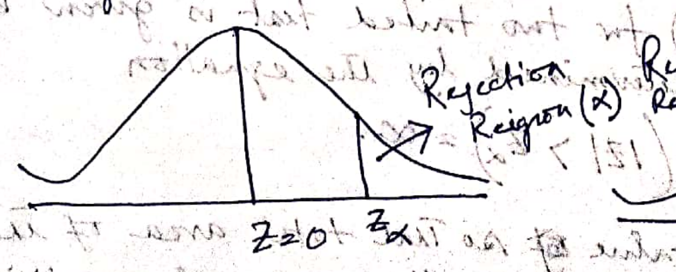
$$\Rightarrow 2P(Z > Z_\alpha) = \alpha.$$

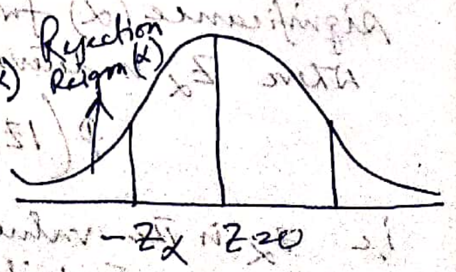
$$\therefore P(Z > Z_\alpha) = \alpha/2$$

i.e the area of each tail is $\alpha/2$. Thus Z_α is the value such that area to the right of Z_α and to the left of $-Z_\alpha$ is $\alpha/2$. (from the above figure).

In case of single-tail alternative, the critical value Z_α is determined so that the total area to the right of it (for right-tailed test) is a α and for left tailed test the total area to the left of $-Z_\alpha$ is —

Right-Tailed Test





Thus we can say that, the significant or critical value of Z for a single ~~test~~-tailed test (left or right) at the level of significance 'α' is same as the critical value of Z for a two-tailed test at the level of significance '2α'.

The following critical values of Z at commonly used of significance for both two-tailed and single-tailed tests.

Critical value of (Z_α) of Z

Critical values (Z_α)	Level of significance (α) 1%	5%	10%
Two-tailed Test	$\lvert Z_\alpha \rvert = 2.58$	$\lvert Z_\alpha \rvert = 1.96$	$\lvert Z_\alpha \rvert = 1.645$
Right-tailed Test	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$	$Z_\alpha = 1.28$
Left-tailed Test	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$	$Z_\alpha = -1.28$

If n is small, then the sampling distribution of the test Statistics Z will not be normal and in that case we can't use the above significant values, which have been obtained from the normal probability curves. In this case viz n small (usually less than 30), we use the significant values based on the exact smapling distribution of Z, which turns ~~to~~ out to be t and F.

Procedure for Testing of Hypothesis

We now summarise below the various steps in testing of a statistical hypothesis in a systematic maner:-

1. Null hypothesis: Set up the Null hypothesis (H_0)

2. Alternative hypothesis:
Set up the alternative hypothesis (H_1). This enables us to decide whether we have to use single-tailed or two-tailed test.

3. Level of significance:
Choose the appropriate level of significance (α). This should be decided before the sample is drawn.

4. Test Statistic or Test Critarion:
Compute the test statistic

$$Z = \frac{t - E(t)}{S.E(t)}$$ under Null hypothesis.

5. Conclusion:
The calculated value of 'Z' in step(4) is compared with the tabulated value of Z at the given level of significance(α).

If the calculated value of 'Z' (in absolute value) is less than the tabulated value, then we say that it is not significant and the null hypothesis is accepted. If the calculated value of 'Z' is more than the tabulated value, then Ho is rejected.

Scanned by CamScanner

Test for Single Proportion

The sampling distribution follows a binomial probability distribution. Hence its standard error is given by the formula $\sqrt{nPQ}$ under null hypothesis.

Where n = sample size

P = probability of Success for each trial

$Q = 1 - P$ = Probability of failure for each trial.

For large n, binomial distribution tends to normal distribution. Hence for large n,

$$Z = \frac{X - E(x)}{\sqrt{nPQ}}$$

Example 1:

In 324 throws of a six-faced die, odd points appeared 181 times. Would you say that die is fair?

Solⁿ:-

Let, the hypothesis be ~~stated~~ that the die is unbiased.

The probability of getting odd points $= \frac{1}{2}$

$\therefore$ Expected ~~probability~~ number of odd points in 324 throws $= 324 \times \frac{1}{2} = 162$

$$X - E(x) = 181 - 162 = 19$$

$$SE = \sqrt{nPQ} = \sqrt{324 \times \frac{1}{2} \times \frac{1}{2}} = 9$$

$$\therefore Z = \frac{X - E(x)}{\sqrt{nPQ}} = \frac{19}{9} = 2.11$$

Tabulated value of Z at 5% level of significance for two tail test is 1.96. Since calculated value is more than tabulated value, the hypothesis is rejected at 5% level of significance. Hence the die is biased.

Scanned by CamScanner

Example-2:

A die is thrown 9000 times and a throw of 3 or 4 is observed 3240 times. Show that the die cannot be regarded as an unbiased one.

Soln:

Let, the die is unbiased.

$\therefore P =$ probability of success

$=$ probability of getting 3 or 4

$= \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

$\therefore Q = 1 - P = 1 - \frac{1}{3} = \frac{2}{3}$

Sample Size $(n) = 9000$

$\therefore$ Expected number $= 9000 \times \frac{1}{3} = 3000$

$\therefore X - E(X) = 3240 - 3000 = 240$

$SE = \sqrt{nPQ} = \sqrt{9000 \times \frac{1}{3} \times \frac{2}{3}} = \sqrt{2000} = 44.73$

$$\therefore Z = \frac{X - E(X)}{\sqrt{nPQ}} = \frac{240}{44.73} = 5.36$$

Since $Z_{calculated} > Z_{tabulated}$ at 1%, 5% and 10% level of significances.

Hence the hypothesis is rejected at all levels and therefore the die is almost certainly biased.

Example- 3:

In a sample of 500 from a town, 280 are tea drinkers and the rest are coeffee drinkers. can we assume that coeffee and tea are equally popular in the town at 1% level of significance?

Solⁿ:

Let us take the hypothesis that tea and coffee are equally popular.

The expected frequency for tea and coffee drinkers would be $\frac{500}{2} = 250$.

$$SE = \sqrt{npQ}$$

$$= \sqrt{500 \times \tfrac{1}{2} \times \tfrac{1}{2}} = 11.2$$

$$\therefore Z = \frac{X - E(x)}{\sqrt{npQ}} = \frac{280-250}{11.2} = 2.68$$

Tabulated value of Z at 1% level of significance for two-tailed test is 2.58. Since calculated value is more than the tabulated value, the hypothesis is rejected at 1% level of significance. Hence tea and coffee are not equally popular.

Large Sample Theory:

The sample size (n) is greater than 30 ($n \geq 30$) it is Known as a large sample. For a large samples the sampling distributions of statistic are Normal (Z test).

Small Sample theory:

If the sample size (n) is less than 30 ($n < 30$). It is known as small sample. For small sample the sampling distributions are t, F and χ^2 distribution.

Test of significance:

* Test of significance for large sample.
* Large sample test or ~~Asym~~ Asymptotic test or Z test ($n \geq 30$)
* Test of significance for small samples ($n < 30$)

Sampling from attributes:

There are two types of test for attributes -

a) Test for single proportion

b) Test for equality of two proportions

Test for single proportion:

In a large sample size (n), we may examine whether the sample would have come from a population having a specified proportion $P = P_0$

We may proceed as follows:-

i) Null hypothesis (H_0):

H_0: The given sample would have come from a population with specified proportion $P = P_0$.

ii) Alternative Hypothesis (H_1):

H_1 : $P \neq P_0$ (Two Sided)

$P > P_0$ (one Sided - right tailed)

$P < P_0$ (one sided - Left tailed).

iii) Test statistic:

Follows the Z statistic with mean $\mu = 0$ and $\sigma^2 = 1$.

iv) Level of significance:

The level of significance may be fixed at either 5% or 1%.

v) Expected value or critical value:

In case of test statistic Z, the expected value is

$Z_e = 1.96$ at 5% level

$= 2.58$ at 1% level two tailed test

$Z_e = 1.65$ at 5% level

2.33 at 1% level of one tailed test.

vi) Inference:

If the observed value of the test statistic (Z_0) exceeds the table of (Z_e), we reject the Ho hypothesis.

Large Sample Tests for a Population Mean

In this section we describe and demonstrate the procedure for conducting a test of hypothesis about the mean of a population in the case sample size(n) is at least 30. The central Limit Theorem states that $\bar{X}$ is approximately normally distributed, Then has Mean $\mu_{\bar{X}} = \mu$ and stard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n}$ where μ and σ are the mean and the standard deviation of the population. This implies that the statistic

$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$ has the standard normal distribution.

Two situation may be arises :-

i) If we know σ Then the statistic is our test statistic.

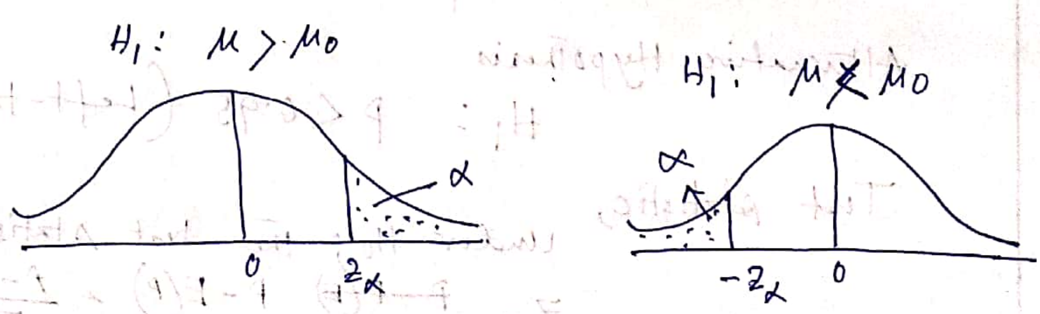
ii). If we donot know σ Then we replace it by the sample standard deviation(s).

Since the ~~lar~~ sample is large the resulting test statistic still has a distribution that is approximately standard normal.

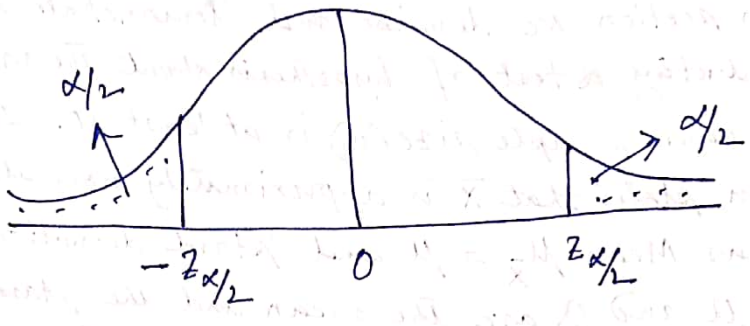
For a single population mean –

- If σ is Known : $Z = \dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$
- If σ is unknown : $Z = \dfrac{\bar{x}-\mu_0}{s/\sqrt{n}}$

The test statistic has the standard normal distribution. The distribution of the standardized test statistic and the corresponding rejection region for each form of the alternative hypothesis (left tailed, right-tailed or two sided tailed). This is shown in the following diagram.



$H_1: \mu \neq \mu_0$



(*) A manufacturer claimed that at least 95% of the equipment which he supplied to a factory conformed to ~~significant~~ specifications. An examination of a sample of 200 pieces of equipment ~~sea~~ revealed that 18 were faulty. Test his claim at a significant level of (i) 0.05 ; (ii) 0.01.

Sol^n:

We are given that,

$n = 200$

X = No. of pieces conforming to specifications in the sample = 200 − 18 = 182

Let p denote the probability of pieces conforming to specifications in the sample $= \frac{182}{200}$

$= 0.91$

Null Hypothesis:

$H_0 : P = 0.95$

i.e. the proportion of pieces conforming to specifications in the lot is 95%

Alternative Hypothesis

$H_1 : P < 0.95$ (Left-tail)

Test statistic,

Under H_0; the test statistic is

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

Now, $Z = \dfrac{0.91 - 0.95}{\sqrt{0.95 \times \frac{0.05}{200}}} = -2.6$

Conclusion:

i) At 5% level of significance, since the alternative ~~test~~ hypothesis is one-tailed (left). we shall apply left-tailed test for testing significance of Z. The significant value of Z at 5% level of significance for left-tail test is -1.645. Since the computed value is less than tabulated value, Then we can say that, Z is significant and we reject the null hypothesis at 5% level of significance. Hence the manufacturer's claim is refuted or rejected at 5% level of significance.

ii) Significance level at 1%:

The critical value of Z at 1% level of significance for single tailed (left tailed) test is -2.33. Since the computed value $Z = -2.6$ is less than -2.33 (i,e $|Z| > 2.33$), H_0 is rejected at 1% level of significance also.

Scanned by CamScanner

(*) In a big city 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smoker? [State the hypothesis clearly]

Ans:

We are given that, $n = 600$

No of smokers $= 325$

Let $p =$ sample proportion of smokers (Sample)

$$= \frac{325}{600} = 0.5417.$$

Null hypothesis (H_0):

H_0 : The number of smokers and non-smokers are equal in the city

so that, $P =$ Population proportion of smokers (Population) in the city

$$= \frac{1}{2} = 0.5.$$

$$\therefore Q = 1 - P = 1 - 0.5 = 0.5$$

Alternative Hypothesis (H_1): $P > 0.5$ (Right Tailed)

Test Statistic Under H_0; the statistic is

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

Since the sample is large.

Now $$Z = \frac{0.5417 - 0.5}{\sqrt{0.5 \times 0.5/600}} = 2.04$$

The significant value of Z for a right-tail test at 5% level of significance is 1.645. Since the calculated value of Z is greater than 1.645. It is significant and null hypothesis is rejected at 5% level of significance. Hence we conclude the majority of men in the city are smokers.

The significant value of Z for ~~raight~~ right-tail test is 2.33. Since the calculated value of Z is less than 2.33, it is not significant 1% level of significance and the null hypothesis may be accepted at 1% level of significance.

(*) In a sample of 400 parts manufactured by a factory, the number of defective parts was found to be 30. The company, however, claimed that only 5% of their product is defective. Is the claim tenable?

Sol^n:- We are given that, $n = 400$

X = No. of defectives in the sample = 30

Let, p = Proportion of defectives in the sample.

$$= \frac{X}{n} = \frac{30}{400} = 0.075$$

Null hypothesis (H_0):

$$H_0: P = 0.075 \Rightarrow Q = 0.95$$

Alternative Hypothesis (H_1):

$$H_1: P > 0.05 \text{ (Right-tailed alternative)}$$

Hence the test is one-tailed (Right-tailed) test.

Test statistics,

Under H_0 for large samples,

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

$$= \frac{0.075 - 0.050}{\sqrt{0.05 \times \frac{0.95}{400}}} = 2.27.$$

For right-tailed test, the significant value of Z at 5% level of significance is 1.645. Since the computed value of $Z = 2.27$ is greater than 1.645, it is significant at 5% level of significance and hence null hypothesis is rejected. However, we conclude that the company's claim of $P = 0.05$ is not tenable.

* Test of Significance for Difference of proportions

Let us compare two large populations, ie A and B with respect to the prevalence of a certain attribute among their members.

Let us suppose we take two independent large sample size n_1 and n_2 from the populations A and B respectively and let X_1 and X_2 be the observed number of successes in these sample respectively.

Then,

P_1 = Observed proportion of success in the sample from population $A = X_1/n_1$.

P_2 = Observed proportion of success in the sample from population $B = X_2/n_2$

If the population proportions are P_1 and P_2 then the corresponding sample proportions provide unbaised ~~estimator~~ estimates for them,

$$E(P_1) = P_1 \quad \text{and} \quad E(P_2) = P_2$$

$$\text{and} \quad Var(P_1) = \frac{P_1 Q_1}{n_1} \quad \text{and} \quad var(P_2) = \frac{P_2 Q_2}{n_2}$$

Since for large samples, P_1 and P_2 are normally distributed, their difference $P_1 - P_2$ is also normally distributed,

Then $t = P_1 - P_2$

$$\therefore E(t) = E(P_1) - E(P_2)$$

$$= P_1 - P_2$$

$$var(t) = Var(P_1 - P_2) = Var(P_1) + var(P_2)$$

because samples are indepedent,

$$Var(t) = \frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}$$

$$\therefore SE(t) = S.E(P_1 - P_2)$$

$$= \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$$

Hence, the standardised variable corresponding to $t = P_1 - P_2$ is given by —

$$Z = \frac{(P_1 - P_2) - E(P_1 - P_2)}{S.E(P_1 - P_2)}$$

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

(*) A machine puts out 16 imperfect articles in a sample of 500. After machine is overhauled it puts out 3 imperfect articles in a batch of 100. Has the machine improved.

We are given that, $n_1 = 500$ and $n_2 = 100$

P_1 = Proportion of defective in the 1st Sample

$= 16/500 = 0.032$

P_2 = Proportion of defective in the 2nd Sample

$= 3/100 = 0.03$

Null hypothesis:

$H_0 : P_1 = P_2$

i.e There is no significant difference in the machine before overhauling and after over hauling.

Alternative hypothesis:

$H_1 : P_2 < P_1$ or $P_1 > P_2$

Test statistic:

Under H_0, Since the sample are large, the test statistic is

$$Z = \frac{P_1 - P_2}{S.E\,(P_1 - P_2)} \sim N(0,1)$$

Now $SE\,(P_1 - P_2) = \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

Where $\hat{P} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{16+3}{500+100} = \frac{19}{600}$

$= 0.032$

$\hat{Q} = 1 - \hat{P} = 0.968$

$$S.E.(P_1 - P_2) = \sqrt{0.032 \times 0.968\left(\frac{1}{500} + \frac{1}{100}\right)}$$

$$= 0.0193$$

$$\therefore \; Z = \frac{0.032 - 0.030}{0.0193}$$

$$= 1.04$$

Since $Z < 1.645$ (Right tailed test), it is not significant at 5% level of significance. Hence we may accept the null hypothesis and conclude that the machine has not improved after overhauling.

(*) Before an increase in excise duty on tea, 400 people out of a sample of 500 persons were found to be tea drinkers. After an increase in duty, 400 people were ~~found~~ tea drinkers in a sample of 600 people. Using ~~under~~ standard error of proportion, state whether there is a significant decrease in consumption of tea.

Sol^n :- We are given that, $n_1 = 500$ and $n_2 = 600$.

P_1 = Sample proportion of tea drinkers before increase in excise duty $= \frac{400}{500} = 0.80$

P_2 = Sample proportion of tea drinkers after increase in excise duty $= \frac{400}{600} = 0.67$.

Null Hypothesis:

$H_0: P_1 = P_2$ i.e. there is no significant difference in the consumption of tea before and after the increase in excise duty.

Alternative Hypothesis: $H_1: P_1 > P_2 \Rightarrow (P_2 < P_1)$ i.e. Right-tailed test.

Test Statistic:

Under the null hypothesis the test statistic is

$$Z = \frac{P_1 - P_2}{S.E(P_1 - P_2)} = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim N(0,1)$$

Where
$$\hat{P} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{400+400}{500+600} = \frac{8}{11}$$

$$\therefore \hat{Q} = 1 - \hat{P} = 1 - 8/11 = 3/11.$$

$$\therefore \quad Z = \frac{0.80 - 0.67}{\sqrt{\frac{8}{11} \times \frac{3}{11}\left(\frac{1}{500}+\frac{1}{600}\right)}} = 4.81$$

Since $Z > 1.645$, as well as $Z > 2.33$ (Since test is one-tailed), it is highly significant at both 5% and 1% level of significance. Hence we reject the null hypothesis, Ho and Conclude that there is a significant decrease in the consumption of tea after increase in the excise duty.