Hypothesis Testing

12·1. Sampling—Introduction.

Before giving the notion of sampling we will first define *population*. In a statistical investigation the interest usually lies in the assessment of the general magnitude and the study of variation with respect to one or more characteristics relating to individuals belonging to a group. This group of individuals under study is called *population or universe*. Thus in statistics, population is an aggregate of objects, animate or inanimate, under study. The population may be finite or infinite.

It is obvious that for any statistical investigation complete enumeration of the population is rather impracticable. For example, if we want to have an idea of the average per capita (monthly) income of the people in India, we will have to enumerate all the earning individuals in the country, which is rather a very difficult task.

If the population is infinite, complete enumeration is not possible. Also if the units are destroyed in the course of inspection (*e.g.*, inspection of crackers, explosive materials, etc.), 100% inspection, though possible, is not at all desirable. But even if the population is finite or the inspection is not destructive, 100% inspection is not taken recourse to because of multiplicity of causes, *viz.*, administrative and financial implications, time factor, etc., and we take the help of *sampling*.

A finite subset of statistical individuals in a population is called a *sample* and the number of individuals in a sample is called the sample size.

12·2. Types of Sampling. Some of the commonly known and frequently used types of sampling are :

(i) *Purposive sampling,* (ii) *Random sampling,* (iii) *Stratified sampling,* (iv) *Systematic Sampling.*

Below we will precisely explain these terms, without entering into detailed discussion.

12·2·1. Purposive Sampling. Purposive sampling is one in which the sample units are selected with definite purpose in view. For example, if we want to give the picture that the standard of living has increased in the city of New Delhi, we may take individuals in the sample from rich and posh localities like Defence Colony, South Extension, Golf Links, Jor Bagh, Chanakyapuri, Greater Kailash etc. and ignore the localities where low income group and the middle class families live. This sampling suffers from the drawback of favouritism and nepotism and does not give a representative sample of the population.

12·2·2 Random Sampling. In this case the sample units are selected at random and the drawback of purposive sampling, *viz.*, favouritism or subjective element, is completely overcome. A *random sample* is one in which each unit of population has an equal chance of being included in it.

Suppose we take a sample of size n from a finite population of size N. Then there are ^NC_n possible samples. A sampling technique in which each of the ^NC_n samples has an equal chance of being selected is known as *random sampling* and the sample obtained by this technique is termed as a *random sample.*

12·2·3. Simple Sampling. Simple sampling is random sampling in which each unit of the population has an equal chance, say p, of being included in the sample and that this probability is independent of the previous drawings. Thus a simple sample of size n from a population may be identified with a series of n independent trials with constant probability 'p' of success for each trial.

Remark. It may be pointed out that random sampling does not necessarily imply simple sampling though, obviously, the converse is true. For example, if

an urn contains 'a' white balls and 'b' black balls, the probability of drawing a white ball at the first draw is $[a/(a + b)] = p_1$, (say) and if this ball is not replaced the probability of getting a white ball in the second draw is $[(a - 1)(a + b - 1)]$ $= p_2 \neq p_1$, the sampling is not simple. But since in the first draw each white ball has the same chance, *viz.*, $a/(a + b)$, of being drawn and in the second draw again each white ball has the same chance, *viz.*, $(a - 1)/(a + b - 1)$, of being drawn, the sampling is random. Hence in this case, the sampling, though random, is not simple. To ensure that sampling is simple, it must be done with replacement, if population is finite. However, in case of infinite population no replacement is necessary.

12·2·4. Stratified Sampling. Here the entire heterogeneous population is divided into a number of homogeneous groups, usually termed as *strata*, which differ from one another but each of these groups is homogeous within itself. Then units are sampled at random from each of these stratum, the sample size in each stratum varies according to the relative importance of the stratum in the population. The sample, which is the aggregate of the sampled units of each of the stratum, is termed as *stratified sample* and the technique of drawing this sample is known as *stratified sampling*. Such a sample is by far the best and can safely be considered as representative of the population from which it has been drawn.

12·3. Parameter and Statistic. In order to avoid verbal confusion with the statistical constants of the population, *viz.*, mean (μ), variance σ^2, etc., which are usually referred to as *parameters*, statistical measures computed from the sample observations alone, *e.g.*, mean ($\bar{x}$), variance (s^2), etc., have been termed by Professor R.A. Fisher as *statistics*.

In practice, parameter values are not known and the estimates based on the sample values are generally used. Thus statistic which may be regarded as an estimate of parameter, obtained from the sample, is a function of the sample values only. It may be pointed out that a statistic, as it is based on sample values and as there are multiple choices of the samples that can be drawn from a population, varies from sample to sample. The determination or the characterisaton of the variation (in the values of the statistic obtained from different samples) that may be attributed to chance or fluctuations of sampling is one of the fundamental problems of the sampling theory.

Remarks 1. Now onwards, μ and σ^2 will refer to the population mean and variance respectively while the sample mean and variance will be denoted by $\bar{x}$ and s^2 respectively.

12·3·1. Sampling Distribution of a Statistic. If we draw a sample of size n from a given finite population of size N, then the total number of possible samples is :

$$^{N}C_{n} = \frac{N\,!}{n\,!\,(N-n)\,!} = k, \quad (say).$$

For each of these k samples we can compute some statistic $t = t(x_1; x_2, \ldots, x_n)$, in particular the mean $\bar{x}$, the variance s^2, etc., as given below :

Sample Number	Statistics		
	t	$\bar{x}$	s^2
1	t_1	$\bar{x}_1$	s_1^2
2	t_2	$\bar{x}_2$	s_2^2
3	t_3	$\bar{x}_3$	s_3^2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	t_k	$\bar{x}_k$	s_k^2

The set of the values of the statistic so obtained, one for each sample, constitutes what is called the *sampling distribution* of the statistic. For example, the values $t_1, t_2, t_3, \ldots, t_k$ determine the sampling distribution of the statistic t. In other words, statistic t may be regarded as a random variable which can take the values $t_1, t_2, t_3, \ldots, t_k$ and we can compute the various statistical constants like mean, variance, skewness, kurtosis etc., for its distribution. For example, the mean and variance of the sampling distribution of the statistic t are given by :

$$\bar{t} = \frac{1}{k}(t_1 + t_2 + \ldots + t_k) = \frac{1}{k}\sum_{i=1}^{k} t_i$$

$$\text{Var}(t) = \frac{1}{k}\left[(t_1 - \bar{t})^2 + (t_2 - \bar{t})^2 + \ldots + (t_k - \bar{t})^2\right]$$

$$= \frac{1}{k}\sum_{i=1}^{k}(t_i - \bar{t})^2$$

12·3·2. Standard Error. The standard deviation of the sampling distribution of a statistic is known as its *Standard Error*, abbreviated as S.E. The standard errors of some of the well known statistics, *for large samples*, are given below, where n is the sample size, σ^2 the population variance, and P the population proportion, and $Q = 1 - P$, n_1 and n_2 represent the sizes of two independent random samples respectively drawn from the given population(s).

S.No.	Statistic	Standard Error
1.	Sample mean : $\bar{x}$	$\sigma/\sqrt{n}$
2.	Observed sample proportion 'p'	$\sqrt{PQ/n}$
3.	Sample s.d. : s	$\sqrt{\sigma^2/2n}$
4.	Sample variance : s^2	$\sigma^2\sqrt{2/n}$
5.	Sample quartiles	$1\cdot36263\ \sigma/\sqrt{n}$
6.	Sample median	$1\cdot25331\ \sigma/\sqrt{n}$

7.	Sample correlation coefficient (r)	$(1 - \rho^2)/\sqrt{n}$, ρ being the population correlation coefficient
8.	Sample moment μ_3	$\sigma^3 \sqrt{96/n}$
9.	Sample moment μ_4	$\sigma^4 \sqrt{96/n}$
10.	Sample coefficient of variation (v)	$\dfrac{v}{\sqrt{2n}} \sqrt{1 + \dfrac{2v^3}{10^4}} \simeq \dfrac{v}{\sqrt{2n}}$
11.	Difference of two sample means : $(\bar{x}_1 - \bar{x}_2)$	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$
12.	Difference of two sample s.d.'s : $(s_1 - s_2)$	$\sqrt{\dfrac{\sigma_1^2}{2n_1} + \dfrac{\sigma_2^2}{2n_2}}$
13.	Difference of two sample proportions $(p_1 - p_2)$	$\sqrt{\dfrac{P_1 Q_1}{n_1} + \dfrac{P_2 Q_2}{n_2}}$

Remark on the Utility of Standard Error. S.E. plays a very important role in the large sample theory and forms the basis of the testing of hypothesis. If t is any statistic, then for large samples

$$Z = \frac{t - E(t)}{\sqrt{V(t)}} \sim N(0, 1) \qquad (c.f. \ \S \ 12\cdot9)$$

$$\Rightarrow \qquad Z = \frac{t - E(t)}{S.E. \ (t)} \sim N(0, 1), \text{ for large samples.}$$

Thus, if the discrepancy between the observed and the expected (hypothetical) value of a statistic is greater than z_α (c.f. $\S$ 12·7·2) times its S.E., the null hypothesis is rejected at α level of significance. Similarly, if

$$| t - E(t) | \leq z_\alpha \times S.E. \ (t),$$

the deviation is not regarded significant at 5% level of significance. In other words, the deviation, $t - E(t)$, could have arisen due to fluctuations of sampling and the data do not provide us any evidence against the null hypothesis which may, therefore, be accepted at α level of significance. [For details see $\S$ 12·7·3]

(*i*) The magnitude of the standard error gives an index of the precision of the estimate of the parameter. The reciprocal of the standard error is taken as the measure of reliability or precision of the statistic.

$$\text{S.E. } (p) = \sqrt{PQ/n} \qquad [c.f. (4b) \ \S \ 12\cdot9\cdot1]$$

and
$$\text{S.E. } (\bar{x}) = \sigma/\sqrt{n} \qquad [c.f. \ \S \ 12\cdot2]$$

In other words, the standard errors of p and $\bar{x}$ vary inversely as the square root of the sample size. Thus in order to double the precision, which amounts to reducing the standard error to one half, the sample size has to be increased four times.

(*ii*) S.E. enables us to determine the probable limits within which the population parameter may be expected to lie. For example, the probable limits for population proportion P are given by

$$p \pm 3\sqrt{pq/n} \qquad (c.f. \text{ Remark } \S \ 12\cdot9\cdot1)$$

Remark. S.E. of a statistic may be reduced by increasing the sample size but this results in corresponding increase in cost, labour and time, etc.

12·4. Tests of Significance. A very important aspect of the sampling theory is the study of the *tests of significance*, which enable us to decide on the basis of the sample results, if

(*i*) the deviation between the observed sample statistic and the hypothetical parameter value, or

(*i*) the deviation between two independent sample statistics;

is significant or might be attributed to chance or the fluctuations of sampling.

Since, for large n, almost all the distributions, *e.g.*, Binomial, Poisson, Negative binomial, Hypergeometric (*c.f.* Chapter 7), t, F (Chapter 14), Chi-square (Chapter 13), can be approximated very closely by a normal probability curve, we use the *Normal Test of Significance* (*c.f.* § 12·9) for large samples. Some of the well known tests of significance for studying such differences for small samples are *t-test*, *F-test* and Fisher's *z-transformation*.

12·5. Null Hypothesis. The technique of randomisation used for the selection of sample units makes the test of significance valid for us. For applying the test of significance we first set up a hypothesis—a definite statement about the population parameter. Such a hypothesis, which is usually a hypothesis of no difference, is called *null hypothesis* and is usually denoted by H_0. According to *Prof. R.A. Fisher, null hypothesis is the hypothesis which is tested for possible rejection under the assumption that it is true.*

For example, in case of a single statistic, H_0 will be that the sample statistic does not differ significantly from the hypothetical parameter value and in the case of two statistics, H_0 will be that the sample statistics do not differ significantly.

Having set up the null hypothesis we compute the probability P that the deviation between the observed sample statistic and the hypothetical parameter value might have occurred due to fluctuations of sampling (c.f. § 12·7). If the deviation comes out to be significant (as measured by a test of significance), null hypothesis is refuted or rejected at the particular level of significance adopted (c.f. § 12·7) and if the deviation is not significant, null hypothesis may be retained at that level.

12·5·1. Alternative Hypothesis.

Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis, usually denoted by H_1. For example, if we want to test the null hypothesis that the population has a specified mean μ_0, (say), *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 alternative* and the alternatives in (*ii*) and (*iii*) are known as *right tailed* and *left-tailed alternatives* respectively. The setting of alternative hypothesis is very important since it enables us to decide whether we have to use a single-tailed (right or left) or two-tailed test [c.f. § 12·7·1].

Type I Error : *Reject H_0 when it is true.*

Type II Error : *Accept H_0 when it is wrong, i.e., accept H_0 when H_1 is true.*

If we write.

$$P\{\text{Reject } H_0 \text{ when it is true}\} = P\ \{\text{Reject } H_0 \mid H_0\} = \alpha \left.\vphantom{\begin{array}{c}1\\1\end{array}}\right\}$$
$$\text{and } P\ \{\text{Accept } H_0 \text{ when it is wrong}\} = P\{\text{Accept } H_0 \mid H_1\} = \beta \left.\vphantom{\begin{array}{c}1\\1\end{array}}\right\} \quad ...(12\cdot2)$$

then α and β are called the *sizes of type I error and type II error*, respectively.

In practice, type I error amounts to rejecting a lot when it is good and type II error may be regarded as accepting the lot when it is bad.

Thus
$$P\{\text{Reject a lot when it is good}\} = \alpha \left.\vphantom{\begin{array}{c}1\\1\end{array}}\right\}$$
and
$$P\{\text{Accept a lot when it is bad}\} = \beta \left.\vphantom{\begin{array}{c}1\\1\end{array}}\right\} \quad ...(12\cdot2a)$$

where α and β are referred to as *Producer's risk* and *Consumer's risk*, respectively.

12·7. Critical Region and Level of Significance.

A region (corresponding to a statistic t) in the sample space S which amounts to rejection of H_0 is termed as *critical region* or *region of rejection*. If ω is the critical region and if $t = t(x_1, x_2, ..., x_n)$ is the value of the statistic based on a random sample of size n, then

$$P\,(t \in \omega \mid H_0) = \alpha, \quad P\,(t \in \bar{\omega} \mid H_1) = \beta \qquad (12\cdot2b)$$

where $\bar{\omega}$, the complementary set of ω, is called the *acceptance region*.

We have $\quad \omega \cup \bar{\omega} = S \;$ and $\; \omega \cap \bar{\omega} = \phi$

The probability 'α' that a random value of the statistic t belongs to the critical region is known as the *level of significance*. In other words, level of significance is the size of the type I error (or the maximum producer's risk). The levels of significance usually employed in testing of hypothesis are 5% and 1%. The level of significance is always fixed in advance before collecting the sample information.

12·7·1. One tailed and Two Tailed Tests. In any test, the critical region is represented by a portion of the area under the probability curve of the sampling distribution of the test statistic.

A test of any statistical hypothesis where the alternative hypothesis is one tailed (right tailed or left tailed) is called a *one tailed test*. For example, a test for testing the mean of a population

$$H_0 : \mu = \mu_0$$

against the alternative hypothesis :

$$H_1 : \mu > \mu_0 \text{ (Right tailed)} \quad \text{or} \quad H_1 : \mu < \mu_0 \text{ (Left tailed)},$$

is a *single tailed test*. In the right tailed test $(H_1 : \mu_1 > \mu_0)$, the critical region lies entirely in the right tail of the sampling distribution of $\bar{x}$, while for the left tail test $(H_1 : \mu < \mu_0)$, the critical region is entirely in the left tail of the distribution.

A test of statistical hypothesis where the alternative hypothesis is two tailed such as :

$H_0 : \mu = \mu_0$, against the alternative hypothesis $H_1 : \mu \neq \mu_0$, $(\mu > \mu_0 \text{ and } \mu < \mu_0)$,

is known as *two tailed test* and in such a case the critical region is given by the portion of the area lying in both the tails of the probability curve of the test statistic.

In a particular problem, whether one tailed or two tailed test is to be applied depends entirely on the nature of the alternative hypothesis. If the alternative hypothesis is two-tailed we apply two-tailed test and if alternative hypothesis is one-tailed, we apply one tailed test.

For example, suppose that there are two population brands of bulbs, one manufactured by standard process (with mean life μ_1) and the other manufactured by some new technique (with mean life μ_2). If we want to test if the bulbs differ significantly, then our null hypothesis is $H_0 : \mu_1 = \mu_2$ and alternative will be $H_1 : \mu_1 \neq \mu_2$, thus giving us a two-tailed test. However, if we want to test if the bulbs produced by new process have higher average life than those produced by standard process, then we have

$$H_0 : \mu_1 = \mu_2 \quad \text{and} \quad H_1 : \mu_1 < \mu_2,$$

thus giving us a left-tail test. Similarly, for testing if the product of new process is inferior to that of standard process, then we have :

$$H_0 : \mu_1 = \mu_2 \quad \text{and} \quad H_1 : \mu_1 > \mu_2,$$

thus giving us a right-tail test. Thus, the decision about applying a two-tail test or a single-tail (right or left) test will depend on the problem under study.

12·7·2. Critical Values or Significant Values. The value of test statistic which separates the critical (or rejection) region and the acceptance region is called the *critical value* or *significant* value. It depends upon :

(*i*) The level of significance used, and

(*ii*) The alternative hypothesis, whether it is two-tailed or single-tailed.

As has been pointed out earlier, for large samples, the standardised variable corresponding to the statistic *t viz.* :

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

asymptotically as $n \to \infty$. The value of Z given by (*) under the null hypothesis is known as *test statistic*. The critical value of the test statistic at level of significance α for a two-tailed test is given by z_α where z_α is determined by the equation

$$P(\,|Z| > z_\alpha) = \alpha \qquad \ldots(12\cdot2c)$$

i.e., z_α is the value so that the total area of the critical region on both tails is α. Since normal probability curve is a symmetrical curve, from (12·2c), we get

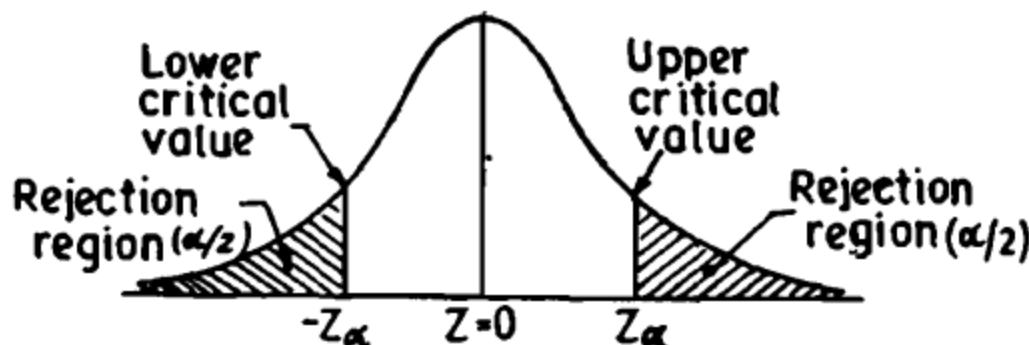$$P(Z > z_\alpha) + P(Z < - z_\alpha) = \alpha \qquad \text{[By symmetry]}$$

$$\Rightarrow \qquad P(Z > z_\alpha) + P(Z > z_\alpha) = \alpha$$

$$\Rightarrow \qquad 2P(Z > z_\alpha) = \alpha$$

$$\Rightarrow \qquad P(Z > z_\alpha) = \frac{\alpha}{2}$$

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 is $\alpha/2$ and to the left of $- z_\alpha$ is $\alpha/2$, as shown in the following diagram.
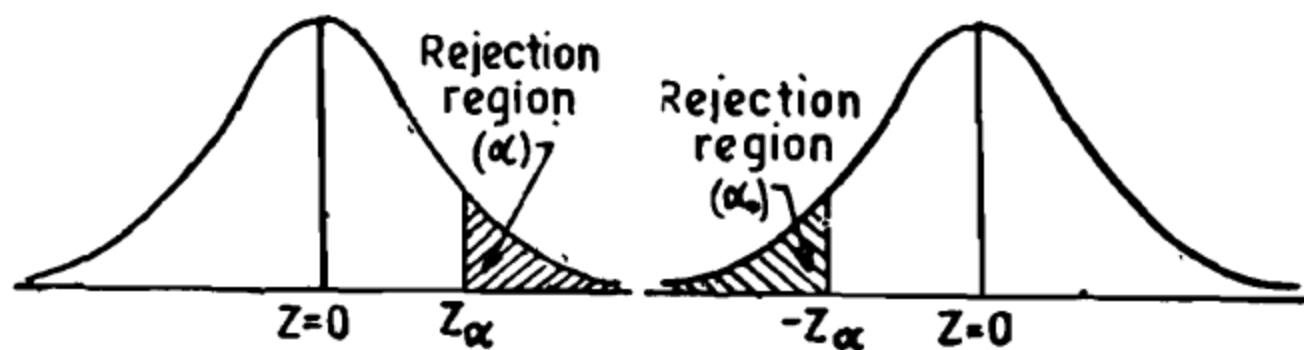
TWO-TAILED TEST
(Level of Significance 'α')

In case of single-tail alternative, the critical value z_α is determined so that total area to the right of it (for right-tailed test) is α and for left-tailed test the total area to the left of $-z_\alpha$ is α (See diagrams below), *i.e.,*

For Right-tail Test : $\qquad P(Z > z_\alpha) = \alpha$ $\qquad\qquad$ ...(12·2d)

For Left-tail Test : $\qquad P(Z < -z_\alpha) = \alpha$ $\qquad\qquad$ ...(12·2e)

RIGHT-TAILED TEST $\qquad\qquad$ LEFT-TAILED TEST
(Level of Signifiance 'α') $\qquad$ *(Level of Significance 'α')*

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

CRITICAL VALUES (z_α) OF Z

Critical Values (z_α)	Level of significance (α)		
	1%	5%	10%
Two-tailed test	$|Z_\alpha| = 2.58$	$|Z_\alpha| = 1.96$	$|Z_\alpha| = 1.645$
Right-tailed test	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$	$Z_\alpha = 1.28$
Left-tailed test	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$	$Z_\alpha = -1.28$

Remark. If n is small, then the sampling distribution of the test statistic Z will not be normal and in that case we can't use the above significant values, which have been obtained from normal probability curves. In this case, viz., n small, (usually less than 30), we use the significant values based on the exact sampling distribution of the statistic Z, [defined in (*), § 12·7·2], which turns out to be t, F, or χ^2 [see Chapters 13, 14]. These significant values have been tabulated for different values of n and α and are given in the Appendix at the end of the book.

12·7·3. Procedure for Testing of Hypothesis. We now summarise below the various steps in testing of a statistical hypothesis in a systematic manner.

1. *Null Hypothesis.* Set up the Null Hypothesis H_0 (see § 12·5, page 12·6).

2. *Alternative Hypothesis.* Set up the Alternative Hypothesis H_1. Ths will enable us to decide whether we have to use a single-tailed (right or left) test or two-tailed test.

3. *Level of Significance.* Choose the appropriate level of significance (α) depending on the reliability of the estimates and permissible risk. This is to be decided before sample is drawn, *i.e.*, α is fixed in advance.

4. *Test Statistic (or Test Criterion).* Compute the test statistic

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

under the null hypothesis.

5. *Conclusion.* We compare z the computed value of Z in step 4 with the significant value (tabulated value) z_α, at the given level of significance, 'α'.

If $|Z| < z_\alpha$, *i.e.*, if the calculated value of Z (in modulus value) is less than z_α we say it is not significant. By this we mean that the difference $t - E(t)$ is just due to fluctuations of sampling and the sample data do not provide us sufficient evidence against the null hypothesis which may therefore, be accepted.

If $|Z| > z_\alpha$, *i.e.*, if the computed value of test statistic is greater than the critical or significant value, then we say that it is significant and the null hypothesis is rejected at level of significance α *i.e.*, with confidence coefficient $(1 - \alpha)$.

12·8. Test of Significance for Large Samples.

In this section we will discuss the tests of significance when samples are large. We have seen that for large values of n, the number of trials, almost all the distributions, *e.g.*, 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 (*area property*) of the normal probability curve.

If $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X - \mu}{\sigma} = \dfrac{X - E(X)}{\sqrt{V(X)}} \sim N(0, 1)$

Thus from the normal probability tables, we have

$$P(-3 \leq Z \leq 3) = 0.9973, \quad i.e., \quad P(|Z| \leq 3) = 0.9973$$

$$\Rightarrow \quad P(|Z| > 3) = 1 - P(|Z| \leq 3) = 0.0027 \qquad \qquad ...(12.3)$$

i.e., in all probability we should expect a standard normal variate to lie between ± 3.

Also from the normal probability tables, we get

$P(-1.96 \leq Z \leq 1.96) = 0.95$ *i.e.,* $P(|Z| \leq 1.96) = 0.95$

$\Rightarrow \qquad\qquad\qquad P(|Z| > 1.96) = 1 - 0.95 = 0.05 \qquad\qquad\qquad ...(12.3a)$

and $\qquad\qquad\qquad P(|Z| \leq 2.58) = 0.99$

$\Rightarrow \qquad\qquad\qquad P(|Z| > 2.58) = 0.01 \qquad\qquad\qquad\qquad ...(12.3b)$

Thus the significant values of Z at 5% and 1% level of significance for a two tailed test are 1.96 and 2.58 respectively.

Thus the steps to be used in the normal test are as follows :

 (*i*) Compute the test statistic Z under H_0.

 (*ii*) If $|Z| > 3$, H_0 is always rejected.

 (*iii*) If $|Z| \leq 3$, we test its signficance at certain level of significance, usually at 5% and sometimes at 1% level of significance. Thus, for a two-tailed test if $|Z| > 1.96$, H_0 is rejected at 5% level of significance.

Similarly if $|Z| > 2.58$, H_0 is contradicted at 1% level of significance and if $|Z| \leq 2.58$, H_0 may be accepted at 1% level of significance.

From the normal probability tables, we have :

$$P(Z > 1.645) \quad = 0.5 - P(0 \leq Z \leq 1.645)$$
$$= 0.5 - 0.45$$
$$= 0.05$$
$$P(Z > 2.33) = 0.5 - P(0 \leq Z \leq 2.33)$$
$$= 0.5 - 0.49$$
$$= 0.01$$

Hence for a single-tail test (Right-tail or Left-tail) we compare the computed value of $|Z|$ with 1.645 (at 5% level) and 2.33 (at 1% level) and accept or reject H_0 accordingly.

Example 12·15. *A sample of 900 members has a mean 3·4 cms,. and s.d. 2·61 cms. Is the sample from a large population of mean 3·25 cms. and s.d. 2·61 cms. ?*

If the population is normal and its mean is unknown, find the 95% and 98% fiducial limits of true mean.

Solution. *Null hypothesis, (H_0)* : The sample has been drawn from the population with mean $\mu = 3.25$ cms., and S.D. $\sigma = 2.61$ cms.

Alternative Hypothesis, H_1 : $\mu \neq 3.25$ (Two-tailed).

Test Statistic. Under H_0, the test statistic is :

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1), \quad \text{(since } n \text{ is large)}$$

Here, we are given

$$\bar{x} = 3.4 \text{ cms.}, \ n = 900 \text{ cms.}, \ \mu = 3.25 \text{ cms. and } \sigma = 2.61 \text{ cms.}$$

$$Z = \frac{3.40 - 3.25}{2.61/\sqrt{900}} = \frac{0.15 \times 30}{2.61} = 1.73$$

Since $|Z| < 1.96$, we conclude that the data don't provide us any evidence against the null hypothesis (H_0) which may, therefore, be accepted at 5% level of significance.

95% *fiducial limits* for the population mean μ are :

$$\bar{x} \pm 1.96 \ \sigma/\sqrt{n} \implies 3.40 \pm 1.96 \times 2.61/\sqrt{900}$$

$$\implies \qquad 3.40 \pm 0.1705, \ i.e., \ 3.5705 \ \text{and} \ 3.2295$$

98% *fiducial limits* for μ are given by :

$$\bar{x} \pm 2.33 \ \frac{\sigma}{\sqrt{n}}, \ i.e., \ 3.40 \pm 2.33 \times \frac{2.61}{30}$$

$$\implies \qquad 3.40 \pm 0.2027 \ i.e., \ 3.6027 \ \text{and} \ 3.1973$$

Remark. 2.33 is the value z_1 of Z from standard normal probability integrals, such that $\quad P(|Z| > z_1) = 0.98 \implies P(Z > z_1) = 0.49.$

12·9·1. Test for Single Proportion. If X is the number of successes in n independent trials with constant probability P of success for each trial (*c.f.* § 7·2·1)

$$E(X) = nP \quad \text{and} \quad V(X) = nPQ,$$

where $Q = 1 - P$, is the probability of failure.

It has been proved that for large n, the binomial distribution tends to normal distribution. Hence for large n, $X \sim N(nP, nPQ)$ *i.e.*,

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - nP}{\sqrt{nPQ}} \sim N(0, 1) \qquad \ldots(12\cdot4)$$

and we can apply the normal test.

Remarks 1. In a sample of size n, let X be the number of persons possessing the given attribute. Then

Observed proportion of successes $= X/n = p$, (say).

$$\therefore \qquad E(p) = E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = \frac{1}{n} nP = P$$

$$\Rightarrow \qquad E(p) = P \qquad\qquad\qquad \ldots(12\text{-}4a)$$

Thus the sample proportion 'p' gives an unbiased estimate of the population proportion P.

$$\text{Also} \qquad V(p) = V\left(\frac{X}{n}\right) = \frac{1}{n^2} V(X) = \frac{1}{n^2} nPQ = \frac{PQ}{n}$$

$$\therefore \qquad S.E.(p) = \sqrt{PQ/n} \qquad\qquad\qquad \ldots(12\text{-}4b)$$

Since X and consequently X/n is asymptotically normal for large n, the normal test for the proportion of successes becomes

$$Z = \frac{p - E(p)}{S.E.\ (p)} = \frac{p - P}{\sqrt{PQ/n}} \sim N(0, 1) \qquad\qquad \ldots(12\text{-}4c)$$

2. If we have sampling from a finite population of size N, then

$$S.E.(p) = \sqrt{\left(\frac{N - n}{N - 1}\right) \cdot \frac{PQ}{n}} \qquad\qquad \ldots(12\text{-}4d)$$

3. Since the probable limits for a normal variate X are $E(X) \pm 3 \sqrt{V(X)}$, the probable limits for the observed proportion of successes are :

$$E(p) \pm 3 \text{ S.E. } (p), \text{ i.e., } P \pm 3 \sqrt{PQ/n}.$$

If P is not known then taking p (the sample proportion) as an estimate of P, the probable limits for the proportion in the population are :

$$p \pm 3 \sqrt{pq/n} \qquad \qquad \text{...(12-4}e)$$

However, the limits for P at level of significance α are given by :

$$p \pm z_\alpha \sqrt{pq/n}, \qquad \qquad \text{...(12-4}f)$$

where z_α is the significant value of Z at level of significance α.

In particular 95% confidence limits for P are given by :

$$p \pm 1 \cdot 96 \sqrt{pq/n}, \qquad \qquad \text{...(12-4}g)$$

and 99% confidence limits for P are given by

$$p \pm 2 \cdot 58 \sqrt{pq/n} \qquad \qquad \text{...(12-4}h)$$

Example 12·1. *A dice is thrown 9,000 times and a throw of 3 or 4 is observed 3,240 times. Show that the dice cannot be regarded as an unbiased one and find the limts between which the probability of a throw of 3 or 4 lies.*

Solution. If the coming of 3 or 4 is called a success, then in usual notations we are given

Solution. If the coming of 3 or 4 is called a success, then in usual notations we are given

$$n = 9,000; \quad X = \text{Number of successes} = 3,240$$

Under the null hypothesis (H_0) that the dice is an unbiased one, we get

$$P = \text{Probability of success} = \text{Probability of getting a 3 or 4} = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

Alternative hypothesis, $H_1 : p \neq \frac{1}{3}$, *(i.e., dice is biased).*

We have $Z = \dfrac{X - nP}{\sqrt{nQP}} \sim N(0, 1)$, since n is large.

Now $\quad Z = \dfrac{3240 - 9000 \times 1/3}{\sqrt{9000 \times (1/3) \times (2/3)}} = \dfrac{240}{\sqrt{2000}} = \dfrac{240}{44 \cdot 73} = 5 \cdot 36$

Since $|Z| > 3$, H_0 is rejected and we conclude that the dice is almost certainly biased.

Since dice is not unbiased, $P \neq \frac{1}{3}$. The probable limits for 'P' are given by :

$$\hat{P} \pm 3 \sqrt{\hat{P}\hat{Q}/n} = p \pm 3 \sqrt{pq/n},$$

where $\hat{P} = p = \frac{3240}{9000} = 0 \cdot 36$ and $\hat{Q} = q = 1 - p = 0 \cdot 64$.

Hence the probable limits for the population proportion of successes may be taken as

$$\hat{P} \pm 3 \sqrt{\hat{P}\hat{Q}/n} = 0 \cdot 36 \pm 3 \cdot \sqrt{\frac{0 \cdot 36 \times 0 \cdot 64}{9000}} \approx 0 \cdot 36 \pm 3 \times \frac{0 \cdot 6 \times 0 \cdot 8}{30 \sqrt{10}}$$

$$= 0 \cdot 360 \pm 0 \cdot 015 = 0 \cdot 345 \text{ and } 0 \cdot 375.$$

Hence the probability of getting 3 or 4 almost certainly lies between $0 \cdot 345$ and $0 \cdot 375$.

Example 12·2. A random sample of 500 pineapples was taken from a large consignment and 65 were found to be bad. Show that the S.E. of the proportion of bad ones in a sample of this size is 0·015 and deduce that the percentage of bad pineapples in the consignment almost certainly lies between 8·5 and 17·5.

Solution. Here we are given $n = 500$

X = Number of bad pineapples in the sample = 65

p = Proportion of bad pineapples in the sample = $\dfrac{65}{500} = 0.13$

$\therefore \qquad q = 1 - p = 0.87$

Since P, the proportion of bad pineapples in the consignment is not known, we may take (as in the last example)

$$\hat{P} = p = 0.13, \quad \hat{Q} = q = 0.87$$

S.E. of proportion = $\sqrt{\hat{P}\hat{Q}/n} = \sqrt{0.13 \times 0.87/500} = 0.015$

Thus, the limits for the proportion of bad pineapples in the consignment are :

$$\hat{P} \pm 3\sqrt{\hat{P}\hat{Q}/n} = 0.130 \pm 3 \times 0.015 = 0.130 \pm 0.045 = (0.085, 0.175)$$

Hence the percentage of bad pineapples in the consignment lies almost certainly between 8.5 and 17.5.