experimental result. Examples include age, sex and body weights of experimental animals (subject-relevant), pH, ionic strength and temperature of the medium or tissue preparation (situational-relevant), and the order of injection of different levels of a hypoglycemic agent (sequence-relevant). Relevant variables should be eliminated or kept constant as far as possible, to exclude their effects on the dependent variable.

## 2.4 Population and samples

Experiments are done with samples drawn from a specific population.

### 2.4.1. Population

For any experiment, the population consists of the entire aggregate of all such living organisms, inanimate objects, cases, events or phenomena as possess or exhibit at that time the specific dependent variable for the experiment or investigation. For example, for an experiment on the blood sugar of pancreatectomized rats, the population consists of all existing pancreatectomized rats on the earth; for working on $O_2$ consumption by a species of dragon flies, all insects of that species living anywhere at that time constitute the population.

Populations belong to two broad classes.

(i) *Infinite populations* are so immensely numerous and so widely dispersed that all the members cannot be reached or counted; e.g., the population of type I diabetic children or that of Jersey breed of cattle.

(ii) *Finite populations* consist of such small limited number of cases located within a given narrow area that all the members of such a population may be reached and counted to get its precise size; e.g., the population of pollutant-affected patients of Bhopal gas disaster, or that of a rare species of salamanders occurring in the waterbodies of a small area of Darjeeling district.

A population, whether finite or infinite, retains its identity with an identical size and unaltered properties, only so long as its members do not undergo any addition, deletion or any other change.

### 2.4.2. Samples

Because of the vast resources and long durations required for covering the entire population intended to be investigated, and also because of chances of unmitigated errors owing to accidental omission of some of its members from the study, an entire population is seldom subjected to any experiment or investigation. Instead, a small group of a limited number of individuals or cases, called a *sample,* is so chosen *at random* from the population by *laws of probability* as to be representative of that population with respect to the variable under investigation; such a representative

sample is then subjected to the intended investigation. The experimental observations of that investigation are then tested statistically to find their significance. If there is adequate probability of significance of the obtained result, the findings are sought to be generalized in the entire population.

Following criteria have to be gratified in a sample, if it is to be used in an investigation or experiment as the *representative* of the relevant population. (i) Individuals or cases must be included in the sample by being chosen at random from the population depending on laws of probability, so as to ensure the conformity between the population and the sample regarding the proportions of different types of cases. (ii) Variations of scores in the sample should closely conform to the variations of such scores in the population. (iii) Values of any statistic (e.g., mean and variance) of scores of different samples from the population should be so closely distributed that their arithmetic average may be identical with the corresponding population value or parameter (e.g., population mean). (iv) Scores of the variable should be distributed in the sample in conformity with their distribution in the population.

## 2.5 Sampling

Evidently, for generalizing the findings in a sample for the entire population, the sample should be representative of the latter. You have already learnt about the criteria to be fulfilled by a representative sample. These criteria depend largely on unbiased sampling. Some methods of sampling are briefly described below.

### 2.5.1. Judgement sampling

In this method, the investigator depends upon his personal judgement in considering some cases with specific properties as representing the population with respect to the intended dependent variable, and chooses arbitrarily some of such cases for inclusion in the sample. Such conscious or unintended subjective preference for some individuals or cases of particular types confines the sampling to only specific types of individuals, excluding other types of them from the chance of getting chosen for the sample. Such *judgement sampling* has a high probability of not drawing a representative sample from the population and is suitable neither for making inferences about the latter, nor for working out sampling errors of statistics computed from sample data.

### 2.5.2. Probability sampling

In this method, the choice of individuals from the population for inclusion in the sample is left entirely to mathematically devised methods of *random sampling* by *laws of probability*. No scope or role is left for the investigator or any other person to choose any case deliberately or arbitrarily; this minimizes the element of bias in

the sample. Instead, cases of different types have the probabilities of random choice commensurate with their respective frequencies or proportions in the population and independent of the choice of each other. Such probability sampling should yield samples, consisting of different types of cases in such relative proportions as in the population, being fairly representative of the population, and suffering from little or no bias. According to the population used and the intended purposes, probability sampling is designed in different ways.

(a) *Simple random sampling* :

If the sample has to be drawn from a *small, finite and homogeneous population,* not divided into distinct strata or sections, random sampling has to be done from the entire population taken together, choosing at random the requisite number of individuals for the sample successively out of all the individuals of the population. Thus, (i) each individual of the population enjoys an *identical probability of choice* at every step of sampling, (ii) each is chosen at random depending on the *laws of probability*, (iii) each gets chosen totally *independent* of the choice or omission of any other individual, (iv) no individual suffers from any subjective selection or rejection of any other individual, and (v) nor does any choice depend on any other quality or property. These should lead to the conformity between the proportions of different types of cases in the sample and those in the entire population.

In the unsophisticated *card drawal method* of simple random sampling, the sample size to be required for the experiment is first worked out statistically; all individuals of the population are then given successive serial numbers which are entered individually on separate cards; all those cards are mixed up in a container, and the requisite number of cards are next successively picked up blindly from that container. Individuals whose cards are so drawn are included in the sample.

Choices may be made for random sampling in two alternative ways. (i) In the *with-replacement method,* an individual once chosen is again included in the cases still left for subsequent choices and is, therefore, again considered for the subsequent choices. Thus, the probability of choice of each individual remains unaltered from choice to choice. However, it would create difficulty for the practical use of a sample if the same individual gets chosen more than once. (ii) In the *without-replacement method,* an individual once chosen is excluded from subsequent choices so that the probability of getting chosen rises progressively at successive choices; however, this rising difference in the probability of choice may be ignored as too small because of the much larger size of the population than the sample.

In a more scientific *random number method,* after giving identity numbers serially to all members of the population, individuals are chosen in the same order as the successive numbers, arranged at random in any chosen part of a random number table.

Simple random sample would not give a representative sample if a small sample is to be drawn from a stratified population with relatively low proportions of cases in one or more strata having different sizes and characters.

(b) *Stratified random sampling* :

If the population is *large and heterogeneous,* divided into distinct strata differing in properties and sizes, a proportional stratified random sampling is used in drawing a representative sample. This consists of the use of simple random sampling separately for each stratum. First, the required total size of the sample and the proportion of each stratum in the population are worked out. Next, simple random sampling is applied separately on each stratum to draw that number of individuals from it as corresponds to its proportional size in the population. All individuals of a stratum have an identical probability of getting chosen for the sample; but this probability varies from stratum to stratum according to their respective proportional sizes in the population.

For example, to draw a sample of 150 cases from a population having three strata (A, B and C) with the respective proportional sizes of 0.50, 0.40 and 0.10, simple random sampling should be undertaken separately from each stratum to draw respectively 150 × 0.50 or 75, 150 × 0.40 or 60, and 150 × 0.10 or 15 cases from A, B and C to constitute the sample.

(c) *Multistage sampling* :

A *vast population, dispersed over a wide area,* may be sampled by this method. Preferably depending on some pre-existing stages, the vast population is arranged stepwise into a number of levels, leading ultimately to the level of individuals. Simple random sampling is then applied at each of these levels. For example, to draw a sample of *Labeo rohita* fishes from waterbodies of West Bengal, three districts are chosen at random at the first stage out of all the districts; at the second stage, three waterbodies are chosen at random from all the waterbodies of the three chosen districts; finally, at the third stage, forty fishes are sampled at random from each of the three chosen waterbodies to constitute a sample of 120 fishes.

(d) *Fixed interval sampling* :

Sometimes, individuals of a population may arrive, occur or get naturally arranged in a systematic sequence; e.g., netting of successive butterflies from the air by an insect collector, angling of successive fishes in the fishing line of an angler, or arrival of successive patients at the out-patients department of a hospital. Fixed interval sampling consists of simple random sampling of cases depending on such a sequence of their random occurrence, appearance or arrival. To start with, any particular individual or case is chosen at random as the first one from the sequence of cases. Simultaneously, an interval is chosen at random as the gap between subsequent successive cases for the purpose of choices. Each subsequent case is next chosen as

it occurs in the given sequence after the preceding chosen one and separated from the latter by the chosen gap. For example, the fifth fish collected may be chosen as the first member of a sample and thereafter, every seventh fish is chosen maintaining a gap of six fishes between each pair of choices. This is continued until the requisite number of cases have been collected for the sample. However, this type of sampling may fail to yield an unbiased representative sample if the cases have been initially arranged in order of a characteristic related in any way to the variable to be investigated.

(e) *Purposive sampling* :

Random sampling is sometimes *deliberately restricted to a particular section* of the population so long as it is justifiable and logical to assume a truly representative nature of that section for the entire population, and the exclusion of its other sections is not anticipated to affect adversely the generalization of obtained results over the whole of the population.

(f) *Incidental sampling* :

In this method, random sampling is kept confined to a particular section or stratum of the population because of reasons like ready availability, easier manipulation and lower cost, instead of attempting to maintain or improve the representative nature of the sample. Such sampling should not be preferred for any investigation because it would seldom turn out a sample representative of the entire population.

## 2.6 Parameter and statistic

You may easily realize that as such, the experimental data consisting of one or more sets of numerical values or *scores* can hardly communicate much of precise and meaningful information or contribute much in comparing, analyzing and interpreting the observations. For these, the individual scores have to be presented, on one hand, in classified, tabulated or graphical forms while, on the other hand, some *summary values* like the mean and the standard deviation have to be worked out from those scores for further analysis and interpretation. While the presentation of data will be described in the next unit, you will be introduced in this unit to the basics of such summary values, also known as *numerical indices*, and will also get an initial idea of their roles in biostatistics.

### 2.6.1. Parameter

*Parameters* serve as measures of different characteristics of a variable in a population, and consist of numerous *summary values* like the mean and the variance, worked out from the scores of the entire population. Parameters of a population remain unchanged so long as the relevant population exists as such, but may differ from population to population. You are aware, however, that seldom do we work

with an entire population (see Sub-section 2.4.2); our investigations are generally undertaken with samples drawn from the population we want to study. Whenever we work with a sample, the summary values of the scores obtained from the latter are used as estimates of the respective parameters of the corresponding population. Two types of such estimates of parameters may be worked out : (i) a *point estimate* is a single summary value (*statistic*) of the sample, directly accepted as an estimate of the population parameter; e.g., a sample mean is the point estimate of the corresponding parametric or population mean; (ii) an *interval estimate* or *confidence interval* consists of a range of scores around a summary value (statistic) of a sample, within which the population summary value (parameter) has a given probability of occurrence; e.g., a 95% confidence interval has a probability of 0.95 for inclusion of the parametric mean.

### 2.6.2. Statistic

*Statistics* (singular : *statistic*) serve as measures of different characteristics of a variable in a sample, and consist of numerous summary values like the mean and the variance, worked out from the scores of that particular sample. As the individual scores vary from time to time in the same sample and also at the same instant from sample to sample drawn from the same population, any statistic varies *temporally* in the same sample and also *spatially* between samples of the same population. Consequently, a particular type of statistics (e.g., the sample mean $\bar{X}$) of different samples differ from the parameter (e.g., the population mean $\mu$) by different amounts called the *sampling errors* ($s_e$) : $s_e = \bar{X} - \mu$. Because of their different sampling errors, the statistics (e.g., $\bar{X}$) of samples lie dispersed around the parameter (e.g., $\mu$) of that population in the form of a *sampling distribution* with the parameter as its mean; e.g., a sampling distribution of sample means ($\bar{X}$) around the population mean ($\mu$). It also follows that the statistics of different samples from a population serve as *different point estimates* of the same population parameter. Statistics belong to different *classes* according to their purposes.

(a) *Descriptive statistics* :

These statistics of a variable measure and describe three different characteristics of a sample in respect of that variable. (i) *Statistics of location* such as mean, median and mode describe the location of a specific point — particularly a central one — of the distribution of the scores of a variable on the scale of the latter. (ii) *Statistics of dispersion* such as variance and standard deviation are the measures of scatter of the scores of a variable around a central point like the mean of the sample. (iii) *Statistics of correlation* or correlation coefficients measure the degree and direction of the association between two or more variables in the sample. Descriptive statistics belong to a particular sample and do not go beyond the limits of the latter.

Table 3.1. A qualitative frequency distribution of phenotypes in a sample of *Drosophila* flies.

| Phenotypes | Frequencies ($f$) | $f/n$ |
|---|---|---|
| Grey-body red-eye | 108 | 0.551 |
| Black-body red-eye | 40 | 0.204 |
| Grey-body scarlet-eye | 36 | 0.184 |
| Black-body scarlet-eye | 12 | 0.061 |
| Total | 196 ($n$) | 1.000 |

### 3.2.2. Quantitative frequency distribution

Both continuous and discontinuous measurement variables can be quantitatively measured and expressed in the form of numerical scores. The scores of such variables in a sample form a *range* which can be divided into groups called *class intervals.* Frequencies of scores belonging to different class intervals are then tabulated in the respective intervals to form a quantitative frequency distribution. The latter not only reveals characteristics of the sample in respect of the variable, but also contributes to the subsequent treatment and interpretation of the data. According to the nature of the variable concerned, quantitative frequency distributions may be either continuous or discontinuous.

1. *Continuous frequency distributions* :

*Continuous measurement variables* have such distributions. Here, the successive class intervals are continuous with each other, without any gap between the contiguous classes. The steps in forming such a frequency distribution are summarized below.

(a) The total *range* from the lowest to the highest score of the sample is divided nto a suitable number of *class intervals* of identical lengths ($i$) covering 3, 5, 7, 10 r 20 scores, separating the successive intervals by their *score limits*, and entered in olumn 1 of the frequency table (Table 3.2). The *interval length* ($i$) may be obtained s the difference between either the lower or the upper limits of the consecutive asses.

(b) To avoid gaps between class intervals, *true limits* or *class boundaries* are rked out, each as the mean of two contiguous score limits of successive intervals, d entered in column 2 of the table. (See Example 3.2.1.)

(c) The *midpoint* ($X_c$) of each interval is worked out as follows and entered in umn 3 of the table. In a frequency distribution with class intervals, all the cases n interval lose their individual identities and are deemed to possess the score of nidpoint ($X_c$). For any interval,

$X_c$ = lower score limit + $\frac{1}{2}$ [(higher score limit) – (lower score limit)].

(d) Each score of the sample is entered as a tally in column 4 of the table against its class interval.

(e) The total number of tallies of each interval is entered in column 5 as the frequency ($f$) of that interval.

(f) Frequencies of all the intervals are finally totalled to give the sample size ($n$) : $\Sigma f = n$.

2. *Discontinuous frequency distributions* :

A discontinuous or discrete measurement variable has such a distribution. Here, the class intervals are separated from each other by gaps because the scores of the variable can be in whole numbers only and not in fractional numbers. A discontinuous frequency distribution is worked out in the same way as a continuous one except that the step (b) for computing the true class limits and the column 2 for entering the latter are omitted to retain the intervening gaps between consecutive class intervals. (See Example 3.2.2.)

**Example 3.2.1.**

(a) Work out a frequency distribution of the following wing length scores (mm) of a sample of insects : 13, 19, 20, 25, 22, 23, 22, 21, 25, 12, 20, 17, 15, 19, 18, 20, 20, 19, 17, 21.

(b) Compute the mean, *SD* and *SE* of the mean of the data.

**Solution :**

(a) *Continuous frequency distribution* :

Highest score = 25. Lowest score = 12. $n = 20$. $i = 3$.

Range = (highest score – lowest score) + 1 = (25 – 12) + 1 = 14 scores.

Number of class intervals = $\frac{\text{range}}{i} = \frac{14}{3} = 4.7 \simeq 5$.

*True limits* are computed as averages of contiguous score limits of successive class intervals and entered in Table 3.2. For example, the true upper limit of the interval 15–17 as well as the true lower limit of the next interval 18–20 is computed as : $\frac{17+18}{2} = 17.5$.

*Midpoint* ($X_c$) of each interval is computed using the higher and lower score limits of that interval, and entered in Table 3.2. For example, for the interval 18–20,

$$X_c = 18 + \frac{20 - 18}{2} = 19.$$

(m) Probability ($P$) of random occurrence of a given $z$ and all other $z$ scores beyond it in *both tails* is known as the *two-tail probability*.

$P = 2$ [0.5000 – (area of unit normal curve from its $\mu$ to the given $z$)].

(n) The normal curve is *mesokurtic*, i.e., has a medium degree of peakedness.

(o) Probability distribution of a continuous measurement variable conforms to the normal distribution, if its scores are determined by the *random effects of many other variables* with no mutual interactions.

(p) Means of samples drawn from a normally distributed population are *distributed normally* around the parametric mean of the latter, forming a sampling distribution of means.

## 3.6 Skewness and kurtosis

These two properties of a distribution determine its form, shape and many other characteristics.

### 3.6.1. Skewness

Skewness is a measure of the degree and direction of *bilateral asymmetry* of a distribution. A symmetrical distribution, e.g., normal and $t$ distributions, has no skewness, has its two tails identically extended and equally pointed, and has its mean, median and mode coincide with the centre and peak of the distribution. But a skewed distribution is bilaterally asymmetric with one of its tails more extended

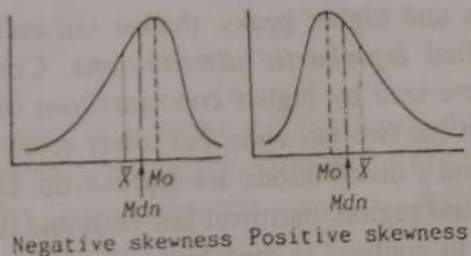


**Fig 3.6.** Skewed distributions. [From D.Das and A. Das, *Statistics in Biology and Psychology*, 4th ed., Academic Publishers, 2003.]

and pointed than the other tail. This results from the presence of more extreme scores in the extended or skewed tail than in the shorter and blunter tail; the scores are more concentrated in the blunter tail than in the skewed one. The skewness is called *positive* if the right or high-value tail is more drawn out than the left or low-value tail, while *negative* skewness consists of a more drawn-out and sharper left or low-value tail compared to the right tail. Poisson distributions are positively skewed

while binomial distributions are mostly either positively or negatively skewed.

*Coefficient of skewness* ($Sk$) is a measure of the magnitude and algebraic sign of skewness, indicating thereby the degree and direction of the skewness, respectively.

$$Sk = \frac{\bar{X} - Mo}{s}; \qquad Sk = \frac{3(\bar{X} - Mdn)}{s}.$$

### 3.6.2. Kurtosis

Kurtosis is a measure of *peakedness* of a ditribution. In assessing kurtosis, the normal distribution is used as the model. Being of a medium degree, its peakedness is known as *mesokurtosis*. Distributions like Student's $t$ distributions possess

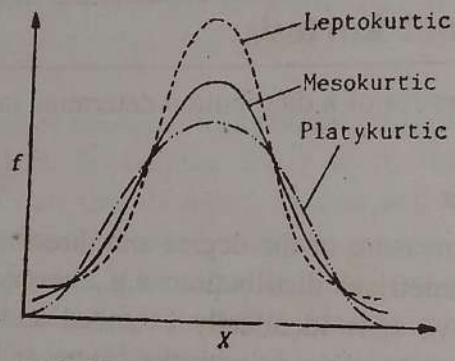


**Fig 3.7.** Different forms of kurtosis. [From D. Das and A. Das, *Statistics in Biology and Psychology,* 4th ed., Academic Publishers, 2003.]

comparatively sharper and higher peaks, thicker tail ends and thinner intervening regions, and are called *leptokurtic distributions*. Compared to mesokurtosis, leptokurtosis is characterized by higher concentrations of scores in a narrow zone around the peak and at the two tail-ends, but lower score concentrations in the area in between. Poisson and $t$ distributions are leptokurtic. Distributions which have a broader and flatter central region, narrower tail-ends and thicker intervening regions, are called *platykurtic*. In contrast to mesokurtosis, platykurtosis is characterized by lower score concentrations in the central region and at the tail-ends, but higher score density in the area in between (Fig. 3.7). Some binomial distributions are platykurtic while some are leptokurtic.

*Percentile coefficient of kurtosis* ($\kappa$) is a measure of kurtosis, worked out using 10th, 25th, 75th and 90th percentiles ($P_{10}$, $P_{25}$, $P_{75}$ and $P_{90}$) which are the scores below which the respective percentages of total scores occur in the sample.

$$\kappa = \frac{P_{75} - P_{25}}{2(P_{90} - P_{10})}.$$

In general, the null hypothesis proposes that the experimental result is not significant or meaningful, that it is the outcome of using a sample drawn at random by laws of probability, and that it would not be obtained if the entire population were used instead of the sample. Its elaborate statement, however, varies according to the assertions of diverse alternative hypotheses it contests, For example, where the $H_a$ proposes that there is a significant difference between two means (i.e., $\bar{X}_1 \neq \bar{X}_2$), the $H_o$ contends that there is *no significant difference* between those means (i.e., $\bar{X}_1 = \bar{X}_2$); but if the $H_a$ states that the mean of sample 1 is significantly higher than that of sample 2 (i.e., $\bar{X}_1 > \bar{X}_2$), the $H_o$ proposes that $\bar{X}_1$ is *not significantly higher* than $\bar{X}_2$ (i.e., $\bar{X}_1 \not> \bar{X}_2$). Where the $H_a$ proposes that there is a significant correlation between two given variables, the $H_o$ contends that there is *no significant correlation* between the two. If the $H_a$ states that the frequency distribution of phenotypes observed in the sample does not fit with Mendel's 9 : 3 : 3 : 1 distribution, the $H_o$ proposes that there is a *significant goodness of fit* between the observed distribution and the Mendelian distribution.

In any significance test of the obtained experimental result, the probability ($P$) of the correctness of $H_o$ is first worked out, and then compared with a chosen probability level called the *level of significance* ($\alpha$). If $P$ is found to exceed the $\alpha$, the $P$ is considered *too high;* so, the $H_o$ is then retained, the $H_a$ is rejected and the observed result is *not significant* ($P > \alpha$). But if $P$ is found to be either equal to or lower than the $\alpha$, the $P$ is considered *too low;* hence, the $H_o$ is then rejected, the $H_a$ is accepted and the observed result is *significant* (either $P = \alpha$ or $P < \alpha$).

The $H_o$ is bound to be tested whenever an experiment is performed with a sample; but the $H_o$ need not be considered or tested when the entire population is subjected to an experiment.

## 4.4 Levels of significance

A level of significance ($\alpha$) is that particular level of the probability ($P$) of correctness of $H_o$, which is compared with the $P$ worked out in a significance test for considering the rejection or acceptance of the $H_o$. It is that maximum level of probability, up to which the $P$ worked-out in the significance test is considered too low, and above which the worked-out $P$ is considered too high. In other words, if $P$ exceeds the chosen $\alpha$, the probability of the $H_o$ being correct is taken as too high so that the $H_o$ cannot be rejected and the observed result is considered *not significant* ($P > \alpha$). But whenever $P$ does not cross the chosen $\alpha$, i.e., whenever the computed $P$ is either equal to or lower than the $\alpha$, it is taken to be so low as to warrant the rejection of the $H_o$ ; the observed result is then considered *significant* ($P \leq \alpha$).

In biostatistics, 0.05, 0.02, 0.01 and 0.001 levels of $\alpha$ are usually used for

comparing with the worked-out $P$. The latter may be compared either with one particular level of $\alpha$ as chosen by the investigator, or successively with different levels of $\alpha$ in a descending order from 0.05. In the second case, the $H_o$ may be rejected and the result considered as significant at or below that lowest level of $\alpha$ which either exceeds or equals the computed $P$ ($P \leq \alpha$). For example; if $P$ is lower than 0.05, it is significant there, but is next compared with the next lower $\alpha$ of 0.02 to find if it is significant even there; this is repeated with successive lower levels of $\alpha$ until that *lowest* $\alpha$ is reached which exceeds or equals $P$. This process is preferable as the lower the $\alpha$ for significance, the lower is the probability of type I error of inference (Section 4.5). For example, if $P$ is lower than 0.01 ($P < 0.01$), then out of 100 such cases, the result may be wrongly considered significant in less than one of the cases; but if $P$ equals 0.001 ($P = 0.001$), in only one in 1000 such cases the result may be wrongly considered significant. Thus, the lower the $\alpha$ at or below which the result is considered significant, the fewer are the cases wrongly declared significant and consequently the lower is the probability of type I error.

## 4.5 Errors of inference

Whether or not the experimental result is considered significant, there are probabilities of errors of inference because the inference derived from either the rejection or the acceptance of the $H_o$ depends in both cases on the probabilities, $P$ and $\alpha$.

*Type I error of inference* results from the wrong rejection of a correct $H_o$, thus inferring an experimental result as significant when it is actually not significant. This error arises from the use of the *level of significance* ($\alpha$) in rejecting the correct $H_o$ and consequently has a probability identical with the $\alpha$ used in considering the computed $P$ as too low ($P \leq \alpha$); so, the probability of type I error has the symbol $\alpha$ identical with that of the level of significance. It follows that the probability of type I error may be lowered by using a lower level of significance in comparing with the probability ($P$) worked out in the significance test. Thus, if $P$ equals the $\alpha$ of 0.05 ($P = 0.05$), there is a probability of 0.05 for the type I error — out of 100 such cases, such results of any 5 cases would actually be not significant, having resulted from mere random sampling; but if the $H_o$ is rejected because $P$ equals 0.01 ($P = 0.01$), there is a much lower probability of 0.01 of the type I error. (See also the last paragraph of Section 4.4 and Sub-section 4.7.2.)

*Type II error of inference* is the opposite of the type I error. It is the error resulting from the wrong acceptance of a wrong $H_o$, thus leading to a wrong inference that the experimental result is not significant when the latter is actually significant. The probability of type II error ($\beta$) has a *reverse relation* with that of the type I error

the observed difference may as well have arisen from the difference between the sampling errors ($s_e$) of the two means, consequent upon random sampling — there might not have been any difference between the means if the levels of the independent variable would have been applied on the entire population. Such a probability would always persist so long as samples are used instead of the population, whatever precision and caution be used in sampling to make the samples truly representative of the population. You would be right to guess that the result of any such experiment using samples would be open to two alternative inferences. One, the obtained result is not meaningful, i.e., *not significant,* has come from chances associated with random sampling, would not have occurred if the population were used instead of samples, and can thus be explained away by sampling errors; the other, the observed result is meaningful and *significant,* is not the outcome of chances of random sampling, and cannot be explained away by sampling errors. To infer which of these two alternatives may be upheld, a *significance test* has to be undertaken to find statistically whether the probability of the observed result occurring by chance is too high or too low. If this probability is *too high*, it is inferred that the obtained result of the experiment is *not significant* and not fit for generalization in the entire population; on the contrary, if the probability of its chance occurrence is *too low*, the result under consideration is *significant* or meaningful and can be generalized in the population.

For a significance test, in most cases, the experimentally obtained result (e.g., a difference between means, a correlation coefficient, etc.) is first transformed into a standard score (e.g., $z$, $t$, $\chi^2$ and $F$) and the latter is referred to the corresponding probability distribution (e.g., normal, $t$, $\chi^2$ or $F$ distribution) to find the probability of its chance occurrence. To judge whether that probability is too high or too low, it is compared with a chosen probability level called the *level of significance* ($\alpha$). You will learn in Sub-sections 4.6.2, 4.7.3, 4.7.4, 5.3.3, 5.5.1, 6.3.4 and 6.4.3 as also in Section 5.6, about significance tests for a number of computed statistics.

## 4.3 Null hypothesis

Each experiment or investigation is intended, designed and performed to substantiate or prove a proposed conjecture called the experimental hypothesis; the latter is generally known in statistics as the *alternative hypothesis* ($H_a$) because it is the alternative to and is contested by another hypothesis ($H_o$) which would be subjected to a significance test. The $H_o$ is called the *null hypothesis* because it contradicts, contests and tries to negate or nullify the assertion of the alternative hypothesis. The *testing of hypothesis* consists basically of the working out of probability of correctness ($P$) of null hypothesis and finding whether that probability is too high or too low.