THE USE OF BIOSTATISTICS IN
THE EVALUATION OF
BIOMEDICAL LITERATURE

PHCL 456
Advanced Drug and Poison Information Services

Dr. Hisham S. Abou-Auda

Department of Clinical Pharmacy
College of Pharmacy
King Saud University

Pharmacists sometimes use biostatistics to describe the results of an experiment or an investigation. This process is referred to as **DATA ANALYSIS** or **DESCRIPTIVE STATISTICS**. We also use statistics another way: if the entire **POPULATION** is not accessible to us when we conduct clinical trials, we often observe only a portion of the population (called **SAMPLE**) and use statistics to answer questions about the whole population. This process is called **INFERENTIAL** STATISTICS. The inferences we make are either decisions or estimates about the population.
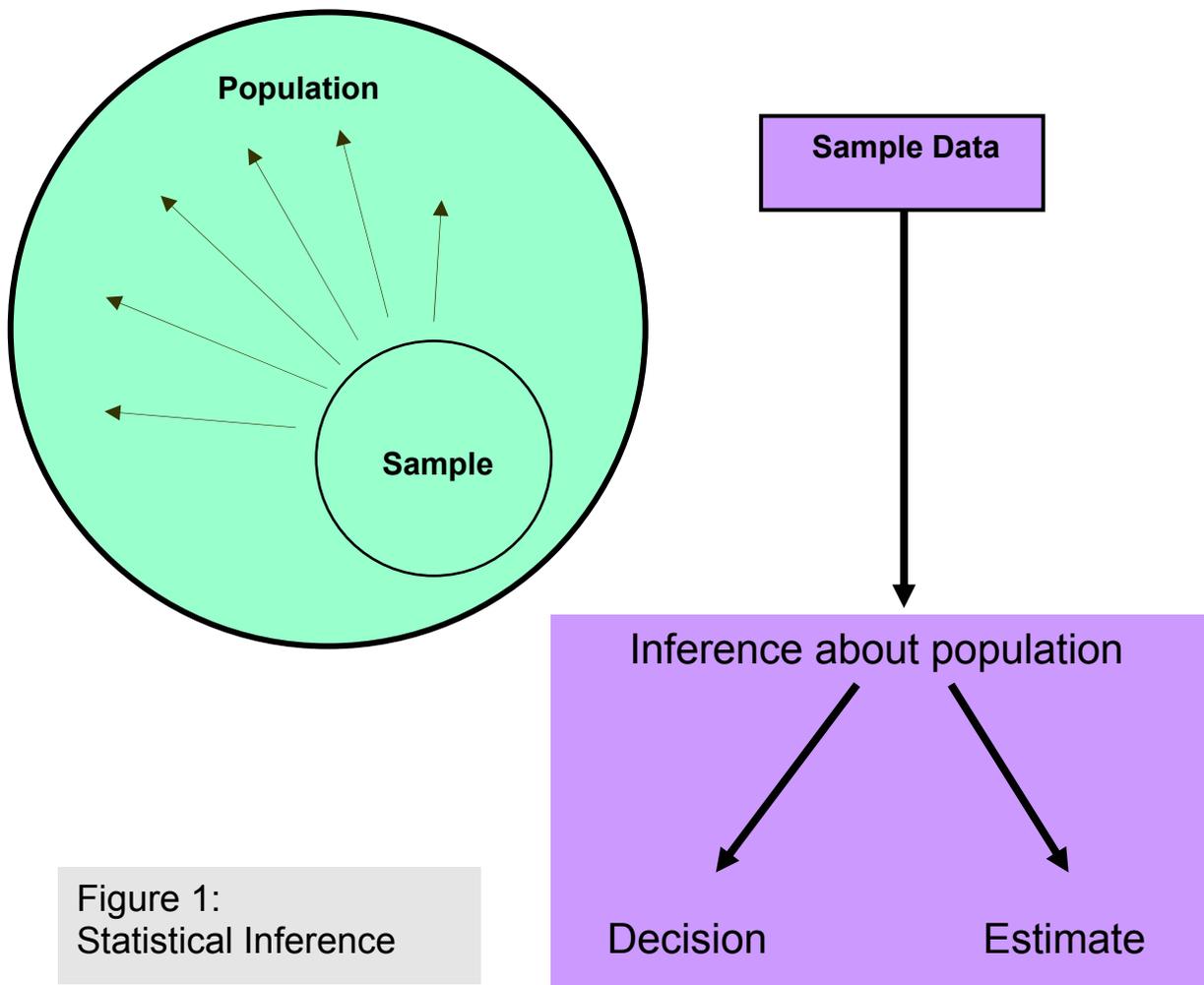

Figure 1:
Statistical Inference

SOURCES AND REPRESENTATION OF DATA

The main sources for collection of medical and pharmaceutical statistics are:

1) Experiments (clinical trials,...etc)

2) Surveys and questionnaires

3) Records (patient records, toxicology records, etc.)

The first and second generate DATA as needed for specific purposes while the records provide ready-made data for routine and continuous information.

The data collected may be analyzed to assess changes in health or disease situations in the community or population by standard parameters. The statistical data obtained from the above sources are of two types:

1. QUALITATIVE (or Discrete) DATA:

They are obtained by counting the individuals having the same characteristic or attribute and not by measurement. There is only one variable; i.e., the number of persons and not the characteristic. Person with the same characteristic are counted to form specific groups or classes such as healthy, sick,..., dead, alive,....,feeling mild pain, moderate pain or severe pain,...,treated, not treated, male, female,...etc. The characteristic such as feeling pain or being treated by a drug is not variable but the frequency of persons treated or feeling pain, alone varies. By one line of treatment 20 survive out of 25 while on the other line 15 may survive out of 25. The characteristic; i.e., survival, remains the same but the number of frequency of survivors varies. Qualitative data are

discrete in nature such as number of deaths in different years, population of different towns, persons with different blood groups in population, and so on.

In biomedical studies such data are mostly collected in pharmacology to find the action of a drug, in clinical practice to test or compare the efficacy of a drug, vaccine, operation or line of treatment, and in demography to find births, deaths, still births, etc. The statistical methods commonly employed in analysis of such data are mostly nonparametric tests such as chi (χ^2) square test (will be explained later)**.**

2.   QUANTITATIVE (or continuous) DATA:

They are obtained by measurement of a character or variable such as time, height, weight, etc. Quantitative data are also called continuous as each individual has one measurement from a continuous spectrum or range; such as pulse rate 70, 73, 74, 80 or height 150 cm, 175 cm, 185 cm. etc. In statistical language any character that varies is called a variable.

In such data, there are two variables, the characteristic such as height and the frequency; i.e., the number of persons with the same characteristic and in the same range. Height varies from person to person, it may be 150 cm in one and 160 cm in another person of the same age and sex. Number of persons with 150 cm or in the range 150 to 152 cm; may be 10 while those with height 160 cm or in the range of 160 to 162 cm; may be 20. Thus we find the characteristic as well as the frequency both vary from person to person as well as from group to group.

In biomedical studies, such statistics are mostly collected to define the normal or to find the limits of deviation from the normal in healthy persons. When

the measurement or counting crosses the normal limits, it becomes unusual and may indicate pathology. Statistical analysis of such data is usually performed by parametric methods because statistical parameters (such as the mean, range, standard deviation, coefficient of variation, etc) can be derived from this type of data.
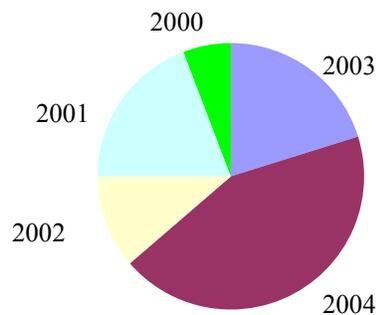
PRESENTATION OF DATA

The student is expected to review and understand <u>on his own</u> the following types of data presentation:

A) <u>Presentation of Qualitative data:</u>

1. **Bar graph**
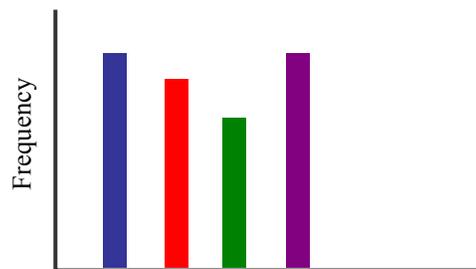
2. **Pie or sector diagram**

3. Pictogram or picture diagram

An attractive alternative to standard bar charts is to replace the bars with pictorial images. There are three ways in which the pictures can be used to represent the data:

- varying both the length and width of the picture - and hence its **area**;

- keeping the width of the picture constant and **stretching** the length;

- keeping the width and length constant but **stacking** multiple copies of the picture.
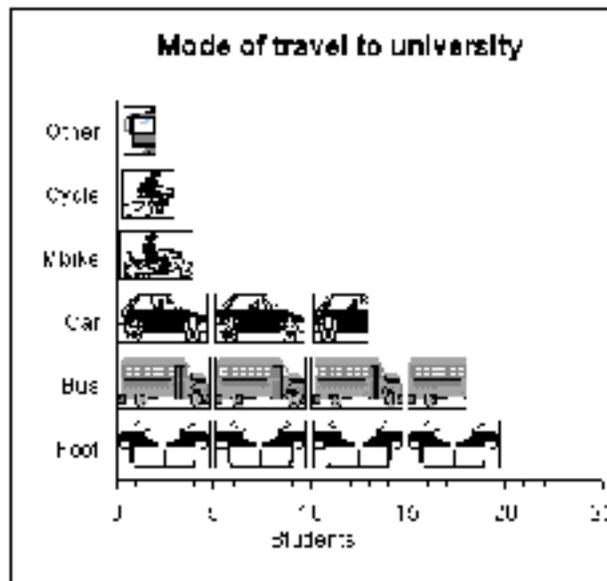
4. Map diagram or spot map

B) Presentation of Quantitative data:

1. Histogram
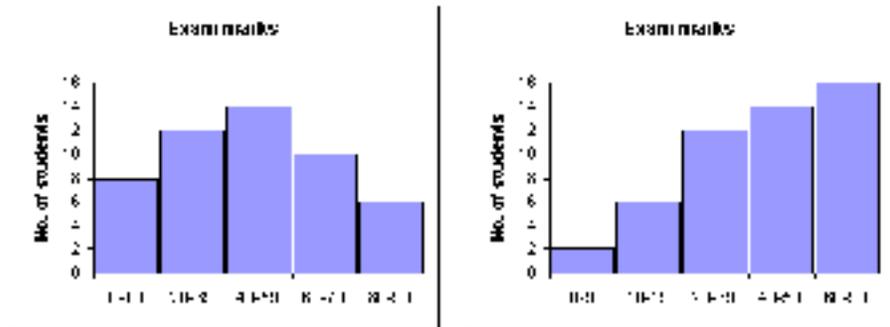
2. Frequency polygon

3. Frequency curve

4. Line chart or graph

Exercise and Pulse Rate

Pulse Rate in Beats Per Minute

Time in Minutes
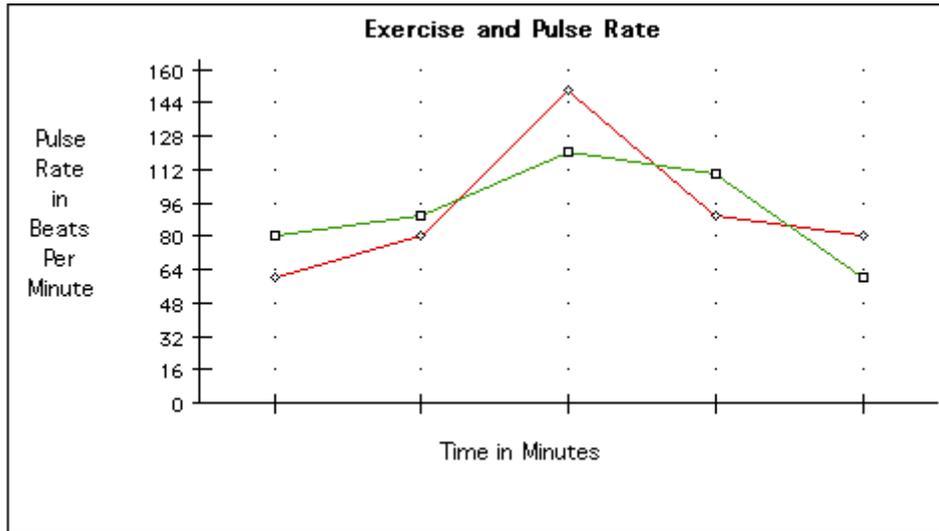
5. Cumulative frequency diagram
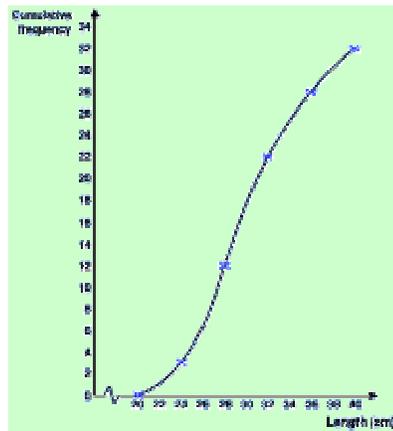
6. Scatter or dot diagram.
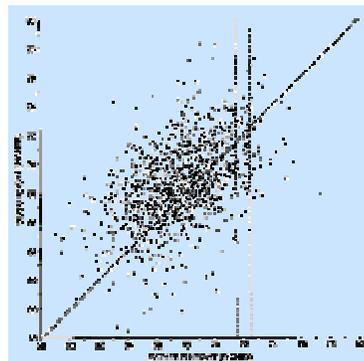
<u>**Statistical analysis can be classified into two types:**</u>

1. Descriptive statistics:

    a.   Measures of central tendency

    b.   Measures of variability

2. Inferential statistics:

    a.   Parametric test

    b.   Nonparametric tests

Notation:

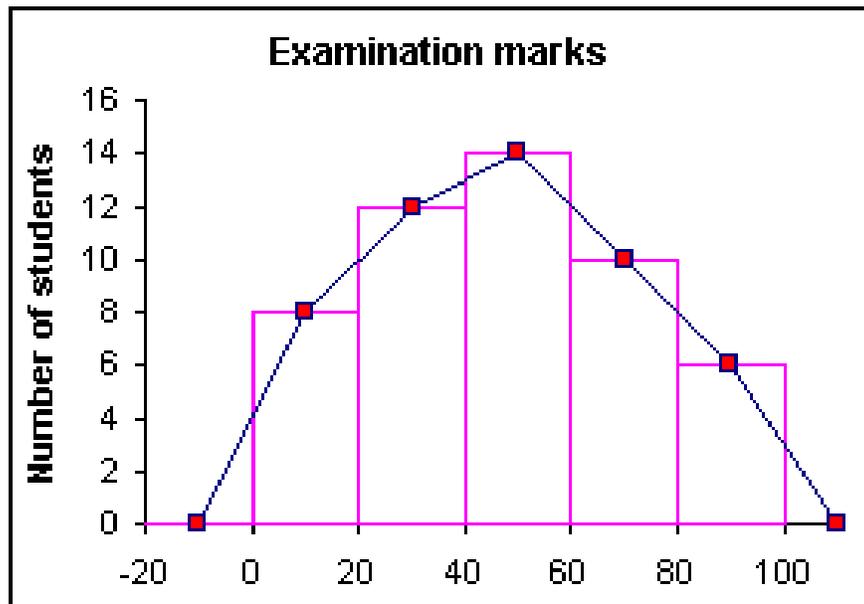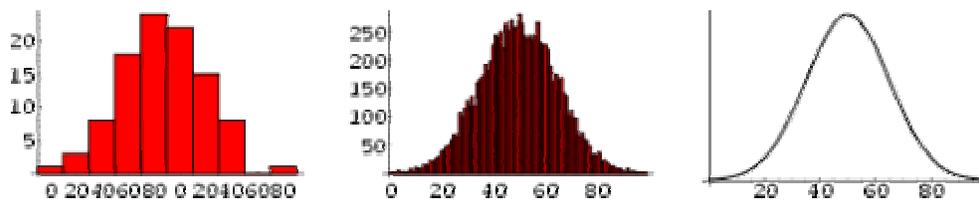**N or n** = Sample size (Number of counts or individuals etc..)

$\sum$ (sigma) = Summation of results or observations

X = A quantity that varies within limits such as height, weight, blood pressure, age, etc. It is denoted as X used for ordered series such as X_1, X_2, ..., X_n.

$\Sigma X \quad = X_1 + X_2 + X_3 + ... + X_n$

$\Sigma X^2 \quad = X_1^2 + X_2^2 + X_3^2 + ... + X_n^2$

$\Sigma XY \quad = X_1.Y_1 + X_2.Y_2 + X_3.Y_3 + ... + X_n.Y_n$

Definitions

<u>**Population**</u>:

It is an entire group of people. It has to be fully defined such as all human beings; all women 15-45 years of age or only married women; all patients with pulmonary tuberculosis; all infants; and so on. Such a population invariably gives qualitative data. It is finite or limited in number and can be easily counted.

A statistical population may also be birth weights, hemoglobin levels,

number of RBC's in the human body, blood cholesterol levels, drug concentrations in all patients receiving a certain treatment, etc. Such a population mostly gives quantitative data. It is infinite, indefinite or unlimited in number that cannot be easily counted. In other words, a population includes all observations that could be made on a specific characteristic. We cannot hope to ever "know" the true parameters of the population (e.g., mean, standard deviation, etc.), the best we can do, therefore, is to estimate them using the procedure of statistical inference from a representative sample.

Sample:

It is a part of population, generally selected so as to be representative of the population whose variables are under study. A randomly selected sample may be defined as a sample that is drawn in such a way as to ensure that (1) each member of the population has the same opportunity to be selected, and (2) each member of the population has the opportunity to be selected independently of any other member. The terms "**random**" and "representative" are thought to be synonymous, when in fact they are not. Random sampling increases the chances that a sample will be unbiased. A biased sample is not representative.

Parameter:

It is the summary value of a variable that describes the population such as mean, standard deviation, variance, correlation coefficient, etc.

Statistic:

It is the summary value such as mean, standard deviation, correlation coefficient, etc. This value is calculated from the sample and is often applied to population but may or may not be a valid estimate of the population.

Though not desirable, parameters and statistic are used as synonyms.

Notations for population and sample values:

Roman letters are used for sample statistics and Greek letters for population parameters. Common notations are:

Summary value	Sample statistic	Population parameter
Mean	$\overline{X}$ Pronounced: X-bar	μ Pronounced: mu
Standard deviation	s	σ
Variance	s^2	σ^2

Other symbols commonly used:

$S_{\overline{X}}$ = Standard error of the mean. (also SEM or $SE_{\overline{X}}$)

CV = Coefficient of variation (or V)

$\widetilde{X}$ = Median

r = Correlation coefficient

z = The number of standard deviation from the mean

df = Degrees of freedom

K = Number of groups or classes

p = probability

α = Significance level

C = Number of Columns in a contingency table

R = Number of Rows in a Contingency table

DESCRIPTIVE STATISTICS

A. Measures of central tendency:

1. Mean:

$$\overline{X} = \frac{\sum X}{N}$$

2. Median ($\widetilde{X}$):

When all the observations of a variable are arranged in either ascending or descending order, the middle observation is known as the median. If the rate sedimentation rate of seven subjects are: 4, -3, 5, 7, 4, 6 and 5 then the median is 5. Generally, if X_1, X_2, X_3,......., X_n represent a random sample of size n, arranged in increasing order of magnitude, then the sample median is defined by the statistic:

$$\widetilde{X} = \begin{cases} X_{n+1}/2 & \text{if n is odd} \\ (X_n/2 + X_n/2)+1 & \text{if n is even} \end{cases}$$

The median is almost equal to the mean if the distribution does not contain extreme values. The mean is affected by extreme values. Median is a better indicator of an average than the mean when one ore more of the lowest or highest observations are wide apart.

3. Mode:

It is the most frequently occurring observation in a series; i.e., the most common or most fashionable. The mode does not always exist. This is certainly true when all observations occur with the same frequency. For some set of data there may be several values occurring with the greatest frequency, in which case we have more than one mode.

REMARKS:

The mean is the most used measure of central location in statistics. It employs all available information. The only real disadvantage of the mean is that it may be affected adversely by extreme values. The median has the advantage of being easy to compute if the number of observations is relatively small. It is not influenced by extreme values and consequently gives a better center of the data. The mode is the least used measure of central tendency. For small sets of data its value is almost useless if, in

fact, it exists at all. Only in case of a large mass of data does it have a significant meaning. It two main advantages are that (1) it requires no calculations, and (2) it can be used for quantitative as well as qualitative data.

B. Variability:

The three measures of central tendency defined above do not by themselves give an adequate description of our data. We need to know how the observations spread out from the average. It is quite possible to have two sets of observations with the same mean or median that differ considerably in the variability of their measurements about the average. In sciences like physics and chemistry there is not as much variability as in biomedical sciences. One chlorine atom is like another but when arranged in biological objects their effects vary.

Biological data, quantitative or qualitative, collected by measurement or counting are very variable. No two measurements in man are absolutely equal, not even the means or proportions of two series in health or disease are equal although we compare like with the like. The cure rate in typhoid with the same drug will vary in different patients, mean height of students of the same age group at one place will differ from that of another place; and birth rates of two towns are seldom equal. Variability is essentially a normal character. In other words, occurrence of variability is a biological phenomenon.

Types of variability:

There are three main types of variability:
1.   Biological variability
2.   Real Variability
3.   Experimental variability

1. Biological Variability:

Individuals, in similar environments differ when compared with regard to age, sex, class and other attributes but the difference noted is small and is said to occur by chance. Such a difference or variability is called biological variability. It may be defined as the normal or natural difference in all

individuals or groups that occurs within certain biological limits. The limits of a variable character from the mean and the proportion of a population can be studied by the application of certain principles of variation such as normal distribution, range, etc. Calculation of such measures usually involves standard mathematical techniques which help us to determine not only the extent of variation but also whether the difference observed from the central value is real (due to some factors other than natural) or not. This is the main if not the sole aim of the study of biostatistics as a science.

A small variation in birth rates in Riyadh from 35 to 37 per thousand population may be considered due to chance; but birth rate, 37 of Riyadh compared to 16 in London would not be a chance variation. Probably it is a true difference due to some human factors interfering with nature such as birth control, raising of marriage age or some other social customs. On the other hand, reduction in typhoid mortality by 2-3% from 30% would be due to chloramphenicol and not due to chance or biological reasons.

Biological variability may be classified as:

a. Individual variability:

one student's height is 160 cm and that of another of the same age is 170 cm. To find whether their height are normal or not, we find the mean and the standard deviation of heights in the population, by taking a large representative sample.

b. Periodical variability:

The same individuals show variation in temperature, pulse rate, blood pressure, WBC count, cholesterol levels, etc., at different times of the day. The variation can be tested by applying suitable statistical methods.

c. Class, group or category variability:

Height, weight, blood pressure, rate of metabolism, etc., vary from class to class depending on age, sex, race, social status, etc. Whether the mean

height of males is more than females or not, is determined by standard error of difference between two means or by applying unpaired t-test if the sample is small.

d. Sampling variability:

To find mean, standard deviation or proportion of certain characters, we do not examine each and every individual in the population but a sample is taken which is much smaller. The values of any sample therefore, will differ from those of the population and further there will be variability from one sample to another sample. This is a biological variability of samples and is called, sampling variability or sampling error. It is measured by finding the standard error. Sampling error is a chance variability which can be reduced by increasing the size of the sample but cannot be totally eliminated.

2. Real variability:

When the difference between two readings, observations or values of classes or samples   is more than the defined limits in universe, it is said to be real. Then the cause may not be natural or inherent in the man or samples of men but lies in external factors, e.g., statistically significant better cure rate may be due to a drug and not by chance. The incidence rate of AIDS in the United States is far more than that in Saudi Arabia due to the difference in social behaviors. Difference in incidence rate of lung cancer among smokers and non-smokers may be due to excessive smoking and not due to chance. If the variability is found to be real, the offending external factors to be found and duly dealt with.

3. Experimental variability:

Error or difference or variation may be due to materials and methods employed in the study or defects in the techniques involved in the experiment. They are of three types:

a. **Observer error: (may be subjective or objective)**

Subjective: An interviewer or interrogator may add a number or of errors while noting human particulars unless trained properly. He may even ask embarrassing questions which the person may not like to answer such as menstrual history, pregnancy, use of family planning methods, sexually transmitted diseases, etc. Some subjects are subjects are very keen while other do not wish to give any information.

Objective: These errors may be added by an untrained observer while recording the measurements, such as blood pressure, pulse rate, etc.

b. **Instrumental error:** defects in the measuring instrument may cause negligible or gross error. Hence all tools employed in a study should be standardized beforehand.

N.B. Observer and experimental errors are sometimes called **Non-sampling errors**.

c. **Sampling defects or errors of bias**:

A sample drawn should not be biased or too small to draw conclusions. As mentioned before, it should be representative and of sufficiently large size to stand statistical tests.

Hospital based studies are mostly biased because the sample of patients under the study is drawn from poor, rich or nearby strata of society. Moreover many come to hospital at a late stage. In many countries, birth weight of children born in a private maternity hospital cannot be compared with those born in a general hospital.

Measures of Variability:

Variability can occur in: (A) individual observations, and (B) in samples:

A) Measures of variability in individual observations:

1. Range

2. Variance

3. Standard deviation

4. Coefficient of variation

B) Measures of variability in samples:

1. Standard error of the mean

2. Standard error of the difference between two means

A. Measures of variability in individual observations:

1. Range:

The range of a random sample of X_1, X_2, X_3,......., X_n is defined by the statistic:

$$\textbf{Range} = \textbf{X}_\textbf{n} - \textbf{X}_1$$

i.e., the largest observation minus the smallest observation in the sample.

2. Standard deviation (s or SD):

The standard deviation is an extremely important measure of variation and its importance will be evident in later topics. It is computed by the following six steps:

i) Calculate the mean ($\bar{x}$)

ii) Find the difference of each observation from the mean. ($x - \bar{x}$)

iii) Square the differences of observations from the mean. ($x - \bar{x}$)^2.

iv) Add the squared values to get the sum of squares. $\sum (x - \bar{x})^2$.

v) Divide this sum by the number of observations minus one to get the mean-squared deviation (called the variance as we will see next).

vi) Find the square root of the variance to get root-mean squared deviation (called standard deviation).

$$S = \sqrt{\frac{\Sigma\left(X - \overline{X}\right)^2}{n - 1}}$$

Example:

Data X	Deviations ($X - \overline{X}$)	Squared deviations ($X - \overline{X}$)^2
10	-30	900
20	-20	400
30	-10	100
50	10	100
90	50	2500

$\overline{X} = 40$

$\Sigma X = 200$ $\Sigma (X - \overline{X})^2 = 4000$

$$S = \sqrt{\frac{\Sigma\left(X - \overline{X}\right)^2}{n - 1}} = \sqrt{\frac{4000}{4}} = 31.62$$

A large standard deviation shows that measurements of the frequency distribution are widely spread out from the mean. Small standard deviation means that the observations are closely spread in the neighborhood of the mean. The standard deviation can also be calculated by a more direct method without

18

determining the mean by the following **Computational Formula**:

$$S = \sqrt{\dfrac{\Sigma X^2 - \dfrac{(\Sigma X)^2}{n}}{n-1}}$$

Solution of the last example using the computational formula:

X	X²
10	100
20	400
30	900
50	2500
90	8100
$\sum X = 200$	$\sum X^2 = 12000$

$$s = \sqrt{\dfrac{\sum X^2 - \dfrac{\left(\sum X\right)^2}{n}}{n-1}}$$

$$= \sqrt{\dfrac{12000 - \dfrac{(200)^2}{5}}{4}}$$

$$= \sqrt{1000} = 31.62$$

Uses of the standard deviation:

1. It summarizes the deviations of a large distribution from mean in one figure used as a unit of variation.

2. It indicates whether the variation of the difference of an individual from the mean is by chance (i.e., natural) OR due to some special reasons.

3. It is used to calculate the standard error which determines whether the difference between two means of similar samples is by chance or real.

4. It is also used to calculate the suitable sample size (n) for any study.

3. Variance (s^2):

The variance is the square of standard deviation. The variance in the last example is 1000.

4. Coefficient of variation:

It is a measure to compare relative variability; i.e., it relates the variability in a series to the average of the series:

$$CV\% = \frac{s}{\overline{X}}.100$$

Example:

In two series of male and female Pharmacy students of the same age, the following values were obtained for systolic blood pressure. Find which series shows greater variation?

<u>Sex</u>	<u>Mean B.P</u>	<u>SD</u>
Boys	120 mm Hg	10 mm Hg
Girls	130 mm Hg	15 mm Hg

Solution:

$$CV\%\,(\mathbf{boys}) = \frac{(10)}{120}.(100) = 8.33\%$$

$$CV\%\,(\mathbf{girls}) = \frac{(15)}{130}.(100) = 11.50\%$$

B. Measures of variability in samples:

1. Standard error of the mean (SEM or $s_{\overline{X}}$):

Whatever be the sampling procedure or the care taken while selecting a sample, the sample statistics or estimates ($\overline{X}$ and **s**) will differ from population parameters (**μ** and **s**) because of chance or biological

variability. Such a difference between sample and population values is measured by a statistic known as <u>sampling error</u> or <u>standard error</u> $s_{\bar{X}}$.

Standard error is thus a measure of chance variation and it does not mean error or mistake. The standard error of means drawn from a particular population could be calculated if means of large number of samples were known. In practice, it is not possible to take large number of samples. Usually only one large sample is drawn. Using the statistics of this sample, the standard error of the mean is then calculated by the following formula:

$$\text{SEM or } s_{\bar{X}} = \sqrt{\frac{s^2}{n}} \qquad \text{or} \qquad \text{SEM or } s_{\bar{X}} = \frac{s}{\sqrt{n}}$$

2. <u>**Standard error of difference between two means**</u>:

$s_{\bar{X}_1 - \bar{X}_2}$ or $SE_{\bar{X}_1 - \bar{X}_2}$

If independent large and random samples are drawn in pairs, repeatedly from the same population and each time the difference between the two means of each pair is calculated, there will be a population of differences or a series of differences of the paired samples as below:

$\left(\bar{X}_{1a} - \bar{X}_{1b}\right), \left(\bar{X}_{2a} - \bar{X}_{2b}\right), \left(\bar{X}_{3a} - \bar{X}_{3b}\right), \ldots\ldots\ldots, \left(\bar{X}_{na} - \bar{X}_{nb}\right)$

In practice, it is <u>NOT</u> possible to find large number of differences in pairs of sample means and then find the S.E. of these differences. Therefore, it is usually calculated using the following formula:

$$\mathbf{s}_{(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)} = \sqrt{\frac{\mathbf{s}_1^2}{\mathbf{n}_1} + \frac{\mathbf{s}_2^2}{\mathbf{n}_2}}$$

This formula means that the standard error of the difference between two means is the square root of the sum of squares of standard errors of those means. This formula applies when the mean and the standard deviation of the population are not known but the samples are large.

NORMAL DISTRIBUTION

In the last lecture, we studied the measures of central tendency and the measures of variability. We have seen how the range is not a satisfactory measure of variation whereas SD summarizes the variation of large distribution, measures the position of an observation with regard to mean, defines the limits of normality and indicates the chances of occurrence of an observation in a population. To know how SD does all this, one has to understand what is a standard distribution or normal distribution of observations in a universe.

Out of all the standard or theoretical distributions, normal distribution is the most important. The following observations were made on fairly large number of normal persons selected at random:

	Mean	SD	CV%	Range
Menstrual cycle in days	28	2	7.98	9-18
Pulse rate in beats	72	3.5	4.86	55-90
Weight at birth in kg	3.05	0.39	12.79	1.8-4.5
Systolic B.P. in mm Hg	115	12	10.43	90-160

Physiological variation, as you can see it under the range, warns us not to pronounce any isolated (extreme) observation as abnormal, in otherwise normal person. Still we have to define normal or critical limits or determine the probable chances of an observation being normal because large deviations from the mean may have unfavorable prognostic value.

If mean menstrual cycle as given above is 28 days and the range is 21 to 35 days, one would like to know how many women will have their cycle between 28 and 30 days or between 26 and 28 days. Similarly the incubation period of gonorrhea studied in large number of persons varies from 2 to 8 days with an average of 5. What proportion of exposed persons develop the disease after 5 days or 6 days or before 4 days or between 3 and 4 days?

Such questions cannot be answered by range and mean. To determine the chances or probability of an observation or observations being normal and also to determine the proportion of observations that lie between mean and a higher or lower observation we think in terms of frequency distribution as a whole taking into account the mean or central position and variability of observations up to the extreme on either side of the mean in terms of SD.

When a large number of observations of any variable characteristic such as height, blood pressure and pulse rate are taken at random to make it a representative sample, a frequency distribution table is prepared, then it will be seen that:

(a) Some observations are above the mean and others are below the mean.

(b) If the observations are arranged in order, deviation towards the extremes from the mean, on plus or minus side, maximum number or frequencies will be seen in the middle around the mean and fewer at the extremes, decreasing smoothly on both sides.

(c) Normally half the observations lie above and half below the mean and all observations are symmetrically distributed on each side of the mean.

A distribution of this nature or shape is called **normal** distribution or **Gaussian** distribution. It is the most important continuous probability distribution in the entire field of biostatistics. Its graph, called the **normal curve**, is the bell-shaped curve of Figure 2.1, which approximately describes many phenomena that occurs in nature and research. Many

Figure 2.1
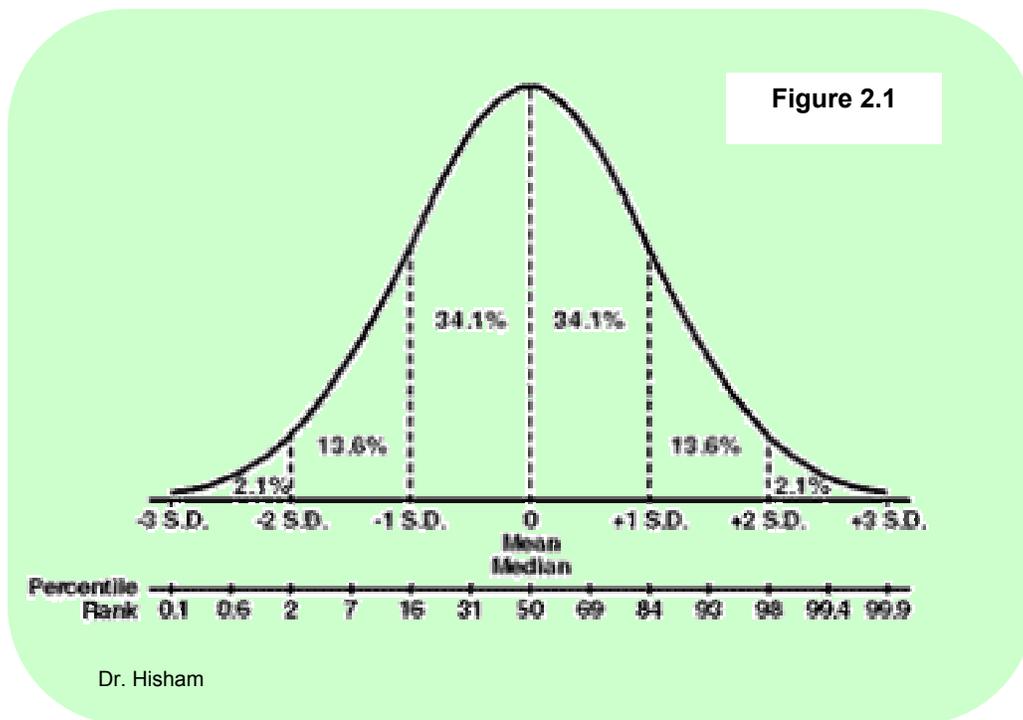
natural attributes such as height, etc., tend to be distributed in a similar pattern. In addition, errors in scientific measurements are extremely well approximated by a normal distribution.

It can be arithmetically expressed as follows in terms of mean and SD, if they are known:

(a) **Mean ± 1 SD:** include **68.27%** or roughly two thirds of all observations. Out of the remaining one third observations, half; i.e., 1/6 will lie below the lower limit of (mean−one SD) and the other half; i.e., 1/6 will lie above the upper limit of (mean + one SD). In other words, about 32% will lie outside the range, (mean−1SD) and (mean+1SD).

(b) **Mean ±2 SD** limits, include **95.45%** of observations while **4.55%** of observations will be outside these limits.
  N.B. (Mean ± 1.96 SD) limits, include exactly **95%** of all observations.

(c) **Mean ± 3 SD** limits, include **99.73%.**
  N.B.   (Mean ± 2.58 SD) limits, include exactly **99%.**

In other words, in any normal distribution, it is found that:

1. Observations larger or smaller than **Mean ± 1 SD** are fairly common, forming nearly one third.
2. Values that differ from the mean by more than twice the standard deviation are rare, being only 4.55%. Their chance of being normal is 4.55 in 100. In 95.45% cases they may be abnormal.
3. Values higher or lower than **mean ± 3 SD** are very rare, being only 0.27%. Their chance of being normal is 0.27 in 100 cases.

Thus, 6 standard deviations, 3 on either side of the mean, cover almost the entire range of a variable character.

Characteristics of normal curve:

1. It is bell shaped
2. It is symmetrical
3. Mean, mode and median coincide

4. It has two inflections. The central part is convex while at the points of inflection the curve changes from convexity to concavity. A perpendicular from the point of inflection will cut the base at a distance of 1 SD from the mean on either side.

Thus, the normal distribution describes the distribution of any variable such as height, weight, blood pressure, bleeding time, clotting time, incubation period, menstrual cycle, etc.

The shape of normal distribution is very useful in clinical practice and makes statistical analysis easy. It tells the probability of occurrence by chance or how often an observation, measured in terms of mean and standard deviation can occur normally in a population.

Relative or standard normal deviate or variate (Z):

Deviation from the mean in a normal distribution is called relative or standard normal deviate and is given the symbol Z. It is measured in terms of SD and indicates how much an observation is bigger or smaller than mean in units of SD. So Z will be a ratio calculated as follows:

$$Z = \frac{\text{Observation} - \text{Mean}}{\text{Standard Deviation}}$$

$$Z = \frac{X - \mu}{\sigma}$$

Example (2.1):

The average weight of baby at birth is 3.05 kg with SD of 0.39 kg. If the birth weights are normally distributed, would you regard:

(a) Weight of 4 kg as abnormal?

(b) Weight of 2.5 kg as normal?

Solution:

(a) Normal limits of weight at 1.96 SD = 3.05 ± 1.96 x 0.39, i.e., 2.29 kg and 3.81 kg. The weight of kg will be abnormal in more than 95% of the cases. (To understand why we used 1.96, read part "b" on page 4 again)

(b) The weight of 2.5 kg lies within the normal limits of 2.29 and 3.81 kg (calculated in part "a" of this problem) so it is not abnormal.

Example (2.2):

Menstrual cycle in women following normal distribution has a mean of 28 days and SD of 2 days. How frequently would you expect a menstrual cycle (MC) of:

    (a) More than 30 days?

    (b) Less than 22 days?

Solution:

(a) About 68% of the women as per normal distribution will have a MC of mean ± 1 SD; i.e., (28 ± 2) or 26 to 30 days. This means that 32% will have longer or shorter MC. Because of symmetry in normal distribution, half the women; i.e., 16% will have a cycle of more than 30 days.

This part also can be solved using standard normal deviate as follows:

$$Z = \frac{X - \mu}{\sigma} = \frac{30 - 28}{2} = 1$$

Referring to the Z table, we find that the probability for Z=1 is 0.1587; i.e., 15.87%

(b)
$$Z = \frac{X - \mu}{\sigma} = \frac{22 - 28}{2} = -3$$

Corresponding value as per the table is 0.0013. Thus, 0.13% of women will have MC less than 22 days.

Confidence Interval

A confidence interval gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data.

If independent samples are taken repeatedly from the same population, and a confidence interval calculated for each sample, then a certain percentage (confidence level) of the intervals will include the unknown population parameter. Confidence intervals are usually calculated so that this percentage is 95%, but we can produce 90%, 99%, 99.9%, confidence intervals for the unknown parameter.

The width of the confidence interval gives us some idea about how uncertain we are about the unknown parameter. A very wide interval may indicate that more data should be collected before anything very definite can be said about the parameter.

The confidence interval can be calculated as follows:

$$\mathbf{CI} = \overline{\mathbf{X}} + \mathbf{z.SEM}$$
$$= \overline{\mathbf{X}} + \mathbf{z.}\frac{\mathbf{s}}{\sqrt{\mathbf{n}}}$$

Examples of Confidence Intervals:

(To be given in class)

Normal Curve Areas
Entries in the Body of the Table Give the Area Under
the Standard Normal Curve from 0 to z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Hypothesis and Significance Testing

We saw how we can use sample statistics to estimate population parameters. Estimation is a useful application of inferential statistics, but perhaps even more important is the application of statistical procedures to hypothesis testing. Clinical research very often involves comparisons between experimental and control groups, such as comparisons; such as comparisons are made in clinical trials, prospective and retrospective studies. Furthermore, we are not interested in differences between the actual groups under study, but in broad, general applications to populations that the groups represent.

For example, if we are interested in the efficacy of a drug on some variable such as heart rate, blood pressure, cholesterol concentration, etc., we might design a study such as a clinical trial in which subjects are randomized to control and drug groups. We can then compare means of the groups to see if they are different. The problem is, they will undoubtedly be different even if the drug has no effect whatever. In other words, we can expect some degree of difference to result from chance alone. Our question, therefore, pertains to whether or not the difference that we observe between the means of the drug and control groups reflects a difference between the populations to which the drug and placebo could hypothetically be given. This is the essence of hypothesis testing.

The Null Hypothesis:

Suppose that a great many measurements taken over a long period of time allow us to assume that the mean level of prothrombin in the normal population is 20 mg/dl of plasma with a standard deviation of 5 mg/dl. Now further suppose that we suspect that patients with vitamin k deficiency also tend to have and abnormally low prothrombin level.

We randomly draw a sample of 100 patients from the population that may be defined as all patients with vitamin k deficiency. Tests performed on the

patients in the sample yield a mean prothrombin level of 19 mg/dl of plasma. Our data are then summarized below.

General population	sample
μ = 20 mg/dl	n = 100
σ = 5 mg/dl	$\overline{X}$ = 19 mg/dl

When performing statistical tests of hypothesis, the hypothesis is always stated in the null form. In other words, a null hypothesis is simply a statement of "no difference." Thus, we begin with the statement that "no statistically significant difference exists between the general population mean (μ_o) and the mean of the population (μ) that consists of all patients with vitamin K deficiency." This is another way to say that the prothrombin level of vitamin k-deficient patients does not differ from the prothrombin level found in the general population. Finally, to put it still another way, we are saying that the difference between the general population mean (μ_o) and our sample mean ($\overline{X}$) occurred by chance alone.

Symbolically, we express the null hypothesis as

$$H_o: \quad \mu_o = \mu$$

Note carefully that the symbols used in the null hypothesis are parametric symbols; that is, they refer to population values. That implies that we are not interested in sample statistics per se; instead, we used them to make "educated guesses" about population. For example, the investigator who tested the Salk vaccine on a sample of children was not primarily concerned about whether the vaccine would protect the sample, but whether or not all children in the world would be protected if given the vaccine.

In most research situations, we would like to show that a difference does exist. No one enjoys spending years developing a drug only to be unable to

demonstrate its efficacy. The researcher with a "cure for cancer" may have only fading dreams of fame and fortune if statistical evidence does not support his or her claims. In general, therefore, the name of the game is to reject the null hypothesis, thus providing evidence that a real difference exists.

Levels of Significance (α level):

In formal hypothesis testing, it is traditional to establish, prior to performing the test, the maximum probability (of a "chance alone" difference) at which the null hypothesis will be rejected. In other words, one traditionally states the maximum probability of b being wrong if the null hypothesis is rejected.

The usual probability levels, or levels of significance, that have been established more or less arbitrarily as acceptable refection points are 0.05 and 0.01. It is therefore traditional to reject the null hypothesis "**if and only if**" the 0.05 (or, in some cases, the 0.01) level is reached.

In general, we see three levels of significance in the literature-0.05, 0.01, and 0.001 instead of computing the actual level of significance, research reports usually state the level of significance obtained as

$$p < 0.05, \qquad p < 0.01, \qquad \text{or} \qquad p < 0.001$$

WHAT DOES "SIGNIFICANT" MEAN?

In the scientific literature there is probably on other word that is more overworked, misused, and misleading than "significant." In the best journals we can see statements such as "The difference between the treatment and the control groups was significant," or "The difference between the two treatments proved significant beyond the 0.05 level."

We should understand without further delay that the term "statistical significance" has a very precise and rather limited meaning. Specifically, it refers to the refection of a null hypothesis at some level, such as 0.05, 0.01, 0.001, etc.

To illustrate, let's suppose that we have new have a new drug that we have reason to believe will effectively lower systolic blood pressure in hypertensive patients to drug and placebo groups so we can test the null hypothesis, $H_O : \mu_D = \mu_P$. Since on one wants to hear about a drug that doesn't work, we naturally hope to reject the null hypothesis.

After treatment with the drug and the placebo, we find that the mean systolic blood pressure of the drug group ($\overline{X}_D$) is lower than the mean systolic blood pressure of the placebo group ($\overline{X}_P$). We are not surprised by that; in fact, we would have been more surprised if $\overline{x}_D$ and $\overline{x}_P$ were equal because we expect some difference by chance alone. What we are really interested in, of course, is whether or not a difference exists between μ_D and μ_P, the means of the populations that the sample groups represent.

Our next step, then is to perform an appropriate statistical test. We do so and find that we can reject the null hypothesis with $p < 0.5$, or, or in other words, we have found statistical significance. Which means that? It means we can claim that a difference does exist between μ_D and μ_P or, we can say that a statistically significant difference exists between μ_D and μ_P. On the other hand, we cannot say what that difference is. The observed difference between $\overline{X}_D$ and $\overline{X}_P$ is only an estimate of the difference between μ_D and μ_P. We do not know the difference between μ_D and μ_P and we will never know it, although we can estimate it with 95% confidence limits using the procedure explained in a previous lecture.

Finally, although a statistically significant difference dose not guarantee that the difference is meaningful or useful, it is necessary to demonstrate

statistical significance in order to claim that a difference exists. Remember, the name of the game is to show (with some small degree of doubt) that sample statistics do not differ by chance alone. In many situations, therefore, claims and conclusions are not taken seriously unless they are supported by appropriate statistical analysis.

Types of Errors in Hypothesis Testing:

As we stated, the researcher usually is hoping to find evidence against the null hypothesis. What constitute such evidence? We said earlier that the goal of hypothesis testing is to answer the question, "Is there a difference between the parameter of the test population and the parameter of the standard population?" If you reject the null hypothesis, you are saying, "yes, there is a difference; my treatment has an effect." If you fail to reject the null hypothesis, you are saying, "I have not found convincing evidence of a difference; my treatment remains unproven." In either case, there is always the possibility that you have made an error. Biostatisticians distinguish between the errors made in each of these two cases. The error of wrongly rejecting the null hypothesis (deciding an ineffective treatment is effective) is called **Type I error**. The error of failing to reject a false null hypothesis (failing to identify that an effective treatment is indeed effective) is called **Type II error**. The following table describes these two types of errors:

	Reality	
	H_O True	H_O False
Accept H_o	Correct Decision	**Type II error**
Decision		
Reject H_o	**Type I error**	Correct Decision

Scheme for Inferential Statistics

(to be given in class)

<p style="text-align:center">**COMPARISIONS BETWEEN MEANS**</p>

<p style="text-align:center">## STUDENT'S t-TESTS</p>

The simplest and the most frequently used statistical test is the Student's t-test. This test is specially suited for any experiment in which the same measurement (e.g., plasma levels of drug) has been made on two separate groups. One group usually is a control (i.e., standard reference) and the other the test group.

There are two types of t-test:

   a)   Independent t-test (for independent or unpaired samples).

   b)   Paired t-test (for dependent or paired samples)

These two tests have the following assumptions:-

   1.   The sample is selected randomly to avoid bias.

   2.   Normal distribution.

   3.   Quantitative data.

   4.   Homogeneity of variance.

A) **Independent t-test**:

This test is applied to unpaired data of independent observations made on individuals of two different or separate groups or samples drawn from two populations, to test if the difference between the means is real or it can be attributed to chance (sampling variability).

*   Null hypothesis (H_o): It is assumed that that there is on real difference between the means of the two samples.

*   Calculation:

For the t-test, the individual measurements (i.e., mean, standard deviation) of one group are calculated followed by the second group of data.

1. Calculate the "t" value; i.e., the ratio between the observed difference between the two means and its standard error $(S_{\overline{X}_1 - \overline{X}_2})$ using the following formula:

$$t = \frac{\left|\overline{X}_1 - \overline{X}_2\right|}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

2. Determine the degrees of freedom associated with the experiment as follows:

$$Df = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$$

3. Consult a book of statistical tables and turn to the corresponding page that contains the table entitled "significance limits of the Student distribution" (See the t-table at the end of this section). Compare the calculated value with the corresponding value in the table. This value can be identified from the table if you know the following:

   a) Significance level, i.e., α-value, (usually 0.05, 0.01 or 0.001.

   b) Degrees of freedom.

   c) One-tail or two-tail (One-sided or two-sided)

   If the calculated t-statistics exceeds the tabulated t-value, then the two groups differ significantly with only a α% (i.e., 5%, 1% or 0.1% depending on chosen a value) probability that an error has been made.

Example:

Twelve healthy male volunteers were given 650 mg does of acetaminophen (product A). An equivalent dose of another acetaminophen

(product B) was given to other 12 subjects. The table below shows the plasma concentration data for both groups after 0.5 hour of drug administration. Perform t-test for statistical significance between the concentrations of the two groups at $\alpha=0.05$.

Plasma concentration	
Product A	Product B
6.4	11.5
8.8	6.3
3.6	6.9
9.8	4.2
9.5	8.5
8.9	5.7
2.9	7.2
8.6	7.6
1.3	3.6
7.0	1.6
8.7	8.0
2.0	8.7

Solution:

- $\overline{X}_A = 6.5,$ $\overline{X}_B = 6.7$
  $S_A = 3.15,$ $S_B = 2.63$
  $n_A = 12,$ $n_B = 12$

- $t = \dfrac{6.5 - 6.7}{\sqrt{\dfrac{3.15^2}{12} + \dfrac{2.63^2}{12}}}$

  $= 0.1688$

- $d£ = b12+12-2 = 22$
- t-value (table) $= 2.074$

The difference between the two groups is not statistically significant (or statistically insignificant) (p<0.05).

Summary:
Based on the means observed in two independent sample, how can you test the hypothesis that the two population means are equal?

- To test the null hypothesis that the two population means are equal, you must calculate the probability of seeing a difference at least as large as the one you've observed in your sample, if there is no difference in the population.
- The hypothesis that there is no difference between the two population means is called null hypothesis.
- The probability of seeing a difference at least as large as the one you've observed, when the null hypothesis is true, is called the observed significance level.
- If you reject the null hypothesis when it is true, you make a type I error. If you don't reject the null hypothesis when it is false, you make a type II error. The t test is used to test the hypothesis that two population means are equal.

Paired t-test:

It is applied to paired data of independent observations from one sample only when each individual gives a pair of observations. Such situations are commonly faced in biomedical sciences such as:

1. To study the role of a factor or cause when the observations are made before and after its play, e.g., effect of exertion on pulse rate, effect of a drug on blood pressure, etc.
2. To compare the effect of two drugs, given to same individuals in the sample at two different occasions.
3. To compare the bioavailabilities of two drug formulations with the same chemical moiety (Bioequivalence studies).
4. To compare the accuracy of two different instruments.
5. To compare results of two different lab techniques.
6. To compare observations made at two different sites in the same body.

In all of these instances, the subject serves as his own control by being measured before and after treatment. Testing by this method eliminates many sampling errors because the sample is one and the same with before and after observation on each person in the sample, while in unpaired t-test there are two samples similar in all respects to eliminate bias. In such designs, the formula for $s_{(\bar{X}_1 - \bar{X}_2)}$ is not appropriate, since there will be correlation between sample means.

For testing the significance of difference using this test, we follow the following steps:

- **Null Hypothesis**: The matched pair t test involves H_O that the average of all difference scores $(\overline{D})$ is zero. This is, of course, what we would expect if the treatment has on effect.

$$H_O : \overline{D} = 0$$

- Find the difference in each set of paired obser4vations before and after:

$$D = X_1 - X_2$$

- Calculate the mean of the difference ($\overline{D}$)

- Calculate the standard deviation of the differences, then, calculate the standard error of the mean difference.

- Calculate the "t" value; i.e., the ratio between the mean difference and its standard error ($s_{\overline{X}}$) using the following formula:

$$t = \frac{\left| \overline{D} \right|}{\sqrt{\dfrac{s^2}{n}}}$$

- Determine the degrees of freedom associated with the experiment as follows:

$$df = (n - 1)$$

- Consult a book of statistical tables and turn to the corresponding page that contains the table entitled "significance limits of the Student distribution" (see the t-table at the end of this section). Compare the calculated value with the corresponding value in the table. This value can be identified from the table if you know the following:
  - a) Significance level, i.e., α - value, (usually 0.05, 0.01, or 0.001)

b) Degrees of freedom.

c) One-tail or two-tail (One-sided or two-sided).

- If the calculated t-statistics exceeds the tabulated t-value, then the two groups differ significantly with only a $\alpha\%$ probability that an error has been made (i.e., 5%, 1% or 0.1% depending on chosen α-value).

Example:

Assume that the data in the example for unpaired t-test is paired, then, the solution will be as follows:

Pair #	Product A	Product B	D
1	6.4	11.5	-5.1
2	8.8	6.3	2.5
3	3.6	6.9	-3.3
4	9.8	4.2	5.6
5	9.5	8.5	1.0
6	8.9	5.7	3.2
7	2.9	7.2	-4.3
8	8.6	7.6	1.0
9	1.3	3.6	-2.3
10	7.0	1.6	5.4
11	8.7	8.0	0.7
12	2.0	8.7	-6.7

Solution:
- Standard deviation of the difference:

$\Sigma D^2 = 185.07$

$\Sigma D = -2.3$

$$s_D = \sqrt{\frac{185.07 - (-2.3)^2 / 12}{12 - 1}}$$

$$= 4.097$$

$\overline{D} = -0.192$

$$* \quad t = \frac{|-0.192|}{\sqrt{\frac{(4.097)^2}{12}}} = 0.162$$

* $df = 12 - 1 = 11$
* t-value (Table) =2.201
* What is your conclusion?

Problem:

Systolic blood pressure (SBP) of 9 normal individuals, who had been recumbent for 5 minutes, was measured. Then 2 ml of 0.5% solution of a hypotensive drug was given and blood pressure recorded again. Did the drug injection lower the SBP?

Pair #	1	2	3	4	5	6	7	8	9
SBP (before)	122	121	120	115	126	130	120	125	128
SBP (after)	120	118	115	110	122	130	116	124	125

Summary:

   Based on the means observed in two related (Paired, matched or dependent) samples, how can you test the null hypothesis that in the population these is on difference between the means?

* In a paired design, the same subject is observed under two conditions, or data are obtained from a pair of subjects that are matched on some basis.

* Paired designs help to make the two groups being compared more similar. Some of the differences between subjects are eliminated.

* If you observe the same subject under two conditions, you must make sure that the effect of one treatment has worn off before the other one is given.

* you compare two related means using a paired–sample t test.

The t Table

Distribution of Giving Both the Two-Sided or Two-Tailed Probability and the One-Sided or One-Tailed Probability According to Degrees of Freedom

				One Tail				
	P 0.4	P 0.3	P 0.2	P 2.1	P 0.05	P 0.025	P 0.01	P 0.005
				Two Tails				
DF	P 0.8	P 0.6	P 0.4	P 0.2	P 0.1	P 0.05	P 0.02	P 0.01
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.684	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.259	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.259	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.563	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.639	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	9.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.630	0.854	1.310	1.697	2.042	2.457	2.750
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576

CHI SQUARE TEST (χ^2)

The test involves the calculation of a quantity, called chi-square from the Greek letter (χ) and pronounced "kye." This test has the following common applications in medical sciences:

1. As an alternative test to find the significance of difference in two or more than two proportions.

2. As a test of association between two events. It measures the probability of association between two discrete attributes. Two events can often be related such as smoking and cancer; treatment and outcome of disease; vaccination and immunity; nutrition and intelligence; cholesterol and coronary heart disease; contraceptives and breast or endometrial cancer in women; alcohol and gastric ulcer; and so on. There are two possibilities, either they influence or affect each other or they do not. In other words, they are independent of each other or they are dependent on each other; i.e., associated. Assumptions of independence of each other or no association between two events is made, unless proved otherwise by χ2 test. Thus the test measures the probability (p) or relative frequency of association due to chance and also if the two events are associated or dependent on each other.

Here again has an added advantaged that it can be applied to find association or relationship between two discrete (qualitative) attributes when there are more than two classes or groups as happens in multinomial samples, e.g., test of association between number of cigarettes, ≤10, 11–20, 21–30, and >30 smoked per day and the incidence of lung cancer; or between incidence of filariasis and social class; e.g., very rich, middle and poor, and so on.

Table 1: Outcome of treatment of drug and placebo

Group	Outcome or result		
	Died	Survived	Total
(control, or placebo)	10	25	35
(Experiment, or Drug)	5	60	65
Total	15	85	100

As you can see Table 1 is prepared by enumeration of qualitative data. Actual frequencies as they are, and not the percentage of the characteristics, are entered. Since the table presents joint occurrence of two sets of events, it is called **CONTINGENCY TABLE** (Latin, con = together, tangere = to touch).

Sometimes we want to know the association between two sets of events, the table is also called **ASSOCIATION TALBE**. Because are only two samples, each divided into two classes, it is also called **four-fold, four-cell** or **2×2- contingency table**.

Chi - square test is also applied to find association between tow events occurring together when there are more than two samples. Then the contingency table or association table be larger than four-cell table as in table 2.

Table 2: Social class and Microfilaria positivity

Social Class	Microfilaria Positivisty		
	Number Positive	Number Negative	Total
I	4	76	80
II	20	180	200
III	60	440	500
IV	144	576	720
Total	228	1272	1500

3. As a test of **GOODNESS OF FIT:** Chi-square test is also applied as a test of "goodness of fit", to determine if actual numbers are similar to the expected or theoretical number; i.e., goodness of fit to a theory. The number of classes in the goodness of fit test is k. we can find wither the observed frequencies of a character differ from the hypothetical or expected ones by chance or due to some factors.

Calculation of χ^2 value:

There are three essential requirements to apply this test. They are:

a) Random sample.

b) Qualitative data.

c) Lowest expected frequency not less than 5.

If these requirements are fulfilled then proceed as follows:

1. Make contingency table as tables 1 and 2 (depending on your data and the comparison you make). The numbers in each class or cell are called observed frequencies (f_o).

2. Determine the expected frequencies (f_e); i.e., the number that we expect to find in each class or cell of the table if the null hypothesis is true. (Remember that Ho states that there is no difference or variation in the proportions of the group from that of the population). The f_e in each cell can be calculated by multiplying the corresponding column and row totals then dividing by the grand total.

$$f_e = \frac{(\text{Row Total})(\text{column Total})}{(\text{Grand Total})}$$

To explain this, the total number in the sample is taken as population, such as 1500 slides, examined for microfilaria in Table 2 out of which 228 were positive. Proportion of total positive in all groups or samples is (228/1500). From this, we can determine the positive number expected out of 80 in the social class I. Therefore, the expected frequency in the cell marked (Positive-class I) is:

$$f\text{e} \quad = \quad \frac{(228)(80)}{(1500)} = 12.16$$

i.e., we expect to find 12.16 positive slides among 80 slides if the null hypothesis is true.

3. Calculate χ^2 using the following formula:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Where,

f_o = observed frequency,

f_e = expected frequency,

k = number of classes or cells

4. Calculate the degrees of freedom (df) which are related, not to the total number of observations but to the total number of categories in both events. An observer expects no difference between observed and expected frequencies in either direction. Consequently, in case of contingency table, the degrees of freedom is given by the following formula:

$$d f \quad = (C - 1)(R - 1)$$

where, C is number of columns and R is number of rows.

5. Refer to χ^2 table and obtain the χ^2 value (the highest obtainable value by chance) at the desired df and level of significance.

6. If the calculated value of χ^2 is higher or lower than χ^2 value given in the table, then it is statistically significant or insignificant, respectively, at that particular level of significance.

Example:

NUMBER OF PATIENTS

	Poison victims	others	Total
Hospital A	2 (10)	48 (40)	50
Hospital B	6 (15)	69 (60)	75
Hospital C	16 (20)	84 (80)	100
Hospital D	56 (35)	119 (140)	175
Total	80	320	400

$$\chi^2 = \frac{(2-10)^2}{10} + \frac{(48-40)^2}{40} + \frac{(6-15)^2}{15} + \frac{(69-60)^2}{60} + \frac{(16-20)^2}{20} + \frac{(84-80)^2}{80}$$

$$+ \frac{(56-35)^2}{35} + \frac{(119-140)^2}{140} = 6.4 + 1.6 + 5.4 + 1.35 + 0.8 + 0.2 + 12.6 + 3.15 = 31.5$$

$$\mathbf{d}f = (4-1)(2-1) = 3$$

$$\chi^2(\mathbf{d}f = 3, \mathbf{P} = 0.05)(\mathbf{from\,table}) = 7.82$$

$$\chi^2_{\mathbf{Calc.}} > \chi^2_{\mathbf{table}}$$

The four hospitals do not admit the same proportion of poison victims in their emergency room. The difference is statistically significant (P < 0.05)

Restrictions in application of χ^2 test:

1. χ^2 in a fourfold table (i.e., 2×2 contingency table) will not give reliable results with one degree of freedom if the expected value in any cell is less than 5. In such cases, to apply χ^2 test, you must perform YATES' correction, i.e., subtract 0.5 from the absolute difference between the observed and expected frequencies and square it then divide by the expected frequency. In this case the formula would be as follows:

$$\chi^2 = \sum \frac{(|f_o - f_e| - 0.5)^2}{f_e}$$

2. Even after Yates' correction, the test may be misleading if any expected frequency is much below 5. In such cases, some other appropriate test may be applied.

3. In tables larger than 2×2, Yates' correction cannot be applied. If any cell frequency is less than 5, this small frequency may be pooled or combined with that in the next group in the table.

4. Interpret χ^2 test results with caution if the grand total is less than 50.

5. χ^2 test tells the presence or absence of an association between two events or characters but does not measure the strength of this association.

6. The statistical finding of relationship does not indicate the cause and effect relationship. χ^2 values do not tell why tonsillectomy and whooping cough; or smoking and cancer are associated; or what is the effect of social class on filariasis. It only tells the probability (p) of occurrence of association by chance.

N.B. In case of a two-by-two contingency table, there is another formula that gives the same results with Yates' correction (i.e., correction for continuity when df=1) while eliminating the need to compute expected values:

Group A	a	b	a+b
Group B	c	d	c+d
	a+c	b+d	N ← Grand Total

PROBLEMS:

1.   All male patients admitted to KKUH with the diagnosis of acute myocardial infarction were included in a study on the evaluation of an anticoagulant, Warfarin sodium. The 147 patients who met the criteria of admission were divided into two groups by whether their hospital unit number was odd or even. The mortality observed in the two groups is given in the table below:

INFLUENCE OF ANTICOAGULANT THERAPY ON MORTALITY

	NO Anticoagulation	Anticoagulation Therapy
No. of patients	70	77
No. of Deaths	15	12

Is there any evidence of difference in mortality between the groups?

A question is raised concerning the possibility of a relationship between smoking history and the severity of a first heart attack. In a retrospective study, 1000 patients are chosen at random from a coronary registry. Severity of attack is categorized as mild, moderate, or severe. Smoking history is categorized on the basis of usage during the year preceding the attack. Four classes are used: (a) Not smokers, (b) intermittent smokers, (c) moderate smokers, and (d) heavy smokers (see table). Compute x^2 as a test of the hypothesis that there is no association between the two variables for the patients in the registry.

	A	b	c	d
Mild	140	120	80	60
Moderate	40	10	100	250
Severe	20	20	70	90

Percentiles of the Chi-Square Distribution
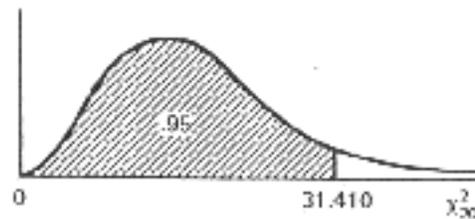
d.f.	$\chi^2_{.005}$	$\chi^2_{.025}$	$\chi^2_{.05}$	$\chi^2_{.90}$	$\chi^2_{.95}$	$\chi^2_{.975}$	$\chi^2_{.99}$	$\chi^2_{.995}$
1	.0000393	.0009982	.00393	2.706	3.841	5.024	6.635	7.879
2	.0100	.0506	.103	4.605	5.991	7.378	9.210	10.597
3	.0717	.216	.352	6.251	7.815	9.348	11.345	12.838
4	.207	.484	.711	7.779	9.488	11.143	13.277	14.860
5	.412	.811	1.145	9.236	11.070	12.832	15.086	16.750
6	.676	1.237	1.635	10.645	12.592	14.449	16.812	18.548
7	.989	1.690	2.167	12.017	14.067	16.013	18.475	20.278
8	1.344	2.180	2.733	13.362	15.507	17.535	20.090	21.955
9	1.735	2.700	3.325	14.684	16.919	19.023	21.666	23.589
10	2.156	3.247	3.940	15.987	18.307	20.483	23.209	25.188
11	2.603	3.816	4.575	17.275	19.675	21.920	24.725	26.757
12	3.074	4.404	5.226	18.549	21.026	23.336	26.217	28.300
13	3.565	5.009	5.892	19.812	22.362	24.736	27.688	29.819
14	4.075	5.629	6.571	21.064	23.685	26.119	29.141	31.319
15	4.601	6.262	7.261	22.307	24.996	27.488	30.578	32.801
16	5.142	6.908	7.962	23.542	26.296	28.845	32.000	34.267
17	5.697	7.564	8.672	24.769	27.587	30.191	33.409	35.718
18	6.265	8.231	9.390	25.989	28.869	31.526	34.805	37.156
19	6.844	8.907	10.117	27.204	30.144	32.852	36.191	38.582
20	7.434	9.591	10.851	28.412	31.410	34.170	37.566	39.997
21	8.034	10.283	11.591	29.615	32.671	35.479	38.932	41.401
22	8.643	10.982	12.338	30.813	33.924	36.781	40.289	42.796
23	9.260	11.688	13.091	32.007	35.172	38.076	41.638	44.181
24	9.886	12.401	13.848	33.196	36.415	39.364	42.980	45.558
25	10.520	13.120	14.611	34.382	37.652	40.646	44.314	46.928
26	11.160	13.844	15.379	35.563	38.885	41.923	45.642	48.290
27	11.808	14.573	16.151	36.741	40.113	43.194	46.963	49.645
28	12.461	15.308	16.928	37.916	41.337	44.461	48.278	50.993
29	13.121	16.047	17.708	39.087	42.557	45.722	49.588	52.336
30	13.787	16.791	18.493	40.256	43.773	46.979	50.892	53.672
35	17.192	20.569	22.465	46.059	49.802	53.203	57.342	60.275
40	20.707	24.433	26.509	51.805	55.758	59.342	63.691	66.766
45	24.311	28.366	30.612	57.505	61.656	65.410	69.957	73.166
50	27.991	32.357	34.764	63.167	67.505	71.420	76.154	79.490
60	35.535	40.482	43.188	74.397	79.082	83.298	88.379	91.952
70	43.275	48.758	51.739	85.527	90.531	95.023	100.425	104.215
80	51.172	57.153	60.391	96.578	101.879	106.629	112.329	116.321
90	59.196	65.647	69.126	107.565	113.145	118.136	124.116	128.299
100	67.328	74.222	77.929	118.498	124.342	129.561	135.807	140.169