Behavioral Science

1) Descriptive statistics

Descriptive statistics describes, summarize or organizes data. A population is the universe about which the investigator wishes to draw conclusions. E.g. if he wishes to draw conclusions about the blood pressure of Americans then that is the population not the Americans themselves. A sample is the subset of the population- the part that is actually being studied. Samples are studied and inferences drawn about the whole population as researcher can rarely study whole populations. A single observation e.g. one's blood pressure is an element and denoted by X. N is the no. of elements in a population, ***n*** is the number of elements in a sample. A population consists of samples from XI to XN and a sample consists of ***n*** of these N elements.

Simple random sample - is the simplest kind of probability sample. It is drawn in such a way that every element in the population has an equal probability of being included.

Stratified random sample - in this sample population is first divided into homogenous groups or strata e.g. male, female or black and white and random samples are drawn from each of the groups. This results in greater representativeness.

Cluster samples - are used when it is too expensive or laborious to use the simple random or stratified random samples. E.g. In a survey of 100 medical schools 10 schools are randomly chosen and then all students at this school are interviewed.

Systematic samples - involves selecting samples in a systematic way - e.g. each fifth patient admitted to the hospital. This provides a random sample without actually doing randomization.

Sampling problems - a researcher advertises to recruit people suffereing from a particular disease. People who respond form self-selected sample, which is probably not representative of the population of people with the disease.

A dermatologist reports the results of a new treatment for acne, which he has been using with his patients. This sample may not be representative of all people with acne (likely that only with severe acne came to see the doctor) In this case the study may be valid for his patients (called internal validity) but may not be valid to be generalized to people with acne in general (so may lack external validity).

Probability - the probability of an event is denoted by p. Probabilities are usually expressed as decimal fractions and lie between 0 (zero probability) and 1 (absolute certainty). They are not expressed as percentages. Probability of events cannot be negative. The probability of an event can also be expressed as a ratio of the number of likely outcomes to the number of likely outcomes (50% heads and 50% tails in a coin toss). Probability of an event not occurring is q. q=l-p.

Key probability Rules -

Addition - The addition rule of probability states that the probability of any one of several particular events occurring is equal to the sum of their individual probabilities, provided the events are ***mutually exclusive.*** Probability of picking a heart from a deck of cards is .25 and the probability of picking a diamond card is also .25 so the probability of picking either a heart or a diamond is ,25+.25= 0.50. These events are mutually exclusive because no card can be both a heart and a diamond.

If two events are ***not mutually exclusive e.g.*** chance of having diabetes is 10% and chance of being obese is 30% then probability for someone obese and having diabetes is calcuted as follows - 0.1+0.3-(0.1*0.3) = 0.4-.003 = 0.37 or 37%.

Multiplication - rule of probability states that the probability of two or more statistically ***independent*** events all occurring is equal to the product of their individual probabilities. E.g. probability of a person having cancer is .25 and having schizophrenia is .01, then the lifetime probability of a person having both is 0.25*0.01 = 0.0025 provided that the two illnesses are independent (having one does not increase or decrease the risk of having the other). If the events are ***dependent*** then multiply the probability of one with the probability of the other assuming that the first has occurred. E.g. A box has 5 white balls and 5 black balls. The chance of picking two black balls is (5/10) * (4/9) = ,5*.44= .22 or 22%.

Binomial distribution - The probability that a specific combination of mutually exclusive independent events will occur can be determined by the use of binomial distribution. Binominal distribution has only two possibilities such as yes/no, male/female. If an experiment has only two possible outcomes then and one is generally termed success then the binomial distribution gives the probability of obtaining an exact number of successes in a series of independent trials. Medical use of this is genetic counseling. It has two possible outcomes - Inheriting the disease or not, mutually exclusive - a person cannot both have and not have the disease and independent - one child inheriting the disorder does not affect the chances of the other child inheriting it.

Types of data - (NOIR) Data will always form one of four scales of measurement - **nominal, ordinal, interval or ratio.** N - data in groups, O - Data in meaningful order with no info on the size of the interval, I - Data in meaningful order with meaningful intervals between them and R - Data as a ratio. Data may also be **discrete or continuous.** Discrete values can take only certain values and none in between e.g. patients maybe 78 or 79 not in between. Continuous variables may take any value (between certain limits). E.g. patient's height, weight, age etc.

Frequency distributions - A set of unorganized data is difficult to understand. A simple way to organize the data is to list all values between the highest and lowest in order recording the frequency with which each score occurs. This forms a frequency distribution.

Grouped Frequency distributions - The above data is unwieldly. It can be made more manageable by creating a grouped frequency distribution where individual scores are grouped into 7-20 groups with an equal class interval of 10.

Relative frequency distributions - Grouped frequency distributions can be transformed into relative frequency distributions, which show the percentage of all the elements that fall in each class interval. This is found by dividing the frequency of the number of elements by the sample size.

Cumulative frequency distributions - shows the percentage of elements lying within and below each class interval.

Graphical representation of frequency distributions - Graphs or histograms. The X-axis shows the grouped scores and the Y-axis shows the frequencies. For nominal data bar graph is typically used. They are identical to histograms except they are separated from each other by a space. For ratio or interval scale data frequency distributions may be drawn as frequency polygons in which the midpoints of each class interval is joined by straight lines. A cumulative frequency distribution can also be presented graphically as a polygon, which forms a characteristic S shaped curve (ogive).

Centiles - The cumulative frequency polygon and the cumulative frequency distribution both illustrate the concept of of centile (or percentile) rank, which states the percentage of observations that fall below any particular score. Centile ranks provide a way of giving information about one individual score in relation to all the other scores.

Normal Distributions - The data presented as distributions are summarized by a central tendency and variation around that center. The most important distribution is the normal or Gaussian curve. This is bell shaped symmetric and one size is the mirror image of the other.
Skewed, J shaped and bimodal distributions - Asymmetrical distributions are called skewed distributions. Positive skew has a tail to the right because it a larger no. of low scores and small no. of high scores. In a positive skew the mean is greater than the median. Negative skew has a tail to the left because it a low no. of low scores and high no. of high scores. In a negative skew the median is greater than the mean.
Measures of central tendency - central tendency is a general term for several characteristics of the distribution of a set of values or measurements around a value at or near the middle of the set. These are Mean, Median and Mode.
Mode(Mo): is the observed value that occurs most frequently in a set of observation. If two scores both occur with the greatest frequency the distribution is bimodal. If more than two scores occur with the greatest frequency then the distribution is multi modal.
Mo is uninfluenced by the small numbers of extreme scores in a distribution.
Median (Md) - is the figure that divides the frequency distribution in half when all the scores are listed in order. When elements are an odd number the median is the central one and when the elements are an even number then the median is the average of two middle scores. Md is uninfluenced by the small numbers of extreme scores in a distribution.
Mean X: a synonym for average. It is the sum of all the elements divided by the total number of elements in the distribution. It is symbolized by p in a population and by X-bar in a sample, p = SX/N in a population and X-bar = EX/n in a sample. Mean is very sensitive to extreme scores. So it is not usually an appropriate measure for characterizing very skewed distributions. But when repeated samples are drawn from the same population they tend to have similar means so mean is a measure of central tendency that best resists the influence of fluctuation between different samples. The relationship between the three measures of central tendency depends on the shape of the distribution. In a unimodal symmetrical distribution all three are identical but in a skewed distribution they will differ. See page 12.
Measures of variability - Two normal distribution that are symmetrical and unimodal may have identical mean, mode and median. But they differ in terms of their variability - the extent to which their scores are clustered together or scattered about. There are 3 important rules of variability - range, variance and standard deviation.
Range - is the simplest measure of variability. It is the difference between the highest and the lowest score. Range is unstable and changes easily. SD is a more stable and useful measure of deviation.
Variance & deviation scores - Calculating variance and SD involves use of deviation scores. The deviation score of an element is found by subtracting the distribution mean from the element. A deviation score is x whereas an element is X. In a distribution with a mean of 16 an element of 23 would have deviation score of 23-16 = 7. The total of deviation scores for all elements will be zero (negative or positive). To get variance find deviation score x for each element, square each deviation score to eliminate the minus signs and then obtain their mean by adding all up and dividing the total by their number. Variance is sigma square for the population and S square for the sample. Variance is known as the mean square. Also the square of the standard deviations equals the variance.
Standard deviation - It is square root of the variance. It is expressed in the same unit of measurement as the original data. The SD is useful in normal deviations because the proportion

of elements in the normal curve a constant proportion of the cases fall within one, two and three SD of the mean. These proportions hold true for every normal distribution.
Within one SD - 68%.
Within two SD - 95%.
Within three SD - 99.7%
Z scores - The location of any element in a normal distribution can be expressed in terms of how many standard deviations it lies above or below the mean of the distribution. This is the z score of the element. If the element lies above the mean it will have a positive z score if it lies below the mean it will have a negative z score. All points in a z score distribution are represented in SD units. In a distribution of heart rates, the mean is 70 and SD is 10. In this distribution a heart rate of 85 lies 1.5 SD above the mean so it has a z score of 1.5 whereas a heart beat of 65 lies -0.5 SD below the mean and has a z score of-0.5. z= (X-Mean of population)/SD of population.

$$z = \frac{X - fl}{o},$$

where:

x is a raw score to be standardized;
p is the mean of the population;
o is the standard deviation of the population

The z score is easy to use for calculations because it has simple values and all points in a z score are represented in SD units.
Tables of Z scores - state what proportion of any normal distribution lies above any given z scores, not just z scores of ± 1,2 or 3.
Using z scores to specify probability - Z scores allow us to specify the probability of a randomly picked element being above or below a particular score. If we know that 5% of the population has a heart rate above 86.5 then the probability of one randomly selected person having a heart rate above 86.5 will be 5% or .05. We can find the probability that a random person will have a pulse less than 50. 50 lies 2 SD below the 70 the mean, so it corresponds to a z score of-2. We know that 95% of distribution lies within the limits of ± 2. Therefore 5% of the distribution lies outside these limits, equally in each of the two tails of the distribution. So 2.5% lies below 50 and the probability that a randomly selected person has a pulse less than 50 is 2.5% or .025.

2) Inferential statistics — involves generalization from the sample to the population as a whole. Because we can rarely observe and study entire populations we try to select samples that are representative of the entire population so that we can generalize the results from the sample to the population. The purpose of inferential statistics is to designate how likely it is that a given finding is simply a result of chance. In these cases the population parameters (population mean and SD) are unknown but sample statistics (sample mean and SD) are known. Inferential statistics uses statistics to estimate parameters. It is unlikely that a sample will perfectly represent the population it is drawn from. There will be sampling error - which is not an error but just natural, expected random variation.
The random sampling distribution of means - If you take many samples from the same population and calculate their respective means, then all the means will vary due to sampling error. If you plot all the means in a frequency distribution the sample means form a distribution called random sampling distribution of means. You will also note that this distribution looks like a normal distribution. It will always tend to be normal per central limit theorem irrespective of

the shape of the original population distribution from which the samples were drawn, even if the underlying population were not normally distributed. This theorem also states that the random sampling distribution of means will become closer to normal as the size of the sample increases. This theorem also states that the mean of the random sampling distribution of means is equal to the mean of the original population. This distribution of random sampling distribution of means also has a mean and a SD. This SD is also called standard error (SE) or standard error of means (SEM). It is a measure of the extent to which the sample means deviate from the true population mean. The standard error of a statistic is the standard deviation of the sampling distribution of that statistic. Standard errors are important because they reflect how much sampling fluctuation a statistic will show. The inferential statistics involved in the construction of confidence intervals and significance testing are based on standard errors. The SE or SEM of a statistic or random sampling distribution of means depends on the sample size. In general, the larger the sample size the smaller the standard error or SEM. If the sample is small the distribution will have a broader fatter shape and higher SE but if the sample size is more than the distribution will have a thinner narrower curve and a lower SE. The standard error of a statistic is usually designated by the Greek letter sigma (o) with a subscript indicating the statistic.

$$se_{\bar{x}} = \frac{s}{\sqrt{n}}$$

$$SD^* = \frac{\sigma}{\sqrt{n}}$$

SE = Population SD/square root of the size of the samples. This shows that SE is inversely related to the square root of the sample size.

Predicting the probability of drawing samples with a given mean -Because the random sampling distribution of means is normal, known facts about the distribution i.e. SD, z score etc can be used to find other inferences and predictions provided the sample is random. Even if the original population is not normal the random sampling distribution of means is normal, and z scores can be used to make predictions.

Using SE - The method used to make predictions about a sample mean involves finding the z score corresponding to the value of interest. In a random sampling distribution of means, the z score is calculated in terms of the number of standard errors by which the sample mean lies above or below the sample mean. E.g. in a population with a mean heart rate of 70 and a SD of 10, the probability that a random sample of 25 people will have mean heart rate of 75 can be calculated. The steps are:

1) Calculate the SE = 10/sq root of 25 = 10/5 = 2.
2) Calculate the z score of the sample mean = 75-70/2 = 5/2 = 2.5.
3) Find the proportion of the normal distribution that lies beyond this z score 2.5. As per z score table in book we can see that it is .0062. Therefore the probability that a random sample of 25 people from this population will have a mean heart rate above 75 is .0062.

We can also find which random sample mean (n=25) is so high that it would occur in only .05 or less in all samples or the probability of obtaining it is .05 or less. Z score table shows that z score that divides 95% from 5% is 1.65. The corresponding heart rate is the population mean plus 1.65 standard errors. As population mean is 70 and the SE is two heart rate will be 70+ (1.65*2) = 73.3. It is also possible to find the limits between which 95% of all possible random samples means would be expected to fall. In any normal distribution 95% of the random sampling distribution of means lie approximately ± 2 SE of the population mean or z= ± 2. z score table shows that the

exact z score that correspond to the middle 95% of any normal distribution are in fact ± 1.96 not ±2. The exact limits are therefore 70 ± (1.96*2) = 66.08 and 73.92. It is apparent that 95% of all possible random sample means will fall between the limits of 66 and 74.

Estimating the mean of a population -z scores are used to find the probability that a random sample will have a mean of above or below a given value. It has been shown that 95% of the population will lie within approx ± 2 or exactly ± 1.96 SE of the population mean, and 95% of all such means will be within ± 2 SE of the mean.

Confidence limits - It *is* possible to use your sample to calculate a range within which the population value is likely to fall. "Likely" is usually taken to be "95% of the time," and the range is called the **95% confidence interval.** The values at each end of the interval are called the **confidence limits.** Logically if the sample mean lies within ± 1.96 SE of the population mean, 95% of the time, then population mean must lie within ± 1.96 SE of sample mean, 95% of the time. These limits of ± 1.96 SE are called confidence limits. In general confidence limits are equal to the sample mean ± the z score obtained from the table (for the appropriate level of confidence) multiplied by the standard error. Therefore 95% confidence limits (ones used in conventional medical research) are approximately equal to the sample mean plus or minus two standard errors. E.g. If mean heart rate of the sample is 74 and the SE of the sample is 2, the researcher can be 95% certain that the population mean lies within 1.96 SE of 74 or 74 ± 1.96*2 = 70.08 and 77.92.

Confidence Intervals - are a way of admitting that any measurement from a sample is only an estimate of the population. Although the estimate given from the sample is likely to be close, the true values for the population may be above or below the sample values. A confidence value specifies how far above or below a sample-based value the population value lies within a given range from a possible high to a possible low. Reality therefore is most likely to be somewhere in the specified range (The difference between the upper and lower confidence limits is called confidence interval or CI). Researchers want the confidence interval to be as narrow as possible. To make this narrower the standard error must be made smaller. The only way to reduce the standard error is to increase the sample size. To halve the confidence interval the sample size must be increased fourfold. See SE.

Confidence Interval of the mean -

The confidence interval contains two parts. Standard error of the mean and z score. The sample mean is the usual estimator of a population mean. However, different samples drawn from that same population would in general have different values of the sample mean. The standard error of the mean (i.e., of using the sample mean as a method of estimating the population mean) is the standard deviation of those sample means over all possible samples (of a given size) drawn from the population. Secondarily, the standard error of the mean can refer to an estimate of that standard deviation, computed from the sample of data being analysed at the time. The confidence interval of the mean can be calculated by - Mean ± appropriate z score * standard error of the mean = X bar ± Z * SE. (SE is equal to SD divided by the square root of the sample size).

Precision - Precision is a degree to which a figure (such as an estimate of a population mean) is immune from random variation. The width of a confidence interval reflects precision - the wider the confidence interval the less precise the estimate.

Accuracy - is the degree to which an estimate is immune from systematic error or bias.

The narrower the distribution the more precise and the more the sample mean of the distribution coincides with the population mean the more accurate it is.

Estimating the standard error - SE = sigma (SD)/square root of sample. Sigma is the population SD and in practice it will not be known. So SE cannot be calculated and therefore z scores cannot

be used. Instead SE can be estimated using data that are available from the sample alone. The resulting statistic is the estimated SE of the mean called estimated SE. Estimated standard error of the mean is = sample SD/ square root of (sample-1).

t scores - The estimated SE is used to find a statistic called t that can be used in place of z. t scores instead of z scores must be used when making inferences about means that are based on estimates of population parameters e.g. estimated SE rather than the population parameters themselves. T score is calculated in the same way as z. Z score tables give the proportions of the normal distributions that lie above and below any z score and t score tables give the same information for any t score. However value of z for any given proportion of distribution is constant but the value of t for any given proportion is not constant but varies from one sample to the next. As the sample gets larger z and t scores get increasingly similar and as the sample gets smaller they get increasingly different.

Degrees of freedom and t tables - Tables of t values do not show sample size directly. Instead they express sample size in degrees of freedom. Degrees of freedom = n-1. Therefore to determine the value of t for a sample size of 15 one would look in the table for the appropriate value of t for df= 14 (15-1). E.g. In a sample of 15, mean heart rate is 74. SD is 8.2 so estimated SE is = 8.2/ sq rt of 15-1 = 8.2/3.74 = 2.2. For a sample of 15, t scores for a df of 14 (15-1) is 2.145 in Table 2-1. (z scores is 2). The 95% confidence intervals are therefore equal to the sample mean plus or minus t times the estimated SE = 74 ± (2.145*2.2) = 69.281 and 78.719.

Because the figure for 95% confidence intervals is almost invariably going to be in the region of 2 it should be noted that in general one can be 95% confident that the true mean of the population lies within approx, plus or minus two estimated SE of the mean of a random sample drawn from that population.

3) Hypothesis testing

To test a hypothesis about a mean the steps are as follows

1. The first step in any hypothesis testing is to state the relevant null and alternative hypotheses to be tested
2. Select the decision criterion oc (or level of significance).
3. Establish the critical values.
4. Draw a random sample from the population and calculate mean of that sample.
5. Calculate the SD and estimated SE of the sample.
6. Calculate the value of the test statistic t that corresponds to the mean of the sample.
7. Campare the calculated value of t with the critical values of t and then accept or reject the null hypothesis. The decision rule is to reject the null hypothesis (HO) if S > CV and vice versa.

Step 1: State the null and alternative hypotheses - E.g. Dean of a medical school states the average IQ of the students of the school is 135. This claim is a hypothesis that can be tested and is a null hypotheses or Ho. If this hypothesis is rejected as false then there is an alternate hypothesis H_a which logically must be accepted. In this case

Null Hypothesis Ho p = 135

Alternative Hypothesis H_a p ≠135

One way of testing the null hypothesis is to test the IQ of each student of the school, i.e. the entire population. This would be expensive and time consuming. It would be more practical to draw a sample of random students, find their mean IQ and draw an inference from this sample.

Step 2: Select the decision criterion - If the null hypothesis is correct then the mean sample of the IQ would not be 135 due to sampling error. To reach a conclusion it must therefore be decided at what point is the difference between the sample mean and 135 not due to sampling error but due to the fact that the population mean is not really 135 as the null hypothesis claims. This point must be set before the sample is drawn and the data are collected. Instead of setting it in terms of IQ scores it is set in terms of probability. The probability level at which it is decided that the null hypothesis is incorrect constitutes a criterion, or significance level known as alpha a. It is unlikely that a random sample mean will be very different from the true population mean. The greater the difference between the sample mean and the population mean, the less probable it is that the sample does come from the specified population. When this probability is very low it can be concluded that the null hypothesis is incorrect. By convention the null hypothesis will be rejected if the probability that the sample mean could have come from the hypothesized population is less than, or equal to .05; Conversely if the probability of obtaining the sample mean is greater than .05, the null hypothesis will be accepted as correct. Although alpha may be set lower than the conventional .05, it is not normally any higher than this.

Step 3: Establish the critical values - Earlier it was shown that if a very large number of random samples are taken from any population their means form a normal distribution - the random sampling distribution of means - which has a mean equal to population mean. It was also shown that one can state what random sample means are so high or so low that they would occur in only 5% or fewer of all posssible random samples. This ability can now be put to use by stating which random sample means are so high or so low that they would occur in only 5% or fewer of all random samples that could be drawn from a population with a mean of 135. If the sample mean falls inside the range within which 95% of random samples means would be expected to fall, the null hypothesis is accepted. The range is therefore called the **area of acceptance.** If the sample mean falls outside this range in the **area of rejection,** the null hypothesis is rejected and the alternative hypothesis is accepted. The limits of this range are called the critical values and they are established by refering to a table of t scores. In the current example the following values can be calculated:

- The sample size is 10 so there are (n-1) =9 degrees of freedom.
- The table of t scores (Table 2-1) shows that when df=9, the values of t that divides the 95% (.95) area of acceptance from the two 2.5% (.025) areas of rejection is ± 2.262. These are critical values. (See Figure 3-1 to understand).

The following have now been established:

- The null and alternative hypotheses
- The criterion that will determine when the null hypothesis will be accepted or rejected.
- The critical values of t associated with this criterion.

A random sample of students can now be drawn from the population; the t score associated with their mean IQ can then be calculated and compared with the critical values of t. This is a t test.

Step 4: Draw a random sample from the population and calculate the mean of that sample - A random sample of ten students are drawn and their IQ's are as follows: 115...140...133...125...120...126...136...124...132...129. The mean of this sample is 128.

Step 5: Calculate the SD and estimated SE of the sample - To calculate the t score corresponding to the sample mean, the estimated SE must first be found. Here the SD of this sample is calculated to be 7.155. The estimated SE is then calculated as follows:

Est. SE = SD/Square root of n-1. = 7.155/3 = 2.385.

Step 6: Calculate the value of t that corresponds to the mean of the sample - After Estimated SE is found the t score corresponding to the sample mean can be found. It is the number of est. SE by which the sample mean lies above or below the hypothesized population mean, t = sample - null hypothesis / est. SE = 128-135/2.385 = -2.935. So the sample mean lies approx. 2.9 est. SE below the hypothesized population mean.

Step 7: Compare the calculated value of t with the critical values of t and then accept or reject the null hypothesis - If the calculated value of t associated with the sample mean falls at or beyond either of the critical values it is within one or two areas of rejection. In this e.g. the t score in this example does fall within the lower area of rejection. Therefore the null hypothesis is rejected and alternative hypothesis is accepted. This is because the difference between the sample mean and the hypothesized population mean is so **statistiscally significant** that the null hypothesis has been rejected at the .05 level i.e. the probability is only .05 or less that the sample mean could be obtained if the null hypothesis was true. This would be reported as follows: " The hypothesis that the mean IQ of the population is 135 was rejected, t= 2.935, df = 9, p≤ .05". If on the other hand the calculated value of t associated with the sample mean fell between the two critical values, in the areas of acceptance, the null hypothesis would be accepted. It would be said that the difference between the sample mean and the hypothesized population mean failed to reach statistical significance (p > .05).

Z -Tests - A Z-Test involves the same steps as a t test and can be used when the sample is large enough (n > 100) for a sample SD to provide a reliable estimate of SE. There are situation when Z test cannot be used but no situations where T tests cannot be used so T tests are more important and widely used of the two.

The meaning of statistical significance - When the result is reported to be significant at p ≤ .05 it only means that the result was unlikely to have occurred by chance and the likelihood of the result having occurred by chance is .05 or less. It does not mean that the result is important or clinically significant. If the sample was large enough and IQ was found to be 134 it could fall in the area of rejection and the null hypothesis could be rejected. In fact any null hypothesis could be rejected if the sample is large enough because there would always be some trivial difference between the sample mean and the hypothesized mean. Therfore the findings would be statistically significant but otherwise insignificant.

Type I and II errors - A statement that the result is significant at p ≤ .05 means that an investigator can be 95% sure that the result was not obtained by chance. It also means that there is a 5% probability that the result could have been obtained by chance. Although the null hypothesis is being rejected it could still be true: there remains a 5% chance that the data did in fact come from the population specified by the null hypothesis. If the null hypothesis is true the decision we are making on the basis of the error could be in error. There are two possible types of error we could make.

Type I error *(a* error) -: rejecting the null hypothesis when it is actually true i.e. assuming a statistically significant effect on the basis of the sample when there is none in the sample e.g. deciding the IQ is not 135 when it is. The chance of type I error is given by p value. If p = .05 then the chance of a type I error is 5 in 100 or 1 in 20. (False - negative error.)

Type II error (B error) -: Failing to reject the null hypothesis when it is actually false, i.e. declaring no significant effect on the basis of the sample when there really one in the population e.g. deciding the IQ is 135 when it is not. The chance of a type II error cannot be directly estimated from the p value. (False - positive error.)

The choice of an appropriate level for the criterion therefore depends on the relative consequences of making a type I or type II error. E.g. if a study is expensive and time consuming and so unlikely to be repeated but has practical implications the researchers may wish to establish a more stringent level of criterion (.01, .005 or even .001) to be more than 95% sure that their conclusions are correct. Although the criterion to be selected need not be .05 by convention it cannot be any higher. Many researchers do not report their results in terms a predetermined criterion but report the actual probability that the obtained result could have occurred by chance if the null hypothesis were true (p = .015). The researchers are showing how unlikely it is that a type I error has been made even though they would have still rejected the null hypothesis if the outcome were only significant at .05 level.

Power of statistical tests - Although it is possible to guard against a Type I error by using more stringent levels of criterion, preventing a Type II error is not so easy. A Type II error involves accepting a false null hypothesis, the ability of a test to avoid a Type II error depends on its ability to detect a null hypothesis is false. This ability is called power of the test and is equal to 1- 0. A tests power or ability to detect a false null hypothesis will increase as

- Alpha increases (e.g. from .01 to .05). This increases the size of the areas of rejection and makes rejection of null hypothesis more likely. Increasing alpha reduces the chance of Type II error but increases the chance of Type I error.
- Increasing the size of the sample. This reduces the estimated SE and increases the calculated value of t. Increasing the sample size is the most practical and important way of increasing the power of a statistical test.

Directional Hypotheses - In the above e.g. the dean claimed that the IQ of the students was 135. This was the null hypothesis and could be rejected if the sample mean turned out to be significantly above or below 135. So there were two areas of rejection. The above hypothesis was a non-directional one. If the dean had said that the IQ of the students was at least 135 then the claim could only be rejected if the sample mean IQ turned out to be significantly lower than 135 and the null hypothesis would be $p \geq 135$ and alternative hypothesis would be $p < 135$. The alternative hypothesis is now a directional one, which specifies that the population mean lies in a particular direction with the null hypothesis. Now there is only one area of rejection, which lies at only one tail of the distribution, rather than in both tails. The steps in conducting a t test are the same as before critical value is now different. The critical value of a one tailed test is less extreme (-1.833) than a two-tailed test (± 2.262). One-tailed tests are more powerful than two tailed tests.

Testing for differences between groups - In biomedical groups' researchers often want to compare two means such as the effect of two drugs or the mean survival times of patients receiving two different treatments. Many research problems involve comparing more than two groups. E.g. comparing the treatment outcomes of three groups of patients with depression - one group taking a placebo, one taking a tricyclic antidepressant and the third a selective serotin reuptake inhibitor (SSRI). Each group contains both sexes and if the researcher wants to make comparisons between sexes then there are total six groups. Multiple t testing comparing one with others and so on but this method has disadvantages. There is a technique that overcomes these problems: analysis of variance or ANOVA. ANOVA is used very commonly when means of more than two groups are being compared.

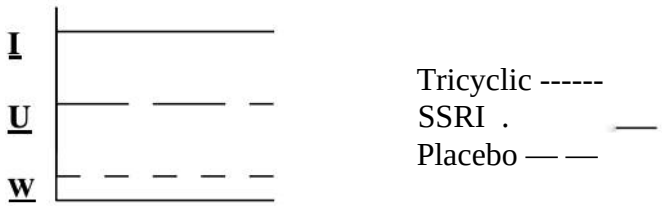
Analysis of Variance (ANOVA) - There will be variability in any set of experimental results which are due to variability resulting from 1) known differences between the groups e.g. gender, use of placebo or SSRI etc. and 2) The ordinary random variability within each group that is to be expected caused by Sampling error and individual differences between patients etc. If the variance

between the groups is large in comparison to random fluctuations then it must be due to some difference between the groups above and beyond the random fluctuations. If the experiment has been performed correctly then the difference must be due to (in this e.g.) the gender of the patients or the difference of treatment because there is no other random difference between groups.

The F-Ratio - ANOVA compares the variance by means of F- Ratio:

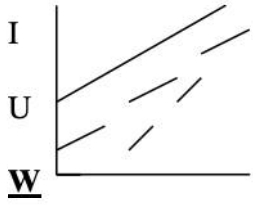
F= variance between groups/variance within groups. The resulting F statistic is then compared with the critical values of F obtained from F tables in the same way as t. If the calculated value exceeds the critical value for the appropriate value of alpha the null hypothesis will be rejected. If the various experimental groups differ in terms of only one factor at a time - such as the type of drug being used - a **one-way** ANOVA is used. If the variance between groups is sufficiently large, compared with the variance within groups, for the F ratio to reach significance it would then be known that the drug type was a significant source of variation in the result. If the various experimental groups differ in terms of only two factors at a time - e.g. the type of drug being used, and the gender - a **two-way** ANOVA is used. This ANOVA will show not only if there are significant sources of variability in the results - it will also show if this variability is attributable to one factor (drug type), to the other factor (gender), or to the two factors in combination with each other. If a single factor is found to have a significant effect it is called the **main effect.** If a combination of factors has a significant effect it is called an **interaction effect.** Interaction effects occur when the effects of two factors together differs from the sum of the individual effects of each alone.

Graphical presentations of the ANOVA data - Main and interaction results are easily understood visually.



I = Improved
U= Unchanged
W= Worse

1) Best outcomes in patients who took tricyclics, followed by SSRI and worst in Placebo. No change due to gender.



2) Best outcomes in patients who took tricyclics, followed by SSRI and worst in Placebo. Female patients had a better outcome.

To look at other graphs see page 44-45

Non parametric and distribution free tests - The t, z, F tests share 3 common features.

- Their hypotheses refer to population parameters: the population mean in case of z and t tests or the population variance in case of F tests. For this reason they are called **parametric** tests.
- Their hypotheses concern interval or ratio scale data, such as weight, blood pressure, IQ, measures of clinical improvement and so on.
- They make certain assumptions about the distribution of the data of interest in the population - principally that the population data are normally distributed.

Some other statistical techniques do not share these features:

- They do not test hypotheses concerning parameters, and hence are known as **non-parametric** tests.
- They do not assume that the population is normally distributed, so they are also called **distribution-free** tests.
- They are used to test nominal or ordinal scale data.

These tests are less powerful than parametric tests.

Chi square (x2)_ This is the most important non-parametric test, which is used for testing hypotheses about nominal scale data. Chi-square is basically a test of proportions telling us whether the proportions of observations falling in different categories differ significantly from the proportions that would be expected by chance. E.g. tossing a coin hundred times we would expect 50% heads. If the result were 59 heads and 41 tails then chi-square would show whether this difference in proportion is too large to be expected by chance i.e. statistically significant. Chi-square can be used to test any number of groups.

Number of students passing Dental Boards by dental school

	School A	SchoolB	School C	
Passing Number	49	112	26	Total 187
Number Failing	12	37	8	Total 57
Total	61	149	34	

Contingency Table

Chi-square might compare the pass rates of Dental Boards Exam results of different schools as shown in the above table. This is called a contingency table, which is the usual way of presenting this kind of data. It expresses the idea that one variable (passing or failing) may be contingent on the other variable (medical school attended). Chi-square can answer the questions - is there a relationship between the school attended and the passing and failing of the examination.

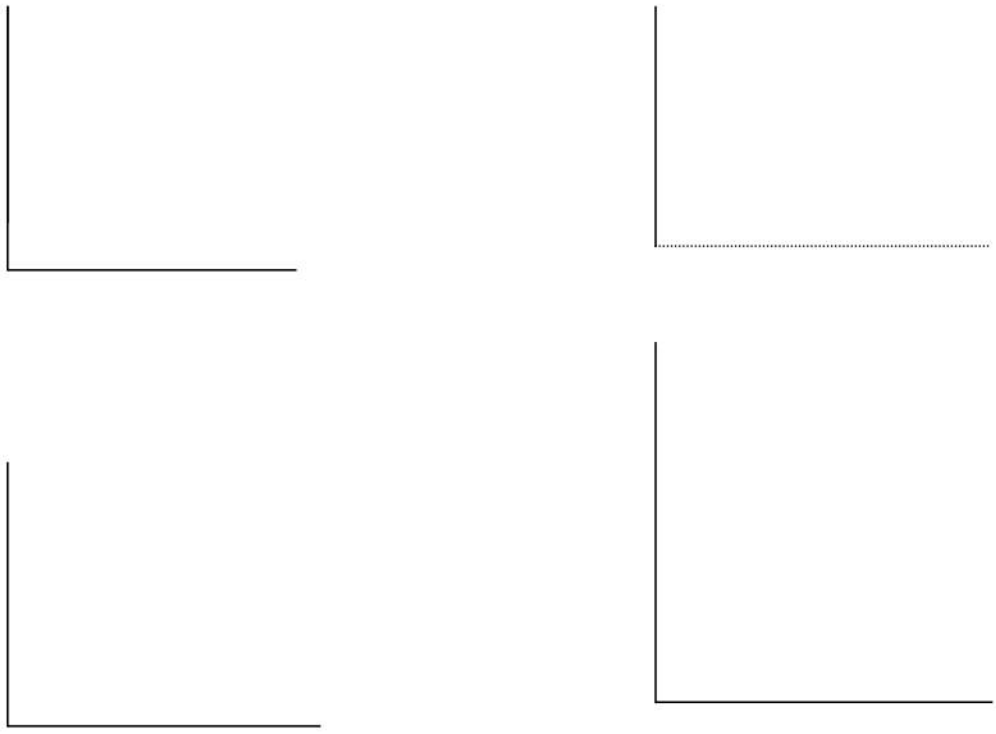
4) Correlational Techniques - Biomedical research often seeks to establish relationship between two variables e.g. salt intake and blood pressure or smoking and life expectancy. **Correlational techniques** are used to do this. They measure the co-relatedness of the two variables. There are two basic kinds of correlational techniques:

- **Correlation,** which is used to establish and quantify the strength and direction of the relationship between two variables.

- **Regression,** which is used to express the functional relationship between two variables so that the value of one variable can be predicted from the knowledge of the other.

Correlation - expresses the strength and direction of the relationship between two variables in terms of a correlation coefficient, signified by r. Values of r vary from -1 to +1; the strength of the relationship is indicated by the size of the coefficient, while its direction is indicated by the sign. A plus sign means there is a **positive correlation** between the two variables e.g. high values of one variable (salt intake) are associated with the high values of the other variable (blood pressure). A minus sign means there is a **negative correlation** between the two variables e.g. high values of one variable (cigarette consumption) are associated with the low values of the other (life expectancy). If there is a **perfect** linear relationship between the two variables it is possible to know the exact value of one variable from the knowledge of the other variable the value of correlation coefficient r will be exactly plus or minus 1. If there is no relation between the two variables so that it is impossible to know anything about one variable based on the knowledge of the other variable then the coefficient r will be zero. Coefficients beyond ± 0.5 are regarded as strong and coefficients between zero and ± 0.5 are usually regarded as weak. The further from zero the stronger the relationship.

Scattergrams and bivariate distributions - The relationship between these two correlated variables form a bivariate distribution, which is commonly presented graphically in the form of a scattergram. The first variable (salt intake, cigarette consumption) is usually plotted on the horizontal (x) axis and the second variable (blood pressure, life expectancy) is plotted on the vertical (y axis). Each plotted data point represents one observation of a pair of values such as one patient's salt intake and blood pressure so the number of plotted points is equal to the sample size n. **See pg 27 kaplan**



First all observations are plotted. The **line of best fit** is found to the plotted data points. The relationship between the appearance of the scattergram and the correlation coefficient can be understood by imagining how well a straight line would fit the plotted points. If it is not possible to draw any straight line that would fit the plotted points at all the correlation coefficient is approximately zero. If a straight line would fit the plotted points perfectly the correlation coefficient is 1.00. Correlation itself does not mean causation. A Correlation coefficient indicates the degree to which the two measures are related not why they are related.

Types of correlation coefficient - there are two important types. **Pearson product -moment correlation (r),** which is used for interval or ratio scale data (e.g. salt intake and blood pressure-both ratio scale data). **Spearman rank-order correlation (p - phi),** which is used for ordinal scale data (e.g. association between birth order and class position at school-both ordinal scale data). Both these corelational techniques are linear: they evaluate the strength of a straight-line relationship between the two variables. So if there is a strong nonlinear relationship between the two variables the Spearman or Pearson will be an underestimate of the true strength of the relationship. E.g. a drug has a strong effect at medium dosages but weak effects at high or low dosages (graph will be bell shaped). Because the relationship is nonlinear r will be low even though there is a strong relationship between the two variables. Visual inspections of scattergrams help in identifying this type of relationship.

Coefficient of determination -Once a correlation coefficient has been determined the coefficient of determination can be found by squaring the value of r. Coefficient of determination r square expresses the proportion of variance in one variable that is accounted for or explained by the variance in the other variable. So if r between salt intake and BP was found to be 0.40 then it can be concluded that .40*.40 = 0.16 or 16% of the variance in BP in this study is accounted for by variance in salt intake. It does not mean that the changes in salt intake caused the change in BP. A correlation between two variables does not demonstrate a causal relationship between the two variables, no matter how strong it is. Correlation is merely a measure of the variables statistical association, not of their causal relationship. Inferring a causal relationship between two variables on the basis of a correlation is a common and fundamental error.

The fact that a correlation is present between two variables in a sample does not necessarily mean that the correlation actually exists in the population. When a correlation has been found between two variables in a sample a researcher will normally wish to test that there is no correlation between the two variables (i.e. r=0) in the population. This is done by a special form of t-test.

REGRESSION - If two variables are highly correlated then it is possible to predict the value of one of them (dependant) from the value of the other (independent) by using regression techniques. In simple linear regression the value of one variable X is used to predict the value of the other variable Y by means of a simple linear mathematical function, the regression equation which quantifies the straight-line relationship between the two variables. The straight line is also called **regression line** and is the line of best fit used to calculate the correlation coefficient. Simple linear regression equation is

Expected value of Y = a + bX

Where x and y are the variables

a = is a constant known as the intercept constant because it is the point on the Y-axis where the Y-axis is intercepted by the regression line.

b = The slope of the regression line and is known as the regression coefficient.

Once the values of a and b have been established the expected value of Y can be predicted for any given value of X.

Multiple regression - In **multiple regression** equations more than one variable is used to predict the expected value of Y.

Choosing an appropriate inferential or correlational technique - The basic choice for a research problem is determined by two factors: the scale of measurement and the type of question being asked.

Nominal scale data - question discussed is - Do the proportions of observations falling in different categories differ significantly from the proportions that would be expected by chance? The technique is **chi-square test.**

Ordinal scale data - question discussed is - Is there an association between an ordinal position on one ranking and ordinal position on another ranking? The technique here is **Spearman rank-order correlation.**

Interval or ratio-scale data - three kinds of questions have been discussed -

1. Question concerning means:
   - What is the true mean of the population?
   - Is one sample mean significantly different from one or more other sample means.?
2. Questions concerning variances:
   - Are the variances in two samples significantly different?
3. Questions concerning association:
   - To what degree are two variables correlated

	Scale of Data		
	Nominal	Ordinal	Interval or Ratio
Differences in **proportion**	**x square**		
One or two Means			t-test (z test if n > 100)
More than two Means			ANOVA with F-tests
Variances			F-test
Association		Spearman p	Pearson r
Predicting the Value of a variable			Regression

Summary for AD AT

5) Research Methods - Research typically attempts to uncover the relationships between independent variables and dependant variables. Independent variables are presumed to be causes of changes in other variables, which are called dependant variables because they are presumed to be dependant on the values of independent variables.

EXPERIMENTAL STUDIES- The relationships between independent variables and dependant variables can be experimented in two ways:

- Experimental or intervention studies in which researcher exercises control over the independent variables, deliberately manipulating them. In a research about the effects of a certain drug effectiveness for a disease the investigator would give the drug to one group of patients and not to the other. Properly conducted experiments are the most powerful way of establishing cause and effect relationship between the independent and the dependant variables. They have disadvantages also, like they may be unethical if they expose the subjects to the risk of physical or mental harm, and they are impractical if the cause and effect relationship takes a long time to appear.
- Nonexperimental or observational studies, in which nature is simply allowed to take its course. In this the investigator would simply observe patients who had or hadn't taken the drug in the normal course of events.

Clinical Trials - attempts to evaluate the effects of a treatment. Clinical Trials aim to isolate one factor (use of a drug) and examine its contribution to a patient's health by holding all other factors as constant as possible. Apart from manipulation or intervention, clinical trials typically have two other characteristics: they utilize control groups and involve randomization. Hence, they are often termed randomized controlled clinical trials.

Control Groups -Patients in clinical trials are divided into two general groups:

- the experimental group which is given the treatment under investigation; and
- the control group which is treated in exactly the same way except that it is not given the treatment

At the end of the study the difference that appears between the two groups can be attributed to the treatment under investigation. Control group helps to eliminate alternative explanations for a study's result. E.g. if a drug eliminates all symptoms of an illness in a group of patients in a month, it may be that the symptoms would have disappeared spontaneously over time even if this drug was not used. But if a similar control group of patients did not receive the drug and experienced no improvement in their symptoms, this alternative explanation is untenable.
There are two main types of medical groups used in medical research.

- The no-treatment control group: where patients receive no treatment at all. This leaves open the possibility that the patients whose symptoms were removed by the drug were responding not to the specific pharmacologic properties of the drug, but to the non-specific placebo effect that is part of any treatment.
- The placebo control group who are given an inert placebic treatment allowing the elimination of the explanation that patients in the treatment group who improved were responding to the placebic component of the treatment; so the effectiveness of the drug would have to be attributed to its pharmacologic properties.

In these studies, it is important that the patients do not know if they are receiving the drug or the placebo. If the patients knew they were taking a placebo the placebo effect would be greatly reduced or eliminated. Also, it is important that the physicians or nurses administering the drug and the researchers who assess the patient's outcome do not know which patients are taking the drug and which are taking the placebo. If they knew then the knowledge could cause conscious or unconscious bias that might affect their interactions with and evaluations of the patients. When this is done the studies are called double-blind studies. It is not always possible to perform a double-blind study. Sometimes it is difficult for the patient to be blind to the treatment he receives but a blind rater can measure the outcome. This would be called a single blind study.

Sometimes truly controlled experiments are not possible e.g. in psychotherapy research patients who are placed in no-treatment control groups may receive help from family, friends etc. As this research would not constitute a true no-treatment control, the study would be called a partially controlled clinical trial. Controlled experiments pose ethical problems if there is a good reason to believe that the treatment under investigation is either beneficial or harmful one. E.g. when it is strongly suspected that one group would be deprived of a beneficial treatment, or subject to harmful treatment. A number of drug trials have been cut short when they appeared to be so effective that it was thought unethical to continue depriving the control group patients of the group (e.g. zidovudine for AIDS, tamoxifen for breast cancer prevention).

Randomization - means that patients are randomly assigned to different groups (experimental or control groups) to equalize the effects of extraneous variables. E.g. if male patients are assigned to drug group and female to the control group then any difference in the outcome between the two groups could be attributed to the difference between the sexes. In this situation patient gender is called the confounding variable because they contribute differently and inextricably to the two groups. To avoid confounding effects patients are normally assigned randomly to the two groups so that independent variables e.g. gender are not systematically related. Allocating patients randomly to the different groups guards against bias. True randomization means that the group should be similar with respect to all variables (e.g. gender, disease severity, age, and occupation) that may differentially affect response to the experimental intervention.

Matching - Randomization cannot guarantee that the two groups are similar in all important ways. Matching is another way to do this: each patient in the experimental group is paired with a patient in the control group who matches him or her closely in all relevant characteristics. Thus any resulting differences between the two groups cannot be attributed to differences in age, race or gender.

Stratified Randomization - This is a combination of randomization and matching techniques. The population under study is first divided or stratified into subgroups that are internally homogenous with respect to the important factors e.g. gender, age, race, etc. Equal numbers of patients from each subgroup are then randomly allocated to the experimental and control groups. The two groups are therefore similar, but their exact membership is still a result of randomization.

Experimental Designs - Comparisons made between patients in one group and patients in the other group as shown above are called between-subjects design. Alternatively, comparisons can be made within each subject called **within-subjects** design and the control group is called **same-subject** control group. A common type of **within-subjects** approach is **crossover** design. Here half the patients receive the placebo for a period followed by the experimental treatment; and the other half receive the treatment first followed by the placebo. If there is a danger of a "carryover" effect, then there can be a **washout** period in between the drug and the placebo phases during which no treatment is given. The effect of the drug is determined by comparing the effect of the drug and the placebo *within each patient.* This is also called a **repeated measures** design because the measurements are repeated within each patient at different times, and the results are analyzed by comparing the measurements that have been repeated on each patient.

NON EXPERIMENTAL STUDIES - or observational studies fall into two general classes: descriptive studies and analytic studies.

Descriptive studies - These aim to describe the occurrence and distribution of disease or other phenomena. They do not try to offer explanations or test a hypothesis. They give a general description of the frequency of the disease and other points of interest about the disease by using descriptive statistics but not inferential statistics. Descriptive studies are often the first method to study a disease - hence they are also called **exploratory** studies - and they may serve to generate hypotheses for analytic studies to test. Well-known examples of descriptive studies include early studies of AIDS showing that male homosexuals, intravenous drug users and persons with hemophilia are at risk. The true nature of disease was unknown at that time, but the findings of descriptive studies generated useful hypotheses.

Analytic studies - These aim to test the hypotheses or to provide explanations about a disease or other phenomena - the hypotheses or explanations are often drawn from earlier descriptive studies. Analytic studies use inferential statistics to test hypotheses. Descriptive and analytic studies are often not entirely distinguishable. E.g. large-scale descriptive studies may find such clear data that it may provide an answer to questions or give clear support to a particular hypothesis.

Nonexperimental designs - Descriptive or analytic studies use one of four principal research designs: cohort studies, case-control studies, case-series studies, or prevalence surveys.

Cohort studies - A cohort (group of people) that does not have the disease of interest is selected and then observed for an extended period. Some members will have been exposed to the suspected factor and others will eventually be exposed; by following them all the relationship between the risk factor and the eventual outcomes can be seen. This study allows the incidence and the natural history of a disease to be studied. Cohort studies may be called **follow-up** or **longitudinal** studies because they observe any changes over a prolonged period of time. They may also be called **prospective studies** because people are followed from a particular point in time. Also called **Incidence studies** because they look for the incidence of new cases of the disease over time.

Cohort studies have some significant advantages:

1. When a true experiment cannot be done for ethical or practical reasons cohort studies are the best form of investigation; their findings are extremely valuable.
2. They are the only method that can establish the absolute risk of contracting a disease and help to answer a clinically relevant question: if someone is exposed to a certain suspected risk factor, is that person more likely to contract the disease? Cohort studies may also reveal the existence of protective factors such as diet and exercise.
3. Cohort studies are **prospective** and so the assessment of risk factors in these studies is unbiased by the outcome. If the study was retrospective, then people's recollection of their diet and smoking habits may be biased because they already have the disease (called **recall-bias).** In addition, the choronologic relationship between the risk factors and the disease is clear.
4. For the individuals in a cohort who ultimately contract the disease of interest, data concerning their exposure to suspected risk factors have already been collected. In a retrospective study this may not be possible. Other types of studies are often

unable to include people who die of the disease in question- often the most important people to study.

5. Information about suspected risk factors collected in cohort studies can be used to examine the relationship between the risk factors and many diseases; therefore, a study designed as an analytic investigation of one disease may simultaneously serve as valuable descriptive study of several other diseases.

Cohort diseases may have some important disadvantages:

1. They are time-consuming, laborious and expensive to conduct; members must be followed for a long time before a sufficient number of them get the disease of interest. It is expensive and it is difficult to keep track of a large number of people for several years, and it takes many years before results are produced, especially in cases of disease that take a long time to appear after the exposure to a risk factor.
2. This may be impractical for rare diseases but if there is an high rate of occupational disease in an occupational cohort then it can be studied by this method e.g. bladder cancer among dye workers.

Case-control studies - Whereas cohort studies examine people who are initially free of the disease of interest; case-control studies compare people who do have the disease (cases) with otherwise similar people who do not have the disease (controls). Case-control studies start with the outcome, or dependant variable (the presence or the absence of disease). They then look back into the past to for possible independent variables that may have caused the disease, to see if a possible risk factor was present more frequently in cases than in controls. Hence these studies are also called **retrospective** studies.

Case-control studies have some significant advantages:

1. They can be performed fairly quickly and cheaply as compared to cohort studies, even for rare diseases or diseases that take a long time to appear. Thus case-control studies are the most important way of investigating rare diseases and are typically used in the early exploration of a disease.
2. They require comparatively fewer subjects.
3. They allow multiple potential causes of a disease to be investigated.

Case-control studies have a number of disadvantages and are particularly subject to bias:

1. People's recall of their past behaviors or risk factor exposure may be biased by the fact that they now have the disease.
2. The only cases that can be investigated are people who have been identified and diagnosed; undiagnosed or asymptomatic cases are missed by this kind of study. People who have already died of the disease cannot be questioned about their past behaviors and exposure to risk factors.
3. Selecting a comparable control group is a difficult task that relies entirely on the researcher's judgment.
4. Case-control studies cannot determine the rates or the risk of the disease in exposed and nonexposed people.
5. They cannot prove a cause-effect relationship.

Case series studies - A case series simply describes the presentation of a disease in a number of patients. It does not follow the patients for a period, and it uses no control or comparison group. Therefore, it cannot establish a cause-effect relationship, and its validity is entirely a matter for the reader to decide. A report that 8 of 10 patients with a certain disease have a history of exposure to a particular risk may be judged to be

extremely useful or almost worthless. Despite these shortcomings case series studies are commonly used to present new information about patients with rare diseases and they may stimulate new hypotheses. These studies can be done by any clinician, who carefully observes and records patient information. A case report is special form of case series in which only one patient is described-this too may be very valuable or virtually worthless.

Prevalence surveys- A prevalence survey or a community survey is a survey of the whole population. It assesses the proportion of people with a certain disease (prevalence of the disease) and examines the relationship between the disease and other characteristics of the population. As prevalence surveys are based on a single examination of the population at a particular point in time and do not follow the population over time, they are also called cross-sectional studies, in distinction to longitudinal (cohort) studies.

Prevalence surveys suffer from a number of disadvantages:

1. Because they look at existing cases of a disease, and not at the occurrence of new cases, they are likely to over-represent chronic diseases and under-represent acute diseases.
2. They may be unusable for acute diseases, which few people suffered from at the moment they were surveyed.
3. People with some types of disease may leave the community, or may be institutionalized, causing them to be excluded from the survey.

Findings of prevalence surveys must be interpreted cautiously; the mere fact that two variables are associated does not mean they are causally related. Although they are expensive and laborious prevalence surveys are common because they can produce valuable data about a wide range of diseases, behaviors and characteristics. These data can be used to generate hypotheses for more analytic studies to examine. See chart on page 13 of Kaplan

6) Statistics in Epidemiology

Epidemiology is the study of rate of disease in a population.

1. **Rate** = No. of people with a particular condition/No of people at risk. The most important rates are **incidence, prevalence** (both measures of **morbidity), mortality** (special form of Incidence) and **case fatality.**

Morbidity = sick rate

Mortality = death rate

Incidence and prevalence are measures of rates of morbidity or illness. Mortality is a special form of Incidence.

Incidence - of a disease is the number of new cases occurring in a particular time period such as a year.

Incidence Rate = $\frac{\text{No of new cases of the disease}}{\text{Total no of people at risk}}$ per unit of time

Incidence rates are found by cohort studies (also called incidence studies) and often stated per 100,000 of the population at risk or as a percentage. E.g. If incidence of shingles in a community is 2,000 per 100,000 per annum then we can say that in one year 2% of the population experiences an episode of shingles.

* Focus on acute conditions** Incidence is new incidents. *** People previously positive for the disease are not considered at risk.

Prevalence - of a disease is the number of people affected by it at a particular moment of time.

Prevalence Rate = No of people with the disease at a particular time
Total number of people at risk.

Prevalence Rate is also stated per 100,000 people or as a percentage. They are found by prevalence surveys. E.g. At a given time 170 out of every 100,000 people (0.17%) may be suffering from shingles

* Focus on relatively stable chronic conditions e.g. hypertension or diabetes.

Relation between Incidence(I) and prevalence(P) = P=I X D where D is Average Duration of the disease. If I goes up Then P goes up if D remains same. Increased P may reflect increased Duration. In e.g. of shingles 1=2% and P= .17%

New treatment is initiated	I =N	P=Down
New treatment widespread use	I=Down	P=Down
People dying from this condition increase	I=N	P=Down
More money added for research	I=N	P=N
Behavioral risk factors reduced in population	I=Down	P=Down
Contacts bet. infected &not infected reduced -infectious disease	I=Down	P=Down
Contacts bet. infected &not infected reduced -not infectious disease	I=N	P=N
Recovery more rapid than a year ago	I=N	P=Down
Long term survival rates are increasing for the disease	I =N	P=Up

* New Cases + Pre-Existing Cases = Prevalence (no of cases at a given time). * Prevalent cases either die or recover to reduce prevalence.

Mortality - is the number of deaths.

Mortality rate = total number of deaths per unit of time
total number of people at risk

Mortality rate may be expressed as a percentage or no of deaths per 1,000 or 100,000 people.

The relationship between incidence prevalence and mortality in any disease can be visualized with the aid of epidemiologist's bathtub. See Page 70

Crude Mortality Rate - Death/population.

Cause-specific mortality rate - Death from cause/ population

Cause fatality rate - Death from cause/ number of persons with the disease/cause.

Proportionate Mortality Rate - Death from cause/All deaths

Case fatality Rate (CFR) = total no. of people dying in an episode of the disease * 100
total number of episodes of the disease

This is expressed as a percentage. It is used to follow the effectiveness of treatment over time or in different places e.g. CFR of acute myocardial infarction in a given hospital.

Attack Rates = number of people contracting a disease * 100
total number of people at risk

Attack rate is expressed as a percentage e.g. 1000 people eat at a barbecue where contaminated food is served and 300 get sick then attack rate is 300/1000 = 30%. Attack rates help in deducing the source of an epidemic. Different people at the barbecue ate different combination of food so attack rate for each food is calculated.

Food	Attack Rate in %
Chicken only	25
ROibs only	12.5
Cole Slaw only	30

Chicken & Ribs	9
Chicken, ribs & Cole slaw	40
Total for all food	30

The lowest is 9 and highest is 40. So Cole slaw is the source.

Crude Rates- Actual measured rate for the whole population. Crude rate can be expressed as a weighted sum of different age-specific rates. See page 5 Kaplan..

Specific Rate- Actual measured rate for the subgroup of population e.g. age or sex specific.

Adjustment or standardization of rates- e.g. A researcher may want to compare the prevalence of AIDS in two cities of equal size but one city has a larger prevalence of elderly people. Due to the effect of different age structures the prevalence rate alone tells nothing about any real underlying difference in the prevalence of AIDS in the two cities. This biasing influence of a confounding variable such as age (or any other variable) can be removed by the technique of adjustment or standardization of rates. Adjustment can also be made for two or more factors simultaneously. Mortality rates are commonly standardized producing a statistic called Standardized Mortality Ratio (SMR).

Measurement of Risk - The knowledge that something is a risk factor for a disease can be used to help in:

- Preventing disease
- Preventing future incidence and prevalence
- Diagnosing it
- Establishing the cause of a disease of unknown etiology.

Absolute risk - The incidence of a risk is the absolute risk of contracting it. If incidence is 10 per 1000 people per annum or 1% then absolute risk is 1%. It is useful to do a number of different comparisons of risk including relative risk, attributable risk and the odds ratio. For cohort studies Relative risk and or attributable risk is used.

Relative risk (RR) (How much more likely?) - states by how many times exposure to the risk factor increases the risk of contracting the disease. It is the ratio of the incidence of the disease among exposed persons to the incidence of the disease among unexposed persons. RR = incidence of the disease among persons exposed to the risk factor/ incidence of the disease among persons not exposed to the risk factor.

Risk	lung cancer	no lung cancer	Total
Heavy Smokers (Exposed)	283	725	1008
Nonsmokers (nonsmokers)	64	1010	1074
Total	347	1735	2082

The incidence of cancer over total time period of the study among people among exposed to the risk factor is 283/1008 or .28 or 28%, while the incidence among those not exposed is 64/1074 or .06 or 6%. So RR is ,28/.06= 4.67 showing that people who smoked cigarettes heavily were 4.67 times more likely to contract lung cancer than non smokers. As RR is a ratio of risks it is also called risk ratio **or morbidity ratio** and when cases of death rather than diseases are involved it may also be called **mortality ratio.** RR reduction due to the use of a drug is 1-RR. Absolute R reduction due to the use of a drug is Mortality rate of those not using the drug - mortality rate of those using the drug. Absolute Risk reduction allows calculation of another important statistic i.e. numbers needed to treat (NNT) which allows comparisons of effectiveness of different treatments e.g. ARR = .07% then NNT = 100/0.7 = 143. So 143 persons should be treated to save one person. NNT =

1/ARR. NNT allows calculation of cost of saving one life with the treatment e.g. 143 persons treated for 58 months save one life and cost per month is $ 100 then cost of saving one life is 143*58*100 = $ 829,400.

Attributable Risk (AR) (How many more cases in one group?) is the additional incidence of the disease attributable to the risk factor in question. AR = incidence of the disease among persons exposed to the risk factor- incidence of the disease among persons not exposed to the risk factor. In the above example of smokers AR is .28-.06 = .22 i.e. 22% get disease due to smoking and 6% get disease due to other factors.

Odds Ratio (OR) - AR and RR both require cohort studies which are expensive and time consuming and therefore impractical. An alternative is to use a case control study, which compares people with the disease (cases) to similar people without the disease (controls). OR looks at the increased odds of getting a disease with exposure to a risk factor versus a non risk factor. OR = Odds that a case was exposed to a risk factor/odds that a control was exposed to a risk factor, (odds that a person with lung cancer was a smoker versus odds that a person without lung cancer was a smoker.) Case control studies cannot give prevalence or Incidence of a disease because the proportion of people in the study who have the disease is determined by the researcher's choice, so they cannot determine the risk of contracting a disease. Using the above data for lung cancer we assume that these data was generated by a case control study.

OR = no. of cases exposed to risk factor (A)* no of controls not exposed (D)
A no. of controls exposed to risk factor (B) * no of cases not exposed (C)

No. of cases exposed to risk factor (A) = 283
No. of controls exposed to risk factor (B) = 725
No of cases not exposed to the risk factor (C) = 64
No. of controls not exposed to the risk factor (D) =1010

OR = 283*1010/725*64 = 6.16. In other words among the people studied a person with lung cancer was 6.16 times more likely to have been exposed to the risk factor. If OR =1 then it suggests that the risk factor is not related to the disease. OR of less than 1 indicates that a person with the disease is less likely to have been exposed to the risk factor than is a person without the disease suggesting that the risk factor may actually be a protective factor against the disease.

7) Statistics in medical decision making - Medical decision-making often involves using various kinds of tests. The qualities of a test can be evaluated in several ways. To assess the quality of a diagonistic test, it is necessary as a minimum to know its

- Validity and reliability,
- Sensitivity and specificity, and
- Positive and negative predictive values.

When using quantitative test results- such as measurement of fasting glucose levels and hematocrit levels- the physician will need to know the accuracy and precision of the measurement and the normal reference for the variable in question.

VALIDITY- of a test is the degree to which a test measures what it claims to test i.e. how closely its results correspond to the real state of affairs. Validity of a test is therefore its ability to show which individuals have the disease in question and which do not. To be truly valid, test should be highly sensitive, specific and unbiased. Validity is synonymous

with **accuracy.** When assessing the validity of a research study as whole two kinds of validity are involved:

- **Internal validity:** are the results of the study valid for the population of patients who were actually studied?
- **External validity:** are the results of the study valid for other patients?

RELIABILITY- Reliability is synonymous with **repeatability and reproductibility.** It is the ability of the test to measure something consistently, either across testing situations (test-retest reliability), within a test (split-half reliability), or across judges (inter-rater reliability). Reliability corresponds to precision. A reliable test is therefore consistent, stable and dependable one.

The repeated reliability of a test influences the extent to which a single measurement may be taken as a definitive guide for making a diagnosis. In the case of a highly reliable test, one measurement alone maybe sufficient to allow a physician to diagnose with confidence; however if the test in unreliable in any way, this may not be possible e.g. BP measurement is required to be repeated many times and the mean is used to obtain a reliable measurement and make a confident diagnosis. In practice standard laboratory tests have been carefully validated, and quality control procedures in the laboratory ensure reliability.

A test measurement may be reliable without necessarily being valid e.g. it would be possible to measure the circumference of the person's skull with reliability and precision, but this would not constitute a valid assessment of the person's intelligence.

Bias may also cause a reliable and precise measurement to be invalid, e.g. a balance may weigh very precisely, and with very little variation repeated weighing's of the same object. However if it has not been zeroed properly, all its measurements may be 3 mg too high, causing all its results to be biased and hence inaccurate and invalid.

Conversely a measurement may be valid yet unreliable e.g. BP of a person may vary considerably, but if all of these measurements cluster around one figure, the findings as a whole may accurately represent the true state of affairs (that a patient is hypersensitive).

REFERENCE VALUES - Even if the measurements are both highly valid and reliable, it does not allow the physician to make a diagnosis. To make a diagnosis the physician must know the measurement's range of values among healthy, normal people. This range is called the reference range or normal range and the physician will use this to interpret the values obtained from the patient. (The range between the reference values is sometimes called the reference interval). The reference range of a biomedical valuable is defined as middle 95% of a normal distribution - or the population mean plus or minus two Standard deviations. The assumptions made are that:

- The 95% of the population that lie within this range are 'normal' and the 5% that lie beyond it are 'abnormal'.
- The 'normal range' for a particular biomedical variable (e.g. serum cholesterol) can be obtained by measuring it in a large representative population of normal, healthy individuals, thus obtaining a normal distribution; the central 95% would then be of 'normal range'.

In practice the normal range and the reference values are a compromise between the statistically derived values and clinical judgment, and may be altered from time to time as the lab gains experience with a given test. The values must be interpreted in the light of other factors that may influence the data obtained about a given patient e.g. age, weight, gender, diet, and the time of the day when the specimen was drawn.

SENSITIVITY AND SPECIFICITY - are both measures of a test's validity- its ability to correctly detect people with or without the disease in question. There are four logical possibilities in diagnostic testing:

TP: A positive result is obtained in the case of a person who has the disease. This is a "true-positive" finding.

FP: A positive result is obtained in the case of a person who does not have the disease. This is a "false-positive" finding which is a type II error.

FN: A negative result is obtained in the case of a person who does have the disease. This is a "false-negative" result and constitutes a Type I error.

TN: A negative result is obtained in the case of a person who does not have the disease. This is a "true-negative" result.

Sensitivity - of the test is its ability to detect people who have the disease (TP). It is the percentage of people with the disease correctly detected or classified:

Sensitivity = TP/(TP+FN) * 100

Thus a test that identifies every person with the given disease has a sensitivity of 100%. An insensitive test leads to missed diagnoses (FN) and a sensitive test produces few FN results. A sensitive test is required in situations in which the consequences of FN are serious e.g. a condition that is treatable or transmissible. Thus high sensitivity is required to screen donated blood for HIV, for pap smears for cervical cancer etc.

Very sensitive tests are therefore used for screening or ruling disease; if the result of a highly sensitive test is negative, it allows the disease to be ruled out with confidence. **(Snout** -Sensitive test wit a Negative result **OUTs** the disease)

Specificity - of a test is its ability to detect people who do not have the disease. It is the percentage of disease free people who are correctly classified or detected:

Specificity = TN/(FP+TN) * 100.

Thus a test that is always negative for healthy individual identifying every nondiseased person has a specificity of 100%. A test low in specificity leads to FP diagnoses and a highly specific test produces few FP results.

High specificity is required in situations in which the consequences of FP diagnoses are serious e.g. a diagnosis that might lead to the initiation of dangerous, painful or expensive treatments (e.g. cancer chemotherapy), or diagnosis may be alarming (HIV, cancer), in which diagnosis may cause a person to make irreversible decisions (Alzheimer's disease); or in which the diagnosis may result in a person being stigmatized (schizophrenia, HIV, tuberculosis).

Very specific tests are therefore appropriate for confirming or ruling in the existence of a disease. If the result of a highly specific test is positive, the disease is almost certainly present. **(Spin =** SPecific test with a Positive result rules **IN** the disease).

In clinical practice sensitivity and specificity are inversely related: an increase in one causes a decrease in the other. This is because diseased and nondiseased patients do not form two discrete groups but overlap each other. The tester therefore has to select a "cutoff point" to make a diagnostic decision. E.g. when a test of fasting glucose levels of people with and without diabetes is done the choice of cutoff point will determine the test's sensitivity and specificity. If the cutoff point was lowered then the test would be 100% sensitive (identifying every diabetic) but it would have low specificity (high FP- less Type I errors but more Type II errors). If the cutoff point is increased then the test's sensitivity would gradually decrease and its specificity would gradually increase until the cutoff point

is reached where it would be 100% specific. At this point it would correctly identify all non-diabetics but it would be highly insensitive giving many FN (Highly specific -Type I errors).
A test can be 100% sensitive and 100% specific if there is no overlap between the population that is normal and the population that has the disease. This kind of situation does not normally occur and when it does the disease may be so obvious that no diagnostic testing is required.
Although there are tests of relatively high sensitivity and specificity for some diseases, it is often best to use a combination of tests when diagnosing a particular disease. A highly sensitive (and usually relatively cheap) test should be used first almost guaranteeing the detection of the disease (albeit at the expense of including a number of false positive results). This should be followed by a more specific (and usually more expensive) test to eliminate the false positive results. This is the usual sequence for testing HIV, Hepatitis B, and many other serious but common disease.
Predictive Values - When a physician or patient is faced with a test result they want to know how likely it is that the disease really is present or absent. This requires the knowledge of predictive values of the test.
Positive predictive values (PPV) - of a test is the proportion of positive results that are true positives i.e. the likelihood that the person with a positive test result actually has the disease. PPV = TP/(TP+FP)
Negative predictive values (NPV) - of a test is the proportion of negative results that are true negatives i.e. the likelihood that the person with a negative test result truly does not have the disease. NPV = TN/(TN+FN).
Sensitivity and specificity of a test depend on the characteristic of the test itself but the predictive values vary according to the prevalence of the disease. Thus predictive values cannot be determined without prior knowledge of the prevalence of the disease-they are not qualities of the test per se, but are a function of the test's characteristics and of the setting in which it is being used.
The higher the prevalence of a disease in a population, the higher the PPV and the lower the NPV of a test for it. If a disease is rare