RESEARCH ARTICLE

SOME APPLICATIONS OF BIOSTATISTICS TO MEDICAL RESEARCH

Dar Nazir Ahmad[1], Khan Nazir Ahmad[2], M Mohib-ul-Haq[3], Nayak Bilal Gul[4] and Lone Mohammad Maqbool[5].

1. Statistical Assistant PBCR Radiation Oncology SKIMS.
2. Professor Radiation Oncology SKIMS.
3. Head Department of Radiological Physics & Bio-Engineering SKIMS.
4. Senior Resident Radiation Oncology SKIMS.
5. Professor & Head Radiation Oncology SKIMS.

………………………………………………………………………………………………………....

Manuscript Info	*Abstract*
……………………..	………………………………………………………………

Biostatistics is well recognized as an essential tool in medical research, clinical research and health research. Biostatistics is the branch of statistics applied to biological areas or medical sciences or medical research. Biostatistics not only covers the health, medicine, and nutrition but also covers the agriculture, genetics, biology, biochemistry, demography, epidemiology, anthropology and many others. Biostatistics helps to evaluate many life and death issues in medicine. The aim of this article is to educate the medical residents about the applications of biostatistics so that the proper and accurate conclusions may be drawn from the study.

………………………………………………………………………………………………………....

Introduction:-

Biostatistics is the science that helps in managing medical uncertainties. It is mainly consists of various steps like generation of hypothesis, collection of data and application of statistical analysis. An ample knowledge of bio-statistics is important for research scholars, medical students, and nursing students so that they can design epidemiological study accurately and draw meaningful conclusions. Biostatistics is the branch of applied statistics and it must be taught with the focus being on its various application in biomedical research[1].The biostatistical literacy of medical students is a problem in all over the world .Medical research is being encouraged in India and biostatistics is a vital tool. An important aspect of a doctor's job is to understand and condense the provided data, analyze findings and engage in research activities. An important tool for all these tasks is biostatistics! Biostatistics is vital in understanding certain epidemiological principles such as incidence, prevalence, odds ratio, risk ratio, and number needed to treat. Biostatistics is also important for postgraduate residents for their dissertation and various research projects. A clear statement was made during the Edinburg Declaration of World Medical Association in 1988, the purpose of the medical education is to produce professional who are able to understand the needs of their communities and respond accordingly. To achieve these medical students must gain professional skills and theoretical knowledge.

Biostatistics gives great importance in both developed and developing countries but the scope of biostatistics varies from school to school in terms of time allocation, scope and depth of topics covered. For instance in the majority of medical schools in the USA biostatistics related courses such as epidemiology , preventive medicine , and

---

Corresponding Author:-Dar Nazir Ahmad.
Address**:-**Statistical Assistant PBCR Radiation Oncology SKIMS.

demography taught in during the first or second year. In Pakistan university of Ziauddin , epidemiology biostatistics and survey methodology courses are taught in the first two years of medical school . In South Africa biostatistics and research methodology are taught in the year 1 and 2 with further reinforcement practiced in year 3 and 4 in some universities and in 44 medical schools in Turkey, biostatistics is usually taught in the ist year.

Statistical analysis of data is not difficult and without proper statistical analysis one can't draw the meaningful conclusion of the study. So it is mandatory for every medical researcher to know the applicability of biostatistics in medical field so that the proper and accurate conclusions may be drawn from the selected samples or population .The following is an example a certain disease when untreated is known to have a mortality of 50% and a new drug is developed for this disease and administered to 6 patients and out of 6 patients 1 dies and 5 recovers , it means that the efficacy of this new drug is much better than the previous treatment.

Some of the application of biostatistics in medical research discussed in this research article because anyone who is engaged in medical research will be able to analyze his own data and draw meaningful conclusions. This research article will cover the following topics:
1. Functions of statistics: method of statistics, sampling, normal distribution, mean, mode, median, variance, and standard deviation.
2. T-test
3. Chi-square
4. Linear Regression and simple correlation

Functions of statistics:
1. To reduce the data
2. To provide methods of test of significance
3. To provide a sound basis for experimental design.

Methods of statistics:
The methods of statistics are broadly categorized into two division's descriptive and inferential statistics. In descriptive statistics we describe the data and give the graphical representation of the data so that layman understands in a quick view and it becomes easy for understanding. Another type of statistics is inferential statistics means we draw inference about the sample selected from population. When we analyze our data it is necessary to use both descriptive and inferential statistics. Descriptive statistics are used to describe our data, summarize our data and organize our data so that we can draw meaningful results. The important descriptive statistical measures that are used to describe, summarize and find out the central value and location of the data are Measures of location and Measures of variability.

Mean, Median and Mode
Mean is the first and simplest measure of location. It is the most frequently used measure of location. It can be defined as sum of observation divided by the number of observation. The most important drawback is, mean of a particular group is affected by very small and very large number. Median is defined as middle of observation. It divides the whole data into two equal parts one part comprising all the values less than median and second part comprising all the values greater than median. Median is not affected by extreme values. Median is the only average used for dealing with the qualitative of data. In median we have two cases odd and even case. In odd case we arrange the distribution into ascending (descending) order and distribute the series into two parts and middle one is median. In even case we arrange the distribution into ascending (descending) order and calculate average between the two middle values and the middle value is median. Mode is the most frequently occurring value in a set of data. Mode is particularly useful in the study of popular sizes. Mode is the average to be used to find the ideal size in a series.

Table 1:-Summary of Central tendency

Measures	Descriptive
Mean	Balance point
Median	Middle value when ordered
Mode	Most frequent

When looking at a set of values or a frequency distribution, it can easily be seen if the observations are widely dispersed from the measure of central value or are scattered fairly closely around it, but it may often be desirable to describe the dispersion in a single summary figure. Some of the important measures used in measures of dispersion are discussed below:

Variance and Standard Deviation
A better measure of dispersion is the sums of squares of the deviations from the mean divided by the number of variates. This is the variance. It is not directly related to the normal probability and as a measure of variability it is difficult to interpret, because the original unit of measurement is squared. This is avoided by taking the square root of variance. This is the standard deviation or standard error and is the most widely used measure of dispersion.

The standard deviation is related to normal distribution in the following way. The mean plus or minus one standard deviation includes 68% of the population. The mean $\pm 2SD$ includes 95% of the population and mean $\pm 3SD$ includes 99% of the population.

Sampling:
Sampling is a tool which enables us to draw conclusion about the characteristics of population .sampling is done when a large population is there for study .An element and group of elements of the population used for drawing a sample is sampling unit. Sampling frame is the list of sampling units with identification addresses in a population which serve as a base for selecting a sample .In study a sample size increases, error chances decreases, but cost increases. It is very important to find the optimal sample size. Sampling will be further categorized into two types' probability and non probability sampling. In probability sampling every unit of population has an equal chance of being included in the sample and non-probability means not having the equal chance of being included in the sample. Probability sampling includes simple random sampling, systematic random sampling, stratified random sampling, multistage random sampling, and cluster sampling. Non-probability sampling includes judgment, quota sampling, convenience, snowball and extensive sampling.

Normal distribution
The normal distribution is the most important and most widely used distribution in statistics. It is sometimes called the "bell curve, " although the tonal qualities of such a bell would be less than pleasing. It is also called the "Gaussian curve" after the mathematician Karl Friedrich Gauss. As you will see in the section on the history of the normal distribution, although Gauss played an important role in its history, de Moivre first discovered the normal distribution. Some of application of the normal distribution are mentioned below:
1.   Normal distributions are symmetric around their mean.
2.   The mean, median, and mode of a normal distribution are equal.
3.   The area under the normal curve is equal to 1.0.
4.   Normal distributions are denser in the center and less dense in the tails.
5.   Normal distributions are defined by two parameters, the mean (μ) and thestandard deviation (σ).
6.   68% of the area of a normal distribution is within one standard deviation of themean.
7.   Approximately 95% of the area of a normal distribution is within two standarddeviations of the mean.

Student's t-test
Student's t-test is a parametric test, it is applicable to find a significant difference between two means. It is used in three situations:-
1.   To test if the sample mean differs significantly from the population mean (one sample test)
2.   To test if the population means estimated by the two independent samples differ significantly (unpaired t-test).
3.   The t-test is used when we have a small samples, i.e. sample size <30 and the sample means are not distributed normally about the population mean .In such cases t-test is used to test the significance. Unpaired t-test is used when the observation are independent .paired t-test is used when the observation are dependent.

Chi-Square
Chi-square test involves the calculation of a quantity called chi-square ($\chi2$). It was developed by Karl Pearson. It is used to compare the observed and expected frequencies, and to test the statistical significance of association between two discrete (qualitative) variables. Categorical data may be displayed in contingency tables. The $\chi2$ statistic compares the observed count in each table cell to the count which would be expected under the assumption of no association between the row and column classifications. The $\chi2$ statistic may be used to test the hypothesis of no

association between two or more groups, populations, or criteria. Observed counts are compared to expected counts. The $\chi2$ statistic is calculated under the assumption of no association. Large value of $\chi2$ statistic $\Rightarrow$ small probability of occurring by chance alone (p<0.05) $\Rightarrow$ conclude that association exists between disease and exposure. Small value of $\chi2$ statistic $\Rightarrow$ large probability of occurring by chance alone (p>0.05) $\Rightarrow$ conclude that no association exists between disease and exposure.

Conclusion:-
Without knowing the application of biostatistics to medical research a researcher cannot do the research and can't draw the conclusion of the research. So learning basic of biostatistics and knowing its applications in medical research is important for researcher, clinician and an ample knowledge of bio-statistics is important for research scholars, medical students, and nursing students so that they can design epidemiological study accurately and draw meaningful conclusions.

Table 2:- Description of various parametric and non parametric tests

	Parametric	**Non parametric**
Assumed distribution	Normal	Any
Assumed variance	Homogeneous	Any
Typical data	Ratio or interval	Ordinal or nominal
Data set relationships	Independent	Any
Usual central measure	Mean	Median
Benefits	Can draw more conclusions	Simplicity, less affected by outlers
Tests		
Choosing	Parametric tests	Non parametric tests
Correlation test	Pearson	Spearman
Independent measures,2 groups	Independent measures, t test	Mann whitney test
Independent measures,>2 groups	One way independent measures ANOVA	Kruskal wallis test
Repeat measure , 2 conditions	Matched pair t- test	Wilcoxon test
Repeated measures, >2 conditions	One way repeated measures ANOVA	Friedman's test
ANOVA= Analysis of variance		

References:-
1. PerrieA,sabin C (Eds).Describing data.Medical statistics at galance.UK: Blackwell Science Ltd;2000 pp 16-9.
2. KuzmaJW ,Bohnenblust SE (Eds). Summarizing data : Basic statistics for the science.London:Mayfield publishing company;2001 pp 44-54.
3. Manikandan S. Measures of central tendency: median and mode. J Pharmacother 2011;2:214-5
4. Sprent p. statistics in medical research.Swiss Med wkly 2003;133:522-9
5. Bewick V , check L,BallJ.statistics review 10:Further non-parametric methods.crit care 2004;8:196-9
6. Altman DG, Bland JM.parametric v non-parametric methods for data analysis.BMJ 2009;338:a3167
7. Kaur SP .variables in research. Indian J Res Rep Med Sci 2013;4:36-8
8. Ali Z,BhaskarSB.Basic statistical tools in research and data analysis.indian J Anaesth 2016;60:662-9
9. Magnello ME karl person and the origin of modern statistics : An elastician becomes statistician , Rutherford J , Vol.1 2005-2006 avaliable online at :http://Rutherford.org
10. RanaR,Singhal R, Chi-square test and its application in hypothesis testing.JpractcradiovascSci 2015;1:69-71
11. Indrayan A ,satyanarayanL.Simple biostatistics for MBBS,PG entrance and USMLE 4th edition Delhi: Academa publisher 2013
12. PerrieA,sabin C (Eds).Describing data.Medical statistics at galance.UK: Blackwell Science Ltd;2000 pp 42-3
13. Park K. Textbook of preventive and social medicine. 21st ed. BanarasidasBhanot; 2011.
14. Mahajan BK. Methods in biostatistics. 7th ed. New Delhi: Jaypee Brothers Medical Publishers; 2010.