REVIEW ARTICLE

STATISTICAL CONCEPTS IN BIOLOGY AND HEALTH SCIENCES

Huma Zahir, Aisha Javaid*, Rehana Rehman*, Zahir Hussain**

King Abdulaziz University (Formerly Department of Statistics, DHA College for Women, Karachi, Pakistan), * Department of Physiology, University of Karachi, Karachi, Pakistan, **Department of Physiology, Faculty of Medicine, Umm Al-Qura University, Makkah, Kingdom of Saudi Arabia

In view of its applied aspects, Statistics serves as a separate mathematical science. In that respect, biostatistics is the application of statistical concepts and methods in biology, public health and medicine. One major task of medical biostatistics is to understand why a disease occurs in certain area and why that disease does not occur in other areas. In general, the advantages for properly applying statistics for a country are to keep the detailed information of people in a country. However, there must in mind be the other face of the task remembering not to adapt these surveys and limited data with entirety for quick applications that might be less advantageous. Some of the programs are much expensive and time consuming and people may feel not comfortable conveying their personal information just for the sake of applying a so called organized procedure. In such conditions, one must consider the moral values as well. Another quite unfortunate fact is that a statistical data can be misused for personal needs of a presenter. There must be ways to eradicate such customs at the governmental level. Basic and higher courses, certificate courses, diploma programs, degree programs, and other opportunities for students can be well organized and can be utilized in various employment areas in industry, government, life sciences, computer science, medicine, public health, education, teaching, research, and survey research. Statisticians, hence, are very important people for establishing various schemes, programs, institutions and organizations in medical and biological sectors.

Keywords: Biostatistics, medical statistics, basic statistics and mathematical concepts, tasks and significance of biostatistics

J Ayub Med Coll Abbottabad 2014;26(1):95–7

INTRODUCTION

Statistics and statistic are two different terms meaning scientific discipline and a quantity (e.g., mean, median or mode) respectively. The discipline of Statistics is considered as a branch of mathematics for collection and interpretation of data.[1] The other view for statistics is that it is a mathematical science for collection, analysis, presentation and interpretation of the data.[2] In view of its applied aspects, statistics is a separate mathematical science instead of being a branch of mathematics.[3,4]

Biostatistics and medical statistics or biomedical statistics or medical biostatistics is other terms used for important applied branches of Statistics. Biostatistics is the application of statistical concepts and methods in biology, public health, and medicine. In other words, biostatistics is applied to living world and it includes clinical trials, demography, and epidemiology.

Biostatistics is also referred to as biometrics or simply biometry as both have a similarity of involving wide range of biological topics. Whereas, applications of statistics in medicine and health sciences comprise medical or clinical statistics. Both of these, however, have similarity with the statistics and statistical sciences[5] in designing the experiments, collecting and categorizing data, and organizing, analyzing and comparing the results that later serve for the interpretation of the clinical, behavioral, biological and biomedical processes and concepts. This article is based on review of statistical and mathematical literature and our main sources have been libraries where mainly books were searched.

BASIC STATISTICAL CONCEPTS

Statistics has two main basic types: descriptive statistics (set of methods to describe the collected data) and inferential statistics (set of methods to generalize, predict and decide by using information from a sample and inferring something about population). Both branches of statistics collectively are sometimes referred to as applied statistics.[6]

Collecting, organizing, and summarizing the data (discrete or continuous) comprise descriptive statistics, and making valid conclusions based on sample data is inferential statistics. Levels of the data measurements are nominal, ordinal, interval level, and ratio level. Complete collection of a data or universal set of all objects of interest for analyzing properties is termed as population, and sub-collection selected as a part of population of interest is a sample. Hence, a sample is any subset of a population. A parameter is considered as the measurement that describes certain characteristic(s) of a population, whereas a statistic is a measurement that describes certain feature of a sample. Sampling is the process of collecting certain information from a sample that is easily manageable and inferences about the population of that sample can be obtained conveniently.

Data collection usually means each member or data points in a set of data. Data organization refers to frequency distribution of the data. Relative frequency is used for comparative purposes. If there are many different values in the data points, various numbers of groups are formed that constitute a grouped data.

The data distribution wherein most of the values fall in the middle (nearly symmetrical) shows a bell-shaped curve called as normal distribution. In normal distribution, most of the values occur in the middle and less number of values farther away from the centre, and the distribution is symmetric, i.e., one side is the mirror image of the other side. In such case, the mean, median and mode (calculated statistically) all occur in the centre. For such normal distribution, 2/3 of the data values constitute one standard±mean, whereas 95.44% and 99.74% data respectively represent two and three standard deviations±the mean.

Hence, statistics basically concerns with the estimation of variability.[7–9] Random variables, probability, normal, binomial, and Poisson distributions are the main mathematical concepts applied to biostatistics. In short, the data in statistics concerns with the variables which could be: qualitative/quantitative, discrete/continuous, dependent/independent. Distributions are presented as histogram, bar graph, scatter plot, box plot etc. Measures of central tendency are: mean, median, and mode. Dispersion in data is measured as: range, inter-quartile range, variance, and standard deviation; and shape is measured as: symmetric or normal distribution, positive and negative skewness. Pertaining to the peakedness of curve, distributions have mesokurtosis, platykurtosis and leptokurtosis. Correlation, regression of various types, Chi-square test, *t*-test, and analysis of variance (ANOVA) are some simple statistical tests applied for inferences.

BASIC MATHEMATICAL CONCEPTS

Set theory (founded by Georg Cantor in 1874)[10] on a characteristic property of all real algebraic numbers[10,11] especially for probability studies[12,13] are an important set of concepts. Universal set (set of all members or elements for specific case, null set or empty set (set having no elements or members), subset and set operations (complement of a set A, denoted by A', intersection and union of a set and mutually exclusive sets, Venn diagrams, cardinality of set A, and multiplication principle) are the basic concepts in set theory.

Permutations, combinations, addition and compliment rule, classical and conditional probability, and independent events are the other concepts used in biostatistics and statistics in general. Mathematically, Statistics is the study of statistics from a mathematical background and applies the probability and other branches of mathematics.[14–21]

BIOSTATISTICS AND MEDICAL STATISTICS

Medical biostatisticians perform the studies comprising surveys, comprehensive reports, laboratory data, and hospital records. These records provide the information about the birth/birth rate, rate of incidence and prevalence of a disease, death/death rate for the comparison of population data for interpreting about the diagnostic, causal, therapeutic and other management purposes.

Developing descriptive statistics in public and population health is recommended essential by using bio-statistical methods. Statistics in biology and health sciences has a variety of applications. Some of the most important statistical issues relate to complex diseases[22], bioinformatics[23], microarray gene expression[24,25] and networking in system biology[26].

The field of biostatistics shares several types of methods with other fields including statistics, computer studies, econometrics, demography, operations research etc. It develops new statistical methodologies where required. Some of the main applications of biostatistics and biomedical statistics relate to: diagnostic, etiological and management of medical disorders; clinical trials; epidemiology; environmental health; human medical genetics; proteomics or microarray analysis in genomics; gene network studies in systems biology; and ecology.

TASKS OF BIOSTATISTICS WITH ADVANTAGES AND DISADVANTAGES

Biostatisticians get especially involved in the ways data should be accumulated or gathered, and the tracking, analytical approaches and application of data before, during and after the biological experimentation wherein a variety of variables[7-9] are interacted in the set of measurements (a population) or few of the measurements taken and done separately from population (a sample). Biostatistics usually helps in building the environmental models and testing drugs as trial for new pharmaceuticals to recommend certain drugs for human use on the basis of its efficacy and comparative effects.

One major task of biostatistics is to understand why a disease occurs in certain area (by conducting experiments in a group who got that disease- a sample) and why that disease does not occur in other areas (rest of population-a sample). This approach helps solving the environmental and biological problems. For that purpose, the biological statisticians hypothesize and use probability theory and modeling methodologies.

Advantages for properly applying statistics for a country are to keep the detailed information of people in a country, to inform the governmental authorities about the population status and to suggest them for

launching programs for balanced population status that keeps the organization of a country more stable. Organizing the matters of a country on sound statistical evaluation grounds make the plans fruitful for further developments.

However, there must in mind be the other face of the task remembering not to adapt these surveys and limited data with entirety for quick applications that might be less advantageous. Some of the programs are much expensive and time consuming and people may feel not comfortable conveying their personal information just for the sake of applying a so called organized procedure. In such conditions, one may prefer the appropriate statistical way for the convenience of public but considering moral and ethical values as well. Another quite unfortunate fact is that a statistical data can be misused for personal need of a presenter.[27]

CONCLUSIONS

Biostatistics is used in conducting clinical research comprising preparing a protocol, selecting the patients and their size, organizing a design of a clinical trial, and efficacy and comparative efficacy of certain treatment. Biostatistics helps in determining gene frequency, population of certain specific species, and is involved in almost all aspects of genetics e.g., Hardy-Weinberg theorem, Mendel's laws etc.

Biostatistics uses methodology and theory of statistics and applies in life and health sciences. The experts in biostatistics via their specific statistical approaches have made possible to understand a variety of the diseases[22] like cancer, diabetes, obesity, AIDS, heart diseases and other disorders. The institutions related to biostatistics should organize basic and higher courses, certificate courses, diploma programs, degree programs (MSc, PhD etc.) and other opportunities for students to later apply this knowledge and methods in various employment areas in industry, government, life sciences, computer science, medicine, public health, education, teaching, research, and survey research. Statisticians are, hence, very important people for establishing various schemes, programs, institutions and organizations in medical and biological sectors.

REFEERENCES

1. Hays WL, editor. Statistics for the Social Sciences. California: Holt, Rinehart and Winston;1973.
2. Moses LE, editor. Think and Explain with Statistics. Boston: Addison-Wesley;1986.
3. Moore D. Teaching statistics as a respectable subject. In: Gordon F, Gordon S, editors. Statistics for the twenty-first century. Washington DC: The Mathematical Association of America; 1992. p. 14–25.
4. Chance BL, Rossman AJ. Preface. In: Chance BL, Rossman AJ, editors. Investigating statistical concepts, applications, and methods. California: Duxbury Press;2005.
5. Dodge Y, editor. The Oxford Dictionary of Statistical Terms. Oxford: Oxford University Press;2003.
6. Anderson DR, Sweeney DJ, Williams TA, editors. Statistics: Concepts and Applications. New Jersey: McGraw-Hill;1995.
7. Wilkis SS, editor. Mathematical Statistics. New York: John Wiley;1962.
8. Thompson SP, Gardner M, editors. Calculus Made Easy. New York: St. Martin's Press;1998.
9. Random House Webster's Unabridged Dictionary. New York: Random House, Inc.;2001.
10. Cantor G. Ueber eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen, Crelles J Math 1874;77:258–62.
11. Johnson P, editor. A History of Set Theory. Boston: Prindle, Weber & Schmidt;1972.
12. Feller W, editor. Introduction to probability theory and its applications. 3rd ed. New York: John Wiley;1968.
13. Stuart A, Ord K, editor. Kendall's advanced theory of statistics, Distribution theory. Vol-1: 6th ed. New York: John Wiley;2009.
14. Wald A, editor. Sequential analysis. New York: John Wiley;1947.
15. Wald A. editor. Statistical Decision Functions. New York: John Wiley;1950.
16. Lucien LC, editor. Asymptotic methods in statistical decision theory. Heidelberg: Springer-Verlag;1986.
17. Erich L, editor. Testing statistical hypotheses. 2rd ed. New York: Springer;1997.
18. Eric L, Cassella G, editors. Theory of point estimation. 2nd ed. New York: Springer;1998.
19. Doksum KA, editor Mathematical Statistics: Basic and Selected Topics. New Jersey: Pearson Prentice-Hall;2001.
20. Freedman DA. editor Statistical Models: Theory and Practice.Cambridge: Cambridge University Press;2005.
21. Friedrich L, Klaus JM, editors. Statistical Decision Theory: Estimation, Testing, and Selection. New York: Springer;2008.
22. Emmert-Streib F, Dehmer M, editors. Medical Biostatistics for Complex Diseases. New York: Wiley-Blackwell; 2010.
23. Ewens WJ, Grant GR, editors. Statistical Methods in Bioinformatics: An Introduction. New York: Springer; 2004.
24. Causton H, Quackenbush J, Brazma A, editor. Statistical Analysis of Gene Expression Microarray Data. New York: Wiley-Blackwell; 2003.
25. Speed T, editor. Microarray Gene Expression Data Analysis: A Beginner's Guide. London: Chapman & Hall/CRC; 2003.
26. Dehmer M, Emmert-Streib F, Graber A, Salvador A, editors. Applied Statistics for Network Biology: Methods in Systems Biology. New York: Wiley-Blackwell; 2011.
27. Huff D, editor. How to lie with statistics. New York: WW Norton & Company, Inc.; 1954.

Address for Correspondence:

Prof. Dr. Zahir Hussain, Department of Physiology, Faculty of Medicine, Umm Al-Qura University, Makkah, Kingdom of Saudi Arabia. **Cell:** +966-536772763

Email: zahussai@yahoo.ca