**IMPORTANCE OF STATISTICS TO COMPUTER SCIENCE**

 **ABSTRACT**

Statistics is a branch of mathematics that deals with the collection, analysis, interpretation, presentation, and organization of data. In computer science, statistics plays a significant role in data analysis, machine learning, artificial intelligence, and many other areas. In this report, we will discuss the application of statistics in computer science.

Conceptually, the current use of computers has taken two forms in the teaching of elementary statistics: integrating the content of statistics, with that of computers; and integrating methods of instruction of statistics by use of computers. 'In the first half of this paper, three computer language textbooks are reviewed. Each uses statistics as a content area presenting programming problems. Also, three textbooks which focus on learning statistics using the computer as an aide are reviewed. The second half of this, paper surveys six published articles that evaluate courses employing "hands-on" computer instruction (CA1) and also, many published articles evaluating ‘courses employing a demonstrational mode of instruction. The generation and use of simulated experimental data and i computerized statistical packages are reviewed. Extensive recommendations for integrating computers into the teaching of statistics

 **INTRODUCTION**

Within the last five years a revolution has occurred in all courses that require calculations. From primary grades to post-doctoral study, the inexpensive electronic pocket calculator has had a pervasive impact upon the curricula. But, just as the introduction of calculators in courses of statistics greatly Influenced the development of the analysis of variance and experimental design (Evans, 1973), so too can we expect the introduction of inexpensive, programmable computers to have a greater influence on the development of statistical theory, practice, and teaching. The impact of computers on statistical theory is best exemplified by the recent work on matrix decompositions, generalized inverses, and multivariate analysis (Golub, 1969; Rao, 1966; Kshirsagar, 1972). Many of the classical, hand-calculator based methods have now become obsolete or have been revised with the advent of computers (Golub, J969). The computer has already changed statistical practice. Extensive plotting of data and residuals is quite common. There has been a shift of emphasis from general tables of statistical functions, to direct evaluation of discrete values, is unusual not to see "p-values" reported in research articles. Whereas, ten years ago ".05", ".01", and "ns" were commonplace. Jack-knifing is an example of a statistical technique who’s widespread" application would not have been seriously considered before the advent of. Computers, but it is now included in the Curriculum. Evans (1973) gives an excellent review of the influence of computers-on modern statistics. .Yet, for all the impact computers have had on the theory find practice

The following are soft wares used n statistics

STATA MICROSOFT EXCEL

TALLY SPSS SAS

MATLAB

.Computers to do use a lot of numerical data in the daily basis of operation of data as thy carry a lot of data thus needed to be organized Statistically ,computer science being a mathematics concepts statistics will be a a necessity to it

Within the last five years a revolution has occurred in all courses that require calculations. From primary grades to post-doctoral study, the inexpensive electronic pocket calculator has had a pervasive impact upon the curricula. But, just as the introduction of calculators in courses of statistics greatly Influenced the development of the analysis of variance and experimental design (Evans, 1973), so too can we expect the introduction of inexpensive, programmable computers to have a greater influence on the development of statistical theory, practice, and teaching. The impact of computers on statistical theory is best exemplified by the recent work on matrix decompositions, generalized inverses, and multivariate analysis (Golub, 1969; Rao, 1966; Kshirsagar, 1972). Many of the classical, hand-calculator based methods have now become obsolete or have been revised with the advent of computers (Golub, J969). The computer has already changed statistical practice. Extensive plotting of data and residuals is quite common. There has been a shift of emphasis from general tables of statistical functions, to direct evaluation of discrete values. is unusual not to see "p-values" reported in research articles. Whereas, ten years ago ".05", ".01", and "ns" were commonplace. Jack-knifing is an example of a statistical technique who’s widespread" application would not have been seriously considered before the advent of. Computers, but it is now included in the. Curriculum. Evans (1973) gives an excellent review of the influence of computers-on modern statistics. .Yet, for all the impact computers have had on the theory find practice of statistics, only recently have there been attempts to integrate the use of computers with the teaching of elementary statistics. The purpose of this article is to explore the current use of computers in the teaching of elementary statistics. Conceptually, this exploration will be .in two forms. First, the various means of integrating computers into the content of elementary statistics will be examined. Criteria for identifying the strengths and weaknesses of published textbooks and computerized statistical packages will be examined from the viewpoint of content relevance. Second, the various methods o£ implementing the 'integrated content of statistics and computerized statistical packages will be surveyed. Although the principal emphasis is on an introductory non-calculus course, most of the methods described in the second part of this paper are useful in higher level courses in statistics and research methodology

 **Importance of statistics to computer science**

1. Data mining is the process of discovering patterns in large datasets. Statistical techniques such as cluster analysis, principal component analysis, and factor analysis are used to identify these patterns. Data mining is used in various fields such as marketing, finance, and healthcare to extract useful information from large datasets.

2. Big data refers to the massive amounts of data that are generated by various computer applications and systems. Computer scientists use statistical techniques to analyze the Big Data Analytics: Big data refers to the massive amounts of data that are generated by various computer applications and systems. Computer scientists use statistical techniques to analyze these datasets and extract useful information. Big data analytics is used in various fields such as finance, healthcare, and marketing to make informed decisions based on the data.

3. Software Development: Statistics is used in software development to test and validate software applications. Statist Software Development: Statistics is used in software development to test and validate software applications. Statistical techniques such as hypothesis testing and confidence intervals are used to ensure that the software performs as expected. Computer scientists also use statistical techniques to identify and fix software bugs and to optimize software performance.

4. STATISTICS IN HARDWARE MANUFACTURING The hardware manufacturing companies are applying statistical approaches to create a plan of action that will work more efficiently for forecasting the future productivity of the hardware enterprise[8]. Adopted statistical approaches for: Forecasting production, when there is a stable demand and uncertain demand. Pinpoint when and which inputs of a specific model will be the cause of uncertainty Calculate summary statistics in order to set sample data. To make market analysis and process optimizations. Statistical tracking

8. STATISTICS IN BIOINFORMATICS Bioinformatics is the application of "computational biology “to the management and analysis of biological data. Concepts from computer science, discrete mathematics and statics are being used increasingly to study and describe biological systems. Bioinformatics would not be possible without advances in computer hardware and software: analysis of algorithms, data structures and software engineering. To elaborate algorithms on computers increased the awareness of more recent statistical methods. Statistical analysis for differently expressed genes are best carried out via hypothesis test. More complex data may require analysis via ANOVA or general linear models

6. STATISTICS IN ARTIFICIAL INTELLIGENCE Artificial intelligence (AI) is the intelligence exhibited by machines or software. Popular AI approaches include statistical methods, computational intelligence, machine learning and traditional symbolic AI. The goals of AI include reasoning, knowledge, planning, learning, natural language processing, perception and the ability to move and manipulate objects. There are a large number of tools used in AI, including versions of search and mathematical optimization, logic, methods based on probability and economics, and many others [4]. The simplest AI applications can be divided into two types: Classifiers: Classifiers are functions that use pattern matching to determine a closest match. A classifier can be trained in various ways; there are many statistical and machine learning approaches. The most widely used classifiers is the neural network. Controllers: Controllers do however also classify conditions before inferring actions, and therefore classification forms a central part of many A

In conclusion

In conclusion, statistics plays a vital role in computer science. It helps in data analysis, machine learning, probability theory, computer networks, software testing, and optimization. As computer science continues to evolve, the role of statistics in this field will become even more critical.

**REFERENCES**

 [1] Lauro, C. (1996). Computational Statistics or Statistical Computing, is that the question? Computational Statistics and Data Analysis, Vol. 23, pp.191–193.

 [2] Billiard, L. and Gentle, J.E. (1993). The middle years of the Interface. Computing Science and Statistics, Vol. 25, pp.19–26. [3] Yates, F (1966). Computers: the second revolution in statistics. Biometrics, Vol. 22.

[4]Cheng, B. and Titterington, D. M. (1994). Neural networks: a review from a statistical perspective. Statistical Science, Vol. 9, pp.2-54.

[5] Elder, J. F. and Pregibon, D. (1996). A statistical perspective on knowledge discovery in databases. Advances in Knowledge Discovery and Data Mining, MIT Press, pp.83-115.

[6] Gentle, J.E. (2004). Courses in statistical computing and computational statistics. The American Statistician, Vol. 58, pp.2–5. [7] Grier, D.A. (1991). Statistics and the introduction of digital computers. Chance, Vol. 4(3), pp.30–36.

 [8] Friedman, J. H. and Fisher, N. I. (1999). Bump hunting in high-dimensional data. Statistics and Computing, Vol. 9

9. Statistics for Computer Scientists: A Bayesian Approach" by Radford M. Neal - This book provides an introduction to Bayesian statistics and its application to computer science.

10."Statistical Methods in Computer Science" by William W. Hwu - This book provides an introduction to statistical methods and their applications in computer science.

11. Introduction to Probability and Statistics for Computer Science" by James M. Steele - This book provides an introduction to probability and statistics, with an emphasis on applications in computer science.

12. Data Science from Scratch: First Principles with Python" by Joel Grus - This book provides an introduction to data science, including statistics, with a focus on Python programming.

13Elements of Statistical Learning: Data Mining, Inference, and Prediction" by Trevor Hastie, Robert Tibshirani, and Jerome Friedman - This book provides an in-depth introduction to statistical learning, with a focus on machine learning algorithms and their applications in computer science.

.