64th ISI World Statistics Congress - Ottawa, Canada

Paper-Statistical Computing and Data Visualization

Author and Organizer-Md. Shariful Islam

Participant

Dr Stephen MacFeely(Chair)

Md Shariful Islam(Presenter/Speaker)

sushanta Paul(Speaker)

Prof. Ying Chen(Discussant)

Category: International Statistical Institute

Brief Description

Statistical computing Statistical analysis involves use of observational data together with domain knowledge to develop a model to study and understand a data-generating process. The data analysis is used to refine the model possibly to select a different model, to determine appropriate values for terms of the model and to use the model to make inference concerning the process. Computer are used in virtually every field of human endeavor such as in business, government, legal profession, medicine, education, industry, entertainment and sports, agriculture, and the home. The following field are indicated where computers are used each and every time. That are software and hardware, Decision support analyst, programmer analyst, MIS Manager, Application programmer, Telecommunication engineer, Data processing specialist, Computer training specialist, Data Processing profession, Computer Engineer, Systems Programmer, Computer marketing professionals and Systems Analyst.

Statistical computational Methodology

Computers are useful in all aspects of statistical data analysis, of course, and generally in computational statistics, and focus on statistical methods that are computationally intensive. Although a theoretical justification of these methods often depends on asymptotic theory, in particular, on the asymptotic of the empirical cumulative distribution function, asymptotic inference is generally replaced by computational inference. Selection of a model implies consideration of more than one model. As suggested above, this is one of the hallmarks of computational statistics: looking at data through a variety of models. Cross validation and its generalizations and resampling are important techniques for addressing the problems. Resampling

methods also have much wider applicability in statistics, from estimating variances and setting confidence regions to larger problems in statistical data analysis. Computational inference depends on simulation of data-generating processes. Any such simulation is an experiment, and principles for design and analysis of experiments using computer models are discussed. This led to the development of more user-friendly programmes (due to the evolvement of windows) like the SPSS (Statistical Package for the Social Sciences), developed in the USA, and GENSTAT (a General Statistical Program) and GLIM, both developed in Britain. The most user-friendly program that could analyze statistical problems is the Microsoft Excel. The programming STATA. The effect of computers is much more noticeable in advanced statistics, such as the Multiple Regression.

Statistical computing in Big Data Context

For statistical computation in respect of Big Data following software engineering and statically engineering are used: Software Engineering: Statistical Engineering Requirement Prior elicitation Design Experimental Design Implementation Data acquisition Verification Model Specification Maintenance Indifference Modern Validation Model Selection Decision making Big Time Big Data Statistics The big Data analytics market is set to reach $103 billion by 2023. Poor data quality costs the US economy up to $3.1 trillion yearly. In 2020, every person generated 1.7 megabytes in just a second. Internet users generate about 2.5 quintillion bytes of data each day. 95% of businesses cite the need to manage unstructured data each day. 97.2% of organizations are investing in big data and AI. Using the big Data Netflix saves $1 billion per year on customer retention. Statistical Computing software(SCS) with diverse decisioning opportunities –such as retail, insurance, manufacturing, public sector, Telecom, media and more areas its activities are seen actively. In retail section SCS helps for decisioning as meeting merchandising demands for specific locations, improving the supply chain with near real time inventory balancing, guiding customers on a path to purchase through personalized digital interaction, and detecting and preventing fraud quickly to minimize margin losses.

Decisioning on the base of Statistical computing software(SCS):

Decisioning capabilities from SCS offer of possibilities for insurance companies-from striving to meet compliance requirements such as IFRS 17, to detecting and preventing claims of underwriting fraud, to using IoT data and AI to generate real-time pricing, to offering health and well-being services. Manufacturers use SCS to become nimble, AI-driven organizations that minimize risk and seize opportunities through deep operational insights and confident decision making on manufacturing quality, connected factories, supply chain optimization and more.

Data Visualization For Statistical computing

Data Visualization is more important at present world. without Visualization computational result of data cannot be shown properly. In this respect it is described that Data visualization is the

process of graphical representation of data in the form of geographic maps, charts, spark lines, infographics, heat maps, or statistical graphs. Data presented through visual elements is easy to understand and analyze, enabling the effective extraction of actionable insights from the data. Relevant stakeholders can then use the findings to make more efficient real-time decisions. Data visualization tools, incorporating support for streaming data, AI integration, embed ability, collaboration, interactive exploration, and self-service capabilities, facilitate the visual representation of data. Data visualization is very critical to market research where both numerical and categorical data can be visualized, which helps in an increase in the impact of insights and also helps in reducing the risk of analysis paralysis. Data visualizations can either stand alone as single visualizations or they can be combined with multiple charts to create dashboards. Humans are programmed to think visually – so the way display visualizations if very important. For example: - 65% of people are Visual Learners - process images 60,000 times faster than text - It takes 13 milliseconds to recognize an image This makes effective visualizations more important than ever. For that need to help users understand complex information faster. Data visualization is also a medium to tell a data story to the viewers. The visualization can be used to present the data facts in an easy-to-understand form while telling a story and leading the viewers to an inevitable conclusion. This data story, like any other type of story, should have a good beginning, a basic plot, and an ending that it is leading towards. For example, if a data analyst has to craft a data visualization for company executives detailing the profits on various products, then the data story can start with the profits and losses of various products and move on to recommendations on how to tackle the losses.

There are 15 Data visualization tools at modern computing arena. That’s 1. Tableau2. Dundas BI3. JupyteR4. Zoho Reports5. Google Charts6. Visual.ly7. RAW8. IBM Watson9. Sisense10. Plotly11. Data Wrapper12. Highcharts13. Fusioncharts14. Power BI15. QlikView. These tools are used for visualization the data in different statistical computation and accelerate the world to visualize the data more and more.

Abstract

Now a days no statistical research can be done without substantial informatics support. My presentation covers overview the computational statistics with visualization by the implication of computerization. However, the role of computational statistics in statistical education and discovery, including the big data analyses, has been under-recognized even by peer statisticians. Especially in the presence of massive data coming with more heterogeneity we need to change our statistical thinking in order to adapt classical statistics, still invaluable in the context of big data analysis, and software developments in statistics that address new challenges. Moreover, the data which come from two or more sources and are characterized by different stochastic mechanisms may sometimes be mixed and registered together, which leads to polluted samples. For all these cases the traditional stochastic models, suitable for primary events (the ones which actually occur) may not be appropriate for description of registered events (observed data) unless necessary modifications are introduce. The adaptation process should be based on incorporating into curriculums the topics as: exploratory data analysis(EDA) and visualization, advanced statistical modelling and forecasting hypothesis testing from randomized experiments, planning of adaptive experiments, to name only a few. They all need special teaching methods based on both statistical and computer-science tools. Similar approach can be make familiar with computationally. My presentation covers all field Computational Statistics and the field of Data Visualization by the computational software at present and future world. In line with the companion volumes, it contains a collection of chapters high volume of computational areana of software which are used for computational work and visualization that run the world at a glance. Data Visualization is an active area of application and research and this is a good time to gather together a summary of current knowledge. Graphic displays are often very effective at communicating information. They are also very often not effective at communicating information. Two important reasons for this state of affairs are that graphics can be produced with a few clicks of the mouse without any thought, and that the design of graphics is not taken seriously in many scientific textbooks. Data visualization is more important part of computational Statistics. This part shows all computational scenario of the calculating arena by the dashboard and users easily determine their acts easily. My presentation covers all area at present world's computational software and describe easily how to work such kind of software with visualization. This volume of the presentation of Computational Statistics takes Data visualization tools which provide an accessible way to see and understand trends, patterns in data, and outliers, data visualization tools and technologies are essential to analyzing massive amounts of information and making data driven decisions, the concept of using pictures is to understand data that has been used for centuries.