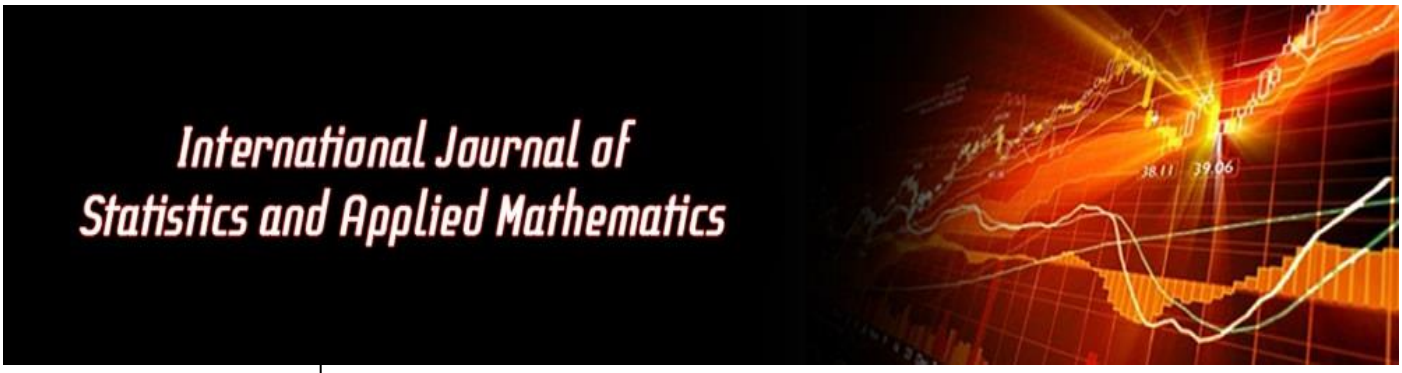
International Journal of Statistics and Applied Mathematics 2018; 3(3): 107-110



ISSN: 2456-1452
Maths 2018; 3(3): 107-110

www.mathsjournal.com
Received: 13-03-2018
Accepted: 14-04-2018

Nduka Ethelbert Chinaka
Department of Mathematics and Statistics, University of Port Harcourt, Port Harcourt, Rivers State, Nigeria

Ogoke Uchenna Petronilla
Department of Mathematics and Statistics, University of Port Harcourt, Port Harcourt, Rivers State, Nigeria

Correspondence
Nduka Ethelbert Chinaka
Department of Mathematics and Statistics, University of Port Harcourt, Port Harcourt, Rivers State, Nigeria

Data analysis: Imperatives for optimal decision making

Nduka Ethelbert Chinaka and Ogoke Uchenna Petronilla

Abstract
The quality of any decision making process largely depends on the data system. In this wise, the triad of experimental design, analysis and decision making is vital for planning. The challenges of data collection, use and mis-use of statistical analyses, software interpretations and opportunities for optimal decision making as well as suggestions for new thinking are highlighted.

Keywords: Imperatives, optimal decision making

Introduction
Data analysis has always dominated discourse among researchers, statisticians and all users of data in their day to day activities. In this wise, the refrain has always been that serious research requires serious statistics. However, the triad of experimental design, analysis and decision making which is a chain has not been much emphasised. Data analysis poses general challenges in the data system (i.e. all men and materials involved at all stages of experimental design, data collection and analysis). As noted by Afonja (1985) [1], the statistician faces the challenge of providing a realistic solution to these problems by devising optimal methods of data collection and analysis compatible with the environment so that at the end, he not only minimises any uncertainty and hence falsehood but assesses the exact degree of such uncertainties.
The interdisciplinary nature of statistics and its near universal application calls for utmost interaction between the statistician (as the consultant) and the experimenter/researcher (as the client) in all stages of experimentation and analysis. In all these, data analysis seems to play a central role for optimal decision making.

Data Analysis as Triad
- What precedes data analysis is design of experiments and what succeeds it is optimal decision. In this chain, the design of experiment embodies as follows:
- Recognition and objective of the problem which involves stating and developing clear experimental objectives.
- Choice of factors (treatments) which involves randomizing the factors and levels of factors, replication, responses, etc.
- Choice of experimental design and experimentation which involves data collection and collation.

Data analysis embodies as follows
- Data cleaning, screening for missing observations, outliers, etc.
- Appropriate statistical methods for numerical accuracy which include graphical, exploratory methods, etc.
- Analytical software-SPSS, SAS, STATA, R, GENSTAT, STATISTICA, etc.

Optimal decision embodies as follows
- Inferences, recommendations, presentations.
- Implementation of policy direction

This can be summarised in Nduka (1997) [7]

Wrong design of experiments + right analysis ⇔ biased result

Right design of experiments + wrong analysis ⇔ biased result

Right design of experiment + right analysis ⇔ unbiased result

Challenges to Statisticians (Data Analysts)

We are living in the e- age where technology has made it easier to collect large data. This data explosion has thrown up its own challenges/implications in analysing and making sense of all the data collected.

The following are also challenges:

- Interdisciplinary nature of statistics calls for utmost interaction between statisticians and experimenters (researchers), Nduka (1997) [7], Asiribo (1991) [4].
- Statistician-researcher-misconception.

Some statisticians are more interested in the secondary aspect of research activity ie data analysis thereby being perceived as data "crunchers", also insufficient exposure to the rudiments of other disciplines. Again, some researchers are not involving the statistician all through the stages of experiments and often come with finished experimental data whose result of analysis may not conform to their expectation and may express disappointment with the results.

- Lack of basic knowledge of statistical principles and methods.
- Dearth of statistician, lack of awareness and sufficient recognition of the role/importance of statistics, etc.
- Student population explosion versus the shortage of personnel, poor training, retraining availability, computing software, etc.
- Mis-use and abuse of statistical methods and software in data analysis.

Statistical Packages in Data Analysis

The existence and availability of an array of software packages has tremendously enhanced data analysis and ease computational difficulties in complex experiments and results. Such include, SPSS, SAS, GENSTAT, STATA, MINITAB, R, E-views, CSPRO, ECONOMETRICA, SOLAS, INSTAT, STATISTICA, S-PLUS, EPI-INFO, SIMULA,
SSC-STAT, etc.

The features of these software packages need to be evaluated in terms of user friendliness, similarity of menus, efficiency, number of iterations, etc. Each software may recommend itself for a particular problem. e.g. CSPRO (mapping census and survey data)

SPSS (general package for data processing)

SAS (general package, with a comprehensive system of component)

GENSTAT (general package for practice in modelling and analysis of designed experiments)

MINITAB (general package for data analysis, popular for teaching)

R (general object oriented, statistical programming language, giving access to cutting- edge statistics)

EPI-INFO (package especially for epidemiological studies)

SSC-STAT (an Excel add-in data manipulation for basic statistical work), see Stern, R. *et al*. (2005) [11].

SIMULA (general package for simulation studies)

SOLAS (general package for missing values in an experiment), etc.

University Data System

University ranks among the generators of large data, but ranks low in the use of data in its operations unlike others; internet providers, mobile communication networks, banks, airlines, etc, who make use of theirs in the competitive environment. As the conception or misconception of its existence as a social institution, it is yet to tap into the use of data for its maximal operations. The Academic Planning Unit has the core mandate to collect, collate, analyse all data generated for the development of the university data bank for better service delivery, NUC (1996) [9]. The core mission is to develop a well-coordinated data system to guide policies for better decision making.

The primary generators of data in Nduka (2015) [8]. Are

I. Vice Chancellor's Office: Physical Planning, Internal Audit, Advancement (Alumini) Office, Protocols, Transport and Logistics, Procurement, Information and Communications Technology, Academic Planning, etc.
II. Registry: Personal Academic, Senior Professional, Administrative to technical staff, Junior staff, Establishment, Exams and Records, Student Affairs, Admissions Office, Legal Unit, etc.
III. Bursary: Students 'Account, Budget Office, Main Account, Treasury, Payroll, Pensions and Welfare, etc.
IV. Library
V. Faculties (Colleges, Schools, etc).
VI. Works and Services: Housing Unit, Maintenance (Electrical, Mechanical, Civil) Utilities (Electricity, Diesel)
VII. Health Services etc.

Challenges

I. Lack of co-ordinated (harmonized)data system in the various units e.g. information from/between units in registry may differ on the same subject (number of students between units)
II. Who should be the custodian of university data?
III. Who should be the disseminator of university data?
IV. Lack of personnel with requisite qualifications and expertise
V. Weak academic planning units
VI. Lack of awareness on the importance of timely, accurate and reliable data for proper planning.

Action plan

I. Sensitization/Awareness programme for all primary generators/users of all forms of data on the importance of timely, accurate and reliable data for effective, efficient planning and decision making.
II. Recruitment, training and retraining of staffs as focal officers for various faculties/units in data collection and collation.
III. Academic planning unit should be the data bank and think-tank like R&D in the industries and parastatals acting like NUC's liaison office that pushes data to and from Abuja.
IV. Periodic briefing to:

- Senate on certain statistical issues (e.g. annual activity analysis)
- Management (e.g. cost analysis, responsibility allowance, staff recruitment, students' admission etc.)
- Principal Officers/Vice Chancellor (on confidential issues)

V. Statistical digest/news in the university news bulletin

VI. Workshop/Lectures on record keeping and data management
VII. Publications of annual statistics digest as reference/ resource material

Opportunities for Data Analyst

There has been a growing statistical awareness, particularly since the establishment of planning, research and statistics in all Ministries, Departments and Agencies (MDA), Statistician General of the Federation, State Statistician General, etc.

- Opportunities exist today for the data analyst in the academia, government, multinationals, industries, the private sector, etc.
- The appreciating application of statistics in diverse areas account for these opportunities.
- In the academia: workforce demands and needs, research and funding education, support and mentorship. Growth in statistics programme enrolments are opportunities to respond to both employer demand, collaboration, skills and institutional pressure to develop new programs. Faculty members are sought after for research collaborations and private sector consulting.
- In Government: new opportunities in many interesting and complex problems in defence, health, education; etc.; integration and harmonization of data sources/agencies e.g. Nigerian population commission, Federal Road Safety Commission, National Identity Management Commission, National Bureau of Statistics, etc.
- In Multinationals/Industries/Private Sector: the emergence of technology, explosion of 'big' or 'digital' data offer opportunities for new statistical methods of analysis, new statistical software and more training in these methods for better decision making processes. Large amount of data exists and enlightened companies who use this as a source of competitive advantage will use data analysts to uncover inherent information in the data for their comparative advantages, even in their SWOT analysis, better programme and data management.
- Opportunities now exist to bridge the gap between developed and developing countries via programs like Statistics Without Borders Outreach Program which helps statisticians work with colleagues across the countries in survey and data analysis, LISA 2020 (Laboratory for Interdisciplinary Statistical Analysis 2020) project for building capacity by training statisticians in developing countries on collaboration and building statistical laboratories, statistical training program for Africa, etc. American Statistical Association (2015, 2016), further buttressed the opportunities for statisticians in its report as the fastest growing course recently.
- Since the 1990's, the number of undergraduate degrees in statistics has risen by more than 300% compared with 72% in other Science and Engineering within the same period, Masters degree 260%, PhD 132%.
- Demand for statisticians is likely to continue to outpace this growth.
- Projected shortage of statisticians is between 140,000 and 190,000 by 2018.
- Employment for statisticians has grown from 28,000 in 2010 to 85,000 in 2014.
- Projected job growth to increase by 27% between 2012 and 2022

Compared with 11% rate for other occupations, annual production of graduates as BSc-2000, MSc-3000, PhD-575 which does not match this demand.

- Statisticians earn stronger wages $80,000.00 for those with less than three years and $150,000.00 for those of 9 years and above.
- Chief Economist at Goggle, Hal Varian (2015) stated ''I keep saying that the sexy job in the next ten years will be statisticians. And I am not kidding''. The salaries for statisticians with doctoral degrees currently start at $125,000. 00, and they will probably continue to increase.

Vii) Suggestions for New Thinking

- The challenges, opportunities and impetus for the statisticians call for new thinking and strategies for the future as follows:
- The impact of '' Big Data'' on research and teachings of statistics.
- It can be classified as ''non-designed'' (complex and unstructured) data which needs to be analysed. Others will analyse them if statisticians do not. It is imperative that we get to the table earlier.
- Statistics should adopt ''Diffuse model''ie more statistics courses in many department programs.
- Centre for data analysis in ministries, governments, etc.
- Alternative/new approach to statistics training of less theory, more practice which by focusing on impact will flourish and transform statistics as a discipline to realise the vision of big data and produce the ''golden era of statistics'' (Rudin, *et al*, 2014) [10].
- Develop statistical games eg survey game to show the benefits of stratification, agricultural practices and yields, basic experimental designs, games theory, etc.
- Start off as a general practice statistician and then specialty area of interest eg survey statistician, medical statistician, biometricians, etc.
- Statistics to reformulate how research is to be done, team work (interdisciplinary eg statistics and computer science need each other) to tackle big problem (ie''all data revolution'').
- Evolve new software packages to tackle emerging problems of ''big data'' and other complex problems in MDAs, industries, etc, because statistical analysis and data mining will be the hottest skills in future.

Conclusion

Data analysis will continue to play central role in our data system. In this wise, statisticians as main practitioners of data analysis need to appreciate their importance in statistical awareness, research, development, consultancy and guidance for better decision making. Major planning and decision making in government, academia, industry and general public in future will be data driven, hence the need to position ourselves for the challenges, opportunities, and benefit therein.

Training, retraining and interdisciplinary exposure hold the key to yawning gaps to meet the surging demands ahead. The 'large data' revolution calls for innovation, creativity, new skills, development of methodologies and software packages to analyse such data for meaningful impact in the society.

Science of the future is data science and such future belongs to the statistician who as the major data analyst will be highly employable. The challenges, opportunities of 'big data' need to be appraised in ethical issues (personal privacy and

information security), growing heterogeneity of data sources, undesigned or undersigned data in all ramifications.

References

1. Afonja B. Facts, Figures and Falsehood Syndrome in the Society. Inaugural lecture, university of Ibadan, 1985.
2. American Statistical Association More students earning statistics degrees not enough to meet surging demand in science. Daily, 2015, 1-5.
3. American Statistical Association: Success, Opportunities and Challenges for Statistics and Biostatistics in the Data Science: A report of July 2016 NSF-sponsored workshop for chairs of Departments of Biostatistics and Statistics, 2016, 1-15.
4. Asiribo OE. Statistical Consulting in Agriculture. Journ. Nig. Stat. Asso. 1991; 6:1-7.
   Https://bi-survey.com/big data.
5. Kettenring JR, Koehler KJ, Mckenzie JD. (Jr). Challenges and Opportunities for Statistics in the Next 25 years. The American Statistician. 2015; 69(2):86-9.
6. Lohr S. Hottest New Jobs: Statistics and Mathematics in Introduction to Data Analysis and Decision Making. Published: Cangage learning, 2010, 1-3.
7. Nduka EC. Statistics in Agricultural Research: the Nigerian Environment. Proc 12th Ann. Conf. Farm Mgt Assoc. Nig, 1997, 129-133.
8. Nduka EC. Source of Data for Academic Planning, paper delivered at Federal University of Technology Owerri: Directorate of Human Resources, 2015.
9. Nigerian Universities Commission: Manual on University Management, 1996, 58.
10. Rudin C, Dunson D, Irizarry JIH, Laber E, Leek J, Mc Cormick T *et al*. Discovery with Data: Leveraging Statistics with Computer Science To Transform Science and Society, Alexandria, VA: ASA, 2014, 87.
11. Stern R. Allen E, Barahona C, Dale I, Grayer C, Leidi S. SSC Statistical Resources Handbook, Statistical Services Center, University of Reading, U.K, 2005.
12. www.en.wikipedia.org/wiki/data (accessed June 4), 2015.