ISSN 2348-1218 (print)
International Journal of Interdisciplinary Research and Innovations ISSN 2348-1226 (online)
Vol. 6, Issue 2, pp: (1-5), Month: April - June 2018, Available at: www.researchpublish.com

REGRESSION INVENTED AS STATISTICS

Chuda Prasad Dhakal

PhD, Tribhuvan University, Institute of Agriculture and Animal Sciences, Rampur Campus, Chitwan, Nepal

Abstract: **The central theme of the paper is to shed light on the history of invention of statistics. In this paper, a comprehensive review of literature in 'statistics invented as regression' has been made. Covered the definition of statistics in the changed context, this paper frameworks sufficient debate and discussions about regression analysis considered to be the invention of statistics as a discipline. In the paper, regression in the context of least square theory invention is discussed. And, enclosed is, the debate among Legendre and Gauss, who to be taken for the inventor of least square theory. Comprehensively put forward is, when and who invented this powerful and most flexible discipline regression, which one day had to be the invention of the sole subject statistics. This paper interests any individual who wants to know how statistics was invented and what that was at the very first, was considered to be the subject statistics.**

Keywords: **definition of statistics, least square theory, regression.**

I. INTRODUCTION

In early times, the meaning of statistics was restricted to information about states. This was later extended to collection of information of all types. In the later days, it was extended to include the analysis and interpretation of such data. Finally, in modern terms, "statistics" means both sets of collected information and analytical work which requires statistical inference (History of statistics, 2016). At present, statistics in common is defined as: science of collecting, summarizing, analysing and, interpreting them related to a variable/s to draw meaningful inferences. Delorme (2006) has defined statistics as "the body of analytical and computational methods by which characteristics of a population are inferred through observations made in a representative sample from that population". Yet, the definition of statistics is not motionless. It keeps on changing.

According to Historical Context (2015) two contrasting definition of statistics at some 60+ years apart are: "Statistics is the branch of scientific method which deals with the data obtained by counting or measuring the properties of populations of natural phenomena. Where, 'natural phenomena' includes, all the happenings of the external world whether human or not." Professor Maurice Kendall, 1943. And, "Statistics is: the fun of finding patterns in data; the pleasure of making discoveries; the import of deep philosophical questions; the power to shed light on important decisions, and the ability to guide decisions in business, science, government, medicine, industry etc." Professor David Hand. These two definitions indicate, the discipline of statistics has moved from being grounded firmly in the world of measurement and scientific analysis into the world of exploration, comprehension and decision-making.

The discipline statistics is an applied science. It requires probability theory to be put on a firm theoretical basis. Historical context (2015) reports, gambling was one of the most important early drivers of research into probability and statistical methods. According to which, 'Abraham De Moivre's book, published in 1718, "The Doctrine of Chance: A method of calculating the probabilities of events in play" was essential reading for any serious gambler at the time. The book contained an explanation of the basic ideas of probability, including permutations and combinations, together with detailed analysis of a variety of games of chance etc.

Usage of statistics at present has ranged between, a relatively small set of specific application areas to almost every lifestyle, as the information available to all sectors in the society is massively expanded. Understanding this information, and making well-informed decisions based on such understanding, is the primary function of modern statistical methods. Besides, some diverse fields where statistics is used are: Economics, agriculture, health sciences etc.

Statistics introduced as a discipline, this paper comprehends when and who, in what context, was the first to use the word 'statistics' as a subject. With the most flexible and powerful statistical technique *regression* defined, this paper has thoroughly reviewed how and, as a case of which, the subject statistics was investigated. Should this be a good read to any statistician and/or a starter, or an expert of the related academia, is the rationale of writing this paper.

II. STATISTICS FIRST USEDE

When was statistics first invented or used? Different views and perspectives are found regarding this question. According to timeline of probability and statistics (2016) in 9th century - Al-Kindi, an Iraqi national was the first to use statistics based on frequency analysis. This [timeline of probability and statistics (2016)] claims, in 1560s (published 1663) – Cardano's Liber de ludo aleae attempted to calculate probabilities of dice throws. Whereas, history of statistics (2016) claims the history of statistics started around 1749.

History of statistics (n.d.) claims; 1) Shakespeare used a word 'Statist' in his drama Hamlet (1602); 2) Gottfried Achenwall used the word 'statistik' at a German University in 1749; 3) In 1771 W. Hooper (Englishman) used the word 'statistics' in his translation of Elements of Universal Erudition written by Baron B.F Bieford and, 4) During the 18th century the English writer have used the word statistics in their works.

The other view is, statistics was originated due to a breakthrough in game of chance in the early 18th century. Aldrich (2000) claims probability and statistics were invented during 1650-1700. But Denis (2000) reports, Galton's discovery of Regression and Correlation technique is to be considered for the origin of the subject. However, historical context (2015) discloses that, much of the foundation work for the subject had been developed in the last 150 years, despite its beginning dated back to the 13th century involving the expansion of the series $(p+q)^n$, for n=0,1,2.... famously known as Pascal's triangle.

Probability and statistics are taken to be two different things. History of probability (2016) claims that, statistics deals with data and infers from it. Whereas, probability deals with the random processes which lie behind data or outcomes. Probability theory began in seventeenth century France when the two great French mathematicians, Blaise Pascal and Pierre de Fermat, corresponded over two problems from games of chance. Probability and statistics became closely connected through the work on hypothesis of R. A. Fisher and Jerzy Neyman, which is now widely applied in biological and psychological experiments and in clinical trials of drugs and, as well as in economics and elsewhere.

III. REGRESSION DEFINED

Statistical concepts have an important impact on a wide range of sciences. These include the design of experiments and approaches to statistical inference such as Bayesian inference, each of which can be considered to have their own sequence in the development of the ideas underlying modern statistics. Most prominent has come to be Regression methods.

Regression is a statistical tool for investigating the relationship between variables. It is frequently used to predict the future and understand which factors cause an outcome. By convention, the variable that we are trying to predict is called the dependent variable and the variables that we are using as predictors of that variable are called independent variables. For example, we might wish to explain the relationship between education, measured in years of formal schooling, and income. In this case, we would typically predict income based on no of years of education. This is one simple example, besides this there are a lot that can be done using regression analysis. Simple regression and correlation form the basis of what is generally referred today, as regression analysis. Modern time regression analysis for many reasons, qualifies as one of the most useful and powerful statistical techniques. This is taken to be the most flexible analytical tool ever developed by statisticians. The historian of statistics Stephen M. Stigler calls regression the " automobile " of statistical analysis.

IV. INVENTION OF REGRESSION AS STATISTICS

Legendre (1805) and Guass (1809) (as cited in Wikipedia, 2013) invented the method of least squares, the earliest form of regression. Also, Wikipedia (2013) reports that later in 19th century the term 'regression' was coined by Francis Galton while describing the biological phenomenon. A similar discussion on the subject by Galton (1886) is given below:

It appears that Sir Francis Galton (1822-1911), a well-known British anthropologist and meteorologist, was responsible for the introduction of the word "regression." Originally, he used the term "reversion" in an unpublished address "Typical laws of heredity in man" to the Royal institution on February 9, 1877. The later term "regression" appears in his Presidential address made before section H of the British Association at Aberdeen, 1885, printed in *Nature*, September 1885, pp.507-510 and in a paper "Regression towards mediocrity in hereditary stature".

Fenberg (1992) reveals that invention of probability should not be taken as the invention of statistics. He argues, probability is the mathematical part but not a statistical method. His idea is Gauss Laplace-synthesis, which combined the normal error theory with the curve fitting method of least square was an inferential approach to the analysis of data using linear models, the first and the foremost event invented in the history of statistics. Hence the argument is, it was the method of least square which came out to be the method of regression, seeded out statistics and, Gauss is the one who should be credited for.

About the origin of regression technique Armstrong (2012) claims:

Regression analysis entered the social sciences in the 1870s with the pioneering work by Francis Galton. But "least squares" goes back at least to the early 1800s and the German mathematician Karl Gauss, who used the technique to predict astronomical phenomena.

Regression method was originated from methods of least squares. Francis Galton initially described biological phenomenon using regression technique. Stanton (2001) claims that Galton's work on inherited characteristics of sweet peas led to the initial conceptualization of linear regression and further work of Galton and Pearson brought the modern notion of correlation and regression.

As such, history of this particular statistical technique is traced back to late nineteenth century. This was possible due to the pursuits of gentleman scientists Francis Galton from England. The term regression was first applied to statistics by him. Galton used the term regression to explain a phenomenon in nature. Galton's most important insight about regression theory came to him while he was dealing with how the human characteristic of height was passed from one generation to the next. His unremarkable conclusion was, 'tall people usually had tall parents and short people usually had short parents.'

Famous mathematician Carl Friedrich Gauss thought his discovery of statistical regression was insignificant when he first invented this. For him this was such simple that, he thought he must not have been first to use it. But, someone else already should have used this. Hence, he did not publicly state his finding until many years later when his contemporary Adrien-Marie Legendre had published on the method. Gauss, then suggested that he had used this method before Legendre. Carter's (Rice University, 1995-test book on linear algebra) (as cited in Talk Stat, 2013) explains, Gauss developed least square regression, supports Legendre's idea of the invention of the method of least square. However, he [Carter] reveals that Gauss did not publish the method until 1809. According to Kopf (2015) the debate who was first to invent the method of least square, between Legendre and Gauss set off "one of the most famous priority disputes in the history of science. Which, at the end, Gauss was given most of the credit as the founder of regression with a big fight.

Legendre was the first to make his discovery of the least squares method public. He supplied the original delivery and example of the use of least squares regression. Legendre was confident that his method was a winner. However, today Gauss receives most of the credit for the invention of least squares, and thus regression. This is primarily because Gauss's explanation was so much more fully realized than Legendre's. Stigler explains , "When Gauss did publish on least squares, he went far beyond Legendre in both conceptual and technical development, linking the method to probability and providing algorithms for the computation of estimates." A detail reading about the discovery of the method of least square can be found in Packett (1972). Here is an extract from it.

Sartorius had no hesitation in assuming that practical need, the observation of nature itself, led Gauss to the method of least squares. The year in which he first applied the method is given variously as 1794 and 1795. Gauss then turned to the problem of establishing a relationship between the principle of minimizing a sum of squares and the calculus of probability. According to a contemporary account of his lecture to the Royal Society of Gottingen on 15 February 1826 (Werke, 4, 98), he formulated in 1797 the problem of selecting from all possible combinations of the observations that one which minimized the uncertainty of the results.

“Regression” came to be associated with the least squares method of prediction by the late 1800s. Karl Pearson, among the founders of mathematical statistics and a colleague of Galton , have then mentioned, the method of least squares and regression were somewhat synonymous.

Regression analysis as we know it today is primarily the work of R.A. Fisher, one the most renowned statisticians of the 20th Century. Fisher combined the work of Gauss and Pearson to develop a fully realized theory of the properties of least squares estimation. Due to Fisher’s work, regression analysis is not just used for prediction and understanding correlations, but is also used for inference about the relationship between a factor and an outcome (sometimes inappropriately). Post Fisher, there have been a variety of important extensions of regression including logistic regression , nonparametric regression, Bayesian regression, and regression that incorporates regularization .

Accordingly, we can say that, despite the popular belief that statistics was originated from the game of chance etc., what else has to be most likely is; 'method of least square popped up at the beginning in 1800 and then, the regression methods were invented from it.' It therefore, was the regression methods which should have given the credit giving the birth to the subject statistics rather than anything else or elsewhere.

Further, regression is not limited to the above-mentioned context. It now has evolved itself as a gigantic subject. Many more advanced regression techniques have been invented and applied for different problem solving. To shed light on some other regression methods Flizmoser (2008); Darper and Smith (1998); and Data Science Central (2015) can be summarized as the followings:

Gauss and Legendre would be amazed at the ubiquity of least squares regression today. Regression analyses are frequently used by academics, policy analysts, journalists and even sports teams to predict the future and understand the past. Even with the development of increasingly sophisticated algorithms for prediction and inference, good old least squares regression is still perhaps the crown jewel of statistical analysis.

V. CONCLUSION

This paper has reviewed how statistics was invented. Discussion covers, who were those pioneers to talk the subject statistics and when and in what context, had they used it. This paper can be taken as an appetiser literature review to further advance and investigate about the history of statistics more precisely. Statistics is one of the indispensable discipline to human endeavour at present. It has a great history evolving, first from game of chance and later further advanced by the origin of least square. Later one day, method of least square turned out to be the method of regression, was the first concept that incorporated hypothesis testing for making inferences. And ultimately this is taken to be the origin of today’s modern statistics.

REFERENCES

[1] Allen M. P. (1997). The origins and uses of regression analysis. In Understanding regression analysis. Plenum Press, New York, and London. DOI: 10.1007/978-0-585-25657-3_1

[2] Finney, David J. (1996) A note on the history of regression Journal of Applied Statistics. Vol 23 (5) pp. 555 – 558 DOI: 10.1080/02664769624099

[3] History of statistics. (2016). Wikipedia the free encyclopaedia. Retrieved from https://en.wikipedia.org/wiki/History_of_statistics Accessed on 20.12.2016

[4] History of probability (2016). Wikipedia the free encyclopaedia. Retrieved from https://en.wikipedia.org/wiki/History_of_probability Accessed on 12.20.2016

[5] *History of Statistics, n.d). MATHZONE. Retrieved from https://www.emathzone.com/tutorials/basic-statistics/history-of-statistics.html Accessed on 01.18.2018*

[6] Historical Context. (2015). Statistical Analysis Handbook- (c) Dr M J de Smith, Retrieved from http://www.statsref.com/HTML/index.html?introduction.html Accessed on 12.21.2016

[7] Kopf, D. (2015). The Discovery of Statistical Regression. PRICEONOMICS. Retrived from https://priceonomics.com/the- discovery-of-statistical-regression/ Accessed on 12.25.2016

[8] Packett, R.L. (1972). Studies in the History of Probability and Statistics.XXIX: The Discovery of the Method of Least Squares. Biometrika, Vol. 59 (2). Pp. 239-251. DOI: 10.1093/biomet/59.2.239

[9] The Discovery of Statistical Regression (2015). PRICEONOMICS. Retrieved from file:///G:/FreshWorK/PAPERS_Aftr_PhD/7.%20ORIGIN%20OF%20REGRESSION%20METHODS/The%20Discovery%20of%20Statistical%20Regression.html Accessed on 12.21.2016

[10] Timeline of Probability and Statistics. (2016). Wikipedia the free encyclopaedia. Retrieved from https://en.wikipedia.org/wiki/Timeline_of_probability_and_statistics Accessed on 20.12.2016