International Journal of Formal Sciences: Current and Future Research Trends (IJFSCFRT)
ISSN (Print) , ISSN (Online)

https://ijfscfrtjournal.isrra.org/index.php/Formal_Sciences_Journal/index

Importance and Challenges of Statistics in Decision Making: It's Effect in Geographical Studies

Chinago Alexander B[a]*, Bright Ajoku C[b]

[a]*Captain Elechi Amadi Polytechnic, School of Environmental Sciences, Rumuola, Port Harcourt 500001, Nigeria*

[b]*Ignatius Ajuru University of Education, Department of Geography and Environmental Management, Port Harcourt 500001, Nigeria*

[a]*Email: chinago.alexander@portharcourtpoly.edu.ng*

Abstract

This paper deals on illustration and explanation of some statistical decision theories as applied in the field of climatology. Among the few reviewed in this work includes, Chi-square, Kruskal-Wallis test of several treatment, Regression analysis and the Seasonality index. The study also shows that there are diverse source of data collection in Geography and its related courses like Climatology. The work discovered among other things that statistical theories were useful in prediction and decision making in climatology and other geography related course. Despite different methods used it was observed that rainfall pattern in Port Harcourt has not change over time. The regression analysis using Microsoft Excel for charts shows that rainfall distribution has reduced over years without change in the pattern and characteristics. However, sources of data may affect decision and result same is true of interpretations.

Keywords: Rainfall; Statistical; Theories; Variability; Decision and Geographical Studies.

1. Introduction

Geography and its related courses are regarded as ideographic science from onset. At that point geography deals with mere description of what is where, how and why. Scholars at that point cannot think outside the box because the entire discipline is governed by the philosophy of determinism. So scholars then try to give the interpretations that will suit the philosophy or school of thought [1, 2]. In quest to improve, scholars turned their attention to cultural determinism with its two pronged theories of possibilism and probabilism.

---

* Corresponding author.

The former submitted that the environment does not simply impose itself on man and his activities rather it provides him with several options and possible choice ,while the letter held that all available choices , one is more advantageous and thereby more probable [3,4, 5] Possibilism and probabilism left geography and is related courses still in a situation where it was not able to explain anything fully - it could not discover any course and effect relationship,- It could not discover any absolute laws or predict the future with any certainty. For geographers to be able to deal with perceived similarities and differences between areas, such questions as what phenomena are observed? What patterns are discernable? What correlation? What alternative possibilities may appear in future and within what frame? It needs precise answers that only the precise language of mathematics may provide [6]. The aforementioned curiosity spurred scholars to yearn for alternative to the old ways of doing things this led to a shift in paradigm [7], defines a paradigm as '' some accepted examples of actual scientific practices which include laws, theories, applications and instrumentation which together provide a model of coherent tradition of scientific research ''. Put in another way, a paradigm is a model or an approach to scientific enquiry which is accepted by the majority of the people in a discipline [8], concludes that the concept of paradigm is useful to us because it expresses something rather important about explanation as a process and as an activity. The impact of quantification which is the language of mathematics began late in geography; however some scholars initiated numbers of statements calling for adoption of quantification. For example, [9] in his paper on corn yield and climate 'argued that' the method of correlation analysis would seem especially promising tools for geographical investigation [10]. Similar statements were made by [11, 12], in his attack on the davisian explanatory descriptive system geomorphology and his endorsement quantitative system of Gilbert. Geography underwent significant but radical transformation in its spirit, its nature and its purpose [13]. This radical change is usually referred to as "Quantitative Revolution". It is a revolution because certain dramatic and fundamental changes have been introduced into the methodology content and activity of geography. These combines have changed the discipline from a purely ideographic science that it was to a nomothetic science that it is today. Nomothetic because geography is relating to involving and dealing with abstract, general or universal statements or laws that govern the distribution of phenomena on the surface of the earth. And the surface of the earth to the geographer means - as deep into the Lithosphere, as high into the Atmosphere and as the deep into the Hydrosphere as man can go or get information from. For geography to successfully investigate the surface of the earth, it has to depend on science as a source of knowledge and use its methodology such as theory too. Climatology is a branch of geography just like Geomorphology, Biogeography etc. Climatology for example studies climate and how climate affects man and his environment. Climatology deals with analysis of accumulated data. A climatologist need data and uses statistics as a scientist needed specimen and uses laboratory. Theory is defined as the rules, principles, and assumptions that govern or shape a practice or something. The Oxford Advance Learner's Dictionary defines theory as a formal set of ideas that is intend to explain why something happens or exist. It is a principle on which a particular subject is based. English dictionary defines theory as a belief that can guide behaviour or a well - sustained explanation of some aspect of the natural world; an organized system of accepted knowledge that applies in a variety of circumstances to explain a specific set of phenomena. It could be seen as doctrine or scheme of thing which terminates in speculation or contemplation, without a view to practice; hypothesis; speculation. Statistics is defined as a branch of applied mathematics concerned with the collection and interpretation of quantitative data and the use of probability theory to estimate population parameters [14],stated that essentially statistics is a scientific

approach to information presenting itself in numerical form this enables us to maximize our understanding of such information. Statistical decision theories are statistical rules, principles that enable one to make decision using statistical postulation or inferences. Among the statistical decision theories include, Regression Analysis, Chi-square, Kruskal-Wallis test of several treatment, Friedman Test, Analysis of variance (ANOVA), Mann-Whitney Test and Wilcoxon Paired Comparison Test, Students 't' Test. So, decision theory in statistics is sets of quantitative methods of reaching optimal decisions in face of uncertainty

1.1. Sources of Geographical Data

Documentary sources of geographic data are information or statistics that are recorded as written materials. Those that produced it have tested validity. It is very useful to Climatologist, it is especially not very expensive. From the point of geography and particularly climatology there are about eight types of documentary sources of data. These include published studies, published statistics, unpublished statistics, maps, aerial photographs, directories, and historical sources, and lecturer's monographs. Published studies include the information derived from textbooks, journals, abstracts, magazines and periodicals. This actually enables scholars to know how best to handle research problems. Since these types of data are numerous, it required real effort on researchers to sort the best information that suits their research. The lot of information's posed a great challenge to researchers. So, care needs to be exercised, in selecting appropriate text, journal and the likes in order that valuable time, energy and other resources are not wasted in going through a welter of irrelevant materials. Secondly scholars may be misguided by junk information in some textbooks, journals and periodicals were quacks that want to make name publish for publishing sake to be recognized in their various institutions. Published statistics, this is divided into official and unofficial statistics. The information emanating from government sources, renowned organizations and ministries are termed official statistics e.g. information from federal government agencies like Central Bank of Nigeria, Bureau of Statistics and world organizations like World Bank, World Health Organization, etc. The unofficial statistics includes those published by private firms, individual research institutes and many technical and professional journals. Unpublished statistics: These are secondary data obtained from industries, government agencies that process internal data about their production, supply, manpower, cost etc., for example Ajaokuta Steel Company Limited (ASCL). Map is a cartographic representation of specifically chose spatial information. The information is transmitted through images constructed from symbols. There are different types of maps, however, all maps contains vital data for geographers. In fact map is the short hand for geographers. According to [15] stated that we tend to restrict the term *(map)* to visual maps, but spatial information may be represented on computer screen, through braille, or verbally through spoken description, and these categories of spatial representation may also be described as maps. The limitation of most maps used in climatology and indeed geography is that the information is not current. They are made several years ago, so most of the information contained is out-dated. Example several years after ceding Bakassi to Cameroun, both countries still use their old map both in school and research. The land mass of the countries will be affected by the new map, but what is on record?

Reading of a map required a special skill, which many lacks, so retrieving the true information from a map is a task not for learners or beginners. Besides producing of map is very expensive, so, many depend on out dated or old map with distorted and faulty information. Aerial photographs can be used for the same interpretative

purposes as maps and they can be used to update an existing map. Though very effective and accurate, aerial photograph is very expensive, it also required specialist interpretation. It is time consuming, stereoscopic viewing and matching of pictures is absolutely time consuming and tiredness may lead to making mistakes or misrepresentation of objects. For instance a grassland many be mistaken for a rice farm, etc. Directories and Historical sources are useful information or data that describe certain aspect of life in a town. For instance taxis and buses movement. Television programmes to be shown, type of number of houses, school and commercial activities. This data is vital for a geographer, but, how can a geographer get in contact with this information? They are not easily available, where they are, they are expensive. Most importantly, most of the information is not reliable and true. The publisher most often sit in the comfort of their houses and write what they felt is the reality on ground. In some cases the information given never existed. For historical data, most of the data were not accurate – this was due to oblate methods of measurement or limited knowledge due to not using the state of arts in science. Most of this information has been distorted, or part eaten by termites, destroyed by fire etc. In human geography for selfish reasons some scholars have tempered with the records to suit themselves and their community, especially in Hausa, Ibo and Yoruba. Census data is a typical; example in Nigeria front. From the foregoing it is obvious that documentary sources of geographic data are very important and relevant. Nevertheless it cannot cover for other sources of geographic data. For instance fieldwork is still necessary to ascertain what history and directories said are in existence. The knowledge of statistical errors, point to the fact that field and atmospheric sources of geographic data are vital. A detailed study of the surface is required to be able to select which or identify the right sources of relevant documentary data and appropriate areas of intensive field study; it is true for sub-terrain sources. A geographer determines what sources of data to use and how to use it. No one source is independent of others; there are relativity and interdependence on the various sources of data, especially the documentary data. The limitation of the various sources of documentary data arises from the fact that they depend on other sources for complete and effective result.

1.2. Statement of Problem

Climatology for example, like every other branch of geography is faced with the problem of identification and definition. A very well explained problem statement should be made with very clear research objectives. Then initial hypothesis for solving the problem will be formulated. In addition relevant information (data) is then collected. This step in planning involves the definition of measurement, population and sampling procedures bearing regards to the type of test. For example using climatology as a case study in the foregoing discussion, climatology deals with primary and secondary data so, the ability to check a suitable statistics is not easy. Both parametric and nonparametric statistics can be used for analysis, similarly both descriptive and inferential statistics are vital for robust decision making in climatology and other areas of geography. One of the greatest problems in climatology is the fact that most of the variables are qualitative. For example, vegetation distribution, climate classification. To mention the relationship between such characteristics, it would be necessary to assume that the topography is divided into unit areas for operationalization. Then the statistics of attribute base on count of individuals would apply. Problem like this, require climatologist who would use their intellect and knowledge to design and or use appropriate statistics theory to answer specific climatological questions and problems, because misapplication of theory will give a result that will rather confuse the general public. The aim of this work is to discuss some statistical decision theories and how they affect analysis in the

field of climatology. To achieve this set goal, the following specific objectives where pursued.

- To statistically test rainfall variability over Port Harcourt specific statistics theories
- To examine how statistical theory works in decision making.

The study explains some statistical decision theories in analysis in climatology. This will enable both professionals and scholars to know how to make prediction and analysis of climatologic problem using quantification methodology. The study cannot exhaust various statistical decision theories in this work. Sometimes, it is obvious that some of the decision reached could be by chance. A major problem in use of statistics in the field of climatology is misapplication of or misuse of statistics. A nonparametric statistics could be used in place of a parametric statistics and verse versa. Another challenge in the use of statistics in Geography and especially in climatology is the type and sources of data. There are diverse sources of data all over the globe. For instance we have about seven sources of documentary data.

- Published studies
- Published statistics
- Unpublished statistics
- Maps
- Aerial photographs
- Directories
- Historical sources.

Published studies include the information derived from Geography textbooks, Journals, abstracts, magazines and periodicals. The problem is that scholars may be misguided or misled by junk information in some predatory or substandard textbooks, journals etc. There is millions of information that get a researcher confused. Published statistics is divided into official and unofficial statistics. Information emanating from government sources, renowned organisations, institutions and ministries are termed official statistics. This includes Local government, State, and Federal government bulletin. Information from reputable organisation such as Federal Government Agencies, World Health Organisation, International Labour Organisation, Food and Agricultural Organisation, International Institute of Tropical Agriculture, etc. are also recognize as official statistics. The unofficial statistics include those published by private firms or institutions, individual researchers and many technical and professional journals. Examples include IOSR Journals, Global journals and IISTE journals. Unpublished statistics are secondary data obtained from industries, government agencies that process internal data about its production, supply, distributions, manpower, cost, etc. These include Nigerian National Petroleum Company, Nigeria Port Authority, Ajaokuta Steel Company Limited, etc. The constraint to these sources of data is that they are relatively inaccessible to all except those who are engaged in advance research

2. Methodology

The statistical decision theories to be used include Kruskal-Wallis test of several treatment, Chi-square, regression an analysis and seasonality index. Krustal-Wallis test for several treatments is an extension of the

method used in the two simple Mann-Whitney test, it allows several treatment to be compared. Since it is concerned with examining variability and treatment effect only, it is sometimes referred to as a one way analysis.

$$\chi^2 = \frac{12}{N(N+1)} \left(\frac{C^2}{n}\right) - 3(N+1) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where N = summation of the all observed outcomes or occurrences ($C_1 + C_2 + C_3 + \ldots C_C$)

n = number of occurrences in a column.

C = summation of the observed items in each column.

12, 3 and +1 are constant in the equation.

Chi-square (χ^2) is a simple technique which works by testing a distribution actually observed in the field against some other distribution decided by the researcher- for example, that object under the study are evenly spread over the landscape or over time

$$\chi^2 = \frac{\sum(0-E)^2}{E} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Where O = the frequently actually observed and E = the frequencies accepted

Regression analysis is simply a line of “best fit” on a scatter gram. It may be merely a summary of the relationship between two variables; or a means of highlighting individual derivation (called residuals) from this general relationship; or it may be used to interpolate predict unknown values of one variable from known values of the other. The trend line plotted through a time series is a special case of regression line, where one variable is time and the other is a set of data of different magnitude

$$Y = a + bx \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (3)$$

Seasonality index show the variation of variable within the season or month

$$SI = \frac{1}{R} \quad \sum \left(Xn - \frac{R}{12}\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots..\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

Where Xn is the mean rainfall of the month and R is the mean annual rainfall of the station.

3. Analysis and discussion

3.1. Chi square is used for analysis

Illustration of how statistical theories are used in the field of climatology is demonstrated using table one. Table one shows the rainfall distribution/occurrence in Port Harcourt from 1989-2012, twenty four year record to

establish pattern by establishing rainfall variability the 24 year record was divided into two groups of twelve years each. Rainfall data are mostly large so I divided the yearly rainfall by 100; same was done to the expected rainfall

The yearly rainfall were then observed and the expected are determinant of fluctuation

The H_o; theirs is no significant change in rainfall pattern over Port Harcourt throughout the study period

Hi; there is significant in rainfall pattern over the year in Port Harcourt

df=(r-1) (k-1) = (12-1 (2-1) =11

At 0.05 =95% degree of freedom χ^2critical value = 19.68 while χ^2calculated is 4.48. χ^2calculated value is 4.48. χ^2calculated < χcritical at 95% degree of freedom the null hypothesis (H_o) there is no significant change in rainfall pattern over port Harcourt is accepted and the alternative rejected reason theχ^2calculated < χ^2 critical at 95%

Table 1: Rainfall Distribution

X	E	$(0\text{-}E)^2$	$\frac{(0-E)^2}{E}$	Y	E	$(0\text{-}E)^2$	$\frac{(0-E)^2}{E}$
18.7	20.3	2.56	0.13	21.5	19.9	2.56	0.13
24.8	23.6	1.44	0.06	21.9	23.1	1.44	0.06
20.9	23.0	4.41	0.19	24.5	22.4	4.41	0.20
19.7	19.5	0.04	0.00	18.8	19.0	0.04	0.00
25.8	23.5	5.29	0.23	20.6	22.9	5.29	0.23
23.7	26.5	7.84	0.30	28.7	25.9	7.84	0.31
25.3	26.8	2.25	0.08	27.8	26.2	2.56	0.10
23.4	22.5	0.81	0.04	21.0	22.0	1.00	0.05
23.3	24.7	1.96	0.08	25.6	24.2	1.96	0.08
25.7	21.6	16.81	0.78	17.0	21.1	16.81	0.80
25.0	22.9	4.41	0.19	20.3	22.4	4.41	0.20
19.9	21.5	2.56	0.12	22.5	21.0	2.25	0.12
Total	2.2						2.28

2.2 + 2.28 = 4.48.

3.2. Kruskal-Wallis Test of Several Treatment is Used for analysis

Table 2: The Ranked Rainfall distribution

S/N	X	Y
1.	2	11
2.	17	11
3.	8	16
4.	4	3
5.	22	7
6.	15	24
7.	19	23
8.	14	9
9.	13	20
10.	21	1
11.	18	6
12.	5	12
	158	142

$\chi^2 = \frac{12}{N(N+1)} \left(\frac{C^2}{n}\right) - 3(N+1)$

Here Port Harcourt Rainfall is also grouped into two of 12years each.

The variables are ranked from the least to the top.

H_0: No difference in Rainfall pattern and characteristics over year.

H_1: There is difference in rainfall pattern and characteristics over time

N = 24

χ^2 is the sum of the ranks in each sample, n is the number of values in each sample.

N is the total number of values and K is the number of sample.

$\therefore \chi^2 = 12/4(24+1) \ (158^{2/}/12 + 1422/12) - 3 \ (24+1)$

$= (12/60) * (24964/12 + 20164/12) - 3(25)$

$= 0.02 \ (2080.3 + 1680.3) – 75$

$0.02 \times 3760.6 – 75$

$75.21 – 75$

$\chi^2 = 0.212$

df = C – 1 = 2 – 1 = 1

At 0.05 0r 95% level of confidence

χ^2critical = 3.84

χ^2 calculated = 0.21

Since χ^2calculated < χ^2critical at 95% the null hypothesis that there is no difference in rainfall pattern over the years in Port Harcourt is accepted.

If the calculate value exceeded the critical value the Ho would have been rejected.

3.3. Regression analysis

This statistical theory enables us to establish relationship among variables or to find trend over time.

Table 3 shows relationship between rainfall (y) and thunderstorm (x)

Table 3: Rainfall and Thunderstorm Distribution

X	Y	XY	X^2
2	60	120	4
5	100	500	25
4	70	280	16
6	90	540	36
3	80	240	9
$\Sigma x = 20$	$\Sigma xy = 400$	$\Sigma xy = 1680$	$\Sigma x^2 = 90$

Note $\Sigma y = an + b\Sigma x$

$\Sigma xy = 9\Sigma x + b\Sigma x^2$

∴ (i) 400 = ax5+ bx 20

(ii) 1680 = 5a+ b x 90

Multiply equation (1) by 4, = 1600 x 20a + 80b

Multiply equation by 1 = 1680 20a + 90b

80 = 10b

10b = 80

∴ Divide both side by 10 = b =8

If 8 is substitute for b in equation

(1) Then a can be found.

400 = 5a + 8 x 20

5a = 160 – 400 = 240

5a = 240

a = 240/5 = 48

a = 48

The solution can enable us to fix the line of best fitting. That will show trend over years.

Note – Regression line (Trend) are now done in excel.

Pearson Product Moment Correlation can also be used for analysis, however to make decisions, validity test will be made using student "test" or coefficient of variation.

3.4. Seasonality index

Seasonality index helps in identifying the rainfall regime based on the monthly distribution of rainfall. It is very important in agriculture and in road construction.

Table 4: shows rainfall distribution over Port Harcourt for just one year (2013) Using SI

Months	**Rainfall**	**Xn – R/12**	**∑(Xn – R/12) /R**
January	40.6	24.91	0.13
February	37.0	21.31	0.11
March	122.1	106.41	0.57
April	210.5	194.81	1.03
May	260.1	244.41	1.30
June	461.7	446.01	2.37
July	274.4	258.31	1.37
August	112.9	97.01	0.52
September	178.8	163.11	0.87
October	304.8	289.11	1.54
November	144.0	128.31	0.68
December	112.7	97.01	0.52
Total	2259.6	1970.72	11.01
Mean	188.3	10.47	0.9175

The seasonality index classified the types of climate in relationship to water availability. The higher the seasonality index of a region the greater the water resources variability and scarcity in time, the more vulnerable the area to desertification. But in terms of monthly rainfall, it shows variation of rainfall distribution over years. In this study, however, the higher the variability, the more rainfall expected within the month. Reason, we are looking at rainfall within just a year (2013). The result is rainfall from zero distribution to the highest amount of rainfall. The lesser the rainfall, the closer it is to zero, and the higher the rainfall the farther it is from zero.

4. Conclusion

The work precisely shows how quantification/statistical decision theories can be used in climatology. The rainfall variability shows that the pattern has not changed over time, there could be the natural ups and downs in distribution know as fluctuation. Chi-square and Kruskal-Wallis test for several treatments had similar result for the same problem. The null hypothesis was accepted because the χ^2calculated did not exceed the χ^2critical value. Whenever χ^2 critical > χ^2 calculated H_0 is accepted at the level of confidence or degree of freedom. The regression analysis provide line through which trend will be achieved. The line of best fit shows to what extent the variables deviate from the normal. The seasonality index is a critical factor affecting the evolution of natural vegetation. The robustness of quantification gives scholar's confidence in his speculation and predictions. The reliability and dependency of the result is not based on the statistical method use, but on the reliability and sources of data. Therefore, data sources must be validated before use in climatology to avert wrong decision and disaster.

References

[1]. P.E. James. All Possible Worlds – A History of Geographical ideas, New York: The Odyssey Press, 1972.

[2]. W.L. Garrison. "Applicability of Statistical Inference to Geographical Research," Geographical Review: 46:427-429, 1956.

[3]. U. Fekadu.. The paradox in environmental determinism and possiblism: A literature review. Journal of Geography and Regional Planning, 7(7), 132-139, 2014.

[4]. L. Schnellert, D. Kazak, and S. Moore. "Professional development that position teachers as inquirers and possibilizers". LEARNing landscapes 9(1), 217-236, 2015.

[5]. L. Schnellert, P. Richardson and S. Cherkowski. "Teachers educator professional development as reflexive inquiry", LEARNing Landscapes, 8(1), 233-250, 2015.

[6]. J.D. Wilhelm.." Engaging readers and write with inquiry". New York: Scholastic, 2007.

[7]. T.S. Kuhn. "The Structure of Scientific Revolution". Chicago Publications in Geography 2, 1962.

[8]. D. Harvey. "Explanation in Geography", London Edward Arnold Ltd. 1971.

[9]. J.K. Rose. "Corn yield and climate in the corn Belt" Geographical Review 26: 88-102, 1936.

[10]. J.M. Rose. "An Analysis of Isopleth and Choropleth Class Intervals" Econ. Geography Vol. 31, 1955.

[11]. A.N.Strahler. "Davis' Concepts of Slope Development viewed in the light of Recent Quantitative Investigation" Annals A.A.G., 40, 1950.

[12]. L.C. Moll " L.S. Vygotsky and education". New York: Routledge, 2014.

[13]. Burton . “The Quantitative Revolution and Theoretical Geography” The Cndin Geographer, 7, 1963.

[14]. W.M. Harper. “Statistics” Published by M&E Handbook. Richard Clay. 1977

[15]. C. Mayhew. “ Dictionary of Geography” University of Oxford press London. 2004

[16]. D. Blurmenstock . “The Reliability factor in the drawing of Isarithm” Annals AAG, Vol. 43, 1953.

[17]. G.K.. Gilberty. “The Transportation of Debris by Running Water” US Geological Survey Professional Raper: No.86 (Washington: USGPO), 1914

[18]. F. MacCannel . “Problems, Techniques and Theory of Isarithm Mapping” Journal of American Statistical Association Vol. 50, 1955.