USMANU DANFODIYO UNIVERSITY, SOKOTO

(POSTGRADUATE SCHOOL)

COMPARISON OF SPURIOUS CORRELATION METHODS USING PROBABILITY DISTRIBUTOINS AND PROPORTION OF REJECTING A TRUE NULL HYPOTHESIS

A Dissertation

Submitted to the

Postgraduate School

USMANU DANFODIYO UNIVERSITY, SOKOTO, NIGERIA

In Partial Fulfillment of the Requirements

For the Award of the Degree of

MASTER OF SCIENCE (STATISTICS)

BY

ALFA, Mohammed Alhaji

ADM. NO: 112103061166

DEPARTMENT OF MATHEMATICS

JUNE, 2015

DEDICATION

I would like to dedicate my research to my beloved parents Alhaji Muhammad Alfa (Late) and Hajiya Aishatu Alfa, my beloved wife Aishatu Mohammed Rokota and my children Muhammad and Ahamadu.

CERTIFICATION

This dissertation by Alfa, Mohammed Alhaji (112103061166) has met the requirements for the award of the Degree of Master of Science (Statistics) of the Usmanu Danfodiyo University, Sokoto, and is approved for its contribution to knowledge.

___________________ _________________

External Examiner Date

___________________ ________________

Prof. S. U. Gulumbe Date

Major supervisor

___________________ _________________

Dr. A. Danbaba Date

Co-supervisor I

___________________ _________________

Dr. H. A. Gwandu Date

Co-supervisor II

___________________ _________________

Prof. I. J. Uwanta

Head of Department Date

ACKNOWLEDGEMENTS

In the name of Allah, Most Gracious, Most Merciful. First of all, I am very thankful to Almighty Allah, Who has empowered me with His guidance throughout my life to achieve my missions in life.

I would like to offer my gratitude to my Major supervisor, Prof. S.U Gulumbe and my co-supervisors, Dr. A. Danbaba and Dr. H.A. Gwandu for their patience, inestimable effort and valuable suggestions that enabled me to work efficiently. Without their kind supervision, I would never have had the chance to complete my research work on time. May Allah bless and grants them a long life, healthy and Aljanatul fridausi, amin .

My special gratitude goes to the Head of Department, Professor. I.J. Uwanta, PG Coordinator Dr Yakubu Musa and all others lecturers of the Department of Mathematics with their advice and guidance, especially Dr Umar Usman, Ahmed Audu and Shamsuddeen Suleiman. May Allah reward you all, amin.

My profound gratitude goes to Allaji Usman Liman (Statistician General) of Niger State, Commissioner, Permanent Secretary, in particular, all the staff of Niger State Planning Commission and Niger State Bureau of Statistics, Minna. May Allah assist you all, amin.

My deeply profound gratitude goes to my beloved parents Alhaji Muhammad Alfa Rokota (Late) and Hajiya Aishatu Alfa Rokota include Alh M.K Abdulkadir, Alh Idiris A. Alfa, Moh’d A. Alfa, Hajiya Mama Ahmed, Hajiya Aminatu Abubakar, Hajiya Fatimatu Alhassan and others for their perseverance, encouragement and constant prayers for me while undergoing my studies. May Allah bless and grants them with Aljanatul fridausi, amin.

Finally, I sincerely grateful to my beloved wife Aishatu Mohammed Rokota and my children Mohammad and Ahamadu for their prayers and patience during my stay in school. May Allah bless and grants them with long life, healthy life and aljanatul fridausi, amin.

TABLE OF CONTENTS

CHAPTER FOUR

CHAPTER FIVE

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

PMCT	Pearson's Product-moment Correlation Test,
SRCT	Spearman's Rank Correlation rho Test
KRCT	Kendall's Rank Correlation tau Test

ABSRACT

The problem of spurious correlation analysis, e.g. Pearson moment-product correlation test is that, the data need to be normally distributed. This research work compares spurious correlation methods using some non- normal probability distributions in order to obtain the method with the best degree of association among them. The methods were compared using proportions of rejecting true null hypothesis obtained from t and z test statistics for testing correlation coefficients. Data from Normal, log-normal, exponential and contaminated normal distributions were generated using simulation method with different sample sizes. The results indicate that, when the data are normal, exponential and contaminated normal random distributions, Pearson's and Spearman's rank have the best proportion of rejecting the true null hypothesis. But, when the data are log-normal distribution, only Spearman's rank correlation coefficient has the best proportion of rejecting the true null hypothesis. Thus, Pearson's and Spearman's rank have the best degree of association under normal, exponential and contaminated normal distributions. While, for log-normal distribution only Spearman's rank has the best degree of association.

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background to the Study

The awareness of problems related to the statistical analysis on spurious correlation began as early as 1897 by Karl Pearson in his seminar paper on spurious correlations, which title began significantly with the words "On a form of spurious correlation" and then repeatedly by a geologist Chayes (1960).

The main source of information about the history of spurious correlation test is that, Pearson used the term spurious correlation to "distinguish the correlations of scientific importance from those that were not." The problem, according to Pearson, was that some correlations did not indicate an "organic relationship." Although this term is never defined, the examples used suggest that spurious correlation was the same as a correlation between two variables that were not causally connected and the term correlation coefficient only measures the strength of linear relationships (Johnson and Kotz 1992). The simplicity and interpretability should be the main ideas when selecting measures of association. Historically, the Pearson correlation has been the main association measure in multivariate analysis. It is simple, as it relates only two variables of a random vector; it concerns only linear transformation in R^n, i.e. change of scale plus a shift. Interpretation relies on the linear regression ideas, which in turn are related to the geometry of R^n, where covariance appears as a Euclidean inner product in the space of samples (Lovell *et al*, 2013). All these desirable properties will be achieved when Pearson correlation is applied to study association. Correlations between variables can be measured with the use of different indices (coefficients). The three most popular are: Pearson's coefficient r, Spearman's rho coefficient ρ and Kendall's tau coefficient τ . Although coming back to the history of developing the idea of measuring correlation strength, Aitchison (1986), created the basis for a proper and correct application and interpretation of correlations (in the modern meaning of the word). Thus, the history was presented by Rodgers and Nicewander (1988). Pearson's coefficient of correlation was discovered by Bravais in 1846, but Karl Pearson was the first to describe it in 1896, by showing the standard method that is, the formula for its calculation, application and interpretation of correlations coefficient. Pearson also offered some comments about an extension of the idea made by Galton (1879), who applied it to anthropometric data. He called this method the "product-moments'' method or the Galton function for the coefficient of correlation r .

In 1904 Spearman adopted Pearson's correlation coefficient as a measure of the strength of the relationship between two variables that cannot be measured quantitatively. He noted: "The most fundamental requisite is to be able to measure our observed correspondence by a plain numerical symbol. There is no reason whatever to be satisfied either with vague generalities such as "large", "medium", "small," or, on the other hand, with complicated tables and compilations. Kendall's tau, introduced by Kendall (1938), which can be used as an alternative to Spearman's rho for data in the form of ranks. It is a simple function of the minimum number of neighbour swaps needed to produce one ordering from another. Its properties were also analyzed by Kendall in his book concerning rank correlation methods, first published in 1948. The main advantages of using Kendall's tau are the fact that its distribution has slightly better statistical properties, and that there is a direct interpretation of this statistics in terms of probabilities of observing concordant and discordant pairs. Nonetheless, coefficient τ has not been used so often in the past (the last sixty years) as Spearman's coefficient in measuring rank correlation, mainly because it was the one more difficult to compute.

Nowadays the calculation of Kendall's τ poses no problem. Kendall's τ is equivalent to Spearman's r**s** in terms of the underlying assumptions, but they are not identical in magnitude, since their underlying logic and computational formulae are quite different. The relationship between the two measures for large numbers of pairs is given by Daniels (1944) as $-1 \leq 3\tau - 2r_s \leq 1$. Properties and comparisons of Kendall's τ and Spearman's r**s** have been analyzed by many researchers and they are still under investigation (Valz and Thompson 1994, Weichao *et al.* 2010). Hence the association between two variables is often of interest in data analysis and methodological research. Pearson's, Spearman's and Kendall's correlation coefficients are the most commonly used measures of monotone association, with the latter two usually suggested for non-normally distributed data. These three correlation coefficients can be represented as the differently weighted averages of the same concordance indicators. The weighting used in the Pearson's correlation coefficient could be preferable for reflecting monotone association in some types of continuous and not necessarily bivariate normal data. (Nian,2010). The Pearson correlation coefficient *r* measuring a linear relationship between the variables is one of the most frequently used tools in statistics (Rodgers and Nicewander, 1988). Generally, correlation indicates how well two normally distributed variables move together in a linear way (Aczel, 1998). When the assumption about the normal distributions of the variables considered is not valid or the data are in the form of ranks, we use other measures of the degree of association between two variables, namely the Spearman rank correlation coefficient rs (Aczel, 1998) or the Kendall τ correlation coefficient. In addition, since the normality assumption of data usually does not provide an adequate approximation to data sets with heavy tail, non-normal distributions are used in practice (Johnson and Kotz, 1992; Kotz, *et.al*, 2000).

1.2 Statement of the Problems

The Spurious correlations coefficient tests (Pearson, Spearman and Kendall) are the major methods used in measuring the degree of association. These methods have been adopted by many researchers' like Nian (2010) and Jan and Tomasz (2011) that used normal data and they found out that, Pearson's product moment correlation coefficient has the best degree of association among the other methods. But, comparisons of these methods using different distributions have received no or little attention. Hence, the present research will study these methods under normal and some non normal distributions.

1.3 Aim and Objectives

The aim of this research work is to compare the different "spurious correlation tests" using data of poverty levels and simulated data sets. The objectives are to:

i. Assess the proportion of rejecting true null hypothesis under the Pearson's product-moment correlation test, Spearman's rank correlation rho test and Kendall's rank correlation tau test.

ii. Identify the method with the best degree of association under normal, log-normal, exponential and contaminated normal random distributions.

1.4 Significance of Study

The realization of the problem in the issue of "spurious correlation", to measure the dependence between the two variables began as early as 1897, these methods have been adopted by many researchers, but mostly of them used Primary and Secondary data, without investigate the nature of data that is, which probability distribution does the data follows. It is our hope that at end of this study, we will come up with some best methods that will be used on some probability distributions when need arise.

1.5 Scope and Limitations

There are many different types of spurious correlation coefficients that reflect somewhat different aspects of association and are interpreted differently in statistical analysis. In this study, focus will be on three popular methods that are often provided next ideal to each other, namely the Pearson's, Spearman's and Kendall's correlation. Only normal, log-normal, exponential and contaminated normal random distributions will be used for study. In addition, proportion of rejecting true null hypothesis (type one error rate) is used in this research work.

1.6 Definition of Terms

Poverty is multifaceted and has no single universally accepted definition. The World Bank defined poverty as a pronounced deprivation of human wellbeing; which includes vulnerability to adverse events outside their control, being badly treated by the institutions of state and society and being excluded from having a voice and power. Any household or individual with insufficient income or expenditure to acquire the basic necessities of life is considered to be poor. Most countries of the world fall under the absolute poverty line, which indicates that they live on less than one U.S Dollar per day. Those that are moderate or relatively poor live on more than one US Dollar but less than two Dollars per day.

NBS (2010).

Food Poverty –is an aspect of absolute Poverty Measure which considers only food expenditure for the affected Household.

Relative Poverty- is defined by reference to the living standards of the majority of people in a given society.

Absolute Poverty- is defined in terms of the minimal requirements for food, clothing, healthcare and shelter.

Dollar Per-Day- Measure of poverty refers to the proportion of people living on less than US$1 per day poverty line based on World Bank's Purchasing Power Parity (PPP) index NSBS (2012).

CHAPTER TWO

2.0 LTERATURE REVIEW

2.1 Introduction

The term spurious correlation coefficient tests (e.g Pearson, Spearman and Kendall) are the major methods used in measuring the degree of association between the two random variables. Nian (2010), compared Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data using type one error rate on the same sets of data, and found that Pearson's product moment correlation coefficient has the best degree of association among the other methods. Jan and Tomasz (2011), compared the values of Pearson's and Spearman's correlation coefficients on the same sets of data, they stated that, when analyzing both Pearson's and Spearman's coefficients, one could logically expect that the significance of one would imply the significance of the other. And concluded that Pearson's product moment correlation coefficient has a better significant measure of the strength of the associations between two variables. Sorana and Lorentz (2006), studied Pearson versus Spearman rho, Kendall's Tau Correlation Analysis on Structure-Activity Relationships of Biologic Active Compounds; these methods were used to evaluate the correlation between measured and estimated by Molecular Descriptors Family on Structure-Activity Relationships (MDF-SAR) model. The result shows that inhibitory activities are statistically significant. Gregory and Roger (2007), have studied the relationship between Spearman's and Kendall's for pairs of continuous random variables. And found that sufficient conditions for determining the direction of the inequality between three times tau and twice rho when the underlying joint distribution is absolutely continuous is $-1 \leq 3\tau - 2r\mathbf{s} \leq 1$ as it has been verified by Daniels (1944).

Eulalia and Janusz (2011), applied Spearman and Kendall rank correlation coefficients between Atanassov's intuitionistic fuzzy sets (A-IFSs,) to measure the degree of association between the A-IFSs when the assumption that the data distributions are normal is not valid that is, the data are in the form of ranks Spearman and Kendall rank correlation coefficients should be used to test the degree of association between the two variables.

Hans and Melberg (2000), investigate on the nature and extent of spurious correlation and its implication for the philosophy of science with special emphasis and stated that, correlation coefficient only measures the strength of linear relationships. Their further research concerned with the more general topic of regular conjunctions of all types – linear or non-linear, which lead them to adopt more general measures of association capable of capturing non-linear associations.

David and Christopher *et.al* (2005), said that an inherent problem in measuring the influence of expert reviews on the demand for experience goods is that a correlation between good reviews and high demand may be spurious, induced by an underlying correlation with unobservable quality signals. Using the timing of the reviews by two popular movie critics, Siskel and Ebert, relative to opening weekend box office revenue, they applied a difference-in-differences approach to circumvent the problem of spurious correlation. After purging the spurious correlation, the measured influence effect is smaller though still detectable. Positive reviews have a particularly large influence on the demand for dramas and narrowly-released movies.

Aldrich (1995), examined spurious Correlations coefficient in Pearson and Yule correlation coefficient and only considers the development of ideals on both genuine and spurious correlation he makes some references to related modern work as follow 'Scientific inference' deals with inference from sample population while ' Statistical inference deals with the interpretation of the population in terms of a theoretical structure.

Theodoros *at el*. 2011 studied the Removal of Spurious Correlations between Spikes and Local Field Potentials they stated that, the existence and magnitude of spurious correlations in such cases will depend on the nature of the data and the types of analyses applied. Particularly when higher frequencies of the Local Field Potentials (LFP) are analyzed with sophisticated nonlinear methods, the effect of even modest contaminations on the results can be significantly magnified.

2.2 Theory of Spurious Correlation Coefficients

The Pearson product moment correlation is the most frequently used coefficient for normal distributed data. On the other hand, nonparametric methods such as Spearman's rank-order and Kendall's tau correlation coefficients are usually suggested for non-normal data. There are numerous guidelines on when to use each of these correlation coefficients. One guideline is based on the type of the data being analyzed. According to Khamis, (2008) Pearson product moment correlation coefficient is appropriate only for interval data while the Spearman's and Kendall's correlation coefficients could be used for either ordinal or interval data. However, all three methods of correlation coefficients could be viewed as weighted averages of concordance indicators, and as proposed by (Snedecor and Cochran (1989). Pearson's type of weighting could be conceptually preferable for continuous, but not necessarily normally distributed data.

2.2.1 The Pearson Correlation Coefficient

The Pearson correlation coefficient (Pearson's r) measures the extent of degree of a linear relationship in the actual value of two variables. The correlation coefficient is 1 in the case of a perfect positive (increasing) linear relationship, -1 in the case of a negative (decreasing) linear relationship, and some value between -1 and 1 in all other cases. The closer the coefficient is to either -1 or 1, the stronger the correlation between the variables (Aczel, 1998).

2.2.2 The Spearman Rank Correlation Coefficients

The Spearman rank correlation coefficient measures the extent of degree of linear relationship in the position or rank of two variables. It is also measures the dependence between two variables by ranks all the observations of the first variable, and independently ranks the values of the second variable. The Spearman rank correlation coefficient is like the Pearson correlation coefficient but different by applying the ranks to the variables (Nian, 2010).

2.2.3 Kendall Rank Correlation Coefficients

Similar to the two previous correlation coefficients, Kendall's tau ranges from -1 to +1, with the absolute value of τ indicating the strength of the monotonic relationship between the two variables. Thus, Spearman's rho (ρ) and Kendall's tau (τ) are the two most commonly used nonparametric measures of association between two random variables (Nelsen, 1992).

CHAPTER THREE

3.0 MATERIALS AND METHODOLOGY

3.1 Data Used for the Study

The data used in this study were two sets, the real live data and simulated data. The real live data were obtained from Nigeria poverty profile 2010 report by National Bureau of Statistics. The data composed of poverty levels for the all thirty six (36) States including F.C.T Nigeria in terms of food poverty, absolute poverty, relative poverty and dollar per day poverty line based on World Bank's Purchasing Power Parity (PPP) index.

This poverty profile has four mutually exclusive and exhaustive groups as follows:

Food poverty – including Food, Alcoholic Beverages and Non Alcoholic Beverages.

Absolute poverty – including food, clothing, healthcare and shelter.

Relative poverty – including Housing, Water, Electricity, Gas and fuel.

Dollar per day – including people living on less than US$1 and US$2.

Simulated data were generating from Normal, Log-Normal, and Exponential and contaminated random distributions.

The results of live data are listed as Appendix A, the simulated data are listed as Appendix B, and the collected live data from National Bureau of Statistics (NBS) were listed as Appendix C.

3.2 Simulation Study

The simulated data used for the comparison of different spurious correlation methods were generate as follows:

1. Generate data sets x and y using probability distribution.

2. Obtain the correlation of x and y

3. Refer to t under $H_0 : \rho$ =0 for single coefficient test or $H_0 : \rho_1 = \rho_2$ for two coefficient dsa4tests.

4. Reject H_0 if $\alpha < 0.05$ otherwise accepts H_1 at the 5% level of significance.

5. Repeat 4 to 5, 1000 times.

6. Count the number of rejection and divide by 1000 this gives proportion of rejecting true null hypothesis.

7. Different sample sizes were selected n=5, 10, 15 and 20 for test single correlation coefficient and for two correlation coefficients with mean $\mu_1 = 1, \mu_2 = 2$ and $\mu_1 = 2$, $\mu_2 = 1$ as well as variance $\sigma_1^2 = 2, \sigma_2^2 = 3$ and $\sigma_1{}^2 = 4$, $\sigma_2{}^2 = 2$

3.3 Probability Distribution Used for Simulation

Normal, log-normal, exponential and contaminated random distributions were used as probability functions for simulation of the data to be used for the analysis.

3.3.1 Normal Distribution

The normal distribution with parameters mean $\mu_1 = 1, \mu_2 = 2$ and $\mu_1 = 2$, $\mu_2 = 1$ as well as variance $\sigma_1^2 = 2, \sigma_2^2 = 3$ and $\sigma_1{}^2 = 4$, $\sigma_2{}^2 = 2$ was used for data simulation.

3.3.2 Log-normal Distribution

A random variable X is said to have lognormal distribution with parameters $\mu \in R$ and σ > 0 if $\ln(X)$ has the normal distribution with μ and standard deviation σ. Equivalently, X = e^Y where Y is normally distribution with mean μ and standard deviation σ.

The lognormal distribution is used to model continuous random quantitative when the data is believed to have heavy tail.

3.3.3 Exponential Distribution

We said that the random variables $X \sim \text{Exp}(\lambda)$.

Exponential distribution is used to represent light tailed and generate data that is, negatively skewed.

3.3.4 Contaminated normal Distribution

We said that the random variables $X \sim \alpha$ x $rnorm\left(\mu,\sigma^2\right)+\left(1-\alpha\right)$x $rnorm\left(\mu,\sigma^2\right)$

Contaminated normal distribution is used as a model for population when outliers occur.

3.4 Software Used

The statistical package used for the analysis was R package version 3.0.3 (2014-03-06). The functions *cor*(x,y), $rnorm(n,\mu,\sigma^2)$, $rlnorm(n,\mu,\sigma^2)$, $rexp(n,\frac{1}{\mu})$ and α x $rnorm\left(n,\mu,\sigma^2\right)+\left(1-\alpha\right)$x $rnorm\left(n,\mu,\sigma^2\right)$ are used for the study.

3.5 Level of Significant

α = 0.01 to 0.1. Were used for each of the figure while, only the result of α = 0.05 are displayed in the tables.

3.6 The Pearson Correlation Coefficient

The Pearson's correlation coefficient is a common measure of association between two continuous variables. It is defined as the ratio of the covariance of the two variables to the product of their respective standard deviations, commonly denoted by the Greek letter ρ (rho):

$$\rho = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$$

The sample correlation coefficient, *r*, can be obtaining by plugging-in the sample covariance and the sample standard deviations into the previous formula, i.e.:

$$r = \frac{\text{cov}\left(xy\right)}{\sqrt{\left(\text{var}\,x\right)}\sqrt{\left(\text{var}\,y\right)}}$$

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \sum_{i=1}^{n} \left(y_i - \bar{y}_i\right)^2}} \qquad 3.2$$

where:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n},\ \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$ (Aczel, 1998).

Where: $\bar{x}$ and $\bar{y}$ are the means of the *x* and *y* values.

3.7 The Spearman Rank Correlation Coefficients

Spearman's rank-order correlation coefficient (denoted ρ_s) is a rank-based version of the Pearson's correlation coefficient. Its estimate or sample correlation coefficient (denoted r_s), can be written as follows:

$$r_k = 1 - \frac{6\sum d^2}{n\left(n^2 - 1\right)} \text{.....3.3}$$

where $d = \left(rank\left(x\right) - rank\left(y\right)\right)^2$

Spearman's correlation coefficient varies from -1 to +1 and the absolute value of ρ_s describes the strength of the monotonic relationship. The closer the absolute value of ρ_s to 0, the weaker is the monotonic relationship between the two variables. However, similar to the Pearson product moment correlation coefficient, Spearman's correlation coefficient can be 0 for variables that are related in a non-monotonic manner. At the same time, unlike the Pearson's correlation coefficient, Spearman's coefficient can be 1 not only for linearly related variables, but also for the variables that are related according to some type of non-linear but monotonic relationship. (Nian, 2010).

3.8 Kendall Rank Correlation Coefficients

Similar to Spearman's rank-order correlation coefficient, Kendall's tau correlation coefficient is designed to capture the association between two ordinal (not necessarily interval) variables. Its estimate (denoted τ) can be expressed as follows:

$$\tau = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \left(x_i - x_j\right)\left(y_i - y_j\right)}{n\left(n-1\right)} \qquad 3.4$$

$$\text{Where: } \left(x_i - x_j\right) = \left(\begin{array}{lll} 1 & if \left(x_i - x_j\right) > 0 & \\ 0 & if \left(x_i - x_j\right) = 0 & ;\left(y_i - y_j\right) \\ -1 & if \left(x_i - x_j\right) < 0 & \end{array} \begin{array}{ll} 1 & if \left(y_i - y_j\right) > 0 \\ 0 & if \left(y_i - y_j\right) = 0 \\ -1 & if \left(y_i - y_j\right) < 0 \end{array} \right)$$

This coefficient quantifies the discrepancy between the number of concordant and discordant pairs. Any two pairs of ranks $\left(x_i, y_i\right)$ and $\left(x_j, y_j\right)$ are said to be concordant when $x_i < x_j$ and $y_i < y_j$, or when $x_i > x_j$ and $y_i > y_j$, or when $\left(x_i - y_i\right)$ and $\left(x_j - y_j\right)$>0. Correspondingly, any two pairs of ranks $\left(x_i, y_i\right)$and $\left(x_j, y_j\right)$ are said to be discordant when $x_i < x_j$ and $y_i < y_j$, or when $x_i > x_j$ and $y_i > y_j$, or when $\left(x_i - y_i\right)$ and $\left(x_j - y_j\right)$<0 (Nelsen, 1992).

For example if x and y are random variables with marginal distribution functions F and G, respectively, then Spearman's ρ is the ordinary (Pearson) correlation coefficient of the transformed random variables $F(x)$ and $G(y)$, while Kendall's τ is the difference between the probability of concordance $P[(x_1 - x_2)(y_1 - y_2)>0]$ and the probability of discordance $P[((x_1 - x_2)(y_1 - y_2)<0]$ for two independent pairs (x_1, y_1) and (x_2, y_2) of observations drawn from the distribution. In terms of dependence properties, Spearman's ρ is a measure of average quadrant dependence, while Kendall's τ is a measure of average likelihood ratio dependence (Nelsen, 1992, 2001).

3.9 Testing a Single Correlation Coefficient

The significance of a sample correlation, *r*, depends on the sample size and it can be tested with a t-test (Snedecor and Cochran 1989; Haan 2002).

If these assumptions are satisfied, the statistic ~ $t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}} \sim t_{n-2}$ 3.5

Follows a t-distribution with *n*-2 degrees of freedom, where *N* is the sample size. The null and alternative hypotheses for a test (two-sided) are:

$H_0 : \rho = 0$ i.e. correlation coefficient is zero

$H_1 : \rho \neq 0$ i.e. correlation coefficient is "greater than" zero

Or

H_0: A significant relationship does not exist between the two random variables.

H_1: A significant relationship exists between the two random variables.

To apply the test, the steps are:

- Compute the statistic *t*
- Decide on an α-level (e.g., for a 95% confidence interval)
- Compare the computed *t* with the $1-\alpha/2$ probability point of the cdf of Student's t-distribution
- If the absolute value of *t* is greater than the probability point from (3) point, reject H_0 otherwise accept H_1

Decision Rule:

Reject H_0 if $\alpha < 0.05$ otherwise accept H_1 at the 5% level of significance.

Or if absolute value of the calculated t-value is greater than or equal to the critical t-value, reject the H_0

3.10 Testing Two Correlation Coefficients

The null and alternative hypotheses for a test (two-sided) are:

$$H_0 : \rho_1 = \rho_2$$

$$H_1 : \rho_1 \neq \rho_2$$

The test statistic

$$Z \sim N\left(\mu, \sigma^2\right)$$

where

$$\sigma = \sqrt{Z_1 + Z_2}$$

$$Z_1 = 1.1513 \log_{10}\left(\frac{1+r_1}{1-r_1}\right)$$

$$Z_2 = 1.1513 \log_{10}\left(\frac{1+r_2}{1-r_2}\right)$$

$$\sigma_{Z_1} = \frac{1}{\sqrt{n_1 - 3}}$$

$$\sigma_{Z_2} = \frac{1}{\sqrt{n_2 - 3}}$$

$$Z = \frac{Z_1 - Z_2}{\sigma} \sim N(\mu, \sigma^2) .....3.6 \quad \text{Haan (2002)}$$

Where μ and σ^2 denote the mean and the variance of the distribution respectively.

Thus, normal curve is completely dependent on these two parameters, with the following properties.

- The mode is a point on the horizontal axis where the curve attains a maximum values. This point is also the same as the mean of the distribution
- The curve is symmetric about a vertical axis through the mean.
- The normal curve approaches the horizontal axis asymptotically as we proceed in direction
- The total area under the curve and above the horizontal axis is equal to one.

Decision Rule:

Reject H_0 if $\alpha < 0.05$ otherwise accept H_0 at the 5% level of significance. Kotz (2000)

Or if absolute value of the calculated Z-value is greater than or equal to the critical Z-value, reject the H_0

3.11 Criteria for Identifying Best Proportion of Rejecting True Null Hypothesis

Bradley (1978) identified three different type 1 error rates of robusteness which he termed fairly stringent, moderate and very liberal. The fairly stringent criterion is the situation when empirical type I error rates lies in (0.009 0.011), (0.045 0.055) for α =0.01 and α =0.05 respectively. Moderate criterion is the situation when empirical type I error rates lies in (0.008 0.012), (0.040 0.064) for α =0.01 and α =0.05 respectively. The very liberal criterion is the situation when empirical type I error rates lies in (0.005 0.015), (0.025 0.075) for α =0.01 and α =0.05 respectively.

In this study Moderate criterion will be used and only results for α =0.05 levels of significant are reported, in addition bolded and bracket values in these tables are less than the lower limit and greater than the upper limit of the confidence interval (0.040 0.064).

CHAPTER FOUR

4.0 RESULTS AND DISCUSSION

4.1 Introduction

This chapter deals with the Presentation, Analysis and Discussion of Findings on the different "spurious correlation tests. The chapter is divided in to four sections: Section one is the introduction, section two deals with the spurious correlation test for poverty levels in Nigeria, section three focuses on the comparison of correlation coefficient tests using different random distributions while, section four deals with the Discussion of findings.

4.2 Spurious Correlation Test for Poverty Levels in Nigeria

The results of the Spurious Correlation Test for Poverty Levels in Nigeria are presented in Table 4.1 below.

Table 4.1: Spurious Correlation Test for Poverty Levels in Nigeria

Correlations	Sample size	correlation coefficient	Correlation test	P-value
Fooodpoverty against Absolutepoverty	37	Pearson	0.846	4.20E-11
		Spearman	0.883	4.98E-13
		Kendall	0.699	1.27E-09
Fooodpoverty against Relativepoverty	37	Pearson	0.807	1.58E-09
		Spearman	0.823	7.13E-09
		Kendall	0.613	9.21E-09
Fooodpoverty against Dollarpoverty	37	Pearson	0.843	5.71E-11
		Spearman	0.874	1.63E-12
		Kendall	0.689	2.08E-09
Absolutepoverty against Relativepoverty	37	Pearson	0.968	2.20E-16
		Spearman	0.982	2.20E-16
		Kendall	0.852	1.30E-13
Absolutepoverty against Dollarpoverty	37	Pearson	0.91	2.20E-16
		Spearman	0.999	2.20E-16
		Kendall	0.991	2.20E-16
Dollarpoverty against Relativepoverty	37	Pearson	0.961	2.20E-16
		Spearman	0.983	2.20E-16
		Kendall	0.845	1.96E-13

Table 4.1 shows that the Pearson's product-moment correlation test (PPMCT), Spearman's rank correlation test (SRCT) and Kendall's rank correlation test (KRCT) results were all positive values and closed to unit value (1) which indicate high correlation between the Poverty levels in Nigeria in terms of Food Poverty, Absolute Poverty, Relative Poverty and Dollar Poverty per day. Hence, the *p-value < 0.05* allows us to accept the value of Pearson’s, Spearman's and Kendall's rank correlation test calculated as being statistically significant.

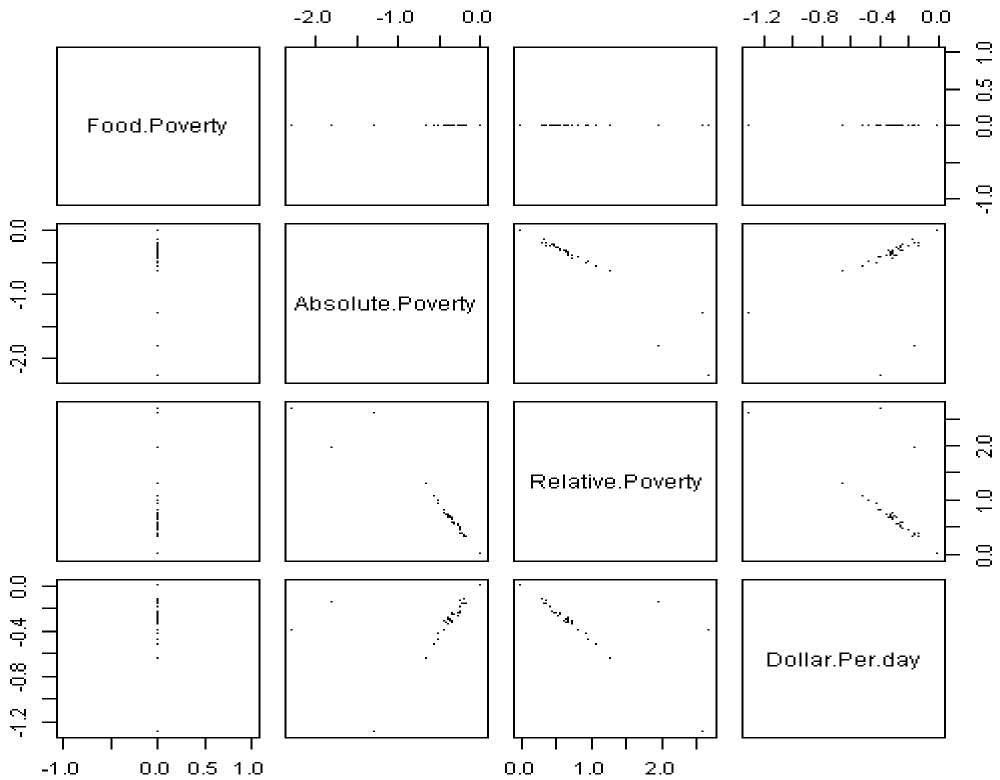

Figure 4.1: Relationship between different Poverty levels profile in Nigeria.

Fig 4.1 shows that increase in Food poverty will constantly increase Absolute poverty, Relative poverty and Dollar per day poverty in Nigeria. Increase in Absolute poverty will decrease Relative poverty and increase Dollar per day poverty in Nigeria. Thus, increase in Relative poverty will decrease Absolute poverty and Dollar per day poverty in Nigeria. Hence increased Dollar per day poverty will Increased Absolute poverty and decreased Relative poverty in Nigeria.

Although, these results are from Poverty levels in Nigeria and they are based on different methods of correlation coefficients. Hence, this conclusion is not enough to be generalized since these methods are different in methodologies for statistical analysis, and the calculated *p-value* of Pearson's, Spearman's and Kendall's rank correlation test are all < 0.05 as being statistically significant. Simulation study under Normal, Log-Normal, Exponential and Contaminated random distributions were conducted with different sample sizes (5, 10, 15 and 20) was carried out. Thus, since these methods have different methodologies for statistical analysis, and their correlation coefficient test cannot be further compared directly unless they are transformed using t and z test statistics for testing correlation coefficients.

The transformed results of these methods will be compared using proportions of rejecting true null hypothesis obtained from t and z test statistics for testing correlation coefficients that is, to be able to justify whether there is a significant difference between the measures of associations of these methods or not.

4.3 Comparison of Correlation Coefficient Tests

Table 4.2 below shows all the proportion of rejecting H_0 under normal, log-normal, exponential and contaminated normal distributions and the method with range proportion of rejecting H_0 close to the nominal $\alpha = 0.05$ or $0.04 \le \alpha \le 0.06$ is the best.

Under normal and contaminated normal distribution, Pearson's was found to have the best degree of association. While, under log-normal and exponential distribution, Spearman's rank was found to have the best degree of association.

Table4.2: Proportion of rejecting $H_0 : \rho = 0$ When $\mu_1 = 1,\ \mu_2 = 2\ \ \sigma_1^{\ 2} = 2,\ \sigma_2^{\ 2} = 3$.

Distributions	Correlations	n1=5	n2=10	n1=15	n2=20
	Pearson	0.059	0.053	0.049	0.058
Normal	Kendall	**0.007**	**0.020**	**0.010**	**0.008**
	Spearman	0.041	(0.065)	0.046	0.062
	Pearson	(0.115)	(0.105)	(0.077)	(0.100)
Log Normal	Kendall	**0.010**	**0.017**	**0.019**	**0.011**
	Spearman	0.044	0.056	0.053	0.049
	Pearson	(0.074)	(0.072)	(0.067)	(0.069)
Exponential	Kendall	**0.007**	**0.016**	**0.008**	**0.011**
	Spearman	0.045	0.052	0.053	0.047
	Pearson	0.044	0.042	0.050	0.051
Contaminated	Kendall	**0.006**	**0.008**	**0.015**	**0.006**
Normal	Spearman	0.037	0.044	0.050	0.052

NB *Bolded values are bellow the lower limit of the 95% confidence interval (0.040, 0.064), while the Bracket values are above the upper limit of the interval.*

Table 4.3: Proportion of rejecting $H_0\ : \rho_1 = \rho_2$ when $\mu_1 = 2,\ \mu_2 = 1\ \ \sigma_1^{\ 2} = 2\ ,\ \sigma_2^{\ 2} = 3$.

Distributions	Correlations	n1=5	n2=10	n1=15	n2=20
	Pearson	0.051	0.048	0.057	0.058
	Kendall	**0.023**	**0.005**	**0.007**	**0.008**
Normal	Spearman	0.059	0.05	0.048	0.056
	Pearson	(0.089)	(0.074)	(0.069)	(0.077)
	Kendall	**0.024**	**0.009**	**0.017**	**0.006**
Log Normal	Spearman	0.053	0.053	0.056	0.052
	Pearson	0.047	0.057	0.053	0.057
	Kendall	**0.027**	**0.009**	**0.009**	**0.008**
Exponential	Spearman	0.052	0.053	0.053	0.051
	Pearson	0.049	0.053	0.054	0.05
Contaminated	Kendall	**0.027**	**0.006**	**0.012**	**0.004**
Normal	Spearman	0.052	0.046	0.053	0.040

NB *Bolded values are bellow the lower limit of the 95% confidence interval (0.040, 0.064), while the Bracket values are above the upper limit of the interval.*

Table 4.3 shows all the proportion of rejecting H_0 under normal, log-normal, exponential and contaminated normal distributions and the method with range proportion of rejecting H_0 close to the nominal α = 0.05 or $0.04 \le \alpha \le 0.06$ is the best.

Under normal, contaminated normal and exponential distribution, Pearson's and Spearman's rank were found to have the best degree of association. While, under log-normal distribution, only Spearman's rank was found to have the best degree of association.

Table4.4: Proportion of rejecting H_0: $\rho = 0$ when $\mu_1 = 2$, $\mu_2 = 1$ $\sigma_1^2 = 4$, $\sigma_2^2 = 2$

Distributions	Correlations	n1=5	n2=10	n1=15	n2=20
Normal Random	Pearson	0.040	0.051	0.043	0.052
	Kendall	**0.003**	**0.016**	**0.006**	**0.008**
	Spearman	0.034	0.057	0.04	0.037
Log Normal	Pearson	(0.157)	(0.114)	(0.091)	(0.069)
	Kendall	**0.006**	**0.010**	**0.007**	**0.008**
	Spearman	0.048	0.053	0.054	0.052
Exponential	Pearson	(0.078)	(0.082)	(0.070)	(0.069)
	Kendall	**0.010**	**0.019**	**0.007**	**0.012**
	Spearman	0.043	0.052	0.051	0.052
Contaminated Normal	Pearson	0.057	0.045	0.053	0.064
	Kendall	**0.005**	**0.016**	**0.010**	**0.018**
	Spearman	0.040	0.045	0.047	(0.067)

NB *Bolded values are bellow the lower limit of the 95% confidence interval (0.040, 0.064), while the Bracket values are above the upper limit of the interval.*

Table 4.4 shows all the proportion of rejecting H_0 under normal, log-Normal, exponential and contaminated normal distributions and the method with range proportion of rejecting H_0 close to the nominal α = 0.05 or $0.04 \le \alpha \le 0.06$ is the best.

Under normal distribution, Pearson's was found to have the best degree of association. While, under log-normal, exponential and contaminated normal distributions, Spearman's rank was found to have the best degree of association.

Table 4.5 Proportion of rejecting H_0 : $\rho_1 = \rho_2$ when $\mu_1 = 2$, $\mu_2 = 1$ $\sigma_1^{\ 2} = 4$, $\sigma_2^{\ 2} = 2$

Distributions	Correlations	n1=5	n2=10	n1=15	n2=20
Normal Random	Pearson	0.044	0.040	0.057	0.054
	Kendall	**0.016**	**0.011**	**0.006**	**0.010**
	Spearman	0.043	0.048	0.051	0.048
Log Normal	Pearson	(0.128)	(0.100)	(0.094)	0.063
	Kendall	**0.026**	**0.010**	**0.019**	**0.006**
	Spearman	0.050	0.049	0.050	0.049
Exponential	Pearson	0.045	0.056	0.055	0.055
	Kendall	**0.027**	**0.014**	**0.006**	**0.012**
	Spearman	0.059	0.056	0.047	0.054
Contaminated Normal	Pearson	0.044	0.038	0.050	0.050
	Kendall	**0.019**	**0.005**	**0.010**	**0.009**
	Spearman	0.052	0.039	0.050	0.063

NB *Bolded values are bellow the lower limit of the 95% confidence interval (0.040, 0.064), while the Bracket values are above the upper limit of the interval.*

Table 4.5 shows all the proportion of rejecting H_0 under normal, log-normal, exponential and contaminated normal distributions and the method with range proportion of rejecting H_0 close to the nominal α = 0.05 or $0.04 \le \alpha \le 0.06$ is the best.

Under normal, contaminated normal and exponential distribution, Pearson's and Spearman's rank were found to have the best degree of association. While, under log-normal distribution, only Spearman's rank was found to have the best degree of association.

Table 4.6 below shows all the proportion of rejecting H_0 under normal, log-normal, exponential and contaminated normal distributions and the method with range proportion of rejecting H_0 close to the nominal α = 0.05 or $0.04 \le \alpha \le 0.06$ is the best.

Under normal distribution, Pearson's was found to be the best degree of association , under log-normal distributions, Spearman's rank was found to have the best degree of association and under exponential and contaminated normal distributions, Pearson's and Spearman's rank were found to have the best degree of association.

Table4.6: Proportion of rejecting H_0: $\rho = 0$ when $\mu_1 = 2$, $\mu_2 = 1$ $\sigma_1^2 = 4$, $\sigma_2^2 = 2$

Distributions	Correlations	n1=20	n2=5	n1=20	n2=20
	Pearson	0.058	0.049	0.058	0.046
Normal Random	Kendall	**0.008**	**0.013**	**0.016**	**0.008**
	Spearman	0.053	0.031	0.058	0.045
	Pearson	(0.069)	(0.155)	(0.089)	0.046
	Kendall	**0.017**	**0.009**	**0.011**	**0.008**
Log Normal	Spearman	0.051	0.058	0.055	0.045
	Pearson	0.063	0.048	0.065	0.044
	Kendall	**0.012**	**0.003**	**0.010**	**0.010**
Exponential	Spearman	0.051	0.051	0.058	0.044
	Pearson	0.032	0.05	0.045	0.044
Contaminated Normal	Kendall	**0.004**	**0.006**	**0.008**	**0.010**
	Spearman	0.037	0.050	0.046	0.044

NB *Bolded values are bellow the lower limit of the 95% confidence interval (0.040, 0.064), while the Bracket values are above the upper limit of the interval.*

Table 4.7 below shows all the proportion of rejecting H_0 under normal, log-normal, exponential and contaminated normal distributions and the method with range proportion of rejecting H_0 close to the nominal α = 0.05 or $0.04 \le \alpha \le 0.06$ is the best. Under normal, contaminated normal and exponential distribution, Pearson's and Spearman's rank were found to have the best degree of association. While, under log-normal distribution, only Spearman's rank was found to have the best degree of association.

Table 4.7 Proportion of rejecting H_0: $\rho_1 = \rho_2$ when $\mu_1 = 2$, $\mu_2 = 1$ $\sigma_1^2 = 4$, $\sigma_2^2 = 2$

Distributions	Correlations	n1=10	n2=10	n1=20	n2=20
Normal Random	Pearson	0.055	0.054	0.046	0.041
	Kendall	**0.010**	**0.022**	**0.014**	**0.004**
	Spearman	0.051	0.05	0.052	0.045
Log Normal	Pearson	(0.071)	(0.139)	(0.072)	0.041
	Kendall	**0.012**	**0.02**	**0.013**	**0.004**
	Spearman	0.051	0.050	0.061	0.045
Exponential	Pearson	0.052	0.063	0.047	0.044
	Kendall	**0.010**	**0.020**	**0.008**	**0.006**
	Spearman	0.048	0.057	0.055	0.050
Contaminated Normal	Pearson	0.046	0.043	0.054	0.044
	Kendall	**0.011**	**0.016**	**0.006**	**0.006**
	Spearman	0.054	0.059	0.048	0.050

NB *Bolded values are bellow the lower limit of the 95% confidence interval (0.040, 0.064), while the Bracket values are above the upper limit of the interval.*

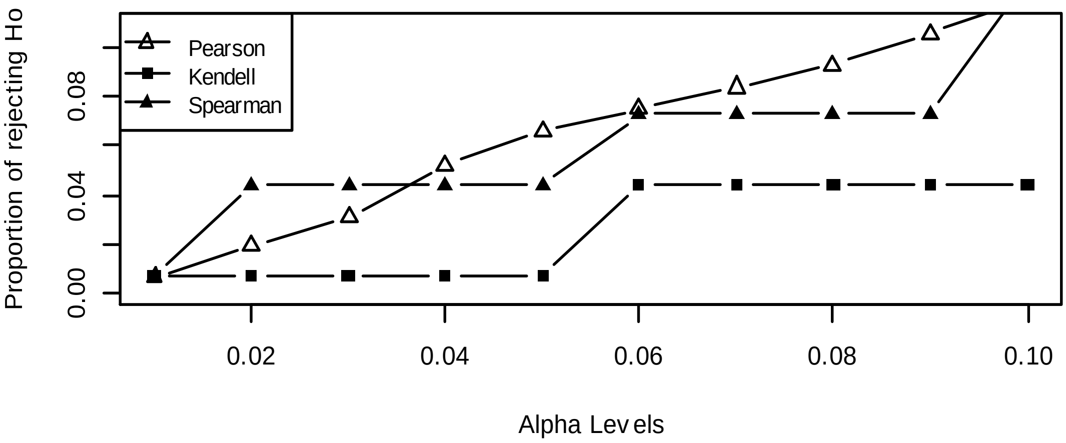

Figure 4.2: Proportion of rejecting $H_0: \ \rho = 0$ under Normal distribution when n=5

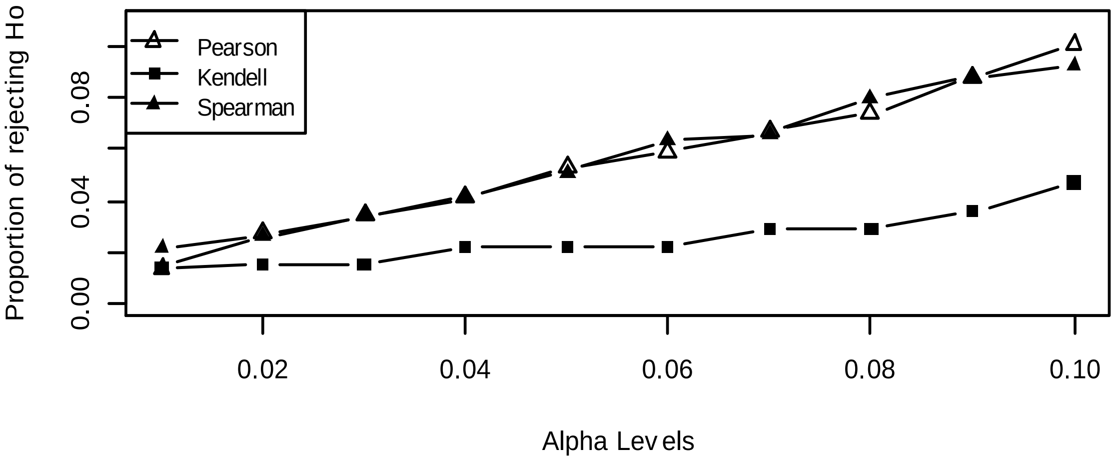

Figure 4.3: Proportion of rejecting H_0: $\rho_1 = \rho_2$ under Normal distribution when n = 5

Figures 4.2 and 4.3 show the graphical representation of Proportion of rejecting H_0 under Normal random distributions with n = 5 for both $\rho = 0$ and $\rho_1 = \rho_2$, Pearson's and Spearman have almost similar Proportion of rejecting H_0 ranging from $0.02 < \alpha > 0.08$ while the Kendall's has the Proportion of rejecting H_0 ranging from $0.00 \leq \alpha < 0.04$

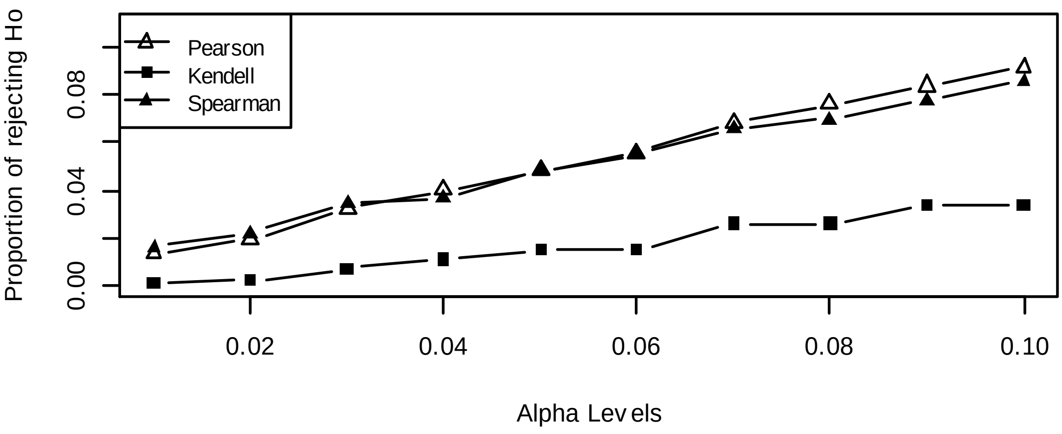

Figure 4.4: Proportion of rejecting H_0: $\rho = 0$ under Normal distribution when n=10

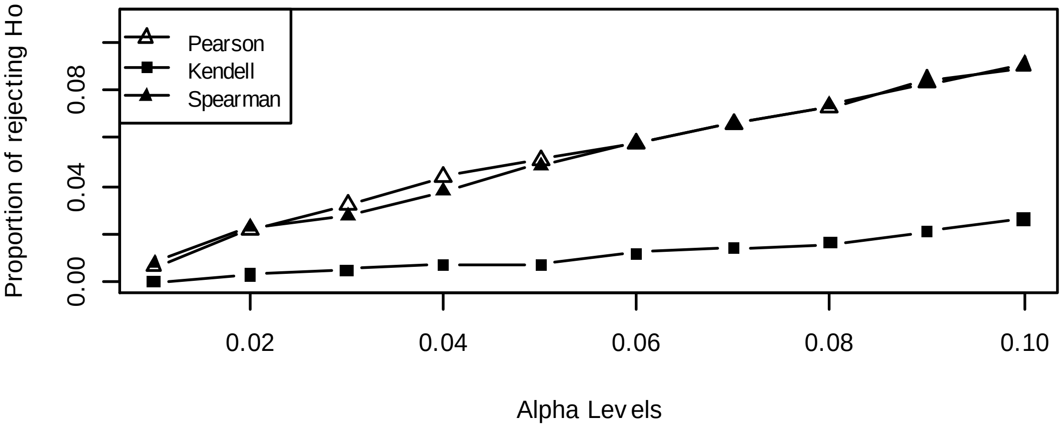

Figure 4.5: Proportion of rejecting H_0: $\rho_1 = \rho_2$ under Normal distribution when n = 10

Figures 4.4 and 4.5 show the graphical representation of Proportion of rejecting H_0 under Normal random distributions with n = 10 for both $\rho = 0$ and $\rho_1 = \rho_2$, Pearson's and Spearman have almost similar Proportion of rejecting H_0 ranging from $0.02 < \alpha > 0.08$ while the Kendall's has the Proportion of rejecting H_0 ranging from $0.00 \le \alpha < 0.04$

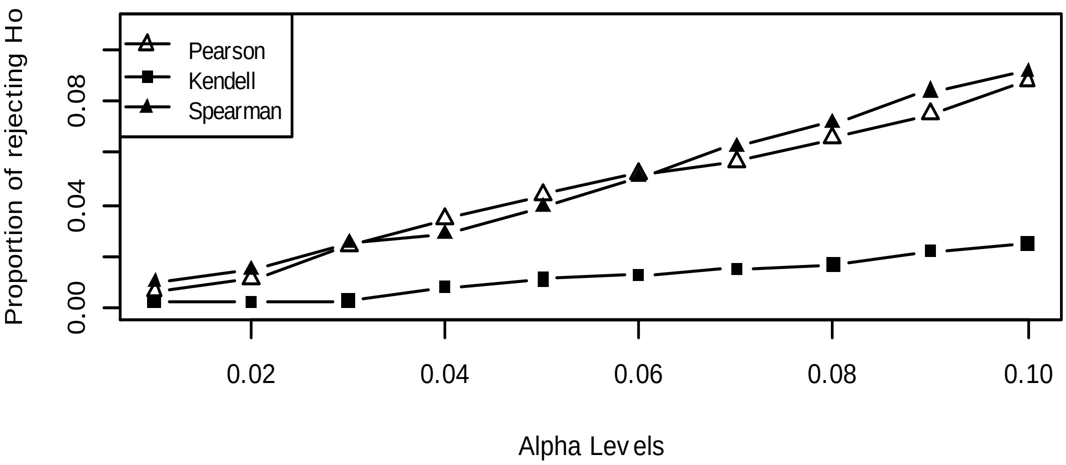

Figure 4.6: Proportion of rejecting H_0: $\rho = 0$ under Normal distribution when n=15

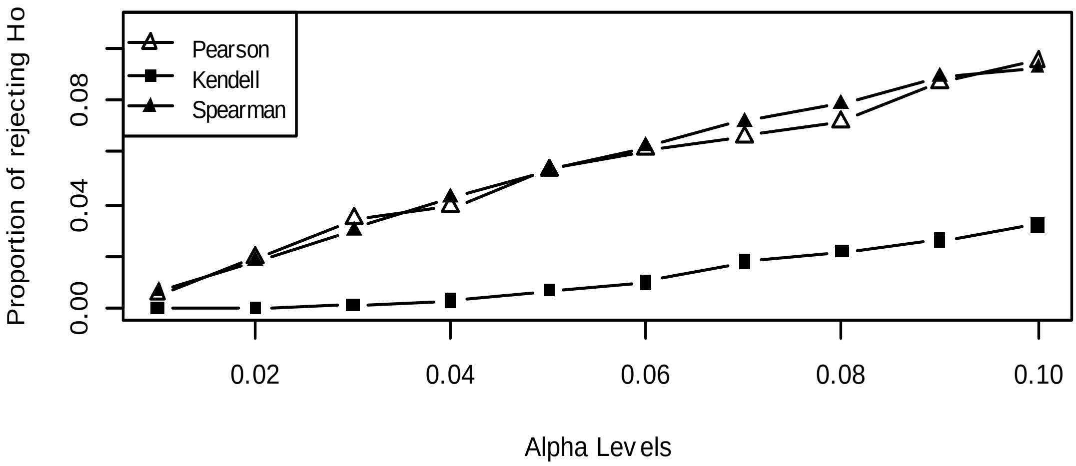

Figure 4.7: Proportion of rejecting H_0: $\rho_1 = \rho_2$ under Normal distribution when n = 15

Figures 4.6 and 4.7 show the graphical representation of Proportion of rejecting H_0 under Normal random distributions with n = 15 for both $\rho = 0$ and $\rho_1 = \rho_2$, Pearson's and Spearman have almost similar Proportion of rejecting H_0 ranging from $0.02 < \alpha > 0.08$ while the Kendall's has the Proportion of rejecting H_0 ranging from $0.00 \leq \alpha < 0.04$

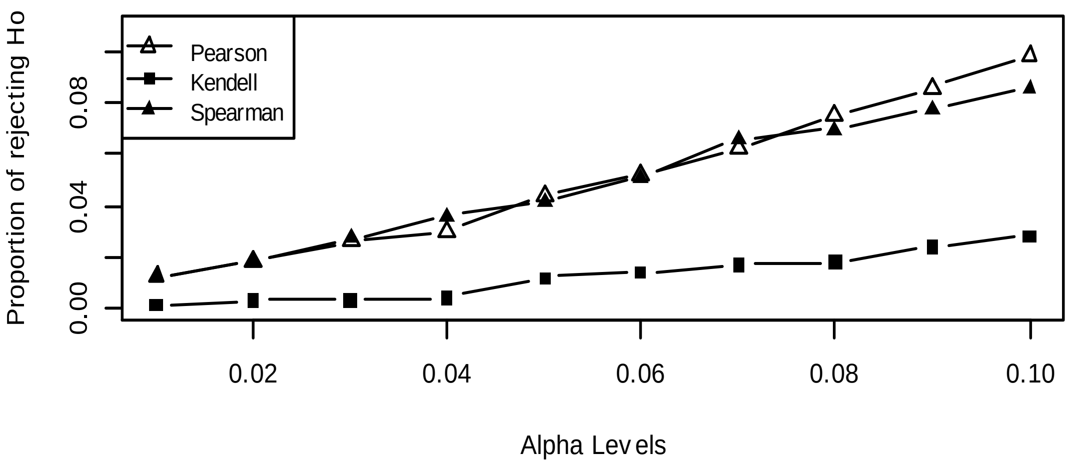

Figure 4.8: Proportion of rejecting H_0: $\rho = 0$ under Normal distribution when n=20

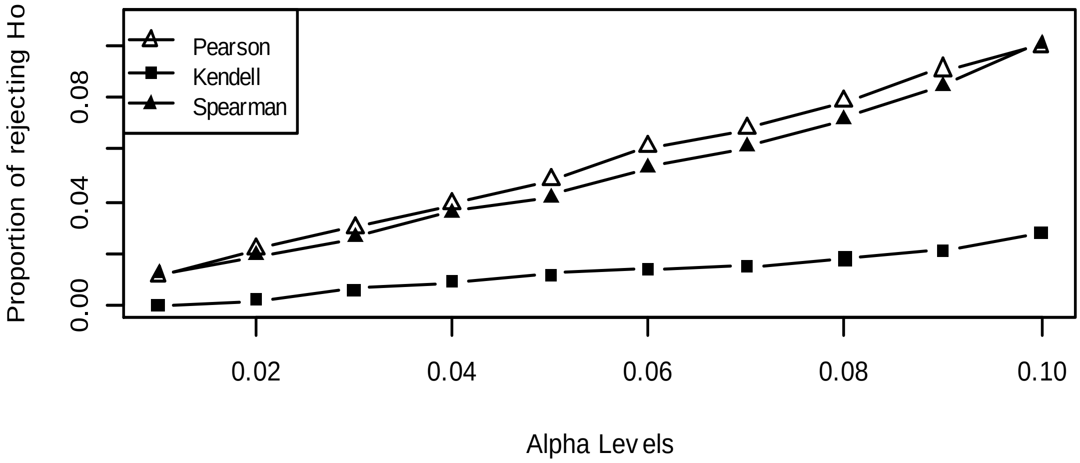

Figure 4.9: Proportion of rejecting H_0: $\rho_1 = \rho_2$ under Normal distribution when n = 20

Figures 4.8 and 4.9 show the graphical representation of Proportion of rejecting H_0 under Normal random distributions with n = 20 for both $\rho = 0$ and $\rho_1 = \rho_2$, Pearson's and Spearman have almost similar Proportion of rejecting H_0 ranging from $0.02 < \alpha > 0.08$ while the Kendall's has the Proportion of rejecting H_0 ranging from $0.00 \le \alpha < 0.04$

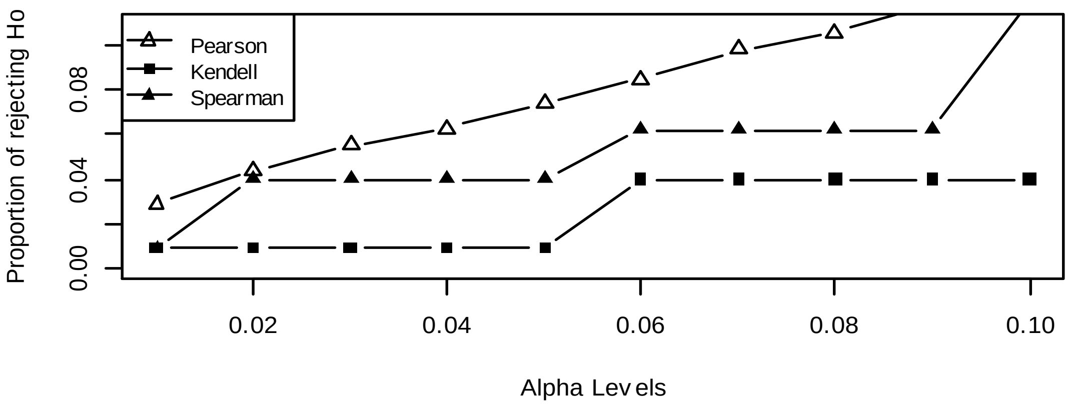

Figure 4.10: Proportion of rejecting H_0: $\rho = 0$ under Log-normal distribution when n=5

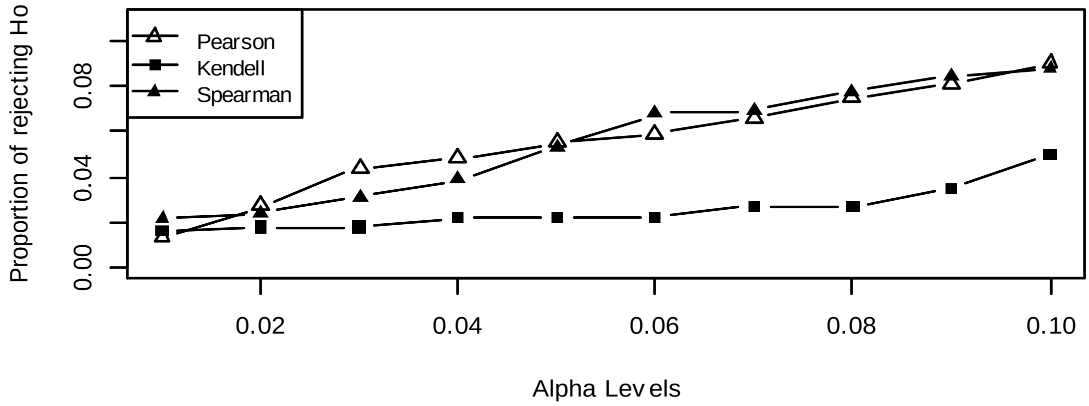

Figure 4.11: Proportion of rejecting H_0: $\rho_1 = \rho_2$ under Log-normal distribution when n = 5

Figures 4.10 and 4.11 show the graphical representation of Proportion of rejecting H_0 under Log-normal random distributions with n = 5 for $\rho = 0$ the Proportion of rejecting H_0 under Pearson's ranging from $0.02 > \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$. While $\rho_1 = \rho_2$, the Proportion of rejecting H_0 under Pearson's ranging from $0.04 < \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$

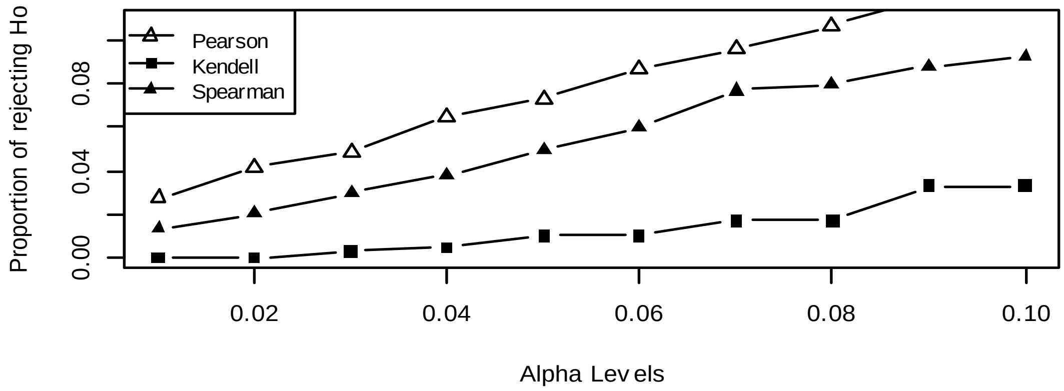

Figure 4.12: Proportion of rejecting H_0: $\rho = 0$ under Log-normal distribution when n=10

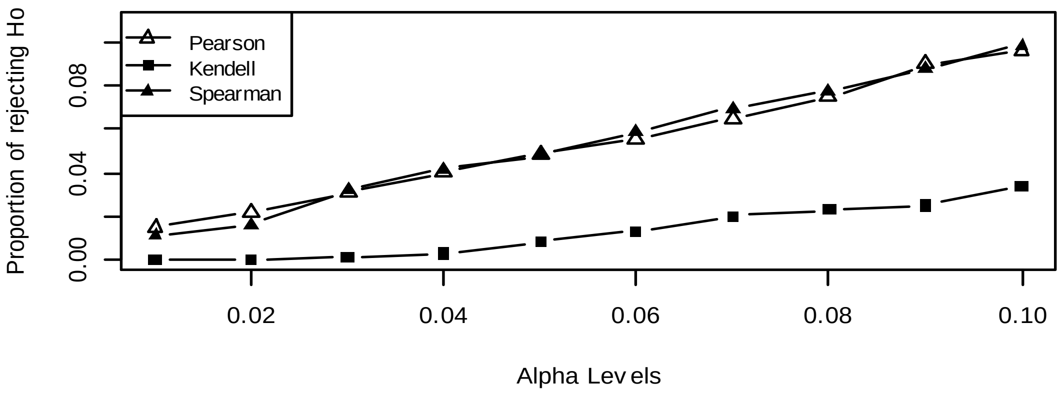

Figure 4.13: Proportion of rejecting H_0: $\rho_1 = \rho_2$ under Log-normal Normal distribution when n = 10

Figures 4.12 and 4.13 show the graphical representation of Proportion of rejecting H_0 under Log normal random distributions with n = 10 for $\rho = 0$ the Proportion of rejecting H_0 under Pearson's ranging from $0.02 > \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$. While $\rho_1 = \rho_2$, the Proportion of rejecting H_0 under Pearson's ranging from $0.02 < \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$

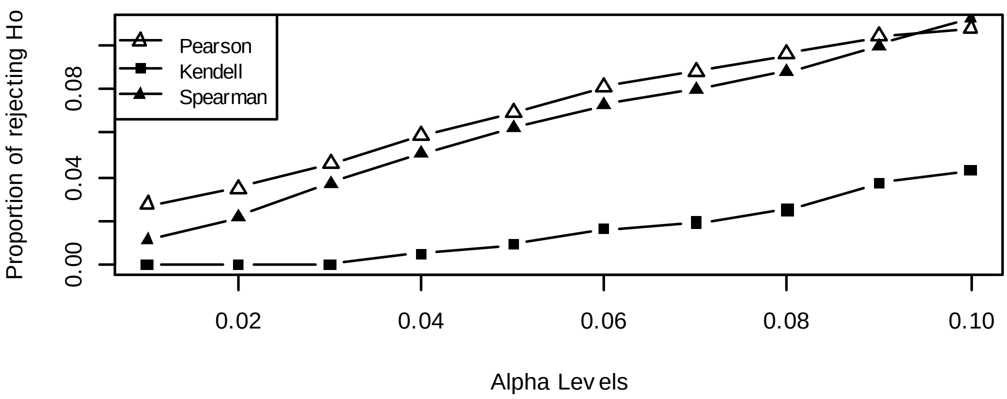

Figure 4.14: Proportion of rejecting H_0: $\rho = 0$ under Log-normal distribution when n=15

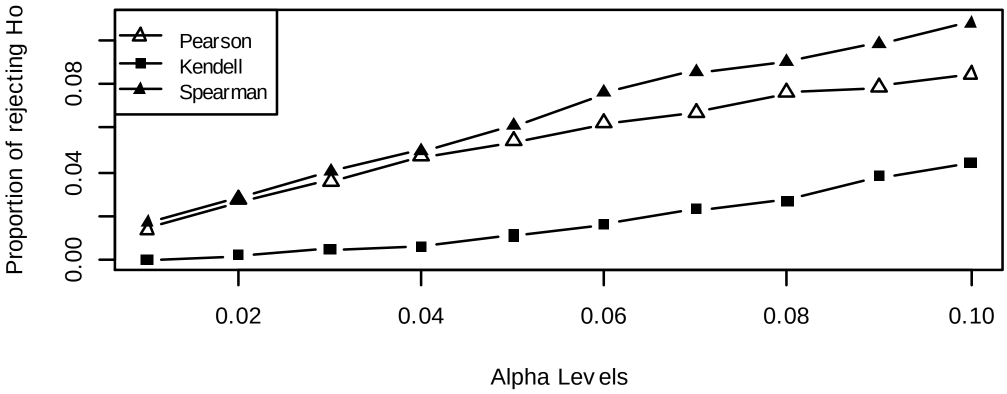

Figure 4.15: Proportion of rejecting H_0: $\rho_1 = \rho_2$ under Log-normal Normal distribution when n = 15

Figures 4.14 and 4.15 show the graphical representation of Proportion of rejecting H_0 under Log normal random distributions with n = 15 for $\rho = 0$ the Proportion of rejecting H_0 under Pearson's ranging from $0.02 > \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$. While $\rho_1 = \rho_2$, the Proportion of rejecting H_0 under Pearson's ranging from $0.02 < \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$

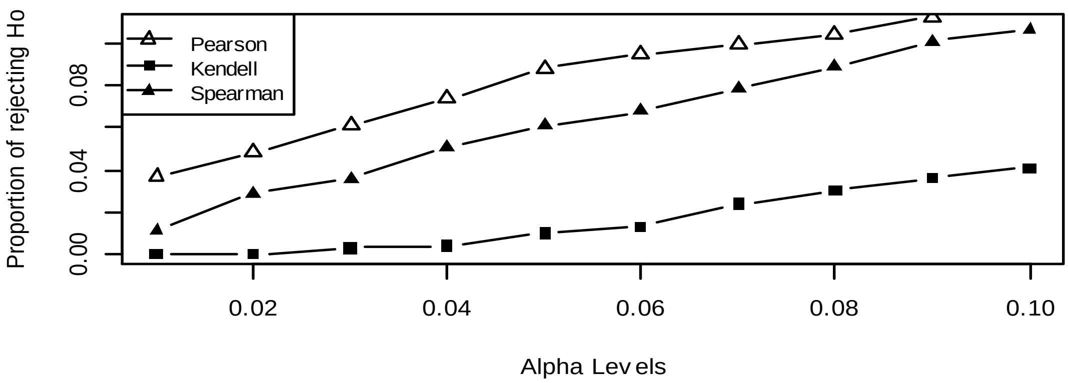

Figure 4.16: Proportion of rejecting H_0: $\rho = 0$ under Log-normal distribution when n=20

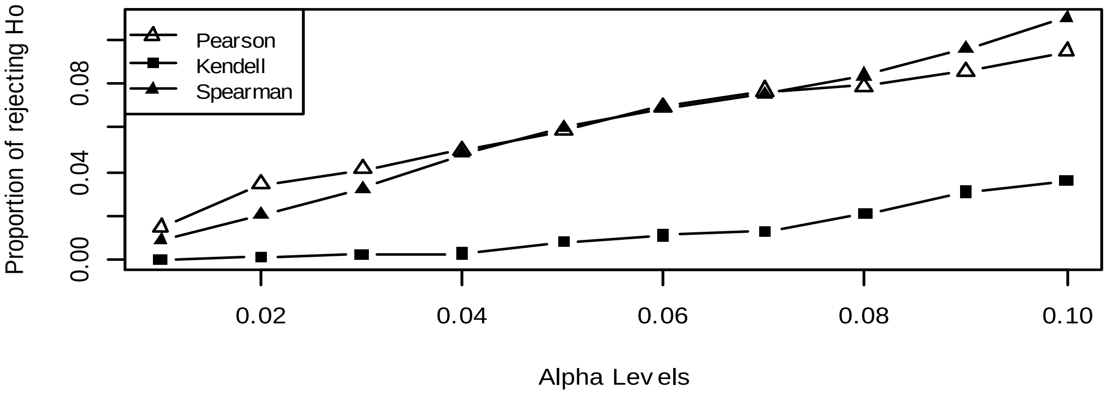

Figure 4.17: Proportion of rejecting $H_0: \rho_1 = \rho_2$ under Log-normal Normal distribution when n = 20.

Figures 4.16 and4.17 show the graphical representation of Proportion of rejecting H_0 under Log normal random distributions with n = 20 for $\rho = 0$ the Proportion of rejecting H_0 under Pearson's ranging from $0.02 > \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$. While $\rho_1 = \rho_2$, the Proportion of rejecting H_0 under Pearson's ranging from $0.02 < \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$

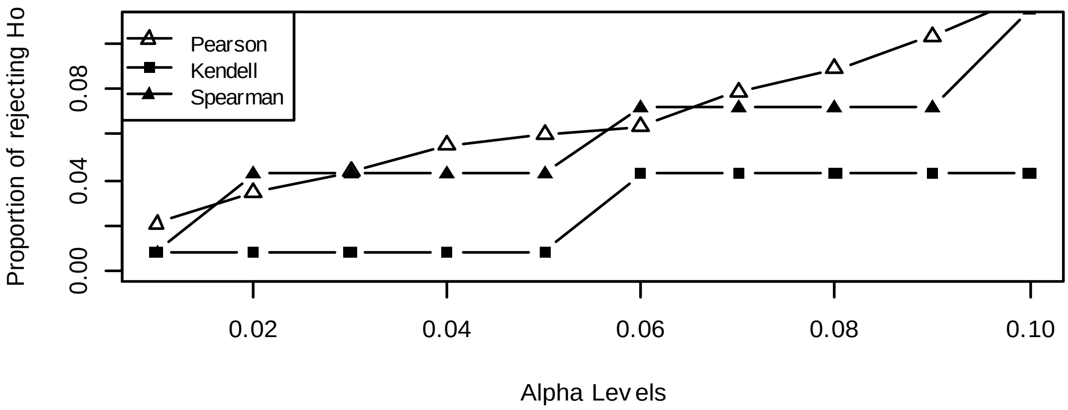

Figure 4.18: Proportion of rejecting $H_0: \rho = 0$ under exponential distribution when n=5

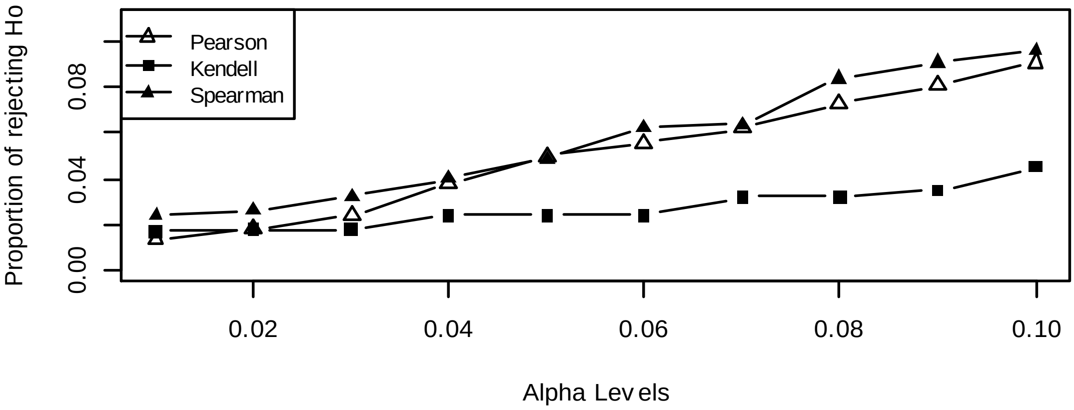

Figure 4.19: Proportion of rejecting H_0: $\rho_1 = \rho_2$ under exponential distribution when n = 5

Figures 4.18 and 4.19 show the graphical representation of Proportion of rejecting H_0 under Exponential random distributions with n = 5 for $\rho = 0$ the Proportion of rejecting H_0 under Pearson's ranging from $0.02 > \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$. While $\rho_1 = \rho_2$, the Proportion of rejecting H_0 under Pearson's ranging from $0.02 < \alpha > 0.08$, Spearman's ranging from $0.02 \leq \alpha > 0.08$, and Kendall's ranging from $0.02 < \alpha < 0.04$

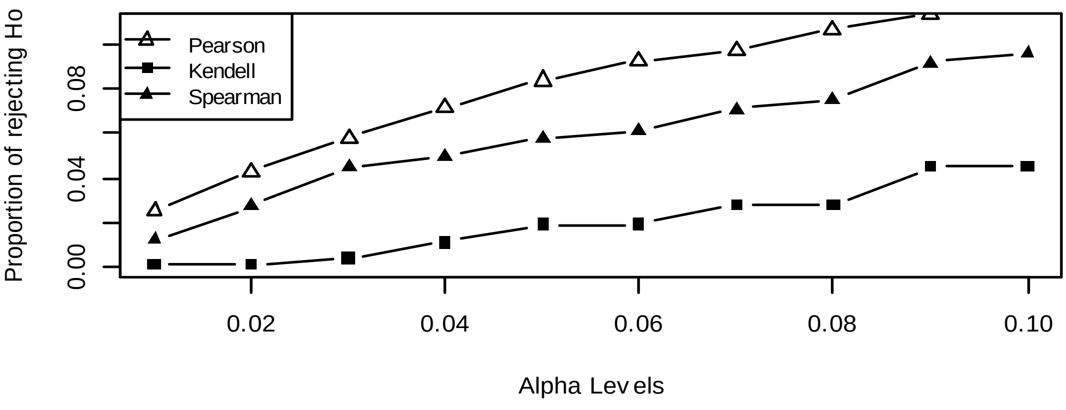

Figure 4.20: Proportion of rejecting H_0: $\rho = 0$ under Exponential distribution when n=10

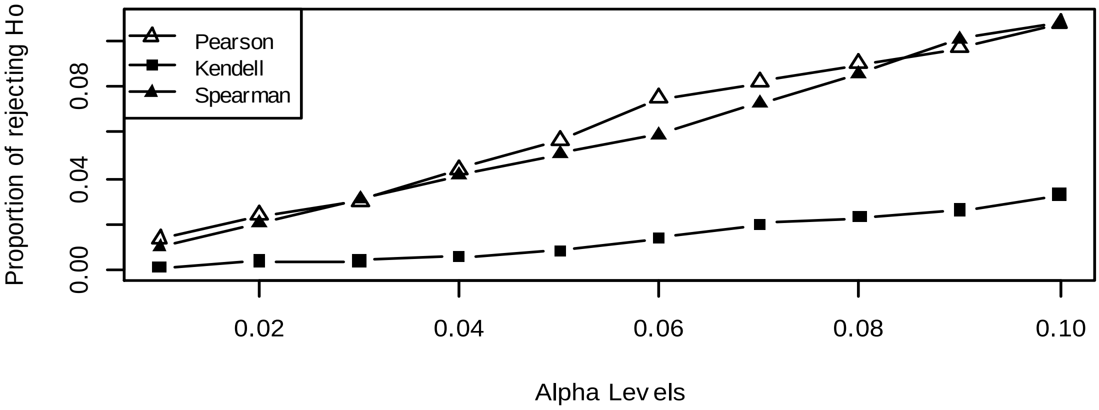

Figure 4.21: Proportion of rejecting H_0: $\rho_1 = \rho_2$ under Exponential distribution when n = 10

Figures 4.20 and 4.21 show the graphical representation of Proportion of rejecting H_0 under Exponential random distributions with n = 10 for $\rho = 0$ the Proportion of rejecting H_0 under Pearson's ranging from $0.02 > \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$. While $\rho_1 = \rho_2$, the Proportion of rejecting H_0 under Pearson's ranging from $0.02 < \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$.

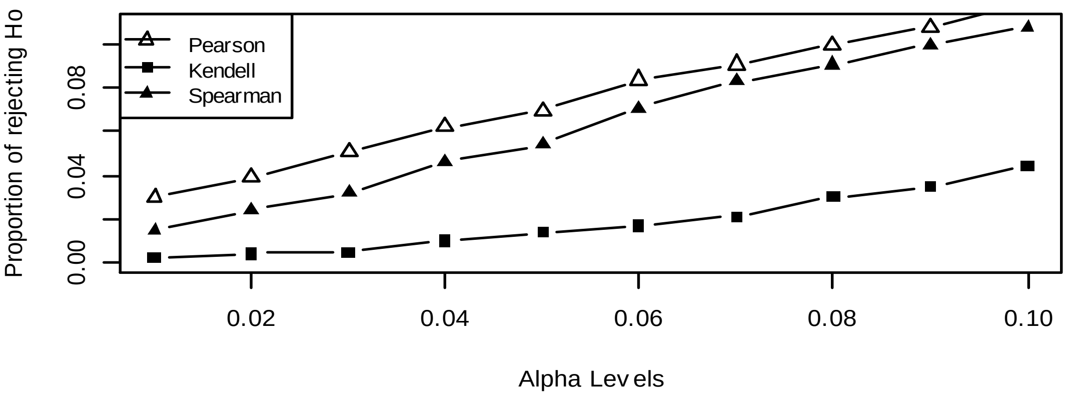

Figure 4.22: Proportion of rejecting H_0: $\rho = 0$ under Exponential distribution when n=15

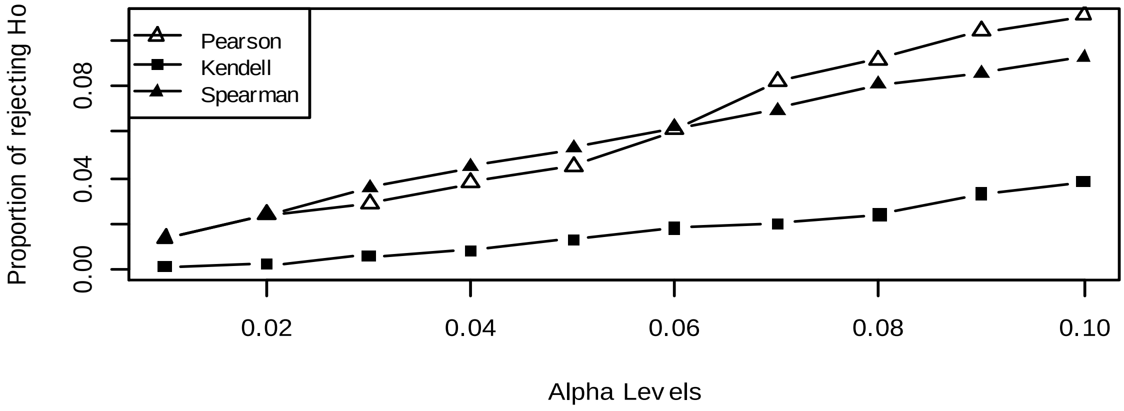

Figure 4.23: Proportion of rejecting H_0 : $\rho_1 = \rho_2$ under Exponential distributions when n = 15

Figures 4.22 and 4.23 show the graphical representation of Proportion of rejecting H_0 under Exponential random distributions with n = 15 for $\rho = 0$ the Proportion of rejecting H_0 under Pearson's ranging from $0.02 > \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$. While $\rho_1 = \rho_2$, the Proportion of rejecting H_0 under Pearson's ranging from $0.02 < \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$.

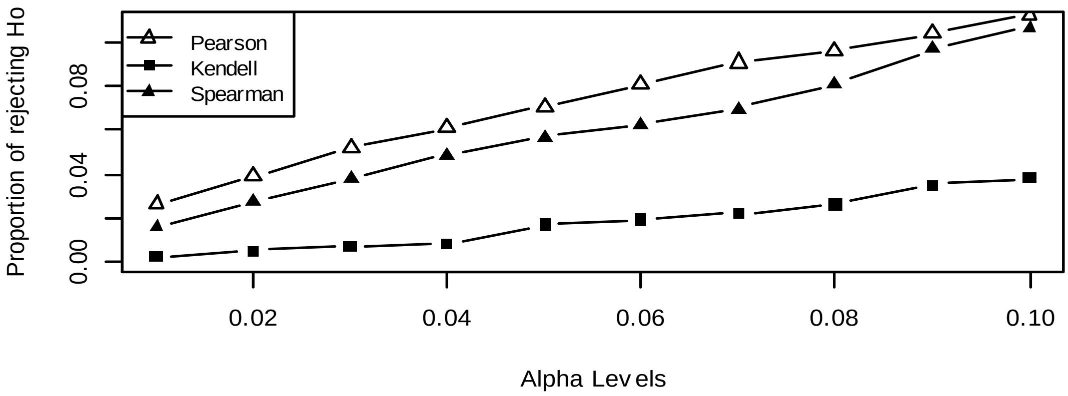

Figure 4.24: Proportion of rejecting H_0 : $\rho = 0$ under Exponential distribution when n=20

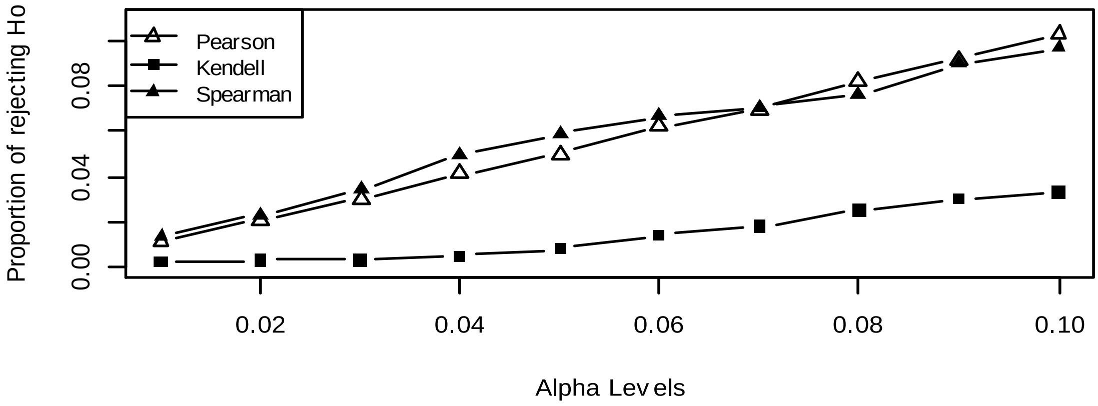

Figure 4.25: Proportion of rejecting H_0: $\rho_1 = \rho_2$ under Exponential distribution when n = 20

Figures 4.24 and 4.25 show the graphical representation of Proportion of rejecting H_0 under Exponential random distributions with n = 20 for $\rho = 0$ $\rho = 0$ the Proportion of rejecting H_0 under Pearson's ranging from $0.02 > \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$. While $\rho_1 = \rho_2$, the Proportion of rejecting H_0 under Pearson's ranging from $0.02 < \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$.

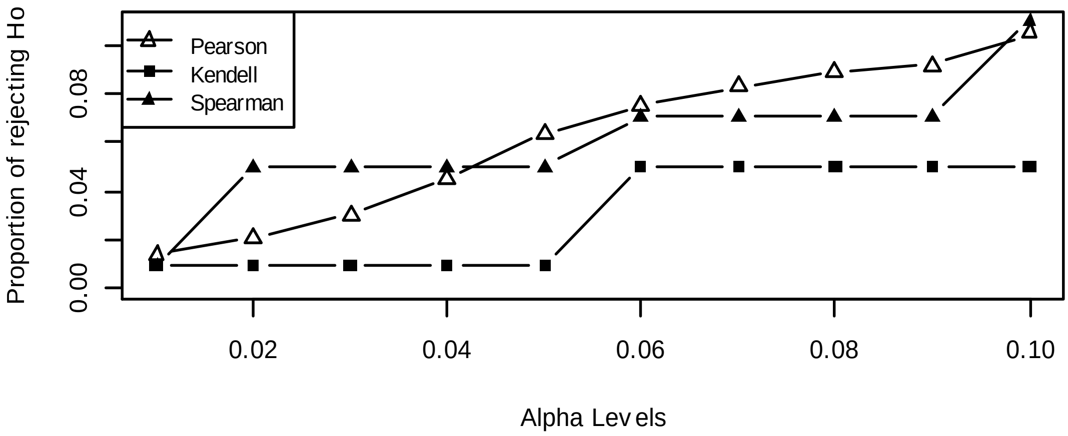

Figure 4.26: Proportion of rejecting H_0: $\rho = 0$ under Contaminated distribution when n=5

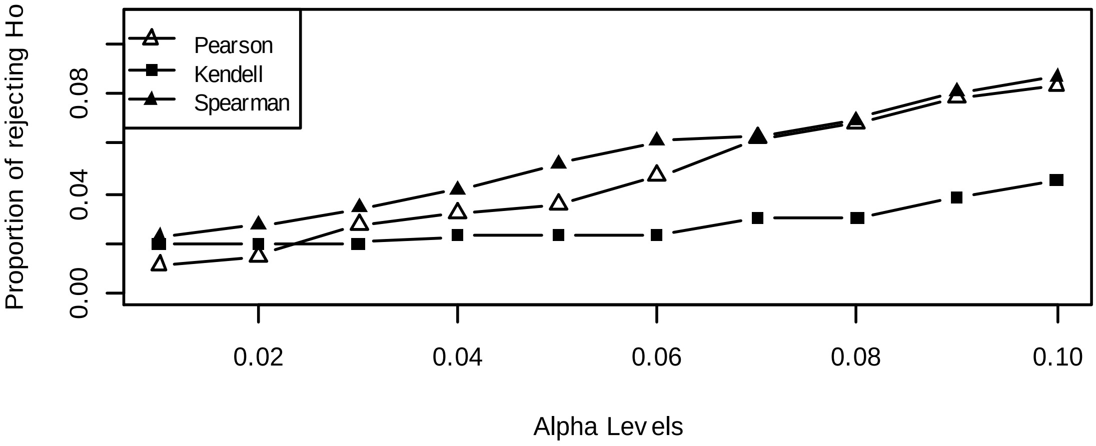

Figure 4.27: Proportion of rejecting H_0: $\rho_1 = \rho_2$ under contaminated distribution when n = 5

Figures 4.26 and 4.27 show the graphical representation of Proportion of rejecting H_0 under Contaminated random distributions with n = 5 for both $\rho = 0$ and $\rho_1 = \rho_2$, Pearson's and Spearman have almost similar Proportion of rejecting H_0 ranging from $0.02 < \alpha > 0.08$ while the Kendall's has the Proportion of rejecting H_0 ranging from $0.00 \leq \alpha < 0.04$

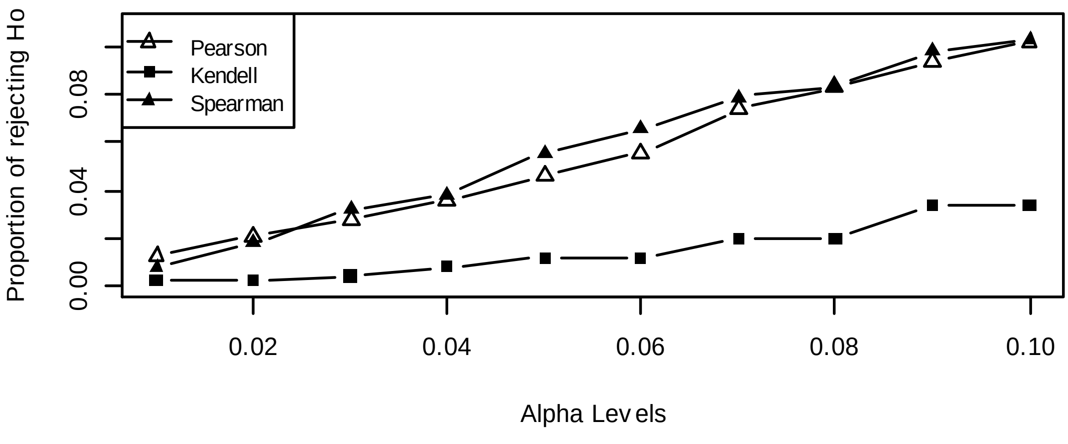

Figure 4.28: Proportion of rejecting H_0: $\rho = 0$ under Contaminated distribution when n=10

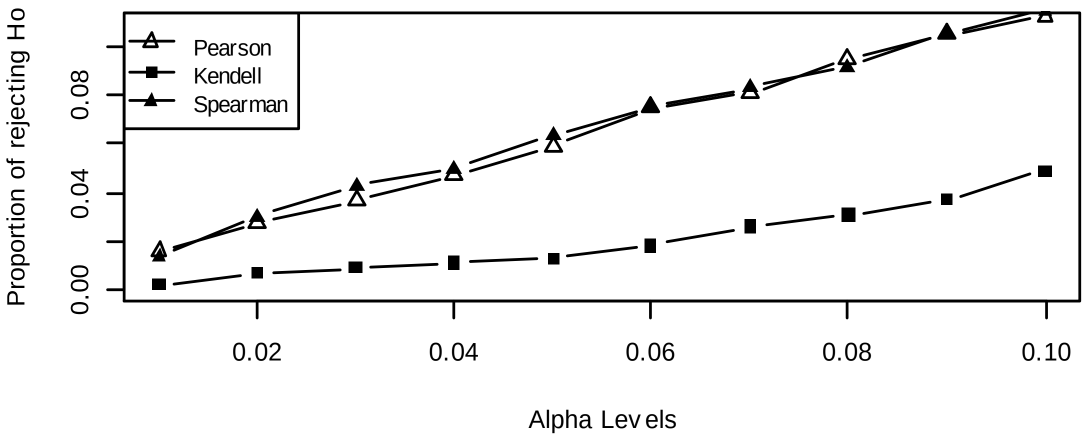

Figure 4.29: Proportion of rejecting H_0: $\rho_1 = \rho_2$ under contaminated distribution when n = 10

Figures 4.28 and 4.29 show the graphical representation of Proportion of rejecting H_0 under Contaminated random distributions with n = 10 for both $\rho = 0$ and $\rho_1 = \rho_2$, Pearson's and Spearman have almost similar Proportion of rejecting H_0 ranging from $0.02 < \alpha > 0.08$ while the Kendall's has the Proportion of rejecting H_0 ranging from $0.00 \leq \alpha < 0.04$

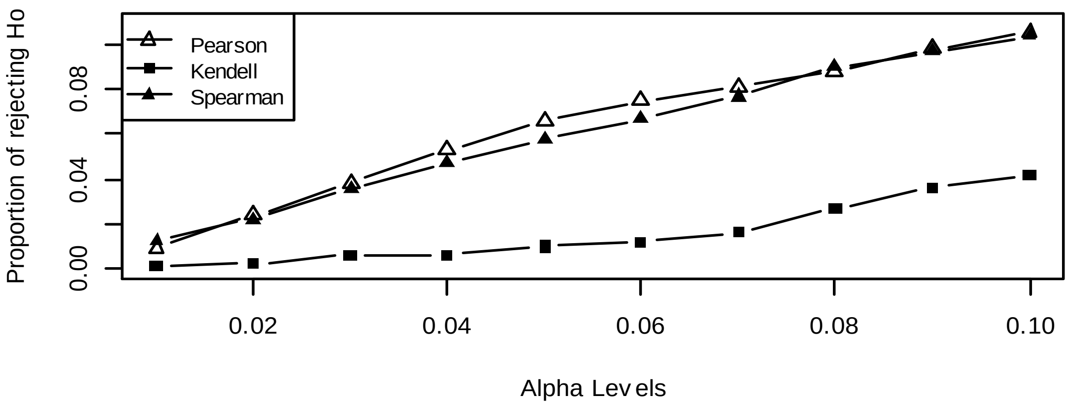

Figure 4.30: Proportion of rejecting H_0: $\rho = 0$ under Contaminated random distribution when n=15

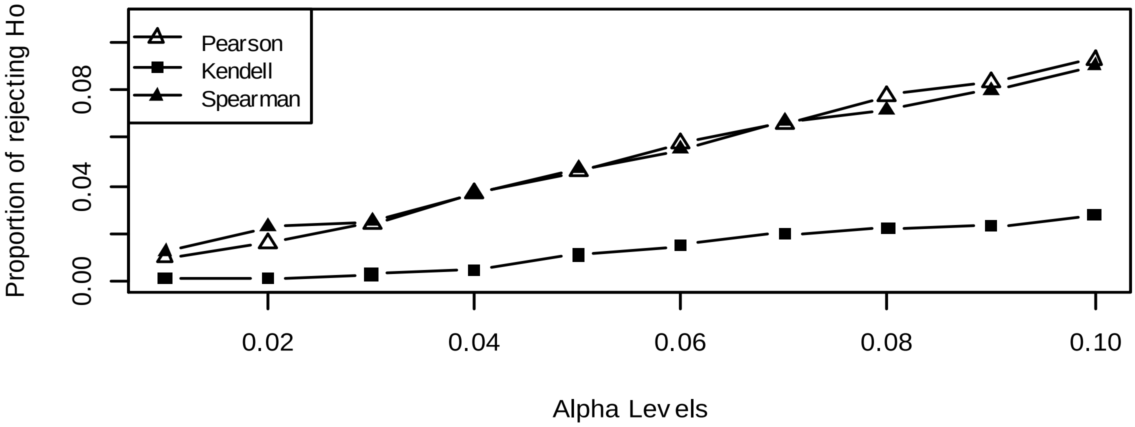

Figure 4.31: Proportion of rejecting H_0: $\rho_1 = \rho_2$ under contaminated distribution when n = 15

Figures 4.30 and 4.31 show the graphical representation of Proportion of rejecting H_0 under Contaminated random distributions with n = 15 for both $\rho = 0$ and $\rho_1 = \rho_2$,Pearson's and Spearman have almost similar Proportion of rejecting H_0 ranging from $0.02 < \alpha > 0.08$ while the Kendall's has the Proportion of rejecting H_0 ranging from $0.00 \le \alpha < 0.04$

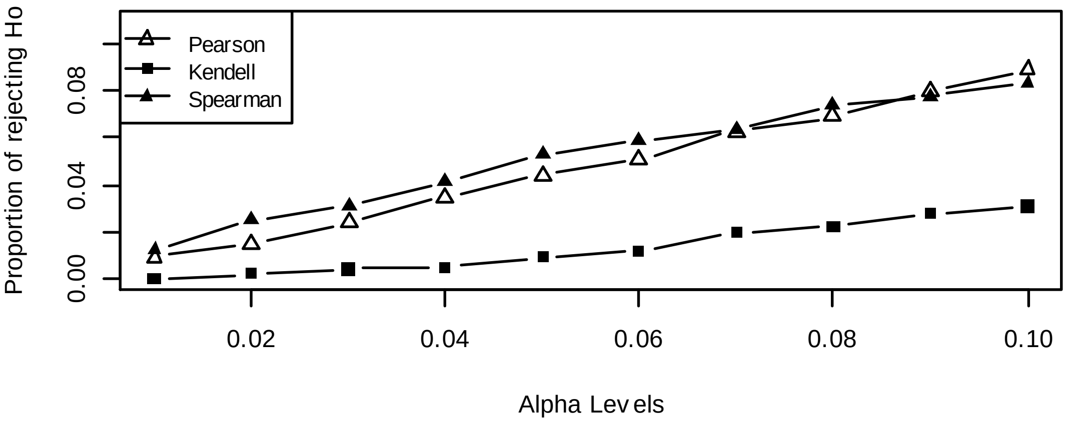

Figure 4.32: Proportion of rejecting H_0: $\rho = 0$ under Contaminated random distribution when n=20

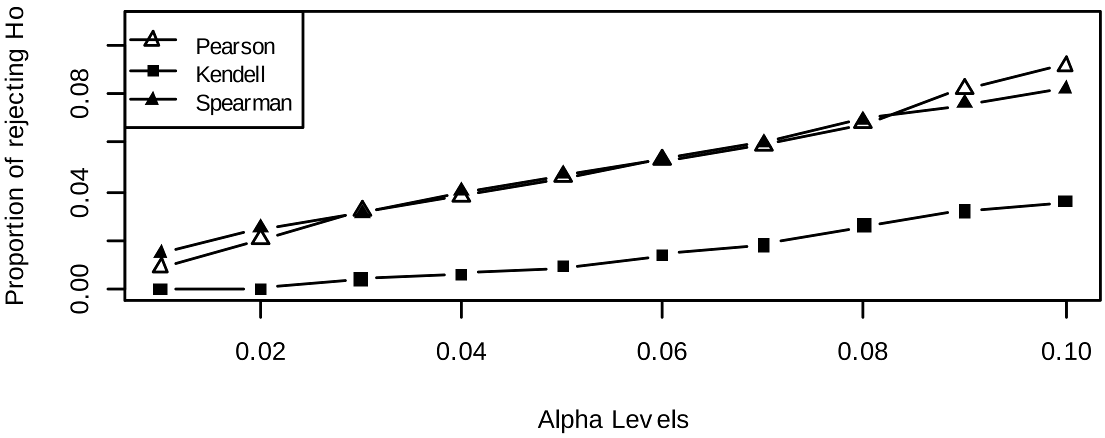

Figure 4.33: Proportion of rejecting $H_0: \rho_1 = \rho_2$ under contaminated distribution when n = 20

Figures 4.32 and 4.33 show the graphical representation of Proportion of rejecting H_0 under Contaminated random distributions with n = 20 for both $\rho = 0$ and $\rho_1 = \rho_2$, Pearson's and Spearman have almost similar Proportion of rejecting H_0 ranging from $0.02 < \alpha > 0.08$ while the Kendall's has the Proportion of rejecting H_0 ranging from $0.00 \le \alpha < 0.04$

4.4 Discussion of Findings

The "spurious correlation test" used in this research are Pearson's product-moment correlation test, Spearman's rank correlation test and Kendall's rank correlation test; and the fact that emanated from them are as follows:

Table 4.1 shows that, Pearson's product-moment correlation test (PPMCT), Spearman's rank correlation test (SRCT) and Kendall's rank correlation test (KRCT) results were all positive values and closed to unit value (1) which indicate high correlation between the Poverty levels in Nigeria in terms of Food Poverty, Absolute Poverty, Relative Poverty and Dollar Poverty per day. Hence, the *p-value < 0.05* allows us to accept the value of Pearson's, Spearman's and Kendall's rank correlation test calculated as being statistically significant.

Furthermore, the result of the Pearson's, Spearman's and Kendall's rank correlation tests on the tables 4.1, 4.2, 4.3, 4.4, 4.5, 4.6 and 4.7 were transformed and compared using proportions of rejecting true null hypothesis obtained from t and z test statistics for testing correlation coefficients, only the result of α = 0.05 are displayed in the tables.

Table 4.2 shows all the proportion of rejecting H_0 under normal, log-normal, exponential and contaminated normal distributions and the method with range proportion of rejecting H_0 close to the nominal α = 0.05 or $0.04 \leq \alpha \leq 0.06$ is the best.

Under normal and contaminated normal distribution, Pearson's was found to give the best degree of association. While, under log-normal and exponential distribution, Spearman's rank was found to give the best degree of association.

Table 4.3 shows all the proportion of rejecting H_0 under normal, log-normal, exponential and contaminated normal distributions and the method with range proportion of rejecting H_0 close to the nominal α = 0.05 or $0.04 \leq \alpha \leq 0.06$ is the best.

Under normal, contaminated normal and exponential distribution, Pearson's and Spearman's rank were found to give the best degree of association. While, under log-normal distribution, only Spearman's rank was found to give the best degree of association.

Table 4.4 shows all the proportion of rejecting H_0 under normal, log-Normal, exponential and contaminated normal distributions and the method with range proportion of rejecting H_0 close to the nominal α = 0.05 or $0.04 \leq \alpha \leq 0.06$ is the best.

Under normal distribution, Pearson's was found to give the best degree of association. While, under log-normal, exponential and contaminated normal distributions; Spearman's rank was found to give the best degree of association.

Table 4.5 shows all the proportion of rejecting H_0 under normal, log-normal, exponential and contaminated normal distributions and the method with range proportion of rejecting H_0 close to the nominal α = 0.05 or $0.04 \le \alpha \le 0.06$ is the best.

Under normal, contaminated normal and exponential distribution, Pearson's and Spearman's rank were found to give the best degree of association. While, under log-normal distribution, only Spearman's rank was found to give the best degree of association.

Table 4.6 shows all the proportion of rejecting H_0 under normal, log-normal, exponential and contaminated normal distributions and the method with range proportion of rejecting H_0 close to the nominal α = 0.05 or $0.04 \le \alpha \le 0.06$ is the best.

Under normal distribution, Pearson's was found to give the best degree of association , under log-normal distributions, Spearman's rank was found to give the best degree of association and under exponential and contaminated normal distributions, Pearson's and Spearman's rank were found to give the best degree of association.

Table 4.7 shows all the proportion of rejecting H_0 under normal, log-normal, exponential and contaminated normal distributions and the method with range proportion of rejecting H_0 close to the nominal α = 0.05 or $0.04 \le \alpha \le 0.06$ is the best.

Under normal, contaminated normal and exponential distribution, Pearson's and Spearman's rank were found to give the best degree of association. While, under log-normal distribution, only Spearman's rank was found to give the best degree of association.

From the figures, α = 0.01 to 0.1 were used for each of the figure and the results were displayed in the figures:

Figures 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8 and 4.9 show the graphical representation of Proportion of rejecting H_0 under Normal random distributions with n = (5,10,15 and 20) for both $\rho = 0$ and $\rho_1 = \rho_2$, Pearson's and Spearman have almost similar Proportion of rejecting H_0 ranging from $0.02 < \alpha > 0.08$ while the Kendall's has the Proportion of rejecting H_0 ranging from $0.00 \leq \alpha < 0.04$

Figures 4.10, 4.11, 4.12, 4.13, 4.14, 4.15, 4.16 and 4.17 show the graphical representation of Proportion of rejecting H_0 under Log-normal random distributions with n = (5,10,15 and 20) for $\rho = 0$ the Proportion of rejecting H_0 under Pearson's ranging from $0.02 > \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$. While $\rho_1 = \rho_2$, the Proportion of rejecting H_0 under Pearson's ranging from $0.04 < \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$

Figures 4.18, 4.19, 4.20, 4.21, 4.22, 4.23, 4.24 and 4.25 show the graphical representation of Proportion of rejecting H_0 under Exponential random distributions with n = (5,10,15 and 20) for $\rho = 0$ the Proportion of rejecting H_0 under Pearson's ranging from $0.02 > \alpha > 0.08$, Spearman's ranging from $0.02 < \alpha > 0.08$ and Kendall's ranging from $0.02 < \alpha < 0.04$. While $\rho_1 = \rho_2$, the Proportion of rejecting H_0 under Pearson's ranging from $0.02 < \alpha > 0.08$, Spearman's ranging from $0.02 \leq \alpha > 0.08$, and Kendall's ranging from $0.02 < \alpha < 0.04$

Figures 4.26, 4.27, 4.28, 4.29, 4.30, 4.31, 4.32 and 4.33 show the graphical representation of Proportion of rejecting H_0 under Contaminated random distributions with n = (5,10,15 and 20) for both $\rho = 0$ and $\rho_1 = \rho_2$, Pearson's and Spearman have almost similar Proportion of rejecting H_0 ranging from $0.02 < \alpha > 0.08$ while the Kendall's has the Proportion of rejecting H_0 ranging from $0.00 \le \alpha < 0.04$.

Hence, testing for a single and double correlation coefficient that is, $\rho = 0$ and $\rho_1 = \rho_2$ under normal, log-normal, exponential and contaminated normal distributions, the method with range proportion of rejecting H_0 close to the nominal α = 0.05 or $0.04 \le \alpha \le 0.06$ is the best.

Thus, under normal, exponential and contaminated normal distributions, Pearson's and Spearman's ranks have the best proportion of rejecting the true null hypothesis. While under log-normal distribution, only Spearman's rank correlation coefficient has the best proportion of rejecting the true null hypothesis.

CHAPTER FIVE

5.0 SUMMARY, CONCLUSION AND SUGGESTION FOR FURTHER STUDIES

5.1 Summary

In this research, different Poverty levels in Nigeria and different simulated data were used to compare different “spurious correlation tests” like Pearson’s, Spearman’s and Kendall’s correlation coefficients in order to obtain the method with the best degree of association among them since they all have different methodologies for statistical analysis. Meanwhile, the real live data used composed of poverty levels for all the thirty six (36) States of Nigeria including F.C.T based on their respective rates of Food poverty, Absolute poverty, Relative poverty and Dollar Poverty per day line based on World Bank’s Purchasing Power Parity (PPP) were analyzed and the results were all positive values directly and closed to unit value (1) which indicate high correlation between the Poverty levels in Nigeria in terms of Food Poverty, Absolute Poverty, Relative Poverty and Dollar Poverty per day. Hence, the *p-value* < 0.05 allows us to accept the value of Pearson’s, Spearman's and Kendall's rank correlation test calculated as being statistically significant.

Thus, simulated data used are normal, log-normal, exponential and contaminated normal random distributions were generated and the result of Pearson’s, Spearman's and Kendall's rank correlation test have been transformed and compared using t and z test statistics for testing correlation coefficients. Hence, from the result, it has been observed that when the data are normal, exponential and contaminated normal distributions, Pearson's and Spearman's rank have the best proportion of rejecting the true null hypothesis. But when the data are log-normal distribution, only Spearman's rank correlation coefficient has the best proportion of rejecting the true null hypothesis.

5.2 Conclusion

In this present study, the "spurious correlation test" Pearson's product-moment correlation test (PPMCT), Spearman's rank correlation test (SRCT) and Kendall's rank correlation test (KRCT) were all produced positive values and closed to unit value (1) which indicate high correlation between the Poverty levels in Nigeria in terms of Food Poverty, Absolute Poverty, Relative Poverty and Dollar Poverty per day. Hence, the $p\text{-}value < 0.05$ allows us to accept the value of Pearson's, Spearman's and Kendall's rank correlation test it's calculated as being statistically significant.

From simulated data, it is observed that when the data are normal, exponential and contaminated normal distributed, Pearson's and Spearman's rank have the best proportion of rejecting the true null hypothesis. But, when the data are log-normal distributed; only Spearman's rank correlation coefficient has the best proportion of rejecting the true null hypothesis. In conclusion, Pearson's and Spearman's rank have the best degree of association under normal, exponential and contaminated normal random distributions. While, for log-normal distribution only Spearman's rank has the best degree of association.

5.3 Suggestion for further Studies

1. Based on the outcome of this study, it was suggested that, if the data are normally distributed Pearson's product moment coefficient should be used to test the degree of association of that data.
2. If the data are contaminated normal or exponentially distributed, both Pearson's and Spearman's rank can be used to test the degree of association of that data.
3. Finally, if the data follow log-normally distribution, only Spearman's rank correlation coefficient should be used to test the degree of association of that data.

4. Since there is high correlation among Food poverty, Absolute poverty, Relative poverty and Dollar Poverty per day line based on World Bank's Purchasing Power Parity (PPP), it is highly recommended that Federal, State and Local Government Area's should provide Employment, security, Economy Growth and Price Stability for Nigerian's in order to reduce the poverty levels in Nigeria.
5. Lastly, these studies will serve as a gateway for further studies and future researchers who wish to engage in similar research work.

REFERENCES

Aczel , A. D. (1998). Complete business statistics *Richard D. Irvin, Inc. (US)* http:www.biblio.com/book/complete-business-statistics-aczel-amir-d/d/647867446

Aitchison, J. (1986). The Statistical Analysis of Compositional Data. *Monographs on Statistics and Applied Probability*. Chapman & Hall Ltd., London (UK). (Reprinted in 2003 with additional material by The Blackburn Press). 416.

Aldrich, J. (1995). Correlations genuine and spurious in Pearson and Yule. *Statistical science* **10** (4):364-376.

Bradley, J. V. (1978). Robustness British *journal of Mathematical and Statistical Pyschology*,**31,**144-152

Bravais, A. (1846). Analyse mathématique surles probabilités deserreurs de situation d'un point. *Mémoires présentés par divers savants à l'Académie Royale des Sciences de l'Institut de France* **9**: 255–332.

Chayes, F. (1960). On correlation between variables of constant sum: *J. Geophys. Res.* **65**, 4185- 4193

Daniels, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika* **33** (2): 129–135.

David, A. R. and Christopher, M. S. (2005). The influence of expert reviews on consumerdemand for experience goods: a case study of movie critics the *journal of industrial economics* **53**: 0022-1821

Eulalia, S. and Janusz , K. (2011). The Spearman and Kendall rank correlation coefficients between intuitionistic fuzzy sets 35- 45.

Galton, F. (1879). The geometric mean, in vital and social statistics. *Proceedings of the Royal Society of London **29**, 365–366.*

Gregory, A. F. and Roger, B. N. (2007). On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables *Journal of Statistical Planning and Inference **137:** 2143 – 2150*

Haan, C. T. (2002). *Statistical methods in Hydrology*, second edition: Ames, Iowa, Iowa State University Press.

Hans and Melberg. (2000). The nature and extent of spurious correlation and its implication for the philosophy of science with special emphasis on positivism 17-19

Jackson, J. B. and James, E. M. (2005). 'Type 1 error of four pairwise mean comparison procedures conducted as protected and unprotected tests' Journal *of Modern Applie statistical Method*,**4**(2):446-459

Jan, H. and Tomasz, K. (2011). Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data quaestiones geographicae **30**(2): 30-45

Johnson, N. L. and Kotz, S. (1992). *Distributions in Statistics: Discrete Multivariate Distributions*. John Wiley & Sons, New York.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika* **30**: 81–89.

Kendall, M. G. (1948). *Rank correlation methods*. 4th ed. Griffin, London.

Khamis, H. (2008). Measures of Association. How to Choose? *Journal of Diagnostic Medical Sonography*. **24**:155-162.

Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). *Continuous Multivariate Distributions.* John Wiley & Sons, New York.

Lovell, D., Pawlowsky, G. V. and Egozcue, .J. J. (2013). Don't correlate proportions!http://www.slideshare.net/AustralianBioinformatics/dont-correlate-proportions.

N. B. S. (2010). National Bureau of Statistics. www.nbs.gov.ng/2010/Nigeria Poverty Rate Profile.

Nelsen, R. B. (1992). On measures of association as measures of positive dependence. Statist. *Probab. Lett*. **14**: 269–274.

Nelsen, R. B. (2001). *Kendall tau metric. In: Hazewinkel, Michiel, Encyclopaedia of Mathematics, Springer*. 28-42.

Nian, S. C. (2010). Pearson's versus spearman's and kendall's correlation coefficients for continuous data. *Master's Thesis, University of Pittsburgh* 1-8

N. S. B. S. (2012). Niger State Bureau of Statistics. www.nigerstats.ni.gov.ng /2012/ Fact and Figure.

Pearson, K. (1897). Mathematical contributions to the theory of evolution*:* on a form of spurious correlation which may arise when indices are used in the measurements of organs: *Proc. Roy. Soc*. **60**: 489-98.

Pearson, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. 67-69.

Rodgers, J. L. and Nicewander, W. W. (1988). Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, **42**(1): 59 – 66.

Snedecor, G. W. and Cochran, W. G. (1989). *Statistical methods*, eighth edition, Iowa State University Press, Ames, Iowa, 803

Sorana, D. B. and Lorentz, J. (2006). Pearson versus Spearman, Kendall's Tau Correlation Analysis on Structure-Activity Relationships of Biologic Active Compounds Issue **9**:179-200

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology* **15** (1): 72–101.

Theodoros, P. Z., Patrick, J. M. and Christopher, C. P. (2011). Removal of Spurious Correlations Between Spikes and Local Field Potentials *Journal of Neurophysiol* 105: 474–486

Valz , P. D. and Thompson, M. E. (1994). Exact inference for Kendall's S and Spearman's rho. *Journal of Computational and Graphical Statistics* **3**: 459–472.

Weichao, Y. H., Hung, Y. S. and Yuexian, Z. (2010). Comparison of Spearman's rho and Kendall's tau in normal and contaminated normal models. *Manuscript submitted to IEEE Transactions on Information Theory*. 88-97.

(http://arxiv.org/PS_cache/arxiv/pdf/1011/1011.2009v1.pdf)

APPENDICES

APPENDIX A

This is the result of the analysis on Table 4.1 using difference Spurious Correlation Test for Testing Poverty Levels in Nigeria. Hence the Correlations and p-values are bolded.

Pearson's product-moment correlation tests

Pearson's product-moment correlation

data: Foodpoverty and Foodpoverty

t = Inf, df = 35, p-value **< 2.2e-16**

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

1 1

sample estimates:

cor

1

> cor.test(Foodpoverty, Absolutepoverty, method="pearson")

Pearson's product-moment correlation

data: Foodpoverty and Absolutepoverty

t = 9.3978, df = 35, p-value = **4.197e-11**

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.7195708 0.9184482

sample estimates:

cor

0.8462765

> cor.test(Foodpoverty, Relativepoverty, method="pearson")

Pearson's product-moment correlation

data: Foodpoverty and Relativepoverty

t = 8.0919, df = 35, p-value = **1.58e-09**

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.6544136 0.8967245

sample estimates:

cor

0.8072593

> cor.test(Foodpoverty, Dollarpoverty, method="pearson")

Pearson's product-moment correlation

data: Foodpoverty and Dollarpoverty

t = 9.2837, df = 35, p-value = **5.713e-11**

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.7145553 0.9168175

sample estimates:

cor

0.8433207

cor.test(Absolutepoverty, Absolutepoverty, method="pearson")

Pearson's product-moment correlation

data: Absolutepoverty and Absolutepoverty

t = Inf, df = 35, p-value < **2.2e-16**

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

1 1

sample estimates:

cor

1

> cor.test(Absolutepoverty, Relativepoverty, method="pearson")

Pearson's product-moment correlation

data: Absolutepoverty and Relativepoverty

t = 30.8487, df = 35, p-value < **2.2e-16**

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.9652439 0.9908224

sample estimates:

cor

0.9821029

> cor.test(Absolutepoverty, Dollarpoverty, method="pearson")

Pearson's product-moment correlation

data: Absolutepoverty and Dollarpoverty

t = 287.8364, df = 35, p-22value < **2.2e-16**

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.9995865 0.9998922

sample estimates:

cor

0.9997888

> cor.test(Relativepoverty, Relativepoverty, method="pearson")

Pearson's product-moment correlation

data: Relativepoverty and Relativepoverty

t = Inf, df = 35, p-value < **2.2e-16**

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

1 1

sample estimates:

cor

1

> cor.test(Relativepoverty, Dollarpoverty, method="pearson")

Pearson's product-moment correlation

data: Relativepoverty and Dollarpoverty

t = 31.4571, df = 35, p-value < **2.2e-16**

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.9665305 0.9911664

sample estimates:

cor

0.9827709

> cor.test(Dollarpoverty, Dollarpoverty, method="pearson")

Pearson's product-moment correlation

data: Dollarpoverty and Dollarpoverty

t = Inf, df = 35, p-value < **2.2e-16**

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

1 1

sample estimates:

cor

1

Spearman's rank correlation rho tests

cor.test(Foodpoverty, Foodpoverty, method="spearman")

Spearman's rank correlation rho

data: Foodpoverty and Foodpoverty

S = 0, p-value < **2.2e-16**

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

1

> cor.test(Foodpoverty, Absolutepoverty, method="spearman")

Spearman's rank correlation rho

data: Foodpoverty and Absolutepoverty

S = 989.2343, p-value = **4.976e-13**

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

0.8827366

> cor.test(Foodpoverty, Relativepoverty, method="spearman")

Spearman's rank correlation rho

data: Foodpoverty and Relativepoverty

S = 1492, p-value = **7.129e-08**

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

0.8231389

> cor.test(Foodpoverty, Dollarpoverty, method="spearman")

Spearman's rank correlation rho

data: Foodpoverty and Dollarpoverty

S = 1063.126, p-value = **1.633e-12**

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

0.8739775

> cor.test(Absolutepoverty, Absolutepoverty, method="spearman")

Spearman's rank correlation rho

data: Absolutepoverty and Absolutepoverty

S = 0, p-value < **2.2e-16**

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

1

> cor.test(Absolutepoverty, Relativepoverty, method="spearman")

Spearman's rank correlation rho

data: Absolutepoverty and Relativepoverty

S = 330.078, p-value < **2.2e-16**

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

0.9608727

> cor.test(Absolutepoverty, Dollarpoverty, method="spearman")

Spearman's rank correlation rho

data: Absolutepoverty and Dollarpoverty

S = 11.0039, p-value < **2.2e-16**

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

0.9986956

> cor.test(Relativepoverty, Relativepoverty, method="spearman")

Spearman's rank correlation rho

data: Relatipoverty and Relativepoverty

S = 0, p-value < **2.2e-16**

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

1

> cor.test(Relativepoverty, Dollarpoverty, method="spearman")

Spearman's rank correlation rho

data: Relativepoverty and Dollarpoverty

S = 341.0404, p-value < **2.2e-16**

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

0.9595732

> cor.test(Dollarpoverty, Dollarpoverty, method="spearman")

Spearman's rank correlation rho

data: Dollarpoverty and Dollarpoverty

S = 0, p-value < **2.2e-16**

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

1

Kendall's rank correlation tau tests

> cor.test(Foodpoverty, Foodpoverty, method="kendall")

Kendall's rank correlation tau

data: Foodpoverty and Foodpoverty

T = 666, p-value = **4.663e-15**

alternative hypothesis: true tau is not equal to 0

sample estimates:

tau

1

> cor.test(Foodpoverty, Absolutepoverty, method="kendall")

Kendall's rank correlation tau

data: Foodpoverty and Absolutepoverty

z = 6.0707, p-value = **1.274e-09**

alternative hypothesis: true tau is not equal to 0

sample estimates:

tau

0.6987984

> cor.test(Foodpoverty, Relativepoverty, method="kendall")

Kendall's rank correlation tau

data: Foodpoverty and Relativepoverty

T = 537, p-value = **9.209e-09**

alternative hypothesis: true tau is not equal to 0

sample estimates:

tau

0.6126126

> cor.test(Foodpoverty, Dollarpoverty, method="kendall")

Kendall's rank correlation tau

data: Foodpoverty and Dollarpoverty

z = 5.9912, p-value = **2.084e-09**

alternative hypothesis: true tau is not equal to 0

sample estimates:

tau

0.6887226

> cor.test(Absolutepoverty, Absolutepoverty, method="kendall")

Kendall's rank correlation tau

data: Absolutepoverty and Absolutepoverty

z = 8.6641, p-value < **2.2e-16**

alternative hypothesis: true tau is not equal to 0

sample estimates:

tau

1

> cor.test(Absolutepoverty, Relativepoverty, method="kendall")

Kendall's rank correlation tau

data: Absolutepoverty and Relativepoverty

z = 7.4052, p-value = **1.31e-13**

alternative hypothesis: true tau is not equal to 0

sample estimates:

tau

0.8524135

> cor.test(Absolutepoverty, Dollarpoverty, method="kendall")

Kendall's rank correlation tau

data: Absolutepoverty and Dollarpoverty z = 8.5972, p-value < **2.2e-16**

alternative hypothesis: true tau is not equal to 0

sample estimates:

tau

0.9909514

> cor.test(Relativepoverty, Relativepoverty, method="kendall")

Kendall's rank correlation tau

data: Relativepoverty and Relativepoverty

T = 666, p-value = **4.663e-15**

alternative hypothesis: true tau is not equal to 0

sample estimates:

tau

1

> cor.test(Relativepoverty, Dollarpoverty, method="kendall")

Kendall's rank correlation tau

data: Relativepoverty and Dollarpoverty

z = 7.3516, p-value = **1.958e-13**

alternative hypothesis: true tau is not equal to 0

sample estimates:

tau

0.8451137

> cor.test(Dollarpoverty, Dollarpoverty, method="kendall")

Kendall's rank correlation tau

data: Dollarpoverty and Dollarpoverty

z = 8.6873, p-value < **2.2e-16**

alternative hypothesis: true tau is not equal to 0

sample estimates:

tau

1

> pairs(clr(poverty),pch=".")

APPENDIX B

This is the program used for comparison of correlation coefficient tests using different sample sizes and α-values ranging from 0.01 to 0.1 while only the result of α = 0.05 are displayed in the tables.

```
n1= 5, 10, 15 and 20
n2= 5, 10, 15 and 20
sim=1000
tp=sim
tk=sim
ts=sim
zp=sim
zk=sim
zs=sim
for (i in 1:sim){
y= rnorm(n1,0,1)
x= rnorm(n1,0,1)
w= rnorm(n1,0,1)
m= rnorm(n1,0,1)
r1p=cor(x, y,method = c("pearson"))
r1k=cor(x, y, method = c("kendall"))
r1s=cor(x, y,method = c("spearman"))
r2p=cor(w, m,method = c("pearson"))
r2k=cor(w, m, method = c("kendall"))
r2s=cor(w, m,method = c("spearman"))
# t test for testing correlation
tp[i]=r1p*sqrt(n1-2)/sqrt(1-r1p^2)
tk[i]=r1k*sqrt(n1-2)/sqrt(1-r1k^2)
ts[i]=r1s*sqrt(n1-2)/sqrt(1-r1s^2)
# Z test for tesing two correlation coeff
z1p=1.1513*log10((1+r1p)/(1-r1p))
z1k=1.1513*log10((1+r1k)/(1-r1k))
```

```
z1s=1.1513*log10((1+r1s)/(1-r1s))
z2p=1.1513*log10((1+r2p)/(1-r2p))
z2k=1.1513*log10((1+r2k)/(1-r2k))
z2s=1.1513*log10((1+r2s)/(1-r2s))
d=sqrt(1/(n1-3)+1/(n2-3))
zp[i]=(z1p-z2p)/d
zk[i]=(z1k-z2k)/d
zs[i]=(z1s-z2s)/d
}
x1=qt(1-0.01,n1-2, lower.tail = TRUE, log.p = FALSE)
x2=qt(1-0.02,n1-2, lower.tail = TRUE, log.p = FALSE)
x3=qt(1-0.03,n1-2, lower.tail = TRUE, log.p = FALSE)
x4=qt(1-0.04,n1-2, lower.tail = TRUE, log.p = FALSE)
x5=qt(1-0.05,n1-2, lower.tail = TRUE, log.p = FALSE)
x6=qt(1-0.06,n1-2, lower.tail = TRUE, log.p = FALSE)
x7=qt(1-0.07,n1-2, lower.tail = TRUE, log.p = FALSE)
x8=qt(1-0.08,n1-2, lower.tail = TRUE, log.p = FALSE)
x9=qt(1-0.09,n1-2, lower.tail = TRUE, log.p = FALSE)
x10=qt(1-0.1,n1-2, lower.tail = TRUE, log.p = FALSE)
y1=qnorm((1-0.01),0,1,lower.tail = TRUE)
y2=qnorm((1-0.02),0,1,lower.tail = TRUE)
y3=qnorm((1-0.03),0,1,lower.tail = TRUE)
y4=qnorm((1-0.04),0,1,lower.tail = TRUE)
y5=qnorm((1-0.05),0,1,lower.tail = TRUE)
y6=qnorm((1-0.06),0,1,lower.tail = TRUE)
y7=qnorm((1-0.07),0,1,lower.tail = TRUE)
y8=qnorm((1-0.08),0,1,lower.tail = TRUE)
y9=qnorm((1-0.09),0,1,lower.tail = TRUE)
y10=qnorm((1-0.1),0,1,lower.tail = TRUE)
xp=c(mean(tp>=x1),mean(tp>=x2),mean(tp>=x3),mean(tp>=x4),mean(tp>=x5),mean(t
p>=x6),
mean(tp>=x7),mean(tp>=x8),mean(tp>=x9),mean(tp>=x10))
xk=c(mean(tk>=x1),mean(tk>=x2),mean(tk>=x3),mean(tk>=x4),mean(tk>=x5),mean(t
k>=x6),
```

```
mean(tk>=x7),mean(tk>=x8),mean(tk>=x9),mean(tk>=x10))

xs=c(mean(ts>=x1),mean(ts>=x2),mean(ts>=x3),mean(ts>=x4),mean(ts>=x5),mean(ts>
=x6),
mean(ts>=x7),mean(ts>=x8),mean(ts>=x9),mean(ts>=x10))
yp=c(mean(zp>=y1),mean(zp>=y2),mean(zp>=y3),mean(zp>=y4),mean(zp>=y5),mean
(zp>=y6),
mean(zp>=y7),mean(zp>=y8),mean(zp>=y9),mean(zp>=y10))
yk=c(mean(zk>=y1),mean(zk>=y2),mean(zk>=y3),mean(zk>=y4),mean(zk>=y5),mean
(zk>=y6),
mean(zk>=y7),mean(zk>=y8),mean(zk>=y9),mean(zk>=y10))
ys=c(mean(zs>=y1),mean(zs>=y2),mean(zs>=y3),mean(zs>=y4),mean(zs>=y5),mean(z
s>=y6),
mean(zs>=y7),mean(zs>=y8),mean(zs>=y9),mean(zs>=y10))
par(mfrow = c(2, 1),cex=0.6)
xx=seq(0.01,0.1,0.01)
plot(xx,xp,type="b",pch=2,ylim=c(0,0.11),ylab="Proportion of rejecting
Ho",xlab="Alpha Levels")
points(xx,xk,type="b",pch=15)
points(xx,xs,type="b",pch=17)
legend("topleft","n",legend=c("Pearson","Kendel","Spearman"),
cex=0.6, lwd=1,pch=c(2,15,17))
plot(xx,yp,type="b",pch=2,ylim=c(0,0.11),ylab="Proportion of rejecting
Ho",xlab="Alpha Levels")
points(xx,yk,type="b",pch=15)
points(xx,ys,type="b",pch=17)
legend("topleft","n",legend=c("Pearson","Kendel","Spearman"),
cex=0.6, lwd=1,pch=c(2,15,17))
pp=matrix(c(xp[5],xk[5],xs[5],yp[5],yk[5],ys[5]),6,1,
dimnames = list(c("Pearson","Kendel","Spearman","2Pearson",
"2Kendel","2Spearman"),c("Rate at alpha=0.05") ))
>Pp
```

APPENDIX C

POVERTY LEVELS DATA IN NIGERIA

The data: it was obtained from Nigeria poverty profile 2010 report by National Bureau of Statistics, composed of poverty levels for the all thirty six (36) States including F.C.T Nigeria.

State	Food Poverty	Absolute Poverty	Relative Poverty	Dollar Per day
Abia	30.5	57.4	63.4	57.8
Adamawa	55.4	74.2	80.7	74.3
Akwa ibom	35.6	53.7	62.8	53.8
Anambra	34.2	56.8	68	57.4
Bauchi	54.1	73	83.7	73.1
Bayelsa	23.3	47	57.9	47
Benue	48.5	67.1	74.1	67.2
Borno	33.2	55.1	61.1	55.1
Cross- Rivers	46.4	52.9	59.7	52.9
Delta	42.8	63.3	70.1	63.6
Ebonyi	63.5	73.6	80.4	73.6
Edo	39.4	65.6	72.5	66
Ekiti	35.8	52.4	59.1	52.6
Enugu	52.7	62.5	72.1	63.4
Gombe	71.5	74.2	79.8	74.2
Imo	33.3	50.5	57.3	50.7
Jigawa	71.1	74.1	79	74.2
Kaduna	41.7	61.5	73	61.8
Kano	48.3	65.6	72.3	66
Katsina	56.2	74.5	82	74.8
Kebbi	47	72	80.5	72.5
Kogi	50.1	67.1	73.5	67.3
Kwara	38.1	61.8	74.3	62
Lagos	14.6	48.6	59.2	49.3
Nasarawa	26.8	60.4	71.2	60.4
Niger	20.4	33.8	43.6	33.9
Ogun	41.8	62.3	69	62.5
Ondo	36.1	45.7	57	46.1
Osun	19.5	37.9	47.5	38.1
Oyo	24.6	51.8	60.7	51.8
Plateau	44	74.1	79.7	74.7
Rivers	26.3	50.4	58.6	50.6
Sokoto	56.6	81.2	86.4	81.9
Taraba	45.2	68.9	76.3	68.9
Yobe	58.5	73.8	79.6	74.1
Zamfara	44.4	70.8	80.2	71.3
FCT	32.5	55.6	59.9	55.6