Diagnostic Test on Ecotourism Development and Welfare of Host Community in Doguwa Local Government Kano, Nigeria

Dr. Ali Adamu Naniya; Abubakar Mohammad and Garba Isma'ila Muhammad

Kano State Polytechnic, kano, Nigeria

Abstract	Original Research Article

After performing the necessary data editing and data cleaning, and confirming the validity and reliability of the instruments, the study subjected the data to further diagnostic test of: normality, linearity, no multi-collinearity and homogeneity tests, to evaluate the accuracy, reliability and appropriateness of data before running the parametric statistics. The study used cross sectional, descriptive and correlation designs. The data were collected using none standardize instruments. The study arrived at sample size of 399 through Sloven's formula. The study used both probability and non-probability of purposive, stratified and simple random. Similarly, in ascertaining internal consistency and trustworthiness, the findings from reliability tests using Cronbach alpha arrived at Cronbach alpha coefficients above 0.7, and validity tests using content validity index with CVIs above 0.7, KMO above 0.7, Bartlets test sig less than 0.05, determinants greater than 0.000, communalities above 0.5 and rotated component matrices discriminated and loading highly on distinct factors. Moreover, the data were analysed at univariate, bivariate and multivariate levels. The findings from diagnostic test reveal that data are appropriate for the conduct of parametric statistics of correlation and regression.

Keywords: Diagnostic test, normality test, linearity test, no multi-collinearity test and homogeneity test.

1.0 Introduction

After performing the necessary data editing and data cleaning, and confirming the validity and reliability of the instruments, the study subjected the data to further diagnostic test of: normality, linearity, multi-collinearity and homogeneity tests to evaluate the accuracy, reliability, and effectiveness of data before running the parametric statistics. According to Panneerselvam (2014), a normal distribution "is a continuous probability distribution" (P. 183). Thus, normality **test** is a statistical procedure used to determine whether a dataset follows a normal distribution. This is important concept in diagnostic test because many commonly used statistical methods such as **t-tests,** ANOVA, and regression analysis, assume that the data is normally distributed. If this assumption is violated, the results of these tests may be misleading (Arker, Kumar &Day, 2005). Thus, normality in this study was tested using histogram, P-P Plot, Q-Q Plot, descriptive statistics and Kolmogorov, which established the data was normal.

Naniya, A. A., Mohammad, A., & Muhammad, G. I. (2026). Diagnostic test on ecotourism development and welfare of host community in Doguwa Local Government Kano, Nigeria. *SSR Journal of Arts, Humanities and Social Sciences*, 3(3), 52-59.

52

Similarly, according to Loftus and Loftus (1988), a linear relationship "is assumed to relate two variables, for instance, Xand Y" (P.444). Thus, it is the property of a relationship between two variables, where the change in one variable is directly proportional to the change in the other variable either positively or negatively (Loftus & Loftus, 1988, P.

444). This relationship can be represented by a straight line on a scatter plot. Moreover, linearity is a key assumption for many statistical techniques, including correlation and regression analysis, and is important for ensuring the accuracy and reliability of quantitative analytical methods. In this study, linearity was tested using scatter plots, correlation matrix and regression (ANOVA).

Moreover, multi-collinearity occurs when two or more independent variables in a regression model are highly correlated. According to Kothari (2004), multi-collinearity refers to "situation where there is a high degree of correlation between independent variables" (P 142). In this study, multicollinearity was tested using variance inflation factor and tolerance value to find out the extent to which independent variables components are highly correlated.

Finally, a homogeneity test is a statistical procedure used to determine the equality of variances for a variable calculated for two or more groups (Aaker, Kumar & Day, 2005). Some common statistical procedures assume variances of the population from which different samples are drawn are equal (Zikmund & Sabin, 2000). Therefore, Lavene's test assesses this assumption to check whether the variances across different groups are equal and share similar characteristics. In this study, homogeneity test was conducted using Lavene's test to establish the homogeneity of variance. However, this paper will focus on the former (normality test).

2.0: Literature Review

Normality Test

According to George and Mallery (1999), "many naturally occurring phenomena produce distributions of data that approximate a normal distribution" (P. 2020). A normality test is a statistical procedure used to determine whether a dataset follows a normal distribution. A normal distribution is a "continuous probability distribution" (Panneerselvam, 2014, P. 183). Thus, a normal distribution is symmetric about the mean, which is based on the assumption that the data follows a normal distribution. This is important because many commonly used parametric statistical methods, such as t-tests, ANOVA, and regression analysis — assume that the data is normally distributed. If this assumption is violated, the results of these tests may be misleading. Moreover, a normality test evaluates the shape of the data's distribution and compares it to the expected shape of a normal distribution. If the result shows significant deviation that shows that the data is not normal. Therefore, it plays an important role in ensuring that the statistical methods used give accurate and trustworthy results. Thus, if this assumption is not met, the results from these methods can be misleading or incorrect. The five most common type of normality tools are:

(a) Histogram

According to Aaker, Kumar and Day (2005), a histogram" is a series of rectangles, each proportional in width to the range of values within a class and proportional in height to the number of items falling in the class" (P. 428). Similarly, a histogram is a graphical view of the given data such that the frequency of each range of the variable of interest is shown as adjacent vertical strip against range of the variable in proportion to the heights of other such vertical strips (Panneerselvam, 2014, P. 54). Thus, a histogram is a histogram is a type of graph used in research and statistics to visually represent the distribution of numerical data and estimate the probability distribution of a continuous variable. If the graph is approximately bell-shaped and symmetric about the mean, we can assume normally distributed data.

(b) A P–P plot (probability–probability plot or percent–percent plot)

A normal probability plot is extremely useful for testing normality assumptions. There are two

versions of normality probability plots: QQ and PP. A P–P plot (probability–probability plot) is a graphical technique for assessing how closely two data sets (observed and expected) agree. It forms an approximate straight line when data are normally distributed. Departures from this straight line indicate departures from normality. Thus, a P-P Plot is a statistical tool used to visually assess how well two probability distributions align are useful for testing normality of the data. Moreover, the p-p plots are vastly used to evaluate the skewness of distribution (Singh, 2021).

(C) Q-Q Plot

A **Q-Q plot** (Quantile-Quantile plot) is a graphical tool widely used in research to check whether a dataset follows a particular theoretical distribution, most commonly the normal distribution. Thus, it is a scatterplot created by plotting two sets of quantiles (observed and expected) against one another.

(c) Descriptive Statistics

Descriptive statistics are a fundamental part of research because they help summarize, organize, and present data in a clear and meaningful way (George & Mallery, 1999). According to Singh (2021), descriptive Statistics refers to a "discipline of quantitatively describing the main features of a collection of data aim at summarizing a sample, rather than making inference about the entire population" P. 12). Therefore, a distribution of responses, going by theory distribution, should have normal distribution from zero. Thus, that coincides with Aaker, Kumar and Day (2005), as "descriptive statistics normally associated with a frequency distribution that helps summarise the information presented in the frequency table" (P. 429). Thus, a descriptive statistics (skewness and Kurtosis) shows whether the data is normally distributed as the value of skewness and Kurtosis is not far away from zero (Field, 2009). That was also in conformity with Pallant (2001) who states any value that is not greater than zero is not far away from zero. This was concurred by (George and Mallery, 1999). Moreover, it is only when skewness and Kurtosis are closer to zero,

represents a symmetric or evenly balanced distribution (P. 87).

(e) Kolmogorov-Smirnov

Parametric test assumes that data from samples are normally distributed. Therefore, Kolmogorov "tests whether the distribution of members of a single group differ significantly from a normal distribution (George and Mallery, 1999, p.194). Although there are various methods for normality testing, however, for small sample size (n <50), Shapiro–Wilk test should be used. Moreover, when the sample size (n) is at least 50, Kolmogorov–Smirnov test for instance, is appropriate. Thus, the Kolmogorov–Smirnov test (K–S test) is a widely used nonparametric statistical test that plays an important role in research, especially when analyzing distributions.

3.0: Methodology

In achieving the study objectives, the study used cross sectional, descriptive and correlation designs, employing both qualitative and quantitative approaches. Similarly, considering the nature of the study hypothesis, qualitative and quantitative nature of study variables, Study measurement, the data were collected using both questionnaires and interviews. The sample size of 399 was generated using Sloven's formular. In considering the acceptability and reliability of the instruments, the study used face, content and construct validity. Moreover, ascertaining internal consistency reliability Cronbach alpha was used. The study subjected the data to further diagnostic test of: normality, linearity, no multi-collinearity and homogeneity, to evaluate the accuracy, reliability, and effectiveness of data before running the parametric statistics, using SPSS software. Finally, the data were analysed at univariate, bivariate and multivariate levels.

4.0 Findings:

The study analyses the data and presents the findings based on the objectives of the study, using normality, linearity, no multi-collinearity and homogeneity tests. The findings are presented below:

a) Histogram

A histogram is a fundamental tool in research for visualizing the distribution of numerical data. It helps researchers quickly identify patterns, trends, and anomalies in datasets using graphical representation as follows:

Figure 4.1: Histogram
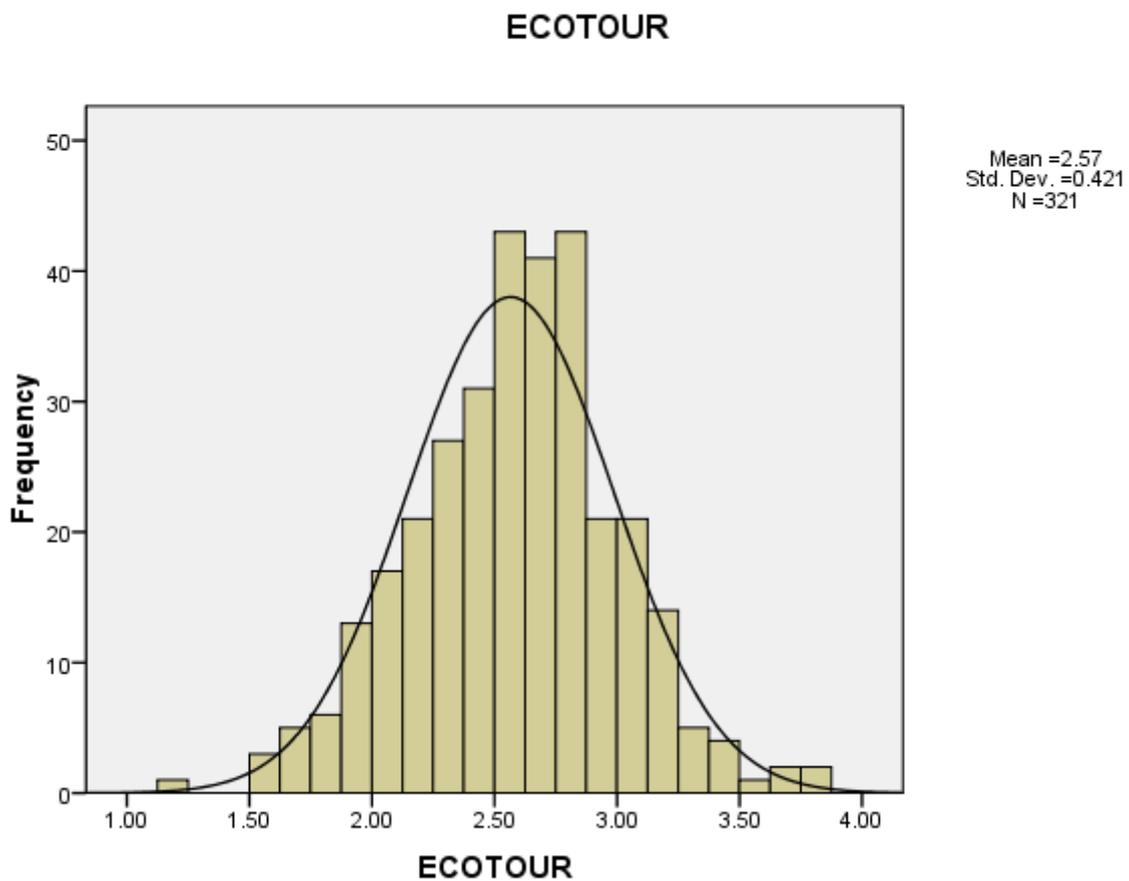
Looking at the overall pattern displayed by the shape and spread of distribution, the histogram shows that the data are normally distributed.

b) **P-P Plots**. Here it shows the positive direction of the relationship where most of the points are located near the straight line, which show the normality of the data.

Figure 4.2: P-P-Plots


Normal P-P Plot of WELFHOST
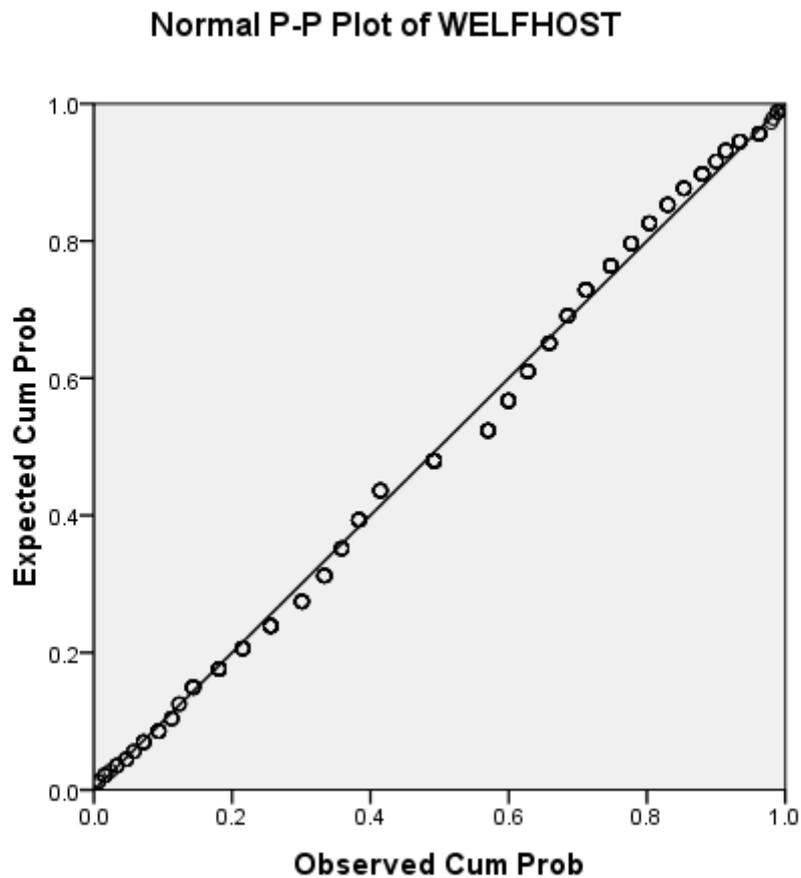
c). Q-Q Plots

A Q-Q plot (Quantile-Quantile plot) is a graphical tool widely used to check whether a dataset follows a particular theoretical distribution, most commonly the normal distribution. Thus, a Q–Q plot is a scatterplot created by plotting two sets of quantiles (observed and expected) against one another.

Q-Q Plots

Normal Q-Q Plot of ECOTOUR
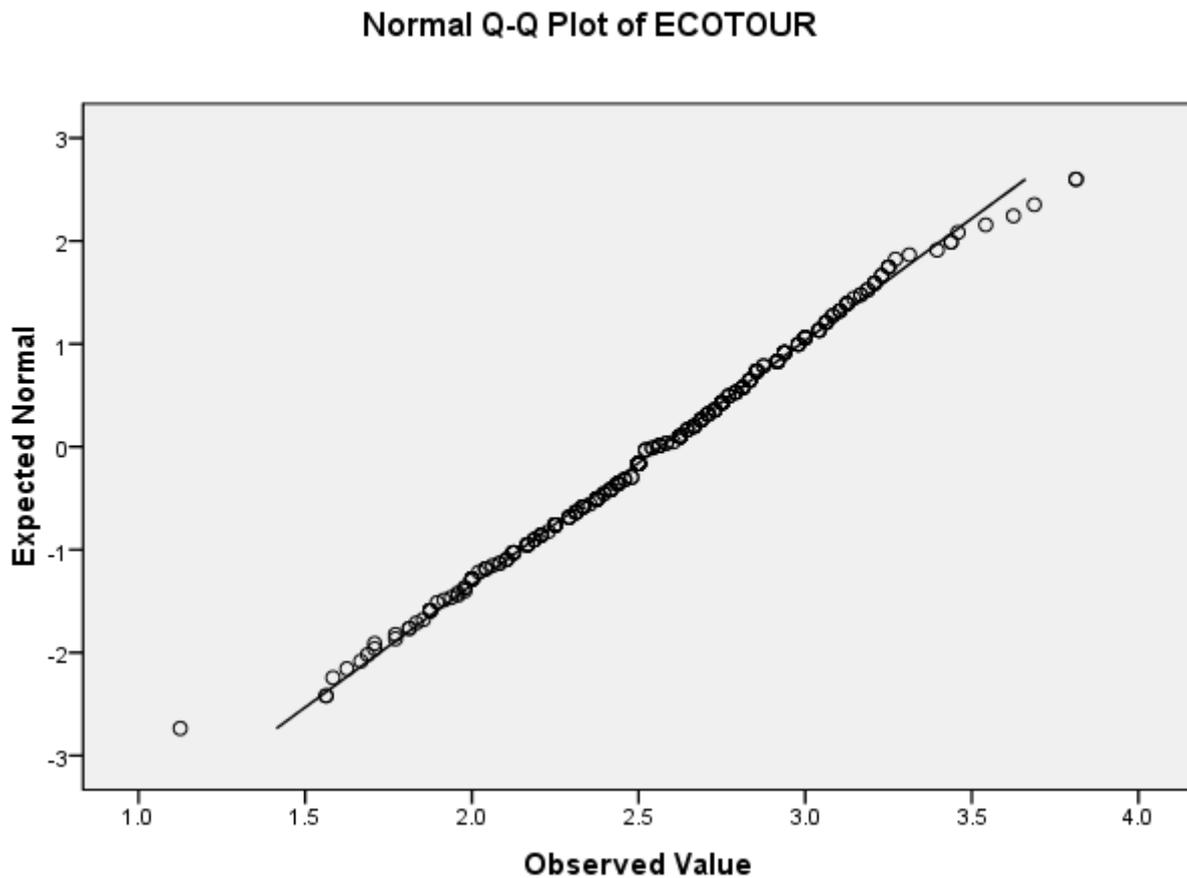
Thus, the data displays above in QQ Plots shows that the data is normally distributed as most points are located near straight line.

d). Descriptive Statistics

Table 4.11: Descriptive Statistics

When checking whether data follow a normal distribution, descriptive statistics play a crucial role before applying formal statistical tests. Here's how they fit into the process:

Table 4.1: Descriptive Statistics

Variables	Mean Statistic	Std. Deviation Statistic	Skewness Statistic	Std. Error	Kurtosis Statistic	Std. Error
Accommodation	2.4838	.63673	.723	.136	2.983	.271
Transportation	2.5822	.51687	-.133	.136	-.530	.271
Infrastructure	2.2966	.51237	.008	.136	-.509	.271
Marketing	2.7701	.48216	.039	.136	.640	.271
Employment	2.4977	.66358	.953	.136	2.978	.271
Income	2.6680	.75922	-.393	.136	-.748	.271

The table 4.1 of descriptive statistics shows that the data on skewness and kurtosis are all normal as manifesting that all their values are not far away from zero. This was in conformity with Pallant (1994) who states that any value that is not greater than 3 is closer to zero.

e). Kolmogorov-Smirnov

Table 4.2: Kolmogorov-Smirnov

Variables	Kolmogorov-Smirnov[a] Statistic	Df	Sig.	Shapiro-Wilk Statistic	df	Sig.
Welfare	.075	321	.000	.991	321	.063
Ecotourism	.051	321	.039	.996	321	.615

a. Lilliefors Significance Correction

In this study, considering the sample size greater than 200, the data was tested using Kolmogorov Smirnov with sig value of greater than 0.05 as seen in table 4.2 which testifies that data was fairly normally distributed for ecotourism and welfare.

5.0: Conclusion

The findings from normality test reveal that the study can safely use further parametric tests manifesting in the overall pattern displayed by the shape and spread of distribution, and the positive direction of the relationship on the P P

and Q Q plots where most of the points are located near the straight line. Moreover, the findings from descriptive statistics reveal that the normality of data as the values of skewness and kurtosis are not far away from zero. Finally, considering the sample size greater than 200, the data was tested using Kolmogorov Smirnov with sig value of greater than 0.05 which confirms that data was normally distributed.

References:

Field, A. (2009). Discovering statistics using SPSS (3rd Ed.). Sage Publication.

Kothari, C. R., (2004). Research Methodology. New Age International, India.

Loftus, G.R., & Loftus, E.F., (1988). Essence of statistics. Alfred A. Knopf, India.

Pallant, J. (2001). Survival MaNual, Open University Press.

Panneerselvam, R., (2014). Research Methodology. Industrial Estate, India.

Singh, R. N. (2021). Statistics. Random Publications, India.

Zikmund, W. G. & Babin, B. J., (2000). Marketing Research. Tenth Edition, South Western, Centgage Learning, India.