International Journal of Commerce and Management Research
ISSN: 2455-1627, Impact Factor: RJIF 5.22
www.managejournal.com
Volume 3; Issue 5; May 2017; Page No. 188-190

Assessing Normality of the Data: A Case Study of FMCG Sector in India

Sahil Narang

Assistant Professor in Commerce, MM PG College, Fatehabad, Haryana, India

Abstract
This study has been designed to check normality of the data set and to sort out the problems related to normality of data set. Skewness, Kurtosis, Histogram, Stem and Leaf plot, Q-Q plot and Shapiro-Wilk test have been used to check the normality of data sets related to FMCG sector. The study helps the readers to sort out the problems related to the normality of the data sets.

Keywords: Skewness, Kurtosis, Histogram

Introduction
Normality is one of the most important concerns in field of research and statistics. It is very imperative to know about the normality status of the data, whether the random variable fit into the Normal distribution or not, in order to analyze the data and to make rational decisions. Normal distribution is the most commonly used distribution in the field of statistics and decision making. It is very important to assess the normality of the data that a researcher is going to analyze in order to decide that what kind of the test and the statistical tool should be applied in order to analyze the data and to reach the rational decision in the concerned area.

Relevance of Assessing Normality
- **To know about the statistical tool to be used for the purpose of study:** There are a wide variety of the statistical tools available for the purpose of data analysis and to reach at any decision by the way of drawing inferences. Every tool has its core property and the property can only be use at a specific type of the data. Few of the tools such as the One sample Z-test, One sample t-test, two-way ANOVA, One-way ANOVA and Pearson's Co-efficient of correlation can only be used on the parametric or the normal data but on the other hand Spearman's Rank Correlation and other statistical tools such as Wilcoxon Signed Rank test, Friedman test, Kruskal-Wallis test, Mann-Whitney test are mainly used in case when the data is non-parametric.
- **To ensure the reliability of decision making:** If the a random variable distribution is normal it will be easy to draw the inference to the basis of the particular type of the data but if data is of non-normal nature then it is not so easy to predict the future values of very and to draw inferences (Abnormal events are difficult to predict).

Review of Literature
Few of the previous studies have been reviewed for the purpose of the present study: S. S. Shapiro; M. B. Wilk (1965) in their research work titled: "An Analysis of Variance Test for Normality (Complete Samples) " According to Shapiro and Wilk assumption testing is one of the most important task in the statistics and the field of research. Shapiro and Wilk has used a variety the statistical tools in order to test the normality of the data and to sort the related problems. Use of W test has been made for assessing Normality with respect to the complete samples. Plackett's "a," approximations have also been used in order to assess the normality of the variances with respect to the data. E. S. Pearson and M. A. Stephens (1964) in their study titled: "The Ratio of Range to Standard Deviation in the Same Normal Sample" In the paper of 1964 and of the 1954 with David and Hartley Pearson discussed about the distribution of the ratio that is equal to u = w/s of (a) and the range w in thea sample of n number of observations from a normally distributed population having standard deviation SD. They introduced a table of few of the upper and the lower points in parentage terms of the ratio and discussed its use in assessment of normality of data. Johnson, Nixon, Amos & Pearson (1963) of standardized the points (%) of Pearson curves more thoroughly and accurately. Thomas Lumley, Paula Diehr, Scott Emerson, and Lu Chen (2002) in their research work titled: "The Importance of the Normality Assumption in Large Public Health Data Sets" attempted to measure and the sort out the problems related to the public health. They majorly emphasized the normality assessment of the large public health data sets. The research conducted in the department of Bio-statistic, Thomas, Paula and Scott concluded that it is a misconception that t-statistics and the linear regression are useful only in case of normal or parametric data. They found that beside the few exceptional cases t-statistics and linear can be used for any kind of data distribution with the large sized samples. They suggested the use of Wilcoxon test and ordinal logistic regression in the cases where the hetrosedasticity is found in the data.

A Case Study of FMCG Sector in India
Monthly return data of various FMCG companies has been used for the purpose of preparing illustration. Same methodology can be used to test the normality of the data of the prices and the returns of any company in FMCG sector. Monthly return on the closing prices of the company's shares has been taken for the purpose of assessing the normality of the data set.

Tools for Assessment of Normality
Use of different tools has been made in order to illustrate the assessment of normality, to test and assess the normality, the

data sets of ATFL and other FMCG companies has been used and to support further data analysis. The same kind of the tools and methodology can be used to test the normality of the prices and returns on other companies in different sectors of the capital markets India and of the other international markets as well. The details of the tools have been given below:

Skewness and Kurtosis: Skewness is one of the most important and most commonly used tool for testing the normality assumptions of any data set. Skewness is the measure of the departure from the symmetry and it can be categorized into different categories: the left tailed or the negatively skewed distribution and the right skewed distribution.

Set of Data

The data set includes the returns computed on the monthly closing prices the share of ATF Limited there are a total no. of the 60 observations, with no missing value, that has been studied in the study in order to test the assumption of normality for the case of single variable data set. The case processing summary has been given in the table 1 which has been given below:

Table: 1: (Case Processing Summary)

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Atfl	60	100.0%	0	0.0%	60	100.0%

*Data compiled through SPSS

Let us see how to check out: Whether a data set is normal or not normal.

The table 2 shows the values of the skewness and the standard error with reference to the value of skewness (as shown in the corresponding cell).

Table 2: (Descriptive Statistics)

			Statistic	Std. Error
Atfl	**Mean**		**.00177**	**.00744**
	95% Confidence Interval for Mean	Lower Bound	-.0131	
		Upper Bound	.0166	
	5% Trimmed Mean		.0010	
	Median		-.0076	
	Variance		.003	
	Std. Deviation		.05813	
	Minimum		-.1104	
	Maximum		.1329	
	Range		.2433	
	Interquartile Range		.0878	
	Skewness		.368	.306
	Kurtosis		-.503	.604

*Data is compiled through SPSS.

Skewness and Kurtosis both are used in case of the interval and ratio level data. Values of the skewness and kurtosis are zero if the data set of the observed items is exactly normal. A positive value of the skewness indicates the positively skewed and negative values of the skewness indicate the negatively skewed data set of the left tailed data set. While positive values of the Kurtosis indicate that the data is peaked and the Leptokurtic distribution has been followed and negative values of the Kurtosis indicates that the data set is of Plato-kurtic nature. Both of the types of the distributions: the Lepto-kurtic and the Platy-Kurtic are considered as the non-normal distributions while Meso-kurtic or the zero value of the Kurtosis is considered as the normal distribution. If the value of the skewness is positive. Here the value of the skewness is positive 0.368 and the value of Kurtosis is negative 0.503. It demonstrates that the observed data set is not normal.

Komlogorov-Smirnov and Shapiro-Wilk Tests of Normality

Komlogorov-Smirnov and **Shapiro-Wilk** Tests of Normality can also be used to test the normality of the data set. The Komlogorov-Smirnov test (with Lilliefors significance level) is constructed with the normal probability and de-trended probability plots in order to test the assumption of normality for the observed set of data. The data set is considered to be normal if significance level of the test is greater than 0.5. The results of the test have been shown in the table 3.

Set of Data

Use of the total 60 observations of the monthly returns of the Bombay Burmah Trading Corporation has been made in order to test illustrate the assessment of Normality of the data set using Komlogorov-Smirnov and Shapiro-Wilk test of Normality.

Hypothesis of the Study

Hypothesis for the purpose of the study are given below:

Null Hypothesis (H0): Monthly returns of the BBTC are normally distributed;

Alternative Hypothesis (H1): Monthly returns of the BBTC are not normally distributed.

Table 3: (Tests of Normality)

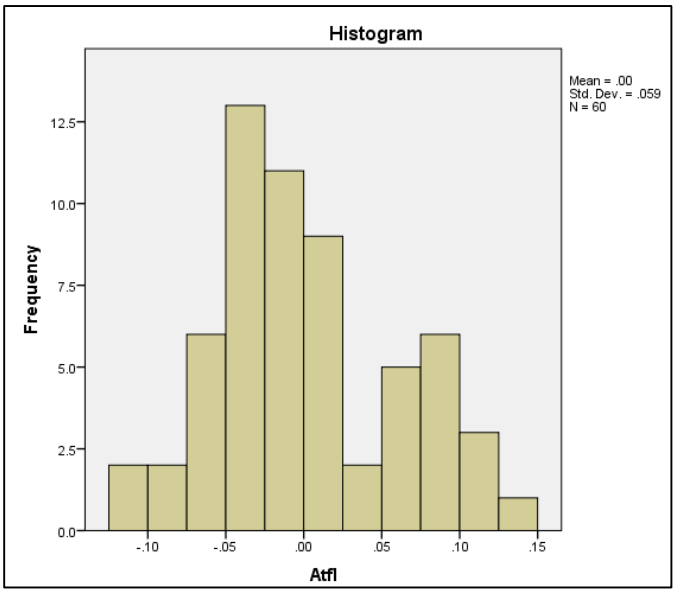
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	df	Sig.
BBTC	.112	60	.054	.963	60	.064
a. Lilliefors Significance Correction						

Table 3 reveals the results of the Komlogorov-Smirnov normality test and Shapiro and Wilk normality test. The value of the significance level is greater than 0.05. Therefore it can be concluded that that data is normal in nature. The same results can be verified with the help of the Shapiro and Wilk test if number of the observations is less than 100. The table shows the value of Shapiro and Wilk test is also greater than 0.05. Therefore the assumption of the normality has been checked up. The data is normally distributed.

Histogram

Histogram has been popularly used in the cases when the researcher wants to know about the shape of the distribution. The values on the vertical axis indicate the frequency of the data value/cases and the values of the horizontal axis indicate the mid-values of the data interval or the data ranges. Use of a data set of the ATFL, an FMCG company, has been made in order to illustrate assessment of normality using histogram. Figure 1 is the histogram of the monthly returns of the ATF

limited for the last five years. The total number of the observations is 60.



Avanti Stem-and-Leaf Plot

Stem and leaf plot is very similar to the histogram. It can be said that it is the more accurate and the improved version of the histogram is explain the exact frequency with respect to a data set and a better tool than the histogram. The stem represents the graph corresponding to the first digit of the score and leaf represents the trailing of the digit. The shape of plot represents the normality of the data set. Following is the stem and leaf plot of the Avanti Feeds Limited.

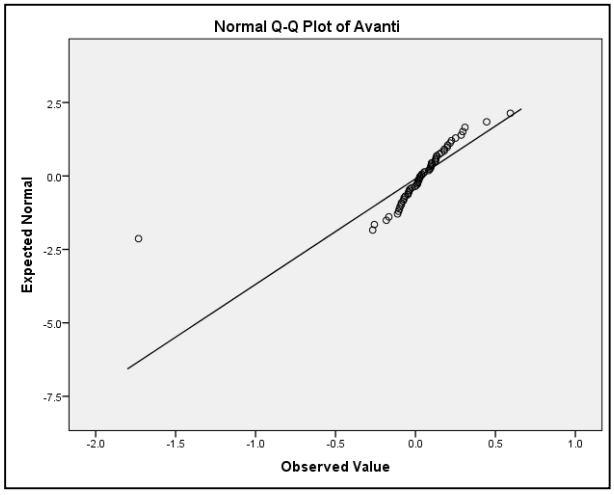
Avanti Stem-and-Leaf Plot

Frequency	Stem & Leaf
1.00 Extremes	(=<-1.73)
2.00	-2. 56
5.00	-1. 00168
14.00	-0. 02334446777889
17.00	0. 01112222345588999
12.00	1. 002223346889
6.00	2. 012589
1.00	3. 0
2.00 Extremes	(>=.45)
Stem width:	.10
Each leaf:	1 case(s)

Normal Q-Q Plot of Avanti Feeds

Q-Q plot refers to the comparison of the observed values of the data set with the expected value of the data corresponding to the observed values. A straight trend line shows the expected values of the data set. if the dotted line coincides the expected values then data is exactly normally distributed. Proximity indicates the extent of the normality of the data set. Given is the QQ plot of the Avanti Feeds Limited a leading FMCG company and a constituent of FMCG index of BSE. The data set includes 60 observations for the monthly returns of the company for last five years.

There are a large variety of the tools that can be used of for the purpose of testing the normally of the set of the data. Only suitable tools, with due care, should be used for the case specific and there must be a careful interpretation of the results of the normality tests through different tools of the normality testing.



Conclusion and Findings

Assessment of the normality testing is very important and the initial step in the process of data analysis and decision making. There are different solutions to the problem when data is found non-normal. Someone may make use of the non-parametric statistical tools for the purpose of data analysis; someone may use natural logarithms in order to sort out the problem of normality and to attain normality of the data set. While use of parametric tests can be used on the non-paramedic of the non-normal data as well if size of the sample is very large (Thomas Lumley, Paula Diehr, Scott Emerson, and Lu Chen, 2002).

References

1. Razali, Nornadiah; Wah, Yap Bee. Power comparisons of Shapiro–Wilk, Kolmogorov–Smirnov, Lilliefors and Anderson–Darling tests (PDF). Journal of Statistical Modeling and Analytics. Archived from the original (PDF) on 2015-06-30. 2011; 2(1):21-33.
2. Judge, George G, Griffiths WE, Hill R, Carter, Lütkepohl, *et al.* Introduction to the Theory and Practice of Econometrics (Second ed.). Wiley. 1988; 890-892. ISBN 0-471-08277-5.
3. Gujarati, Damodar N. Basic Econometrics (Fourth ed.). McGraw Hill. 2002; 147-148. ISBN 0-07-123017-3.
4. Lin CC, Mudholkar GS. A simple test for normality against asymmetric alternatives. Biometrika. 1980; 67(2):455-461. doi:10.1093/biomet/67.2.455. Retrieved.
5. Mardia KV. Measures of multivariate skewness and kurtosis with applications. Biometrika. 1970; 57:519-530.
6. Epps TW, Pulley LB. A test for normality based on the empirical characteristic function. Biometrika. 1983; 70: 723-726.
7. Henze N, Zirkler B. A class of invariant and consistent tests for multivariate normality. Communications in Statistics - Theory and Methods. 1990; 19:3595-617.
8. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965; 52(34):591-
611. doi:10.1093/biomet/52.34.591. JSTOR 2333709. M R 205384. 593