VOL. 11, NO. 18, SEPTEMBER 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

ASSESSING NORMALITY FOR DATA WITH DIFFERENT SAMPLE SIZES USING SAS, MINITAB AND R

Anwar Fitrianto[1,2,3] and Low Yen Chin[1]
[1] Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, UPM Serdang, Selangor, Malaysia
[2]Laboratory of Computational Statistics and Operations Research, Institute for Mathematical Research, Universiti Putra, Malaysia
[3]Department of Statistics, Faculty of Mathematics and Natural Resources, Bogor Agricultural University, Indonesia
E-Mail: anwarstat@gmail.com

ABSTRACT

Different statistical packages may produce different results of statistical analysis such as normality test. One of possible sources of the difference is the computational approach. This study tries to explore results of normality tests based on different statistical packages. Empirical data with varied sample sizes were conducted. It was found that SAS, Minitab, and R produced different conclusion in normality test. Meanwhile, sample size also has effect on the test of normality where larger sample size tends to produce different conclusion of normality. But, all the three statistical packages produced similar results of normality test for AD and KS tests.

Keywords: normality test, statistical package, anderson-darling, kolmogorov-smirnov, saphiro-wilk.

INTRODUCTION

In statistics, normality refers to lack of significant deviation from the average and the fact of being normal. Normality is a measure of how well an observed distribution approximates to a normal distribution. Most of statistical procedures such as *t*-test and linear regression analysis require assessing the assumption of the normality. Normality should be tested before interpreting results of a statistical analysis since the inference may not be valid if the normality is violated. The easiest way for the normality test is using graphical methods such as histogram, normal quantile-quantile plot (Q-Q plot) and box-plot.

But sometimes graphical method to assess normality is not enough since the visual assessment are subjective especially when the pattern is not clear. There are many formal approaches of normality tests to support the graphical method and provide conclusive evidence that the normal assumption holds. There have been many research conducted of normality test in literatures. Nornadiah and Wah (2010) found that Shapiro-Wilk (SW) test is the most powerful normality test against all alternatives and Kolmogorov-Smirnov (KS) test is the least powerful test in all situations. According to them, the performance of Anderson-Darling (AD) test is comparable with SW test and Lilliefors (LF) test which always excel the KS test. So, the selection of different normality test should be given attention based on the sample size and selected distribution.

There are few normality tests available in statistical packages. In this study, the focus is on the comparisons of normality tests such as SW, KS and AD test based on few available statistical packages such as MINITAB, SAS, and R-Language. The performances of those tests are compared based on a numerical example. Different sample sizes will be determined by simple random sampling from the original data. The objectives of the study are: 1) to evaluate the performances of Anderson-Darling test (AD), Kolmogorov-Smirnov test (KS), and Shapiro-Wilk test (SW) for testing normality by using different statistical packages, 2) to compare normality tests on small, moderate and large sample data set, and 3) to compare the skewness and kurtosis on small, moderate and large sample data set.

LITERATURE REVIEW

Normal distribution is a continuous probability distribution with parameter mean, μ and standard deviation, σ. Normal distribution is symmetrical with a single central peak at the mean of the data. It forms a bell-shaped curve, a shape that is created when a line is plotted to fit the data points. A normal distribution is typically the ideal in research and data that every research strives for. Keskin (2006) mentioned that Type I error rate and power of normality tests may be difficult to assess due to many possibilities when choosing a particular alternative hypothesis for those samples with increased sample size. The performance of those normality tests are strongly related with distribution type and sample size of that sample data. Jason (2010) stated that data transformation is used for improving the normality of the observed distribution and equalizing variance to meet the normality assumption while preparing for statistical analyses since the interpretation and inferences are not valid if the normality assumption is violated. Graphical methods such as histogram and Q-Q-plots are used to test the normality assumption. Even though the graphical method can serve as a useful tool in checking normality, they are still not sufficient to provide conclusive evidence that normal assumption holds. So, numerical method is used to support the graphical method in normality testing. Nornadiah and Wah (2010) conducted a study on comparing the tests of normality via Monte Carlo simulation of data generated from alternative distribution that follow symmetric and asymmetric distribution with respective critical values.

Anderson-darling (AD) test

Anderson-Darling test (AD) is a general test to compare the fit of an observed cumulative distribution function to an expected cumulative distribution function. It

ARPN Journal of Engineering and Applied Sciences

is used to test a sample of data from a population which from a specific distribution. It is a modification of the Kolmogorov-Smirmov test (KS) and gives more weight to the tails than KS test. Critical value of the AD test is based on the specific distribution being tested. The test statistic for AD test as following (Equation 1),

$$W_n^2 = -n - \frac{1}{n}\sum(2i-1)\left\{\log F^*(x_i) + \log\left\{1 - F^*(x_{n+1-i})\right\}\right\} \quad (1)$$

where $F^*(x_i)$= cumulative density function of the specified distribution and x_i are ordered observations. Arshad *et al.* (2003) discussed about the parameters of generalized Pareto distribution which were estimated by probability of weighted moment's method and critical points.

Kolmogorov-smirnov (KS) test

Kolmogorov-Smirnov test (KS) is based on the comparisons between the observed and expected frequencies which are estimated using *z*-scores. The KS test is a nonparametric test used to compare a sample with reference probability distribution. The test statistic for KS tests as follows (Equation 2):

$$D = \max\left\{\max\left(\left|D^+\right|, \left|D^-\right|\right)\right\} \quad (2)$$

where:

$$D^+ = \max_{1\le i\le n}\left\{\left(\frac{i}{n}\right) - \hat{z}_i\right\}, \text{and}$$

$$D^- = \max_{1\le i\le n}\left\{\hat{z}_i - \left(\frac{i-1}{n}\right)\right\}.$$

Related to the KS test, Mendes and Pala (2003) investigated and evaluated few normality tests for Type 1 error rate and their power on normality testing. They concluded that KS had the smallest rejection rates and it should be used with strong caution when test the normality. They mentioned Lilliefors Test (LF) is different with Kolmogorov-Smirnov test because the parameter estimated which result in different decisions, but the formula of test statistic is the same.

Shapiro-wilk (SW) test

Shapiro-Wilktest (SW) was first developed by Shapiro and Wilk (1965) and it is the most powerful and omnibus test in many situations. Unfortunately, it depends on the correlation between given data and their corresponding normal scores. The test statistic of SW test is written as follows,

$$W = \frac{\left\{\sum_{i=1}^{n} a_i x_{(i)}\right\}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad (3)$$

where $x_{(i)}$ are ordered statistics and n is number of observation. In the Equation 3, the W statistic requires that the sample size is greater than or equal to 7 and less than or equal to 2,000 (Shapiro and Wilk 1965).

There have been studies related to the SW test. A research conducted by Srivastava and Hui (1987) extended the application of SW statistic to test the multivariate normality without any approximation. They proposed two test statistics for multivariate normality which depend on principal components. It can also be considered as generalization of SW statistics. Mendes and Pala (2003) found that the SW test is most powerful test in most situations. Turk (2006) found that the SW test can also be used for practical purposes and it achieves sufficient power at small sample sizes except for t distribution. Nornadiah and Wah (2010) claimed that small value of test statistic of SW test tends to the rejection of normality. Meanwhile, Nor *et al.* (2011) evaluated the performance of normality tests under different kind of non-normal distribution and different sample sizes. They concluded that SW test is a best performing normality test since it rejects the null hypothesis at smallest sample size data when comparing with KS test, AD test and Cramer-von Mises test (CVM).

Skewness and kurtosis coefficients

Skewness is a measure of symmetry and asymmetry of a data set and how symmetric the data set is about the mean. The value of skewness can be positive or negative, or even undefined. Jean-Marie *et al.* (1998) investigated that deviation from normality can be assess by sample moments because the moment test are derived from recognition that the 3rd and 4th moments of ideal normal distribution are equal to 0 and 3. For calculating skewness and kurtosis, we use the following equations (Equation 4, and Equation 5):

$$skewness = \frac{n\sum(x-\bar{x})^3}{(n-1)(n-2)s^3} \quad (4)$$

And

$$kurtosis = \frac{n(n+1)\sum(x-\bar{x})^4}{(n-1)(n-2)(n-3)s^4} - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (5)$$

where $s = \frac{\sum(x-\bar{x})^2}{n-1}$.

 ISSN 1819-6608

ARPN Journal of Engineering and Applied Sciences

The use of statistical packages for normality testing

It is much easier way to use a computer statistical package to solve some mathematics problems compared to using hand calculation. Existing statistical packages allow us to change a set of data and can observe the impact of those changes on the data immediately and it will save much time when comparing with hand calculation. There are different normality test available in each different statistical packages. The R-Language, SAS and MINITAB provide almost all the major normality test.

Table-1. Availability test of normality in statistical packages.

Normality test	Statistical packages		
	R	SAS	MINITAB
KS	Yes *	Yes	Yes
SW	Yes	Yes	Yes *
AD	Yes	Yes	Yes
JB	Yes	No	Yes

The * at first row means that the test statistic for Lillefors (LF) test is considered the same with KS test although LF test is different from KS test since the parameter are estimated in R-Language. The * at the second row means that the test statistic for Ryan Joiner (RJ) test is considered similar to the Shapiro-Wilk (SW) in MINITAB. The test statistic of RJ test is written in the following Equation (6):

$$R_J = \frac{\sum x_{(i)} b_i}{\sqrt{s^2 (n-1) \sum b_i^2}} , \tag{6}$$

where $x_{(i)}$ is ordered statistics and b_i is normal score of the ith order.

Rob and Yanan (1996) summarized sample categories by using some major statistical packages such as MINITAB and SAS. But, they only relied on commands line to compute quantiles and ignore hidden quantile definition that is used in probability plot since different packages use different definition of sample quantile to plot the probability plot when computing the quantiles. Hun (2008) used graphical methods such as histogram and numerical method assess measure of skewness and kurtosis for univariate analysis and normality tests. He illustrated the way of normality testing by using SAS, STATA and SPSS. Meanwhile, Jason (2010) discussed about normalizing data through transformations and the way how the Box-Cox improved on normalizing data. Meanwhile, Umaporn (2011) studied on the efficiency comparison of selected normality tests by using the statistical packages such as SPSS and MINITAB. He found that RJ Test which considers as SW test in MINITAB had the highest power in all cases and sample sizes and it was able to control probability of Type I error.

NUMERICAL EXAMPLE

A numerical example about university admissions which is available in Michael *et al.* (2005) is used in this study. The data set is about the director of admissions at a state university wanted to determine how accurately student's grade point average (GPA) at the end of their freshman year could be predicted by the entrance examination score (ACT Score) and the high school rank. The academic year which freshman enters the university covers year 1996 to year 2000. There are 705 students in total and their identity is represented by using identification number. There are 3 variables in the data set which we going to analysis such as GPA, high school class rank and ACT score.

The performance of KS, AD, and SW test on the small, moderate and large sample size were conducted by using SAS, R-Language and MINITAB. Sample sizes of 20, 50, and 250 observations from the dataset were taken to represent small, medium, and large sample size. Selecting random sample was conducted by using MINITAB.

Descriptive statistics of the data

Assessing normality assumption is the important step before we proceed to further statistical procedures like linear regression analysis and discriminant analysis. The interpretation will be reliable when the normality assumption is not violated. Normal probability plot is used to investigate whether the certain data set follows a normal distribution. Cumulative density function will be plotted on a straight line and the mean and standard deviation are calculated from the data. Then the data points are plotted along the fitted normal line. Normal probability plot for the numerical example with small size sample is displayed in Figure-1.

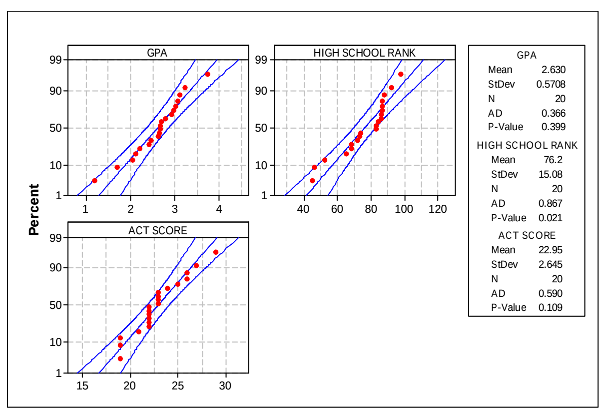


Figure-1. Normal probability plots of the data for small sample size using Minitab.

The Figure-1 reveals that the two variables, namely GPA and High School Rank for the small sample size are not normally distributed. But, the Act Score seems

VOL. 11, NO. 18, SEPTEMBER 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

more normally distributed based on visual assessment. The probability plot for the GPA is right skewed since the plotted points are curl up and to the left of the normal line which indicates a long tail to the right size. Another variable, ACT Score is assumed as normal since the plotted points are fit to the normal line.

Different graphical methods were used to display distribution of data with different sample size. Figure-2 displays histogram for the sample data for moderate sample size. Variable GPA and High School Rank positively skewed since the histogram is slightly skewed to the right. On the others hand, the variable ACT score seems as symmetric since its histogram are in the middle and it does not skewed to any side.

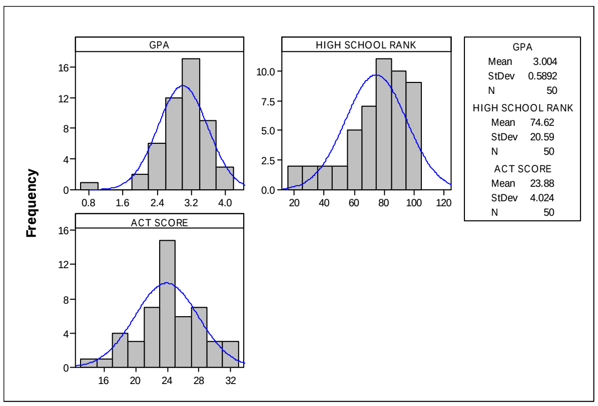


Figure-2. Histograms of the data for moderate sample size using Minitab.

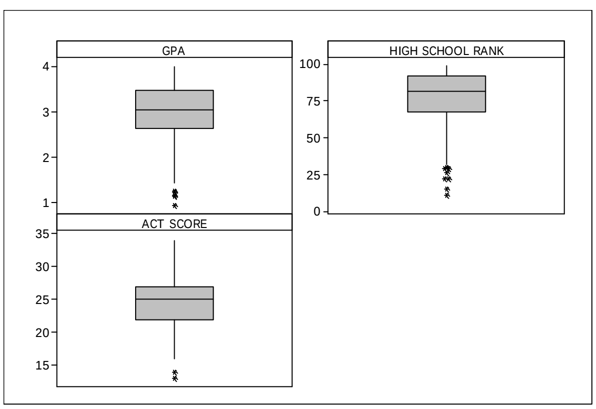
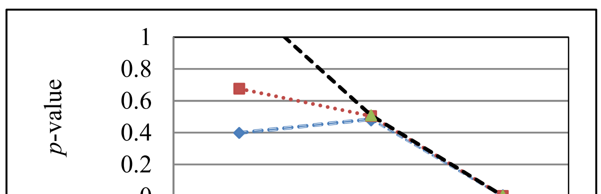


Figure-3. Boxplots of the data for large sample size using Minitab.

For the sample data with sample size, we use boxplot to graphically assess the normality (Figure-3). Boxplot is an easier way to summarize a set of data measured based on an interval scale and it is used to show the shape of the distribution. The boxplot can provide information about the normality and skew of the data. Based on the boxplot we can clearly see that the variable GPA and High School Rank are left skewed distribution since the box is shifted significantly to the high end. The variables ACT Score is assumed as normal since that box is centered between the whiskers

Comparisons of normality tests on the numerical example

Normality data were assessed by using AD, KS, and SW test. Summarize of the *p*-value of normality test of the variables for all samples based on different statistical software and different sample sizes are displayed in Figure-4. All the *p*-values of KS test for all variables are approximately equal although different statistical packages are used such as variable High School Rank. There is similar phenomenon can be observed for other variables in different normality test and statistical packages. All tests based on all statistical packages consistently produce the very low *p*-value for large sample. It is obvious that all variables of large sample size are not normally distributed. To observe the effects of sample sizes on normality test easily, we can focus on observing the *p*-values produced by certain software. For our discussion, let's focus on normality test for variable CGPA which is produced by R in Figure-4. The *p*-value of variable GPA for all samples which we obtained by using R. From Figure 4, we look at the *p*-values of variable GPA decreases as the sample size increases AD, SW and KS test. It means that normality is tended to be concluded as the simple size increases.



	20	50	250
SW	0.5808	0.004461	3.72E-06
KS	0.2782	0.01945	0.003129
AD	0.3988	0.4851	4.26E-05

Figure-4. Plot of *p*-value against sample size by using R for variable GPA.

According to Table-2, all the tests show that for small sample size, all variables are approximately normally distributed. The SW test and AD test produces higher test statistic value while the KS test produces lower test statistic value.

For moderate sample, all the tests consistently signify the normality of variables GPA, High School Rank and ACT Score.

Table-2. Test Statistics Values of Variables in SAS, MINITAB and R.

Sample size	Variable	Statistical packages								
		SAS			MINITAB			R		
		AD	KS	SW	AD	KS	SW	AD	KS	SW
20	GPA	0.366	0.150	0.962	0.366	0.150	0.974	0.366	0.150	0.962
	H.S.R	0.867	0.224	0.895	0.867	0.224	0.951	0.867	0.224	0.895
	ACT.S	0.590	0.192	0.934	0.590	0.192	0.988	0.591	0.193	0.934
50	GPA	0.747	0.137	0.928	0.747	0.137	0.961	0.747	0.137	0.928
	H.S.R	1.420	0.148	0.901	1.424	0.148	0.957	1.423	0.148	0.901
	ACT.S	0.390	0.108	0.985	0.390	0.108	0.997	0.390	0.108	0.985
250	GPA	1.996	0.072	0.962	1.996	0.072	0.984	1.996	0.072	0.962
	H.S.R	6.750	0.123	0.897	6.750	0.123	0.955	6.750	0.123	0.897
	ACT.S	0.866	0.074	0.989	0.866	0.074	0.998	0.866	0.074	0.989

Comparisons of skewness and kurtosis

The graphical method sometimes can't provide sufficient evidence that the normality assumption holds. Skewness and kurtosis coefficients are another way to access the normality assumption. The different statistical packages will have the different definition on the skewness and kurtosis. SAS and MINITAB define that if the skewness and kurtosis value close to zero, then the data could be assumed as normal distribution. Another statistical package which will be discussed in this paper such as R-Language gives a normal distribution a skewness of 0 and a kurtosis of 3.

Table-3. Skewness and kurtosis values for all variables.

Sample Size	Variable	SAS		MINITAB		R	
		SK	KU	SK	KU	SK	KU
20	GPA	-0.659	1.253	-0.660	1.250	-0.609	3.675
	H.S.R	-0.886	-0.020	-0.890	-0.02	-0.818	2.699
	ACT	0.496	0.224	0.500	0.220	0.458	2.886
50	GPA	-1.128	2.576	-1.130	2.500	-1.094	5.208
	H.S.R	-1.070	0.619	-1.070	0.620	-1.037	3.441
	ACT.S	-0.114	0.047	-0.110	0.050	-0.111	2.925
250	GPA	-0.631	0.245	-0.700	0.250	-0.691	3.217
	H.S.R	-1.153	1.034	-1.150	1.030	-1.146	3.990
	ACT.S	-0.147	-0.363	-0.150	-0.360	-0.147	2.620

Results for calculation of skewness and kurtosis coefficients are displayed in Table-3. It is found that the kurtosis values for all variables are similar for SAS and MINITAB, but relatively different if we use R. For the small sample size, the skewness value for GPA and High School Rank is between -0.5 to -1.0 which means the distribution is moderately right skewed. The variable ACT Score for small sample size is very close to symmetric distribution with the skewness value between -0.5 to 0.5. By looking on their kurtosis value, it can be concluded that the variable GPA and ACT Score have heavier and thicker tails since the kurtosis value is positive. For moderate sample, the variable GPA and High School Rank are highly right skewed and the variable ACT Score is normal distributed. It can be concluded that the variables GPA, High School Rank and ACT Score have heavier and thicker tails by looking on their kurtosis value. And for large sample size, variable ACT Score is approximately normal by looking on its skewness value which close to 0. The variable GPA and High School Rank have heavier and thicker tails since the kurtosis value is positive. Otherwise, the variable ACT Score has a lighter and thinner tails with lower peak than normal since its kurtosis value is negative.

CONCLUSIONS

In this paper, three types of normality tests are compared using SAS, MINITAB and R-Language. Performance of the selected normality (AD, KS, and SW) test using numerical example on university admission data were conducted. The comparisons for the normality were done against small sample size, moderate sample size and large sample size. It was found that regardless the statistical package, for small sample size, the results of normality test were varied where the SW test tended to produce higher *p* values compared to the KS and AD tests. But, for medium and large sample size, all the tests produced approximately the same results.

Once the statistical package were considered, it was found that all the three statistical packages produced similar results for AD and KS tests. But, for the SW test, Minitab tended to produce higher SW statistics compared to the other two software.

REFERENCES

Arshad, M., Rasool, M. T. and Ahmad, M. I. 2003. Anderson Darling and Modified Anderson Darling Tests for Generalized Pareto Distribution. Pakistan Journal of Applied Sciences, 3(2), pp. 85-88.

Currell, C. and Dowman, A. 2006. Essential Mathematics and Statistics for Science: Wiley, West Sussex, U.K.

Hun, M.P. 2008. Univariate Analysis and Normality Test Using SAS, Stata, and SPSS*. Ph.D. Thesis, Indianan University, India.

Jason, W.O. 2010. Practical Assessment, Research and Evaluation. Practical Assessment, Research and Evaluation, 15(12), pp. 1-9.

Jean-Marie, D., Abdeljelil, F., Lucien, G., and Lynda, K. 1998. Simulation-based Finite Sample Normality Tests in Linear Regressions. Econometrics Jorna, 1, pp. 154-173.

Jushan, B. and Serena, N. 2005. Tests for Skewness, Kurtosis, and Normality for Time Series Data. American Statistical Association Journal of Business &Economic Statistics. 23(1), pp. 49-60.

Keskin, S. 2006. Comparison of Several Univariate Normality Tests Regarding Type1Error Rate and Power of the Test in Simulation based Small Samples. Journal of Applied Science Research, 2(5), pp. 296-300.

Massey, F. J., Jr. 1951. The Kolmogorov-Smirnov test of goodness of fit, J. Amer. Statist. Ass. 46, pp. 68-78.

Mendes, M., and Pala, A. 2003. Type 1 Error Rate and Power of Three Normality Tests, Pakistan Journal of Information and Technology, 2(2), pp. 135-139.

Michael, H.K., Christopher, J.N., John,N. and William, L. 2005. Applied Linear Statistical Models: McGraw-Hill/Irwin, New York.

Michael, J.C. 2007. The R Book: John Wiley and Sons Ltd, United Kingdom.

Nor, A. A., Teh, S.Y., Abdul, R.O. and Che, R.Y. 2011. Sensitivity of Normality Tests to Non-normal Data. Sains Malaysian, 40(6), pp. 637-641.

Nornadiah, M.R., Yap, B.W. 2010. Power Comparison of Some Selected Normality Tests. Proceedings of the Regional Conference on Statistical Sciences (RCSS'10): 126-138.

Rob, J. H. and Yanan, F. 1996. Statistical Computing. The American Statistical, 50(4), pp. 361-365.

Shapiro, S.S. and Wilk, M.B. 1965. An Analysis of Variance Test for Normality (Complete Samples). Biometrika, 52(3/4), pp. 591-611.

Srivastava, M.S., Hui, T.K. 1987. On Assessing Multivariate Normality Based on Shapiro-Wilk W Statistics. Statistic and Probability Letters, 5, pp. 15-18.

Tanweeer-Ul-Islam. 2011. Normality Testing - A New Direction. International Journal of Business and Social Science, 2(3), pp. 115-118.

Umaporn,C. 2011. Efficiency Comparisons of Normality Test Using Statistical Packages. Thammasat Int. J. Sc. Tech, 16(3), pp. 9-25.

Yosef, C. and Jeremiah, Y.C. 2008. Statistics and Data with R: An applied approach through examples: John Wiley and Sons Ltd, United Kingdom.