

# Exploring Postgraduate Students' Research Knowledge and Skills in Normality Tests and Verifications

Clement Ayarebilla Ali, Peter Akayuure  
University of Education, Winneba, Ghana

The study explored research knowledge and skills of postgraduate students in testing and verifying normality of data in order to boost their confidence and credibility of educational research findings. This exploratory survey randomly sampled 66 postgraduate students, out of about 150 postgraduate students in five faculties of the University of Education, Winneba, Ghana. The results of Statistical Package for Social Sciences (SPSS) Version 16 revealed that the postgraduate students require these innovative skills in order to test and verify their educational research data. We therefore recommended continuous inter-faculty collaborations and regular research conferences participations by teaching staff and postgraduate students to beef up their knowledge and skills in research analyses.

*Keywords:* descriptive statistics, normality tests and verifications, parametric tests, research knowledge and skills

## Introduction

Russell (2006) described the two main types of analysis as descriptive and inferential. The descriptive analysis involves graphical displays, central tendencies, and dispersions while the inferential analysis involves hypothesis testing and verifications. And in exploring research knowledge and skills in statistics, Russell (2006) added that physical, educational, intellectual, and occupational characteristics must be taken of and included in both statistics. Particularly, Joreskog and Sorbom (2004) suggested that the four important characteristics of postgraduate students' research knowledge and skills are demographic, descriptive, normality, and verifications of tests.

Apart from the demographic, studies of Myoung (2008), Elnabris (2011), and Avioli (2012) show that the descriptive, normality, and verification tests can be assessed with the normal distribution. The normal distribution is a bell-shaped curve, in which the mean, median, and mode are equal, and are located at the centre. Also, it is symmetric about the mean, continuous, asymptotic to the horizontal, and has a total area of 1.00 or 100% (50% lies to the left and the other 50% to the right of the mean).

Narrowing down the context to inferential tests, the two main forms are parametric and nonparametric. Parametric tests require assumptions while nonparametric do not. However, nonparametric require even stronger evidence to reject the null hypothesis (Spector, 2004; GraphPad, 2007; Harmon, 2011; Ali, Amihere, Adzifome, & Ngman-Wara, 2014). If these tests are non-normality, then the data either have outliers, multiple modes, incorrect measuring tools, incorrect distributions, zero/infinite limits, or scanty collections.

### Normality Tests

Assumptions and technicality for the application of statistical tools and suitability of the tests are important in research work and normality is one of the most important aspects for statistical analysis (Singh & Masuku, 2014). Normality and verification tests are divided into three categories—graphical, numerical, and verifications. The commonest graphical methods are histogram, box plot, and quintile plots (probability-probability [P-P] or quantile-quantile [Q-Q]). For small or moderate sized data sets, the box plot is more suitable, and for large data sets, the histogram or polygon is more appropriate. The quintile plots can be used for any sample side of data (Spector, 2004; Russell, 2006; GraphPad, 2007; Ghasemi & Zahadisi, 2012). Myoung (2008), Horst (2010), Elnabris (2011), and Motulsky (2013) contended that normality of the data is determined by the shape of the histogram, the outliers in the box plot, and overlapping of straight lines in the Q-Q plot.

Russell (2006) and Elnabris (2011) named the numerical ones as measures of central tendency (mean, median, and mode), dispersions (inter-quartile range, range, and standard deviation), and positions (skewness and kurtosis). In the normal data, the values of the mean, median, and mode are the same; the inter-quartile range is approximately 1.33 standard deviations, the range is approximately 6 of the standard deviations, and the number of observations within 1, 2, and 3 standard deviations falls in the empirical rule of 68%, 95%, and 99.7% or errors of 1.64, 1.96, and 2.58 respectively. Russell (2006), Myoung (2008), Horst (2010), and Elnabris (2011) explained that a perfectly normal distribution has zero skewness (negative is left skewed and positive is right), and that of perfectly normal distribution has zero statistic (flat top is platykurtic and thin is leptokurtic).

Myoung (2008), Horst (2010), Arthur (2011), Elnabris (2011), Ghasemi and Zahadisi (2012), and Shafer and Zhang (2012) agreed that the most appropriate verification tests are Kolmogorov-Smirnov, Shapiro-Wilk, and Anderson-Darling tests verifications of normality. Even though Anderson-Darling is the best, it is mostly unavailable in Statistical Package for Social Sciences (SPSS) software that most postgraduate students use. And, while Kolmogorov-Smirnov is best for exploring independence, Shapiro-Wilk is most preferred for both independence and identically distributed small samples.

### Statement of the Problem

Normality tests and verifications should be taken seriously, for when these are not clarified, it may result into inaccurate and unreliable conclusions about the reality (Oztuna, Elhan, & Tuccar, 2006). Ali et al. (2014) observed that in exploring normality tests and verifications to research data, errors and mistakes are often detected in the presentation, interpretation, labelling, computation, and discussion of results. It is not uncommon to observe the pie charts being compared with each other, when bar charts and percentages are suitable for such comparisons. The researchers observed that postgraduate students have difficulties in exploring and reporting normality tests and verifications to their research data. The commonest tools often applied by postgraduate students were bar charts, pie charts, and percentage frequency tables. Again, even though some postgraduate students clearly stated mixed research methodologies, the analyses depicted otherwise. The research topics, statements, research questions, and tools of analysis never matched the research protocol for mixed methodologies. And where some students tested their research questions and hypotheses well, they failed to test and verify normality, due to possible constraints in the testing and verification processes.

Worse still, some postgraduate students concluded the statistical significances of their results but never made any efforts to justify such decisions. It must be stressed that rejecting or accepting the null hypotheses

have never been any issues in modern educational research designs. The most importance issues now centre on the justifications for rejecting or failing to reject the hypotheses.

Just like other universities and research institutions, the University of Education, Winneba, has revised the Postgraduate Research Methods and Theses Reporting with a clear demand for normality tests and verifications of research data and theses reporting. Thus, the continuous disregard and violations of this research demand could hinder strong focus, inference, trust, and policy direction. The need to explore various techniques postgraduate students could employ in testing and verifying normality of research data is warranted.

### **Purpose of Study**

The purpose of this study was to explore the techniques of testing and verifying for normality of research data. Data were drawn from the postgraduate students for the purpose of illustrating various techniques of normality testing and verification.

### **Research Question**

In order to show various techniques of normality testing and verification, the following research question was formulated: How do gender, highest academic levels, and type of programme influence postgraduate students' research knowledge and skills in normality tests and verifications?

### **Tests of Hypothesis**

Null hypothesis ( $H_0$ ): Postgraduate students' research knowledge and skills in normality tests and verifications are the same.

Alternative hypothesis ( $H_a$ ): Postgraduate students have different research knowledge and skills in normality.

Level of significance ( $\alpha$ ): 5%.

Rejection criteria: Reject  $H_0$  when  $p$ -values are less than  $\alpha$ .

### **Methodology**

The exploratory survey design was used to assess and explore postgraduate students' knowledge and skills in normality tests and verifications of their research results. The postgraduate students' population across five faculties of the University of Education, Winneba, was estimated at 150. The simple random sampling procedure was employed to select 66 postgraduate students, on the assumptions of large population, homogeneity, and identically distributed in that academic year. Three demographic variables were gender, highest academic levels, and faculty of study programme. The study variables were graduate students' knowledge of types of statistics, measurement scales, data forms, and instruments of data collection. The others were their understanding in confidence levels, numerical, theoretical, and graphical tests. The rest were their challenges in presentation, interpretation, labelling, computation, and discussion of research results.

A simple coded questionnaire was designed with 13 items to avoid verbose, monotonous, and repeated responses from the students. The items were made simple and straight to the points to ensure that the relevant information as well as the quality of the responses was explicitly and adequately understood and responded to. It further ensured that the required information was obtained within the period, under the same conditions of all the groups of students. Permission was sought from the Office of Graduate School and individual students. Additionally, the researchers collected the data from graduate students who were enrolled in the 2013-2014 academic year and had started writing their theses. The traditional true-score theory and item-response test theory

provided us appropriate frameworks for assessing reliability and internal validity of the instruments (Ghasemi & Zahadial, 2012). The various coefficients of associations were used to assess external validity of 70%. Out of the 200 questionnaire distributed, about 66 were recovered due to high attrition in the forms of non-compliance and persistent absenteeism. However, these attritions may not affect the generalizability of our findings.

The compositions of the samples were from the Faculties of Science (18), Education (18), Social Science (9), Languages (9), Creative Arts (9), and Institute for Educational Development and Extension (3). These sample sizes might have affected the hypotheses and compelled the two-tailed level of significance. On gender, the researchers recovered 64 male and 36 female students. On highest academic achievement of the graduate students, the data comprised 85% four-year and 15% two-year undergraduate certificates. The other types of certificates anticipated never featured. The main tools of data analyses were histogram, Q-Q, P-P, box plot, and central tendencies. The others were dispersions, positions, Kolmogorov-Smirnov, and Shapiro-Wilk tests (Best & Kahn, 2014; Ghasemi & Zahadial, 2012; Russell, 2006).

## Results

Figure 1 shows the histogram of graduate students and their academic faculties in the university. We observed that the histogram was symmetric, despite slightly skewed to the right. However, the histogram only counts number of categories, compares shape, and assesses similarity of bar widths and centres, but does not depict large differences in precise normality of data.

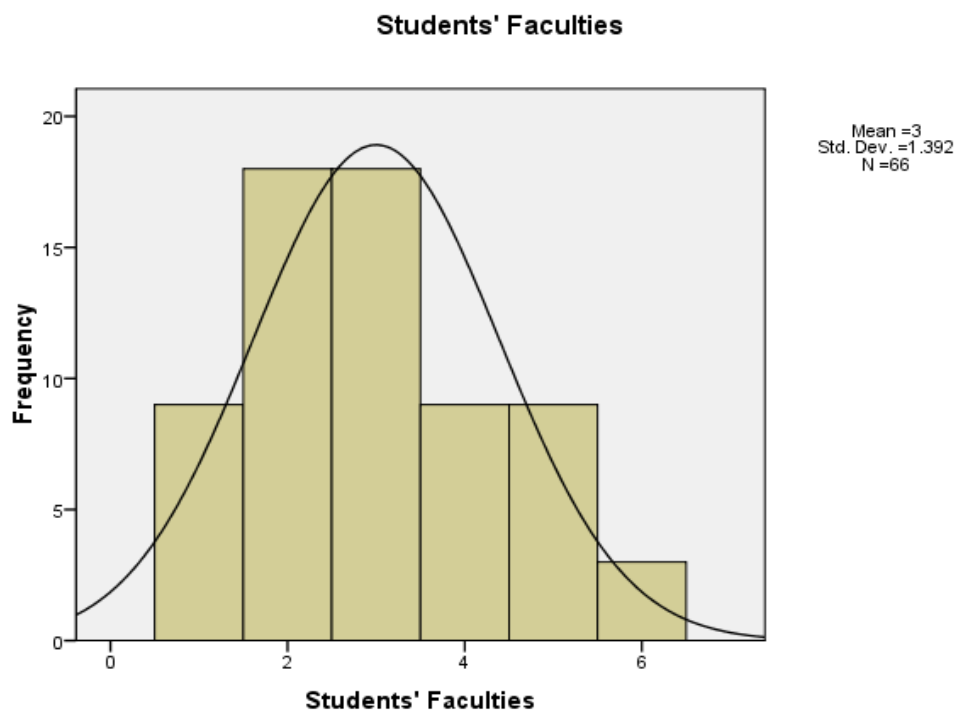


Figure 1. Histogram of students' faculties.

Figure 2 shows the normal Q-Q plot of the students' faculties. We observed that most plotted points fell approximately along a straight line (a 45 degree line). However, some points trailed off from a straight-line pack in a curve at the top end to suggest a slightly or approximately normal distribution.

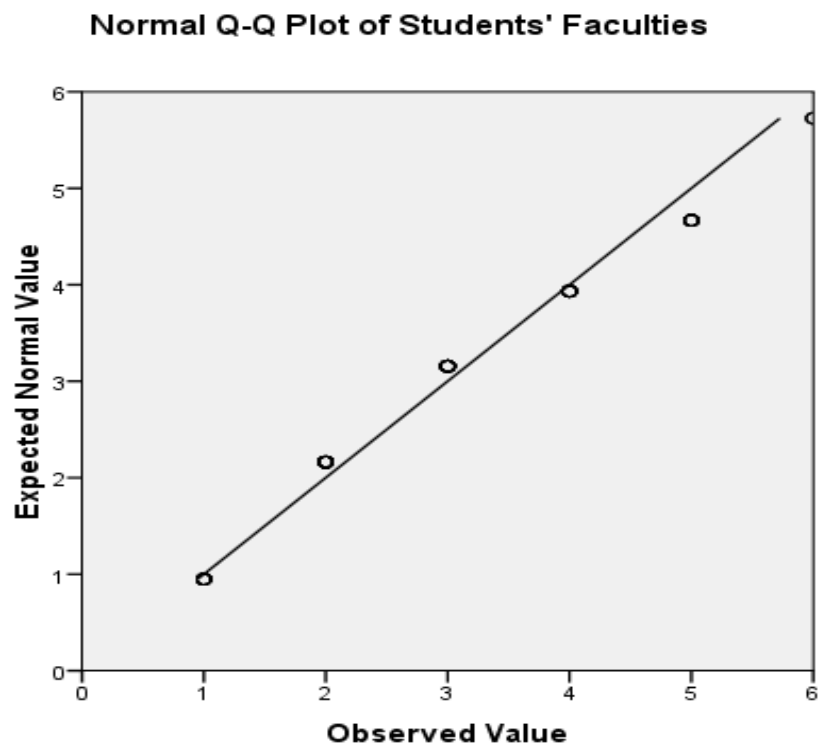
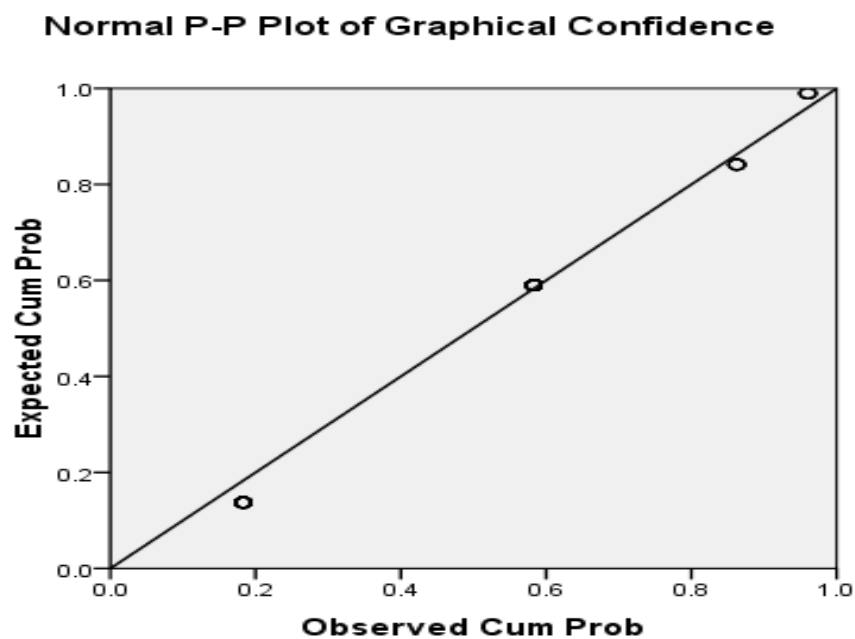


Figure 2. Normal Q-Q plot of students' faculties.

Figure 3 shows the normal P-P plot of students' faculties. Like the Q-Q plot, we equally observed that most plotted points fell approximately along a straight line at 45 degree line, but some points trailed off from the straight-line pack at the top end.



Transforms: natural log

Figure 3. Normal P-P plot of students' faculties.

Figure 4 shows the box plot of students' faculties. We observed that, apart from the ratio scale, the median lines did not evenly divide the boxes. Also, apart from the interval and the ratio scales, the upper tails of the boxes were consistently longer than the lower tails. These indicate that the population distribution of graduate students in the faculties from which the data were sampled was slightly skewed to the right.

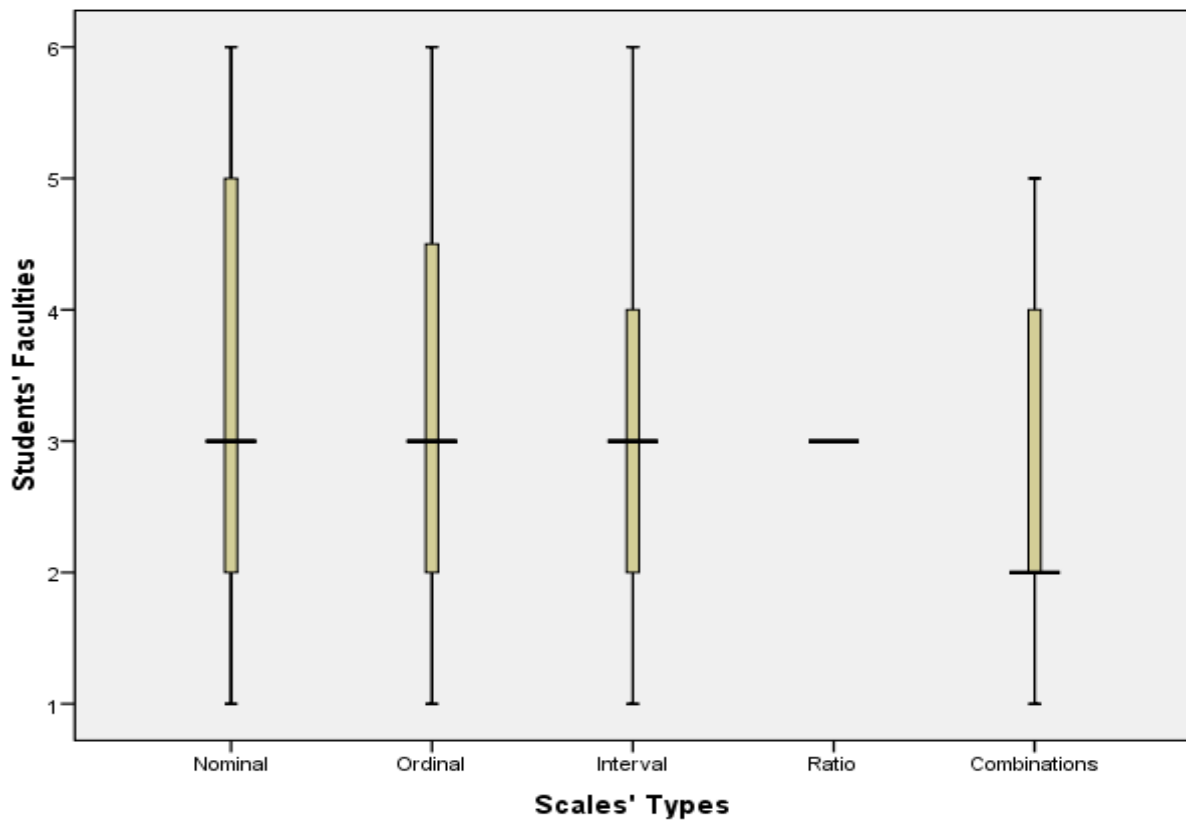


Figure 4. Box plot of students' faculties.

Table 1 shows the measures of central tendencies of students' knowledge in basic data concepts. We discovered that the mode, the median, and the mean were approximately the same. This indicates that the three measures of central tendencies came from a nearly normal distribution. However, the mean was slightly pulled to the right of the median to confirm the direction of the histogram in Figure 1.

Table 1

*Central Tendency Statistics of Students' Knowledge in Basic Concepts*

Statistics	Statistics types	Scales' types	Data types	Collection types
Mean	1.17	3.30	2.48	3.79
Median	1.00	3.00	3.00	4.00
Mode	1	3	3	5

Table 2 shows the spread statistics of students' knowledge in basic hypothesis testing concepts. We discovered that the inter-quartile range was approximately 1.33 standard deviations, the range was approximately 6 standard deviations, and the number of observations within 1, 2, and 3 standard deviations felt within the empirical rule of 68%, 95%, and 99.7% respectively.

Table 2

*Spread Statistics of Students' Knowledge in Basic Hypothesis Testing Concepts*

Statistics	Range	Minimum	Maximum	Mean	Std. deviation	Variance
Confidence levels	2	2	4	2.88	0.621	0.385
Numerical confidence	4	1	5	1.98	1.398	1.954
Theoretical confidence	3	1	4	2.06	0.802	0.642
Graphical confidence	5	1	6	2.06	1.311	1.719

Table 3 describes the position statistics of students' challenges in research report writing. Perfectly normal distributions have skewness and kurtosis statistics of zero. We discovered that the coefficient of skewness was positive (0.282) to suggest right skewness. Also, the kurtosis statistic of -1.667 indicates that our distribution was relatively flatter (platykurtic) than the normal distribution.

Table 3

*Position Statistics of Students' Challenges in Basic Hypothesis Testing*

Challenge	Mean		Skewness		Kurtosis	
	Statistic	Std. error	Statistic	Std. error	Statistic	Std. error
Challenges	3.21	0.156	0.282	0.295	-1.667	0.582

Table 4 shows the tests of the demographic information of postgraduate students. We have observed that both Shapiro-Wilk and Kolmogorov-Smirnov statistics were statistically significant at 5%. We concluded that gender, academic levels, and students' faculties were not normally distributed.

Table 4

*Tests of Normality Demographic Information and Descriptives*

Variable	Statistics types	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	Df	Sig.	Statistic	Df	Sig.
Gender	Descriptive diagrams	0.400	55	0.000	0.616	55	0.000
	Descriptive numbers	0.448	11	0.000	0.572	11	0.000
Highest academic levels	Descriptive diagrams	0.506	55	0.000	0.446	55	0.000
	Descriptive numbers	0.528	11	0.000	0.345	11	0.000
Students' faculties	Descriptive diagrams	0.235	55	0.000	0.897	55	0.000

Table 5 shows the tests of the demographic information of postgraduate students and the types of scales they use for research data. We have observed that even though both Kolmogorov-Smirnov and Shapiro-Wilk fail to reject students' faculties, Shapiro-Wilk did quite better. We concluded while gender and highest academic levels were normally distributed, the students' faculties were not.

Table 6 shows the tests of the demographic information of postgraduate students and the types of confidence levels they use for research data. We have observed that the Shapiro-Wilk statistic estimated the sample better than the Kolmogorov-Smirnov, but both rejected the null hypothesis of normality. We concluded that the research knowledge and skills in confidence levels of students' gender, academic levels, and students' faculties were not normally distributed.

Table 7 shows the tests of the demographic information of postgraduate students and the types of theoretical confidence levels they use for research data. We have observed that the Shapiro-Wilk statistic estimated the sample better than the Kolmogorov-Smirnov, but both rejected the null hypothesis of normality.

We concluded that the research knowledge and skills in theoretical confidence levels of students' gender, academic levels, and students' faculties were not normally distributed.

Table 5

*Tests of Normality of Demographic Information and Scale Types*

Variable	Scales' types	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	Df	Sig.	Statistic	Df	Sig.
Gender	Nominal	0.407	6	0.002	0.640	6	0.001
	Interval	0.469	29	0.000	0.533	29	0.000
	Ratio	0.414	9	0.000	0.617	9	0.000
	Combinations	0.419	15	0.000	0.603	15	0.000
Highest academic levels	Ordinal	0.504	7	0.000	0.453	7	0.000
	Interval	0.539	29	0.000	0.184	29	0.000
	Ratio	0.471	9	0.000	0.536	9	0.000
Students' faculties	Nominal	0.251	6	0.200	0.927	6	0.557
	Ordinal	0.191	7	0.200	0.955	7	0.772
	Interval	0.180	29	0.018	0.924	29	0.039
	Combinations	0.301	15	0.001	0.835	15	0.011

Table 6

*Tests of Normality of Demographic Information and Confidence Levels*

Variable	Confidence levels	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	Df	Sig.	Statistic	Df	Sig.
Gender	Ninety	0.521	17	0.000	0.385	17	0.000
	Ninety-five	0.351	40	0.000	0.636	40	0.000
	Ninety-nine	0.519	9	0.000	0.390	9	0.000
Highest academic levels	Ninety	0.380	17	0.000	0.632	17	0.000
	Ninety-five	0.538	40	0.000	0.147	40	0.000
	Ninety-nine	0.471	9	0.000	0.536	9	0.000
Students' faculties	Ninety	0.276	17	0.001	0.871	17	0.022
	Ninety-five	0.198	40	0.000	0.915	40	0.005

Table 7

*Tests of Normality of Demographic Information and Theoretical Confidence*

Variable	Theoretical confidence	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	Df	Sig.	Statistic	Df	Sig.
Gender	Critical values	0.431	13	0.000	0.592	13	0.000
	P-values	0.424	42	0.000	0.595	42	0.000
Highest academic levels	Critical values	0.392	13	0.000	0.628	13	0.000
	P-values	0.538	42	0.000	0.142	42	0.000
	Residual plots	0.492	6	0.000	0.496	6	0.000
Students' faculties	Critical values	0.317	13	0.001	0.878	13	0.067
	P-values	0.183	42	0.001	0.919	42	0.006
	Normal plots	0.287	5	0.200*	0.914	5	0.490
	Residual plots	0.251	6	0.200*	0.927	6	0.557

Table 8 shows the tests of the demographic information of postgraduate students and the challenges students encounter in the analysis of results. We observed that the Shapiro-Wilk statistic estimated higher



statistics than the Kolmogorov-Smirnov, but both rejected the null hypothesis of normality. We therefore concluded that students' challenges in gender, academic levels, and students' faculties were not normally distributed.

Table 8

*Tests of Normality and Challenges in Analysis of Results*

Variable	Analysis challenges	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	Df	Sig.	Statistic	Df	Sig.
Gender	Interpretation	0.354	32	0.000	0.637	32	0.000
	Labelling	0.385	3	0.000	0.750	3	0.000
	Computation	0.462	16	0.000	0.546	16	0.000
	Discussion	0.453	15	0.000	0.561	15	0.000
Highest academic levels	Interpretation	0.465	32	0.000	0.540	32	0.000
	Computation	0.518	16	0.000	0.398	16	0.000
Students' faculties	Interpretation	0.247	32	0.000	0.869	32	0.001
	Computation	0.319	16	0.000	0.837	16	0.009
	Discussion	0.305	15	0.001	0.766	15	0.001

### Conclusion and Recommendations

The results of SPSS Version 16 revealed that the postgraduate students require innovative skills in order to test and verify their research data. This would inspire stakeholder confidence, and give reliable inferences and trustworthy interpretations for policy-making. We therefore recommended continuous inter-faculty collaborations and university research conferences to beef up their knowledge and skills in educational research analyses.

The three stages of conducting normality explored include visualizing shape, detection of centrality, and verifying normality of distribution of research data. All the graphs passed the normal tests. Therefore, we recommended these graphs should be the first step of testing for normality in any educational research.

Also, the central tendencies, the dispersions, and positions of the descriptive statistics were approximately the equal. These statistics supported the graphical results of normality. We suggest that the central tendencies should always be employed as supports, and not as separate entities in any educational research at the postgraduate level.

Again, we discovered that the theoretical underpinnings of the inter-quartile range, the standard deviation, the range, and empirical rule were in tandems with our empirical research. Therefore, universities and other research institutions must not leave out measures of dispersions in normality testing in any scientific educational research.

Furthermore, the coefficients of skewness and kurtosis simply justified the relative locations of the mean, the median, and the mode. This revelation is necessary for teaching and analyzing normality of postgraduate research in education.

Finally, both Shapiro-Wilk and Kolmogorov-Smirnov statistics were statistically significant at 5% in most of the responses. However, the Shapiro-Wilk verifications exceedingly verified the normality than that of the Kolmogorov-Smirnov. Therefore, even though our results of gender, academic levels, and students' faculties were statistically significant, these two verification tools must be explored to assess the direction of the normality. But where significance differences exist, the statistics of Shapiro-Wilk should be more preferred.

## References

- Ali, A. C., Amihire, A. K., Adzifome, N. S., & Ngman-Wara, E. I. D. (2014). *Introduction to educational research methods assessment and statistics for diploma and degree students*. Winneba: Payless Ltd..
- Arthur, J. (2011). *Anderson Darling A2: P-value and critical values*. Denver: KnowWare International, Inc.. Retrieved from <http://www.qimacros.com/qiwizard/data-normality-test.html>
- Avioli, J. J. (2012). *Chapter 6: The normal distribution*. Virginia: Christopher Newport University.
- Best, J. W., & Kahn, J. V. (2014). *Research in education* (8th ed.). New York, N.Y.: Allyn and Bacon. Retrieved from <http://ww2.odu.edu/~jritz/attachments/reined.pdf>
- Elnabris, K. J. A. (2011). *BIOL 4243: Assessing normality*. New York, N.Y.: Mayfield Publishing Company.
- Ghasemi, A., & Zahadial, S. (2012). Normality test for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology & Metabolism*, 10(2), 486-489.
- GraphPad. (2007). *Column statistics, descriptive statistics and normality tests*. San Diego, C.A.: GaphPad Software Inc..
- Harmon, M. (2011). *Normality testing in Excel: The Excel statistical master*. Retrieved from <http://www.ExcelMasterSeries.com>
- Horst, H. (2010). *Is my data normally distributed? Making a decision based on visualizing data, finding skewness and kurtosis, and performing formal tests for normality*. Bren: Bren School of Environmental Science & Management, UCSB.
- Joreskog, K. G., & Sorbom, D. (2004). *LESREL 8.7 for Window, the diagnosis and treatment of non-normality*. Lincoldwood: Scientific Software International, Inc..
- Motulsky, H. J. (2013). *Analyzing data with GraphPad Prism*. San Diego, C.A.: GraphPad Software Inc.. Retrieved from <http://www.graphpad.com>
- Myoung, P. H. (2008). *Univariate analysis and normality test using SAS, Stata, and SPSS*. Bloomington: University Information Technology Services, Center for Statistical and Mathematical Computing, Indiana University.
- Oztuna, D., Elhan, A. H., & Tuccar, E. (2006). Investigation of four different normality tests in terms of type 1 error rate and power under different distributions. *Turkish Journal of Medical Sciences*, 36(3), 171-176.
- Russell, B. H. (2006). *Research methods in anthropology: Qualitative and quantitative approaches* (4th ed.) Oxford: AltaMira Press/Rowman & Littlefield Publishers, Inc.. Retrieved from <http://www.altamirapress.com>
- Shafer, D. S., & Zhang, Z. (2012). *Beginning statistics V.1.0*. New York, N.Y.: Environmental Protection Agency.
- Singh, S. A., & Masuku, B. M. (2014). Assumption and testing of normality for statistical analysis. *American Journal of Mathematics and Mathematical Sciences*, 3(1), 169-175.
- Spector, P. (2004). *An introduction to R*. Berkeley, C.A.: Statistical Computing Facility, University of California.