UNIVERSITI TEKNOLOGI MARA

# TEST OF NORMALITY: A POWER COMPARISON OF KOLMOGOROV-SMIRNOV, ANDERSON-DARLING, SHAPIRO-WILK AND LILLIEFORS TESTS

NORNADIAH BINTI MOHD RAZALI

Project submitted in fulfillment of the requirements
for the degree of
**Master of Science (Applied Statistics)**

**Faculty of Computer and Mathematical Sciences**

**November 2009**

# CANDIDATE'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the result of my own work, unless otherwise indicated or acknowledged as referenced work. This dissertation has not been submitted to any other academic institution or non-academic institution for any other degree or qualification.

In the event that my dissertation be found to violate the conditions mentioned above, I voluntarily waive the right of conferment of my degree and agree to be subjected to the disciplinary rules and regulations of Universiti Teknologi MARA.

| | |
|---|---|
| Name of candidate | Nornadiah Binti Mohd Razali |
| Candidate's ID No. | 2007131509 |
| Programme | Master of Science (Applied Statistics) |
| Faculty | Faculty of Computer and Mathematical Sciences |
| Dissertation Title | Test of Normality: A Power Comparison of Kolmogorov-Smirnov, Anderson-Darling, Shapiro-Wilk and Lilliefors Tests |

| | |
|---|---|
| Signature of Candidate | |
| Date | 9 DECEMBER 2009 |

**APPROVED BY:**

**ASSOC. PROF. DR. YAP BEE WAH**
**Supervisor**
**Faculty of Computer and Mathematical Sciences**
**Universiti Teknologi MARA**

# ABSTRACT

The importance of normal distribution is undeniable since it is an underlying assumption of many statistical procedures such as t-tests, linear regression analysis, discriminant analysis and Analysis of Variance (ANOVA). When the normality assumption is violated, interpretation and inference may not be reliable or valid. The three common procedures in assessing whether a random sample of independent observations of size $n$ come from a population with a normal $N(\mu,\sigma^2)$ distribution are: *graphical methods* (histograms, box-plots, Q-Q plots), *numerical methods* (skewness and kurtosis indices) and formal *normality tests*. This study compares the power of four tests of normality: Shapiro-Wilk (SW) test, Kolmogorov-Smirnov (KS) test, Lilliefors (LF) test and Anderson-darling (AD) test. Power comparisons of these four tests were obtained via Monte Carlo simulation of sample data generated from alternative distributions that follow symmetric and asymmetric distributions. The significance levels considered are 5% and 10%. First, critical values for power comparisons were obtained based on 50000 simulated samples from a standard normal distribution. As the SW test is a left-tailed test, the critical values are the $100(\alpha)^{th}$ percentiles of the empirical distributions of the SW test statistic. The AD, KS, and LF tests are right-tailed tests and the critical values are the $100(1-\alpha)^{th}$ percentiles of the empirical distribution of the respective test statistics. Then, 10000 samples each of size $n$ = 10, 15, 20, 25, 30, 40, 50, 100, 200, 300, 400, 500, 1000, 1500 and 2000 were generated from each of the given alternative symmetric and symmetric distributions. The power of each test was then obtained by comparing the test of normality statistics with the respective critical values. Simulation results show that SW test is the most powerful test followed by AD and LF tests in detecting departures from the normality assumption while KS test is the least powerful test. This study also shows that LF test performs better than the KS test. For sample sizes $n \geq 50$, the performance of SW and AD tests are quite similar. Results also show that KS and LF tests require large sample size (at least 2000 or more) to achieve similar power with SW and AD tests.

# ACKNOWLEDGEMENTS

In The Name of Allah, The Most Gracious, The Most Merciful

Praise to Allah for His blessing until I managed to complete this report successfully.

My highest gratitude and appreciation goes to my supervisor, Assoc. Prof. Dr. Yap Bee Wah for her guidance throughout the period of study. She is always there to assist me whenever I need help. She helped me to go through many obstacles that I have faced while completing this study. I feel honored to be given an opportunity to work with her.

I would also like to extend my appreciation to all Msc. (Applied Statistics) lecturers who have been very supportive and encouraging. I would also like to thank them for their opinions and suggestions during the proposal defense to further improve my dissertation.

To my dearest fiancé and family, thanks for being so understanding that I have to sacrifice the valuable time which is supposed to be spent with them to complete this study. Thanks also for their prayers to see me succeed in this study.

Finally, I would like to thank my friends for their assistance, support and comments on my dissertation. I appreciate every single help from them whether directly or indirectly.

# TABLE OF CONTENTS

# LIST OF TABLES

## LIST OF FIGURES

# LIST OF APPENDICES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| SW | Shapiro-Wilk |
| KS | Kolmogorov-Smirnov |
| AD | Anderson-Darling |
| LF | Lilliefors |
| JB | Jarque-Bera |
| EDF | Empirical distribution function |
| CDF | Cumulative distribution function |
| PDF | Probability density function |
| SES | Standard error of skewness |
| SEK | Standard error of kurtosis |

# CHAPTER 1

# INTRODUCTION

## 1.1 Background of Study

When carrying out statistical analysis using parametric methods, the assumption of normality is a fundamental concern for the analyst. As a statistician, we often conclude that 'the data are normal' or 'not normal' based on some test of normality results. For those without statistical background, this statement might be questionable. By definition, normal data are data that come from a population that has a normal distribution. The normal distribution is also referred to as Gaussian or bell-shaped distribution.

If X is a random variable which comes from a population with normal distribution, or in notation, $X \sim N(\mu, \sigma^2)$, then the probability distribution of X is (Hogg & Tanis, 2006),

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} ; \quad -\infty < x < \infty \tag{1.1}$$

where $\mu$ and $\sigma$ are the mean and the standard deviation of the distribution, respectively. This well-known distribution takes the form of a symmetric bell-shaped curve. The illustration of this distribution can be seen in Figure 1.1. As can be seen from the figure, the mean of the distribution is the point at the centre of the curve whereas the standard deviation describes the spread of the curve or the variation of the data points around the mean.

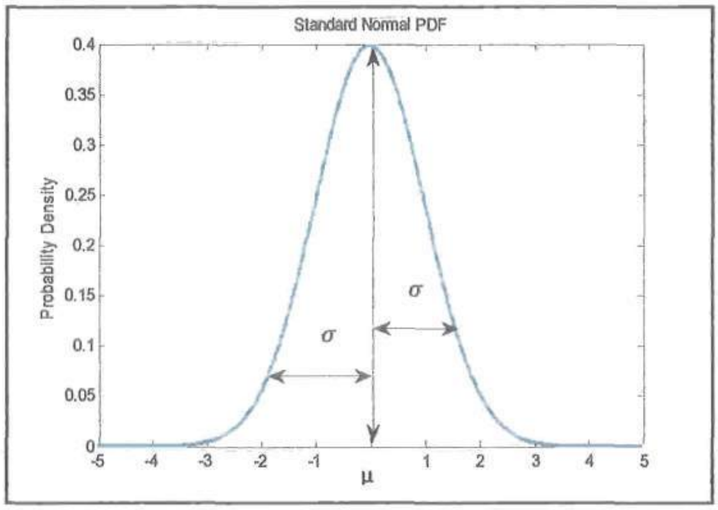

Figure 1.1: The Shape of the Standard Normal Distribution

The importance of normal distribution is undeniable since it is an underlying assumption assumed of many statistical procedures. It is also the most frequently used distribution in statistical theory and applications. Though it is important for certain statistical procedures to assume that data should come from a normal distribution, but in real life it is indeed impossible for the data to be perfectly normal. Geary (1947) suggested that in front of all statistical texts should be printed,

"Normality is a myth. There never was and will never be a normal distribution."

Hart and Hart (2002) agreed with Geary and emphasized that the normal distribution never reflects real life data since it yields values that range from minus infinity to plus infinity. They also added that the data should only necessary to be approximately normal instead of perfectly normal.

Assessing the assumption of normality is required by most statistical procedures. Parametric statistical analysis is one of the best examples to show the importance of assessing the normality assumption. Parametric statistics are those which assume a certain distribution of the data, usually the normal distribution. If the assumption of normality is violated, interpretation and inference may not be reliable or valid. Some of the most well-known statistical tests such as t-test, F-test and Analysis of Variance (ANOVA) are parametric tests. Parametric statistical analysis is more powerful compared to non-parametric statistical analysis.

Since the assumption of normality is fundamental to the use of many statistical tests and inferences, therefore it is important to check for this assumption before proceeding with any relevant statistical procedures. Testing the normality of the data should be the first step of any analysis which requires the assumption of normality. Basically, there are three common ways to check the normality assumption. The easiest way is by using the *graphical method*. The normal quantile-quantile plot (Q-Q plot) is the most commonly used and effective diagnostic tool for checking normality of the data. Other common graphical methods that can be used to assess the normality assumption include histogram, box-plot and stem-and-leaf plot. Even though the graphical methods can serve as a useful tool in checking normality for sample of $n$ independent observations, they are still not sufficient to provide conclusive evidence that the normal assumption holds. This method is very subjective as it lies in the eyes of the analyst. What seems 'normal' to one analyst may not necessarily be so for others. In fact, experience and good statistical knowledge are needed in order to interpret the graph.

Therefore, to support the graphical methods, a more formal method which is the *numerical methods* and formal *normality tests* should be performed before making any conclusion about the normality of the data. Our judgement on the normality of the data will be much improved by combining the graphical methods, numerical methods and normality tests. The numerical methods include the skewness and kurtosis coefficients whereas for the normality tests, there are various procedures available for testing the assumption of normality. However, the most common normality test procedures available in most statistical software are the Shapiro-Wilk test, Kolmogorov-Smirnov test, Lilliefors test and Anderson-Darling test. The focus of this study is to compare the power of several normality tests via Monte Carlo simulation. Further discussions on these normality tests are available in the following chapter. The problem statement is given in Section 1.2 while the research questions and objectives are stated in Section 1.3 and Section 1.4, respectively. Section 1.5 states the significance of the study while the scope and limitations are explained in Section 1.6.

## 1.2 Problem Statement

There are significant amount of tests of normality available in the literature. Some of these tests can only be applied under a certain condition or assumption. Moreover, different test of normality often produce different results i.e. some test reject while others fail to reject the null hypothesis of normality. The contradicting results are misleading and often confuse practitioners. Therefore, the choice of test of normality to be used should indisputably be given tremendous attention. The preparation of some guidelines will be very helpful to solve this problem. However, the guidelines provided in the literature especially in terms of the power of the test are still ambiguous and contradictory. Therefore, the purpose of this study is to provide knowledge and guidelines on the choice of normality tests which will be useful for practitioners.

## 1.3 Research Questions

This study seeks answers to the following questions:

1. What are the characteristics of various tests of normality available in the statistical literature?

2. How does the performance of Kolmogorov-Smirnov test, Anderson-Darling test, Shapiro-Wilk test and Lilliefors test varies among different sample sizes and alternative distributions?

3. What are the guidelines that should be followed in choosing the test of normality?

## 1.4 Objectives of Study

The main objectives of the study include:

1. To investigate the characteristics of the tests of normality identified in statistical literature.

2. To perform a simulation study to compare the performance of the Kolmogorov-Smirnov test, Anderson-Darling test, Shapiro-Wilk test and Lilliefors test of normality.

3. To provide guidelines to practitioners on the choice of normality test.

## 1.5 Significance of Study

The various tests of normality revealed through this study may increase the awareness of the practitioners on more recent and perhaps better test of normality than the ones most commonly used. At the end of this study, it is hoped that the results will be able to provide some guidelines to practitioners on the choice of test of normality. This study is also expected to provide a clearer idea on the test of normality that should be used under different sample size conditions. Finally, this study should be able to provide some idea to software developers such as the SAS Institute and SPSS Inc. on new tests of normality that should be included in their software.

## 1.6 Scope and Limitations of Study

Due to the time constraints, the scope of this study has been narrowed down. For the first objective of the study, only some normality tests that were found to be most well-known in the statistical literature were explained in the report. For the simulation study, only four most commonly available tests of normality in statistical software packages were considered: Kolmogorov-Smirnov test, Anderson-Darling test, Shapiro-Wilk test and Lilliefors test. The Monte Carlo simulation was carried out using FORTRAN programming language since it was the only software which had the subroutines for all the four tests.

## 1.7 Layout of Report

This chapter provides the background of the study and defines the problem or main issues which are the main concern of this study. In addition, the research questions and objectives are also stated. The significance as well as the scope and limitations of the study are also discussed. Chapter 2 presents a review of relevant literature, including a summarization of previous studies on normality test and the development of normality test statistics. Chapter 3 discusses in detail the Monte Carlo simulation methodology for power comparisons of the normality tests. The algorithms involved in this simulation study are described here. The simulation results are presented in Chapter 4. Some analyses to rank the power of the tests are also conducted and presented in this chapter. Finally, Chapter 5 includes a discussion of the results and a conclusion based on the findings obtained. The strategy for assessing normality and recommendations for future research are also proposed in this final chapter.

# CHAPTER 2

# LITERATURE REVIEW

## 2.1 Introduction

There are nearly 40 tests of normality available in the statistical literature (Dufour *et al.*, 1998). The effort of developing techniques to detect departures from normality has begun as early as the late $19^{th}$ century. This effort was initiated by Pearson (1895) who worked on the skewness and kurtosis coefficients (Althouse *et al.*, 1998). In 1900, Pearson extended his work and introduced the chi-square test of normality. Kolmogorov and Smirnov then introduced the Kolmogorov-Smirnov test of normality in 1933. Conover (1999) stated that the Cramer-von Mises test was developed based on the contributions by Cramer (1928), von Mises (1931) and Smirnov (1936). In 1954, Anderson and Darling proposed their test which was the modification of the Cramer-von Mises test (Farrel & Stewart, 2006).

According to Yazici and Yolacan (2007), another test of normality was proposed in 1962 which was called the Kuiper test. This was followed by the introduction of the Shapiro-Wilk test in 1965. In 1967, the Kolmogorov-Smirnov test was modified by Lilliefors. Lilliefors test of normality differs from that of the Kolmogorov-Smirnov whereby the parameters of the hypothesized distribution are estimated rather than initially specified (Abdi & Molin, 2007).

Then, in 1971, D'Agostino proposed an omnibus test for moderate and large size samples, or widely known as D'Agostino test. This test is similar to the Shapiro-Wilk test which is based on regression (Coin and Corradetti, 2006). However, it requires no table of weights and it can be used for sample size of more than 50 (D'Agostino, 1971). Shapiro and Francia (1972) then developed another test of normality which was a modification of the Shapiro-Wilk test. The D'Agostino-Pearson (1973) test that took into consideration both the skewness and kurtosis values was proposed in the following year. Another test of normality which was known as Vasicek test emerged in 1976. This test was constructed based on the sample entropy of the data. Jarque-Bera test which was also based on skewness and kurtosis was introduced in 1987.

Royston (1982a) modified the Shapiro-Wilk test to broaden the restriction of the sample size to 2000 as the original test was only limited to sample size only up to 50. Royston (1982b, c) provided algorithm AS 181 in FORTRAN 66 for computing the *SW* test statistic and $p$-value for sample sizes 3 to 2000. Later, Royston (1992) observed that Shapiro-Wilk's (1965) approximation for the weights $\boldsymbol{a}$ used in the algorithms was inadequate for $n > 50$. He then gave an improved approximation to the weights and provided algorithm AS R94 (Royston, 1995) which can be used for any $n$ in the range $3 \leq n \leq 5000$. Rahman and Govindarajulu (1997) proposed a modification of the Shapiro-Wilk test and claimed that the computation of their statistic was much simpler than computing Royston's (1982a) approximation.

In the same year, Declercq and Duvaut introduced another test which was based on the Hermite polynomial. The history of the test of normality discussed above is based on the articles acquired during the period of this study. It should be noted that there may be more tests of normality that emerged in between or after the period mentioned above. Figure 2.1 shows the development of tests of normality in chronological order.

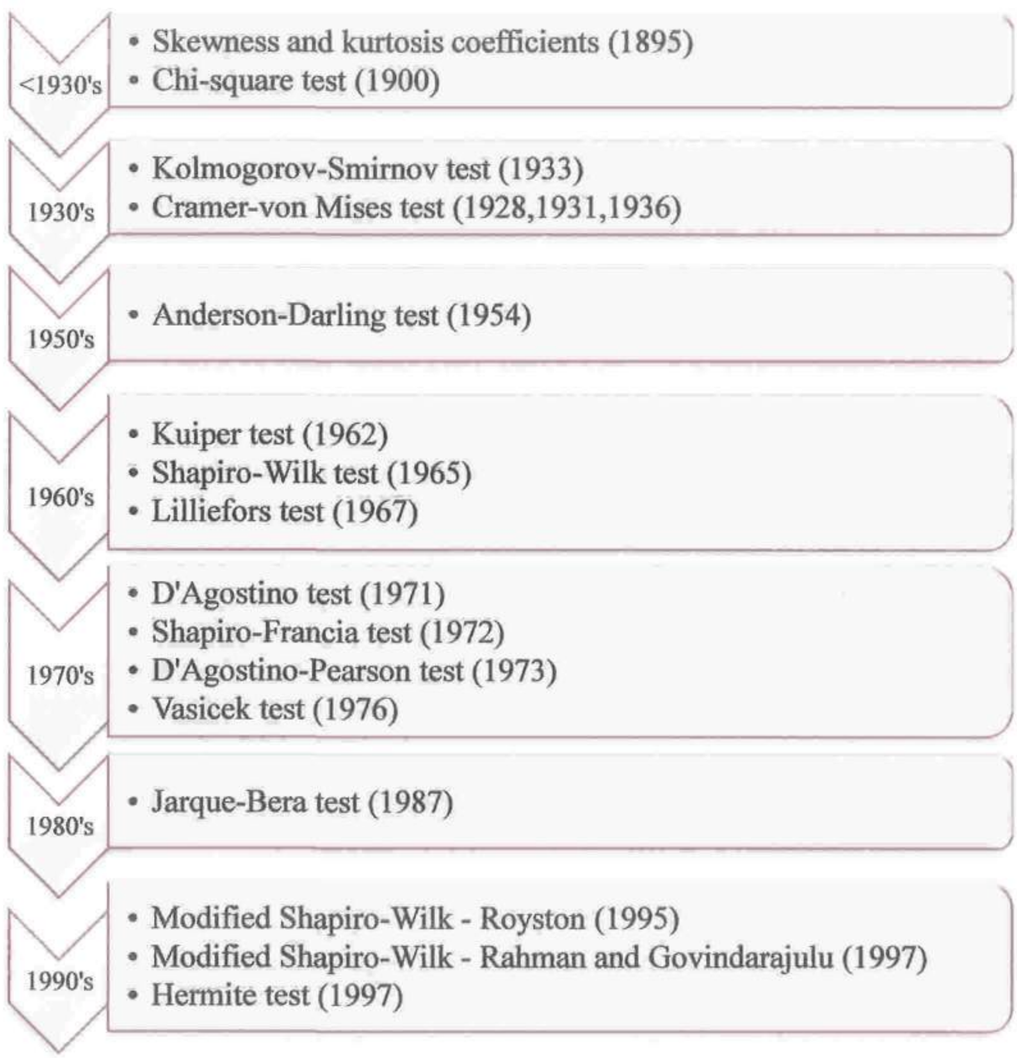


Figure 2.1: The Development of Tests of Normality in Chronological Order