Honam Mathematical J. **29** (2007), No. 1, pp. 83–100

STATISTICAL CONCEPTS AND TECHNIQUES FOR TESTING DEPARTURES FROM NORMALITY IN THE MATHEMATICS TEACHER PREPARATION

Sang-Gone Lee

Abstract. Normality is one of the most common assumptions made in sampling and statistical inference procedures without suffering from lack of attention. Its results may lead to an invalid conclusion. We present several testing procedures that can be used to evaluate the effects of departure from normality using concrete examples by hand or with the aid of Minitab. The goal is to influence prospective teachers in order to learn statistical concepts and techniques for testing normality on the basis of the didactical theory.

1. Introduction

Statistics is part of the mathematics curriculum for secondary school. This teaching includes the usefulness of statistics for daily life, its instrumental role in other disciplines, and its role in developing a critical reasoning. Secondary mathematics teachers frequently feel to lack specific preparation in statistics education. Most of them teach statistics concepts with the same methods presented in the textbooks they are using. Also, prospective teachers with a major in mathematics education do not receive specific training in statistics education; but the National Council of Teachers of Mathematics' Curriculum and Evaluation Standards for School Mathematics(1989) call for opportunities for students to describe the normal curve and use its properties to answer questions about sets of data that are assumed to be normally distributed. It is premised on the problem of testing whether a sample of observations come from a normal distribution. Whereas not all bell-shaped curves

Received Feb. 7, 2007. Accepted Mar. 31, 2007.

2000 Mathematics Subject Classification: 62G10 97U70.

Key words and phrases: Testing Normality, Problem Solving, Didactical Theory.

are normal, many statistical methods are designed for data from normal distributions. Unfortunately, many students and teachers want to believe that the distributions of data from many real-world phenomena can be closely approximated by the normal curve. They can arise from bias in the presentation and interpretation of data gathered. Their expectation may strongly affect the message derived from the statistical results. Testing normality gives teachers a feel for the practical struggles and small decision needed in real data gathering.

Normality is one of the most common assumptions made in the development and use of statistical procedures. A review of tests of normality shows many contributions from the most outstanding theorists of statistics. We consider some formal testing procedures that have been proposed to test specifically for normality, as well as plotting methods, and general goodness of fit tests that are useful in detecting non-normality. Several demonstrations and examples of testing normality reviewed in this paper will support many important concepts and ideas for prospective teachers in our introductory statistics courses for university students who have majored in mathematics education. We begin our quest for normality by suggesting basic plotting procedures for the display of a sample of data. Since the use of graphs may lead to a spurious conclusion if the sample size is small, the use of formal numerical techniques to generate estimates will avoid this. We focus on simple to use graphical procedures involving readily available sample skewness and kurtosis statistics and special probability plotting. Next we provide the Kolmogorov-Smirov test and the Shapiro-Wilk test as common tests of normality, because the graphical analyses are also less formal than the numerical analyses in understanding statistical properties of the underlying phenomena.

The validity of the parametric statistical inference procedure is focused on the shape of populations from which the samples have been drawn. Since populations do not always meet the assumptions underlying parametric tests, we frequently need inferential procedures whose validity does not depend on rigid assumptions. Although modest departures from this assumption do not seriously affect our conclusions, we have to seek an alternative method of analysis. One such alternative is a non-parametric procedure. A non-parametric test may be probably more appropriate and easier to use with small samples. Therefore, in testing normality we try to apply a goodness of fit method as the Kolmogorov-Smirov test, which is distribution free in the sense that the critical values do not depend on the specific distribution being tested.

Since the Kolmogorov-Smirov test has power so low that prospective teachers should consider testing for normality, we have simultaneously used the Shapiro-Wilk test and two other tests(i.e. the Anderson-Daring test and the Ryan-Joiner test) with the aid of Minitab, which have various advantages but also some drawbacks.

Our 7th educational curriculum in 2000 is not allowed to teach testing normality in the classroom. The general approach in statistics education is to define a null hypothesis to test hypothesis that asserts the data did come from the specified distribution, then to calculate some suitable test statistic whose distribution is known if the null hypothesis is true. The null hypothesis is rejected if the value of the statistic falls outside a certain range of values based on a pre-chosen significance level. But a theoretical framework of teachers to a better statistical knowledge base may result in a clear performance to make coherent connections among the content knowledge, and link the concepts within statistics and other areas. In this paper, the assumption of normality requires an effective test of whether the assumption holds, or a careful argument showing that the violation of the assumption does not invalidate the procedure used. Through testing non-normality, prospective teachers will be aware the reason how and why it is appropriate to use a normal distribution in statistical analysis, will be familiar with probability tables of the test, and will be able to use either testing techniques associated with normality or computer software to test normality. We believe that beyond recall of concepts and algorithms or recitation of axioms and theorems, prospective teachers must have theoretical knowledge of statistics that can be adapted to different teaching levels, predict learners' learning difficulties, errors, obstacles, and strategies in problem solving, and show examples of teaching situations. As the vehicle for improving teaching as well as teacher education, university educators should develop contents and procedures supporting statistical beliefs and activities which can engage prospective teachers so that they have early vivid experience of didactical situations as a means to enhance statistical knowledge, and the content in which learning and teaching occurs. Therefore, we present the effects of violations of normality with the aid of Minitab and at the same time, review the major concepts and procedures under didactical situations using concrete examples in the mathematics teacher preparation.

2. Concepts and Techniques for Testing Normality

A data set is used to teach both descriptive and inferential statistics in the classroom, including calculating measures of central tendency and variation, statistical plotting, construction of confidence intervals, and the methods of hypothesis testing. There are many types of data display available including histograms, stem-and-leaf plots, and box plots. Particularly these three graphical displays can be implemented easily with the use of graphs or simple computer programs. Histograms display spread, symmetry, outliers and multi-modality of the distribution of a data set. Stem-and-leaf plots show the frequency of observations within bins, but also show the data values, allowing identification of the median, extremes and other sample percentiles directly from the plot. Box plots show less data detail than histograms, but give indication of the spread and symmetry of a data set, and are slightly more formal than histograms in identifying outliers in a data set.The width of the box is the fourth spread, which contains 75 percent of the observations. The bar inside the box indicates the location of the median. The whiskers extend to the observation in each tail of the data which is the furthest from the fourths but less than or equal to 1.5 times the fourth spread. Observations outside of the whiskers may be considered outliers with different levels depending on the distance from the fourths. Symmetry or skewness of the data can be ascertained from the location of the median within the box and the length of the whiskers on either side.
Example 1. Sample values are rainfall amounts(mm) at Seoul, Korea for the month of April over 30 years (1977-2006).[1] The resulting data from Korea Meteorological Administration is given in the following; 228.8 13.0 138.6 216.7 55.4 8.1 113.4 41.8 69.0 20.6 55.3 60.5 14.0 94.2 48.5 76.5 85.5 44.9 44.4 62.0 56.1 120.2 97.2 30.7 12.3 155.1 139.6 74.1 94.7 51.8. Figure 1 using the menu path *Graph > Character Graph > Stem-and-Leaf* displays the plot of rainfall amounts. We identify the median and fourths, as well as values of the minimum and maximum. The stem-and-leaf looks more or less skewed bell-shaped. In figure 2 a box plot using the menu path *Graph > Box Plot* shows that the main body is asymptotically symmetric, the large values cause the sample to be skewed, while two values of 216.7 and 228.8 are identified as outliers. It's difficult to judge normality using a box plot.

The features that cause the non-normality cannot be determined solely on the basis of the data display. More formal testing for normality could be accomplished by evaluating sample moments and comparing

Leaf Unit = 10 N=30

4	0	0111
6	0	23
14	0	44445555
(5)	0	66677
11	0	8999
7	1	1
6	1	233
3	1	5
2	1	
2	1	
2	2	1
1	2	2

FIGURE 1. Stem and Leaf Plot of Rainfall Amounts

them to theoretical moments. We briefly review them with emphasis on skewness and kurtosis. Moreover, let testing for distributional assumptions have a normalizing transformation and alternative methods such as distribution-free techniques in order to investigate the goodness of fit of the data to the normal distribution. We can provide the most common type of probability plot given by the Quantitle-Quantitle(Q-Q) plot, which is a plot of sample order statistics against some expected values of order statistics. At this time the special transformation has been applied to the vertical scale of a graph of the assumed cumulative distribution function, which can transform into straight line. The Q-Q plots can display a sense of pattern and a level of detail not available in a single test statistic. This allows subjective judgments of distributional shape. Although these plotting methods show the intuitive reason for testing normality, more formal testing procedures allow an objective judgment of normality for rejecting or accepting a null hypothesis. Thus we describe formal testing methods specific to the assessment of normality

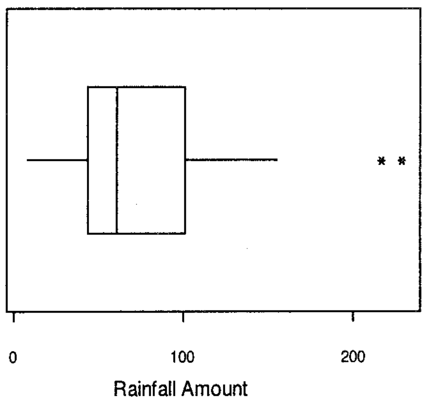

FIGURE 2. Box Plot of Rainfall Amounts

by suggesting the Kolmogorov-Smirnov test, and the Anderson-Daring test, for the empirical distribution function, and the Shapiro-Wilk test (similar to the Ryan-Joiner) for regression and correlation.

In this paper, we provide contents and examples for testing normality, and carry out activities with technology in order to train prospective teachers in statistics. At the beginning, they will have the difficulty that leads themselves to think; but, statistical graphs and procedures will make them act, speak, and evolve by their own motivation. These activities can provoke prospective teacher's reflection about the data analysis, the learners' difficulties and obstacles, and motive action. By carrying out these preliminary testing procedures, they can only reach an informal situation for normality testing based on visual inspection and intuition. To devolve this environment in an upper level, we lead prospective teachers to seek more formal testing of normality. We present more formal testing followed by emphasizing the potentiality in solving problems and mentioning the usefulness for giving concepts. Testing normality will give interest to the activity itself and encourages problem solving to foster reflection and motivation. Moreover, Minitab may reduce prospective teachers' cognitive load, replacing complex algorithmic procedures by simpler commands, thus allowing them to focus on high-level understanding. The nature of statistical applications makes co-operate problem solving out of a spontaneous approach to didactics. The cooperation among individuals results in the formation of grouped teamwork. Each team will explain the strategy to the solution and argue toward the validity of their solution. Teamwork allows standardization of knowledge among peers; fosters discussion of different solutions and strategies of solution; develop the ability of communicate mathematical ideas. At the end, all the teams will have a valid solution for all the problems and the validation of results' stages will be completed. Consequently, the test for normality using of technology may have a motivating aspect and change the way they study statistics. Therefore, prospective teachers build their knowledge in an active way, by solving problems and interacting with their classmates. Finally, they identify what the type of testing serves to validate the best normality as support for their decision. During the testing of normality, prospective teachers will naturally reach the transformed situations implemented by adidectical situations based on didectical situations regardless of the educator. At that time, we can summarize the results and give appropriate ideas for the concepts used. If we want prospective teachers to follow a didectical approach in their teaching, we need to try this same approach through their training. Also, that is why University educators should give prospective teachers more responsibility in their training, encourage their problem solving as a mean to foster reflection and motivate action, and help them to develop creative and critical thinking.

Preliminary Testing Normality

Skewness and Kurtosis

When samples have been drawn from a population, testing for normality may be accomplished by evaluating sample moments and comparing them to theoretical moments. Deviation in distributions from the normal could be characterized by differences in the moments. Particularly, there may be many variations from these values for small samples from a population. In fact, none of these similar test has yet proven to be better than moment tests under general situation in terms of having higher power.

Let X be a random variable from a population with mean μ and variance σ^2 and a probability density function f(x). Define the k^{th} population standardized moment as $\mu_k/{\mu_2}^{k/2}$ under the k^{th} population moment $\mu_k = \int_{-\infty}^{\infty}(x-\mu)^k f(x)dx$. The third and fourth standardized moments denoted as $\sqrt{\beta_1}$ and β_2 respectively, measure the skewness and kurtosis of a distribution. These values in a standard normal distribution are 0 and 3, respectively. The k^{th} sample moment is defined as

$m_k = \Sigma_{i=1}^n (x_i - \bar{x})^k/n$ where x_i is one of the n observations of X, and $\bar{x}$ is the sample mean for $k \geq 2$. The most commonly used moment test is based on the test statistic $m_k/{m_2}^{k/2}$.

The third sample moment test for skewness is given by the test statistic $\sqrt{b_1} = m_3/m_2^{3/2}$. This sample statistic provides formal numerical measure for investigating symmetry. When $\sqrt{b_1} > 0$, the data are skewed to the right, and when $\sqrt{b_1} < 0$, the data are skewed to the left. Under the null hypothesis of normality, $\sqrt{b_1}$ is asymptotically normal with mean 0 and variance 6/n. Shapiro, Wilk and Chen(1968) showed that $\sqrt{b_1}$ is one of the best tests for detecting skewed distributions and also a powerful test for symmetric long tailed alternatives when the sample size is small. A table of Pearson and Hartley(1966) contains 0.5 percent, 1 percent, 2.5 percent, and 5 percent lower and upper tail critical values for $\sqrt{b_1}$.

The fourth sample moment test for kurtosis is obtained by $b_2 = m_4/m_2^2$. The test statistic b_2 along with $\sqrt{b_1}$ is one of the oldest and most powerful tests for normality. If b_2 is greater than 3, a symmetric distribution is called 'heavy' or 'long' tailed and if b_2 is less than 3, it is called 'light' or 'short' tailed. Under the null hypothesis of normality, b_2 is asymptotically normal with mean 3 and variance 24/n. In sampling from fairly symmetric distributions, we might expect the kurtosis to reflect the non-normality. Thus we must remark that a combination of the test statistics $\sqrt{b_1}$ and b_2 may provide a more efficient test than either taken by itself. Under most circumstances, a two-tailed test is used, unless the direction of the tail is assumed known. Most tests can be broken down into directional and bidirectional tests. A directional test for skewness is used when a left or right skewed distribution is known to be the alternative of concern. A bidirectional skewness test is used when a skewed distribution is of concern but the direction of skewness is not known. But the use of kurtosis as a measure of the degree of non-normality in a distribution is somewhat arbitrary. (F. J. Anscombe and W. J. Glynn, 1983)
Example 2. 53 weekly average of winning numbers of Kookmin Bank, Korea in the Lotto 6/45 game was collected from 7 December 2002 to 6 December 2003.[4] Figure 3 using the menu path *Stat > Basic Statistics > Display Descriptive Statistics* is a histogram of the frequency distribution of these data with normal curve. In this sample, the moments are $m_2 \approx 24.563, m_3 \approx 8.583$, and $m_4 \approx 1380.823$ when $\bar{x} \approx 23.302$. We get $\sqrt{b_1} \approx 0.071$ and $b_2 \approx 2.289$. Since the expected

value of $\sqrt{b_1}$ under normality is 0, we reject the null hypothesis of normality if $\sqrt{b_1}$ is sufficiently far from 0. The 0.05 level critical value is 0.631. Since the absolute value of $\sqrt{b_1} = m_3/m_2{}^{3/2} = 0.071$ is smaller than 0.631, we do not reject the hypothesis of normality from the standpoint of skewness. The value of the Kurtosis test statistic for this data is $b_2 = m_4/m_2{}^2 = 2.289$. The lower and upper 2.5 percentiles for samples of size 53 (for a two-sided 0.05 test) are 2.06 and 4.32, respectively. Since $2.06 < 2.289 < 4.32$, we do not reject the null hypothesis of normality due to a symmetric departure based on these observations.[16]

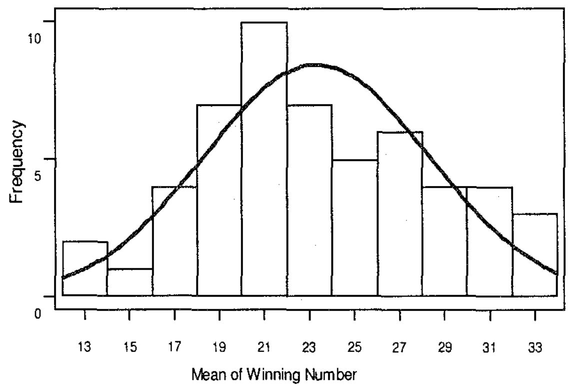

FIGURE 3. Histogram of Sample Mean in L 6/45

Quantile-Quantile Plot

Any statistical analysis always includes a graphical inspection of the data. Graphical techniques play a critical role in student's understanding of statistical reasoning. One such technique which can be used for evaluating a distributional assumption is the Q-Q plot. The Q-Q plot is a plot of the ordered observations from a sample against the corresponding percentage points from the standard normal distribution. The plot will be linear except for random fluctuations in the data under the null hypothesis. Any systematic non-linearity in a plot indicates that the data are not normal. Thus the Q-Q plot is used to check whether or not samples could have come from normal population.

Let $X_1, \cdots, X_n$ denote a random sample drawn from a normal distribution $N(\mu, \sigma^2)$ with cumulative distribution function (cdf) F. The empirical cumulative distribution function (ecdf) is defined as $F_n(x) = N[X_i \leq x]/n$, $-\infty < x < \infty$ where $N[X_i \leq x]$ denote the number of $X_i^{'}s$ less than or equal to x. The ecdf plot in attempting to judge visually the correctness of a specific hypothesized distribution depends on the curvature of the ecdf and cdf plots. Suppose that for all x $F_n(x)$ converges for large samples to $F(x)$. Let $X_{(1)}, \cdots, X_{(n)}$ be the order statistics of a random sample from the normal distribution. Choose $p_i = i/(n+1)$ as one of the most common plotting positions to use in constructing the Q-Q plot. If Φ is the cdf of the standard normal distribution and $z = (x-\mu)/\sigma$, the cdf F can be written as $F(x) = \Phi((x-\mu)/\sigma) = \Phi(z)$. For the n ordered observations $x_{(1)} \leq \cdots \leq x_{(n)}$, we have $z_{(i)} = \Phi^{-1}(F(x_{(i)})) = (x_{(i)} - \mu)/\sigma = (x_{(i)}/\sigma) - (\mu/\sigma)$. Letting $F(x) \approx F_n(x_{(i)})$, this plotting can be described as a plot of $z_{(i)} = \Phi^{-1}(F_n(x_{(i)}) \approx \Phi^{-1}(p_i) = \Phi^{-1}(i/(n+1))$ on $x_{(i)}$ for $i = 1, \cdots, n$. This task will entail some inverse probability calculations involving the theoretical population, and is best left to a computer.

By transforming the vertical scale, the curved line of the cumulative standard normal distribution can be transformed in a straight line. If the assumed distribution is correct, the corresponding quantities when plotted against the sample order statistics will be nearly linear. The expected relation between the $x_{(i)}$ and $\Phi^{-1}(p_i)$ in the Q-Q plot suggests the use of a measure of the linear correlation as a location and scale free test for normality. There are multiple ways to produce normal probability plot in Minitab. If the Q-Q plot is done on prepared normal probability paper, the ordered observations are plotted arithmetically on the X-axis in the probability paper and the values of p_i are plotted logarithmically on the Y-axis for use in determining whether data is normally distributed. But the Q-Q plot can be led by using arithmetic graph paper where the z is prepared approximately by the formula.(H. C. Hamaker, 1978) If we are attempting to reach a decision based on visual inspection it is probably easiest to judge if a set of points deviates from a straight line. The assumed normal distribution can be accepted or rejected on the basis of a judgment about the linearity of the data points. While this technique does not yield a formal testing procedure, it is sensitive in detecting such features as differences in symmetry, length of tails, and the existence of outliers.

The Q-Q plot is obtained by plotting the quantiles of the data of an empirical distribution against the quantiles of a theoretical distribution.

This is involves four steps;

- *Sort the observations to obtain the order statistics* $X_{(1)} \leq \cdots \leq X_{(n)}$. *These are the empirical quantiles used as the abscissa on the plot.*
- *Choose the plotting position* $p_i = i/(n+1)$ *for* $i = 1, \cdots, n$.
- *Plot the pairs* $(x_{(i)}, p_i)$ *on normal probability paper. Using this special graph paper is equivalent to plotting* $x_{(i)}$ *on arithmetic graph paper against* $\Phi^{-1}(p_i)$
- *Examine the plot for departures from a straight line.*

Example 3. Sample values are monthly average temperatures ($C°$) at Pohang, Korea in 2005.[1] The resulting data from Korea Meteorological Administration is given in the following; 1.1 1.6 6.8 15.2 17.6 23.7 25.8 26.4 22.5 16.8 10.9 0.1 We order the observations from the smallest to the largest, and plot the set of $x_{(i)}{}'s$ against $p_i = i/(n+1)$. The plotted values using the menu path *Stat > Regression > Fitted Line Plot* and adding confidence intervals are shown in figure 4. We only see a couple of data values outside the limits in the tail. A point on the left corresponds to a data value more extreme than expected from a normal distribution. Some points on the left are above the line. But several points on the right are below the line. We expect around 5 percent outside the limits, so there is no indication of non-normality.

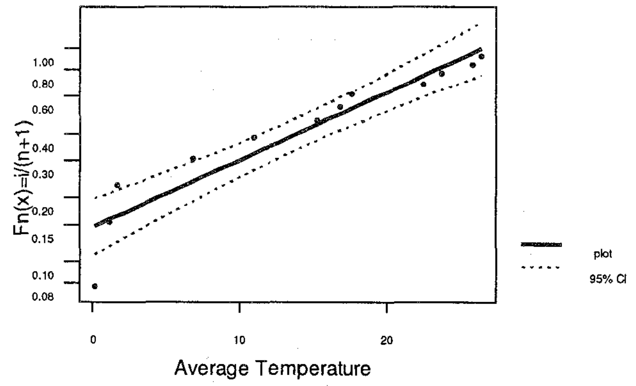

FIGURE 4. Q-Q Plot of Monthly Average Temperatures

Let's turn to Example 3 to see how two other plots identify features. Figure 5 shows that samples are skewed to the left. The box plot has a median 14.04 between the first quartile value of 2.9 and the third quartile value of 23.4. In figure 6 the Q-Q plot in the opposite direction displays very little difference except little flip on the middle.

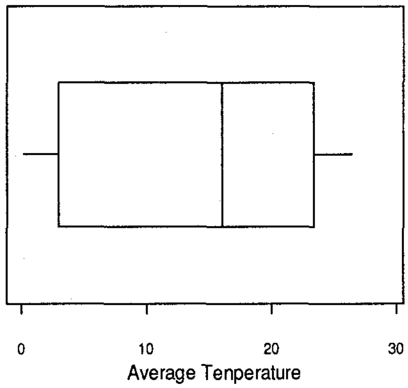

FIGURE 5. Box Plot of Monthly Average Temperatures

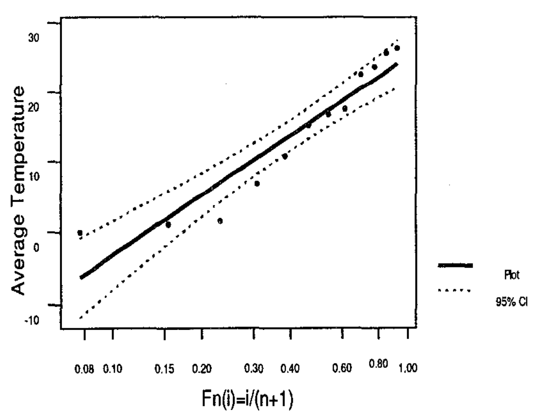

FIGURE 6. Q-Q Plot of Monthly Average Temperatures

Formal Testing Normality

Kolmogorov-Smirnov Test

This goodness-of-fit test was first introduced in 1933 by the Russian mathematician A. N. Kolmogorov. The Kolmogorov-Smirnov(K-S) test is the non-parametric or distribution-free test whose sampling distribution does not determine whether or not a set of observations are from some specified distribution. The K-S test is to used to compare an empirical distribution function with a theoretical distribution function in order to decide whether or not to reject the hypothesis that $F(x) = \hat{F}(x)$ for all x, where $\hat{F}(x)$ is the unknown distribution function which is estimated from the sample. Since the K-S statistics are based on the maximum differences between the ecdf $F_n(x)$ and the $\hat{F}(x)$, these are given by $D = max_x \mid F_n(x) - F(x) \mid$ on the null hypothesis $H_0 : F(x) = \hat{F}(x)$. If the value of the K-S test statistic exceeds the critical value, then we reject the hypothesis that the observations came from a specified population. The K-S statistic provides a means of testing whether a set of observations are from a normal population when the mean and variance are specified but must be estimated from the sample. Even though many goodness of fit tests with higher statistical power have been developed in recent decades, the K-S test remains popular because it is simple and intuitive, comparing the ecdf with the cdf and basing the test on the maximum deviation.

Example 4. The following data are the ordered observations of a random number of size 20 on a random device in MS-Excel; 10.75 6.37 2.39 17.12 2.59 4.18 3.06 17.38 0.75 14.52 8.5 17.89 19.32 13.74 17.07 5.19 11.84 4.5 3.02 0.97. The result using the menu path *Stat* > *Basic Statistics* > *Normality Test* in figure 7 is in the bottom right corner. The K-S test statistic D is 0.174. This value does not exceed the 0.05 level critical value of 0.294.[15] The p-value is $0.111 > 0.1$, so we do not reject the hypothesis of normality.

In Minitab we use the Anderson-Darling(A-D) test that is a modification of the K-S test and gives more weight to the tails than does the K-S test. The A-D test can be used to test whether data represent a sample from population with any specified distribution. The A-D test has the advantage of allowing a more sensitive test and the disadvantage that critical values must be calculated for each distribution. Suppose we have a random sample of n observations, and wish to test whether they are from a normal population with pre-specified mean and standard deviation. The A-D test statistic is computed as

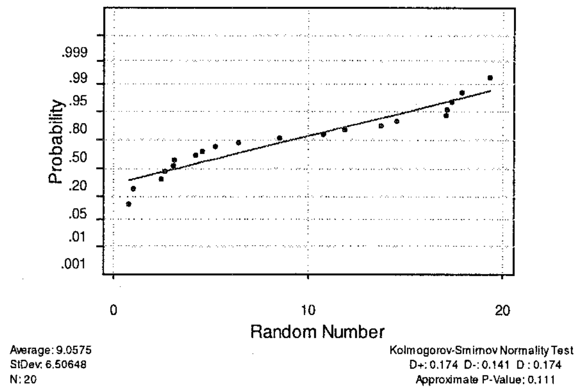

FIGURE 7. Kolmogorov-Smirnov Test of Random Numbers

$A^2 = -n - \Sigma_{i=1}^{n} \frac{(2i-1)}{n}[\ln(F(X_{(i)})) + \ln(1 - F(X_{(n-i+1)})]$

Example 5. Assume that we use random samples obtained from Example 4. The result of the A-D test statistic using the menu path *Stat > Basic Statistics > Normality* is 0.821. This value does not exceed the 0.05 level critical value of 2.492.[8] But the p-value is slightly below 0.1. It calls into questions in a normality assumption.

Shapiro and Wilk Test

The Shapiro-Wilk(S-W) test was developed in 1965, a procedure used to compare the generalized least squares estimate with the estimate given by the sample variance. The S-W statistic considers the regression of ordered sample on corresponding expected normal order statistics, which is linear for a sample from a normal distribution. This test has good power against a wide range of alternatives but the rationale is not so easily described by intuitive argument.

If the random samples $X_1, \cdots, X_n$ come from a normal distribution $N(\mu, \sigma^2)$, the straight line $z_i = -\mu/\sigma + (1/\sigma)x_i$ would be on normal probability paper. Let $U' = (u_1, \cdots, u_n)$ denote the vector of expected values of the i^{th} order statistics $u_i = E(X_{(i)})$ and let $\mathbb{V}$ be the covariance matrix of the order statistics $x_{(1)}, \cdots, x_{(n)}$. Letting $a = (U'\mathbb{V}^{-1})/[(U'\mathbb{V}^{-1})(\mathbb{V}^{-1}U)]^{1/2}$, the best linear unbiased estimate of unknown variance σ^2 is obtained from the least squares regression of the sample order statistics on their expected values if the statistic b yields the slop of the regression line as the following; $b = \sum_{i=1}^{k} a_{n-i+1}(x_{(n-i+1)} -$

$x_{(i)}$) where k=n/2 if n is even or k=(n-1)/2 if n is odd for the appropriate values a_i in the table.[22] Then the S-W test statistic is defined by $W = b^2/(n-1)s^2$ where s^2 is the sample variance. If the S-W value is less than the value W_α in the special table that is required in employing the test, we reject the null hypothesis H_0 : the X_i is a random sample from $N(\mu, \sigma^2)$ with a significance level α.

The S-W test statistic is calculated as follows;

- *Order the observations to obtain an ordered sample*
- *Compute b with k=n/2 if n is even or with k=(n-1)/2 if n is odd.*
- *Compute the test statistic*
- *If the computed value of S-W test statistic is less than the W_α , we reject the hypothesis of normality on a significance level α.*

Example 6. There was 27 undergraduate students in a class of our basic probability theory in 2006. At end of fall semester, we selected random students of size 15 and looked at their scores; 17 76 46 51 27 45 61 40 34 15 67 77 43 65 7. We compute $(n-1)s^2 = \Sigma_{i=1}^{n} {x_i}^2 - (\Sigma_{i=1}^{n})^2/n$=6762.9333. Also we get a_9 =0.0433, a_{10}=0.0880, a_{11}=0.1353, a_{12} =0.1878, a_{13}=0.2495, a_{14}=0.3306, and a_{15}=0.5150. Since we have b = 80.579, the test statistic W is approximately 0.96. Thus, for 5 percent test with n=15, we get $W_{0.05}$=0.881 and 0.96 is greater than 0.881. Thus we do not reject the hypothesis of normality.

In Minitab the Ryan-Joiner(R-J) test (similar to Shapiro-Wilk test) is based on regression and correlation. First the graph is constructed by plotting the normal scores versus the data for a sample of the given size. Second the R-J test is done by storing the residuals with the null hypothesis: the data $X_1, \cdots, X_n$ are a random sample of size n from a normal distribution.

Assume that the observations, $x_{(1)}, \cdots, x_{(n)}$ are numbered in rising magnitude. Compute the correlation percentiles $Y_{(i)} = \Phi^{-1}(p_i)$. Then the R-J test statistic is defined as the correlation coefficient r between the data and the normal scores for the $(x_{(i)}, y_{(i)})$ pairs. If the data are samples from a normal distribution then the plot of normal scores against the data will be close to a straight line, and the correlation r will be close to 1. If the data set are samples from a non-normal distribution then the plot may show a marked deviation from a straight line, resulting in a smaller correlation r. Smaller values of r are regarded as stronger evidence against the null hypothesis.

Example 7. Suppose that the data of random samples are obtained from Example 6. In Minitab the R-J test is done by storing the residuals, using the menu path *Stat* > *Basic Statistics* > *Normality*, entering the residual column as the variable, and selecting the R-J test. In figure 8 the result is in the bottom right corner. The correlation between residuals and normal scores is 0.9552. The p-value is > 0.1000, so we do not reject the null hypothesis.

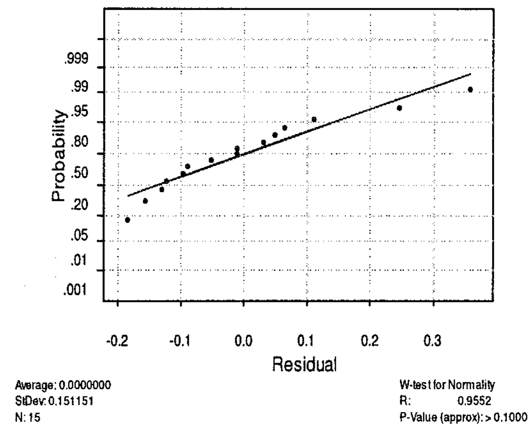

FIGURE 8. Ryan-Joiner Test of Random Scores

3. Conclusion

University courses must reflect the changes in emphasis and rapidly broadening scope of statistics itself. Prospective teachers need to have opportunities to easily implement procedures that can assess the normality assumption in ways different from their previous experiences and can emphasize enough to learners the main statistical concepts without neglecting aspects like the analysis of results and validation. We recommend that formal tests for normality in statistics education are added to the visual tool and used alongside the other exploratory tests available by hand or with the aid of a computer.

Graphical displays are used for informal preliminary judgment as adjuncts to formal numerical tests. If a sample data displays non-normal features and the sample size is relatively small, classical inference procedures that assume normality could not be more used. We provide non-parametric procedures offering rich alternatives to the classical inference tests. The Kolmogorov-Smirnov test and the Anderson-Daring test both compare the empirical cumulative distribution function to that of the best fitting normal curve. The Shapiro-Wilk test(the Ryan-Joiner test in Minitab) uses correlation observed in the probability plot. Thus we hope that non-parametric techniques will become an integral part of the evolving the statistics education for secondary school.

While testing normality is followed, prospective teachers can connect with the motivation as the cooperation among individuals' results, can build their knowledge in an active way, by solving problems and interacting with their classmates. Therefore, we expect that adidectical situations can be evolved by didectical situations. Since prospective teachers will teach using beliefs and methods they have remembered as being effective in their own learning, we present a didectical propersal for testing normality with the aid of Minitab and review the major concepts and procedures of these testing using a concrete example.

References

[1] Korea Meteorological Administration. Http://www.kma.go.kr. 11 2006.

[2] T. W. Anderson and D. A. Daring. A Test of Goodness of Fit. *JASA*, 49:765–769, 1954.

[3] F. J. Anscombe and W. J. Glynn. Distribution of the Kurtosis Statistic b_2 for Normal Samples. *Biometrika*, pages 227–234, 1983.

[4] KookMin Bank. Http://www.lot.kbstar.com. Web site, 11 2003.

[5] R. B. D'Agostino. Transformation to Normality of the Null Distribution of g_1. *Biometrika*, 57:679–681, 1970.

[6] R. B. D'Agostino. Approaches to the Null Distribution of $\sqrt{b}_1$. *Biometrika*, 60:169–173, 1973.

[7] R. B. D'Agostino. Tests for Departure from Normality. Empirical Results for the Distribution of b_2 and $\sqrt{b}_1$. *Biometrika*, 60:613–622, 1973.

[8] D.E.A. Giles. A Saddle Point Approximation to the Distribution Function of the Anderson-Daring Test Statistic. *Econometrics Working Paper*, 2000. Univ. of Victoria.

[9] R. C. Grary. Testing for Normality. *Biometrika*, 34:209–242, 1947.

[10] H. C. Hamaker. Approximating the Cumulative Normal Distribution and Its Inverse. *Appl Stat*, 27(1):76–77, 1978.

[11] H. W. Lilliefors. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *JASA*, 1:399–402, 1967.

[12] S. W. Looney and Jr T. R. Gulledage. Probability Plotting Positions and Goodness-of-fit for the Normal Distribution. *Stat*, 34:297–303, 1985.

[13] F.J. Massey. The Kolomogorov-Smirnov Test for Goodness of Fit. *JASA*, 46:68–78, 1951.

[14] National Council of Teachers of Mathematics. *Curriculum and Evaluation Standards for School Mathematics.* NCTM, 1989.
[15] D. B. Owen. *The Handbook of Statistical Tables. Reading Massachusetts:.* Addison Wesley, 1962.
[16] E. S. Pearson and H. O. Hartley. Tables for Statisticians. *Biometrika*, 1966.
[17] M. Rouncefield. Is it Normal ?, How to Use Probability Paper. *Teaching Stat*, 12(1):6–8, 1990.
[18] T. A. Ryan. Note on a Test for Normality. *Technometrics*, 10:825–839, 1990.
[19] T. A. Ryan and B. L. Joiner. Normal Probability Plots and Tests for Normality. *Technical Report*, 1976.
[20] M. B. Wilk S. S. Shapiro and H. J. Chan. A Comparative Study of Various Tests for Normality. *JASA*, 63:1343–1372, 1968.
[21] K. Sandrock. A Common Pitfall in the Use of the Kologorov-Smirnov One Sample Test. *Teaching Stat*, 12(1):9–11, 1990.
[22] S. S. Shapiro and M. B. Wilk. An Analysis of Variance Test for Normality. *Biometrika*, 52:591–606, 1965.

Sang-Gone Lee
Department of Mathematics Education,
Chonbuk National University,
Jeonju, Korea
E-mail: sgl@chonbuk.ac.kr