

Visual EDF Software to Check the Normality Assumption

Maria E. Calzada and Stephen M. Scariano
Loyola University New Orleans
New Orleans, LA 70118
calzada@loyno.edu

Abstract:

A host of materials in an introductory level statistics course relies on the “normality assumption.” However, assessing normality is a very subtle and difficult task, even for expert data analysts. To aid our student’s understanding we have written a program that implements a visual version of the Lilliefors test for normality and the powerful Anderson-Darling test for normality. In this paper we introduce these tests and show their usefulness through examples. The software, which is freely available to colleagues and students at <http://www.loyno.edu/~calzada>, should be used as a complement to other statistical techniques, like histograms and probability plots.

Introduction:

The following problem comes from David S. Moore’s *The Basic Practice of Statistics* [1], page 382:

Our subjects are 11 people diagnosed as being dependent on caffeine. Each subject was barred from coffee, colas, and other substances containing caffeine. Instead, they took capsules containing their normal caffeine intake. During a different time period, they took placebo capsules. The order in which subjects took caffeine and the placebo was randomized. Table 7.3 (Table 1 here) contains data on two of several tests given to the subjects. “Depression” is the score on the Beck Depression Inventory. Higher scores show more symptoms of depression. “Beats” is the beats per minute the subject achieved when asked to press a button 200 times as quickly as possible. We are interested in whether being deprived of caffeine affects these outcomes.

Table 1. Information from Table 7.3 in Moore [1].

Subjects	Depression (Caffeine)	Depression (Placebo)	Difference Depression	Beats (Caffeine)	Beats (Placebo)	Difference Beats
1	5	16	-11	281	201	80
2	5	23	-18	284	262	22
3	4	5	-1	300	283	17
4	3	7	-4	421	290	131
5	8	14	-6	240	259	-19
6	5	24	-19	294	291	3
7	0	6	-6	377	354	23
8	0	3	-3	345	346	-1
9	2	15	-13	303	283	20
10	11	12	-1	340	391	-51
11	1	0	1	408	411	-3

The Moore text is generally very effective in explaining basic statistical concepts. We have enjoyed teaching from this book, and the choice of this example is not meant as a critique of the book but as an illustration of the difficulties our introductory students have with the “normality assumption,” the assumption that the data set they are studying comes from a normal distribution. The cited problem asks the students to perform a hypothesis test for the difference in “Depression” scores (column 4 in Table 1), since these differences are not strikingly nonnormal. However, the problem guides the students to discard a hypothesis test for the difference in the “Beats” scores (column 7 in Table 1) because these data are “clearly” nonnormal. Figure 1 presents the default histogram a student would get for the difference in “Depression” scores using the statistical software SPSS. Figure 2 presents the default histogram a student would get for the difference in “Beats” scores using the statistical software SPSS. While most statistics instructors would take a closer look at these histograms, we believe most statistics students would say that Figure 2 is more “bell-shaped” than Figure 1. The argument that the t procedures are valid for the “Depression” data and not for the “Beats” data would make them very uneasy.

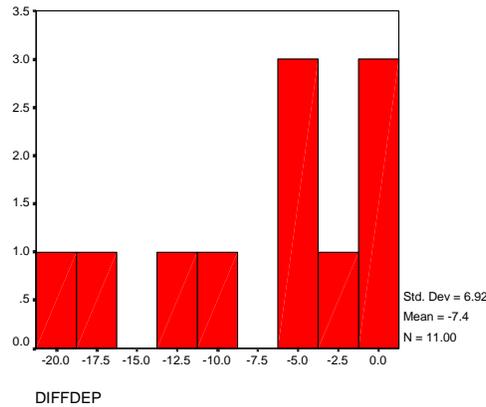


Figure 1. Histogram of the difference in “Depression” scores from Table 1.

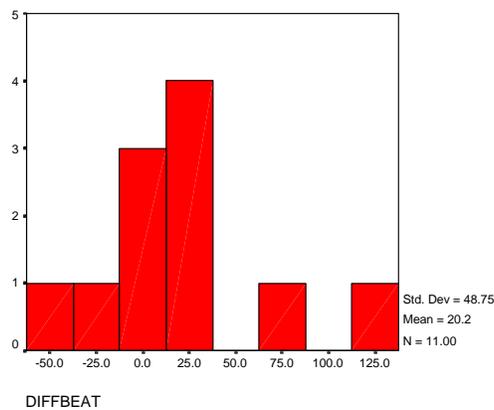


Figure 2. Histogram of the difference in “Beats” scores from Table 1.

Figures 3 and 4 present normal probability plots for the difference in “Depression” scores and the difference in “Beats” scores, respectively, as produced by SPSS. Strong departures from the line $y = x$ correspond to nonnormal characteristics. While Figure 4 shows a stronger departure from this straight line, Figure 3 also shows some non-linear characteristics. Since no real data set will produce a perfectly straight line, students are often called to make “educated-guesses” on their plots. We have observed that they sometimes overreact to minor departures from a linear trend, while in other instances they miss important departures from normality.

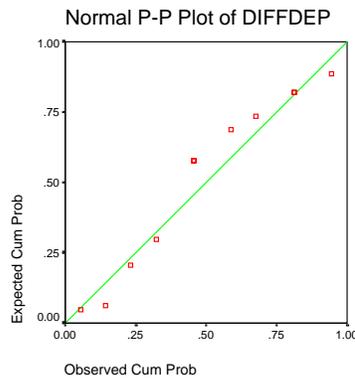


Figure 3. Normal Probability Plot of the difference in “Depression” scores from Table 1.

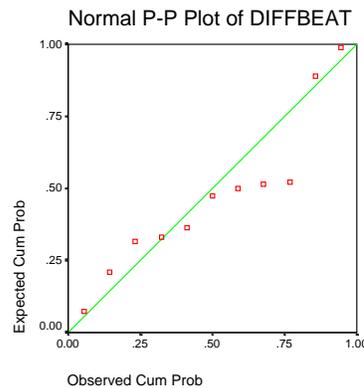


Figure 4. Normal Probability Plot of the difference in “Beats” scores from Table 1.

To aid our students in their analysis of data, we have written a Visual Basic program (EDF.exe) that implements two Empirical Distribution Function (EDF) distance tests: the Lilliefors and the Anderson-Darling tests for normality. These are presented in the form of hypothesis tests, so they can be introduced in the statistics class as soon as hypothesis testing has been covered. We have found that students can easily interpret the results from these tests, which are presented next, with the aid of our program. Interested readers can learn more about the Lilliefors and Anderson-Darling tests from

the sources listed in the references.

The Lilliefors Test for Normality

Lilliefors [2] suggested a modification of the Kolmorov-Smirnov goodness-of-fit test that can be used to test the normality assumption when the population parameters are not specified. The Lilliefors test is very intuitive and has the added advantage of having a visual implementation in our program.

To test the hypotheses

H_0 : The data distribution is a normal distribution

H_1 : The data distribution is some nonnormal distribution,

perform the following steps.

- Given data values x_1, x_2, \dots, x_n , form the order statistics (smallest to largest) $x_{(1)}, x_{(2)}, \dots, x_{(n)}$. Let \bar{x} and s^2 denote the usual unbiased sample estimators of μ and σ^2 , that is

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \text{ and}$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Standardize each data value using $z_{(k)} = (x_{(k)} - \bar{x})/s$.

- Let $F_n(z)$ be the EDF of the data, that is $F_n(z) = (\# \text{ of } z_{(k)}\text{'s } \leq z)/n$, for every z . Plot $F_n(z)$ and $\Phi(z)$ in the same coordinate system, where $\Phi(z)$ is the standard normal cumulative distribution function. The Lilliefors test statistic, D^* , is the maximum vertical distance between $F_n(z)$ and $\Phi(z)$, that is

$$D^* = \sup_z \{|F_n(z) - \Phi(z)|, -\infty \leq z \leq \infty\}.$$

- For sample size n , reject H_0 in favor of H_1 at significance level α when and only when D^* exceeds the upper percentage point $D_\alpha^*(n)$. The percentile points $D_\alpha^*(n)$ have been approximated using Monte Carlo simulation for each sample size n . Our program uses the tabulations produced by Mason and Bell [3], and a simple analytic approximation to the significance achieved which was developed by Dallal and Wilkinson [4].

We can quickly decide between H_0 and H_1 by plotting $\Phi(z) - D_\alpha^*(n)$, $\Phi(z) + D_\alpha^*(n)$, $\Phi(z)$, and $F_n(z)$ in the same coordinate system. We fail to reject H_0 if the graph of $F_n(z)$ always lies between the graphs of $\Phi(z) - D_\alpha^*(n)$ and $\Phi(z) + D_\alpha^*(n)$, but reject H_0 in favor of H_1 otherwise.

Figure 5 presents the output of our program for the difference in ‘‘Depression’’ scores. We see that $F_n(z)$ is between $\Phi(z) - D_{.1}^*(11)$ and $\Phi(z) + D_{.1}^*(11)$ bands. The hypothesis that the data came from a normal distribution cannot be rejected, which agrees with the assessment given in Moore [1] concerning these data.

Figure 5 presents the output of our program for the difference in ‘‘Beats’’ scores. We see that $F_n(z)$ exceeds the $\Phi(z) - D_{.01}^*(11)$ and $\Phi(z) + D_{.01}^*(11)$ bands so we reject the null hypothesis of normality at the .01 significance level. In fact, the approximated significance level is about .008. This also agrees with Moore’s assessment that the

use of the t procedures for these data is suspect.

The Anderson-Darling Test for Normality:

Anderson and Darling [5,6] formulated their distance test based on the accumulating square distance function,

$$D(\psi) = n \int_{-\infty}^{\infty} [F_n(x) - \Phi(x)]^2 \psi(x) d\Phi(x),$$

which we have adapted for our purposes by using $F_n(x)$, the EDF of the standardized data, and $\Phi(x)$, the standard normal cumulative distribution function. Here, $\psi(x)$ is a weight function which accentuates the importance of different subsets of the x -axis.

The usual weight function chosen for the Anderson-Darling test for normality is $\psi(x) = [\Phi(x)(1 - \Phi(x))]^{-1}$. For a wide variety of alternatives with unknown parameters, Stephens [7] demonstrated that the Anderson-Darling test is typically more powerful than other “distance” tests when testing for normality, including the Lilliefors test for normality. Therefore, we decided to give our students this power by including the discrete sample Anderson-Darling test developed by Stephens [7] in our program. Shapiro [8] provides an excellent description of this test.

To test the hypotheses

H_0 : The data distribution is a normal distribution

H_1 : The data distribution is some nonnormal distribution,

perform the following steps.

- As before, given data values x_1, x_2, \dots, x_n , form the order statistics $x_{(1)}, x_{(2)}, \dots, x_{(n)}$, and standardize each data value using $z_{(k)} = (x_{(k)} - \bar{x})/s$.
- The $z_{(k)}$ provide standard normal cumulative probabilities using $p_k = \Phi(z_{(k)})$.
- The discrete Anderson-Darling sample test statistic is

$$A = \frac{-1}{n} \left[\sum_{i=1}^n (2i - 1) [\ln(p_i) + \ln(1 - p_{n+1-i})] \right] - n$$

which is modified by computing $A^* = A(1 + \frac{0.75}{n} + \frac{2.25}{n^2})$.

- If A^* exceeds the selected critical values tabulated in Stephens [7] or Shapiro [8], H_0 is rejected in favor of H_1 at the appropriate significance level. Nelson [9] gives the formula

$$\alpha \approx 3.6789468e^{-A^*/0.1749916}$$

for approximating the actual significance level achieved and we have incorporated this into the program.

Looking again at Figures 4 we see that, as expected, the Anderson-Darling Test does not reject the assumption that the difference in “Depression” scores are normally distributed. Figure 5 confirms, again as expected, that the difference in “Beats” scores are nonnormal based on the Anderson-Darling Test.

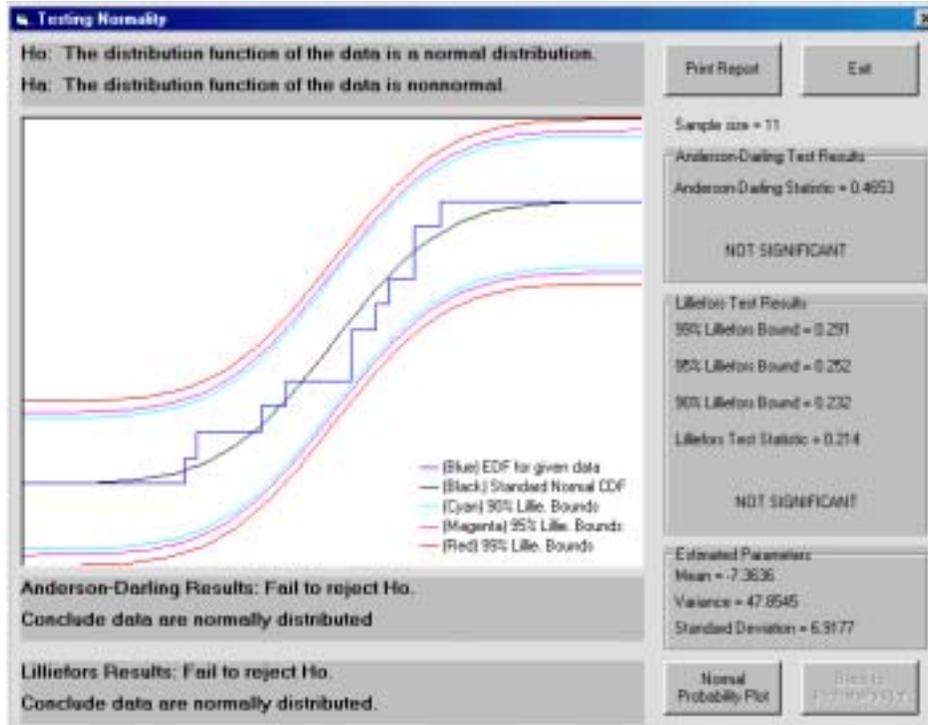


Figure 5. Output from EDF.exe for the difference in “Depression” scores from Table 1.

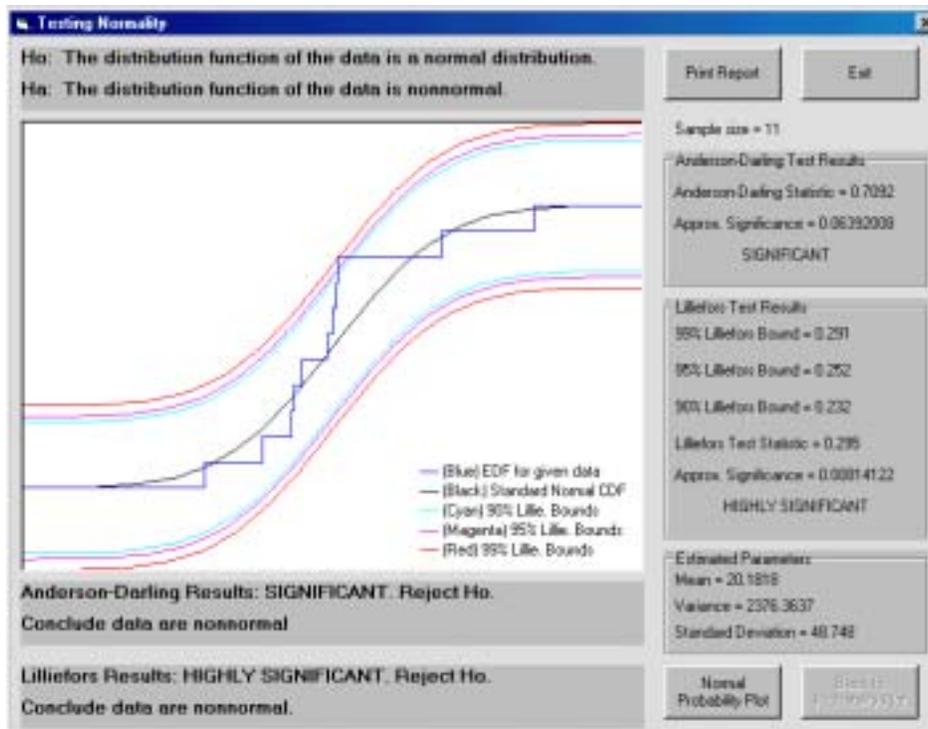


Figure 6. Output from EDF.exe for the difference in “Beats” scores from Table 1.

Conclusions

The Lilliefors and Anderson-Darling tests are important tools that should be routinely used alongside probability plots and histograms to check the “normality assumption” for sample data. These tests, like other statistical techniques, must be used sensibly and carefully, not in a vacuum. The stand-alone Windows application we offer enhances the craft of exploratory data analysis by encouraging students to gain deeper insight into the nature of distributional assumptions.

References

- [1] Moore, D. S. *The Basic Practice of Statistics, 2nd ed.*, W. H. Freeman and Company, New York, **2000**.
- [2] Lilliefors, H. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of the American Statistical Association*, **1967**, 62, 399-402.
- [3] Mason, A. L.; Bell, C. B. New Lilliefors and Srinivasan Tables with Applications. *Communications in Statistics—Simulation and Computation*, **1986**, 15, 451-477.
- [4] Dallal, G. E.; Wilkinsons, L. An Analytic Approximation to the Distribution of Lilliefors’s Test Statistic for Normality. *The American Statistician*, **1986**, 40, 294-296.
- [5] Anderson, T. W.; Darling, D. A. Asymptotic Theory of Certain Goodness of Fit Criteria Based on Stochastic Processes. *Annals of Mathematical Statistics*, **1952**, 23, 193-212.
- [6] Anderson, T. W.; Darling, D. A. A Test of Goodness of Fit. *Journal of the American Statistical Association*, **1954**, 49, 765-769.
- [7] Stephens, M. A. EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association*, **1974**, 69, 730-737.
- [8] Shapiro, S. S. *How to Test Normality and Other Distributional Assumptions*. American Society for Quality Control, Statistics Division, **1990**.
- [9] Nelson, L. S. The Anderson-Darling Test for Normality, *Journal of Quality Technology*, **1998**, 30, 298-299.