Journal of ISOSS
2017 Vol. 3(1), 35-44

NORMALITY TESTS: A BRIEF REVIEW

Munir Ahmad[1] and **Alya Al-Mutairi**[2]

[1] National College of Business Administration & Economics
Lahore, Pakistan. Email: munirahmaddr@yahoo.co.uk

[2] Department of Mathematics, Faculty of Science, Taibah University
Madina, Kingdom of Saudi Arabia Email: afaaq99@ hotmail.com

1. INTRODUCTION

> "*Normality is one of the most common assumptions made in the development and use of statistical procedures*."
>
> (Thode, 2002).

Researchers have been developing normality tests to check if observations follow a normal model (See Pearson et al., 1977; White and Macdonald, 1980; Pierce and Gray, 1982; Kuiper, 1960; Keya and Imon, 2016 and Jarque and Bera, 1980, 1987).

In the derivation of some normality tests, researchers have used the third and fourth moments of the random variables (See D'Agostino and Pearson, 1973; and Bowman and Shenton, 1975). Jarque and Bera (1980, 1987) and White and MacDonald (1980) used their tests for residuals when the errors are identically and independently distributed.

2. TESTS OF NORMALITY

Some of the tests that are being used in many fields are

- i) Anderson–Darling Test, (1952, 1962),
- ii) Cramér–Von Mises Criterion, (1928)
- iii) Watson test, (1961)
- iv) D'Agostino's K-Squared Test, (1986)
- v) Finniben Test, (1975)
- vi) Jarque–Berta Test, (1980)
- vii) Kolmogorov–Smirnov Test, (1933 & 1948)
- viii) Lilliefors test, (1967)
- ix) Normal Probability Plot, (1983)
- x) Pearson's Chi-Squared Test, (1900)
- xi) Quantile–Quantile Plot (Q-Q Plot), (1968)
- xii) Shapiro–Wilk Test, (1965)
- xiii) Shapiro–Francia Test, (2007)
- xiv) Weisberg-Bingham Test, (1975)
- xv) Kuiper Test, (1960)
- xvi) $Z-$, $t-$, χ^2 and F-Tests

Short notes on the tests are given below. References are also provided for consultations of readers.

2.1 Anderson–Darling Test

The Anderson–Darling test identifies departures from normality and is useful for the empirical distribution function (EDF). The EDF statistic measures the distance between the assumed CDF F and EDF F_n by

$$n\int_{-\infty}^{\infty} \left(F_n(x)-F(x)\right)^2 w(x)\,dF(x),$$

where $w(x)$ is a weight function. When the weight function is $w(x)=1$, then it becomes the Cramér–von Mises statistic. When the weight function is $w(x)=\left[F(x)\left(1-F(x)\right)\right]^{-1}$, then the Anderson–Darling (1954) test is

$$A=n\int_{-\infty}^{\infty} \frac{\left(F_n(x)-F(x)\right)^2}{F(x)\left(1-F(x)\right)}\,dF(x),$$

Stephens [1986] considers A^2 to be a good EDF statistic for detecting departure from normality and finds that the test provides better results if the parameters of models are calculated from the observations.

2.2 Cramér–Von Mises (C-vM) Criterion (1928)

The Cramér–von Mises criterion is a goodness of fit test of F^* compared to an empirical F_n (called one sample case), or for comparing two empirical distributions (called 2 sample case). It is defined as

$$w^2=\int_{-\infty}^{\infty} \left[F_n(x)-F^*(x)\right]^2 dF^*(x)$$

Let $x_1, x_2, \ldots, x_n$ be the data set, arranged in increasing order. Then the statistic is

$$T=nw^2=\frac{1}{12n}+\sum_{i=1}^{n}\left[\frac{2i-1}{2n}-F(x_i)\right]^2.$$

If T a calculated value from the data set is larger than the tabulated value (using Anderson & Darling's table), then the hypothesis is rejected. The C–vM criterion is similar to that of the Kolmogorov–Smirnov test.

2.2.1 Cramér–von Mises Test (Two Samples)

Let $x_1, x_2, \ldots, x_N$ be a data set of the first sample and $y_1, y_2, \ldots, y_M$ be the data set of the second sample. Both data sets are arranged in ascending order. Consider $T_1, T_2, \ldots, T_N$ to be the positions of the $x's$ and $s_1, s_2, \ldots, s_M$ be the positions of the $y's$ in the combined sample. Anderson (1961) derived

$$T = Nw^2 = \frac{U}{NM\left(N+M\right)} - \frac{4MN-1}{6\left(M+N\right)}$$

where U is

$$U = N\sum_{i=1}^{N} \left(r_i - i\right)^2 + M\sum_{j=1}^{M} \left(s_j - j\right)^2$$

If the computed value of T is larger than the table value of T (See Anderson, 1961), then the hypothesis that the two samples do not come from the same distribution is not accepted.

2.3 Watson Test

The Watson test [1961] is an altered form of the Cramér–von Mises test where

$$U^2 = T - n\left(\bar{F} - \tfrac{1}{2}\right)^2,$$

and

$$\bar{F} - \frac{1}{n}\sum F\left(x_i\right).$$

2.4 D'Agostino's K-Squared Test (1986)

The D'Agostino's K-squared test is derived using sample kurtosis and skewness and considered the transformation for sample skewness, g_1

$$Z_1\left(g_1\right) = \delta.\ln\left(\frac{g_1}{\alpha\sqrt{\mu_2}} + \sqrt{\frac{g_1^2}{\alpha^2\mu_2} + 1}\right),$$

where α and δ are constants and are calculated as

$$W^2 = \sqrt{2\gamma_2 + 4} - 1,$$

$$\delta = 1\big/\sqrt{\ln W},$$

$$\alpha^2 = 2\big/\left(W^2 - 1\right),$$

$\mu_2 = \mu_2\left(g_1\right)$ and $\gamma_2 = \gamma_2\left(g_1\right)$ are the variance and the kurtosis of g_1 respectively. Similarly, Anscombe & Glynn (1983) propose $Z_2\left(g_2\right)$ for g_2, and good for large samples.

D’Agostino et al. (1990) showed that Z_1 and Z_2 can be pooled as $K^2 = Z_1\left(g_1\right)^2 + Z_2\left(g_2\right)^2$ to test for normality.

K^2 is $\approx \chi^2 -$ with 2 degrees of freedom when the null hypothesis is normal.

2.5 Filliben Test, (1975)

The Filliben test statistic is defined as

$$r = \frac{\sum\left(X_i - \bar{X}\right)\left(M_i - \bar{M}\right)}{\left[\sum\left(X_i - \bar{X}\right)^2 \sum\left(M_i - \bar{M}\right)^2\right]^{1/2}},$$

for testing normality, where X_i $(i = 1, 2, 3, \ldots, n)$ is observed order statistics, r correlation coefficient and $M_i\,(i = 1, 2, 3, \ldots, n)$ is the median from a $N(0,1)$ distribution.

Filliben (1975) used the median of i^{th} order statistic Y_i instead of mean. It has advantage that location and scale parameters may not be known. The test is also applicable to non-normal distributions.

2.6 Jarque–Berta (JB) Test, (1980)

The Jarque-Bera (1980, 1987) test is normal testing procedure. The test involves sample skewness and kurtosis. The test has a reasonable power for small and large samples but it has low power for short tail distributions.

The J-B statistic is $JB = n\left[(b_1)/6 + (b_2 - 3)^2 / 24\right]$, where n denotes the sample size, b_1 the skewness and b_2 denotes the kurtosis. A large value of JB may show rejection of hypothesis.

2.7 Kolmogorov–Smirnov Test, (1933 & 1948)

The Kolmogorov–Smirnov test (K–S test) is a nonparametric test for comparing an empirical distribution with an assumed distribution (one-sample case), or comparing two empirical distributions. The K-S test may show whether the empirical distribution and the assumed distribution are similar or not. The empirical distribution function F_n is

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{[-\infty, x]}(X_i)$$

where $I_{[-\infty, x]}(X_i) = 1$ if $X_i \le x$ and = 0 otherwise.

The K-S test is

$$D_n = \sup_x \left|F_n(x) - F(x)\right|,$$

where $\sup_x$ is the supremum of the differences.

D_n tends to 0 almost surely. In practice, large observations are needed to reject the null hypothesis.

2.8 Lilliefors Test (1967)

The one-sample K-S is not advantageous in practice as it needs known mean and variance. In many cases mean and variance are unknown and as such K-S test is not a good test. Lilliefors (1967, 1969) established the test which is similar to K-S test except that mean and variance are unknown. The test is defined as

$$D_n = Max\left(\left|F(z) - S(z)\right|,\ \left|F(z) - S'(z)\right|\right),$$

where $Z = \left(X - \bar{X}\right)/S$, $S(z)$ is the proportion of values less than or equal to z. If D_n is greater than the critical value, the null hypothesis is prone to rejection.

Lilliefors (1967, 1969) computed critical values and is available in the Lilliefors table (See www.real-statistics.com/statistics-tables/lilliefors-test-table**)**.

2.9 Normal Probability Plot (1983)

The normal probability plot (also called ‘Normal plot’) is used to test normality. If the data when plotted on a special graph paper, look close to a straight line, the observations follow the normal model otherwise the departure from the line displays deviation from normality.

The ‘Normal probability plot’ is like Q-Q plot. Researchers may plot the data on the vertical axis (Chambers, et al., 1983) or plot on the horizontal axis (Box and Draper, 2007; and Titterington et al., 1985).

Normal plots are used on 7 or lower points but it is difficult to differentiate between random variability and departure from normality.

2.10 Pearson's Chi-Squared Test (1900)

Pearson's chi-squared test $\left(\chi^2\right)$ is generally used to check two types of tests of hypotheses viz.

i) test of goodness of fit.
ii) test of independence.

It is a widely used test and its properties are considered by Pearson in 1900.

The test-statistic is

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(O_i - E_i\right)^2}{E_i} = N\sum_{i=1}^{n} p_i \left(\frac{O_i/N - p_i}{p_i}\right)^2$$

where

O_i = the i^{th} observations,
N = total number of observations, and
$E_i = Np_i$ = the i^{th} expected frequency.

If the calculated χ^2-value is larger than its table value then the null hypothesis is rejected.

2.11 Quantile-Quantile Plot (QQ Plot) (1968)

In the Q–Q plot, the quantiles of two distributions are plotted against each other and if the points of two distributions are linearly related, or lie on the line $y = x$ or approximately lie on the line, they will be assumed to be similar.

The Q–Q plot is a non-parametric method. A Q–Q plot is assumed to be more powerful than method of comparing the two distributions. The Q–Q plots are also used to compare two data sets.

2.12 Shapiro-Wilk Test

The Shapiro–Wilk test uses the frequency or proportion of the data and tests its normality.

The Shapiro–Wilk test statistic is:

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

where

$x_{(i)}$ is the i^{th} order statistic,

$\bar{x} = \left(x_1 + ... + x_n\right)/n$ is the sample mean,

$a_i, i = 1, 2, \ldots n$ are given by $\left(a_1, ..., a_n\right) = \dfrac{m^T V^{-1}}{\left(m^T V^{-1} V^{-1} m\right)^{1/2}}$

where $m = \left(m_1, ..., m_n\right)^T$ and $m_1, ..., m_n$ are the expected order statistics and V is the covariance matrix of the order statistics sampled from the normal population. Shapiro–Wilk test has the best power as compared to Anderson–Darling, Kolmogorov–Smirnov, Lilliefors, and Anderson–Darling tests (Razali and Bee, 2011).

Sometimes a Q–Q plot is used for confirmation besides the Shapiro–Wilk test.

2.13 Shapiro–Francia test (1972)

The Shapiro–Francia test is comparable to the Shapiro-Wilk test for large data.

The test statistic is

$$W' = \left(\sum_i c_i X_i\right)^2 \Big/ \sum_i \left(X_i - \bar{X}\right)^2,$$

where $\bar{X} = \sum_i X_i / n$ is the sample mean; c_i is $c_i = \tilde{m}_i \sqrt{\left(\tilde{m}'\tilde{m}\right)}$, $\hat{m}_i$ is $\hat{m}_i = \Phi^{-1}\left\{\left(i - 3/8\right)/\left(n + 1/4\right)\right\}$ and $\Phi^{-1}\{.\}$ is the inverse normal cdf.

2.14 The Weisberg-Bingham Test

The Weisberg-Bingham test is a modified form of the Shapiro-Francia test to make it appropriate for machine calculation (Weisberg and Bingham, 1975).

2.15 The Kuiper's Test

Kuiper's test is close to the Kolmogorov–Smirnov test (Kuiper, (1960). Kuiper uses $D^{+} + D^{-}$ as the test statistic and tests circular probability distributions.

The Kuiper test statistic, V, is defined as

$$V = D^{+} + D^{-} \text{ where } D^{+} = \max\left[i/n - z_i\right], \text{ and } D^{-} = \max\left[z_i - (i-1)/n\right],$$

$$\text{and } z_i = F(x_i), \quad x_i\,(i = 1,\ldots,n),$$

F is the continuous cdf under the null hypothesis. Tables for the critical points of the test statistic are available at Pearson and Hartley, (1972).

2.16 $Z, t-, F-$ and χ^2 Tests

The $Z, t-, F-$ and χ^2 tests are the classical tests widely used in statistics. These are available in any under graduate books.

3. CONCLUSION

All tests are useable in testing normality, but we shall be vigilant when using the JB test in small-sample cases. In such cases, the significance level of the test can be erroneous.

The JB test may have low power biased in finite samples.

If the regression errors are auto-correlated or heteroskedastic, Royston (1992)'s approach may overcome this shortcoming.

The K-S test has benefits over the chi square test. The K-S statistic can use small data set where the chi square test would be uncertain for small data set. However K-S test has more power than the chi-square test for any sample size.

REFERENCES

1. Ahmad, F. and Khan, R.A. (2015). Power Comparison of Various Normality Tests. *Pak. J. Stat. Oper. Res.*, XI(3), 331-345.
2. Anderson, T.W. (1962). On the Distribution of the Two-Sample Cramer-von Mises Criterion. *Ann. Math. Statist.*, 33(3), 1148-1159.
3. Anderson, T.W. and Darling, D.A. (1952). Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23, 193-212.
4. Anscombe, F.J. and Glynn, W.J. (1983). Distribution of the kurtosis statistic b_2 for normal statistics. *Biometrika*, 70(1), 227-234.
5. Bera, A.K. and Jarque, C.M. (1982). Model specification tests: A simultaneous approach. *J. Econometrics*, 20, 59-82.

6. Blom G. (1958). *Statistical estimates and transformed beta-variables*. Wiley; New York.
7. Bowman, K.O. and Shenton, L.R. (1975). Omnibus contours for departures from normality based on V/bl and b2. *Biometrika*, 62, 243-250.
8. Box, G.E. and Draper, N.R. (2007). *Response surfaces, mixtures, and ridge analyses* (Vol. 649). John Wiley & Sons.
9. Chambers, J.M., Cleveland, W.S., Kleiner, B. and Tukey, P.A. (1983). *Graphical methods for data analysis* (Vol. 5, No. 1). Belmont, CA: Wadsworth.
10. Cramér, H. (1928). On the Composition of Elementary Errors. *Scandinavian Actuarial Journal*, 1928(1), 13-74.
11. D’Agostino, R.B. and Stephens, M.A. (1986). Goodness of fit Technique, Marcel Dekker Inc., New York.
12. D'Agostino, R.B. and Pearson, E.S. (1973). Tests for departure from normality: Empirical results for the distributions of b_2 and $\sqrt{b_1}$. *Biometrika* 60, 613-622. Correction (1974), 61, 647.
13. D'Agostino, R.B. (1970). Transformation to normality of the null distribution of g1. *Biometrika*, 57(3), 679-681.
14. D'Agostino, R.B. (1971). An omnibus test for normality for moderate and large size samples. *Biometrika*, 58, 341-348.
15. D'Agostino, R.B. (1986). Tests for the Normal Distribution. In D'Agostino, R.B. and Stephens, M.A. *Goodness-of-Fit Techniques*. Marcel Dekker, New York.
16. D'agostino, R.B., Belanger, A. and D'Agostino Jr, R.B. (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44(4), 316-321.
17. Das, K.R. and Imon, A.H.M.R. (2016). A brief review of tests for normality. *American Journal of Theoretical and Applied Statistics*, 5, 5-12.
18. Field, A. (2009). Discovering statistics using SPSS (3rd ed.). Los Angeles [i.e. Thousand Oaks, Calif.]: SAGE Publications. p. 143.
19. Filliben, J.J. (1972). Techniques for tail length analysis. In *Proceedings of the Eighteenth Conference on the Design of Experiments in Army Research and Testing*, Durham: US Army Research Office.
20. Filliben, J.J. (1975). The probability plot correlation coefficient test for normality. *Technometrics*, 17(1), 111-117.
21. Gnanadesikan, R. (1977). *Methods for statistical data analysis of multivariate observations*. New York: Wiley, p 199.
22. Jarque, C.M. and Bera, A.K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3), 255-259.
23. Jarque, C.M. and Bera, A.K. (1987). A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, 55(2), 163-172.
24. Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione, *G. Ist Ital. Attuari.*, 4, 83-91.
25. Kuiper, N.H. (1960). Tests concerning random points on a circle. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen*, Series A 63, 38-47.

26. Lilliefors, H. (1967). On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62. 399-402.
27. Lilliefors, H. (1969). On the Kolmogorov–Smirnov test for the exponential distribution with mean unknown. *Journal of the American Statistical Association*, 64, 387-389.
28. McDonald, J.H. (2014). G–test of goodness-of-fit. *Handbook of Biological Statistics* (Third Ed.). Baltimore, Maryland: Sparky House Publishing, 53-58.
29. Pearson, E.S., D'Agostino, R.B. and Bowman, K.O. (1977). Tests for departure from normality: Comparison of powers. *Biometrika*, 64, 231-246.
30. Pearson, E.S. and Hartley, H.O. (1972). *Biometrika Tables for Statisticians*, Volume 2, CUP. ISBN 0-521-06937-8.
31. Pearson, Egon S. (1931). Note on tests for normality. *Biometrika*, 22(3/4), 423-424.
32. Pierce, D.A. and Gray, R.J. (1982). Testing normality of errors in regression models. *Biometrika*, 69, 233-236.
33. Razali, N.M. and Wah, Y.B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21-33.
34. Royston, P. (1992). Approximating the Shapiro–Wilk W test for non-normality. *Statist. Comput*., 2, 117-119.
35. Shapiro, S.S. and Francia, R.S. (1973). An approximate analysis of variance test for normality. *J. Amer*. *Statist. Assoc*., 67, 215-216.
36. Shapiro, S.S. and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591-611.
37. Shapiro, S.S., Wilk, M.B. and Chen, H.J. (1968). A comparative study of various tests for normality. *J. Am. Statist. Assoc*., 63, 1343-1372.
38. Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.
39. Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *Annals of Statistics*, 19, 279-281.
40. Sokal, R.R. and Rohlf, F.J. (1981). Biometry: The Principles and Practice of Statistics in Biological Research (Second ed.). New York: Freeman. ISBN 0-7167-2411-1.
41. Stephens, M.A. (1986). Tests Based on EDF Statistics. In D'Agostino, R.B. and Stephens, M.A. *Goodness-of-Fit Techniques*. New York: Marcel Dekker. ISBN 0-8247-7487-6.
42. Thode, H.C. Jr. (2002). Quantile-Quantile Plots. Section 2.2.2, *Testing for Normality*, Marcel Dekker, New York, USA. p. 21.
43. Titterington, D.M., Smith, A.F.M. and Makov, V.E. (1985). Learning about the parameters of a mixture. *Statistical analysis of fixed mixture distribution*. John Wily & Sons, Chichester New York Brisbane Toronto Singapore, 52-147.
44. Von Mises, R.E. (1928). Wahrscheinlichkeit, Statistik und Wahrheit. Julius Springer.
45. Watson, G.S. (1961). Goodness-Of-Fit Tests on a Circle. *Biometrika*, 48(1/2), 109-114.
46. Weisberg, S. (1980). Comment on paper by H. White and G.M. MacDonald. *J. Am. Statist. Assoc*., 75, 28-31.

47. Weisberg, S. and Bingham, C. (1975). An approximate analysis of variance test for non-normality suitable for machine calculation. *Technometrics*, 17, 133-134.
48. White, H. and MacDonald, G.M. (1980). Some large sample tests for non-normality in the linear regression model. *J. Am. Statist. Assoc*., 75, 16-28.
49. Wilk, M.B. and Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data, *Biometrika*, 55(1), 1-17.