Shapiro Wilk Test of Normality

Asal Rahmani Lahoot
Arman Jahangiri
Shiyu Sun
Jessica Grant

November 24^{th}, 2023

UNIVERSITY OF CALGARY

Outline

Outline

Introduction

- **Foundation:** Introduced in 1965 by Shapiro and Wilk for small sample sizes ($n \leq 50$)
- **Evolution:** Enhanced in 1968 to improve the W statistic's usability across statistical domains.
- **Comparative Analysis:** The 1968 comparative study underscored the test's sensitivity against nine other normality tests.
- **Contextual Expansion:** D'Agostino's 1971 test offered an alternative for moderate to large samples, enriching the methodologies for normality testing.

Introduction

- Random sample $X_1, X_2, ..., X_n$ with some unknown distribution function $F(x)$
- Test Statistic:

$$W = \frac{1}{D}\left[\sum_{i=1}^{k} a_i(X^{(n-i+1)} - X^{(i)})\right]^2$$

or equivalently

$$W = \frac{1}{D}\left[\sum_{i=1}^{n} a_i X^{(i)}\right]^2$$

where $D = \sum_{i=1}^{n}(X_i - \bar{X})^2$

Introduction

- Hypotheses:
  H_0: The sample comes from a normally distributed population with unspecified mean and variance
  H_1: The sample does not come from a normally distributed population
- Reject Region:
  If $W \leq$ Critical value found in Table A17, you would reject H_0 at a particular level of significance α
- p-value:
  Use Table A18 to approximate z-value $\rightarrow$ Use Table A1

Outline

Derivation of Test Statistic

- We have already seen that

$$W = (\sum_{i=1}^{n} a_i X_{(i)})^2 / \sum_{i=1}^{n} (X_{(i)} - \overline{X})^2$$

Two important questions:

1. What is W? (Range, Extreme values, concept, etc.)

2. How to find $a_i's$? (disregarding the table given in the textbook)

Derivation of Test Statistic

Answer to the second question

- Define

$$a' = (a_1, \cdots, a_n) = \frac{m'V^{-1}}{(m'V^{-1}V^{-1}m)^{\frac{1}{2}}}$$

where m and V are the mean vector and variance-covariance matrix of an ordered standard normal random sample.

- **Lemma:**

$$\sum_{i=1}^{n} a_i^2 = a'a = 1 \quad \text{(proof available in the report)}$$

- **Corollary:** The maximum value of W is 1; hence, we have $0 \leq W \leq 1$.

$$\begin{aligned} 0 \leq W &= \frac{(\sum_{i=1}^{n} a_i X_{(i)})^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2} && \text{(Cauchy-Schwarz inequality)} \\ &\leq \frac{\sum_{i=1}^{n} a_i^2 \times \sum_{i=1}^{n} X_{(i)}^2}{\sum_{i=1}^{n} X_i^2 - n\overline{X}} && \text{(by the above Lemma)} \\ &= \frac{1 \times \sum_{i=1}^{n} X_{(i)}^2}{\sum_{i=1}^{n} X_{(i)}^2} = 1 && \text{(assuming a population with zero mean)} \end{aligned}$$

Interpretation of the Test Statistic

Answer to the first question

- W Statistic is basically the square of a correlation coefficient, where the Pearson correlation coefficient is computed between the order statistics $X_{(i)}$ in the sample and the scores a_i.

Question: Why?

$$\rho = \frac{\sum_{i=1}^{n}(X_{(i)} - \overline{X})(a_i - \overline{a})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(a_i - \overline{a})^2}}$$

where

$$\sum_{i=1}^{n}(a_i - \overline{a})^2 = \sum_{i=1}^{n} a_i^2 - n\overline{a}^2 = \sum_{i=1}^{n} a_i^2 - n(\sum_{i=1}^{n} a_i)^2/n^2 = 1 - 0 = 1$$

$$\Rightarrow W = \rho^2$$

Interpretation of the Test Statistic

- $0 \leq W \leq 1$;

- W is maximized when $X_{(i)} = \gamma . a_i$ for some γ (Cauchy-Schwartz)

- $W = \rho^2$,

  $\Rightarrow$ **Greater W implies evidence of normal distribution for the data.**

- shapiro.test(Random Sample)

Data	**W Value**	**p-value**
Normal $(\mu = 5, \sigma = 3)$	0.98826	0.9951
Uniform $(a = 2, b = 4)$	0.90965	0.06277
Poisson$(\lambda = 2)$	0.90417	0.0494
Binomial$(n = 100, p = 0.5)$	0.97223	0.8011
t(df=100)	0.97139	0.7839

Shapiro-Wilk Test Results

Reproduction Table A16

- Shapiro and Wilk have found an approximation for a' to resolve the problem of computational complexity of matrix inversion, provided below

$$a' = (a_1, \cdots, a_n) = \frac{m'V^{-1}}{(m'V^{-1}V^{-1}m)^{\frac{1}{2}}} = \frac{a^\star}{\sqrt{a^{\star T} a^\star}}$$

approximate $\hat{a}_i^\star$ using $2m_i$ for $i \in \{2, \cdots, n-1\}$, and using

$$\hat{a}_1^2 = \hat{a}_n^2 = \begin{cases} \frac{\Gamma(\frac{1}{2}n)}{\sqrt{2}\Gamma\frac{1}{2}(n+1)} & \text{if } n \leq 20 \\ \frac{\Gamma(\frac{1}{2}(n+1))}{\sqrt{2}\Gamma(\frac{1}{2}n+1)} & \text{if } n \geq 20 \end{cases}$$

0.3881448	0.3074076	0.2607904	0.2255492	0.1964327
0.1710355	0.1487100	0.1278294	0.1083968	0.0904921
0.0729026	0.0561627	0.0399988	0.0239285	0.0080125

Approximated Values Table

Outline

1. Student Math Scores Example

Student performances in a math exam, n=42, $k = \frac{n}{2} = 21$.

- The observations are as follows:

88	65	60	91	51	59	61	68	84	61
52	58	60	88	81	69	85	65	63	72
90	73	64	94	54	73	71	66	65	50
62	60	87	69	76	32	94	77	52	82
91	79								

- Testing the following:

H_0 : The sample comes from a normally distributed population vs
H_1 : The sample does not come from a normally distributed population

1. Student Math Scores Example

- Sorting data by increasing trend to get denominator

$$D = \sum_{i=1}^{n} (X^{(i)} - \bar{X})^2 = 8327.905$$

- For $n = 42$, $k = \frac{n}{2} = 21$, coefficients $a_1, a_2, ..., a_{21}$ are given by Table A16.

$$W = \frac{1}{D} \left[\sum_{i=1}^{k} a_i (X^{(n-i+1)} - X^{(i)}) \right]^2 = 0.9683587$$

From Table A17, the test statistic lies between the 0.10 and 0.50 quantile. By using interpolation, p-value is 0.38 approximately.

1. Student Math Scores Example

- To calculate G for a more precise p-value, use Table A18 to get coefficients (b_n, c_n, d_n) for $n = 42$.

$$G = b_n + c_n \ln\left(\frac{W - d_n}{1 - W}\right) = b_{42} + c_{42} \ln\left(\frac{W - d_{42}}{1 - W}\right) = -0.2953535$$

- From Table A1, it corresponds to $p = 0.384$, we do not reject H_0.
- Conclusion: There is insufficient evidence to conclude that the students math scores do not come from a normally distributed population.

2. Resident Income in California Example

The income of residents in California (thousand US dollars),$n = 48$, $k = 24$.

- The observations are as follows:

150.527	40.577	53.397	73.443	67.388	84.943	127.232	28.889
44.894	51.402	52.961	72.545	81.935	46.563	49.596	36.991
57.961	53.802	35.096	36.806	49.986	66.912	54.286	40.286
40.851	111.850	62.283	134.625	24.611	35.154	42.716	49.181
78.576	55.589	103.263	60.474	42.500	52.143	57.917	94.044
56.850	27.891	41.034	48.199	17.339	100.952	58.107	50.727

- Testing the following:

H_0 : The sample comes from a normally distributed population vs
H_1 : The sample does not come from a normally distributed population

2. Resident Income in California Example

- Use Table A16 to obtain coefficients, use Table A18 to calculate G for a more precise p-value.

$$W = \frac{1}{D}\left[\sum_{i=1}^{k} a_i(X^{(n-i+1)} - X^{(i)})\right]^2 = 0.8744595$$

$$G = b_{48} + c_{48} \ln \frac{W - d_{48}}{1 - W} = -3.718423$$

  From Table A1, the corresponding p-value is less than 0.001, so we reject H_0.

- Conclusion: There is sufficient evidence to conclude that the income of residents in California do not come from a normally distributed population.

Outline

Comparison with Other Methods

- **Goodness of Fit Tests**

  H_0: Data follows a specified distribution vs
  H_1: Data does not follow the specified distribution

- **Normality Tests**

  H_0 : Data follows a normal distribution vs
  H_1 : Data does not follow a normal distribution

- **Graphical Methods**

Goodness of Fit Tests

- Pearsons's Chi-Squared Goodness of Fit Test
- The Kolmogorov-Smirnov Test
- The Cramér–von Mises Test
- The Anderson-Darling Test

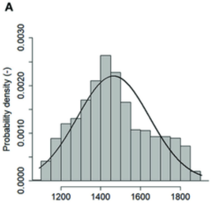
- **Advantages**
  - Can handle sample sizes $n \geq 50$
  - KS test rivals the SW test in terms of popularity
- **Disadvantages**
  - Requires fully specified distributions
  - CSQ is not recommended for continuous distributions

Normality Tests

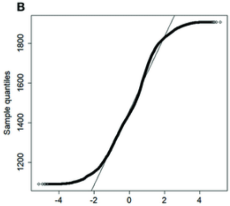
- Lilliefors Test
- D'Agostino-Pearson's K^2Test and Jarque–Bera Test
- Extensions Shapiro–Wilk Test
  - Shapiro–Wilk Test modified by Roysten
  - Shapiro–Francia Test

- **Advantages**
  - Can handle sample sizes $n \geq 50$
  - Tests for an unspecified normal distribution
  - SF test is a simplification of SW
- **Disadvantages**
  - Lilliefors known to perform worse than the SW, AD and CVM tests

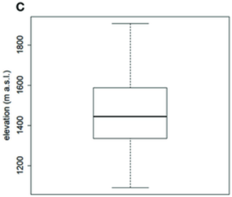
Graphical Methods



Histograms



QQ Plots



Boxplots

- **Advantages**
  - Simple to implement
- **Disadvantages**
  - Subjective

Outline

Conclusion

- The Shapiro Wilk test is a widely used test for normality.
- The test is simple to implement through the use of software or by hand with the use of table A16 (test statistic) and one of table A17 (RR) and table A18 (p-value).
- The test has been improved upon by from its original form to remedy key limitations, particularly the limitation of sample size.

Outline

References

[1] Anderson-Darling Test. 1.3.5.14. Anderson-Darling Test. (n.d.).

[2] Conover, W. J. (1971). Practical Nonparametric Statistics. (Illustrated). Wiley.

[3] D'Agostino, R. B. (1971). An Omnibus Test of Normality for Moderate and Large Size Samples. Biometrika, 58(2), 341–348.

[4] Dodge, Y. (n.d.). The concise encyclopedia of statistics. SpringerLink.

[5] García-Portugués, E. (n.d.). Notes for Nonparametric Statistics. 6.1 Goodness-of-fit tests for distribution models.

[6] Kolmogorov-Smirnov Goodness-of-Fit Test. 1.3.5.16. Kolmogorov-Smirnov goodness-of-fit test. (n.d.).

[7] MuonNeutrino. (2019, March 3). US Census Demographic Data. Kaggle.

[8] Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., and Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. Annals of cardiac anaesthesia, 22(1), 67–72.

References

[9] Shapiro, S. S., Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). Biometrika, 52, 3 and 4, p. 591.

[10] Shapiro, S. S., and Wilk, M. B. (1968). Approximations for the Null Distribution of the W Statistic. Technometrics, 10(4), 861–866. https://doi.org/10.2307/1267467

[11] Shapiro, S. S., Wilk, M. B., and Chen, H. J. (1968). A Comparative Study of Various Tests for Normality. Journal of the American Statistical Association, 63(324), 1343–1372.

[12] Schoonjans, F. (2023, August 13). Tests for normal distribution. MedCalc.

[13] Seshapanpu, J. (2018, November 9). Students performance in exams. Kaggle.

[14] Stephanie. (2017, November 2). Lilliefors test for normality and exponential distributions. Statistics How To.

[15] Yap, B. W., and Sim, C. H. (2011). Comparisons of various types of normality tests. Journal of Statistical Computation and Simulation, 81(12), 2141–2155.

Thank You!