THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

CHE384, From Data to Decisions: Measurement, Uncertainty, Analysis, and Modeling

Lecture 16
Shapiro-Wilk Test for Normality

Chris A. Mack
Adjunct Associate Professor

http://www.lithoguru.com/scientist/statistics/

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Shapiro-Wilk Test

- Null Hypothesis: The sample (of size n) comes from a normal distribution
- Test statistic:

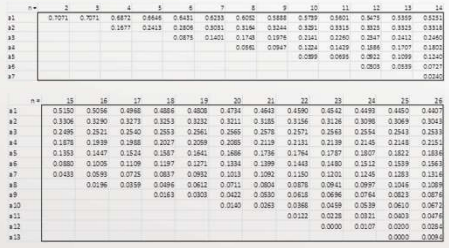
$$W = \frac{(\sum_{i=1}^{n} a_i x_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad x_i = i^{th}\ smallest\ value\ of\ x \qquad a_i = Shapiro - Wil\ \ constant$$

– Reject null hypothesis if $W < W_\alpha$ (p-value < α)

S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)", *Biometrika*, **52**(3–4), 591–611 (1965).

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Shapiro-Wilk Coefficients Table

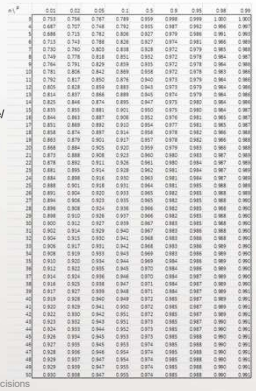


For more tables, see http://www.real-statistics.com/statistics-tables/shapiro-wilk-table/

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Shapiro-Wilk Critical Values

www.real-statistics.com/statistics-tables/shapiro-wilk-table/



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Shapiro-Wilk in R

- It is best to run the Shapiro-Wilk test using a statistical software package
- In R,
  - Let X = (X_1, . . . ,X_n) be the data vector
  - shapiro.test(X)
    - calculates W and the p-value
    - reject the null hypothesis of a normal distribution if the p-value is less than α, your significance level

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Properties of the Test

- $0 < W \leq 1$, defined for $n \geq 3$
- W is very similar to the correlation coefficient of a normal probability plot
- W is independent of location and scale (add or multiply the data by a constant and W doesn't change)
- The Shapiro–Wilk test has the best power for a given significance compared to other popular tests
- This test is sample-size biased
  - A normal probability plot should be used to confirm any test results

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Interpreting the Test

- Failure to reject the null hypothesis is not proof that the distribution is Normal
  - The Default is that the distribution is normal
  - It means that the data does not give enough evidence to reject this default assumption
  - Small sample sizes often don't have enough information to conclude much of anything
- Rejecting the null hypothesis does not tell you how much the distribution differs from normal
  - For a large data set, even a very small departure from normality will trigger a rejection

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Lecture 16: What have we learned?

- Why do people like the Shapiro-Wilk test for normality?
- Be able to run the Shapiro-Wilk test in R
- What are the difficulties in interpreting the results of any hypothesis test for normality?