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

Lecture 14
Testing for Kurtosis

Chris A. Mack
Adjunct Associate Professor

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

Kurtosis

- For any distribution, the kurtosis (sometimes called the excess kurtosis) is defined as

$$\gamma_2 = \varphi_4 - 3 \qquad \text{(old notation = } \beta_2\text{)}$$

- For a unimodal, symmetric distribution,
  - a positive kurtosis means "heavy tails" and a more peaked center compared to a normal distribution
  - a negative kurtosis means "light tails" and a more spread center compared to a normal distribution

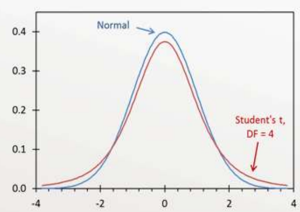
Kurtosis Examples

- For the Student's t distribution, the excess kurtosis is

$$\gamma_2 = \frac{6}{DF - 4}$$

for DF > 4 (for DF ≤ 4 the kurtosis is infinite)

- For a uniform distribution, $\gamma_2 = -\frac{6}{5}$



One Impact of Excess Kurtosis

- For a normal distribution, the sample variance will have an expected value of σ^2, and a variance of

$$var(s^2) = \frac{2\sigma^4}{n-1}$$

- For a distribution with excess kurtosis γ_2

$$var(s^2) = \frac{2\sigma^4}{n-1}\left(1 + \frac{n-1}{2n}\gamma_2\right)$$

Sample Kurtosis

- For a sample of size n, the sample kurtosis is

$$g_2 = \frac{m_4}{m_2^2} - 3 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^2\right)^2} - 3$$

- For large n, the sampling distribution of g_2 approaches Normal with mean 0 and variance of $24/n$
- For small samples, this estimator is biased

Sample Kurtosis

- An unbiased estimator of the sample excess kurtosis is

$$G_2 = \frac{n-1}{(n-2)(n-3)}[(n+1)g_2 + 6]$$

Standard Error:

$$SE(G_2) = 2\,SE(G_1)\sqrt{\frac{n^2-1}{(n-3)(n+5)}} \approx \sqrt{\frac{24}{n}}\left(1 - \frac{2}{n} + O\left(\frac{1}{n^2}\right)\right)$$

D. N. Joanes and C. A. Gill, "Comparing Measures of Sample Skewness and Kurtosis", *The Statistician*, **47**(1),183–189 (1998).

Sample Kurtosis Test

- Only perform the kurtosis test if the skewness test fails to reject the null hypothesis
- Null hypothesis: $\gamma_2 = 0$ (Normal distribution)
- Test statistic: $G_2/SE(G_2)$ is approximately standard Normal for $n > 20$
  - We generally perform a two-tailed test
- If the test statistic $G_2/SE(G_2)$ is beyond the critical z-value for our significance level we reject the null hypothesis that the distribution is Normal

Jarque-Berra Test

- The tests for skewness and kurtosis can be combined into one

$$\left(\frac{G_1}{SE(G_1)}\right)^2 + \left(\frac{G_2}{SE(G_2)}\right)^2 \sim \chi^2(2)$$

- For example, the 95% ($\alpha = 0.05$) critical value for $\chi^2(2)$ is 5.99 and the 99% critical value is 9.21

Impact of Outliers

- Both the Skewness test and the Kurtosis test are very sensitive outlier detectors
  - One outlier will make the distribution appear skewed
  - Two symmetric outliers will make the tails appear heavy
- More on outlier detection in the next lectures

Lecture 14: What have we learned?

- How is kurtosis defined?
- For positive excess Kurtosis, what is the shape of the pdf?
- Be able to test a sample data set for excess kurtosis. What test statistic is used? What is its sampling distribution?