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

Lecture 13
Testing for Skewness

Chris A. Mack
Adjunct Associate Professor

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

Moment Testing

- The most common normality test is to test for the third and fourth standardized moments

The k^{th} centered moment: $$\mu_k = \int_{-\infty}^{\infty} (x-\mu)^k f_X(x)dx$$ ($f_X(x)$: PDF)

$$\mu_1 = 0$$

$$variance = \mu_2 = \sigma^2$$

The k^{th} standardized moment: $$\varphi_k = \frac{\mu_k}{\sigma^k}$$

Sample Moments

- The k^{th} centered moment is estimated for a sample by

$$m_k = \frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^k$$

- The standard deviation σ is estimated by the square root of the second centered moment

$$s = \sqrt{m_2} = \left(\frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^2\right)^{1/2}$$

Normal Standardized Moments

- For a normal distribution, the standardized moments are

$$\varphi_k = \frac{k!}{\left(\frac{k}{2}\right)!\ 2^{k/2}} \quad for\ k\ even$$

$$\varphi_k = 0 \quad for\ k\ odd$$

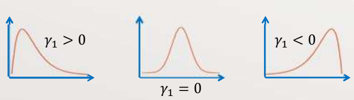
So, $\varphi_3 = 0$ and $\varphi_4 = 3$

Skewness

- For any distribution, the skewness is defined as

$$\gamma_1 = \varphi_3 = \frac{\mu_3}{\sigma^3}$$ (old notation = $\sqrt{\beta_1}$)

- For a unimodal distribution,
  - positive skewness: longer tail to the right
  - negative skewness: longer tail to the left



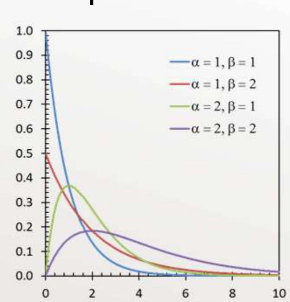
Skew Example

- Gamma Distribution (parameters α, β)

$$f(x) = \frac{\beta^\alpha x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)}$$

for $x \geq 0$ and $\alpha,\beta > 0$

- Mean = α/β
- Variance = α/β^2
- Skewness = $2/\sqrt{\alpha}$
- Excess kurtosis = $6/\alpha$



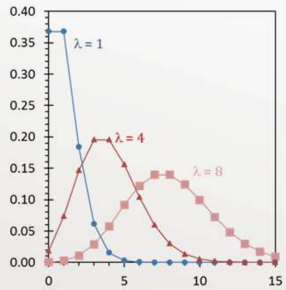
Skew Example

- Poisson Distribution (parameter λ)

$$f(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

for $k \geq 0$ and $\lambda > 0$

- Mean = λ
- Variance = λ
- Skewness = $1/\sqrt{\lambda}$
- Excess kurtosis = $1/\lambda$



Sample Skewness

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

$$g_1 = \frac{m_3}{m_2^{3/2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^2\right)^{3/2}}$$

- For large n, the sampling distribution of g_1 approaches Normal with mean 0 and variance of $6/n$
- For small samples, this estimator is biased

Sample Skewness

- An unbiased estimator of the sample skewness:

$$G_1 = \frac{\sqrt{n(n-1)}}{n-2} g_1$$

Standard Error:

$$SE(G_1) = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \approx \sqrt{\frac{6}{n}}\left(1 - \frac{3}{2n} + O\left(\frac{1}{n^2}\right)\right)$$

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

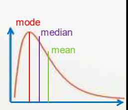
Sample Skewness Test

- Null hypothesis: $\gamma_1 = 0$ (Normal distribution)
- Test statistic: $G_1/SE(G_1)$ is approximately standard Normal for $n > 20$
  - We generally perform a two-tailed test
- If the test statistic $G_1/SE(G_1)$ is beyond the critical z-value for our significance level we reject the null hypothesis that the distribution is Normal

Other Skewness Measures

- Pearson's first & second skewness coefficients

$$Sk_1 = \frac{mean - mode}{\sigma}, \quad Sk_2 = \frac{3(mean - median)}{\sigma}$$



- Bowley's (Galton) robust skewness coefficient:

$$Sk_3 = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

quartiles

Lecture 13: What have we learned?

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