STAT 3600 Notes on Skewness and Kurtosis S. Maghsoodloo

1. Skewness

For a data of size n > 2, the 3rd central moment of the sample is defined as

$m_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3$, and the unbiased measure of skewness is given by

$$\hat{\mu}_3 = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}(x_i - \bar{x})^3 , \quad n > 2.$$

The standardized (or unitless) measure of skewness is given by

$$\hat{\alpha}_3 = \frac{\hat{\mu}_3}{S^3} - 0 = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left[\frac{x_i - \bar{x}}{S}\right]^3 = \hat{\beta}_3 , \; n > 2$$

If the data is symmetric, then for certain $\hat{\alpha}_3 \equiv 0$, but the converse is not always true.

In Statistical literature, $\hat{\alpha}_3$ or $\hat{\beta}_3$, is called the coefficient of skewness, or simply skewness. Furthermore, the amount of normal (or Gaussian) skewness is identically zero and hence the sample standardized 3rd moment is compared against zero.

2. Kurtosis

For a data of size n > 3, the 4th central moment of the sample is $m_4 =$ $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4$, and the unbiased measure of Kurtosis is given by

$$\hat{\mu}_4 = \frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}(x_i - \bar{x})^4 , \quad n > 3.$$

The standardized (or unitless) measure of Kurtosis is given by

$$\text{Kurtosis} = \hat{\beta}_4 = \frac{\hat{\mu}_4}{S^4} - \frac{3(n-1)^2}{(n-2)(n-3)} , \; n > 3.$$

If n > 20, the above sample kurtosis measure is approximately equal to $\hat{\beta}_4 \cong \hat{\alpha}_4 - 3$. In Statistical literature, $\hat{\alpha}_4 = \frac{\hat{\mu}_4}{S^4}$ is called the sample standardized 4th central moment. Kurtosis $\hat{\beta}_4$ ($\cong \hat{\alpha}_4 - 3$) measures peakedness in the middle (or roughly the height of the mode if it is in the middle) and the amount of (probability) heaviness at the tails of a distribution. Furthermore, for a normal (or Gaussian) pdf the value of $\alpha_4 = E\{[(X - \mu)/\sigma]^4\} \equiv E(Z^4) = 3$, and hence the sample standardized 4th moment is compared roughly against three. Because $\alpha_4 = E\{[(X - \mu)/\sigma]^4\} \equiv E(Z^4) = 3$ for any normal universe, the kurtosis of all normal distributions is zero.

Note that the kth central moment of the sample $m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r$ so that $m_1 \equiv 0$ and the 2nd central moment of the sample $m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is simply the sample variance, i.e., $\hat{\sigma}^2 = m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$. Similarly, the rth sample origin moment (i.e., rth moment about zero) is defined as $m'_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - 0)^r = \frac{1}{n}\sum_{i=1}^{n}x_i^r$. Thus, the 1st sample origin moment is $m'_1 = \frac{1}{n}\sum_{i=1}^{n}(x_i - 0)^1 = \bar{x}$, and the 2nd sample origin moment $m'_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - 0)^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 = \text{USS}$. Simple binomial expansion will show that

$$m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r = \sum_{j=0}^{r}(-1)^j({}_rC_j)(m'_{r-j})(m'_1)^j, \text{ where } m'_0 \equiv 1.$$