Descriptive Statistics

Jacek Wallusch

Institute of Cliometrics and Transition Studies

What do we need the descriptive statistics for?

To describe:

1. location (mean values, median)
2. dispersion (variance, std. deviation)
3. skewness of the distribution (skewness)
4. pickedness of the distribution (kurtosis)

Mean values

Arithmetic mean:

$$\bar{x} = T^{-1} \sum_{i=1}^{T} x_i;$$

Geometric mean:

$$\bar{x}^g = \left(\prod_{i=1}^{T} x_i \right)^{\frac{1}{T}} = \sqrt[T]{x_1 \cdot x_2 \cdot \ldots \cdot x_{T-1} \cdot x_T};$$

Weighted average:

$$\bar{x}^w = \frac{\sum_{i=1}^{T} \omega_i x_i}{\sum_{i=1}^{T} \omega_i};$$

where: T - sample size, $\sum$ - summation operator, $\prod$ - product operator, ω - weights

Median

Median separates the lower half of a sample from the higher half. For the odd number of observations, the median is estimated as:

$$m = x_{\frac{T+1}{2}},$$

whilst for the even number of observations:

$$m = \frac{x_{\frac{T}{2}} + x_{\frac{T+1}{2}}}{2}.$$

Variance

Dispersion: Variation or scattering of data around some average or central value.

Population variance:

$$\sigma_P^2 = T^{-1} \sum_{i=1}^{T} (x_i - \mu)^2 ;$$

Sample variance:

$$\sigma^2 = (T-1)^{-1} \sum_{i=1}^{T} (x_i - \bar{x})^2 ,$$

where: μ population mean

Standard Deviation

Standard deviation (square root of variance):

$$\sigma_P = \sqrt{\sigma_P^2},$$

$$\sigma = \sqrt{\sigma^2}.$$

Important difference between variance and std. deviation: Standard deviation is measured in the same units as the original data.

Skewness

The MS Excel formula is as follows:

$$s_k = \frac{T}{(T-1)(T-2)} \frac{\sum_{i=1}^{T} (x_i - \bar{x})^3}{\sigma^3}.$$

Note that if mean = median = mode, the skewness is equal to 0.

Kurtosis

Kurtosis measures the peackedness of a distribution.
Non-zero values of the excess kurtosis:

1. $s_k > 0$: leptokurtic distribution (fatter tails)
2. $s_k < 0$: platykurtic distribution (thinner tails)

The MS Excel formula for excess kurtosis is as follows:

$$s_k = \frac{T(T+1)}{(T-1)(T-2)(T-3)} \frac{\sum_{i=1}^{T}(x_i - \bar{x})^4}{\sigma^4} - \frac{3(T-1)^2}{(T-2)(T-3)}.$$