

Descriptive statistics (distribution)...

Indicator	Definition	Formula	In Excel	In Stata	In R
Standard error (deviation) of the mean	Indicates how close the sample mean is from the 'true' population mean. It increases as the variation increases and it decreases as the sample size goes up. It provides a measure of uncertainty.	$SE_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$	=(STDEV(range of cells))/(SQRT(COUNT(same range of cells))).	tabstat var1, s(semicolon)	sem=sd(x)/sqrt(length(x)); sem
Confidence intervals for the mean	The range where the 'true' value of the mean is likely to fall most of the time	$CI_{\bar{X}} = \bar{X} \pm SE_{\bar{X}} * Z$	Use "Descriptive Statistics" in the "Data Analysis" tab (1)	ci var1	Use package "pastecs"
Skewness	Measures the symmetry of the distribution (whether the mean is at the center of the distribution). The skewness value of a normal distribution is 0. A negative value indicates a skew to the left (left tail is longer than the right tail) and a positive value indicates a skew to the right (right tail is longer than the left one)	$Sk = \frac{\sum (X_i - \bar{X})^3}{(n-1)s^3}$	=SKEW(range of cells)	-tabstat var1, s(skew) - sum var1, detail	Custom estimation
Kurtosis	Measures the peakedness (or flatness) of a distribution. A normal distribution has a value of 3. A kurtosis >3 indicates a sharp peak with heavy tails closer to the mean (leptokurtic). A kurtosis < 3 indicates the opposite a flat top (platykurtic).	$K = \frac{\sum (X_i - \bar{X})^4}{(n-1)s^4}$	=KURT(range of cells)	-tabstat var1, s(k) - sum var1, detail	Custom estimation kurtosis(x)

Notation:

X_i = individual value of X

\bar{X} = mean of X

n = sample size

s^2 = variance

s = standard deviation

$SE_{\bar{X}}$ = standard error of the mean

Z = critical value (Z=1.96 give a 95% certainty)

For more info check the module "Descriptive Statistics with Excel/Stata" in <http://dss.princeton.edu/training/>