

---

MSD (ALL, BYVAR, CORR, COVA, MOMENT, PAIRWISE,  
PRINT, SILENT, TERSE, WEIGHT=*series name*) *list of series* ;

---

**Function:**

MSD produces a table of means, standard deviations, minima, maxima, sums, variances, skewness, and kurtosis for all the variables listed. Only observations in the current sample with no missing values are included. Weighted versions of the statistics may be computed.

**Usage:**

For univariate statistics on a set of variables, use the MSD command with no options. The CORR, MOMENT, and COVA options enable you to get several forms of descriptive statistics on the variables at once, saving on computation time. The option ALL allows you to obtain additional statistics such as the median. The TERSE option restricts the statistics computed in order save space and time.

**Options:**

ALL/NOALL computes the median, first and third quartiles, and the interquartile range, in addition to the normal statistics. The median, etc. are computed ignoring any weight.

BYVAR/NOBYVAR treats missing values for each series separately, so that the maximum possible number of observations for each series is used. @NOBMSD will be stored in this case. Normally, if any series has missing values for any observation, that observation is dropped for all series.

CORR tells MSD to compute and print the correlation matrix of the variables.

COVA tells MSD to compute and print a covariance matrix.

MOMENT tells MSD to compute and print an uncentered moment matrix also. This matrix is divided by the number of observations with positive weights to scale it conveniently.

PAIRWISE/NOPAIRWISE treats missing values for each pair of series separately from other series. It applies to CORR, COVA and MOMENT matrices. @NOBCOVA will be stored.

PRINT/NOPRINT specifies whether the results of the procedure are to be printed, or just stored in data storage.

SILENT/NOSILENT turns off all printed output (equivalent to NOPRINT).

TERSE/NOTERSE specifies that only the means, standard deviations, minima, and maxima are computed and printed.

WEIGHT= the name of series which will be used to weight the calculations. The weighting series should be proportional to the inverses of the variances of the variables. For first order statistics like mean and sum, the series are multiplied directly by the weight. For higher order statistics like standard deviation, variance, covariances, moments, correlations, skewness and kurtosis, the mean-differenced data are multiplied by the square root of the weight. For min, max, median and quartiles, the weights are ignored. If the weight is zero for a particular observation, that observation is not included in the computations nor is it counted in determining degrees of freedom.

# MSD

## Examples:

LIST VARS PAT RND ASSETS DRND DPAT ;  
MSD (CORR) VARS ;

MSD (CORR,COVA,WEIGHT=POP) INCOME PHONES NEWBUS ;

## Output:

MSD stores any of the statistics which are requested as well as printing them.

Name	Type	Length	Variable Description
@NOBMSD	vector	#vars	Number of non-missing observations (for BYVAR).
@MEAN	vector	#vars	Means.
@STDDEV	vector	#vars	Standard Deviations.
@MIN	vectpr	#vars	Minimums.
@MAX	vector	#vars	Maximums.
@SUM	vector	#vars	Sums.
@VAR	vector	#vars	Variances.
@SKEW	vector	#vars	Skewness .
@KURT	vector	#vars	Kurtosis .
@MEDIAN	vector	#vars	Median (for ALL option).
@Q1	vector	#vars	1st quartile.
@Q3	vector	#vars	3rd quartile.
@IQR	vecotr	#vars	Inter-quartile range.
@CORR	matrix	#vars*#vars	Correlation matrix.
@COVA	matrix	#vars*#vars	Covariance matrix.
@MOM	matrix	#vars*#vars	Moment matrix divided by number of observations.
@NOBCOVA	matrix	#vars*#vars	Number of non-missing observations for each pair of variables (for PAIRWISE).
@MSD	matrix	#vars*(4 to 13)	Combined table of num. obs., means, std. dev.s, min, max, [sums, variances, skewness, kurtosis, median, Q1, Q3, IQR].

## Method:

The mean, minimum, maximum, variance, and standard deviation are computed in the usual way. The estimated variance and covariance are computed by small sample formulas (division by N-1 instead of N). The formulas for the skewness and kurtosis are the following (also small sample):

$$\text{Skewness} = \frac{N^2}{(N-1)(N-2)} \frac{M3}{S^3} \quad \text{Kurtosis} = \frac{N^2}{(N-1)(N-2)(N-3)} \frac{(N+1)M4 - 3(N-1)M2^2}{S^4}$$

where M3 and M4 are the centered third and fourth moments and S is the estimated standard deviation. These statistics can be used to test for normality of the variables. The skewness multiplied by  $\sqrt{N/6}$  and the kurtosis multiplied by  $\sqrt{N/24}$  both have a normal(0,1) distribution under the null (when the mean and standard deviation have been estimated; see Davidson and MacKinnon for a derivation).

The median is the value of the series at the  $(N+1)/2$  observation (after sorting from low to high). The first and third quartiles are the values of the series at the  $(N+1)/4$  and  $(3N+3)/4$  observations respectively and the interquartile range is the difference between these two values. If these observation numbers are not integers, the values are a weighted average of the bracketing observations.

**References:**

Davidson, Russell, and James G. MacKinnon, **Estimation and Inference in Econometrics**, Oxford University Press, New York, NY, 1993, Chapter 16.

Godfrey, L.G., **Misspecification Tests in Econometrics**, Econometric Society Monograph, Cambridge University Press, Cambridge, England, 1988, pp. 143-145.