

mean() — Means, variances, and correlations

Syntax Diagnostics	Description Also see	Remarks and examples	Conformability
-----------------------	-------------------------	----------------------	----------------

Syntax

real rowvector `mean(X[, w])`

real matrix `variance(X[, w])`

real matrix `quadvariance(X[, w])`

real matrix `meanvariance(X[, w])`

real matrix `quadmeanvariance(X[, w])`

real matrix `correlation(X[, w])`

real matrix `quadcorrelation(X[, w])`

where

X: *real matrix X* (rows are observations, columns are variables)

w: *real colvector w* and is optional

Description

`mean(X, w)` returns the weighted-or-unweighted column means of data matrix *X*. `mean()` uses quad precision in forming sums and so is very accurate.

`variance(X, w)` returns the weighted-or-unweighted variance matrix of *X*. In the calculation, means are removed and those means are calculated in quad precision, but quad precision is not otherwise used.

`quadvariance(X, w)` returns the weighted-or-unweighted variance matrix of *X*. Calculation is highly accurate; quad precision is used in forming all sums.

`meanvariance(X, w)` returns `mean(X, w) \ variance(X, w)`.

`quadmeanvariance(X, w)` returns `mean(X, w) \ quadvariance(X, w)`.

`correlation(X, w)` returns the weighted-or-unweighted correlation matrix of *X*. `correlation()` obtains the variance matrix from `variance()`.

`quadcorrelation(X, w)` returns the weighted-or-unweighted correlation matrix of *X*. `quadcorrelation()` obtains the variance matrix from `quadvariance()`.

In all cases, *w* specifies the weight. Omit *w*, or specify *w* as 1 to obtain unweighted means.

In all cases, rows of X or w that contain missing values are omitted from the calculation, which amounts to casewise deletion.

Remarks and examples

stata.com

1. There is no `quadmean()` function because `mean()`, in fact, is `quadmean()`. The fact that `mean()` defaults to the quad-precision calculation reflects our judgment that the extra computational cost in computing means in quad precision is typically justified.
2. The fact that `variance()` and `correlation()` do not default to using quad precision for their calculations reflects our judgment that the extra computational cost is typically not justified. The emphasis on this last sentence is on the word typically.

It is easier to justify means in part because the extra computational cost is less: there are only k means but $k(k+1)/2$ variances and covariances.

3. If you need just the mean or just the variance matrix, call `mean()` or `variance()` (or `quadvariance()`). If you need both, there is a CPU-time savings to be had by calling `meanvariance()` instead of the two functions separately (or `quadmeanvariance()` instead of calling `mean()` and `quadvariance()`).

The savings is not great—one `mean()` calculation is saved—but the greater rows(X), the greater the savings.

Upon getting back the combined result, it can be efficiently broken into its components via

```
: var   = meanvariance(X)
: means = var[1,.]
: var   = var[2,1 ..,1]
```

Conformability

`mean(X, w)`:

X : $n \times k$
 w : $n \times 1$ or 1×1 (optional, $w = 1$ assumed)
result: $1 \times k$

`variance(X, w)`, `quadvariance(X, w)`, `correlation(X, w)`, `quadcorrelation(X, w)`:

X : $n \times k$
 w : $n \times 1$ or 1×1 (optional, $w = 1$ assumed)
result: $k \times k$

`meanvariance(X, w)`, `quadmeanvariance(X, w)`:

X : $n \times k$
 w : $n \times 1$ or 1×1 (optional, $w = 1$ assumed)
result: $(k+1) \times k$

Diagnostics

All functions omit from the calculation rows that contain missing values unless all rows contain missing values. Then the returned result contains all missing values.

Also see

[M-4] [statistical](#) — Statistical functions