ameans — Arithmetic, geometric, and harmonic means

Description

`ameans` computes the arithmetic, geometric, and harmonic means, with their corresponding confidence intervals, for each variable in *varlist* or for all the variables in the data if *varlist* is not specified. `gmeans` and `hmeans` are synonyms for `ameans`.

Quick start

Arithmetic, geometric, and harmonic means of variable `v1`

```
ameans v1
```

Same as above, but for variables `v1`, `v2`, and `v3`

```
ameans v1 v2 v3
```

Means for all variables in the dataset

```
ameans
```

Add *n* to each observation before calculating means

```
ameans v1, add(n)
```

Add *n* to each observation only for variables with at least 1 nonpositive value

```
ameans v1 v2 v3, add(n) only
```

Request 99% confidence intervals

```
ameans v1, level(99)
```

Menu

Statistics > Summaries, tables, and tests > Summary and descriptive statistics > Arith./geometric/harmonic means

Syntax

ameans [*varlist*] [*if*] [*in*] [*weight*] [, *options*]

options	Description
Main	
add(#)	add # to each variable in *varlist*
only	add # only to variables with nonpositive values
level(#)	set confidence level; default is level(95)

by and collect are allowed; see [D] **by**.

aweights and fweights are allowed; see [U] **11.1.6 weight**.

Options

Main

add(#) adds the value # to each variable in *varlist* before computing the means and confidence intervals. This option is useful when analyzing variables with nonpositive values.

only modifies the action of the add(#) option so that it adds # only to variables with at least one nonpositive value.

level(#) specifies the confidence level, as a percentage, for confidence intervals. The default is level(95) or as set by set level; see [U] **20.8 Specifying the width of confidence intervals**.

Remarks and examples

stata.com

▷ Example 1

We have a dataset containing 8 observations on a variable named x. The eight values are 5, 4, −4, −5, 0, 0, *missing*, and 7.

```
. ameans x

    Variable |    Type             Obs        Mean       [95% conf. interval]
-------------+---------------------------------------------------------------
           x | Arithmetic            7           1       -3.204405   5.204405
             |  Geometric            3    5.192494         2.57899   10.45448
             |   Harmonic            3    5.060241        3.023008    15.5179
-----------------------------------------------------------------------------

. ameans x, add(5)

    Variable |    Type             Obs        Mean       [95% conf. interval]
-------------+---------------------------------------------------------------
           x | Arithmetic            7           6        1.795595   10.2044*
             |  Geometric            6    5.477226          2.1096   14.22071*
             |   Harmonic            6    3.540984               .          .*
-----------------------------------------------------------------------------
* 5 was added to the variables prior to calculating the results.
Note: Missing values in confidence intervals for harmonic mean indicate that
      confidence interval is undefined for corresponding variables.
```

The number of observations displayed for the arithmetic mean is the number of nonmissing observations. The number of observations displayed for the geometric and harmonic means is the number of nonmissing, positive observations. Specifying the `add(5)` option produces 3 more positive observations. The confidence interval for the harmonic mean is not reported; see *Methods and formulas* below. ◁

Video example

Descriptive statistics in Stata

Stored results

`ameans` stores the following in `r()`:

Scalars

`r(N)`	number of nonmissing observations; used for arithmetic mean
`r(N_pos)`	number of nonmissing positive observations; used for geometric and harmonic means
`r(mean)`	arithmetic mean
`r(lb)`	lower bound of confidence interval for arithmetic mean
`r(ub)`	upper bound of confidence interval for arithmetic mean
`r(Var)`	variance of untransformed data
`r(mean_g)`	geometric mean
`r(lb_g)`	lower bound of confidence interval for geometric mean
`r(ub_g)`	upper bound of confidence interval for geometric mean
`r(Var_g)`	variance of $\ln x_i$
`r(mean_h)`	harmonic mean
`r(lb_h)`	lower bound of confidence interval for harmonic mean
`r(ub_h)`	upper bound of confidence interval for harmonic mean
`r(Var_h)`	variance of $1/x_i$
`r(level)`	confidence level of confidence interval

Methods and formulas

See Armitage, Berry, and Matthews (2002) or Snedecor and Cochran (1989). For a history of the concept of the mean, see Plackett (1958).

When restricted to the same set of values (that is, to positive values), the arithmetic mean ($\overline{x}$) is greater than or equal to the geometric mean, which in turn is greater than or equal to the harmonic mean. Equality holds only if all values within a sample are equal to a positive constant.

The arithmetic mean and its confidence interval are identical to those provided by `ci`; see [R] **ci**.

To compute the geometric mean, `ameans` first creates $u_j = \ln x_j$ for all positive x_j. The arithmetic mean of the u_j and its confidence interval are then computed as in `ci`. Let $\overline{u}$ be the resulting mean, and let $[\,L, U\,]$ be the corresponding confidence interval. The geometric mean is then $\exp(\overline{u})$, and its confidence interval is $[\,\exp(L), \exp(U)\,]$.

The same procedure is followed for the harmonic mean, except that then $u_j = 1/x_j$. The harmonic mean is then $1/\overline{u}$, and its confidence interval is $[\,1/U, 1/L\,]$ if L is greater than zero. If L is not greater than zero, this confidence interval is not defined, and missing values are reported.

When weights are specified, `ameans` applies the weights to the transformed values, $u_j = \ln x_j$ and $u_j = 1/x_j$, respectively, when computing the geometric and harmonic means. For details on how the weights are used to compute the mean and variance of the u_j, see [R] **summarize**. Without weights, the formula for the geometric mean reduces to

$$\exp\left\{\frac{1}{n}\sum_j \ln(x_j)\right\}$$

Without weights, the formula for the harmonic mean is

$$\frac{n}{\displaystyle\sum_j \frac{1}{x_j}}$$

Acknowledgments

This improved version of `ameans` is based on the `gmci` command (Carlin, Vidmar, and Ramalheira 1998) and was written by John Carlin of the Murdoch Children's Research Institute and the University of Melbourne; Suzanna Vidmar of the University of Melbourne; and Carlos Ramalheira of Coimbra University Hospital, Portugal.

References

Armitage, P., G. Berry, and J. N. S. Matthews. 2002. *Statistical Methods in Medical Research*. 4th ed. Oxford: Blackwell.

Carlin, J. B., S. Vidmar, and C. Ramalheira. 1998. sg75: Geometric means and confidence intervals. *Stata Technical Bulletin* 41: 23–25. Reprinted in *Stata Technical Bulletin Reprints*, vol. 7, pp. 197–199. College Station, TX: Stata Press.

Keynes, J. M. 1911. The principal averages and the laws of error which lead to them. *Journal of the Royal Statistical Society* 74: 322–331. https://doi.org/10.2307/2340444.

Plackett, R. L. 1958. Studies in the history of probability and statistics: VII. The principle of the arithmetic mean. *Biometrika* 45: 130–135. https://doi.org/10.2307/2333051.

Snedecor, G. W., and W. G. Cochran. 1989. *Statistical Methods*. 8th ed. Ames, IA: Iowa State University Press.

Stigler, S. M. 1985. Arithmetic means. In Vol. 1 of *Encyclopedia of Statistical Sciences*, ed. S. Kotz and N. L. Johnson, 126–129. New York: Wiley.

Vogel, R. M. 2022. The geometric mean? *Communications in Statistics—Theory and Methods* 51: 82–94. https://doi.org/10.1080/03610926.2020.1743313.

Also see

[R] **ci** — Confidence intervals for means, proportions, and variances

[R] **mean** — Estimate means

[R] **summarize** — Summary statistics

[SVY] **svy estimation** — Estimation commands for survey data