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

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

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main</strong></td>
</tr>
<tr>
<td><strong>add(#)</strong></td>
</tr>
<tr>
<td><strong>only</strong></td>
</tr>
<tr>
<td><strong>level(#)</strong></td>
</tr>
</tbody>
</table>

by is 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

#### 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
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Obs</th>
<th>Mean</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>Arithmetic</td>
<td>7</td>
<td>1</td>
<td>-3.204405 5.204405</td>
</tr>
<tr>
<td></td>
<td>Geometric</td>
<td>3</td>
<td>5.192494</td>
<td>2.57899 10.45448</td>
</tr>
<tr>
<td></td>
<td>Harmonic</td>
<td>3</td>
<td>5.060241</td>
<td>3.023008 15.5179</td>
</tr>
</tbody>
</table>

```
. ameans x, add(5)
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Obs</th>
<th>Mean</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>Arithmetic</td>
<td>7</td>
<td>6</td>
<td>1.795595 10.2044 *</td>
</tr>
<tr>
<td></td>
<td>Geometric</td>
<td>6</td>
<td>5.477226</td>
<td>2.1096 14.22071 *</td>
</tr>
<tr>
<td></td>
<td>Harmonic</td>
<td>6</td>
<td>3.540984</td>
<td>. . *</td>
</tr>
</tbody>
</table>

(*) 5 was added to the variables prior to calculating the results.

Missing values in confidence intervals for harmonic mean indicate that confidence interval is undefined for corresponding variables. Consult Reference Manual for details.
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 \( \ln(x) \)
- `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 \( 1/x \)
- `r(level)` confidence level of confidence interval

**Methods and formulas**


When restricted to the same set of values (that is, to positive values), the arithmetic mean (\( \bar{x} \)) 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 \( \bar{u} \) be the resulting mean, and let \( [L, U] \) be the corresponding confidence interval. The geometric mean is then \( \exp(\bar{u}) \), 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 \( 1/\bar{u} \), 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

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

Acknowledgments

This improved version of `ameans` is based on the `gmc1` 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


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