jjackknife — Jackknife estimation

Description

Jackknife performs jackknife estimation of the specified statistics (or expressions) for a Stata command or a user-written program. Statistics are jackknifed by estimating the command once for each observation or cluster in the dataset, leaving the associated observation or cluster out of the calculations. jackknife is designed for use with nonestimation commands, functions of coefficients, or user-written programs. To jackknife coefficients, we recommend using the vce(jackknife) option when allowed by the estimation command.

Quick start

Jackknife estimate of the standard deviation of v1 returned by summarize in r(sd)
   jackknife sd=r(sd), rclass: summarize v1

Jackknife estimate of the statistic r(mystat) returned by rclass program myprog1 that returns the sample size in r(N)
   jackknife stat=r(mystat), rclass: myprog1 v1

As above, and save the results from each replication in mydata.dta
   jackknife stat=r(mystat), rclass saving(mydata): myprog1 v1

Jackknife estimate of a difference in coefficients estimated by regress
   jackknife diff=(_b[x2]-_b[x1]): regress y x1 x2 x3

Jackknife estimate of the statistic e(mystat) returned by eclass program myprog2 that returns the sample size in e(N)
   jackknife stat=e(mystat), eclass: myprog2 y x1 x2 x3

Jackknife estimates of coefficients stored in e(b) by myprog2
   jackknife _b, eclass: myprog2 y x1 x2 x3

Add variables containing the pseudovalues of the coefficients to the dataset
   jackknife _b, eclass keep: myprog2 y x1 x2 x3

Menu

Statistics > Resampling > Jackknife estimation
Syntax

```
jackknife exp_list [ , options eform_option ] : command
```

<table>
<thead>
<tr>
<th>options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main</strong></td>
<td></td>
</tr>
<tr>
<td><em>eclass</em></td>
<td>number of observations used is stored in e(N)</td>
</tr>
<tr>
<td><em>rclass</em></td>
<td>number of observations used is stored in r(N)</td>
</tr>
<tr>
<td><em>n(exp)</em></td>
<td>specify exp that evaluates to the number of observations used</td>
</tr>
<tr>
<td><strong>Options</strong></td>
<td></td>
</tr>
<tr>
<td><em>cluster(varlist)</em></td>
<td>variables identifying sample clusters</td>
</tr>
<tr>
<td><em>idcluster(newvar)</em></td>
<td>create new cluster ID variable</td>
</tr>
<tr>
<td><em>saving(filename,...)</em></td>
<td>save results to filename; save statistics in double precision; save results to filename every # replications</td>
</tr>
<tr>
<td><em>keep</em></td>
<td>keep pseudovalues</td>
</tr>
<tr>
<td><em>mse</em></td>
<td>use MSE formula for variance estimation</td>
</tr>
<tr>
<td><strong>Reporting</strong></td>
<td></td>
</tr>
<tr>
<td>_level(#)</td>
<td>set confidence level; default is level(95)</td>
</tr>
<tr>
<td><em>notable</em></td>
<td>suppress table of results</td>
</tr>
<tr>
<td><em>noheader</em></td>
<td>suppress table header</td>
</tr>
<tr>
<td><em>nolegend</em></td>
<td>suppress table legend</td>
</tr>
<tr>
<td><em>verbose</em></td>
<td>display the full table legend</td>
</tr>
<tr>
<td><em>nodots</em></td>
<td>suppress replication dots</td>
</tr>
<tr>
<td>_dots(#)</td>
<td>display dots every # replications</td>
</tr>
<tr>
<td><em>noisily</em></td>
<td>display any output from command</td>
</tr>
<tr>
<td><em>trace</em></td>
<td>trace command</td>
</tr>
<tr>
<td><em>title(text)</em></td>
<td>use text as title for jackknife results</td>
</tr>
<tr>
<td><strong>display_options</strong></td>
<td>control columns and column formats, row spacing,</td>
</tr>
<tr>
<td></td>
<td>line width, display of omitted variables and base</td>
</tr>
<tr>
<td></td>
<td>and empty cells, and factor-variable labeling</td>
</tr>
<tr>
<td><strong>eform_option</strong></td>
<td>display coefficient table in exponentiated form</td>
</tr>
<tr>
<td><strong>Advanced</strong></td>
<td></td>
</tr>
<tr>
<td><em>nodrop</em></td>
<td>do not drop observations</td>
</tr>
<tr>
<td><em>reject(exp)</em></td>
<td>identify invalid results</td>
</tr>
<tr>
<td><em>coeflegend</em></td>
<td>display legend instead of statistics</td>
</tr>
</tbody>
</table>

-svy- is allowed; see [SVY] svy jackknife.

_command_ is any command that follows standard Stata syntax. All weight types supported by command are allowed except _aweight_; see [U] 11.1.6 weight.

_coeflegend_ does not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.
exp_list contains (name: elist)
  elist
eexp
elist contains newvar = (exp)
  (exp)
eexp is
  specname
  [eqno]specname
specname is
  _b
  _b[]
  _se
  _se[]
eqno is
  ##
  name

exp is a standard Stata expression; see [U] 13 Functions and expressions.
Distinguish between [], which are to be typed, and [], which indicate optional arguments.

Options

Main

eclass, rclass, and n(exp) specify where command stores the number of observations on which it based the calculated results. We strongly advise you to specify one of these options.

eclass specifies that command store the number of observations in e(N).

rclass specifies that command store the number of observations in r(N).

n(exp) specifies an expression that evaluates to the number of observations used. Specifying n(r(N)) is equivalent to specifying the rclass option. Specifying n(e(N)) is equivalent to specifying the eclass option. If command stores the number of observations in r(N1), specify n(r(N1)).

If you specify no options, jackknife will assume eclass or rclass, depending on which of e(N) and r(N) is not missing (in that order). If both e(N) and r(N) are missing, jackknife assumes that all observations in the dataset contribute to the calculated result. If that assumption is incorrect, the reported standard errors will be incorrect. For instance, say that you specify

    . jackknife coef=_b[x2]: myreg y x1 x2 x3

where myreg uses e(n) instead of e(N) to identify the number of observations used in calculations. Further assume that observation 42 in the dataset has x3 equal to missing. The 42nd observation plays no role in obtaining the estimates, but jackknife has no way of knowing that and will use the wrong N. If, on the other hand, you specify

    . jackknife coef=_b[x2], n(e(n)): myreg y x1 x2 x3

jackknife will notice that observation 42 plays no role. The n(e(n)) option is specified because myreg is an estimation command but it stores the number of observations used in e(n) (instead of the standard e(N)). When jackknife runs the regression omitting the 42nd observation, jackknife will observe that e(n) has the same value as when jackknife previously ran the regression using all the observations. Thus jackknife will know that myreg did not use the observation.
Options

cluster(varlist) specifies the variables identifying sample clusters. If cluster() is specified, one cluster is left out of each call to command, instead of 1 observation.

idcluster(newvar) creates a new variable containing a unique integer identifier for each resampled cluster, starting at 1 and leading up to the number of clusters. This option may be specified only when the cluster() option is specified. idcluster() helps identify the cluster to which a pseudovalue belongs.

saving(filename[, suboptions]) creates a Stata data file (.dta file) consisting of (for each statistic in exp_list) a variable containing the replicates.

double specifies that the results for each replication be saved as doubles, meaning 8-byte reals. By default, they are saved as floats, meaning 4-byte reals. This option may be used without the saving() option to compute the variance estimates by using double precision.

every(#) specifies that results be written to disk every #th replication. every() should be specified only in conjunction with saving() when command takes a long time for each replication. This option will allow recovery of partial results should some other software crash your computer. See [P] postfile.

replace specifies that filename be overwritten if it exists. This option does not appear in the dialog box.

keep specifies that new variables be added to the dataset containing the pseudovalues of the requested statistics. For instance, if you typed

. jackknife coef=_b[x2], eclass keep: regress y x1 x2 x3

new variable coef would be added to the dataset containing the pseudovalues for _b[x2]. Let b be the value of _b[x2] when all observations are used to fit the model, and let b(j) be the value when the jth observation is omitted. The pseudovalues are defined as

\[ \text{pseudovalue}_j = N \{b - b(j)\} + b(j) \]

where N is the number of observations used to produce b.

When the cluster() option is specified, each cluster is given at most one nonmissing pseudovalue. The keep option implies the nodrop option.

mse specifies that jackknife compute the variance by using deviations of the replicates from the observed value of the statistics based on the entire dataset. By default, jackknife computes the variance by using deviations of the pseudovalues from their mean.

Reporting

level(#) ; see [R] Estimation options.

notable suppresses the display of the table of results.

noheader suppresses the display of the table header. This option implies nolegend.

nolegend suppresses the display of the table legend. The table legend identifies the rows of the table with the expressions they represent.

verbose specifies that the full table legend be displayed. By default, coefficients and standard errors are not displayed.

nodots and dots(#) specify whether to display replication dots. By default, one dot character is displayed for each successful replication. A red ‘x’ is displayed if command returns an error or if
any value in exp_list is missing. You can also control whether dots are displayed using set dots; see [R] set.

nodots suppresses display of the replication dots.

dots(#) displays dots every # replications. dots(0) is a synonym for nodots.

noisily specifies that any output from command be displayed. This option implies the nodots option.

trace causes a trace of the execution of command to be displayed. This option implies the noisily option.

title(text) specifies a title to be displayed above the table of jackknife results; the default title is Jackknife results or what is produced in e(title) by an estimation command.

display_options: noci, nopvalues, noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofylabel, fvwrap(#), fvwrappon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch; see [R] Estimation options.

eform_option causes the coefficient table to be displayed in exponentiated form; see [R] eform_option. command determines which eform_option is allowed (eform(string) and eform are always allowed).

command determines which of the following are allowed (eform(string) and eform are always allowed):

<table>
<thead>
<tr>
<th>eform_option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>eform(string)</td>
<td>use string for the column title</td>
</tr>
<tr>
<td>eform</td>
<td>exponentiated coefficient, string is exp(b)</td>
</tr>
<tr>
<td>hr</td>
<td>hazard ratio, string is Haz. Ratio</td>
</tr>
<tr>
<td>shr</td>
<td>subhazard ratio, string is SHR</td>
</tr>
<tr>
<td>irr</td>
<td>incidence-rate ratio, string is IRR</td>
</tr>
<tr>
<td>or</td>
<td>odds ratio, string is Odds Ratio</td>
</tr>
<tr>
<td>rrr</td>
<td>relative-risk ratio, string is RRR</td>
</tr>
</tbody>
</table>

Advanced

nodrop prevents observations outside e(sample) and the if and in qualifiers from being dropped before the data are resampled.

reject(exp) identifies an expression that indicates when results should be rejected. When exp is true, the resulting values are reset to missing values.

The following option is available with jackknife but is not shown in the dialog box:

coefflegend; see [R] Estimation options.

Remarks and examples

Remarks are presented under the following headings:
Introduction

Although the jackknife—developed in the late 1940s and early 1950s—is of largely historical interest today, it is still useful in searching for overly influential observations. This feature is often forgotten. In any case, the jackknife is

- an alternative, first-order unbiased estimator for a statistic;
- a data-dependent way to calculate the standard error of the statistic and to obtain significance levels and confidence intervals; and
- a way of producing measures called pseudovalues for each observation, reflecting the observation’s influence on the overall statistic.

The idea behind the simplest form of the jackknife—the one implemented here—is to repeatedly calculate the statistic in question, each time omitting just one of the dataset’s observations. Assume that our statistic of interest is the sample mean. Let $y_j$ be the $j$th observation of our data on some measurement $y$, where $j = 1, \ldots, N$ and $N$ is the sample size. If $\bar{y}$ is the sample mean of $y$ using the entire dataset and $\bar{y}(j)$ is the mean when the $j$th observation is omitted, then

$$\bar{y} = \frac{(N-1) \bar{y}(j) + y_j}{N}$$

Solving for $y_j$, we obtain

$$y_j = N \bar{y} - (N-1) \bar{y}(j)$$

These are the pseudovalues that jackknife calculates. To move this discussion beyond the sample mean, let $\hat{\theta}$ be the value of our statistic (not necessarily the sample mean) using the entire dataset, and let $\hat{\theta}(j)$ be the computed value of our statistic with the $j$th observation omitted. The pseudovalue for the $j$th observation is

$$\hat{\theta}_j^* = N \hat{\theta} - (N-1) \hat{\theta}(j)$$

The mean of the pseudovalues is the alternative, first-order unbiased estimator mentioned above, and the standard error of the mean of the pseudovalues is an estimator for the standard error of $\hat{\theta}$ (Tukey 1958).

The jackknife estimate of variance has been largely replaced by the bootstrap estimate (see [R] bootstrap), which is widely viewed as more efficient and robust. The use of jackknife pseudovalues to detect outliers is too often forgotten and is something the bootstrap does not provide. See Mosteller and Tukey (1977, 133–163) and Mooney and Duval (1993, 22–27) for more information.

Using jackknife

Typing

```
. jackknife exp_list: command
```

executes `command` once for each observation in the dataset, leaving the associated observation out of the calculations that make up `exp_list`.

`command` defines the statistical command to be executed. Most Stata commands and user-written programs can be used with `jackknife`, as long as they follow standard Stata syntax and allow the `if` qualifier; see [U] 11 Language syntax. The by prefix may not be part of `command`.

`exp_list` specifies the statistics to be collected from the execution of `command`. If `command` changes the contents in `e(b)`, `exp_list` is optional and defaults to `_b`.
When the `cluster()` option is given, clusters are omitted instead of observations, and \( N \) is the number of clusters instead of the sample size.

### Example 1

As our first example, we will show that the jackknife standard error of the sample mean is equivalent to the standard error of the sample mean computed using the classical formula in the `ci means` command. We use the `double` option to compute the standard errors with the same precision as the `ci means` command.

``` stata
jackknife r(mean), double: summarize mpg
```

```
Jackknife replications (74)

|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |

Jackknife results

<table>
<thead>
<tr>
<th>Command</th>
<th>Number of obs</th>
<th>Replications</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>summarize mpg</code></td>
<td>74</td>
<td>74</td>
</tr>
</tbody>
</table>

| Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
|-------|-----------|---|-------|------------------|
| `_jk_1` | 21.2973 | .6725511 | 31.67 | 0.000 | 19.9569 | 22.63769 |

. `ci means mpg`
```
Variable | Obs | Mean | Std. Err. | [95% Conf. Interval] |
----------|-----|------|-----------|---------------------|
mpg       | 74  | 21.2973 | .6725511 | 19.9569 | 22.63769 |
```

### Jackknifed standard deviation

### Example 2

Mosteller and Tukey (1977, 139–140) request a 95% confidence interval for the standard deviation of the 11 values:

\[ 0.1, \ 0.1, \ 0.1, \ 0.4, \ 0.5, \ 1.0, \ 1.1, \ 1.3, \ 1.9, \ 1.9, \ 4.7 \]

Stata’s `summarize` command calculates the mean and standard deviation and stores them as `r(mean)` and `r(sd)`. To obtain the jackknifed standard deviation of the 11 values and save the pseudovalues as a new variable, `sd`, we would type

``` stata
jackknife r(sd), double: summarize mpg
```
. clear
. input x
   x
   1. 0.1
   2. 0.1
   3. 0.1
   4. 0.4
   5. 0.5
   6. 1.0
   7. 1.1
   8. 1.3
   9. 1.9
  10. 1.9
  11. 4.7
  12. end
. jackknife sd=r(sd), rclass keep: summarize x
   (running summarize on estimation sample)
Jackknife replications (11)
1 2 3 4 5
...........
Jackknife results
   Number of obs = 11
   Replications = 11
command:  summarize x
sd:  r(sd)
n():  r(N)

|       | Jackknife Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
|-------|-----------------|-----------|---|------|----------------------|
| sd    | 1.343469        | 0.624405  | 2.15 | 0.057 | -0.047792 to 2.73473 |

Interpreting the output, the standard deviation reported by summarize mpg is 1.34. The jackknife standard error is 0.62. The 95% confidence interval for the standard deviation is −0.048 to 2.73.

By specifying keep, jackknife creates in our dataset a new variable, sd, for the pseudovalues.

. list, sep(4)

<table>
<thead>
<tr>
<th>x</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>1.139977</td>
</tr>
<tr>
<td>2.0</td>
<td>1.139977</td>
</tr>
<tr>
<td>3.0</td>
<td>1.139977</td>
</tr>
<tr>
<td>4.0</td>
<td>0.8993147</td>
</tr>
<tr>
<td>5.0</td>
<td>0.824267</td>
</tr>
<tr>
<td>6.0</td>
<td>0.632489</td>
</tr>
<tr>
<td>7.0</td>
<td>0.6203189</td>
</tr>
<tr>
<td>8.0</td>
<td>0.6218889</td>
</tr>
<tr>
<td>9.0</td>
<td>0.835419</td>
</tr>
<tr>
<td>10.0</td>
<td>0.835419</td>
</tr>
<tr>
<td>11.0</td>
<td>7.703949</td>
</tr>
</tbody>
</table>

The jackknife estimate is the average of the sd variable, so sd contains the individual values of our statistic. We can see that the last observation is substantially larger than the others. The last observation is certainly an outlier, but whether that reflects the considerable information it contains or indicates that it should be excluded from analysis depends on the context of the problem. Here Mosteller
and Tukey created the dataset by sampling from an exponential distribution, so the observation is informative.

### Example 3

Let’s repeat the example above using the automobile dataset, obtaining the standard error of the standard deviation of mpg.

```stata
. use https://www.stata-press.com/data/r16/auto, clear
(1978 Automobile Data)
. jackknife sd=r(sd), rclass keep: summarize mpg
(running summarize on estimation sample)
Jackknife replications (74)

Jackknife results Number of obs = 74
Replications = 74
command:  summarize mpg
sd:  r(sd)
n():  r(N)

|       | Jackknife Coef. | Std. Err. | t   | P>|t|  | [95% Conf. Interval] |
|-------|-----------------|-----------|-----|-------|----------------------|
| sd    | 5.785503        | .6072509  | 9.53| 0.000 | 4.575254 6.995753   |
```

Let’s look at sd more carefully:

```stata
. summarize sd, detail
```

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>Smallest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>2.870471</td>
</tr>
<tr>
<td>5%</td>
<td>2.870471</td>
</tr>
<tr>
<td>10%</td>
<td>2.906255</td>
</tr>
<tr>
<td>25%</td>
<td>3.328489</td>
</tr>
<tr>
<td>50%</td>
<td>3.948335</td>
</tr>
<tr>
<td>Largest</td>
<td>Mean</td>
</tr>
<tr>
<td>95%</td>
<td>38.60905</td>
</tr>
<tr>
<td>99%</td>
<td>38.60905</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>75%</td>
<td>17.34316</td>
</tr>
<tr>
<td>90%</td>
<td>19.7617</td>
</tr>
<tr>
<td>95%</td>
<td>19.7617</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
</tr>
<tr>
<td>99%</td>
<td>23.37823</td>
</tr>
</tbody>
</table>

```stata
. list make mpg sd if sd > 30
```

<table>
<thead>
<tr>
<th>make</th>
<th>mpg</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>71.</td>
<td>VW Diesel</td>
<td>41</td>
</tr>
</tbody>
</table>
```

Here the VW Diesel is the only diesel car in our dataset.
Collecting multiple statistics

Example 4

jackknife is not limited to collecting just one statistic. For instance, we can use `summarize, detail` and then obtain the jackknife estimate of the standard deviation and skewness. `summarize, detail` stores the standard deviation in `r(sd)` and the skewness in `r(skewness)`, so we might type

```
use https://www.stata-press.com/data/r16/auto, clear
(1978 Automobile Data)
jackknife sd=r(sd) skew=r(skewness), rclass: summarize mpg, detail
```

```
running summarize on estimation sample
```

```
Jackknife replications (74)
```

```
........................
```

```
Jackknife results Number of obs = 74
Replications = 74
command: summarize mpg, detail
sd: r(sd)
skew: r(skewness)
n(): r(N)
```

```
| Coef.  | Std. Err. | t    | P>|t|  | [95% Conf. Interval] |
|--------|-----------|------|------|---------------------|
| sd     | 5.785503  | .6072509 | 9.53 | 0.000  | 4.575254  | 6.995753 |
| skew   | .9487176  | .3367242 | 2.82 | 0.006  | .2776272  | 1.619808 |
```

Collecting coefficients

Example 5

jackknife can also collect coefficients from estimation commands. For instance, using `auto.dta`, we might wish to obtain the jackknife standard errors of the coefficients from a regression in which we model the mileage of a car by its weight and trunk space. To do this, we could refer to the coefficients as `b[weight]`, `b[trunk]`, `se[weight]`, and `se[trunk]` in the `exp_list`, or we could simply use the extended expressions `_b`. In fact, `jackknife` assumes `_b` by default when used with estimation commands.
. use https://www.stata-press.com/data/r16/auto
   (1978 Automobile Data)
. jackknife: regress mpg weight trunk
   (running regress on estimation sample)
Jackknife replications (74)

.................................................. 50
........................
Linear regression

|                | Coef.  | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|----------------|--------|-----------|-------|------|----------------------|
| mpg            |        |           |       |      |                      |
| weight         | -.0056527 | .0010216  | -5.53 | 0.000 | -.0076887 -.0036167 |
| trunk          | -.096229 | .1486236  | -0.65 | 0.519 | -.3924354 .1999773  |
| _cons          | 39.68913 | 1.873324  | 21.19 | 0.000 | 35.9556 43.42266   |

If you are going to use jackknife to estimate standard errors of model coefficients, we recommend using the vce(jackknife) option when it is allowed with the estimation command; see [R] vce_option.

. regress mpg weight trunk, vce(jackknife, nodots)
Linear regression

|                | Coef.  | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|----------------|--------|-----------|-------|------|----------------------|
| mpg            |        |           |       |      |                      |
| weight         | -.0056527 | .0010216  | -5.53 | 0.000 | -.0076887 -.0036167 |
| trunk          | -.096229 | .1486236  | -0.65 | 0.519 | -.3924354 .1999773  |
| _cons          | 39.68913 | 1.873324  | 21.19 | 0.000 | 35.9556 43.42266   |

John Wilder Tukey (1915–2000) was born in Massachusetts. He studied chemistry at Brown and mathematics at Princeton and afterward worked at both Princeton and Bell Labs, as well as being involved in a great many government projects, consultancies, and committees. He made outstanding contributions to several areas of statistics, including time series, multiple comparisons, robust statistics, and exploratory data analysis. Tukey was extraordinarily energetic and inventive, not least in his use of terminology: he is credited with inventing the terms bit and software, in addition to ANOVA, boxplot, data analysis, hat matrix, jackknife, stem-and-leaf plot, trimming, and winsorizing, among many others. Tukey’s direct and indirect impacts mark him as one of the greatest statisticians of all time.
 Stored results

`jackknife` stores the following in `e()`:

Scalars
- `e(N)` sample size
- `e(N_reps)` number of complete replications
- `e(N_missreps)` number of incomplete replications
- `e(N_clust)` number of clusters
- `e(k_eq)` number of equations in `e(b)`
- `e(k_extra)` number of extra equations
- `e(k_exp)` number of extended expressions (`_b` or `_se`)
- `e(df_r)` degrees of freedom

Macros
- `e(cmdname)` command name from `command`
- `e(cmd)` same as `e(cmdname)` or `jackknife`
- `e(command)` command as typed
- `e(cmdline)` command as typed
- `e(wtype)` weight type
- `e(wexp)` weight expression
- `e(title)` title in estimation output
- `e(cluster)` cluster variables
- `e(pseudo)` new variables containing pseudovalues
- `e(nfunction)` `e(N), r(N), n()` option, or empty
- `e(exp#)` expression for the `#`th statistic
- `e(mse)` from `mse` option
- `e(vc)` `jackknife`
- `e(vcetype)` title used to label Std. Err.
- `e(properties)` `b V`

Matrices
- `e(b)` observed statistics
- `e(b_jk)` jackknife estimates
- `e(V)` jackknife variance–covariance matrix
- `e(V_modelbased)` model-based variance

In addition to the above, the following is stored in `r()`:

Matrices
- `r(table)` matrix containing the coefficients with their standard errors, test statistics, \( p \)-values, and confidence intervals

Note that results stored in `r()` are updated when the command is replayed and will be replaced when any `r-class` command is run after the estimation command.

When `exp_list` is `_b`, `jackknife` will also carry forward most of the results already in `e()` from `command`.

Methods and formulas

Let \( \hat{\theta} \) be the observed value of the statistic, that is, the value of the statistic calculated using the original dataset. Let \( \hat{\theta}_{(j)} \) be the value of the statistic computed by leaving out the \( j \)th observation (or cluster); thus \( j = 1, 2, \ldots, N \) identifies an individual observation (or cluster), and \( N \) is the total number of observations (or clusters). The \( j \)th pseudovalue is given by

\[
\hat{\theta}^*_j = \hat{\theta}_{(j)} + N\{\hat{\theta} - \hat{\theta}_{(j)}\}
\]
When the `mse` option is specified, the standard error is estimated as

\[
\hat{se} = \left\{ \frac{N-1}{N} \sum_{j=1}^{N} (\hat{\theta}(j) - \hat{\theta})^2 \right\}^{1/2}
\]

and the jackknife estimate is

\[
\bar{\theta}(.) = \frac{1}{N} \sum_{j=1}^{N} \hat{\theta}(j)
\]

Otherwise, the standard error is estimated as

\[
\hat{se} = \left\{ \frac{1}{N(N-1)} \sum_{j=1}^{N} (\hat{\theta}^*_j - \bar{\theta}^*)^2 \right\}^{1/2}
\]

\[
\bar{\theta}^* = \frac{1}{N} \sum_{j=1}^{N} \hat{\theta}^*_j
\]

where \( \bar{\theta}^* \) is the jackknife estimate. The variance–covariance matrix is similarly computed.

References


Also see

[R] jackknife postestimation — Postestimation tools for jackknife

[R] bootstrap — Bootstrap sampling and estimation

[R] permute — Monte Carlo permutation tests

[R] simulate — Monte Carlo simulations

[SVY] svy jackknife — Jackknife estimation for survey data

[U] 13.5 Accessing coefficients and standard errors

[U] 13.6 Accessing results from Stata commands

[U] 20 Estimation and postestimation commands