The Stata Journal (2022)
22, Number 2, pp. 416–429 DOI: 10.1177/1536867X221106417

Interactively building table reports with basetable

Niels Henrik Bruun
Unit of Clinical Biostatistics
Aalborg University Hospital
Aalborg, Denmark
nbru@rn.dk

Abstract. In statistical work, it is essential to have an overview of the data used. In, for example, biomedical articles, a standardized way of reporting summaries of continuous and categorical variables is "table 1". This standardized way of reporting can be useful in most cases of statistical work. The `basetable` command is a flexible and straightforward way to build and format such table reports. The final reports are easy to style into Stata Markup and Control Language, comma-separated values, HyperText Markup Language, LaTeX or TeX, or Markdown and, for example, save into a file specified by the `using` modifier. Also, it is possible to export the reports created by `basetable` into Excel worksheets. Because of the General Data Protection Regulation, it has become necessary to blur information on individuals when making reports; in `basetable`, there are options to blur both categorical and continuous data.

Keywords: st0678, basetable, table reports, table 1

1 Introduction

In statistical work, it is essential to have an overview of the data used. When writing articles in biomedical research, it is often formalized into the concept "table 1". The "table 1" approach is also a simple and standardized way to get and give a quick overview of a new dataset.

In Stata, commands for summarizing data are, for example, `summarize`, `codebook`, `ds`, `tabulate`, and `table`. But with these commands, output has to be summarized once more to be publication ready. In Stata 17, the command `tables` has been greatly improved and integrated into a series of commands on customizable tables; see [TABLES] *Customizable Tables and Collected Results Reference Manual*.

Also, there are community-contributed commands like `balancetable` (Chiapello 2017), `table1` (Clayton 2013), `tabout` (Watson 2019), and `table1_mc` (Chatfield 2017).

In this article, I present a flexible yet straightforward command, `basetable`, which can be used to describe a variety of study designs (cohort, case–control, cross-sectional) where the user presents descriptive statistics by exposure or outcome status.

With `basetable`, you can report summary measures like the mean and quartiles for continuous data, and counts, percentages, or both for categorical data. Continuous data can be reported as mean with standard deviation (the default), confidence interval, or prediction interval; as geometric mean with confidence interval; or as median with interquartile range, interquartile interval, interdecentile range, interdecentile interval, range, or min plus max. In the output from `basetable`, a test is reported by default. When the mean is reported, the test is a one-way analysis of variance, and when the median is reported, it is the Kruskal–Wallis test. When count data are reported, a χ^2 test is used as default. However, there is an option for Fisher's exact test. There is also an option for adding a missing data report for each variable. The layout matches the typical "table 1" biomedical research. This layout is usable in other cases, for example, when making adverse event summary tables.

An essential feature of `basetable` is the type of styling and export. If labels and value labels are correctly set, the reports created by `basetable` are submission ready. The exported reports are text based and, hence, formatting like setting decimals is not needed after the export. The reports can be styled as Stata Markup and Control Language (SMCL), comma-separated values (CSV), HyperText Markup Language (HTML), LaTeX or TeX, or Markdown. Alternatively, several tables can be gathered into one Excel file.

When working with, for example, registry data, it is sometimes required to blur information on individuals. Guidelines from Statistics Denmark require that downloads "be in a format such that they are submission-ready", and "it is not allowed to download results or aggregated data where it is possible to identify individuals or companies". Similar rules likely apply to other registries. Sometimes, not all authors are allowed access to registry datasets. Hence, standardized summary tables are a relatively safe and effective way to download and distribute and discuss datasets and data analysis. The `basetable` command makes it possible to blur publication-ready reports such that it is harder or impossible to identify individuals from numbers in small groups.

Section 2 describes the `basetable` command. Section 3 demonstrates how to build a table interactively using Stata example datasets. Section 4 shows how to export reports from `basetable`. Section 5 describes how information on individuals can be blurred in the output of `basetable`. Finally, section 6 is the conclusion of this article.

2 The basetable command

This section describes the syntax and options for `basetable`.

2.1 Syntax

The syntax is

`basetable` *column_variable* [*summary_variables*] [*if*] [*in*] [`using` *filename*]
[, *options*]

column_variable must be a (categorical) variable for subgrouping or the text `_none` for no column variable.

summary_variables are as defined in the next section.

2.1.1 *summary_variables*

summary_variables are categorical or continuous variables or variable lists followed by suboptions written inside parentheses or are headers. These are detailed in the following subsections.

Categorical variables

Categorical variables are specified by their name followed by *categorical_options* in parentheses.

varname`(`*categorical_options*`)`

Available *categorical_options* are

- `r` or `R` (row percentages)
- `c` or `C` (column percentages)
- *label_value* (show only the row for this value); after a comma, a second argument can be added: `r`, `R`, `c`, or `C` as above, or `ci` for a Wald confidence interval

The *p*-value is the default from a χ^2 test. To instead choose Fisher's exact test, specify the option `exact(`*#*`)` (detailed in section 2.2).

Continuous variables

Continuous variables are specified by their name followed by *continuous_options* in parentheses.

varname`(`*continuous_options*`)`

continuous_options are either a numeric format (for example, `%6.2f`) or, after a comma, one of the following report specifications:

- `sd` (mean and standard deviation; the default)
- `ci` (mean and confidence interval)
- `gci` (geometric mean and confidence interval)

- `pi` (mean and prediction interval)
- `iqr` (median and interquartile range)
- `iqi` (median and interquartile interval)
- `idr` (median and interdecentile range)
- `idi` (median and interdecentile interval)
- `imr` (median and range)
- `imi` (median, minimum, and maximum)

When the mean is reported, the p-value is based on an analysis of variance test, and when the median is reported, the p-value is based on a Kruskal–Wallis test.

Centile calculations are similar to Stata's `centile` command, not the `summarize` command.

Headers

Headers are written in square brackets to group variables. Headers can be used to make multidimensional tables by conditioning inside the headers.

`[`*header_text* [`#`] [`,` *cell_text local_if*]`]`

where

- *header_text* is a plain-text string to be used as the column header.
- `#` (hashtag) is optional for adding a subcount; the subcount matches the subcondition.
- *cell_text* is the text to be used in each cell.
- *local_if* is a Stata `if` expression.

2.2 Options

`log` specifies to show the underlying Stata output.

`nthousands` specifies to add the thousands separator to `n` values.

`pctformat(`*string*`)` specifies to alter the format used for the percentages for the categorical summary variables. *string* must be a numeric format.

`pvformat(`*string*[`, top`]`)` specifies to alter the format used for the p-value. *string* must be a numeric format. The suboption `top` places the p-value at the top.

`exact(#)` specifies to report Fisher's exact test instead of the χ^2 tests. The recommended value is `exact(1)`, and `exact(0)` means no exact test. If an error occurs, try with 5 and higher numbers.

`continuousreport(`*contreport*`)` specifies the overall default continuous report. *contreport* may be one of `sd`, `iqr`, `iqi`, `idr`, `idi`, `imr`, `imi`, `ci`, `gci`, or `pi`; see section 2.1.1 for definitions of these terms.

`categoricalreport(`*catreport*`)` specifies the overall default categorical report. *catreport* may be `n` for count or `p` for percentages. The default is `categoricalreport(n)` (count), followed by the percentages in parentheses.

`notopcount` specifies to exclude the first row with count and percentages of row totals.

`caption(`*string*`)` specifies the caption for the output. This is the same as the `title()` option; if both are specified, the `title()` option will overwrite the `caption()` option.

`title(`*string*`)` specifies the title for the output. This is the same as the `caption()` option; if both are specified, the `title()` option will overwrite the `caption()` option.

`top(`*text*`)` specifies text to place before table content. The default is dependent on the value of the `style()` option.

`undertop(`*text*`)` specifies text to place between the header and the table content. The default is dependent on the value of the `style()` option.

`bottom(`*text*`)` specifies text to place after the table content. The default is dependent on the value of the `style()` option.

`missing` specifies to show the missing report to the right of the table.

`small(#)` specifies the limit for being small with regard to `hidesmall`. The default is `small(5)`.

`hidesmall` hides data when count values are less than the values specified in `small(#)`, where the default is `small(5)`. Note that the values less than "small" can sometimes be deduced from surrounding values.

`pseudo` specifies to use pseudopercentiles, which are found by sorting the values and averaging some values around each position. The averages are used to get percentiles. The `small()` option is used to set the average size.

`style(`*format*`)` allows the output to be shown in various formats. *format* is one of the following: `smcl`, `csv`, `html`, `latex` or `tex`, or `md` (Markdown). The default is `style(smcl)`.

`replace` specifies to replace the styled output that was previously saved with the `using` modifier.

`toxl(`*name*[`,` *xloptions*]`)` specifies a string containing up to five values, separated by commas. *name* may be

- the path and filename of the Excel file in which to save the output; the Excel file suffix is set or reset to `.xls` for Stata 13 and to `.xlsx` for Stata 14 and above, or
- the sheet name in which to save the output.

xloptions are the following:

`replace` specifies to replace or overwrite the content in the sheet.

`row` specifies to add column numbers to the upper-right corner of the table in the sheet.

`column` specifies to write the widths in brackets. If there are more columns than widths, the last column width is used for the rest.

3 Formatting and building tables

`lbw.dta` contains information on 189 women's behavior during pregnancy.

```
. webuse lbw
```

To start building a table with nonsmoker and smoker subgroups, use the `basetable` command with the variable `smoke` for columns.

```
. basetable smoke

────────────────────────────────────────────────────────────────────────────────
Columns by: Smoked during pregnancy   Nonsmoker     Smoker        Total  P-value
────────────────────────────────────────────────────────────────────────────────
n (%)                               115 (60.8)  74 (39.2)  189 (100.0)
────────────────────────────────────────────────────────────────────────────────
```

Summary variables are added by name with a suffix of options in parentheses. There are two types of summary variables, categorical and continuous.

Categorical summary variables are presented by counts and row or column percentages. They are added to the table by name and with, for example, `c` for column percentages in parentheses.

```
. basetable smoke race(c)

────────────────────────────────────────────────────────────────────────────────
Columns by: Smoked during pregnancy   Nonsmoker     Smoker        Total  P-value
────────────────────────────────────────────────────────────────────────────────
n (%)                               115 (60.8)  74 (39.2)  189 (100.0)
Race, n (%)
  White                              44 (38.3)  52 (70.3)    96 (50.8)
  Black                              16 (13.9)  10 (13.5)    26 (13.8)
  Other                              55 (47.8)  12 (16.2)    67 (35.4)     0.00
────────────────────────────────────────────────────────────────────────────────
```

The p-value refers to a χ^2 test. For a p-value from Fisher's exact test, use the option `exact(1)` (see `help tabulate twoway`).

Some might prefer row percentages (replace `c` with `r`) or Fisher's exact test p-value with three decimals at the top instead of at the bottom:

```
. basetable smoke race(r), exact(1) pvformat(%6.3f, top)

────────────────────────────────────────────────────────────────────────────────

Columns by: Smoked during pregnancy   Nonsmoker     Smoker        Total  P-value
────────────────────────────────────────────────────────────────────────────────

n (%)                                115 (60.8)  74 (39.2)  189 (100.0)
Race, n (%)                                                                0.000
  White                               44 (45.8)  52 (54.2)   96 (100.0)
  Black                               16 (61.5)  10 (38.5)   26 (100.0)
  Other                               55 (82.1)  12 (17.9)   67 (100.0)
────────────────────────────────────────────────────────────────────────────────
```

The default presentation of categorical variables is numbers followed by percentages in parentheses, that is, "n (p)". Alternatively, numbers (n) or percentages (p) alone can be chosen.

```
. basetable smoke race(c), categoricalreport(n)

──────────────────────────────────────────────────────────────────────

Columns by: Smoked during pregnancy  Nonsmoker  Smoker  Total  P-value
──────────────────────────────────────────────────────────────────────

n                                          115      74    189
Race, n
  White                                     44      52     96
  Black                                     16      10     26
  Other                                     55      12     67     0.00
──────────────────────────────────────────────────────────────────────
```

For a categorical variable, like `race`, it might be best to report just one value, for example, "White" and a Wald-type confidence interval.

```
. basetable smoke race(White, ci), nopvalue nototal

─────────────────────────────────────────────────────────────────────────

Columns by: Smoked during pregnancy           Nonsmoker            Smoker
─────────────────────────────────────────────────────────────────────────

n (%)                                        115 (60.8)         74 (39.2)
Race (White), % (95% CI)             38.3 (29.4; 47.1)  70.3 (59.9; 80.7)
─────────────────────────────────────────────────────────────────────────
```

Continuous variables are by default presented as the mean followed by a standard deviation in parentheses. They are added by variable name with a suffix of options in parentheses. The simplest option is a number format:

```
. basetable smoke age(%5.1f)

─────────────────────────────────────────────────────────────────────────────────
Columns by: Smoked during pregnancy   Nonsmoker      Smoker        Total  P-value
─────────────────────────────────────────────────────────────────────────────────
n (%)                                115 (60.8)   74 (39.2)  189 (100.0)
Age of mother, mean (sd)             23.4 (5.5)  22.9 (5.0)   23.2 (5.3)     0.54
─────────────────────────────────────────────────────────────────────────────────
```

When a mean is reported, the *p*-value is from an analysis of variance test; when a median is reported, the *p*-value is from a Kruskal–Wallis test. When a variable is nonnormal, it can be specified to report the median and the interquartile interval (`iqi`, lower and upper quartile):

```
. basetable smoke age(%5.1f, iqi), nototal

─────────────────────────────────────────────────────────────────────────────────
Columns by: Smoked during pregnancy         Nonsmoker             Smoker  P-value
─────────────────────────────────────────────────────────────────────────────────
n (%)                                      115 (60.8)          74 (39.2)
Age of mother, median (iqi)         23.0 (20.0; 26.0)  22.0 (19.0; 26.2)     0.51
─────────────────────────────────────────────────────────────────────────────────
```

If most or all continuous variables should be reported as the median and the interquartile interval (`iqi`, lower and upper quartile), use the `continuousreport()` option to set this:

```
. basetable smoke age(%5.1f) bwt(%5.1f), nototal nopvalue continuousreport(iqi)

─────────────────────────────────────────────────────────────────────────────────
Columns by: Smoked during pregnancy             Nonsmoker                  Smoker
─────────────────────────────────────────────────────────────────────────────────
n (%)                                          115 (60.8)               74 (39.2)
Age of mother, median (iqi)             23.0 (20.0; 26.0)       22.0 (19.0; 26.2)
Birthweight (grams), median (iqi) 3100.0 (2495.0; 3629.0) 2775.5 (2363.5; 3270.8)
─────────────────────────────────────────────────────────────────────────────────
```

Variable labels and value labels are preferred if defined.

The `pvformat()` option sets the format of the *p*-value, and the `pctformat()` option sets the format of percentages. The `nthousands` option is for splitting counts into thousands. Sometimes the top count is not necessary, and so the `notopcount` option can exclude it. The `missing` option adds a summary of missings to the table report:

```
. basetable smoke age(%5.1f) race(c), nototal nopvalue missing

───────────────────────────────────────────────────────────────────────────────
Columns by: Smoked during pregnancy   Nonsmoker      Smoker  Missings / N (Pct)
───────────────────────────────────────────────────────────────────────────────
n (%)                                115 (60.8)   74 (39.2)       0 / 189 (0.0)
Age of mother, mean (sd)             23.4 (5.5)  22.9 (5.0)       0 / 189 (0.0)
Race, n (%)
  White                               44 (38.3)   52 (70.3)
  Black                               16 (13.9)   10 (13.5)
  Other                               55 (47.8)   12 (16.2)       0 / 189 (0.0)
───────────────────────────────────────────────────────────────────────────────
```

Headers in square brackets can group variables. Headers can be used to make multidimensional tables by conditioning inside the headers. Let us assume that mothers of age below 22 need special reporting.

```
. basetable smoke race(c) age(%5.1f) [age < 22 #, if age < 22] race(c),
> nototal nopvalue missing

───────────────────────────────────────────────────────────────────────────────
Columns by: Smoked during pregnancy   Nonsmoker      Smoker  Missings / N (Pct)
───────────────────────────────────────────────────────────────────────────────
n (%)                                115 (60.8)   74 (39.2)       0 / 189 (0.0)
Race, n (%)
  White                               44 (38.3)   52 (70.3)
  Black                               16 (13.9)   10 (13.5)
  Other                               55 (47.8)   12 (16.2)       0 / 189 (0.0)
Age of mother, mean (sd)             23.4 (5.5)  22.9 (5.0)       0 / 189 (0.0)
age < 22
n (%)                                 45 (55.6)   36 (44.4)        0 / 81 (0.0)
Race, n (%)
  White                                8 (17.8)   26 (72.2)
  Black                               10 (22.2)    5 (13.9)
  Other                               27 (60.0)    5 (13.9)        0 / 81 (0.0)
───────────────────────────────────────────────────────────────────────────────
```

To see the age mean and standard deviation by smoking behavior and races limited to White and Black, we type

```
. basetable smoke race(c) age(%5.1f) [white #, ** if race == 1] age(%5.1f)
> [black #, ** if race > 1] age(%5.1f) if race < 3

─────────────────────────────────────────────────────────────────────────────────
Columns by: Smoked during pregnancy   Nonsmoker      Smoker        Total  P-value
─────────────────────────────────────────────────────────────────────────────────
n (%)                                 60 (49.2)   62 (50.8)  122 (100.0)
Race, n (%)
  White                               44 (73.3)   52 (83.9)    96 (78.7)
  Black                               16 (26.7)   10 (16.1)    26 (21.3)     0.16
Age of mother, mean (sd)             24.4 (6.1)  23.0 (5.1)   23.7 (5.6)     0.18
white                                       **          **           **       **
n (%)                                 44 (45.8)   52 (54.2)   96 (100.0)
Age of mother, mean (sd)             26.0 (6.0)  22.8 (4.9)   24.3 (5.7)     0.01
black                                       **          **           **       **
n (%)                                 16 (61.5)   10 (38.5)   26 (100.0)
Age of mother, mean (sd)             19.9 (3.9)  24.1 (6.0)   21.5 (5.1)     0.04
─────────────────────────────────────────────────────────────────────────────────
```

The `basetable` command can be made shorter by using *varlists* and, for example, the commands `order` and `rename` (groups). Consider `charity.dta`:

```
. webuse charity, clear
```

This dataset comprises five questions measuring the public's faith and trust in charity organizations.

Agreement with the five statements is given on a scale: "strongly agree", "somewhat agree", "somewhat disagree", and "strongly disagree".

To get a summary report for all (categorical) variables named `ta1` to `ta5` using *varlists*, we could type

```
basetable _none ta*(c)
```

To limit the output to questions 2 to 4 (this possibly needs a new ordering of variables to work), type

```
. basetable _none ta2-ta4(c)

───────────────────────────────────────────────────────────
Variables                                           Summary
───────────────────────────────────────────────────────────
n (%)                                           949 (100.0)
Degree of trust, n (%)
  Strongly agree                                 185 (20.3)
  Agree                                          263 (28.8)
  Disagree                                       362 (39.7)
  Strongly disagree                              102 (11.2)
Charitable organizations honest/ethical, n (%)
  Strongly agree                                 205 (22.1)
  Agree                                          511 (55.0)
  Disagree                                       158 (17.0)
  Strongly disagree                                55 (5.9)
Role improving communities, n (%)
  Strongly agree                                 372 (39.8)
  Agree                                          438 (46.9)
  Disagree                                         88 (9.4)
  Strongly disagree                                36 (3.9)
───────────────────────────────────────────────────────────
```

4 Styling and exporting basetable reports

The generated reports can be saved in different styles by using the `style()` option. The default is `style(smcl)`, but you can also set the style to `csv`, `latex` or `tex`, `html`, or `md` (Markdown).

The idea in `basetable` output is to be as close as possible to a final table; hence, numbers are formatted into fixed strings.

Default layouts are defined, but the `top()`, `undertop()`, and `bottom()` options can modify the output.

Here is the latex or tex style.

```
. basetable _none ta2-ta4(c), style(tex)
\begin{table}[h]
\centering
\begin{tabular}{lr}
\hline
\hline
Variables                                      &     Summary \\
\hline
n (\%)                                          & 949 (100.0) \\
Degree of trust, n (\%)                         &             \\
\quad Strongly agree                               &  185 (20.3) \\
\quad Agree                                        &  263 (28.8) \\
\quad Disagree                                     &  362 (39.7) \\
\quad Strongly disagree                            &  102 (11.2) \\
Charitable organizations honest/ethical, n (\%) &            \\
\quad Strongly agree                               &  205 (22.1) \\
\quad Agree                                        &  511 (55.0) \\
\quad Disagree                                     &  158 (17.0) \\
\quad Strongly disagree                            &    55 (5.9) \\
Role improving communities, n (\%)              &            \\
\quad Strongly agree                               &  372 (39.8) \\
\quad Agree                                        &  438 (46.9) \\
\quad Disagree                                     &    88 (9.4) \\
\quad Strongly disagree                            &    36 (3.9) \\
\hline
\hline
\end{tabular}
\end{table}
```

We now switch back to using lbw.dta.

```
webuse lbw, clear
```

By specifying a filename in the using modifier, the generated basetable report will be saved in that file. A caption or title (which are the same thing) for the table can be added with the caption() or title() option.

```
basetable smoke race(c) age(%5.1f) [age < 22 #, if age < 22] race(c) ///
     using tbl.html, style(html) replace caption(Saved as html)
```

To alter the default layout, the top(), undertop(), and bottom() options can be used to specify lines (each added line in quotation marks) above table content, between table header and table body, and below table content, respectively. To get, for example, plain HTML without formatting (adding HTML tags before and after makes the saved file readable without formatting in MS Word):

```
basetable smoke race(c) age(%5.1f) [age < 22 #, if age < 22] race(c) ///
     using plaintbl.html, style(html) top("<html>" "<table>" "<thead>") ///
     undertop("</thead>" "<tbody>") bottom("</tbody>" "</table>" "</html>")
```

To get the output body only, use a space within quotes (" ") for the top(), undertop(), and bottom() options.

From Stata 13 and up, the toxl() option is another way of exporting tables.

Just specify an Excel workbook, name a sheet to save in, possibly select upper-left corner to place the table on a sheet by row and column number, and specify whether you want to replace the sheet. This way, several tables can be saved into one Excel book.

```
basetable smoke race(c) age(%5.1f) [age < 22 #, if age < 22] race(c), ///
     toxl(tables.xlsx, Table1, replace)
```

A Mata string matrix with all current basetable content can be used with putdocx after saving in Mata.

```
mata out = tbl.output
```

Adding the basetable content is done by typing

```
putdocx begin
putdocx table tablename = mata(out)
putdocx save basetable_report, replace
```

5 Working with sensitive data

When working with registry data in, for example, Denmark, it is not allowed to present data that make it possible to identify the individuals of that study. This means that groups less than 5 are not to be reported directly.

Values like minimum, maximum, and percentiles are only to be reported if it can be documented that the values are based on group sizes greater than five; for example, there must be at least five individuals having the same minimum value. In reality, these values are impossible to report directly.

The solution to this in basetable is to implement a pseudodataset, that is, the average of the five nearest neighbors as the base of the reporting. The pseudominimum reported is the average of the five lowest original values, the median is the average of the five center values, etc. In many cases, numerical differences between actual values and pseudovalues are minimal.

The pseudo option makes reporting of continuous variables “pseudomized”.

For categorical data, basetable offers the possibility of (in part) hiding group sizes less than 5 with the hidesmall option. If hidesmall is specified, group sizes less than 5 are reported as “< 5 (.)”, and row totals are reported as if the size less than 5 were 5.

The minimal group size can be altered from the default 5 by using the small() option. If pseudo is specified and small() is set to an even number, then 1 is added, so, for example, 6 in the pseudopercentiles is a 7.

Here is raw `basetable` output:

```
. basetable smoke age(%5.1f, iqi) ht(c), nototal

─────────────────────────────────────────────────────────────────────────────────

Columns by: Smoked during pregnancy         Nonsmoker             Smoker  P-value
─────────────────────────────────────────────────────────────────────────────────

n (%)                                      115 (60.8)          74 (39.2)
Age of mother, median (iqi)         23.0 (20.0; 26.0)  22.0 (19.0; 26.2)     0.51
Has history of hypertension, n (%)
  0                                        108 (93.9)          69 (93.2)
  1                                           7 (6.1)            5 (6.8)     0.85
─────────────────────────────────────────────────────────────────────────────────
```

Let us compare it with output where both `hidesmall` and `pseudo` are set and where minimal group size is set to 6:

```
. basetable smoke age(%5.1f, iqi) ht(c), nototal hidesmall small(6) pseudo

─────────────────────────────────────────────────────────────────────────────────

Columns by: Smoked during pregnancy         Nonsmoker             Smoker  P-value
─────────────────────────────────────────────────────────────────────────────────

n (%)                                      115 (60.8)          74 (39.2)
Age of mother, median (iqi)         23.0 (19.6; 26.4)  21.6 (19.0; 26.8)     0.51
Has history of hypertension, n (%)
  0                                        108 (93.9)          69 (93.2)
  1                                           7 (6.1)            < 6 (.)     0.85
─────────────────────────────────────────────────────────────────────────────────
```

There is little numerical difference between the raw quartiles and the pseudoquartiles.

Sometimes, it is possible to find out the actual count in cells from the surrounding values and hence may not be acceptable to download. In that case, it is visually clearer where the problems are, and combining rows can be a solution.

6 Conclusion

The need for efficiently summarizing data has never been more critical. Downloading and distributing standardized summary tables is a safe and effective way to discuss datasets and data analysis.

There are three main features in the presented command, `basetable`.

First, it is easy to interactively build summarizing tables. The interactiveness of `basetable` resembles that of pivot tables, which makes `basetable` a convenient tool to get an overview of a dataset. The command `basetable` used the layout of “table 1” from reports in biomedical subjects. This layout has been developed over the years, which makes it easy to grasp for most readers.

Second, distributing summary results is made simple using `basetable` because the generated tables can be exported into standard formats like Markdown, LaTeX, CSV, and HTML. Alternatively, one or more tables can be saved at different worksheets in the same workbook. From there, tables can be copied into the final article if labels are set correctly. The output is text based, so decimals in the basetable output remain the same. Table content is based on variable and value labels. The right choices of labels make the output publication ready.

Finally, when using registry data, it has always been necessary to blur data. This necessity has become more critical because of the General Data Protection Regulation. Options in basetable make it possible to blur categorical data by recoding numbers for small groups and their related totals. Continuous data are blurred by reporting results from pseudo-observations (moving averages on the ordered data).

7 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 22-2
. net install st0678      (to install program files, if available)
. net get st0678          (to install ancillary files, if available)
```

8 References

Chatfield, M. 2017. table1_mc: Stata module to create "table 1" of baseline characteristics for a manuscript. Statistical Software Components S458351, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s458351.html.

Chiapello, M. 2017. balancetable: Stata module to build a balance table and print it in a LaTeX file or an Excel file. Statistical Software Components S458424, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s458424.html.

Clayton, P. 2013. table1: Stata module to create "table 1" of baseline characteristics for a manuscript. Statistical Software Components S457730, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s457730.html.

Watson, I. 2019. tabout. Version 3 beta. http://tabout.net.au.

About the author

Niels Henrik Bruun has a master's degree in mathematics and statistics and a bachelor's degree in microeconomics, both from the University of Aarhus in Denmark. He is currently working as a statistical consultant at the Aalborg University Hospital. He maintains two websites: Stata Hacks (http://www.bruunisejs.dk/StataHacks/) and Python Hacks (http://www.bruunisejs.dk/PythonHacks/).