gtsummary Overview

Lupe Antonio Lopez

March 27, 2024

Why {gtsummary}

- It provides publication & presentation -ready analytical and summary tables using R.
- It allows you summarize datasets, regression models, customize tables, and more in order to generate reproducible reports.
- The code to create tables is consise and customizable.
- Can summarize datasets/dataframes using mean(), median(), min(), max(), etc.

Installation

```
install.packages('gtsummary')
```

Once installed, load package for use in file:

```
library(gtsummary)
```

Tutorial

Load dataset

```
#other libraries used:
library(tidyverse)


#included in the package:
data("trial")
```

```
#quick view
head(trial)
```

```
## # A tibble: 6 x 8
##   trt      age marker stage grade response death ttdeath
##   <chr>  <dbl>  <dbl> <fct> <fct>    <int> <int>   <dbl>
## 1 Drug A    23  0.16  T1    II           0     0    24
```

```
## 2 Drug B     9  1.11  T2    I            1     0    24
## 3 Drug A    31  0.277 T1    II           0     0    24
## 4 Drug A    NA  2.07  T3    III          1     1    17.6
## 5 Drug A    51  2.77  T4    III          1     1    16.4
## 6 Drug B    39  0.613 T4    I            0     1    15.6
```

The dataset contains information on 200 patients receiving two types of chemotherapy treatments (Drug A or B).

Creating summary tables

Within the `gtsummary` package we can create and customize summary tables using `tbl_summary()`.

Below is a basic template:

```
tbl_summary(by =,
            statistic = list(all_continuous() ~ '{mean} ({sd})',
                             all_categorical() ~ '{n} / {N} ({p}%)'),
            digits = all_continuous() ~ 3)
```

The basics:

- `by`: input the groups to split the data into
- `statistic`: indicate the summary statistics to be generated
  differs by data type (i.e., continuous, categorical, etc.)
- `digits`: number of digits to appear in table
  can customize by variables

some other modifiable input options include:

- `label`: specify variable names to be presented
- `missing`: whether to display row with number of missing observations
- `missing_text`: text label for the missing number row
- `include`: list of variables to include in summary table
  (can also do this beforehand with `select()`)

Important: use `list()` to pass more than 2 sets of instructions to each input

Functions to add information and format tables *added using pipes after tbl_summary()*

- `add_p()`: adds p-values to table comparing values across groups (detects variable type and uses appropriate statistical tests)
- `add_overall()`: adds a column with overall summary statistics
- `add_n()` adds a column with N for non-missing observations for each variable
- `add_stat_label()`: adds label for summary statistics shown in each row

- `modify_header()`: updates column headers
- `bold/italicize_labels()`: bolds/italicizes variable labels within table
- `bold/italicize_levels()`: bolds/italicizes variable levels
- `modify_caption()`: updates table caption/title
- `modify_spanning_header()`: adds/updates a spanning header (i.e., btwn 2 groups)
- `modify_footnote()`: updates footnote

Using regression models

Within the `gtsummary` package we can create and customize tables from model outputs using `tbl_regression()`.

Below is a basic template:

```
#fit model:
mod <- glm(response ~ trt + age + grade + stage, data = trial,
           family = binomial(link = 'logit'))


#table:
 tbl_name <- tbl_regression(mod, exponentiate = TRUE)
```

can add functions:

- `add_global_p()`: adds global p-values
- `add_glance_table()`: adds various model statistics

. . . and many more.

Reporting results in-line:

`gtsummary` allows for results from tables to be reported efficiently using `inline_text()`. The default pattern within this function is: `{estimate} ({conf.level*100}% CI {conf.low}, {conf.high}; {p.value})`.

Example:

Code: The odds ratio for age is 1.02 (95% CI 1.00, 1.04; p=0.092)

Examples

```
trial %>%
  tbl_summary()
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	N = 200
Chemotherapy Treatment	
Drug A	98 (49%)
Drug B	102 (51%)
Age	47 (38, 57)
Unknown	11
Marker Level (ng/mL)	0.64 (0.22, 1.39)
Unknown	10
T Stage	
T1	53 (27%)
T2	54 (27%)
T3	43 (22%)
T4	50 (25%)
Grade	
I	68 (34%)
II	68 (34%)
III	64 (32%)
Tumor Response	61 (32%)
Unknown	7
Patient Died	112 (56%)
Months to Death/Censor	22.4 (16.0, 24.0)

```
trial %>%
  tbl_summary(
    include = c(trt, age, grade, marker),
    by = trt,
    missing = 'no',
    statistic = list(age ~ '{mean}',
                     c(marker) ~ '{mean} ({sd})',
                     all_categorical() ~ '{n} / {N} ({p}%)'),
    digits = list(all_continuous() ~ 3,
                  c(grade) ~ 2),
    label = grade ~ 'Tumor Grade')
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	**Drug A**, N = 98	**Drug B**, N = 102
Age	47.011	47.449
Tumor Grade		
I	35.00 / 98.00 (35.71%)	33.00 / 102.00 (32.35%)
II	32.00 / 98.00 (32.65%)	36.00 / 102.00 (35.29%)
III	31.00 / 98.00 (31.63%)	33.00 / 102.00 (32.35%)
Marker Level (ng/mL)	1.017 (0.885)	0.821 (0.828)

```
#adding in other functions
trial %>%
  tbl_summary(
    include = c(trt, age, grade, marker),
```

```
    by = trt,
    missing = 'no',
    statistic = list(c(age) ~ '{mean}',
                     c(marker) ~ '{mean} ({sd})',
                     all_categorical() ~ '{n} / {N} ({p}%)'),
    digits = list(all_continuous() ~ 3,
                  c(grade) ~ 2),
    label = grade ~ 'Tumor Grade') %>%
  add_p() %>%
  add_n() %>%
  add_overall() %>%
  modify_spanning_header(c('stat_1', 'stat_2') ~ '**Treatment**') %>%
  bold_labels() %>%
  modify_header(label = '**Variable**') %>%
  modify_caption('Table 1')
```

Table 3: Table 1

Variable	**N**	**Overall**, N = 200	**Drug A**, N = 98	**Drug B**, N = 102	**p-value**
Age	189	47.238	47.011	47.449	0.7
Tumor Grade	200				0.9
I		68.00 / 200.00 (34.00%)	35.00 / 98.00 (35.71%)	33.00 / 102.00 (32.35%)	
II		68.00 / 200.00 (34.00%)	32.00 / 98.00 (32.65%)	36.00 / 102.00 (35.29%)	
III		64.00 / 200.00 (32.00%)	31.00 / 98.00 (31.63%)	33.00 / 102.00 (32.35%)	
Marker Level (ng/mL)	190	0.916 (0.859)	1.017 (0.885)	0.821 (0.828)	0.085

```
#change header: label = 'Variable'
#bold spanning header
#change caption/title
```

Additional Resources:

- https://www.danieldsjoberg.com/gtsummary
- https://www.danieldsjoberg.com/gtsummary/articles/tbl_summary.html
- https://www.danieldsjoberg.com/gtsummary-Columbia-Computing-Club/#12