

# An {arsenal} of R Functions for Statistical Summaries

<https://github.com/mayoverse/arsenal>

Ethan Heinzen, Beth Atkinson, Jason Sinnwell

2021-08-26



- ▶ Mayo Clinic: 3-site research hospital
- ▶ Historically a SAS shop
- ▶ Now ~500 R users
- ▶ A SAS license negotiation revealed our dependency on SAS
- ▶ How to port in-house macros and other SAS procedures to R?

- ▶ First started with an internal-only package: `rlocal`
- ▶ Contained some “private” functions, some “public” functions
- ▶ Decided to separate “public” into a CRAN package

- ▶ Goal: mimic and improve SAS functionality “easily”
- ▶ v0.1.2 on CRAN Dec 30, 2016
- ▶ Major releases Feb 2018, Jan 2019, and March 2019
- ▶ Minor releases every couple of months

- ▶ `tableby() ≡ %table()`: create a table 1
- ▶ `paired() ≡ %paired()`: create a table 1 for paired data
- ▶ `modelsum() ≡ %modelsum()`: fit models over a set of independent variables
- ▶ `freqlist() ≡ PROC FREQ`: compile frequency table
- ▶ `comparedf() ≡ PROC COMPARE`: compare two datasets
- ▶ `write2*() ≈ ODS OUTPUT`: output results to a file

# The dataset

```
library(arsenal)
data(mockstudy)
str(mockstudy)
```

```
## 'data.frame':    1499 obs. of  14 variables:
## $ case      : int  110754 99706 105271 105001 112263 86205 99508 90158 88989 90515 ...
## $ age       : int   67 74 50 71 69 56 50 57 51 63 ...
## .. attr(*, "label")= chr "Age in Years"
## $ arm        : chr   "F: FOLFOX" "A: IFL" "A: IFL" "G: IROX" ...
## .. attr(*, "label")= chr "Treatment Arm"
## $ sex        : Factor w/ 2 levels "Male","Female": 1 2 2 2 2 1 1 1 2 1 ...
## $ race       : chr   "Caucasian" "Caucasian" "Caucasian" "Caucasian" ...
## .. attr(*, "label")= chr "Race"
## $ fu.time    : int   922 270 175 128 233 120 369 421 387 363 ...
## $ fu.stat    : int    2 2 2 2 2 2 2 2 2 2 ...
## $ ps         : int    0 1 1 1 0 0 0 0 1 1 ...
## $ hgb        : num   11.5 10.7 11.1 12.6 13 10.2 13.3 12.1 13.8 12.1 ...
## $ bmi        : num   25.1 19.5 NA 29.4 26.4 ...
## .. attr(*, "label")= chr "Body Mass Index (kg/m^2)"
## $ alk.phos   : int   160 290 700 771 350 569 162 152 231 492 ...
## $ ast        : int    35 52 100 68 35 27 16 12 25 18 ...
## $ mdquality.s: int    NA 1 1 1 NA 1 1 1 1 1 ...
## $ age.ord    : Ord.factor w/ 8 levels "10-19"<"20-29"<...: 6 7 4 7 6 5 4 5 5 6 ...
```

```
tb <- tableby(arm ~ sex + age, data = mockstudy)
summary(tb, text = TRUE)
```

```
##
##
## |           | A: IFL (N=428) | F: FOLFOX (N=691) | G: IROX (N=380) | Total (N=1499) | p value|
## |-----| :-----: | :-----: | :-----: | :-----: |-----:|
## |sex      |           |           |           |           |         |
## |- Male   |    277 (64.7%) |    411 (59.5%) |    228 (60.0%) |    916 (61.1%) |         |
## |- Female  |    151 (35.3%) |    280 (40.5%) |    152 (40.0%) |    583 (38.9%) |         |
## |Age in Years |           |           |           |           |         |
## |- Mean (SD) | 59.673 (11.365) | 60.301 (11.632) | 59.763 (11.499) | 59.985 (11.519) |         |
## |- Range    | 27.000 - 88.000 | 19.000 - 88.000 | 26.000 - 85.000 | 19.000 - 88.000 |         |
```

```
tb <- tableby(arm ~ sex + age, data = mockstudy)
summary(tb)
```

	A: IFL (N=428)	F: FOLFOX (N=691)	G: IROX (N=380)	Total (N=1499)	p value
<b>sex</b>					0.190
Male	277 (64.7%)	411 (59.5%)	228 (60.0%)	916 (61.1%)	
Female	151 (35.3%)	280 (40.5%)	152 (40.0%)	583 (38.9%)	
<b>Age in Years</b>					0.614
Mean (SD)	59.673 (11.365)	60.301 (11.632)	59.763 (11.499)	59.985 (11.519)	
Range	27.000 - 88.000	19.000 - 88.000	26.000 - 85.000	19.000 - 88.000	



Common requests:

- ▶ Change labels: `set_labels()`, `labels()<-`, `labelTranslations=`
- ▶ Change summary statistics: `tableby.control()` or `inline`
- ▶ Change decimal points: `tableby.control()` or `inline`

```
tb <- tableby(arm ~ fe(sex, digits.pct = 0) + notest(age, digits = 1, "median", "q1q3"), data = mockstudy,
summary(tb, pfootnote = TRUE)
```

	A: IFL (N=428)	F: FOLFOX (N=691)	G: IROX (N=380)	Total (N=1499)	p value
<b>sex</b>					0.190 <sup>1</sup>
Male	277 (65%)	411 (59%)	228 (60%)	916 (61%)	
Female	151 (35%)	280 (41%)	152 (40%)	583 (39%)	
<b>Age in Years</b>					
Median	61.0	61.0	61.0	61.0	
Q1, Q3	53.0, 68.0	52.0, 69.0	52.0, 68.0	52.0, 68.0	

# 1. Fisher's Exact Test for Count Data

## Overall summary:

```
tb <- tableby( ~ sex + age, data = mockstudy)
summary(tb)
```

Overall (N=1499)	
<b>sex</b>	
Male	916 (61.1%)
Female	583 (38.9%)
<b>Age in Years</b>	
Mean (SD)	59.985 (11.519)
Range	19.000 - 88.000

## Stratified, subsetted, and multiple endpoint summary:

```
tb <- tableby(list(arm, sex) ~ age, strata = ps, data = mockstudy, subset = ps %in% 0:1)
summary(tb)
```

ps		A: IFL (N=340)	F: FOLFOX (N=521)	G: IROX (N=305)	Total (N=1166)	p value
0	<b>Age in Years</b>					0.740
	Mean (SD)	60.101 (10.948)	60.173 (11.096)	59.361 (11.904)	59.935 (11.261)	
	Range	27.000 - 81.000	22.000 - 82.000	26.000 - 85.000	22.000 - 85.000	
1	<b>Age in Years</b>					0.582
	Mean (SD)	60.579 (12.026)	61.342 (11.918)	60.081 (11.037)	60.800 (11.721)	
	Range	28.000 - 88.000	26.000 - 88.000	28.000 - 84.000	26.000 - 88.000	

ps		Male (N=720)	Female (N=446)	Total (N=1166)	p value
0	<b>Age in Years</b>				0.614
	Mean (SD)	59.757 (11.031)	60.221 (11.637)	59.935 (11.261)	
	Range	27.000 - 85.000	22.000 - 82.000	22.000 - 85.000	
1	<b>Age in Years</b>				0.045
	Mean (SD)	61.599 (11.748)	59.500 (11.588)	60.800 (11.721)	
	Range	26.000 - 88.000	28.000 - 88.000	26.000 - 88.000	

Other features:

- ▶ `as.data.frame()`, `as.data.frame(summary())`
- ▶ Subset variables, change the order, delete variable: `[`, `head()`, `tail()`
- ▶ Sort by p-value: `sort()`
- ▶ Filter by p-value: `<`, `<=`, `>`, `>=`, etc.
- ▶ Merge two tables: `merge()`
- ▶ Custom p-values and user statistics

```
ms <- modelsum(alk.phos ~ arm + ps, data = mockstudy)
summary(ms)
```

```
##
##
## |               | estimate | std.error | p.value | adj.r.squared | Nmiss |
## |:-:-----|:-:-----|:-:-----|:-:-----|:-:-----|:-:-----|
## |(Intercept)   | 175.577 | 6.779     | < 0.001 | 0.001         | 266   |
## |**Treatment Arm F: FOLFOX** | -13.593 | 8.715     | 0.119   |               |       |
## |**Treatment Arm G: IROX**  | -2.070  | 9.842     | 0.833   |               |       |
## |(Intercept)   | 143.772 | 4.813     | < 0.001 | 0.046         | 266   |
## |**ps**         | 46.719  | 15.979    | < 0.001 |               |       |
```

## Add common adjustors:

```
ms <- modelsum(alk.phos ~ arm + ps, data = mockstudy, adjust = ~ sex + age)
summary(ms)
```

```
##
##
## | estimate |std.error |p.value |adj.r.squared |Nmiss |
## |-----|:-----|:-----|:-----|:-----|
## |(Intercept) |175.548 |20.587 |< 0.001 |-0.001 |266 |
## |**Treatment Arm F: FOLFOX** |-13.701 |8.730 |0.117 | | |
## |**Treatment Arm G: IROX** |-2.245 |9.860 |0.820 | | |
## |**sex Female** |3.016 |7.521 |0.688 | | |
## |**Age in Years** |-0.017 |0.319 |0.956 | | |
## |(Intercept) |148.391 |19.585 |< 0.001 |0.045 |266 |
## |**ps** |46.721 |5.987 |< 0.001 | | |
## |**sex Female** |1.169 |7.343 |0.874 | | |
## |**Age in Years** |-0.084 |0.311 |0.787 | | |
```

## Don't show common adjustors:

```
ms <- modelsum(alk.phos ~ arm + ps, data = mockstudy, adjust = ~ sex + age)
summary(ms, show.adjust = FALSE, show.intercept = FALSE)
```

```
##
##
## |               |estimate |std.error |p.value |adj.r.squared |Nmiss |
## |:-----| |:-----| |:-----| |:-----| |:-----|
## |**Treatment Arm F: FOLFOX** | -13.701 | 8.730 | 0.117 | -0.001 | 266 |
## |**Treatment Arm G: IROX** | -2.245 | 9.860 | 0.820 | | |
## |**ps** | 46.721 | 5.987 | < 0.001 | 0.045 | 266 |
```



## Other options:

- ▶ Change model “family”: Poisson, Binomial, Survival, Negative Binomial, Ordinal, Conditional Logistic, Relative Risk
- ▶ Change labels: `labels()<-`, `set_labels()`, `labelTranslations=`
- ▶ Change decimal places: `modelsum.control()`
- ▶ Change summary statistics: `modelsum.control()`
- ▶ `as.data.frame()`, `as.data.frame(summary())`
- ▶ Subset variables, change the order, delete variable: `[`, `head()`, `tail()`
- ▶ Merge two tables: `merge()`

# freqlist

```
fl <- freqlist(~ sex + arm + ps, data = mockstudy)
summary(fl)
```

```
##
##
## |sex      |Treatment Arm|ps | Freq| Cumulative Freq| Percent| Cumulative Percent|
## |-----|:-----|:--|:--|:-----|:-----|:-----|
## |Male     |A: IFL       |0  | 122|      122|    8.14|         8.14|
## |         |             |1  | 101|      223|    6.74|        14.88|
## |         |             |2  |   8|      231|    0.53|        15.41|
## |         |             |NA |  46|      277|    3.07|        18.48|
## |         |F: FOLFOX    |0  | 168|      445|   11.21|        29.69|
## |         |             |1  | 148|      593|    9.87|        39.56|
## |         |             |2  |  16|      609|    1.07|        40.63|
## |         |             |NA |  79|      688|    5.27|        45.90|
## |         |G: IROX      |0  | 101|      789|    6.74|        52.64|
## |         |             |1  |  80|      869|    5.34|        57.97|
## |         |             |2  |  10|      879|    0.67|        58.64|
## |         |             |NA |  37|      916|    2.47|        61.11|
## |Female   |A: IFL       |0  |  66|      982|    4.40|        65.51|
## |         |             |1  |  51|     1033|    3.40|        68.91|
## |         |             |2  |  11|     1044|    0.73|        69.65|
## |         |             |NA |  23|     1067|    1.53|        71.18|
## |         |F: FOLFOX    |0  | 110|     1177|    7.34|        78.52|
## |         |             |1  |  95|     1272|    6.34|        84.86|
## |         |             |2  |  13|     1285|    0.87|        85.72|
## |         |             |NA |  62|     1347|    4.14|        89.86|
## |         |G: IROX      |0  |  68|     1415|    4.54|        94.40|
## |         |             |1  |  56|     1471|    3.74|        98.13|
## |         |             |2  |   9|     1480|    0.60|        98.73|
## |         |             |NA |  19|     1499|    1.27|       100.00|
```

Other options:

- ▶ `as.data.frame()`, `as.data.frame(summary())`
- ▶ Change labels: `labels()<-`, `set_labels()`,  
`labelTranslations=`
- ▶ Subset variables, change the order, delete variable: `[`, `head()`,  
`tail()`
- ▶ Sort: `sort()`
- ▶ Merge two tables: `merge()`

Three main functions: `write2word()`, `write2pdf()`,  
`write2html()`

Can use other output formats supported by R Markdown.

```
fp <- tempfile(fileext = "pdf")
write2pdf(list(
  tb,
  summary(lm(age ~ sex, data = mockstudy)),
  "\\newpage",
  "# My modelsum table",
  ms,
  code.chunk(1 + 1)
), file = "test.pdf")
```

```
##
##
## processing file: test.pdf.Rmd
```

```
## |
## ordinary text without R code
##
## |
## label: unnamed-chunk-12
## |
## ordinary text without R code
```



Docs: <https://mayoverse.github.io/arsenal/>

Issues: <https://github.com/mayoverse/arsenal/issues>

This presentation:

[https://github.com/eheinzen/2021\\_arsenal\\_RMedicine](https://github.com/eheinzen/2021_arsenal_RMedicine)

Connect with us on Github: @eheinzen, @bethatkinson,  
@sinnweja, <https://github.com/mayoverse/>