Building Table One for a Clinical Study

Peter D.R. Higgins

We will start by loading libraries and data.

library(arsenal)  
library(knitr)  
library(survival)  
library(tidyverse)

## ── Attaching packages ───────────────────────────────────────────────── tidyverse 1.2.1 ──

## ✔ ggplot2 3.2.0.9000 ✔ purrr 0.3.2.9000  
## ✔ tibble 2.1.3 ✔ dplyr 0.8.3.9000  
## ✔ tidyr 0.8.3.9000 ✔ stringr 1.4.0   
## ✔ readr 1.3.1 ✔ forcats 0.4.0

## ── Conflicts ──────────────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

data(mockstudy)  
library(magrittr)

##   
## Attaching package: 'magrittr'

## The following object is masked from 'package:purrr':  
##   
## set\_names

## The following object is masked from 'package:tidyr':  
##   
## extract

library(keyring)  
library(REDCapR)

dim(mockstudy)

## [1] 1499 14

glimpse(mockstudy)

## Observations: 1,499  
## Variables: 14  
## $ case <int> 110754, 99706, 105271, 105001, 112263, 86205, 99508,…  
## $ age <int> 67, 74, 50, 71, 69, 56, 50, 57, 51, 63, 61, 59, 61, …  
## $ arm <chr> "F: FOLFOX", "A: IFL", "A: IFL", "G: IROX", "F: FOLF…  
## $ sex <fct> Male, Female, Female, Female, Female, Male, Male, Ma…  
## $ race <chr> "Caucasian", "Caucasian", "Caucasian", "Caucasian", …  
## $ fu.time <int> 922, 270, 175, 128, 233, 120, 369, 421, 387, 363, 16…  
## $ fu.stat <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2…  
## $ ps <int> 0, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1…  
## $ hgb <dbl> 11.5, 10.7, 11.1, 12.6, 13.0, 10.2, 13.3, 12.1, 13.8…  
## $ bmi <dbl> 25.09861, 19.49786, NA, 29.42922, 26.35352, 19.03673…  
## $ alk.phos <int> 160, 290, 700, 771, 350, 569, 162, 152, 231, 492, 74…  
## $ ast <int> 35, 52, 100, 68, 35, 27, 16, 12, 25, 18, 45, 16, 50,…  
## $ mdquality.s <int> NA, 1, 1, 1, NA, 1, 1, 1, 1, 1, NA, NA, 1, 0, 1, 1, …  
## $ age.ord <ord> 60-69, 70-79, 40-49, 70-79, 60-69, 50-59, 40-49, 50-…

### Basic Table 1

Let’s make a basic Table 1 grouped by arm with details on sex and age in each group.

#summary by groups  
tab1 <- tableby(arm ~ sex + age, data = mockstudy)  
summary(tab1, text=TRUE)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A: IFL (N=428) | F: FOLFOX (N=691) | G: IROX (N=380) | Total (N=1499) | p value |
| sex |  |  |  |  | 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%) |  |
| Age in Years |  |  |  |  | 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 |  |

### Table 1 without groups

Let’s make a Table 1 - but ungrouped, with stats on BMI, sex, Age in each group.

#summary without groups  
tab.noby <- tableby(~ bmi + sex +age, data = mockstudy)  
summary(tab.noby)

##   
##   
## | | Overall (N=1499) |  
## |:----------------------------|:----------------:|  
## |\*\*Body Mass Index (kg/m^2)\*\* | |  
## |&nbsp;&nbsp;&nbsp;N-Miss | 33 |  
## |&nbsp;&nbsp;&nbsp;Mean (SD) | 27.206 (5.432) |  
## |&nbsp;&nbsp;&nbsp;Range | 14.053 - 60.243 |  
## |\*\*sex\*\* | |  
## |&nbsp;&nbsp;&nbsp;Male | 916 (61.1%) |  
## |&nbsp;&nbsp;&nbsp;Female | 583 (38.9%) |  
## |\*\*Age in Years\*\* | |  
## |&nbsp;&nbsp;&nbsp;Mean (SD) | 59.985 (11.519) |  
## |&nbsp;&nbsp;&nbsp;Range | 19.000 - 88.000 |

### Table 1 grouped, build right hand side

Let’s make a Table 1 by arm, with stats on ps, sex, Age in each group.

myvars <- names(mockstudy)  
rhs <- paste(myvars[8:10], collapse = '+')  
rhs

## [1] "ps+hgb+bmi"

as.formula(paste('arm ~', rhs))

## arm ~ ps + hgb + bmi

summary(tableby(as.formula(paste('arm ~', rhs)), data=mockstudy))

##   
##   
## | | A: IFL (N=428) | F: FOLFOX (N=691) | G: IROX (N=380) | Total (N=1499) | p value|  
## |:----------------------------|:---------------:|:-----------------:|:---------------:|:---------------:|-------:|  
## |\*\*ps\*\* | | | | | 0.903|  
## |&nbsp;&nbsp;&nbsp;N-Miss | 69 | 141 | 56 | 266 | |  
## |&nbsp;&nbsp;&nbsp;Mean (SD) | 0.529 (0.597) | 0.547 (0.595) | 0.537 (0.606) | 0.539 (0.598) | |  
## |&nbsp;&nbsp;&nbsp;Range | 0.000 - 2.000 | 0.000 - 2.000 | 0.000 - 2.000 | 0.000 - 2.000 | |  
## |\*\*hgb\*\* | | | | | 0.639|  
## |&nbsp;&nbsp;&nbsp;N-Miss | 69 | 141 | 56 | 266 | |  
## |&nbsp;&nbsp;&nbsp;Mean (SD) | 12.276 (1.686) | 12.381 (1.763) | 12.373 (1.680) | 12.348 (1.719) | |  
## |&nbsp;&nbsp;&nbsp;Range | 9.060 - 17.300 | 9.000 - 18.200 | 9.000 - 17.000 | 9.000 - 18.200 | |  
## |\*\*Body Mass Index (kg/m^2)\*\* | | | | | 0.892|  
## |&nbsp;&nbsp;&nbsp;N-Miss | 9 | 20 | 4 | 33 | |  
## |&nbsp;&nbsp;&nbsp;Mean (SD) | 27.290 (5.552) | 27.210 (5.173) | 27.106 (5.751) | 27.206 (5.432) | |  
## |&nbsp;&nbsp;&nbsp;Range | 14.053 - 53.008 | 16.649 - 49.130 | 15.430 - 60.243 | 14.053 - 60.243 | |

### Table 1 digit control

Let’s make a Table 1 but now control # of digits

summary(tableby(arm ~ sex + fu.time, data=mockstudy), digits=4, digits.p =2, digits.pct =1)

##   
##   
## | | A: IFL (N=428) | F: FOLFOX (N=691) | G: IROX (N=380) | Total (N=1499) | p value|  
## |:---------------------------|:-------------------:|:-------------------:|:-------------------:|:-------------------:|-------:|  
## |\*\*sex\*\* | | | | | 0.19|  
## |&nbsp;&nbsp;&nbsp;Male | 277 (64.7%) | 411 (59.5%) | 228 (60.0%) | 916 (61.1%) | |  
## |&nbsp;&nbsp;&nbsp;Female | 151 (35.3%) | 280 (40.5%) | 152 (40.0%) | 583 (38.9%) | |  
## |\*\*fu.time\*\* | | | | | < 0.01|  
## |&nbsp;&nbsp;&nbsp;Mean (SD) | 553.5841 (419.6065) | 731.2460 (487.7443) | 607.2421 (435.5092) | 649.0841 (462.5109) | |  
## |&nbsp;&nbsp;&nbsp;Range | 9.0000 - 2170.0000 | 0.0000 - 2472.0000 | 17.0000 - 2118.0000 | 0.0000 - 2472.0000 | |

### Table 1 out to word document

Let’s make a Table 1 and write to MS Word

tab1 <- tableby(arm ~ sex + age, data = mockstudy)  
write2word(tab1, file = 'table1.docx')

##   
##   
## processing file: table1.docx.Rmd

##   
 |   
 | | 0%  
 |   
 |.................................................................| 100%  
## ordinary text without R code

## output file: table1.docx.knit.md

## /Applications/RStudio.app/Contents/MacOS/pandoc/pandoc +RTS -K512m -RTS table1.docx.utf8.md --to docx --from markdown+autolink\_bare\_uris+ascii\_identifiers+tex\_math\_single\_backslash+smart --output table1.docx --highlight-style tango

##   
## Output created: table1.docx

### Table 1 with new data

Let’s make a Table 1 with a new dataset from REDCap We will assign arms 1 and 2, then make a table one

fake\_df <- REDCapR::redcap\_read\_oneshot(  
 redcap\_uri = "https://bbmc.ouhsc.edu/redcap/api/",   
 token = "F304DEC3793FECC3B6DEEFF66302CAD3"  
)$data

## 500 records and 13 columns were read from REDCap in 1.2 seconds. The http status code was 200.

#assign arms  
fake\_df$arm <- c(rep(1,250), rep(2, 250))  
fake\_df$race<- as.factor(fake\_df$race)

rhs <- paste(names(fake\_df[c(8:11)]), collapse = '+')  
tab2 <- summary(tableby(as.formula(paste('arm ~', rhs)), data=fake\_df), pfootnote=TRUE)  
tab2

##   
##   
## | | 1 (N=250) | 2 (N=250) | Total (N=500) | p value|  
## |:---------------------------|:-----------------:|:-----------------:|:-----------------:|--------:|  
## |\*\*race\*\* | | | | 0.625^1^|  
## |&nbsp;&nbsp;&nbsp;1 | 12 (4.8%) | 7 (2.8%) | 19 (3.8%) | |  
## |&nbsp;&nbsp;&nbsp;3 | 24 (9.6%) | 32 (12.8%) | 56 (11.2%) | |  
## |&nbsp;&nbsp;&nbsp;4 | 176 (70.4%) | 176 (70.4%) | 352 (70.4%) | |  
## |&nbsp;&nbsp;&nbsp;5 | 31 (12.4%) | 28 (11.2%) | 59 (11.8%) | |  
## |&nbsp;&nbsp;&nbsp;6 | 7 (2.8%) | 7 (2.8%) | 14 (2.8%) | |  
## |\*\*gender\*\* | | | | 0.372^2^|  
## |&nbsp;&nbsp;&nbsp;Mean (SD) | 0.460 (0.499) | 0.500 (0.501) | 0.480 (0.500) | |  
## |&nbsp;&nbsp;&nbsp;Range | 0.000 - 1.000 | 0.000 - 1.000 | 0.000 - 1.000 | |  
## |\*\*height\*\* | | | | 0.498^2^|  
## |&nbsp;&nbsp;&nbsp;Mean (SD) | 173.092 (10.610) | 172.476 (9.725) | 172.784 (10.171) | |  
## |&nbsp;&nbsp;&nbsp;Range | 142.500 - 204.300 | 142.900 - 205.300 | 142.500 - 205.300 | |  
## |\*\*weight\*\* | | | | 0.850^2^|  
## |&nbsp;&nbsp;&nbsp;Mean (SD) | 110.356 (22.806) | 109.940 (26.311) | 110.148 (24.597) | |  
## |&nbsp;&nbsp;&nbsp;Range | 48.000 - 171.000 | 36.000 - 189.000 | 36.000 - 189.000 | |  
## 1. Pearson's Chi-squared test  
## 2. Linear Model ANOVA

# write2word(tab2, file = 'table3.docx')  
# write2html(tab2, "~/table3.html")