Data Science for Biological, Medical and Health Research: Notes for 432

Thomas E. Love, Ph.D.

Version: 2018-01-15

Contents

Introduction

These Notes provide a series of examples using R to work through issues that are likely to come up in PQHS/CRSP/MPHP 432.

While these Notes share some of the features of a textbook, they are neither comprehensive nor completely original. The main purpose is to give 432 students a set of common materials on which to draw during the course. In class, we will sometimes:

- reiterate points made in this document,
- amplify what is here,
- simplify the presentation of things done here,
- use new examples to show some of the same techniques,
- refer to issues not mentioned in this document,

but what we don't do is follow these notes very precisely. We assume instead that you will read the materials and try to learn from them, just as you will attend classes and try to learn from them. We welcome feedback of all kinds on this document or anything else. Just email us at `431-help at case dot edu`, or submit a pull request.

What you will mostly find are brief explanations of a key idea or summary, accompanied (most of the time) by R code and a demonstration of the results of applying that code.

Everything you see here is available to you as HTML or PDF. You will also have access to the R Markdown files, which contain the code which generates everything in the document, including all of the R results. We will demonstrate the use of R Markdown (this document is generated with the additional help of an R package called bookdown) and R Studio (the "program" which we use to interface with the R language) in class.

To download the data and R code related to these notes, visit the Data and Code section of the 432 course website.

R Packages used in these notes

Here, we'll load in the packages used in these notes.

```r
library(tableone); library(tidyverse)
```

Data used in these notes

Here, we'll load in the data sets used in these notes.

```r
fakestroke <- read.csv("data/fakestroke.csv") %>% tbl_df
```

Chapter 1

Building Table 1

Many scientific articles involve direct comparison of results from various exposures, perhaps treatments. In 431, we studied numerous methods, including various sorts of hypothesis tests, confidence intervals, and descriptive summaries, which can help us to understand and compare outcomes in such a setting. One common approach is to present what's often called Table 1. Table 1 provides a summary of the characteristics of a sample, or of groups of samples, which is most commonly used to help understand the nature of the data being compared.

1.1 Two examples from the New England Journal of Medicine

1.1.1 A simple Table 1

Table 1 is especially common in the context of clinical research. Consider the excerpt below, from a January 2015 article in the *New England Journal of Medicine* (Tolaney et al., 2015).

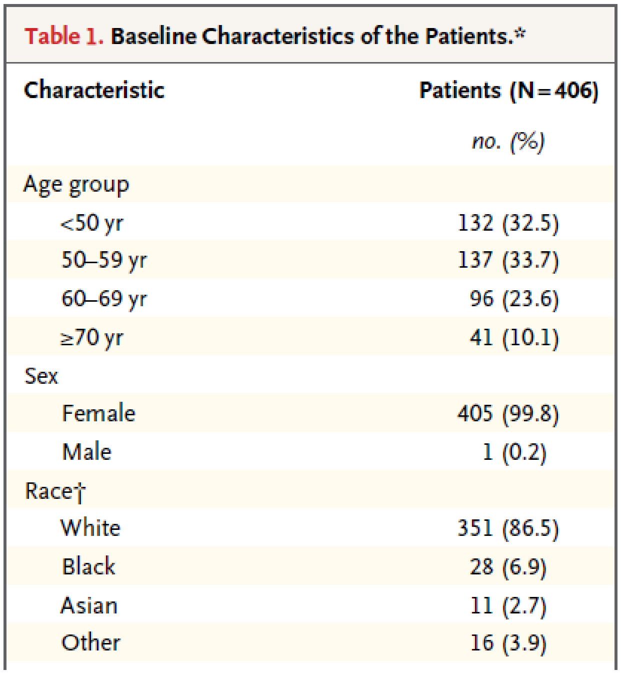

Table 1. Baseline Characteristics of the Patients.*

Characteristic	**Patients (N=406)**
	no. (%)
Age group	
<50 yr	132 (32.5)
50–59 yr	137 (33.7)
60–69 yr	96 (23.6)
≥70 yr	41 (10.1)
Sex	
Female	405 (99.8)
Male	1 (0.2)
Race†	
White	351 (86.5)
Black	28 (6.9)
Asian	11 (2.7)
Other	16 (3.9)

This (partial) table reports baseline characteristics on age group, sex and race, describing 406 patients with

HER2-positive[1] invasive breast cancer that began the protocol therapy. Age, sex and race (along with severity of illness) are the most commonly identified characteristics in a Table 1.

In addition to the measures shown in this excerpt, the full Table also includes detailed information on the primary tumor for each patient, including its size, nodal status and histologic grade. Footnotes tell us that the percentages shown are subject to rounding, and may not total 100, and that the race information was self-reported.

1.1.2 Table 1 showing a group comparison

A more typical Table 1 involves a group comparison, for example in this excerpt from Roy et al. (2008). This Table 1 describes a multi-center randomized clinical trial comparing two different approaches to caring for patients with heart failure and atrial fibrillation[2].

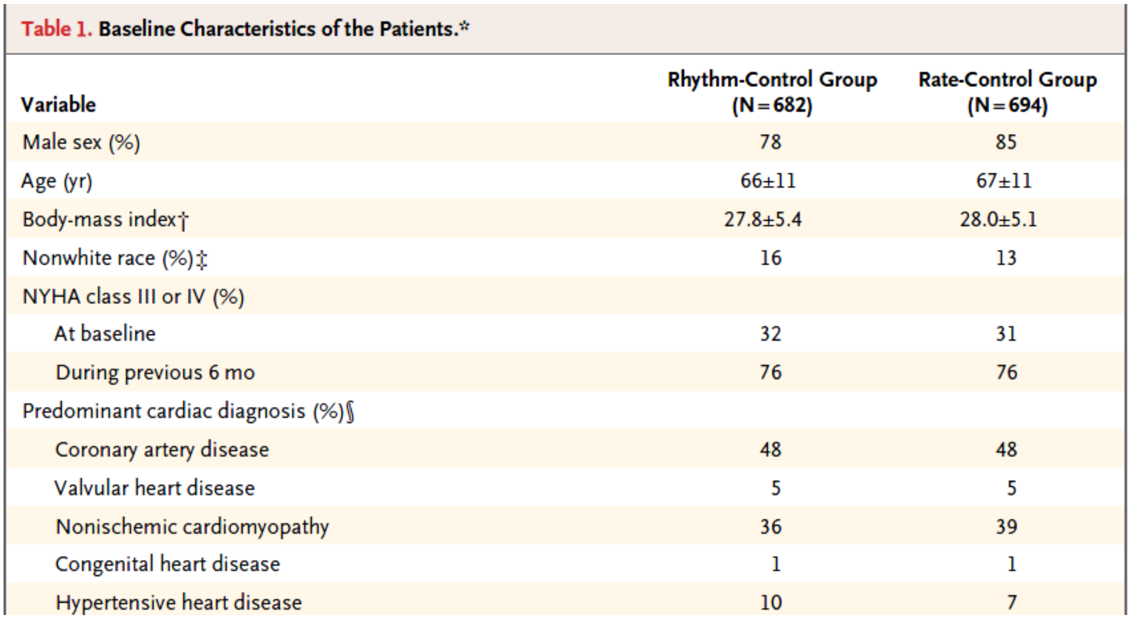

Table 1. Baseline Characteristics of the Patients.*

Variable	Rhythm-Control Group (N = 682)	Rate-Control Group (N = 694)
Male sex (%)	78	85
Age (yr)	66±11	67±11
Body-mass index†	27.8±5.4	28.0±5.1
Nonwhite race (%)‡	16	13
NYHA class III or IV (%)		
At baseline	32	31
During previous 6 mo	76	76
Predominant cardiac diagnosis (%)§		
Coronary artery disease	48	48
Valvular heart disease	5	5
Nonischemic cardiomyopathy	36	39
Congenital heart disease	1	1
Hypertensive heart disease	10	7

The article provides percentages, means and standard deviations across groups, but note that it does not provide p values for the comparison of baseline characteristics. This is a common feature of NEJM reports on randomized clinical trials, where we anticipate that the two groups will be well matched at baseline. Note that the patients in this study were *randomly* assigned to either the rhythm-control group or to the rate-control group, using blocked randomizations stratified by study center.

1.2 Simulating Data from a Clinical Trial

Consider the following simulated data, available on the Data and Code page of our course website in the `fakestroke.csv` file, which I built to let us mirror the Table 1 for a real randomized clinical trial, called MR CLEAN (Berkheimer et al., 2015).

The MR CLEAN trial report describes 500 patients with acute ischemic stroke at 16 medical centers in the Netherlands, where 233 were randomly assigned to the intervention (intraarterial treatment plus usual care) and 267 to control (usual care alone.)

[1] HER2 = human epidermal growth factor receptor type 2. Over-expression of this occurs in 15-20% of invasive breast cancers, and has been associated with poor outcomes.

[2] The complete Table 1 appears on pages 2668-2669 of Roy et al. (2008), but I have only reproduced the first page and the footnote in this excerpt.

1.2.1 The `fakestroke` data

Here's a quick look at the simulated data in `fakestroke`.

```
fakestroke
```

```
# A tibble: 500 x 18
   studyid trt        age sex   nihss location hx.isch  afib    dm mrankin
   <fct>   <fct>    <dbl> <fct> <int> <fct>    <fct>   <int> <int> <fct>
 1 z001    Control   53.0 Male     21 Right    No          0     0 2
 2 z002    Interve~  51.0 Male     23 Left     No          1     0 0
 3 z003    Control   68.0 Fema~    11 Right    No          0     0 0
 4 z004    Control   28.0 Male     22 Left     No          0     0 0
 5 z005    Control   91.0 Male     24 Right    No          0     0 0
 6 z006    Control   34.0 Fema~    18 Left     No          0     0 2
 7 z007    Interve~  75.0 Male     25 Right    No          0     0 0
 8 z008    Control   89.0 Fema~    18 Right    No          0     0 0
 9 z009    Control   75.0 Male     25 Left     No          1     0 2
10 z010    Interve~  26.0 Fema~    27 Right    No          0     0 0
# ... with 490 more rows, and 8 more variables: sbp <int>, iv.altep <fct>,
#   time.iv <int>, aspects <int>, ia.occlus <fct>, extra.ica <int>,
#   time.rand <int>, time.punc <int>
```

The `fakestroke.csv` file contains the following 18 variables for 500 patients.

Variable	Description
`studyid`	Study ID # (z001 through z500)
`trt`	Treatment group (Intervention or Control)
`age`	Age in years
`sex`	Male or Female
`nihss`	NIH Stroke Scale Score (can range from 0-42; higher scores indicate more severe neurological deficits)
`location`	Stroke Location - Left or Right Hemisphere
`hx.isch`	History of Ischemic Stroke (Yes/No)
`afib`	Atrial Fibrillation (1 = Yes, 0 = No)
`dm`	Diabetes Mellitus (1 = Yes, 0 = No)
`mrankin`	Pre-stroke modified Rankin scale score (0, 1, 2 or > 2) indicating functional disability - complete range is 0 (no symptoms) to 6 (death)
`sbp`	Systolic blood pressure, in mm Hg
`iv.altep`	Treatment with IV alteplase (Yes/No)
`time.iv`	Time from stroke onset to start of IV altepase (minutes) if iv.altep=Yes
`aspects`	Alberta Stroke Program Early Computed Tomography score, which measures extent of stroke from 0 - 10; higher scores indicate fewer early ischemic changes
`ia.occlus`	Intracranial arterial occlusion, based on vessel imaging - five categories[3]
`extra.ica`	Extracranial ICA occlusion (1 = Yes, 0 = No)
`time.rand`	Time from stroke onset to study randomization, in minutes
`time.punc`	Time from stroke onset to groin puncture, in minutes (only if Intervention)

[3] The five categories are Intracranial ICA, ICA with involvement of the M1 middle cerebral artery segment, M1 middle cerebral artery segment, M2 middle cerebral artery segment, A1 or A2 anterior cerebral artery segment

1.2.2 fakestroke Table 1: Attempt 1

Our goal, then, is to take the data in fakestroke.csv and use it to generate a Table 1 for the study that compares the 233 patients in the Intervention group to the 267 patients in the Control group, on all of the other variables (except study ID #) available. I'll use the tableone package of functions available in R to help me complete this task. We'll make a first attempt, using the CreateTableOne function in the tableone package. To use the function, we'll need to specify:

- the vars or variables we want to place in the rows of our Table 1 (which will include just about everything in the fakestroke data except the studyid code and the trt variable for which we have other plans)
  - A useful trick here is to use the dput function, specifically something like dput(names(fakestroke)) can be used to generate a list of all of the variables included in the fakestroke tibble, and then this can be copied and pasted into the vars specification, saving some typing.
- the strata which indicates the levels want to use in the columns of our Table 1 (for us, that's trt)

```
fs.vars <- c("age", "sex", "nihss", "location",
           "hx.isch", "afib", "dm", "mrankin", "sbp",
           "iv.altep", "time.iv", "aspects",
           "ia.occlus", "extra.ica", "time.rand",
           "time.punc")

fs.trt <- c("trt")

att1 <- CreateTableOne(data = fakestroke,
                       vars = fs.vars,
                       strata = fs.trt)
print(att1)
```

```
                       Stratified by trt
                        Control        Intervention   p      test
  n                        267            233
  age (mean (sd))        65.38 (16.10)  63.93 (18.09)  0.343
  sex = Male (%)           157 (58.8)     135 (57.9)   0.917
  nihss (mean (sd))      18.08 (4.32)   17.97 (5.04)   0.787
  location = Right (%)     114 (42.7)     117 (50.2)   0.111
  hx.isch = Yes (%)         25 ( 9.4)      29 (12.4)   0.335
  afib (mean (sd))        0.26 (0.44)    0.28 (0.45)   0.534
  dm (mean (sd))          0.13 (0.33)    0.12 (0.33)   0.923
  mrankin (%)                                          0.922
    > 2                     11 ( 4.1)      10 ( 4.3)
    0                      214 (80.1)     190 (81.5)
    1                       29 (10.9)      21 ( 9.0)
    2                       13 ( 4.9)      12 ( 5.2)
  sbp (mean (sd))       145.00 (24.40) 146.03 (26.00)  0.647
  iv.altep = Yes (%)       242 (90.6)     203 (87.1)   0.267
  time.iv (mean (sd))    87.96 (26.01)  98.22 (45.48)  0.003
  aspects (mean (sd))     8.65 (1.47)    8.35 (1.64)   0.033
  ia.occlus (%)                                        0.795
    A1 or A2                 2 ( 0.8)       1 ( 0.4)
    ICA with M1             75 (28.2)      59 (25.3)
    Intracranial ICA         3 ( 1.1)       1 ( 0.4)
    M1                     165 (62.0)     154 (66.1)
    M2                      21 ( 7.9)      18 ( 7.7)
  extra.ica (mean (sd))   0.26 (0.44)    0.32 (0.47)   0.150
  time.rand (mean (sd)) 213.88 (70.29) 202.51 (57.33)  0.051
  time.punc (mean (sd))    NaN (NA)    263.02 (54.23)  NA
```

Some of this is very useful, and other parts need to be fixed.

1. The 1/0 variables (`afib`, `dm`, `extra.ica`) might be better if they were treated as the factors they are, and reported as the Yes/No variables are reported, with counts and percentages rather than with means and standard deviations.
2. In some cases, we may prefer to re-order the levels of the categorical (factor) variables, particularly the `mrankin` variable, but also the `ia.occlus` variable. It would also be more typical to put the Intervention group to the left and the Control group to the right, so we may need to adjust our `trt` variable's levels accordingly.
3. For each of the quantitative variables (`age`, `nihss`, `sbp`, `time.iv`, `aspects`, `extra.ica`, `time.rand` and `time.punc`) we should make a decision whether a summary with mean and standard deviation is appropriate, or whether we should instead summarize with, say, the median and quartiles. A mean and standard deviation really only yields an appropriate summary when the data are least approximately Normally distributed. This will make the p values a bit more reasonable, too. The `test` column in the first attempt will soon have something useful to tell us.
4. We've got some warnings (which I've silenced here), having to do with the fact that `time.punc` is only relevant to patients in the Intervention group. We might consider removing that variable from this table, as a result, and summarizing those data separately.

1.2.3 `fakestroke` Cleaning Up Categorical Variables

Let's specify each of the categorical variables as categorical explicitly. This helps the `CreateTableOne` function treat them appropriately, and display them with counts and percentages. This includes all of the 1/0, Yes/No and multi-categorical variables.

```
fs.factorvars <- c("sex", "location", "hx.isch", "afib", "dm",
                   "mrankin", "iv.altep", "ia.occlus", "extra.ica")
```

Then we simply add a `factorVars = fs.factorvars` call to the `CreateTableOne` function.

We also want to re-order some of those categorical variables, so that the levels are more useful to us. Specifically, we want to:

- place Intervention before Control in the `trt` variable,
- reorder the `mrankin` scale as 0, 1, 2, > 2, and
- rearrange the `ia.occlus` variable to the order[4] presented in Berkheimer et al. (2015).

To accomplish this, we'll use the `fct_relevel` function from the `forcats` package (loaded with the rest of the core `tidyverse` packages) to reorder our levels manually.

```
fakestroke <- fakestroke %>%
    mutate(trt = fct_relevel(trt, "Intervention", "Control"),
           mrankin = fct_relevel(mrankin, "0", "1", "2", "> 2"),
           ia.occlus = fct_relevel(ia.occlus, "Intracranial ICA",
                                   "ICA with M1", "M1", "M2",
                                   "A1 or A2")
           )
```

1.2.4 `fakestroke` Table 1: Attempt 2

```
att2 <- CreateTableOne(data = fakestroke,
                       vars = fs.vars,
```

[4] We might also have considered reordering the `ia.occlus` factor by its frequency, using the `fct_infreq` function

```
                       factorVars = fs.factorvars,
                       strata = fs.trt)
print(att2)
```

```
                       Stratified by trt
                        Intervention   Control        p      test
  n                        233            267                    
  age (mean (sd))        63.93 (18.09)  65.38 (16.10)  0.343
  sex = Male (%)           135 (57.9)     157 (58.8)   0.917
  nihss (mean (sd))      17.97 (5.04)   18.08 (4.32)   0.787
  location = Right (%)     117 (50.2)     114 (42.7)   0.111
  hx.isch = Yes (%)         29 (12.4)      25 ( 9.4)   0.335
  afib = 1 (%)              66 (28.3)      69 (25.8)   0.601
  dm = 1 (%)                29 (12.4)      34 (12.7)   1.000
  mrankin (%)                                          0.922
     0                     190 (81.5)     214 (80.1)
     1                      21 ( 9.0)      29 (10.9)
     2                      12 ( 5.2)      13 ( 4.9)
     > 2                    10 ( 4.3)      11 ( 4.1)
  sbp (mean (sd))       146.03 (26.00) 145.00 (24.40)  0.647
  iv.altep = Yes (%)       203 (87.1)     242 (90.6)   0.267
  time.iv (mean (sd))    98.22 (45.48)  87.96 (26.01)  0.003
  aspects (mean (sd))     8.35 (1.64)    8.65 (1.47)   0.033
  ia.occlus (%)                                        0.795
     Intracranial ICA        1 ( 0.4)       3 ( 1.1)
     ICA with M1            59 (25.3)      75 (28.2)
     M1                    154 (66.1)     165 (62.0)
     M2                     18 ( 7.7)      21 ( 7.9)
     A1 or A2                1 ( 0.4)       2 ( 0.8)
  extra.ica = 1 (%)         75 (32.2)      70 (26.3)   0.179
  time.rand (mean (sd)) 202.51 (57.33) 213.88 (70.29)  0.051
  time.punc (mean (sd)) 263.02 (54.23)    NaN (NA)     NA
```

The categorical data presentation looks much improved.

1.2.5 What summaries should we show?

Now, we'll move on to the issue of making a decision about what type of summary to show for the quantitative variables. Since the `fakestroke` data are just simulated and only match the summary statistics of the original results, not the details, we'll adopt the decisions made by Berkheimer et al. (2015), which was to use medians and interquartile ranges to summarize the distributions of all of the continuous variables **except** systolic blood pressure.

- Specifying certain quantitative variables as *non-normal* causes R to show them with medians and the 25th and 75th percentiles, rather than means and standard deviations, and also causes those variables to be tested using non-parametric tests, like the Wilcoxon signed rank test, rather than the t test. The `test` column indicates this with the word `nonnorm`.
- Specifying *exact* tests for certain categorical variables (we'll try this for the `location` and `mrankin` variables) can be done, and these changes will be noted in the `test` column, as well.

To accomplish this, we need to specify which variables should be treated as non-Normal in the `print` statement - notice that we don't need to redo the `CreateTableOne` for this change.

```
print(att2,
      nonnormal = c("age", "nihss", "time.iv", "aspects", "time.rand",
                    "time.punc"),
      exact = c("location", "mrankin"))
```

```
                        Stratified by trt
                         Intervention            Control
n                           233                     267
age (median [IQR])        65.80 [54.50, 76.00]    65.70 [55.75, 76.20]
sex = Male (%)              135 (57.9)              157 (58.8)
nihss (median [IQR])      17.00 [14.00, 21.00]    18.00 [14.00, 22.00]
location = Right (%)        117 (50.2)              114 (42.7)
hx.isch = Yes (%)            29 (12.4)               25 ( 9.4)
afib = 1 (%)                 66 (28.3)               69 (25.8)
dm = 1 (%)                   29 (12.4)               34 (12.7)
mrankin (%)
   0                        190 (81.5)              214 (80.1)
   1                         21 ( 9.0)               29 (10.9)
   2                         12 ( 5.2)               13 ( 4.9)
   > 2                       10 ( 4.3)               11 ( 4.1)
sbp (mean (sd))          146.03 (26.00)          145.00 (24.40)
iv.altep = Yes (%)          203 (87.1)              242 (90.6)
time.iv (median [IQR])    85.00 [67.00, 110.00]   87.00 [65.00, 116.00]
aspects (median [IQR])     9.00 [7.00, 10.00]      9.00 [8.00, 10.00]
ia.occlus (%)
   Intracranial ICA           1 ( 0.4)                3 ( 1.1)
   ICA with M1               59 (25.3)               75 (28.2)
   M1                       154 (66.1)              165 (62.0)
   M2                        18 ( 7.7)               21 ( 7.9)
   A1 or A2                   1 ( 0.4)                2 ( 0.8)
extra.ica = 1 (%)            75 (32.2)               70 (26.3)
time.rand (median [IQR]) 204.00 [152.00, 249.50] 196.00 [149.00, 266.00]
time.punc (median [IQR]) 260.00 [212.00, 313.00]     NA [NA, NA]
                        Stratified by trt
                         p      test
n
age (median [IQR])        0.579 nonnorm
sex = Male (%)            0.917
nihss (median [IQR])      0.453 nonnorm
location = Right (%)      0.106 exact
hx.isch = Yes (%)         0.335
afib = 1 (%)              0.601
dm = 1 (%)                1.000
mrankin (%)               0.917 exact
   0
   1
   2
   > 2
sbp (mean (sd))           0.647
iv.altep = Yes (%)        0.267
time.iv (median [IQR])    0.596 nonnorm
aspects (median [IQR])    0.075 nonnorm
ia.occlus (%)             0.795
   Intracranial ICA
   ICA with M1
   M1
```

```
     M2
     A1 or A2
  extra.ica = 1 (%)         0.179
  time.rand (median [IQR])  0.251 nonnorm
  time.punc (median [IQR])  NA    nonnorm
```

1.2.6 Obtaining a Detailed Summary

If this was a real data set, we'd want to get a more detailed description of the data to make decisions about things like potentially collapsing categories of a variable, or whether or not a normal distribution was useful for a particular continuous variable, etc. You can do this with the `summary` command applied to a created Table 1, which shows, among other things, the effect of changing from normal to non-normal *p* values for continuous variables, and from approximate to "exact" *p* values for categorical factors.

Note in the summary below that we have some missing values here. Often, we'll present this information within the Table 1, as well.

```r
summary(att2)
```

```
     ### Summary of continuous variables ###

trt: Intervention
            n miss p.miss mean sd median p25 p75 min max  skew  kurt
age       233    0    0.0   64 18     66  54  76  23  96 -0.34 -0.52
nihss     233    0    0.0   18  5     17  14  21  10  28  0.48 -0.74
sbp       233    0    0.0  146 26    146 129 164  78 214 -0.07 -0.22
time.iv   233   30   12.9   98 45     85  67 110  42 218  1.03  0.08
aspects   233    0    0.0    8  2      9   7  10   5  10 -0.56 -0.98
time.rand 233    2    0.9  203 57    204 152 250 100 300  0.01 -1.16
time.punc 233    0    0.0  263 54    260 212 313 180 360  0.11 -1.33
--------------------------------------------------------
trt: Control
            n miss p.miss mean sd median p25 p75 min  max   skew  kurt
age       267    0    0.0   65 16     66  56  76  24   94 -0.296 -0.28
nihss     267    0    0.0   18  4     18  14  22  11   25  0.017 -1.24
sbp       267    1    0.4  145 24    145 128 161  82  231  0.156  0.08
time.iv   267   25    9.4   88 26     87  65 116  44  130  0.001 -1.32
aspects   267    4    1.5    9  1      9   8  10   5   10 -1.071  0.36
time.rand 267    0    0.0  214 70    196 149 266 120  360  0.508 -0.93
time.punc 267  267  100.0  NaN NA     NA  NA  NA Inf -Inf    NaN   NaN

p-values
              pNormal pNonNormal
age       0.342813660 0.57856976
nihss     0.787487252 0.45311695
sbp       0.647157646 0.51346132
time.iv   0.003073372 0.59641104
aspects   0.032662901 0.07464683
time.rand 0.050803672 0.25134327
time.punc          NA         NA

Standardize mean differences
              1 vs 2
age       0.08478764
```

```
nihss     0.02405390
sbp       0.04100833
time.iv   0.27691223
aspects   0.19210662
time.rand 0.17720957
time.punc         NA

=======================================================================================

     ### Summary of categorical variables ### 

trt: Intervention
       var   n miss p.miss            level freq percent cum.percent
       sex 233    0    0.0           Female   98    42.1        42.1
                                       Male  135    57.9       100.0
                                                                    
  location 233    0    0.0             Left  116    49.8        49.8
                                      Right  117    50.2       100.0
                                                                    
   hx.isch 233    0    0.0               No  204    87.6        87.6
                                        Yes   29    12.4       100.0
                                                                    
      afib 233    0    0.0                0  167    71.7        71.7
                                          1   66    28.3       100.0
                                                                    
        dm 233    0    0.0                0  204    87.6        87.6
                                          1   29    12.4       100.0
                                                                    
   mrankin 233    0    0.0                0  190    81.5        81.5
                                          1   21     9.0        90.6
                                          2   12     5.2        95.7
                                        > 2   10     4.3       100.0
                                                                    
  iv.altep 233    0    0.0               No   30    12.9        12.9
                                        Yes  203    87.1       100.0
                                                                    
 ia.occlus 233    0    0.0 Intracranial ICA    1     0.4         0.4
                                ICA with M1   59    25.3        25.8
                                         M1  154    66.1        91.8
                                         M2   18     7.7        99.6
                                   A1 or A2    1     0.4       100.0
                                                                    
 extra.ica 233    0    0.0                0  158    67.8        67.8
                                          1   75    32.2       100.0
                                                                    
--------------------------------------------------------
trt: Control
       var   n miss p.miss            level freq percent cum.percent
       sex 267    0    0.0           Female  110    41.2        41.2
                                       Male  157    58.8       100.0
                                                                    
  location 267    0    0.0             Left  153    57.3        57.3
                                      Right  114    42.7       100.0
   hx.isch 267    0    0.0               No  242    90.6        90.6
                                        Yes   25     9.4       100.0

      afib 267    0    0.0                0  198    74.2        74.2
                                          1   69    25.8       100.0

        dm 267    0    0.0                0  233    87.3        87.3
                                          1   34    12.7       100.0

   mrankin 267    0    0.0                0  214    80.1        80.1
                                          1   29    10.9        91.0
                                          2   13     4.9        95.9
                                        > 2   11     4.1       100.0

  iv.altep 267    0    0.0               No   25     9.4         9.4
                                        Yes  242    90.6       100.0

 ia.occlus 267    1    0.4 Intracranial ICA    3     1.1         1.1
                                ICA with M1   75    28.2        29.3
                                         M1  165    62.0        91.4
                                         M2   21     7.9        99.2
                                   A1 or A2    2     0.8       100.0

 extra.ica 267    1    0.4                0  196    73.7        73.7
                                          1   70    26.3       100.0


p-values
            pApprox    pExact
sex       0.9171387 0.8561188
location  0.1113553 0.1056020
hx.isch   0.3352617 0.3124683
afib      0.6009691 0.5460206
dm        1.0000000 1.0000000
mrankin   0.9224798 0.9173657
iv.altep  0.2674968 0.2518374
ia.occlus 0.7945580 0.8189090
extra.ica 0.1793385 0.1667574

Standardize mean differences
               1 vs 2
sex       0.017479025
location  0.151168444
hx.isch   0.099032275
afib      0.055906317
dm        0.008673478
mrankin   0.062543164
iv.altep  0.111897009
ia.occlus 0.117394890
extra.ica 0.129370206
```

Again, I have simulated the data to mirror the results in the published Table 1 for this study. In no way have I captured the full range of the real data, or any of the relationships in that data, so it’s more important here to see what’s available in the analysis, rather than to interpret it closely in the clinical context.

Bibliography

Berkheimer, O. A., Fransen, P. S. S., Buerner, D., et al. (2015). A randomized trial of intraarterial treatment for acute ischemic stroke. *New England Journal of Medicine*, 372:11–20.

Roy, D., Talajic, M., Nattel, S., et al. (2008). Rhythm control versus rate control for atrial fibrillation and heart failure. *New England Journal of Medicine*, 358:2667–2677.

Tolaney, S. M., Barry, W. T., Chau, T. D., et al. (2015). Adjuvant paclitaxel and trastuzumab for node-negative, her2-positive breast cancer. *New England Journal of Medicine*, 372:134–141.