Development document for creating Proc. Freq. in R

# Exhaustive Requirements:

1. Code must work in R
2. output in table format
3. Displays a table of stroke incidence (only new cases for each wave), with wave (fu\_year) at the top of the table and a sum of the number of unique people with each combination. [0=no stroke, NA (or represented by “.”)= missing, 1=stroke]

code in SAS:

proc Freq;

Tables W1\*W2\*W3 / list missing;

Run;

SAS: Bottom of page: http://www.ats.ucla.edu/stat/sas/faq/nummiss\_sas.htm

| Number of individuals  Frequency | Wave 1 | Wave 2 | Wave 3 |
| --- | --- | --- | --- |
| 0 | 0 | 0 | 1 |
| 75 | 0 | 1 | 1 |
| 100 | 1 | 1 | 1 |
| 30 | 0 | 0 | 0 |
| 55 | NA | NA | NA |

# Define Requirements in English:

get a table with the frequency of new occurrences of stroke from baseline to last wave of data collection.

There are n groups representing:

1. no stroke at any level

2. stroke at all levels

2. missing for all levels

3. no stroke at baseline, stroke at all other levels

4. no stroke at baseline or wave 2, stroke at all other levels etc.

5 …continuing for all waves of data in the data set.

Calculate the frequency of each unique projid that has stroke at baseline

Calculate the frequency of each unique projid that has stroke at wave 1 minus baseline

# Code searches

1. <http://stats.stackexchange.com/questions/13855/is-there-an-r-equivalent-of-sas-proc-freq>

Summarising data in base R is just a headache. This is one of the areas where SAS works quite well. For R, I recommend the plyr package.

In SAS:

/\* tabulate by a and b, with summary stats for x and y in each cell \*/

proc summary data=dat nway;

class a b;

var x y;

output out=smry mean(x)=xmean mean(y)=ymean var(y)=yvar;

run;

with plyr:

smry <- ddply(dat, .(a, b), summarise, xmean=mean(x), ymean=mean(y), yvar=var(y))