Advanced data manipulation in R

(STATS240.1: A short course in R)

Gunnar Stefánsson, Ásta Jenný Sigurðardóttir and Lorna Taylor

October 12, 2012

Tabular summaries in R

table	Simple frequencies
apply	Simple operations on a table
tapply	Arbitrarily complex operations on data in columns

Frequency tables

Example: If M= $\begin{pmatrix} 1 & 5 & 7 \\ 5 & 5 & 1 \end{pmatrix}$ then

```
> table(M)
M
1 5 7
2 3 1
```

returns a table showing the frequencies of each element in M.

Row and column summaries

Example: If M= $\begin{pmatrix} 1 & 5 & 7 \\ 5 & 5 & 1 \end{pmatrix}$ then

```
> apply(M,2,mean)
[1] 3 5 4
```

calculates the mean of each column in M

Using 1 instead of 2 in apply will return the mean of each row in M in place of the columns

Operations on data columns

If $age = (1, 2, 3, 4, 5, 2, 3, 4, 5, 6)$ and $length = (33, 43, 52, 37, 28, 39, 41, 32, 54, 25)$

then

```
> tapply(length,age,mean)
   1    2    3    4    5    6
33.0 41.0 46.5 34.5 41.0 25.0
```

returns an array containing the mean length in each age group.

Other tabular functions

sapply(X,FUN, ...)
lapply(X,FUN, ...)
Applies function fun to each element of X.