

Extracting subsets of data

Often we want to examine a subset of our data. This handout describes two ways to extract a subset of data from a vector and a data frame.

The first way of extracting subsets of numbers is to index elements of a vector. The following code creates a numerical vector, `n`, that contains 100 integers, and then prints out subsets of the numbers:

```
> n <- seq(2,200,2) # create sequence of even integers from 2 to 200
> length(n)
[1] 100

> (n1 <- n[1:10]) # extract the first 10 numbers in n
[1] 2 4 6 8 10 12 14 16 18 20

> (n2 <- n[11:20]) # extract the 2nd set of 10 numbers in n
[1] 22 24 26 28 30 32 34 36 38 40

> n1[c(1,2,6)] # list the first, second, and sixth numbers in n1
[1] 2 4 12

> n1[c(-1,-3)] # list all but the first and third numbers in n1
[1] 4 8 10 12 14 16 18 20
```

The second way of extracting subsets of numbers is to use the `subset` command. The following code creates the data frame `theData`, and then uses `subset` to extract the data separately for each group. Note that the resulting variables are data frames:

```
> score <- c(1,2,7,10,3,5)
> group <- as.factor(c("g1","g1","g2","g2","g3","g3"))
> (theData <- data.frame(group,score))

  group score
1   g1     1
2   g1     2
3   g2     7
4   g2    10
5   g3     3
6   g3     5

> g1.data <- subset(theData,group=="g1")
> class(g1.data)
[1] "data.frame"

> g1.data
  group score
1   g1     1
2   g1     2

> g2.data <- subset(theData,group=="g2")
> class(g2.data)
[1] "data.frame"

> g2.data
```

```

group score
3   g2     7
4   g2    10

```

```

> g3.data <- subset(theData,group=="g3")
> class(g3.data)

```

```
[1] "data.frame"
```

```
> g3.data
```

```

group score
5   g3     3
6   g3     5

```

The following code illustrates how to combine multiple criteria for extracting subsets of data:

```

> score <- round(rnorm(n=12,mean=100,sd=10))
> group <- as.factor(x=c(rep("group1",6),rep("group2",6)))
> sex <- as.factor(x=c(rep("male",3),rep("female",3),rep("male",3),rep("female",3)))
> (myData <- data.frame(score,group,sex))

```

```

score group sex
1   111 group1 male
2    93 group1 male
3   100 group1 male
4   100 group1 female
5   102 group1 female
6   118 group1 female
7   103 group2 male
8    99 group2 male
9   106 group2 male
10   95 group2 female
11   76 group2 female
12   98 group2 female

```

```
> (g1.data <- subset(myData,group=="group1")) ) # extract data from group 1
```

```

score group sex
1   111 group1 male
2    93 group1 male
3   100 group1 male
4   100 group1 female
5   102 group1 female
6   118 group1 female

```

```
> (male.data <- subset(myData,sex=="male")) ) # extract data from males
```

```

score group sex
1   111 group1 male
2    93 group1 male
3   100 group1 male
7   103 group2 male
8    99 group2 male
9   106 group2 male

```

```
> (g1.male.data <- subset(myData,group=="group1"&sex=="male")) ) # extract data from males in group 1
```

```

score group sex
1   111 group1 male
2    93 group1 male
3   100 group1 male

```