```r
library(MASS)
library(help=MASS)        # functions and datasets in MASS - also go online
```

```r
d1=Cars93
str(d1)
```

```
## 'data.frame':    93 obs. of  27 variables:
##  $ Manufacturer      : Factor w/ 32 levels "Acura","Audi",..: 1 1 2 2 3 4 4 4 4 5 ...
##  $ Model             : Factor w/ 93 levels "100","190E","240",..: 49 56 9 1 6 24 54 74 73 35 ...
##  $ Type              : Factor w/ 6 levels "Compact","Large",..: 4 3 1 3 3 3 2 2 3 2 ...
##  $ Min.Price         : num  12.9 29.2 25.9 30.8 23.7 14.2 19.9 22.6 26.3 33 ...
##  $ Price             : num  15.9 33.9 29.1 37.7 30 15.7 20.8 23.7 26.3 34.7 ...
##  $ Max.Price         : num  18.8 38.7 32.3 44.6 36.2 17.3 21.7 24.9 26.3 36.3 ...
##  $ MPG.city          : int  25 18 20 19 22 22 19 16 19 16 ...
##  $ MPG.highway       : int  31 25 26 26 30 31 28 25 27 25 ...
##  $ AirBags           : Factor w/ 3 levels "Driver & Passenger",..: 3 1 2 1 2 2 2 2 2 2 ...
##  $ DriveTrain        : Factor w/ 3 levels "4WD","Front",..: 2 2 2 2 3 2 2 3 2 2 ...
##  $ Cylinders         : Factor w/ 6 levels "3","4","5","6",..: 2 4 4 4 2 2 4 4 4 5 ...
##  $ EngineSize        : num  1.8 3.2 2.8 2.8 3.5 2.2 3.8 5.7 3.8 4.9 ...
##  $ Horsepower        : int  140 200 172 172 208 110 170 180 170 200 ...
##  $ RPM               : int  6300 5500 5500 5500 5700 5200 4800 4000 4800 4100 ...
##  $ Rev.per.mile      : int  2890 2335 2280 2535 2545 2565 1570 1320 1690 1510 ...
##  $ Man.trans.avail   : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 1 1 1 1 1 ...
##  $ Fuel.tank.capacity: num  13.2 18 16.9 21.1 21.1 16.4 18 23 18.8 18 ...
##  $ Passengers        : int  5 5 5 6 4 6 6 6 5 6 ...
##  $ Length            : int  177 195 180 193 186 189 200 216 198 206 ...
##  $ Wheelbase         : int  102 115 102 106 109 105 111 116 108 114 ...
##  $ Width             : int  68 71 67 70 69 69 74 78 73 73 ...
##  $ Turn.circle       : int  37 38 37 37 39 41 42 45 41 43 ...
##  $ Rear.seat.room    : num  26.5 30 28 31 27 28 30.5 30.5 26.5 35 ...
##  $ Luggage.room      : int  11 15 14 17 13 16 17 21 14 18 ...
##  $ Weight            : int  2705 3560 3375 3405 3640 2880 3470 4105 3495 3620 ...
##  $ Origin            : Factor w/ 2 levels "USA","non-USA": 2 2 2 2 2 1 1 1 1 1 ...
##  $ Make              : Factor w/ 93 levels "Acura Integra",..: 1 2 4 3 5 6 7 9 8 10 ...
```

```r
dim(d1)
```

```
## [1] 93 27
```

```r
nrow(d1)
```

```
## [1] 93
```

SUBSETTING

```r
d2 = d1[c(1:10),c(2,4,9)]
```

```r
d2
```

```
##         Model Min.Price            AirBags
## 1     Integra      12.9               None
## 2      Legend      29.2 Driver & Passenger
## 3          90      25.9        Driver only
## 4         100      30.8 Driver & Passenger
## 5        535i      23.7        Driver only
## 6     Century      14.2        Driver only
## 7     LeSabre      19.9        Driver only
## 8  Roadmaster      22.6        Driver only
## 9     Riviera      26.3        Driver only
## 10    DeVille      33.0        Driver only
```

```r
# you may also try d2 = d1[c(1:10),]  and  d2 = d1[,c(2,4,9)]
```

```r
# Manufacturers and Prices
d2 = data.frame(d1$Manufacturer,d1$Price)
head(d2)
```

```
##   d1.Manufacturer d1.Price
## 1           Acura     15.9
## 2           Acura     33.9
## 3            Audi     29.1
## 4            Audi     37.7
## 5             BMW     30.0
## 6           Buick     15.7
```

```r
d2 = subset(d1,select=c(Manufacturer,Price))    # must use select explicitly
head(d2)
```

```
##   Manufacturer Price
## 1        Acura  15.9
## 2        Acura  33.9
## 3         Audi  29.1
## 4         Audi  37.7
## 5          BMW  30.0
## 6        Buick  15.7
```

```r
# Ford cars
d2 = subset(d1,subset = Manufacturer=="Ford")
d2[,1:7]
```

```
##    Manufacturer          Model    Type Min.Price Price Max.Price MPG.city
## 31         Ford        Festiva   Small       6.9   7.4       7.9       31
## 32         Ford         Escort   Small       8.4  10.1      11.9       23
## 33         Ford          Tempo Compact      10.4  11.3      12.2       22
## 34         Ford        Mustang  Sporty      10.8  15.9      21.0       22
## 35         Ford          Probe  Sporty      12.8  14.0      15.2       24
## 36         Ford       Aerostar     Van      14.5  19.9      25.3       15
## 37         Ford         Taurus Midsize      15.6  20.2      24.8       21
## 38         Ford Crown_Victoria   Large      20.1  20.9      21.7       18
```

```r
d2 = subset(d1,Price,subset = Manufacturer=="Ford")     # only prices
d2
```

```
##    Price
## 31   7.4
## 32  10.1
## 33  11.3
## 34  15.9
## 35  14.0
## 36  19.9
## 37  20.2
## 38  20.9
```

```r
d2 = subset(d1,c(Manufacturer,Price),subset = Manufacturer=="Ford")
d2
```

```
##    Manufacturer Price
## 31         Ford   7.4
## 32         Ford  10.1
## 33         Ford  11.3
## 34         Ford  15.9
## 35         Ford  14.0
## 36         Ford  19.9
## 37         Ford  20.2
## 38         Ford  20.9
```

```r
# Ford and Nissan cars
d2 = subset(d1,subset=Manufacturer=="Ford"|Manufacturer=="Nissan")
```

```
d2[,1:7]
```

```
##    Manufacturer          Model    Type Min.Price Price Max.Price MPG.city
## 31         Ford        Festiva   Small       6.9   7.4       7.9       31
## 32         Ford         Escort   Small       8.4  10.1      11.9       23
## 33         Ford          Tempo Compact      10.4  11.3      12.2       22
## 34         Ford        Mustang  Sporty      10.8  15.9      21.0       22
## 35         Ford          Probe  Sporty      12.8  14.0      15.2       24
## 36         Ford       Aerostar     Van      14.5  19.9      25.3       15
## 37         Ford         Taurus Midsize      15.6  20.2      24.8       21
## 38         Ford Crown_Victoria   Large      20.1  20.9      21.7       18
## 64       Nissan         Sentra   Small       8.7  11.8      14.9       29
## 65       Nissan         Altima Compact      13.0  15.7      18.3       24
## 66       Nissan          Quest     Van      16.7  19.1      21.5       17
## 67       Nissan         Maxima Midsize      21.0  21.5      22.0       21
```

```
d2 = subset(d1,c(Manufacturer,Price),subset=Manufacturer=="Ford"|Manufacturer=="Nissan")
d2
```

```
##    Manufacturer Price
## 31         Ford   7.4
## 32         Ford  10.1
## 33         Ford  11.3
## 34         Ford  15.9
## 35         Ford  14.0
## 36         Ford  19.9
## 37         Ford  20.2
## 38         Ford  20.9
## 64       Nissan  11.8
## 65       Nissan  15.7
## 66       Nissan  19.1
## 67       Nissan  21.5
```

```
# cars weighting > 3500
d2 = d1[d1$Weight>4000,]   # there are 4
d2[,1:7]
```

```
##    Manufacturer      Model  Type Min.Price Price Max.Price MPG.city
## 8         Buick Roadmaster Large      22.6  23.7      24.9       16
## 17    Chevrolet      Astro   Van      14.7  16.6      18.6       15
## 52      Lincoln   Town_Car Large      34.4  36.1      37.8       18
## 66       Nissan      Quest   Van      16.7  19.1      21.5       17
```

COUNTING

```
# how many exceeding 3000 lbs?
aux = d1$Weight
cars1=aux[aux>3000]
length(cars1)
```

```
## [1] 48
```

```
# there are 48 cars exceeding 3000 lbs

# number of cars by DriveTrain?
table(d1$DriveTrain)
```

```
##
##   4WD Front  Rear
##    10    67    16
```

```
# relative freq
prop.table(table(d1$DriveTrain))
```

```
##
```

```
##       4WD     Front      Rear
## 0.1075269 0.7204301 0.1720430
```

```
# by two factors
table(d1$AirBags,d1$DriveTrain)
```

```
##
##                      4WD Front Rear
##   Driver & Passenger   0    11    5
##   Driver only          5    28   10
##   None                 5    28    1
```

```
# how many cars by AirBags & DriveTrain & Passengers?
ftable(d1$AirBags,d1$DriveTrain,d1$Passengers)
```

```
##                           2  4  5  6  7  8
##
## Driver & Passenger 4WD    0  0  0  0  0  0
##                    Front  0  2  3  6  0  0
##                    Rear   0  3  1  1  0  0
## Driver only        4WD    0  1  1  0  3  0
##                    Front  0  5 16  7  0  0
##                    Rear   2  2  3  3  0  0
## None               4WD    0  2  1  0  1  1
##                    Front  0  8 15  1  4  0
##                    Rear   0  0  1  0  0  0
```

MEASURING

```
# median weight per DriveTrain
aux1=tapply(d1$Weight,d1$DriveTrain,median)
sort1=aux1[order(aux1)]   # in ascending order
sort1
```

```
## Front  Rear   4WD
##  2910  3520  3720
```

```
# relative freq
rel1=prop.table(aux1)
rel1
```

```
##       4WD     Front      Rear
## 0.3665025 0.2866995 0.3467980
```

```
# median weight per Airbags & DriveTrain
aux = list(d1$AirBags,d1$DriveTrain)
tapply(d1$Weight,aux,median)  # factors in a list()
```

```
##                      4WD  Front Rear
## Driver & Passenger    NA 3490.0 3515
## Driver only         3735 2970.0 3510
## None                2640 2552.5 3610
```

```
# change NA to 0
m1=tapply(d1$Weight,aux,median)
m1[is.na(m1)]=0
m1
```

```
##                      4WD  Front Rear
## Driver & Passenger     0 3490.0 3515
## Driver only         3735 2970.0 3510
## None                2640 2552.5 3610
```

SORTING

```
d2 = subset(d1,select=c(Manufacturer,Price,Weight,Width))
head(d2)
```

```
##   Manufacturer Price Weight Width
## 1        Acura  15.9   2705    68
## 2        Acura  33.9   3560    71
## 3         Audi  29.1   3375    67
## 4         Audi  37.7   3405    70
## 5          BMW  30.0   3640    69
## 6        Buick  15.7   2880    69
```

```
# sort by Width
d3 = d2[order(d2$Width),]
head(d3)
```

```
##    Manufacturer Price Weight Width
## 80       Subaru   8.4   2045    60
## 31         Ford   7.4   1845    63
## 39          Geo   8.4   1695    63
## 44      Hyundai   8.0   2345    63
## 83       Suzuki   8.6   1965    63
## 88   Volkswagen   9.1   2240    63
```

```
tail(d3)
```

```
##    Manufacturer Price Weight Width
## 75      Pontiac  17.7   3240    75
## 18    Chevrolet  18.8   3910    77
## 52      Lincoln  36.1   4055    77
## 8         Buick  23.7   4105    78
## 17    Chevrolet  16.6   4025    78
## 38         Ford  20.9   3950    78
```

```
# sort by Width and break ties by Weight
d3 = d2[order(d2$Width,d2$Weight),]
head(d3)
```

```
##    Manufacturer Price Weight Width
## 80       Subaru   8.4   2045    60
## 39          Geo   8.4   1695    63
## 31         Ford   7.4   1845    63
## 83       Suzuki   8.6   1965    63
## 88   Volkswagen   9.1   2240    63
## 44      Hyundai   8.0   2345    63
```

NAs

```
# rows with NAs
totals = rowSums(is.na(d1))
totals[totals>0]
```

```
## 16 17 19 26 36 56 57 66 70 87 89
##  1  1  2  1  1  1  2  1  1  1  1
```

```
d2 = d1[totals > 0,]
rownames(d2)
```

```
##  [1] "16" "17" "19" "26" "36" "56" "57" "66" "70" "87" "89"
```

```
# rows excluding NAs
index = as.integer(rownames(d2))
index
```

```
##  [1] 16 17 19 26 36 56 57 66 70 87 89
```

```
d3 = d1[-index,]    # wo NAs
totals2 = rowSums(is.na(d3))
totals2[totals2>0]
```

```
## named numeric(0)
```

SAMPLING

$\hat{y} = 0.934 - 0.25\,x_1 + 1.76\,x_2$

```
# Choose 4 cars at random
set.seed(1928)
x   = nrow(d1)
idx = sample(x,4)          # [1] 63 51 36 41
d1[idx,1:8]
```

```
##    Manufacturer    Model    Type Min.Price Price Max.Price MPG.city MPG.highway
## 6         Buick  Century Midsize      14.2  15.7      17.3       22          31
## 72     Plymouth    Laser  Sporty      11.4  14.4      17.4       23          30
## 68   Oldsmobile  Achieva Compact      13.0  13.5      14.0       24          31
## 75      Pontiac Firebird  Sporty      14.0  17.7      21.4       19          28
```