Numerical explorations with R

means, median, SD; use R as calc

Amos Thairu

January 17, 2015

Import data

```
> #import dataset
> data <- read.csv("bwmal.csv")
```

Import data

```
> #import dataset
> data <- read.csv("bwmal.csv")

> dim(data) #Returns the dimension of the dataset

[1] 791  12
```

Import data

```
> #import dataset
> data <- read.csv("bwmal.csv")

> dim(data) #Returns the dimension of the dataset

[1] 791  12

> names(data) #Returns variable names of the dataset

 [1] "X"         "matage"    "mheight"   "gestwks"   "sex"       "b
 [7] "smoke"     "pfplacen"  "parity"    "workload"  "matagegp"  "g
```

View data

```
> head(data) #Returns first six rows of dataset
```

	X	matage	mheight	gestwks	sex	bweight	smoke	pfplacen	parity	w
1	1	26	1.575	40	0	3.11	0	0	3	
2	2	23	1.529	40	0	2.65	0	0	1	
3	3	18	1.540	40	1	3.41	0	0	0	
4	4	25	1.581	40	1	2.99	0	0	2	
5	5	25	1.555	40	1	3.16	0	0	1	
6	6	21	1.561	40	1	2.82	0	0	1	

	gestcat
1	2
2	2
3	2
4	2
5	2
6	2

Some data explorations

```
> #some data explorations
> mean(data$mheight)

[1] 1.543273
```

Some data explorations

```
> #some data explorations
> mean(data$mheight)

[1] 1.543273

> var(data$mheight)

[1] 0.002884892
```

Some data explorations

```
> #some data explorations
> mean(data$mheight)
[1] 1.543273
> var(data$mheight)
[1] 0.002884892
> sd(data$matage)
[1] 5.139645
```

Some data explorations

```
> #some data explorations
> mean(data$mheight)
[1] 1.543273
> var(data$mheight)
[1] 0.002884892
> sd(data$matage)
[1] 5.139645
> median(data$matage)
[1] 23
```

More data explorations

```
> summary(data$matage)  #sumarize continous variable

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  13.00   20.00   23.00   23.78   27.00   46.00
```

Explorations for categorical variables

```
> table(data$sex)  #summarize categorical variable

  0   1
381 410
```

Explorations for categorical variables

```
> table(data$sex)  #summarize categorical variable

  0   1
381 410

> table(data$sex,data$smoke) #cross-tabulation of two categori

     0   1
  0 346  35
  1 378  32
```

Explorations for categorical variables

```
> table(data$sex)  #summarize categorical variable

  0   1
381 410

> table(data$sex,data$smoke) #cross-tabulation of two categori

     0   1
  0 346  35
  1 378  32

> with(data, table(sex,smoke)) #Alternatively, with variable l

    smoke
sex   0   1
  0 346  35
  1 378  32
```

Use R as calculator

```
> #enter the expression that we want evaluated and hit enter
> 1000-2*10^2/(8+2)

[1] 980

> #Built-in functions:
> log(1.4)  #returns the natural logarithm of the number 1.4

[1] 0.3364722

> log10(1.4)  # returns the log to the base of 10

[1] 0.146128

> sqrt(16)  #returns the square root of 16

[1] 4
```

Calculations with assignment statements

```
> #we can store a value(s) under some variable name using the
> #("less than" followed by a hyphen)
> x <- 2.5
> #to know what is in a variable type the variable name
> x
[1] 2.5
> #can store a computation under a new variable name or change
> y <- 3*log(x)
```