Lab2_Answer

Anyi Guo

09/10/2018

Lab 2 - Basic Statistics

1. Use the sample of following observations of variable x to find the values below.

```
x <- c(10, 2, 15, 6, 4, 9, 12, 11, 3, 0, 12, 10, 9, 7, 11, 10, 8, 5, 10, 6)
```

    a. n (number of observations)
    b. sum of all the observations in y
    c. mean
    d. median
    e. mode
    f. five number summary - Min, Q1, M, Q3, Max
    g. s2 (sample variance) Is it biased or unbiased?
    h. s (sample standard deviation)

```
n<-length(x)
s<-sum(x)
m1<-mean(x)
m2<-median(x)
mode <- function(v) {
   uniqv <- unique(v)
   uniqv[which.max(tabulate(match(v, uniqv)))]
}
mode(x)
```

```
## [1] 10
```

```
summary(x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    5.75    9.00    8.00   10.25   15.00
```

```
v<-var(x)
s<-sd(x)
```

2. Create a vector y of random normal variables. Let y be of length 10, with the same mean as x, and standard deviation 1.

```
y<-rnorm(10,mean=mean(x),sd=1)
```

a. Calculate the covariance and correlation between x and y. Can you do that? Why or why not?

Cannot do cov() or cor() as the dimensions of x and y don't match.

```
#cov(x,y)
#cor(x,y)
```

b. Now change your y so that it has the same length as x. The mean and standard deviation stay the same. Calculate the covariance and correlation between x and y again.

```
y<-rnorm(20,mean=mean(x),sd=1)
cov(x,y)
```

```
## [1] 1.3685
```

```
cor(x,y)
```

```
## [1] 0.4023529
```

c. Repeat b. several times. Did you get the same result every time? If not, why? What can you do make your result repeatable?

Use set.seed() so that the same group of random numbers may be generated every time.

d. Now pick an integer, say, 63. Run set.seed(63) before the rnorm function. Repeat the two functions for several times and check whether you get the same results from rnorm every time.

```
set.seed(63)
y<-rnorm(20,mean=mean(x),sd=1)
cov(x,y)
```

```
## [1] 1.192533
```

```
cor(x,y)
```

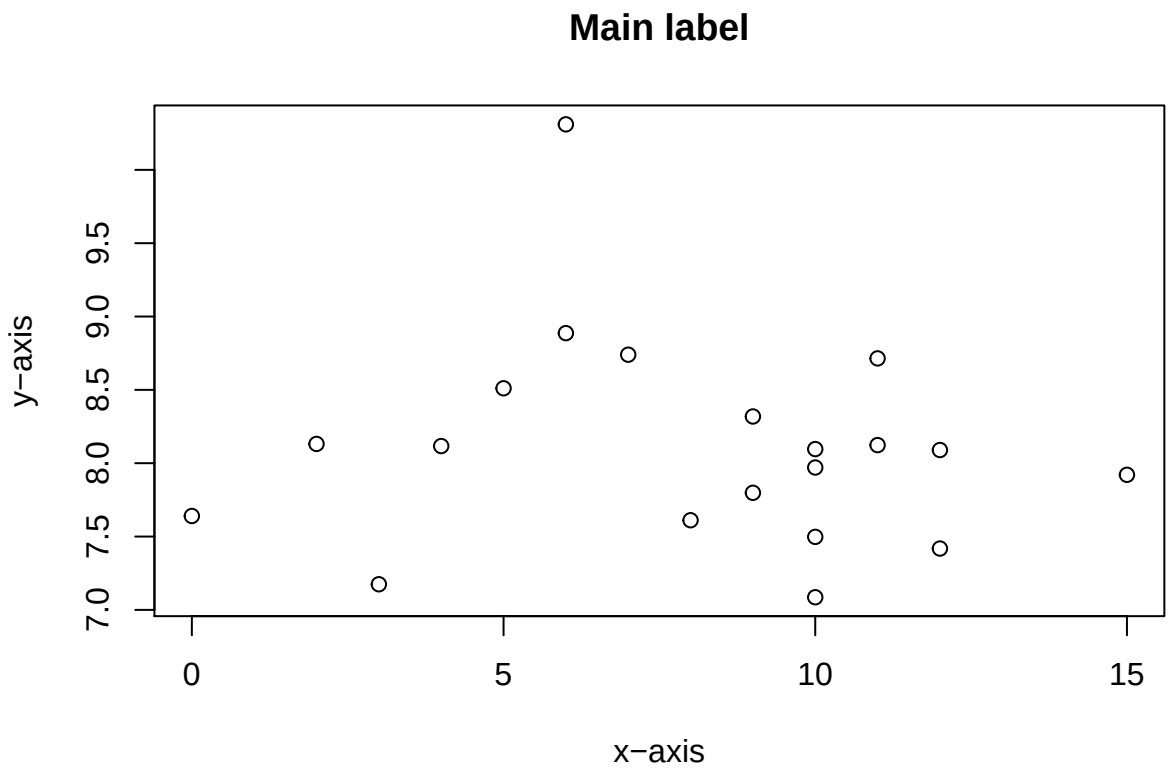
```
## [1] 0.3270855
```

The results are the same every time after using `set.seed()`

3. In this question, always set seed to be 100.

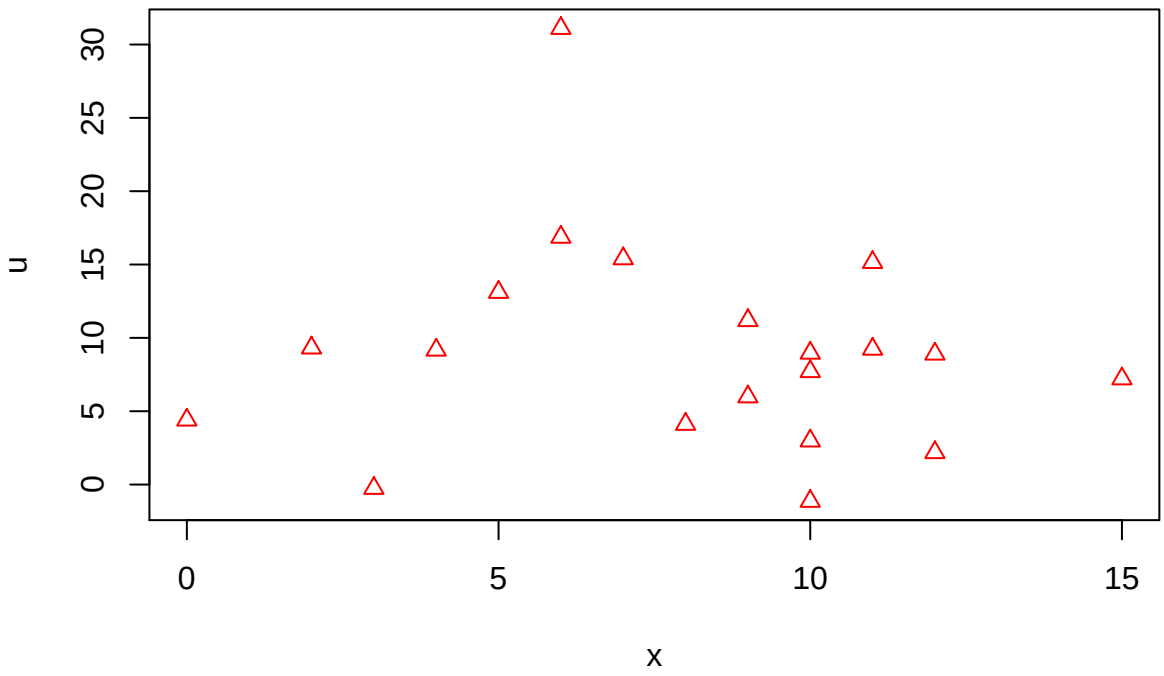
a. Create z of random normal variables of length 20, mean 8 and SD 1. Plot the scatter plot of x and z. Add the main label, x-label and z-label.

```
set.seed(100)
z<-rnorm(20,mean=8,sd=1)
plot(x,z,xlab="x-axis",ylab="y-axis",main="Main label")
```



b. Change the SD of z to 10 and obtain a new vector u. Plot the corresponding x and u. Draw the points onto the same plot as in 3a, and change the colour and symbol of the points.

```
set.seed(100)
u<-rnorm(20,mean=8,sd=10)
plot(x,u,col="red",pch=2)
```

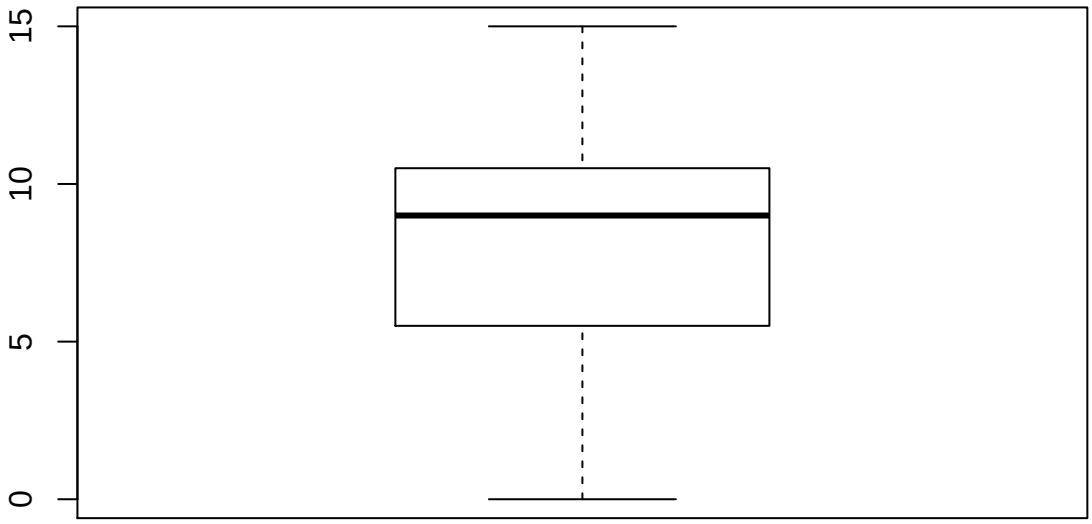


`pch=2` uses triangle as symbol of the points

4. Explore yourself how to plot a boxplot in R.

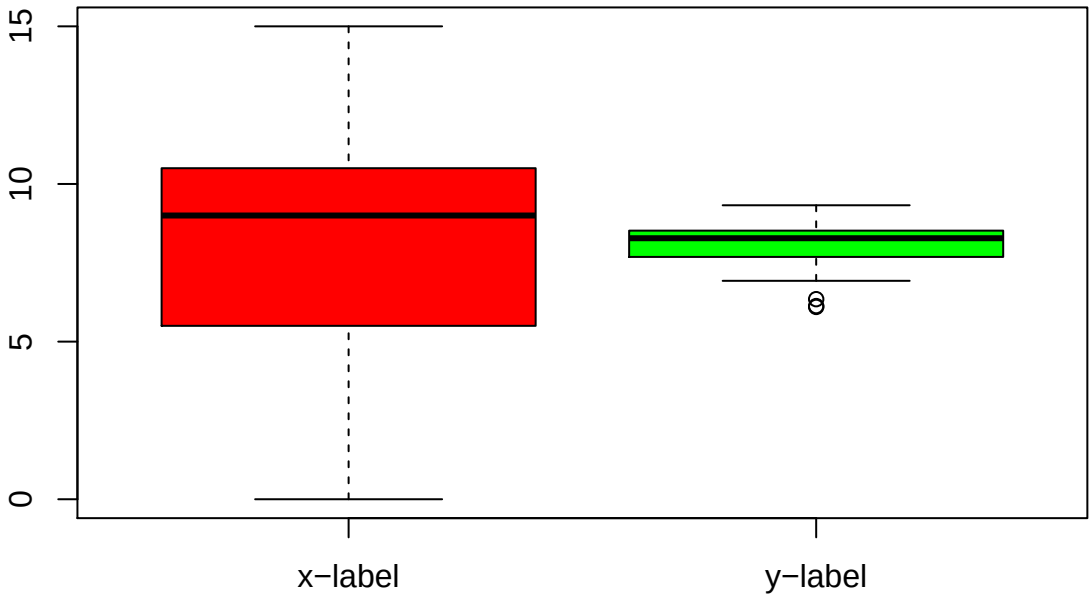
a. Plot vector x in boxplot.

```
boxplot(x)
```



b. Plot vector x and y in boxplot and display the result in one plot. See if you can add a label under each boxplot, and add some colour to each box.

```
boxplot(x,y,names=c("x-label","y-label"),col=c("red","green"))
```


15
10
5
0
x−label
y−label