September 11, 2014

R- Stat Examples

http://wiki.ubc.ca/LFS:Seminars/R

Note: The “>” is the command prompt line in R. What follows is the syntax commands you enter for computing stats.

> x <- c(1,2,5,7,9,10)

Reads sample data into x.

> x

[1] 1 2 5 7 9 10

Output or displays values in x array on screen.

> y <- c(3,4,8,1,12,20)

Read sample data into y array.

> y

[1] 3 4 8 1 12 20

Ouputs or displays values in y array on screen.

> mean(x)

[1] 5.666667

The computed mean or average of sample x. This is equal is to the sum of x values divided by the sample size.

> mean(y)

[1] 8

> min(x)

[1] 1

The minimum value of sample x.

> max(x)

[1] 10

The maximum value of sample x.

> sd(x)

[1] 3.669696

The standard deviation of x. the amount of variation or dispersion of the sample.

> median(x)

[1] 6

The value in sample x which falls in the middle when you order the sample values from smallest to largest.

>x <- read.table("c:\\rdata\\a.txt",header=T)

Reads in text file with 1 data value per row and first row is the header row

>mean(x[[1]])

Calculates the man for column 1 in the table. Mean is defined as the average value across the values in columns. To calculate sum all the values in column 1 and divide by the sample size or number of data values.

>sd(x[[1]])

Amount of variation or dispersion from the average (mean). Higher value indicates a larger range or spread of values.

>median(x[[1]])

The middle value in the distribution of your data when sorted from lowest to highest value.

x <- read.table("c:\\rdata\\c.txt",sep=" ",header=T)

Reads a table with 3 columns or variables of data.

> x[[1]]

[1] 32.2 37.4 35.9 28.1 41.0 44.5 35.6 31.2 34.8

Column 1 values

> x[[2]]

[1] 35.0 31.2 29.3 28.4 34.1 38.1 32.6 31.5 27.7

Column 2 values

> mean(x[[2]])

[1] 31.98889

> mean(x[[1]])

[1] 35.63333

> mean(x[[3]])

[1] 39.81111

> max(x[[3]])

[1] 48.9

> min(x[[1]])

[1] 28.1

> sum(x[[1]])

[1] 320.7

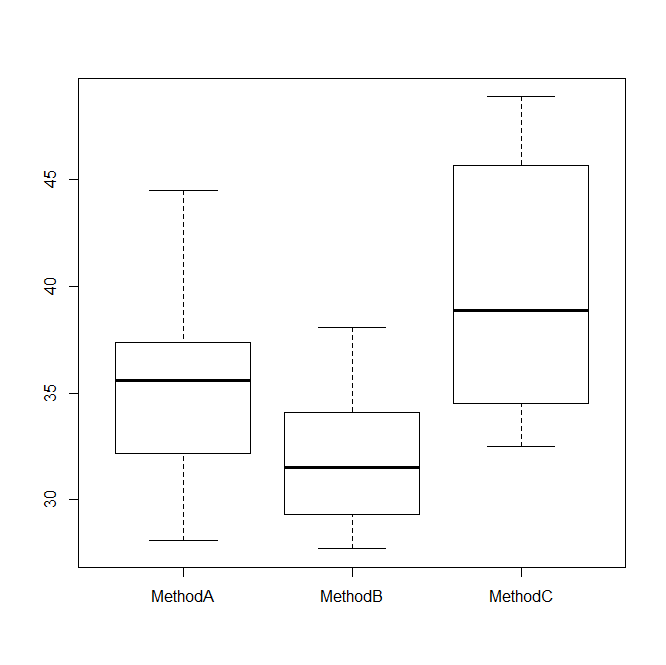
> sd(x[[1]])/mean(x[[1]])\*100

[1] 14.01425

This is the coefficient of variation, the standard deviation divided by the mean. Shows the extent of variability in relation to mean of the sample. The higher the value, the probability is higher the sample is farther from the mean.

> boxplot(x)

This will display boxplot, a visual representation of your data set.



> t.test(x[[1]],x[[2]])

This is a hypothesis testing, testing whether the means of Method A and Method B are equal and how likely they are equal.

Welch Two Sample t-test

data: x[[1]] and x[[2]]

t = 1.8156, df = 14.024, p-value = 0.09087

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.6601084 7.9489973

sample estimates:

mean of x mean of y

35.63333 31.98889

> t.test(x[[2]],x[[3]])

Welch Two Sample t-test

data: x[[2]] and x[[3]]

t = -3.4266, df = 12.625, p-value = 0.004685

A negative value for t indicates the mean for MethodA is less than the mean of MethodB.

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-12.768786 -2.875658

sample estimates:

mean of x mean of y

31.98889 39.81111

> var.test(x[[1]],x[[2]])

This tests whether the variances of MethodA and MethodB are equal by calculating the ratio of variances.

F test to compare two variances

data: x[[1]] and x[[2]]

F = 2.2018, num df = 8, denom df = 8, p-value = 0.2852

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.4966482 9.7610219

sample estimates:

ratio of variances

2.201771