September 12

September 12, 2014

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

> x

 MethodA MethodB MethodC

1 32.2 35.0 34.5

2 37.4 31.2 38.9

3 35.9 29.3 32.5

4 28.1 28.4 33.3

5 41.0 34.1 45.8

6 44.5 38.1 48.9

7 35.6 32.6 41.1

8 31.2 31.5 37.6

9 34.8 27.7 45.7

> Material = c(x[[1]])

> Material

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

> Material = c(x[[1]],x[[2]],x[[3]])

> Material

 [1] 32.2 37.4 35.9 28.1 41.0 44.5 35.6 31.2 34.8 35.0 31.2 29.3 28.4 34.1 38.1

[16] 32.6 31.5 27.7 34.5 38.9 32.5 33.3 45.8 48.9 41.1 37.6 45.7

> Method = c(rep("MethodA",9),rep("MethodB",9),rep("MethodC",9))

> mdata = data.frame(Method,Material)

> mdata

 Method Material

1 MethodA 32.2

2 MethodA 37.4

3 MethodA 35.9

4 MethodA 28.1

5 MethodA 41.0

6 MethodA 44.5

7 MethodA 35.6

8 MethodA 31.2

9 MethodA 34.8

10 MethodB 35.0

11 MethodB 31.2

12 MethodB 29.3

13 MethodB 28.4

14 MethodB 34.1

15 MethodB 38.1

16 MethodB 32.6

17 MethodB 31.5

18 MethodB 27.7

19 MethodC 34.5

20 MethodC 38.9

21 MethodC 32.5

22 MethodC 33.3

23 MethodC 45.8

24 MethodC 48.9

25 MethodC 41.1

26 MethodC 37.6

27 MethodC 45.7

> anovareport = aov(Material~Method,data=mdata)

> summary(anovareport)

 Df Sum Sq Mean Sq F value Pr(>F)

Method 2 275.8 137.88 5.758 0.00907 \*\*

Residuals 24 574.7 23.95

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> TukeyHSD(aov(lm(Material~Method),data=mdata),conf.level=0.95)

 Tukey multiple comparisons of means

 95% family-wise confidence level

Fit: aov(formula = lm(Material ~ Method), data = mdata)

$Method

 diff lwr upr p adj

MethodB-MethodA -3.644444 -9.405159 2.116270 0.2734463

MethodC-MethodA 4.177778 -1.582937 9.938492 0.1873878

MethodC-MethodB 7.822222 2.061508 13.582937 0.0065662

> summary(mdata)

 Method Material

 MethodA:9 Min. :27.70

 MethodB:9 1st Qu.:31.85

 MethodC:9 Median :34.80

 Mean :35.81

 3rd Qu.:38.50

 Max. :48.90

> mean(MethodA)

Error in mean(MethodA) : object 'MethodA' not found

> summary(x)

 MethodA MethodB MethodC

 Min. :28.10 Min. :27.70 Min. :32.50

 1st Qu.:32.20 1st Qu.:29.30 1st Qu.:34.50

 Median :35.60 Median :31.50 Median :38.90

 Mean :35.63 Mean :31.99 Mean :39.81

 3rd Qu.:37.40 3rd Qu.:34.10 3rd Qu.:45.70

 Max. :44.50 Max. :38.10 Max. :48.90

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