**Problem1:**

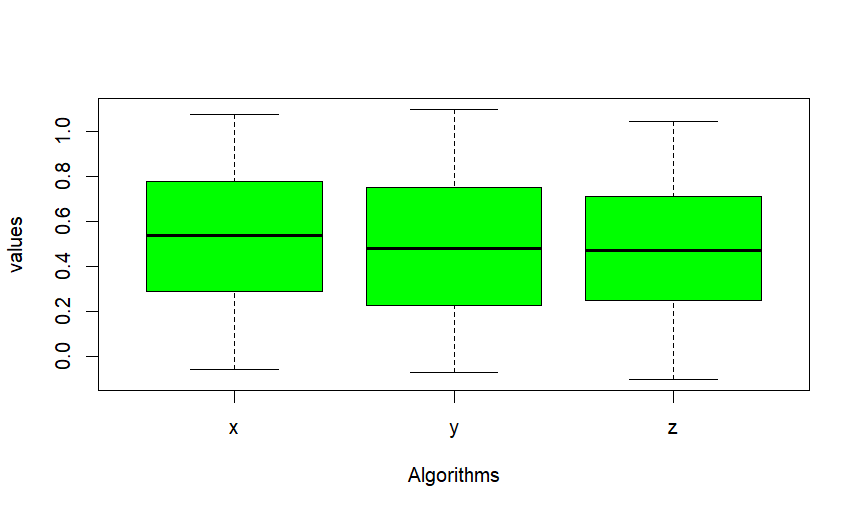
#Srinidhi Acharla

>

> set.seed(6476)

> data<- randu + rnorm(400)/ 23 #Random numbers with 3 alternative algorithms

> boxplot(data , xlab= "Algorithms" , ylab = "values", col = "Green") #Using boxplot and coloring



# As per the boxplot above, it is seen that the values for the three algorithms are showing similar random values falling in the range of 0.4 to 0.6

> cor(data)

x y z

x 1.00000000 -0.03023752 0.08214679

y -0.03023752 1.00000000 -0.03423609

z 0.08214679 -0.03423609 1.00000000

# The correlation matrix between the three algorithms x, y, z are as shown that the values are <= 1

> mean(data$x)

[1] 0.5260905

> mean(data$y) #Finding the mean for the algorithms x, y, z

[1] 0.4857142

> mean(data$z)

[1] 0.4806158

> t.test(data$x, data$y , data$z, paired = TRUE, mu=0, alternative = "two.sided", conf.level = 0.95) #It can be concluded that the algorithms have statistically different means

Paired t-test

data: data$x and data$y

t = 1.927, df = 399, p-value = 0.05469

alternative hypothesis: true mean difference is not equal to 0

95 percent confidence interval:

-0.0008165636 0.0815690336

sample estimates:

mean difference

0.04037623

#As per the above the p value is 0.05 and given the confidence level of 95%, the p-value is equal to that of the alpha value of 0.05. Hence the Null hypothesis can be rejected that the algorithms having the random numbers are similar.

#On top of this the mean difference in the values are 0.40. And the confidence interval is between -0.000 and 0.08

#T-value is below the 2 range which is of 1.92 and is closer to zero.

**Problem2:**

> set.seed(6476) #last 4 digits of your U-number

> before <-c(230.1, 220.9, 212.7, 293, 341.4, 296.9, 192.2, 175.5, 255.2, 293.7)

> # Weight of the mice after treatment

> after <-c(392.9, 393.2, 345.1, 393, 434, 427.9, 422, 383.9, 392.3, 352.2)

> # Create a data frame

> mydata <- data.frame(

+ group = rep(c("before", "after"), each = 10),

+ unit = c(seq(1:10),seq(1:10)),

+ score = c(before, after)

+ )

> mydata2 <- mydata[sample(1:nrow(mydata)),]

> mydata2 <- mydata2[1:19,]

#Above code selects just the length of 19. It takes the sample ranging from 1 to the length of 19 from the data present in the dataframe mydata

#Let us assume the level of confidence of 0.095 to find out the confidence intervals of before and after

> t.test(x = before, y= after, conf.level = 0.95)

Welch Two Sample t-test

data: before and after

t = -7.3714, df = 13.966, p-value = 3.553e-06

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

95 percent confidence interval:

-183.9586 -101.0214

sample estimates:

mean of x mean of y

251.16 393.65

>

> #As per the above test, it can be seen that the confidence intervals fall in the range of -183.95 to -101.02 for the 95% confidence levels.

> summary(mydata2)

group unit score

Length:19 Min. : 1.000 Min. :175.5

Class :character 1st Qu.: 3.000 1st Qu.:242.7

Mode :character Median : 5.000 Median :341.4

Mean : 5.421 Mean :317.2

3rd Qu.: 8.000 3rd Qu.:392.6

Max. :10.000 Max. :434.0

#As per the results it can be seen that the dataframe above cannot be tested as this is considered to be a corrupted dataframe. It has a length of 19 and the class is of characters and you cannot consider to test this with this type and requires preprocessing.