**Mann Whitney U Test (Wilcoxon Rank Sum Test)**

This module presents hypothesis testing techniques for situations with small sample sizes and outcomes that are ordinal, ranked or continuous and cannot be assumed to be normally distributed.

We are doing analysis if two population(security of personal information & security of payment information) are equal/similar to each other.

i. e. Does poeple give same rating of security to those two variable.

**Step1**: We extract the data and convert data in suitable format for testing purpose

**Input in R**

data = read.csv("/home/jayprakash/Videos/College/Project/Project\_data.csv")

#View(data) # To indentfiy index of colume we want

dim(data) #To see dimension of data

data1 = data[c(24,25)] #selecting data we want

colnames(data1) = c("Security\_Of\_Payment\_Info","Security\_Of\_Personal\_Info")

View(data1) #changing colume names and Viewing data

**Output in R**

[1] 250 32 #dimension of data

|  |  |  |
| --- | --- | --- |
|  | Security\_Of\_Payment\_Info | Security\_Of\_Personal\_Info |
| 1 | 4 | 4 |
| 2 | 4 | 4 |
| 3 | 3 | 3 |

. . .

|  |  |  |
| --- | --- | --- |
| 249 | 5 | 5 |
| 250 | 3 | 3 |

#output of View(data1)

**Step2:** checking normality of data

**Input in R**

shapiro.test(data1$Security\_Of\_Payment\_Info)

shapiro.test(data1$Security\_Of\_Personal\_Info)

**Output in R**

**#For 1st line**

**Shapiro-Wilk normality test**

data: data1$Security\_Of\_Payment\_Info

W = 0.87312, p-value = 1.479e-13 = 0.00003343

**#For 2nd line**  
**Shapiro-Wilk normality test**

data: data1$Security\_Of\_Personal\_Info

W = 0.89663, p-value = 4.572e-12 = 0.0000280913

Since p-value < alpha(0.05), we reject H0, i.e. our data in non-normal

so we can’t use any parameteric test. We have to use non-parametric test.

for this kind of problem mann whitney U test is suitable. Since it is ordered data.

**Step3:** using mann whitney U test

**Input in R**

wilcox.test(data1$Security\_Of\_Payment\_Info, data1$Security\_Of\_Personal\_Info)

**Output in R**

**Wilcoxon rank sum test with continuity correction**

data: data1$Security\_Of\_Payment\_Info and data1$Security\_Of\_Personal\_Info

W = 33804, p-value = 0.1002

alternative hypothesis: true location shift is not equal to 0

Since p-value > alpha(0.05), we accept H0.

**Conculsion:**

We have statistically significant evidence at α =0.05, to show that the two populations of security of Payment information and Security of Personal information are equal/same/similar.