Objective

1. To find out the significant difference exit or not
2. To find out the optimal rank order list
3. To find out the conditional probability of each subject

Rank Order analysis in R

Let's upload relevant libraries and data points into R environment.

```
rm(list=ls())
library(RankAggreg)
library(e1071)
library(xlsx)
library(stringr)
data<-read.xlsx("D:/RStudio/Website/data.xlsx",sheetName="Sheet1")
head(data)
```

Outcome

CID Product Rank

CID	Product	Rank
5	A	1
1	B	1
6	C	1
1	D	1
1	E	1
1	A	2

What is business forecasting?

Approach 1-Rank Aggregation

Rank aggregation help you to find the optimal rank order. In other way aggregation converts multi rank list into single rank list.

Brute aggregate is used for find out the optimal list.

Rank Aggregation-Rank Order analysis in R

```
data$Product<-toupper(data$Product)
colnames(data)<-c("CID","Product","Rank")
data<-data[order(data$Rank),]
wide2<-reshape(data,v.names="Product",idvar=c("CID"),
timevar="Rank",direction="wide")
wide2<-as.matrix(wide2[,-1])
rownames(wide2)<-NULL
colnames(wide2)<-str_replace_all(colnames(wide2),
pattern="Product.",repl="")
```

```
BruteAggreg(wide2,length(unique(data$Product)))
```
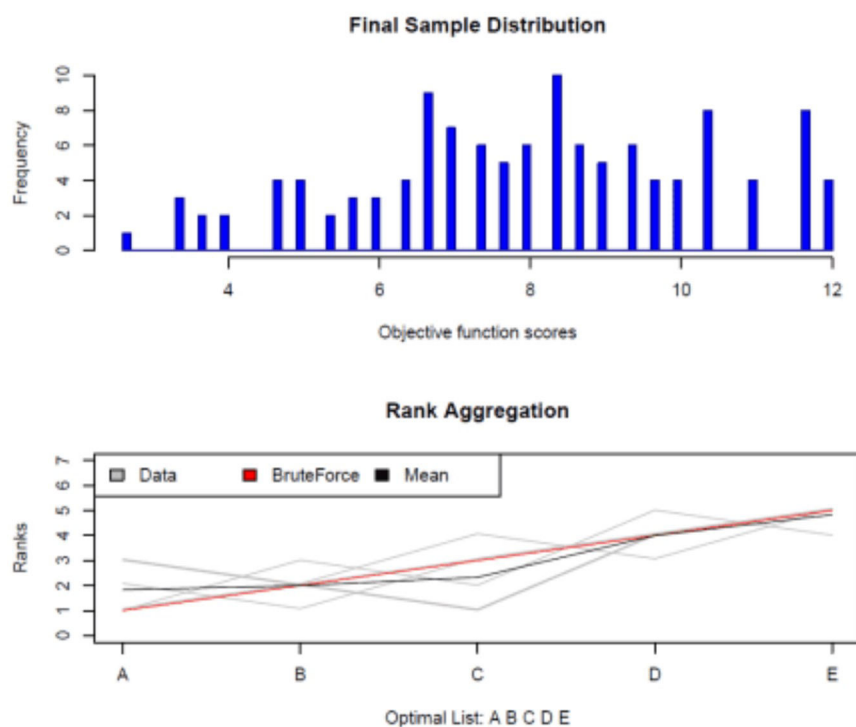
Output

Optimal List

The optimal list is: A B C D E

  Algorithm: BruteForce

  Distance: Spearman

  Score: 2.666667

Rank Aggregate Plot

Significance of Spearman's Rank Correlation

Final Sample Distribution

Rank Aggregation

Optimal List: A B C D E

Approach2- Contingency table analysis

Based on Fisher/Chi square method is used for identify the significance difference between products.

Contingency table provides the integer count for the measurements with respect to the categorical variables.

You can use compare group package here, for easy application, compare group will execute fisher exact or chi-square test automatically, depends on the number of contingency table cell frequencies.

```
dataA<-data[,c(-1)]
library(Hmisc)
ansA <- compareGroups(Product~ ., data = dataA,max.ylev = 30,max.xlev
=15)
```

```
ansA <- update(ansA, method =3)
anstabA<-createTable(ansA,show.p.mul = TRUE)
```

Output

	A	B	C	D	E	p.overall	p.A vs B	p.A vs C	p.A vs D	p.A vs E	p.B vs C	p.B vs D	p.B vs E	p.C vs D	p.C vs E	p.D vs E
	N=6	N=6	N=6	N=6	N=6											
Rank:						<0.001	0.438	1	0.082	0.032	0.462	0.038	0.032	0.346	0.038	0.133
1	3 (50.0%)	1 (16.7%)	2 (33.3%)	0 (0.00%)	0 (0.00%)											
2	1 (16.7%)	4 (66.7%)	1 (16.7%)	0 (0.00%)	0 (0.00%)											
3	2 (33.3%)	1 (16.7%)	2 (33.3%)	1 (16.7%)	0 (0.00%)											
4	0 (0.00%)	0 (0.00%)	1 (16.7%)	4 (66.7%)	1 (16.7%)											
5	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (16.7%)	5 (83.3%)											

Approach 3-Conditional Probability

If you try to calculate probability of a particular condition is called as conditional probability. In other words, Probability of happening of an event A when event B has already taken care.

The example of naive conditional probabilities are automatic identification of spam mails, rain prediction based on particular conditions etc…

Bayes theorem is P(A/B) = [p(B/A) * p(A)] / p(B)

```
naiveBayes(Rank ~., data=data)
```

output

 Product

Y	A	B	C	D	E
1	**0.5000000**	0.1666667	**0.3333333**	0.0000000	0.0000000
2	0.1666667	**0.6666667**	0.1666667	0.0000000	0.0000000
3	0.3333333	0.1666667	**0.3333333**	0.1666667	0.0000000
4	0.0000000	0.0000000	0.1666667	**0.6666667**	0.1666667
5	0.0000000	0.0000000	0.0000000	0.1666667	**0.8333333**

Conclusion

Based on compare group analysis able to identify the significant difference between samples.

Significant difference was observed between A vs E, B vs D, B vs E, C vs E.

Based on Rank aggregation method able to identify the optimal order is A B C D E.