MULTIVARIATE ANALYSIS (LAB) [MSC523]
(L-T-P: 0-0-3)

LAB MANUAL

Department of Management Studies
IIT(ISM)-Dhanbad

II-MBA (BA)

Course Type	Course Code	Name of Course	L	T	P	Credit
DC	MSC 523	Multivariate Analysis (Lab)	0	0	3	3

Course Objective
This course will demonstrate the theory of multivariate statistical methods and their applications.
Learning Outcomes
Upon successful completion of this course, students will:

- Learn about the application of Multivariate Statistical Methods in Data Analytics.
- Get familiar with the application software like SPSS, SYSTAT for solving multivariate data analysis problems.

Unit No.	Topics to be Covered	Lecture Hours	Learning Outcome
1	Overview of Multivariate Methods: Types of Multivariate Technique, Structured approach to Multivariate Model Building, Examination of Data	3	Understand the fundamental concept of Multivariate Methods
2	Exploratory Factor Analysis: Factor Analysis Decision Process and applications,	6	Understanding the use of factor analysis
3	Classification Technique: Multiple Discriminant Analysis, Cluster analysis – Hierarchal and non-hierarchal clustering procedures (K-Means Clustering algorithm) and applications	9	Understanding the use of classification methods
4	Multiple Regression Analysis: Decision process and illustrations; Logistic regression (Binary Logit analysis) with illustrations	9	Understanding regression models and their applications
5	Likelihood ratio principles, Hotelling's T2 and MANOVA: Decision Process and illustrations	6	Using Multivariate Analysis of Variance technique
6	Conjoint Analysis: Choice–based conjoint approach, Part worth estimation, Decision process and illustrations	6	Understand conjoint analysis and its application
	TOTAL	**39**	

Text Books:

1. Hair et al. (2015), Multivariate Data Analysis. 7th Ed. Pearson
2. R. A. Johnson and D. W. Wichern (2013), Applied Multivariate Statistical Analysis. 6th Ed. Pearson

Reference Books:

1. C. R. Rao (2002), Linear Statistical Inference and its Applications. 2nd Ed. Wiley
2. M. S. Srivastava and C. G. Khatri (1979), An Introduction to Multivariate Statistics, Elsevier North Holland, Inc., New York
3. R. J. Muirhead (2009). Aspects of Multivariate Statistical Theory. 2nd Ed. Wiley- Interscience.

SL. NO	EXPERIMENT NO.
1	EXPERIMENT NO. 1
2	EXPERIMENT NO. 2
3	EXPERIMENT NO. 3
4	EXPERIMENT NO. 4
5	EXPERIMENT NO. 5
6	EXPERIMENT NO. 6
7	EXPERIMENT NO. 7
8	EXPERIMENT NO. 8
9	EXPERIMENT NO. 9
10	EXPERIMENT NO. 10
11	EXPERIMENT NO. 11

Experiment 1

Objective: To understand and perform Simple factor analysis on the given data

Details: Do the Factor Analysis to understand its importance in data analytics

Tool/Software: R-Studio

Procedure:

1. Create a Working Directory (setwd("C:/Users/acer/Desktop/AMA")
2. Download and open libraries (library(psych), library(GPArotation))
3. Perform the following R-Script for entering and viewing the data
   data <- read.csv("FA.csv")
   View(data)
   str(data)

4. Perform the following R-Script for generating number of possible factors from data
   NIF <- fa.parallel(data, fm = "minres", fa = 'fa')

5. Perform the following R-Script for Factor Analysis
   FactorAna <- fa(data, nfactors = 4, rotate = "oblimin", fm = "minres")
   print(FactorAna$loadings,cutoff = 0.3)

Experiment 2

Objective: To understand and perform in-depth factor analysis on the given data

Details: Do the Factor Analysis and important tests to understand the key concept and data adequacy for factor analysis

Tool/Software: R-Studio

Procedure:

1. Create a Working Directory (setwd("C:/Users/acer/Desktop/AMA")
2. Download and open libraries (library(psych),library(corrplot), library(HH),library(ppcor))
3. Perform the following R-Script for entering and viewing and description of the data

   ```
   data <- read.csv("Factor-Hair-Revised.csv")
   head(data)
   tail(data)
   dim(data)
   str(data)
   summary(data)
   boxplot(data)
   describe(data)
   ```
4. Perform the following R-Script for doing in-depth Factor Analysis

   ```
   describe(data)
   data$ID <- NULL
   cor(data)
   library(corrplot)
   a <- cor(data)
   corrplot(a,method = "number")
   pcor(data,method = "pearson")
   r <- lm(Satisfaction~.,data )
   summary(r)
   ```

#variable inflation factor >2.5 multi-collinearity is high

```
vif(r)
data$DelSpeed <- NULL
b <- lm(Satisfaction~.,data )
vif(b)
data$CompRes <- NULL
c <- lm(Satisfaction~.,data )
vif(c)
X <- data[,-c(1,13)]
Y <- data[,13]
Cor <- cor(X)
KMO(X)
```

if MSA value >0.5 data is adequate so factor analysis can be done.

```
NOF<-fa.parallel(X, fm = "minres", fa = 'fa')
```

```r
Fact_Ana = fa(data,nfactors = 4,rotate = "oblimin", fm = "minres")
print(Fact_Ana$loadings,cutoff = 0.5)
fa.diagram(Fact_Ana)
Fact_Ana$scores
finaldata<- cbind(Y,Fact_Ana$scores)
names(finaldata)<-
c("satisfaction","purchase","marketing","post_purchase","product_positioning")
```

Experiment 3

Objective: With reference to the factors obtained from Experiment 4 do a multiple regression analysis

Details: Do the multiple-regression and its interpretations

Tool/Software: R-Studio

Procedure:

1. Create a Working Directory (setwd("C:/Users/acer/Desktop/AMA")
2. Download and open libraries (library(psych))
3. Perform the following R-Script for doing multiple regression
   reg <- lm(satisfaction~.,finaldata)
   finaldata <- as.data.frame(finaldata)
   summary(reg)

Experiment 4

Objective: To perform different test to obtained number of possible cluster for factor analysis

Details: To test silhouette, elbow and gapstat method to obtain number of cluster

Tool/Software: R-Studio

Procedure:

1. Create a Working Directory (setwd("C:/Users/acer/Desktop/AMA")
2. Download and open libraries library(factoextra), library(cluster), library(animation)
3. Perform the following R-Script for entering and scaling the data
   ```
   data('USArrests')
   Data <- na.omit(USArrests)
   Data <- scale(Data)
   str(Data)
   ```

4. Perform the following R-Script to test silhouette and gap-stat method to obtain number of cluster
   ```
   KM1 <- eclust(Data, "kmeans", nstart = 25)
   fviz_gap_stat(KM1$gap_stat)
   fviz_silhouette(KM1)
   ```

Experiment 5

Objective: To perform hierarchical clustering

Details: do the hierarchical clustering and visualization for given data with interpretation

Tool/Software: R-Studio

Procedure:

1. Create a Working Directory (setwd("C:/Users/acer/Desktop/AMA")
2. Download and open libraries library(factoextra), library(cluster), library(tidyverse), library(dendextend)
3. Perform the following R-Script for entering and scaling the data
   ```
   data('USArrests')
   Data <- na.omit(USArrests)
   Data <- scale(Data)
   str(Data)
   ```
4. Perform the following R-script for hierarchical clustering
   ```
   distance <- dist(data,method = 'euclidean')
   view(distance)
   #single linkage
   hcluster <- hclust(distance,method = 'single')
   plot(hcluster,cex = 0.8,hang = -1)
   hcluster1 <- agnes(distance,method = 'single')
   plot(hcluster1,cex = 0.5,hang = -1)
   #complete Linkage
   hcluster2 <- hclust(distance,method = 'complete')
   plot(hcluster2,cex = 0.8,hang = -1)
   hcluster3 <- agnes(distance,method = 'complete')
   plot(hcluster3,cex = 0.5,hang = -1)
   #centroid
   hcluster4 <- hclust(distance,method = 'centroid')
   plot(hcluster4,cex = 0.8,hang = -1)
   hcluster5 <- agnes(distance,method = 'centroid')
   plot(hcluster5,cex = 0.5,hang = -1)
   #wards
   hcluster6 <- hclust(distance,method = 'ward.D')
   plot(hcluster6,cex = 0.8,hang = -1)
   hcluster7 <- agnes(distance,method = 'ward')
   plot(hcluster7,cex= 0.5,hang = -1)
   #average
   hcluster8 <- hclust(distance,method = 'average')
   plot(hcluster8,cex = 0.8,hang = -1)
   hcluster9 <- agnes(distance,method = 'average')
   plot(hcluster9,cex =0.5,hang =-1)
   #cophenetic corelation
   hcluster_co <- cophenetic(hcluster)
   cor(distance,hcluster_co)
   ```

```r
hcluster_co1 <- cophenetic(hcluster1)
cor(distance,hcluster_co1)
hcluster_co2 <- cophenetic(hcluster2)
cor(distance,hcluster_co2)
hcluster_co4 <- cophenetic(hcluster4)
cor(distance,hcluster_co4)
hcluster_co6 <- cophenetic(hcluster6)
cor(distance,hcluster_co6)
hcluster_co8 <- cophenetic(hcluster8)
cor(distance,hcluster_co8)
plot(distance,hcluster_co8)
plot(distance,hcluster_co)
#if cophenetic correlation is high means clusters are seaprable
#if gower distance is lower clustering method is better
godists <- sum((distance-hcluster_co)^2)
godists1 <- sum((distance-hcluster_co8)^2)
```

Experiment 6

Objective: To perform K-Mean clustering

Details: do the K-Mean clustering and visualization for given data with interpretation

Tool/Software: R-Studio

Procedure:

1. Create a Working Directory (setwd("C:/Users/acer/Desktop/AMA")
2. Download and open libraries library(cluster),library(factoextra) library(animation)
3. Perform the following R-Script for entering and scaling the data
   data('USArrests')
   Data <- na.omit(USArrests)
   Data <- scale(Data)
   str(Data)
4. Perform the following R-script for K-Meanclustering
   kmeans(Data, 4)
   kmeans.ani(Data, 4)
   KM <- kmeans(Data, 4, nstart = 50)

Experiment 7

Objective: To perform Logistic Regression for Classification

Details: do the Logistic Regression for Classification for given data with interpretation

Tool/Software: R-Studio

Procedure:

1. Create a Working Directory (setwd("C:/Users/acer/Desktop/AMA")
2. Download and open libraries library(caret)
3. Enter the data using following R-Script
   data <- read.csv("Intention.csv")
   str(data)
   View(data)
   data$Revenue <- as.factor(data$Revenue)
4. Split the data for training and testing
   split <- floor(0.66*nrow(data))
   set.seed(123)
   sample<- sample(seq_len(nrow(data)),size = split)
   train <-data[sample,]
   test <- data[-sample,]
5. Building Logistic model and interpretation of parameters
   control <- trainControl(method = "cv" , summaryFunction = twoClassSummary ,
   classProbs = T ,savePredictions = T)
   model_log <- train(Revenue~.,method = 'glm', data = train ,trControl = control)
   predict <- predict(model_log,test,type = 'raw')
   CM_log <- table(predict,test$Revenue,dnn = c('Predicted','Actual'))
   confusionMatrix(CM_log)

Experiment 8

Objective: To perform Linear Discriminant Analysis (LDA) for Classification

Details: do the Discriminant Analysis (LDA) for Classification for given data with interpretation

Tool/Software: R-Studio

Procedure:

1. Create a Working Directory (setwd("C:/Users/acer/Desktop/AMA")
2. Download and open libraries library(caret)
3. Enter the data using following R-Script
   data <- read.csv("Intention.csv")
   str(data)
   View(data)
   data$Revenue <- as.factor(data$Revenue)
4. Split the data for training and testing
   split <- floor(0.66*nrow(data))
   set.seed(123)
   sample<- sample(seq_len(nrow(data)),size = split)
   train <-data[sample,]
   test <- data[-sample,]
5. Building Discriminant Analysis (LDA) and interpretation of parameters
   control <- trainControl(method = "cv" , summaryFunction = twoClassSummary ,
   classProbs = T ,savePredictions = T)
   model_lda <- train(Revenue~.,method = 'lda', data = train ,trControl = control)
   predict1 <- predict(model_lda,test,type = 'raw')
   CM_lda <- table(predict,test$Revenue,dnn = c('Predicted','Actual'))
   confusionMatrix(CM_lda)

Experiment 9

Objective: To prepare data for conjoint analysis

Details: Make data profiles for conjoint analysis

Tool/Software: R-Studio

Procedure:

1. Create a Working Directory (setwd("C:/Users/acer/Desktop/AMA")
2. Data Design
   #car =Engine,Price,Color,Mileage)
   #Engine = Low,Medium,High
   #Price = Low,Medium,High
   #Color =Red,Green,Silver
   #Mileage = Average,High
3. No. of profile may be obtained
   #No of profiles = 3^3*2^1 = 54
4. Run the following R-Script for Coinjoint analysis

   ```
   Engine <- c('Low','high','Medium')
   Price <- c('Low','high','Medium')
   Color <- c('Red','green','Silver')
   Mileage <- c('Average','high')

   Attr <- expand.grid(Engine,Price,Color,Mileage)
   Design = caFactorialDesign(Attr, type = 'full')
   Design1 = caFactorialDesign(Attr,type = 'orthogonal', cards = 18)
   Design2 = caFactorialDesign(Attr, type = 'fractional')
   Design3 = caFactorialDesign(Attr, type = 'ca')
   Design4 = caFactorialDesign(Attr, type = 'aca')

   cor(caEncodedDesign(Design))
   cor(caEncodedDesign(Design1))
   cor(caEncodedDesign(Design2))
   cor(caEncodedDesign(Design3))
   cor(caEncodedDesign(Design4))
   ```

Experiment 10

Objective: To do the conjoint analysis

Details: perform conjoint analysis with interpretation

Tool/Software: R-Studio (Library)

Procedure:

1. Create a Working Directory (setwd("C:/Users/acer/Desktop/AMA")
2. Open and view the data
   data(tea)
   View(tprof)
   View(tsimp)
   View(tlevn)
   View(tprefm)
   View(tpref)
3. Perform conjoint analysis with following R-Script
   imp <- caImportance(tprefm,tprof)
   summary(imp)
   logit <- caLogit(tsimp,tpref,tprof)
   summary(logit)
   sum(logit)
   maxutility <- caMaxUtility(tsimp,tpref,tprof)
   summary(maxutility)
   model <- caModel(tprefm[2,],tprof)
   utility <- caPartUtilities(tprefm,tprof,tlevn)
   total <- caTotalUtilities(tprefm,tprof)
   uti <- caUtilities(tprefm,tprof,tlevn)
   decision <- Conjoint(tprefm,tprof,tlevn,y.type = "score")

Experiment 11

Objective: To perform ANOVA and MANOVA

Details: ANOVA and MANOVA with interpretation

Tool/Software: R-Studio

Procedure:

1. Create a Working Directory (setwd("C:/Users/acer/Desktop/AMA")
2. Read Data
   data1 <- read.csv("DietWeightLoss.csv")
   View(data1)
   names(data1)
   class(data1$WeightLoss)
   class(Diet)
   data1$Diet <- as.factor(data1$Diet)
   levels(data1$Diet)
   boxplot(data1$WeightLoss~data1$Diet)
3. R-Script for ANOVA
   #H0 : Mean weight loss is same for all diets
   #Ha : Mean weight loss is not same for all diets

   model <- aov(data1$WeightLoss~data1$Diet)
   summary(model)

   #Null hypothesis is rejected

   #tukey test
   plot(TukeyHSD(model))

   #kruskal test
   kruskal.test(data1$WeightLoss~data1$Diet)
4. R-Script for MANOVA

   data2 <- read.csv("test.csv")
   View(data2)
   str(data2)
   data2$teaching_method <- as.factor(data2$teaching_method)
   summary(data2)

   m1 <- lm(data2$t1 ~ data2$teaching_method)
   m2 <- lm(data2$s1 ~ data2$teaching_method)

   summary(m1)
   summary(m2)

   y <- cbind(data2$s1,data2$t1)

```
model1 <- manova(y~data2$teaching_method)
summary(model1)

###################

data3 <- read.csv("skulls.csv")
View(data3)
str(data3)
data3$epoch <- as.factor(data3$epoch)
levels(data3$epoch)
m1 <- lm(data3$mb~data3$epoch)
m2 <- lm(data3$bh~data3$epoch)
m3 <- lm(data3$bl~data3$epoch)
m4 <- lm(data3$nh~data3$epoch)

summary(m1)
## epochc4000BC is significant
summary(m2)
## epochcAD150 is significant
summary(m3)
## epochc3300BC,epochc4000BC,epochcAD150 are significant
summary(m4)
## no significant variables

y <- cbind(data3$mb,data3$bh,data3$bl,data3$nh)
model <- manova(y~data3$epoch)
summary(model)
```