Chapter 5 DICA

The main idea behind DICA is to represent each group by the sum of its observations and to perform a simple CA on the groups by variables matrix. The original observations are then projected as supplementary elements and each observation is assigned to the closest group. The comparison between the a priori and the a posteriori classifications can be used to assess the quality of the discrimination. A similar procedure can be used to assign new observations to categories. The stability of the analysis can be evaluated using cross-validation techniques such as jackknifing or bootstrapping

NOTE(In this project we are using DICA for analysis not for prediction)

LETS START BY CLEARING THE ENVIRONMENT.

```
knitr::opts_chunk$set(echo = TRUE)
rm(list = ls())
graphics.off()
```

5.1 DATASET

The dataset is the same which is used in BADA

Let us check if we get the same interpretation or not.

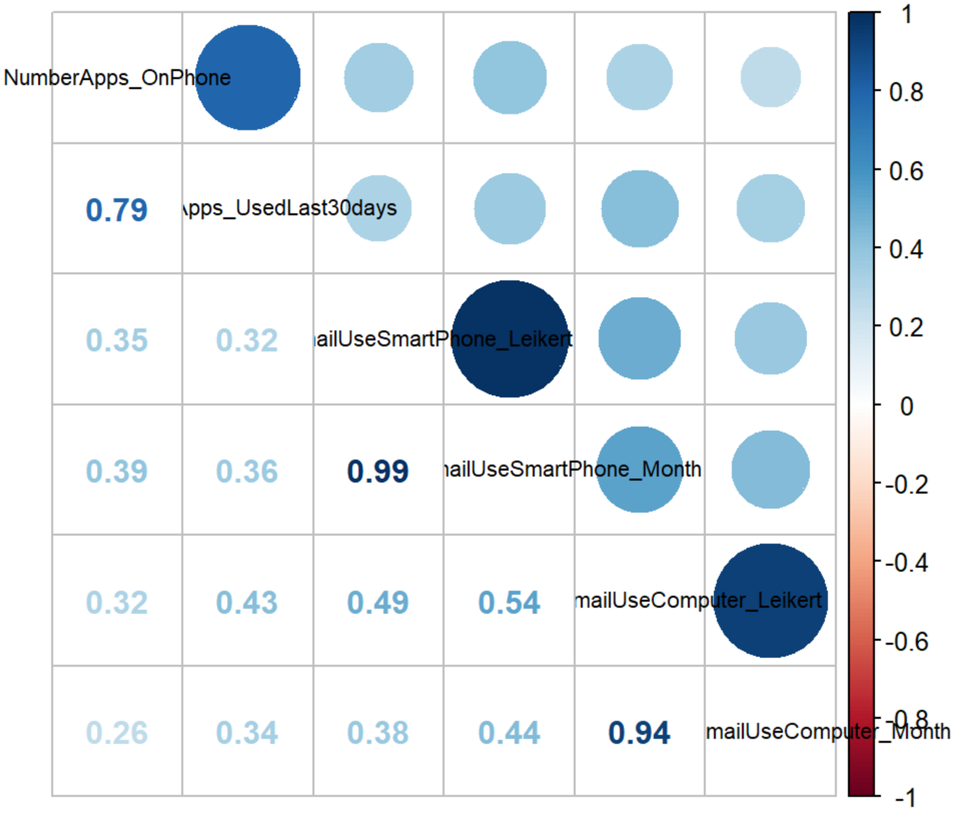
We have excluded purchase decision same as bada. (it makes more sense to exclude the variable which does not correlate significantly with anyother variable)

```
load("C:/Users/jeevan/Desktop/RM2/R-M/SmartphoneUsage (1).RData")
#install.packages('tidyverse')
library(tidyverse)
library(ExPosition)
#install.packages('TExPosition') # if needed
library(TExPosition)
library(TInPosition)
library(PTCA4CATA)
# devtools::install_github('HerveAbdi/data4PCCAR')
library(data4PCCAR)
b=data.more
b["C"]=c("Am","E","Am","A","E","E","E","E","E","A","A","E","E","A","A","Am","E","E","A","A'

rawdata=data.pca[-c(22),]
Xmat <-  rawdata[, 2:4]
Ymat <-  rawdata[, 5:7]
XYmat <- rawdata[, 2:7]
```

5.2 COR PLOT

Just as bada we will use the corplot to confirm our correlations between variables.

```
library(corrplot)


cor.res <- cor(XYmat)


corrplot.mixed(cor.res, tl.cex = 0.7, tl.col = "black")
```



5.3 SCREE PLOT

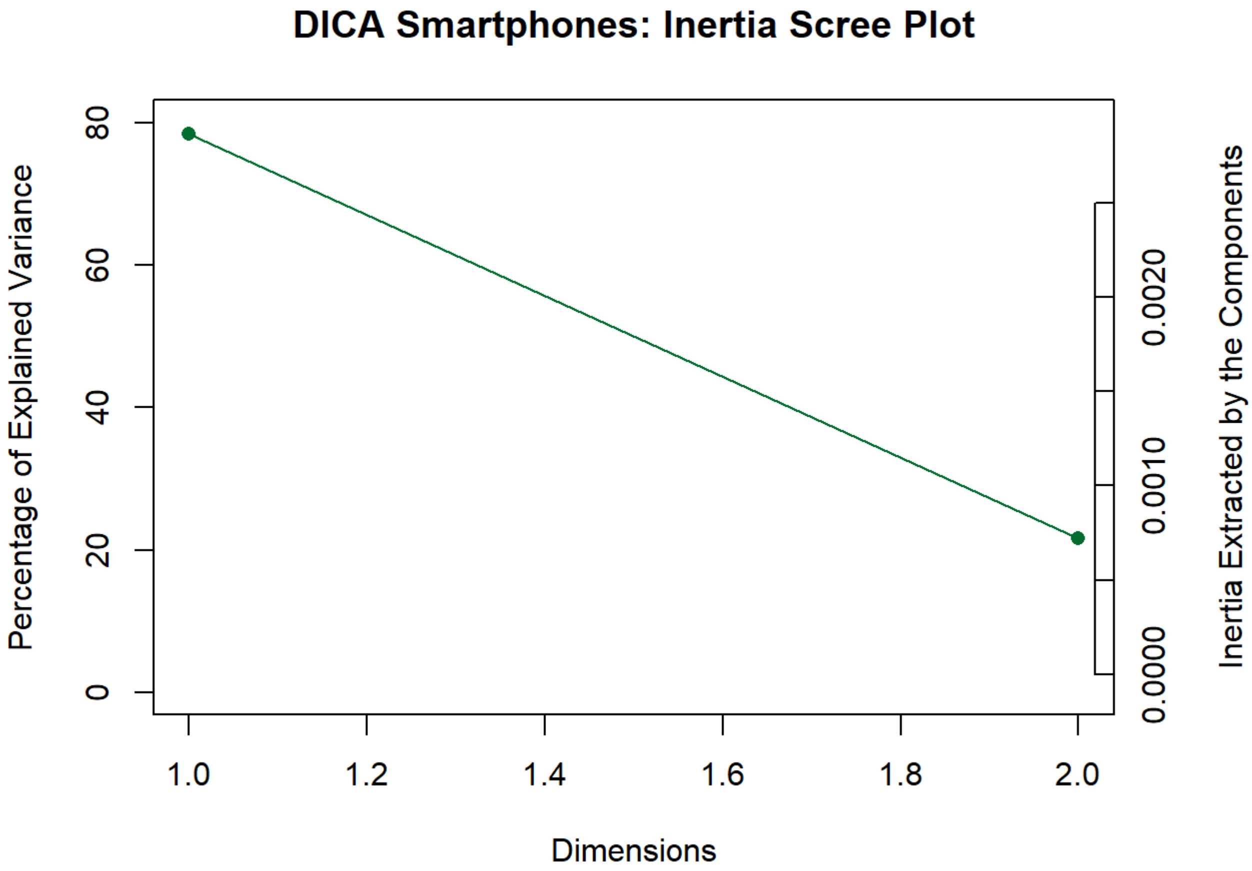
Component 1 shows max variance.

Component 2 shows max variance.

```
resDICA <- tepDICA(XYmat, DESIGN =  b$C[which(b$C!="AF")], make_design_nominal = TRUE,
                   graphs = FALSE)
# Inferences ----
set.seed(70301) # we have a problem with the inference part
  # it will be addressed soon. In the meantime we fix the seed
  # for random
nIter = 50
resDICA.inf <- tepDICA.inference.battery(XYmat,
                  DESIGN =  b$C[which(b$C!="AF")],
                  test.iters = nIter,
                  graphs = FALSE)
```

```
## [1] "It is estimated that your iterations will take 0.02 minutes."
## [1] "R is not in interactive() mode. Resample-based tests will be conducted. Please take
## ===========================================================================
```

```
PlotScree(ev = resDICA$TExPosition.Data$eigs,
          title = 'DICA Smartphones: Inertia Scree Plot',
          plotKaiser = FALSE,
          color4Kaiser = ggplot2::alpha('darkorchid4', .5),
          lwd4Kaiser  = 2)
```



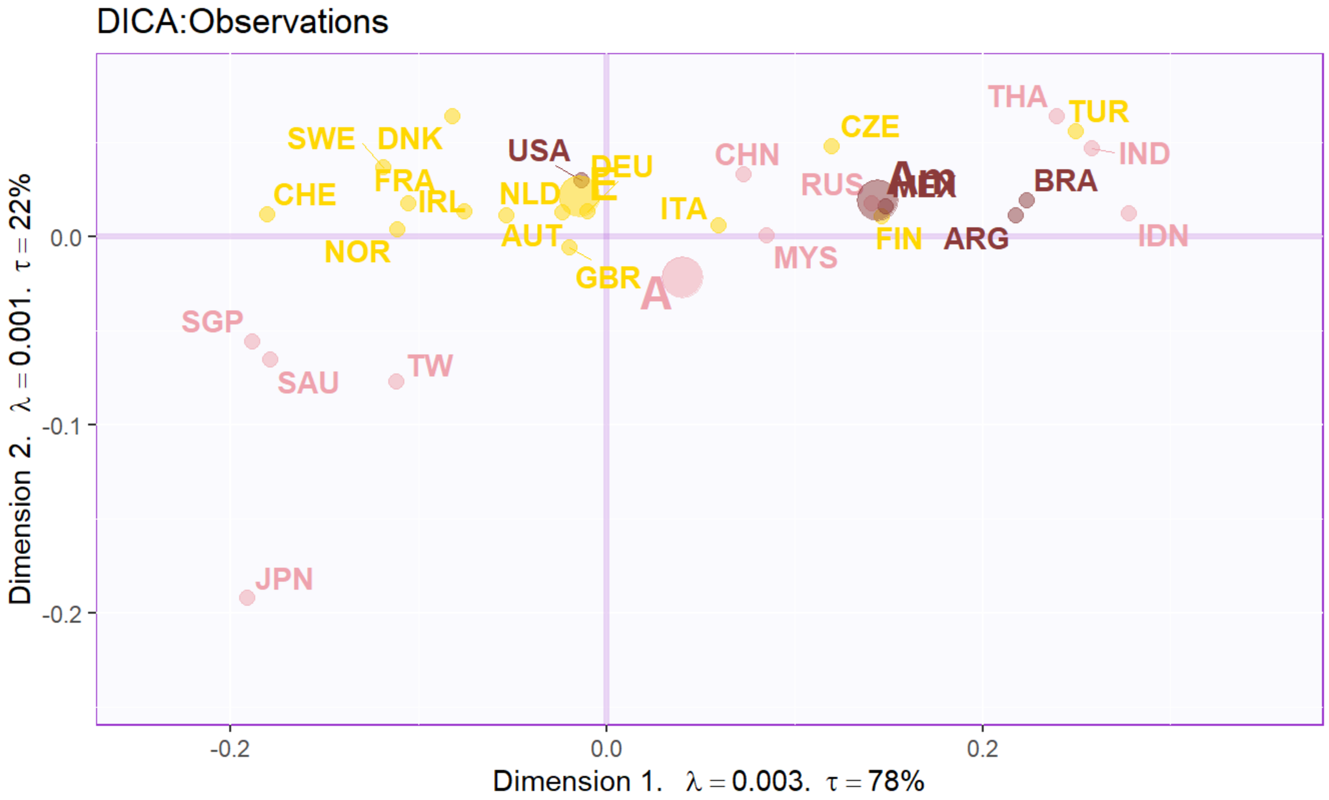
5.4 PROJECTIONS

Most of the countries are spread out across the first component.

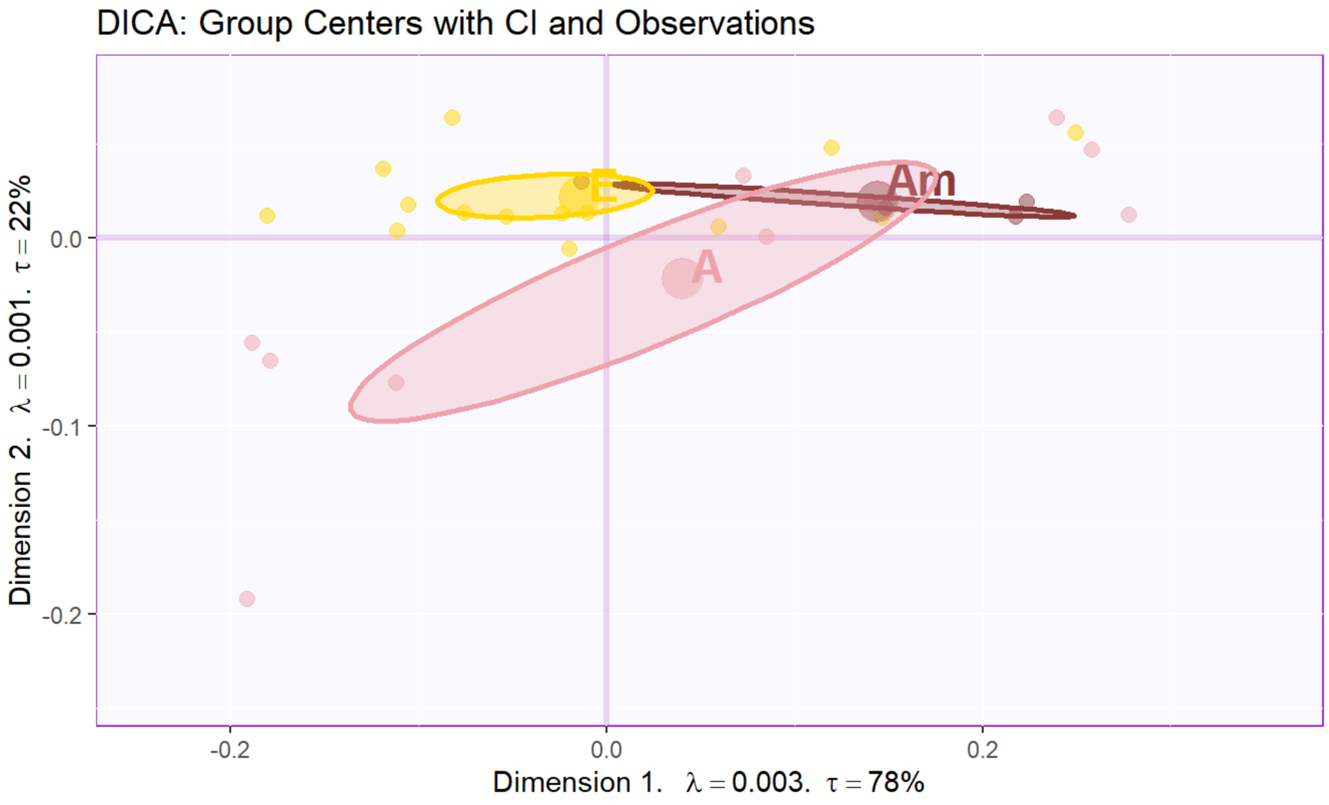
Lets group them to better interpret.

```r
wineColors <-  b$C[which(b$C!="AF")]
wineColors <- recode(wineColors, Am = 'indianred4',
                        E = 'gold', A = 'lightpink2')

Imap <- PTCA4CATA::createFactorMap(
  resDICA$TExPosition.Data$fii,
  col.points = wineColors,
  col.labels = wineColors,
  alpha.points = .5
)

label4Map <- createxyLabels.gen(1,2,
                                lambda = resDICA$TExPosition.Data$eigs,
                                tau = resDICA$TExPosition.Data$t)




wineMeans <- PTCA4CATA::getMeans(resDICA$TExPosition.Data$fii,
                                                  b$C[which(b$C!="AF")])

col4Means <- recode(rownames(wineMeans),Am = 'indianred4',
                        E = 'gold', A = 'lightpink2' )


names(col4Means) <- rownames(wineMeans)


MapGroup <- PTCA4CATA::createFactorMap(wineMeans,
                             # use the constraint from the main map
                             constraints = Imap$constraints,
                             col.points = col4Means,
                             cex = 7,  # size of the dot (bigger)
                             col.labels = col4Means,
                             text.cex = 6)
```

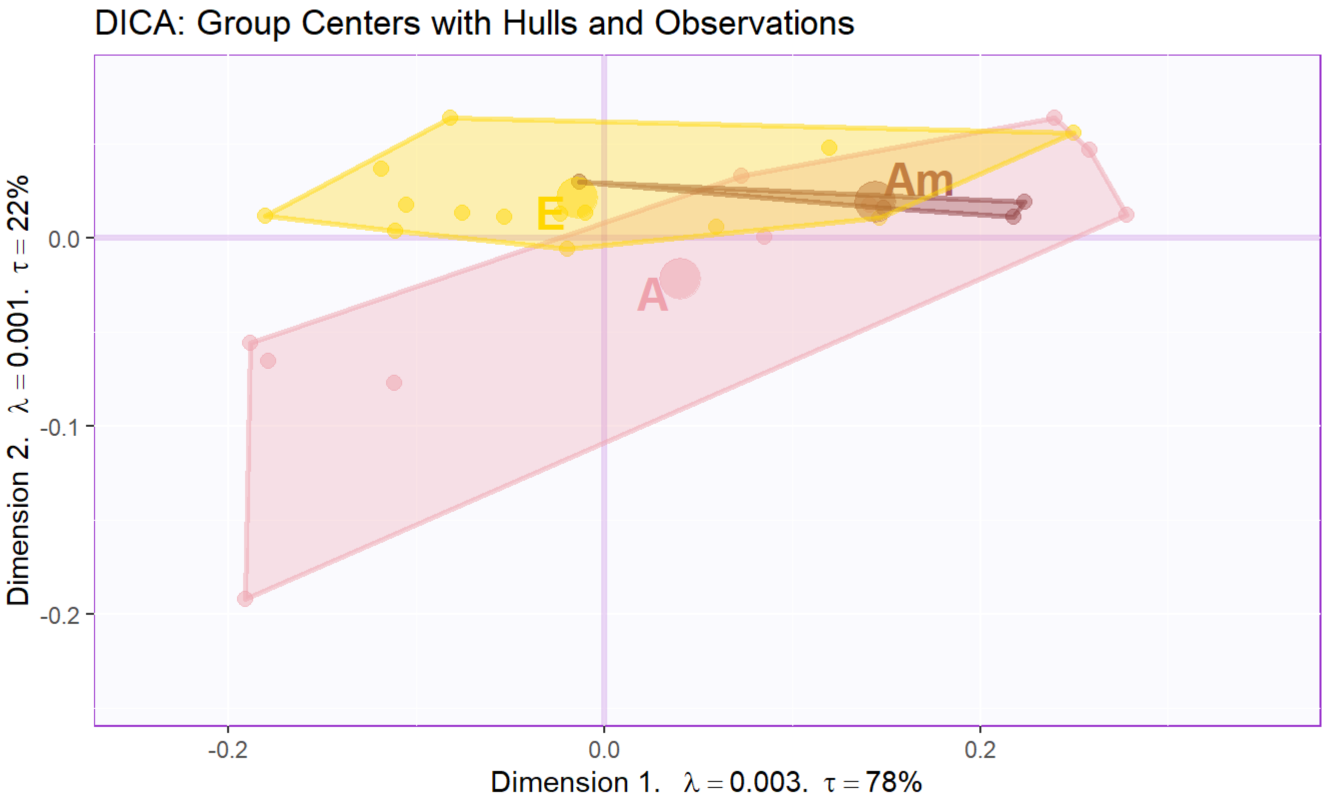
```
# The map with observations and group means
a003.bada <- Imap$zeMap + label4Map +
    MapGroup$zeMap_dots + MapGroup$zeMap_text+ ggtitle('DICA:Observations')
a003.bada
```



```
fi.boot <- resDICA.inf$Inference.Data$boot.data$fi.boot.data$boots
# We want to use the rownames of fi.boot as reference to get the correct
# color. However, the original rownames include "." and don't match with
# the original row names. So, the `sub` function was used to get rid of
# the "." by replacing all "." in the rownames of fi.boot as an empty
# string.
rownames(fi.boot) <- sub("[[:punct:]]","",rownames(fi.boot))
# use function MakeCIEllipses from package PTCA4CATA
GraphElli <- PTCA4CATA::MakeCIEllipses(
            resDICA.inf$Inference.Data$boot.data$fi.boot.data$boots,
                                        col = col4Means[rownames(fi.boot)], # use rownames c
                                        p.level = .95
)
#______________________________________________________________________
# create the I-map with Observations, means and confidence intervals
#
a004.bada.withCI <-  Imap$zeMap_background + Imap$zeMap_dots +
                     MapGroup$zeMap_dots + MapGroup$zeMap_text +
                     GraphElli + label4Map +
                     ggtitle('DICA: Group Centers with CI and Observations')
#______________________________________________________________________
                                                                                           #

a004.bada.withCI
```


DICA: Group Centers with CI and Observations
E
Am
A
Dimension 1. $\lambda = 0.003$. $\tau = 78\%$
Dimension 2. $\lambda = 0.001$. $\tau = 22\%$
-0.2
0.0
0.2
0.0
-0.1
-0.2

```
Fii <- resDICA$TExPosition.Data$fii
colnames(Fii) <- paste0('Dimension ', 1:ncol(Fii))
# getting the color correct: an ugly trick
col4Hull <- col4Means[match(names(col4Means),
                              levels( b$C[which(b$C!="AF")]) )]
GraphHull <- PTCA4CATA::MakeToleranceIntervals(Fii,
                                 design =  b$C[which(b$C!="AF")],
                                 col = col4Means,
                                 # the next line is required
                                 # for some strange unknown reasons
                                 names.of.factors =  c("Dim1","Dim2"),
                                 p.level = 1.00)
#
a006.bada.withHull <-  Imap$zeMap_background + Imap$zeMap_dots +
            MapGroup$zeMap_dots + MapGroup$zeMap_text +
            GraphHull + label4Map +
            ggtitle('DICA: Group Centers with Hulls and Observations')


col4X <- prettyGraphsColorSelection(n.colors = ncol(Xmat),
                                    starting.color = 42)

col4Y <- prettyGraphsColorSelection(n.colors = ncol(Ymat),
                                    starting.color = 13)
col4Var = c(col4X,col4Y)
#_____________________________________________________________________
a006.bada.withHull
```



5.5 LOADINGS

Use cosine interpretations between the variables

```r
# J-set ----
# gt colors
col4X <- prettyGraphsColorSelection(n.colors = ncol(Xmat),
                                    starting.color = 42)

col4Y <- prettyGraphsColorSelection(n.colors = ncol(Ymat),
                                    starting.color = 13)
col4Var = c(col4X,col4Y)
Fj <- resDICA$TExPosition.Data$fj
baseMap.j <- PTCA4CATA::createFactorMap(Fj,
                                        col.points   = col4Var,
                                        alpha.points =  .3,
                                        col.labels   = col4Var)
# arrows
zeArrows <- addArrows(Fj, color = col4Var)
# A graph for the J-set
# A graph for the J-set
b001.aggMap.j <- baseMap.j$zeMap_background + # background layer
  baseMap.j$zeMap_dots + baseMap.j$zeMap_text +  # dots & labels
  label4Map
b002.aggMap.j <- b001.aggMap.j + zeArrows
# We print this Map with the following code
dev.new()
print(b002.aggMap.j)
```

5.6 ANALYSIS/HYPOTHESIS

DICA: Group Centers with CI and Observations

European countries are away from american and asian countries

Few countries from american continent show similar usage of smart phones like asian and european as there developed and developing countries in america

DICA: Group Centers with Hulls and Observations

Here few european countries can be seen interracting with american and asian countries.

The countries which are have similar gdp are overlapping for example Turkey Thailand India , ITLAY Maylasia.

USA is with the european economies.

DICA: LOADINGS

If apps are used more on the phone email app will be used less and vice versa

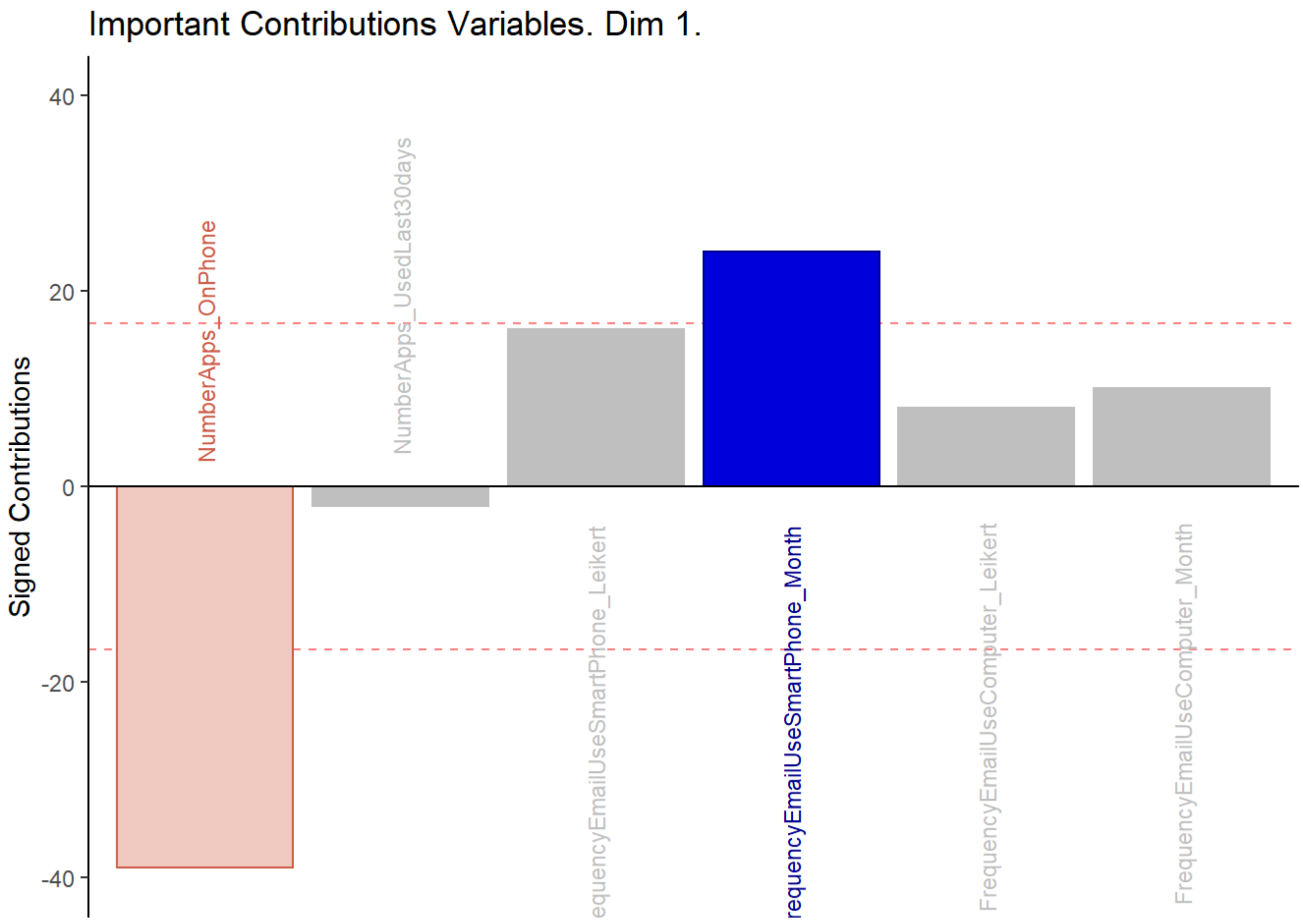
If more apps are there on the phone computer email usage would be less and vice versa.

5.7 CONTRIBUTIONS:

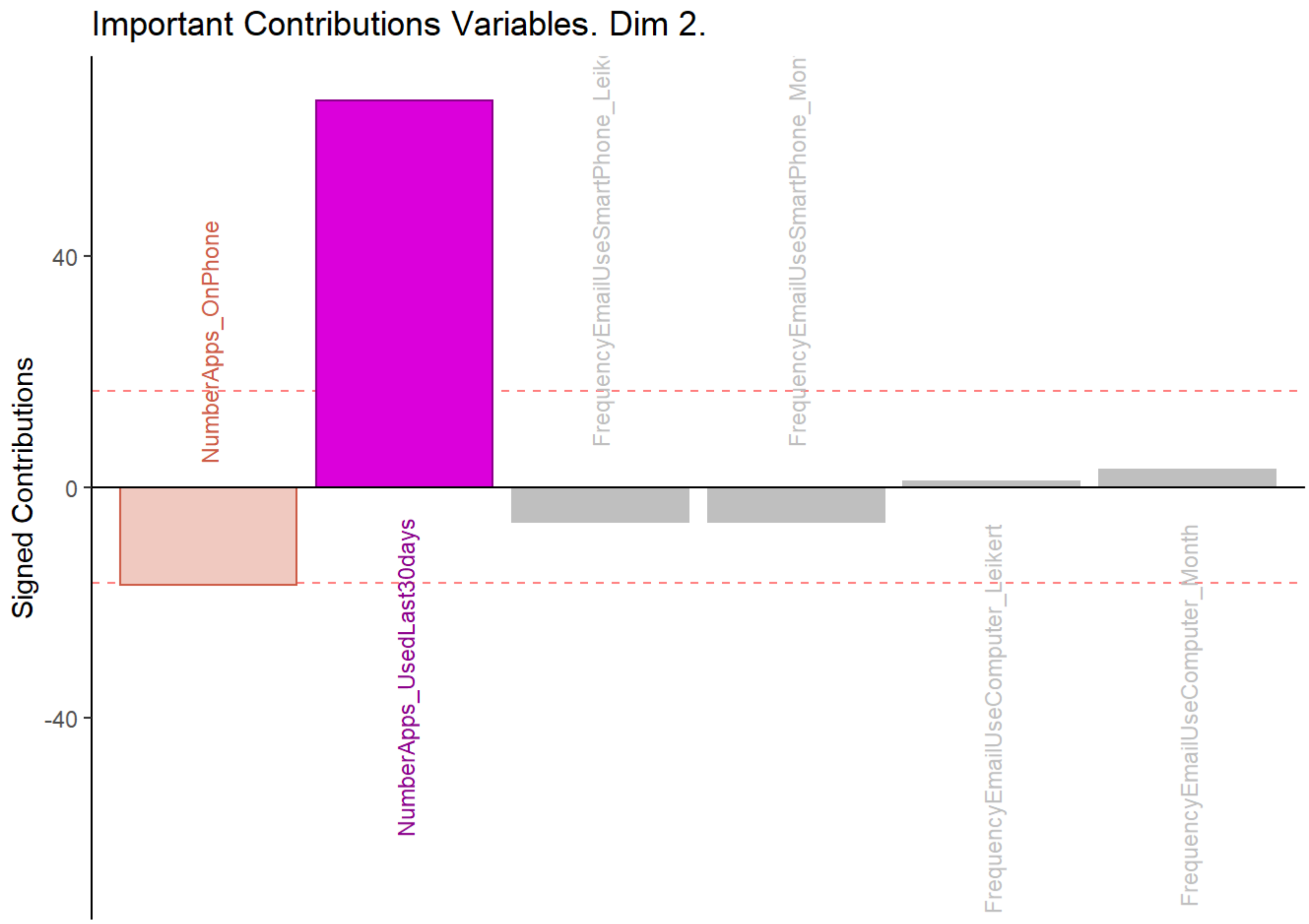
For component 1, 2 variables show significant contributions.

For component 2, 2 variables show significant contributions.

```r
#  Contributions ####
#_____________________________________________________________________
#_____________________________________________________________________
# Ctr J-set
###### 1 ====
#
ctrj <- resDICA$TExPosition.Data$cj
signed.ctrj <- ctrj * sign(Fj)
# BR1
c001.plotCtrj.1 <- PrettyBarPlot2(
                       bootratio = round(100*signed.ctrj[,1]),
                       threshold = 100 / nrow(signed.ctrj),
                       ylim = NULL,
                       color4bar = gplots::col2hex(col4Var),
                       color4ns = "gray75",
                       plotnames = TRUE,
                       main = 'Important Contributions Variables. Dim 1.',
                       ylab = "Signed Contributions")

###### 2 ====
#
c002.plotCtrj.2 <- PrettyBarPlot2(
  bootratio = round(100*signed.ctrj[,2]),
  threshold = 100 / nrow(signed.ctrj),
  ylim = NULL,
  color4bar = gplots::col2hex(col4Var),
  color4ns = "gray75",
  plotnames = TRUE,
  main = 'Important Contributions Variables. Dim 2.',
  ylab = "Signed Contributions")
c001.plotCtrj.1
```



```
c002.plotCtrj.2
```



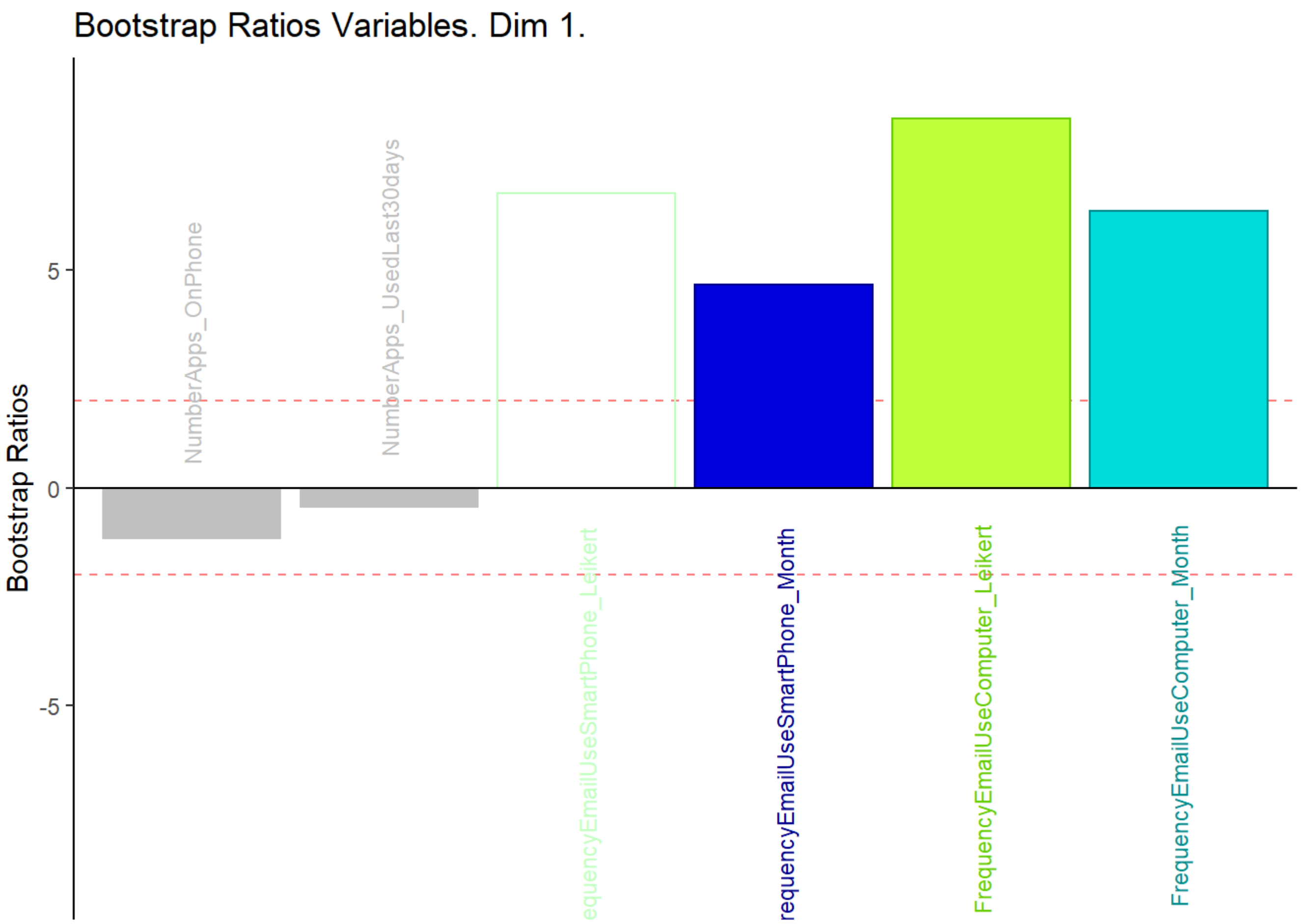
5.8 BOOTSTRAP

The information conveyed by the scree plot is confirmed by the boostrap contributions.

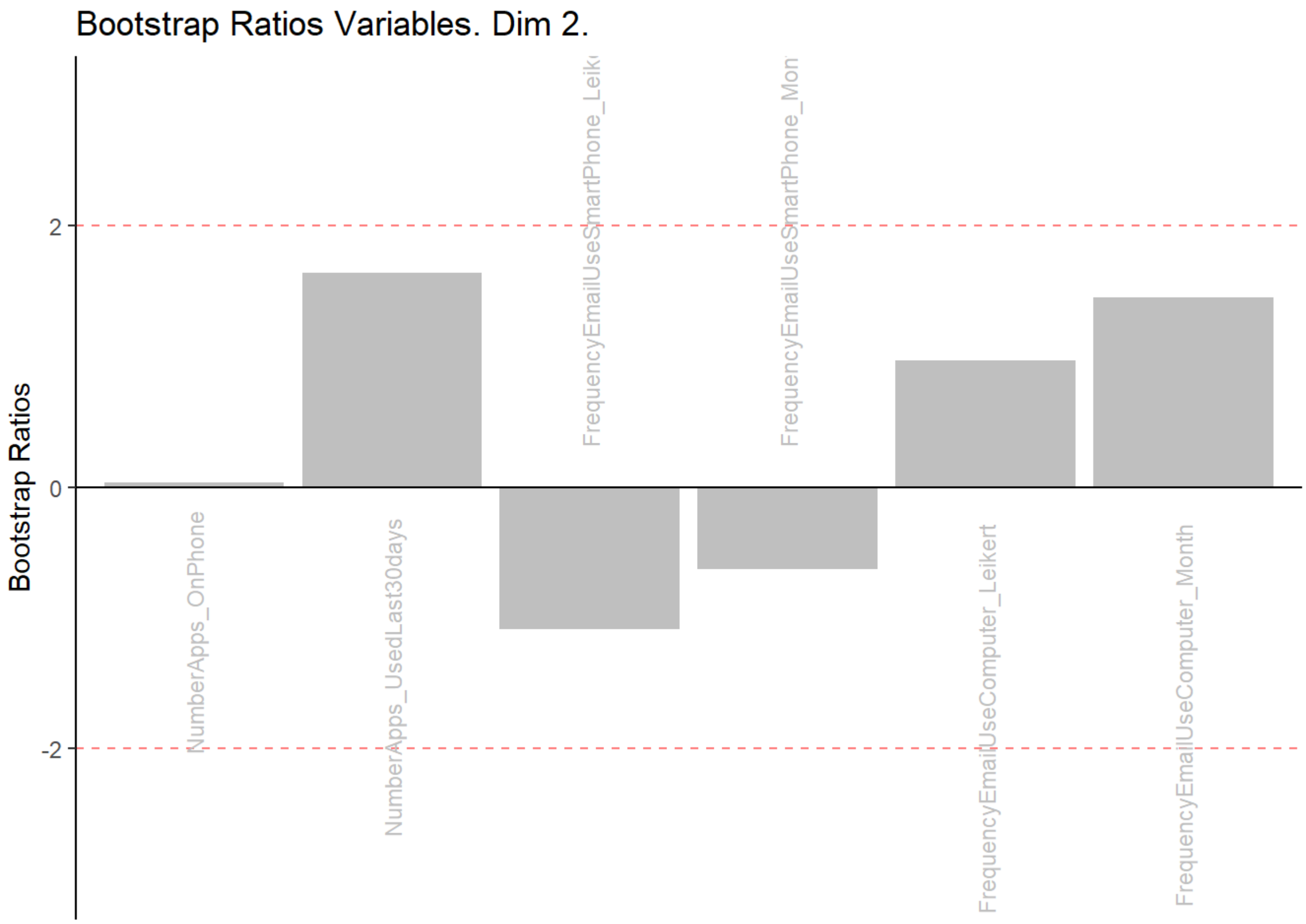
Variables have shown major contributions for component 1 than component 2.

We can thereby reject the null hypothesis

```
#BR. 1 ====
#
BRj <- resDICA.inf$Inference.Data$boot.data$fj.boot.data$tests$boot.ratios
# BR1
d001.plotBRj.1 <- PrettyBarPlot2(
  bootratio = BRj[,1],
  threshold = 2,
  ylim = NULL,
  color4bar = gplots::col2hex(col4Var),
  color4ns = "gray75",
  plotnames = TRUE,
  main = 'Bootstrap Ratios Variables. Dim 1.',
  ylab = "Bootstrap Ratios")

print(d001.plotBRj.1)
```



```
###### 2 ====
#
d003.plotBRj.2 <- PrettyBarPlot2(
  bootratio = BRj[,2],
  threshold = 2,
  ylim = NULL,
  color4bar = gplots::col2hex(col4Var),
  color4ns = "gray75",
  plotnames = TRUE,
  main = 'Bootstrap Ratios Variables. Dim 2.',
  ylab = "Bootstrap Ratios")

print(d003.plotBRj.2)
```


Bootstrap Ratios Variables. Dim 2.
Bootstrap Ratios
NumberApps_OnPhone
NumberApps_UsedLast30days
FrequencyEmailUseSmartPhone_Leik
FrequencyEmailUseSmartPhone_Mon
FrequencyEmailUseComputer_Leikert
FrequencyEmailUseComputer_Month