HeatMap

Nitish Malhotra

9 February 2018

# Load of data File

rm(list=ls())  
setwd('C:\\Users\\10382\\Downloads')  
df <- read.csv('data.csv',header = TRUE)  
head(df)

## Dividend.rate... Dividend.per.share Dividend.type BSE.Closing.price  
## 1 19 0.78 1 127.55  
## 2 12 0.40 1 146.40  
## 3 17 0.62 1 213.00  
## 4 26 1.52 1 223.90  
## 5 9 3.29 1 141.55  
## 6 11 3.68 1 201.35  
## Announcement\_year Record\_year BSE.Ex\_year Announcement\_month  
## 1 2017 2017 2017 5  
## 2 2016 2016 2016 5  
## 3 2015 2015 2015 5  
## 4 2014 2014 2014 5  
## 5 2013 2013 2013 5  
## 6 2012 2012 2012 5  
## Record\_month BSE.Ex\_month Announcement\_dec\_date Record\_dec\_date  
## 1 8 8 2017.405 2017.649  
## 2 9 9 2016.402 2016.708  
## 3 9 9 2015.397 2015.707  
## 4 9 9 2014.405 2014.690  
## 5 9 9 2013.389 2013.693  
## 6 9 9 2012.391 2012.691  
## BSE\_Ex\_dec\_date  
## 1 2017.641  
## 2 2016.702  
## 3 2015.701  
## 4 2014.685  
## 5 2013.679  
## 6 2012.680

# Inspecting Data

head(df) #first five row of data

## Dividend.rate... Dividend.per.share Dividend.type BSE.Closing.price  
## 1 19 0.78 1 127.55  
## 2 12 0.40 1 146.40  
## 3 17 0.62 1 213.00  
## 4 26 1.52 1 223.90  
## 5 9 3.29 1 141.55  
## 6 11 3.68 1 201.35  
## Announcement\_year Record\_year BSE.Ex\_year Announcement\_month  
## 1 2017 2017 2017 5  
## 2 2016 2016 2016 5  
## 3 2015 2015 2015 5  
## 4 2014 2014 2014 5  
## 5 2013 2013 2013 5  
## 6 2012 2012 2012 5  
## Record\_month BSE.Ex\_month Announcement\_dec\_date Record\_dec\_date  
## 1 8 8 2017.405 2017.649  
## 2 9 9 2016.402 2016.708  
## 3 9 9 2015.397 2015.707  
## 4 9 9 2014.405 2014.690  
## 5 9 9 2013.389 2013.693  
## 6 9 9 2012.391 2012.691  
## BSE\_Ex\_dec\_date  
## 1 2017.641  
## 2 2016.702  
## 3 2015.701  
## 4 2014.685  
## 5 2013.679  
## 6 2012.680

dim(data) #Dimension of data

## NULL

names(df) #Column names of data

## [1] "Dividend.rate..." "Dividend.per.share"   
## [3] "Dividend.type" "BSE.Closing.price"   
## [5] "Announcement\_year" "Record\_year"   
## [7] "BSE.Ex\_year" "Announcement\_month"   
## [9] "Record\_month" "BSE.Ex\_month"   
## [11] "Announcement\_dec\_date" "Record\_dec\_date"   
## [13] "BSE\_Ex\_dec\_date"

str(df) #Structure of data

## 'data.frame': 43 obs. of 13 variables:  
## $ Dividend.rate... : int 19 12 17 26 9 11 10 4 27 24 ...  
## $ Dividend.per.share : num 0.78 0.4 0.62 1.52 3.29 3.68 17.9 12.3 8 6.25 ...  
## $ Dividend.type : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ BSE.Closing.price : num 128 146 213 224 142 ...  
## $ Announcement\_year : int 2017 2016 2015 2014 2013 2012 2011 2010 2009 2008 ...  
## $ Record\_year : int 2017 2016 2015 2014 2013 2012 2011 2010 2009 2008 ...  
## $ BSE.Ex\_year : int 2017 2016 2015 2014 2013 2012 2011 2010 2009 2008 ...  
## $ Announcement\_month : int 5 5 5 5 5 5 5 5 5 5 ...  
## $ Record\_month : int 8 9 9 9 9 9 8 9 9 9 ...  
## $ BSE.Ex\_month : int 8 9 9 9 9 9 8 9 9 9 ...  
## $ Announcement\_dec\_date: num 2017 2016 2015 2014 2013 ...  
## $ Record\_dec\_date : num 2018 2017 2016 2015 2014 ...  
## $ BSE\_Ex\_dec\_date : num 2018 2017 2016 2015 2014 ...

# Plotting of missing data

#library(VIM)  
#aggr\_plot <- aggr(data, col=c('navyblue','red'), numbers=TRUE, sortVars=TRUE, labels=names(data), cex.axis=.7, gap=3, ylab=c("Histogram of missing data","Pattern"))

# Imputing Missing data

# Filled data with median   
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.3.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

df <- df %>% mutate\_all(~ifelse(is.na(.), median(., na.rm = TRUE), .))

## Warning: package 'bindrcpp' was built under R version 3.3.3

# converting factor into numeric

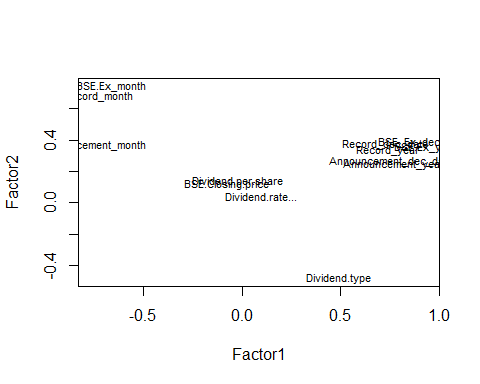
df$Dividend.type <- as.numeric(df$Dividend.type)

# Factor Analysis

mydata <- df  
n.factors <- 2   
fit <- factanal(mydata,   
 n.factors, # number of factors to extract  
 scores=c("regression"),  
 rotation="none")  
  
print(fit, digits=2, cutoff=.3, sort=TRUE)

##   
## Call:  
## factanal(x = mydata, factors = n.factors, scores = c("regression"), rotation = "none")  
##   
## Uniquenesses:  
## Dividend.rate... Dividend.per.share Dividend.type   
## 0.99 0.98 0.52   
## BSE.Closing.price Announcement\_year Record\_year   
## 0.98 0.34 0.34   
## BSE.Ex\_year Announcement\_month Record\_month   
## 0.00 0.28 0.00   
## BSE.Ex\_month Announcement\_dec\_date Record\_dec\_date   
## 0.01 0.36 0.33   
## BSE\_Ex\_dec\_date   
## 0.00   
##   
## Loadings:  
## Factor1 Factor2  
## Announcement\_year 0.77   
## Record\_year 0.74 0.34   
## BSE.Ex\_year 0.93 0.36   
## Announcement\_month -0.76 0.37   
## Record\_month -0.73 0.68   
## Announcement\_dec\_date 0.76   
## Record\_dec\_date 0.73 0.38   
## BSE\_Ex\_dec\_date 0.92 0.39   
## BSE.Ex\_month -0.66 0.74   
## Dividend.rate...   
## Dividend.per.share   
## Dividend.type 0.49 -0.48   
## BSE.Closing.price   
##   
## Factor1 Factor2  
## SS loadings 5.76 2.09  
## Proportion Var 0.44 0.16  
## Cumulative Var 0.44 0.60  
##   
## Test of the hypothesis that 2 factors are sufficient.  
## The chi square statistic is 1093.45 on 53 degrees of freedom.  
## The p-value is 3.16e-194

# plot factor 1 by factor 2   
load <- fit$loadings[,1:2]   
plot(load,type="n") # set up plot   
text(load,labels=names(mydata),cex=.7) # add variable names



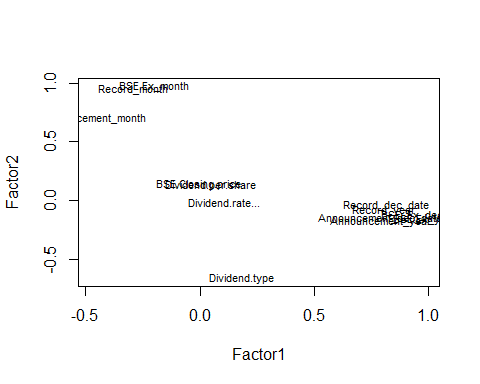
# Factor Analysis using Varimax

fit <- factanal(mydata,   
 n.factors, # number of factors to extract  
 rotation="varimax") # 'varimax' is an ortho rotation  
  
load <- fit$loadings[,1:2]   
load

## Factor1 Factor2  
## Dividend.rate... 0.107295969 -0.01208121  
## Dividend.per.share 0.052006307 0.13267346  
## Dividend.type 0.184814149 -0.66420828  
## BSE.Closing.price -0.002935692 0.14068762  
## Announcement\_year 0.794581258 -0.17075044  
## Record\_year 0.808736233 -0.07652917  
## BSE.Ex\_year 0.986538060 -0.15571206  
## Announcement\_month -0.473213170 0.70136262  
## Record\_month -0.290631378 0.95426147  
## BSE.Ex\_month -0.200087239 0.97358797  
## Announcement\_dec\_date 0.789101582 -0.14976341  
## Record\_dec\_date 0.817686838 -0.03503179  
## BSE\_Ex\_dec\_date 0.991386591 -0.12046888

# Plot of varimax

plot(load,type="n") # set up plot   
text(load,labels=names(mydata),cex=.7) # add variable names



library(psych)

## Warning: package 'psych' was built under R version 3.3.3

solution <- fa(r = cor(mydata), nfactors = 2, rotate = "oblimin", fm = "pa")

## Loading required namespace: GPArotation

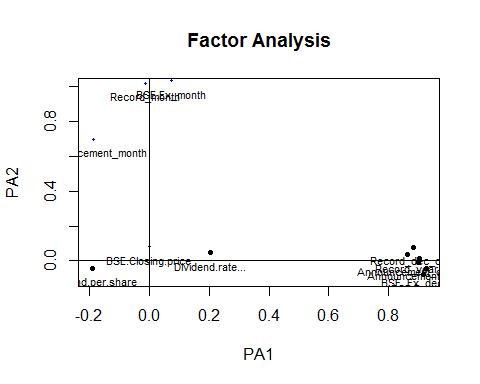
## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate =  
## rotate, : A loading greater than abs(1) was detected. Examine the loadings  
## carefully.

## Warning in cor.smooth(model): Matrix was not positive definite, smoothing  
## was done

## The estimated weights for the factor scores are probably incorrect. Try a different factor extraction method.

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate =  
## rotate, : An ultra-Heywood case was detected. Examine the results carefully

plot(solution,labels=names(mydata),cex=.7, ylim=c(-.1,1))



solution

## Factor Analysis using method = pa  
## Call: fa(r = cor(mydata), nfactors = 2, rotate = "oblimin", fm = "pa")  
##   
## Warning: A Heywood case was detected.   
## Standardized loadings (pattern matrix) based upon correlation matrix  
## PA1 PA2 h2 u2 com  
## Dividend.rate... 0.20 0.05 0.0363 0.964 1.1  
## Dividend.per.share -0.19 -0.04 0.0324 0.968 1.1  
## Dividend.type 0.08 -0.58 0.3868 0.613 1.0  
## BSE.Closing.price 0.00 0.08 0.0065 0.994 1.0  
## Announcement\_year 0.90 -0.01 0.8134 0.187 1.0  
## Record\_year 0.86 0.04 0.7196 0.280 1.0  
## BSE.Ex\_year 0.91 -0.08 0.8994 0.101 1.0  
## Announcement\_month -0.19 0.70 0.6249 0.375 1.1  
## Record\_month -0.01 1.02 1.0504 -0.050 1.0  
## BSE.Ex\_month 0.07 1.03 1.0158 -0.016 1.0  
## Announcement\_dec\_date 0.90 0.02 0.8019 0.198 1.0  
## Record\_dec\_date 0.88 0.08 0.7313 0.269 1.0  
## BSE\_Ex\_dec\_date 0.93 -0.04 0.8893 0.111 1.0  
##   
## PA1 PA2  
## SS loadings 5.00 3.00  
## Proportion Var 0.38 0.23  
## Cumulative Var 0.38 0.62  
## Proportion Explained 0.62 0.38  
## Cumulative Proportion 0.62 1.00  
##   
## With factor correlations of   
## PA1 PA2  
## PA1 1.00 -0.39  
## PA2 -0.39 1.00  
##   
## Mean item complexity = 1  
## Test of the hypothesis that 2 factors are sufficient.  
##   
## The degrees of freedom for the null model are 78 and the objective function was 46.87  
## The degrees of freedom for the model are 53 and the objective function was 83607921   
##   
## The root mean square of the residuals (RMSR) is 0.13   
## The df corrected root mean square of the residuals is 0.16   
##   
## Fit based upon off diagonal values = 0.92

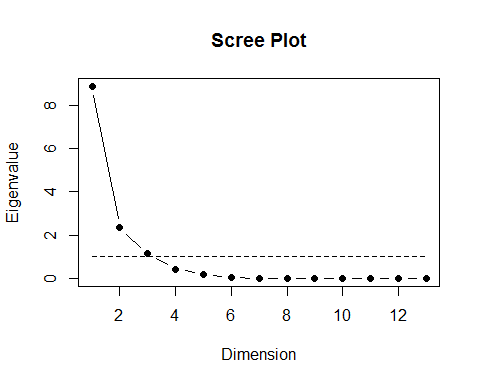
#Determining the Number of Factors to Extract  
  
# install.packages("psy")  
library(psy)

## Warning: package 'psy' was built under R version 3.3.2

##   
## Attaching package: 'psy'

## The following object is masked from 'package:psych':  
##   
## wkappa

scree.plot(fit$correlation)



# Determine Number of Factors to Extract  
# install.packages("nFactors")  
library(nFactors)

## Warning: package 'nFactors' was built under R version 3.3.3

## Loading required package: MASS

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

## Loading required package: boot

##   
## Attaching package: 'boot'

## The following object is masked from 'package:psych':  
##   
## logit

## Loading required package: lattice

## Warning: package 'lattice' was built under R version 3.3.3

##   
## Attaching package: 'lattice'

## The following object is masked from 'package:boot':  
##   
## melanoma

##   
## Attaching package: 'nFactors'

## The following object is masked from 'package:lattice':  
##   
## parallel

ev <- eigen(cor(mydata)) # get eigenvalues  
ap <- parallel(subject=nrow(mydata),var=ncol(mydata), rep=100, cent=.05)  
nS <- nScree(x=ev$values, aparallel=ap$eigen$qevpea)  
plotnScree(nS)

