Assignment7: Factor Analysis

Sri Seshadri

8/5/2017

Stotzel's liquor preference paper

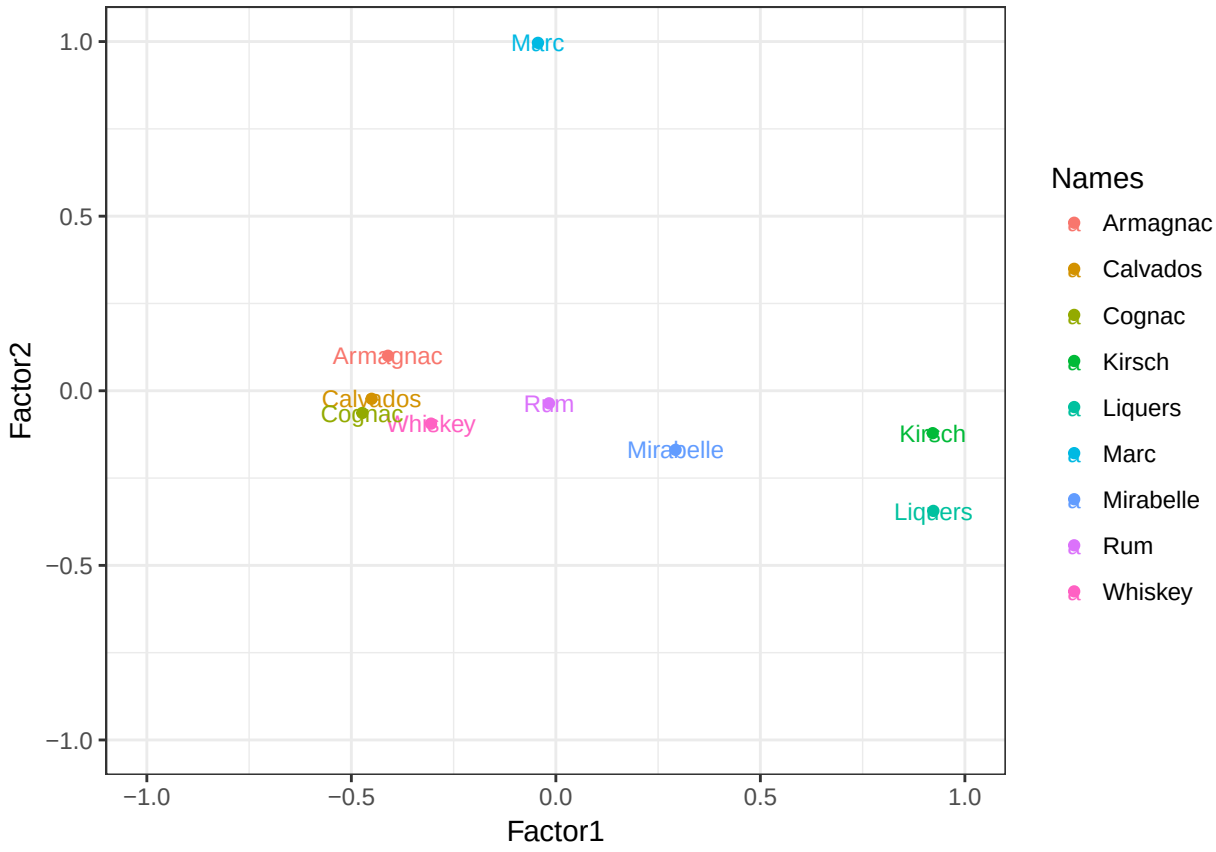
There are 1442 survey results on liquor preference

```
cor.values <- c(1, 0.21, 0.37, -0.32, 0, -0.31, -0.26, 0.09,
    -0.38, 0.21, 1, 0.09, -0.29, 0.12, -0.3, -0.14, 0.01, -0.39,
    0.37, 0.09, 1, -0.31, -0.04, -0.3, -0.11, 0.12, -0.39, -0.32,
    -0.29, -0.31, 1, -0.16, 0.25, -0.13, -0.14, 0.9, 0, 0.12,
    -0.04, -0.16, 1, -0.2, -0.03, -0.08, -0.38, -0.31, -0.3,
    -0.3, 0.25, -0.2, 1, -0.24, -0.16, 0.18, -0.26, -0.14, -0.11,
    -0.13, -0.03, -0.24, 1, -0.2, 0.04, 0.09, 0.01, 0.12, -0.14,
    -0.08, -0.16, -0.2, 1, -0.24, -0.38, -0.39, -0.39, 0.9, -0.38,
    0.18, 0.04, -0.24, 1)

# Get correlation values into a matrix
cor.matrix <- matrix(cor.values, nrow = 9, ncol = 9, byrow = T)
colnames(cor.matrix) <- c("Calvados", "Armagnac", "Cognac", "Kirsch",
    "Marc", "Mirabelle", "Rum", "Whiskey", "Liquers")
rownames(cor.matrix) <- c("Calvados", "Armagnac", "Cognac", "Kirsch",
    "Marc", "Mirabelle", "Rum", "Whiskey", "Liquers")
cor.matrix
```

```
##           Calvados Armagnac Cognac Kirsch  Marc Mirabelle   Rum Whiskey
## Calvados      1.00     0.21   0.37  -0.32  0.00     -0.31 -0.26    0.09
## Armagnac      0.21     1.00   0.09  -0.29  0.12     -0.30 -0.14    0.01
## Cognac        0.37     0.09   1.00  -0.31 -0.04     -0.30 -0.11    0.12
## Kirsch       -0.32    -0.29  -0.31   1.00 -0.16      0.25 -0.13   -0.14
## Marc          0.00     0.12  -0.04  -0.16  1.00     -0.20 -0.03   -0.08
## Mirabelle    -0.31    -0.30  -0.30   0.25 -0.20      1.00 -0.24   -0.16
## Rum          -0.26    -0.14  -0.11  -0.13 -0.03     -0.24  1.00   -0.20
## Whiskey       0.09     0.01   0.12  -0.14 -0.08     -0.16 -0.20    1.00
## Liquers      -0.38    -0.39  -0.39   0.90 -0.38      0.18  0.04   -0.24
##           Liquers
## Calvados    -0.38
## Armagnac    -0.39
## Cognac      -0.39
## Kirsch       0.90
## Marc        -0.38
## Mirabelle    0.18
## Rum          0.04
## Whiskey     -0.24
## Liquers      1.00
```

```
f.1 <- factanal(covmat = cor.matrix, n.obs = 1442, factors = 3,
    rotation = "varimax")
loadingsasdf <- function(x) {
    if (class(x) != "factanal") {
        stop("input must be of the class factanal")
```

```
    }
    i <- 1:x$factor
    df <- as.data.frame(purrr::map(.x = i, .f = function(y) (x$loadings[,
        y])))
    colnames(df) <- sapply(i, FUN = function(x) {
        paste0("Factor", x)
    })
    df$Names <- row.names(df)
    return(df[, c(x$factor + 1, 1:x$factor)])

}

calculateMAE <- function(x, y = cor.matrix) {
    loading <- loadingsasdf(x)
    CorrMatrixHat <- as.matrix(loading[, -1]) %*% t(as.matrix(loading[,
        -1])) + diag(x$uniquenesses)
    return(mean(abs(y - CorrMatrixHat)))

}

df <- loadingsasdf(f.1) %>% dplyr::arrange(desc(Factor1))

library(ggplot2)
ggplot(data = df, mapping = aes(x = Factor1, y = Factor2, label = Names,
    color = Names)) + ggplot2::geom_point() + geom_text(size = 3) +
    theme_bw() + xlim(c(-1, 1)) + ylim(c(-1, 1))
```



```r
# create arguments with number of factors and rotation and
# put them in a tibble (data frame)

FA.Arguments <- dplyr::tibble(factors = rep(1:5, 2), rotation = rep(c("varimax",
    "promax"), 1, each = 5))
# simultaneously run all factanal for all the arguments
FA <- purrr::map2(.x = FA.Arguments$factors, .y = FA.Arguments$rotation,
    .f = function(x, y) factanal(covmat = cor.matrix, n.obs = 1442,
        factors = x, rotation = y))
FactorAdequacy <- ifelse(sapply(X = 1:nrow(FA.Arguments), FUN = function(x) FA[[x]]$PVAL) >=
    0.05, "yes", "no")

MAEs <- sapply(FA, calculateMAE)
FactorAdequacy <- cbind(FA.Arguments, FactorAdequacy, MAE = MAEs)
knitr::kable(FactorAdequacy, caption = "Factor Adequacy & MAE")
```

Table 1: Factor Adequacy & MAE

factors	rotation	FactorAdequacy	MAE
1	varimax	no	0.0861562
2	varimax	no	0.0704521
3	varimax	no	0.0485786
4	varimax	no	0.0325464
5	varimax	no	0.0200446
1	promax	no	0.0861562
2	promax	no	0.0815585
3	promax	no	0.1102698
4	promax	no	0.1472274
5	promax	no	0.1238722

Closing points:

1. The loadings are not hte same as mentioned in the paper ... the loadings do not have a unique solution
2. Got prety much the same interpretation as Stotzel, but the factors selections based on MLE wasn't sufficient.
3. Null hypothesis is, the K number of chosen parameters is sufficient for explaining the variation
4. Oblique roration (promax) has made interpretation easy by making the loadings disappear into one column. i.e. made the loadings bigger in one row of a column than spread them out.
5. Rotation does not affect the statistical inference for the number of factors

Note: Grouping can be seen using factor scores computation, but due to lack of raw data factor scores were not computed.