R practical – Factor Analysis

1. Grades

   The file `grades.txt` contains the exam results for 120 students on 5 examinations: two in mathematics, two in literature and one "comprehensive" exam.

   Download the text file and save it in your R working directory. You can read in the data set using `grades <- read.table("grades.txt", header=T, sep="")`.

   Use a factor model with one factor `fa_grades1 <- factanal(grades, 1)` to analyse the data and comment on the result. Answer

   Now use a two factor model `fa_grades2 <- factanal(grades, 2)` and comment on the result. Answer

   Interpret the factors of your chosen model. Answer

   Produce a biplot using (note that we are rerunning the factor analysis to ask R to calculate the scores. We are using the "regression" method here, check ?factanal to see more if you are interested and/or read `https://pareonline.net/getvn.asp?v=14&n=20` for more details on scores.)

```
fa_grades2 <- factanal(grades, 2, scores="regression")
biplot(fa_grades2$scores, loadings(fa_grades2), cex=0.5, col=c("black","red"),
 xlim = c(-3, 3))
```

   What can you see? Answer

   Package psych provides different tests for selecting the number of factors. As mentioned in the documentation of the package "Determining the most interpretable number of factors from a factor analysis is perhaps one of the greatest challenges in factor analysis. There are many solutions to this problem, none of which is uniformly the best."

   One approach that has been suggested is the so-called parallel approach, where factors are extracted until the eigenvalues of the real data are less than the corresponding eigenvalues of a random data set of the same size. The output looks at both the number of components in a PCA and the number of factors in FA. Run the following code a few times and each time make a note of the suggested number of factors. What do you see and why?

```
fa.parallel(grades, fm = "mle")
```

   Answer

2. Life expectancy

The data in `life.txt` show life expectancy in years by country, age and gender and related to life expectancies in the 1960s (Keyfitz and Flieger, 1971).

The columns are

1 male expectancy at age 0
2 male expectancy at age 25
3 male expectancy at age 50
4 male expectancy at age 75
5 female expectancy at age 0
6 female expectancy at age 25
7 female expectancy at age 50
8 female expectancy at age 75

Download the text file and save it in your R working directory. You can read in the data set using `life <- read.table("life.txt", header=F)`.

Use a χ^2 test to select the number of factors and comment on the result. Answer

Plot the scores and comment on the result

```
scores <- factanal(life[,-1], factors = 3, method = "mle",
      scores = "regression")$scores

plot(scores[,1], scores[,2], type = "n", xlab = "Factor 1",
   ylab = "Factor 2")
text(scores[,1], scores[,2], abbreviate(rownames(life), 5), cex = 0.7)

plot(scores[,1], scores[,3], type = "n", xlab = "Factor 1",
    ylab = "Factor 3")
text(scores[,1], scores[,3], abbreviate(rownames(life), 5), cex = 0.7)

plot(scores[,2], scores[,3], type = "n", xlab = "Factor 2",
  ylab = "Factor 3")
text(scores[,2], scores[,3], abbreviate(rownames(life), 5),
  cex = 0.7)
```

Answer