Naissances

Axel-Cleris Gailloty

2023-03-09

# Introduction

L’ analyse factorielle des correspondances (AFC ou CA pour correspondence analysis en anglais) est une extension de l’analyse en composantes principales (chapitre @ref(analyse-en-composantes-principales)) pour analyser l’association entre deux variables qualitatives (ou catégorielles).

L’AFC permet de résumer et de visualiser l’information contenue dans le tableau de contingence formé par les deux variables catégorielles. Le tableau de contingence contient les fréquences formées par les deux variables.

L’AFC retourne les coordonnées des éléments des colonnes et des lignes du tableau de contingence. Ces coordonnées permettent de visualiser graphiquement l’association entre les éléments de lignes et de colonnes dans un graphique à deux dimensions.

L’analyse factorielle par correspondance est une approche géométrique pour visualiser les lignes et les colonnes d’une table de contingence dans un graphique en nuage de points, de sorte que les positions des points lignes et celles des points colonnes correspondent à leurs associations dans le tableau.

## Cas pratique sur le jeu de données naissances

naissances <- readRDS("naissances.RDS")

library(FactoMineR)  
library(Factoshiny)

## Le chargement a nécessité le package : shiny

## Le chargement a nécessité le package : FactoInvestigate

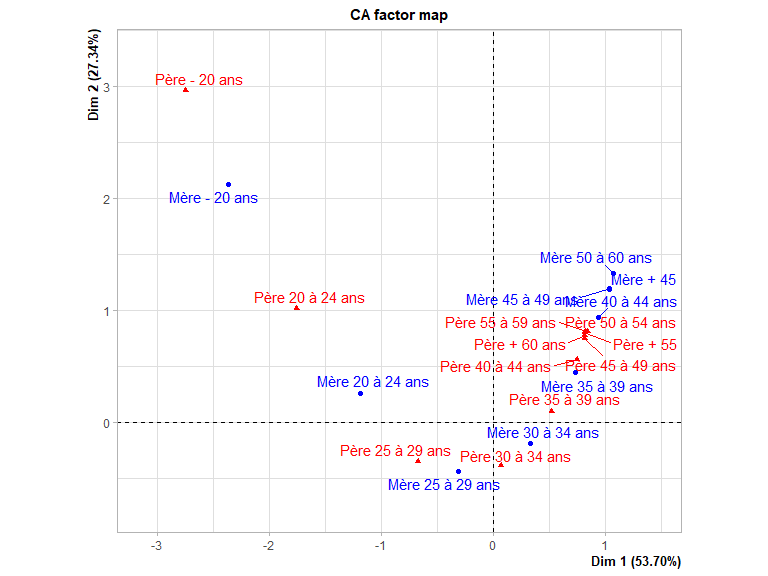
## Le chargement a nécessité le package : ggplot2

#Factoshiny(naissances)

res <- CA(naissances, graph = FALSE)  
summary(res)

##   
## Call:  
## CA(X = naissances, graph = FALSE)   
##   
## The chi square of independence between the two variables is equal to 731895.7 (p-value = 0 ).  
##   
## Eigenvalues  
## Dim.1 Dim.2 Dim.3 Dim.4 Dim.5 Dim.6 Dim.7  
## Variance 0.480 0.244 0.109 0.037 0.016 0.006 0.001  
## % of var. 53.703 27.343 12.237 4.138 1.774 0.720 0.085  
## Cumulative % of var. 53.703 81.046 93.284 97.421 99.195 99.915 100.000  
## Dim.8  
## Variance 0.000  
## % of var. 0.000  
## Cumulative % of var. 100.000  
##   
## Rows  
## Iner\*1000 Dim.1 ctr cos2 Dim.2 ctr  
## Mère - 20 ans | 209.664 | -2.367 20.712 0.474 | 2.126 32.816  
## Mère 20 à 24 ans | 199.746 | -1.195 36.337 0.873 | 0.259 3.347  
## Mère 25 à 29 ans | 105.556 | -0.319 6.410 0.291 | -0.441 24.031  
## Mère 30 à 34 ans | 85.916 | 0.324 7.297 0.408 | -0.193 5.089  
## Mère 35 à 39 ans | 137.963 | 0.733 19.026 0.662 | 0.444 13.724  
## Mère 40 à 44 ans | 112.386 | 0.934 8.723 0.372 | 0.932 17.054  
## Mère 45 à 49 ans | 18.442 | 1.029 0.701 0.183 | 1.186 1.830  
## Mère 50 à 60 ans | 3.334 | 1.068 0.046 0.066 | 1.327 0.140  
## Mère + 45 | 20.504 | 1.031 0.748 0.175 | 1.194 1.969  
## cos2 Dim.3 ctr cos2   
## Mère - 20 ans 0.382 | -1.169 22.182 0.116 |  
## Mère 20 à 24 ans 0.041 | 0.244 6.671 0.037 |  
## Mère 25 à 29 ans 0.556 | 0.188 9.737 0.101 |  
## Mère 30 à 34 ans 0.145 | -0.322 31.637 0.403 |  
## Mère 35 à 39 ans 0.243 | 0.002 0.000 0.000 |  
## Mère 40 à 44 ans 0.371 | 0.689 20.865 0.203 |  
## Mère 45 à 49 ans 0.242 | 1.181 4.052 0.240 |  
## Mère 50 à 60 ans 0.103 | 1.524 0.412 0.135 |  
## Mère + 45 0.235 | 1.200 4.444 0.237 |  
##   
## Columns (the 10 first)  
## Iner\*1000 Dim.1 ctr cos2 Dim.2 ctr  
## Père - 20 ans | 93.762 | -2.751 6.054 0.310 | 2.960 13.768  
## Père 20 à 24 ans | 257.451 | -1.761 39.207 0.731 | 1.016 25.623  
## Père 25 à 29 ans | 163.523 | -0.676 21.059 0.618 | -0.355 11.384  
## Père 30 à 34 ans | 73.750 | 0.067 0.299 0.019 | -0.385 19.547  
## Père 35 à 39 ans | 86.098 | 0.515 12.294 0.685 | 0.098 0.874  
## Père 40 à 44 ans | 102.975 | 0.745 12.321 0.574 | 0.557 13.562  
## Père 45 à 49 ans | 61.924 | 0.813 5.179 0.401 | 0.748 8.608  
## Père 50 à 54 ans | 26.310 | 0.838 1.867 0.341 | 0.806 3.394  
## Père 55 à 59 ans | 10.695 | 0.817 0.618 0.277 | 0.810 1.195  
## Père + 60 ans | 3.260 | 0.812 0.241 0.354 | 0.770 0.425  
## cos2 Dim.3 ctr cos2   
## Père - 20 ans 0.359 | -2.128 15.892 0.185 |  
## Père 20 à 24 ans 0.243 | -0.177 1.729 0.007 |  
## Père 25 à 29 ans 0.170 | 0.382 29.467 0.197 |  
## Père 30 à 34 ans 0.648 | -0.224 14.874 0.221 |  
## Père 35 à 39 ans 0.025 | -0.247 12.405 0.158 |  
## Père 40 à 44 ans 0.322 | 0.230 5.167 0.055 |  
## Père 45 à 49 ans 0.340 | 0.553 10.525 0.186 |  
## Père 50 à 54 ans 0.315 | 0.646 4.863 0.202 |  
## Père 55 à 59 ans 0.273 | 0.694 1.956 0.200 |  
## Père + 60 ans 0.318 | 0.605 0.586 0.197 |

plot(res, axes = c(1,2))



Dans le graphique ci-dessus, les lignes sont représentées par des points bleus et des colonnes par des triangles rouges. La distance entre les points lignes ou entre les points colonnes donne une mesure de leur similitude (ou dissemblance). Les points lignes avec un profil similaire sont proches sur le graphique. Il en va de même pour les points colonnes.

dimdesc(res, axes = c(1,2))

## $`Dim 1`  
## $`Dim 1`$row  
## coord  
## Mère - 20 ans -2.3671127  
## Mère 20 à 24 ans -1.1949875  
## Mère 25 à 29 ans -0.3189009  
## Mère 30 à 34 ans 0.3242924  
## Mère 35 à 39 ans 0.7327401  
## Mère 40 à 44 ans 0.9337979  
## Mère 45 à 49 ans 1.0292017  
## Mère + 45 1.0314511  
## Mère 50 à 60 ans 1.0682757  
##   
## $`Dim 1`$col  
## coord  
## Père - 20 ans -2.7512005  
## Père 20 à 24 ans -1.7608104  
## Père 25 à 29 ans -0.6763055  
## Père 30 à 34 ans 0.0666460  
## Père 35 à 39 ans 0.5150954  
## Père 40 à 44 ans 0.7446557  
## Père + 60 ans 0.8124167  
## Père 45 à 49 ans 0.8125884  
## Père + 55 0.8155595  
## Père 55 à 59 ans 0.8167967  
## Père 50 à 54 ans 0.8379469