

User's Guide to the Exploratory Multivariate Data Analysis MOOC

How the MOOC works



1. **Course**: access to all parts of the coursework
2. **Information and announcements**: announcements and supporting documents
3. **Discussion**: a place to discuss the course
4. **Wiki**: a wiki you can contribute to
5. **Progress**: visualize your progress (quiz and exercise results)

Course content

Section

Sub-section

- Principal Component Analysis
- Correspondence Analysis
- Multiple Correspondence Analysis
 - 1. Data - Issues
 - 2. Visualizing the point cloud of individuals
 - 3. Visualizing the point cloud of categories - simultaneous representation
 - 4. Interpretation aids
- Tutorial FactoMineR
- Exercises
 - Exercise due May 02, 2017 at 23:30 UTC
- To go further

Units

VIDEO: VISUALIZING THE POINT CLOUD OF CATEGORIES - SIMULTANEOUS REPRESENTATION

Barycentric representation – simultaneous representation
Optimal representation of individuals
Categories at the pseudo-barycenter :
$$G_k(k) = \frac{1}{\sqrt{\lambda_k}} \sum_{j=1}^J \frac{y_{jk}}{k} F_k(j)$$

Optimal representation of categories
Individuals at the pseudo-barycenter :
$$F_k(j) = \frac{1}{\sqrt{\lambda_k}} \sum_{k=1}^K \frac{y_{jk}}{j} G_k(k)$$

MCA gives only 1 graph: the graph of the simultaneous representation

Télécharger la vidéo en qualité : Haute (1080p) / Normale (720p) / Mobile (480p)

The course takes place over 5 weeks. Each week, one *section* deals with one data analysis method, for example: multiple correspondence analysis.




Each *section* has several *subsections* corresponding to parts of the course, a tutorial using the R package FactoMineR, and exercises.

■ For example, the multiple correspondence analysis coursework has 4 *subsections*:

1. Data – issues
2. Visualizing the point cloud of individuals
3. Visualizing the point cloud of categories – simultaneous representation
4. Interpretation aids.

You should work through each section in the order it is presented.

Each part of the course contains 3 units, which can be accessed using the menu bar:

- a course video: 
- supporting material including slides and audio transcriptions of the videos: 
- a quiz with a few questions on the content of the course video: 

Example from the quiz:

▼ Principal Component Analysis


1. Data - Practicalities

2. Studying individuals and variables






3. Aids for interpretation

Tutorial FactoMineR

Exercises

Exercise due May 02, 2017 at 23:30 UTC 

To go further: handling missing values



QUIZ (10 points possible)

For each question, tick the correct answer or answers.

Q1) In principal component analysis, the variables are

☐ qualitative and quantitative

☐ qualitative

☐ quantitative

?

Q2) In principal component analysis,

☐ we try to summarize the similarities between individuals

☐ two individuals are close if they have similar values for at least one variable

☐ two individuals are close if they have similar values for all variables

?

Q3) In principal component analysis,

- The **FactoMineR tutorial** is a video (or videos) showing how to run each data analysis method using FactoMineR in R.

The data sets and R scripts corresponding to the coursework and FactoMineR tutorials are available for download. You can use them to run the code in R yourself to obtain the outputs seen in the course and tutorial.

Principal Component Analysis

Correspondence Analysis

- 1. Data - Introduction and independence model
- 2. Visualizing the row and column clouds
- 3. Inertia and percentage of inertia
- 4. Simultaneous representation
- 5. Interpretation aids

Tutorial FactoMineR


Case study: text mining

Exercises
Exercise due May 02, 2017 at 23:30 UTC

- Nobel data set (data used in the course)
 - [The data set](#)
 - [The script file \(lines of code R\)](#)
 - [The script with the outputs \(pdf\)](#)
- Birth data set (data used in the tutorial of FactoMineR)
 - [The data set](#)
 - [The script file \(lines of code R\)](#)
 - [The script with the outputs \(pdf\)](#)

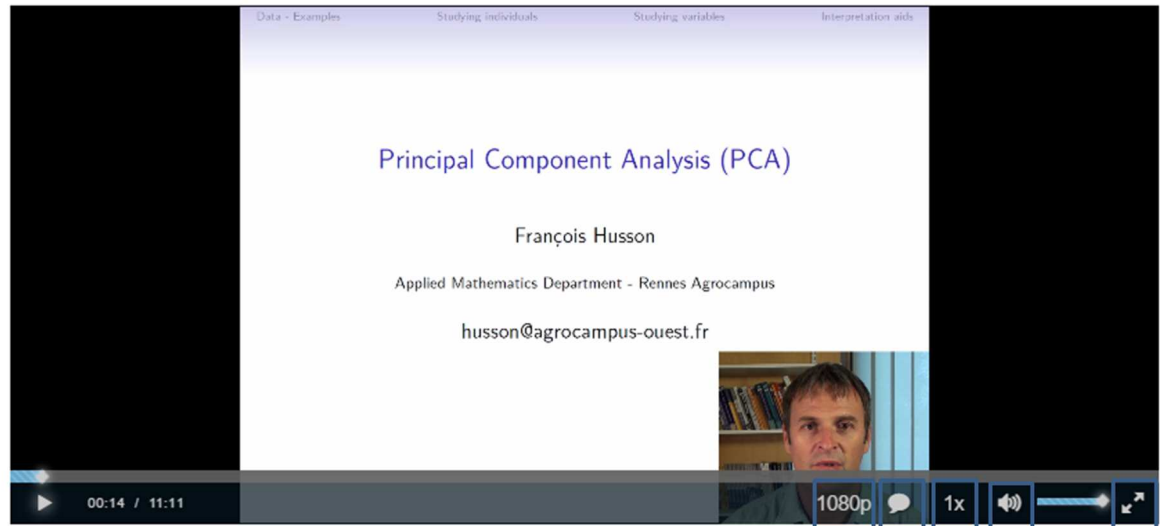
- Each week, there will be course and/or computer exercises proposed. These are for the whole week's course. The computer questions require you to run the data analysis method using the data set provided. Like for the quiz, these exercises are self-evaluated.

Two other subsections called *Case study* and *Extensions* (for example, on how to deal with missing data) are also included with some parts of the course.

 This icon indicates self-evaluation at this location in the course.

The videos

VIDEO: DATA - PRACTICALITIES



1 Télécharger la vidéo en qualité : Haute (1080p) / Normale (720p) / Mobile (480p) 2 3 4 5 6

1. There is a link to download each video.
2. Click on the 1080p button to watch in HD.
3. For greater accessibility, subtitles are available. If you don't want to see them, you can turn them off. Subtitles appear along the bottom of the video and also to the right.
4. You can speed up or slow down the video speed.
5. Here you can change the volume.
6. Go to full screen. To get back, click on "Escape".