Audio transcription of the Multiple Factor Analysis course

Part 1. Data - introduction

(Slides 1 to 6)

Slide 1:

This week, we are going to look at a method to study more complex data tables, where a group of individuals is characterized by variables structured as groups, and possibly coming from different information sources. Our interest in the method is due to it being able to analyze a data table as a whole, but also its ability to compare information provided by the various information sources.

Slide 2 (outline):

There are four course videos for you to watch, which present the main details of MFA, that is: multiple factor analysis.

The outline of the videos is as follows: first, we describe the nature of the data we are working with in MFA. We'll present several examples of datasets, and consider the kinds of questions we want to answer using MFA.

Then, we will see that MFA can be seen as a type of PCA on a weighted matrix that balances the information provided by different groups of variables. We will describe this weighting and explain why it's of interest. MFA provides results on individuals and variables just like PCA does for quantitative variables, and MCA does for qualitative variables.

The most important feature of MFA is that it can take into account several groups of variables. We will see how to compare information provided by each of these groups, what information is common to several groups, and what information is specific to certain groups.

Lastly, we will describe several interpretation aids that are useful for analyzing MFA results.

Slide 2b (outline):

Let's start in this video by defining what kinds of data we are working with, and what the associated issues are with such data.

Slide 3:

Let's first look at the dataset which we will use to present the method. This dataset concerns a sensory evaluation of 10 white wines from the Loire Valley. During this sensory evaluation, 5 Vouvray and 5 Sauvignons were tasted and compared, using sensory descriptors such as acidity, bitterness and citrus odor.

To do this, the judges were required to evaluate each wine and give a score for each descriptor. Scores vary between 0 and 10, with 0 corresponding to a very low or non-existent flavor or odor, and 10 to a very strong flavor or odor. In the picture, we see the judges hard at work, shall we say, evaluating flavors and odors.

When all judges have evaluated all wines according to each descriptor, a table of averages is constructed with each row corresponding to a wine, and each column to a variable (that is, descriptor); so a given entry in the table corresponds to the average value of the given descriptor for the given wine over the panel of judges.

The table shown here concerns the evaluation made by a jury of experts (oenologists). We can see to the right of the table a qualitative variable that characterizes wines: grape variety, that can be either Vouvray or Sauvignon.

This rectangular table, with individuals as rows, and variables as columns, can be analyzed by PCA, with the qualitative variable: grape variety, treated as a supplementary variable.

Slide 4:

However, the problem here is a bit more complex because we have obtained a sensory description of the same wines by different juries: the jury of experts for whom we saw an excerpt before, but also a jury of students and a jury of consumers. Also, the variables used to describe the wines are not the same from one jury to the next. The expert jury evaluated twenty-seven sensory descriptors, whereas the student and consumer juries evaluated only fifteen sensory descriptors.

The problems associated with this dataset are multiple: on the one hand, to describe the ten wines from a sensory point of view using evaluations from the three juries, but also to see if the juries describe the wines in a similar way, or if certain juries describe certain wines in specific ways. To deal with such problems, we will focus on the data table that brings together the data from the three juries.

We also have an evaluation of the appreciation of each wine by a jury of sixty consumers. In this table, rows correspond to wines, and columns to consumers, and each entry corresponds to the score for a given wine for a given consumer. We consider this set of variables for illustrative purposes, because here we are focusing our attention on sensory descriptions only. However, we can relate sensory descriptions and wine appreciation by using these variables as supplementary ones in the analysis. Similarly, the qualitative variable corresponding to grape variety will be considered as a qualitative supplementary variable.

Slide 5:

The structure of our sample data set is as follows. The same set of individuals (for us the ten wines) is described by several groups of variables (in our example a group corresponds to a jury). We consider that we have I individuals in each group, that we have J groups of variables, and that in each of these groups we have K_j variables. The value taken by individual i for variable k is denoted x_{ik}. In each group, variables can be quantitative or qualitative.

In many applications, data tables like this can be found, with the same set of individuals described by several groups of variables. In fact, the groups of variables may correspond to different sources of information. Here are some examples of different applications. In genomics, a group of patients with brain cancer can be described according to three different and possibly complementary sources of information: DNA measurements corresponding to measurements on genes, expression measurements, and protein measurements. The goal is then to understand which genes are involved in the disease and which are differentially expressed, according to whether the patient is ill or not, and the tumor type. The aim is also to know whether the three sources of information are consistent, or whether sources provide discordant information.

Another example could be a survey of young students who are asked several groups of questions about their consumption of products such as alcohol and soft drugs, their psychological state, sleep quality, and lastly, basic details like sex, age-range, etc. All questions are qualitative in this survey. The aim is to broadly understand the links between sleep quality, psychological state, and drug use. We are not interested in connections between pairs of variable, but rather in more global similarities.

Another example is when the different groups of variables correspond to different moments of time. This makes it possible to see overall the evolution of a set of variables over time. In economics, this could correspond to a set of economic indices, from one period to another or from one year to the next.

Examples of multitables like these are increasingly numerous because, on the one hand, it is increasingly easy to collect and store data, and on the other, increasingly, we are trying to analyze more complex problems involving multiple sources of information.

Slide 6:

In the majority of examples, the problem, or problems, will be the following: to study and describe the set of individuals using all variables, and describe the relationships between variables (in one group, or from one group to another). These goals coincide with those of PCA and MCA.

Here we will also take advantage of the group structure of the variables to study globally the similarities and differences between groups: do the variables of one group provide the same information as the variables of the other groups, or is there partly shared information and partly group-specific information provided by that group? Or even: is there no link between the information provided by that group, and that provided by the other groups?

To do this, we can see if a particular individual is described in the same way by all groups, or if one of the groups describes it in a particular way. For example, is wine number two described in the same way by the three juries (experts, consumers and students)? If so, we can say that there is a strong consensus between the 3 juries. Or inversely, is the wine described in a specific or different way by the jury of experts, for example?

Lastly, we can compare the characteristics of individuals for each group. In the wine example, this amounts to comparing the three configurations of wines obtained through separate PCAs constructed from the data of each of the juries.

Before we go on: a crucial point in multi-table analysis is that of equilibrium of the information. One group's information should not overwrite or mask information provided by other groups, so we have to try to balance the information provided by each group. We'll see in the next video which criterion to use to balance the groups in the MFA, and we'll also see that this balancing act can be achieved by simply weighting the variables.

Part 2. Equilibrium and global PCA

(Slides 7 to 14)

Slide 7:

We saw in the previous video that the data to which Multiple Factor Analysis can be applied is that where the variables are structured as groups. We've seen that it is important to balance the influence of each group of variables in the analysis. Let's now look at how to do this, and how to implement it in reality.

Slide 8:

First, why is it important to balance the influence of each group of variables in the analysis?

Note that we have already tried to balance the influence of variables when doing factor analysis before. Indeed, when constructing a standardized PCA, the influence of each variable is balanced to avoid that variables with the greatest variance have more influence in the analysis, that is, more influence in the calculation of distances between individuals.

In a similar way, with several groups of variables, we want to balance the groups in order to avoid certain groups having a too large contribution to the calculation of distances between individuals.

An initial idea, which approximates what happens with PCA, consists in balancing the influence of each group of variables by dividing by its inertia. Indeed, the inertia is a kind of variance in the multidimensional framework. We could therefore weight the groups so that each group has an equivalent inertia. To do this, it's sufficient to weight each variable by the total inertia of the group to which it belongs. With such a weighting, if two groups don't have the same number of variables, the group with less variables will have the same influence as the group with more. From this point of view, balance is attained.

However, this weighting doesn't take into account the method used here: factor analysis. Indeed, factor analysis is sensitive to the distribution of the inertia from one group to another.

In the example shown in this small diagram, a first group is composed of eight highly-correlated variables. The second group is composed of 3 orthogonal variables. Lastly, the third group is identical to the second. With the weighting by the total inertia of each group, the first dimension of group 1 will have a weight close to 1 (this group is uni-dimensional and nearly all of its inertia is located in only one dimension), while the first dimension of groups 2 and 3 will have a weight of 1/3 (if the total inertia of the group is 1, each variable represents 1/3 of the inertia). Consequently, the first dimension for the overall analysis will coincide with the first dimension of the first group. BUT, we are looking for some kind of balance, in order to find common structures from one group to another, so we would in fact like to have appear here the shared structure of groups 2 and 3.

So, a second possible idea consists in balancing the groups, no longer as a function of total inertia, but as a function of the maximum inertia of a dimension. Thus, we can avoid a uni-dimensional group having too much importance in the analysis, and the richness provided by a multi-dimensional group, that is, a group with several axes of variability, is not missed or hidden.

To construct such an equilibrium, we divide the values of a variable by the largest eigenvalue of the group to which it belongs; in fact, we divide by the square root of the largest eigenvalue. In our little example, this weighting makes it possible to see the shared structure generated by groups 2 and 3. This small made-up example corresponds to a situation that is relatively common in practice: some uni-dimensional groups along with some multidimensional groups. The proposed weighting avoids giving too much importance to unidimensional groups, which contain less rich information than the others.

Slide 9:

Here, we have found a weighting that we like, and it remains only to look for the main principal components, as we have seen for PCA, MCA and CA. The difficulty lies in the choice of weighting, as has been explained by JP. Bénzécri, who has said that doing factor analysis, from a mathematics point of view, is just matrix diagonalization, but the science and art of it is to find the right matrix to diagonalize.

The weighting we have found amounts to using, for each variable, a weight equal to the inverse of the first eigenvalue of the group to which it belongs. Thus, the first step of the analysis consists in calculating the first eigenvalue of each group. Then, weighting a variable by the inverse of the inertia of the first eigenvalue of the group to which it belongs, amounts to dividing each variable by the square root of the first eigenvalue of its group. In the notation used here, X_j corresponds to the centered, or centered and standardized, table j, depending on whether the variables of the group j are normalized or not. Lastly, we just have to do a PCA of the table which juxtaposes the (weighted) variables of all the tables.

Slide 10:

For our wine dataset, the PCAs constructed on the data of each jury give the following eigenvalues. The first eigenvalue of the jury of experts, equal to 11.74, is greater than the first eigenvalues of the student and consumer juries. This can be explained by the fact that a greater number of descriptors were evaluated by the jury of experts.

We also note that the following eigenvalues (eigenvalues 2 and 3) are larger for the jury of experts. If a PCA were to be constructed on the whole dataset without first weighting the juries, the jury of experts would strongly influence the results, in the sense that the first dimensions of the analysis would be very much related to the description of the jury of experts. By balancing the influence of each jury through a weighting by each group's first eigenvalue, each jury will make the same contribution to the construction of the MFA dimensions.

The first point to notice, is that the weighting chosen is the same for all variables in a group. This leads to preserving the group's structure. In other words, within a group, the equilibrium between variables is preserved: the ratio lambda_2 / lambda_1 remains the same: 6.78 / 11.74 = 0.58 / 1. Since the equilibrium within a group remains the same, it is also possible to standardize, if desired, the variables in a group, depending on the equilibrium we want to have within the group. This won't upset the balance between groups.

More precisely, thanks to the weighting, the first eigenvalue of each group is made equal to 1, and therefore the main axis of variation has the same variation from one group to the next.

Thus, one group cannot generate on its own the first MFA dimension, even if it has many variables and is highly structured.

One last important thing to note: this weighting sets to 1 the main dimension of variability, and the following eigenvalues are decreased proportionally. Thus, a more multidimensional group will contribute to more dimensions, because its overall inertia, after reweighting, is greater. A more multidimensional group contains more information, so it's normal that it contributes to more dimensions.

Now, is it a problem that a group has a greater total inertia than the others? In fact, no, because this total inertia is spread out over more dimensions. We rediscover here a feature that we have already seen in multiple correspondence analysis.

Slide 11:

We have seen that an MFA can be considered a PCA on the weighted table that juxtaposes the data from all the groups. Since it is a global PCA, we'll first obtain the same plots as in PCA. An individuals' plot and variables' plot. The individuals' plot helps us to visualize the overall similarities between individuals, taking into account information on all of the variables. The variables' plot makes it possible to visualize the relationships between all the variables, and of course, these two plots can be studied simultaneously in order to understand the similarities and differences between individuals due to the variables. In addition to the plots, indices for quality of representation and contribution are also the same as in PCA. Also, as with any PCA, it's possible to add supplementary elements. For instance, supplementary individuals, or additional variables, be they quantitative and/or qualitative.

Slide 12:

For our example, here is the individuals plot. The wines are colored according to grape variety: red for Sauvignon and green for Vouvray. Coloring the individuals according to grape variety shows a contrast between the two grape varieties. The Sauvignon are in the upper left corner, and the Vouvray in the lower right. Recall that the "grape variety" variable was not an active one, and was therefore not used to construct the dimensions. This contrast between grape varieties is therefore closely related to the sensory description variables evaluated by the juries.

We also note that the Vouvray are more spread out in the plane, which can be interpreted as follows: the Vouvray have more variety from a sensory point of view, as compared with the Sauvignon, which are more homogeneous.

One final note: we can see several groups of wines: for example, the Aubuisières Marigny and Fontainerie Coteau Vouvrays are very similar, sensorially speaking, as are Fontainerie Domaine and Fontainerie Brûlés.

Slide 13:

The variables plot, with the circle of correlations, is shown here. The variables were colored according to the group they belong to: in red, the variables used by the experts (with just the name of the descriptor), in green the variables used by the students (with the name of the descriptor followed by "index S"); and in blue, the variables used by the consumers (with the name of the descriptor followed by "index C"). Clearly, this plot is overloaded with information and difficult to read because there are so many variables. But remember, we have several groups of variables and we have represented them all. So this graph summarizes a LOT of information! That's why it's so overloaded.

Slide 14:

So, in order to study this graph more easily, we can select certain variables for better visualizing. The graph is therefore the same as before; we have just masked some labels, highlighting a few of them shared by the different juries.

We can see for example that the variables "passionfruit odor" are quite related, meaning that from one jury to the next, this variable was experienced in a similar way. The same is true for "sweetness". On the other hand, the "acidity" variable was experienced differently: the consumers and students experience this variable in the same way, but differently to the experts. The correlation coefficient between the acidity variable for the expert jury and the consumer jury is 0.34, while that between the acidity variable for the student jury and the consumer jury is 0.76. We cannot know who "correctly understood" the acidity variable, but it's possible that the consumer and student juries had more difficulty in assessing acidity when the sweetness-acidity balance changed.

We're not going to go further into the interpretation of this variables plot, with all its richness, and which, combined with the individuals plot, makes it possible to have a sensory description of the wines obtained from the three juries. These two plots are entirely equivalent to those obtained using PCA, but MFA benefits from the group structure, which we have not taken advantage of before, except in weighting the groups. We will see in the next video the contribution this group structure brings to the story, a contribution which is specific to MFA and which provides much of its interest.

Part 3. Studying groups

(Slides 15 to 27)

We saw in the previous video that MFA is a specifically-weighted PCA that provides an individuals plot and a variables plot, entirely equivalent to those obtained with PCA. Let's now have a look at what the group structure in MFA brings to the table.

Slide 15 (outline):

To begin with, we'll see how to obtain a certain representation of the groups of variables that will allow to globally compare the information provided by each group. Then, we will use the common framework provided by MFA to compare the results of the PCAs constructed from the data of each group. This means comparing the point cloud of individuals from one PCA to the other, and then the dimensions of the separate PCAs.

Slide 16:

First, let's go back to the construction of the principal components. In PCA, the first principal component is the variable (with norm 1 which we call v1 here) belonging to the space R I which is most related to all of the variables, with "most linked" meaning in terms of squared covariance. If the PCA is normalized, the first component corresponds to the most related variable in terms of the squared correlation coefficient.

As the MFA is a PCA on a weighted table, we can write the previous criterion in a way that shows the weighting used by MFA. Each variable x_k is divided by the square root of the first eigenvalue of the group to which it belongs. We have to sum over all the variables, but since the variables are organized in groups, we can write the criterion by making a sum over the groups appear, and a sum over all the variables of the j-th group.

We can extract the first eigenvalue lambda_1 ^ j and show that the first component of the MFA is the variable (with norm 1) which maximizes the squared covariance with the variables of all the groups, with the covariance being divided by the maximum axial inertia of each group. The "maximum axial inertia" corresponds to the inertia of the group's strongest axis.

If Lg (Kj, v1) is the term in red, this coefficient Lg (Kj, v1) corresponds to the projected inertia of all the variables of the j-th group onto the variable v1. It is a measure of the connection between a group of variables and a variable. There is another well-known measure of the relationship between a variable and a group of variables, known as the R-squared, which is used in regression. The interest in Lg is that it is based on the inertia, and in factor analysis, we always try to maximize inertia. The formula tells us that the first component of the MFA is the variable that maximizes the link with all the groups, in the Lg sense. In order to find the second principal component of the MFA, we must look for, among the variables orthogonal to the variable v1, the variable that is most linked to the whole set of groups, according to the Lg criterion. And so on for the following principal components.

The coefficient Lg varies between 0 and 1. It is equal to 0 if all the variables of the j-th group are not correlated with the variable v1. It is 1 if the variable v_1 is the same as the first principal component of the group. Indeed, the first principal component of the group is, by definition, the variable most linked to all the variables in the group, and thus maximizes the sum of the squared covariances. This sum of the squared covariances is equal to the inertia of the first principal component, which is the first eigenvalue of the group. Since the variables are weighted by the first eigenvalue of the group, the maximum Lg corresponds to the value 1.

Slide 17:

This Lg that we have just defined makes it possible to quantify the link between a variable and a group of variables. This measure can therefore be used to quantify the link between a group of variables and a principal component. This suggests that we can use this measure to construct a graph of the groups. A given group will have on the x-axis, the coefficient Lg between this group and the first component, and on the y-axis, the coefficient Lg with the second component.

In our example, the representation of the groups in relation to the principal components of the MFA is as follows.

We can see that all the juries have an x-axis value close to 1, and therefore a high Lg with the first dimension of the MFA. However, a high Lg means that the dimension of the MFA is linked to a strong dimension of the group (the first dimension or one almost as large as the first one). Thus, in our example, the first dimension of the MFA is present in all groups, so it's one that's common to all groups.

We can also see that the second dimension is much more linked to the expert group than to the other groups. This is not really surprising, since we saw that the expert group had a larger second eigenvalue (relative to its 1st eigenvalue) than the other two groups.

Lastly, the points representing the student and consumer groups are close to each other, which means that these groups have several dimensions in common, and thus induce more or less the same structure on the individuals.

This visual view of the data provides a kind of summary representation of the groups, and it's very easy to see which groups are broadly similar. Two groups are similar if they are close on all the dimensions, but since the first dimensions summarize most of the information, it is often enough to compare the groups over the first few (3 or 4 are generally sufficient) .

This graph makes it possible to globally compare the groups and therefore to see if the distances between individuals are similar from one group to another. In other words, if the point clouds seen by one group of variables and another group of variables, are similar.

Slide 18:

We have defined the measure Lg between a component and a group of variables, but this measure Lg can be extended very easily to the connection between two groups of variables. We just have to calculate the sum of all the covariances between the variables of one group and the variables of the other; the variables being of course weighted by the weighting in the MFA.

If we calculate the coefficient Lg between a group and itself, this amounts to calculating the dimensionality of this group. By rewriting the criterion, the ratios of the squared eigenvalues of each component, divided by the largest squared eigenvalue of the group, appear. If the first eigenvalue is much larger than the others, then Lg will be close to 1 and the group is almost one-dimensional. On the contrary, if many eigenvalues are close to the value of the first eigenvalue of the group, then several dimenions of variability are important, and the group will be multidimensional. This coefficient can thus be seen as a dimensionality indicator for a group.

Lastly, the disadvantage of calculating the coefficient Lg between two groups is that this criterion is not bounded. If instead this Lg between two groups is normalized by the dimensionality of each of the groups, what we get is called the RV coefficient. This RV coefficient varies between 0 and 1, and it's easier to use to know if two groups are linked or not, since it's bounded by 1. However, this coefficient tends to be larger when the number of individuals is small, or when the number of variables in each group is large.

Slide 19:

Coming back to our example, the Lg and the RV between the groups are as follows. As the matrix is symmetric, we show here only the lower diagonal terms.

From the Lg, it can be seen that the expert jury gives a more multidimensional description, and thus richer, of the wines than the other juries, because the Lg is higher.

If we look at the RV coefficients, we see that the student and expert juries are close to each other, because the RV is close to 1 (0.85).

Lastly, we can consider the mean configuration of the MFA and look at the link between this and each group. The mean configuration of the MFA is here a group, corresponding to the set of coordinates of the individuals on all the dimensions. It's the "shared" or mean configuration. These RV coefficients are given in the MFA row. We can see that the student jury has a configuration that is very close to the mean one, since the RV is very close to 1 (it's 0.96).

Slide 20:

Let's now look at another way of comparing the groups, starting from the configurations of the individuals in each group. For this, the typology provided by each group will be compared in a shared framework, more precisely, in the shared reference frame given by the MFA. Comparing the typologies means asking if individuals are seen in the same way by the different groups, or if certain individuals are specific to some groups.

Slide 21:

To do this, we'll proceed as follows. Let's illustrate the approach with a data set containing 3 groups of variables. The MFA is first built from these 3 groups, as we have seen earlier. This gives us a representation of the individuals on the principal dimensions of the MFA. The individuals are in a space R K, with K dimensions, where this space is the direct sum of the spaces R K j generated by each group of variables. Let's now try to represent each individual in the dataset using only the data from one group of variables. We'll call partial individual i to the power of j, the i-th individual seen by only the variables in group j. How can we represent this partial individual? And above all, how can we represent all of the partial individuals associated with the i-th individual (namely, the partial individuals i 1, i 2, up to i J) in the same plot?

The shared plot is the space provided by the MFA using data from all of the groups. We juxtapose one table per group with the main data table. In the sub-table for group j to be projected, all variables that do not belong to group j take the value zero (because the array is centered, and otherwise the values would be the means of the variables). Also, the values of the variables belonging to group j are centered (and possibly standardized if we have standardized the variables of this group), and divided by the square root of the first eigenvalue of the group. In other words, the values used in the overall MFA analysis. These partial individuals are then used as supplementary individuals, and thus projected onto the axes of the global analysis.

We thus get the projection of the i-th individual seen by the variables of group 1, shown here in red. An important point, however, is that since we have 0 for all groups other than group 1, the partial point of group 1 tends to be found near the center of the point cloud. This is true for all partial individuals in each group. To be clear about this, let's take the trivial example where the three groups of variables are identical. All the partial points of the same individual should be superimposed, and also superimposed on the "average" individual. Now, if we calculate the projection of a partial individual of a group, as it takes the value zero on the other 2 groups, its coordinates will be close to the point cloud's barycenter. It will in fact be 3 times too close to the barycenter. Thus, the projection of all the partial points must be dilated by a factor of 3 in our example. When the data set has J groups, the partial points must be projected as supplementary points and then dilated by a factor of J.

With this dilation, we have for each individual as many partial points as there are groups, and the mean point, corresponding to the individual seen in the global analysis, is at the barycenter of the partial points.

Slide 22:

The partial points representation is therefore very useful, since it makes it possible to compare the representation of individuals from one group to another. Take the following example. Individuals were first questioned on several issues related to their opinions. Then they were interviewed on several issues related to their behavior. These two sets of questions can be considered as two sets of variables in the MFA. The MFA then provides a representation of the individuals, as seen by all of the variables. This means the point corresponding to the global analysis, that is the mean point here, in purple. An interpretation of the positions of the mean points on the dimensions can be made as is done in PCA. We can also represent the partial points corresponding to an individual seen only in terms of the opinion variables, and then only in terms of the behavioral variables. If the two partial points are close to each other, and therefore close to the mean point, then the individual behaves in accordance with their own opinions, whereas if the partial points are far apart, then the individual doesn't behave in accordance with their own opinions. The distance between the 2 partial points is a way of visualizing this discordance.

Let's take another example involving a questionnaire which participants in a course answered before starting the course. Several questions relate to their expectations of the course. A second questionnaire was sent to the same people at the end of the course. The questions in it relate to what the participants have learned during the course. The individual on the right has its two partial points close together, while the individual on the left, has them far apart.

The individual on the right therefore got from the course what they expected from it, so we can suppose that they were satisfied. The individual on the left, on the other hand, didn't learn what they expected to learn from the course. You could even say that they were surprised by the course. But what type of surprise? In other words, were they disappointed? Or were they surprised, but in a good way? To find out, we can return to the interpretation of the dimensions, and more specifically here, to the interpretation of the 1st dimension, which separates the 2 partial points. If positive coordinates on the axis are related to a strong positive appreciation, then the individual has learned more than they expected from the course, and can be said to have been pleasantly surprised. And if not, we can suppose that they were disappointed. This shows that the position of the mean points, that is, of the individuals seen by all of the variables, and the positions of the partial points, can be interpreted using just a single plot of the shared dimensions of the MFA.

Slide 23:

The transition formulas that we have seen in PCA can be directly applied to the mean points in the MFA. Remember that these transition formulas make it possible to interpret an individual's position thanks to the variables plots. An individual is to be found on the side of the variables for which they have large values, and opposite from variables for which they have small values.

These transition formulas can also be used for the partial points. We just have to restrict the relevant formula to the variables of group j. As mentioned earlier, we multiply the coordinates of the partial points by J, the number of groups, so that the coordinates of the partial points and the mean points can be interpreted all at once and on the same scale. The dilation by a factor of J applied to all the partial points puts the mean point at the barycenter of its partial points, so that mean and partial points are represented on the same scale.

Slide 24:

In our wine example, this superimposed representation looks as follows. The points in black correspond to the wines as seen by all the juries. The points in color correspond to the partial points: in red the wines seen only by the expert jury, in green by only the student jury, and in blue by only the jury of consumers.

Each black dot corresponding to the mean wine is of course at the barycenter of the colored points.

We can see that all the wines are interpreted fairly homogeneously by the three sensory juries, because all the partial points for each given wine are close to the mean point. Except perhaps the Aubuisières Silex wine, lower left, which was perceived as more "extreme" by consumers than by experts for variables related to the first dimension. Indeed, on this dimension, the value for this wine is strongly negative for consumers, and fairly close to 0 for the experts.

Slide 25:

Visually, we can see that the coordinates of the partial points are relatively similar from one group to another for the first two dimensions. In fact, we can define an index measure to quickly and globally calculate whether the coordinates of the partial points are close to those of the mean points, axis by axis.

To do this, for a given axis, we are going to break down the total inertia of the partial points. For the s-th dimension, the total inertia of the partial points can be broken down into the mean between-points inertia (that is, between-individuals), and the within-individual inertia. Also, we can calculate the ratio of the between-point inertia to the total inertia.

In our example, for the 1st dimension, the ratio is 0.93, and is therefore very close to 1, which means that the coordinates of the partial points with respect to a given individual are very close to each other, and therefore very close to the coordinates of the corresponding mean point.

The within-inertia, which measures the similarity between partial point clouds, dimension by dimension, can also be broken down by individual. This decomposition by individual is very useful for sorting individuals in terms of decreasing within-individual inertia, and for identifying which individuals are perceived differently by each group of variables. When interpreting differences between groups, these individuals can be examined more closely.

Slide 26:

We have compared the groups of variables using the partial points plot, which represents the features of the individuals of each of the groups in a shared global space. It is also possible to compare groups using the dimensions of separate analyses, in other words, by comparing the principal components of the PCAs for each group.

To do this, we are going to take a closer look at the principal components of the PCA of a given group.

And in fact, we are going to use them like supplementary variables in the global analysis, that is, in the MFA. This will allow us to project the principal components of this group, and see how these components are related to the MFA's dimensions. If the principal components of the group are highly correlated with the dimensions of the MFA, then the individuals plot in the PCA of this group will have the same form as the individuals' plot of the mean points of the MFA.

Slide 27:

For our example, here is the plot which gives the projection of the principal components of the individual PCAs onto the first two dimensions of the MFA.

The first dimension of the students' PCA is highly correlated with the first dimension of the MFA. Therefore, wines with a small coordinate value (or a large one) on the x-axis of the PCA for the student data have a small coordinate value (or a large one) on the x-axis of the MFA. The same is true for the second dimension of the students' PCA, which is also highly correlated with the second dimension of the MFA. This means that the description of the wines provided by the students is very similar to the description of the wines obtained by the MFA, and therefore that of all of the juries.

For the experts, dimensions 1 and 2 of the PCA do not exactly coincide with dimensions 1 and 2 of the MFA. However, the PCA dimensions are both well-projected, which means that the PCA's 1st-2nd dimensions, built using the expert's data, is close to the MFA's 1st-2nd dimensions, but with a rotation. The same thing has happened with the consumers data.

Part 4. Further topics

(Slides 28 to 39)

We have seen in the previous videos how to take into account group structure using multiple factor analysis when all of the groups contain quantitative variables.

Slide 28:

We're now going to have a look at some additional stuff, starting with how to take into account groups of qualitative variables. Then, we'll see what to do when one or more groups of variables correspond to one or more contingency tables. And lastly, we'll see which interpretation aids are useful for interpreting the results of an MFA. We'll focus on ones which are specifically made for MFA.

Slide 29:

Consider a data table where all variables are qualitative and all are structured by group. This situation is then similar to that in which the variables are all quantitative. Indeed, we'll try to balance the influence each group of variables has in the global analysis, and then we'll construct specific MFA plots to study the similarities and differences between groups. The individuals plot will be identical to the one obtained in Multiple Correspondence Analysis, and thus it will also show the categories. Remember that a category will be found at the barycentre of the individuals that are part of it. Then, we'll be able to construct the plot showing the superimposed representations, the groups plot, and the partial axes plot, which will show connections between the global analysis and the separate ones.

The methodological approach is therefore the same as before except that, on each group of variables, we use MCA rather than PCA. Thus, for each group, we'll construct a complete disjunctive table, then we'll weight it as in MCA as a function of the column margins. Then, all that remains is to balance the groups by weighting by the first eigenvalue of their MCA.

Slide 30:

In our example, we can add the group of qualitative variables, restricted just to the grape variety variable. This will help us to illustrate the outputs obtained when certain groups are qualitative, depending on whether they are active or supplementary. Here we have the groups plot, where we can see that the "grape variety" group has coordinates close to 0.4 on both dimensions. This means that this group is related to the first 2 dimensions of the MFA, and therefore that the wines of the two grape varieties are distinguished from each other on both dimensions.

We can check this on the individuals plot by coloring the wines according to the qualitative variable: grape variety.

The partial axes plot shows a single dimension for the qualitative group. This is normal because since there is only one variable with two categories, the space generated has only one dimension, that is, the total number of categories, here 2, minus the total number of variables, here 1. This dimension is well projected in terms of MFA, so the variability induced by the group of qualitative variables is recovered well by the MFA plane.

Next, we can construct the graph of partial points of individuals and categories. Here, we show only the partial points for the categories to avoid overloading the plot. We can see that the expert jury tends to separate to a greater extent the Vouvray from the Sauvignon, because the partial points are much further apart compared to those for the consumers, which are close to the center of the plot.

Slide 31:

It's also possible to mix the variable types from one group to another, and to study, using MFA, data tables in which certain sub-tables are made up of quantitative variables, and others of qualitative variables. Within each group, the relationships between variables are studied like in PCA when the variables are quantitative, and like in MCA when the variables are qualitative.

The influence of each group is then balanced using the MFA weighting *i.e.* the first eigenvalue of the principal component method constructed on the data of each group. So: first eigenvalue of the PCA, or first eigenvalue of the MCA, according to the group's variable type. The associated plots will be the same as those we have already seen, that is, plots with individuals and categories, variables plots, and then specific representations including the groups plot, the partial points plot, and the partial axes plot.

Let's also take note of the following specific situation: if each group of variables consists of only one variable, quantitative or qualitative, then we arrive at a method called factor analysis of mixed data. This method allows us to analyze simple data tables consisting of quantitative and qualitative variables. The visual outputs are a mix between PCA ones for the quantitative variables, and MCA ones with an individuals and categories plot, along with a variables plot with the squares of the relationships. That is, the square of the correlation ratios for the qualitative variables, and the square of the correlation for the quantitative variables. This multiple factor analysis of mixed data thus balances the influence of each quantitative and qualitative variable. It's quite a useful way to proceed, because many data tables contain both quantitative and qualitative variables.

Slide 32:

MFA can also be extended to contingency tables. Thus, some groups of columns can be contingency tables. All these tables therefore need to have one dimension which is the same.

The main difficulty is then to find a weighting for the rows of the table, since the weights of the rows are different from one contingency table to the next. To choose the shared weight for a given row, we can take the sum of the entries of this row across all contingency tables, divided by the total sum of the entries of all of the contingency tables. When working with contingency tables, the analysis of the results and the interpretation of the MFA are exactly the same as for the usual MFA.

Let's take a look at a few applications of this. Firstly, there are many examples of this in text analysis, which corresponds to an extension of studying text data with correspondence analysis. The following example involves the same survey, but carried out in several countries. Each column corresponds to a question, and the rows correspond to the categories of a qualitative variable (age groups, for example) and the table entry corresponds to the number of respondents who answered “yes” to that question. Each country corresponds to a table, and we want to look for information common to all countries, or specific to certain countries, using MFA.

Another example is in ecology, with a table matching sites with species, gathered over several years. We'd be interested in looking at the evolution of associations between site and species over time. The sites, corresponding to rows, could be the same from year to year, while some species - the columns - may disappear and no longer be present some years.

We also note that MFA can simultaneously deal with frequency tables, tables with quantitative variables, and tables with qualitative variables. The row weights are then calculated using the frequency tables.

Slide 33:

Now, let's go back and take another look at interpretation aids which are useful for all types of principal component methods, including PCA, MCA, and MFA.

First, it's possible to add supplementary information, supplementary individuals, or supplementary variables, and of course, in MFA, supplementary groups of variables. These additional elements do not participate in the construction of the dimensions, but may be of help in the interpretation.

In our wine example, we had 3 groups of variables corresponding to the tastings of the 3 juries, but we also had preferences data. A new group of consumers tasted the wines and gave them appreciation scores (0 for: I do not like at all this wine, up to 10 for: I very much appreciate this wine). A data table can then be constructed on this, with the ten wines as rows, and the 60 tasters as columns. Each table entry corresponds to the score given to a certain wine by a certain taster. This table can then be juxtaposed with the data table used in the MFA, since the rows are the same.

We can then ask ourselves the following question: is the appreciation of wines linked to the sensory descriptions? To answer this question, we are going to redo an MFA by taking into account the preferences group, made up of 60 additional variables.

We can also use the "grape variety" variable as supplementary information, but we have already seen how to analyze the results of this qualitative group, so we will content ourselves with analyzing the results of the supplementary quantitative group.

Slide 34:

The preferences group having been added as supplementary information, the plot of the mean individuals and the variables plot obtained by MFA are the same as before. They give the sensory description of the wines constructed from the tastings of the three juries.

The preferences group has relatively strong coordinate values on the first two dimensions of the MFA. This means that overall this group is linked to these two dimensions, and therefore to the sensory description of the wines. That is, wine appreciation is not independent of the sensory descriptions.

Now, going into more detail, all of the preferences variables can be represented on the variables plot, but for the sake of readability, we have put them on a second plot. Each arrow corresponds to a taster, and we can see that most of the arrows are pointing towards the bottom left, and are rather well represented. This means that the wine to the bottom left is highly appreciated by many tasters. This is the Aubuisières Silex wine, which is a wine described as "sweet" by the 3 juries, as can be seen on the sensory variables plot.

Slide 35:

With MFA, we again find the classic indices for contributions and representation quality. For individuals and variables, these indices are calculated exactly as in PCA.

For groups, we can also calculate these indices. First, the contribution of a group is equal to its coordinate value (and not its coordinate value squared), divided by the sum of the coordinate values, if we want to express relative contributions. We can therefore directly read off a group's contribution to the construction of an axis using the groups plot.

Lastly, the representation quality of a group is measured as usual by the squared cosine of the angle between the vector starting from the center of the cloud and going to the point representing the group in the space, and the vector starting from the center of the cloud and going the projection of the group onto the dimension.

We can then sum the squared cosines over several dimensions to calculate the representation quality in a plane or a subspace.

Slide 36:

As with other principal component methods, the dimensions can be characterized automatically by calculating the links between each variable and each dimension. This is particularly useful in MFA because the number of variables is often quite large. To describe each dimension using the quantitative variables, we can calculate the correlation coefficient between each variable and the dimension, and then sort them from largest to smallest, keeping only those that are significantly different to zero.

In our example here, we only use the active variables, that is, the sensory descriptors, to characterize the axes. We could also characterize the axes using the supplementary variables, but here the preferences are not very interesting in terms of being able to understand the dimensions. The variable most linked to the first dimension, in fact negatively linked to it, is the typicity of the "Chenin" taste, as evaluated by the students. Next up is "vanilla odor", as evaluated by the experts, which is positively correlated with the first dimension. In this way, we have characterized the first dimension. We have therefore found a way to see which of the fifty-seven variables best characterize the dimension.

Slide 37:

To describe the dimensions using the qualitative variables, we construct a one-way analysis of variance model, in which we characterize the coordinates of the individuals (that is, the wines) on a dimension according to the qualitative variables. The qualitative variables are then sorted according to the correlation ratios, and we keep the ones whose correlation ratio is significantly different to zero. This makes it possible to see overall which variables are most related to the dimensions. In addition, Student *t*-tests are performed in order to sort the categories. We can then test whether the coordinate value of a given category is significantly different to zero or not. Again, here, the categories are then sorted according to the p-values, and as a function of the sign of the coefficient.

In our example, the "grape variety" variable is related to the first two dimensions. The p-value is very slightly less than 5%, and therefore we reject the hypothesis that the correlation ratio is equal to 0. Recall that here the "grape variety" variable is a supplementary one, and therefore was not used to construct the dimensions. Thus, the test we have built is valid. On the other hand, for the sensory variables, the tests must be used with more caution since the variables themselves were used to construct the dimensions, and are therefore automatically related to them. If we look at the categories, we can see that the Vouvray wines have significantly positive coordinates on the first dimension, and significantly negative ones on the second dimension. They are therefore found bottom right, while the Sauvignon are found top left.

Slide 38:

To conclude, here is a slide summarizing the important steps required to implement MFA.

To begin, the first question to ask is: is there a group structure in the dataset? Or should we do a simple analysis of the table using PCA (if the data is quantitative) or MCA (if the data is qualitative)? If there **is** a group structure, which groups should we build? A group corresponds to a subset of the table, and therefore most often to a certain type of information. This choice of grouping is essential because it is used to define the weightings and balance between groups.

Then, we have to choose the groups that will be active, and the groups that will be supplementary.

For the groups of quantitative variables, we then have to decide whether the variables in each group should be normalized or not, depending on the importance that we want to give to each variable within each group. This really means looking at the importance of the variables **within** each group, and not **between** groups, because the weighting of the MFA by the first eigenvalue of each group means granting the same importance to each group of variables.

Then, we run the MFA.

Next, it's necessary to choose the number of dimensions to be interpreted, by looking for example at a plot of the eigenvalues of the MFA.

Then, we begin by interpreting simultaneously the individuals plot and the variables plot, before using the specific outputs of the MFA to study the groups overall, to see: first, which groups resemble each other (with the help of the groups plot), and second: which are the individuals not "seen" in the same way by each group, with the help of the partial points plot.

The separate analyses can then be examined in detail to see which are the dimensions of each group that are related to the dimensions of the MFA.

Lastly, the indices for quality of representation, contribution, Lg or RV coefficients, can be used to enrich our interpretations.

Some software packages have a function for running MFA. The most advanced one is the MFA function in the FactoMineR package, since it allows us to do MFAs with any type of group: groups of quantitative variables, groups of qualitative variables, and contingency tables. Each group can be either active or supplementary. Via this function, you can also access all of the interpretation aids described in this course.

Slide 39:

So, let's conclude. MFA is a multi-table method useful for analyzing tables with the same rows. The groups, or tables, must have quantitative or qualitative variables, or represent frequency tables.

What MFA does well is to balance the influence of each group, then represent the information provided by each group in a shared frame of reference. An MFA provides two main plots: an individuals plot and a variables plot, which are classical outputs in principal component methods. These plots are analyzed the same way as in PCA and MCA. But MFA also provides its own specific outputs, which contribute to its importance and relevance. These specific outputs make it possible to compare information provided by each table. This information can be compared in a very general way thanks to the groups plot.

We can then see if overall the groups have the same dimensions, and if the dimensions provided by the MFA are present in each group. The information provided by each group can then be compared more precisely by the representation of dimensions in the separate analyses. Essentially, we can see whether the dimensions of variability of a given table (obtained by PCA, MCA or CA) are related or not to the shared dimensions, that is, to the MFA dimensions. Lastly, the information provided by each group can also be compared in terms of each individual, by means of the partial points plot.

Here is a book written by Jérôme Pagès that is dedicated to multiple factor analysis. In this book, you can find more details on the mathematical justifications of the methods we've presented here, and also plenty of examples which use MFA.

We've now reached the end of the multiple factor analysis course, so now you can go on and see how to implement multiple factor analysis using FactoMineR. Don't forget to have a go at the quizzes and exercises that come with the course. Good luck!