# Introduction to Latent Class Analysis

## Video 1 transcript

https://www.ncrm.ac.uk/resources/online/all/?id=20806

Oliver Perra: So, in this first presentation, I will provide an example of a research problem that we can solve using Latent Class Analysis. I will then talk about the main characteristics of Latent Class Analysis and also describe the main goals and the type of research questions that Latent Class Analysis can answer. And I will then briefly talk about the main assumptions of Latent Class Analysis.

So, I will start with simulated example. Assume that I am observing a series of people and, just for the sake of the example, I will assume they fall into two genders. I will observe these people over a period of two weeks and check if they have experienced any of these problems. Low mood, lack of interest or pleasure, fatigue, sleep problems. These four issues are symptoms of major depression.

And let's assume that every individual in this figure represents a number of individuals that report the same pattern of symptoms. In this example, there is heterogeneity. People report different responses and I want to make sense of this variability, the variability I observe in these symptoms and find ways to categorise people into fewer groups.

For example, I might be interested in identifying a group of people are more likely to be depressed and people that are less likely to be depressed, and I might decide arbitrarily that if people show all the symptoms, they are more likely to be depressed. But I might also arbitrarily decide to divide my participants into three categories, the most likely depressed, the unlikely to be depressed and the certainly not depressed. But ideally I'd like to find a method that is more robust than just the arbitrary choices I'm describing.

Data class analysis provides a method to decide how many categories I need to make sense of the variability I observe in people's behaviour. And Latent Class Analysis also allows to categorise individuals according into categories according to their observed behaviour. But rather than arbitrary rules, Latent Class Analysis uses probability rules, therefore providing a more robust, systematic and transparent method to categorise individuals into groups based on the behaviour I observe.

So, I will now illustrate the main characteristics of Latent Class Analysis.

Use the example of different symptoms of depression to highlight that Latent Class Analysis is a person-centred approach. It assumes that the pool of people we observe is made up of a mixture of individuals, different categories of individuals that differ in distinct ways each from the other.

Latent Class Analysis is a measurement model. What does it mean? Well, the example of symptoms of depression highlights that Latent Class Analysis assumes that what explains the differences we observe in behaviour patterns are some unobservable latent characteristics.

And the underlying depression status basically explains why some people report low mood, loss of interest, fatigue and sleep problems over a period of time and why other people do not. So, we cannot observe depression directly, but we can infer the latent status of being depressed or not by observing symptoms of depression. So, Latent Class Analysis is conceptually very similar to factor analysis, but the key difference is that the underlying constructs that explain differences in behaviour are in the case of Latent Class Analysis categorical differences rather than quantitative differences.

In this example, rather than representing the depression as a continuum where people may be more or less depressed, we are representing the underlying condition of being depressed or being likely to be depressed as two different categories. Whether we should represent the underlying depression condition as a categorical or continuous variable is probably a question that we should answer considering theory and substantive knowledge, and not just statistics. However, if we think that it makes sense to suggest the categorical underlying variable to explain individual differences in symptoms of depression, then Latent Class Analysis is the approach we should use.

[0:05:13]

So, the basic situation in the example I provided is a situation where I'm observing participants’ behaviour in a set of variables. And observing these variables I can see that they are correlated. In the example, this correlation means that people who are in low mood most of the time also tend to report loss of pleasure and fatigue most of the time, sleep problems and so on. I call these behaviours indicators because we are assuming that they are signs of some underlying construct, for example, depression.

I also want to emphasise that it's not, Latent Class Analysis is usually described considering observed categorical variables. But Latent Class Analysis can be applied to any type of variables. For example, it could have considered how many days during two weeks people were in low mood, were displaying lack of pleasure or interest and so on. So, it's possible to use any type of variable to run Latent Class Analysis. But I will stick to categorical variables for simplicity.

So, the latent class model assumes that what explains the correlation between these indicators, these variables I observe, is an underlying construct. In the example, the categorical dimension of being depressed or not.

And note that when I introduced the latent classes in this frame, I have removed the correlations between symptoms. And this is to represent one of the main assumptions of Latent Class Analysis, that of conditional independence. This means that we are assuming that the variables we observe are conditionally independent of each other given the level of the latent variable. Put it in another way, the association between being in low mood most of the time, inability to experience pleasure most of the time, and so on happened because these behaviours are caused by an underlying status, that of being depressed. If I control for the underlying status, I should not observe residual associations between these variables.

So, I will now talk about the main goals of Latent Class Analysis and provide some examples of research questions that can be answered using Latent Class Analysis.

And Latent Class Analysis is similar to a factor analysis as I highlighted. The main difference is that what explains differences in behaviour patterns we observe is not a continuous dimension but a categorical variable. So, Latent Class Analysis can be used to identify the number of categories that are necessary to explain differences in the behaviour patterns we observe. And it's also used to characterise these underlying categories based on their associations with observed behaviours.

So, I will provide an example of a study that I have selected because emphasises the categorical nature of the underlying variables. In this study, Grunow and others used data from the European Values Study to investigate citizens’ views about gender roles. The participants reported how much they agreed with statements such as preschool children suffer if the mother works and so on. The answers were dichotomised into two categories of response. Agree versus disagree. And the researchers expected people to differ along the dimension of traditional versus egalitarian attitudes to gender roles. But they also hypothesised that more nuanced attitudes may be emerging.

And indeed, the results of this study identified three additional categories of individuals beyond the egalitarian or traditional views on gender roles. For example, the study identified the category of egalitarian essentialists, people who, despite agreeing that women and men can have similar roles, still believed that women mostly aspire to dedicate themselves to their home and children. But the researchers highlighted how more nuanced views have emerged, even in countries where policies still endorse separate roles for males and females.

But the latent class approach can also be extended to answer other types of questions and particularly questions about the precursors and the consequences of latent classes of people being different latent classes. Here I reported in the picture the five classes of gender attitudes that were identified in the study by Grunow others.

[0:10:48]

And we can investigate what could be the factors that influence individuals’ membership to different categories, for example, where there are significant differences across countries in people's likelihood of being egalitarian or traditionalist or essentialist egalitarian and so on or where differences across socioeconomic status so that different socioeconomic states are where associated with different categories of gender attitudes.

But we can also investigate if class affiliation is related to or predicts other behaviours. For example, two different classes of gender roles attitudes influence intentions to vote in referendum, in referenda, for example, referenda on abortion laws. So, it's possible to also include Latent Class Analysis as dependent variables and check how these categorisations are influenced by covariate and investigate how different categories are related to these sort of outcomes, as in this example here.

So, now I will briefly introduce some of the main assumptions of Latent Class Analysis and in the next presentations, I will provide more formal definitions of these assumptions.

The first assumption is that latent classes are mutually exclusive. So, each individual in our sample belongs to only one of the underlying classes, for example, in this picture, depression or no depression. The latent classes are also exhaustive. So, each individual sample belongs to one of the classes identified. So, every individual has one class.

Individuals within each class are supposed to behave in the same way, or in other words, there are supposed to be no differences in the level of behaviour between individuals within the same class. The differences we observe in the behaviour of people within the same class is due to variables in measuring the behaviour. In other words, the assumption of a latent class model is that individuals within one class are homogeneous in their behaviour and they differ from individuals in the other classes. More formally, individuals are supposed to have the same parameters when they are in the same class.

And the associations between the indicator behaviours and the underlying classes, the underlying latent classes, is therefore probabilistic and not deterministic. We estimate the probability that people in one class display a specific pattern of behaviour.

But this also means that class affiliation is also estimated with error. Individuals’ class membership is not certain but estimated as a probability of belonging to a latent class based on the pattern of behaviour we observe.

And here, for example, represent different probabilities of people being in the depression class and different probabilities of different people being in the no depression class.

But the main point is that there is uncertainty about membership to latent classes and different latent class outputs can vary in the degree of uncertainty concerning individuals’ latent class membership. And this is important to consider when we use Latent Class Analysis.

So, to summarise, Latent Class Analysis is a person-centred approach. It uses probability to estimate the number of underlying categorical types that can explain variability in behaviour patterns we observe. Latent Class Analysis also allows to assign individuals to these underlying latent classes or categories, but this assignment is probabilistic, not certain. Latent class analysis is a measurement model. It assumes that the patterns of behaviour we observe in our participants is explained by the fact the participants belong to different categories. And since the underlying latent variable that explained participants behaviours are categorical, latent class analysis also allows to identify interindividual differences that may go beyond unidimensional differences. In fact, individuals can differ in ways that are as it is categorical quite distinct. And how participants can differ in qualitative way across different dimensions as an example of a study that I described before illustrated.

National Centre for Research Methods (NCRM)  
Social Sciences  
Murray Building (Bldg 58)  
University of Southampton  
Southampton SO17 1BJ  
United Kingdom

**Web** www.ncrm.ac.uk   
**Email** info@ncrm.ac.uk  
**Tel** +44 23 8059 4539  
**Twitter** @NCRMUK