

Promoting the use of Latent Class Analysis across the GSS to improve survey quality and analysis.

Debbie Cooper, Bethan Russ and Matt Greenaway, Survey Methodology Division, Office for National Statistics

1. Introduction

In the 2015/16 financial year the Office for National Statistics (ONS) was granted funding by the Quality Improvement Fund (QIF) to promote the use of Latent Class Analysis (LCA) across the Government Statistical Service (GSS).

The key points of the bid were to disseminate the LCA technique to colleagues throughout the GSS; allowing them to use the technique to produce better quality and more relevant outputs for users of Official Statistics.

The aim of the project was to disseminate the technique via a series of seminars in Newport, Titchfield and London; thus making the seminars more accessible to colleagues working across a number of departments.

2. What is Latent Class Analysis?

Latent Class Analysis provides a flexible and powerful approach to categorical data analysis (McCutcheon and Hagenars, 1997). It is a type of model-based cluster analysis.

In numerous studies, particularly in social research, researchers are interested in latent variables (variables that cannot be measured directly) e.g. personal well-being or quality of life. These variables tend to be measured by means of a number of indicator (observed) variables. For example, the ONS uses four indicator variables to measure personal well-being (a latent variable) in the UK: Anxious, Happy, Satisfied and Worthwhile.

In LCA, the indicator variables are categorical variables. LCA is used to identify patterns of responses to the indicator variables to create a set of mutually exclusive latent classes, that is, groups of individuals or other units of analysis. Individuals in the same latent class will have similar response patterns to the indicator variables whilst individuals across latent

classes tend to have different response patterns to each other. In other words, LCA splits respondents into homogenous groups (latent classes).

LCA has numerous advantages over traditional cluster analysis techniques such as hierarchical cluster analysis and K-means clustering. Some of these advantages include:

- It is model-based unlike other types of cluster analysis which tend to be distance-based. An advantage of this is that there are more formal criteria for choosing the final model when using LCA (for further information see Vermunt & Magidson, 2002).
- It is relatively easy to deal with variables having different scale types (Vermunt & Magidson, 2002).
- In traditional cluster analysis techniques persons are assigned to clusters on an all-or-none basis. On the other hand, LCA allows membership of a person to each cluster to a certain degree allowing for fractional cluster membership (captured by posterior possibilities).

3. Dissemination of Latent Class Analysis

During the outset of the project the aim was to run a series of seminars in Newport, Titchfield and London, promoting the use of LCA across the country to encourage attendees from a range of government departments. Due to unforeseen circumstances these seminars were unable to go ahead. To overcome this problem a PowerPoint presentation was developed and disseminated to the GSS via their website.

This presentation provided a high level description of what LCA is, the advantages of LCA, types of LCA and real life examples of where LCA has been used previously. The presentation was originally developed as a narrated presentation, providing commentary on each slide. Due to publishing restraints on the GSS website, namely the ability to upload large sound files, the narrated version of the slides was not published and a regular version of the presentation was published instead. A note on this issue was published alongside the presentation, encouraging users to contact the ONS Methodology department if they required a narrated version.

To aid interpretation of the presentation a full paper titled "*A Short Guide to using Latent Class Analysis*" has been produced and will be published in the following months. The paper builds on the information presented in the narrated presentation, with the additional content providing a full worked example of LCA using R software (R Development Core Team, 2011). The paper was originally planned to be published alongside the presentation,

however the applied example uses unpublished personal well-being data and is therefore subject to pre-release constraints and the constraints set out due to purdah.

To access the published presentation please go to the news pages of the GSS website using this [link](#). The paper will become available shortly and will be published alongside the presentation.

4. References

McCutcheon A.L. and Hagenaars, J.A.,1997. Simultaneous Latent Class Models for Comparative Social Research. In: Langeheine, R. and Rost, J. (eds) Applications of Latent Trait and Latent Class Models. New York: Waxmann. Pgs. 266-277.

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. 3.0.2. Available at: <<http://www.R-project.org>> [Accessed 19 January 2016].

Vermunt, J. K., & Magidson, J. 2002. Latent class cluster analysis. In: J. A. Hagenaars, & A. L. McCutcheon eds. *Applied latent class analysis*. Cambridge University Press, New York. pp.89-106.