# Latent Profile Model

**Jeroen K. Vermunt**

The latent profile model is a LATENT VARIABLE model with a categorical latent variable and continuous manifest indicators. It was introduced in 1968 by Lazarsfeld and Henry. Although under different names, very similar models were proposed in the same period by Day and Wolfe.

Over the past ten years, there has been renewed interest in this type of latent variable model, especially as a tool for cluster analysis. The latent profile model can be seen as a probabilistic or model-based variant of traditional non-hierarchical cluster analysis procedures such as the K-means method. It has been shown that such a model-based clustering procedure outperforms the more ad hoc traditional methods. It should be noted that only a few authors use the term latent profile model. More common names are mixture of normal components, mixture model clustering, model-based clustering, latent discriminant analysis, and latent class clustering.

Possible social sciences applications of the latent profile model include building typologies and constructing diagnostic instruments. A sociologist might use it to build a typology of countries based on a set of socio-economic and political indicators. A psychologist may apply the method to combine various test scores into a single diagnostic instrument.

As in LATENT CLASS ANALYSIS, latent profile analysis assumes that the population consists of $C$ unobserved subgroups that can be referred to as latent profiles, latent classes, or mixture components. Because the indicators are continuous variables, it is most natural to assume that their conditional distribution is normal. The most general model is obtained with unrestricted multivariate normal distributions; that is,

$$f(\mathbf{y}) = \sum_{x=1}^{C} P(x)\, f(\mathbf{y}|\mu_x, \mathbf{\Sigma}_x).$$

The equation states that the joint density of the $L$ indicators, $f(\mathbf{y})$, is a mixture of class-specific densities. Each latent class $x$ has its own mean vector $\mu_x$ and covariance matrix $\mathbf{\Sigma}_x$. The proportion of persons in each of the components is denoted by $P(x)$. It should be noted that the model structures resembles quadratic discriminant analysis, with the important difference, of course, that the classes (groups) are unknown.

Several special cases are obtained by restricting the covariance matrix $\mathbf{\Sigma}_x$. Common restrictions are equal covariance matrices across classes, diagonal covariance matrices, and both equal and diagonal covariance matrices. The assumption of equal covariance matrices is similar to the basic assumption of linear discriminant analysis. Specifying the covariance matrices to be diagonal amounts to assuming LOCAL INDEPENDENCE. Such a model can also be formulated as follows:

$$f(\mathbf{y}) = \sum_{x=1}^{C} P(x) \prod_{\ell=1}^{L} f(y_\ell | \mu_{\ell x}, \sigma_{\ell x}^2).$$

Assuming local independence and equal error variances, $\sigma_{\ell x}^2 = \sigma_\ell^2$, yields a specification similar to K-means clustering. In most applications, however, this local independence specification is much too restrictive.

Recently, various methods have been proposed for structuring the class-specific covariance matrices. A relative simple one, which is a compromise between a full and a diagonal covariance matrix, is the use of block diagonal matrices. This amounts to relaxing the local independence assumption for subsets of indicators. More sophisticated methods make use principal component and factor analysis decompositions of the covariance matrix, as well as structural equation model type restrictions on the $\mathbf{\Sigma}_x$.

Another recent development are models for combinations of continuous and categorical indicators. This is actually a combination of the classical LATENT CLASS MODEL with the latent profile model. Another recent extension is the inclusion of covariates to predict class membership.

Software packages that can be used for estimating latent profile models are EMMIX, Mclust, Mplus, and Latent GOLD.

# References

Lazarsfeld, P.F., and Henry, N.W. (1968). *Latent Structure Analysis*. Boston: Houghton Mill.

McLachlan, G.J., and Basford, K.E. (1988). *Mixture Models: Inference and Application to Clustering*. New York: Marcel Dekker.

Vermunt, J.K., and Magidson, J. (2002). Latent class cluster analysis. In J.A. Hagenaars and A.L. McCutcheon (eds.), *Applied Latent Class Analysis*, 89-106. Cambridge University Press.