

Latent Class Models:

The Latent Class Logit Model

Accounting for unobserved heterogeneity:

➤ Random Parameters

- Parametric assumption: must specify the functional form of the mixing distribution (for example, normal, log-normal, etc.).

➤ Latent Class (Finite Mixture)

- Semi-parametric: requires a parametric base model (logit), but seeks unobserved groups in the data that have the same betas.

Latent Class (Finite Mixture)

- Latent class approach is less restrictive in that unobserved classes are identified without distributional assumptions.
- Drawback is that the number of classes can be quite small so there is a very coarse approximation of the distribution of heterogeneity.

Latent Class (continued)

- To resolve this, a have combined latent class and random parameter model can be estimated.
- Procedure: identify latent classes and then allowing the parameters to be random in each class (see Xiong and Mannering, 2013, for an application of this approach in the accident literature).

Latent Class Multinomial Logit

➤ Define a function that determines the probability of a discrete outcomes as,

$$S_{in} = \beta_i X_{in} + \varepsilon_{in} ,$$

where:

- S_{in} is a function that is used to determine the probability of discrete outcome i in for observation n ,
- β_i is a vector of estimable parameters for outcome i ,
- X_{in} is a vector of the observable characteristics that affect the outcome for observation n , and
- ε_{in} is a disturbance terms that is assumed to be extreme-value distributed (McFadden, 1981) which give rise to the multinomial logit form.

Introducing latent classes:

- The idea is that data can be divided into C distinct classes and that each of these classes will have their own parameter values.
- The resulting outcome probabilities are (see Greene and Hensher, 2003)

$$P_n(i/c) = \frac{\text{EXP}(\boldsymbol{\beta}_{ic} \mathbf{X}_{in})}{\sum_{\forall I} \text{EXP}(\boldsymbol{\beta}_{ic} \mathbf{X}_{in})}$$

Where:

$P_n(i/c)$ is the probability of discrete outcome i , for observation n , which is a member of unobserved class c .

➤ The unconditional class probabilities $P_n(c)$ are also determined by the multinomial logit form as:

$$P_n(c) = \frac{\text{EXP}(\mathbf{a}_c \mathbf{Z}_n)}{\sum_{\forall C} \text{EXP}(\mathbf{a}_c \mathbf{Z}_n)}$$

where:

\mathbf{Z}_n is a vector of characteristics that determine class c probabilities for observation n and \mathbf{a}_c is a corresponding vector of estimable parameters (class probabilities can be determined by a variety of characteristics for observation n).

So, the unconditional probability of observation n having discrete outcome i is simply,

$$P_n(i) = \sum_{\forall C} P_n(c) \times P_n(i/c)$$

- Latent class models can be readily estimated with maximum likelihood procedures (see Greene and Hensher, 2003).
- marginal effects, which capture the effect that a one-unit change in x has on the unconditional injury-category $P_n(i)$ also can be readily computed.