

Class counts for the most likely latent class variable in the presence of sampling weights

May 7, 2020

In Mixture analysis the most likely latent class variable is computed for each observation as the latent class for which the posterior probability (i.e. the probability that the observation belongs to a given class conditional on the observed data and the estimated model) is the largest. The most likely latent class variable can be obtained using the command

```
SAVEDATA: FILE=1.DAT; SAVE=CPROB;
```

The total number of observations with most likely latent class variable in a particular class, i.e., the distribution of the most likely latent class variable can be found in the output under the heading

CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP

These number are directly computed from the savedata file. When sampling weights are used in the Mixture analysis, however, the total number of observations with most likely latent class variable in a particular class (i.e. the distribution of the most likely latent class variable) is computed not as a simple count of the observations but it is weighted by the sampling weights. For example, the total number of observations with the most likely latent class variable equal to class 1 is the sum of the standardized weights (standardized here means that the weights are scaled to add up to the total sample size) for all observations where the most likely latent class variable equals 1, rounded to the nearest integer. The distribution, i.e., the proportions, are

also computed from these weighted values.

Thus, when sampling weights are used in the analysis, the total values in the above table will not match the simple count for the most likely latent class variable from the savedata file. They will match the weighted count.