Latent Class Model

Latent class analysis is a useful tool that is used to identify groups within multivariate categorical data. These groups are known as latent classes. As a simple comparison this can be compared to the k-means multivariate cluster analysis. There are several key difference between the two methods. First, latent classes analysis assigns observations to groups based on probability while k-means cluster analysis absolutely assigns observations to groups. While k-means is readily available in many software packages it is only appropriate for continuous data. Latent class analysis is not as widely available in many software packages but is designed to handle categorical data.

There are a handful of latent class analysis software packages. Probably the best and most common is Latent Gold. However, the license can be somewhat cost prohibitive. This is particularly true if one's daily routine does not include latent classes. Currently, SPSS does not include latent class analysis. IBM, the company that owns SPSS, has indicated that the enhancement request for latent class analysis has been filed with SPSS Development. On the open source side of things there are the R packages poLCA and MCLUST. Unless one needs the many features available in Latent Gold these R packages will generally be sufficient for data analysis.

This simple example below is a generalized form for latent class analysis. This example uses only one covariate however multiple covariates and their interactions can be included. Be sure to install the poLCA package available from CRAN.

```
library("poLCA");
f <- cbind(VAR1,VAR2,VAR3)~COVAR1;

lca.fit <- poLCA(f,raw,nclass=3,nrep=5);
covmat <- cbind(1,c(1:10));
exb <- exp(covmat %*% lca.fit$coeff);
matplot(c(35:98),(cbind(1,exb)/(1+rowSums(exb))),ylim=c(0,1),type="l",
        main="COVAR1 as a predictor of tech and social media class",
        xlab="COVAR1",
        ylab="Probability of class membership",lwd=1,col=c(1,2,3,4))
```