Finite mixture models (FMMs)

Do you have two or more groups in your data? Is group membership unobserved? Do your regression models differ across those groups? Or do the groups have different means? Or do they follow different distributions?

- Mixture models of two, three, or more unobserved groups (classes)
- Mixtures of a single estimator
- Mixtures combining different estimators or distributions
- `fmm`: prefix that can be used with 17 estimators
- Outcomes and mixture distributions can be continuous, binary, ordinal, count, categorical, fractional, censored, truncated, or even survival
- **Predict**
  - Group membership probabilities
  - Expected outcomes from the whole model
  - Expected outcomes assuming membership in a specific group
- **Select the number of groups using AIC or BIC**

**Examples**

- **Mixture of two normal distributions**
  
  `fmm 2: regress y`

- **Mixture of three normal distributions**
  
  `fmm 3: regress y`

- **Mixture of three linear regression models**
  
  `fmm 3: regress y x1 x2`

- **Mixture of three Poisson regression models**
  
  `fmm 3: poisson y x1 x2`

- **Mixture of two logistic regression models**
  
  `fmm 2: logistic y x1 x2`

- **Add predictors of group membership**
  
  `fmm 2, lcprob(x1 x3): logistic y x1 x2`
Fit your model
Say we are interested in modeling yearly number of doctor visits based on age, education level, whether the individual has private insurance or Medicaid, and whether the individual has a chronic illness. We believe the model may vary across two unobserved groups, so we type

```
fmm 2: poisson drvisits i.private i.medicaid age educ chronic
```

Estimate group means
```
estat lcmean
```
reports the estimated mean—in this case the estimated number of doctor visits—in each of the unobserved groups, that is, latent classes.

Expected doctor visits are much higher for the first group.

Estimate group proportions
```
estat lcprob
```
estimates the proportion of the population in each group.

About 28% are in the first group—those expected to visit the doctor more frequently.

Advanced inferences
- Make inferences about
  - Expected proportion of population in each group
  - Expected outcomes (means, probabilities, or counts) assuming membership in a group
  - Expected outcomes using model of all groups and membership probabilities
  - How covariates affect outcomes assuming membership in a group
  - How covariates affect outcomes considering all groups and membership probabilities
- Make inferences for groups or individuals:
  - Full population
  - Subpopulations
  - Expected outcomes for specific covariate values

Predictions
We can predict the number of doctor visits,
```
predict drvisit, mu
```
and compare the resulting distributions using `twoway histogram`.

And this is just the beginning

Want different covariates in different groups?
```
fmm: (regress y x1 x2)
(regress y x1 x3)
```

Want different models in different groups?
```
fmm: (poisson y x1 x2) (nbreg y x1 x2)
```

Want to fit a zero-inflated model?
```
fmm: (poisson y x1 x2) (pointmass y)