meprobit postestimation — Postestimation tools for meprobit

Description

The following postestimation commands are of special interest after `meprobit`:

Command	Description
`estat group`	summarize the composition of the nested groups
`estat icc`	estimate intraclass correlations

The following standard postestimation commands are also available:

Command	Description
`contrast`	contrasts and ANOVA-style joint tests of estimates
`estat ic`	Akaike's and Schwarz's Bayesian information criteria (AIC and BIC)
`estat summarize`	summary statistics for the estimation sample
`estat vce`	variance–covariance matrix of the estimators (VCE)
`estimates`	cataloging estimation results
`lincom`	point estimates, standard errors, testing, and inference for linear combinations of coefficients
`lrtest`	likelihood-ratio test
`margins`	marginal means, predictive margins, marginal effects, and average marginal effects
`marginsplot`	graph the results from margins (profile plots, interaction plots, etc.)
`nlcom`	point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients
`predict`	predictions, residuals, influence statistics, and other diagnostic measures
`predictnl`	point estimates, standard errors, testing, and inference for generalized predictions
`pwcompare`	pairwise comparisons of estimates
`test`	Wald tests of simple and composite linear hypotheses
`testnl`	Wald tests of nonlinear hypotheses

Special-interest postestimation commands

`estat group` reports the number of groups and minimum, average, and maximum group sizes for each level of the model. Model levels are identified by the corresponding group variable in the data. Because groups are treated as nested, the information in this summary may differ from what you would get if you used the `tabulate` command on each group variable individually.

`estat icc` displays the intraclass correlation for pairs of latent linear responses at each nested level of the model. Intraclass correlations are available for random-intercept models or for random-coefficient models conditional on random-effects covariates being equal to 0. They are not available for crossed-effects models.

Syntax for predict

Syntax for obtaining predictions of random effects and their standard errors

`predict` [*type*] *newvarsspec* [*if*] [*in*] , { `remeans` | `remodes` } [`reses(`*newvarsspec*`)`]

Syntax for obtaining other predictions

`predict` [*type*] *newvarsspec* [*if*] [*in*] [, *statistic* *options*]

newvarsspec is *stub*`*` or *newvarlist*.

statistic	Description
Main	
`mu`	predicted mean; the default
`fitted`	fitted linear predictor
`xb`	linear predictor for the fixed portion of the model only
`stdp`	standard error of the fixed-portion linear prediction
`pearson`	Pearson residuals
`deviance`	deviance residuals
`anscombe`	Anscombe residuals

These statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample.

options	Description
Main	
means	compute *statistic* using empirical Bayes means; the default
modes	compute *statistic* using empirical Bayes modes
nooffset	ignore the offset variable in calculating predictions; relevant only if you specified offset() when you fit the model
fixedonly	prediction for the fixed portion of the model only
Integration	
intpoints(#)	use # quadrature points to compute empirical Bayes means
iterate(#)	set maximum number of iterations in computing statistics involving empirical Bayes estimators
tolerance(#)	set convergence tolerance for computing statistics involving empirical Bayes estimators

Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

Options for predict

Main

remeans, remodes, reses(); see [ME] **meglm postestimation**.

mu, the default, calculates the predicted mean (the probability of a positive outcome), that is, the inverse link function applied to the linear prediction. By default, this is based on a linear predictor that includes both the fixed effects and the random effects, and the predicted mean is conditional on the values of the random effects. Use the fixedonly option if you want predictions that include only the fixed portion of the model, that is, if you want random effects set to 0.

fitted, xb, stdp, pearson, deviance, anscombe, means, modes, nooffset, fixedonly; see [ME] **meglm postestimation**.

By default or if the means option is specified, statistics mu, fitted, xb, stdp, pearson, deviance, and anscombe are based on the posterior mean estimates of random effects. If the modes option is specified, these statistics are based on the posterior mode estimates of random effects.

Integration

intpoints(), iterate(), tolerance(); see [ME] **meglm postestimation**.

Syntax for estat

Summarize the composition of the nested groups

```
estat group
```

Estimate intraclass correlations

```
estat icc [, level(#)]
```

Menu for estat

Statistics > Postestimation > Reports and statistics

Option for estat icc

`level(#)` specifies the confidence level, as a percentage, for confidence intervals. The default is `level(95)` or as set by `set level`; see [U] **20.7 Specifying the width of confidence intervals**.

Remarks and examples

stata.com

Various predictions, statistics, and diagnostic measures are available after fitting a mixed-effects probit model using `meprobit`. Here we show a short example of predicted probabilities and predicted random effects; refer to [ME] **meglm postestimation** for additional examples.

▷ Example 1

In example 2 of [ME] **meprobit**, we analyzed the cognitive ability (`dtlm`) of patients with schizophrenia compared with their relatives and control subjects, by using a three-level probit model with random effects at the family and subject levels. Cognitive ability was measured as the successful completion of the "Tower of London", a computerized task, measured at three levels of difficulty.

```
. use http://www.stata-press.com/data/r13/towerlondon
(Tower of London data)
. meprobit dtlm difficulty i.group || family: || subject:
  (output omitted)
```

We obtain predicted probabilities based on the contribution of both fixed effects and random effects by typing

```
. predict pr
(predictions based on fixed effects and posterior means of random effects)
(option mu assumed)
(using 7 quadrature points)
```

As the note says, the predicted values are based on the posterior means of random effects. You can use the `modes` option to obtain predictions based on the posterior modes of random effects.

We obtain predictions of the posterior means themselves by typing

```
. predict re*, remeans
(calculating posterior means of random effects)
(using 7 quadrature points)
```

Because we have one random effect at the family level and another random effect at the subject level, Stata saved the predicted posterior means in the variables `re1` and `re2`, respectively. If you are not sure which prediction corresponds to which level, you can use the `describe` command to show the variable labels.

Here we list the data for family 16:

```
. list family subject dtlm pr re1 re2 if family==16, sepby(subject)
```

	family	subject	dtlm	pr	re1	re2
208.	16	5	1	.5301687	.5051272	.1001124
209.	16	5	0	.1956408	.5051272	.1001124
210.	16	5	0	.0367041	.5051272	.1001124
211.	16	34	1	.8876646	.5051272	.7798247
212.	16	34	1	.6107262	.5051272	.7798247
213.	16	34	1	.2572725	.5051272	.7798247
214.	16	35	0	.6561904	.5051272	-.0322885
215.	16	35	1	.2977437	.5051272	-.0322885
216.	16	35	0	.071612	.5051272	-.0322885

The predicted random effects at the family level (`re1`) are the same for all members of the family. Similarly, the predicted random effects at the individual level (`re2`) are constant within each individual. The predicted probabilities (`pr`) for this family seem to be in fair agreement with the response (`dtlm`) based on a cutoff of 0.5.

We can use `estat icc` to estimate the residual intraclass correlation (conditional on the difficulty level and the individual's category) between the latent responses of subjects within the same family or between the latent responses of the same subject and family:

```
. estat icc
Residual intraclass correlation
```

Level	ICC	Std. Err.	[95% Conf.	Interval]
family	.1352637	.1050492	.0261998	.4762821
subject|family	.3622485	.0877459	.2124808	.5445812

`estat icc` reports two intraclass correlations for this three-level nested model. The first is the level-3 intraclass correlation at the family level, the correlation between latent measurements of the cognitive ability in the same family. The second is the level-2 intraclass correlation at the subject-within-family level, the correlation between the latent measurements of cognitive ability in the same subject and family.

There is not a strong correlation between individual realizations of the latent response, even within the same subject.

◁

Stored results

estat icc stores the following in r():

Scalars

r(icc#)	level-# intraclass correlation
r(se#)	standard errors of level-# intraclass correlation
r(level)	confidence level of confidence intervals

Macros

r(label#)	label for level #

Matrices

r(ci#)	vector of confidence intervals (lower and upper) for level-# intraclass correlation

For a G-level nested model, # can be any integer between 2 and G.

Methods and formulas

Methods and formulas are presented under the following headings:

Prediction
Intraclass correlations

Prediction

Methods and formulas for predicting random effects and other statistics are given in *Methods and formulas* of [ME] **meglm postestimation**.

Intraclass correlations

Consider a simple, two-level random-intercept model, stated in terms of a latent linear response, where only $y_{ij} = I(y^*_{ij} > 0)$ is observed for the latent variable,

$$y^*_{ij} = \beta + u^{(2)}_j + \epsilon^{(1)}_{ij}$$

with $i = 1, \dots, n_j$ and level-2 groups $j = 1, \dots, M$. Here β is an unknown fixed intercept, $u^{(2)}_j$ is a level-2 random intercept, and $\epsilon^{(1)}_{ij}$ is a level-1 error term. Errors are assumed to be distributed as standard normal with mean 0 and variance 1; random intercepts are assumed to be normally distributed with mean 0 and variance σ^2_2 and to be independent of error terms.

The intraclass correlation for this model is

$$\rho = \mathrm{Corr}(y^*_{ij}, y^*_{i'j}) = \frac{\sigma^2_2}{1 + \sigma^2_2}$$

It corresponds to the correlation between the latent responses i and i' from the same group j.

Now consider a three-level nested random-intercept model,

$$y^*_{ijk} = \beta + u^{(2)}_{jk} + u^{(3)}_k + \epsilon^{(1)}_{ijk}$$

for measurements $i = 1, \dots, n_{jk}$ and level-2 groups $j = 1, \dots, M_{1k}$ nested within level-3 groups $k = 1, \dots, M_2$. Here $u_{jk}^{(2)}$ is a level-2 random intercept, $u_k^{(3)}$ is a level-3 random intercept, and $\epsilon_{ijk}^{(1)}$ is a level-1 error term. The error terms have a standard normal distribution with mean 0 and variance 1. The random intercepts are assumed to be normally distributed with mean 0 and variances σ_2^2 and σ_3^2, respectively, and to be mutually independent. The error terms are also independent of the random intercepts.

We can consider two types of intraclass correlations for this model. We will refer to them as level-2 and level-3 intraclass correlations. The level-3 intraclass correlation is

$$\rho^{(3)} = \mathrm{Corr}(y_{ijk}^*, y_{i'j'k}^*) = \frac{\sigma_3^2}{1 + \sigma_2^2 + \sigma_3^2}$$

This is the correlation between latent responses i and i' from the same level-3 group k and from different level-2 groups j and j'.

The level-2 intraclass correlation is

$$\rho^{(2)} = \mathrm{Corr}(y_{ijk}^*, y_{i'jk}^*) = \frac{\sigma_2^2 + \sigma_3^2}{1 + \sigma_2^2 + \sigma_3^2}$$

This is the correlation between latent responses i and i' from the same level-3 group k and level-2 group j. (Note that level-1 intraclass correlation is undefined.)

More generally, for a G-level nested random-intercept model, the g-level intraclass correlation is defined as

$$\rho^{(g)} = \frac{\sum_{l=g}^{G} \sigma_l^2}{1 + \sum_{l=2}^{G} \sigma_l^2}$$

The above formulas also apply in the presence of fixed-effects covariates $\mathbf{X}$ in a random-effects model. In this case, intraclass correlations are conditional on fixed-effects covariates and are referred to as residual intraclass correlations. `estat icc` also uses the same formulas to compute intraclass correlations for random-coefficient models, assuming 0 baseline values for the random-effects covariates, and labels them as conditional intraclass correlations.

Intraclass correlations will always fall in [0,1] because variance components are nonnegative. To accommodate the range of an intraclass correlation, we use the logit transformation to obtain confidence intervals. We use the delta method to estimate the standard errors of the intraclass correlations.

Let $\widehat{\rho}^{(g)}$ be a point estimate of the intraclass correlation and $\widehat{\mathrm{SE}}(\widehat{\rho}^{(g)})$ be its standard error. The $(1-\alpha)\times 100\%$ confidence interval for $\mathrm{logit}(\rho^{(g)})$ is

$$\mathrm{logit}(\widehat{\rho}^{(g)}) \pm z_{\alpha/2} \frac{\widehat{\mathrm{SE}}(\widehat{\rho}^{(g)})}{\widehat{\rho}^{(g)}(1-\widehat{\rho}^{(g)})}$$

where $z_{\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution and $\mathrm{logit}(x) = \ln\{x/(1-x)\}$. Let k_u be the upper endpoint of this interval, and let k_l be the lower. The $(1-\alpha)\times 100\%$ confidence interval for $\rho^{(g)}$ is then given by

$$\left(\frac{1}{1+e^{-k_l}}, \frac{1}{1+e^{-k_u}} \right)$$

Also see

[ME] **meprobit** — Multilevel mixed-effects probit regression

[ME] **meglm postestimation** — Postestimation tools for meglm

[U] 20 Estimation and postestimation commands