Title stata.com

meglm postestimation — Postestimation tools for meglm

Postestimation commands

The following postestimation command is of special interest after `meglm`:

Command	Description
`estat group`	summarize the composition of the nested groups

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)
`estat` (svy)	postestimation statistics for survey data
`estimates`	cataloging estimation results
* `hausman`	Hausman's specification test
`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

* `hausman` and `lrtest` are not appropriate with `svy` estimation results.

predict

Description for predict

`predict` creates a new variable containing predictions such as mean responses; linear predictions; density and distribution functions; standard errors; and raw, Pearson, deviance, and Anscombe residuals.

Menu for predict

Statistics > Postestimation

Syntax for predict

Syntax for obtaining predictions of the outcome and other statistics

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

Syntax for obtaining estimated random effects and their standard errors

`predict` [*type*] *newvarsspec* [*if*] [*in*], `reffects` [*re_options*]

Syntax for obtaining ML scores

`predict` [*type*] *newvarsspec* [*if*] [*in*], `scores`

newvarsspec is *stub** or *newvarlist*.

statistic	Description
Main	
`mu`	mean response; the default
`pr`	synonym for `mu` for ordinal and binary response models
`eta`	fitted linear predictor
`xb`	linear predictor for the fixed portion of the model only
`stdp`	standard error of the fixed-portion linear prediction
`density`	predicted density function
`distribution`	predicted distribution function
`residuals`	raw residuals; available only with the Gaussian family
`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	
conditional(*ctype*)	compute *statistic* conditional on estimated random effects; default is conditional(ebmeans)
marginal	compute *statistic* marginally with respect to the random effects
nooffset	make calculation ignoring offset or exposure
outcome(*outcome*)	outcome category for predicted probabilities for ordinal models
Integration	
int_options	integration options

pearson, deviance, anscombe may not be combined with marginal.

For ordinal outcomes, you specify one or k new variables in *newvarlist* with mu and pr, where k is the number of outcomes. If you do not specify outcome(), these options assume outcome(#1).

ctype	Description
ebmeans	empirical Bayes means of random effects; the default
ebmodes	empirical Bayes modes of random effects
fixedonly	prediction for the fixed portion of the model only

re_options	Description
Main	
ebmeans	use empirical Bayes means of random effects; the default
ebmodes	use empirical Bayes modes of random effects
reses(*stub** | *newvarlist*)	calculate standard errors of empirical Bayes estimates
Integration	
int_options	integration options

int_options	Description
intpoints(#)	use # quadrature points to compute marginal predictions and 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

Options for predict

Main

mu, the default, calculates the expected value of the outcome.

pr calculates predicted probabilities and is a synonym for mu. This option is available only for ordinal and binary response models.

eta calculates the fitted linear prediction.

`xb` calculates the linear prediction $\mathbf{x}\boldsymbol{\beta}$ using the estimated fixed effects (coefficients) in the model. This is equivalent to fixing all random effects in the model to their theoretical (prior) mean value of 0.

`stdp` calculates the standard error of the fixed-effects linear predictor $\mathbf{x}\boldsymbol{\beta}$.

`density` calculates the density function. This prediction is computed using the current values of the observed variables, including the dependent variable.

`distribution` calculates the distribution function. This prediction is computed using the current values of the observed variables, including the dependent variable.

`residuals` calculates raw residuals, that is, responses minus the fitted values. This option is available only for the Gaussian family.

`pearson` calculates Pearson residuals. Pearson residuals that are large in absolute value may indicate a lack of fit.

`deviance` calculates deviance residuals. Deviance residuals are recommended by McCullagh and Nelder (1989) as having the best properties for examining the goodness of fit of a GLM. They are approximately normally distributed if the model is correctly specified. They can be plotted against the fitted values or against a covariate to inspect the model fit.

`anscombe` calculates Anscombe residuals, which are designed to closely follow a normal distribution.

`conditional(`*ctype*`)` and `marginal` specify how random effects are handled in computing *statistic*.

`conditional()` specifies that *statistic* will be computed conditional on specified or estimated random effects.

`conditional(ebmeans)`, the default, specifies that empirical Bayes means be used as the estimates of the random effects. These estimates are also known as posterior mean estimates of the random effects.

`conditional(ebmodes)` specifies that empirical Bayes modes be used as the estimates of the random effects. These estimates are also known as posterior mode estimates of the random effects.

`conditional(fixedonly)` specifies that all random effects be set to zero, equivalent to using only the fixed portion of the model.

`marginal` specifies that the predicted *statistic* be computed marginally with respect to the random effects, which means that *statistic* is calculated by integrating the prediction function with respect to all the random effects over their entire support.

Although this is not the default, marginal predictions are often very useful in applied analysis. They produce what are commonly called population-averaged estimates. They are also required by `margins`.

`nooffset` is relevant only if you specified `offset(`*varname*$_o$`)` or `exposure(`*varname*$_e$`)` with `meglm`. It modifies the calculations made by `predict` so that they ignore the offset or the exposure variable; the linear prediction is treated as $\mathbf{X}\beta + \mathbf{Zu}$ rather than $\mathbf{X}\beta + \mathbf{Zu} + \text{offset}$, or $\mathbf{X}\beta + \mathbf{Zu} + \ln(\text{exposure})$, whichever is relevant.

`outcome(`*outcome*`)` specifies the outcome for which the predicted probabilities are to be calculated. `outcome()` should contain either one value of the dependent variable or one of `#1`, `#2`, ..., with `#1` meaning the first category of the dependent variable, `#2` meaning the second category, etc.

`reffects` calculates estimates of the random effects using empirical Bayes predictions. By default, or if the `ebmeans` option is specified, empirical Bayes means are computed. With the `ebmodes` option, empirical Bayes modes are computed. You must specify q new variables, where q is the number of random-effects terms in the model. However, it is much easier to just specify *stub*`*` and let Stata name the variables *stub*`1`, *stub*`2`, ..., *stubq* for you.

`ebmeans` specifies that empirical Bayes means be used to predict the random effects.

`ebmodes` specifies that empirical Bayes modes be used to predict the random effects.

`reses(`*stub*`*` | *newvarlist*`)` calculates standard errors of the empirical Bayes estimators and stores the result in *newvarlist*. This option requires the `reffects` option. You must specify q new variables, where q is the number of random-effects terms in the model. However, it is much easier to just specify *stub*`*` and let Stata name the variables *stub*`1`, *stub*`2`, ..., *stubq* for you.

The `reffects` and `reses()` options often generate multiple new variables at once. When this occurs, the random effects (and standard errors) contained in the generated variables correspond to the order in which the variance components are listed in the output of `meglm`. The generated variables are also labeled to identify their associated random effect.

`scores` calculates the scores for each coefficient in `e(b)`. This option requires a new variable list of length equal to the number of columns in `e(b)`. Otherwise, use the *stub*`*` option to have `predict` generate enumerated variables with prefix *stub*.

Integration

`intpoints(`*#*`)` specifies the number of quadrature points used to compute marginal predictions and the empirical Bayes means; the default is the value from estimation.

`iterate(`*#*`)` specifies the maximum number of iterations when computing statistics involving empirical Bayes estimators; the default is the value from estimation.

`tolerance(`*#*`)` specifies convergence tolerance when computing statistics involving empirical Bayes estimators; the default is the value from estimation.

margins

Description for margins

`margins` estimates margins of response for mean responses and linear predictions.

Menu for margins

Statistics > Postestimation

Syntax for margins

margins [*marginlist*] [, *options*]

margins [*marginlist*] , predict(*statistic* ...) [predict(*statistic* ...) ...] [*options*]

statistic	Description
`mu`	mean response; the default
`pr`	synonym for `mu` for ordinal and binary response models
`eta`	fitted linear predictor
`xb`	linear predictor for the fixed portion of the model only
`stdp`	not allowed with `margins`
`density`	not allowed with `margins`
`distribution`	not allowed with `margins`
`residuals`	not allowed with `margins`
`pearson`	not allowed with `margins`
`deviance`	not allowed with `margins`
`anscombe`	not allowed with `margins`
`reffects`	not allowed with `margins`
`scores`	not allowed with `margins`

Options `conditional(ebmeans)` and `conditional(ebmodes)` are not allowed with `margins`.
Option `marginal` is assumed where applicable if `conditional(fixedonly)` is not specified.

Statistics not allowed with `margins` are functions of stochastic quantities other than `e(b)`.

For the full syntax, see [R] **margins**.

estat

Description for estat

`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.

Menu for estat

Statistics > Postestimation

Syntax for estat

```
estat group
```

Remarks and examples

stata.com

Various predictions, statistics, and diagnostic measures are available after fitting a mixed-effects model using `meglm`. For the most part, calculation centers around obtaining predictions of the random effects. Random effects are not estimated when the model is fit but instead need to be predicted after estimation.

▷ Example 1

In example 2 of [ME] **meglm**, we modeled the probability of contraceptive use among Bangladeshi women by fitting a mixed-effects logistic regression model. To facilitate a more direct comparison between urban and rural women, we express rural status in terms of urban status and eliminate the constant from both the fixed-effects part and the random-effects part.

```
. use http://www.stata-press.com/data/r14/bangladesh
(Bangladesh Fertility Survey, 1989)
. generate byte rural = 1 - urban
. meglm c_use rural urban age child*, nocons || district: rural urban, nocons
> family(bernoulli) link(logit)
Fitting fixed-effects model:
Iteration 0:   log likelihood = -1229.5485
Iteration 1:   log likelihood = -1228.5268
Iteration 2:   log likelihood = -1228.5263
Iteration 3:   log likelihood = -1228.5263
Refining starting values:
Grid node 0:   log likelihood = -1208.3922
Fitting full model:
Iteration 0:   log likelihood = -1208.3922  (not concave)
Iteration 1:   log likelihood = -1203.6498  (not concave)
Iteration 2:   log likelihood = -1200.6662
Iteration 3:   log likelihood = -1199.9939
Iteration 4:   log likelihood = -1199.3784
Iteration 5:   log likelihood = -1199.3272
Iteration 6:   log likelihood = -1199.3268
Iteration 7:   log likelihood = -1199.3268
Mixed-effects GLM                               Number of obs     =      1,934
Family:                 Bernoulli
Link:                       logit
Group variable:          district               Number of groups  =         60
                                                Obs per group:
                                                              min =          2
                                                              avg =       32.2
                                                              max =        118
Integration method: mvaghermite                 Integration pts.  =          7
                                                Wald chi2(6)      =     120.59
Log likelihood = -1199.3268                     Prob > chi2       =     0.0000
 ( 1)  [c_use]_cons = 0
```

c_use	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
rural	-1.712549	.1603689	-10.68	0.000	-2.026866	-1.398232
urban	-.9004495	.1674683	-5.38	0.000	-1.228681	-.5722176
age	-.0264472	.0080196	-3.30	0.001	-.0421652	-.0107291
child1	1.132291	.1603052	7.06	0.000	.8180983	1.446483
child2	1.358692	.1769369	7.68	0.000	1.011902	1.705482
child3	1.354788	.1827459	7.41	0.000	.9966122	1.712963
_cons	0	(omitted)				
district						
var(rural)	.3882825	.1284858			.2029918	.7427064
var(urban)	.239777	.1403374			.0761401	.7550947

```
LR test vs. logistic model: chi2(2) = 58.40                Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
```

We used the binary variables, `rural` and `urban`, instead of the factor notation `i.urban` because, although supported in the fixed-effects specification of the model, such notation is not supported in random-effects specifications.

This particular model allows for district random effects that are specific to the rural and urban areas of that district and that can be interpreted as such. We can obtain predictions of posterior means of the random effects and their standard errors by typing

```
. predict re_rural re_urban, reffects reses(se_rural se_urban)
(calculating posterior means of random effects)
(using 7 quadrature points)
```

The order in which we specified the variables to be generated corresponds to the order in which the variance components are listed in `meglm` output. If in doubt, a simple `describe` will show how these newly generated variables are labeled just to be sure.

Having generated estimated random effects and standard errors, we can now list them for the first 10 districts:

```
. by district, sort: generate tag = (_n==1)
. list district re_rural se_rural re_urban se_urban if district <= 10 & tag,
> sep(0)
```

	district	re_rural	se_rural	re_urban	se_urban
1.	1	-.9523374	.316291	-.5619418	.2329456
118.	2	-.0425217	.3819309	3.80e-18	.4896702
138.	3	-9.33e-17	.6231232	.2229486	.4658747
140.	4	-.2703357	.3980832	.574464	.3962131
170.	5	.0691029	.3101591	.0074569	.4650451
209.	6	-.3939819	.2759802	.2622263	.4177785
274.	7	-.1904756	.4043461	8.69e-18	.4896702
292.	8	.0382993	.3177392	.2250237	.4654329
329.	9	-.3715211	.3919996	.0628076	.453568
352.	10	-.5624707	.4763545	-5.67e-18	.4896702

The estimated standard errors are conditional on the values of the estimated model parameters: β and the components of Σ. Their interpretation is therefore not one of standard sample-to-sample variability but instead one that does not incorporate uncertainty in the estimated model parameters; see *Methods and formulas*. That stated, conditional standard errors can still be used as a measure of relative precision, provided that you keep this caveat in mind.

You can also obtain predictions of posterior modes and compare them with the posterior means:

```
. predict mod_rural mod_urban, reffects ebmodes
(calculating posterior modes of random effects)
. list district re_rural mod_rural re_urban mod_urban if district <= 10 & tag,
> sep(0)
```

	district	re_rural	mod_rural	re_urban	mod_urban
1.	1	-.9523374	-.9295366	-.5619418	-.5584528
118.	2	-.0425217	-.0306312	3.80e-18	0
138.	3	-9.33e-17	0	.2229486	.2223551
140.	4	-.2703357	-.2529507	.574464	.5644512
170.	5	.0691029	.0789803	.0074569	.0077525
209.	6	-.3939819	-.3803784	.2622263	.2595116
274.	7	-.1904756	-.1737696	8.69e-18	0
292.	8	.0382993	.0488528	.2250237	.2244676
329.	9	-.3715211	-.3540084	.0628076	.0605462
352.	10	-.5624707	-.535444	-5.67e-18	0

The two set of predictions are fairly close.

Because not all districts contain both urban and rural areas, some of the posterior modes are 0 and some of the posterior means are practically 0. A closer examination of the data reveals that district 3 has no rural areas, and districts 2, 7, and 10 have no urban areas.

Had we imposed an unstructured covariance structure in our model, the estimated posterior modes and posterior means in the cases in question would not be exactly 0 because of the correlation between urban and rural effects. For instance, if a district has no urban areas, it can still yield a nonzero (albeit small) random-effects estimate for a nonexistent urban area because of the correlation with its rural counterpart; see example 1 of [ME] **meqrlogit postestimation** for details.

◁

▷ Example 2

Continuing with the model from example 1, we can obtain predicted probabilities, and unless we specify the `fixedonly` option, these predictions will incorporate the estimated subject-specific random effects $\widetilde{\mathbf{u}}_j$.

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

The predicted probabilities for observation i in cluster j are obtained by applying the inverse link function to the linear predictor, $\widehat{p}_{ij} = g^{-1}(\mathbf{x}_{ij}\widehat{\boldsymbol{\beta}} + \mathbf{z}_{ij}\widetilde{\mathbf{u}}_j)$; see *Methods and formulas* for details. Because we specified the `means` option, the calculation uses posterior means for $\widetilde{\mathbf{u}}_j$. You can use the `modes` option to obtain predictions based on the posterior modes for $\widetilde{\widetilde{\mathbf{u}}}_j$.

```
. predict prm, pr conditional(ebmodes)
(predictions based on fixed effects and posterior modes of random effects)
```

We can list the two sets of predicted probabilities together with the actual outcome for some district, let's say district 38:

```
. list c_use pr prm if district == 38
```

	c_use	pr	prm
1228.	yes	.5783408	.5780864
1229.	no	.5326623	.5324027
1230.	yes	.6411679	.6409279
1231.	yes	.5326623	.5324027
1232.	yes	.5718783	.5716228
1233.	no	.3447686	.344533
1234.	no	.4507973	.4505391
1235.	no	.1940524	.1976133
1236.	no	.2846738	.2893007
1237.	no	.1264883	.1290078
1238.	no	.206763	.2104961
1239.	no	.202459	.2061346
1240.	no	.206763	.2104961
1241.	no	.1179788	.1203522

The two sets of predicted probabilities are fairly close.

For mixed-effects models with many levels or many random effects, the calculation of the posterior means of random effects or any quantities that are based on the posterior means of random effects may take a long time. This is because we must resort to numerical integration to obtain the posterior means. In contrast, the calculation of the posterior modes of random effects is usually orders of magnitude faster because there is no numerical integration involved. For this reason, empirical modes are often used in practice as an approximation to empirical means. Note that for linear mixed-effects models, the two predictors are the same.

We can compare the observed values with the predicted values by constructing a classification table. Defining success as $\widehat{y}_{ij} = 1$ if $\widehat{p}_{ij} > 0.5$ and defining $\widehat{y}_{ij} = 0$ otherwise, we obtain the following table.

```
. generate p_use = pr > .5
. label var p_use "Predicted outcome"
. tab2 c_use p_use, row
-> tabulation of c_use by p_use
```

Key
frequency   *row percentage*

Use contraception	Predicted outcome 0	1	Total
no	991 84.34	184 15.66	1,175 100.00
yes	423 55.73	336 44.27	759 100.00
Total	1,414 73.11	520 26.89	1,934 100.00

The model correctly classified 84% of women who did not use contraceptives but only 44% of women who did. In the next example, we will look at some residual diagnostics.

◁

❑ Technical note

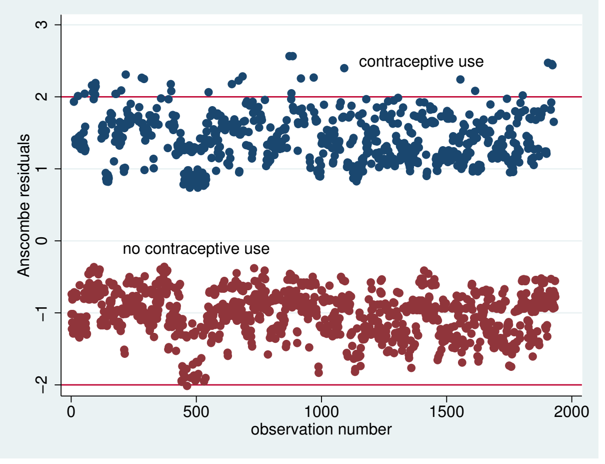
Out-of-sample predictions are permitted after `meglm`, but if these predictions involve estimated random effects, the integrity of the estimation data must be preserved. If the estimation data have changed since the model was fit, `predict` will be unable to obtain predicted random effects that are appropriate for the fitted model and will give an error. Thus to obtain out-of-sample predictions that contain random-effects terms, be sure that the data for these predictions are in observations that augment the estimation data.

❑

▷ Example 3

Continuing our discussion from example 2, here we look at residual diagnostics. `meglm` offers three kinds of predicted residuals for nonlinear responses—Pearson, Anscombe, and deviance. Of the three, Anscombe residuals are designed to be approximately normally distributed; thus we can check for outliers by plotting Anscombe residuals against observation numbers and seeing which residuals are greater than 2 in absolute value.

```
. predict anscombe, anscombe
(predictions based on fixed effects and posterior means of random effects)
(using 7 quadrature points)

. generate n = _n

. label var n "observation number"

. twoway (scatter anscombe n if c_use) (scatter anscombe n if !c_use),
> yline(-2 2) legend(off) text(2.5 1400 "contraceptive use")
> text(-.1 500 "no contraceptive use")
```



There seem to be some outliers among residuals that identify women who use contraceptives. We could examine the observations corresponding to the outliers, or we could try fitting a model with perhaps a different covariance structure, which we leave as an exercise.

◁

▷ Example 4

In example 3 of [ME] **meglm**, we estimated the effects of two treatments on the tobacco and health knowledge (THK) scale score of students in 28 schools. The dependent variable was collapsed into four ordered categories, and we fit a three-level ordinal logistic regression.

```
. use http://www.stata-press.com/data/r14/tvsfpors, clear
. meologit thk prethk i.cc##i.tv || school: || class:
Fitting fixed-effects model:
Iteration 0:   log likelihood =  -2212.775
Iteration 1:   log likelihood =  -2125.509
Iteration 2:   log likelihood = -2125.1034
Iteration 3:   log likelihood = -2125.1032
Refining starting values:
Grid node 0:   log likelihood = -2152.1514
Fitting full model:
```

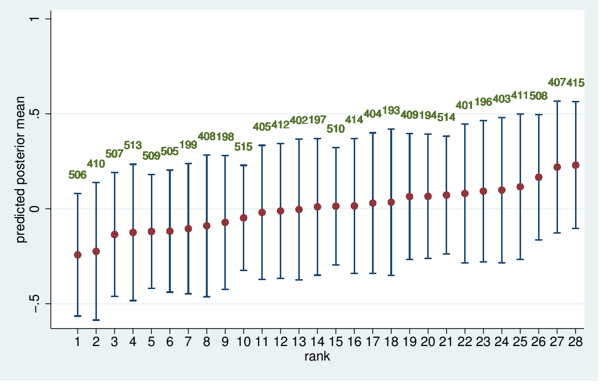
(output omitted)

```
Mixed-effects ologit regression                 Number of obs     =      1,600

-------------------------------------------------------------
                |   No. of       Observations per Group
 Group Variable |   Groups    Minimum    Average    Maximum
----------------+--------------------------------------------
         school |       28         18       57.1        137
          class |      135          1       11.9         28
-------------------------------------------------------------

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(4)      =     124.39
Log likelihood = -2114.5881                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
         thk |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      prethk |   .4085273    .039616    10.31   0.000     .3308814    .4861731
        1.cc |   .8844369   .2099124     4.21   0.000     .4730161    1.295858
        1.tv |    .236448   .2049065     1.15   0.249    -.1651614    .6380575
             |
       cc#tv |
        1 1  |  -.3717699   .2958887    -1.26   0.209     -.951701    .2081612
-------------+----------------------------------------------------------------
       /cut1 |  -.0959459   .1688988    -0.57   0.570    -.4269815    .2350896
       /cut2 |   1.177478   .1704946     6.91   0.000     .8433151    1.511642
       /cut3 |   2.383672   .1786736    13.34   0.000     2.033478    2.733865
-------------+----------------------------------------------------------------
school       |
   var(_cons)|   .0448735   .0425387                      .0069997    .2876749
-------------+----------------------------------------------------------------
school>class |
   var(_cons)|   .1482157   .0637521                       .063792    .3443674
------------------------------------------------------------------------------
LR test vs. ologit model: chi2(2) = 21.03                 Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
```

Not surprisingly, the level of knowledge before the intervention is a good predictor of the level of knowledge after the intervention. The social resistance classroom curriculum is effective in raising the knowledge score, but the TV intervention and the interaction term are not.

We can rank schools by their institutional effectiveness by plotting the random effects at the school level.

```
. predict re_school re_class, reffects reses(se_school se_class)
(calculating posterior means of random effects)
(using 7 quadrature points)

. generate lower = re_school - 1.96*se_school

. generate upper = re_school + 1.96*se_school

. egen tag = tag(school)

. gsort +re_school -tag

. generate rank = sum(tag)

. generate labpos = re_school + 1.96*se_school + .1

. twoway (rcap lower upper rank) (scatter re_school rank)
> (scatter labpos rank, mlabel(school) msymbol(none) mlabpos(0)),
> xtitle(rank) ytitle(predicted posterior mean) legend(off)
> xscale(range(0 28)) xlabel(1/28) ysize(2)
```



Although there is some variability in the predicted posterior means, we cannot see significant differences among the schools in this example.

◁

Methods and formulas

Continuing the discussion in *Methods and formulas* of [ME] **meglm** and using the definitions and formulas defined there, we begin by considering the prediction of the random effects $\mathbf{u}_j$ for the jth cluster in a two-level model. Prediction of random effects in multilevel generalized linear models involves assigning values to random effects, and there are many methods for doing so; see Skrondal and Rabe-Hesketh (2009) and Skrondal and Rabe-Hesketh (2004, chap. 7) for a comprehensive review. Stata offers two methods of predicting random effects: empirical Bayes means (also known as posterior means) and empirical Bayes modes (also known as posterior modes). Below we provide more details about the two methods.

Let $\widehat{\boldsymbol\theta}$ denote the estimated model parameters comprising $\widehat{\boldsymbol\beta}$ and the unique elements of $\widehat{\boldsymbol\Sigma}$. Empirical Bayes (EB) predictors of the random effects are the means or modes of the empirical posterior distribution with the parameter estimates $\boldsymbol\theta$ replaced with their estimates $\widehat{\boldsymbol\theta}$. The method is called "empirical" because $\widehat{\boldsymbol\theta}$ is treated as known. EB combines the prior information about the random effects with the likelihood to obtain the conditional posterior distribution of random effects. Using Bayes's theorem, the empirical conditional posterior distribution of random effects for cluster j is

$$\omega(\mathbf{u}_j|\mathbf{y}_j,\mathbf{X}_j,\mathbf{Z}_j;\widehat{\boldsymbol\theta}) = \frac{\Pr(\mathbf{y}_j,\mathbf{u}_j|\mathbf{X}_j,\mathbf{Z}_j;\widehat{\boldsymbol\theta})}{\Pr(\mathbf{y}_j|\mathbf{X}_j,\mathbf{Z}_j;\widehat{\boldsymbol\theta})}$$
$$= \frac{f(\mathbf{y}_j|\mathbf{u}_j,\mathbf{X}_j,\mathbf{Z}_j;\widehat{\boldsymbol\beta})\,\phi(\mathbf{u}_j;\widehat{\boldsymbol\Sigma})}{\int f(\mathbf{y}_j|\mathbf{u}_j)\,\phi(\mathbf{u}_j)\,d\mathbf{u}_j}$$
$$= \frac{f(\mathbf{y}_j|\mathbf{u}_j,\mathbf{X}_j,\mathbf{Z}_j;\widehat{\boldsymbol\beta})\,\phi(\mathbf{u}_j;\widehat{\boldsymbol\Sigma})}{\mathcal{L}_j(\widehat{\boldsymbol\theta})}$$

The denominator is just the likelihood contribution of the jth cluster.

EB mean predictions of random effects, $\widetilde{\mathbf{u}}$, also known as posterior means, are calculated as

$$\widetilde{\mathbf{u}} = \int_{\Re^q} \mathbf{u}_j\,\omega(\mathbf{u}_j|\mathbf{y}_j,\mathbf{X}_j,\mathbf{Z}_j;\widehat{\boldsymbol\theta})\,d\mathbf{u}_j$$
$$= \frac{\int_{\Re^q} \mathbf{u}_j\, f(\mathbf{y}_j|\mathbf{u}_j,\mathbf{X}_j,\mathbf{Z}_j;\widehat{\boldsymbol\beta})\,\phi(\mathbf{u}_j;\widehat{\boldsymbol\Sigma})\,d\mathbf{u}_j}{\int_{\Re^q} f(\mathbf{y}_j|\mathbf{u}_j)\,\phi(\mathbf{u}_j)\,d\mathbf{u}_j}$$

where we use the notation $\widetilde{\mathbf{u}}$ rather than $\widehat{\mathbf{u}}$ to distinguish predicted values from estimates. This multivariate integral is approximated by the mean–variance adaptive Gaussian quadrature; see *Methods and formulas* of [ME] **meglm** for details about the quadrature. If you have multiple random effects within a level or random effects across levels, the calculation involves orthogonalizing transformations using the Cholesky transformation because the random effects are no longer independent under the posterior distribution.

In a linear mixed-effects model, the posterior density is multivariate normal, and EB means are also best linear unbiased predictors (BLUPs); see Skrondal and Rabe-Hesketh (2004, 227). In generalized mixed-effects models, the posterior density tends to multivariate normal as cluster size increases.

EB modal predictions can be approximated by solving for the mode $\widetilde{\widetilde{\mathbf{u}}}_j$ in

$$\frac{\partial}{\partial \mathbf{u}_j} \log \omega(\widetilde{\widetilde{\mathbf{u}}}_j|\mathbf{y}_j,\mathbf{X}_j,\mathbf{Z}_j;\widehat{\boldsymbol\theta}) = \mathbf{0}$$

Because the denominator in $\omega(\cdot)$ does not depend on $\mathbf{u}$, we can omit it from the calculation to obtain

$$\frac{\partial}{\partial \mathbf{u}_j} \log\left\{ f(\mathbf{y}_j|\mathbf{u}_j,\mathbf{X}_j,\mathbf{Z}_j;\widehat{\boldsymbol\beta})\,\phi(\mathbf{u}_j;\widehat{\boldsymbol\Sigma})\right\}$$
$$= \frac{\partial}{\partial \mathbf{u}_j} \log f\left(\mathbf{y}_j|\mathbf{u}_j,\mathbf{X}_j,\mathbf{Z}_j;\widehat{\boldsymbol\beta}\right) + \frac{\partial}{\partial \mathbf{u}_j} \log\phi\left(\mathbf{u}_j;\widehat{\boldsymbol\Sigma}\right) = 0$$

The calculation of EB modes does not require numerical integration, and for that reason they are often used in place of EB means. As the posterior density gets closer to being multivariate normal, EB modes get closer and closer to EB means.

Just like there are many methods of assigning values to the random effects, there exist many methods of calculating standard errors of the predicted random effects; see Skrondal and Rabe-Hesketh (2009) for a comprehensive review.

Stata uses the posterior standard deviation as the standard error of the posterior means predictor of random effects. The EB posterior covariance matrix of the random effects is given by

$$\text{cov}(\widetilde{\mathbf{u}}_j|\mathbf{y}_j,\mathbf{X}_j,\mathbf{Z}_j;\widehat{\boldsymbol{\theta}}) = \int_{\Re^q} (\mathbf{u}_j - \widetilde{\mathbf{u}}_j)(\mathbf{u}_j - \widetilde{\mathbf{u}}_j)'\,\omega(\mathbf{u}_j|\mathbf{y}_j,\mathbf{X}_j,\mathbf{Z}_j;\widehat{\boldsymbol{\theta}})\,d\mathbf{u}_j$$

The posterior covariance matrix and the integrals are approximated by the mean–variance adaptive Gaussian quadrature; see *Methods and formulas* of [ME] **meglm** for details about the quadrature.

Conditional standard errors for the estimated posterior modes are derived from standard theory of maximum likelihood, which dictates that the asymptotic variance matrix of $\widetilde{\widetilde{\mathbf{u}}}_j$ is the negative inverse of the Hessian, $g''(\boldsymbol{\beta},\boldsymbol{\Sigma},\widetilde{\widetilde{\mathbf{u}}}_j)$.

In what follows, we show formulas using the posterior means estimates of random effects $\widetilde{\mathbf{u}}_j$, which are used by default or if the `means` option is specified. If the `modes` option is specified, $\widetilde{\mathbf{u}}_j$ are simply replaced with the posterior modes $\widetilde{\widetilde{\mathbf{u}}}_j$ in these formulas.

For any ith observation in the jth cluster in a two-level model, define the linear predictor as

$$\widehat{\eta}_{ij} = \mathbf{x}_{ij}\widehat{\boldsymbol{\beta}} + \mathbf{z}_{ij}\widetilde{\mathbf{u}}_j$$

The linear predictor includes the offset or exposure variable if one was specified during estimation, unless the `nooffset` option is specified. If the `fixedonly` option is specified, $\widehat{\eta}$ contains the linear predictor for only the fixed portion of the model, $\widehat{\eta}_{ij} = \mathbf{x}_{ij}\widehat{\boldsymbol{\beta}}$.

The predicted mean, conditional on the random effects $\widetilde{\mathbf{u}}_j$, is

$$\widehat{\mu}_{ij} = g^{-1}(\widehat{\eta}_{ij})$$

where $g^{-1}(\cdot)$ is the inverse link function for $\mu_{ij} = g^{-1}(\eta_{ij})$ defined as follows for the available links in `link(`*link*`)`:

link	Inverse link
`identity`	η_{ij}
`logit`	$1/\{1+\exp(-\eta_{ij})\}$
`probit`	$\Phi(\eta_{ij})$
`log`	$\exp(\eta_{ij})$
`cloglog`	$1-\exp\{-\exp(\eta_{ij})\}$

By default, random effects and any statistic based on them—`mu`, `fitted`, `pearson`, `deviance`, `anscombe`—are calculated using posterior means of random effects unless option `modes` is specified, in which case the calculations are based on posterior modes of random effects.

Raw residuals are calculated as the difference between the observed and fitted outcomes,

$$\nu_{ij} = y_{ij} - \widehat{\mu}_{ij}$$

and are only defined for the Gaussian family.

Let r_{ij} be the number of Bernoulli trials in a binomial model, α be the conditional overdispersion parameter under the mean parameterization of the negative binomial model, and δ be the conditional overdispersion parameter under the constant parameterization of the negative binomial model.

Pearson residuals are raw residuals divided by the square root of the variance function

$$\nu_{ij}^P = \frac{\nu_{ij}}{\{V(\widehat{\mu}_{ij})\}^{1/2}}$$

where $V(\widehat{\mu}_{ij})$ is the family-specific variance function defined as follows for the available families in `family(`*family*`)`:

family	Variance function $V(\widehat{\mu}_{ij})$
`bernoulli`	$\widehat{\mu}_{ij}(1-\widehat{\mu}_{ij})$
`binomial`	$\widehat{\mu}_{ij}(1-\widehat{\mu}_{ij}/r_{ij})$
`gamma`	$\widehat{\mu}_{ij}^2$
`gaussian`	1
`nbinomial mean`	$\widehat{\mu}_{ij}(1+\alpha\widehat{\mu}_{ij})$
`nbinomial constant`	$\widehat{\mu}_{ij}(1+\delta)$
`ordinal`	not defined
`poisson`	$\widehat{\mu}_{ij}$

Deviance residuals are calculated as

$$\nu_{ij}^{D} = \text{sign}(\nu_{ij})\sqrt{\widehat{d}_{ij}^{2}}$$

where the squared deviance residual $\widehat{d}_{ij}^{2}$ is defined as follows:

family	Squared deviance $\widehat{d}_{ij}^{2}$
`bernoulli`	$-2\log(1-\widehat{\mu}_{ij})$ if $y_{ij}=0$
	$-2\log(\widehat{\mu}_{ij})$ if $y_{ij}=1$
`binomial`	$2r_{ij}\log\left(\dfrac{r_{ij}}{r_{ij}-\widehat{\mu}_{ij}}\right)$ if $y_{ij}=0$
	$2y_{ij}\log\left(\dfrac{y_{ij}}{\widehat{\mu}_{ij}}\right)+2(r_{ij}-y_{ij})\log\left(\dfrac{r_{ij}-y_{ij}}{r_{ij}-\widehat{\mu}_{ij}}\right)$ if $0<y_{ij}<r_{ij}$
	$2r_{ij}\log\left(\dfrac{r_{ij}}{\widehat{\mu}_{ij}}\right)$ if $y_{ij}=r_{ij}$
`gamma`	$-2\left\{\log\left(\dfrac{y_{ij}}{\widehat{\mu}_{ij}}\right)-\dfrac{\widehat{\nu}_{ij}}{\widehat{\mu}_{ij}}\right\}$
`gaussian`	$\widehat{\nu}_{ij}^{2}$
`nbinomial mean`	$2\log\left(1+\alpha\widehat{\mu}_{ij}\right)\alpha$ if $y_{ij}=0$
	$2y_{ij}\log\left(\dfrac{y_{ij}}{\widehat{\mu}_{ij}}\right)-\frac{2}{\alpha}(1+\alpha y_{ij})\log\left(\dfrac{1+\alpha y_{ij}}{1+\alpha\widehat{\mu}_{ij}}\right)$ otherwise
`nbinomial constant`	not defined
`ordinal`	not defined
`poisson`	$2\widehat{\mu}_{ij}$ if $y_{ij}=0$
	$2y_{ij}\log\left(\dfrac{y_{ij}}{\widehat{\mu}_{ij}}\right)-2\widehat{\nu}_{ij}$ otherwise

Anscombe residuals, denoted ν_{ij}^A, are calculated as follows:

family	Anscombe residual ν_{ij}^A
bernoulli	$\dfrac{3\left\{y_{ij}^{2/3}\mathcal{H}(y_{ij}) - \widehat{\mu}_{ij}^{2/3}\mathcal{H}(\widehat{\mu}_{ij})\right\}}{2\left(\widehat{\mu}_{ij} - \widehat{\mu}_{ij}^2\right)^{1/6}}$
binomial	$\dfrac{3\left\{y_{ij}^{2/3}\mathcal{H}(y_{ij}/r_{ij}) - \widehat{\mu}_{ij}^{2/3}\mathcal{H}(\widehat{\mu}_{ij}/r_{ij})\right\}}{2\left(\widehat{\mu}_{ij} - \widehat{\mu}_{ij}^2/r_{ij}\right)^{1/6}}$
gamma	$\dfrac{3(y_{ij}^{1/3} - \widehat{\mu}_{ij}^{1/3})}{\widehat{\mu}_{ij}^{1/3}}$
gaussian	ν_{ij}
nbinomial mean	$\dfrac{\mathcal{H}(-\alpha y_{ij}) - \mathcal{H}(-\alpha\widehat{\mu}_{ij}) + 1.5(y_{ij}^{2/3} - \widehat{\mu}_{ij}^{2/3})}{(\widehat{\mu}_{ij} + \alpha\widehat{\mu}_{ij}^2)^{1/6}}$
nbinomial constant	not defined
ordinal	not defined
poisson	$\dfrac{3(y_{ij}^{2/3} - \widehat{\mu}_{ij}^{2/3})}{2\widehat{\mu}_{ij}^{1/6}}$

where $\mathcal{H}(t)$ is a specific univariate case of the Hypergeometric2F1 function (Wolfram 1999, 771–772), defined here as $\mathcal{H}(t) = {}_2F_1(2/3, 1/3, 5/3, t)$.

For a discussion of the general properties of the various residuals, see Hardin and Hilbe (2012, chap. 4).

References

Hardin, J. W., and J. M. Hilbe. 2012. *Generalized Linear Models and Extensions*. 3rd ed. College Station, TX: Stata Press.

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*. 2nd ed. London: Chapman & Hall/CRC.

Skrondal, A., and S. Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman & Hall/CRC.

——. 2009. Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society, Series A* 172: 659–687.

Wolfram, S. 1999. *The Mathematica Book*. 4th ed. Cambridge: Cambridge University Press.

Also see

[ME] **meglm** — Multilevel mixed-effects generalized linear model

[U] 20 Estimation and postestimation commands