Title stata.com

irt hybrid postestimation — Postestimation tools for irt hybrid

Postestimation commands

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

Command	Description
`estat report`	report estimated IRT parameters
`irtgraph icc`	plot item characteristic curve (ICC)
`irtgraph iif`	plot item information function (IIF)
`irtgraph tcc`	plot test characteristic curve (TCC)
`irtgraph tif`	plot test information function (TIF)

The following standard postestimation commands are also available:

Command	Description
`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
`lincom`	point estimates, standard errors, testing, and inference for linear combinations of coefficients
* `lrtest`	likelihood-ratio test
`nlcom`	point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients
`predict`	predictions
`predictnl`	point estimates, standard errors, testing, and inference for generalized predictions
`test`	Wald tests of simple and composite linear hypotheses
`testnl`	Wald tests of nonlinear hypotheses

* `lrtest` is not appropriate with `svy` estimation results.

predict

Description for predict

predict creates a new variable containing predictions such as probabilities, linear predictions, and parameter-level scores.

Menu for predict

Statistics > Postestimation

Syntax for predict

Syntax for obtaining predictions of item probabilities and other statistics

predict [*type*] *newvarsspec* [*if*] [*in*] [, *statistic* *item_options*]

Syntax for obtaining estimated latent variables and their standard errors

predict [*type*] *newvarsspec* [*if*] [*in*], latent [*latent_options*]

Syntax for obtaining parameter-level scores

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

newvarsspec is *stub** or *newvarlist*.

statistic	Description
Main	
pr	probabilities; the default
xb	linear prediction

item_options	Description
Main	
† outcome(*item* [#])	specify item variable; default is all variables
conditional(*ctype*)	compute *statistic* conditional on estimated latent variables; default is conditional(ebmeans)
marginal	compute *statistic* marginally with respect to the latent variables
Integration	
int_options	integration options

† outcome(*item* #) may also be specified as outcome(*#.item*) or outcome(*item* ##).
outcome(*item* #3) means the third outcome value. outcome(*item* #3) would mean the same as outcome(*item* 4) if outcomes were 1, 3, and 4.

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

latent_options	Description
Main	
ebmeans	use empirical Bayes means of latent trait; the default
ebmodes	use empirical Bayes modes of latent trait
se(*newvar*)	calculate standard errors
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

pr, the default, calculates the predicted probability.

xb specifies that the linear predictor be calculated.

outcome(*item* [#]) specifies that predictions for *item* be calculated. Use # to specify which outcome level to predict. Predictions for all observed response variables are computed by default.

conditional(*ctype*) and marginal specify how latent variables are handled in computing *statistic*.

conditional() specifies that *statistic* will be computed conditional on specified or estimated latent variables.

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

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

conditional(fixedonly) specifies that all latent variables 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 latent variables, which means that *statistic* is calculated by integrating the prediction function with respect to all the latent variables 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.

`latent` specifies that the latent trait is predicted using an empirical Bayes estimator; see options `ebmeans` and `ebmodes`.

`ebmeans` specifies that empirical Bayes means are used to predict the latent variables.

`ebmodes` specifies that empirical Bayes modes are used to predict the latent variables.

`se(`*newvar*`)` calculates standard errors of the empirical Bayes estimator and stores the result in *newvar*. This option requires the `latent` option.

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

Methods and formulas

This section builds on the notation introduced in *Methods and formulas* of each of the other IRT postestimation entries.

We begin by considering the prediction of the latent trait θ for a given person. Prediction of the latent trait in IRT models involves assigning a value to the latent trait, 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 latent traits: empirical Bayes means (also known as posterior means) and empirical Bayes modes (also known as posterior modes).

Methods and formulas are presented under the following headings:

Empirical Bayes
Other predictions

Empirical Bayes

Let $\widehat{\boldsymbol{B}}$ denote the estimated model parameters. Empirical Bayes (EB) predictors of the latent trait are the means or modes of the empirical posterior distribution with the parameter estimates $\boldsymbol{B}$ replaced with their estimates $\widehat{\boldsymbol{B}}$. The method is called "empirical" because $\widehat{\boldsymbol{B}}$ is treated as known. EB combines the prior information about the latent trait with the likelihood to obtain the conditional posterior distribution of the latent trait. Using Bayes's theorem, we see that the empirical conditional posterior distribution of the latent trait for person j is

$$\omega(\theta_j|\mathbf{y}_j;\widehat{\boldsymbol{B}}) = \frac{f(\mathbf{y}_j|\widehat{\boldsymbol{B}},\theta_j)\,\phi(\theta_j)}{\int_{-\infty}^{\infty} f(\mathbf{y}_j|\widehat{\boldsymbol{B}},\theta_j)\,\phi(\theta_j)\,d\theta_j} = \frac{f(\mathbf{y}_j|\widehat{\boldsymbol{B}},\theta_j)\,\phi(\theta_j)}{L_j(\widehat{\boldsymbol{B}})}$$

The denominator is just the likelihood contribution for person j.

EB mean predictions of the latent trait, also known as posterior means, are calculated as

$$\widetilde{\theta}_j = \int_{-\infty}^{\infty} \theta_j\,\omega(\theta_j|\mathbf{y}_j;\widehat{\boldsymbol{B}})\,d\theta_j$$

where we use the notation $\widetilde{\theta}_j$ rather than $\widehat{\theta}_j$ to distinguish predicted values from estimates. This integral is approximated by MVAGHQ.

EB modal predictions can be approximated by solving for $\widetilde{\widetilde{\theta}}_j$ such that

$$\frac{\partial}{\partial\theta_j}\log\omega(\theta_j|\mathbf{y}_j;\widehat{\boldsymbol{B}})\Big|_{\theta_j=\widetilde{\widetilde{\theta}}_j} = \mathbf{0}$$

Because the denominator in $\omega(\cdot)$ does not depend on θ_j, we can omit it from the calculation to obtain the EB mode. The calculation of EB modes does not require numerical integration; thus, they are often used in place of EB means. As the posterior density gets closer to the normal distribution, EB modes get closer and closer to EB means.

Just as there are many methods of assigning values to the latent trait, there are many methods of calculating standard errors of the predicted latent trait; 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 the latent trait. The EB posterior variance of the latent trait is given by

$$\mathrm{Var}(\widetilde{\theta}_j|\mathbf{y}_j;\widehat{\boldsymbol{B}}) = \int_{-\infty}^{\infty} (\theta_j - \widetilde{\theta}_j)^2\,\omega(\theta_j|\mathbf{y}_j;\widehat{\boldsymbol{B}})\,d\theta_j$$

The posterior variance and the integrals are approximated by MVAGHQ.

Conditional standard errors for the estimated posterior modes are derived from the standard theory of maximum likelihood, which dictates that the asymptotic variance matrix of the posterior mode is the negative inverse of the Hessian matrix.

Other predictions

The other predictions are governed by the model selected for the specified item response variable. For binary items, see *Methods and formulas* of the postestimation entries for `irt 1pl`, `irt 2pl`, and `irt 3pl`. For ordered items, see *Methods and formulas* of the postestimation entries for `irt grm`, `irt pcm`, and `irt rsm`. For nominal items, see *Methods and formulas* of the postestimation entry for `irt nrm`.

References

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.

Also see

[IRT] **irt hybrid** — Hybrid IRT models

[IRT] **estat report** — Report estimated IRT parameters

[IRT] **irtgraph icc** — Item characteristic curve plot

[IRT] **irtgraph iif** — Item information function plot

[IRT] **irtgraph tcc** — Test characteristic curve plot

[IRT] **irtgraph tif** — Test information function plot

[U] **20 Estimation and postestimation commands**